AN ABSTRACT OF THE DISSERTATION OF

Franklin A. Hays for the degree of Doctor of Philosophy in Biochemistry and Biophysics presented on April 14, 2005.

Title: Sequence Dependent Conformational Variations in DNA Holliday Junctions

Redacted for privacy

Abstract approved: _____

Pui Shing Ho

Four-stranded DNA junctions (also known as Holliday junctions) are structural intermediates involved in a growing number of biological processes including DNA repair, genetic recombination, and viral integration. Although previous studies have focused on understanding the conformational variability and sequence-dependent formation of Holliday junctions in solution there have been relatively few insights into junction structure at the atomic level. Recent crystallographic studies have demonstrated that the more compact stacked-X junction form has an antiparallel alignment of DNA strands and standard Watson-Crick base pairs across the central crossover region. Junction formation within this crystallographic system was seen to be dependent on a common trinucleotide sequence motif ("ACC-triplet" at the 6th, 7th and 8th positions of the decanucleotide sequence d(CCnnn$N_6N_7N_8$GG)) containing a series of stabilizing direct and solvent-mediated hydrogen bonding interactions. This thesis addresses questions concerning the nucleotide sequence-dependent formation

and conformational variability of DNA Holliday junctions as determined by single crystal x-ray diffraction.

We have used the modified bases 2,6-diaminopurine and inosine to demonstrate that minor groove interactions adjacent to the trinucleotide junction core are not major contributors to overall conformation. In addition, incorporation of guanine into the sixth position of this core does not have a significant effect on junction geometry. Meanwhile, incorporation of 5-bromouracil into the eighth position perturbs the geometry in terms of the interduplex angle as well as the defined conformational variables, J_{roll} and J_{slide}. These novel junction structures demonstrate that the nucleotide sequence within the central core generates a position specific relationship between molecular interactions at the junction crossover and overall structural geometry.

A systematic crystallographic screen of the trinucleotide core region is presented here as an unbiased, comprehensive, search for sequences that stabilize junctions. As the result of this screen, we can extend the core sequence motif to '$N_6Y_7C_8$' where N_6 is an adenine, guanine, or cytosine nucleotide and Y_7 is either a cytosine or thymine (if N_6 = adenine) nucleotide. Using these novel junction structures, we demonstrate that base sequence within the central core has a significant effect on the overall geometry of the junction. Thus, this central region of the structure may serve as a linchpin for determining the local and global conformation and overall variability of the four-stranded DNA Holliday junction. These observations raise some interesting questions regarding the importance of this core region in biological processes such as genetic recombination.

© Copyright by Franklin A. Hays

April 14, 2005

All Rights Reserved

Sequence Dependent Conformational Variations in DNA Holliday Junctions

by

Franklin A. Hays

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented April 14, 2005
Commencement June 2005

Doctor of Philosophy dissertation of Franklin A. Hays presented on April 14, 2005

APPROVED:

Redacted for privacy

Major Professor, Representing the Department of Biochemistry and Biophysics

Redacted for privacy

Chair 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

Franklin A. Hays, Author

ACKNOWLEDGMENTS

I thank my graduate advisor and mentor, Dr. P. Shing Ho, for his patient and dedicated support throughout my training in his laboratory. Shing, your commitment to training students within your lab and teaching them how to engage the process of scientific discovery and description is impressive. I thrived in the environment you established and I appreciate your efforts. I would also like to thank all current and former members of the Ho lab for establishing the foundation on which a lot of my work is based. I would like to thank Jeffrey Vargason for supporting me during the steep learning curve of the first two years of graduate school. You taught me that any hurdle can be overcome through focused effort. Thank you Amy Teegarden, Zebulon Jones, Michael Harms, Dustin Raup and Emily Cavaliere for your help and assistance with the crystallographic screen project ("just one more setup!"). I would like to thank Andrea Voth and Patricia Khuu for joining the lab and pursuing some very interesting scientific prospects. Both of you will do great!

To the Karplus lab, thank you for constantly providing support on all aspects of crystallographic data collection, refinement and analysis. Dr. Andrew Karplus, thank you for improving the quality of my oral presentations and teaching x-ray crystallography. Though all of my structures are still double helical I would like to thank Zachary Wood, I admire your zest and love for science Zac! In Rick Faber I found a great friend and somebody that I enjoyed many conversations with on all aspects of Oregon life and science. Thank you for taking care of the diffractometer Rick and constantly putting up with my "need it now" pushiness! I owe a significant

thank you to Ganapathy N. Sarma for always having a colorful quote. And finally, I would like to thank the remaining members of my doctoral committee for their support: Drs. Joe Beckman, Mike Schimerlik and Mary Slabaugh.

I would like to thank my undergraduate advisor - Dr. H. Olin Spivey. Thank you Dr. Spivey for always putting up with my lab antics and providing an atmosphere to pursue what I enjoy.

Thank you to my family and friends for their support through many years of schooling. I would like to especially thank my parents, Ken and Mary Hays, for providing space to explore and freedom to make my own decisions. As for the Thibodeaux's (Jackie, Morris, David, Micayla and Mariah) - you guys are always a source of enjoyment and I look forward to what the future holds. Thanks to my little sister Tonya for being around in my absence and to my brother David for turning into an amazing friend. You are doing great David (including Autumn and Aleeca)!

To my extended family, thank you for always loving and supporting me. Sharon and Allen - thank you so much for all that you have done during my "never-ending" pursuit of higher education. Geoff and Jo- thanks for providing the jump drive on which this thesis took shape! Thank you Tina for pushing me to excel - nothing goes unnoticed. Thank you Robert Kaster for introducing me to Miles Davis and Charles Mingus - they both proved to be invaluable during graduate school.

And finally, I want to thank my wife and best friend Susan. You are an amazing wife and friend for which I will always be indebted. You, above all else, have experienced every step along the path to this degree and I sincerely thank you for every ounce of support. Thank you for saying yes.

CONTRIBUTION OF AUTHORS

P. Shing Ho was involved in the design, analysis, and writing of each manuscript. Jeffrey Watson assisted in some aspects of manuscript preparation for Chapter 2. Jeffrey M. Vargason designed and determined initial crystallization conditions for the d(CCAGTACTGG) sequence in Chapter 3 and played a consultative role in the subsequent structure solution and refinement. Zebulon J. R. Jones provided some assistance with refinement for the d(CCDGTACTGG) structure in Chapter 4. Amy Teegarden, Zebulon J. R. Jones, Michael Harms and Dustin Raup were involved in DNA purification, crystallization, and structure refinement for data presented in Chapter 5. Emily Cavaliere helped with some aspects of crystallization and structure refinement for material in Chapter 5. Dustin Raup was involved in the initial stages of crystallization for the d(CCGATATCGG) sequence in Chapter 6.

TABLE OF CONTENTS

Page

1. Introduction 1

2. CAUTION! DNA Crossing: The crystal structures of Holliday Junctions 7

2.1 Summary 8

2.2 B.C.—Before Crystal Structures 10

2.3 Single Crystal Structures of DNA Holliday Junctions 11

2.3 Sequence and ions affect the geometry of DNA junctions 14

2.4 Conclusions and Perspectives 20

2.5 Acknowledgement 23

3. Effect of Sequence on the Conformational Geometry of DNA Holliday Junctions 24

3.1 Summary 25

3.2 Introduction 26

3.3 Materials and Methods 32

3.3.1 Crystallization and structure of d(CCGGCGCCGG). 32
3.3.2 Crystallization and structure determination of d(CCAGTACTGG) and d(CCAGTACbr^{5}UGG) 33

3.4 Results 36

3.4.1 Structure of d(CCGGCGCCGG) as a Holliday Junction. 37
3.4.2 Structure of d(CCAGTACTGG) as B-DNA duplexes. 47
3.4.3 Structure of d(CCAGTACbr^{5}UGG) as a Holliday Junction. 47

3.5 Discussion 53

3.6 Acknowledgements 62

4. Influence of Minor Groove Substituents on the Structure of DNA Holliday Junctions 63

TABLE OF CONTENTS (Continued)

Page

4.1 Summary 64

4.2 Introduction 65

4.3 Materials and Methods 71

4.3.1 Crystallization and structure solution of d(CCIGTACm^{5}CGG). 71
4.3.2 Crystallization and structure solution of d(CCDGTACTGG. 72

4.4 Results 75

4.4.1 Structure of d(CCIGTACm^{5}CGG) as a Holliday Junction. 76
4.4.2 B-DNA Structure of d(CCDGTACTGG). 88

4.5 Conclusions 93

4.6 Acknowledgements 95

5. How sequence defines structure? A crystallographic map of DNA structure and conformation 96

5.1 Summary 97

5.2 Introduction 97

5.3 Materials and Methods 101

5.4.1 Structural Classes 103
5.4.2 Crystallization Conditions 104
5.4.3 Crystal Lattices 106
5.4.4 Crystallographic map of DNA structural space 107

5.5 Discussion 110

5.6 Acknowledgements 118

6. Cation Effects on the Geometry and Formation of the DNA Junction in d(CCGATATCGG) 119

6.1 Summary 120

TABLE OF CONTENTS (Continued)

Page

6.2 Introduction 120

6.3 Results 125

6.3.1 Structure of d(CCGATATCGG) as resolved B-DNA duplexes 127
6.3.2 Overall Structures of d(CCGATATCGG) as DNA Holliday junctions. 130
6.3.3 Comparison of Junction Geometry. 131
6.3.4 Comparison of Junction Crossover Regions. 135

6.4 Discussion 139

6.5 Materials and Methods 149

6.5.1 Crystallization and structure determination of B-DNA form d(CCGATATCGG) 150
6.5.2 Crystallization and structure determination of d(CCGATATCGG) low salt junction 151
6.5.3 Crystallization and structure determination of d(CCGATATCGG) high salt junction 152

6.6 Acknowledgements 154

7. Conclusion and Discussion 155

8. Bibliography 162

9. APPENDIX 176

LIST OF FIGURES

Figure Page

2.1 DNA Holliday junctions....9

2.2 Core interactions in the stacked-X junction....13

3.1 Open-X and stacked-X forms of the Holliday junction 27

3.2 Crystal structure of d(CCGGCGCCGG) as a DNA Holliday junction 38

3.3 Superposition of d(CCGGCGCCGG) and d(CCGGTACCGG) junctions 43

3.4 Atomic interactions within the trinucleotide core of the d(CCGGCGCCGG) junction viewed perpendicular to the junction dyad axis 45

3.5 Crystal structure of d(CCAGTACbr^{5}UGG) as a DNA Holliday Junction 48

3.6 Atomic interactions within the trinucleotide core of the d(CCAGTACbr^{5}UGG) junction viewed perpendicular to the junction dyad axis 51

3.7 Superposition of the d(CCAGTACbr^{5}UGG) junction (blue) on the d(CCGGTACCGG) junction (red) 58

4.1 Three general forms of DNA Holliday junctions: (a) parallel stacked-X, (b) open-X, and (c) antiparallel stacked-X 66

4.2 Standard and modified nucleotide base pairs located in the R_3-Y_8 base pair step .70

4.3 DNA Holliday junction crystal structure of d(CCIGTACm^{5}CGG) 77

4.4 Stabilizing interactions within the stacked-X core region 80

4.5 Minor groove solvent interactions adjacent to $R_3 \cdot Y_8$ base step in DNA Holliday junctions 83

4.6 Structure of d(CCDGTACTGG) 89

5.1 Structures from the crystallographic screen of the IR sequence d(CCnnn$N_6N_7N_8$GG), where all 64 combinations of the $N_6N_7N_8$ trinucleotide are sampled....105

LIST OF FIGURES (Continued)

Figure Page

5.2 Map of DNA structure space sampled by crystals of d(CCnnnN$_6$N$_7$N$_8$GG).....108

5.3 Correlating sequence effects to atomic interactions in junctions.....111

6.1 Conformational variability of DNA Holliday junctions.....122

6.2 Structure of the d(CCGATATCGG) sequence as a resolved B-DNA duplex in the $P2_12_12_1$ crystal form.....128

6.3 Structures of the sequence d(CCGATATCGG) as four-stranded Holliday junctions in the C2 crystal form.....132

6.4 Stabilizing interactions within the central core of ATC-LS and ATC-HS.....136

6.5 Schematic of conformational changes associated with decreasing salt concentration for the d(CCGATATCGG) sequence.....146

LIST OF TABLES

Table Page

2.1 Structures of DNA Holliday junctions....16

3.1 Data collection and refinement statistics....34

3.2 Comparison of global conformational geometry values for d(CCGGCGCCGG) and d(CCAGTACBr^{5}UGG) with previous junction structures of d(CCGGTACCGG) and d(CCGGTACm^{5}CGG)....40

3.3 Minor groove widths measured as the phosphorus P to P distance for the phosphodiester linkage between dinucleotides....41

3.4 Comparison of helical twist....42

4.1 Data collection and refinement statistics....74

4.2 Comparison of conformational parameters for DNA Holliday junction structures of the d(CCGGTACm^{5}CGG), d(CCIGTACm^{5}CGG) and d(CCAGTACBr^{5}UGG) sequences, and B-DNA structures of d(CCAGTACTGG) in the monoclinic and hexagonal crystal forms, d(CCDGTACTGG) and d(CCGGCGCCGG)....86

5.1 Conformations observed in the single-crystals of d(CCnnn$N_6N_7N_8$GG)....100

5.2 Helical parameters for structures from the crystallographic screen of the sequence motif d(CCnnn$N_6N_7N_8$GG)....115

6.1 Data collection and refinement statistics for three conformations of the sequence d(CCGATATCGG)....126

6.2 Global conformation of DNA Holliday junctions....143

7.1 Comparison of global conformational geometry for all currently known inverted-repeat stacked-X DNA Holliday junctions....157

LIST OF APPENDIX TABLES

Table	Page
Supplemental Table 1	177
Supplemental Table 2	184
Supplemental Table 3	189

DEDICATION

I dedicate this thesis to my wife Susan. I am eternally grateful for her love, friendship, advice and steadfast support - qualities that I realize were instrumental to the completion of this work.

Sequence Dependent Conformational Variations in DNA Holliday Junctions

Chapter 1

Introduction

Four-stranded DNA junctions were first proposed by Robin Holliday in 1964 to explain abnormal segregation of alleles in fungi and yeast (*1, 2*). These junctions, now commonly referred to as Holliday junctions, have since been identified as structural intermediates in a number of biological processes including homologous recombination, viral integration and DNA repair. At some point, in most of these cellular processes, the junction structure is recognized by proteins which may perform any number of roles, from stabilizing a specific junction conformation to resolving the intermediate into B-DNA duplexes. Holliday junctions, composed of four DNA strands with duplex arms projecting from a central core, are known to be highly variable in global conformation. Understanding this conformational variability in both the presence, and absence, of bound protein is instrumental to better understanding how junctions participate in biological systems. This thesis aims to better characterize how nucleotide sequence within the core region of DNA-only junction structures affects their conformation.

It has become very clear that four-stranded Holliday junctions play an integral part in many recombination specific pathways. Following DNA replication, which

generates a duplicate copy of each allele, a single strand nick is introduced into each DNA molecule allowing a reciprocal exchange of strands between adjacent duplexes. In prokaryotic cells this process is facilitated by the RecA protein while eukaryotic systems have a RecA homolog called Rad51. Following ligation of the exchanged strands, a four-stranded Holliday junction is formed that is capable of migrating along the linked DNA arms in a process called branch migration. Branch migration is typically facilitated by a host of cellular recombination specific proteins (*3-6*), though it can also occur spontaneously (*7*). Following junction migration, the heteroduplex structure is resolved back into duplex DNA by the action of junction specific enzymes (integrases and nucleases) which cleave specific junction strands. This process of branch migration between adjoined DNA duplexes followed by junction resolution is a hallmark of the Holliday model of genetic recombination (*2*). Overall the general features of this model have remained intact, though it has undergone some revisions in detail (*8-11*).

Relatively recent advancements in our understanding of homologous recombination, especially at the atomic level, have provided significant insights into this fundamental cellular process. For instance, crystal structures of all three components of the prokaryotic RuvABC "resolvasome" have now been determined either in complex with DNA (RuvA), in complex with each other (RuvA and RuvB) or the protein by itself in the absence of DNA (RuvA, RuvB and RuvC) (*3, 12-15*). RuvA is a large tetrameric protein that binds to and stabilizes the junction structure which facilitates ATP-dependent branch migration by the RuvB helicase. The junction resolvase RuvC completes the recombination process in this system by

cleaving the junction structure, producing two duplex DNA's (*15, 16*). Thus, we currently have a significant amount of detailed information on the process of homologous recombination in prokaryotic systems, especially in regards to the RuvABC resolvasome (*5*), yet relatively little is currently known about the same processes in eukaryotic organisms (*6*). Analogous branch migration and junction resolution activities have been observed in mammalian extracts (*17-19*), although, to date, very few proteins have actually been identified (*6, 20*). Relatively recent studies have demonstrated that the eukaryotic helicases WRN, BLM and RecQ5β are capable of catalyzing branch migration *in vitro*, though it still isn't clear as to what their functionality may actually be (*21-23*). Indeed, in the case of BLM, the actual function of the protein may be to suppress crossover formation (*24*). Thus, it seems fairly clear that four-stranded junctions are important intermediates in homologous recombination, yet there is still a lot to be learned about their function and structure. Specifically, we now know that four-stranded DNA junctions serve as substrates for a wide range of cellular proteins, such as RuvA or WRN, and understanding the intrinsic structural properties of these junctions could provide valuable insights into their recognition and function.

A wealth of solution studies have demonstrated that Holliday junctions exist in two general conformations: open-X and stacked-X (*25-28*). In low salt solutions, the junction adopts the extended open-X form to minimize the strong repulsion between negatively charged phosphates at the cross-over strands that exchange between DNA double-helices (Figure 2.1b). With increasing cation concentrations the four duplex arms of the junction structure coaxially stack on one another to form semi-continuous

B-DNA duplexes (Figure 2.1a and 2.1c). This latter junction conformation, generally referred to as the stacked-X form, has two continuous (noncrossover) and two non-continuous (crossover) strands which can be oriented in either a parallel or antiparallel fashion. Chapter 2 of this thesis provides a more detailed review of the initial characterization and formation of these two junction conformations. The open-X form is the primary conformation that undergoes branch migration between two joined duplex DNA's and has been observed in crystal structures with RuvA (*3*) and the site specific recombinase Cre (*4, 29, 30*). Stacked-X Holliday junctions on the other hand have been proposed as the predominant form recognized by junction resolvases such as Hjc (*31, 32*) or T4 endonuclease VII (*33, 34*), although no co-crystal structure has yet been determined of a resolvase-junction complex. RuvC, which cleaves junctions in a sequence specific manner (*35*), seems to be an exception as it has been shown to induce an unfolded, symmetric, form of the junction upon binding (*36, 37*). Thus, there is a large disparity in the conformation of four-stranded junctions in biological systems and precisely how junction recognition occurs is still largely unknown.

This conformational variability of the junction structure has been the focus of a large number of studies using an assortment of techniques (such as gel electrophoresis (*26, 38, 39*), NMR (*40*), molecular modeling (*41*) and dynamics (*42*), fluorescence resonance energy transfer (*28, 43*) and chemical probing (*44*), to name a few) with a common objective - to understand how sequence and solvent affect formation and conformation of DNA Holliday junctions (*45*). These studies produced a general consensus on the arrangement of DNA strands in the open-X conformation but also

generated a significant amount of controversy regarding the stacked-X conformation (*41*). The stacked-X junction model that evolved from these studies had standard Watson-Crick base pairs within the junction core and antiparallel arrangement of DNA strands (*28, 41*), instead of parallel as proposed by the original Holliday model. This controversy surrounding the stacked-X conformation led to a significant amount of effort to try and obtain atomic structures in order to shed light on properties of the junction crossover and arrangement of DNA strands. Surprisingly, it wasn't until very recently that crystal structures of the stacked-X junction were determined using the DNA sequences d(CCGGTACCGG) and d(CCGGGACCGG) (*46, 47*). Chapter 2 of this thesis provides a more detailed overview of the initial crystallographic results, but generally the structures corresponded to the stacked-X model proposed by von Kitzing, et al. (*41*) with both sharing a common ACC trinucleotide motif in the central crossover region of the junction (underlined above). Furthermore, a series of stabilizing interactions were identified within the central core of both structures and a subsequent study demonstrated that disruption of these core interactions produced a change in global junction conformation (*48*). Thus, as the result of these novel junction structures a system is in place to study the effects of sequence within the central crossover of the stacked-X junction on global conformation.

One of the first questions we addressed in the studies described in the current thesis was to determine if a central $A_6C_7C_8$ sequence was truly required in the general inverted repeat sequence d(CCnnn$N_6N_7N_8$GG), where $N_6N_7N_8$ can be any of the four common nucleotides and *'nnn'* are specified accordingly to maintain the inverted repeat motif of the sequences, for junction formation within the present system. In

Chapter 3 we describe the initial expansion of this core sequence motif to include a guanine nucleotide in position N_6 (GCC core) and a 5-bromouracil but not thymine nucleotide in position N_8 (ACBr^{5}U core). Additionally, in Chapter 4 we demonstrate that the major groove interactions within the $N_3 \cdot N_8$ base step are key to junction formation within this crystallographic system. In order to determine the sequence dependence on junction formation in a more systematic fashion we pursued a crystallographic screen of all 64 potential sequence possibilities within the junction core. These results are described in more detail within Chapter 5 along with three supplementary tables in Appendix 1 which outline the current status of the screen. One surprising feature of this crystallographic screen was the identification of four amphimorphic sequences which crystallize in two different structural classes. In Chapter 6 we describe in greater detail one of these amphimorphic sequences, d(CCGATATCGG), which crystallized as both B-form DNA and four-stranded Holliday junction in a salt dependent manner. This result is very interesting and novel in that it allows us to characterize the affects of decreasing calcium chloride concentration on the atomic structure of the four-stranded junction which leads, ultimately, to resolved B-DNA duplexes of the same DNA sequence. The results presented in this thesis provide a detailed characterization, at the atomic level, of how sequence variation within the central core of the junction structure affects its global conformation. These results are discussed and summarized in broader terms within Chapter 7.

Chapter 2

CAUTION! DNA Crossing: The crystal structures of Holliday Junctions

Franklin A. Hays, Jeffrey Watson and P. Shing Ho

Published in *Journal of Biological Chemistry*,

American Society for Biochemistry and Molecular Biology, Bethesda, MD, USA

2003, **278** (50), 49663-49666

2.1 Summary

A four-stranded DNA junction (Figure 2.1a) was first proposed by Robin Holliday in 1964 as a structural intermediate in a mechanistic model to account for how genetic information is exchanged in fungi (*1*). This mechanism for genetic exchange is now generally known as homologous recombination and the four-stranded intermediate as the Holliday junction. The general mechanism for recombination has undergone a number of revisions in detail, but the Holliday junction remains a key component in the process and in a growing number of analogous cellular mechanisms (reviewed in (*49, 50*)), including site-specific recombination (*51*), resolution of stalled replication forks (*52, 53*), DNA repair (*5, 54*), and phage integration (*55*). Consequently, the structural and dynamic properties of the Holliday junction have been the focus of intense study since the mid-1960s. As is often the case in science, a multitude of single crystal structures of Holliday junctions in various forms have been solved over a relatively short period of time, but only after decades of disappointment. The first structures of junctions in complexes with recombination and DNA repair proteins were reported in 1997 and 1998 (*30, 56*), while junctions as RNA/DNA complexes (*57, 58*) and in DNA only constructs (*46, 47*) emerged just as the twentieth century came to a close. In this review, we will focus on the structures of DNA-only junctions and their geometries, as defined by sequence and ion dependent interactions.

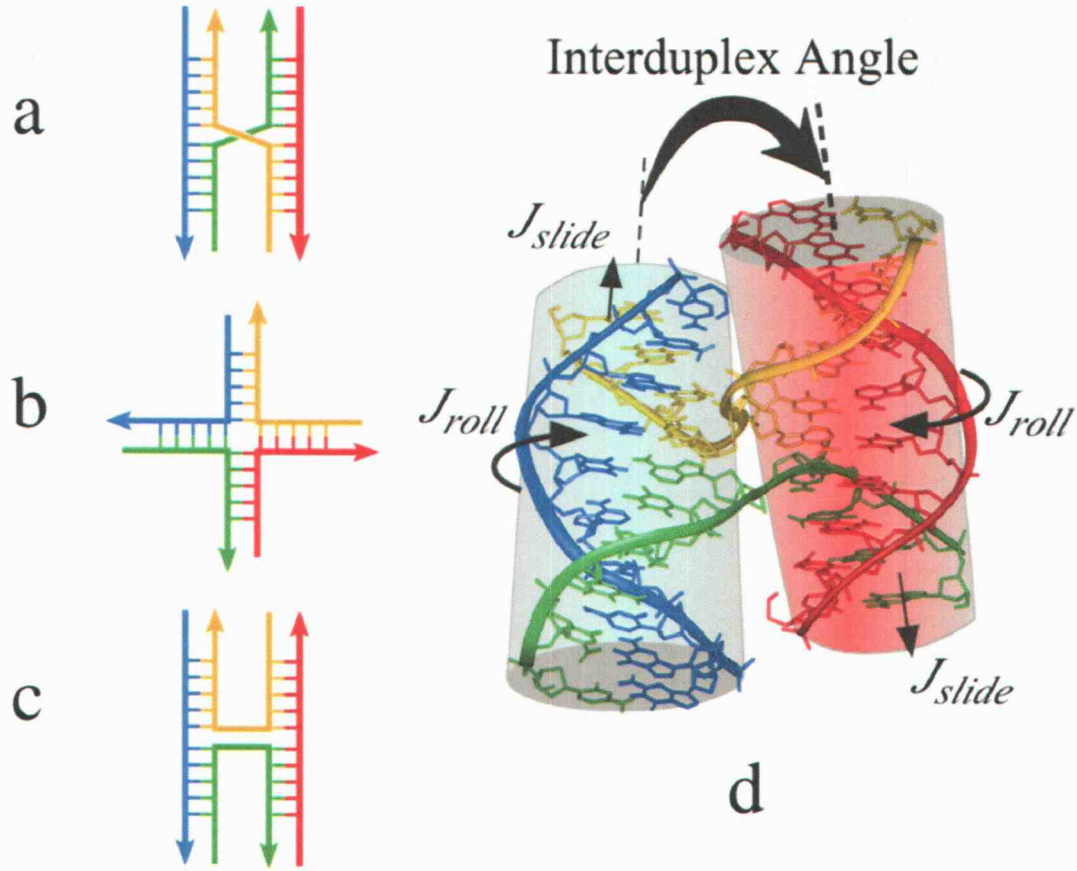

Figure 2.1. DNA Holliday junctions. Schematics of four-stranded DNA junctions in the parallel stacked-X (a), open-X (b), and antiparallel stacked-X (c) configurations, and definitions for the geometric parameters that vary in the stacked-X form (d). The parallel stacked-X is the topology of the junction initially proposed by Holliday (*1*), while the open-X form is seen in protein complexes and the antiparallel stacked-X is seen in the DNA and RNA/DNA junctions. The geometry of the stacked-X structure, as seen in single-crystal structures, varies according to the angle between the stacked duplex arms (interduplex angle), and the sliding (J_{slide}) and rolling (J_{roll}) of these arms along their respective helical axes.

2.2 B.C.—Before Crystal Structures

Before any of the crystal structures of junctions were available, the general architecture and dynamic properties of the Holliday junction were already being defined through biochemical and biophysical studies on immobilized junctions assembled from asymmetric sequences (reviewed in (*50*)). From comparative gel and solution studies, the junction was known to undergo ion dependent conformational transitions. In low salt solutions, the junction adopts an extended "open-X" form to minimize the strong repulsion between negatively charged phosphates at the cross-over strands that exchange between DNA double-helices (Figure 2.1b). The junction collapses to a stacked-X form when the phosphates are screened by cations, particularly by magnesium (Mg^{2+}) and other divalent ions. This more compact junction pairs the arms into coaxially stacked, semicontinuous duplexes, with each stacked duplex having an outside non-exchanging strand and an inside cross-over strand. The crossed-over strands can be oriented either in parallel or antiparallel directions.

Although the parallel configuration of the stacked-X junction (Figure 2.1a) has been suggested to be formed by extrusion of cruciform structures in negatively supercoiled closed circular DNAs (*59, 60*) and in synthetic DNA arrays (*61*), it is the antiparallel form that is seen by gel (*26, 38, 62*) and, subsequently, fluorescence resonance energy transfer studies (*27, 28, 63*), and by atomic force microscopy in interlinked sheets of DNA (*64, 65*). The basic model that emerged for the antiparallel

stacked-X junction (*26, 41*) has the arms paired and stacked to form near continuous B-type DNA duplexes, interrupted only by the U-turn of the junction cross-over. The two opposing stacked duplexes are related by a right-handed twist to give an interduplex angle of +60° across the junction. One of the problems with the antiparallel configuration, however, is that it does not allow migration of the junction along homologous DNA sequences. A resolution to this problem is that the stacked-X junction is not static, but can undergo dynamic conformational rearrangements: a conformational isomerization can occur in which the stacked duplex arms switch partners in a sequence dependent manner (*40, 43, 66*). This isomerization transition has recently been observed in single molecules (*25*), and has been proposed to involve the open-X structure (which does support junction migration) as a transient intermediate conformation (*46*). Additionally, stacked-X junction isomerization can affect the genetic outcome of homologous recombination through production of either patched or spliced gene products. The single crystal structures of DNA junctions and junctions in complex with proteins are now providing conformational details of the two junction forms at the atomic level, and, particularly with the stacked-X form, an understanding of how sequence and ions affect these structural details.

2.3 Single Crystal Structures of DNA Holliday Junctions

The first single-crystal structure of a DNA junction was seen in the Cre recombination complex (*30*). There are now structures of at least three unique

complexes with recombination proteins (*56, 67, 68*) and, in all cases, the junctions adopt the extended open-X form, with the DNA geometries determined primarily by the topology of the binding surfaces of the proteins (reviewed in (*49*)). The question of how DNA sequence and other factors affect the inherent structure of the Holliday junction remains and is being answered by single-crystal structures in the absence of proteins.

The first DNA Holliday junction crystal structures came not by design, but by happenstance while attacking other problems. The sequence d(CCGGGACCGG) was the first structure of a DNA Holliday junction, but the initial intent was to study the effects of tandem GA-mismatches on B-DNA (*46*). Shortly afterwards, the structure of a junction with all standard Watson-Crick base pairs (*47*) was solved as a control study for the psoralen cross-linked structure of d(CCGGTACCGG) (*69*). In these and all subsequent DNA-only studies, the junctions adopt the antiparallel stacked-X conformation (Figure 2.2) with general structural features that are remarkably similar to the model proposed in 1988 from solution and gel studies (*26, 41*). There were, however, some surprising observations that could not be predicted from earlier studies on asymmetric junctions, first being that these decanucleotide sequences could be crystallized as junctions at all. The single-crystal structures of many 10 base-pair sequences have been reported since 1987 (reviewed in (*70*)) as B-DNA, A-DNA and left-handed Z-DNA. In fact, early models of junctions were constructed using structures of decanucleotides in crystal lattices that resemble the general geometry of

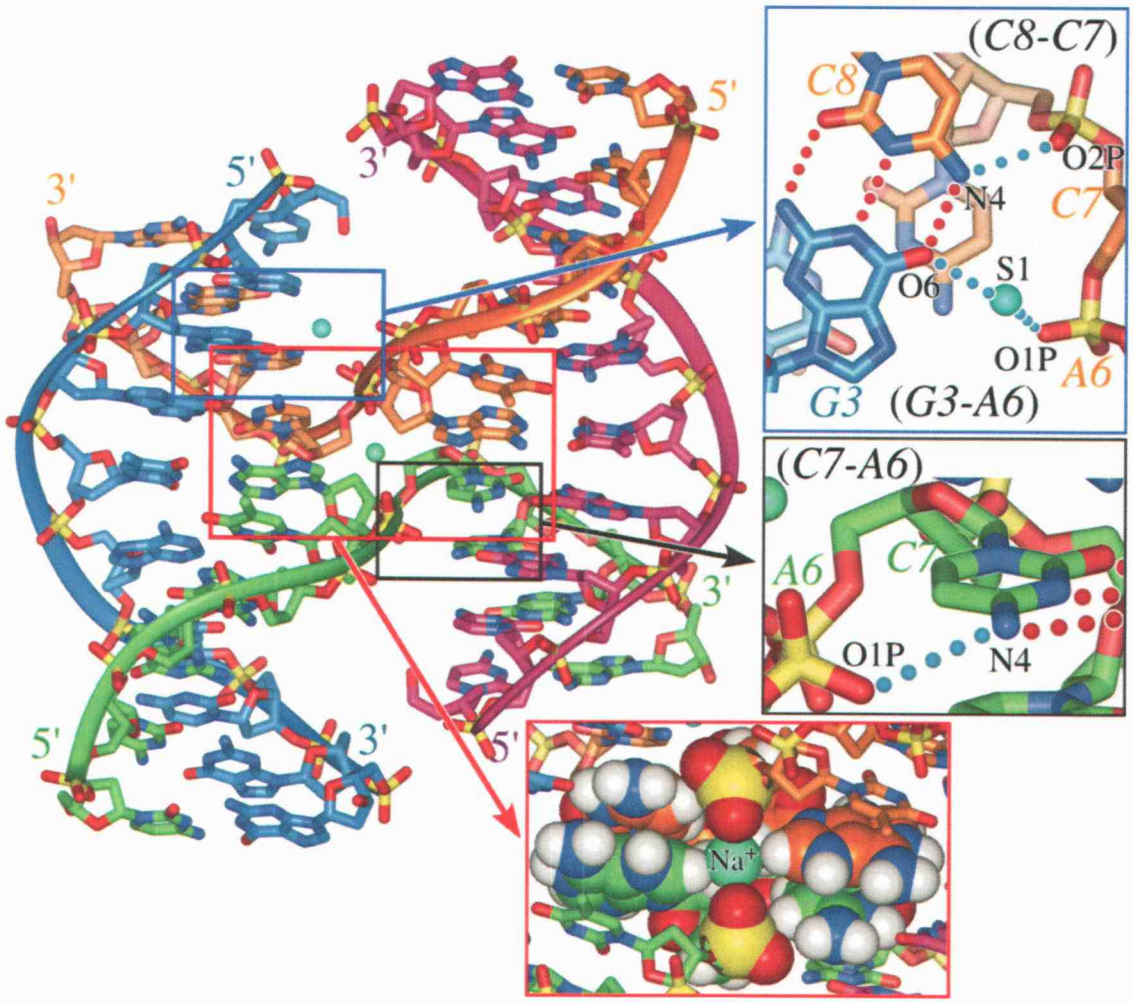

Figure 2.2. Core interactions in the stacked-X junction. The structure of d(CCGGTACCGG) as a junction (ACC-2Na structure) (*47*) is shown along with the core interactions at the junction cross-over that account for the formation and geometry of this DNA conformation. The locations of the core interactions (see Table 2.1) in the overall structure are indicated by boxes, and the regions in the boxes enlarged in each inset panel. The C_8-C_7 and G_3-A_6 interactions are shown in the top inset panel while the C_7-A_6 interaction is shown in the middle panel. Blue dots highlight hydrogen bonding interactions that are important within these interactions, while red dots are used to show hydrogen bonds in the standard Watson-Crick base pairs. The bottom inset panel shows the environment surrounding the sodium ion (Na^+) that sits in a cavity formed by the phosphates of two A_6 and the bases of the A_6 and C_7 nucleotides.

the stacked-X junction (*71, 72*). What became apparent when the actual Holliday junction structures were solved was that a common $A_6C_7C_8$ trinucleotide sequence and an associated set of unique intramolecular interactions define a common core in the two junctions (*47, 73*). Since these two early breakthroughs, we are learning that both sequence and ions at and near this ACC-core affect these interactions, and consequently the geometry of the stacked-X junction, in both subtle and profound ways.

2.3 Sequence and ions affect the geometry of DNA junctions

The mismatched and inverted repeat junctions are very similar in structure (*73*), but show minor structural differences that could be attributed to an unusual base interaction at the G·A mismatched base pairs (*46*). The first indication that the standard geometry of Holliday junctions could be perturbed in a systematic way was seen in the structure of the d(CCGGTACm^{5}CGG) (*48*), where m^{5}C is a 5-methylcytosine. This methylated structure demonstrated that, in addition to the interduplex angle, the stacked duplexes can be rotated and shifted relative to each other across the junction cross-over (we label these J_{roll} and J_{slide}, respectively, 2.1d), and that these two geometric parameters were dependent on the interactions at the junction core. The J_{slide} parameter measures the extent to which the stacked helices translate about their respective helix axes relative to the center of the junction cross-over, while J_{roll} rotates these duplexes along the helix axes (a large J_{roll} opens the

major groove surfaces of the junction cross-over, with a value of J_{roll} = 180° indicating that the major and minor grooves of the junction are equally accessible).

There are now 12 junction structures (as of the date of publication) deposited in the Nucleic Acid Database (*74*) representing 6 unique sequence types. All of the structures are in the antiparallel stacked-X form, and vary according to their interduplex angles, J_{roll} and J_{slide} (Table 2.1). What have we learned from these structures? Our first lesson is that the sequence requirements at the junction core are not exclusive to ACC, but can be expanded to $Pu_6C_7Py_8$, where Pu_6 can be either the purine bases adenine or guanine, and Py_8 can be cytosine, 5-methylcytosine, or 5-bromouracil (br^5U), but not thymine. This may be expanded further as more sequences become crystallized as junctions. The terminal base-pairs ($C_1 \cdot G_{10}$) are not critical for formation of junctions, and can be replaced by T_1A_{10} base pairs, as long as the trinucleotide core sequence remains intact (*75*).

As the sequence is changed and/or base substituent groups are added at the trinucleotide core, it is becoming increasingly clear that the core intramolecular interactions not only dictate the formation of the junction, but also its geometry as defined by the interduplex angle and, more significantly, J_{roll} and J_{slide}. In this discussion, we will start by defining a standard set of interactions that are associated with the reference "unperturbed" d(CCGGTACCGG) junction (ACC-4Na, see Table 2.1 for description of this nomenclature) (*47*) (Figure 2.2). In this case, $J_{slide} \approx 0$ Å and $J_{roll} \approx 160°$. The key interactions in this reference structure involve base substituents in the major grooves of the stacked duplex arms and the phosphate oxygens of the

Table 2.1. Structures of DNA Holliday junctions. Each published structure and available single crystal structures of DNA only Holliday junctions are listed with the abbreviation used in the review, a description (the sequence, number of DNA strands in the asymmetric unit (asu), important cation clusters in structure and reference), the geometric parameters (interduplex angle (IDA), J_{roll} and J_{slide}), and core interactions.

Structure[1]	Description of Structure				Geometry[2]			Core Interactions[3]			
	Sequence	asu	Cation	Ref[4]	IDA	J_{roll}	J_{slide}	C_8-C_7	G_3-A_6	C_7-A_6	Na-Cavity
ACC-4Na	CCGGTACCGG	4	Na	(46)	41.4°	159.6°	0 Å	Direct	$1H_2O$	Direct	Na^+
ACC-2Na	CCGGTACCGG	2	Na	(78)	42.7	163.5°	2.0 Å	Direct	$1H_2O$	Direct	Absent
ACC-2Ca1	CCGGTACCGG	2	Ca	UD0024	39.6°	158.5°	0 Å	Direct	$1H_2O$	Absent	H_2O
ACC-2Ca2	CCGGTACCGG	2	Ca	UD0025	39.4°	158.8°	0 Å	Direct	$1H_2O$	$1H_2O$	H_2O
tACC-4Ca	TCGGTACCGA	4	Ca	(74)	39.6°	159.9°	0 Å	Direct	$1H_2O$	Direct	H_2O
tACC-2Ca	TCGGTACCGA	2	Ca	UD0023	39.6°	168°	3.6 Å	Direct	Absent	Absent	H_2O
tACC-2Sr1	TCGGTACCGA	2	Sr	(74)	43.0°	169.0°	2.4 Å	Direct	$2H_2O$	$1H_2O$	Absent
tACC-2Sr2	TCGGTACCGA	2	Sr	UD0026	43.3°	169.5°	2.4 Å	Direct	$2H_2O$	$1H_2O$	H_2O
gACC-2Na	CCGGGACCGG	2	Na	(45)	40.6°	174.5°	0 Å	Direct	$1H_2O$	Absent	H_2O
ACmC-2Ca	CCGGTACm^{5}CGG	2	Ca	(47)	41.3°	170.4°	3.4 Å	$1H_2O$	$2H_2O$	$1H_2O$	Absent
GCC-2Na	CCGGCGCCGG	2	Na	(75)	40.0°	159.8°	0 Å	Direct	$1H_2O$	Direct	Absent
ACbU-2Ca	CCAGTACbr^{5}UGG	2	Ca	(75)	38.8°	164.6°	2.4 Å	Absent	Absent	$1H_2O$	Na^+

Table 2.1 (Continuted)

[1]Junctions are designated by the core sequence at $N_6N_7N_8$ positions, number of DNA stands in the asu, the primary cation type located in the structure, and, for multiple identical entries, order in which they were listed.

[2]The geometric parameters for the junctions (interduplex angle or IDA, J_{roll}, and J_{slide}) were calculated as previously described (38).

[3]The core interactions are C_8-C_7 (from the N4 nitrogen of the C_8 nucleotide base to phosphate oxygen of C_7 of the cross-over strands), G_3-A_6 (from the G_3 nucleotide of an outside strand to phosphate oxygen of A_6), and C_7-A_6 (from the N4 nitrogen of the C_7 nucleotide base to phosphate oxygen of A_6 of the cross-over strands). These can be direct hydrogen bonds, or meditated by one or more solvent molecules. In addition, the cavity of the junction sandwiched by the phosphate oxygens of the N_6 nucleotides of the crossover strands, may contain a sodium ion.

[4]References are given for published structures, while those that are unpublished are identified by the Nucleic Acid Database accession codes.

junction cross-over, and include 1) a direct hydrogen bond between the N4 nitrogen of the C_8 base and the C_7 nucleotide phosphate oxygen (C_8-C_7 interaction), 2) a solvent mediated interaction from the complementary G_3 nucleotide and the A_6 nucleotide phosphate oxygen (G_3-A_6 interaction) and 3) a direct hydrogen bond from the N4 nitrogen of the C_7 base to the A_6 phosphate oxygen (C_7-A_6 interaction) (2.2, insets). The first interaction helps explain the requirement for a cytosine at the C_8 position, while the last interaction, for the first time, provides a rationale for the C_7 cytosine of the original ACC-core trinucleotide. The involvement of the A_6 phosphate in two of these three interactions may explain the importance of having a purine base at the Pu_6 position of the core.

As we change this reference sequence, we see losses in the core interactions that affect both the structure and stability of the junction. For the most part (with one exception) structures that maintain all three key interactions effectively adopt the unperturbed reference structure of the ACC-4Na junction. The GA mismatched (gACC-2Na) junction (*46*) has lost the C_7-A_6 interaction and, consequently, we see a significant J_{roll} to open the major groove surfaces. Similarly, the methylated d(CCGGTACm^{5}CGG) junction (*48*) (ACmC-2Ca) places additional intervening solvent into the C_8-C_7, C_7-A_6, and G_3-A_6 interactions, and these perturbations are associated with large values for both J_{slide} and J_{roll}.

In the case of the br^{5}U-containing junction of ACbU-2Ca, the first two core interactions (C_8-C_7 and G_3-A_6) are entirely missing, and the C_7-A_6 interaction becomes solvent mediated (*76*). Why then is this not the most perturbed structure? In this structure, a new water-mediated interaction is introduced between the N7 nitrogen of

the A_6 nucleotide along one crossing strand and the phosphate oxygen of the A_6 nucleotide of the opposite crossing strand (A_6-A_6 interaction). This compensating A_6-A_6 interaction does not, however, diminish the importance of the three core interactions for the formation of junctions. Indeed, the ACbU-2Ca sequence, under our conditions, has been crystallized as both a junction (*76*) and as resolved B-DNA duplexes (in a different crystal system), suggesting that the thermodynamic difference between the two forms in this particular sequence is very small and is associated with the loss of these core interactions. The sequence d(CCGGCGCCGG) (GCC-2Na) has also been crystallized in both conformations (*76-78*), but the crystallization conditions for the junction versus B-DNA were very different and may account for the conformational differences. The GCC-2Na junction maintains the three core interactions and, consequently, is nearly identical to the reference ACC-4Na structure, even though the adenine A_6 of the core trinucleotide has been replaced by a guanine base.

