AN ABSTRACT OF THE DISSERTATION OF

Brandt F. Eichman for the degree of Doctor of Philosophy in Biochemistry and Biophysics presented on September 14, 2000. Title: Crystal Structures of DNA Four-way Junctions.

Redacted for Privacy

Abstract approved _____

Pui Shing Ho

DNA four-way junctions (also known as Holliday junctions) are the primary structural intermediate during recombination, an important process responsible for biological evolution and maintenance of genomes. These junctions arise from the assembly of four nucleic acid strands to produce double-helical regions extending from a central point. Although much progress has been made in understanding the general structural and functional properties of four-way junctions, we are only now beginning to elucidate the atomic interactions that stabilize such structures. This thesis deals with the sequence- and drug-dependent formation of Holliday junctions at the atomic level, as determined by single-crystal x-ray diffraction.

The first crystal structure of a DNA four-way junction from an inverted repeat sequence is described here. The conformation of the junction corresponds to that of the stacked-X form previously characterized in solution, in which helical arms are coaxially stacked to form two pseudo-continuous duplexes related in a right-handed sense. The junction structure is held rigidly in place by a sequence-dependent network of hydrogen

bonds at the point of strand exchange, and the relative geometries of the arms are influenced by contacts further removed from the junction. Comparison of this structure with a nearly identical nucleotide sequence but containing base mismatches details both the effect of non-standard base pairing on the junction and a general structural motif in a consensus ACC-triplet that forms the core of a stable junction.

We also describe the first atomic resolution structures of oligonucleotides cross-linked with the photochemotheraputic agent psoralen, a drug used in the clinical treatment of skin disorders and one of the oldest known natural products. We find that the psoralen cross-link can induce the formation of a four-way junction in the crystal by destabilizing the duplex at the six-membered pyrone ring of the drug. These structures also further detail the nature of the junction seen in the uncross-linked sequence. In total, the structures presented in this thesis shed light on a number of important issues regarding the intrinsic stability of four-way junctions, such as branch migration and the sequence-dependent distribution of stacking conformers, as well as raise interesting questions regarding the role of recombination during repair of DNA lesions.

Crystal Structures of DNA Four-way Junctions

by

Brandt F. Eichman

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented September 14, 2000
Commencement June 2001

Doctor of Philosophy dissertation of Brandt F. Eichman presented on September 14, 2000

APPROVED:

Redacted for Privacy

Major Professor, Representing the Department of Biochemistry and Biophysics

Redacted for Privacy

Head of the Department of Biochemistry and Biophysics

Redacted for Privacy

Dean of Graduate School

I understand that my 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

Brandt F. Eichman, Author

ACKNOWLEDGMENTS

In the P. Shing Ho Lab, I thank my mentor, Dr. P. Shing Ho, for his guidance and support, and for teaching me how to ask the right questions. I thank Blaine H. M. Mooers for sharing everything he knows about experimental crystallography and nucleic acids, and for his enthusiasm towards my research. Thanks to Jeffrey M. Vargason for many good discussions about crystallography and everything except crystallogrphy in front of the graphics workstations, as well as for his computer technical support. Thanks to Beth E. Basham for much valuable advice and computer expertise. Also, I gratefully thank Christine Nguyen for her help around the lab.

I would also like to thank Dr. P. Andrew Karplus, as well as the rest of my committee for their helpful comments and discussions. Many thanks to Savvas Savvides and Zac Wood of the Karplus lab for sharing synchrotron beamtime and data collection tips. I am also indebted to my fellow colleagues and graduate students in the Department of Biochemistry and Biophysics for their friendship and love for science. Very special thanks go to Dr. John E. Hearst for his enthusiasm and support with the psoralen project.

In addition, I thank my high school chemistry teacher, Jon Valasek, as well as Dr. Maurice Eftink, Dr. Daniel Mattern, Dr. Walter Cleland, and the rest of the chemistry faculty at the University of Mississippi, who first inspired me to want to learn chemistry.

Lastly, I owe very special thanks to my parents, Frank and Judy Eichman, for my up-bringing and teaching me to always strive for excellence.

My sincere apologies to the many other people not mentioned here who also deserve my thanks.

CONTRIBUTION OF AUTHORS

P. Shing Ho was involved in the design, analysis, and writing of each manuscript. Blaine H. M. Mooers designed and determined initial crystallization conditions for the oligonucleotide sequences used in Chapters 2 and 4, and helped with the solution to the psoralen cross-linked structures. Jeffery M. Vargason was involved in data collection, structure refinement, and manuscript preparation for the material in Chapter 2, and developed the X-PLOR script used to solve the first psoralen structure. Marie Alberti performed all the psoralen cross-linking in the John Hearst lab at U.C. Berkeley.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF TABLES

DEDICATION

I dedicate this thesis to my wife Jessica. I am eternally grateful for her love, support, friendship, and advice.

Crystal Structures of DNA Four-way Junctions

Chapter 1

Introduction

DNA recombination is an important biological process which occurs during DNA replication, viral integration, and DNA repair, thus ensuring genetic diversity and genomic integrity. The four-way DNA junction, first proposed by Robin Holliday in 1964 to explain homologous recombination, is now known to be the key intermediate formed during a variety of recombination events (Broker and Lehman, 1971; Holliday, 1964; Orr-Weaver *et al.*, 1981; Potter and Dressler, 1976; Potter and Dressler, 1978; Sigal and Alberts, 1972). Holliday junctions, and four-way junctions in general, are helical intersections in which four nucleic acid strands are assembled to generate four helical arms projecting from a central location. These DNA structures are substrates for a number of enzymes that catalyze different reactions, such as branch migration and junction resolution. In RNA, four-stranded structures are important architectural tertiary interactions that ensure functionality of the molecule. Although Holliday junctions generally are known to form in a biological environment through the action of proteins, the study of four-way junctions in the absence of proteins is necessary to understand their intrinsic structural properties.

The process of homologous recombination involves the formation of heteroduplex regions of DNA through the exchange of identical, or nearly identical, DNA segments. For this to occur, one strand from each of two duplexes becomes associated with the opposite duplex (Figure 1.1). Strands that have been nicked can dissociate from their respective duplexes, cross over, and base pair with the complementary region of the adjacent duplex. Homologous pairing and strand exchange are catalyzed by the RecA protein in prokaryotes

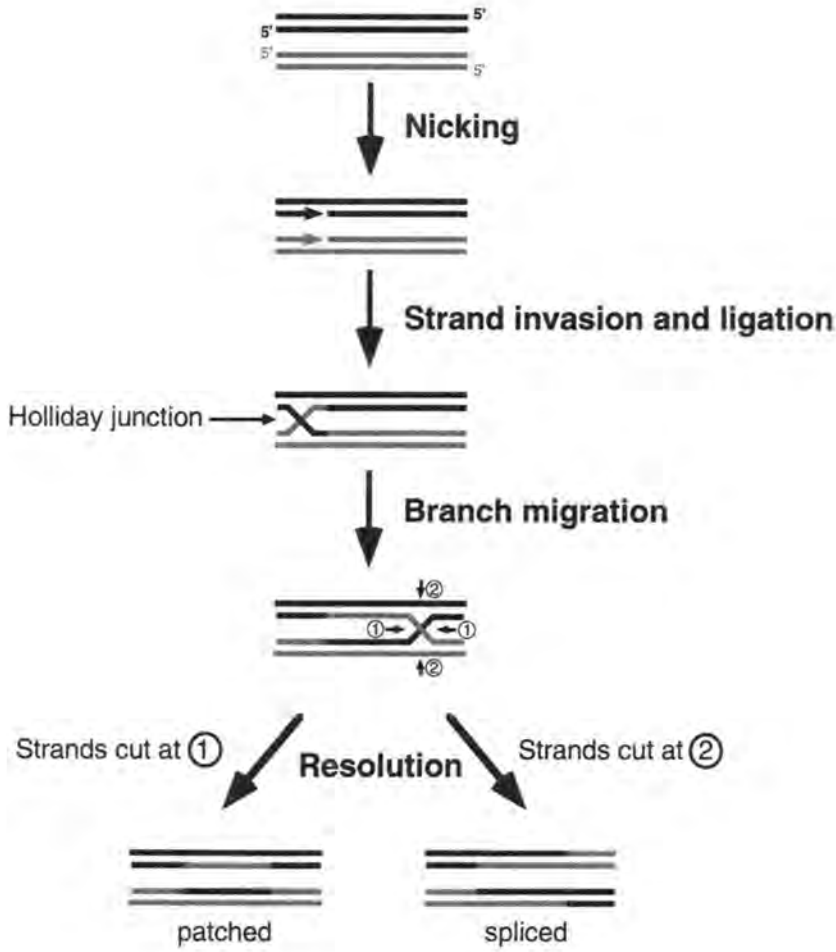


Figure 1.1. The Holliday model of homologous recombination (Holliday, 1964). Black and grey lines represent homologous DNA duplexes. Exchange of DNA strands is initiated by strand breaks in both of the homologous strands. Nicked strands dissociate from their respective duplexes and base pair with the other partner to produce heteroduplex DNA. Heteroduplex regions are elongated by movement of the junction along the DNA (branch migration). Recombination is completed by resolving the four-way junction into duplexes through nuclease cleavage of strands. Depending on the strands that are cut, recombinant (spliced) or nonrecombinant (patched) heteroduplex products are produced.

(Kowalczykowski *et al.*, 1994) and a RecA homolog, Rad51, in eukaryotes (Benson *et al.*, 1994; Shinohara *et al.*, 1992). The re-associated crossed-over strands are ligated to the new partner duplex to complete formation of the four-way junction. The heteroduplex regions are elongated by translation of the junction along the DNA, a process known as branch migration. Branch migration can occur spontaneously through a random walk process (Robinson and Seeman, 1987; Thompson *et al.*, 1976), but in biological systems is usually driven by proteins such as the RuvAB repair complex (West, 1997). Following branch migration, the two duplexes are resolved through nuclease cleavage of the junction. Depending on which strands are cut, either recombinant (spliced) or nonrecombinant (patched) products may be generated.

Four-way junctions are also formed by other mechanisms besides homologous recombination. Under conditions of negative supercoiling, they can be formed as a consequence of extruding a cruciform structure from an inverted-repeat DNA sequence (Gellert *et al.*, 1979; Lilley, 1980). Many natural RNA species, such as U1 snRNA (Branlant *et al.*, 1981) and the hairpin ribozyme (Hampel and Tritz, 1989), contain four-way helical junctions which help to define RNA function (Murchie *et al.*, 1998) presumably by providing architectural stability to the folded structure. The four-way junction is also an intermediate in site-specific recombination events, such as those catalyzed by integrase (Hoess *et al.*, 1987; Kitts and Nash, 1987). Because recombination plays an important role in DNA replication and repair, Holliday junctions are also involved in these processes (reviewed in Kowalczykowski, 2000).

A large body of work has been published to elucidate the structure and intrinsic properties of four-way junctions. The first observations of these structures were from electron microscopy, where figure-8 molecules and chi structures in a variety of systems showed four helical regions extending from a central point (Potter and Dressler, 1976; Potter and Dressler, 1978; Thompson *et al.*, 1976) (Figure 1.2a). Most of the progress in understanding the molecular properties of helical junctions has been achieved through the

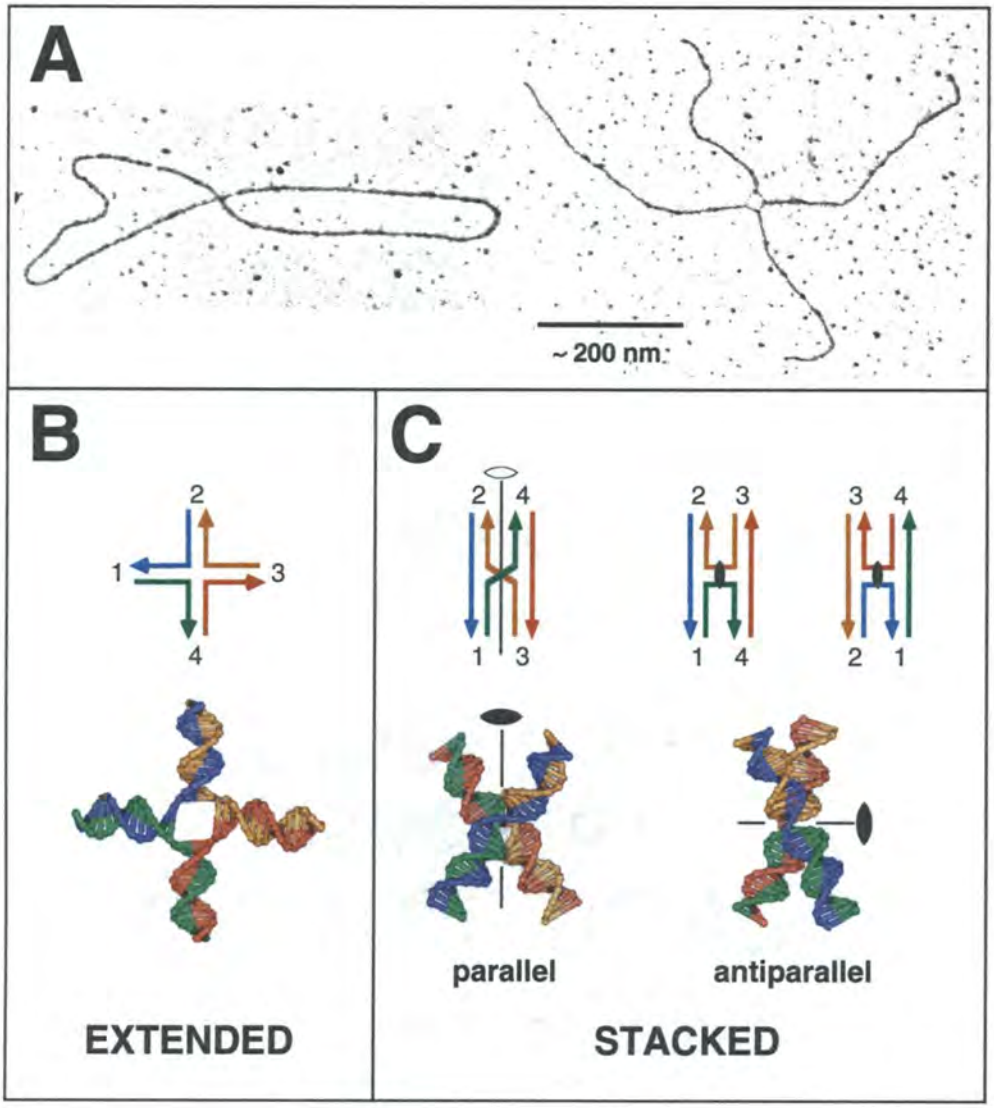


Figure 1.2. Different conformations of four-way junctions. A. Traces of original electron micrographs of plasmid DNA connected by four way junctions, showing figure-8 structures of closed circular DNA (left) and chi structures after restriction nuclease cleavage of circles (right). (Adapted from Voet and Voet, 1995). B. The unfolded, extended form of the junction. (Upper) Four DNA strands (colored blue, green, red, and orange) associate to produce a structure (lower) with four helical arms (numbered 1-4) with approximate four-fold symmetry. C. The folded, stacked-X junction. Pairs of arms stack to produce pseudo-continuous helices related by two-fold symmetry. Cross-over strands can have the same (parallel, left) or opposite (antiparallel, right) polarity. The antiparallel form is generated from a 180° rotation of one stacked duplex in the parallel form. Two conformers exist for the stacked-X junction, depending on the choice of stacking partners (far right).

design of synthetic junctions formed by the hybridization of four unique DNA strands, which traps a normally transient structure by inhibiting branch migration (Cooper and Hagerman, 1987).

Studies of these synthetic constructs in solution show that there are two predominant forms of the Holliday junction which differ in the geometry of the four helical arms. The open-X form has the four helices extended in a square-planar arrangement (Figure 1.2b), held together only by the continuity of the four DNA strands. The stacked-X form is a more compact, folded conformation with coaxial stacking of pairs of arms and two-fold symmetry (Figure 1.2c). The open formation undergoes a salt-induced transition to the stacked form (for a recent review, see Lilley, 1999). Comparative gel electrophoresis (Duckett *et al.*, 1988) and fluorescence resonance energy transfer (FRET) studies (Clegg *et al.*, 1994) have shown that at low ionic strength the open-X form predominates, and that addition of cations allow the arms to fold into the stacked form by shielding the electrostatic repulsion of phosphates at the junction.

This reduction from four-fold to two-fold symmetry upon transition to the stacked-X junction creates two structurally distinct strands (Figure 1.2c). In this structure, two stands are continuous along the same pair of stacked arms, while the other two strands cross over between the stacked duplexes. Theoretically, the arms should be able to exist in a continuum of relative orientations between a parallel arrangement, in which the uncrossed strands are lying in the same 5'-3' direction (180° interhelical angle), and an antiparallel orientation, with opposite polarity of the uncrossed strands (0° interhelical angle). In spite of modeling studies which have shown that the parallel form is sterically allowed (Sigal and Alberts, 1972; Srinivasan and Olson, 1994), a parallel stacked junction has never been experimentally observed in unconstrained constructs. On the other hand, the antiparallel arrangement is generally accepted to be the more stable form based on a large body of structural evidence.

The existence of a right-handed antiparallel stacked-X junction (Figure 1.2c) has been supported by comparative gel electrophoresis (Duckett *et al.*, 1988), FRET (Clegg *et al.*, 1992; Murchie *et al.*, 1989), and electric birefringence measurements (Cooper and Hagerman, 1989). These solution experiments, as well as atomic force microscopy studies of junctions constructed from two-dimensional crystal lattices (Mao *et al.*, 1999), estimated that the angle relating the antiparallel stacked arms is ~60° in a right-handed sense. Early modeling studies suggested that a 60° interhelical angle is the most favorable arrangement of the phosphoribose backbone of one duplex relative to the grooves of the adjacent duplex (von Kitzing *et al.*, 1990). This right-handed, antiparallel orientation was also easily modeled from duplexes packed into crystal lattices (Goodsell *et al.*, 1995; Timsit and Moras, 1991; Timsit and Moras, 1994; Wood *et al.*, 1997). These modeling studies have also shown that the Watson-Crick base pairing of nucleotides at the junction are preserved. NMR studies have confirmed this by showing that the bases across the cross-over point are indeed stacked and remain base paired, and that the DNA conformation of the arms is in the B-form (Carlström and Chazin, 1996; Chen *et al.*, 1993; Pikkemaat *et al.*, 1994).

For any particular stacked-X junction there are two conformers possible, depending on the alternative choices of stacking partners. An extended junction with arms 1-4 can fold into an antiparallel stacked structure with arm 1 stacked on arm 2, and 3 stacked on 4. Alternatively, arm 1 can be stacked on 4, and 2 stacked on 3 (Figure 1.2c). The choice of stacking partners has been shown to be dependent on the nucleotide sequence at the junction (Duckett *et al.*, 1988; Murchie *et al.*, 1989), as well as the sequence further down the arms (Miick *et al.*, 1997). The effect of the base pairs further removed from the junction suggests that the small energy difference in base stacking does not entirely account for the distribution of conformers. Generally, a given sequence tends to adopt predominantly one conformer (Carlström and Chazin, 1996), although a case has been reported in which both conformers were equally populated (Grainger *et al.*, 1998)

In the absence of proteins, branch migration of the junction is considered to be a random walk process, and is dependent on the nucleotide sequence and base stacking preferences (Robinson and Seeman, 1987; Sun *et al.*, 1998; Thompson *et al.*, 1976). This process has been shown to convert chi structures to resolved linear duplexes (Thompson *et al.*, 1976). Any movement of the arms through the junction requires breaking and re-forming base pairs. The rate of branch migration has been shown to be dependent on the concentration of cations present (Panyutin *et al.*, 1995; Panyutin and Hsieh, 1994). Branch migration occurs faster under salt conditions that favor the open-X form, suggesting that unstacking the junction facilitates this process. Enzymes that catalyze branch migration, do so by stabilizing the open-X form, as illustrated in the crystal structures of RuvA bound to four way junctions (Ariyoshi *et al.*, 2000; Hargreaves *et al.*, 1998; Roe *et al.*, 1998).

Four-way junctions are substrates for branch migration and junction resolving enzymes. These proteins recognize a particular conformation of the junction, and in most cases distort the DNA structure to an unfolded form (reviewed in White *et al.*, 1997). However, the reason for this specificity, as well as the particular DNA structure that is recognized, is not clear. It is known that resolving enzymes T7 endonuclease I, T4 endonuclease VII and RuvC bind four-way junctions as dimers, consistent with nuclease cleavage of two strands. Nucleotide sequence recognition differs among these enzymes. While T7 endonuclease I and T4 endonuclease VII cleave almost any junction, RuvC prefers to cleave at specific sequences (Fogg *et al.*, 1999). The junctions seen in cocrystals with branch migration enzymes RuvA (Ariyoshi *et al.*, 2000; Hargreaves *et al.*, 1998; Roe *et al.*, 1998) and resolving enzymes Cre (Gopaul *et al.*, 1998) are in the extended-X form, while models of junctions bound to T4 endonuclease VII (Bhattacharyya *et al.*, 1991; Raaijmakers *et al.*, 1999) suggest that these enzymes stabilize a stacked-X structure. In addition, DNAase I footprinting has established that RuvC induces an unfolded form of the junction upon binding (Bennett and West, 1995). Clearly, detailed structures of four-way

junctions both in the absence and presence of branch migration and resolving enzymes are needed to clarify the nature of junction recognition and distortion.

Although the structure of four-way junctions has been extensively studied in solution, and much effort has been spent trying to crystallize a four-way junction with the hopes of determining the structure at atomic resolution, the first crystal structures of four-way junctions in the absence of proteins have only been solved within the past year. In the process of elucidating the structure of an RNA-cleaving DNA enzyme (the DNAzyme), Nowakowski, *et al.* solved to a resolution of 3.0 Å the first structure of a four-way junction which formed from both RNA and DNA strands (Nowakowski *et al.*, 1999). In this structure, one coaxially stacked A-form helix was related to the other stacked B-DNA helix by a 55° angle. The second junction structure to be solved (to 2.16 Å resolution) was formed from four copies of the deoxyoligonucleotide sequence d(CCGGGACCGG), which was being used to study tandem d(G·A) mismatches in a B-DNA duplex (Ortiz-Lombardía *et al.*, 1999). This was the first all-DNA four-way junction to be reported, and shows stacked B-DNA duplexes related by 41°. Both of these structures contain mismatched G·A base pairs at the junction, raising a question regarding the effect of the mismatches on the ability of these sequences to crystallize. Another RNA-DNA junction from an identical sequence as the DNAzyme junction, but with a -80° interhelical angle, has recently been reported and shows the conformational variability that can occur in four-way junctions (Nowakowski *et al.*, 2000).