Cation interactions are expected to be very important in the structure and formation of the junction, particularly divalent cations. In the crystal structures, a sodium ion was originally identified in the central cavity of the ACC-4Na structure (*47*), and a similar solvent has been observed in several other structures (Table 2.1); however, its presence or absence in itself does not appear to be a major determinant of the junction structure or stability. In the case of the ACbU-2Ca junction (*76*), its presence may be an added factor that contributes to the stability of the junction in the absence of 2/3 of the core interactions, while its absence may help account for the perturbations to J_{slide} and J_{roll} in the closely related ACC-2Na junction (*79*).

The other significant cation interactions involve aqueous calcium (*48*) and strontium clusters (*75*) that sit in the grooves of the stacked duplexes in several junctions. Again, the effects of these cation complexes on the structure of the junction depend on how they perturb the core interactions. For example, five strontium-water complexes were located in the minor grooves of each stacked duplex of the tACC-2Sr1 (*75*) and tACC-2Sr2 junctions. In both cases, additional waters intervene in the core G_3-A_6 and C_7-A_6 interactions, resulting in nearly identical perturbations to J_{roll} and J_{slide}.

2.4 Conclusions and Perspectives

We are now starting to understand how sequence, ions, and drugs (this latter effect, though not discussed here, is covered previously in (*69, 80*)) affect the detailed conformation of the Holliday junction in single-crystals. As with any set of crystallographic structures, however, we need to address the question of whether the conclusions that we draw from these studies correspond to how DNA behaves in solution and in the cell. In most respects, the crystal structures do indeed conform to the general features of Holliday junctions as defined from previous studies, with one exception. In all cases, the interduplex angles of the crystal structures are about 20° more shallow than the ~60° observed from earlier gel mobility, FRET and AFM studies (*26-28, 38, 62-64*). One reason for this discrepancy may be that the crystallographic studies all use symmetric sequences where the junction is fixed in place by specific intermolecular interactions. In contrast, the non-crystallographic

studies on junctions were nearly all applied to immobilized junctions that are assembled from four unique asymmetric sequences and are thus locked in place by sequence design rather than by any defined set of molecular interactions, other than standard Watson-Crick base pairing. Consequently, we would expect these sequence immobilized constructs to be missing the core and even solvent mediated interactions between the stacked arms and the phosphates at the junction cross-over seen in the crystal structures and therefore would show significant deviation from the reference ACC-4Na junction. This is supported by AFM studies which show that indeed the interduplex angle of asymmetrically locked junctions are ~60°, while the angle in similar constructs that incorporate the symmetric ACC core trinucleotide is ~43°, as seen in the crystal structures (*65*). In addition, these same ACC-containing constructs were shown by hydroxyl radical footprinting to cross-over at the nucleotide positions seen in the crystal structures (*65*). Thus, we can assert that the sequence dependent interactions at the junction core, as identified in the single crystal structures, are relevant in other systems and that these interactions do indeed define the conformational details of the stacked-X junction even in solution.

Are these molecular details important for the participation of Holliday junctions in recombination-dependent cellular processes? As we know, junctions in the antiparallel stacked-X form cannot migrate along the DNA sequence, but can do so if it adopts the open-X form. In solution, stacked-X stability and conformational switching rates have been shown to be dependent on the population of open-X junction available for branch migration (*7, 81*). We see that disrupting one or more of the core interactions allows the stacked-X junction to roll, slide, and twist. We

suggest, therefore, that sequences in which this set of core interactions are disrupted or compromised by intervening solvent will be more conformationally malleable and thus be more amenable to switching to an open-X form and, consequently, promote branch migration.

How might these core interactions affect protein binding? As we have already seen, the junctions in protein complexes all adopt the open-X and not the stacked-X form. Thus, the barriers to conformational switching that we have discussed above should also contribute to the energetics of protein binding—any effect that facilitates formation of the open-X structure should also favor protein binding. In addition, we see that these core interactions, as they are affected by sequence and ions, specify the accessibility of the major and minor grooves (as measured by the parameter J_{roll}). Many junction-resolving enzymes, such as T4 endonuclease VII (*33*), are known to recognize the stacked-X structure from the minor groove face and therefore we see that accessibility also plays an important role in protein recognition and binding. In short, it appears that a stacked-X junction that is structurally malleable, which we propose is determined by a set of core interactions, will also have a lower barrier to conformational switching and be more accessible, both of which are important for binding and recognition by junction resolving and DNA repair enzymes. At this point, however, it is unclear exactly how specific proteins take advantage of the structural and dynamic features of the stacked-X junction to define their sequence dependent functions.

2.5 Acknowledgement

The work from P. S. Ho's laboratory has been supported by the National Institutes of Health (R1GM62957A), the National Science Foundation (MCB0090615), and the Environmental Health Sciences Center at Oregon State University (NIEHS ES00210).

Chapter 3

Effect of Sequence on the Conformational Geometry of DNA Holliday Junctions

Franklin A. Hays, Jeffrey M. Vargason, and P. Shing Ho

Published in *Biochemistry*,

American Chemical Society, Washington, D.C., USA

2003, **42**, 9586-9597

3.1 Summary

Structures of the DNA sequences d(CCGGCGCCGG) and d(CCAGTACBr^{5}UGG) are presented here as four-way Holliday junctions in their compact stacked-X forms, with antiparallel alignment of the DNA strands. Thus the *ACC*-trinucleotide motif, previously identified as important for stabilizing the junction, is now extended to PuCPy, where Pu is either an adenine or guanine, and Py is either a cytosine, 5-methyl cytosine, or 5-bromouracil, but not thymine nucleotide. We see that both sequence and base substituents affect the geometry of the junction in terms of the interduplex angle as well as the previously defined conformational variables, J_{roll} (the rotation of the stacked duplexes about their respective helical axis) and J_{slide} (the translational displacement of the stacked duplexes along their respective helical axis). The structures of the GCC and parent ACC containing junctions fall into a distinct conformational class that is relatively undistorted in terms of J_{slide} and J_{roll}, with interduplex angles of 40°-43°. The substituted ACBr^{5}U structure, however, is more akin to that of the distorted methylated ACm^{5}C containing junction, with J_{slide} ($\geq$ 2.3 Å), similar J_{roll} (2.5°) opening the major groove-side of the junction, but shows a reduced interduplex angle. In contrast, the analogous d(CCAGTACTGG) sequence has to date been crystallized only as resolved B-DNA duplexes. This suggests that there is an electronic effect of substituents at the pyrimidine Py position on the stability of four-stranded junctions. The single crystal structures presented here,

therefore, show how sequence affects the detailed geometry and, subsequently, the associated stability and conformational dynamics of the Holliday junction.

3.2 Introduction

The exchange of genetic information across double-helical DNA through recombination is an important process in DNA biochemistry, and has been implicated in the mechanisms of DNA repair, replication restart, and viral integration (*49*). The critical role of a four-stranded junction as a DNA intermediate in recombination was proposed nearly forty years ago by Robin Holliday (*1*). The structure of this four-stranded complex has been elucidated in detail through a series of single crystal structures as complexes with proteins (*49*) and, more recently, as isolated DNA constructs (*80*). All of the DNA junctions that have been crystallized to date are decanucleotides with a common ACC trinucleotide motif (or its methylated variant ACm^5C) at the $N_6N_7N_8$ positions (represented by the sequence d(CCGGT*ACC*GG), where the underlined nucleotides are the ACC junction core). Not surprisingly, these structures have many common structural features. All adopt the compact antiparallel stacked-X form in which pairs of the duplex arms stack collinearly into nearly continuous double-helices (broken only at the cross-over point of the junction on the inside strand of each pair, Figure 3.1). The structure with a methylated analogue of the ACC-core shows that the substituent group perturbs the stacked-X geometry. Here, we present the single crystal structures as Holliday junctions of the two

Figure 3.1. Open-X (*a*) and stacked-X (*b*) forms of the Holliday junction. The phosphodeoxyribose backbone of the DNAs are traced with ribbons, and color-coded to distinguish between strands of the junction. *a.* The open-X form modeled from the DNA in the single-crystal structure of the Cre-loxP complex (*67*). The protein is shown in grey, with the α-helices rendered as cylinders. *b.* The compact stacked-X structure of the reference d(CCGGTACCGG) single-crystal structure (*47*). The coaxially stacked arms in each column of pseudocontinuous duplexes are shown as cylinders. The conformational parameters of the junction are shown as the twist of the stacked duplexes across the junction (the interduplex angle), the translation of each set of stacked duplexes along their respective helix axes (J_{slide}), and the rotation of the duplexes about the axes (J_{roll}).

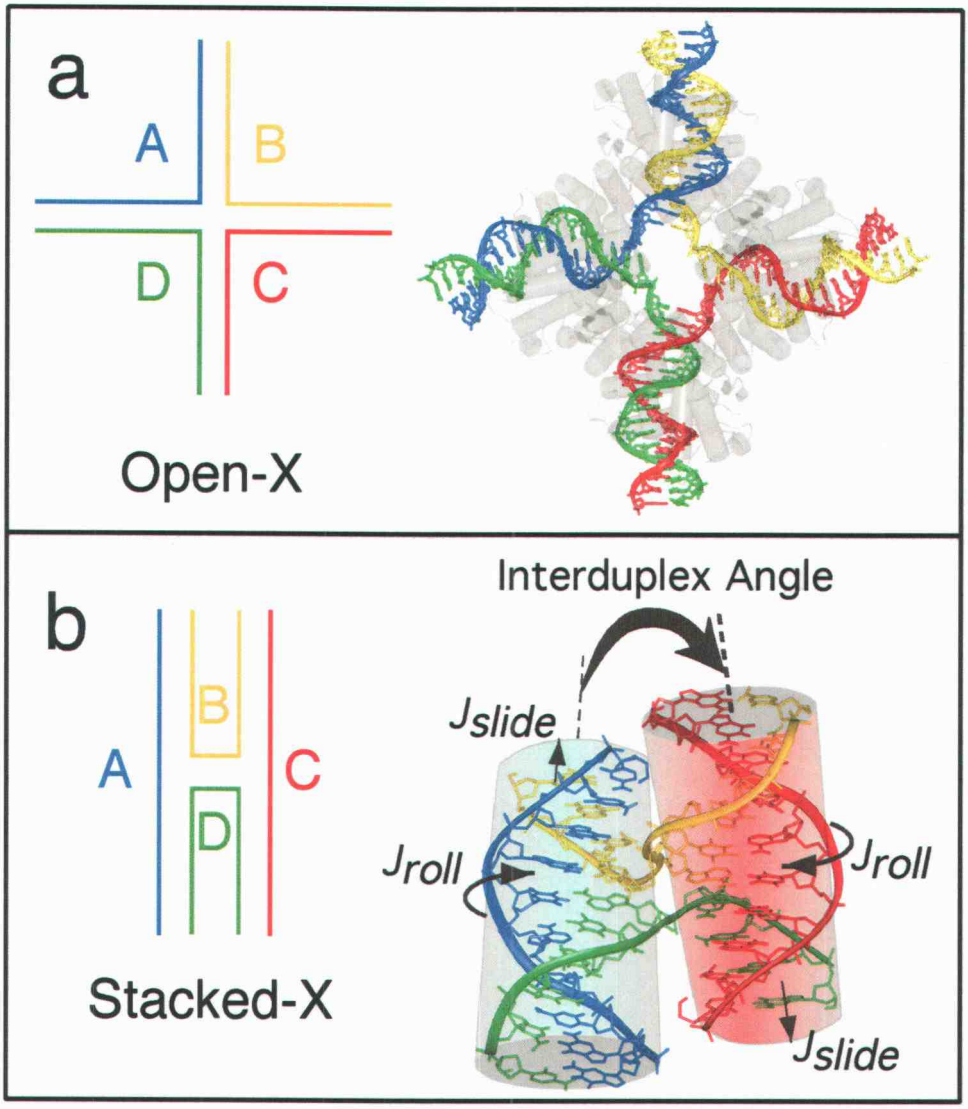
a
A
B
D
C
Open-X
b
A
B
C
D
Stacked-X
Interduplex Angle
Jslide
Jroll
Jroll
Jslide

Figure 3.1

sequences d(CCGGC*GCC*GG) and d(CCAGT*AC*br^{5}UGG) which do not contain the ACC junction motif at the core trinucleotide (underlined) positions.

The compact stacked-X form of the Holliday junction was first proposed from gel mobility assays (*38, 62*) and fluorescence resonance energy transfer studies in solution (*27, 28, 63*) to form under high salt conditions in nonhomologous sequence constructs that immobilize the four-way junction at specific sites. The general model from these studies are that high concentrations of cations help to screen the negatively charged phosphates of the backbone allowing the four arms, which would be splayed out from the junction in the open-X form (Figure 3.1a), to pair and coaxially stack to form two sets of semicontinuous double-helices (Figure 3.1b). The solution studies also showed that these pairs of stacked duplexes are related across the junction cross-over by a right-handed interduplex angle of ~60°. This general geometric relationship was confirmed by atomic force microscopy (*64*), but the most detailed molecular details of this conformation have come only from single-crystal X-ray diffraction studies.

The structure of the Holliday junction has now been determined in a number of nucleic acid constructs, including RNA-DNA complexes (*57, 58*) and DNA constructs with true inverted repeat patterns (*47*), mismatched base-pairs (*46*), the cross-linking drug psoralen (*69*), and methylated cytosines (*48*). The DNA structures show that the geometry of the compact stacked-X junction (Figure 3.1b) was basically that determined in solution, with the coaxially stacked duplexes related by a right-hand rotation across the junction cross-over. In all but one case so far, the DNA junction structures were determined from decanucleotide sequences with a core ACC

trinucleotide (or its methylated analogue, ACm^5C) at the $N_6N_7N_8$ positions, as in the sequences d(CCGGG*ACC*GG) (*46*), d(CCGGT*ACC*GG) (*47*), and d(TCGGT*ACC*GA) (*75*). The one exception to date is the structure of the psoralen cross-linked sequence d(CCGCT*AGCGG), which was proposed to form a junction because of the destabilization of the duplex by the drug cross-linked thymine base (T*) (*69*). Otherwise, the crystal structures demonstrate the significance of this core ACC trinucleotide, particularly through the interactions from the cytosine base at the C_8 cytosine to the phosphate backbone at the junction cross-overs.

The methylated junction seen in d($CCGGTACm^5CGG$) first showed that, in addition to the interduplex angle, the geometry of the junction can also be affected along and around the helix axes of the collinear duplex arms (*48*). The methyl substituent was shown to disrupt the direct hydrogen bonding interactions between the cytosine C_8 base and the phosphate backbone of the cross-over. Furthermore, the accessibility of the major and minor grooves of the stacked duplex arms were shown to depend on the rotation of these duplexes along their helix axes (J_{roll}) and the arms were seen to slide along their axes (J_{slide}) to affect the overall symmetry of the complex. Similar distortions to J_{roll} were seen to be affected by the types of cation interactions localized at the junction (*75*).

The interduplex angle relating the orientation of the stacked duplex arms across all of these ACC-type junctions are relatively constant, at ~40°, which appears to be more shallow than the 60° seen in solution (*50*). However, a study of immobilized junctions that incorporated the ACC trinucleotide motif showed that this angle, as well as many details concerning the position of the cross-over did not result

from crystal lattice distortions, but can directly be attributed to the sequence at the junction (*65*). This raises the question of how sequence variations would affect the structure and geometry of the junction. From solution studies, it has been shown that the sequence around an immobilized junction defines which arms pair in the coaxially stacked duplexes (*40, 43, 66*). In addition, single-molecule FRET studies show that the exchange between coaxially stacked arms is dynamic, with the junction undergoing rapid conformational rearrangements in a sequence dependent manner (*25*). However, the strong apparent requirement of the ACC trinucleotide for their crystallization has not allowed us to study the effect of sequence at the critical positions on the conformational features of the Holliday junction.

The question is thus raised as to whether the ACC core is an absolute requirement for the formation of DNA Holliday junctions, particularly in crystals. The sequence d(CCGGCGCCGG) had earlier been solved as standard B-DNA duplexes, but recently had been reported to crystallize in a lattice similar to that of the mismatched d(CCGGGACCGG) junction (*77*). We confirm in this study that indeed this DNA crystal form is that of a four-stranded junction, indicating therefore the ACC trinucleotide sequence motif is not an absolute requirement for crystallization of the junction. We can thus compare the effects of N_6 being either an adenine or guanine on the four-stranded conformation. In addition, we show that the sequence d(CCAGTACBr^{5}UGG), but not the analogous d(CCAGTACTGG) sequence crystallizes as a junction, showing that the substituent at this Py_8 position strongly influences the formation and geometry of the Holliday junction.

3.3 Materials and Methods

All deoxyoligonucleotides were synthesized on an Applied Biosystems DNA synthesizer in the Center for Gene Research and Biotechnology at Oregon State University using phosphoramidite chemistry, with the trityl-protecting group left intact at the 5'-terminal nucleotide for subsequent purification by reverse-phase HPLC. The purified DNA was deprotected by treatment with 3% acetic acid for fifteen minutes, neutralized with ammonium hydroxide, and desalted on a Sigma G-25 Sephadex column. Samples were lyophilized and stored at –80°C, then resuspended in deionized double-distilled water prior to crystallization. All crystals were grown using the sitting drop vapor diffusion method.

3.3.1 Crystallization and structure of d(CCGGCGCCGG). Crystals were grown at 25°C from solutions containing 0.5 mM DNA, 25mM sodium cacodylate buffer (pH 7.0), 7.5mM $CaCl_2$, and 0.1mM spermine tetrahydrochloride in the crystallization drop, and equilibrated against a reservoir solution of 28% (v/v) 2-methyl-2,4-pentanediol (MPD). These are very different conditions compared to those that yielded crystals of the B-DNA duplex form of the sequence (*82*) (B-DNA duplex crystals were grown at 4°C by microdialysis, with solutions containing 2mM DNA, 10mM Tris-HCl, pH 7.0, 150mM MgCl and equilibrated against the same buffer with 24% MPD).

A thin diamond plate crystal measuring 0.25 mm x 0.25mm x 0.1 mm was used for data collection and found to be in the monoclinic *C2* space group, with unit cell dimensions a = 66.51 Å, b = 24.18 Å, c = 37.00 Å, and β = 110.01°. X-ray

diffraction data were collected to 1.7 Å resolution at liquid nitrogen temperatures using CuKα radiation from a RUH3R generator with an RAXIS-IV image plate detector. The volume of the unit cell in this space group indicated that there are two strands of DNA in the asymmetric unit of the crystal lattice. Thus a search model was constructed using two strands of the d(CCGGTACCGG) Holliday junction (*47*), one crossover and one non-crossover, for molecular replacement using EPMR (*83*) against 2.5Å data. A distinct structural solution with two unique DNA strands adjacent to the crystallographic 2-fold axis yielded a correlation coefficient of 59% and R_{cryst}= 52%. Subsequent refinement in CNS (*84*) using rigid body refinement, followed by simulated annealing, several rounds of positional and individual B-factor refinement, and addition of solvent produced final values of R_{cryst} = 22.9% and R_{free} = 26.0% (Table 2.1). The coordinates and structure factors have been deposited in the Protein Data Bank (*74*) with accession number 1P4Y.

3.3.2 Crystallization and structure determination of d(CCAGTACTGG) and d(CCAGTACBr^{5}UGG). Crystals of d(CCAGTACTGG) were grown at 20°C from solutions containing 0.6mM DNA in 5mM Tris-HCl buffer (pH=7.5), 25mM calcium acetate, and 7% MPD, and equilibrated against 30% MPD. A single crystal measuring 0.2mm x 0.2mm x 0.3 mm was used for data collection under liquid nitrogen temperatures on BIO-CARS beamline 14-BMC at the Advanced Photon Source, Argonne National Labs, with 1.0Å radiation. Crystals of d(CCAGTACBr^{5}UGG) were grown at room temperature from solutions containing 0.6mM DNA in 5mM Tris-HCl

Table 3.1. Data collection and refinement statistics

	d(CCGGCGCCGG)	d(CCAGTACBr^{5}UGG)	d(CCAGTACTGG)
Data Collection			
Unit Cell Parameters	a = 66.5Å, b = 24.2Å, c = 37.0Å, β = 110.0°	a = 64.9Å, b = 22.6Å, c = 38.1Å, β = 106.8°	$a = b$ = 33.4Å, c = 87.7Å
Space group	*C2*	*C2*	$P6_122$
Total reflections	50,581	13,013	32,690
Unique reflections	5,137	4,011	2,143
Resolution	50 Å - 1.6 Å	50 Å - 1.8 Å	50 Å - 2.0Å
Completeness[a]	78.7% (48.0%)	82.0% (34.9%)	94.6% (53.2%)
<I/sigma>[a]	22.1 (2.7)	23.8 (4.1)	25.3 (3.3)
R_{merge}[a,b]	4.9% (27.8%)	4.1% (18.9%)	4.8% (12.4%)
Refinement			
Resolution	20 Å-1.7Å	20 Å-1.9 Å	20 Å-2.0 Å
R_{cryst} (R_{free})[c]	22.9% (26.0%)	21.5% (24.3%)	23.6% (26.8%)
DNA atoms	404	404	202
Solvent atoms	116	89	67
RMSD bond lengths[d]	0.007 Å	0.005 Å	0.015 Å
RMSD bond angles[d]	1.24°	0.88°	1.75°
Conformation	Junction	Junction	B-DNA

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

[b]$R_{merge} = \sum_{hkl} \sum_i | I_{hkl,i} - \langle I \rangle_{hkl} | / \sum_{hkl} \sum_i | I_{hkl,i} |$ 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.

[c]$R_{cryst} = \sum_{hkl} | F_{obs} - kF_{calc} | / \sum_{hkl} |F_{obs}|$. $R_{free} = R_{cryst}$ for 10% of reflections that were not used in refinement (*84*). All refinements were performed targeting maximum likelihood.

[d] Root mean square deviation of bond lengths and angles from ideal values.

buffer (pH=7.5), 120mM calcium acetate, and 16% MPD, and equilibrated against 20% MPD. A 1.9Å dataset was collected on this crystal at liquid nitrogen temperatures using CuKα radiation from the in-house RUH3R generator with an RAXIS-IV image plate detector. Crystals of the d(CCAGTACTGG) are in the hexagonal space group $P6_122$ with unit cell dimensions a = 33.38 Å, b = 33.38 Å, and c = 87.67 Å, while those of d(CCAGTACbr^{5}UGG) are in the monoclinic space group *C2* with unit cell dimensions a = 64.84 Å, b = 22.58 Å, c = 38.07 Å, and β=106.78°.

The structure of d(CCAGTACTGG) was solved by molecular replacement against a 2.0 Å dataset using a previously solved B-DNA duplex model in a similar hexagonal lattice. Resulting solutions from an EPMR search yielded a correlation coefficient of 64.2% and R_{cryst} = 47.4%. Subsequent refinements in CNS using rigid body refinement, simulated annealing followed by standard positional and individual B-factor refinement, and addition of solvent resulted in final values of R_{cryst} = 23.6% and R_{free} = 26.8%.

The lattice of d(CCAGTACbr^{5}UGG) crystal was similar to previously crystallized DNA junctions, suggesting that this br^{5}U analogue of the d(CCAGTACTGG) sequence could in fact be a four-stranded complex. Thus, the structure of this sequence was solved by molecular replacement using the two unique strands of the d(CCGGTACm^{5}CGG) junction structure (*48*) as the initial search model in an EPMR search. The search resulted in a position and orientation of the model having a correlation coefficient of 74% and R-Factor of 44.5% for the best solution. An initial round of rigid body refinement and simulated annealing in CNS yielded values of R_{cryst}=42.5% and R_{free}= 43.5%. Omit maps calculated with the crossover

phosphates removed and after simulated annealing of the edited model showed distinct F_o-F_c density consistent with the cross-overs of the junction and discontinuities in phosphodiester backbones along the stacked DNA duplexes. Thus, the structure of d(CCAGTACbr^{5}UGG) was refined as a DNA Holliday junction with two strands, one crossover and one non-crossover, in the asymmetric unit, with the two additional related strands generated by a crystallographic two-fold axis to form a complete four-stranded complex. Refinement was carried out in CNS using rigid body and simulated annealing routines followed by standard positional and individual B-factor refinement producing a final R_{cryst}=21.5% and R_{free}= 24.3% after addition of solvent. The coordinates and structure factors have been deposited in the Protein Data Bank (*74*) with accession number 1P4Z for d(CCAGTACTGG) and 1P54 for d(CCAGTACbr^{5}UGG).

All data was reduced using the HKL suite of programs (*85*). Root-mean-square-deviation (RMSD) values were calculated using the algorithm (*86*) implemented in the program ProFit v2.2 (http://www.bioinf.org.uk/software/profit/). Structural analysis was performed with CURVES 5.2 (*87*) and X3DNA (*88*).

3.4 Results

The crystal structures presented here show that d(CCGGCGCCGG) and d(CCAGTACbr^{5}UGG) both crystallize as four-stranded Holliday junctions in the compact antiparallel stacked-X form, with the former nearly identical to that of the

parent d(CCGGTACCGG) junction (*47*) and the latter most closely related to that of the methylated sequence d(CCGGTACm^{5}CGG) (*48*). The sequence d(CCAGTACTGG), however, was determined to be resolved B-DNA duplexes, now in two unrelated crystal forms. Thus, this study shows that the $A_6C_7C_8$ trinucleotide motif identified as important for stabilizing the junction can accommodate other nucleotide bases at the 6th and 8th positions. In addition, the two bases have their own distinct effect on the conformation of the detailed intramolecular interactions and, consequently, the overall geometries relative to previously determined structures. Thus, the resulting junction structures will be interpreted through comparisons between the current and previously reported structures in their respective structural classes, with the d(CCGGTACCGG) structure serving as the reference for the unsubstituted and d(CCGGTACm^{5}CGG) for the substituted junctions. These structures therefore allow us to define each component of the junction geometry, including the interduplex angle, rolling of the stacked duplex columns relative to each other (J_{roll}) and sliding of one junction relative to the other along their helical axes (J_{slide}), and how they can be affected by the sequences at the core of the junction.

3.4.1 Structure of d(CCGGCGCCGG) as a Holliday Junction. The single crystal structure of the sequence d(CCGGC*GCC*GG) (which we will call the GCC-junction), originally solved as a B-DNA duplex (*82*) under different crystallization conditions, was recently reported to crystallize in a crystal lattice that is isomorphous with that of a DNA four-way junction (*77*). We show here that, under our crystallization conditions, the sequence indeed is a Holliday junction. The complete

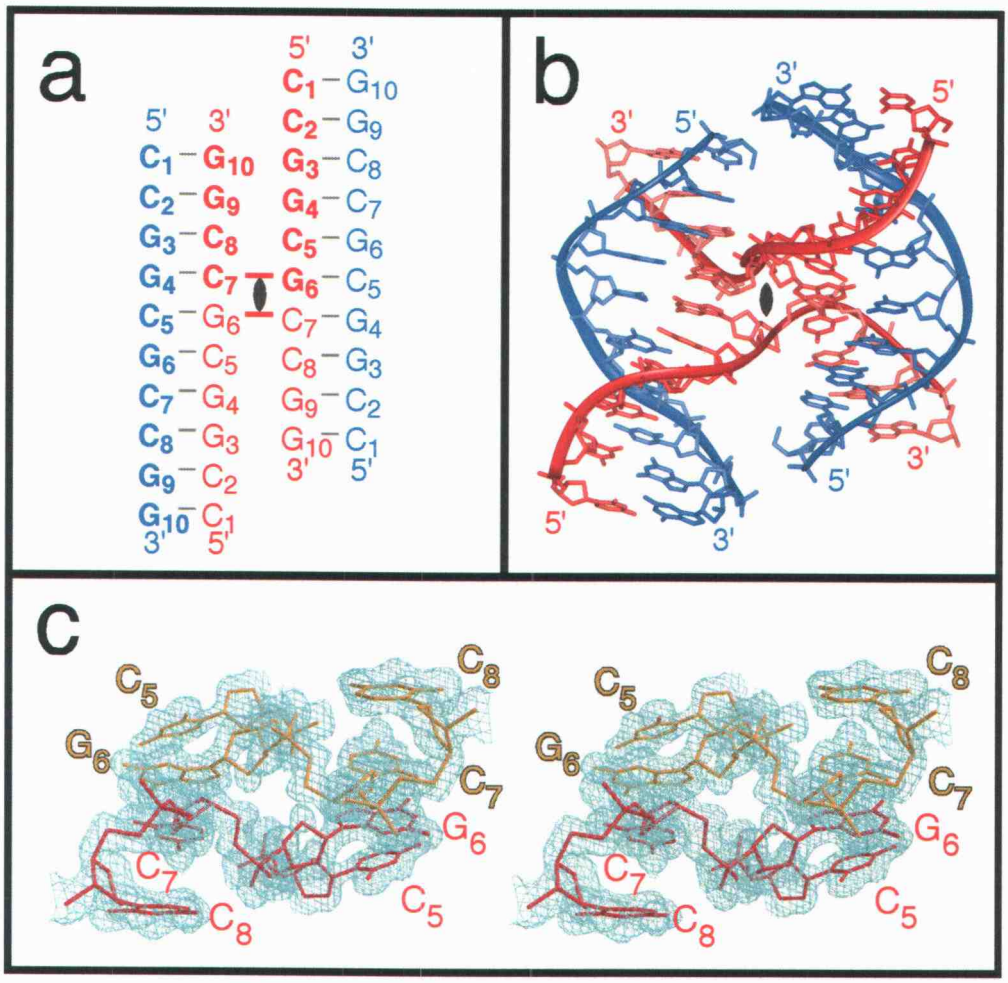

Figure 3.2. Crystal structure of d(CCGGCGCCGG) as a DNA Holliday junction. *a*. Sequence topology of the d(CCGGCGCCGG) junction. The four-stranded antiparallel stacked-X Holliday junction is generated by applying the crystallographic two-fold symmetry to the two unique strands (bold). Strands are numbered from 1 to 10 in the 5' to 3' direction, with the inside crossing strands colored red and the outside noncrossing strands colored blue. *b*, The atomic structure of d(CCGGCGCCGG). Chemical bonds in the structure are rendered as sticks and the phosphodeoxyribose backbone rendered as a solid ribbon (colors and strand designations are as in *a*, figure rendered with InsightII from MSI/Biosym, Inc.). *c*, Electron density map. The $2F_o$-F_c map (contoured at 1σ) shows the discontinuity in electron density between nucleotides G_6 and C_7 in the stacked DNA duplexes, but bridging between adjacent stacked duplexes to form the junction cross-overs (panel created with Bobscript (*89*)).

junction is generated by 2-fold symmetry applied to the two unique strands (one outside noncrossover and one inside crossover strand) of the asymmetric unit. This four-stranded complex (Figure 3.2a) has four arms—two longer six base pair arms (from the C_1 to G_6 of one outside strand paired with G_{10} to C_5 of one cross-over strand) and two short four base pair arms (pairing the C_7 to G_{10} of the same outside strand to G_4 to C_1 of an alternative cross-over strand. Each six base-pair duplex domain is stacked over a four base-pair duplex to form the near continuous ten base pair helices of the stacked-X junction (Figure 3.2b), with the cross-over occurring between nucleotides G_6 and C_7 of the inside cross-over strands (Figure 3.2c).

In nearly all respects, this GCC-junction is identical to that of the reference d(CCGGTACCGG) structure (ACC-junction, Figure 3.3, RMSD = 0.43 Å for all common DNA atoms). The stacked double-helical arms are related by an interduplex angle of 40.0° in the GCC-junction as compared to the 41.4° angle in ACC (Table 3.2). Furthermore, there is no sliding of the duplexes across the junction cross-over in GCC ($J_{slide} \approx$ 0 Å) relative to the ACC-junction. The exposure of the major and minor grooves are also nearly identical between the sequences, with J_{roll} values = 159° in both junction structures.

One interesting difference between the GCC-junction and other DNA junctions to date is that the width of the minor groove (as measured by the phosphorus to phosphorus distance, Table 3.3) is wider at the $C_5{\cdot}G_6/G_6{\cdot}C_5$ and $G_6{\cdot}C_5/C_7{\cdot}G_4$ dinucleotide steps (18.5 Å and 19.5 Å respectively compared to the averages of 17.5 Å (0.26Å S.D.) and 18.8 (0.35 Å S.D.) for ACN type junctions). In addition, we see that the $G_6{\cdot}C_5/C_7{\cdot}G_4$ dinucleotide steps are underwound by 3° as compared to the

Table 3.2. Comparison of global conformational geometry values for d(CCGGCGCCGG) (GCC) and d(CCAGTACBr^{5}UGG) (ACBr^{5}U) with previous junction structures of d(CCGGTACCGG) (ACC, (*47*)) and d(CCGGTACm^{5}CGG) (ACm^{5}C, (*48*)).

Junction[a]	Interduplex Angle	J_{slide}[b]	J_{roll}[c]
ACC	41.4°	_	159.6°
ACm^{5}C	41.3°	~1.7Å per duplex	170.4°
GCC	40.0°	0 Å	159.8°
ACBr^{5}U	38.8°	~1.2 Å per duplex	164.0°

[a] Trinucleotide core region corresponding to N_6-N_7-N_8 positions of the general sequence d(CC$N_3N_4N_5N_6N_7N_8$GG)

[b] Relative to d(CCGGTACCGG) junction, as estimated by the slide seen for one set of duplexes when the opposing stacked duplexes are superimposed with those of the ACC-junction. Values are for the duplexes on both sides of the junction, sliding symmetrically. If one set of stacked duplexes are superimposed, then the opposing stacked double-helices would show a slide of twice the value reported in this table (see, for example, Figure 3.7).

[c] Angle between vectors extended from the center of the junction to points that bisect the phosphates on the two outside strands that complement the nucleotides at the gap of the junctions (see for example Figure 3.4.)

Table 3.3. Minor groove widths (Å) measured as the phosphorus P to P distance for the phosphodiester linkage between dinucleotides. The widths are compared for the junction structures of d(CCGGTACCGG) (ACC-J), d(CCGGCGCCGG) (GCC-J), d(CCGGTACm^{5}CGG) (ACm^{5}C-J), and d(CCAGTACBr^{5}UGG) (ACBr^{5}U-J), and the B-DNA duplex structures of d(CCAGTACTGG) in the monoclinic crystal form (ACT-B *C2*) and the hexagonal form (ACT-B *P6$_1$22*), and d(CCGGCGCCGG) (GCC-B).

Dinucleotide (reference)	ACC-J *(12)*	GCC-J (this work)	ACm^{5}C-J *(15)*	ACBr^{5}U-J (this work)	ACT-B *C2 (32)*	ACT-B *P6$_1$22* (this work)	GCC-B *(31)*
$C_1{\cdot}G_{10}/C_2{\cdot}G_9$	16.9	16.9	16.3	17.2	18.1	18.3	18.1
$C_2{\cdot}G_9/Pu_3{\cdot}Py_8$	17.4	17.2	17.1	17.0	16.7	17.0	17.4
$Pu_3{\cdot}Py_8/G_4{\cdot}C_7$	17.2	17.5	17.6	17.4	18.0	18.3	17.8
$G_4{\cdot}C_7/Py_5{\cdot}Pu_6$	17.9	17.9	18.1	18.1	17.8	18.5	18.3
$Py_5{\cdot}Pu_6/Pu_6{\cdot}Py_5$	17.8	18.5	17.3	17.4	17.0	17.2	18.0
$Pu_6{\cdot}Py_5/C_7{\cdot}G_4$	**19.1**	**19.5**	**18.4**	**18.8**	17.8	18.5	18.2
$C_7{\cdot}G_4/Py_8{\cdot}Pu_3$	16.8	16.9	17.0	17.9	18.0	18.3	17.9
$Py_8{\cdot}Pu_3/G_9{\cdot}C_2$	16.8	17.1	17.2	16.7	16.7	17.0	18.1
$G_9{\cdot}C_2/G_{10}{\cdot}C_1$	16.8	16.4	16.8	17.8	18.1	18.3	17.1
Average (S.D.)[a]	17.41 (0.76)	17.54 (0.96)	17.31 (0.65)	17.59 (0.64)	17.58 (0.60)	17.93 (0.66)	17.88 (0.39)

Values for the dinucleotide $Pu_6{\cdot}Py_5/C_7{\cdot}G_4$ in bold represent the point where the phosphodeoxyribose backbone departs from the duplex to form the junction cross-over in the junction structures.

[a] $S.D = \sqrt{((\sum (x - \langle x \rangle)^2) / (n - 1))}$ where n = number of observations, x = calculated minor groove width for each dinucleotide, and $\langle x \rangle$ is the average of all observations.

Table 3.4. Comparison of helical twist. Measured using a global axis in CURVES 5.1 (28), for the junction structures of d(CCGGTACCGG) (ACC-J), d(CCGGCGCCGG) (GCC-J), d(CCGGTACm^{5}CGG) (ACm^{5}C-J), and d(CCAGTACbr^{5}UGG) (ACbr^{5}U-J), and the B-DNA duplex structures of d(CCAGTACTGG) in the monoclinic crystal form (ACT-B *C2*) and the hexagonal form (ACT-B *P6$_1$22*), and d(CCGGCGCCGG) (GCC-B).

Dinucleotide (reference)	ACC-J *(12)*	GCC-J (this work)	ACm^{5}C-J *(15)*	ACbr^{5}U-J (this work)	ACT-B *C2* *(32)*	ACT-B *P6$_1$22* (this work)	GCC-B *(31)*
$C_1 \cdot G_{10}/C_2 \cdot G_9$	38.0°	38.0°	37.4°	45.1°	30.1°	26.5°	36.3°
$C_2 \cdot G_9/Pu_3 \cdot Py_8$	41.1°	36.7°	38.5°	43.9°	50.5°	51.8°	37.3°
$Pu_3 \cdot Py_8/G_4 \cdot C_7$	39.0°	41.8°	38.1°	34.6°	22.1°	26.9°	36.4°
$G_4 \cdot C_7/Py_5 \cdot Pu_6$	34.3°	33.4°	37.0°	28.8°	34.9°	37.2°	35.5°
$Py_5 \cdot Pu_6/Pu_6 \cdot Py_5$	32.1°	33.7°	41.5°	37.9°	47.8°	32.7°	30.1°
$Pu_6 \cdot Py_5/C_7 \cdot G_4$	**33.6°**	**30.6°**	**26.5°**	**32.6°**	34.9°	37.2°	36.1°
$C_7 \cdot G_4/Py_8 \cdot Pu_3$	40.9°	43.0°	40.0°	32.9°	22.1°	26.9°	42.6°
$Py_8 \cdot Pu_3/G_9 \cdot C_2$	39.9°	37.4°	38.6°	51.7°	50.5°	51.8°	24.8°
$G_9 \cdot C_2/G_{10} \cdot C_1$	41.3°	42.0°	38.9°	30.9°	30.1°	26.5°	44.0°
Average (S.D.)[a]	37.80° (3.55)	37.40° (4.31)	37.39° (4.30)	37.6° (7.69)	35.89° (11.28)	35.28° (10.33)	35.90° (5.81)

Values for the dinucleotide $Pu_6 \cdot Py_5/C_7 \cdot G_4$ in bold represent the point where the phosphodeoxyribose backbone departs from the duplex to form the junction cross-over in the junction structures.

[a] $S.D = \sqrt{((\sum (x - \langle x \rangle)^2) / (n - 1))}$ where n = number of observations, x = calculated helical twist for each dinucleotide, and $\langle x \rangle$ is the average of all observations.

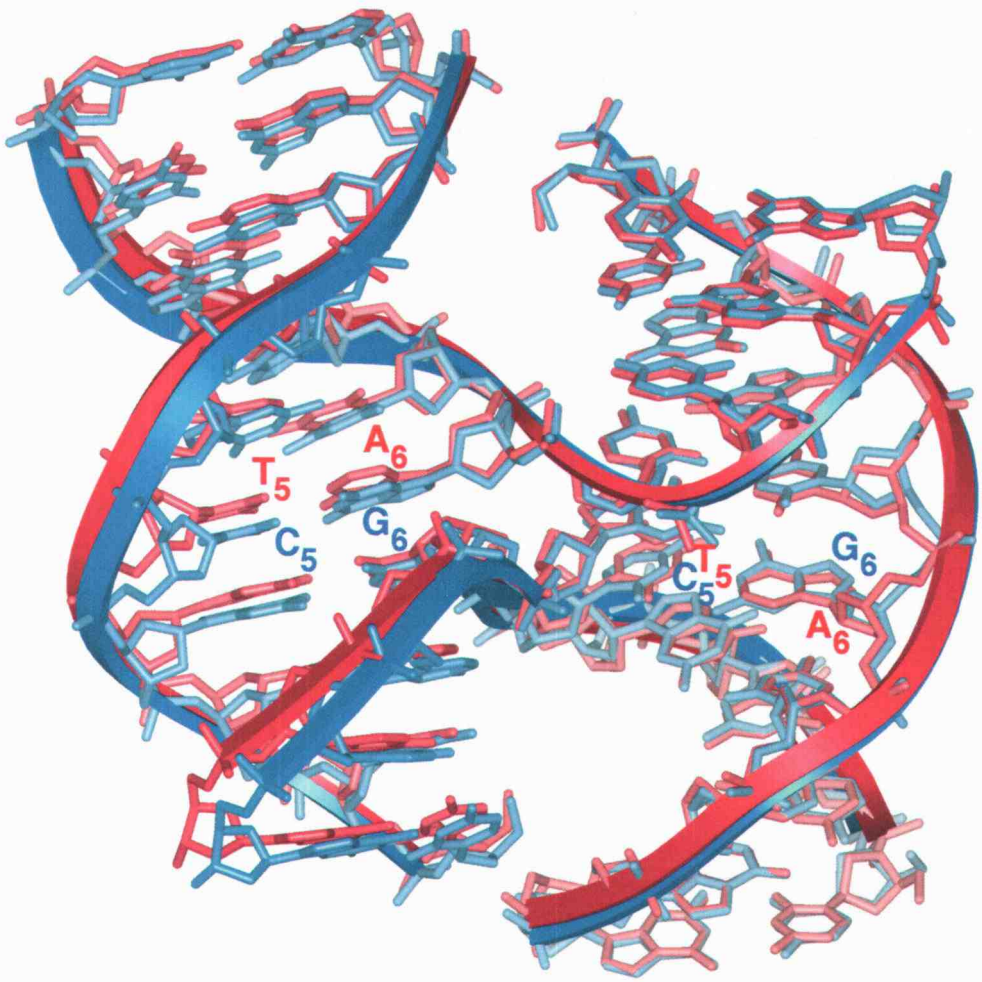

Figure 3.3. Superposition of d(CCGGCGCCGG) and d(CCGGTACCGG) junctions. The structures of the two junctions are superimposed using the common atoms of the stacked arms on the right of the junction in this figure. In superposition of the overall structure, RMSD = 0.426 Å for all common DNA atoms.

analogous $A_6 \cdot T_5/C_7 \cdot G_4$ dinucleotide step in the junctions (2.4), and significantly underwound (by 5.5°) compared to the comparable steps in the B-DNA duplex of the same sequence. In fact, the duplexes in all of the junction structures are underwound by an average of ~2° (37.54°, S.D. = 4.99°) compared to their B-DNA counterparts (35.69°, S.D. 9.08°) (Table 3.4), and all of the dinucleotides that span the junction in these structures are underwound by ~8° relative to the identical or analogous sequences as standard B-DNA (Table 3.4).

Not surprisingly, many of the structural details seen in the GCC-junction are similar to those in the ACC-junction. For example, the direct hydrogen bonding interaction between the amino N4-nitrogen at the major groove surface of the cytosine C_8 base to the phosphate oxygen of C_7 that was first identified as helping to stabilize the ACC-junction (Figure 3.4a) is also seen in the current GCC-junction. The associated solvent mediated hydrogen-bonding interaction that bridges the keto-oxygen of the complementary G_3 nucleotide to the cross-over phosphate oxygen at A_6 in the ACC-junction (S1 in Figure 3.4a), however, is now a network in the current GCC-junction (Figure 3.4b), similar to that seen previously in the ACm^5C-junction structure. Thus, the commonality of the direct C_8 amino to C_7 phosphate hydrogen bond and the similarities of the junction geometries across these structures suggest a correlation between this particular interaction at the atomic level and the general conformational features for this class of structures. However, the solvent that was assigned as a sodium ion in the center of the ACC-junction is not present in the current structure, suggesting that this solvent interaction is not crucial to the overall geometry of this structural class.