We describe here the structures of three four-way DNA junctions solved in this laboratory. The first crystal structure of a junction formed from a true inverted repeat DNA sequence, d(CCGGTACCGG), is described in Chapter 2, and shows the atomic interactions between the duplexes, as well as confirms several properties of four-way junctions deduced from the solution studies. A detailed comparison of this structure with the DNA mismatch junction (Chapter 3) further shows the effects of the d(G·A) mismatches and allows us to make some important generalizations regarding DNA junctions. The structures of two

Holliday junctions formed from DNA sequences cross-linked with the photochemotheraputic drug psoralen (Chapter 4) show exactly how DNA cross-linking agents are capable of inducing the formation of junctions in a crystalline environment, as well as how sequence-dependent properties of junctions can overcome perturbation from such cross-links. By studying the detailed structures of four-way junctions at the nucleic acid level, we have gained insight into the intrinsic interactions that stabilize such a structure in the absence of proteins. Several generalizations regarding the intrinsic properties of four-way junctions emerge from the crystal structures presented here, and are discussed in Chapter 5.

Chapter 2

The Holliday Junction in an Inverted Repeat DNA Sequence: Sequence Effects on the Structure of Four-way Junctions

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

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

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

2000, **97**, 3971-3976

2.1 Summary

Holliday junctions are important structural intermediates in recombination, viral integration and DNA repair. We present here the single-crystal structure of the inverted repeat sequence d(CCGGTACCGG) as a Holliday junction at the nominal resolution of 2.1 Å. Unlike the previous crystal structures, this DNA junction has B-DNA arms with all standard Watson-Crick base pairs; it therefore represents the intermediate proposed by Holliday as being involved in homologous recombination. The junction is in the stacked-X conformation, with two interconnected duplexes formed by coaxially stacked arms, and crossed at an angle of 41.4° as a right-handed X. A sequence comparison with previous B-DNA and junction crystal structures shows that an ACC trinucleotide forms the core of a stable junction in this system. The 3'-C·G base pair of this ACC core forms direct and water mediated hydrogen bonds to the phosphates at the cross-over strands. Interactions within this core define the conformation of the Holliday junction, including the angle relating the stacked duplexes and how the base pairs are stacked in the stable form of the junction.

2.2 Introduction

When genetic information is exchanged, *e.g.*, during recombination between homologous regions of chromosomes or integration of viral DNA into host genomes, the DNA double-helix is disrupted. Holliday proposed that the intermediate formed during homologous recombination is a four-way junction (Figure 2.1) (Holliday, 1964). Similar junctions form in cruciform DNAs extruded from inverted repeat sequences. Recently, the crystal structures of junctions in a DNA-RNA complex (Nowakowski *et al.*, 1999) and in the sequence d(CCGGGACCGG) (Ortiz-Lombardía *et al.*, 1999) have been reported. In

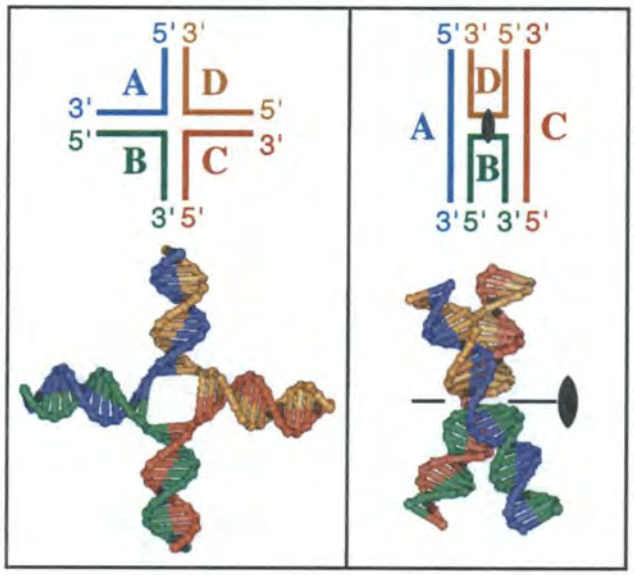


Figure 2.1. Conformations of four-way junctions. The left-hand panel shows the association of DNA strands A (blue), B (green), C (red) and D (yellow) to form a junction (top) with four duplex arms extended in a square planar geometry (extended-X form, bottom). The right-hand panel shows these same strands (top) associated to form the stacked-X structure of the junction, with pairs of arms coaxially stacked as double-helices related by two-fold symmetry (bottom).

the first structure, the DNA/RNA arms are in the A-conformation, while in the latter two G·A mismatched base pairs sit adjacent to the cross-over between the duplexes. Here, we present the structure of a Holliday junction in a true inverted repeat DNA sequence d(CCGGTACCGG) in which all the nucleotides are in B-type helices with Watson-Crick base pairs.

Studies on synthetic four-stranded complexes and DNA cruciforms show that four-way junctions can adopt either an open extended-X or the more compact stacked-X conformations (for a recent review, see reference Lilley, 1999). In the presence of monovalent cations, the four arms of the junction are extended into a square planar geometry (Figure 2.1) to minimize electrostatic repulsion between phosphates. Divalent cations and polyvalent polyamines help shield the phosphate charges (Møllegaard *et al.*, 1994), allowing the junction to adopt a more compact structure with pairs of arms coaxially stacked as duplexes (Figure 2.1) and the duplexes related by ~60° (Churchill *et al.*, 1988; Cooper and Hagerman, 1989; Duckett *et al.*, 1988; Murchie *et al.*, 1989). A 63° angle is estimated from atomic force microscopy studies on arrays of such junctions (Mao *et al.*, 1999). During recombination, four-way junctions are resolved by enzymes to complete the process of strand exchange between duplexes. The junctions seen in co-crystals with the resolving enzymes RuvA (Hargreaves *et al.*, 1998; Roe *et al.*, 1998) and Cre (Gopaul *et al.*, 1998) are in the extended-X form, while T4 endonuclease VII (Bhattacharyya *et al.*, 1991; Raaijmakers *et al.*, 1999) and T7 endonuclease I (de Massy *et al.*, 1987; Dickie *et al.*, 1987) seem to maintain the relationship of the stacked-X arms.

Despite repeated efforts over the years to crystallize a four-way DNA junction, the recent crystal structures have all been serendipitous. The RNA/DNA junction resulting from studies on an RNA-cleaving DNA motif, or DNAzyme, complexed with its RNA substrate (Nowakowski *et al.*, 1999) have arms that adopt an A-RNA conformation. The sequence d(CCGGGACCGG) designed to study tandem G·A mismatched base pairs in B-DNA also crystallized as a junction (Ortiz-Lombardía *et al.*, 1999); however, the structure

around the junction is perturbed by the mismatches. Thus we are left asking what is the structure of a Holliday junction with B-DNA arms and standard base pairs.

We had designed the sequence d(CCGGTACCGG) to study the d(TA) dinucleotide in B-DNA that is a target for the photochemotheraputic drug psoralen. Surprisingly, four strands of this DNA assembled to crystallize as a four-way junction with all Watson-Crick base pairs. We can thus examine the detailed structure, and define the nucleotides and intramolecular interactions that help to stabilize the Holliday junction.

2.3 Materials and Methods

2.3.1 Crystallization and x-ray data collection

DNA sequences were synthesized on an Applied Biosystems DNA synthesizer and purified by reverse phase HPLC. Very thin diamond shaped crystals of the sequence d(CCGGTACCGG) (dimensions of 300 x 100 x 20 μm^3) were grown at room temperature from solutions containing 0.25 mM DNA, 75 mM sodium cacodylate buffer (pH 7), 15 mM $CaCl_2$, 2.5% 2-methyl-2,4-pentanediol (MPD) and equilibrated against 30% v:v MPD. X-ray diffraction data was collected at liquid nitrogen temperatures using beamline 5.0.2 (λ = 1.1 Å) at the Advanced Light Source (ALS) in Berkeley, CA. The crystal is in the monoclinic *C2* space group, with unit cell dimensions a = 66.5 Å, b = 23.5 Å, c = 76.9 Å, and β = 114.8°. Diffraction images were processed and reflections merged and scaled using DENZO and Scalepack from the HKL package (Otwinowski and Minor, 1997). The data was limited to a nominal resolution of 2.1 Å according to $\langle I/\sigma \rangle$, completeness, and R_{merge} statistics (Table 2.1).

Similar crystals of the sequence d(CCGCTAGCGG), with dimensions of 400 x 200 x 60 μm^3, were grown at 4° C from solutions containing 0.5 mM DNA, 20 mM sodium cacodylate buffer (pH 7), 50 mM $CaCl_2$, 10% PEG 200 and equilibrated against 30% v:v

PEG 200. Data from these crystals were collected at 4° C using Cu-K_α radiation from a Rigaku RU H3R rotating anode generator and an R-AXIS IV image plate detector, with images processed and reflections merged and scaled using d*TREK from Molecular Structure Corporation (Pflugrath, 1999). This crystal is also *C2*, with unit cell dimensions a = 64.1 Å, b = 25.9 Å, c = 39.9 Å, and β = 122.0°.

2.3.2 Structure solution of d(CCGGTACCGG)

This structure was solved by molecular replacement. The native Patterson map showed that two B-type double helices (helical rise = 3.4 Å) were in the asymmetric unit (asu) with the helix axes aligned in the *xy*-plane. Thus, two structures of this sequence as B-DNA were used as the search models in EPMR (Kissinger *et al.*, 1999), resulting in correlations of 60% and R-factors of 53% for the two molecules in the asu. The models were refined using X-PLOR 3.851 (Brünger, 1992b). We observed anisotropic diffraction from these very thin crystals; therefore, anisotropic B-factor scaling was applied to properly weigh the calculated (F_c) to the observed structure factors (F_o) during refinement. The resulting $2F_o$-F_c, F_o-F_c, and annealed omit maps showed clean breaks in the electron density between A6 and C7 in each duplex, and connectivity across adjacent duplexes. The models were therefore rebuilt with the cross-overs of a four-way junction. Minimal non-crystallographic symmetry restraints were subsequently applied between the two cross-linked duplexes. The final values of R_{cryst} and R_{free} converged to 23.0% and 31.8%, respectively (Table 2.1). These values increase to 23.9% and 32.7% when the structure is refined as two non-crossed over B-DNAs.

2.3.3 Structure solution of d(CCGCTAGCGG)

The structure of d(CCGCTAGCGG) was also solved by molecular replacement. The solution from EPMR (Kissinger *et al.*, 1999) was refined using X-PLOR 3.851

Table 2.1. Data collection and refinement statistics

Data Collection	d(CCGGTACCGG)	d(CCGCTAGCGG)
Resolution Range (Å)	30.16-2.10	33.81-2.49
Measured (unique) reflections	18370 (6322)	7532 (2009)
Completeness (%)[a]	97.4 (96.7)	98.2 (91.6)
R_{merge} (%)[a,b]	4.5 (21.8)	5.2 (41.8)
$\langle I/\sigma_I \rangle$ [a]	10.6 (2.7)	12.8 (3.1)
Refinement		
Resolution Range (Å)	8-2.10	10-2.50
No. of reflections (F/σ_F cutoff)	5723 (3.0)	1943 (2.0)
Completeness[a]	88.9 (68.0)	98.6 (95.4)
R_{cryst} (R_{free})[c]	23.0 (31.8)	20.7 (31.7)
DNA atoms (solvent molecules)	808 (92)	404 (23)
Ave. B-factors (Å^2)		
DNA atoms (water atoms)	32.2 (44.6)	55.4 (63.3)
r.m.s deviation from ideality		
Bond lengths (Å) / Bond angles (°)	0.017 (1.90)	0.005 (1.00)

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

[b] $R_{merge}(I) = \Sigma_{hkl} \Sigma_i | I_{hkl,i} - \langle I \rangle_{hkl} | / \Sigma_{hkl} \Sigma_i | I_{hkl,i} |$ 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 (Brünger, 1992a). The minimum converged values of R_{free} are reported.

(Brünger, 1992b), again with anisotropic B-factor scaling applied to F_c during all steps. R_{cryst} converged to 20.7% and R_{free} to 31.7%.

The coordinates and structure factors of d(CCGGTACCGG) have been deposited in the Protein Data Bank under accession number 1DCW, and d(CCGCTAGCGG) as 1DCV.

2.4 Results

The sequence d(CCGGTACCGG) was observed to crystallize as a four-stranded Holliday junction (Figure 2.2) even though it was originally designed to study d(T·A) base pairs in B-DNA. Despite attempts to solve the structure as B-DNA double-helices, the electron density maps indicated that the duplexes were connected by the crossed-over strands of a junction. The electron density was consistently observed at all resolution limits to be discontinuous between nucleotides A6 and C7 in one strand of each duplex, but to connect these nucleotides across adjacent duplexes. The F_o-F_c maps drawn with the backbone atoms of nucleotides A6 and C7 omitted showed these same connections regardless of whether the structure was refined as two resolved double-helices or as a four-way junction (Figure 2.3a). The structure refined as a junction clearly shows the backbone trace between the adjoining duplex (Figure 2.3b).

This unusual tracing of the electron density across adjacent duplexes was neither an artifact of the crystal lattice nor of the close approach of the phosphoribose backbone in two B-DNA duplexes. The structure of the nearly identical sequence d(CCGCTAGCGG) places two symmetry related B-DNA duplexes very close to each other (with the phosphates between duplexes approaching 6.2 Å) and in nearly the same orientation as the duplexes of the junction (Figure 2.4). The electron density maps could be traced continuously through the backbone of this double-helix with no evidence for a junction (Figure 2.3c). This is

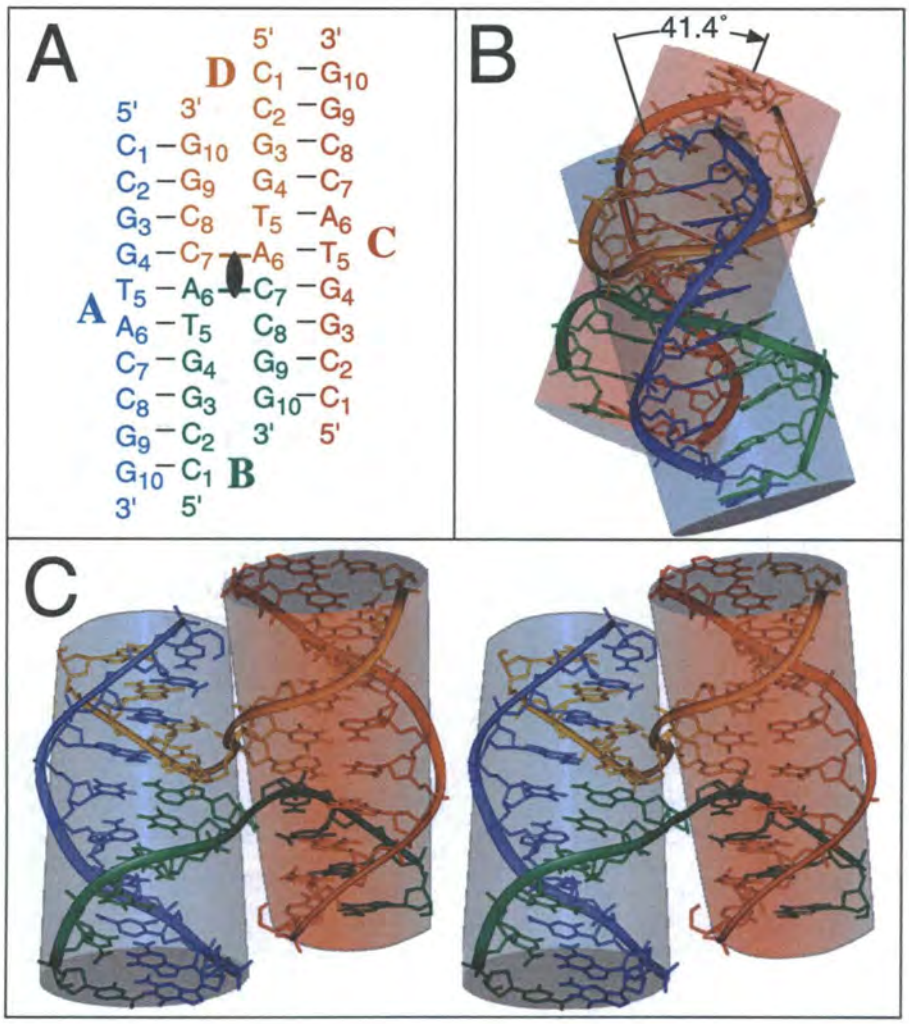


Figure 2.2. The Holliday junction structure of d(CCGGTACCGG). A. Four strands of the sequence assemble into the stacked-X conformation of a four-way junction. The strands are numbered 1 to 10 from the 5' to the 3'-termini. B. View along the Holliday junction. Two duplexes, formed by stacking arms D-A over A-B and C-D over B-C, are related by a right-handed twist of 41.4°. C. Stereoview down the two-fold axis of the junction. The B and D strands pass from one set of stacked duplexes to the neighboring duplexes.

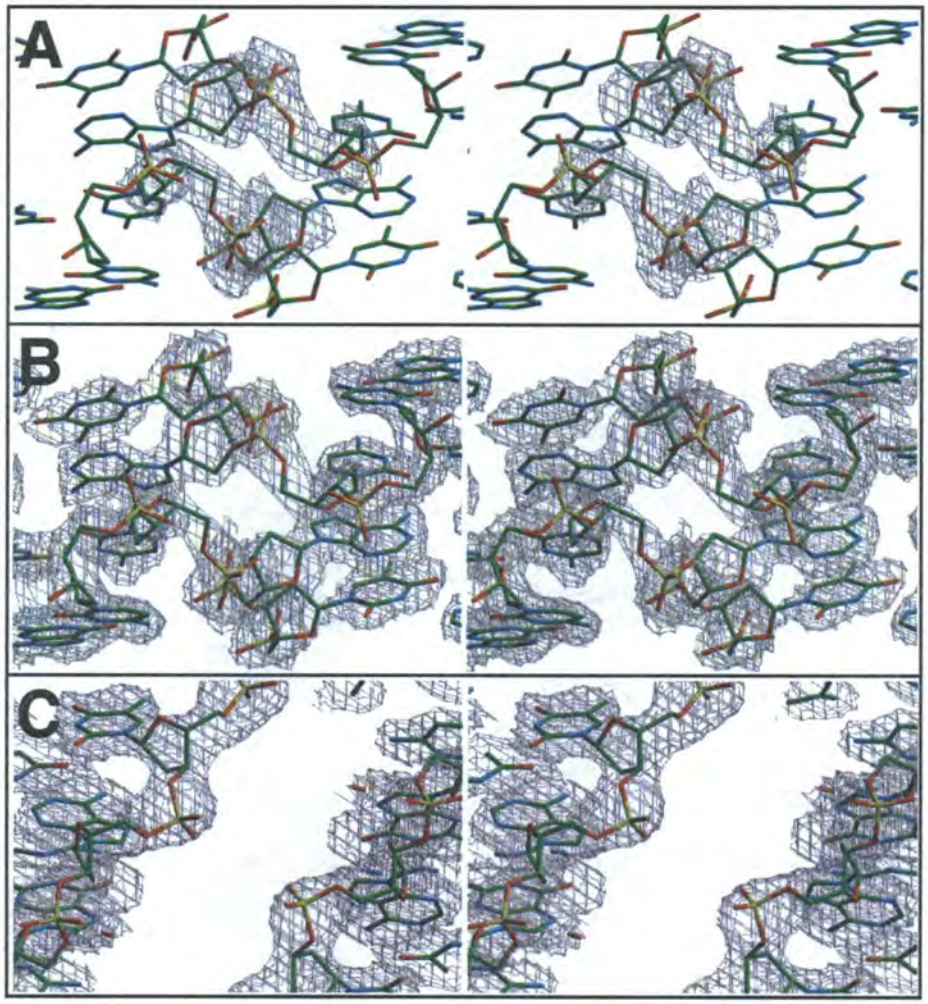


Figure 2.3. Stereoview of electron density maps from the d(CCGGTACCGG) and d(CCGCTAGCGG) structures. The atoms are colored as green carbons, red oxygens, blue nitrogens, and yellow phosphoruses. A. F_o-F_c annealed omit map of d(CCGGTACCGG) centered at nucleotides T5 to C8 of strands B and D. The map (contoured at 1.5σ) was calculated with atoms of the phosphoribose backbone of A6 and C7 omitted (but included in the figure for clarity), and the remainder of the DNA subjected to simulated annealing to eliminate model bias. This is indistinguishable from the analogous omit map calculated with a refined non-crossed-over model. B. $2F_o$-F_c electron density map calculated from the refined model of d(CCGGTACCGG) as a Holliday junction (contoured at 1.0 σ). C. $2F_o$-F_c electron density map of the refined d(CCGCTAGCGG) model as double-stranded DNA. The map is centered at nucleotides T5, A6, and C7, showing the continuity of the trace at the 1.0σ level of each strand.

further evidence that the sequence d(CCGGTACCGG) had indeed crystallized as a Holliday junction.

The structure of d(CCGGTACCGG) is a junction formed by the cross-over of strands between homologous duplexes. In a standard duplex, this would be an inverted repeat sequence. The junction resulting from the assembly of the four identical DNA strands has approximate two-fold symmetry, but with the dyad sitting between A6 and C7 rather than bisecting the sequence. Pairing the complementary nucleotides of the DNA strands A, B, C, and D results in six-base pair A-B and C-D arms, and four-base pair C-D and D-A arms.

The A-B arm is coaxially stacked on the D-A arm and B-C stacked on C-D to form the two continuous antiparallel duplexes in the stacked-X form of the junction. The conformation resembles an H with the two arms twisted 41.4° in a right-handed sense (Figure 2.2b). This angle is not as steep as the ~60° estimated from solution studies and in DNA arrays (Churchill *et al.*, 1988; Cooper and Hagerman, 1989; Duckett *et al.*, 1988; Mao *et al.*, 1999; Murchie *et al.*, 1989), and is not a consequence of the crystal lattice. The sequence $d(CCGCTAGCGG)_2$ crystallizes in a similar lattice but the B-DNA duplexes are related by a 44.1° angle.

The two double-helices across the junction are nearly identical to each other, with a root-mean-square deviation of 0.2 Å between atoms in the duplexes (0.16 Å for backbone and 0.14 Å for base atoms). Similarly, the twist, tilt, and roll angles between nucleotides are nearly identical in the two duplexes. The two duplexes are also indistinguishable from standard B-DNA, with the exception of the cross-overs at the junction. The stacking between base pairs (Figure 2.5) in and around the junction are nearly identical to those observed previously in B-DNA. The stacking of the G3·C8 and G4·C7 base pairs within the B-C and within the D-A arms are nearly identical to R·Y over R·Y base pairs (R for purine and Y for pyrimidine bases) in standard B-DNA (Dickerson, 1999). Furthermore, the stacking between the G4·C7 (of the D-A and B-C arms) and T5·A6 base pairs (of the A-

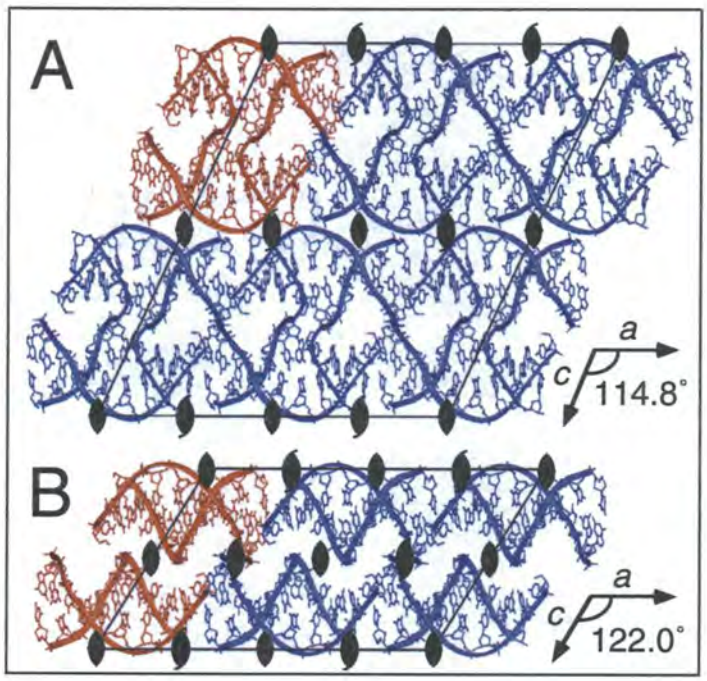


Figure 2.4. Crystal packing of the d(CCGGTACCGG) Holliday junction (A) and d(CCGCTAGCGG) B-DNA (B) structures in the *a-c* plane. Four strands in each panel are colored in red, with symmetry related DNAs in blue. In A, four DNA strands associate as a single Holliday junction in the asymmetric unit. A two-fold axis runs through the junction, but is slightly shifted from the edge of the unit cell. Since this is a non-crystallographic symmetry axis, the *c*-axis of the d(CCGGTACCGG) unit cell is extended compared to that of d(CCGCTAGCGG). In panel B, the four red strands associate to form two B-DNA duplexes related by true crystallographic two-fold symmetry.

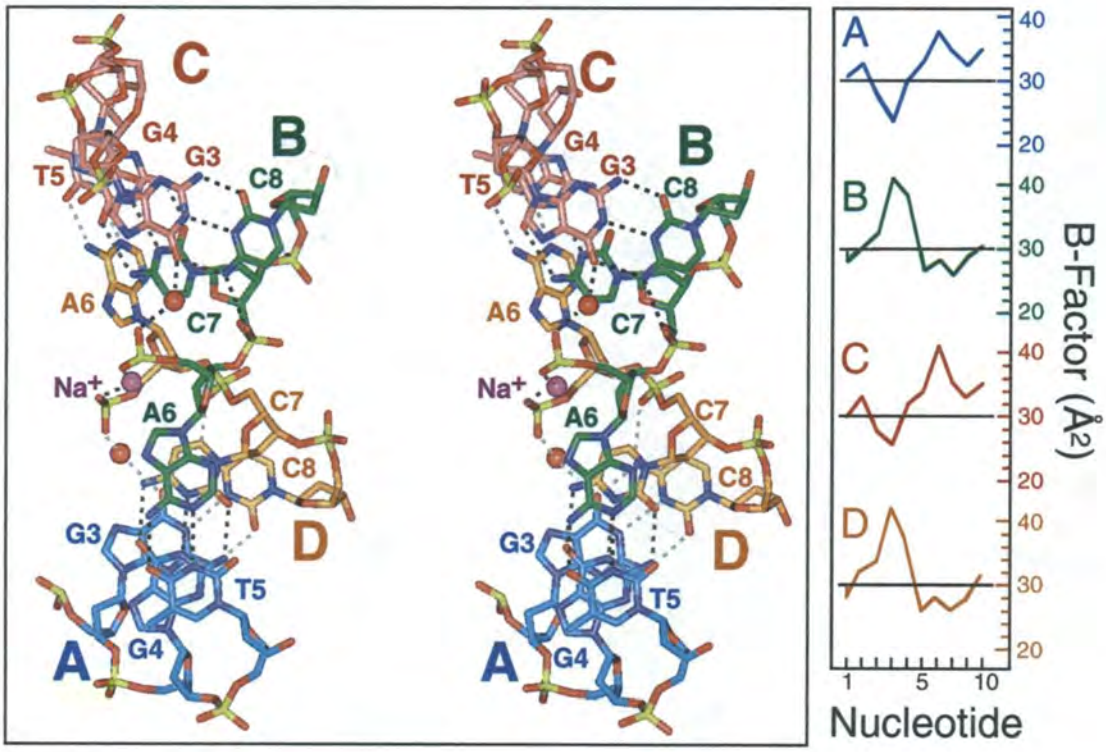


Figure 2.5. Structure of the Holliday junction. The left panel shows a stereoview of the cross-over at the junction formed by atoms of the G3·C8, G4·C7, and T5·A6 base pairs from strands A (blue carbons), B (green carbons), C (red carbons), and D (yellow carbons). The stacked base pairs of the A-B and D-A arms on the lower half are shown looking down the helix axis. The N4 nitrogens of nucleotides C8 in strands B and D are hydrogen bonded (dashed lines) to the oxygens of the phosphate linking A6 to C7 at the junction cross-over. Waters bridging the O6 keto oxygens of guanine G3 in the A and C stands to the 5'-phosphates of A6 are shown as red spheres, while the sodium ion that links the 5'-phosphates of the two A6 nucleotides is indigo. The right panel is a plot of the average B-factor ($Å^2$) for each nucleotide along strand A (blue), B (green), C (red), and D (yellow). B-factors that are below 30 $Å^2$ are colored solid. The average B-factor for the DNA model is 32.2 $Å^2$.

B and C-D arms) across the gap of the junction appears identical to previous R·Y over Y·R stacking motifs. This theme is continued in the hydration of the DNA bases. Each duplex of the structure has a network of waters in the minor groove. Although not as extensive as the spine of waters observed previously in B-DNA (this may be limited by the resolution of the current structure), the networks are continuous through the crossed-over strands. Thus, the stacked arms form B-DNA duplexes with very little disruption in base pairing, base stacking and solvent interactions by the junction. The only significant deviations from B-DNA are seen as the rotation of the χ-angles of the A6 nucleotides to place them in the high-*anti* conformation, and the nearly complete rotation in the β-dihedral angles (P-O5'-C5'-C4') from an average 146.2° ± 17.0° (calculated from the B-type nucleotides of the current structure) to -157.8° ± 3.2° of the C7 nucleotides. These perturbations extend the phosphates linking the A6 and C7 nucleotides away from their respective duplexes to form the cross-over of the junction.

The junction itself is compact and relatively rigid. The temperature factors (B-factors) of the trinucleotides A6-C7-C8 in strands B and D that span the junction and the complementary G3 and G4 nucleotides are generally lower than the remainder of the DNA (Figure 2.5). If we accept that B-factors reflect the thermal disorder of atoms in a crystal, the low B-factors of the nucleotides reflect a relatively inflexible junction. This local rigidity can be attributed to specific hydrogen bonds between the C8·G3 base pairs of the B-C and D-A arms and the phosphates at the strands that cross-over. The N4 amino group of each C8 is directly hydrogen bonded to the O2P oxygen of the phosphate linking nucleotides A6 and C7 of the crossed-over strands. In addition, a water mediates the interaction between the O6 keto oxygen of G3 and the O1P phosphate oxygen of adenine A6 (Figure 2.5).

The compactness of the junction results in a close approach of four phosphates (within 6.5 Å of each other) at the cross-over and thus a highly negative electrostatic potential that must be compensated by counterions. A single well-ordered solvent molecule (B-factor = 18.6 $Å^2$ *versus* an average of 44.6 $Å^2$ for other added solvent) sits very close (<

2.6 Å) to the oxygens of the two A6 nucleotides spanning the junction (Figure 2.5). This is closer than standard hydrogen bonding distances, suggesting that an Na^+ ion is directly coordinated to the phosphate oxygens. Modeling this as a Ca^{2+} (the other possible cation) increased both R_{cryst} and R_{free}, while a Na^+ had no effect on these values. It is well accepted that Na^+ ions are difficult to distinguish from waters at this resolution (Shui *et al.*, 1998); therefore, we relied on the short distances to the phosphates to make this assignment. We would expect additional ion interactions to help compensate for the phosphate charges, and footprinting studies show that divalent cations such as Mg^{2+} specifically bind to junctions (Murchie *et al.*, 1989). The specificity of these interactions revealed at the nucleotide level, however, may not translate to the atomic level. This may explain why no additional ions were located in the current or the previous (Ortiz-Lombardía *et al.*, 1999) mismatched junction structures. Clearly this is not dependent on the types of cations present in the crystal since no well-defined cation complexes could be definitively located at the cross-over of the mismatched junction (Ortiz-Lombardía *et al.*, 1999), even though the sequence was crystallized in the presence of Mg^{2+} rather than Ca^{2+} solutions.

2.5 Discussion

We have solved a structure in which four strands of the sequence d(CCGGTACCGG) assemble to form a Holliday junction. This junction is similar to that previously reported for d(CCGGGACCGG) (Ortiz-Lombardía *et al.*, 1999). Both junctions are in the stacked-X conformation, where pairs of helical arms stack coaxially to form nearly continuous criss-crossed duplexes. There are, however, significant differences between the two structures. The current structure has all standard Watson-Crick base pairs, making it representative of the intermediate involved in homologous recombination. Furthermore, the

junction does not fall on a crystallographic symmetry axis; therefore, each arm around the junction is unique.

Comparing the sequences of similar decanucleotide crystals, we see that an A6-C7-C8 trinucleotide is common to the junction forming sequences (Table 2.2). Changing any one of these nucleotides results in a B-DNA duplex. Immediately preceding A6 is either a guanine or thymine; however, since guanine G5 of d(CCGGGACCGG) is mispaired, this position can be any nucleotide. The CC/GG dinucleotide at the 5' and 3'-termini are common to both the junction and B-DNA sequences, while G3 and G4 complement the cytosines of the ACC trinucleotide. This ACC trinucleotide forms the cross-over of the four-way junction. The hydrogen bonds from cytosine C8 to an adjacent phosphate at the cross-over and the water mediated hydrogen bonds from G3 to the phosphates of adenine A6 across the junction play important roles in defining the geometry and stability of the junction in both sequences. For example, these well defined interactions fix the orientation between the two duplexes across the junction, and thus account for the 41.4° twist of these duplexes. We therefore define this ACC trinucleotide as the core of the Holliday junction in these crystal systems.

In the structure of d(CCGGGACCGG), the mismatched guanines form unusual hydrogen bonds to cytosines of the flanking C·G base pairs (Ortiz-Lombardía *et al.*, 1999), including cytosine C7 of the core ACC trinucleotide. This interbase-pair hydrogen bonding results in a local unwinding of the G4-G5 dinucleotide by 10°. In a B-DNA duplex, G·A mismatches overwind the flanking base pairs (Privé *et al.*, 1987). We see no distortion to the B-DNA geometry at or near the junction in the present structure. Therefore, the distortions in the d(CCGGGACCGG) structure result from the tandem G·A mismatches, and not the junction itself.

The geometry of nucleotides A6 and C7 of the core ACC trinucleotide do not differ significantly from that of B-DNA and, therefore, their role in stabilizing the junction is not entirely clear. Previous biochemical studies on synthetic junctions suggest that base

Table 2.2. Comparison of unique DNA decamer sequences with d(CC)/d(GG) ends and standard nucleotide bases.

Conformation	Sequence[a]	Space Group	Reference
4-way junction	CCGGT**ACC**GG	*C2*	this work
	CCGGG**ACC**GG	*C2*	(Ortiz-Lombardía *et al.*, 1999)
B-DNA duplex	CCGCT**A**G**C**GG	*C2*	this work
	CCAGT**AC**TGG	*C2*	(Kielkopf *et al.*, 2000)
	CCGGCG**CC**GG	*R3*	(Heinemann *et al.*, 1992)
	CCAGG**CC**TGG	*C2*	(Heinemann and Alings, 1989)
	CCGCCGG**C**GG	*R3*	(Timsit and Moras, 1994)
	CCA**AC**GTTGG	*C2*	(Privé *et al.*, 1991)
	CC**AC**TAGTGG	$P3_221$	(Shakked *et al.*, 1994)
	CCAAGATTGG	*C2*	(Privé *et al.*, 1987)
	CCAAGCTTGG	*P6*	(Grzeskowiak *et al.*, 1993)
	CCATTAATGG	$P3_221$	(Goodsell *et al.*, 1994)

[a] Nucleotides in the ACC core in the junctions and similar nucleotides in other sequences are in bold, while mismatched base pairs are underlined

stacking could account for the significance of this AC step in the core. In junctions assembled from four unique DNA strands, there are two different ways to stack four arms into two duplexes. Minor perturbations to the base pairs at the junction have dramatic effects on which arms become coaxially stacked (Duckett *et al.*, 1988). For example, a junction defined by C·G, A·T, T·A, and G·C base pairs, in this order, has the A-B arm stacked on the B-C arm and C-D stacked over D-A. Inverting the C·G to a G·C base pair at the junction results in a duplex with A-B stacked on D-A and B-C on C-D. These effects, which have been quantitated in symmetric sequences by Zhang, *et al* (Zhang and Seeman, 1994), have been attributed to the unique base stacking across the junction. The A6 and C7 nucleotides in the common ACC motif may play a similar role. In addition, the specific hydrogen bonds from the C8·G3 base pair of the ACC core to the phosphates at the cross-over help define which stacked isomer is favored in this DNA sequence. Thus, we can start to correlate the sequence dependent structure with specific structural interactions in the junction.

The structure of the Holliday junction described here and previously (Ortiz-Lombardía *et al.*, 1999) is the stacked-X conformation. It is remarkable that both these crystal structures are so similar to the models previously proposed for junctions assembled with nonhomologous sequences (Lilley, 1999). One would expect that a junction between two truly homologous duplexes would migrate and, therefore, would not specifically locate at one position. Migration of the junction through the sequence d(CCGGGACCGG) would presumably be blocked by the two mismatched G·A base pairs. However, in the current structure, we observe that the interactions of the base pairs within the ACC core and with the phosphates at the cross-over define the overall conformation of the junction. This includes the angle between the two stacked duplexes and the coaxial stacking of arms in the duplexes within the relatively rigid junction. Thus, this ACC core may help to define a stable Holliday junction that prevents this migration and, therefore, allows the structure to be crystallized.

2.6 Acknowledgments

We thank Prof. P. A. Karplus and his group for helpful suggestions, and Drs. K. Henderson and T. Earnest at the ALS for help during data collection. This work was supported by grants from the National Science Foundation (MCB-9728240), the American Cancer Society, Oregon Division (J0159880), and the National Institutes of Environmental Health Sciences (ES00210). The X-ray diffraction facility has been supported by the M. J. Murdock Charitable Trust and the Proteins and Nucleic Acids Research Core of the Environmental Health Sciences Center at Oregon State University.

Chapter 3

The Effects of Mismatched Base Pairs on the Structure of the Holliday Junction: A Study of the Inherent Properties of DNA Four-way Junctions

Brandt F. Eichman and P. Shing Ho

Formatted for submission

3.1 Summary

Holliday junctions are four-stranded DNA complexes that are formed during recombination and DNA repair events. Much work has focused on the overall structure and properties of four-way junctions in solution, but we are just now beginning to understand these complexes at the atomic level. The crystal structures of two all-DNA Holliday junctions formed from the sequences d(CCGGGACCGG) and d(CCGGTACCGG) have recently been reported. A detailed analysis of these structures shows that the junction does not perturb the sequence-dependent B-DNA duplexes or hydration pattern, and that the distortions to the phosphoribose backbone and base pairs in d(CCGGGACCGG) are a consequence of the mispaired d(G·A) base pairs rather than the strand cross-overs. Both structures show a concerted rotation of the adjacent duplex arms relative to B-DNA, and this is discussed in terms of the conserved interactions between the duplexes at the junctions and further down the helical arms. These interactions distant from the strand cross-overs of the junction appear to be significant in defining its macroscopic properties, including the angle relating the stacked duplexes across the junction.

3.2 Introduction

Holliday junctions are important biological structures that are formed during homologous recombination, a process that is important in DNA repair, and that maintains genomic integrity and genetic diversity. Nucleic acid junctions are formed when strands are shared between two different double-helical segments. The overall structure and properties of four-way junctions have been extensively studied in solution (reviewed in Lilley, 1999). In the presence of divalent cations, these junctions exist predominantly as the stacked-X form in which the double-helical segments are coaxially stacked and related by a 60° angle

in a right-handed sense. The stacked arms resemble two adjacent duplexes that are linked at the junction by two common strands. The overall features of several different types of four-way junctions from recent crystal structures (Eichman *et al.*, 200X; Eichman *et al.*, 2000; Nowakowski *et al.*, 1999; Nowakowski *et al.*, 2000; Ortiz-Lombardía *et al.*, 1999) are in good agreement with the solution studies.

The first crystal structures of four-way junctions containing all deoxyribonucleotides were serendipitously solved in two different laboratories within six months of one another. Ortiz-Lombardía, *et al.* solved the first of these structures from the sequence d(CCGGGACCGG) while studying the effects of tandem d(G·A) mismatched base pairs on B-DNA (Ortiz-Lombardía *et al.*, 1999). The second all-DNA Holliday junction was solved in this laboratory from the sequence d(CCGGTACCGG), which was being used to study the effects of the photochemotheraputic drug psoralen on the DNA double-helix (Eichman *et al.*, 2000). The latter structure represents the original Holliday model to explain homologous recombination (Holliday, 1964) because it is a true inverted-repeat sequence, while the former shows how mismatched base pairs are accommodated by the structure. It should be noted that the two sequences (hereafter referred to as the GA and TA sequences) are identical except for the nucleotide at position 5 in the DNA (Figure 3.1a). Therefore, we can define d(CCGGNACCGG) as the sequence motif for the Holliday junction in single crystals.

This chapter focuses on the differences and similarities between the structures of the inverted-repeat TA and mismatched GA Holliday junctions in order to determine which structural properties are truly inherent to the DNA junction, and which are effects of the base mismatches. We find that the DNA structure across the stacked arms at the junction is influenced by base stacking, and that the interactions at the ACC core sequence and at the ends of the duplex arms are conserved in both structures. In addition, the mismatched d(G·A) base pairs slightly distort the backbone in the same manner as in duplex DNA.

Figure 3.1. Comparison of the Holliday junction crystal structures d(CCGGTACCGG) (red) (Eichman *et al.*, 2000) and d(CCGGGACCGG) (blue) (Ortiz-Lombardía *et al.*, 1999). All distances shown are in Å. A. Association of four strands A (blue), B (green), C (red), and D (orange) into the stacked-X form of the four-way junctions. For simplicity, strands A-D of d(CCGGGACCGG) (bottom) were re-assigned relative to the published sequence in order to correspond to d(CCGGTACCGG) (top). B. Overlay of the two structures, viewed into the major groove face (top) and along (bottom) the junction. All atoms except those of nucleotides T5 and G5 have an r.m.s.d. of 1.15 Å. Phosphates along each strand are traced with a ribbon. C. Stereoview of atoms at the ACC core, showing the contacts formed between the bases and phosphates across the junction. Images are rotated 90° in the plane of the page with respect to the top structure in B. Water molecules are rendered as spheres.