Figure 3.4. Atomic interactions within the trinucleotide core of the d(CCGGCGCCGG) junction viewed perpendicular to the junction dyad axis. Hydrogen bonds are shown as dotted lines and solvent molecules as spheres. The nucleotides in the junction cross-over are labeled, along with the major (M) and minor (m) grooves. The J_{roll} angles are shown as the angle between the vectors extending from the center of the junction to points that bisect the phosphates of the outside strands that complement the nucleotides of the junction gap in the inside strands. *a*, The reference d(CCGGTACCGG) junction. The important interactions had been identified as the direct hydrogen bond between the amino N4-nitrogen of C_8 to the C_7 phosphate oxygen O2P, a water mediated hydrogen bond between the keto O6-oxygen of G_3 and G_6 phosphate oxygen O1P, and solvent (sodium ion) coordinated to the O2P phosphate oxygens of the G_6 nucleotides (*47*). *b*, The d(CCGGCGCCGG) junction. This structure shows the direct hydrogen bond between nucleotide C_8 and the crossing phosphate, and a set of interactions mediated by solvent (S1 to S3) from the major groove of the G_3 nucleotide to the opposing crossing phosphates. The solvent are labeled as follows: S1 is a water molecule that is hydrogen bonded to the Pu3 nucleotide that complements the Py8 of the junction trinucleotide core; S2 is a water that is hydrogen bonded directly to the Py8 position of the junction core; S3 is the solvent (proposed as the Na+ ion displaced from the center of the ACC-junction) that sits between the phosphates of the junction cross-over strands.

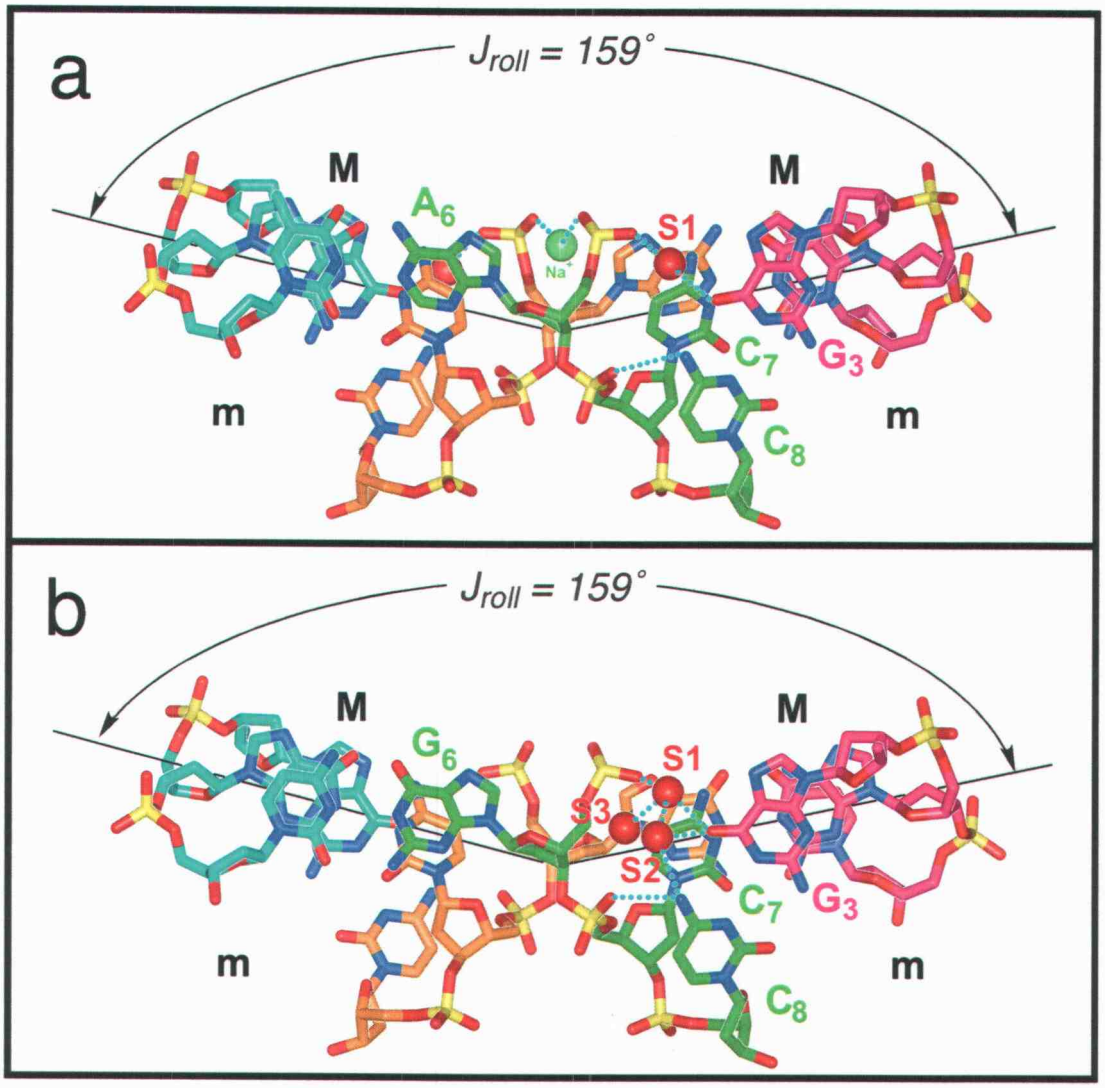
a
J_{roll} = 159°
M
M
A_6
S1
Na+
C_7
G_3
m
m
C_8
b
J_{roll} = 159°
M
M
G_6
S1
S3
S2
C_7
G_3
m
m
C_8

Figure 3.4

3.4.2 Structure of d(CCAGTACTGG) as B-DNA duplexes. Our experience with the d(CCGGCGCCGG) sequence indicated that, although originally reported to be a B-DNA duplex (*90*), we should study the structure of the sequence d(CCAGTACTGG) in our crystallization solutions. However, even under conditions that have yielded junctions, this sequence crystallized as double-stranded B-DNA, although in a different crystal form than was previously reported. The space group of the current crystals is hexagonal $P6_122$, as compared to the earlier monoclinic *C2* form (*90*). In both crystal forms, the asymmetric unit is one strand of the decanucleotide, with the second strand generated by the crystal symmetry. This precludes the sequence in either crystal form from being a Holliday junction since at least two unique strands are required to form the crossed four-stranded complex. The current and previous B-DNA structures of d(CCAGTACTGG) are very similar, RMSD = 1.03Å for all nonhydrogen atoms, and similar minor groove widths and helical twists at each dinucleotide step along the double-helix (Tables 3.3 & 3.4). The noticeable differences result from differences in the duplex-duplex interactions of the two crystal lattices. For example, the duplexes in the monoclinic crystal form have the nonstandard BII conformations (as defined by a ζ dihedral angle $\approx 180°$) at the phosphoribose spanning the C_2pA_3, G_4pT_5, and T_8pG_9 dinucleotide steps along the DNA chain. In contrast, only the C_2pA_3 dinucleotide step shows this BII conformation in the current hexagonal crystal form.

3.4.3 Structure of d(CCAGTACbr^{5}UGG) as a Holliday Junction. The sequence d(CCAGTACbr^{5}UGG) was initially designed as a means to introduce a bromine heavy

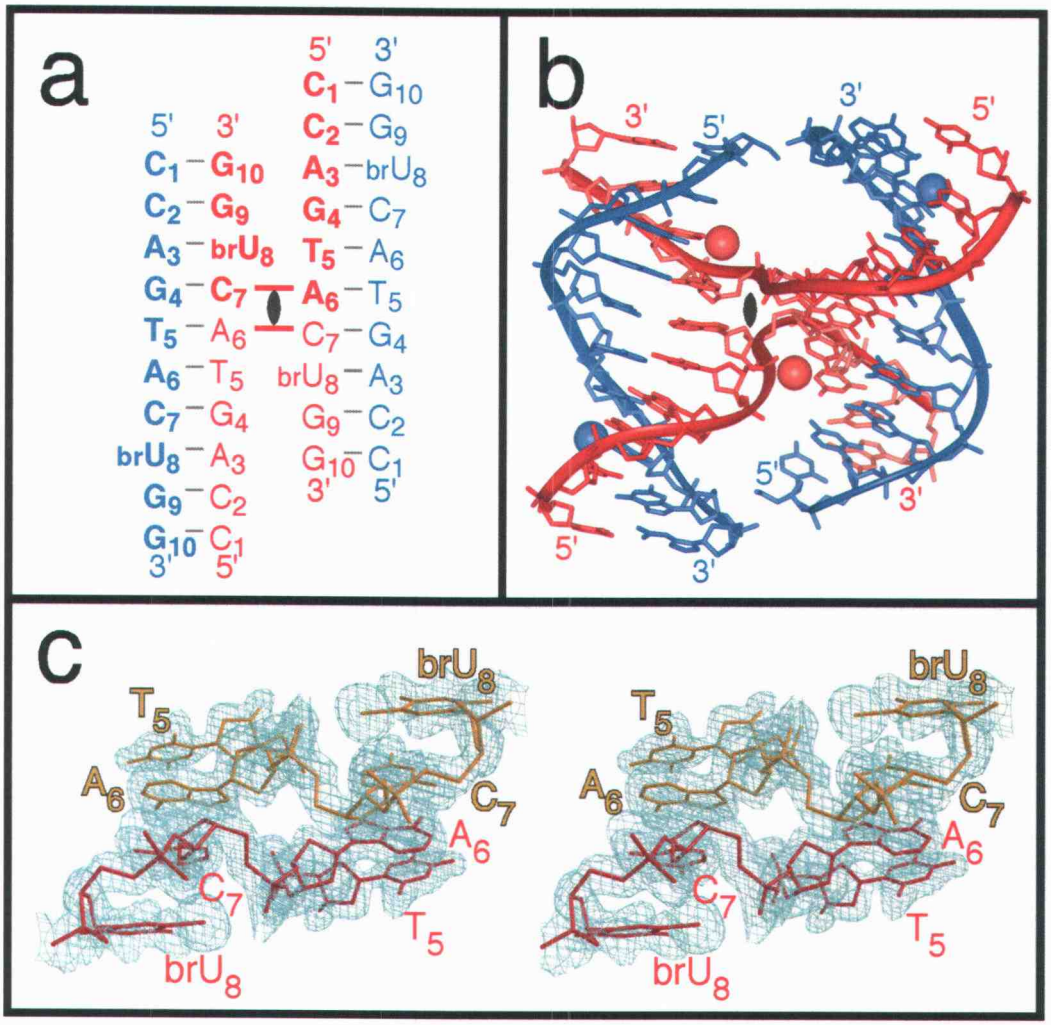

Figure 3.5. Crystal structure of d(CCAGTACbr^{5}UGG) as a DNA Holliday Junction. *a*, Sequence topology of the d(CCAGTACbr^{5}UGG) junction. The four-stranded antiparallel stacked-X Holliday junction is generated by applying the crystallographic two-fold symmetry to the two unique strands (bold). Strands are numbered from 1 to 10 in the 5' to 3' direction, with the inside crossing strands colored red and the outside noncrossing strands colored blue. *b*, The atomic structure of d(CCAGTACbr^{5}UGG). Chemical bonds in the structure are rendered as sticks, the phosphodeoxyribose backbone rendered as a solid ribbon, and bromines rendered as spheres (colors and strand designations are as in *a*, figure rendered with InsightII from MSI/Biosym, Inc.). *c*, Electron density map. The $2F_o$-F_c map (contoured at 1σ) shows the discontinuity in electron density between nucleotides G_6 and C_7 in the stacked DNA duplexes, but bridging between adjacent stacked duplexes to form the junction cross-overs (figure in panel created with Bobscript (*89*)).

atom at the nucleotide-8 position of d(CCAGTACTGG) to help solve this structure in the hexagonal crystal form. Although to date d(CCAGTACTGG) has only crystallized as resolved B-DNA duplexes, this brominated analogue is presented here as a four-stranded Holliday junction (Figure 3.5). The resulting structure of d(CCAGTACbr^{5}UGG) (ACbr^{5}U-junction) is most closely related to that of the methylated d(CCGGTACm^{5}CGG) (ACm^{5}C-junction) structure. Although the RMSD of 1.57 Å for common DNA atoms between the ACbr^{5}U- and ACm^{5}C-junctions is relatively high, both show similar structural perturbations relative to the ACC-type junctions in terms of the J_{roll} and J_{slide} (Table 3.2) around and along the axes of the stacked duplexes. The J_{slide} of ~3.4 Å and J_{roll} of 170.4° rendered the ACm^{5}C-junction more symmetric in terms of the relative positions of the ends and accessibility of the major and minor grooves of the stacked duplex arms. The current ACbr^{5}U-junction is intermediate between this distorted ACm^{5}C and the parent ACC-junctions. The interduplex angle of the ACbr^{5}U-junction (38.8°) is the most shallow of any DNA-only junction (but not as shallow as the psoralenated junctions (*69*), where this angle varies from 36.4°-37.2°).

The detailed interactions at the ACbr^{5}U-junction cross-over (Figure 3.6b) show some similarities to the ACm^{5}C-structure (Figure 3.6a). In both cases, the direct hydrogen bond from the Py_8 nucleotide to the crossing phosphate of C_7 has been disrupted. In the ACm^{5}C-structure, this hydrogen bond from the amino N4 of cytosine C_8 to O2P oxygen of the phosphate was disrupted by addition of the methyl group at the C_8 nucleotide, but the two atoms are now bridged by a solvent molecule (S2 in Figure 3.6a). In replacing the cytosine C_8 with the substituted uracil base in the

ACbr^{5}U-junction, the hydrogen bond donating amino group of the cytosine has been replaced by a hydrogen bond accepting keto-oxygen. This precludes the formation of the direct hydrogen bonding interaction, regardless of the substituent group added to the C5 carbon of the base. A solvent (S2, Figure 3.6b) remains hydrogen bonded to this keto-oxygen, but does not span the gap to the phosphate oxygen of the junction cross-over.

The ACC-, GCC-, and ACm^{5}C-junctions all show a network of one or more solvent molecules that help bridge the complementary G_3 nucleotide base of the $C_8{\cdot}G_3$ base pair to the crossing phosphate of A_6. The analogous A_3 nucleotide of the br$^5U_8$-A_3 base pair shows no such network of bridging solvent molecules. We had previously suggested that one of the bridging solvent molecules in the ACm^{5}C network (S3, Figure 3.6a) was the sodium ion in the center of the ACC-junction, but displaced by the collapse of the junction core. The loss of this series of solvent between A_3 and the A_6 phosphate would suggest that the ion could have buried itself back into the center of the ACbr^{5}U-junction, and the electron density in the center of the junction supports that assertion. A solvent molecule was seen to be sandwiched between, and located very close to (< 2.5Å), the O1P phosphate oxygen atoms of the crossing A_6 nucleotides (Na^+, Figure 3.6b). It is interesting, however, that the volume estimated for this binding site corresponds to a sphere with radius 1.25 Å, which is now larger than that of the ACm^{5}C junction (at 0.9 Å radius) and even the unmodified ACC junction (at 1.13 Å radius) (*91*). Finally, a new solvent (S8, Figure 3.6b) is seen spanning the O1P oxygens of the C_7 cytosines in both crossing strands.

Figure 3.6. Atomic interactions within the trinucleotide core of the d(CCAGTACbr^{5}UGG) junction viewed perpendicular to the junction dyad axis. Hydrogen bonds are shown as dotted lines and solvent molecules as spheres. The nucleotides in the junction cross-over are labeled, along with the major (M) and minor (m) grooves. The J_{roll} angles are shown as the angle between the vectors extending from the center of the junction to points that bisect the phosphates of the outside strands that complement the nucleotides of the junction gap in the inside strands. *a*, The reference d(CCGGTACm^{5}CGG) junction. The important interactions had previously been identified as the network of solvent (S1 to S4) mediated hydrogen bonds from the C_8·G_3 base pair to the C_7 O2P and G_6 O1P phosphate oxygens (*48*). *b*, The d(CCAGTACbr^{5}UGG) junction. The solvent (S2) at the br$^5U_8$·A_3 base pair is no longer continuous with the solvent at the phosphates of nucleotide C_7. A sodium ion (Na^+) is seen sandwiched between the phosphate oxygens of nucleotide A_6, and a solvent (S9) bridges the N7 nitrogen of A_6 to the phosphate of the A_6 nucleotide of the opposing cross-over strand. The labels for solvent S1-S3 occupy positions similar to those described for Figure 3.4. S4 is a water that bridges S1 to the phosphate of the cross-over strand of the ACm^{5}C-junction. The solvent molecules labeled S5-S9 are new to the core of the ACbr^{5}U-junction.

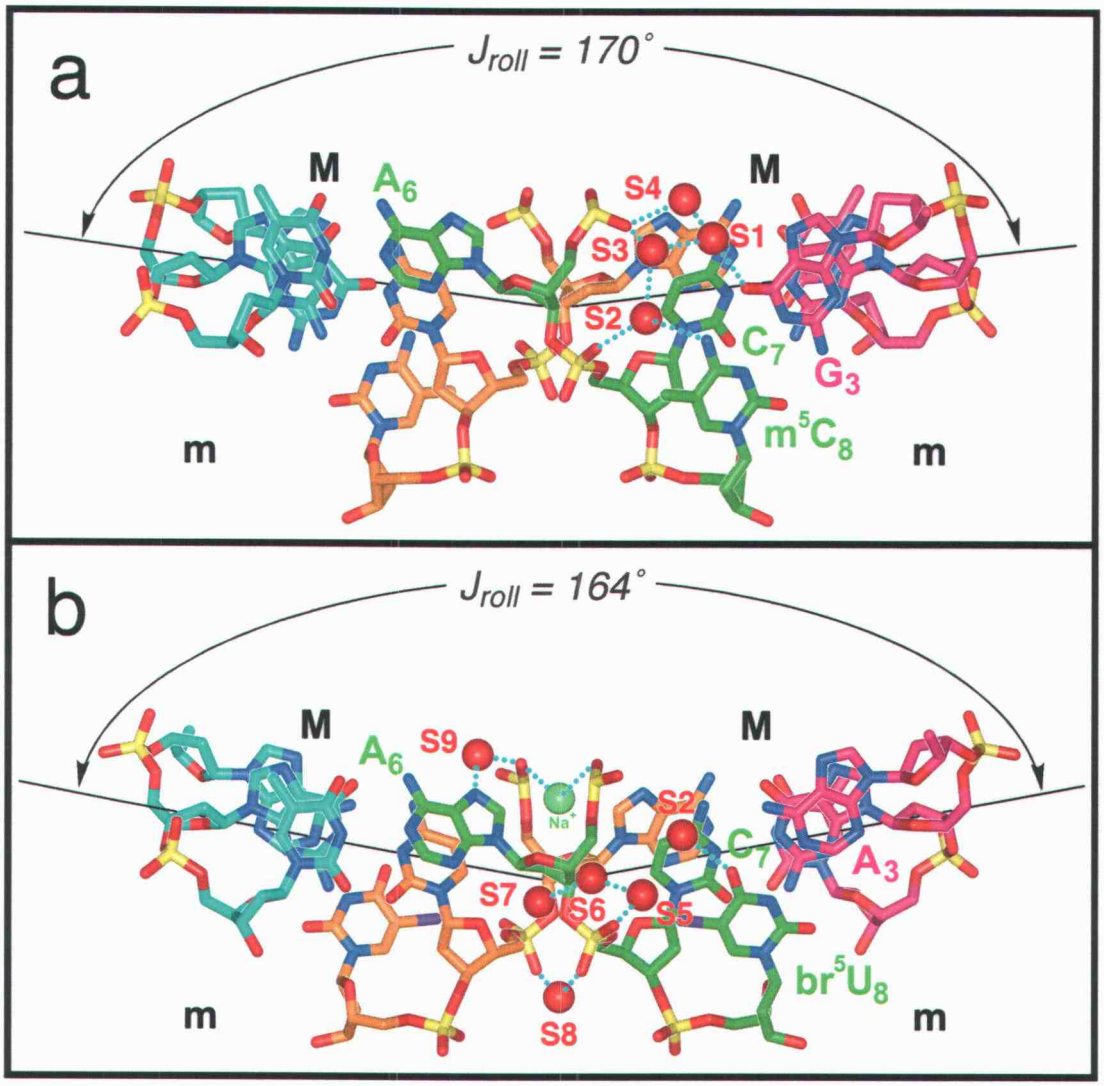
a
J_{roll} = 170°
M
A_6
S4
S3
S1
S2
M
C_7
G_3
m^5C_8
m
m
b
J_{roll} = 164°
M
A_6
S9
Na^+
S2
M
C_7
A_3
S7
S6
S5
br^5U_8
m
S8
m

Figure 3.6

Thus, in the current structure of the ACbr^{5}U-junction, there are no direct or solvent mediated hydrogen bonds between the major groove surface of the br^{5}U$_8$·A$_3$ base pair and the phosphate oxygens of the junction cross-over, even though many of the solvent molecules remain hydrogen bonded to the bases at the major groove surface. There is, however, an interaction at the other end of the Pu$_6$C$_7$Py$_8$ trinucleotide motif, with the N7 nitrogen of the A$_6$ nucleotide base now bridged through a solvent molecule (S9, Figure 3.6b) to the O2P oxygen of the A$_6$ nucleotide of the opposite crossing strand. This would seem to compensate for the loss of interactions between the stacked arms and the core of the junction that appears to be critical to the stability of the junction in this system.

3.5 Discussion

In this study, we have determined the single crystal structures of the sequences d(CCGGCGCCGG) and d(CCAGTACbr^{5}UGG) as four-stranded Holliday junctions in the antiparallel stacked-X conformation, and the structure of d(CCAGTACTGG) as resolved B-DNA duplexes. The two new junctions are the first such structures from sequences that do not contain the previously identified ACC-core trinucleotide sequence at the N$_6$N$_7$N$_8$ positions of the general sequence d(CCN$_3$N$_4$N$_5$*N$_6$N$_7$N$_8$*GG). Thus, we have now expanded this central sequence motif for the crystallization of junctions to Pu$_6$C$_7$Py$_8$, where Pu is either a guanine or adenine purine-containing nucleotide and Py is a cytosine, 5-methylcytosine, or 5-bromouracil pyrimidine

nucleotide, but not a thymine nucleotide. Our previous study with the methylated d(CCGGTACm^{5}CGG) sequence (*48*) showed that perturbations to the overall geometry of the Holliday junction can include a slide in the stacked duplexes across the junction (J_{slide}) and a roll of the duplexes along their helix axes (J_{roll}). With the current study, we see that sequence and substituent effects define distinct classes of stacked-X junctions: the ACC-type structure (the original junction structure), and ACm^{5}C class (which shows significant J_{slide} and J_{roll} relative to ACC). In addition, these effects are seen to perturb the interduplex angle relating the orientation of the stacked duplex arms across the junction.

A comparison of the GCC- and ACC-junctions show that the Pu_6 position of the $Pu_6C_7Py_8$ motif does not dramatically perturb the conformation of the Holliday junction. The only noticeable effect on the general geometry of the conformation in replacing the adenine with a guanine base at this position is to make the interduplex angle shallower (from 41.4° in ACC to 40° in the GCC-structure). It is not surprising that the two structures are so similar. The nucleotide at the N_6 position is the initial point of departure of the DNA strand that crosses-over to form the junction. However, the purine base at this position remains paired with the complementary strand of the stacked duplex arms at this point and is not directly involved in the interactions of the cross-over.

In contrast to Pu_6, the base and the substituent at the Py_8 position of the now expanded $Pu_6C_7Py_8$ trinucleotide motif has a significant affect on the conformation of the junction. The sliding (J_{slide}) and rolling of the duplexes along and around their respective helix axes (J_{roll}) previously defined in the d(CCGGTACm^{5}CGG) structure

are also observed here with the br^{5}U deoxynucleotide at the Py_8 position, but are less exaggerated. Thus, the J_{slide} distortions are primarily induced by steric interactions between the methyl or bromo-substituent and the deoxyphosphoribose backbone of the cross-over strands. In the current structure, the bromine is in direct contact with one of these phosphates. Thus, we can think of the junction as the pivot point not only to twist the stacked duplexes around (in defining the interduplex angle), but to push against and thus slide these columns of base pairs relative to each other.

The rolling the stacked duplexes about their respective helical axes (J_{roll}) has the effect of opening the major groove surface of the junction. What we see by comparison of these structures is that the direct interaction between the C_8 cytosine and the cross-over phosphates fixes the orientation of the opposing stacked duplexes of the ACC- and GCC-junctions. This is not surprising, since hydrogen bonding interactions involving nucleotide bases have well defined geometric constraints in terms of distances and angles between hydrogen bond donors and acceptors. Thus, such interactions predetermine the relationships between the major groove surface of the arms and the core phosphates of the junction. The addition of intervening solvent molecules, as in the ACm^{5}C and ACbr^{5}U junctions, can be thought of as "lubricating" the interaction surfaces, relieving the constraints imposed by the ACC- and GCC-interactions and thereby allowing more variations within the geometries of the junctions, allowing variability in the interduplex angle (as seen in the more shallow angle of the ACbr^{5}U-junction), as well as perturbations to J_{slide} and J_{roll} (evident in both the substituted structures). This does not, however, suggest that the structure of the junction itself is more flexible. A comparison of all the junction crystal structures

show that the nucleotides directly at and those paired to nucleotides at the cross-over have lower average temperature or B-factors than those nucleotides that are distant from the junction core.

The other distortion seen in the d(CCAGTACbr^{5}UGG) structure, relative to all other junctions, is the shallower interduplex angle. This angle, which is approximately 40° (ranging from 40° to 41.4°) for all PuCC type junction structures, becomes 38.4° with the ACbr^{5}U structure. The distortions to J_{slide} and J_{roll} means that the best comparison is between the two structures with substituents at the Py_8 position (ACm^{5}C and ACBr^{5}U), which places these two structures in their own class. In this comparison, we see that the br^{5}U base induces a 3° rotation about the junction relative to the ACm^{5}C-junction, which results from substitution of an A_3·br$^5U_8$ for the G_3·m$^5C_8$ base pair. The rotation of the interduplex angle is thus not simply a substituent effect, but more likely is associated with the A·br^{5}U base pair at this position.

Why does sequence have such an effect on the structure of the junctions? The answer to this question may be seen in the interactions at the ACbr^{5}U trinucleotide that compensate again for the loss of interactions between the major groove of the Py_8·Pu_3 base pair to junction phosphates seen in all of the other structures of the complex. There were no direct or even solvent mediated bridges at this surface. However, there is a compensating solvent bridge from A_6 across to the phosphate of the opposite crossing strand. We can now imagine this as being a factor that both stabilizes and defines the interduplex angle of the ACbr^{5}U junction.

Alternatively or in addition, we had previously proposed that the interduplex angle is defined by a short phosphate-phosphate interaction that is distant (three base pairs away) from the actual junction cross-over (*69*). We again see this very short oxygen to oxygen distance (3.69 Å) and, in the GCC-junction structure, observe two solvent molecules that are ~2.4 Å from one set of the phosphate oxygen atoms. If these two closely interacting solvents are assigned as sodium ions rather than waters, then the net charge at the phosphate of the N_{10} nucleotide would be +1, which would provide a rational for this short interarm phosphate-phosphate distance. This short phosphate-phosphate distance remains even in the ACm^{5}C (at 3.5 Å) and ACbr^{5}U (at 3.6 Å) structures. In both junctions, the direct amino nitrogen to the cross-over phosphate of C_7 in the ACC-junction has been disrupted (by the methyl group of ACm^{5}C and, additionally, by the replacement of the amino nitrogen with a keto oxygen in ACbr^{5}U). This supports a model that the components of the helical structure of the duplex arms (twist, rise, tilt, etc.) help to define this angle. For example, in the current ACPy structures, we see that both the average helical twist overall and across the unique trinucleotide steps at C_2-Pu_3-G_4 and C_7-Py_8-G_9 are significantly shallower in the ACbr^{5}U-junction as compared to the ACm^{5}C-junction. The orientation of the phosphates in close approach would in fact result in a shallower interduplex angle when the double-helical arms are unwound in this class of shifted and rolled junctions (Figure 3.7).

In the original ACC-junction, we had observed a solvent molecule (which was assigned as a sodium cation) imbedded in a central cavity formed by the phosphates of the crossing strands. We now observe this same solvent molecule in the center of the

Figure 3.7. Superposition of the d(CCAGTACbr^{5}UGG) junction (blue) on the d(CCGGTACCGG) junction (red). Atoms are rendered as stick models with the phosphodiester backbone traced by ribbons. One set of stacked duplex arms (yellow cylinder) from each structure was used for superposition. *a*, View along the junctions with the nonsuperimposed stacked duplex arms in front. The ACbr^{5}U-junction shows ~ 2.3 Å J_{slide} along the helix axis. *b*, View into the junctions with the superimposed stacked arms to the right. The non-superimposed stacked duplexes show the J_{slide} and J_{roll} effects imposed by the br^{5}U·A base pair of the ACbr^{5}U core trinucleotide on the overall geometry of the Holliday junction.

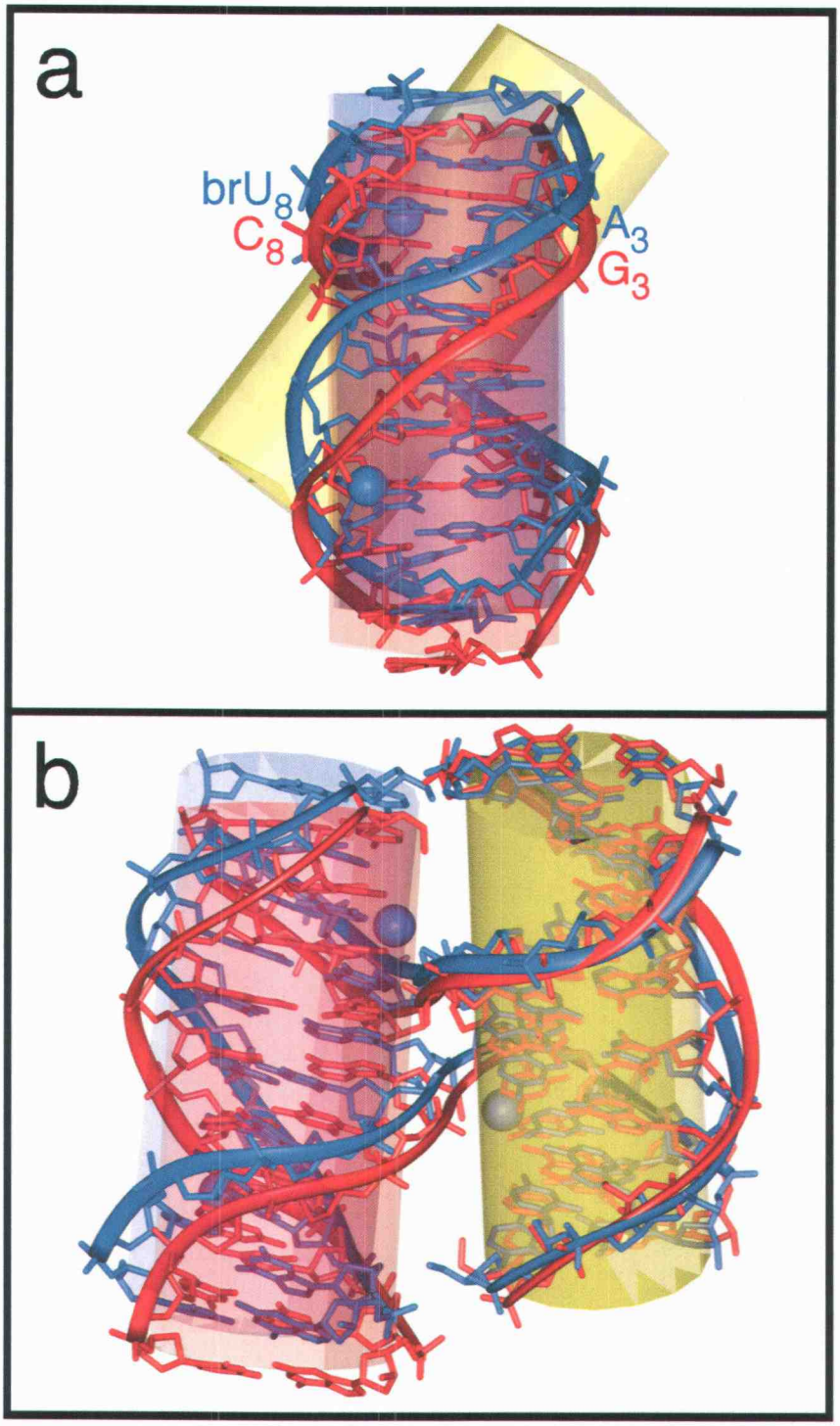

Figure 3.7

ACbr^{5}U-junction, but not the ACm^{5}C or GCC structures. We can thus conclude that neither the J_{roll} nor J_{slide} distortions alone account for the accessibility of the junction to the intruding ion. Why does the ACbr^{5}U trinucleotide favor formation of crystals of the junction while ACT does not? The difference is obviously the bromine versus the methyl group at the C5 carbon of the Py_8 nucleotide base. Since the two substituents are nearly identical in size (both approximately 2 Å radius) and hydrophobicity, the difference in behavior likely arises from differences in their electronic properties—bromine is electron withdrawing, making it slightly electron rich, while methyl groups are electron donating to an aromatic ring system and thus are slightly electron poor. One can suggest then that the electrostatic interactions between the substituent and the close phosphate group from the junction cross-over might be the distinguishing factor; however, this would predict that the methyl rather than bromo-group should stabilize the junction, which is the opposite of what is observed here. We suggest, therefore, that the electrostatic effect is indirect, affecting for example the base pair stacking. The helical twist at the C·G/br^{5}U·A dinucleotide steps in the ACbr^{5}U-junction are 8° to 10° larger than the comparable C·G/T·A steps of the B-DNA structures (Table 3.4). This is true regardless of whether the base pairs are at the junction core (C_7·G_4/br$^5U_8$·A_3) or not (A_3·br$^5U_8$/G_4·C_7), indicating again that the effects of nucleotide sequence and substituents on the structure of the stacked duplex arms of the junction mirror those seen in standard B-DNA duplexes (*69*). This difference in the helical twist at the junction core would dramatically affect the position of the substituent groups relative to the crossing phosphates. In the ACbr^{5}U-junction, the bromine sits against the phosphate oxygen and above the deoxyribose

ring of the preceding C_7 nucleotide. A rotation of this base pair by –10° in the stack (as would be expected in a theoretical ACT-junction) would place the analogous methyl group of an ACT junction core nearly equidistant from the two crossing phosphates, and in contact with neither. One possible explanation therefore for why the thymine does not facilitate crystallization of the junction is that the effects of the methyl group on the electron distribution of the thymine base results in a base-stacking geometry that does not force the substituent to push against the phosphates of the junction and, consequently, would not induce a J_{slide} of the stacked helices. The result may then be that the solvent bridged interaction between adenine A_6 (across the gap of the inside strands of the complex) could not be formed. Other factors, including solvent effects, may contribute to or may be more important than the electronic effect on base stacking described here.

Thus, we see from this series of structures, that the overall geometry of the junctions varies according to the nucleotides within the trinucleotide core sequence motif. This is consistent with the understanding that the nucleotides immediately flanking the junction determines how helical arms pair to form the stacked duplexes (*40, 66*), and the rates of conformational isomerization to switch these pairing partners (*25*). The current structures therefore provide an atomic level view of how these sequence variations would affect the stability of the four-way stacked-X junction, which also defines the ability of such junctions to undergo conformation dependent dynamic processes, including branch migration and protein binding.

3.6 Acknowledgements

We thank Drs. Z. Wood and R. Faber for helpful discussion during data refinement, and Dr. J. Watson for critical reading of the manuscript. This work was funded by grants from the National Institutes of Health (R1GM62957A) and the National Science Foundation (MCB0090615). The X-ray diffraction facilities were supported by the Proteins and Nucleic Acids Facility Core of the Environmental Health Sciences Center at OSU (NIEHS ES00210) and a grant from the Murdock Charitable Trust. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under Contract No. W-31-109-Eng-38. Use of BioCARS Sector 14 was supported by the National Institutes of Health, National Center for Research Resources, under grant number RR07707.

Chapter 4

Influence of Minor Groove Substituents on the Structure of DNA Holliday Junctions

Franklin A. Hays, Zebulon J.R. Jones and P. Shing Ho

Published in *Biochemistry*,

American Chemical Society, Washington, D.C., USA

2004, **43** (30), 9813-9822

4.1 Summary

The inosine (I) containing sequence d(CCIGT*ACm*^{5}*C*GG) is shown to crystallize as a four-stranded DNA junction. This structure is nearly identical to the antiparallel junction formed by the parent d(CCGGT*ACm*^{5}*C*GG) sequence (*48*) in terms of its conformational geometry, and inter- and intramolecular interactions within the DNA and between the DNA and solvent, even though the 2-amino group in the minor groove of the important $G_3 \cdot m^5C_8$ base pair of the junction core trinucleotide (italicized) has been removed. In contrast, the analogous 2,6-diaminopurine sequence d(CCDGTACTGG) crystallizes as resolved duplex DNAs, just like its parent sequence d(CCAGTACTGG) (*76*). These results demonstrate that it is not the presence or absence of the 2-amino group in the minor groove of the $R_3 \cdot Y_8$ base pair that specifies whether a sequence forms a junction, but the positions of the extracyclic amino and keto groups in the major groove. Finally, the study shows that the arms of the junction can accommodate perturbations to the B-DNA conformation of the stacked duplex arms associated with the loss of the 2-amino substituent, and that two hydrogen bonding interactions from the C_7 and Y_8 pyrimidine nucleotides to phosphate oxygens of the junction cross-over specify the geometry of the Holliday junction.

4.2 Introduction

DNA Holliday junctions are ephemeral intermediates that have been implicated in the mechanisms of a large number of biological processes. A four-stranded DNA junction was proposed by Holliday over 40 years ago (*1*) to help explain the process of gene conversion in yeast and has since been implicated in a diverse array of related mechanisms including genetic recombination, DNA repair, and viral integration (*5, 17, 24, 55, 92*). The involvement of these structural intermediates over such a broad range of biological processes, typically serving to protect and expand the coding potential of genomes, has made junctions the focus of intense study. A recent series of single crystal structures reveal how interactions between the major groove of the double-helical arms and the phosphates at the junction cross-over help to define the formation of these complexes. We present here a systematic study to elucidate the contribution of the minor groove 2-amino substituent groups on the formation of four-stranded junctions in crystals.

The conformation of DNA Holliday junctions is now understood to be highly variable. Solution studies have shown that junctions exist in two general conformations—the open-X and stacked-X forms (Figure 4.1). The open-X form predominates under low salt conditions, with the four B-DNA arms arranged in essentially a square planar configuration to minimize the electrostatic interactions between the negatively charged phosphates of the phosphoribose backbones. High concentrations of monovalent and divalent cations help to shield these phosphates,

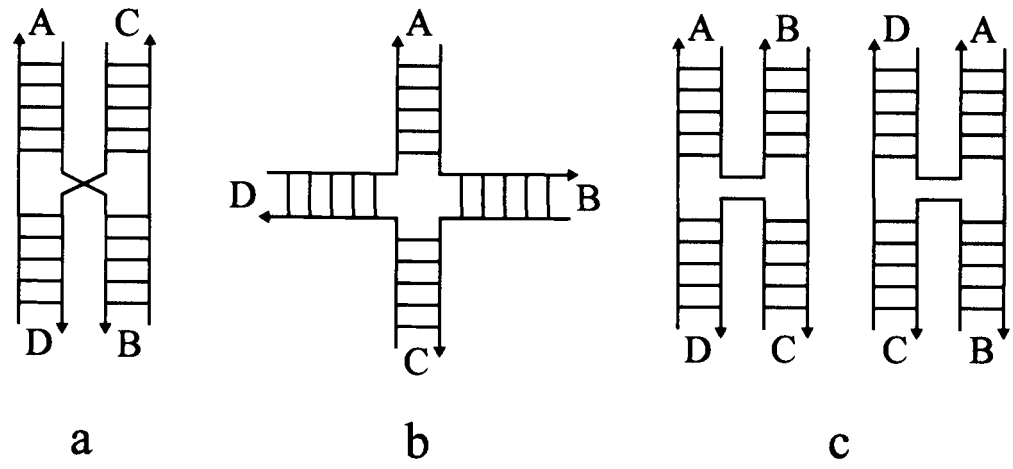

Figure 4.1: Three general forms of DNA Holliday junctions: (a) parallel stacked-X, (b) open-X, and (c) antiparallel stacked-X. The stacked-X form is generated by the coaxial stacking of B-DNA arms of the open-X form onto one another to form pseudo-continuous B-DNA duplexes.

allowing the junction to assume the more compact stacked-X structure in which the arms are coaxially stacked to form two semi-continuous B-DNA duplexes. In the stacked-X structure, the strands of the B-DNA arms can, in theory, assume either a parallel or antiparallel relationship (*26, 62*). To date, only the antiparallel form has been definitively observed in solution (*26, 41*). Furthermore, the duplexes of the antiparallel stacked-X junction can be related in a left or right handed orientation (Figure 4.1c). The antiparallel junction can be locked in one specific isomeric form or can be dynamic and readily switch between forms with the open-X structure serving as an intermediate (*25, 43, 66*).

Several single crystal structures of these important biological intermediates have now been determined at the atomic level (reviewed in (*80, 93*)). Comparison of junction forming sequences has identified a trinucleotide core motif as being instrumental to the formation of these four-stranded structures. The sequence of this trinucleotide core was originally defined as $A_6C_7C_8$ (*47*), but later expanded to $Pu_6C_7Py_8$, where Pu is either the adenosine or guanosine purine nucleotides, and Py can be a cytosine, 5-methylcytosine (m^5C) or 5-bromouridine (br^5U) pyrimidine-type nucleotide, but not thymidine (*76*). These nucleotides are involved in a number of crucial interactions that link the major groove surfaces of the stacked duplex arms to the cross-over phosphates. Atomic force microscopy (AFM) and nuclease digestion studies on synthetic junctions that incorporate the core ACC trinucleotide show the same overall geometry and cross-over topology as the crystal structures (*65*).

The conformation of junctions in solution and in the crystals appears to be related to these sequence dependent intramolecular interactions. A detailed

comparison of the available crystal structures show that the junctions can be grouped into two classes. Those structures that show a set of direct hydrogen bonding interactions, between i) the Py_8 base and the O2P phosphate oxygen of C_7, ii) the outside G_3 nucleotide with the O1P phosphate oxygen of Pu_6, and iii) the N4 amino group of C_7 and the O1P phosphate oxygen of Pu_6, have a well-defined geometry, while structures in which one or more of these interactions are missing or mediated by solvent show greater variability in conformation (*93*). Thus, it has become increasingly clear that the nucleotides within, and adjacent to the cross-over region of DNA Holliday junctions play key roles in defining the global geometry and formation of the stacked-X conformation.

Up to this point, it has not been clear what effect interactions in the minor groove may have on the formation and geometry of the junction. For example, a hexaaquo-calcium (II) ion has been observed in the minor groove as a common feature of the junction arms in several of the higher resolution structures, which may play an electrostatic role in stabilizing the complex. The currently available structures, however, do not directly tell us how important the substituent interactions are in this narrow groove. The sequence d(CCGGTACm^{5}CGG) forms a junction (*48*), but d(CCAGTACTGG) does not (*76*). In this comparison, we see that the extracyclic amino and keto substituents at the major groove surface are switched between the G_3-m^5C_8 and A_3-T_6 base pairs, but the difference in conformation may arise instead from the lack of a 2-amino group in the minor groove of the A_3-T_6 base pairs of the latter sequence. To distinguish between the effects of minor groove and major groove substituents, we have studied the single crystal structures of the sequences

d(CCIGTACm^{5}CGG), which replaces the G_3 nucleotide of d(CCGGTACm^{5}CGG) with an inosine (I) and thus removes the amino group from the minor groove, and the sequence d(CCDGTACTGG), where an amino group is added to the A_3 nucleotide of d(CCAGTACTGG) to introduce a 2,6-diaminopurine (D) at this position.

The modified bases inosine and 2,6-diaminopurine have long been used to study the effect of minor groove interactions on DNA conformation, showing how the 2-amino group affects the flexibility (*94, 95*), minor groove width and solvent interactions (*96-98*), endonuclease cleavage (*99, 100*) and molecular recognition (*101-103*) of the DNA duplex. In the current study, I and D nucleotides are placed in key positions of the conserved trinucleotide core sequence, where we have identified important junction stabilizing interactions. The $D_3 \cdot T_8$ base pairs in the d(CCDGTACTGG) sequence are A·T like in the major groove and G·C like in the minor groove. Conversely, the $I_3 \cdot m^5C_8$ base pairs of d(CCIGTACm^{5}CGG) are G·C like in the major groove and A·T like in the minor groove, both with standard Watson-Crick hydrogen bonding interactions and methyl groups at the C5-carbon of the pyrimidine base (Figure 4.2). We observe that d(CCDGTACTGG) crystallized as resolved B-DNA duplexes while d(CCIGTACm^{5}CGG) formed a DNA Holliday junction, thus further supporting the model that it is the interactions between the major groove and cross-over phosphates that determine junction formation in the crystal system (*47, 48, 76, 93*). Additionally, we see that minor groove interactions do not significantly contribute to the global geometry of the resulting junction structure.