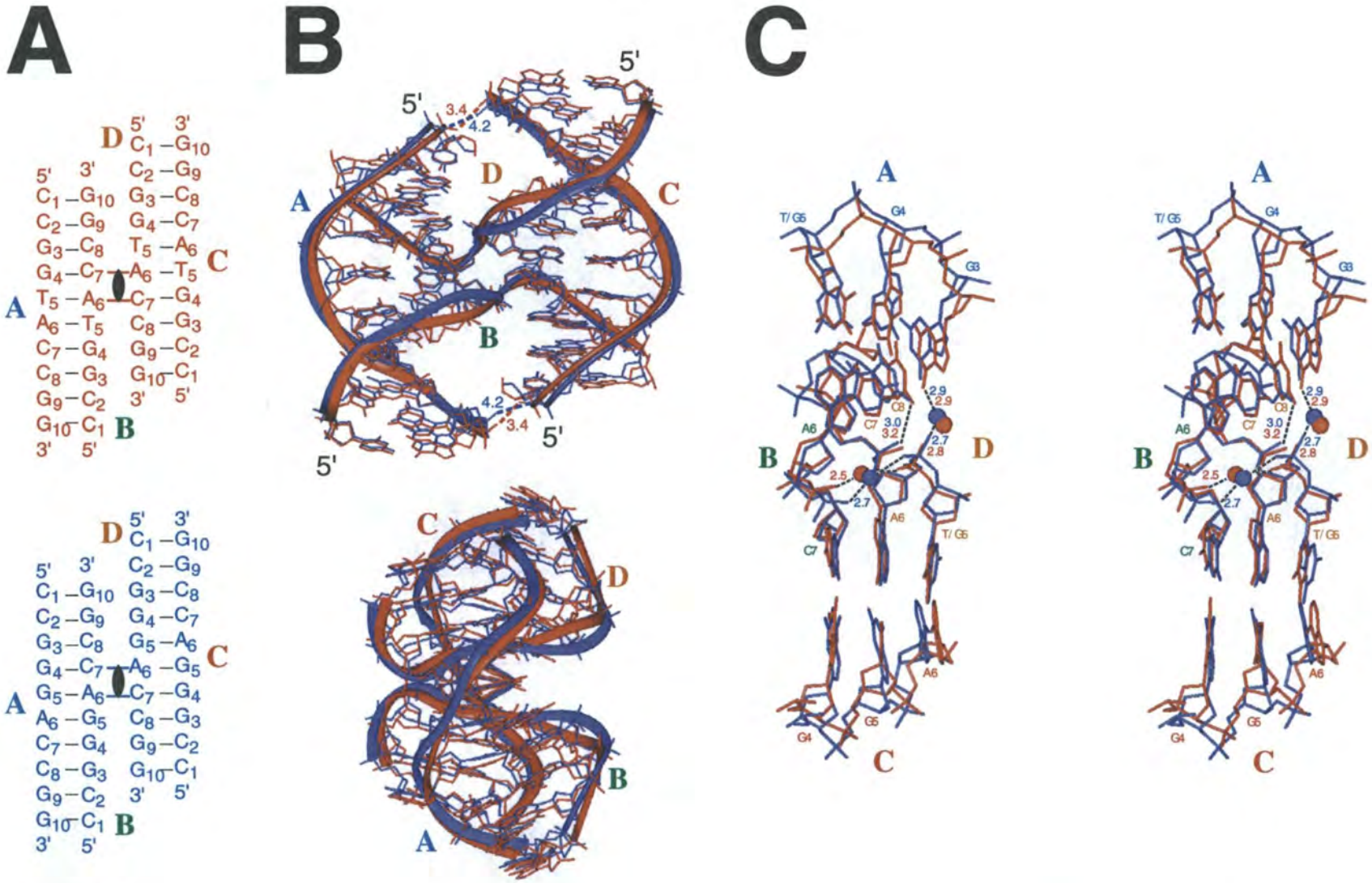


Figure 3.1

These results indicate a strong nucleotide sequence-dependence in the structure of Holliday junctions. The interactions between adjacent duplex arms impose slight distortions to the helical twist at the base steps flanking the junction, showing how the interactions removed from the junction influence the overall geometry of the four-stranded complex.

3.3 Crystallography

The growth conditions, morphology, and symmetry of the crystals were virtually identical between the two sequences. The GA sequence formed thin, plate-like crystals from a solution containing 0.33 mM DNA, 133.3 mM $MgCl_2$, 6.7% 2-methyl-2,4-pentanediol (MPD), and which was equilibrated against 45% MPD. Thin plates were also obtained from the TA sequence using 0.25 mM DNA, 75 mM sodium cacodylate buffer (pH 7), 15 mM $CaCl_2$, 2.5% MPD, and equilibrated against 30% MPD. Diffraction quality crystals of the TA sequence could also be obtained with $MgCl_2$ instead of $CaCl_2$ in the drop, in concentrations ranging from 50 to 150 mM.

Both crystals belong to the space group *C2*. The GA crystals had unit cell dimensions of a = 64.2 Å, b = 23.7 Å, c = 38.3 Å, and β = 112.43° with 2 DNA strands in the asymmetric unit, while the TA structure had unit cell dimensions of a = 66.50 Å, b = 23.5 Å, c = 76.9 Å, and β = 114.8° and an asymmetric unit consisting of four DNA strands. Therefore, the mismatch junction is formed from two duplexes related by two-fold crystallographic symmetry, while the junction in the TA sequence is composed of four structurally unique DNA strands. This loss of crystallographic symmetry in the TA sequence is a result of a ~2 Å shift in the DNA along the a-axis (away from the c-axis). Even with this lattice shift, the DNA crystal packing is identical between the two structures.

The structures were solved and refined using different x-ray diffraction and refinement methods. The GA structure was solved with experimental phases obtained from

a multiwavelength anomolous dispersion (MAD) experiment using a brominated sequence as a derivative. This model was refined to 2.16 Å using REFMAC (Murshudov *et al.*, 1999). In contrast, the TA structure was solved using molecular replacement with idealized B-DNA helices as search models, and was refined to 2.10 Å with X-PLOR 3.851 (Brünger, 1992b).

3.4 Results

3.4.1 The overall structure of the DNA Holliday junction

The structures formed from the sequences d(CCGGTACCGG) (Eichman *et al.*, 2000) and d(CCGGGACCGG) (Ortiz-Lombardía *et al.*, 1999) are stacked-X type Holliday junctions, and, overall, are virtually identical (Figure 3.1). The r.m.s.d. between the two structures is 1.15 Å for all atoms except nucleotides T5 and G5. DNA strands B and D each cross over between adenine A6 and cytosine C7 and base pair with complementary strands A and C to form two pseudo-continuous B-DNA helices, each composed of a 6-base pair (bp) arm stacked on a 4-bp arm (Figure 3.1a,b). These helices are related in a right-handed sense by ~41° in both structures. The interactions between the adjacent duplexes, those at the junction A6-C7-C8 triplet sequence and between the ends of the 4- and 6-bp arms are conserved in both structures. The hydrogen-bonding network and van der Waals interactions at the ACC junction core (described in detail in Eichman *et al.*, 2000; Ortiz-Lombardía *et al.*, 1999) are evident in both structures, showing that the ACC sequence is a defining characteristic of these junctions (Figure 3.1c). In addition, there are contacts between adjacent duplex arms that are further removed from the ACC core. The phosphate oxygens of cytosine C2 on the 4-bp arms and guanine G10 on the 6-bp arms are within 3.4 Å and 4.2 Å in the TA and GA structures, respectively (Figure 3.1b). Although these interactions are not at the junction core, they are equally important in defining the overall

geometry of the DNA junctions. Therefore, the overall structures are extremely similar and have the same sequence-dependent stabilization of the junction.

A more detailed analysis of the helical parameters of both structures shows that at the backbone and base pair levels, there are minor distortions from B-DNA at the junction. The distortions that are common to both structures are inherent properties of these junction, and differences between the two structures highlight the effects of the mismatched d(G·A) base pairs.

3.4.2 Similarities between the crystal structures of d(CCGGTACCGG) and d(CCGGGACCGG): The effect of the Holliday junction on B-DNA

The junction does not dramatically affect the B-DNA nature of the helical arms. The phosphoribose backbone is surprisingly unperturbed, even with the sharp re-direction of the strands at the cross-overs. The phosphate positions in both structures are the same as in two adjacent, resolved duplexes. Besides the deviation in backbone trajectory imposed by the mismatched d(G·A) base pairs (described below), the only differences in backbone torsion angles from canonical B-DNA occur as a result of the sharp direction change in the backbone at the junction. This direction change can be described primarily by rotations around χ (glycosidic bond), ε (C4'-C3'-O3'-P), and β (P-O5'-C5'-C4') of nucleotides A6 and C7 (Table 3.1).

As a result of the strand cross-overs, the largest distortions to base stacking occur at the base steps *flanking*, but not *at* the junction (Figure 3.2). The twist angles and base stacking at the junction d(G4pT5/A6*C7) base step, where the asterisk (*) refers to the strand cross-over, are the same between the d(CCGGTACCGG) junction structure and the B-DNA structure of d(CCAGTACTGG) (Kielkopf *et al.*, 2000) (Figure 3.2a,b). Thus, the base stacking is sequence-dependent at this dinucleotide step, and shows that the stacking characteristic of B-DNA is a dominant interaction even across the junction. There is no analogous d(GpG/ApC) base step in a regular B-DNA double helix to compare to the GA

Table 3.1. Backbone torsion angles for nucleotides at the junction cross-over in d(CCGGTACCGG) and d(CCGGGACCGG).

		d(CCGGTACCGG)[a]		d(CCGGGACCGG)[b]		Canonical B-DNA	
Angle	nucleotide	Junction	Arms	Junction	Arms	Drew dodecamer	High-resolution decamer
δ	A6	147.1± 1.5	143.2 ± 8.9	137.1	148.0 ± 12.2	123 ± 21	133 ± 19
ε	A6	**-89.0 ± 0.9**	-124.6 ± 33.9	**-73.6**	-132.4 ± 28.8	-169 ± 25	-151 ± 34
ζ	A6	-77.8 ± 1.2	-164.5 ± 41.7	-96.6	-156.1 ± 51.0	-108 ± 34	-130 ± 52
α	C7	-70.1 ± 7.0	-55.4 ± 20.5	-47.1	-62.4 ± 16.5	-63 ± 8	-68 ± 5
β	C7	**-156.8 ± 1.4**	144.7 ± 17.7	**-177.1**	152.5 ± 15.9	171 ± 14	162 ± 14
γ	C7	57.8 ± 5.4	41.5 ± 11.0	47.5	41.7 ± 9.1	54 ± 8	50 ± 6
χ	C7	**-147.3 ± 0.9**	-81.7 ± 14.2	**-152.1**	-83.3 ± 15.7	-117 ± 14	-96 ± 18

Torsion angles are defined as P-(α)-O5'-(β)-C5'-(γ)-C4'-(δ)-C3'-(ε)-O3'-(ζ)-P (Saenger, 1984). Angles at the junction are from the specified nucleotide, and angles from the arms were averaged across all remaining nucleotides. In the d(CCGGTACCGG) structure, junction angles were averaged across the specified nucleotide in strands B and D. Canonical B-DNA torsion angles are reported for the Drew dodecamer d(CGCGAATTCGCG) (Drew *et al.*, 1981) and the 0.74 Å structure of d(CCAGTACTGG) (Kielkopf *et al.*, 2000), and are averages for all nucleotides in those structures except cytosine C1. Values in bold-face are greater than 2 standard deviations from the corresponding angle in the Drew dodecamer.

[a] (Eichman *et al.*, 2000)
[b] (Ortiz-Lombardía *et al.*, 1999)

Figure 3.2. Similarity in base stacking between the DNA junctions and corresponding B-DNA sequences. Closed symbols refer to the junction structures, and open symbols refer to B-DNA structures with central d(TpA) and d(GpA) steps. Data is shown for d(CCGGTACCGG) (red squares), d(CCGGGACCGG) (dark blue circles), d(CCAGTACTGG) (orange squares) (Kielkopf *et al.*, 2000), and d(CCAAGATTGG) (light blue circles) (Privé *et al.*, 1987). A. Twist angles for the nine dinucleotide steps, numbered 1-9 from the 5'-end of strand A. The strand cross-over occurs at base step 4. B. Sequence-dependence of the base stacking at the d(G4pT5/A6*C7) base step across the junction in d(CCGGTACCGG) (red) and in d(CCAGTACTGG) (orange), where the asterisk (*) refers to the strand cross-over. Subscripts refer to the DNA strands A, B, and D. C. Opening angle for the 10 base pairs, numbered as in A. Parameters were calculated using CURVES 5.2 (Lavery and Sklenar, 1989).

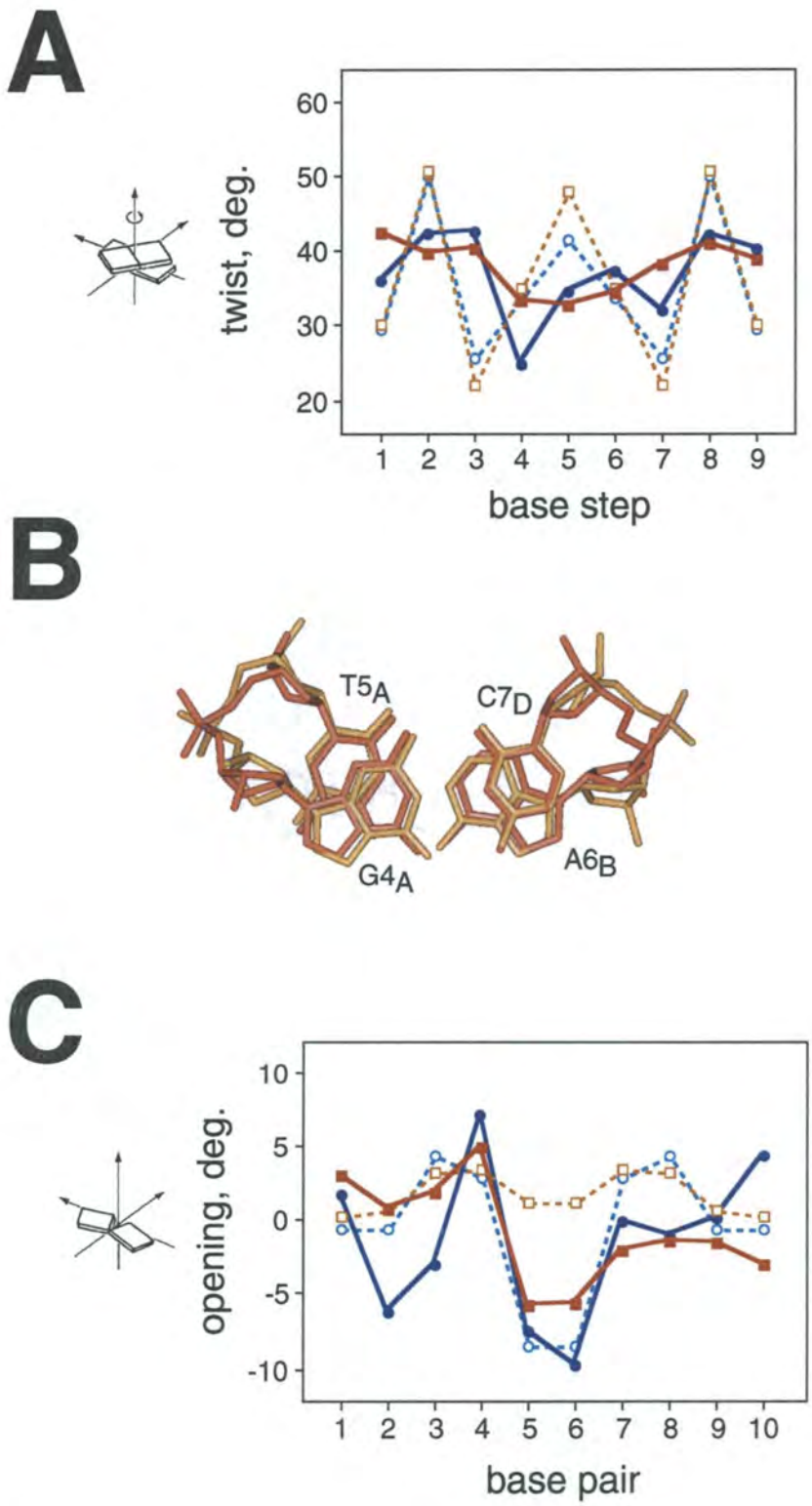

A
twist, deg.
60
50
40
30
20
1 2 3 4 5 6 7 8 9
base step
B
T5A
C7D
G4A
A6B
C
opening, deg.
10
5
0
-5
-10
1 2 3 4 5 6 7 8 9 10
base pair

Figure 3.2

junction. Interestingly, the base steps flanking the junction along the 4-bp and 6-bp arms are over- and underwound, respectively (Figure 3.2a). The twist angles at the 4-bp arm d(G3pG4)/C7pC8) steps are 40.2° (TA) and 42.6° (GA), compared to 36.9° at the d(GG/CC) step in the B-DNA structure of d(CCAGGCCTGG) (Heinemann and Alings, 1989). Conversely, the twist angles at the central d(TpA) and d(GpA) base steps, along the 6-bp arms, are less than the corresponding steps in regular B-DNA structures (Figure 3.2a). Thus, the overwinding on one side of the junction is compensated for by an underwinding on the other side, and can be explained by the interactions between the adjacent arms.

In addition to the twist angles of the base steps flanking the junction, the d(T·A) base pairs seem to be susceptible to distortion as a result of the strand cross-overs. The opening angle of these base pairs is less than expected in a B-DNA duplex (Figure 3.2c). The mismatched d(G·A) base pairs exhibit the same negative opening, but those distortions are inherent to d(G·A) mismatches since the same angles are observed in the B-DNA structure of d(CCAAGATTGG) (Privé *et al.*, 1987). However, this negative opening angle cannot directly be explained in the context of the rest of the structure, and therefore could be a result of the flexible nature of d(T·A) base pairs (reviewed in Dickerson, 1999), and not a result of the junction. In summary, the largest distortions to the B-DNA arms as a result of the junction are the backbone torsion angles between adenine A6 and cytosine C7 of the cross-over strands, and the over- and underwinding of the 4- and 6-bp arms, respectively.

3.4.3 Differences between the crystal structures of d(CCGGTACCGG) and d(CCGGGACCGG): The effect of mismatched base pairs on Holliday junctions

The most obvious difference between the TA and GA Holliday junction crystal structures is the distortion to the backbone near the strand cross-overs (Figure 3.1b). The phosphate positions between the two structures overlay nicely, except to a large degree between the G4 and G6 phosphate groups on the cross-over strands, and to a lesser degree between these nucleotides on the non-cross-over strands. It is perhaps not surprising that the main differences in the backbone trace between the two structures occur at the

uncommon nucleotide. Superimposition of crystal structures of B-DNA decanucleotides containing central d(TpA) and mismatched d(GpA) base steps onto two stacked arms of the Holliday structures shows that the backbone traces are a feature of the nucleotide sequence (Figure 3.3). Thus, the distortions to the backbone in the GA structure are a direct result of the mismatched d(G·A) base pairs, and not of the junction itself.

In addition, distortion to the base pairs in the GA structure are a result of the mismatches, and not an inherent feature of DNA junctions. Inspection of base pair parameters stretch, stagger, *y*-displacement, propeller twist, and inclination in the two Holliday structures and in two analogous B-DNA sequences shows that the central base pairs are different between the TA and GA structures, and that these differences are sequence-dependent (Figure 3.4). Distortions to the base pairs in the GA structure occur only at the mismatched d(G·A) base pairs, and are identical to those observed in the mismatched d(G·A) B-DNA structure (Privé *et al.*, 1987). Base pair parameters in the TA junction structure are remarkably similar to a d(T·A) containing B-DNA structure (Kielkopf *et al.*, 2000), again illustrating the fact that the junction does not significantly distort the B-DNA nature of the duplexes. In conclusion, the differences in backbone trajectory and base pair geometries between the TA and GA junctions structures are a direct result of the d(G·A) mismatched base pairs, and are not due to the presence of the junction.

3.4.4 Solvent structure of the Holliday junction

Thus far, the similarities and differences between the TA and GA junction structures have been described at the DNA level. How does the surface of a four-stranded DNA Holliday junction complex appear to the surrounding solvent? The solvent accessible surface (SAS) shows that the major and minor grooves are as they appear in standard B-DNA double helices, with the minor grooves on one face of the complex separated by the raised phosphoribose backbone ridge at the junction (Figure 3.5a), and the major groove

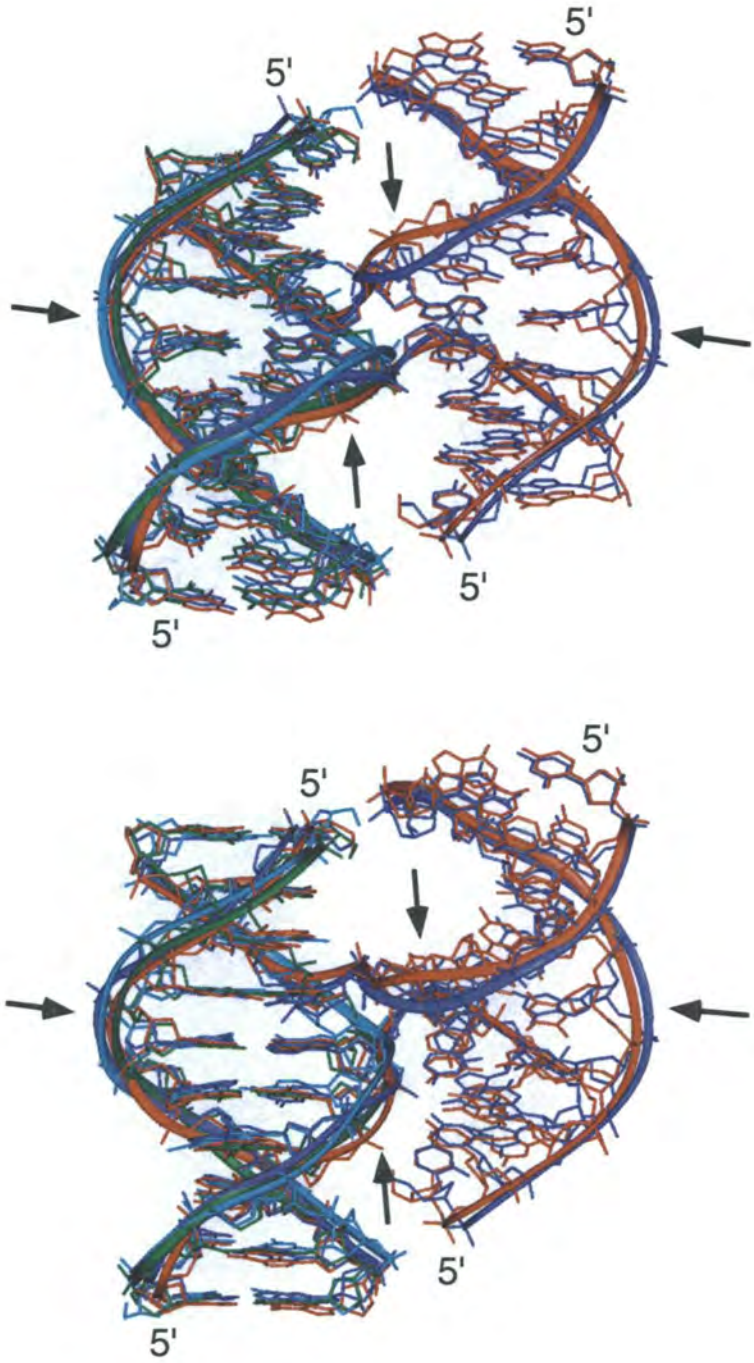


Figure 3.3. Distortion to the phosphoribose backbone as a result of the mismatched d(G·A) base pairs. Junction structures were superimposed and are colored as in Figure 3.1. Central d(T·A)- and d(G·A)-containing B-DNA duplexes d(CCGCTAGCGG) (green) (Eichman *et al.*, 2000) and d(CCAAGATTGG) (light blue) (Privé *et al.*, 1987) are superimposed on one stacked duplex of the junction structures. Phosphate positions along each DNA strand are traced with a ribbon. Arrows indicate the positions of the d(G·A) mispairs. Two views are shown, and are related by a 20° rotation along the junction.