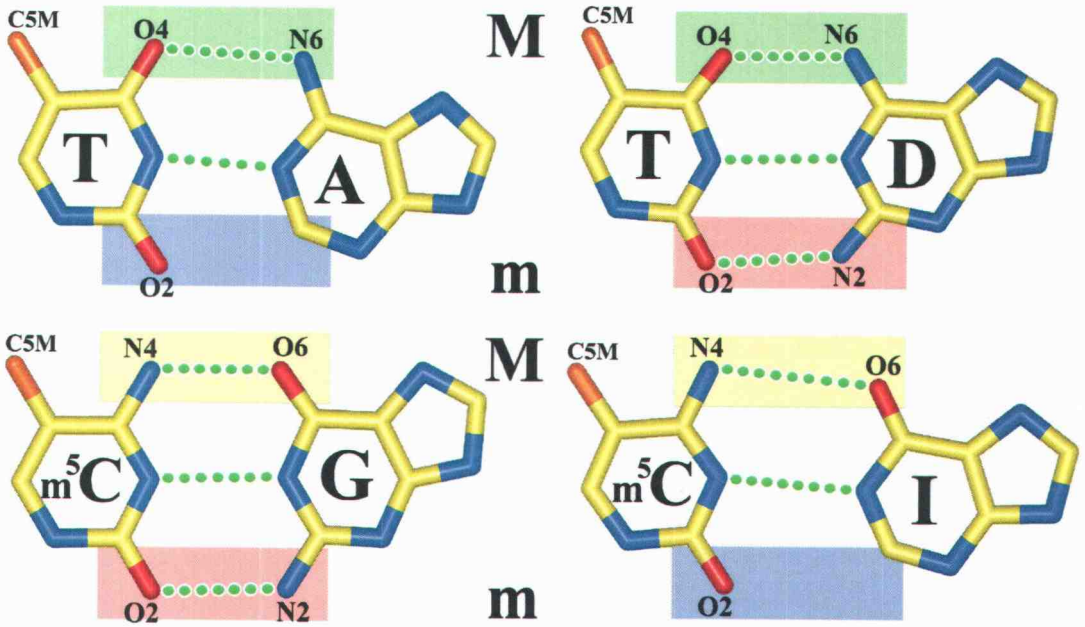

Figure 4.2: Standard and modified nucleotide base pairs located in the R_3-Y_8 base pair step. Exocyclic substituents are labeled with the location of the major (M) and minor (m) groove faces denoted adjacent to each base pair. I·m^5C base pair is similar to a G·m^5C base pair from the major groove face (*light yellow*) and an A·T base pair from the minor groove face (*light blue*). A D·T base pair is similar to an A·T base pair from the major groove (*light green*) and a G·m^5C base pair from the minor groove (*light red*). All base pairs have standard Watson-Crick hydrogen bonding interactions as indicated by *green circles* between each base. 5-methyl cytosine bases are used for consistency with the parent d(CCGGTACm^{5}CGG) junction structure and thymidine nucleotides in the Y_8 position.

4.3 Materials and Methods

All deoxyoligonucleotides were synthesized using solid-phase phosphoramidite chemistry with the trityl-protecting group left attached to the 5'-terminal hydroxyl group to facilitate subsequent purification from failure sequences via reverse phase HPLC. Preparative reverse phase HPLC was performed using a Microsorb 300 Å pore size C-18 column measuring 250 mm X 21.4 mm ID at a flow rate of 8 ml/min in 0.1 mM triethylamine acetate with an acetonitrile gradient. The full length sequence typically eluted at ~31 minutes post-injection in a mobile phase concentration of 20% acetonitrile. Purified tritylated DNA was deprotected by treatment with 3% aqueous acetic acid for fifteen minutes with gentle agitation, neutralized with concentrated ammonium hydroxide, and desalted on a Sigma G-25 Sephadex column yielding purified DNA. Purified DNA was stored at -80° C as a lyophilized powder and resuspended in deionized double-distilled water prior to use. All crystals were grown in 10 μl sitting drops by the vapor diffusion method and equilibrated against a 30 ml reservoir volume.

4.3.1 Crystallization and structure solution of d(CCIGTACm^{5}CGG). Thin diamond crystals of d(CCIGTACm^{5}CGG), in the monoclinic *C2* space group with dimensions a = 64.9 Å, b = 24.9 Å, c = 37.1 Å, and β = 109.2°, were grown from solutions containing 25 mM Na-Cacodylate (pH 7.0), 0.5 mM DNA, 180 mM $CaCl_2$ and 2.4 mM spermine tetra-hydrochloride equilibrated against 20% vol/vol 2-methyl-2,4-pentanediol (MPD) at 25° C. X-ray diffraction data to 2.0 Å resolution was

collected at liquid nitrogen temperatures using CuKα radiation from a Rigaku (Tokyo) RU-H3R rotating anode generator operating at 50kV and 100mA with an Raxis-IV image plate detector on a single crystal measuring 0.55 X 0.25 X 0.05 mm. An artificial mother liquor containing 25% MPD was used for cryoprotection. The space group and unit cell dimensions of d(CCIGTACm^{5}CGG) are very similar to previous junction crystals of the parent d(CCGGTACm^{5}CGG) sequence, with two strands in the asymmetric unit (a.s.u.). Thus a search model using the d(CCGGTACm^{5}CGG) junction (*48*) was used in EPMR (*83*) to solve the structure by molecular replacement, initially to 3.0 Å resolution. A distinct solution with a correlation coefficient of 77.5% and R_{cryst} = 36.2% was obtained with two unique strands in the a.s.u. adjacent to a crystallographic 2-fold symmetry axis. A simulated annealing omit-map with the C_7 crossover phosphate removed showed distinct F_o-F_c density within the crossover region indicative of a DNA Holliday junction. Subsequent refinement of this initial junction structure in CNS (*84*) using rigid body refinement, simulated annealing, several rounds of positional and individual B-factor refinement, as well as addition of solvent resulted in a final R_{cryst} of 23.6% and R_{free} of 25.8% (Table 4.1).

4.3.2 Crystallization and structure solution of d(CCDGTACTGG. Long rod-like crystals of d(CCDGTACTGG) were grown at room temperature in 5 mM TRIS(HCl) (pH 7.5), 23 mM calcium acetate, 0.6 mM DNA, and 16% MPD equilibrated against a reservoir solution of 28% MPD. Crystals were flash frozen directly out of the drop prior to data collection. Data for a crystal measuring 0.30 X 0.1 X 0.1 mm in size was collected to 1.85 Å resolution at liquid nitrogen

temperatures on BioCARS beamline 14-BMC at the Advanced Photon Source, Argonne National Labs, with 0.9 Å radiation. The crystals were indexed in the hexagonal space group $P6_122$ with unit cell dimensions a = 33.0 Å, b = 33.0 Å, and c = 88.6 Å, and are isomorphous to the crystals of the parent d(CCAGTACTGG) sequence. The structure of d(CCDGTACTGG) was solved by molecular replacement using the coordinates of the d(CCAGTACTGG) structure (UD0030). An EPMR search (against 3.0 Å data) using this initial search model of one strand adjacent to a crystallographic symmetry axis produced a correlation coefficient of 77.1% and R_{cryst} of 37.5%. Initial round of rigid body and rigid parts refinement on this model followed by simulated annealing produced a starting R_{cryst} of 35.3% and R_{free} of 38.8%. Structure refinement was carried out in CNS using standard positional and individual B-factor refinement followed by assignment of solvent molecules producing a final R_{cryst} of 23.2% and R_{free} of 25.8% (Table 4.1).

All data were processed and reflections merged and scaled using DENZO and SCALEPACK from the HKL suite of programs (*85*). Resolution cutoffs were determined according to $\langle I/\sigma(I) \rangle$, completeness, R_{merge}, and R_{meas} statistics. R_{meas} was determined using the program available at http://loop8.biologie.uni-konstanz.de/~kay/ and root-mean-square-deviation (RMSD) values were calculated using the McLachlan algorithm as implemented in the program ProFit v2.2 (http://www.bioinf.org.uk/software/profit/) (*86*). Structural analysis was performed using CURVES 5.2 (*87*). The coordinates and structure factors have been deposited in

Table 4.1. Data collection and refinement statistics

	d(CCIGTACm^{5}CGG)	d(CCDGTACTGG)
Data Collection		
Unit Cell		
a (Å)	64.9	33.0
b (Å)	24.9	33.0
c (Å)	37.1	88.6
β angle (°)	109.22	-
Space group	*C2*	*P6(1)22*
Total reflections	9439	34659
Unique reflections	3492	2528
Resolution (Å)	18.2-2.0	17.5-1.85
Completeness(%)[a]	93.4 (61.8)	96.3 (72.2)
I/sigma[a]	15.8 (2.4)	29.9 (4.98)
R_{merge}(%)[a,b]	5.0 (29.1)	5.6 (24.5)
Refinement		
Resolution	18.2-2.2	17.5-1.90
R_{cryst} (R_{free})(%)[c]	23.57 (25.82)	23.21 (25.83)
DNA toms	404	203
Solvent atoms	62	45
Average B-Factors (Å^2)	11.1 (11.7)	10.9 (20.6)
DNA atoms (Solvent atoms)		
RMSD bond lengths (Å)[d]	0.003	0.067
RMSD bond angles (°)[d]	0.8	6.2
Structural conformation	**Holliday Junction**	**B-DNA**

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

[b]$R_{merge} = \Sigma_{hkl} \Sigma_i | I_{hkl,i} - \langle I \rangle_{hkl} | / \Sigma_{hkl} \Sigma_i | I_{hkl,i} |$ 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.

[c]$R_{cryst} = \Sigma_{hkl} | F_{obs} - kF_{calc} | / \Sigma_{hkl} |F_{obs}|$. R_{free} = R_{cryst} for 10% of reflections that were not used in refinement (*104*).

[d] Root mean square deviation of bond lengths and angles from ideal values.

the Protein Data Bank (*105*) with accession number 1S1L for d(CCIGTACm^{5}CGG) and 1S1K for d(CCDGTACTGG).

4.4 Results

Major and minor groove interactions are known to affect the local and global conformation of nucleic acid structures (*106-109*). The crystal structures presented here elucidate the role of nucleotide substituent groups in the minor groove and, by extension, the major groove in the formation of DNA Holliday junctions. The sequence d(CCIGTACm^{5}CGG) was seen to crystallize as a DNA Holliday junction while d(CCDGTACTGG) crystallized as resolved B-DNA duplexes. In terms of base substituents, D·T base pairs are A·T like from the major groove and G·C like from the minor groove while I·m^{5}C base pairs are G·C like from the major groove and A·T like from the minor groove. Thus, this study demonstrates that specific atomic interactions which have been identified at the major groove surface of the trinucleotide core region of DNA Holliday junctions are specific for G·C base pairs in the major groove and do not accommodate A·T base pair substituents. Furthermore, the inosine junction structure provides an opportunity for direct comparison with the parent d(CCGGTACm^{5}CGG) junction structure (ACm^{5}C junction) to determine whether the minor groove 2-amino group in the R_3-Y_8 base step helps to define the geometry of the junction conformation.

4.4.1 Structure of d(CCIGTACm^{5}CGG) as a Holliday Junction. The single-crystal structure of the d(CCIGTACm^{5}CGG) sequence (iACm^{5}C junction) is a DNA Holliday junction in the right-handed antiparallel stacked-X conformation. The full four-stranded structure consists of two unique strands—one crossover and one noncrossover—related by a crystallographic 2-fold symmetry axis with the crossover occurring between nucleotides A_6 and C_7 of each set of pseudocontinuous B-DNA duplexes. These duplexes are formed by the stacking of a shorter four base pair arm (nucleotides C_1 to G_4 of an outside non-crossover strand with C_7 to G_{10} from a crossover strand) onto a longer six base pair arm (nucleotides T_5 to G_{10} of an outside strand stacked on nucleotides C_1 to A_6 of a crossover strand) (Figure 4.3).

The overall conformation of this junction, indeed of any crystal structure of four-stranded junctions, can be described by the geometric relationships between the set of stacked B-DNA arms and the center of the junction. We have defined these relationships and how they can be accurately measured very explicitly elsewhere (*48, 110*), and will briefly describe them here. The most easily recognized of these relationships is the interduplex angle (IDA), which is the angle that relates the ends of the stacked arms to the center of the junction cross-over. This is the angle that was estimated to be ~60° in solution and by AFM for synthetic junctions (*26, 38, 62, 64, 111*) that are fixed by sequence complementarity of four unique strands. The IDA when measured in this way, however, is a length dependent parameter; therefore, we have defined the parameter J_{twist} as a measure of the angle formed by the helical axes projected onto the resolving plane of the junction. This is length independent and, thus, is a more general descriptor for the angle relating the stacked arms of this

Figure 4.3: DNA Holliday junction crystal structure of d(CCIGTACm^{5}CGG). (a) Nomenclature used for the d(CCIGTACm^{5}CGG) junction with the full four-stranded complex generated by the application of a crystallographic two-fold symmetry axis to the two unique strands (*blue* and *orange*). An *asterisk* adjacent to the eighth nucleotide in each decamer sequence denotes the presence of a 5-methyl cytosine base. (b) Atomic structure of the d(CCIGTACm^{5}CGG) DNA Holliday junction. Nucleotides are rendered as *sticks* with the phosphodeoxyribose backbone rendered as a *solid ribbon* with the 5' and 3' ends labeled accordingly (panel was rendered with InsightII from MSI/Biosym, Inc.). (c) $2F_o$-F_c electron density (contoured at 1σ) depicting the junction crossover region with continuous density between nucleotides A_6 and C_7 of the junction crossover strands. Nucleotides T_5, A_6, C_7, and m^5C_8 rendered as *sticks*. Coloring scheme and strand designations in all panels are the same (panel created with Bobscript (*89*) and rendered in Raster3D (*112*)).

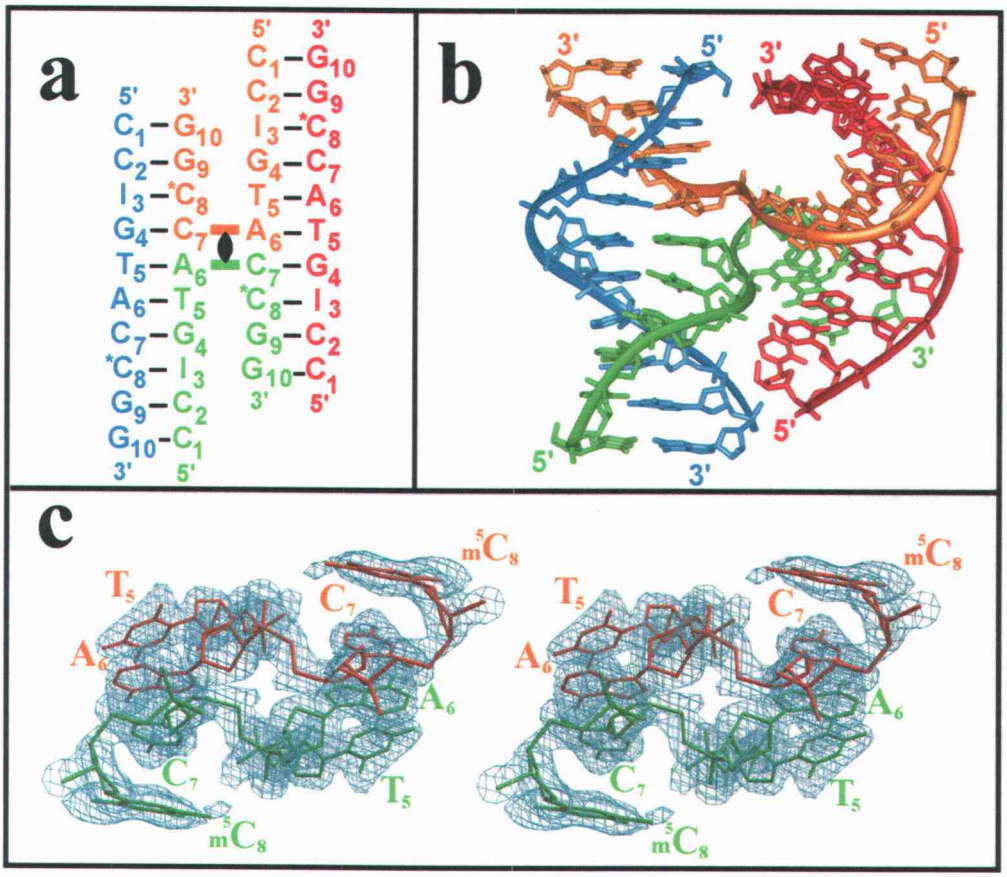
a
5' 3'
C1–G10
C2–G9
I3–*C8
G4–C7
T5–A6
A6–T5
C7–G4
*C8–I3
G9–C2
G10–C1
3' 5'
b
3'
5'
3'
5'
3'
5'
5'
3'
c
T5
A6
C7
5mC8
A6
C7
T5
5mC8
T5
A6
C7
5mC8
A6
C7
T5
5mC8

Figure 4.3

structure. In addition, the stacked B-DNA arms can translate along and rotate about their respective helix axes. These are reflected in the parameters J_{slide} and J_{roll}, respectively. The actual IDA for any crystal structure is thus a function of J_{twist}, J_{slide}, J_{roll}, and the length of each arm.

The structure of the iACm^{5}C junction is nearly identical to that of the parent ACm^{5}C junction in terms of its global geometry and detailed interactions. The global conformation of the iACm^{5}C junction structure (J_{slide} = 2.23 Å, J_{twist} = 42.5°, J_{roll} = 143.6° and IDA = 61.4°) is nearly identical to that of the parent ACm^{5}C junction (J_{slide} = 2.25 Å, J_{twist} = 44.3°, J_{roll} = 145.9° and IDA = 60.0°) in terms of these geometric parameters and as measured by the RMSD between like atoms (0.510 Å). The detailed interactions within the iACm^{5}C junction are also very similar to the parent structure, with minimal deviations. We have previously observed (*47, 48, 76*) that the global junction conformation is dependent on a set of key atomic interactions at the trinucleotide core region (Figure 4.4). These interactions include: i) direct hydrogen bond between the exocyclic amine of C_8 and a phosphate oxygen on the C_7 backbone phosphate (C_7-Y_8 interaction), ii) a solvent mediated interaction between outside R_3 (R is either an adenine or guanine nucleotide) keto oxygen and a phosphate oxygen of N_6 (R_3-N_6 interaction), and iii) a direct hydrogen bond between the O1P phosphate oxygen of N_6 and the exocyclic amine of C_7 (N_6-C_7 interaction). Consistent with this observation we see here that the solvent mediated interaction between the I_3 and A_6 nucleotides (R_3-N_6 interaction) is indeed present and mediated by a solvent molecule similar to that seen in the original d(CCGGTACCGG) junction (*47*). The N_6-C_7 interaction is also present and mediated by a solvent molecule similar to the parent

Figure 4.4: Stabilizing interactions within the stacked-X core region. Structure of the parent ACm^{5}C junction is shown with the location the $R_6C_7Y_8$ trinucleotide core region hydrogen bonding interactions indicated by *boxes*. The area within these boxes is enlarged, as indicated by the corresponding box colors (*red* and *blue*), providing more detailed information on each stabilizing interaction. For comparison, the core region interactions are also shown for the unperturbed junction structure of the sequence d(CCGGTACCGG) (*47*). Within these inset *boxes* Watson-Crick hydrogen bonding interactions are indicated by *green circles* while core region atomic interactions are shown as *orange circles*. The C_7-Y_8 and R_3-N_6 interactions are shown at the top while the A_6-C_7 interaction is shown in the bottom *box*. DNA atoms are rendered as *sticks* and colored by strand with the respective 3' and 5' termini noted. No direct or solvent mediated interaction is observed between the C_7 O2P and m$^5C_8$ N4 atoms (distance between these atoms is 4.4 Å as indicated by the *black arrow*) within the iACm^{5}C-J structure.

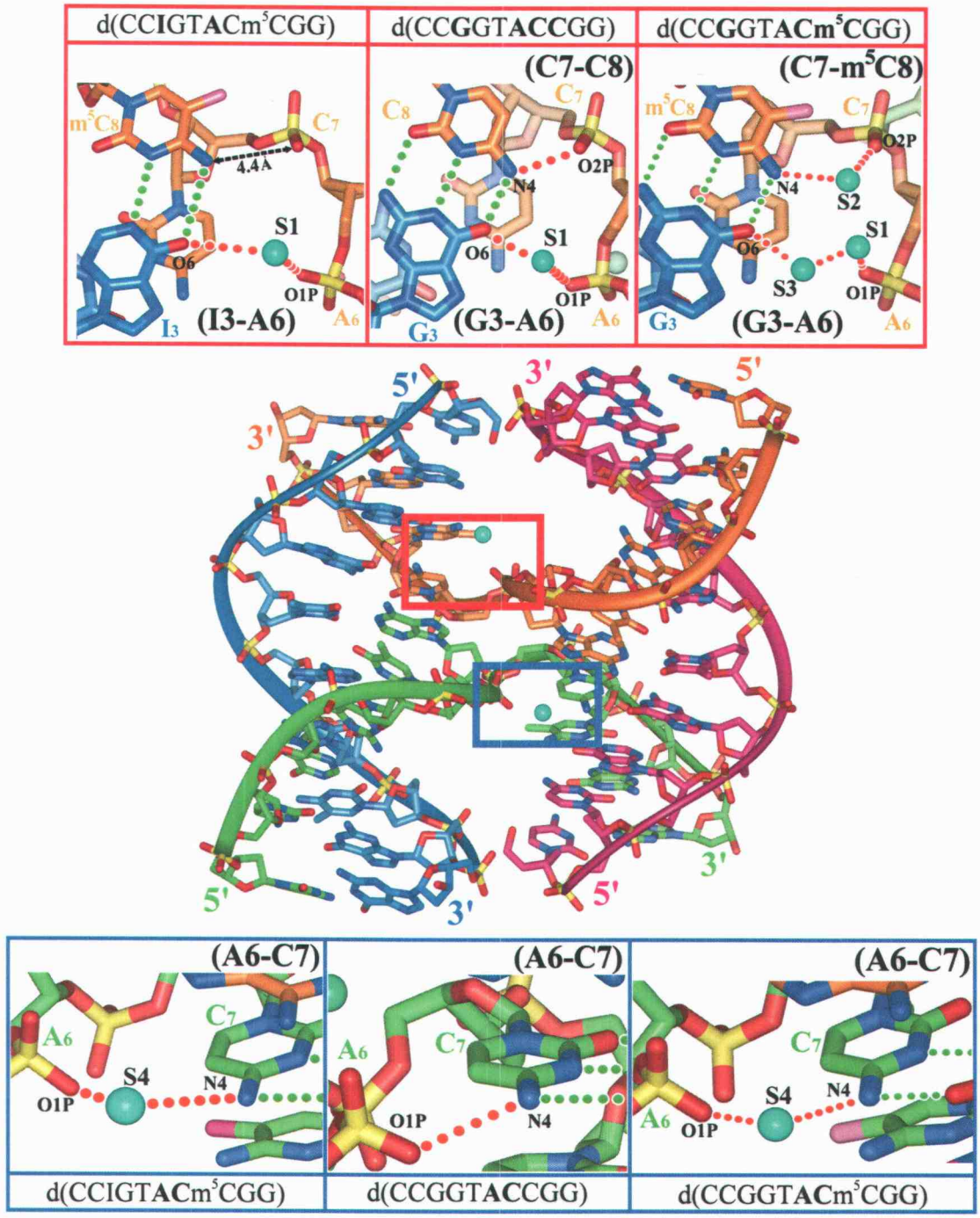
d(CCIGTACm5CGG)
d(CCGGTACCGG)
d(CCGGTACm5CGG)
(C7-C8)
(C7-m5C8)
m5C8
C7
4.4 Å
S1
O6
O1P
I3
(I3-A6)
A6
C8
O2P
N4
G3
(G3-A6)
S2
S3
5'
3'
(A6-C7)
S4

Figure 4.4

ACm^{5}C junction structure (*48*). It was surprising that no direct or solvent mediated interaction was observed between the N4 exocyclic amine of C_8 and the O1P phosphate oxygen of C_7 (C_7-Y_8 interaction) in this inosine containing structure. This C_7-Y_8 interaction had been observed as a direct or solvent mediated set of hydrogen bonding interactions in all but one of the previous DNA Holliday junction structures, suggesting that it is significant. In the case of the one previous exception and the current structure, the distance between the O2P of C_7 and N4 of C_8 is similar to that when solvent mediated interactions have been observed, suggesting that solvent is present, but was not resolved in the crystal structures.

Inosine has been shown to affect structural details of the DNA duplex, including groove widths, solvent interactions, and flexibility. In this case, we observe similar effects on the B-DNA arms of the iACm^{5}C junction. The width of the minor groove (as measured by the distance from the I_3 to G_9 backbone phosphorous atoms) adjacent to the I_3·m$^5C_8$ base pair is narrowed by 0.6 Å relative to the parent ACm^{5}C junction. This had previously been attributed to the loss of the minor groove 2-amino group (*106, 107, 113, 114*). This narrower groove disrupts the solvent structure within the groove. Most notably, the hexaaquo-calcium (II) ion that sits in the minor groove at nucleotide positions G_3 and C_8 in the ACm^{5}C junction is no longer present in the current structure (Figure 4.5), yet the global conformation of the iACm^{5}C junction is nearly identical to the ACm^{5}C junction. Thus, we can conclude that the presence of the minor groove calcium ion in the ACm^{5}C junction does not contribute significantly to the perturbed conformation. The absence of this calcium ion was rather surprising given the twenty-fold higher $CaCl_2$ concentration in the crystallization solutions of the

Figure 4.5: Minor groove solvent interactions adjacent to $R_3 \cdot Y_8$ base step in DNA Holliday junctions. (a) Minor groove of d(CCIGTACm^{5}CGG) structure between the $C_2 \cdot G_9$ and $G_4 \cdot C_7$ base pairs. $2F_o$-F_c electron density (contoured at 1σ) is shown in *blue* for minor groove solvent molecules assigned within this region. Solvent molecules are labeled with distances between solvent waters S7-S9 and S10 being: S7-S10 = 2.9 Å, S8-S10 = 4.2 Å, and S9-S10 = 2.6 Å. (b) Minor groove of the d(CCGGTACm^{5}CGG) Holliday junction structure. (c) Minor groove of the d(CCAGTACbr^{5}UGG) Holliday junction structure. View for all panels is looking into the minor groove with solvent water rendered as *green spheres* and calcium ions rendered as *purple spheres*. Interactions between calcium ions and solvent water are denoted by solid lines colored *green* and *purple*. All nucleotides are rendered as sticks and colored according to atom type (figure created with Bobscript (*89*) and rendered in Raster3D (*112*)).

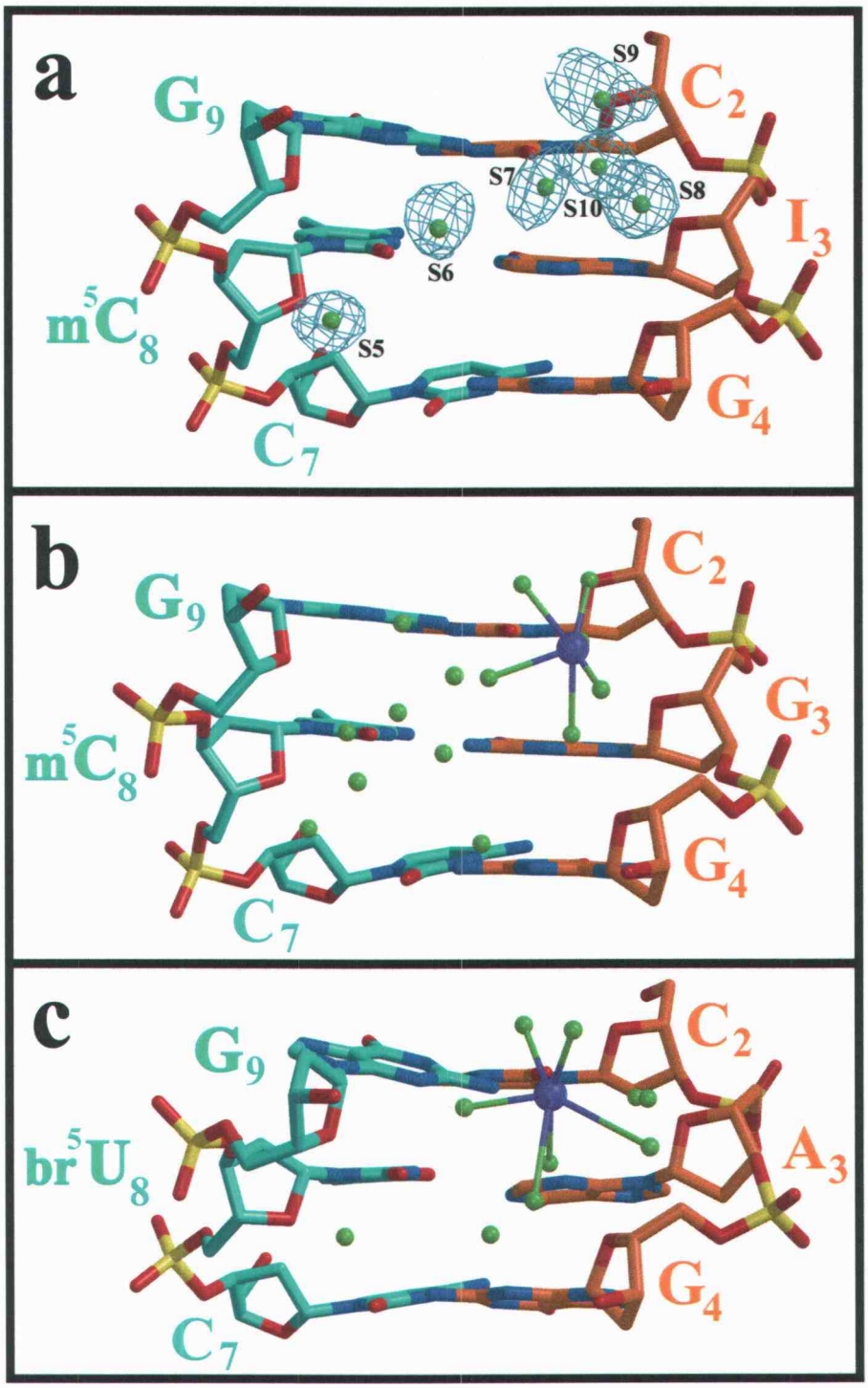
a
G9
S9
C2
S7
S10
S8
I3
S6
m5C8
S5
G4
C7
b
G9
C2
G3
m5C8
G4
C7
c
G9
C2
br5U8
A3
G4
C7

Figure 4.5

iACm^{5}C junction as compared to the ACm^{5}C junction. In fact, no cations were identified in the structure of the iACm^{5}C junction, although solvent is located in close proximity to the position where the calcium ion has been identified in previous junction structures (Figure 4.5). Any attempt to model this solvent as a calcium ion, even with reduced occupancies, resulted in a significant increase in R_{free}. Additionally, there is no ordered solvent around this location in the electron density maps that indicate anything other than water molecules in the minor groove.

The loss of the minor groove 2-amino group not only disrupts the solvent interactions but increases the propeller twist of the $I_3 \cdot m^5C_8$ base pair and the $G_4 \cdot C_7$ base pair 3' to the inosine base (Table 4.2). These perturbations are thought to increase duplex flexibility in the region missing a minor groove 2-amino group (*95, 107*). Greater flexibility is often manifested by an increase in propeller twist (*115*) and alteration of the backbone as observed by the presence of the BII backbone conformation (*113, 116*). It is interesting to note that both inosine residues within the iACm^{5}C junction adopt the unusual BII conformation, as is consistent with these previous studies. Comparison of the inosine junction with the parent ACm^{5}C junction shows that both guanine residues in the corresponding positions of the ACm^{5}C junction also adopt this unusual backbone conformation. Upon closer inspection of the junction structures it is apparent that the unusual backbone conformation in this location of the structures results from crystal packing interactions and not a loss of the 2-amino group in the minor groove. Additionally, a BII conformation observed in the inosine structure for G_{19}, yet not observed in the ACm^{5}C junction, is also caused by crystal packing interactions (the O1P of G_{19} is 3.3 Å from a symmetry related O3' of

Table 4.2. Comparison of conformational parameters for DNA Holliday junction structures of the d(CCGGTACm^{5}CGG) (ACm^{5}C-J), d(CCIGTACm^{5}CGG) (iACm^{5}C-J) and d(CCAGTACbr^{5}UGG) (ACbr^{5}U-J) sequences, and B-DNA structures of d(CCAGTACTGG) in the monoclinic (ACT-B *C2*) and hexagonal (ACT-B *P6$_1$22*) crystal forms, d(CCDGTACTGG) (dACT-B) and d(CCGGCGCCGG) (GCC-B).

		ACm^{5}C-J	iACm^{5}C-J	ACbr^{5}U-J	ACT-B		dACT-B	GCC-B
space group (ref)[a]		*C2 (17)*	*C2 (this paper)*	*C2 (15)*	*C2 (46)*	*P6$_1$22 (15)*	*P6$_1$22 (this paper)*	*C2 (15)*
Minor Groove Width (Å)[b]	C_2·G_9/Pu_3·Py_8	17.14	16.80	17.04	16.66	17.06	**17.41**	17.37
	Pu_3·Py_8/G_4·C_7	17.62	**17.00**	17.40	18.00	18.20	18.10	17.80
ζ dihedral angle (°)	Pu_3	171.1 (BII)	134.0 (**BII**)	-105.4 (BI)	-86.2 (BI)	-81.5 (BI)	-100.4 (BI)	152.4 (BII)
	Pu_{13}	167.0 (BII)	170.2 (**BII**)	-94.6 (BI)	N/A	N/A	N/A	128.5 (BII)
	Py_8	167.9 (BII)	-173.4 (BI)	152.9 (BII)	164.9 (BII)	-158.4 (BI)	-102.6 (BI)	-127.9
	G_{19}	-173.0 (BI)	179.3 (**BII**)	-145.7 (BI)	N/A	N/A	N/A	-114.1
HELICAL								
X-Displacement (Å)	C_2·G_9	2.11	1.61	0.98	1.01	0.72	**-1.19**	0.02
	Pu_3·Py_8	2.11	1.69	0.70	0.83	-0.30	**-1.13**	0.51
Tip (°)	C_2·G_9	-0.56	-1.19	2.46	2.88	-5.17	-6.95	-0.84
	Pu_3·Py_8	-6.37	4.30	-23.46	-18.02	-16.36	**-11.57**	-0.68
BASE PAIR								
Propeller Twist (°)	C_2·G_9	-7.6	-8.7	-22.8	-30.4	-22.3	-7.1	-14.4
	Pu_3·Py_8	-2.8	-8.7	-10.0	-20.7	-23.1	-18.0	3.8
	G_4·C_7	-15. 8	**-40.0**	-16.5	-8.6	-20.5	-16.5	-11.8
Stagger (Å)	C_2·G_9	0.28	0.70	0.59	0.27	-0.10	**0.41**	-0.57
	Pu_3·Py_8	0.15	-0.46	0.35	0.24	0.14	-0.63	-0.15
	G_4·C_7	-0.05	-1.39	-0.22	0.34	-0.01	-0.02	0.17
Buckle (°)	C_2·G_9	-0.8	-7.8	-0.5	-2.3	3.7	-1.3	3.1
	Pu_3·Py_8	4.6	7.3	-2.9	-3.1	-4.2	4.3	0.0
	G_4·C_7	0.6	**-6.1**	7.5	14.9	11.6	-10.2	4.6

Table 4.2 (Continued).

DIMER STEP								
Helical Twist (°)	C_2·G_9/Pu_3·Py_8	37.0	**25.0**	43.4	50.5	49.0	39.0	36.3
	Pu_3·Py_8/G_4·C_7	36.7	42.4	32.6	23.2	26.9	**37.7**	35.8
Roll (°)	C_2·G_9/Pu_3·Py_9	-5.8	5.5	-25.9	-20.9	-11.2	**-4.6**	0.2
	Pu_3·Py_8/G_4·C_8	2.4	**-12.4**	7.4	17.9	5.5	3. 6	-2.9
Slide (Å)	C_2·G_9/Pu_3·Py_{10}	0.27	0.51	0.91	1.05	0.98	**0.45**	0.35
	Pu_3·Py_8/G_4·C_9	-0.05	0.35	-0.57	0.19	0.08	-0.42	0.08
Shift (Å)	C_2·G_9/Pu_3·Py_{11}	0.00	0.08	-0.28	-0.18	-1.01	0.05	0.49
	Pu_3·Py_8/G_4·C_{10}	-1.03	-0.80	-0.66	0.35	0.26	**0.81**	-0.62
Rise (Å)	C_2·G_9/Pu_3·Py_{12}	3.47	3.10	3.54	3.56	3.57	3.35	3.59
	Pu_3·Py_8/G_4·C_{11}	3.50	3.70	3.11	3.03	3.26	**3.85**	3.42
Tilt (°)	C_2·G_9/Pu_3·Py_{13}	-0.99	**8.25**	-5.87	-6.88	-12.69	**3.85**	-2.51
	Pu_3·Py_8/G_4·C_{12}	-8.03	**-0.25**	-12.17	-8.15	-11.49	-10.54	-3.28

[a]References are provided for published structures.

[b]Minor groove interstrand width as measured by the P-P distance across adjacent dinucleotide steps (*117*). Sum of the van der Waals radii (5.8 Å) for the two phosphate groups have not been subtracted from the reported values.

G_{20}). Thus, there is no clear indication from the iACm^{5}C junction structure that loss of the minor groove 2-amino groove results in greater flexibility of the B-DNA arms. Comparison of crystallographic B-factors between these two junctions shows the same general patterns with no apparent increase in mobility or disorder of the inosine bases (average B-factor for DNA atoms in the iACm^{5}C structure is 11.1 Å^2 vs. 13.9 Å^2 in the ACm^{5}C junction).

4.4.2 B-DNA Structure of d(CCDGTACTGG). The d(CCDGTACTGG) sequence (dACT structure) crystallized as resolved B-DNA duplexes with one unique strand in the a.s.u., and the second strand of the duplex generated by a crystallographic two-fold symmetry axis (Figure 4.6). Having a single DNA strand in the a.s.u. precludes this structure from being a DNA Holliday junction, since a junction requires a minimum of two strands in the a.s.u.—one crossover and one non-crossover strand. This B-DNA duplex contains a D·T base pair which has base substituents in the major groove that are identical to an A·T base pair and substituents in the minor groove that are identical to a G·C base pair (Figure 4.2). The crystallization of this D·T base pair containing sequence as resolved B-DNA duplexes demonstrates that the 2-amino group at the purine R_3 base is not sufficient to define a junction in this lattice system and, therefore, further supports the model that it is the major groove interactions at the trinucleotide core that are responsible for stabilization of the four-stranded conformation in crystals

The B-DNA structure of dACT is very similar, but not identical to those of the parent sequence d(CCAGTACTGG) in the hexagonal (RMSD = 1.17 Å for like

Figure 4.6: Structure of d(CCDGTACTGG). (a) Structure of d(CCDGTACTGG) as a B-DNA double-helix in the *$P6_122$* crystal form. $2F_o$-F_c electron density (contoured at 1σ) depicting the unique asymmetric unit strand (rendered as *red sticks*) which sits adjacent to a crystallographic symmetry axis. (b) $2F_o$-F_c electron density (contoured at 1σ) depicting the closest approach between two duplexes. *Arrow*s indicate where the phosphoribose backbone would deviate if this sequence were to form a junction. Since the structure of d(CCDGTACTGG) has only one strand in the a.s.u. we can conclude with certainty that this sequence does not form a DNA Holliday junction, as a minimum of two unique strands are required for junction structures (figure created with Bobscript (*89*) and rendered in Raster3D (*112*)).

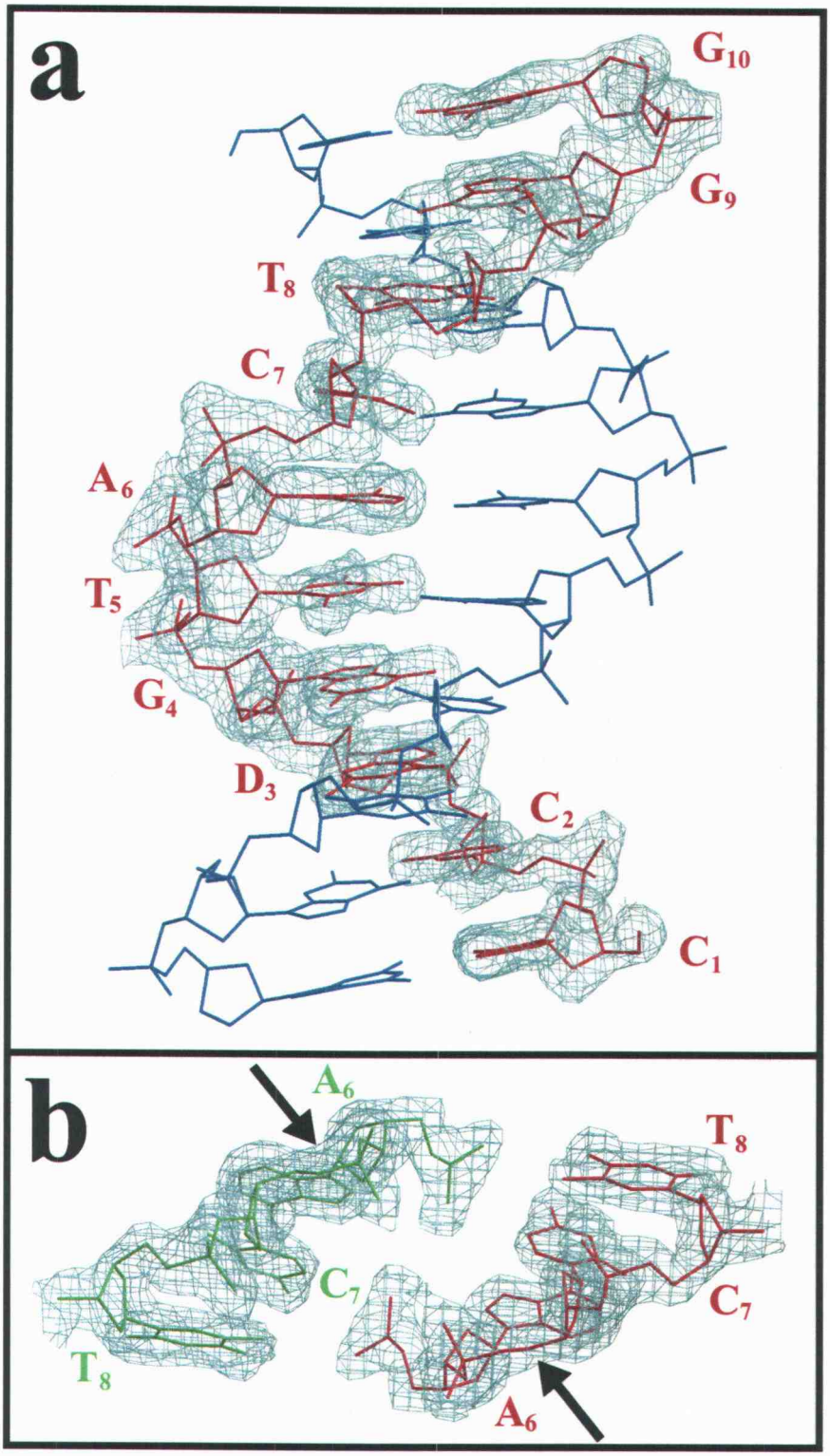

Figure 4.6

nonhydrogen atoms, ACT-$P6_122$) (*76*) and monoclinic (RMSD = 2.01 Å for like nonhydrogen atoms, ACT-*C2*) space groups (*90*). This observation that the ACT B-DNA structure can be crystallized in essentially the same monoclinic *C2* space group as nearly all of the crystals of junctions suggests that the lattice system does not preclude, and indeed can accommodate either DNA conformation. Thus, by extension, the effects on the structure of the B-DNA duplex from introducing a 2-amino group into the minor groove appear to be independent of crystal type. The dACT structure reported here and ACT-$P6_122$, are in the same crystal lattice ($P6_122$) with very similar crystal packing interactions. We do, however, observe crystal packing interactions that cause minor perturbations to the local base geometry (such as the close approach of the C_7 base with a symmetry related phosphate group of A_6).