Figure 3.4. Distortion to the bases as a result of the mismatched d(G·A) base pairs. Local base pair parameters stretch, stagger, propeller twist, inclination, and *y*-displacement are shown for the same structures and colored as in Figure 3.2. Data were calculated using CURVES 5.2 (Lavery and Sklenar, 1989).

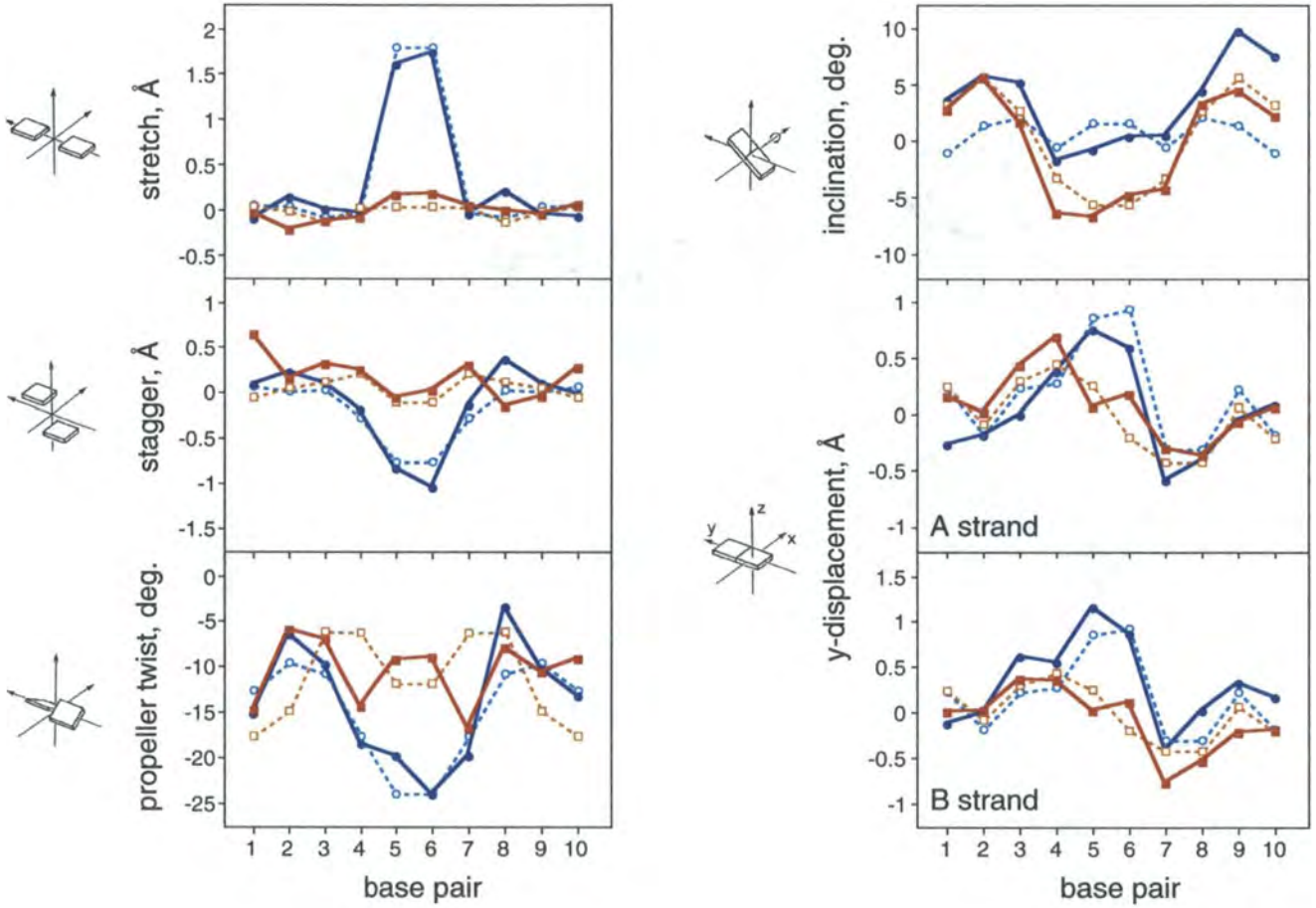

stretch, Å
stagger, Å
propeller twist, deg.
inclination, deg.
y-displacement, Å
A strand
B strand
base pair
base pair

Figure 3.4

surfaces smoothly connected by the junction on the other face of the complex (Figure 3.5b). Two cavities are formed on either side of the strand cross-overs by the interactions between two adjacent arms (Figure 3.5a,b). The cavity floor is formed from the cross-over strands at the junction, the ceiling from the interaction between phosphates at the ends of the arms, and the walls from the major groove of the 4-bp arm facing the minor groove of the 6-bp arm. This cavity is the site of 6 out of the 13 conserved waters between the TA and GA structures (Figure 3.5c), including an essential sequence-dependent bridging water at the ACC core (Eichman *et al.*, 2000; Ortiz-Lombardía *et al.*, 1999).

Although several water molecules partially penetrate the SAS, perhaps most interesting is the sodium ion at the center of the complex in the TA structure (Figure 3.5c,d). This cation helps to counter the negatively charged phosphates at the junction, and is completely buried beneath the SAS (Figure 3.5d). This burial suggests that this ion was bound initially to a less compact form of the junction that had a more accessible cation binding pocket, and consequently stabilized the formation of the compact junction through favorable electrostatic interactions. Thus, it is possible to gain further insight into the nature of four-way junctions by comparing the solvent patterns observed in the crystal structures. Most importantly, given the similarity of the helical arms to B-DNA, is the hydration of the stacked arms what would be expected for B-DNA, or does the junction have a unique hydration pattern?

Because the asymmetric unit of the TA structure consists of two pair of duplexes related by non-crystallographic symmetry, some waters that were observed at one duplex were not observed at the other. Therefore, to accurately compare the waters between the two structures, one complete solvation pattern for the four-stranded TA complex was generated by applying two-fold symmetry to map the waters from one duplex onto the other.

DNA structure is highly dependent on solvent and each particular conformation has been shown to have a unique pattern of hydration (reviewed in Berman and Schneider, 1999). In B-DNA, the "spine of hydration" is a hydrogen-bonded network of waters in

Figure 3.5. Solvent-accessible surface and electrostatic potential of d(CCGGTACCGG). Surfaces and potentials (red negative and blue positive) were calculated using the Delphi module of the program InsightII (Molecular Simulation, Inc., San Diego, CA). Conserved water molecules (purple spheres) between the two junction structures and the sodium ion (blue sphere) at the center of the junction are shown. A, B. Views into the minor groove (A) and major groove (B) face show the cavity formed between the minor groove of the 6-bp arm and the major groove of the 4-bp arm. C. Same view as in A, with a transparent surface, showing the position of the DNA atoms and waters. D. Cross-section at the junction, revealing the sodium ion buried beneath the solvent-accessible surface. The molecule is rotated 90° with respect to the views in A-C.

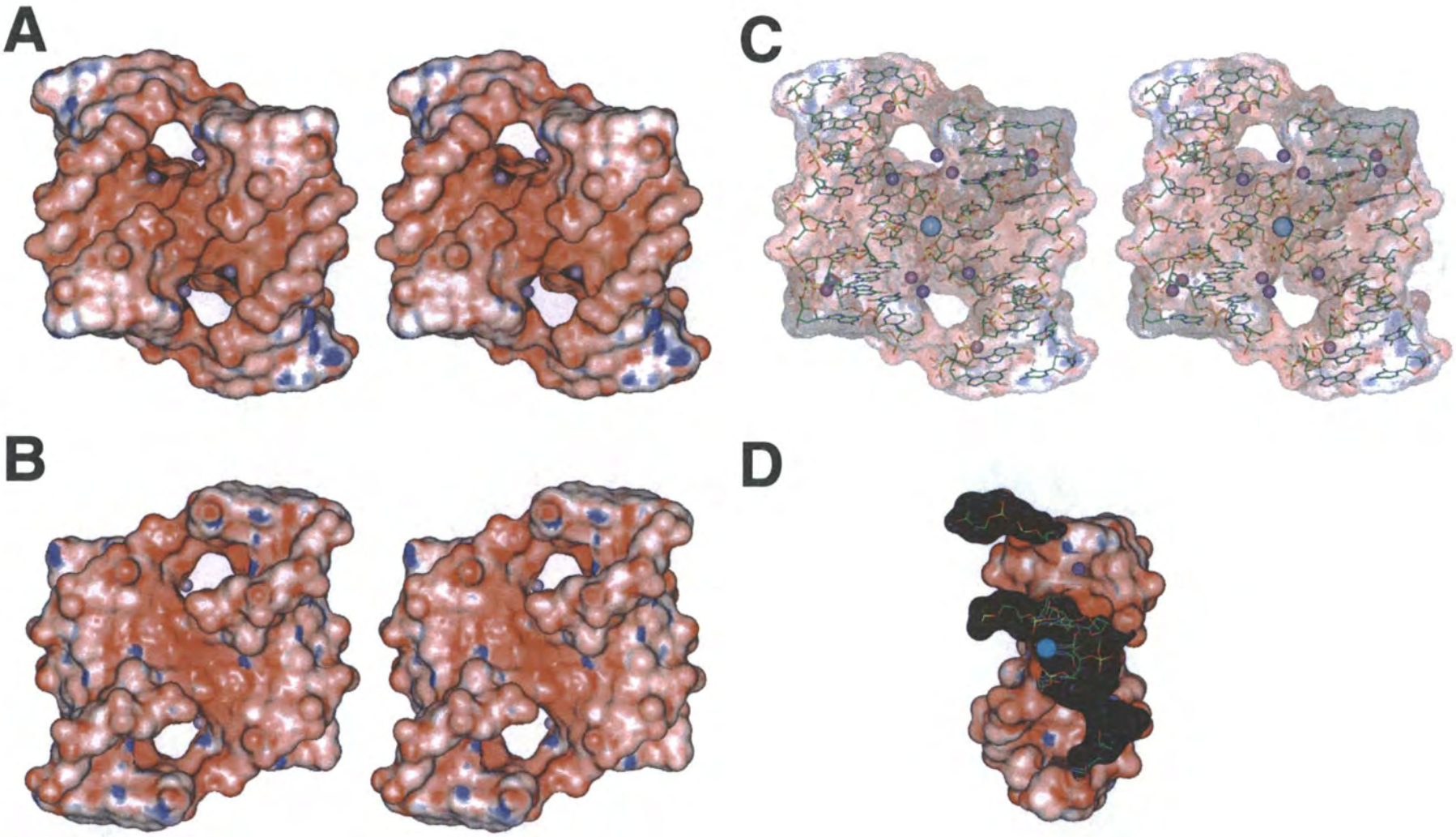


Figure 3.5

the minor groove with well-defined, sequence-dependent positions. These waters are hydrogen-bonded to purine N3 imino nitrogens, pyrimidine O2 keto oxygens, and to a lesser extent guanine N2 amino nitrogens and O4' ribose oxygens. GC-rich sequences tend to have a double spine of waters, whereas AT stretches have a single spine due to the more narrow minor groove in these sequences. In the major groove, waters tend to hydrogen-bond with purine N7 imino (as well as guanine O6 keto and adenine N6), cytosine N4 amino, and thymine O4 keto groups.

With the exception of a few solvent molecules at the strand cross-overs, all of the first-shell waters at the arms of the junction structures are indicative of B-DNA. However, not all of the DNA hydrogen-bond donor and acceptor sites are occupied with waters. For example, the minor groove spine of hydration typical of B-DNA is not fully extended along the length of the duplex arms. The waters seem to be clustered at the ends of the arms and absent from the central base pairs, especially in the minor grooves of the mismatch junction structure (Figure 3.6a). This apparent lack of waters at the central base steps is most likely a function of the resolution of the two structures, and not a result of the junction. The total number of waters observed in the two junction structures is not unusual for structures at this resolution (Figure 3.7). Thus, higher resolution structures will be required to assess the effect of the junction on the hydration pattern at the base pairs across the junction.

The four-stranded complexes of both structures contain 49 first-shell waters that were observed in the electron density maps, 13 of which (6 per duplex plus 1 at the center of the junction) have identical positions between the two structures. Almost all of these conserved waters lie in the cavity formed from the interface of adjacent duplexes (Figures 3.5c and 3.6b). In the minor groove, there are four conserved interactions. The waters at guanines G3 (strand A) and G9 (strand D), at the ends of the 4-bp arms, are the only two waters that do not face the adjacent duplex. The remaining two are located at the 6-bp arms, one at guanine G9 (strand A), and the other at the d(T·A) and d(G·A) base pairs which is

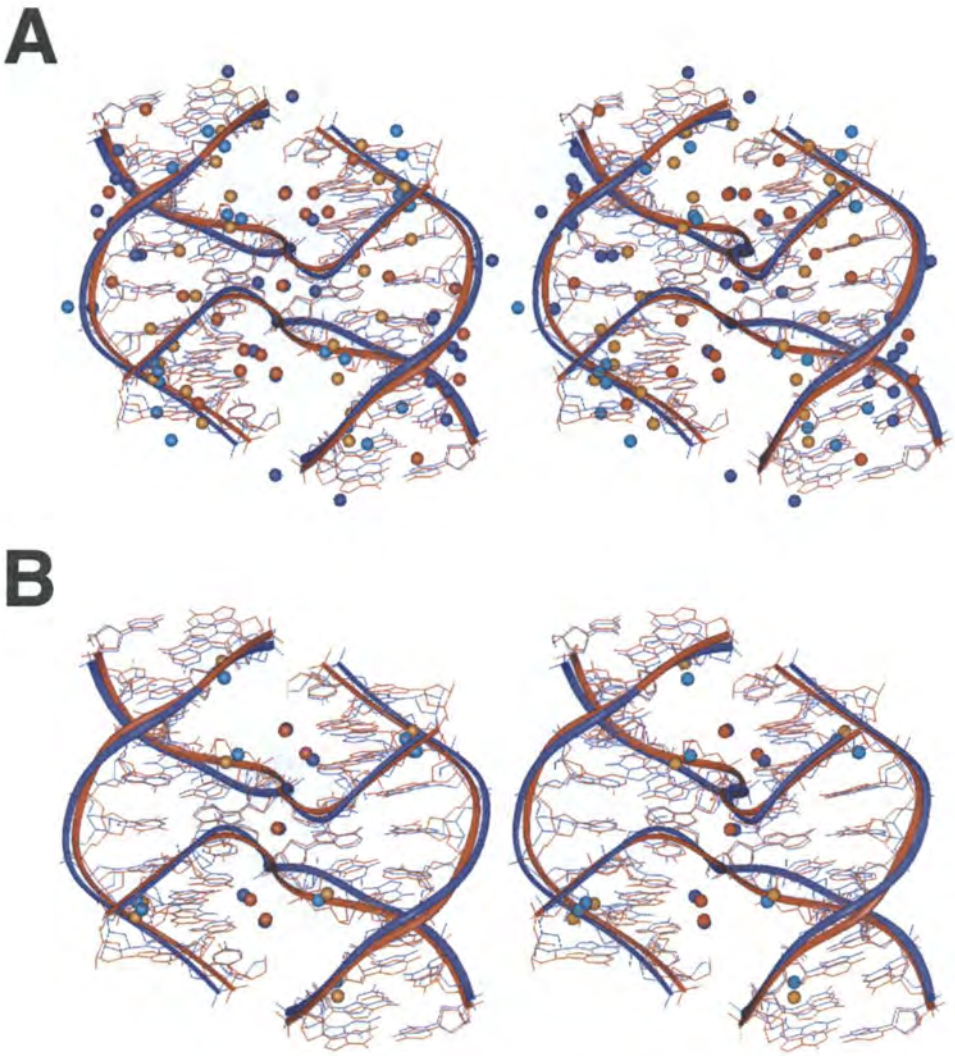


Figure 3.6. Hydration patterns in d(CCGGTACCGG) (red atoms) and d(CCGGGACCGG) (blue atoms). First-shell water molecules (spheres) are colored according to the DNA structure and to their position in the major and minor groove. Red and orange waters are in the major (red) and minor (orange) grooves of d(CCGGTACCGG), while blue waters are in the major (dark blue) and minor (light blue) grooves of d(CCGGGACCGG). Phosphate positions along each DNA strand are traced with a ribbon.

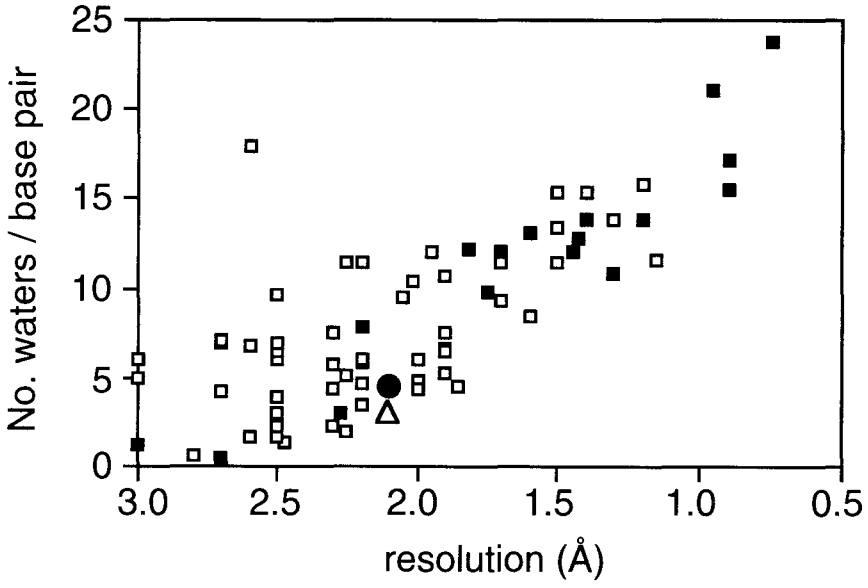


Figure 3.7. Number of waters in B-DNA crystal structures as a function of resolution. Data are shown for all B-DNA structures in the Nucleic Acid Database (Berman *et al.*, 1992), and were normalized by calculating the number of waters per base pair in each structure. Data for the d(CCGGTACCGG) (circle) and d(CCGGGACCGG) (triangle) junctions are compared with all other structures (squares). Closed symbols refer to structures cross-validated with the free-R calculation (Brünger, 1992a), which tends to reduce over-fitting the crystallographic data by adding extraneous water molecules to the atomic model.

hydrogen-bonded to the O2 keto oxygen of thymine T5 (strand B) and the corresponding N3 imino nitrogen of guanine G5 (strand B). In the major groove, the conserved waters are located on the 4-bp arms at cytosine C2 and guanine G3. The water at guanine G3 bridges the O6 oxygen of this guanine with the O1P phosphate oxygen of adenine A6 at the strand cross-overs (Eichman *et al.*, 2000; Ortiz-Lombardía *et al.*, 1999). Thus, the majority of the waters with identical positions between the two structures are lining the duplex interface cavity.

The conformation of four-way DNA junctions have been shown to be dependent on the particular cations present; the stacked-X form is stabilized by polyvalent cations (reviewed in Lilley, 1999). The electrostatic potential surface shows that, as expected, there is highly negative charge potential at the junction cross-over, where four phosphates are closely packed within 6.5 Å of each other (Figure 3.5a,b). Although crystals of both sequences were grown in the presence of divalent metals (Mg^{2+} or Ca^{2+}), none were observed in either structure, even at the junction (Eichman *et al.*, 2000; Ortiz-Lombardía *et al.*, 1999). A single solvent molecule was observed bridging the O2P phosphate oxygens of adenine A6 at the junction (strands B and D) in both structures. This electron density peak was assigned as a sodium ion in the TA structure based on the short distance to each oxygen. The Na^{+}-O bond length was observed to be 2.55 Å and 2.50 Å, with a O-Na^{+}-O angle of 125.8°, compared with the GA structure, in which the OH···O2P hydrogen bonding lengths each were 2.71 Å and the O-OH_2-O angle was 138.8°. Thus, the geometry of this interaction is significantly different between the two structures, and further studies are required to elucidate the true identity of this solvent molecule.

3.5 Discussion

We can conclude from the correlation between each junction structure and similar B-DNA sequences that the distortions to the phosphoribose backbone and to the base pairs across the junction in the GA sequence are induced by the mismatches and not the strand cross-overs. Thus, the conformation of the DNA at the junction is sequence-dependent. The conserved interactions at the ACC core of the junction and between the ends of the arms strongly suggest that these define the overall structure of the four-stranded complex, and thus are responsible for fixing the Holliday junction at a particular point within the d(CCGGNACCGG) motif.

The crystal structure of the GA sequence shows that the distortion to the DNA induced by the mismatches is not sufficient to prevent formation of a stacked junction, which is more than likely due to extra hydrogen bonding across the stacked arms between mismatched guanine G5 (strand A) and cytosine C7 (strand D) (Ortiz-Lombardía *et al.*, 1999). It has been shown by comparative gel electrophoresis that stacked-X junctions in solution are capable of accommodating d(G·A) mismatches at the cross-over, and that in some cases mismatched sequences require increased magnesium ion concentrations to form the stacked junction (Duckett and Lilley, 1991). Indeed, the GA crystals were grown in 9-fold greater divalent cation concentration than the TA crystals, suggesting that formation of a mismatched junction in the crystal is less favorable than a junction consisting of Watson-Crick base pairs.

The strong stacking of the d(GpT/A*C) base steps and the unique hydrogen-bonding within the mispaired d(GpG/A*C) steps (where the asterisk indicates the position of the strand cross-overs) are the dominant interactions that define the DNA structure at the junction. These interactions, as well as the network of contacts at the ACC core, provide an explanation for the ability of these sequences to crystallize as immobile junctions. Panyutin, *et al.* have observed that conditions that favor the stacked-X form of the junction

dramatically decrease the rate of branch migration (Panyutin *et al.*, 1995). This has led to the theory that disruption to the base pairing and stacking at the junction via the open-X form is necessary for each step of branch migration. In the crystal structures, migration of the junction along the duplex was most likely arrested by the unique ACC sequence. It is interesting to speculate that prior to crystallization, all oligonucleotides at such high concentrations are capable of forming four-stranded structures, and eventually form resolved duplexes due to branch migration past the end of a sequence lacking junction-stabilizing interactions. Indeed, linear duplexes have been observed to form from chi structures that undergo branch migration past the ends of the sequence (Thompson *et al.*, 1976).

The 41° interhelical angle is not a consequence of the crystal lattice. Identical packing has been observed in several crystal structures of DNA duplexes of different lengths and in different space groups, showing that this is a common feature of the DNA molecule (Table 3.2). The crystal packing in these B-DNA duplex and the Holliday junction structures involves pseudo-continuous stacked helices that span the length of the crystal and that cross each other in an X, with the backbone of one duplex sitting in the major groove of the adjacent duplex. Timsit, *et al.* have shown that this backbone-groove packing is sequence-dependent (Timsit and Moras, 1991; Timsit and Moras, 1994). The interhelical angles of these duplex crystal structures range from 42° to 90° (Table 3.2). In addition, the brominated derivative of the GA sequence used in the structure determination of the mismatched junction crystallizes in a different space group, and has a geometry that is identical to the native TA and GA structures. This further suggests that the DNA sequence, and not the crystal lattice, determines the overall structure of the four-stranded complex.

It now appears that interactions at the ends of the arms and at the ACC core define the 41° orientation of adjacent helices. The 60° angle measured in solution (Cooper and Hagerman, 1989; Mao *et al.*, 1999; Murchie et al., 1989) is too wide to accommodate the specific interduplex contacts observed in the DNA crystal structures. The overwinding of

Table 3.2. Angles between B-DNA duplexes packed into X-type crystal lattices

Length (bps)	Sequence	Space Group	Angle (°)	Reference
12	ACCGGCGCCACA/ TGTGGCGCCGGT	*R3*	74	(Timsit *et al.*, 1989)
	ACCGGCGGCGCC/ GGCGCCGCCGGT	*R3*	74	(Timsit *et al.*, 1991)
10	CCGGCGCCGG	*R3*	77	(Heinemann *et al.*, 1992)
	CCGCCGGCGG	*R3*	77	(Timsit and Moras, 1994)
	CGATCG6mATCG	$P3_221$	60	(Baikalov *et al.*, 1993)
	CCAACITTGG	$P3_221$	60	(Lipanov *et al.*, 1993)
	CCACTAGTGG	$P3_221$	60	(Shakked *et al.*, 1994)
	CCATTAATGG	$P3_221$	60	(Goodsell *et al.*, 1994)
	CAAAGAAAAG/ CTTTTCTTTG	*C2*	51	(Han *et al.*, 1997)
	CGCAATTGCG	*C2*	51	(Wood *et al.*, 1997)
	CCGCTAGCGG	*C2*	44	(Eichman *et al.*, 2000)
	CTCTCGAGAG	*C2*	43	(Goodsell *et al.*, 1995)
6	GGCGCC	$P4_12_12$	90	(Vargason *et al.*, 2000)

Angles between duplexes are shown for B-DNA duplex crystal structures that pack end-to-end to form columns of pseudo-continuous helices throughout the crystal. In all cases, the contact point between helices is defined by the phosphoribose backbone of one column lying in the major groove of the adjacent column.

the 4-bp arms and underwinding of 6-bp arm at the base steps flanking the junction, the largest perturbations to the DNA common to both structures, act in concert to preserve the interhelical contacts. A clockwise rotation of one arm results in a counter-clockwise rotation of the other, much like two connected gears rotating in opposite directions of one another. Thus, the interactions between the arms are strong enough to effect the rotation of one helix from the other. In addition, elimination of one of the arms results in increased conformation variability of the junction. When one of the arms consists of only one base pair, as in the crystal structures of four-way junctions formed from RNA and DNA strands, the angle between the arms can vary from 55° to -80° (Nowakowski *et al.*, 1999; Nowakowski *et al.*, 2000). Thus, the interactions between the arms that are further removed from the junction seem to play a special role in defining the overall structure.

It is surprising that no polyvalent cations were specifically located in either structure. The floor of the interduplex cavity directly at the strand cross-overs is the most likely binding site for cations because of the high charge potential and proximity to potential ligands. Most of the conserved waters reside in this cavity, and are more than likely partially occupied cations. A single well-ordered Co^{3+} ion was observed in the crystal structure of the 55° RNA-DNA junction, in an identical position to the bridging water observed between guanine G3 and the phosphate of adenine A6 in the DNA junctions (Nowakowski *et al.*, 1999). Presumably, the RNA-DNA junction is able to accommodate the larger cobalt ion because of the wider angle between the adjacent arms, although the Co^{3+}-DNA coordination bonds in that structure were longer than the expected 2 Å. Crystals of d(CCGGTACCGG) could not be obtained in the presence of Co^{3+}, and soaking very small concentrations of Co^{3+} into existing crystals resulted in very fragile or cracked crystals (B. F. Eichman, unpublished data), suggesting that the cobalt severely distorted the DNA structure. Therefore, the compact junction formed as a result of the interactions between the arms does not seem to provide an optimum binding site for polyvalent cations. It is interesting that no Mg^{2+} ion was observed in the RNA-DNA structure solved in the

absence of Co^{3+} (Nowakowski *et al.*, 1999). These data suggest that magnesium and cobalt have different binding sites at the junction. Indeed, a magnesium ion was bound in the major groove of the stacked base pairs at the strand cross-over in the -80° DNA-RNA junction structure. Very few waters were observed in this region in the DNA junctions, suggesting that a diffuse, poorly ordered cloud of counterions could be present in these structures.

Chapter 4

The Crystal Structures of Psoralen Cross-linked DNAs: Drug Dependent Formation of Holliday Junctions

Brandt F. Eichman, Blaine H.M. Mooers, Marie Alberti, John E. Hearst, and P. Shing Ho

Formatted for submission

4.1 Summary

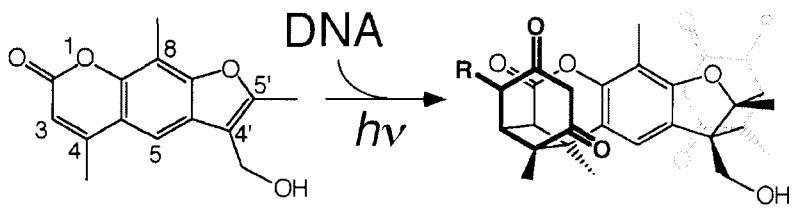
The crystal structures at 2.2 Å resolution are presented for two DNA sequences with the thymines covalently cross-linked across the complementary strands by 4'-hydroxymethyl-4,5',8-trimethylpsoralen (HMT). The adduct of d(CCGCTAGCGG) forms a drug-induced Holliday junction. In contrast, HMT-d(CCGGTACCGG) forms a sequence-dependent junction. In both structures, the DNA duplex is highly distorted at the thymine linked to the six-member pyrone ring of the drug. The psoralen cross-link defines the intramolecular interactions of the drug-induced junction, while the sequence-dependent structure is nearly identical to the native Holliday junction of d(CCGGTACCGG) alone (Eichman, *et al.* (2000) *Proc. Natl. Acad. Sci. USA* 97, 3971-3976). The two structures contrast the effects of drug- and sequence-dependent interactions on the structure of a Holliday junction, suggesting a role for psoralen in the mechanism to initiate repair of psoralen adducts in mammalian DNA.

4.2 Introduction

Psoralens are a class of furocoumarin compounds that have been used as photochemotherapeutic treatments for psoriasis, vitiligo, and other skin disorders as early as ~2000 B.C. (reviewed in Bethea *et al.*, 1999; Helm *et al.*, 1991). They have been shown to inhibit replication (Baden *et al.*, 1972; Lüftl *et al.*, 1998; Trosko and Isoun, 1971) and transcription (Sastry and Hearst, 1991a; Sastry and Hearst, 1991b; Shi *et al.*, 1988a) by covalently modifying DNA, and can act as photosensitizers to affect cell differentiation through interaction with cell surface membranes (Lanskin, 1994; Lanskin and Lee, 1991). Psoralens have also been powerful tools for studying DNA structure and function (reviewed in Cimino *et al.*, 1985) because of their ability to cross-link the two strands of a DNA

duplex. The detailed structures of psoralen-DNA adducts have, therefore, been of great interest because of their broad use as chemotherapeutics and as diagnostic tools. We present here the first single crystal structures of two psoralen cross-linked DNA sequences which, interestingly, form unique four-way Holliday junctions.

Psoralen preferentially forms interstrand DNA cross-links at d(TpA) dinucleotides upon exposure to light (Scheme 4.1). A model has been proposed that describes the mechanism by which psoralen cross-links the two strands of DNA (Isaacs *et al.*, 1977). The drug initially interacts noncovalently with the DNA duplex by intercalating between base pairs. Upon absorption of a single long wavelength photon (320-410 nm), either the furan-side or pyrone-side of the drug reacts with a pyrimidine base to form a cyclobutanyl monoadduct. Furan-side monoadducts can absorb a second photon to cross-link the thymine on the complementary strand of the DNA duplex.



Scheme 4.1

The effect of a psoralen cross-link on the conformation of a DNA duplex remains unresolved. The NMR structures of monoadducts between deoxythymidine and several different psoralen derivatives (Kanne *et al.*, 1982b; Straub *et al.*, 1981) showed that the adducted thymine bases were in *cis-syn* geometry relative to psoralen. This stereochemistry was confirmed by the NMR structure of two deoxythymidine nucleosides cross-linked by psoralen derivatives (Kanne *et al.*, 1982a) and by the crystal structure of a thymine base bonded to the furan-side of 8-methoxypsoralen (Peckler *et al.*, 1982). Models constructed from these simple structures suggested that the drug could kink the DNA helix by as much

as 70°. Indeed, the first NMR structure of a psoralen cross-linked octanucleotide reported a 53° helical bend toward the major groove of the DNA duplex (Tomic *et al.*, 1987). Furthermore, cross-linked, but not furan-side monoadducted DNAs were observed by electron microscopy to be significantly kinked (Shi *et al.*, 1988b). However, studies that probe the ability of the drug to kink or bend DNA using anomalous gel migration (Sinden and Hagerman, 1984) and differential decay of birefringence methods (Haran and Crothers, 1988) showed the DNA duplex to remain essentially straight. Finally, more recently determined NMR structures of psoralen-modified octanucleotides indicate that psoralen cross-links do not place a kink in the DNA at the site of intercalation (Hwang *et al.*, 1996; Spielmann *et al.*, 1995).

In an attempt to detail the effect that psoralen cross-links have on the structure of DNA, we have solved to a resolution of 2.2 Å the single crystal structures of the sequences d(CCGGTACCGG) and d(CCGCTAGCGG) cross-linked at the thymine bases by 4'-hydroxymethyl-4,5',8-trimethylpsoralen (HMT). Interestingly, the cross-linked sequences crystallize as distinct four-stranded Holliday junctions. Interstrand cross-links in themselves have not been shown to induce junctions. The sequence d(CGCGAATTCGCG) with a dithiobis-propane cross-link (Chiu *et al.*, 1999) and the sequence d(CCTCGCTCTC)/d(GAGAGCGAGG) containing a cisplatin interstrand cross-link (Coste *et al.*, 1999) both remain as B-type duplexes. Comparing each of the junctions presented here with structures of their respective unmodified sequences shows that the junction formed by d(CCGCTAGCGG) is induced by the drug, while that of d(CCGGTACCGG) is a consequence of the DNA sequence (Eichman *et al.*, 2000). The drug-induced junction structure represents the most dramatic perturbation to the DNA helix by an intercalator observed in a crystal structure to date, and suggests a mechanism by which the drug destabilizes the DNA duplex to allow initiation of cross-link repair in mammalian systems.

4.3 Materials and Methods

4.3.1 Crystallization and x-ray data collection

DNA sequences were synthesized on an Applied Biosystems DNA synthesizer, purified by HPLC, and photocross-linked with HMT according to the published protocol (Spielmann *et al.*, 1992), with the following modifications. Irradiation was performed using either an argon (360 nm) or a krypton laser (336 and 356 nm) operating at 200 mW. The DNA solution was irradiated for 3 hours using a flow cell while stirring in an ice bath. The cross-linked DNA was extracted three times with 0.25 to 0.5 volume chloroform and ethanol precipitated after removal of chloroform. After volume reduction using a rotary evaporator, the cross-linked oligonucleotides were further purified by HPLC.

Very thin diamond-shaped crystals of HMT-d(CCGGTACCGG)$_2$ (400 x 200 x 20 μm^3) were grown at 4° C from solutions containing 0.5 mM DNA, 30 mM sodium cacodylate buffer (pH 7), 150 mM $MgCl_2$, 0.5 mM spermine·4HCl, 10% 2-methyl-2,4-pentanediol (MPD) and equilibrated against 20% v:v MPD. Slightly thicker crystals of HMT-d(CCGCTAGCGG)$_2$ (300 x 150 x 50 μm^3) were obtained from solutions containing 0.5 mM DNA, 30 mM sodium cacodylate buffer (pH 7), 50 mM $CaCl_2$, 10% 2-methyl-2,4-pentanediol (MPD) and equilibrated against 20% v:v MPD. X-ray diffraction data were collected at liquid nitrogen temperatures using 1.1 Å radiation on beamline 5.0.2 at the Advanced Light Source in Berkeley, CA. Diffraction images were processed and reflections merged and scaled using DENZO and Scalepack from the HKL package (Otwinowski and Minor, 1997). Both crystals are in the monoclinic *C2* space group. Data collection statistics are shown in Table 4.1.

4.3.2 Structure solution and refinement

The HMT-d(CCGGTACCGG) sequence was solved using a directed real space translation/rotation/rigid body search (X-PLOR 3.851 (Brünger, 1992b) script written in

Table 4.1. Data collection and refinement statistics.

	HMT-d(CCGCTAGCGG)$_2$	HMT-d(CCGGTACCGG)$_2$
Data Collection		
Unit Cell		
a	70.7	72.1
b	23.8	23.7
c	44.3	36.2
β	130.1	112.6
Space Group	*C2*	*C2*
Resolution Range (Å)	34.65-2.2	33.39-2.2
Measured (unique) reflections	10418 (2937)	9596 (2900)
Completeness (%)[a]	99.2 (99.3)	96.7 (98.3)
R_{merge} (%)[a,b]	5.9 (8.0)	6.7 (10.0)
$\langle I/\sigma_I \rangle$ [a]	16.9 (8.2)	12.3 (8.1)
Refinement		
Resolution Range (Å)	34.65-2.2	33.39-2.2
R_{cryst} (R_{free})[c]	24.3 (27.7)	22.2 (31.5)
DNA atoms (solvent molecules)	423 (22)	423 (41)
Ave. B-factors (Å^2)		
DNA atoms (water atoms)	32.6 (47.1)	37.6 (60.5)
RMSD from ideality		
Bond lengths (Å) / angles (°)	0.013 / 1.717	0.005 / 1.229

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

[b] $R_{merge}(I) = \Sigma_{hkl} \Sigma_i | I_{hkl,i} - \langle I \rangle_{hkl} | / \Sigma_{hkl} \Sigma_i | I_{hkl,i} |$ 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 (Brünger, 1992a). The minimum converged values of R_{free} are reported.

this lab). Briefly, the structure of d(CCGGCGCCGG) (Heinemann *et al.*, 1992) was used as a search model by modifying the central d(CpG) dinucleotide to d(TpA), splitting the two halves (pentamers) of the duplex apart by 3.4 Å to accommodate the drug, and deleting the phosphate groups at this central dinucleotide step. The center of mass of this model was positioned in the unit cell at x=0, y=0, and z=1/4 and inclined 6° from the xy plane and 17° from the xz plane as indicated by the native Patterson maps. The model was incrementally translated in the x direction, rotating the whole duplex and each pentamer about the helical axis, and performing rigid body refinement against 2.5 Å data at each step. A clear solution was obtained with an R_{cryst} of 48.2% and R_{free} of 48.3%. With the duplex properly positioned, the HMT was unambiguously modeled into 2σ F_o-F_c density. A break in the electron density was observed between adenine 16 and cytosine 17, but was continuous between symmetry-related residues, as was previously observed in the junction structure of the unmodified sequence (Eichman *et al.*, 2000). The model was therefore rebuilt to accommodate the strand cross-overs.

HMT-d(CCGGTACCGG) was refined to a nominal resolution of 2.2 Å using X-PLOR 3.851 (Brünger, 1992b) (Table 4.1). Data from both crystals were highly anisotropic; the intensity of reflections in the b direction fell to a minimum at 2.2 Å. Therefore, refinement was only carried out to this resolution, ensuring completeness in all three directions. The fall-off of intensity along this direction correlates well with the lack of lattice contacts in the real-space y direction. Anisotropic B-factor scaling was applied to properly weigh the calculated (F_c) to the observed structure factors (F_o) during refinement (X-PLOR 3.851 script written by Ethan Merritt). No σ-cutoffs were applied to the data during refinement. R_{cryst} and R_{free} converged to 22.2% and 31.5%, respectively (Table 4.1).

The structure of HMT-d(CCGCTAGCGG) was solved by molecular replacement using EPMR (Kissinger *et al.*, 1999). Two stacked B-DNA arms from the refined HMT-d(CCGGTACCGG) structure were used as a search model against 2.8 Å data, resulting in a correlation of 63% and an R-factor of 49%. This sequence was refined using X-PLOR

3.851 (Brünger, 1992b) to a nominal resolution of 2.2 Å due to the high degree of anisotropy. Anisotropic B-factor scaling and no σ-cutoffs were applied to the data during refinement. Simulated annealing against 2.2 Å data resulted in an R_{cryst} of 30.7% and R_{free} of 37.5%. At this stage the drug could be unabiguously positioned into F_o-F_c density. After successive rounds of conjugate gradient and B-factor refinement, and addition of waters and one calcium ion, R_{cryst} and R_{free} converged to 24.3% and 27.7%, respectively.

4.4 Results

4.4.1 The psoralen-induced Holliday junction in HMT-d(CCGCTAGCGG)

The psoralen adduct of d(CCGCTAGCGG) was initially crystallized to study the effects of the drug on a DNA double helix. The structure unexpectedly was solved as a Holliday junction (Figure 4.1a). Two symmetry-related molecules assembled in the crystal to form this four-stranded complex in an antiparallel stacked-X structure. Since d(CCGCTAGCGG) in the absence of the psoralen cross-link crystallizes as a regular B-DNA double helix (Eichman *et al.*, 2000), the junction in HMT-d(CCGCTAGCGG) is a direct consequence of the drug cross-link.

In this structure, the psoralen cross-link is seen to destabilize the DNA duplex at the strand linked to the HMT pyrone-side, thereby allowing this strand to exchange across duplexes to form a junction (Figure 4.1a). This destabilization directly results from distortions to the base pair with its thymine linked to the drug's six-membered pyrone-ring. It was possible that the asymmetric psoralen intercalated into a symmetric DNA duplex would be disordered; however, the HMT molecule is well ordered in one orientation in the crystal. The F_o-F_c annealed omit electron density map clearly distinguishes the two sides of the drug (Figure 4.2), with both the pyrone-side keto oxygen and the hydroxymethyl group at the furan 4' carbon well resolved in the maps. The electron density maps also clearly

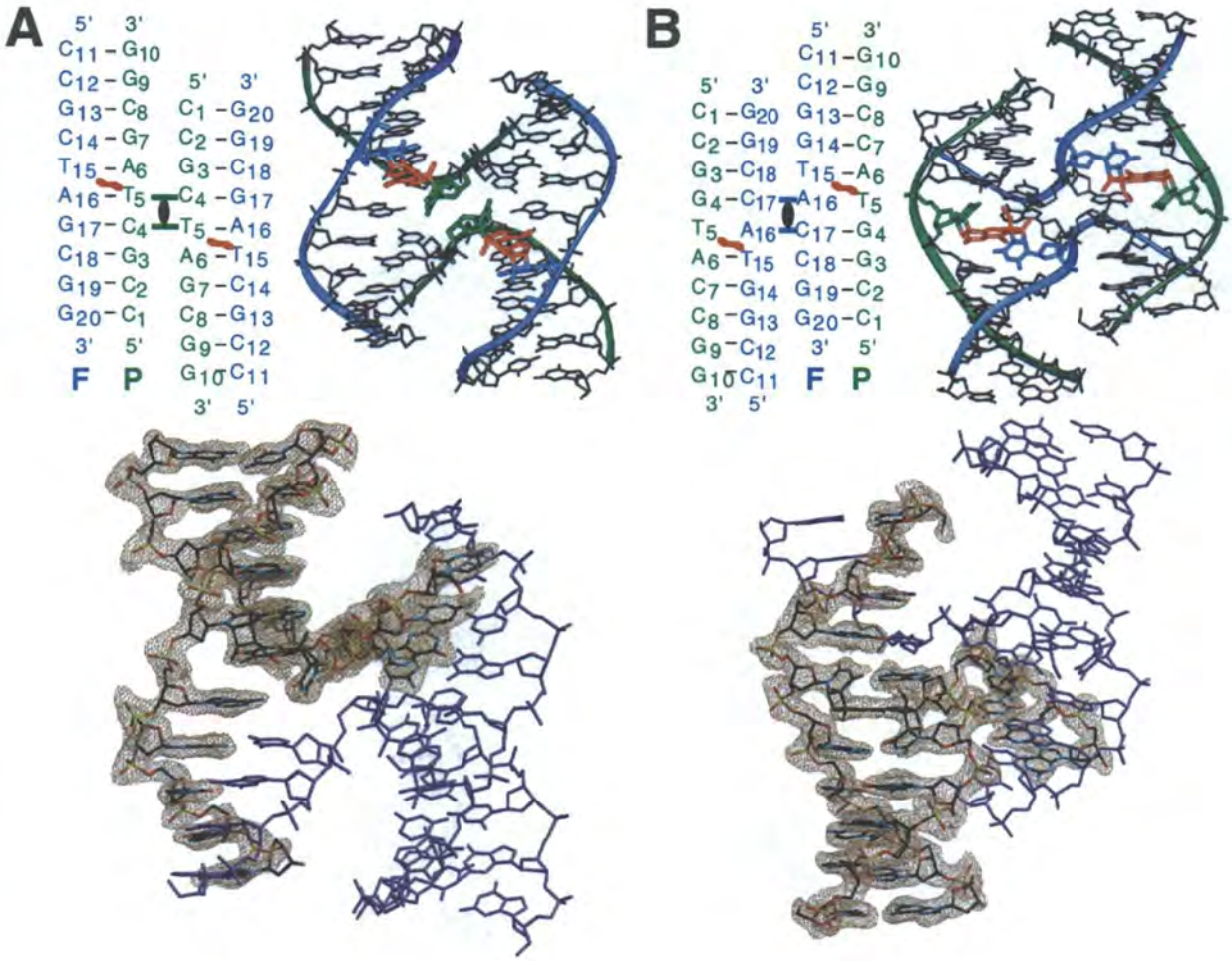


Figure 4.1. Single crystal structures of HMT-d(CCGCTAGCGG) (A) and HMT-d(CCGGTACCGG) (B) as Holliday junctions. In each panel, four DNA strands are shown (top left) to form a junction, with two strands crossing over to connect the stacked helical arms. The thymine that is adducted to the pyrone-side of the psoralen drug is labeled P, and the DNA strand of that thymine is designated the P-strand (green). Similarly, the DNA strand containing the furan-side adducted thymine is designated the F-strand (blue). The two stacked DNA helical arms are related by crystallographic two-fold symmetry so that two cross-linked strands define the asymmetric unit in the crystal. Models of each HMT-adducted structure (top right) are shown with the drug (red) covalently bound to the thymine bases along the P- and F-strands of the DNAs. The electron density contoured at 1σ in a $2F_o$-F_c map (bottom) is shown around the psoralen cross-linked strands of the asymmetric unit. The electron densities of the symmetry-related strands (purple) are omitted for clarity.