We should note that there are noticeable deviations from ideal geometry in the dACT structure as noted by the high RMSD values for the bond lengths and angles in this structure (Table 4.1) which can not be entirely accounted for by crystal packing interactions (the analogous ACT-$P6_122$ shows overall RMSD values for bond lengths and angles that are very close to ideal (*76*)). Divalent cations have previously been shown to cause significant perturbations to DNA structures (*109*). Thus, it may be the presence of two hexaaquo-calcium (II) ions interacting with the C_7 and A_6 nucleobases that are responsible for significant distortions to the bases from their ideal geometries.

Introduction of a minor groove 2-amino group has been shown to increase the minor groove width, increase helical twist, and alter the general base characteristics such as shift, slide, rise, tilt, roll, and propeller twist of nucleotides within B-DNA duplexes (*106, 107, 113, 114*). Indeed, we see a significant increase in the helical twist

for the D·T base pair by ~10.8° when compared to the corresponding step in the ACT structures, and 2° greater compared to the corresponding step in the B-DNA structure of d(CCGGCGCCGG) (GCC-B) (*82*). This increase in helical twist does not extend to the overall twist of the duplex. The average helical twist for the dACT duplex is 34.5° (SD = 5.1°), which is similar to 35.3° for the ACT-$P6_122$ (SD = 8.8°, (*76*)) structure and 36.3° for the ACT-*C2* (SD = 10.3°, (*90*)) structure. Thus, the dACT B-DNA structure is not over wound overall even though the individual D·T base pairs display greater helical twist. Furthermore, there is a 0.75 Å increase in the minor groove width around the D·T base pair when compared to the ACT-*C2* structure (*90*) and 0.35 Å increase when compared to the ACT-$P6_122$ structure. This widening is expected from the addition of the 2-amino group in this location.

The presence of a 2-amino group in the minor groove is also known to disrupt the spine of hydration which is commonly observed in the minor groove of B-DNA duplexes (*118*). Indeed, we see a disruption of solvent in the distal regions of the minor groove adjacent to the 2-amino group of the 2,6-diaminopurine relative to the parent d(CCAGTACTGG) structures. Thus, we can see that the introduction of the 2,6-diaminopurine base into the d(CCDGTACTGG) B-DNA structure results in a D·T base pair which is more G·C like then A·T like in terms of general characteristics of the minor groove (Table 4.2) yet still forms resolved B-DNA duplexes and not a four way junction.

4.5 Conclusions

We solved the single crystal structures of the DNA sequences d(CCIGTACm^{5}CGG) and d(CCDGTACTGG) in order to elucidate the effect of major and minor groove substituents on the formation of four-stranded Holliday junctions. The two sequences differ from their parent sequences d(CCGGTACm^{5}CGG) and d(CCAGTACTGG) by the presence of an I·m^{5}C or D·T base pair in the third step, thus either eliminating or adding an amino substituent in the purine 2-position of the $G_3 \cdot C_8$ base pair or the $A_3 \cdot T_8$ base pair of the respective structures. This renders the I·m^{5}C base pairs to be G·C like in the major groove and A·T like in the minor groove, while D·T base pairs are A·T like in the major groove and G·C like in the minor groove (Figure 4.2). The sequence d(CCIGTACm^{5}CGG) was seen to crystallize as a DNA Holliday junction in a right handed antiparallel stacked-X configuration just like its parent sequence, while the d(CCDGTACTGG) sequence formed resolved B-DNA duplexes, again like its parent sequence. Thus, we can conclude that the minor groove substituents at the $N_6C_7Y_8$ trinucleotide core motif do not play a significant role in the formation of four-stranded junctions.

There are, however, effects from the 2-amino substituent on the conformation and solvent structure of the duplex arms and, indirectly, on the major groove interactions that help us to focus on those structural features that are important and not important in defining the conformation of the Holliday junction in this crystal system. It is now clear that the width of the helix in and around the cross-over point does not affect the formation or geometry of the junction. Narrowing the minor groove by

removing the 2-amino group does not interfere with the junction, while widening it by the introduction of this group does not induce its formation. The associated solvent rearrangements, including displacement of cations resulting from the narrowing of the minor groove has little effect on the geometry of the junction. The overwinding or underwinding of the helix and buckling or propeller twist of the base pairs also has little effect. This study, therefore, reinforces our contention that the conformation of the helical arms are determined primarily by the sequence dependent features of the arms and is only minimally perturbed by the junction, and that the junction can accommodate a number of conformational variations in these arms. The compensatory over and underwinding of the helix (as seen in the increased helical twist around the inosine bases, but associated underwinding of adjacent base pairs to give an average helical twist that is similar to the parent junction) suggests that the long range interactions between arms, as previously proposed (*73*), may remain an important component of the junction.

Perhaps the most significant result from this study is that we can now distinguish two specific interactions at the major groove surface that are the primary determinants of the junction geometry. Specifically, the hydrogen bonds from the Y_8 amino nitrogen to the cross-over phosphate oxygen at C_7 and the amino nitrogen of C_7 to the A_6 phosphate oxygen of the opposing duplex arm, whether they are direct or solvent mediated, are the two defining interactions. The hydrogen bond from R_3 to the A_6 phosphate oxygen of the adjacent duplex can be mediated by one or two intervening solvent molecules without significantly affecting the conformation. Additionally, the monovalent cation at the center of the original d(CCGGTACCGG) junction structure

was previously shown to not play a significant role in defining the junction conformation (*76*). Thus, the conformation and structural variability of the antiparallel stacked-X Holliday junction appears to hinge only on these two specific interactions, and it is the perturbation of these interactions by sequence, base substituents, or drugs and ions that define the gross geometric features of this recombination intermediate.

4.6 Acknowledgements

We thank Ganapathy Sarma for helpful discussions during manuscript preparation. This work was supported by grants from the National Institutes of Health (R1GM62957A) and the National Science Foundation (MCB0090615). The X-ray diffraction facilities are supported by the Proteins and Nucleic Acids Facility Core of the Environmental Health Sciences Center at OSU (NIEHS ES00210) and a grant from the Murdock Charitable Trust. Access to BioCARS Sector 14 was supported by the National Institutes of Health, National Center for Research Resources, under Grant RR07707.

Chapter 5

How sequence defines structure? A crystallographic map of DNA structure and conformation

Franklin A. Hays, Amy Teegarden, Zebulon J.R. Jones, Michael Harms, Dustin Raup, Jeffrey Watson, Emily Cavaliere and P. Shing Ho

In press, Proc. Natl. Acad. Sci. USA,

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

5.1 Summary

The fundamental question of how sequence defines conformation is explicitly answered if the structures of all possible sequences of a macromolecule are determined. We present here a crystallographic screen of all permutations of the inverted repeat DNA sequence d(CCnnn$N_6N_7N_8$GG), where N_6, N_7 and N_8 are any of the four naturally occurring nucleotides. At this point, 63 of the 64 possible permutations have been crystallized from a defined set of solutions. When combined with previous work, we have assembled a dataset of 37 single-crystal structures from 29 of the sequences in this motif, representing three structural classes of DNA (B-DNA, A-DNA, and four-stranded Holliday junctions). This dataset includes a unique set of amphimorphic sequences, those that crystallize in two different conformations and serve to bridge the three structural phases. We have thus constructed a map of DNA structures which can be walked through in single nucleotide steps. Finally, the resulting dataset allows us to dissect in detail the stabilization of and conformational variations within structural classes, and identify significant conformational deviations within a particular structural class that result from sequence rather than crystal or crystallization effects.

5.2 Introduction

The basic principle that sequence defines the three-dimensional structure of a macromolecule was first established in 1957 by Anfinsen (*119*), who showed that

ribonuclease A can be reversibly denatured and renatured in solution. The exact relationship between sequence and conformation, however, remains elusive—the "protein-folding problem", the long coveted "Holy Grail" in protein chemistry, is yet to be solved. This is not from lack of effort—the question of how sequence determines structure has been attacked by nearly every conceivable experimental and theoretical approach. The effect of sequence on the structure-stability relationship in T4 lysozyme has been extensively studied by crystallographic and thermodynamic analyses (*120*), while the propensity of single amino acids to effect formation of isolated α-helices have been studied using host-guest peptides (*121, 122*). However, if the structures of all possible sequence combinations of a macromolecule are determined, then this problem is solved explicitly. We present here the results from a crystallographic screen of all possible sequence permutations within a defined inverted repeat (IR) sequence motif to construct a map of DNA structures that are available to this sequence motif.

DNA is highly polymorphic, capable of adopting a large variety of structures in crystals and in solution, including right and left-handed double-helices, triple-helices, and four-stranded G-quartets and Holliday junctions (*123, 124*). The current dataset of DNA structures has grown over the years, but not in a systematic manner; therefore, it has been difficult, if not impossible, to relate the structures within the framework of a common lineage of sequence or environment. Our attempt to crystallize all the possible combinations of a defined DNA sequence motif from a common set of crystallization solutions initiated with the serendipitous findings that the sequences d(CCGGG***ACC***GG) (*46*) and d(CCGGT***ACC***GG) (*47*) crystallize as

four-stranded Holliday junctions, the central intermediate in recombination and recombination dependent cellular processes (*6*). A common ACC trinucleotide core at nucleotides $N_6N_7N_8$ and an associated set of intramolecular interactions were subsequently identified that fix the junction (where the phosphoribose backbone crosses-over between B-DNA duplexes) and thus allow its crystallization in these sequences (*93*). In order to search for other trinucleotides that stabilize junctions in an unbiased manner, we designed a crystallographic screen to solve the crystal structures of all 64 permutations of the sequence d(CCnnn$N_6N_7N_8$GG), where $N_6N_7N_8$ can be any of the four common nucleotides and *nnn* are specified accordingly to maintain the IR motif and thus self-complementarity of the sequences (Table 5.1). The sequences in this study will be referred to by the unique $N_6N_7N_8$ trinucleotide motif. Although, when isolated, there are only 32 unique trinucleotides, once placed in the context of this motif, each of the 64 possible trinucleotides becomes unique. For example, the TTT sequence in this motif defines the overall sequence d(CCAAATTTGG)—the AAATTT central core of this sequence is associated with highly curved B-DNA (*125, 126*). In contrast, the complementary AAA sequence would be found in d(CCTTTAAAGG), where the TTTAAA core is known to not show significant curvature of the DNA helix.

This IR sequence motif also has the potential to adopt other DNA structures—the trinucleotide has been suggested to be the minimum motif to distinguish between the double-helical forms of B- and A-DNA (*127, 128*). Thus, we expected the current crystallographic screen to sample at least three different DNA structures (Figure 5.1a).

Table 5.1. Conformations observed in the single-crystals of d(CCnnn$N_6N_7N_8$GG). The trinucleotides $N_6N_7N_8$ are presented as a triplet table, with conformations that have been determined from this screen (either new or repeats of prior structures) in bold and conformations that were determined by other groups but have not been repeated in the screen in lower case. B-DNA structures are labeled as "B", A-DNA as "A", and four-stranded Holliday junction as "J". Sequences that have been crystallized in the screen, but whose structures have not yet been determined are labeled as "x".

N_6	N_7				N_8
	G	**C**	**A**	**T**	
G	**A**	x	x	x	**G**
	B/*a*	**b/J**	**B**	**B**	**C**
	A	**B**	x	x	**A**
	A	**B**	x	*b*	**T**
C	**A**	**A**	x	x	**G**
	x	**A/J**	x	**B**	**C**
	-	x	x	x	**A**
	x	*b*	x	x	**T**
A	x	x	x	x	**G**
	B	**J**	**B**	**B/J**	**C**
	B	x	**B**	x	**A**
	b	**B**	*b*	x	**T**
T	**A**	*a*	x	x	**G**
	x	x	x	**B**	**C**
	x	x	x	**B**	**A**
	x	*b*	x	x	**T**

The IR motif, however, is limited in that it samples only those structures that are available to self-complementary sequences with Watson-Crick base pairs, thereby effectively excluding, for example, G-quartets and I-motifs. In addition, the non-alternating CC/GG dinucleotides at the two ends effectively exclude left-handed Z-DNA from this screen. We, therefore, consider this study as the first step towards developing a set of structures under a common framework that, upon extending this motif, will eventually allow us to distinguish the effects of sequence on all possible DNA forms. For this study, we will distinguish between overall structure and the detailed conformational variations that can occur within a particular structural class.

The advantage of this crystallographic approach is that sequence effects on structure and conformation can be directly related to specific molecular interactions. The potential problems, however, are those purported to be inherent in DNA crystallography, including the potential that lattice interactions greatly influence or actually induce the conformation observed in the crystal (*129-131*). The results demonstrate that the sequence motif designed for this study is highly crystallizable and samples DNA structures and conformational variations within these structural classes broadly, apparently independent of such overt crystal lattice effects.

5.3 Materials and Methods

Deoxyoligonucleotides were synthesized with the dimethoxytrityl protecting group left intact at the 5'-terminus to facilitate purification by preparative reverse-phase HPLC. Sequences were detritylated by treatment with 3% acetic acid and passed

over a gel filtration column to yield purified DNA stocks. The DNAs were stored at -80°C as lyophilized powders and redissolved in Millipore water prior to use without any further purification.

All sequences were crystallized by sitting drop vapor diffusion, with setups of 10 μL total sample drops containing 25-100 mM sodium cacodylate buffer at pH 7.0, 0-325 mM $CaCl_2$ and 0-3 mM spermine equilibrated against 30 mL of 2-35% aqueous 2-methyl-2,4-pentanediol in the reservoir (Table 3 in Appendix). These solutions are similar to previous conditions used to crystallize DNA oligomers in this laboratory (*47, 76*). Each sequence was subjected to the full range of $CaCl_2$ and spermine concentrations, with each crystal form refined individually to obtain diffraction quality single crystals. X-ray diffraction data were collected in-house on a Rigaku diffractometer with an R-AXIS IV detector, or at the Advanced Photon Source and Advanced Light Source synchrotrons. Data were reduced using the HKL suite of programs. Structures were solved by molecular replacement using isomorphous structures available in the Nucleic Acid Database (NDB) (*74*) as initial models or by multiple wavelength anomalous dispersion phasing. Structures have all undergone initial refinement with addition of solvent, with most having $R_{free} < 30\%$. All structures have been deposited into the Protein Data Bank and the NDB (Supplemental Table 2 in Appendix), and will be updated as they become fully refined.

5.4 Results

In order to study how sequence defines the structure and conformation of DNA, we attempted to determine the crystal structures of all 64 permutations of the IR sequence d(CCnnn$N_6N_7N_8$GG), where N_6, N_7 and N_8 are any of the four naturally occurring nucleotides. At this point in the study, a remarkable 63 of the possible 64 sequences (the lone exception being CGA) have been crystallized from a common set of crystallization solutions (Supplemental Table 1 in Appendix), indicating that this sequence motif is readily crystallizable and thus ideal for this type of exhaustive structural screen. We currently have solved the structures of 23 sequences from the screen and, along with similar structures from previous studies on sequences in this same motif, have a total of 29 sequences for the study. Crystals that were previously solved and have been reproduced using the current solutions are included in the 23. The screen itself yielded 24 new structures from 21 previously unreported sequences. Although incomplete, the current structures from the screen are sufficient to define a set of general rules that show how individual nucleotides distinguish between various DNA structures and affect their detailed conformations.

5.4.1 Structural Classes: The structures resulting from this screen fall into three classes—the right-handed double-helical forms of B- and A-DNA and four-stranded Holliday junction (Figure 5.1a). Not surprisingly, a majority of these sequences (17, including 5 from previous reports) crystallize as B-DNA duplexes, 7 form only A-DNA and one sequence (ACC) forms only the junction (Table 5.1). A unique aspect of the results, however, is that a set of amphimorphic sequences (those

that crystallize as two different structures) have been identified that link each of these structural phases. They include ATC, which crystallized as both B-DNA and junction, and CCC, which crystallized as both A-DNA and junction under the crystallization conditions of the screen. In addition, GCC, which was previously reported as B-DNA (*82*), was crystallized under our conditions as a junction (*76*) (the B-form was crystallized with Mg^{2+}, while the junction was with Ca^{2+} cations). These amphimorphic sequences, therefore, sit at the interfaces between the junction and the two duplex DNA forms. Finally, GGC, which was previously reported as A-DNA (*132*), was crystallized as B-DNA in the current study (the difference being the alcoholic precipitant used to crystallize the A-DNA form) and represents a sequence at the B-A interface.

5.4.2 Crystallization Conditions: In comparing crystallization solutions, we see first that double-stranded A- and B-DNAs crystallize across nearly the entire range of divalent and polyvalent cations (Figure 5.1b). B-DNAs are seen to crystallize at lower spermine concentrations, but, interestingly, at higher calcium (II) concentrations than A-DNAs. Still, all sequences were subjected to the entire range of solutions in this screen and, therefore, there was no bias towards any structural form designed into the experiment. Four-stranded junctions generally crystallize under lower Ca^{+2} solutions (below 15 mM) than the B-DNAs. A comparison of the crystallization solutions of the amphimorphic sequences, however, confirm our expectations that high concentrations of divalent cations are required to shield the negative electrostatic potential at the phosphates of the compact stacked-X junction (*26*) and to prevent migration of the

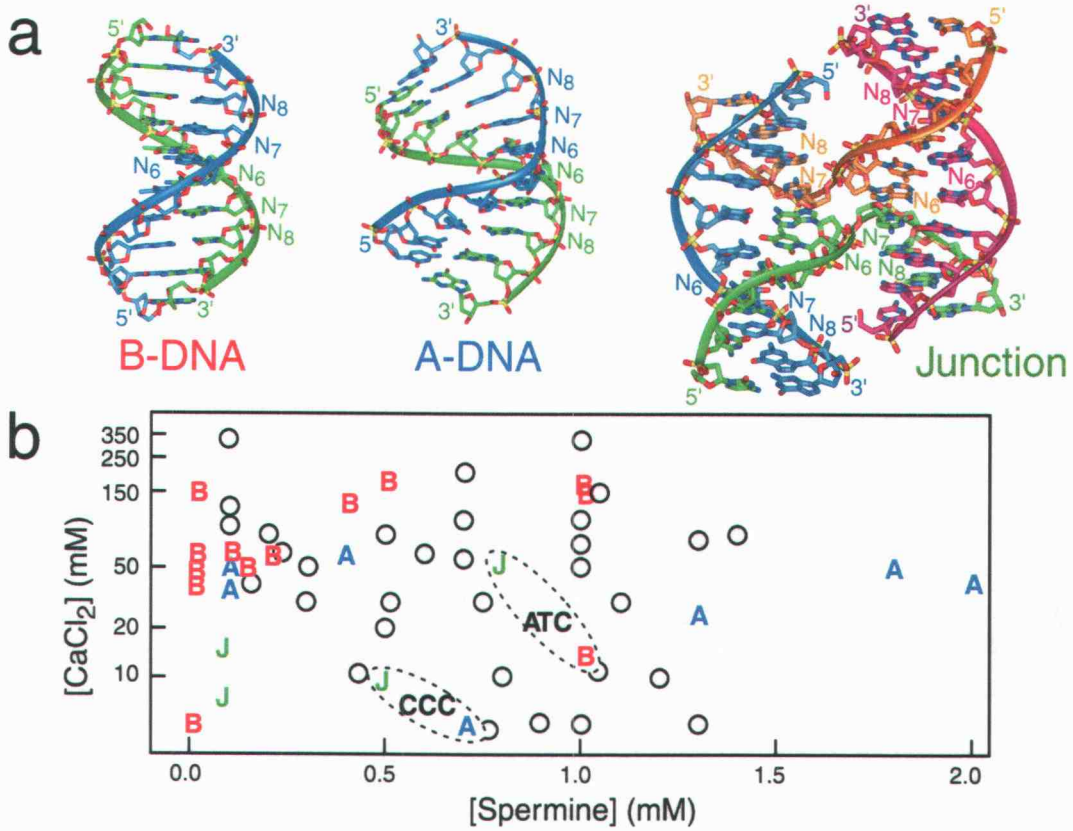

Figure 5.1. Structures from the crystallographic screen of the IR sequence d(CCnnn$N_6N_7N_8$GG), where all 64 combinations of the $N_6N_7N_8$ trinucleotide are sampled. **a**. The conformations observed in the single crystal structures of this sequence include standard B-DNA (AGC structure shown), the altered A-DNA duplex (GGG structure shown), and the four-stranded Holliday junction (ACC structure shown). The positions of the $N_6N_7N_8$ trinucleotide are labeled in each of the structures. **b**. The $CaCl_2$ and spermine concentrations yielding crystals of B-DNA (B), A-DNA (A) and junctions (J) are compared, with the concentration of $CaCl_2$ plotted on a logarithmic scale. Open circles indicate conditions that yielded crystals, but where the conformation has not been determined. Only one label for each form is denoted in cases where crystallization conditions overlap. The conditions for crystallization of the amphimorphic sequences ATC and CCC are encompassed in ovals and labeled by the trinucleotide sequence.

junction along the DNA strands (*81*) of a given sequence. ATC was seen to form junctions at higher concentrations of Ca^{2+} and B-DNA duplexes at lower cation concentrations. Interestingly, CCC forms a junction at higher Ca^{2+} concentrations, but is A-DNA at lower concentrations. Thus, for any particular sequence, the junction is stabilized by higher concentrations of divalent cations, as expected for B-DNA (*26*), but we now also see that this is true also for the A-DNA duplex. We note, however, that in both these cases, the duplex and junction forms can coexist at the intermediate Ca^{2+} crystallization solutions.

A comparison of the two forms of GCC shows that Ca^{2+} is more effective than Mg^{2+} at stabilizing the junction over B-DNA (it should be noted that ACC crystallizes as a junction with either Ca^{2+} or Mg^{2+} (*46, 47*). Finally, GGC is seen to be induced by alcohol to form A-DNA. Although one expects alcohols to favor the A-form (*133*), this study directly implicates alcoholic precipitants on the structural class in crystals. Thus, the solution conditions that favor crystallization of each conformation are generally consistent with what has been observed for the behavior of DNA in solution.

5.4.3 Crystal Lattices: The crystals for which we have diffraction data (43 sequences) fall into 15 unique space groups (including 7 of the 14 Bravais lattice types) and 27 associated crystal forms, demonstrating that the structures of this sequence motif are not restricted by the crystal lattice, but are free to assume a variety of crystal forms (Supplemental Table 2 in Appendix). In addition, the two amphimorphic sequences ATC and CCC have crystal forms that are specific for their respective conformations, but that these crystal forms can coexist under certain conditions. Furthermore, at least one space group (monoclinic *C2*) is common for both

B-DNA and the junction, and a second (orthorhombic $P2_12_12_1$) is seen for both A-DNA and B-DNA. Our supposition, therefore, is that each sequence adopts a crystal lattice that can accommodate the particular structure(s) formed in the crystallization solution. Thus, the crystal lattice serves less as a tyrant (*129, 130*) here than as an experimental facilitator that leaves no doubt concerning the structure(s) of each sequence.

5.4.4 Crystallographic map of DNA structural space: The dataset resulting from this screen defines a phase map that relates DNA structures to sequence and environment (Figure 5.2). We can walk through this map of DNA structures in single nucleotide steps starting with the ACC sequence, thus far the only core trinucleotide that uniquely forms the Holliday junction. A transition of the central C of ACC to T generates the amphimorphic ATC sequence that sits at a junction/B-DNA interface that is dependent on the concentration of divalent cations. The transition from ACC to GCC defines a similar interface, but one that is dependent on the type of divalent cation. To fully enter the B-phase, additional transitions or transversions to convert ATC to (G/C/T)TC, or GCC to GCT or GTC are required. Alternatively, the transversion of ACC to CCC yields an amphimorphic sequence at the junction/A-DNA interface. The trinucleotide is pulled further into the fully A-DNA phase by systematic transitions and transversions from CCC that lead towards GGG, the classic A-DNA trinucleotide. Finally, the interchange between A- and B-DNA duplexes is seen to occur through the amphimorphic sequence GGC in a solvent dependent manner, or, more directly, from GGT to AGT, which is consistent with the understanding that A/T favors B-DNA (*134*).

Figure 5.2. Map of DNA structure space sampled by crystals of d(CCnnn$N_6N_7N_8$GG). The map is divided into three specific structural classes (labeled "B" for B-DNA, "A" for A-DNA and "J" for junctions) and the interfaces between each conformational phase. The sequences in upper case letters define those that have been uniquely solved or reproduced in the current study, while those in lowercase are structures from previous studies, but not reproduced here. The rectangle around GCC indicates that the structure is induced by a change in divalent cations (from Ca^{2+} to Mg^{2+}). Similarly, the oval around GGC indicates that the A-form is induced by alcohol. Arrows trace paths through the conformational map as the $N_6N_7N_8$ trinucleotide undergoes single-nucleotide transitions or transversions. These are not unique paths, but show one set of consistent single-nucleotide steps through the conformational space.

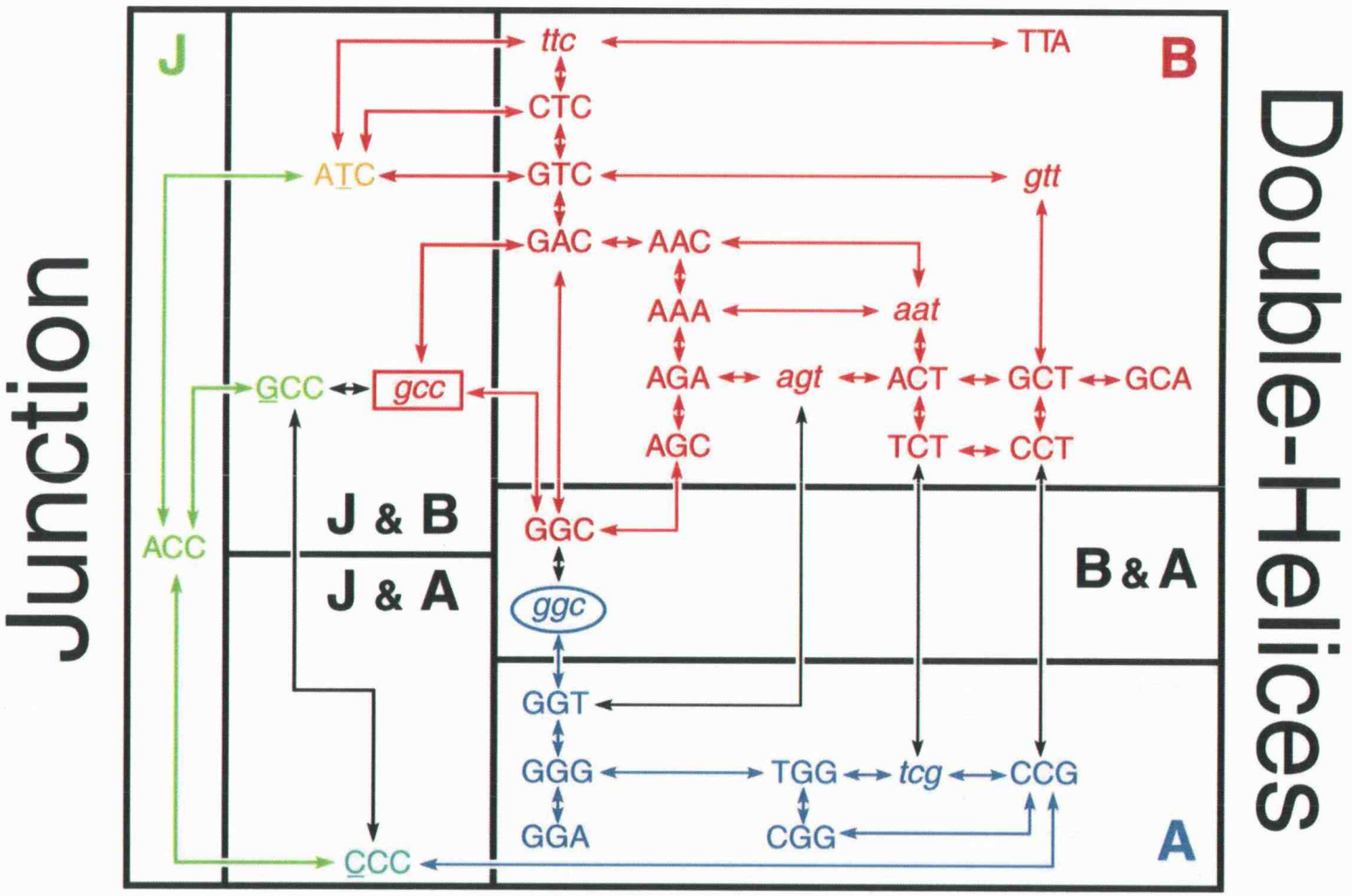
Junction
Double-Helices
J
B
A
J & B
J & A
B & A
ttc
TTA
CTC
ATC
GTC
gtt
GAC
AAC
AAA
aat
AGA
agt
ACT
GCT
GCA
AGC
TCT
CCT
GCC
gcc
ACC
GGC
ggc
GGT
GGG
TGG
tcg
CCG
GGA
CGG
CCC

Figure 5.2

5.5 Discussion

This study started with the goal of identifying trinucleotides in the d(CCnnn$N_6N_7N_8$GG) sequence motif that form four-stranded DNA Holliday junctions, and indeed the screen has done that. The trinucleotides $N_6N_7N_8$ = ACC, GCC, ATC, and CCC that are now identified as junction-forming are associated with specific interactions observed in the four-stranded complex. One unique aspect of the study is that each phase of the structure map is linked by amphimorphic sequences, which allows us to delineate the effects of single nucleotides at each position of the trinucleotide core on the stability of the junction relative to both A- and B-DNA duplexes. The parent ACC trinucleotide has been shown to stabilize DNA junctions in the presence of various cations (*75*), with G·A mismatches (*46*), in drug cross-linked constructs (*69*), and with the terminal C·G base pairs of decanucleotide motif replaced by T·A base pairs (*75*). Thus, ACC is defined as the most stabilizing of the junction forming trinucleotides. A common feature of all of the junction trinucleotides is the cytosine at the N_8 position. The current study shows that not all NNC type trinucleotides form junctions; thus, the cytosine at N_8 appears to be essential, but not sufficient to define a junction. Its importance can directly be attributed to the hydrogen bond from the cytosine N4 amino to the phosphate oxygen at the junction cross-over (Figure 5.3). We note, however, that this interaction can be partially replaced by an analogous Br···O halogen bond, as in the structure of the ACbr^{5}U junction (*76, 135*).

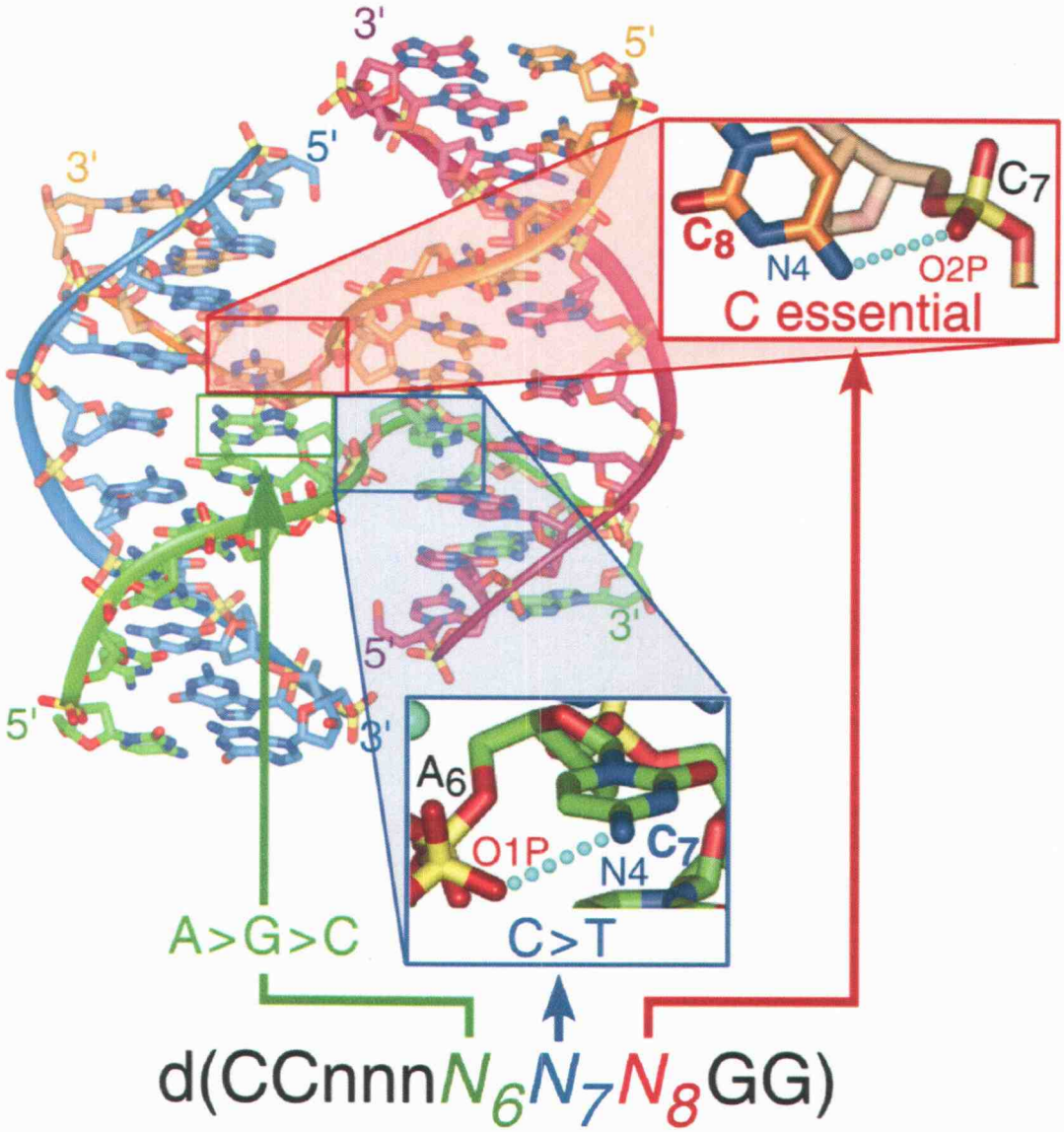

Figure 5.3. Correlating sequence effects to atomic interactions in junctions. The interactions that are identified as being important for fixing the junction in ACC are shown in the insets. General rules for junction forming sequences are noted in green, red, and blue for the nucleotides N_6, N_7, and N_8, respectively. The inset for the cytosine C_8 to phosphate of N_7 is rotated relative to the orientation of the overall structure.

In addition, all of the current junction-forming trinucleotides have a pyrimidine at the central N_7 position. Comparing ACC to the amphimorphic ATC trinucleotide, however, indicates that a cytosine is more effective than thymine at stabilizing the junction. There is a potential interaction between the N4 amino group of cytosine and a phosphate oxygen on the same strand but at the DNA duplex across the junction. This interaction, however, is longer (3.2-3.5 Å) than would be expected for an effective hydrogen bond. In addition, we see that the methyl group of the ATC structure is oriented directly towards and is within 4.1-4.5 Å of the oxygen atoms of this same phosphate. This observation suggests that the stabilizing effect of the pyrimidine base is primarily electrostatic, with the N4 amino group of cytosine being more effective than the methyl group of thymine as a counter to the phosphate oxygen. Thus, the general sequence rule is NCC > NTC in forming the junction for electrostatic reasons.

With a single exception, N_6 of the $N_6N_7N_8$ trinucleotide is a purine, with A > G in stabilizing the junction. This order is evident from the observation that ATC is amphimorphic and capable of forming a junction, but GTC forms B-DNA. The exception to a purine at N_6 is CCC. Again, the results indicate that a cytosine at N_6 is less stabilizing to the junction than either A or G through the argument that CCC is amphimorphic and CTC is B-DNA. Thus, the series at the N_6 nucleotide is A > G > C.

It is interesting that the amphimorphic CCC sequence crystallizes as a duplex in the A-form, but as a four-stranded junction with arms that adopt the B-DNA structure. This suggests that, at least for DNA, the four-stranded junction favors B-type double-helices even if the sequence has a strong propensity for A-DNA. Again,

this results from the hydrogen bonding interaction of the cytosine at N_8 that is required to stabilize the junction in the IR motif (this interaction would not be available with deep major groove that one would expect with A-DNA arms).

The conformation map shows that A-DNA is associated with the trinucleotide motifs GGN, NGG, and CC(C/G). B-DNA is favored as individual C/G base pairs of these A-DNA triplets are replaced by T/A base pairs (Figure 5.2). When applying previous trinucleotide rules to distinguish A- from B-form DNA, those derived from calculations of hydrophobic surfaces (*128*) correctly predicted 16/24 (67%) of the sequences crystallized as A- or B-DNA, while those from experimental alcohol titrations (*127*) correctly predicted 21/28 (75%) sequences (not all trinucleotides are represented in the respective scales). The relatively poor showings reflect the fundamental differences between these two former methods for predicting A- and B-DNA and the design of the current experiment. The hydrophobicity scale was derived from a structural dataset that was highly variable in how the sequences were crystallized (relying on the available structures at the time) and which included a large number of lattice distorted A-DNAs. On the other hand, the alcohol titration scale was determined spectroscopically as the amount of trifluoroethanol required to induce a B- to A-DNA transition. The current study, in contrast, relates the two conformations through sequences that are explicitly determined from nearly identical crystallization solutions. Thus, structures derived by the crystallographic screen provide a potential means to derive a set of rules to predict the sequence formation of A- and B-DNA from a consistent dataset.

In addition to the structural map, the structures from this crystallographic screen also allow us to relate sequence to conformational variations that are important for recognition and function of each structural class (9 such parameters are summarized in Table 5.2). A comparison of the base pairs and base step parameters of the structures in this study shows that A-DNAs are conformationally homogenous compared with B-DNA (Supplemental Table 3 in Appendix). We would not draw this conclusion from a comprehensive analysis of all A-DNA crystal structures available in the NDB (*74*). Although the A-DNA structures detailed here show some sequence dependent variations in their double-helical conformations, they do not vary dramatically from the canonical form. We attribute this to the consistency in the current dataset of A-DNA structures, where solution conditions and sequence end-effects have been controlled.

The B-DNA duplexes and junction arms show similar sequence-dependent conformational variation, with the mean and degree of variability of each helical parameter being very similar between the two structural forms. This supports the model that the arms of the junction mirror the properties of B-DNA in general (*73*). When the helical parameters for the amphimorphic sequence ATC are compared between the junction and B-DNA structures, the helical arms of the junction are seen to be more typical, in many respects, of standard B-DNA than the actual B-DNA structure for this sequence. For example, the B-form of ATC shows significant buckling within and slide and roll between stacked base pairs from the average B-DNA duplex of the dataset. In contrast, the arms of the junction more or less fall into the norms of the B-DNA structures, with the exception that they are slightly

Table 5.2. Helical parameters for structures from the crystallographic screen of the sequence motif d(CCnnn$N_6N_7N_8$GG). The rotational and translational parameters that characterize the helical conformations of nucleic acid structures (as defined in (109), and calculated by the program 3DNA (87)) are compared for the mean values of all base pairs (<Base pairs> (standard deviations)) and as means averaged across the structures (<Structure> (standard deviation of mean)) for structures in the screen that are B-DNA, four-stranded junctions, and A-DNA.

Structures	Rotational Parameters (°)					Translational Parameters (Å)			
	Helical Twist	Propeller Twist	Tilt	Roll	Buckle	Rise	Slide	X-displacement	Z_p
B-DNA									
<Base Pairs>	34.7 (15.5)	-12.0 (8.1)	-0.62 (10.9)	1.74 (7.98)	-0.23 (8.57)	3.30 (0.4)	0.66 (1.02)	0.53 (1.56)	-0.67 (0.59)
<Structures>	35.6 (1.2)	-11.6 (2.8)	-0.05 (0.52)	2.18 (1.51)	-0.31 (2.62)	3.31 (0.06)	0.71 (0.40)	0.49 (0.48)	-0.68 (0.24)
GCA	38.2[1]	-12.9	-0.43	-0.12[1]	-4.94[1]	3.39	1.32[1]	1.43[1]	-0.9
Junctions									
<Base Pairs>	37.5 (4.3)	-11.2 (8.1)	-0.14 (4.69)	1.90 (4.47)	-1.16 (6.09)	3.40 (0.23)	1.63 (1.11)	2.11 (1.89)	-0.97 (1.07)
<Structures>	37.5 (0.33)	-11.2 (3.0)	-0.14 (0.65)	1.90 (0.52)	-1.16 (2.93)	3.40 (0.03)	1.63 (0.11)	2.11 (0.13)	-0.97 (0.14)
A-DNA									
<Base Pairs>	30.4 (4.2)	-7.82 (8.47)	0.62 (3.91)	6.88 (5.69)	-0.57 (7.78)	3.30 (0.24)	-1.74 (0.31)	-4.47 (1.25)	2.27 (0.39)
<Structures>	30.4 (0.8)	-7.82 (3.57)	0.62 (0.67)	6.88 (1.71)	-0.57 (2.42)	3.30 (0.06)	-1.74 (0.15)	-4.47 (0.25)	2.27 (0.15)

Table 5.2 (Continued)

Amphimorphic Junction Structures									
ATC (B-DNA)	36.8	-17.3	-0.06	-0.08	7.35	3.34	-0.07	-0.16	-0.26
ATC (Junction-LS)	37.8	-13.0	0.47	1.53	0.91	3.34	1.63	1.99	-1.08
ATC (Junction-HS)	37.2	-16.5	0.20	2.28	0.37	3.38	1.77	2.11	-1.15
CCC (Junction)	38.0	-14.2	-1.28	2.15	-7.19	3.37	1.60	2.60	-0.97
CCC (A-DNA)	30.5	-8.09	-0.29	7.1	0.83	3.24	-1.76	13.66	2.38

[1]Values that fall at least 1 S.D. outside the mean of the average structural class.

overwound, show a positive slide, and a displacement of the phosphate group (Z_P) that is more negative than B-DNA duplexes. It is clear from an analysis of the overall dataset, however, that these features are particular to the cross-over of the DNA strands of the junctions.

There are, however, specific sequences that fall well outside the standard conformational variations of the B-DNA double-helix. The overall structure of GCA, for example, deviates significantly from standard B-DNA, with 5 of the 9 parameters in Table 5.2 falling at least 1 S.D. from the mean values for the overall structures of this class. In particular, the duplex is highly overwound (at ~9.4 base pairs per turn as estimated from the helical twist), with the stacked base pairs showing significant roll, buckling and slide. The experimental design of the study indicates that these variations are defined by the sequence rather than by the crystal lattice or crystallization conditions. Interestingly, the sequence TGCGCA is the repeating binding motif for at least one eukaryotic promoter (*136*) and, therefore, such conformational perturbations may play a role in protein recognition.

In summary, a crystallographic dataset of DNA structures is being assembled from a well-defined sequence motif and a relatively consistent set of crystallization solutions. This set allows us to correlate sequence and environment with structural classes and conformational variability within structural classes. Thus, it is clear that the strategy of broadly sampling structures by crystallographic screening of a specific sequence motif directly defines the effects of sequence on macroscopic behavior at the level of detailed molecular interactions. Although currently limited to DNA structures of self-complementary sequences, the results of the study shows that the basic premise

is correct: if the structures of all permutations of a molecule can be determined, the sequence effects on the overall structure and the details of their conformation are explicitly known.

5.6 Acknowledgements

The authors thank members of the Karplus lab and Dr. J. Vargason for help with crystallographic problems. This work was supported by grants from the National Institutes of Health (R1GM62957A) and the National Science Foundation (MCB0090615). X-ray diffraction facilities were supported by the Proteins and Nucleic Acids Facility Core of the Environmental Health Sciences Center at OSU (NIEHS ES00210) and by the Murdock Charitable Trust. Several DNA oligomers were synthesized in the Center for Gene Research and Biotechnology at Oregon State University. The BioCARS Sector 14 beamline is supported by the National Institutes of Health, National Center for Research Resources under grant number RR07707, and the Advanced Photon Source supported by the U.S. Department of Energy under Contract No. W-31-109-Eng-38. The Advanced Light Source at Lawrence Berkeley National Laboratory is supported by the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

Chapter 6

Cation Effects on the Geometry and Formation of the DNA Junction in d(CCGATATCGG)

Franklin A. Hays, Dustin Raup and P. Shing Ho

Formatted for Submission

6.1 Summary

The folding and function of four-stranded DNA junctions is highly dependent on cation concentration. In this study, we present crystal structures of the inverted repeat decamer DNA sequence d(CCGATATCGG) crystallized in three different concentrations of calcium (II) chloride. Under high and intermediate concentrations of calcium ions this sequence forms two distinct conformations of four-stranded stacked-X Holliday junction with an antiparallel alignment of DNA strands. Surprisingly, this sequence also crystallized as resolved B-DNA duplexes under intermediate and low divalent ion concentrations. Comparison of the folded junction structures shows a translation and rotation of each duplex arm of the junction with decreasing concentration of calcium chloride. This displacement of the helical junction arms generates a more symmetric central core that increases in volume and contains fewer stabilizing hydrogen bonding interactions. Thus, these structures provide a quantitative view of the affects of calcium concentration on junction geometry that lead, at low calcium concentrations, to resolved B-DNA duplexes for the same decamer DNA sequence.