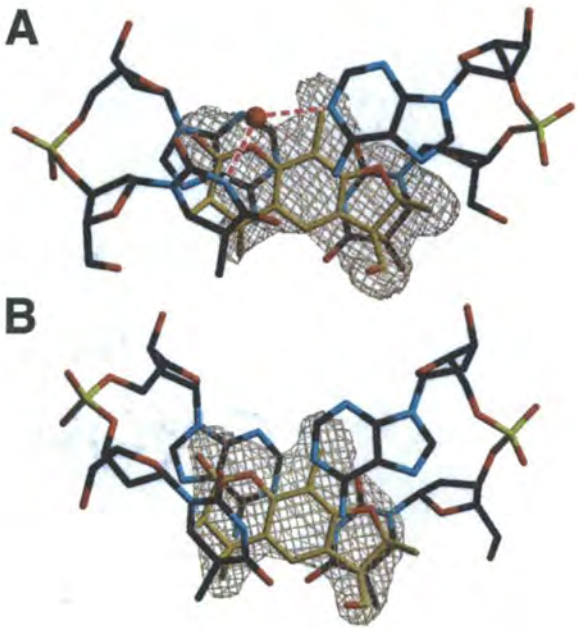


Figure 4.2. Single crystal structures of the HMT-adducted d(T·A) base pairs of the d(CCGCTAGCGG) (A) and d(CCGGTACCGG) (B) sequences. The F_o-F_c omit electron density map (contoured at 3σ) around the HMT moiety shows that the exocyclic substituents can be readily assigned to orient the drug in both DNA structures. A water molecule (colored red) is shown bridging the N3 nitrogen of the pyrone-side thymine with the N1 nitrogen of the complementary adenine base in HMT-d(CCGCTAGCGG).

show cyclobutanyl rings linking the HMT to the thymine bases (Figure 4.1), with carbons C5 and C6 of the thymine bases bonded to carbons C4 and C3 of the pyrone ring, and to carbons C4' and C5' of the furan ring of HMT. Thus, the structure at the present resolution allows us to define the orientation of the psoralen and resolve its effects on the DNA strands at the pyrone-side (P-strand) and at the furan-side (F-strand) of the drug.

The d(T·A) base pair with the thymine on the DNA F-strand shows the least distortion by the drug. Both hydrogen bonds remain within this base pair, even though the thymine base shows a large increase in the inclination angles (η) between the native and HMT-cross-linked sequences ($\Delta\eta = 34.1°$). This furan-side base pair also shows strong π-stacking interactions between the adenine and the HMT pyrone ring.

The most dramatic distortion is seen as a large (55.8°) buckle in the d(T·A) base pair with the thymine on the P-strand (Figure 4.1a), resulting from a steric collision between the O4 keto oxygen of the thymine base and the HMT pyrone ring. Consequently, the base pair no longer maintains two Watson-Crick hydrogen bonds. The thymine O4 keto to the adenine N6 amino hydrogen bond remains. However, the hydrogen bond from thymine N3 to adenine N1 is lost and replaced by a compensating water bridge (Figure 4.2a). Furthermore, there is virtually no stacking interactions between the adenine of this base pair and the HMT furan ring.

The perturbations at the pyrone-side d(T·A) base pair account for the destabilization of the DNA duplex at the P-strand in the HMT adduct as compared to the native B-DNA structure of d(CCGCTAGCGG). In the HMT structure, the direction of the phosphoribose backbone of the crossed-over P-strand deviates from normal B-DNA starting at the ribose of the thymidine nucleotide. As a direct result of the inclined thymine base, the deoxyribose ring is rotated outward, moving C4' away from the helix axis. This displaces the 5'-end of the P-strand away from its normal helical trajectory, with the thymidine phosphate pushed ≥ 3.7 Å away from its position in the native duplex structure. Thus, the distortion to the pyrone-side thymine base is propagated to the DNA backbone, consequently disrupting the

double-helix in this region, and allowing the resulting frayed ends of two molecules to associate into a Holliday junction in the crystal.

Although the structure shows that a Holliday junction is induced by psoralen, are the physical characteristics of the junction defined by the drug? If the cross-over of the junction were not at the drug cross-linked base pairs, would the DNA duplex be similarly affected? These questions are addressed by comparing the psoralen-induced junction to the sequence-dependent Holliday junction of HMT-d(CCGGTACCGG).

4.4.2 Sequence-dependent Holliday junction in HMT-d(CCGGTACCGG)

The sequence d(CCGGTACCGG) cross-linked with HMT also crystallized as a four-stranded Holliday junction (Figure 4.1b). We had previously shown that the unmodified d(CCGGTACCGG) sequence crystallizes in a stacked-X structure (Eichman *et al.*, 2000), and is stabilized by hydrogen bonding and base stacking interactions at the central ACC trinucleotide that forms the stable core of this native junction. The four-stranded Holliday junction and the stabilizing interactions at the cross-over strands formed from two symmetry-related molecules in the HMT-d(CCGGTACCGG) structure are nearly identical to that formed by the native DNA (Figure 4.3a). In particular, we see a direct hydrogen bond from the N4 amino nitrogen of cytosine 18 to the O2P phosphate oxygen of cytosine 17, as well as water-mediated hydrogen bonds from the O6 keto oxygen of guanine 3 to the O2P phosphate oxygen of adenine 16 within the stabilizing ACC core. Also, the two stacked arms flanking the cross-link are indistinguishable from a regular B-DNA duplex in terms of phosphate positions and base stacking, except at the strand cross-overs and at the psoralen cross-links. It is clear that the junction in HMT-d(CCGGTACCGG) results from the inherent ability of this DNA sequence to form a junction and, in this sequence, cross-linking the thymine bases with HMT does not significantly distort the overall conformation of the junction.

Figure 4.3. Similarity between the ACC junctions in the native (black bonds) and HMT (grey bonds) forms of d(CCGGTACCGG). Distances shown are in Å. A. Stereoview of the detailed interactions at the ACC core. B. Overlay of the native and HMT models at the short (4-base pair) arms, showing the conservation of contacts between the ends of the arms.

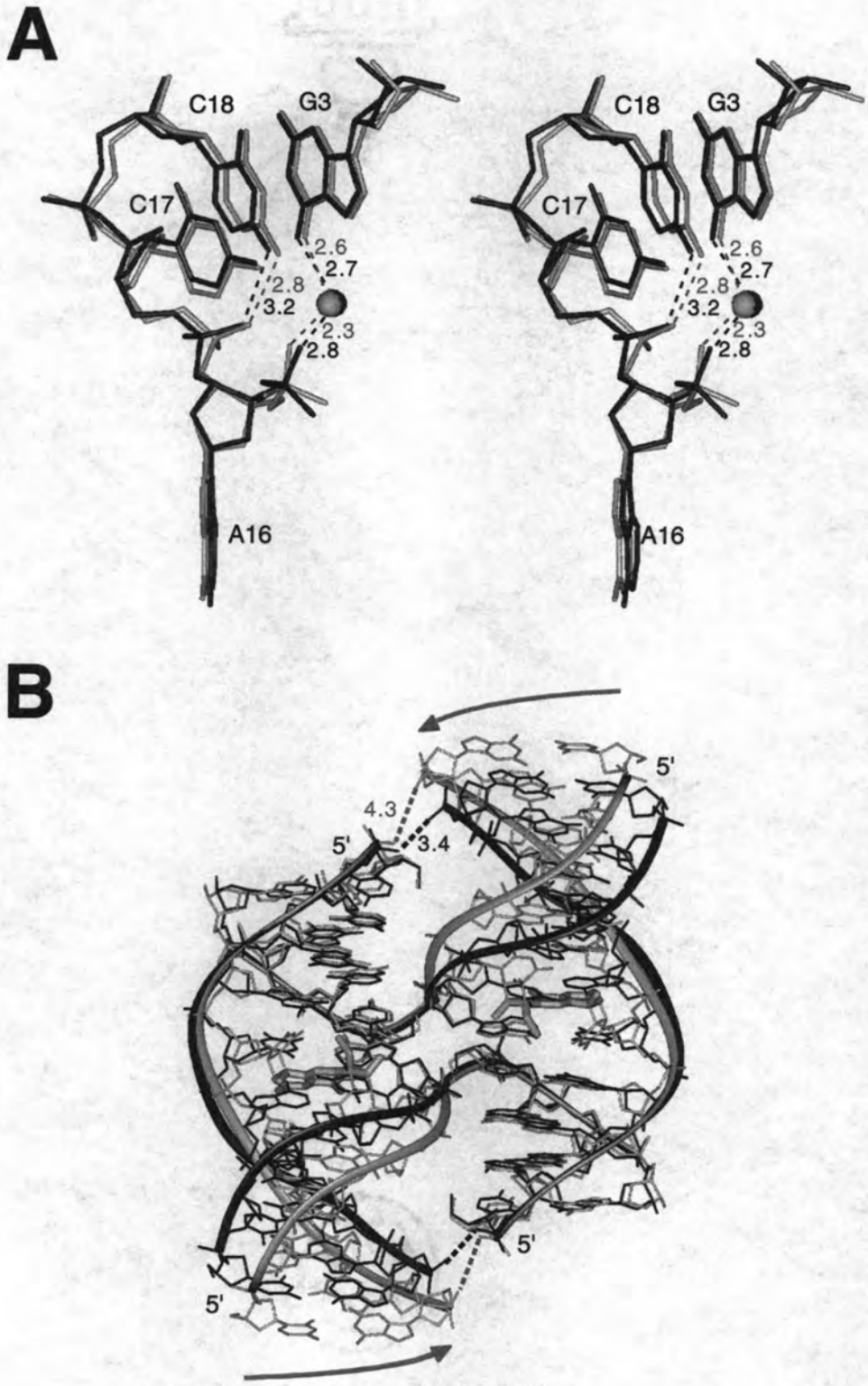

A
C18
G3
C17
2.6
2.7
2.8
3.2
2.3
2.8
A16
C18
G3
C17
2.6
2.7
2.8
3.2
2.3
2.8
A16
B
5'
4.3
5'
3.4
5'
5'

Figure 4.3

As in the psoralen-induced junction structure the cross-linked thymidine nucleotides at the furan- and pyrone-sides of the drug show different degrees of distortion. Both hydrogen bonds remain within the furan-side d(T·A) base pair, although it is slightly buckled as a result of the cross-link ($\Delta\kappa$ = 9.6° between the native and cross-linked structures). The highly inclined and buckled P-strand thymine base, however, is no longer base-paired with its adenine, but shows an elongated 3.2 Å hydrogen-bond from the O4 keto oxygen to the hydroxymethyl O14 oxygen of psoralen. This now unpaired adenine is slightly more inclined and tipped than it is in the native d(CCGGTACCGG) structure, and shows only minimal stacking against the psoralen furan ring.

4.4.3 Effects of psoralen on B-DNA

Because of the similarities between the native and HMT forms of d(CCGGTACCGG), this structure allows us to determine the effects of the drug on the B-DNA arms. As expected from a DNA-intercalating drug, psoralen unwinds and extends the B-DNA helix by one base pair; the differences in helical twist and length of each decanucleotide duplex between the native and HMT structures is -35.7° and 3.4 Å. An overlay of the short 4-base-pair arms of the two structures shows that the psoralen in effect pushes the 6-base-pair arms further from the junction (Figure 4.3b). The phosphoribose backbones of these longer arms clearly show a deviation in these helices starting at the HMT cross-link. Interestingly, the interactions at the ends of the adjacent arms (see Chapter 3) are preserved in spite of the helix elongation. The 3.4 Å distance between the phosphate oxygens of cytosine C2 and guanine G10 on adjacent short and long arms of the native sequence is increased to 4.3 Å in the HMT structure. This distance is expected to be at least 5.9 Å given a 3.4 Å increase in helix length. Thus, there is a slight tilt (~7°) of the helix axes in the HMT structure relative to the native strucutre in order to preserve the contacts at the ends of the arms. Consequently, the interhelical angle in this structure is slightly more shallow (~5°) than that of the native junction.

4.4.4 Comparison of the sequence- and drug-dependent Holliday junctions

How do these two psoralen cross-linked Holliday junctions compare? In both structures, the strand cross-overs of the Holliday junction sit adjacent to the d(T·A) base pair cross-linked to the pyrone-side of the drug. However, the strands that form the cross-over of the junction are switched between the two HMT sequences (Figure 4.1). In the HMT-d(CCGGTACCGG) structure, the junction is 3' to the adenine base on the F-strand (Figure 4.1b), placing one residue and two phosphates between the psoralen cross-link and the junction cross-over. In contrast, the junction in the HMT-d(CCGCTAGCGG) structure is immediately 5' to the P-strand thymine base, placing the backbone of this thymine residue directly at the cross-over (Figure 4.1a). The closest atomic interaction at this junction is between two symmetry-related thymine ribose O4' oxygens. Thus, the junctions are conformational isomers of each other, and can be interconverted by switching the arms that are paired to form the stacked duplexes on each side of the junction.

As a result of the strand cross-overs occuring at different positions in the nucleotide sequences, the specific contacts at the junctions are very different between the two strucutres (Figure 4.4). In contrast to the more rigid ACC junction, the drug-induced junction is stabilized by relatively weak interactions. Two contacts occur between the O2 keto oxygens and the C5' carbons of the symmetry related pyrone-side thymines, and between the C5M methyl carbon of this thymine and the O5' oxygen of cytosine C14 on the same P-strand (Figure 4.4a). Both of these weaker interactions (3.2 - 3.3 Å) can be considered C-H···O hydrogen-bonds (Taylor and Kennard, 1982). The contact between the O4' ribose oxygens of the symmetry-related pyrone-side thymines across the junction is most likely a longer van der Waals interaction, which is not represented by the current model. Although the electron density at the junction clearly shows the sharp turns in the cross-over strands, the density at the current resolution accommodates a 2.5-3.3 Å range between the O4' oxygens. The contacts at the sequence-dependent junction in HMT-d(CCGGTACCGG) are those expected for the ACC core sequence (Eichman *et al.*, 2000), with one exception (Figure

4.4b). The bridging sodium ion between phosphates across the cross-over strands is occupied by two hydrogen-bonded, phosphate-bridging waters in the current HMT structure. In addition to the hydrogen-bonds at this junction, a van der Waals contact is formed between the C5' carbons of the F-strand cytosines across the symmetry-related cross-over strands. Thus, both structures are right-handed Holliday junctions, with different stabilizing contacts at the strand cross-overs.

The immediate effects of psoralen on the cross-linked thymine bases and their paired adenines are significant, but are very similar between the two sequences (RMSD = 0.76 Å for the atoms of psoralen and the thymine and adenine bases between these structures). Thus, although the two HMT-DNA sequences form different isomeric forms of the Holliday junction, the local perturbations to the DNA are nearly identical. In the context of a sequence-dependent structure such as the ACC core-stabilized Holliday junction, psoralen perturbs the DNA duplex only locally, and interactions that stabilize the Holliday junction in the sequence-dependent structure overshadow the destabilization to the B-DNA helix induced by the drug. However, in the absence of such sequence-dependent stabilization, the local distortions can propagate to longer range effects to destabilize the DNA double-helix at the P-strand, which then allows for strand exchange and the formation of a drug-induced junction.

4.5 Discussion

The crystal structures presented here show that psoralen is capable of inducing both very minor and very dramatic distortions to the double helix, depending on the DNA sequence that is cross-linked. The sequence d(CCGGTACCGG) forms nearly identical Holliday junctions in both the HMT cross-linked and unmodified forms, suggesting that

Figure 4.4. Differences between the interactions at the HMT-d(CCGCTAGCGG) (A) and HMT-d(CCGGTACCGG) (B) Holliday junctions. Distances between atoms are shown in Å. Both models are colored by atom type (green carbon, blue nitrogen, red oxygen, yellow phosphorus). Psoralen carbons are colored gold. Nucleotides in each strand are labeled and colored as in Figure 4.1.

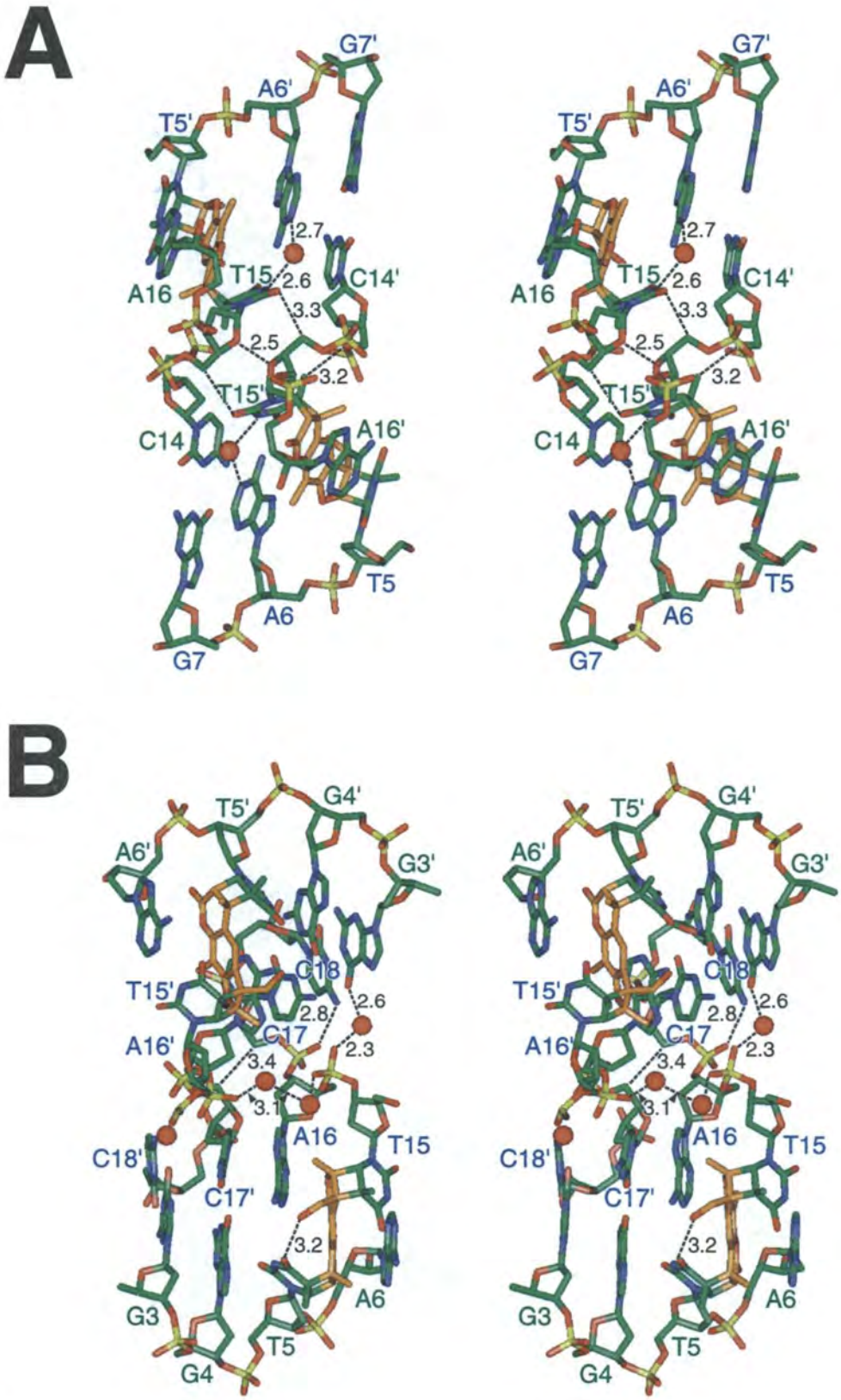

A
G7'
A6'
T5'
2.7
T15
2.6
C14'
A16
3.3
2.5
3.2
T15'
C14
A16'
T5
A6
G7
B
T5'
G4'
A6'
G3'
C18
T15'
2.8
2.6
C17
A16'
2.3
3.4
3.1
A16
T15
C18'
C17'
3.2
A6
G3
T5
G4

Figure 4.4

interstrand psoralen cross-links do not significantly perturb the overall conformation of a stable DNA structure. On the other hand, the structures of cross-linked and unmodified d(CCGCTAGCGG) show that psoralen can greatly perturb the double-helix by destabilizing the DNA P-strand. The distortion is so dramatic that we observe the drug inducing the displacement of four nucleotides from one duplex to cross-over to an adjacent duplex, forming a four-stranded junction.