6.2 Introduction

Four-stranded DNA (Holliday) junctions are now known to be involved in a growing number of cellular processes that generally serve to protect or expand the

coding potential of organisms. These processes include re-initiation of stalled replication forks, DNA repair, viral integration and genetic recombination. A common feature in these disparate biological functions is the conformational variability of the junction structure which can adopt different global conformations in a sequence and solution dependent manner. Understanding this variability is instrumental to gaining insights into how junctions function in biological systems.

Holliday junctions are generally formed by the reciprocal exchange of DNA strands between two adjacent B-DNA duplexes. A number of experimental approaches have demonstrated that the global junction conformation is highly dependent on cation concentration (*26, 27, 66, 137*). Under low cation concentrations the negatively charged phosphodiester DNA backbone produces a square-planar structure containing four duplex arms oriented towards the corners of a box ("open-X" conformation, Figure 6.1b); interestingly, recent molecular dynamics simulations have suggested the possibility of an additional tetrahedral conformation (*138*). As shielding of the negative phosphodiester backbone occurs with increasing cation concentration the arms collapse into a coaxially stacked compact form commonly referred to as the stacked-X structure (Figure 6.1a and 6.1c). Recent studies using single-molecule FRET have provided a new view of the junction structure as being very dynamic and undergoing a range of conformational transitions (*25*). These transitions are dependent on nucleotide sequence within the central crossover and cation concentration surrounding the junction structure. Though these elegant experiments have provided a wealth of information on junction dynamics they have not provided

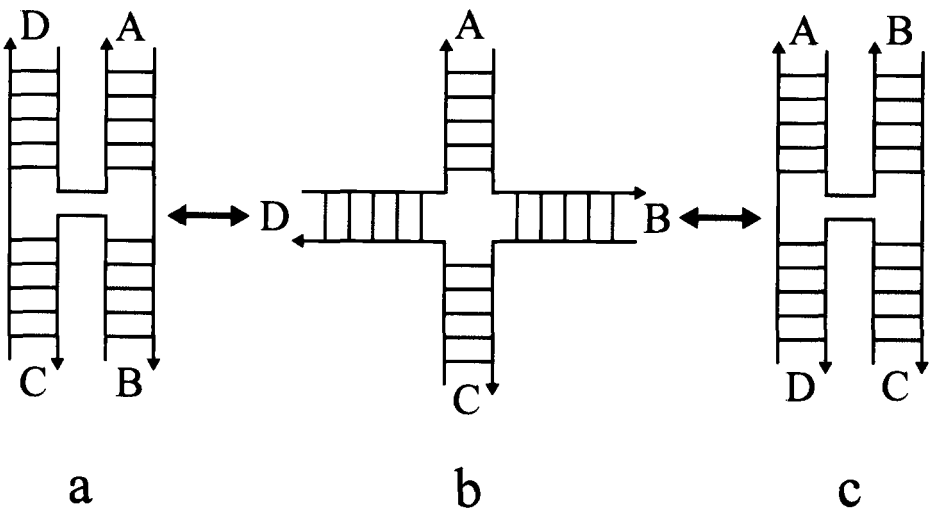

Figure 6.1. Conformational variability of DNA Holliday junctions. The four-stranded DNA Holliday junction can undergo a range of transitions in conformation in a salt and sequence dependent manner. Under low salt conditions the junction favors the square-planer open-X conformation (b) with the arms oriented towards the four corners of a square. At higher salt concentrations, where shielding of the negative phosphodiester backbone can occur, the duplex arms of the open-X structure (A, B, C and D) coaxially stack to form the stacked-X structure (a and c).

insights into the atomic details of what occurs during transitions between stacking conformers - details such as does varying cation concentration influence nucleotide geometry within the stacked-X junction core? Atomic details of junction structures are biologically important since the stacked-X form is thought to be recognized by certain junction resolving enzymes, such as Hjc or T4 endonuclease VII, and the dynamic nature of the junction may play a role in its recognition by cellular recombination and DNA repair enzymes. Indeed, branch migration has been shown to be inhibited by concentrations of divalent cations known to stabilize the stacked-X junction conformation (*81*).

Relatively recent advances using crystallographic methods are providing a great deal of information on the atomic structure of 4H-junctions which have been observed in both the presence (*3, 29*) and absence of bound proteins (*46, 47, 75, 76*). The latter, protein free forms all crystallize in the more compact stacked-X conformation with an antiparallel alignment of DNA strands formed by two continuous and two non-continuous (crossover) strands. Arrangement of the DNA strands in this manner produces two distinct arms within the pseudocontinuous B-DNA duplexes of the junction structure (each duplex is composed of a short four base pair arm sitting on top of a longer six base pair arm) which are joined by the central crossover region (Figure 6.1). Surprisingly, junction structures have now been solved using a variety of decamer sequences containing true inverted-repeat patterns (*47, 75*), base mismatches (*46*), cytosine methylation (*48*), psoralen crosslinking (*69*), and halogen bonding interactions (*76, 135*). All of these structures have a common general sequence motif in the central crossover region with a series of highly

conserved inter- and intra-strand hydrogen bonding interactions that stabilize the junction conformation (*93*). Previous solution and structural studies have shown that disruption of these core interactions, which reside within the major groove face of the junction structure, have a significant effect on junction conformation (*43, 93*). Most of these interactions likely serve to shield the large density of negative charge within the junction crossover region resulting from the close proximity of backbone phosphate oxygens of the exchanging DNA strands. Substitution of these partially negative phosphate oxygen atoms with neutral methyl groups results in a loss, or significant reduction, in the need for cations to fold the stacked-X structure (*139*).

A surprising feature of the stacked-X junction structures solved by crystallographic methods is the lack of divalent cations which have been observed around the central core of the junction, a region previously described as containing a very high electrostatic potential (*140, 141*). Brownian dynamics simulations proposed a site-specific metal ion binding adjacent to the central core of the junction within the minor groove face (*142*) which was later observed to contain a sodium ion in the junction crystal structure of d(TCGGTACCGA) (*75*). Additionally, several structures contain a sodium ion within the central cavity of the junction crossover which may help shield the electrostatic potential within the core (*47, 76*). Although site-specific binding has been proposed for specific ions (*142, 143*) the general consensus seems to be that group I and group II metal ions interact with the junction center in a rapidly exchanging and noncooperative manner (*63, 137*) which affects the energetic barriers to conformer transitions (*25*). This lack of site-specific divalent ion coordination sites within the central crossover region of junction structures may lend credence to the

proposition of "atmospheric binding" in which a cloud of highly mobile ions serve to shield the negative charge density within the junction center (*144*).

The question is then raised as to whether or not cation concentration is having the same effect on the sequences used for crystallographic studies as those used in previous solution experiments. In the present study we have solved three structures of the inverted-repeat decamer DNA sequence d(CCGATATCGG) which crystallizes as resolved B-DNA duplexes or 4H-junction in a salt dependent manner. These structures provide the first observation at atomic resolution of the effects of decreasing cation concentration on the structure of a four-stranded DNA Holliday junction that leads ultimately to resolved B-DNA duplexes of the same sequence.

6.3 Results

The crystal structures presented here demonstrate how decreasing calcium concentration induces a conformational transition between a four-stranded Holliday junction and resolved B-DNA duplexes for the inverted-repeat sequence d(CCGATATCGG). Under relatively high salt conditions this sequences crystallizes as a four-stranded junction and, as salt concentration is decreased, it forms resolved B-DNA duplexes. Interestingly, we identified a junction under intermediate salt conditions that was conformationally perturbed relative to the junction crystallized in high salt solutions. This intermediate structure is very informative since the duplex arms of the junction have translated along their respective helix axes and twisted about

Table 6.1. Data collection and refinement statistics for three conformations of the sequence d(CCGATATCGG) ad B-DNA (ATC-B) and as junctions at low salt (ATC-LS) and high salt (ATC-HS).

	ATC-B	ATC-LS	ATC-HS
Data			
Unit Cell			
a (Å)	35.2	64.9	65.0
b (Å)	39.1	24.0	24.6
c (Å)	33.0	40.3	37.5
β angle (°)	-	118.69	112.09
Space group	$P2_12_12_1$	*C2*	*C2*
Total reflections	47521	94804	86918
Unique reflections	5378	4894	4063
Resolution (Å)	16.0-1.61	12.70 - 1.65	30.1 - 1.84
Completeness(%)[a]	98.9 (90.3)	90.2 (75.3)	98.6 (96.9)
I/σ(I)[a]	6.5 (2.3)	14.1 (2.9)	15.0 (1.1)
R_{merge}(%)[a,b]	5.1 (33.0)	5.7 (28.3)	6.1 (62.6)
Refinement			
Resolution	16.0 - 1.61	12.70 - 1.65	30.1 - 1.84
R_{cryst} (R_{free})(%)[c]	26.0 (27.0)	23.1 (27.5)	20.2 (23.2)
DNA atoms	404	404	404
Solvent atoms	133	119	172
Average B-Factors (Å^2) DNA atoms (Solvent atoms)	7.2 (10.6)	9.1 (15.5)	7.5 (15.3)
RMSD bond lengths (Å)[d]	0.014	0.010	0.010
RMSD bond angles (°)[d]	2.5	1.8	1.7
Structural conformation	B-DNA	4H-Junction	4H-Junction
[$CaCl_2$] in crystallization solution	5 mM	55 mM	110 mM

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

[b]$R_{merge} = \Sigma_{hkl} \Sigma_i | I_{hkl,i} - \langle I \rangle_{hkl} | / \Sigma_{hkl} \Sigma_i | I_{hkl,i} |$ 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.

[c]$R_{cryst} = \Sigma_{hkl} | F_{obs} - kF_{calc} | / \Sigma_{hkl} |F_{obs}|$. $R_{free} = R_{cryst}$ for 10% of reflections that were not used in refinement (*104*).

[d]Root mean square deviation of bond lengths and angles from ideal values.

the junction core, both of which are reflected by a decrease in the interduplex angle, when compared to the high salt junction. These perturbations indicate, as expected, that a decrease in divalent cation concentration perturbs junction conformation and leads to a more symmetric structure. This tendency to become more symmetric is reasonable considering the requirement of an open-X junction intermediate for the formation of resolved B-DNA duplexes from a stacked-X junction. Using both junction forms of this sequence motif we can quantitate the cation-dependent conformational transition between the stacked-X and open-X junction forms that lead, ultimately, to resolved B-DNA duplexes.

6.3.1 Structure of d(CCGATATCGG) as resolved B-DNA duplexes. In 5mM $CaCl_2$ the sequence d(CCGATATCGG) crystallized as resolved B-DNA duplexes (ATC-B structure) with one duplex in the crystallographic asymmetric unit (Figure 6.2). To verify that this crystal form was indeed B-DNA and not a four-stranded junction we created a simulated annealing omit map for the phosphate group of N_7, the branching position for all current crystal structures of unmodified stacked-X junctions, which showed clear Fo-Fc density for the assigned T7 phosphate position indicating that the phosphodiester backbone continues along the duplex structure instead of branching over to a neighboring duplex. The ATC-B structure has a number of similarities to the B-DNA forming sequence d(CCTTTAAAGG), described previously by Dickerson and coworkers (BD0051, (*145*)), which was used to solve the present structure via molecular replacement (RMSD for like non-hydrogen atoms between ATC-B and BD0051 is 0.41 Å). As previously observed for AT rich B-DNA

Figure 6.2. Structure of the d(CCGATATCGG) sequence as a resolved B-DNA duplex in the $P2_12_12_1$ crystal form. The final 2*Fo*-*Fc* electron density map (contoured at 1σ) depicting the contents of the asymmetric unit (two strands, rendered as sticks with coloring according to atom type). A red or blue cylinder connects the phosphorous atoms along the backbone of each strand with each nucleotide along the blue strand labeled by base type and number in the 5' to 3' direction. The bottom panel shows the region where the DNA backbone would deviate between two adjacent duplexes if this was a four-stranded junction. The symmetry related strand is rendered as grey sticks with no clear electron density between the two strands. This figure was created with Bobscript (*89*) and rendered in Raster3D (*112*).

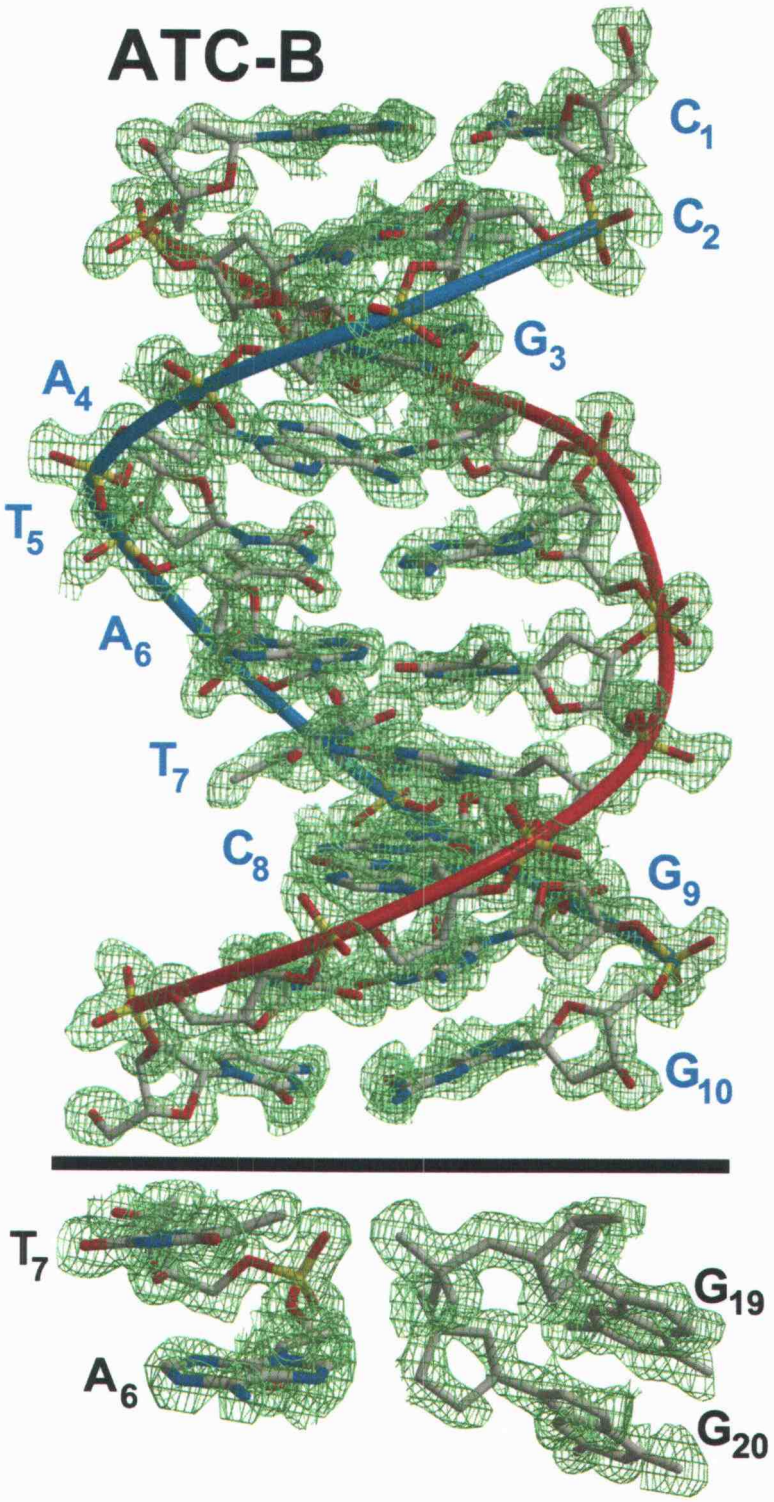

Figure 6.2

sequences (*146, 147*), ATC-B contains a semi-continuous "spine of hydration" within the minor groove of the duplex structure, arranged in the classic "zig-zag" pattern between hydration shells, with solvent waters interacting primarily with the O2 atoms of pyrimidines and N3 atoms of purines. No divalent or monovalent cations were identified in the solvent model of this structure, which has approximately thirteen solvent atoms per base pair (133 preferred hydration sites were identified).

The ATC-B structure has a significant widening of the minor groove (average phosphate-corrected minor groove width is 5.34 Å), with the widest point at the T_5 nucleotide (7.9 Å), that has previously been attributed to deformability of T·A base steps resulting in a bending of the DNA duplex and a concomitant narrowing of the major groove (*148*). This widening of the minor groove does not translate to significant backbone perturbations in the form of BII phosphate conformations or uncommon sugar puckers. The phosphate group of A_6 is the only BII (*113, 116*) conformation present in the ATC-B structure resulting from crystal packing interactions between the O2P atoms of G_{20} and A_6. Likewise, backbone sugar puckers found in ATC-B (C2'-endo, C1'-exo, O1'-endo and C3'-exo) are typical for standard B-DNA duplexes with no C3'-endo family puckers present (*149*). Analysis of individual base pair and base step geometry within the ATC-B structure shows many of the expected characteristics for T·A rich central regions of DNA structures (increased helical twist, propeller twist, roll, slide, etc.) and has been discussed in great detail elsewhere (*149-151*) for similar structures.

6.3.2 Overall Structures of d(CCGATATCGG) as DNA Holliday junctions. Two conformationally unique DNA Holliday junction structures of the sequence

d(CCGATATCGG) were determined from solutions containing 55 mM (ATC-LS structure) and 110 mM (ATC-HS structure) $CaCl_2$ (Figure 6.3). The ATC-LS and ATC-HS structures will be discussed relative to each other since they are very similar in terms of nomenclature and general arrangement of the DNA strands (Figure 6.3). These right-handed stacked-X junctions have an antiparallel alignment of the DNA strands (RMSD for like non-hydrogen atoms between ATC-LS and ATC-HS junctions is 0.57 Å). Each junction is formed by a 2-fold rotation of the two unique strands (one noncrossover and one crossover) in the crystallographic asymmetric unit to generate a four-stranded DNA Holliday junction with pseudo-continuous duplex arms on each side of the central crossover. These adjoined duplexes contain two longer six base pair arms (from the C_1 to A_6 nucleotides of the crossover arm paired with G_{20} to T_{15} of a noncrossover arm) on top of shorter four base pair arms (pairing T_7 to G_{10} of a crossover strand with A_{14} to C_{11} of a noncrossover strand) with the crossover occurring between the A_6 and T_7 nucleotides of the inside strands (Figure 6.3). The central crossover region is composed of the nucleotides A_6, T_7 and C_8 of the noncontinuous DNA strands.

6.3.3 Comparison of Junction Geometry. As evident from above, analyzing the global conformation of these four-stranded junctions is a fundamental part of comparing numerous structures and gaining a better understanding of how variability in conformation may play a role in biological processes. Recent efforts to explicitly define junction geometry (*48, 110*) have led to the development of quantitative parameters which describe the spatial arrangement of the junction duplexes, relative to

Figure 6.3. Structures of the sequence d(CCGATATCGG) as four-stranded Holliday junctions in the *C2* crystal form. The final 2*Fo*-*Fc* electron density (contoured at 1σ) depicting the contents of the asymmetric unit (two strands, one crossover and one non-crossover) rendered as sticks with coloring according to atom type (panels created with Bobscript (*89*) and rendered in Raster3D (*112*)). The phosphodeoxyribose backbone for each strand is rendered as a continuous ribbon. Nomenclature, backbone coloring and orientation for each junction structure is the same as the central diagram with the full four-stranded structure formed by a crystallographic 2-fold symmetry axis in the center. The ATC-LS junction was crystallized in the presence of 55 mM $CaCl_2$ while ATC-HS crystallized in the presence of 110 mM $CaCl_2$.

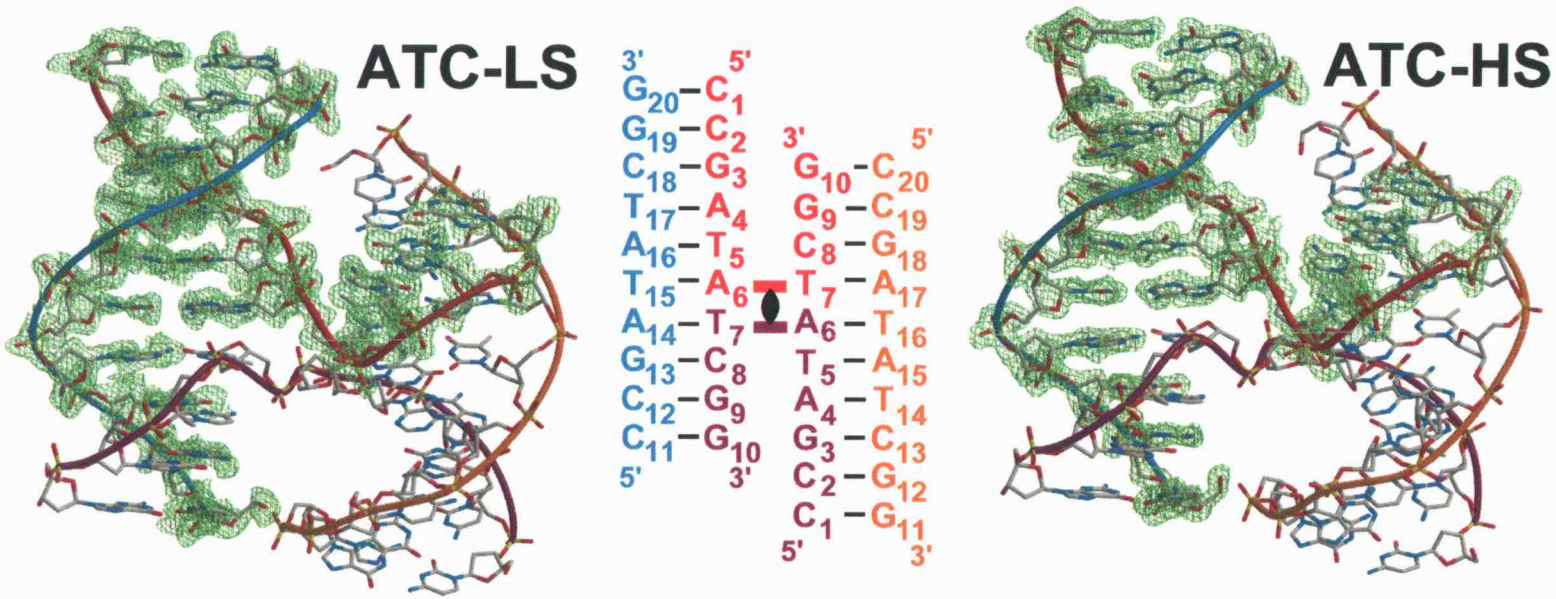
ATC-LS
3' 5'
G20 – C1
G19 – C2
C18 – G3
T17 – A4
A16 – T5
T15 – A6
A14 – T7
G13 – C8
C12 – G9
C11 – G10
5' 3'
3' 5'
G10 – C20
G9 – C19
C8 – G18
T7 – A17
A6 – T16
T5 – A15
A4 – T14
G3 – C13
C2 – G12
C1 – G11
5' 3'
ATC-HS

Figure 6.3

the central core region, in three-dimensional space. J_{slide} defines the translational movement of the duplexes along their helix axes, J_{roll} is rotation of the duplexes relative to the junction center, and J_{twist} defines the twist of each duplex relative to the resolving plane (*110*). Taken together, these parameters affect the angle relating each duplex end to the junction center, generally referred to as the interduplex angle (IDA). Comparison of the IDA's between ATC-LS (IDA = 69.0°) and ATC-HS (IDA = 69.9°) indicates a small conformational difference between the two junction forms. Indeed, a closer inspection of each conformational parameter for ATC-LS (J_{slide} = 0.39 Å, J_{twist} = 41.0° and J_{roll} = 149.7°) and ATC-HS (J_{slide} = 1.00 Å, J_{twist} = 44.0° and J_{roll} = 148.7°) demonstrates that the duplex arms have translated towards the six base duplex ends and twisted towards a more parallel junction structure with decreasing salt concentration. Additionally, the major groove face of the ATC-HS junction has become smaller relative to ATC-LS. These perturbations to the overall junction geometry produce a more symmetric, planar, structure as the salt concentration decreases and a more compact junction core at higher salt concentrations.

In addition to the above perturbations there are a number of additional differences in structure when comparing the low and high salt junctions to each other and the ATC-B structure. The average minor groove width of both junctions (4.54 Å for ATC-LS and 4.26 Å for ATC-HS) is considerably narrower then the minor groove of ATC-B (7.02 Å) between the fourth and sixth base steps of the duplex arms. Though the minor groove in this region is very similar for both junctions there is a noticeable difference in major groove width for the same region, ATC-HS has an

average major groove width between the $A_4 \cdot T_7$ and $A_6 \cdot T_{15}$ base steps of 10.63 Å compared to 11.20 Å for ATC-LS.

6.3.4 Comparison of Junction Crossover Regions. Previous results have demonstrated that specific hydrogen bonding interactions within the junction core are important factors in determining the global conformation and formation of these stacked-X junctions. These central core interactions include (i) a direct hydrogen bond between the C_7 phosphate oxygen and extracyclic amine of C_8 (N_7-C_8 interaction), (ii) a solvent-mediated interaction between the outside G_3 keto oxygen and A_6 phosphate oxygen (G_3-A_6 interaction), and (iii) a direct hydrogen bond between the O1P phosphate oxygen of A_6 and the N4 extracyclic substituent of N_7 (*47, 93*). Consistent with these observations we observe the direct N_7-C_8 hydrogen bonding interaction (2.85 Å in ATC-LS vs. 2.61 Å in ATC-HS) in both junction structures (Figure 6.4). Both structures have additional solvent waters within the G_3-A_6 and A_6-T_7 stabilizing interactions, similar to the previously solved ACm^{5}C junction structure (*48*). It is interesting to point out that the C5 methyl group of T_7 for both ATC-LS and ATC-HS is pointed directly at the A_6 phosphate group which may help shield negative electrostatic potential within the core (Figure 6.4). A notable difference between the two junction structures is the presence of additional compensating interactions within the ATC-HS junction core (such as the N_6-N_6 interaction in Figure 6.4). These interactions are more extensive within the ATC-HS structure and have been observed in other conformationally perturbed junction structures such as ACbr^{5}U in Chapter 3 (*76*).

Figure 6.4. Stabilizing interactions within the central core of ATC-LS and ATC-HS. The central T_5, A_6, T_7 and C_8 nucleotides are shown for each strand of the central crossover for the ATC-LS and ATC-HS junction structures (numbering of nucleotides on the symmetry related strand is incremented by ten). DNA is rendered as ball and sticks with coloring according to atom type (blue = nitrogen, red = oxygen, grey = carbon and yellow = phosphorous). The phosphodeoxyribose backbone is rendered as a continuous ribbon connecting phosphorous atoms along the same strand. The N_6-N_7 interaction is shown at the top of each panel with the corresponding N_7-C_8 interaction shown at the bottom (hydrogen bonding distance in parentheses). The volume of the central cavity is noted for each core and was calculated as described in the text. This figure was created with Bobscript (*89*) and rendered in Raster3D (*112*).

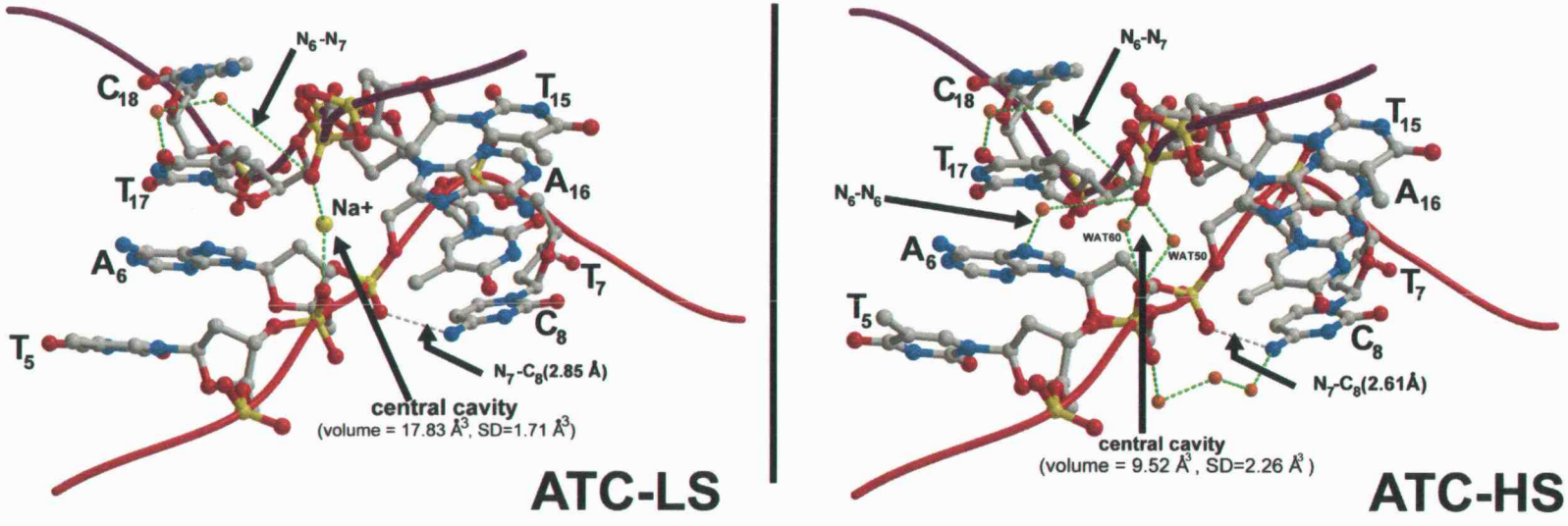
N6-N7
C18
T15
T17
A16
Na+
A6
T7
C8
T5
N7-C8(2.85 Å)
central cavity
(volume = 17.83 Å3, SD=1.71 Å3)
ATC-LS
N6-N7
C18
T15
T17
A16
N6-N6
WAT60
WAT50
A6
T7
T5
C8
N7-C8(2.61Å)
central cavity
(volume = 9.52 Å3, SD=2.26 Å3)
ATC-HS

Figure 6.4

In addition to these hydrogen bonding interactions, which have been identified in some form in every stacked-X structure solved to date (*93*), there is a central cavity located between the backbone phosphate atoms of A_6 that contained a sodium ion in the parent ACC junction structure (*47*). Not surprisingly, the central cavity of ATC-LS contains a sodium ion in the same location while the central cavity of ATC-HS contains a partially occupied water molecule adjacent to the crystallographic two-fold symmetry axis (the ATC-LS sodium is directly on the crystallographic two-fold). This displacement may result from a compression, or pinching, of the central cavity in the ATC-HS structure as indicated by the distance between A_6 phosphate oxygens (distance between phosphate oxygens is equal to 5.9 Å in ATC-HS and 6.7 Å in ATC-LS). Indeed, the central core region of the ATC-HS structure is more compact compared to ATC-LS. The average phosphorous-phosphorous distance within the junction core is 0.54 Å greater in ATC-HS then in ATC-LS (8.73 Å vs. 8.19 Å using $T_5A_6T_7$ phosphorous atoms adjacent to the cavity). If the average distance between backbone atoms within the junction core decreases with increasing salt concentration then one would expect a corresponding decrease in the size of the central cavity that has been previously described (*48*). Indeed, the volume of the central cavity, calculated using the program VOIDOO (*91*) (standard van der Waals radii for DNA atoms; multiple rounds calculated with different molecule orientations, probe radius, grid spacing, and grid-shrink factor), within the ATC-LS structure is almost twice the size of the same cavity within the ATC-HS structure (9.52 $Å^3$ with SD = 2.26 $Å^3$ for ATC-HS vs. 17.83 $Å^3$ with SD = 1.71 $Å^3$ for ATC-LS). Is the compression of this cavity due to local distortions in the backbone

atoms of the core or is it related to the global conformational change of the junction? This cavity, formed by the $T_5A_6T_7$ nucleotides, resides on the major groove face of the junction structure. If it was simply local distortions then we could expect an increase in the major groove width of the ATC-HS structure relative to ATC-LS, assuming compaction is not related to a global conformational change of the junction (i.e. a change in the orientation of the duplex arms relative to each other). Instead of an increase, we observe a 0.68 Å decrease in major groove width of ATC-HS (10.61 Å for ATC-HS vs. 11.29 Å for ATC-LS) which indicates that conformational changes must play some role in compressing the central cavity.

6.4 Discussion

In this study, we determined the salt-dependent single crystal structures of the sequence d(CCGATATCGG) from two crystal forms yielding both four-stranded DNA Holliday junctions, in a stacked-X antiparallel conformation, and B-DNA duplexes. The $N_6N_7N_8$ trinucleotide core of the general inverted-repeat sequence d(CCnnn$N_6N_7N_8$GG) has been previously described and contains a series of inter- and intra-strand hydrogen bonding interactions that stabilize the stacked-X junction conformation (*47, 93*). The current junction structures represent the first from sequences that do not contain a common cytosine nucleotide in the N_7 position of this core region. Furthermore, previous studies have demonstrated that divalent cations can have a significant effect on global junction conformation (*26, 38, 66, 137*) with the square-planar conformation being favored in low cation concentrations while the

stacked-X form predominates under conditions that provide shielding of the negatively charged phosphodiester backbone (Figure 6.1). Subsequent studies on crystal structures of stacked-X junctions derived from a number of different sequences has yielded a set of quantitative parameters describing the conformation of the stacked-X junction (*110*)). A unique aspect of the present study is that a comparison of the two d(CCGATATCGG) junction forms (ATC-HS and ATC-LS) revealed a salt-dependent change in global conformation producing a less compact core and more symmetric junction structure as the salt concentration is decreased. We propose this change in conformation is associated with the transformation of the stacked-X junction towards the open-X intermediate which can undergo spontaneous branch migration producing, as we also observe for this sequence, resolved B-DNA duplexes. Thus, these crystal structures provide the first detailed view of junction conformational changes resulting from decreasing cation concentration which, eventually, leads to the duplex form of the same sequence.

The current studies were initiated by a systematic crystallographic screen of the sequence dependence on junction formation where we expanded the core sequence motif to include the ACC, GCC, CCC and ATC trinucleotides (*152*). The ATC junctions are the only currently available structures where the common cytosine in the N_7 position of the junction core has been replaced by another base, in this case a thymine nucleotide. Incorporation of a thymine into the central core drives this sequence motif into an amphimorphic class that forms two DNA conformations, unlike the parent ACC junction that has only been observed as a DNA Holliday junction (*152*). If electrostatic interactions are a significant driving force in junction

folding then the presence of a thymine in the N_7 position may be less efficient in stabilizing the electrostatic charge from the A_6 phosphate atoms when compared to the extracyclic N4 amine of a cytosine base. Indeed, in every current ACC junction structure there is a long range, or solvent mediated, hydrogen bonding interaction between the N4 substituent of C_7 and the A_6 phosphate oxygen. In the ATC junctions this shielding may occur as the result of the C5 methyl groups of T_7 on the crossover strands which are oriented directly towards these A_6 phosphate oxygens. Furthermore, A·T base pairs are known to produce more significant local distortions when compared to G·C base pairs by increasing parameters such as propeller twist and helical twist as well as a concomitant narrowing of the minor groove. This becomes more evident by the incorporation of a thymine residue into the N_8 position of the parent ACC core (ACT sequence) which has only been observed as resolved B-DNA duplexes (*76*). Overall the trinucleotide core interactions within ATC-LS are very similar to the parent ACC junction structure. The ATC-HS structure has a number of additional compensating interactions which have been observed previously in the conformationally perturbed d(CCAGTACbr^{5}UGG) junction structure (*76*). Previous studies have also shown the importance of sequence affects in stabilizing the stacked-X form or dictating a preference for a specific stacking conformer (*25, 38, 40, 43*). This interplay between cation neutralization of highly negatively charged regions and specific intra- and inter-strand hydrogen bonding interactions between base substituents may have a concerted affect in determining junction conformation. Thus, the current structures provide a means to not only expand the trinucleotide core

sequence motif but also allow us to refine our understanding of how interactions within the central junction core stabilize and influence junction conformation.

The conformation and local geometric characteristics of stacked-X junctions are known to be perturbed by cation binding. When the junction forming sequence d(TCGGTACCGA) is crystallized in the presence of high concentrations of strontium ions the resulting structure is perturbed relative to the parent d(CCGGTACCGG) structure with an increase in the average J_{twist}, J_{slide}, J_{roll} and IDA parameters (Table 6.2). The authors proposed that these conformational changes were the result of a "strontium spine" in the minor groove of the junction structure (*75*). This minor groove binding of strontium ions is consistent with the overall trend of calcium ions, the only other divalent cation identified in junctions to date, being localized to the minor groove of the four base pair arm of junction crystal structures (*48, 76*). A study of minor groove substituent interactions on junction formation suggested that calcium binding in the minor groove may not have a significant affect on junction structure even though it is adjacent to the N_{13}-N_8 base pair of the trinucleotide core (*153*)). Only one calcium ion was identified in the current ATC junction structures and it is located in a similar location within the minor groove of the four base pair arm of ATC-HS. This isn't entirely surprising given that, excluding the strontium junction (UD0021), this calcium position is the only location which has been previously reported to contain a divalent cation in all stacked-X structures solved to-date. Sodium is the only metal ion which has been observed in the immediate vicinity of the junction crossover, typically in the cavity of the major groove face (*47, 76*). This raises an intriguing question - if divalent cations (especially calcium and magnesium)

Table 6.2. Global conformation of DNA Holliday junctions. Comparison of global conformational geometry for junctions containing the parent $A_6C_7C_8$ trinucleotide core sequence motif and the present junction structures of d(CCGATATCGG). The calculation and definition of each conformational parameter has been previously described (*110*).

Junction	J_{twist} (°)	J_{slide} (Å)	J_{roll} (°)	IDA (°)
UD0008 (*47*)	37.8	0.15	139.9	65.9
UD0015 (*79*)	39.5	-0.36	145.3	68.3
UD0018 (*75*)	39.3	0.34	142.1	66.6
UD0023	40.6	1.06	146.6	67.4
UD0024	38.5	0.15	142.2	67.2
UD0025	37.7	0.11	142.7	66.7
UD0021 (*75*)	44.5	0.58	150.1	70.2
UD0026	45.4	0.46	150.3	70.7
ATC-LS	41.0	0.39	149.7	69.0
ATC-HS	43.9	0.90	148.4	70.0

facilitate conformational changes between the ATC-LS and ATC-HS junction structures then why are no calcium ions identified around the central core which is where previous studies have suggested they have the greatest impact on junction folding and conformation (*50, 143*)? In order for identification to occur through crystallographic methods the solvent ions would need to be fairly well ordered or highly occupied in specific positions of the atomic structure. Indeed, FRET studies have demonstrated that ion binding to the center of stacked-X junctions occurs in a very rapid and noncooperative manner (*63, 137*). This rapid exchange could conceivably preclude cation identification using crystallographic methods. It should be noted that we have identified solvent waters within each junction structure with relatively low B-factors which may indicate that they are disordered ions that can not be definitively assigned with the current data. Taken in context these results are more indicative of group II ion binding to the junction center occurring in a passive manner and not to specific sites where phosphates are in close approach.

One of the two locations where a metal ion has been identified within the major groove face of the junction crossover is the central cavity created by the N_6 phosphate oxygens. This central cavity in the initial d(CCGGTACCGG) junction structure contained a well defined sodium ion which has since been identified in other junction structures (*47, 76*). Interestingly, the ATC-LS structure also has a well defined sodium ion in this central cavity, which is the only metal ion identified in the structure, while ATC-HS has two partially occupied preferred hydration sites (*154*) adjacent to the crystallographic two-fold within the cavity center. Displacement of solvent in this central ion binding site within the ATC-HS structure is likely due to the

change in overall geometry of the structure which produces a more compact central core (average P-P distance for ATC-HS is 8.19 Å vs. 8.73 Å in ATC-LS using $T_5A_6T_7$ phosphorous atoms adjacent to the cavity). Compaction of this central cavity is evident from a comparison of the cavity volumes for ATC-HS (9.52 $Å^3$ with SD = 2.26 $Å^3$) and ATC-LS (17.83 $Å^3$ with SD = 1.71 $Å^3$). As compression of the central cavity occurs the solvent atom is likely extruded to either side of the cavity where it can still shield the phosphate oxygens, a similar solvent extrusion model was proposed previously with the compressed cavity of d(CCGGTACm^{5}CGG) (*48*).

The present results shows that the conformation of the d(CCGATATCGG) junction is perturbed with decreasing salt concentration. The lower [$CaCl_2$] generates a 0.51 Å decrease in J_{slide}, 2.9° decrease in J_{twist} and a 1.3° increase in J_{roll} which combine to produce a one degree decrease in the interduplex angle. Thus, a decrease in salt concentration causes the junction structure to open up and expose a greater amount of the major groove face generating a more symmetric overall core (Figure 6.5). This opening, or roll, of the duplex arms is combined with a decrease in the twist of the duplex arms. If a perfectly parallel junction is considered to be fully symmetric in terms of twist then this decrease is a movement towards the more symmetric form where the helix axes of each duplex arm would be aligned coincident with each other. Simply twisting the arms of the stacked-X junction to make them more aligned with one another is not sufficient to produce a fully symmetric structure - the arms would have to translate along the helix axis of each duplex, since they are offset in the compact stacked-X form, to produce a more

Figure 6.5. Schematic of conformational changes associated with decreasing salt concentration for the d(CCGATATCGG) sequence. The calcium chloride concentration corresponding to specific DNA conformations within this study is noted along the top. B-DNA duplexes are shown as cylinders with individual DNA strands as solid black arrows pointing in a 5' to 3' direction. Translational or rotational displacement of cylinders is denoted by dashed grey arrows pointing in the direction of duplex movement. The conformational variability of four-stranded junctions can be described by the parameters J_{twist}, J_{slide} and J_{roll} (*110*). With decreasing [$CaCl_2$] the angle between the duplex arms of the junction (J_{twist}) and the translational displacement of each arm along the duplex helix axis (J_{slide}) decreases. When combined with an increase in J_{roll}, increasing exposure of the major groove junction face, the overall junction core becomes more symmetric and less compact as the [$CaCl_2$] decreases. If each of these trajectories is continued to its extreme the hypothetical structures within the grey box would result. We should note that conformations within the grey box are not observed in the present study, rather they are simply a continuation of the trajectories for each conformational parameter. Presence of the open-X junction form (denoted in brackets), though not observed in the current study, is suggested by the presence of resolved B-DNA duplexes at lower concentrations of calcium chloride.

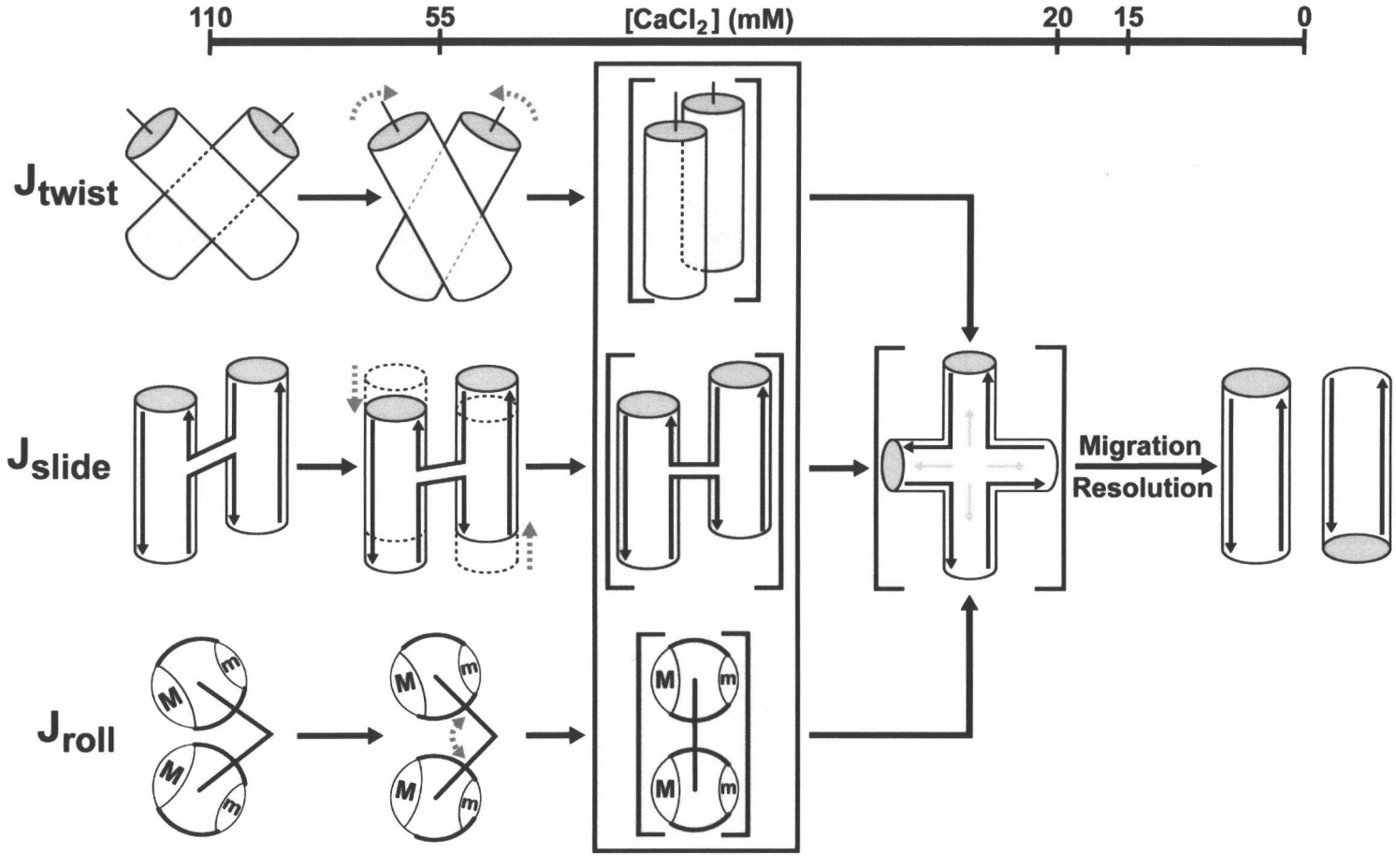
110
55
[CaCl2] (mM)
20
15
0
Jtwist
Jslide
Jroll
M
E
Migration
Resolution

Figure 6.5

symmetric junction center. Indeed, we see a significant decrease (0.51 Å) in J_{slide} as the salt concentration is decreased which generates a less compact and more open junction core. As the stacked-X junction becomes more symmetric, or planar, with decreasing salt concentration the duplex arms will eventually repel away from one another into the fully extended open-X conformation. With inverted-repeat DNA sequences we would expect the open-X form to rapidly produce resolved B-DNA duplexes since branch migration in these sequences is an isoenergetic process (the resolved B-DNA duplexes will have an identical sequence). Thus, even though the open-X structure is not observed in the present studies we anticipate that this form exists as the link between the stacked-X and resolved structures.

Overall these results suggest that the transition between a stacked-X and open-X junction structure occurs via the rotation and translation of the duplex arms of the stacked-X junction producing a more symmetric and less compact core. Crossover stabilization likely occurs through hydrogen bonding networks between solvent water, extracyclic base substituents, phosphate oxygens, and ordered cation hydration shells. As these interactions are disrupted, either by specific modifications (*48, 76, 135*) or through changes in salt concentration (*75*), the conformation of the junction structure becomes perturbed. Thus, the central core may serve as a linchpin for determining, or stabilizing, the conformation of four-stranded DNA junctions. This has biological implications since the bacterial recombination protein RuvA is known to interact with the crossover region of four-stranded junctions through two acidic-pins (Glu55 and Asp56) which repel the phosphodiester backbone of the crossover strands away from each other through electrostatic repulsion and help facilitate base-pair rearrangements

(*3, 155*). These interactions are essential to the activity of RuvA and may play an additional role in targeting the RuvC endonuclease to the junction cleavage site within the RuvABC resolvasome (*155*). RuvC may have particular importance in the current discussion since it may recognize the stacked-X conformation of the Holliday junction (*36*). Thus, changes in junction conformation could conceivably affect its recognition by protein machinery *in vivo*. Arguably there is still a lot to be understood, especially in eukaryotic organisms, regarding the exact role of four-stranded junctions in biological systems. What has become clear is that the stacked-X form of the four-stranded junction is conformationally malleable and significantly affected by sequence and atomic interactions within the central core region.

6.5 Materials and Methods

Deoxyoligonucleotides of the sequence 5'-CCGATATCGG-3' were synthesized using β-cyanoethyl phosphoramidite chemistry and obtained from Midland Certified Reagent Company (Midland, TX). Upon arrival the tritylated oligomers were evaporated to a neutral pH following removal of protecting groups by hydrolysis with concentrated ammonium hydroxide for 24 hours at room temperature. The 5'-O-dimethoxytrityl (DMT) group was left attached to facilitate separation from failure sequences and reaction by-products using preparative reverse-phase high performance liquid chromatography (Microsorb 300 Å pore size C-18 column measuring 250 mm X 21.4 mm at a flow rate of 8 mL / min in 0.1 mM triethylamine acetate with an acetonitrile gradient). The DMT-decamer sequence typically eluted

~31 min postinjection in a mobile phase concentration of 20% acetonitrile. DMT was removed by treatment with 3% aqueous acetic acid for 15 minutes with gentle agitation, neutralized with concentrated ammonium hydroxide, and desalted on a Sephadex G-25 (Sigma) column. Purified DNA was stored at -80 °C as a lyophilized powder and resuspended in deionized double-distilled water prior to use, without strand annealing. Crystallization was carried out at room temperature in 10 μL sitting drops by the vapor diffusion method and equilibrated against a 30 mL reservoir volume. Crystal morphologies consistent with junctions (thin diamond plates vs. long rods for the B-DNA form) for the sequence d(CCGATATCGG) were observed under a range of salt concentrations (20 mM to 110 mM $CaCl_2$), though between 20 mM and 55 mM $CaCl_2$ the B-DNA crystal form would also typically be observed in the same drop. B-DNA only crystals were observed in solutions containing 0 to 15 mM $CaCl_2$.

6.5.1 Crystallization and structure determination of B-DNA form d(CCGATATCGG). The B-DNA form of d(CCGATATCGG) crystallized as long rods in solutions containing 0.65 mM DNA, 5 mM $CaCl_2$, 75 mM NaCacodylate (pH 7.0) and 0.9 mM spermine-4HCl against a reservoir solution of 30% 2-methyl-2,4-pentanediol (MPD). X-ray diffraction data to 1.61 Å was collected at liquid nitrogen temperatures using CuKα radiation from a Rigaku (Tokyo) RU-H3R rotating anode generator operating at 50 kV and 100 mA ($\Delta\Phi = 1°$, 180 15 minute exposures) with an Raxis-IV image plate detector on a single crystal measuring 250 mm X 50 mm X 50 mm. The crystal was indexed in the orthorhombic spacegroup $P2_12_12_1$ with unit cell dimensions (a = 35.2 Å, b = 39.1 Å, and c = 33.0 Å) similar to the previously solved structure of d(CCAAATTTGG) (BD0051, (*145*)). A search model derived from this

structure was used for molecular replacement in EPMR (*83*) which produced a unique solution with a correlation coefficient of 68.3% and R_{cryst} of 39.7%. Subsequent refinement of this model in Refmac5 (*156*) produced a final R_{cryst} of 25.7% and R_{free} of 27.5% (Table 6.1).

6.5.2 Crystallization and structure determination of d(CCGATATCGG) low salt junction. Crystals were grown in solutions containing 0.7 mM DNA, 55 mM $CaCl_2$, 0.8 mM spermine·4HCl, and 75 mM NaCacodylate (pH 7.0) against a reservoir solution containing 30% by volume MPD. Initial small diamond plate crystals appeared after 30 days with optimal crystals appearing after two additional months. A single crystal measuring 250 mm X 375 mm X 50 mm was used for data collection. Initial data collection was performed at liquid nitrogen temperatures using CuKα radiation from a Rigaku (Tokyo) RU-H3R rotating anode generator operating at 50 kV and 100 mA ($\Delta\Phi$ = 1°, 180 15 minute exposures) with an Raxis-IV image plate detector. This data was later merged with data collected from BioCARS beamline 14-IDB at the Advanced Photon Source (Argonne National Laboratory, Argonne, IL) using 1.07 Å radiation, $\Delta\Phi$ = 1°, 100 20 second exposures, and a MAR CCD detector (Table 6.1). The final, merged, dataset extended to usable resolution of 1.65 Å. The crystal was indexed in monoclinic spacegroup C2 with unit cell dimensions (a=64.9 Å, b= 24.0 Å, c= 40.3 Å, α=γ=90°, and β=118.69°) isomorphous to a previously solved high salt ATC junction. Thus, a search model was derived from this previously solved structure and used in EPMR for molecular replacement. A distinct solution was obtained with a correlation coefficient of 44.8% and R_{cryst} of 49.6% and two unique strands adjacent to a crystallographic 2-fold symmetry axis. To verify that this was

indeed a 4H-junction we generated a simulated annealing omit map for the C_7 crossover phosphate which produced distinct F_o-F_c density within the trinucleotide core region. The initial round of rigid body and rigid parts refinement on this model followed by simulated annealing produced a starting R_{cryst} = 30.9% and R_{free} = 37.0%. Subsequent refinement was performed in CNS (*84*) using standard positional and individual B-factor refinement along with assignment of solvent atoms. This model was then completed in Refmac5 (*156*) using TLS refinement (*157, 158*), with each strand assigned as a TLS group, resulting in a final R_{cryst} of 23.7% and R_{free} of 27.3%.

6.5.3 Crystallization and structure determination of d(CCGATATCGG) high salt junction. Crystals of the high salt junction form were grown in solutions containing 0.7 mM DNA, 110 mM $CaCl_2$, 0.8 mM spermine·4HCl, and 75 mM NaCacodylate (pH 7.0) against a reservoir solution containing 30% by volume MPD. Small diamond plate crystals grew over a period of approximately 30 days with diffraction quality crystals appearing after an additional two months. A single crystal measuring 500 mm X 375 mm X 50 mm was used for data collection. Collection of synchrotron diffraction data (to 1.85 Å) was performed on BioCARS beamline 14-BMD at the Advanced Photon Source (Argonne National Laboratory, Argonne, IL) using 0.98 Å radiation, $\Delta\Phi = 1°$, 120 45 second exposures and an ADSC Quantum 4 detector (Table 6.1). The crystal was indexed in monoclinic spacegroup C2 with unit cell dimensions (a=65.0 Å, b= 25.0 Å, c= 37.5 Å, $\alpha=\gamma=90°$, and $\beta=112.09°$) isomorphous to the junction form of d(CCGGGCCCGG) (unpublished data). Thus, a search model was derived from this previously solved structure and used in EPMR for molecular replacement. A distinct solution was obtained with a correlation coefficient

of 75.2% and R_{cryst} of 34.0% and two unique strands adjacent to a crystallographic 2-fold symmetry axis. To verify that this was indeed a 4H-junction we generated a simulated annealing omit map for the C_7 crossover phosphate which produced distinct F_o-F_c density within the trinucleotide core region. The initial round of rigid body and rigid parts refinement on this model followed by simulated annealing produced a starting R_{cryst} = 31.0% and R_{free} = 35.0%. Subsequent refinement was performed in CNS (*84*) using standard positional and individual B-factor refinement along with assignment of solvent atoms. This model was then completed in Refmac5 (*156*) using TLS refinement (*157, 158*), with each strand assigned as a TLS group, resulting in a final R_{cryst} of 20.0% and R_{free} of 23.6%.

Prior to data collection, all crystals were transferred to artificial mother liquor containing 25% MPD and flash cooled to -180°C in liquid nitrogen for cryoprotection. All data were processed and reflections merged and scaled using DENZO and SCALEPACK from the HKL suite of programs (*85*). Resolution cutoffs were determined based on data completeness, I/σ(I), and R_{merge}. Room-mean-square deviation (RMSD) values were calculated using the McLachlan algorithm as implemented in the ProFit version 2.2 program (http://www.bioinf.org.uk/software/profit) (*86*)). Structural analysis was performed using CURVES 5.2 (*87*) and 3DNA (*88*). All assigned solvent atoms within the present structures are listed according to their densities in the final $2F_o$-F_c map, with the first solvent atom having the highest density, using a program provided by the Karplus lab at Oregon State. The coordinates and structure factors have been deposited

in the Protein Data Bank (*105*) as entries 1PNY for ATC-B, 1PNY for ATC-LS, and 1PNY for ATC-HS.

6.6 Acknowledgements

This work was funded by grants from the National Institutes of Health (R1GM62957A) and the National Science Foundation (MCB0090615). The X-ray diffraction facilities are supported by the Proteins and Nucleic Acids Facility Core of the Environmental Health Sciences Center at OSU (NIEHS ES00210) and a grant from the Murdock Charitable Trust. Access to the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under Contract W-31-109-Eng-38. Access to BioCARS Sector 14 was supported by the National Institutes of Health, National Center for Research Resources, under Grant RR07707.

Chapter 7

Conclusion and Discussion

Conformational variability, including the ability to undergo branch migration and subsequent resolution, is a fundamental characteristic of four-stranded DNA Holliday junctions in biological systems. Although DNA crossovers are known to be recognized by a growing number of proteins *in vivo*, it isn't very clear how intrinsic junction properties influence this recognition and subsequent resolution. The current studies were initiated to better understand the role that the sequence within the central crossover region may play in determining junction conformation and formation. As a result of these studies, four general conclusions about junction structure and function have emerged.

Stacked-X junction formation and conformational variability occurs primarily in a sequence-dependent manner within the current system. Both of the initial stacked-X junction structures were stabilized by the presence of a common ACC trinucleotide sequence in the $N_6N_7N_8$ position of the general sequence motif d(CCnnn$N_6N_7N_8$GG). Work described in this thesis expands the core sequence motif to $N_6N_7C_8$ where N_6 is an adenine, guanine, or cytosine nucleotide and N_7 is either a cytosine or thymine (when N_6 = adenine) nucleotide (excluding non-standard bases). Stabilization of these structures as stacked-X junctions is surprising since inverted-

repeat sequences have been shown to undergo spontaneous branch migration to resolved B-DNA duplexes (*7*). Stacked-X junction formation within the current system occurs through favorable base stacking and major groove hydrogen bonding interactions that fix the junction crossover at a specific location within the inverted-repeat sequence. These stabilizing major groove interactions also have a significant effect on junction conformation, producing translational and rotational displacement of the helical junction arms upon disruption (Table 7.1). Indeed, previous solution studies have shown that nucleotide sequence at or near the junction crossover has a significant effect on the formation and conformer preference of stacked-X junctions (*25, 43, 66*). Since junctions are expected to form across a broad range of sequences *in vivo*, these data may suggest that specific trinucleotide core sequences can stabilize the stacked-X junction in biological processes. Many junction resolvases, such as RuvC or CCE1, have strong sequence specificities for cleavage while remaining highly specific for four-stranded junctions in terms of recognition. RusA, which is a λ-phage junction resolvase, has a strong preference for junction cleavage at ↓CC sites within the central crossover region (*159, 160*). Thus, the sequence-dependent formation within the current crystallographic system may serve in biological systems to stabilize specific junction conformers and facilitate subsequent resolution by junction resolvases to duplex DNA products. In addition, the sequence effects on global conformation may play a significant role in junction recognition by cellular proteins, including resolvases and DNA repair proteins.

Table 7.1. Comparison of global conformational geometry for all currently known inverted-repeat stacked-X DNA Holliday junctions. The calculation and definition of each conformational parameter has been previously described (*110*). Each junction is designated by its corresponding trinucleotide core sequence (for structures resulting from this thesis) or by an NDB accession code (for previously published structures). Perturbations to global junction geometry are generally related to sequence or cation dependent disruption of stabilizing interactions within major groove face of the junction core.

Junction	J_{twist} (°)	J_{slide} (Å)	J_{roll} (°)	IDA (°)
iACm^{5}C (*153*)	42.5	2.23	143.6	61.4
ACbr^{5}U (*76*)	37.1	1.27	130.4	68.5
GCC (*76*)	37.0	-0.04	143.6	62.0
CCC	43.1	1.03	147.0	67.4
ATC-LS	41.0	0.39	149.7	69.0
ATC-HS	43.9	0.90	148.4	70.0
UD0008 (*47*)	37.8	0.15	139.9	65.9
UD0015 (*48*)	44.3	2.25	145.9	60.0
UD0015 (*79*)	39.5	-0.36	145.3	68.3
UD0018 (*75*)	39.3	0.34	142.1	66.6
UD0023	40.6	1.06	146.6	67.4
UD0024	38.5	0.15	142.2	67.2
UD0025	37.7	0.11	142.7	66.7
UD0021 (*75*))	44.5	0.58	150.1	70.2
UD0026	45.4	0.46	150.3	70.7

Major groove interactions within the junction core have a significant influence on junction formation and conformation. A common theme throughout this thesis has been the focus on specific stabilizing hydrogen bonding interactions within the junction core. Chapter 2 discusses the nature of these interactions in detail (*93*). With the exception of a direct or solvent mediated interaction between the extracyclic amine of C_8 and phosphate oxygen of N_7, all of the interactions can be replaced by compensating interactions. Indeed, the number of stabilizing interactions within the major groove of the junction core increases as the conformation becomes more distorted (see ACm^{5}C, ACbr^{5}U, and ATC-HS structures for examples). In terms of stabilizing interactions within the core, the key may therefore be the ability of the core nucleotides to form a range of stabilizing interactions and not just the specific interactions proposed in the initial d(CCGGTACCGG) paper (*47*) (excluding the direct or solvent mediated N_7-C_8 interaction which appears to be essential for junction formation within the current system). The results of the minor groove substituent study in Chapter 4 demonstrate that minor groove interactions within the region of the central junction core are not a determining factor in junction formation within the current system (*76*). Furthermore, perturbations to local displacement (rotational and translational) parameters of core nucleotides appear to be tolerated as long as they do not significantly disrupt the ability to form major groove stabilizing interactions. Two examples of this principle are the ACT sequence in Chapter 3 (*76*), which forms resolved B-DNA duplexes, and the methylated junctions in Chapter 4. These methylated junctions (ACm^{5}C and iACm^{5}C) (*48, 153*) are almost identical in global conformation yet the inosine base in the $N_3 \cdot N_8$ base step induces a number of local

distortions to the helical arm (such as the BII backbone conformation of the inosine nucleotides or a narrowing of the minor groove and disruption of the minor groove solvent network) of the junction. Thus, major groove stabilizing interactions play a significant role in the formation and stabilization of the stacked-X junction.

Cation binding to the major groove face of the junction core does not occur in a site-specific manner. A surprising aspect in the current body of stacked-X DNA Holliday junction structures is the lack of group (II) divalent cations assigned at, or near, the central junction crossover. This crossover region contains a high electrostatic potential resulting from the close proximity of backbone phosphate atoms (*50, 140, 161*). Previous studies have tried to ascertain whether specific cation binding sites existed within the junction core and to account for the drastic salt-dependent conformational changes in junction conformation (*63, 137, 142, 143*) and preference for stacking conformer (*43, 161*). The data presented in this thesis suggest that site-specific binding of divalent cations to the junction core does not occur for the present constructs. This is especially evident with the ATC junction structures where a decrease in calcium (II) chloride concentration induced a transition in conformation that leads to resolved B-DNA duplexes. The only calcium cation identified in either junction structure is in the minor groove of the four base pair arm, a location shown in Chapter 4 to not significantly affect junction conformation. These results suggest that cation binding to the junction core does not occur in a site-specific manner. Previous fluorescence studies suggested that cation interaction with the junction core occurred in a rapid and highly mobile manner. Thus, diffuse clouds of ions may shield phosphate oxygens within the junction core. Cation interactions through these diffuse

clouds may play a significant role in the conformational variability and transitions of the folded junction.

The central core, or crossover region, of the junction structure serves as a focal point in determining overall junction conformation. A key observation resulting from work presented in this thesis is that perturbations such as alteration of trinucleotide core sequence (base substituents) or disruption of major groove hydrogen bonding interactions within the central core of the folded junction structure can have a significant effect on junction conformation (compare junction structures) and formation (compare crystallographic screen results). Thus, the central core may have a key role in the function of four-stranded junctions in biological systems. Not surprisingly, a number of cellular proteins interact with the junction through contacts within the central crossover region. RuvA, which targets four-stranded junctions as part of the RuvABC resolvasome, has two highly conserved amino acids (Glu55 and Asp56) that interact with the junction core and stabilize the open conformation by repelling the negative DNA backbone (*3, 13, 155*). Recent modeling studies of the junction resolvase T7 endonuclease I with a four-stranded junction has an extended β-bridge passing across the major groove face of the junction core. Protein-junction interactions distal to the central core are likely to be involved in these assorted biological systems as well though the central junction core is highly correlated with junction conformation in the folded form. Conformational details of the junction arms within the stacked-X junction (such as helical twist, groove width, propeller twist, etc.) are generally determined in a sequence-dependent manner and not induced as a result of junction formation. This indicates that the stacked-X junction structure can

tolerate disruptions to base geometry distal to the central core without significant changes in global conformation (as demonstrated in Chapter 4). On the other hand, perturbations within the central core region can have significant affects on junction geometry, especially within two nucleotides 3' of the junction crossover (7th and 8th positions in current studies). The current results both provide indirect evidence for the importance of the central core region in biological systems and provide direct evidence for a correlation between base sequence and atomic interactions within the core region and junction conformation.

Bibliography

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

2. Holliday, R. (1974) Molecular aspects of genetic exchange and gene conversion, *Genetics 78*, 273-87.

3. Ariyoshi, M., Nishino, T., Iwasaki, H., Shinagawa, H., and Morikawa, K. (2000) Crystal structure of the holliday junction DNA in complex with a single RuvA tetramer, *Proc. Natl. Acad. Sci. U S A 97*, 8257-62.

4. Van Duyne, G. D. (2001) A structural view of cre-loxp site-specific recombination, *Annu. Rev. Biophys. Biomol. Struct. 30*, 87-104.

5. Dickman, M. J., Ingleston, S. M., Sedelnikova, S. E., Rafferty, J. B., Lloyd, R. G., Grasby, J. A., and Hornby, D. P. (2002) The RuvABC resolvasome, *Eur. J. Biochem. 269*, 5492-501.

6. Liu, Y., and West, S. C. (2004) Happy Hollidays: 40th anniversary of the Holliday junction, *Nat. Rev. Mol. Cell Biol. 5*, 937-44.

7. Panyutin, I. G., and Hsieh, P. (1994) The kinetics of spontaneous DNA branch migration, *Proc. Natl. Acad. Sci. U S A 91*, 2021-5.

8. Meselson, M. S., and Radding, C. M. (1975) A general model for genetic recombination, *Proc. Natl. Acad. Sci. U S A 72*, 358-61.

9. Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J., and Stahl, F. W. (1983) The double-strand-break repair model for recombination, *Cell 33*, 25-35.

10. Lin, F. L., Sperle, K., and Sternberg, N. (1984) Model for homologous recombination during transfer of DNA into mouse L cells: role for DNA ends in the recombination process, *Mol. Cell. Biol. 4*, 1020-34.

11. Haber, J. E., Ira, G., Malkova, A., and Sugawara, N. (2004) Repairing a double-strand chromosome break by homologous recombination: revisiting Robin Holliday's model, *Philos. Trans. R. Soc. Lond., B Biol. Sci. 359*, 79-86.

12. Yamada, K., Miyata, T., Tsuchiya, D., Oyama, T., Fujiwara, Y., Ohnishi, T., Iwasaki, H., Shinagawa, H., Ariyoshi, M., Mayanagi, K., and Morikawa, K. (2002) Crystal structure of the RuvA-RuvB complex: a structural basis for the Holliday junction migrating motor machinery, *Mol. Cell 10*, 671-81.

13. Rafferty, J. B., Sedelnikova, S. E., Hargreaves, D., Artymiuk, P. J., Baker, P. J., Sharples, G. J., Mahdi, A. A., Lloyd, R. G., and Rice, D. W. (1996) Crystal structure of DNA recombination protein RuvA and a model for its binding to the Holliday junction, *Science 274*, 415-21.

14. Yamada, K., Kunishima, N., Mayanagi, K., Ohnishi, T., Nishino, T., Iwasaki, H., Shinagawa, H., and Morikawa, K. (2001) Crystal structure of the Holliday junction migration motor protein RuvB from Thermus thermophilus HB8, *Proc. Natl. Acad. Sci. U S A 98*, 1442-7.

15. Ariyoshi, M., Vassylyev, D. G., Iwasaki, H., Nakamura, H., Shinagawa, H., and Morikawa, K. (1994) Atomic structure of the RuvC resolvase: a holliday junction-specific endonuclease from E. coli, *Cell 78*, 1063-72.

16. Sha, R., Liu, F., Iwasaki, H., and Seeman, N. C. (2002) Parallel symmetric immobile DNA junctions as substrates for E. coli RuvC Holliday junction resolvase, *Biochemistry 41*, 10985-93.

17. Liu, Y., Masson, J. Y., Shah, R., O'Regan, P., and West, S. C. (2004) RAD51C is required for Holliday junction processing in mammalian cells, *Science 303*, 243-6.

18. Constantinou, A., Davies, A. A., and West, S. C. (2001) Branch migration and Holliday junction resolution catalyzed by activities from mammalian cells, *Cell 104*, 259-68.

19. Constantinou, A., Chen, X. B., McGowan, C. H., and West, S. C. (2002) Holliday junction resolution in human cells: two junction endonucleases with distinct substrate specificities, *EMBO J. 21*, 5577-85.

20. Ceschini, S., Keeley, A., McAlister, M. S., Oram, M., Phelan, J., Pearl, L. H., Tsaneva, I. R., and Barrett, T. E. (2001) Crystal structure of the fission yeast mitochondrial Holliday junction resolvase Ydc2, *EMBO J. 20*, 6601-11.

21. Garcia, P. L., Liu, Y., Jiricny, J., West, S. C., and Janscak, P. (2004) Human RECQ5beta, a protein with DNA helicase and strand-annealing activities in a single polypeptide, *EMBO J. 23*, 2882-91.

22. Constantinou, A., Tarsounas, M., Karow, J. K., Brosh, R. M., Bohr, V. A., Hickson, I. D., and West, S. C. (2000) Werner's syndrome protein (WRN) migrates Holliday junctions and co-localizes with RPA upon replication arrest, *EMBO Rep. 1*, 80-4.

23. 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. U S A 97*, 6504-8.

24. Wu, L., and Hickson, I. D. (2003) The Bloom's syndrome helicase suppresses crossing over during homologous recombination, *Nature 426*, 870-4.

25. McKinney, S. A., Declais, A. C., Lilley, D. M., and Ha, T. (2003) Structural dynamics of individual Holliday junctions, *Nat. Struct. Biol. 10*, 93-7.

26. Duckett, D. R., Murchie, A. I., Diekmann, S., von Kitzing, E., Kemper, B., and Lilley, D. M. (1988) The structure of the Holliday junction, and its resolution, *Cell 55*, 79-89.

27. Clegg, R. M., Murchie, A. I., and Lilley, D. M. (1994) The solution structure of the four-way DNA junction at low-salt conditions: a fluorescence resonance energy transfer analysis, *Biophys. J. 66*, 99-109.

28. 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*, 763-6.

29. Gopaul, D. N., Guo, F., and Van Duyne, G. D. (1998) Structure of the Holliday junction intermediate in Cre-loxP site-specific recombination, *EMBO J. 17*, 4175-87.

30. Guo, F., Gopaul, D. N., and van Duyne, G. D. (1997) Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse, *Nature 389*, 40-6.

31. Middleton, C. L., Parker, J. L., Richard, D. J., White, M. F., and Bond, C. S. (2004) Substrate recognition and catalysis by the Holliday junction resolving enzyme Hje, *Nucleic Acids Res. 32*, 5442-51.

32. Nishino, T., Komori, K., Tsuchiya, D., Ishino, Y., and Morikawa, K. (2001) Crystal structure of the archaeal holliday junction resolvase Hjc and implications for DNA recognition, *Structure (Camb) 9*, 197-204.

33. Raaijmakers, H., Vix, O., Toro, I., Golz, S., Kemper, B., and Suck, D. (1999) X-ray structure of T4 endonuclease VII: a DNA junction resolvase with a novel fold and unusual domain-swapped dimer architecture, *EMBO J. 18*, 1447-58.

34. Bhattacharyya, A., Murchie, A. I., von Kitzing, E., Diekmann, S., Kemper, B., and Lilley, D. M. (1991) Model for the interaction of DNA junctions and resolving enzymes, *J. Mol. Biol. 221*, 1191-207.

35. Fogg, J. M., Schofield, M. J., White, M. F., and Lilley, D. M. (1999) Sequence and functional-group specificity for cleavage of DNA junctions by RuvC of Escherichia coli, *Biochemistry 38*, 11349-58.

36. Bennett, R. J., and West, S. C. (1996) Resolution of Holliday junctions in genetic recombination: RuvC protein nicks DNA at the point of strand exchange, *Proc. Natl. Acad. Sci. U S A 93*, 12217-22.

37. Bennett, R. J., and West, S. C. (1995) Structural analysis of the RuvC-Holliday junction complex reveals an unfolded junction, *J. Mol. Biol. 252*, 213-26.

38. Cooper, J. P., and Hagerman, P. J. (1987) Gel electrophoretic analysis of the geometry of a DNA four-way junction, *J. Mol. Biol. 198*, 711-9.

39. Gough, G. W., and Lilley, D. M. (1985) DNA bending induced by cruciform formation, *Nature 313*, 154-6.

40. Carlstrom, G., and Chazin, W. J. (1996) Sequence dependence and direct measurement of crossover isomer distribution in model Holliday junctions using NMR spectroscopy, *Biochemistry 35*, 3534-44.

41. von Kitzing, E., Lilley, D. M., and Diekmann, S. (1990) The stereochemistry of a four-way DNA junction: a theoretical study, *Nucleic Acids Res. 18*, 2671-83.

42. Yu, J., Ha, T., and Schulten, K. (2004) Conformational model of the Holliday junction transition deduced from molecular dynamics simulations, *Nucleic Acids Res. 32*, 6683-95.

43. Miick, S. M., Fee, R. S., Millar, D. P., and Chazin, W. J. (1997) Crossover isomer bias is the primary sequence-dependent property of immobilized Holliday junctions, *Proc. Natl. Acad. Sci. U S A 94*, 9080-4.

44. Churchill, M. E., Tullius, T. D., Kallenbach, N. R., and Seeman, N. C. (1988) A Holliday recombination intermediate is twofold symmetric, *Proc. Natl. Acad. Sci. U S A 85*, 4653-6.

45. Lilley, D. M., and Norman, D. G. (1999) The Holliday junction is finally seen with crystal clarity, *Nat. Struct. Biol. 6*, 897-9.

46. Ortiz-Lombardia, M., Gonzalez, A., Eritja, R., Aymami, J., Azorin, F., and Coll, M. (1999) Crystal structure of a DNA Holliday junction, *Nat. Struct. Biol. 6*, 913-7.

47. 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. U S A 97*, 3971-6.

48. Vargason, J. M., and Ho, P. S. (2002) The effect of cytosine methylation on the structure and geometry of the Holliday junction: the structure of d(CCGGTACm5CGG) at 1.5 A resolution, *J. Biol. Chem. 277*, 21041-9.

49. Lilley, D. M., and White, M. F. (2001) The junction-resolving enzymes, *Nat. Rev. Mol. Cell. Biol. 2*, 433-43.

50. Lilley, D. M. (2000) Structures of helical junctions in nucleic acids, *Q. Rev. Biophys. 33*, 109-59.

51. Nunes-Duby, S. E., Matsumoto, L., and Landy, A. (1987) Site-specific recombination intermediates trapped with suicide substrates, *Cell 50*, 779-88.

52. Cox, M. M., Goodman, M. F., Kreuzer, K. N., Sherratt, D. J., Sandler, S. J., and Marians, K. J. (2000) The importance of repairing stalled replication forks, *Nature 404*, 37-41.

53. Haber, J. E., and Heyer, W. D. (2001) The fuss about Mus81, *Cell 107*, 551-4.

54. Declais, A. C., Fogg, J. M., Freeman, A. D., Coste, F., Hadden, J. M., Phillips, S. E., and Lilley, D. M. (2003) The complex between a four-way DNA junction and T7 endonuclease I, *EMBO J. 22*, 1398-409.

55. Subramaniam, S., Tewari, A. K., Nunes-Duby, S. E., and Foster, M. P. (2003) Dynamics and DNA substrate recognition by the catalytic domain of lambda integrase, *J. Mol. Biol. 329*, 423-39.

56. Hargreaves, D., Rice, D. W., Sedelnikova, S. E., Artymiuk, P. J., Lloyd, R. G., and Rafferty, J. B. (1998) Crystal structure of E.coli RuvA with bound DNA Holliday junction at 6 A resolution, *Nat. Struct. Biol. 5*, 441-6.

57. 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.

58. 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, *Nat. Struct. Biol. 6*, 151-6.

59. Kato, M., Hokabe, S., Itakura, S., Minoshima, S., Lyubchenko, Y. L., Gurkov, T. D., Okawara, H., Nagayama, K., and Shimizu, N. (2003) Interarm interaction of DNA cruciform forming at a short inverted repeat sequence, *Biophys. J. 85*, 402-8.

60. Shlyakhtenko, L. S., Potaman, V. N., Sinden, R. R., and Lyubchenko, Y. L. (1998) Structure and dynamics of supercoil-stabilized DNA cruciforms, *J. Mol. Biol. 280*, 61-72.

61. Sha, R., Liu, F., Millar, D. P., and Seeman, N. C. (2000) Atomic force microscopy of parallel DNA branched junction arrays, *Chem. Biol. 7*, 743-51.

62. Cooper, J. P., and Hagerman, P. J. (1989) Geometry of a branched DNA structure in solution, *Proc. Natl. Acad. Sci. U S A 86*, 7336-40.

63. 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*, 4846-56.

64. 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.

65. Sha, R., Liu, F., and Seeman, N. C. (2002) Atomic force microscopic measurement of the interdomain angle in symmetric Holliday junctions, *Biochemistry 41*, 5950-5.

66. Grainger, R. J., Murchie, A. I., and Lilley, D. M. (1998) Exchange between stacking conformers in a four-Way DNA junction, *Biochemistry 37*, 23-32.

67. Gopaul, D. N., Guo, F., and van Duyne, G. D. (1998) Structure of the Holliday junction intermediate in Cre-loxP site-specific recombination., *EMBO J. 17*, 4175-4187.

68. Chen, Y., Narendra, U., Iype, L. E., Cox, M. M., and Rice, P. A. (2000) Crystal structure of a Flp recombinase-Holliday junction complex: assembly of an active oligomer by helix swapping, *Mol. Cell 6*, 885-897.

69. 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.

70. Dickerson, R. E. (1999) in *Oxford Handbook of Nucleic Acid Structure* (Neidle, S., Ed.) pp 145-197, Oxford University Press, New York.

71. Wood, A. A., Nunn, C. M., Trent, J. O., and Neidle, S. (1997) Sequence-dependent crossed helix packing in the crystal structure of a B-DNA decamer yields a detailed model for the Holliday junction, *J. Mol. Biol. 269*, 827-41.

72. Goodsell, D. S., Grzeskowiak, K., and Dickerson, R. E. (1995) Crystal structure of C-T-C-T-C-G-A-G-A-G. Implications for the structure of the Holliday junction, *Biochemistry 34*, 1022-9.

73. Eichman, B. F., Ortiz-Lombardia, M., Aymami, J., Coll, M., and Ho, P. S. (2002) The inherent properties of DNA four-way junctions: comparing the crystal structures of holliday junctions, *J. Mol. Biol. 320*, 1037-51.

74. 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*, 751-9.

75. Thorpe, J. H., Gale, B. C., Teixeira, S. C., and Cardin, C. J. (2003) Conformational and hydration effects of site-selective sodium, calcium and strontium ion binding to the DNA Holliday junction structure d(TCGGTACCGA)(4), *J. Mol. Biol. 327*, 97-109.

76. Hays, F. A., Vargason, J. M., and Ho, P. S. (2003) Effect of sequence on the conformation of DNA holliday junctions, *Biochemistry 42*, 9586-97.

77. Aymami, J., Pous, J., Lisgarten, J. N., and Coll, M. (2002) Crystallization and preliminary X-ray analysis of the DNA decamers d(CCGGATCCGG) and d(CCGGCGCCGG), *Acta Crystallogr. D Biol. Crystallogr. 58*, 310-1.

78. Heinemann, U., and Alings, C. (1989) Crystallographic study of one turn of G/C-rich B-DNA, *J. Mol. Biol. 210*, 369-81.

79. Thorpe, J. H., Teixeira, S. C., Gale, B. C., and Cardin, C. J. (2002) Structural characterization of a new crystal form of the four-way Holliday junction formed by the DNA sequence d(CCGGTACCGG)2: sequence versus lattice?, *Acta Crystallogr. Sect. D Biol. Crystallogr. 58*, 567-569.

80. Ho, P. S., and Eichman, B. F. (2001) The crystal structures of DNA Holliday junctions, *Curr. Opin. Struct. Biol. 11*, 302-8.

81. Panyutin, I. G., Biswas, I., and Hsieh, P. (1995) A pivotal role for the structure of the Holliday junction in DNA branch migration, *EMBO J. 14*, 1819-26.

82. Heinemann, U., Alings, C., and Bansal, M. (1992) Double helix conformation, groove dimensions and ligand binding potential of a G/C stretch in B-DNA, *EMBO J. 11*, 1931-9.

83. Kissinger, C. R., Gehlhaar, D. K., and Fogel, D. B. (1999) Rapid automated molecular replacement by evolutionary search, *Acta Crystallogr. D Biol. Crystallogr. 55 (Pt 2)*, 484-91.

84. Brunger, 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. D Biol. Crystallogr. 54 (Pt 5)*, 905-21.

85. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode, *Methods Enzymol. 276*, 307-326.

86. McLachlan, A. D. (1982) Rapid Comparison of Protein Structures, *Acta Crystallogr. A38*, 871-873.

87. Lavery, R., and Sklenar, H. (1989) Defining the structure of irregular nucleic acids: conventions and principles, *J. Biomol. Struct. Dyn. 6*, 655-67.

88. Lu, X. J., and Olson, W. K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures, *Nucleic Acids Res. 31*, 5108-21.

89. Esnouf, R. M. (1999) Further additions to MolScript version 1.4, including reading and contouring of electron-density maps, *Acta Crystallogr. D Biol. Crystallogr. 55 (Pt 4)*, 938-40.

90. Kielkopf, C. L., Ding, S., Kuhn, P., and Rees, D. C. (2000) Conformational flexibility of B-DNA at 0.74 A resolution: d(CCAGTACTGG)(2), *J. Mol. Biol. 296*, 787-801.

91. Kleywegt, G. J., and Jones, T. A. (1994), *Acta Crystallogr. D Biol. Crystallogr. D50*, 178-185.

92. Rafferty, J. B., Bolt, E. L., Muranova, T. A., Sedelnikova, S. E., Leonard, P., Pasquo, A., Baker, P. J., Rice, D. W., Sharples, G. J., and Lloyd, R. G. (2003) The structure of Escherichia coli RusA endonuclease reveals a new Holliday junction DNA binding fold, *Structure (Camb) 11*, 1557-67.

93. Hays, F. A., Watson, J., and Ho, P. S. (2003) Caution! DNA crossing: crystal structures of Holliday junctions, *J. Biol. Chem. 278*, 49663-6.

94. Lankas, F., Cheatham, T. E., 3rd, Spackova, N., Hobza, P., Langowski, J., and Sponer, J. (2002) Critical effect of the N2 amino group on structure, dynamics, and elasticity of DNA polypurine tracts, *Biophys. J. 82*, 2592-609.

95. Cubero, E., Guimil-Garcia, R., Luque, F. J., Eritja, R., and Orozco, M. (2001) The effect of amino groups on the stability of DNA duplexes and triplexes based on purines derived from inosine, *Nucleic Acids Res. 29*, 2522-34.

96. Lavigne, M., and Buc, H. (1999) Compression of the DNA minor groove is responsible for termination of DNA synthesis by HIV-1 reverse transcriptase, *J. Mol. Biol. 285*, 977-95.

97. Teplukhin, A. V., Poltev, V. I., and Chuprina, V. P. (1991) Dependence of the hydration shell structure in the minor groove of the DNA double helix on the groove width as revealed by Monte Carlo simulation, *Biopolymers 31*, 1445-53.

98. Woods, K. K., Lan, T., McLaughlin, L. W., and Williams, L. D. (2003) The role of minor groove functional groups in DNA hydration, *Nucleic Acids Res. 31*, 1536-40.

99. Brennan, C. A., Van Cleve, M. D., and Gumport, R. I. (1986) The effects of base analogue substitutions on the methylation by the EcoRI modification methylase of octadeoxyribonucleotides containing modified EcoRI recognition sequences, *J. Biol. Chem. 261*, 7279-86.

100. Brennan, C. A., Van Cleve, M. D., and Gumport, R. I. (1986) The effects of base analogue substitutions on the cleavage by the EcoRI restriction endonuclease of octadeoxyribonucleotides containing modified EcoRI recognition sequences, *J. Biol. Chem. 261*, 7270-8.

101. Hamy, F., Asseline, U., Grasby, J., Iwai, S., Pritchard, C., Slim, G., Butler, P. J., Karn, J., and Gait, M. J. (1993) Hydrogen-bonding contacts in the major groove are required for human immunodeficiency virus type-1 tat protein recognition of TAR RNA, *J. Mol. Biol. 230*, 111-23.

102. Pastor, N., MacKerell, A. D., Jr., and Weinstein, H. (1999) TIT for TAT: the properties of inosine and adenosine in TATA box DNA, *J. Biomol. Struct. Dyn. 16*, 787-810.

103. Tuschl, T., Ng, M. M., Pieken, W., Benseler, F., and Eckstein, F. (1993) Importance of exocyclic base functional groups of central core guanosines for hammerhead ribozyme activity, *Biochemistry 32*, 11658-68.

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

105. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank, *Nucleic Acids Res. 28*, 235-42.

106. Lan, T., and McLaughlin, L. W. (2001) Minor groove functional groups are critical for the B-form conformation of duplex DNA, *Biochemistry 40*, 968-76.

107. Sun, Z., Chen, D., Lan, T., and McLaughlin, L. W. (2002) Importance of minor groove functional groups for the stability of DNA duplexes, *Biopolymers 65*, 211-7.

108. Chiu, T. K., and Dickerson, R. E. (2000) 1 A crystal structures of B-DNA reveal sequence-specific binding and groove-specific bending of DNA by magnesium and calcium, *J. Mol. Biol. 301*, 915-45.

109. Minasov, G., Tereshko, V., and Egli, M. (1999) Atomic-resolution crystal structures of B-DNA reveal specific influences of divalent metal ions on conformation and packing, *J. Mol. Biol. 291*, 83-99.

110. Watson, J., Hays, F. A., and Ho, P. S. (2004) Definitions and Analysis of DNA Holliday Junction Geometry, *Nucleic Acids Res. 32*, 1-11.

111. Clegg, R. M., Murchie, A. I., Zechel, A., and Lilley, D. M. (1993) Observing the helical geometry of double-stranded DNA in solution by fluorescence resonance energy transfer, *Proc. Natl. Acad. Sci. U S A 90*, 2994-8.

112. Merritt, E. A., and Bacon, D. J. (1997) Raster3D: Photorealistic Molecular Graphics, *Methods Enzymol. 277*, 505-524.

113. Wellenzohn, B., Flader, W., Winger, R. H., Hallbrucker, A., Mayer, E., and Liedl, K. R. (2001) Exocyclic groups in the minor groove influence the backbone conformation of DNA, *Nucleic Acids Res. 29*, 5036-43.

114. Xuan, J. C., and Weber, I. T. (1992) Crystal structure of a B-DNA dodecamer containing inosine, d(CGCIAATTCGCG), at 2.4 A resolution and its comparison with other B-DNA dodecamers, *Nucleic Acids Res. 20*, 5457-64.

115. 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.

116. Hartmann, B., Piazzola, D., and Lavery, R. (1993) BI-BII transitions in B-DNA, *Nucleic Acids Res. 21*, 561-8.

117. El Hassan, M. A., and Calladine, C. R. (1998) Two distinct modes of protein-induced bending in DNA, *J. Mol. Biol. 282*, 331-43.

118. Howard, F. B., and Miles, H. T. (1984) 2NH2A X T helices in the ribo- and deoxypolynucleotide series. Structural and energetic consequences of 2NH2A substitution, *Biochemistry 23*, 6723-32.

119. Anfinsen, C. B. (1973) Principles that govern the folding of protein chains, *Science 181*, 223-30.

120. Matthews, B. W. (1996) Structural and genetic analysis of the folding and function of T4 lysozyme, *FASEB J. 10*, 35-41.

121. Ramshaw, J. A., Shah, N. K., and Brodsky, B. (1998) Gly-X-Y tripeptide frequencies in collagen: a context for host-guest triple-helical peptides, *J. Struct. Biol. 122*, 86-91.

122. Persikov, A. V., Ramshaw, J. A., Kirkpatrick, A., and Brodsky, B. (2000) Amino acid propensities for the collagen triple-helix, *Biochemistry 39*, 14960-7.

123. Lebrun, A., and Lavery, R. (1997) Unusual DNA conformations, *Curr. Opin. Struct. Biol. 7*, 348-54.

124. Minsky, A. (2004) Information content and complexity in the high-order organization of DNA, *Annu. Rev. Biophys. Biomol. Struct. 33*, 317-42.

125. Stefl, R., Wu, H., Ravindranathan, S., Sklenar, V., and Feigon, J. (2004) DNA A-tract bending in three dimensions: solving the dA4T4 vs. dT4A4 conundrum, *Proc. Natl. Acad. Sci. U S A 101*, 1177-82.

126. Marini, J. C., Levene, S. D., Crothers, D. M., and Englund, P. T. (1982) Bent Helical Structure in Kinetoplast DNA, *Proc. Natl. Acad. Sci. U S A 79*, 7664-7668.

127. Tolstorukov, M. Y., Ivanov, V. I., Malenkov, G. G., Jernigan, R. L., and Zhurkin, V. B. (2001) Sequence-dependent B<-->A transition in DNA evaluated with dimeric and trimeric scales, *Biophys. J. 81*, 3409-21.

128. 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. U S A 92*, 6464-8.

129. DiGabriele, A. D., Sanderson, M. R., and Steitz, T. A. (1989) Crystal lattice packing is important in determining the bend of a DNA dodecamer containing an adenine tract, *Proc. Natl. Acad. Sci. U S A 86*, 1816-20.

130. Dickerson, R. E., Goodsell, D. S., and Neidle, S. (1994) "...the tyranny of the lattice..." *Proc. Natl. Acad. Sci. U S A 91*, 3579-83.

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

132. Mayer-Jung, C., Moras, D., and Timsit, Y. (1998) Hydration and recognition of methylated CpG steps in DNA, *EMBO J. 17*, 2709-18.

133. Sprecher, C. A., Baase, W. A., and Johnson, W. C., Jr. (1979) Conformation and circular dichroism of DNA, *Biopolymers 18*, 1009-19.

134. Peticolas, W. L., Wang, Y., and Thomas, G. A. (1988) Some rules for predicting the base-sequence dependence of DNA conformation, *Proc. Natl. Acad. Sci. U S A 85*, 2579-83.

135. Auffinger, P., Hays, F. A., Westhof, E., and Ho, P. S. (2004) Halogen bonds in biological molecules, *Proc. Natl. Acad. Sci. USA 101*, 16789-94.

136. Okano, R., Mita, T., and Matsui, T. (1992) Characterization of a novel promoter structure and its transcriptional regulation of the murine laminin B1 gene, *Biochim Biophys. Acta. 1132*, 49-57.

137. Duckett, D. R., Murchie, A. I., and Lilley, D. M. (1990) The role of metal ions in the conformation of the four-way DNA junction, *EMBO. J. 9*, 583-90.

138. Joo, C., McKinney, S. A., Lilley, D. M., and Ha, T. (2004) Exploring rare conformational species and ionic effects in DNA Holliday junctions using single-molecule spectroscopy, *J. Mol. Biol. 341*, 739-51.

139. Liu, J., Declais, A. C., and Lilley, D. M. (2004) Electrostatic interactions and the folding of the four-way DNA junction: analysis by selective methyl phosphonate substitution, *J. Mol. Biol. 343*, 851-64.

140. Olmsted, M. C., and Hagerman, P. J. (1994) Excess counterion accumulation around branched nucleic acids, *J. Mol. Biol. 243*, 919-29.

141. Fenley, M. O., Manning, G. S., Marky, N. L., and Olson, W. K. (1998) Excess counterion binding and ionic stability of kinked and branched DNA, *Biophys. Chem. 74*, 135-52.

142. van Buuren, B. N., Hermann, T., Wijmenga, S. S., and Westhof, E. (2002) Brownian-dynamics simulations of metal-ion binding to four-way junctions, *Nucleic Acids Res. 30*, 507-14.

143. Mollegaard, N. E., Murchie, A. I., Lilley, D. M., and Nielsen, P. E. (1994) Uranyl photoprobing of a four-way DNA junction: evidence for specific metal ion binding, *EMBO J. 13*, 1508-13.

144. Porschke, D. (1976) Thermodynamic and kinetic parameters of ion condensation to polynucleotides. Outer sphere complex formed by Mg++ions, *Biophys. Chem. 4*, 383-94.

145. Mack, D. R., Chiu, T. K., and Dickerson, R. E. (2001) Intrinsic bending and deformability at the T-A step of CCTTTAAAGG: a comparative analysis of T-A and A-T steps within A-tracts, *J. Mol. Biol. 312*, 1037-49.

146. Shui, X., McFail-Isom, L., Hu, G. G., and Williams, L. D. (1998) The B-DNA dodecamer at high resolution reveals a spine of water on sodium, *Biochemistry 37*, 8341-55.

147. Drew, H. R., and Dickerson, R. E. (1981) Structure of a B-DNA dodecamer. III. Geometry of hydration, *J. Mol. Biol. 151*, 535-56.

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

149. Dickerson, R. E., and Drew, H. R. (1981) Structure of a B-DNA dodecamer. II. Influence of base sequence on helix structure, *J. Mol. Biol. 149*, 761-86.

150. Olson, W. K., Gorin, A. A., Lu, X. J., Hock, L. M., and Zhurkin, V. B. (1998) DNA sequence-dependent deformability deduced from protein-DNA crystal complexes, *Proc. Natl. Acad. Sci. U S A 95*, 11163-8.

151. Lu, X. J., El Hassan, M. A., and Hunter, C. A. (1997) Structure and conformation of helical nucleic acids: analysis program (SCHNAaP), *J. Mol. Biol. 273*, 668-80.

152. Hays, F. A., Teegarden, A., Jones, Z. J. R., Harms, M., Raup, D., Watson, J., Cavaliere, E., and Ho, P. S. (2005) How sequence defines structure: A crystallographic map of DNA structure and conformation, *Proc. Natl. Acad. Sci. U.S.A., in press*.

153. Hays, F. A., Jones, Z. J., and Ho, P. S. (2004) Influence of minor groove substituents on the structure of DNA Holliday junctions, *Biochemistry 43*, 9813-22.

154. Faerman, C. H., and Karplus, P. A. (1995) Consensus preferred hydration sites in six FKBP12-drug complexes, *Proteins 23*, 1-11.

155. Ingleston, S. M., Sharples, G. J., and Lloyd, R. G. (2000) The acidic pin of RuvA modulates Holliday junction binding and processing by the RuvABC resolvasome, *EMBO J. 19*, 6266-74.

156. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method, *Acta Crystallogr. D Biol. Crystallogr. 53*, 240-55.

157. Winn, M. D., Isupov, M. N., and Murshudov, G. N. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement, *Acta Crystallogr. D Biol. Crystallogr. 57*, 122-33.

158. Winn, M. D., Murshudov, G. N., and Papiz, M. Z. (2003) Macromolecular TLS refinement in REFMAC at moderate resolutions, *Methods Enzymol. 374*, 300-21.

159. Chan, S. N., Harris, L., Bolt, E. L., Whitby, M. C., and Lloyd, R. G. (1997) Sequence specificity and biochemical characterization of the RusA Holliday junction resolvase of Escherichia coli, *J. Biol. Chem. 272*, 14873-82.

160. Giraud-Panis, M. J., and Lilley, D. M. (1998) Structural recognition and distortion by the DNA junction-resolving enzyme RusA, *J. Mol. Biol. 278*, 117-33.

161. Fenley, M. O., Olson, W. K., Tobias, I., and Manning, G. S. (1994) Electrostatic effects in short superhelical DNA, *Biophys. Chem. 50*, 255-71.

162. Tippin, D. B., and Sundaralingam, M. (1997) Nine polymorphic crystal structures of d(CCGGGCCCGG), d(CCGGGCCm5CGG), d(Cm5CGGGCCm5CGG) and d(CCGGGCC(Br)5CGG) in three different conformations: effects of spermine binding and methylation on the bending and condensation of A-DNA, *J. Mol. Biol. 267*, 1171-85.

163. Timsit, Y., and Moras, D. (1994) DNA self-fitting: the double helix directs the geometry of its supramolecular assembly, *EMBO J. 13*, 2737-46.

164. Kielkopf, C. L., Bremer, R. E., White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., Dervan, P. B., and Rees, D. C. (2000) Structural effects of DNA sequence on T.A recognition by hydroxypyrrole/pyrrole pairs in the minor groove, *J. Mol. Biol. 295*, 557-67.

165. Prive, G. G., Yanagi, K., and Dickerson, R. E. (1991) Structure of the B-DNA decamer C-C-A-A-C-G-T-T-G-G and comparison with isomorphous decamers C-C-A-A-G-A-T-T-G-G and C-C-A-G-G-C-C-T-G-G, *J. Mol. Biol. 217*, 177-99.

166. Shakked, Z., Guzikevich-Guerstein, G., Frolow, F., Rabinovich, D., Joachimiak, A., and Sigler, P. B. (1994) Determinants of repressor/operator recognition from the structure of the trp operator binding site, *Nature 368*, 469-73.

167. van Aalten, D. M., Erlanson, D. A., Verdine, G. L., and Joshua-Tor, L. (1999) A structural snapshot of base-pair opening in DNA, *Proc. Natl. Acad. Sci. U S A 96*, 11809-14.

168. Ramakrishnan, B., and Sundaralingam, M. (1993) High resolution crystal structure of the A-DNA decamer d(CCCGGCCGGG). Novel intermolecular base-paired G*(G.C) triplets, *J. Mol. Biol. 231*, 431-44.

169. Ramakrishnan, B., Sekharudu, C., Pan, B., and Sundaralingam, M. (2003) Near-atomic resolution crystal structure of an A-DNA decamer d(CCCGATCGGG): cobalt hexammine interaction with A-DNA, *Acta Crystallogr. D Biol. Crystallogr. 59*, 67-72.

APPENDIX

Supplemental Table 1: Conformations and crystallization conditions for sequences of the type d(CCnnn$N_6N_7N_8$GG), where the $N_6N_7N_8$ core trinucleotide is varied to be any of the four standard nucleotides and nnn are nucleotides that maintain the self-complementarity of the sequences. Sequences are listed according to the unique $N_6N_7N_8$ trinucleotide sequence, along with the conformation (and reference if not from the current study), composition of the crystallization solution in terms of monovalent (M^{+1}), divalent (M^{+2}), multivalent cations (M^{+n}) and any precipitating agent in the crystallization solution. Sequences not assigned to a structural class are listed last (core sequences 30-63) and within these sequences a space group is assigned when known.

Core	Structure	Composition of Crystallization Solution				Reservoir Precipitant	Space Group	Comments (reference)
		Buffer	Salt	Spermine	Precipitant			
1. ACC	**Junction**	**75 mM NaCac**	**15 mM $CaCl_2$**	**0.1 mM**	-	**30% MPD**	***C2***	**Repeat of (*47*)**
	Junction	75 mM NaCac	15 mM $CaCl_2$	-	-	2.5% MPD	*C2*	(*47*)
	Junction	40 mM NaCac	80 mM KCl	12 mM	10% MPD	35% MPD	*C2*	2 strands asu (*79*)
2. GCC	**Junction**	**25 mM NaCac**	**7.5 mM $CaCl_2$**	**0.1 mM**	-	**28% MPD**	***C2***	**Junction (*76*)**
	B-DNA	10 mM TRIS·HCl	150 mM $MgCl_2$	-	-	(see comment)	*R3*	Microdialysis -buffer, salt, MPD (*82*)
3. ATC	**Junction**	**75 mM NaCac**	**55 mM $CaCl_2$**	**0.8 mM**	-	**30% MPD**	***C2***	**Novel junction**
	Junction	**75 mM NaCac**	**110 mM $CaCl_2$**	**0.8 mM**	-	**30% MPD**	***C2***	**Novel junction**
	B-DNA	**75 mM NaCac**	**5 mM $CaCl_2$**	**0.9 mM**	-	**30% MPD**	***$P2_12_12_1$***	**Isomorphous to AAA (*145*)**
		60 mM NaCac	**5 mM $CaCl_2$**	**1.4 mM**	-	**35% MPD**	***$P2_12_12_1$***	

Supplemental Table 1 (Continued)

4. CCC	**Junction**	**50 mM NaCac**	**10 mM $CaCl_2$**	**0.5 mM**	-	**26% MPD**	***C2***	**Novel junction**
	A-DNA	**80 mM NaCac**	**5 mM $CaCl_2$**	**0.7 mM**	-	**26% MPD**	***$P2_12_12_1$***	**Repeat**
	A-DNA	30 mM NaCac	25 mM $La(NO_3)_2$	1 mM	-	40% MPD	*$P2_12_12_1$*	(*162*)
5. AGA	**B-DNA**	**25 mM NaCac**	**40 mM $CaCl_2$**	-	-	**28% MPD**	***C2***	**Isomorphous to AGC (*47*)**
6. AGC	**B-DNA**	**25 mM NaCac**	**50 mM $CaCl_2$**	-	-	**28% MPD**	***C2***	**Repeat**
	B-DNA	20 mM NaCac.	50 mM $CaCl_2$	-	10% PEG 200	30% PEG 200	*C2*	(*47*)
	-	**25 mM NaCac**	**5 mM $CaCl_2$**	-	-	**35% MPD**	***$P3_1$***	**unsolved**
7. TTA	**B-DNA**	**25 mM NaCac**	**60 mM $CaCl_2$**	**0.2 mM**	-	**28% MPD**	***C2***	**Isomorphous w/AGA**
8. TTC	**B-DNA**	**25 mM NaCac**	**60 mM $CaCl_2$**	-	-	**28% MPD**	***C2***	
9. GCT	**B-DNA**	**40 mM NaCac**	**175 mM $CaCl_2$**	**0.5 mM**	-	**24% MPD**	***C2***	**Repeat**
	B-DNA	-	61.2 mM $CaCl_2$	-	-	30% MPD	*C2*	(*108*)
10. GCA	**B-DNA**	**25 mM NaCac**	**130 mM $CaCl_2$**	**0.4 mM**	-	**25% MPD**	***C2***	
11. GAC	**B-DNA**	**25 mM NaCac**	**50 mM $CaCl_2$**	**0.1 mM**	-	**30% MPD**	***R3***	**Isomorphous to GCC-B (*82*)**

Supplemental Table 1 (Continued)

12. CTC	**B-DNA**	**25 mM NaCac**	**50 mM $CaCl_2$**	**0.1 mM**	-	**28% MPD**	***$P3_1$***	
13. AAC	**B-DNA**	**25 mM NaCac**	**5 mM $CaCl_2$**	-	-	**28% MPD**	***$P3_221$***	
14. ACT	**B-DNA**	**50 mM NaCac**	**150 mM $CaCl_2$**	-	-	**35% MPD**	***$P6_122$***	**Crystallized at 4°C (*76*)**
	B-DNA	5 mM TRIS·HCl	50mM CaAcetate		14% MPD	28% MPD	*C2*	Crystallized at 4°C (*90*)
15. GTC	**(B-DNA)**	**25 mM NaCac**	**150 mM $CaCl_2$**	**1.0 mM**		**30% MPD**	***R3***	**Isomorphous w/GGC (*163*)**
16. AAA	**(B-DNA)**	**25 mM NaCac**	**150 mM $CaCl_2$**	**1 mM**	-	**26% MPD**	***$P2_12_12_1$***	**Isomorphous w/TCT (*164*)**
	B-DNA	-	5 mM MgAcetate	0.35 mM	12.5% MPD	30% MPD	$P2_12_12_1$	(*145*)
17. GTT	B-DNA	0.11 mM Streptongrin	8.57 mM MgAcetate	-	10-15% MPD	45% MPD	*C2*	(*108, 165*)
18. AAT	B-DNA	-	15 mM $CaCl_2$	-	15% MPD	35% MPD	$P3_221$	(*148*)
19. AGT	B-DNA	20 mM NaCac	30 mM $MgCl_2$	4 mM	10% PEG 200	30% PEG 200 + 100 mM buffer	$P3_221$	(*166*)
20. CCT	B-DNA	25 mM NaCac	100 mM $CaCl_2$	0.5 mM	2.5% MPD	35% MPD	$P2_12_12_1$	(*167*)
	B-DNA						*C2*	(*78*)
21. TCT	B-DNA	25 mM NaCac	100 mM $CaCl_2$	6 mM	2.5% MPD	35% MPD	$P3_1$	(*164*)

Supplemental Table 1 (Continued)

	B-DNA	26 mM NaCac	101 mM $CaCl_2$	7 mM	2.5% MPD	35% MPD	$P2_12_12_1$	(*164*)
22. GGC	**B-DNA**	**25 mM NaCac**	**60 mM $CaCl_2$**	**0.1 mM**	-	**28% MPD**	***R3***	**Repeat of (*163*)**
	B-DNA	50 mM NaCac	20 mM MgAcetate	1 M	-	30% MPD	*R3*	pH 6.0 (*163*)
	A-DNA	25 mM NaCac	3.5 mM MgAcetate	-	15% MPD	30% MPD, + cryst. solution	$P6_522$	pH 6.0 (*132*)
23. CCG	**A-DNA**	**25 mM NaCac**	**25 mM $CaCl_2$**	**1.3 mM**	-	**20% MPD**	**$P2_12_12_1$**	**Repeat**
	A-DNA	40 mM NaCac	-	5 mM	-	40% MPD	$P2_12_12_1$	(*168*)
24. CGG	**A-DNA**	**25 mM NaCac**	**60 mM $CaCl_2$**	**0.4 mM**	-	**28% MPD**	**$P2_12_12_1$**	**Isomorphous w/GGA**
25. GGA	**A-DNA**	**25 mM NaCac**	**50 mM $CaCl_2$**	**0.1 mM**	-	**28% MPD**	**$P2_12_12_1$**	
26. GGG	**A-DNA**	**25 mM NaCac**	**40 mM $CaCl_2$**	**2 mM**	-	**28% MPD**	**$P2_12_12_1$**	**Isomorphous w/GGA**
27. GGT	**A-DNA**	**25 mM NaCac**	**50 mM $CaCl_2$**	**1.8 mM**	-	**20% MPD**	**$P2_12_12_1$**	**Isomorphous w/GGA**
28. TGG	**A-DNA**	**25 mM NaCac**	**40 mM $CaCl_2$**	**0.1 mM**	-	**28% MPD**	**$P2_12_12_1$**	**Isomorphous w/CCG**
29. TCG	A-DNA	40 mM NaCac	0.5 mM $Co(NH_3)_6$	-	-	50% MPD	$P4_32_12$	pH 6.0 (*169*)
30. AGG		**25 mM NaCac**	**50 mM $CaCl_2$**	**0.1 mM**	-	**28% MPD**	***$P4_12_12$***	**data collected**

Supplemental Table 1 (Continued)

31. GAA		**25 mM NaCac**	**10 mM $CaCl_2$**	**0.5 mM**	**-**	**26% MPD**	***C2***	**data collected**
32. GTG		**25 mM NaCac**	**-**	**1.3 mM**	**-**	**20% MPD**	***$P4_1$***	**data collected**
33. TGC		**25 mM NaCac**	**50 mM $CaCl_2$**	**1.0 mM**	**-**	**26% MPD**	***$P6_222$***	**data collected**
34. TTT		**25 mM NaCac**	**180 mM $CaCl_2$**	**1 mM**	**-**	**28% MPD**	***$P2_12_12_1$***	**data collected**
35. CAC		**25 mM NaCac**	**200 mM $CaCl_2$**	**0.7 mM**	**-**	**20% MPD**	***$P2_12_12_1$***	**data collected**
36. CTA		**25 mM NaCac**	**60 mM $CaCl_2$**	**0.2 mM**	**-**	**28% MPD**	***$C222_1$***	**data collected**
37. AAG		**25 mM NaCac**	**325 mM $CaCl_2$**	**1.0 mM**	**-**	**26% MPD**		**crystallized**
38. ATA		**25 mM NaCac**	**55 mM $CaCl_2$**	**0.7 mM**	**-**	**26% MPD**		**crystallized**
39. ATG		**25 mM NaCac**	**10 mM $CaCl_2$**	**1.2 mM**	**-**	**20% MPD**		**crystallized**
40. ATT		**25 mM NaCac**	**50 mM $CaCl_2$**	**1.0 mM**	**-**	**26% MPD**	***$P2_1$***	**crystallized**
41. ACA		**50 mM NaCac**	**100 mM $CaCl_2$**	**1.0 mM**	**-**	**24% MPD**		**crystallized**
42. ACG		**40 mM NaCac**	**30 mM $CaCl_2$**	**0.3 mM**	**-**	**22% MPD**		**crystallized**
43. GAG		**25 mM NaCac**	**120 mM $CaCl_2$**	**0.1 mM**	**-**	**28% MPD**		**crystallized**

Supplemental Table 1 (Continued)

44. GAT		**25 mM NaCac**	**30 mM $CaCl_2$**	**1.1 mM**	**-**	**30% MPD**		**crystallized**
45. GTA		**25 mM NaCac**	**70 mM $CaCl_2$**	**1.0 mM**	**-**	**12% MPD**		**crystallized**
46. GCG		**50 mM NaCac**	**10 mM $CaCl_2$**	**0.8 mM**	**-**	**28% MPD**		**crystallized**
47. TAA		**30 mM NaCac**	**10 mM $CaCl_2$**	**0.5 mM**	**-**	**22% MPD**		**crystallized**
48. TAG		**30 mM NaCac**	**30 mM $CaCl_2$**	**0.3 mM**	**-**	**26% MPD**		**crystallized**
49. TAT		**25 mM NaCac**	**150 mM $CaCl_2$**	**1.0 mM**	**-**	**30% MPD**	***$P2_1$***	**crystallized**
50. TAC		**25 mM NaCac**	**20 mM $CaCl_2$**	**0.5 mM**	**-**	**28% MPD**	***C2***	**crystallized**
51. TGA		**25 mM NaCac**	**80 mM $CaCl_2$**	**0.5 mM**	**-**	**22% MPD**		**crystallized**
52. TGT		**25 mM NaCac**	**80 mM $CaCl_2$**	**0.2 mM**	**-**	**24% MPD**		**crystallized**
53. TTG		**50 mM NaCac**	**5 mM $CaCl_2$**	**0.7 mM**	**-**	**26% MPD**		**crystallized**
54. TCA		**25 mM NaCac**	**5 mM $CaCl_2$**	**0.1 mM**	**-**	**28% MPD**		**crystallized**
55. TCC		**40 mM NaCac**	**12 mM $CaCl_2$**	**1.0 mM**	**-**	**10% MPD**		**crystallized**
56. CAA		**25 mM NaCac**	**75 mM $CaCl_2$**	**1.3 mM**	**-**	**26% MPD**	***$P2_1$***	**crystallized**

Supplemental Table 1 (Continued)

57. CAG		**25 mM NaCac**	**90 mM $CaCl_2$**	**0.1 mM**	**-**	**20% MPD**		**crystallized**
58. CAT		**25 mM NaCac**	**30 mM $CaCl_2$**	**0.8 mM**	**-**	**30% MPD**		**crystallized**
59. CGT		**25 mM NaCac**	**60 mM $CaCl_2$**	**0.6 mM**	**-**	**20% MPD**		**crystallized**
60. CGC		**25 mM NaCac**	**325 mM $CaCl_2$**	**0.1 mM**	**-**	**26% MPD**		**crystallized**
61. CTG		**25 mM NaCac**	**-**	**0.9 mM**	**-**	**20% MPD**	***$P2_1$***	**crystallized**
62. CTT		**25 mM NaCac**	**50 mM $CaCl_2$**	**0.1 mM**	**-**	**28% MPD**		**crystallized**
63. CCA		**60 mM NaCac**	**100 mM $CaCl_2$**	**0.7 mM**	**-**	**24% MPD**		**crystallized**

Supplemental Table 2: Crystallographic parameters for sequences of the type d(CCnnn$N_6N_7N_8$GG), where the $N_6N_7N_8$ core trinucleotide is varied to be any of the four standard nucleotides and nnn are nucleotides that maintain the self-complementarity of the sequences. Sequences are listed according to the unique $N_6N_7N_8$ trinucleotide sequence, along with the Protein Data Bank accession number, structural class (and reference if not from the current study), crystal morphology as specified by the space group, unit cell parameters (only the unique nonorthorhombic angle is given), resolution, overall completeness of the data, R_{sym} of the data, the final values of R_{cryst} and R_{free} for the refined structure, and any comments concerning the assignment of the conformation. Structures determined in this study or from this laboratory are in bold. Most new structures (those without references) have not been fully refined to their lowest R_{cryst} and R_{free} values. Structures from this study that are isomorphous (within 8% of unit cell parameters) with structures that have already been determined in this or previous studies may not be fully refined. For sequences without an assigned structural class we have collected full datasets and are in the process of obtaining a structural solution.

Core	PDB Accession	Structure	Space Group	Unit Cell Parameters	Crystallographic data			R_{cryst} (%)	R_{free} (%)	Comments (reference)
					Resolution	Completeness	R_{merge}			
1. ACC	**1ZEZ**	**Junction**	***C2***	**a=66.0 Å, b=24.1 Å, c=36.9 Å, β= 114.8°**	**1.95 Å**	**91.8%**	**7.4**	**22.9**	**27.9**	**Repeat of (1)**
	1DCW	Junction	*C2*	a=66.5Å, b=23.5 Å, c=76.9 Å, β=114.8°	2.1 Å	97.4%	4.5	23.0	31.7	(1)
	1JUC	Junction	*C2*	a=64.9 Å, b=25.4 Å, c=37.4 Å, β=110.57°	2.35 Å	96.4%	6.4	22.6	26.8	2 strands asu (2)
2. GCC	**1P4Y**	**Junction**	***C2***	**a=66.5 Å, b=24.2 Å, c=37.0 Å, β=110.0°**	**1.70 Å**	**78.7%**	**4.9**	**22.9**	**26.0**	**Junction (3)**
	1CGC	B-DNA	*R3*	a=53.7 Å, b=53.7 Å, c=44.6 Å, γ=120.0°	2.2 Å	-	-	16.7	-	(4)
3. ATC	**1ZF4**	**Junction**	***C2***	**a=64.9 Å, b=24.0 Å, c=40.3 Å, β=118.7°**	**1.65 Å**	**79.3%**	**5.5**	**23.1**	**27.5**	**Novel junction**
	1ZF3	**Junction**	***C2***	**a=65.0 Å, b=24.6 Å, c=37.5 Å, β=112.1°**	**1.85 Å**	**98.6%**	**6.1**	**20.2**	**23.1**	**Novel junction**

Supplemental Table 2 (Continued)

	1ZFC	**B-DNA**	***$P2_12_12_1$***	**a=35.2 Å, b=39.1 Å, c=33.0 Å**	**1.61 Å**	**98.9%**	**5.1**	**26.0**	**27.0**	**Isomorphous to AAA (5)**
	-	-	***$P2_12_12_1$***	**a=33.1 Å, b=40.2 Å, c=46.3 Å**	**1.85 Å**	**91.7%**	**15.5**	-	-	**unsolved**
4. CCC	**1ZF2**	**Junction**	***C2***	**a=65.0 Å, b=24.7 Å, c=36.8 Å, β=110.6°**	**1.95 Å**	**88.9%**	**10.3**	**23.1**	**27.3**	**Novel junction**
	1ZF1	**A-DNA**	***$P2_12_12_1$***	**a=24.8 Å, b=44.4 Å, c=46.3 Å**	**1.18 Å**	**93.5%**	**2.8**	**22.2**	**24.5**	**Repeat**
	321D	A-DNA	*$P2_12_12_1$*	a=24.9 Å, b=44.9 Å, c=48.1 Å	2.15 Å	77%	3.4	16.2	18.6	(6)
5. AGA	**1ZEW**	**B-DNA**	***C2***	**a=63.7 Å, b=24.6 Å, c=40.6 Å, β=122.6°**	**2.25 Å**	**88.6%**	**5.7**	**24.8**	**29.8**	**Isomorphous to AGC (1)**
6. AGC	**1ZFM**	**B-DNA**	***C2***	**a=63.9 Å, b=24.3 Å, c=40.4 Å, β=122.6°**	**2.20 Å**	**92.3%**	**4.8**	**25.9**	**34.0**	**Repeat**
	1DCV	B-DNA	*C2*	a=64.1 Å, b=25.9 Å, c=39.9 Å, β=122.0°	2.5 Å	98.2%	5.0	20.7	31.7	(1)
	-	-	***$P3_1$***	**a=b=111.3 Å, c=33.4 Å, γ=122.0°**	**2.0 Å**	**96.8%**	**7.5**	-	-	**unsolved**
7. TTA	**1ZFH**	**B-DNA**	***C2***	**a=61.8 Å, b=24.3 Å, c=41.9 Å, β=122.7°**	**2.51 Å**	**99.0%**	**3.8**	**32.0**	**34.1**	**Isomorphous w/AGA**
8. TTC	**1ZFF**	**B-DNA**	***C2***	**a=34.3 Å, b=20.3 Å, c=33.5 Å, β=112.0°**	**0.94 Å**	**81.1%**	**4.6**	**42.0**	**41.2**	**Resolution Extended**
	-	-	***C2***	**a=62.1 Å, b=26.5 Å, c=67.4 Å, β=95.2°**	**2.0 Å**	**94.0%**	**5.5**	-	-	**unsolved**
9. GCT	**1ZF5**	**B-DNA**	***C2***	**a=31.7 Å, b=25.7 Å, c=34.0 Å, β=116.7°**	**0.99 Å**	**89.2%**	**3.3**	**25.5**	**24.2**	**Repeat**
	1EN9	B-DNA	*C2*	a=32.2 Å, b=25.4 Å, c=34.2 Å, β=117.2°	0.98 Å	93.3%	3.1	14.1	15.1	(7)

Supplemental Table 2 (Continued)

10. GCA	**1ZFE**	**B-DNA**	***C2***	**a=64.0 Å, b=23.8 Å, c=46.1 Å, β=126.1°**	**2.35 Å**	**90.9%**	**5.6**	**27.5**	**30.7**	
11. GAC	**1ZF7**	**B-DNA**	***R3***	**a=53.4 Å, b=53.4 Å, c=43.9 Å, γ=120.0°**	**1.05 Å**	**93.8%**	**5.1**	**27.6**	**28.8**	**Isomorphous to GCC-B (4)**
12. CTC	**1ZFG**	**B-DNA**	***$P3_1$***	**a=33.8 Å, b=33.8 Å, c=47.1 Å, γ=120.0°**	**1.71 Å**	**92.3%**	**25.2**	**30.0**	**31.8**	
13. AAC	**1ZF0**	**B-DNA**	***$P3_221$***	**a=33.3 Å, b=33.3 Å, c=93.6 Å, γ=120.0°**	**1.50 Å**	**67.6%**	**7.9**	**23.7**	**26.0**	
14. ACT	**1P4Z**	**B-DNA**	***$P6_122$***	**a=33.4 Å, b=33.4 Å, c=87.7 Å, γ=120.0°**	**2.00 Å**	**94.6%**	**4.8**	**23.6**	**26.8**	**(3)**
	1D8G	B-DNA	*C2*	a=31.7 Å, b=25.7 Å, c=34.2 Å, β=116.9°	0.74 Å	90.0%	5.6	10.5	13.1	(8)
15. GTC	**-**	**(B-DNA)**	***R3***	**a=52.2 Å, b=52.2 Å, c=46.25 Å**	**3.00 Å**	**79.6%**	**17.7**	**-**	**-**	**Isomorphous w/GGC (9)**
16. AAA	**-**	**(B-DNA)**	***$P2_12_12_1$***	**a=33.3 Å, b=40.0 Å, c=44.4 Å**	**3.00 Å**	**64.5%**	**21.2**	**-**	**-**	**Isomorphous w/TCT (10)**
	1IKK	B-DNA	*$P2_12_12_1$*	a=35.6 Å, b=38.7 Å, c=32.5 Å	1.60 Å	99.2%	8.4	17.9	23.6	(5)
17. GTT	1EN3	B-DNA	*C2*	a=32.0 Å, b=25.4 Å, c=33.6 Å, β=113.0°	0.98 Å	95.0%	4.7	14.1	15.9	(7, 11)
18. AAT	167D	B-DNA	*$P3_221$*	a=33.2 Å, b=33.2 Å, c=96.0 Å, γ=120.0°	2.30 Å	91.0%	-	20.0	-	(12)
19. AGT	BDJ061	B-DNA	*$P3_221$*	a=24.9 Å, b=32.9 Å, c=95.1 Å, γ=120.0°	1.95 Å	-	8.7	17.0	-	(13)
20. CCT	1CW9	B-DNA	*$P2_12_12_1$*	a=34.4 Å, b=43.2 Å, c=72.4 Å	1.55 Å	95.5%	7.8	20.3	23.2	(14)
	1BD1	B-DNA	*C2*	a=32.2 Å, b=25.5 Å, c=34.8 Å, β=116.7°	1.60 Å	-	-	16.0	-	(15)

Supplemental Table 2 (Continued)

21. TCT	1CVX	B-DNA	$P3_1$	a=34.2 Å, b=34.2 Å, c=46.1 Å, γ=120.0°	2.27 Å	96.1%	2.3	23.7	23.8	(10)
	1CVY	B-DNA	$P2_12_12_1$	a=34.5 Å, b=38.6 Å, c=46.1 Å	2.15 Å	97.2%	4.1	22.9	22.9	(10)
22. GGC	**1ZFB**	**B-DNA**	***R3***	**a=53.6 Å, b=53.6 Å, c=44.8 Å, γ=120.0°**	**1.75 Å**	**99.5%**	**5.7**	**23.3**	**28.3**	**Repeat of (9)**
	1QC1	B-DNA	*R3*	a=53.7 Å, b=53.7 Å, c=45.0 Å, γ=120.0°	2.5 Å	-	7.8	14.7	-	(9)
	382D	A-DNA	$P6_522$	a=41.5 Å, b=41.5 Å, c=57.5 Å, γ=120.0°	2.2 Å	88.9%	5.6	19.7	-	(16)
23. CCG	**1ZEX**	**A-DNA**	***$P2_12_12_1$***	**a=24.0 Å, b=45.1 Å, c=44.9 Å**	**1.65 Å**	**93.8%**	**4.9**	**26.2**	**26.1**	**Repeat**
	160D	A-DNA	$P2_12_12_1$	a=24.9 Å, b=44.6 Å, c=46.9 Å	1.7 Å	-	-	18.5	-	(17)
24. CGG	**1ZEY**	**A-DNA**	***$P2_12_12_1$***	**a=24.5 Å, b=44.3 Å, c=47.7 Å**	**1.65 Å**	**84.8%**	**7.8**	**26.6**	**32.7**	**Isomorphous w/GGA**
25. GGA	**1ZFA**	**A-DNA**	***$P2_12_12_1$***	**a=23.9 Å, b=44.6 Å, c=46.8 Å**	**1.50 Å**	**97.9%**	**11.2**	**23.5**	**28.3**	
26. GGG	**1ZF9**	**A-DNA**	***$P2_12_12_1$***	**a=23.2 Å, b=45.3 Å, c=47.6 Å**	**1.38 Å**	**92.2%**	**3.7**	**23.7**	**25.9**	**Isomorphous w/GGA**
27. GGT	**1ZF8**	**A-DNA**	***$P2_12_12_1$***	**a=23.8 Å, b=44.9 Å, c=49.6 Å**	**1.48 Å**	**90.2%**	**7.3**	**21.9**	**26.3**	**Isomorphous w/GGA**
28. TGG	**1ZF6**	**A-DNA**	***$P2_12_12_1$***	**a=25.2 Å, b=41.2 Å, c=44.4 Å**	**1.45 Å**	**93.0%**	**5.1**	**25.6**	**30.6**	**Isomorphous w/AHJ060**
29. TCG	1M77	A-DNA	$P4_32_12$	a=44.3 Å, b=44.3 Å, c=57.5 Å, γ=120.0°	1.25 Å	90.0%	5.5	16.3	18.5	(18)
30. AGG	-	-	***$P4_12_12$***	**a=b= 45.0 Å, c= 77.1 Å**	**1.80 Å**	**99.5 %**	**6.2**	-	-	**data collected**

Supplemental Table 2 (Continued)

31. GAA	-	-	*C2*	a=61.9 Å, b=26.0 Å, c=41.1 Å, β=124.9°	3.10 Å	67.0 %	17.8	-	-	data collected
32. GTG	-	-	$P4_1$	a=b=36.8 Å, c=87.1 Å	3.50 Å	92.6 %	7.9	-	-	data collected
33. TGC	-	-	$P6_222$	a=b=356.3 Å, c=32.6 Å	2.00 Å	97.9 %	8.5	-	-	data collected
34. TTT	-	-	$P2_12_12_1$	a=33.0 Å, b=39.4 Å, c=46.8 Å	1.40 Å	88.1 %	26.3	-	-	data collected
35. CAC	-	-	$P2_12_12_1$	a=50.4 Å, b=53.5 Å, c=56.7 Å	2.30 Å	54.7 %	9.4	-	-	data collected
36. CTA	-	-	$C222_1$	a=62.0 Å, b=68.7 Å, c=68.6 Å	1.60 Å	86.0 %	4.2	-	-	data collected

Supplemental Table 3: Conformational parameters for sequences of the type d(CCnnn$N_6N_7N_8$GG), where the $N_6N_7N_8$ core trinucleotide is varied to be any of the four standard nucleotides and 'nnn' are nucleotides that maintain the self-complementarity of the sequences. Sequences are listed according to the unique $N_6N_7N_8$ trinucleotide sequence, along with the Protein Data Bank accession number and the corresponding translational or rotational displacement parameters. Values for each parameter were calculated using the 3DNA program (*88*) and then averaged across the entire structure. The overall average and standard deviation for each displacement parameter is listed for the B-DNA, A-DNA and Holliday junction structural classes.

CORE	NDB Accession	Helical Twist (°)	Roll (°)	Tilt (°)	Propeller Twist (°)	Buckle (°)	Rise (Å)	X-disp (Å)	Slide (Å)	Inclination (°)	Zp (Å)
B-DNA Duplexes											
GCA	1ZFE	38.2	-0.12	-0.43	-12.9	-4.94	3.39	1.43	1.32	1.71	-0.90
TCT	1CVX	35.0	4.24	0.11	-15.1	-0.55	3.29	0.37	0.76	7.36	-0.77
	1CVY	35.0	4.63	0.35	-13.4	-1.14	3.31	0.21	0.82	9.19	-0.77
GTT	5DNB	35.3	2.23	0.00	-10.3	0.00	3.31	0.45	0.78	5.10	-0.85
	1EN33	35.1	2.67	0.00	-10.5	0.00	3.24	0.25	0.73	6.14	-0.86
GGC	1QC1	34.7	2.84	0.08	-16.9	1.07	3.37	0.45	0.54	4.70	-0.42
	1ZFB	35.2	3.29	-0.21	-10.4	-3.74	3.33	0.71	0.86	6.31	-0.69
GCT	1ENE6	35.6	2.63	0.00	-10.2	0.00	3.27	0.62	1.07	7.45	-1.08
	1EN9	35.2	2.28	0.00	-10.2	0.00	3.29	0.65	1.07	6.76	-1.03
	1ZF5	35.8	1.86	0.00	-10.5	0.00	3.29	0.71	1.01	5.88	-0.86
GCC	1CGC	35.7	3.40	1.03	-9.26	0.38	3.39	0.62	0.76	6.25	-0.40
GAC	1ZF7	35.3	1.77	0.23	-12	-0.44	3.35	0.45	0.72	5.04	-0.66
CTC	1ZFG	36.4	5.33	0.62	-10.6	-2.27	3.32	0.39	0.61	7.35	-0.65
CCT	1CW9	35.4	1.68	0.00	-8.49	0.00	3.35	0.80	1.02	5.22	-0.81
	1BD1	35.4	1.68	0.00	-8.49	0.00	3.35	0.80	1.02	5.22	-0.81
ATC	1ZFC	36.8	-0.08	-0.06	-17.3	7.35	3.34	-0.16	-0.07	0.86	-0.26
AGT	167D	33.7	1.13	0.08	-7.84	-1.06	3.24	-0.31	-0.08	1.75	-0.33
AGC	1DCV	36.8	0.26	-0.70	-8.76	-1.50	3.40	0.94	0.72	0.57	-0.35
AGA	1ZEW	38.2	-0.58	-0.79	-15.3	-3.00	3.39	1.68	1.30	0.54	-0.82
ACT	1D8G	35.5	2.63	0.00	-9.26	0.00	3.29	0.55	1.04	7.66	-1.02
	1P4Z	35.1	3.14	0.00	-12.1	0.00	3.17	0.09	0.51	5.43	-0.53
AAT	BDJ055	33.7	2.96	0.45	-14.2	1.05	3.29	-0.42	0.04	5.03	-0.44

Supplemental Table 3 (Continued)

AAC	1ZF0	34.5	2.07	-1.78	-9.68	-4.13	3.22	0.43	0.45	3.41	-0.49
AAA	1IKK	37.3	0.31	-0.10	-15.5	5.37	3.29	0.05	0.09	0.91	-0.42
OVERALL AVERAGE		**35.6**	**2.18**	**-0.05**	**-11.6**	**-0.31**	**3.31**	**0.49**	**0.71**	**4.83**	**-0.68**
STANDARD DEVIATION		**1.16**	**1.51**	**0.52**	**2.81**	**2.62**	**0.06**	**0.48**	**0.40**	**2.53**	**0.24**
Holliday Junctions											
GCC	1P4Y	37.2	1.77	-0.29	-8.86	-3.27	3.43	2.54	1.59	2.15	-0.72
CCC	1ZF2	38.0	2.15	-1.28	-14.2	-7.19	3.37	2.60	1.60	2.10	-0.97
ATC	1ZF4	37.8	1.53	0.47	-13.0	0.91	3.34	2.83	1.63	1.99	-1.08
	1ZF3	37.2	2.28	0.20	-16.5	0.37	3.38	4.17	1.77	2.11	-1.15
ACC	1JUC	37.6	0.92	0.72	-7.52	2.19	3.40	1.01	1.40	1.89	-0.86
	1DCW	37.3	2.40	-0.23	-9.61	-0.93	3.40	3.64	1.67	2.24	-1.05
	1DCW	37.6	1.71	0.02	-9.52	0.00	3.40	2.53	1.67	2.29	-0.99
	1ZEZ	37.1	2.41	-0.75	-10.4	-1.39	3.44	3.63	1.67	2.12	-0.93
OVERALL AVERAGE		**37.5**	**1.90**	**-0.14**	**-11.2**	**-1.16**	**3.40**	**2.87**	**1.63**	**2.11**	**-0.97**
STANDARD DEVIATION		**0.3**	**0.52**	**0.65**	**3.04**	**2.93**	**0.03**	**0.97**	**0.11**	**0.13**	**0.14**
A-DNA Duplexes											
TGG	1ZFE	30.9	9.35	1.28	-9.95	-3.61	3.27	17.3	-1.67	-4.48	2.32
TCG	1M77	31.4	9.39	0.00	-13.0	0.00	3.35	17.2	-1.48	-4.15	2.17
GGT	1ZF8	29.4	5.76	-0.12	-4.03	3.24	3.27	11.4	-1.77	-4.49	2.45
GGG	1ZF9	28.8	6.43	1.37	-9.8	-3.99	3.23	13.2	-1.88	-4.95	2.32
GGC	382D	30.9	9.23	0.00	-11.7	0.00	3.24	17.1	-1.53	-4.38	2.35
GGA	1ZFA	30.4	4.94	1.22	-2.63	2.56	3.38	9.51	-1.91	-4.48	2.01
CGG	1ZEY	30.4	4.87	1.48	-1.87	1.10	3.39	9.31	-1.90	-4.40	1.98
CCG	160D	30.2	6.25	0.77	-8.1	-1.42	3.39	12.1	-1.79	-4.57	2.29
	1ZEX	30.1	6.79	0.89	-8.27	-2.58	3.30	12.9	-1.81	-4.68	2.39
CCC	321D	31.3	5.57	0.22	-8.6	-2.37	3.30	10.2	-1.63	-3.94	2.30

Supplemental Table 3 (Continued)

	1ZF1	30.5	7.10	-0.29	-8.09	0.83	3.24	13.7	-1.76	-4.66	2.38
OVERALL AVERAGE		**30.4**	**6.88**	**0.62**	**-7.82**	**-0.57**	**3.30**	**13.1**	**-1.74**	**-4.47**	**2.27**
STANDARD DEVIATION		**0.77**	**1.71**	**0.67**	**3.57**	**2.42**	**0.06**	**3.01**	**0.15**	**0.27**	**0.15**