The psoralenated sequence d(CCGGTACCGG) forms a junction not because of the drug, but in spite of it. This DNA sequence had been shown to crystallize as Holliday junctions in its unmodified form (Eichman *et al.*, 2000) and with the thymine replaced with a guanine, so that two d(G·A) mismatch base pairs are formed across the cross-over of the junction (Ortiz-Lombardía *et al.*, 1999). We now show that even with the thymines cross-linked, the junction remains stable and effectively unaltered. These results further support our previous model that the ACC trinucleotide is the stabilizing core of the junction in this sequence (Eichman *et al.*, 2000). More importantly for this study, we see that the psoralen drug has virtually no effect on the structure of the junction or the double-helical arms that extend from the junction. In particular, we see that the helical arms are not kinked or bent. Even though the psoralenated thymines sit adjacent to the cross-over of the junction, the d(T·A) base pair on the furan-side of the drug remains effectively as a hydrogen bonded pair, and the base planes of the adenine nucleotides are lying nearly perpendicular to the axis of the stacked helical arms. Thus, although we did not explicitly determine the structure of a psoralen-adducted DNA resolved as a duplex, the current structure strongly suggests that the helical conformation of the DNA is not dramatically affected by the drug.

We can readily construct a model for a psoralenated DNA duplex by simply breaking the phosphate bond between the A6 and C7 nucleotides of one cross-over strand, rotating ~180° about the C4'-C5' bond, and reforming the phosphodiester bond (Figure 4.5a). The resulting duplex formed from the two stacked arms of the four stranded assembly is now resolved into two distinct DNA duplexes, with each duplex showing an

Figure 4.5. Models of the psoralen adducted DNAs in their potentially biologically relevant states. In each panel, green thymine bases and DNA strands correspond to the pyrone-side, and blue to the furan-side of the drug. The psoralen molecule is colored red. A. A duplex model of a DNA cross-linked by HMT was constructed by rotating the C4'-C5' bond of cytosine C17 in the crystal structure of HMT-d(CCGGTACCGG) (dark bonds). The subsequent structure was geometry optimized to form a structure having a continuous phosphodeoxyribose backbone (light bonds). The resulting model is shown looking into (left) and perpendicular to (right) the major groove of the DNA helix. The helix axis was calculated using the program CURVES 5.2 (Lavery and Sklenar, 1989). B. The model of an HMT duplex (left), with its destabilized P-strand allowing the single F-strand to serve as a temple for synthesis of a new DNA strand (gold), was constructed from the two strands of the asymmetric unit of HMT-d(CCGCTAGCGG) along with nucleotides C1 to G4 from the symmetry-related molecule. The position of the water mediating the hydrogen bonds of the P-strand d(T·A) base pair is shown in purple. This water as well as the adducted thymine would be displaced by extension of the third strand with a complementary thymine. A possible mechanism for initiation of psoralen cross-link repair in mammalian systems is shown on the right. The nucleotide excision repair (NER) system creates gaps 5' to the cross-link (red bar) on either strand with equal efficiency (Bessho *et al.*, 1997). Futile repair synthesis to fill the gap occurs 10 times more efficiently on the pyrone-side (Mu *et al.*, 2000).

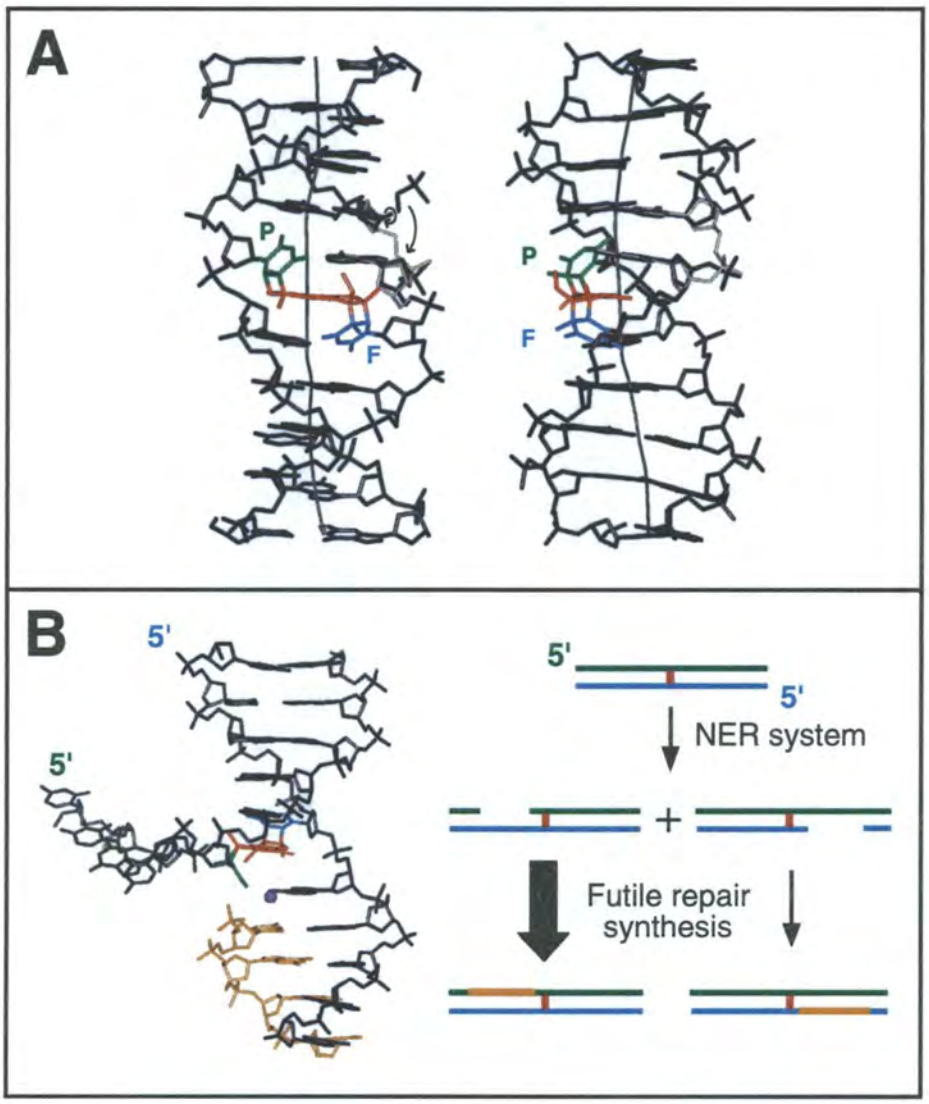


Figure 4.5

unbent trajectory of the helical axis. The double-helical model and the structure of the psoralenated junction in d(CCGGTACCGG) simply show that the drug need not induce a bend, *i.e.*, a straight DNA duplex can readily accommodate the distortions resulting from the cross-link. This is consistent with the more recent unbent NMR structures of the psoralenated DNA duplexes (Hwang *et al.*, 1996; Spielmann *et al.*, 1995). Thus, the psoralenated DNA duplex apparently presents no bend or kink that can serve as a signal for recognition by proteins. However, the drug's effect on the stability of the duplex, as seen in the d(CCGCTAGCGG) junction, may provide such a signal for repair enzymes.

In the drug-induced Holliday junction, we see that the DNA strand cross-linked at the pyrone-side of psoralen becomes destabilized and displaced from the duplex. The four-stranded complex seen here is formed when two such molecules with frayed ends associate. Why does the distorted pyrone-side thymine destabilize the DNA duplex formed from this particular sequence and not from d(CCGGTACCGG)? This seemingly contradictory result can possibly be explained by the small difference between the pyrone-side d(T·A) base pairs in the two structures. In the sequence-dependent junction, there are no hydrogen bonds between the pyrone-side thymine and its complementary adenine. The orientation of this thymine base is fixed through hydrogen bonding with the psoralen hydroxymethyl group. In contrast, the drug-induced structure shows a water-bridging hydrogen bond between the pyrone-side thymine and its partner adenine, and no interaction with the drug substituent groups. This water-mediated hydrogen bond helps to lock the thymine base in its outwardly rotated orientation, which can then destabilize the helix, unlike the pyrone-side thymine of the HMT-ACC junction.

It is unlikely that *two* psoralen cross-links would be found in this type of arrangement in genomic DNAs and, therefore, the junction would appear initially to be an artifact of the sequence design. We can thus ask under what circumstance might this structure be relevant in the cell? The structure of psoralen cross-linked d(CCGCTAGCGG) can serve as a model to explain in several different ways how these types of lesions in the

cell can be recognized by repair proteins. On one hand, the destabilization of the P-strand oligonucleotide is supported by the mechanism to initiate repair of psoralen damaged DNAs in mammalian cells. On the other hand, the Holliday junction itself can serve as a model for the recombinagenic restart of replication after collapse of the replication fork in *E. coli.*

The junction formed by the HMT-d(CCGCTAGCGG) sequence can be construed to be a model of a DNA fragment that has been incised on either side of the psoralen cross-link, *i.e.*, the short F-strand would represent incisions at the furan side of the drug and the P-strand with incisions at the pyrone side (Figure 4.5b). Thus, in a situation where a psoralenated DNA sequence can be incised at either DNA strand, this particular structure tells us that the P-strand is unstable in the duplex and can be displaced. The repair of psoralenated adducts in both *E. coli* and mammalian systems is initiated by incisions to one DNA strand. Mammalian cell extracts show no preference as to which strand is initially incised (Bessho *et al.*, 1997). Despite this, the incised P-strand shows a 10-fold higher probability of initiating futile repair of the adduct (Mu *et al.*, 2000). The current structure suggests that the drug-induced destabilization of the P-strand can facilitate repair from this side of the drug by allowing the complementary F-strand to be recognized as a template for the mammalian repair system (Figure 4.5b). Thus, following its incision during the initiation of repair in mammalian systems, the destabilized P-strand alone may serve as a signal to enzymes to complete the repair process.

4.6 Acknowledgments

We wish to thank Dr. R. Mathies and his lab at UC Berkeley for help with the photocross-linking, G. McDermott at the ALS for help during data collection, J. M. Vargason in this lab for developing the real-space search X-PLOR script, and A. W. Taylor and J. F. Stevens for use of their preparative HPLC. We also thank Professors P. A.

Karplus, C. K. Mathews and K. E. van Holde for critical reading of the manuscript. This work was supported by grants from the National Science Foundation (MCB-9728240), the Oregon Division of the American Cancer Society (J0159880), and the National Institutes of Environmental Health Sciences (ES00210).

Chapter 5

Discussion

The process of homologous recombination requires formation of a four-stranded intermediate, a Holliday junction, in which strands exchange and pair between homologous duplexes (Holliday, 1964). Although strand exchange, branch migration, and junction resolution are known to be predominantly catalyzed by proteins (reviewed in Kowalczykowski *et al.*, 1994; West, 1997), it still is not clear what role DNA structure plays in these mechanisms. Within the past year, six crystal structures of four-way junctions have been solved in the absence of proteins, three of which have been presented here. All of these structures are the stacked-X form of the junction. These detailed structures allow us to better understand the intrinsic interactions that stabilize four-way junctions at the nucleic acid level. From these structural studies, five main principles regarding DNA junctions emerge.

The strand cross-overs at the junction do not distort the nucleic acid conformation of the stacked arms. Except for the gaps in the phosphoribose backbones across the junction, the stacked duplex arms are virtually identical to two continuous double helices. As in standard double-helical structures of RNA and DNA, the conformation of the double helix across the junction is determined by the particular nucleotide sequence and ribose conformation along each strand. For example, the helical parameters describing the structure of the d(GpT/A*C) base step (where A* is the nucleotide at the junction cross-over) are identical across the junction in d(CCGGTACCGG) (Eichman *et al.*, 2000) and in the B-DNA duplex of d(CCAGTACTGG) (Kielkopf *et al.*, 2000). In addition, the mismatched d(G·A) base pairs in the d(CCGGGACCGG) junction (Ortiz-Lombardía *et al.*, 1999) impose the same distortions to the duplex arms as in the B-DNA duplex structure of

d(CCAAGATTGG) (Privé *et al.*, 1987). In the RNA-DNA junction structures (Nowakowski *et al.*, 1999; Nowakowski *et al.*, 2000), the stacked arms containing the RNA strand (and thus 3'-endo sugar puckers) are in the A conformation, which has been shown to be true for all DNA-RNA duplexes (Saenger, 1984). Finally, the structure of the drug-induced junction in HMT-d(CCGCTAGCGG) shows that the detailed interactions at the strand cross-overs can be dramatically different without distorting the DNA conformation across the junction. Thus, the nucleic acid conformation across the junction is sequence-dependent.

The stability of the Holliday junction, with respect to the position of the strand cross-overs along the nucleotide sequence and the rigidity of the junction, is determined by sequence-dependent base stacking and hydrogen bonding. For any given stacked-X junction, there are two possible ways in which to pair the four arms into stacked duplexes. In addition, branch migration of the strand cross-over point can take place in homologous sequences. Studies of immobile synthetic junctions formed from four unique DNA strands have shown that the specific nucleotides both directly at and further removed from the junction determine the distribution of the two possible stacked conformers, suggesting that base stacking is not the only determinant of junction stability. In support of this, we find in the d(CCGGNACCGG) crystal structures that the specific hydrogen bonds from the d(C8·G3) base pair to the phosphates at the cross-over are more important than base stacking in defining which conformer is favored. Conformational re-arrangement of the stacked arms would preserve the d(G*T/ApC) stacking across the junction, but would destroy the hydrogen bonding between duplexes. Therefore, the sequence-dependent contacts between the major groove of one arm and the backbone at the cross-over play an important role in defining the junction conformation, as they do in B-DNA single crystals (Timsit and Moras, 1991; Timsit and Moras, 1994).

On the other hand, base stacking seems to be more important in defining the particular position of the cross-over in the d(CCGGTACCGG) crystal structure, which has

a fixed junction position in spite of its ability to migrate along the nucleotide sequence. This fixed junction cannot be entirely accounted for by the specific hydrogen bonding at the ACC core for two reasons. First, migration of the junction to the d(CCGGT*ACCGG) position in this sequence would be favored by the cytosine C7 at the hydrogen bonding position, but would place the junction at a d(TpA) step, which has weaker base stacking than d(GpT/ApC) steps (Dickerson, 1999). Secondly, although position 8 of d(CCGCTAGCGG) is a cytosine, this sequence did not form a junction. A cross-over 3' to adenine A6 in this sequence would place the junction at a d(CpT/ApG) step, which again has weaker stacking interactions than d(GpT/ApC) steps (Dickerson, 1999). In addition, the sequence-dependent ACC junction is able to accommodate a psoralen (HMT) cross-link adjacent to the cross-overs, in spite of the ability of HMT to induce the formation of a Holliday junction at a different position in the sequence. Therefore, both the base stacking and hydrogen bonding interactions at the ACC triplet contribute to the ability of these sequences to crystallize with the junction held rigidly at this position.

The geometry of the junction is influenced by the interduplex interactions further removed from the junction. Although a 60° interduplex angle has been estimated from several different methods (Cooper and Hagerman, 1989; Mao *et al.*, 1999; Murchie *et al.*, 1989), a range of angles have been observed in the junction crystal structures (Table 5.1). A van der Waals contact was observed between phosphate oxygens at the ends of the arms in the native, mismatched, and psoralenated ACC junction structures. Although no ordered solvent molecules were found to bridge these negatively charged groups, these interactions are more than likely stabilized by cations. Nevertheless, several lines of evidence suggest that these interactions are important in the geometry of the four-way junction. The 55° and -80° RNA-DNA junctions have identical nucleotides at the strand cross-overs, with a one-base-pair arm extending from the junction. Consequently, the interduplex phosphate bridges do not exist, and the arms are capable of more conformational variation. In addition, in the ACC structures, the slight overwinding of the four-base-pair arms is associated with

Table 5.1. Angles between duplex arms in four-way junction crystal structures.

	Angle (°)	Space Group	Reference
RNA-DNA			
DNAzyme	55	$P6_122$	(Nowakowski *et al.*, 1999)
DNAzyme analog	-80	*I222*	(Nowakowski *et al.*, 2000)
DNA-DNA			
d(CCGGGACCGG)	41	*C2*	(Ortiz-Lombardía *et al.*, 1999)
d(CCGGTACCGG)	41	*C2*	(Eichman *et al.*, 2000)
d(CCGGTACCGG)-HMT [a]	36	*C2*	(Eichman *et al.*, 200X)
d(CCGCTAGCGG)-HMT [a]	37	*C2*	(Eichman *et al.*, 200X)

[a] HMT: 4'-hydroxymethyl-4,5',8-trimethylpsoralen

an underwinding of the adjacent six-base-pair arms, which maintains the interduplex contacts. Likewise, there is a slight tilt of one duplex in HMT-d(CCGGTACCGG) relative to the native sequence that preserves these interduplex contacts, in spite of the elongation of the helix as a result of HMT intercalation. It appears that the sequence-dependent groove-backbone interactions at the ACC core alone do not define this interhelical angle, because a 37° angle was observed in the HMT-d(CCGCTAGCGG) structure, which has dramatically different contacts at the strand cross-overs than the ACC junctions. Thus, the interactions distant from the junction cross-overs play important roles in defining the overall geometry of the stacked arms.

Specific binding of divalent cations do not seem to stabilize the junction. It is surprising that no divalent cations were observed in any of the junction crystal structures, with the exception of the -80° RNA-DNA junction (Nowakowski *et al.*, 2000), given the highly negative electrostatic potential at the junctions and that the crystals were grown in the presence of Mg^{2+} (Eichman *et al.*, 200X; Nowakowski *et al.*, 1999; Nowakowski *et al.*, 2000; Ortiz-Lombardía *et al.*, 1999) or Ca^{2+} (Eichman *et al.*, 2000). In the -80° RNA-DNA structure, the stacking of the base pairs surrounding the point of strand exchange is disrupted relative to A-form RNA. A hydrated Mg^{2+} ion was bound in the major groove of this region, potentially stabilizing the distorted junction. In the d(CCGGTACCGG) junction, very few ordered waters were observed in the major groove. However, this DNA junction is stabilized by a sodium ion completely buried beneath the solvent accessible surface at the center of the strand cross-overs. In the 55° RNA-DNA junction, a Co^{3+} was bound in the same location as the major groove-phosphate bridging water in the DNA junctions. The geometry of the DNA structures is not able to accommodate the larger cobalt ion at this position. Thus, we can speculate that specific binding of cations play an active role during junction formation, while diffuse clouds stabilize the completely folded structure.

Psoralen cross-links induce the formation of Holliday junctions in DNA crystals. Four-way junctions are important intermediates during DNA repair events. The drug-induced junction in the HMT-d(CCGCTAGCGG) structure is an important model for the recombination processes required to restart replication after an encounter with a DNA lesion (Figure 5.1). This structure shows that psoralen is capable of accommodating, and even influencing the formation of a Holliday junction. This is consistent with psoralen cross-links being highly recombinagenic (Cassier *et al.*, 1984; Cole, 1973; Vos and Hanawalt, 1989). During replication in bacterial cells, DNA lesions and strand breaks can lead to inactivation of the replication fork, creating DNA gaps or double-strand breaks. Non-mutagenic re-activation of the replication fork is now thought to require homologous recombination and the repair enzymes involved in these mechanisms (reviewed in Cox *et al.*, 2000). From the crystal structure, we can easily construct a junction consisting of a psoralenated DNA duplex and a homologous non-cross-linked DNA duplex (Figure 5.1). This is a better model for the recombination intermediate under cellular conditions than the double cross-link in the crystal structure. Such a structure could be a signal to repair enzymes, leading ultimately to non-mutagenic replication of the damaged DNA, and even to full repair of the lesion through recombination. However, how any lesion, particularly a double-strand cross-link, would be dealt with before or after resolution of the Holliday junction is still unclear. Nonetheless, the structures of the psoralen cross-linked DNAs as Holliday junctions raise interesting questions regarding the processing of such lesions in the cell.

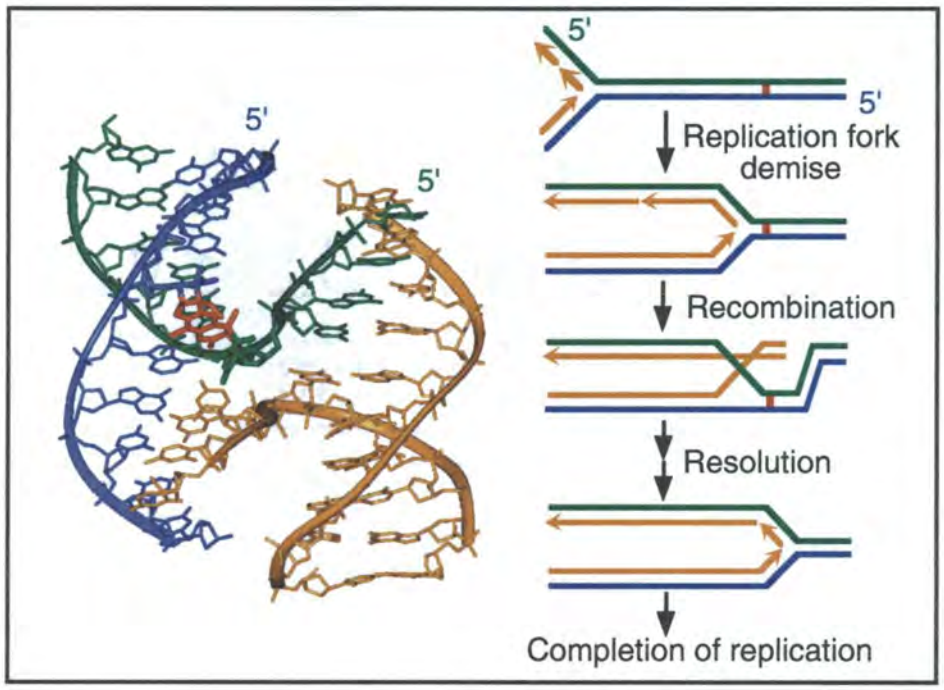


Figure 5.1. Model of a junction containing a single psoralen adduct (left) as might be formed by homologous recombination to rescue replication fork collapse (right). A Holliday junction formed from a psoralen cross-linked duplex assembled with the homologous unmodified DNA duplex (gold strands) was assembled with the asymmetric unit of the junction in the psoralen adduct complexed with the unmodified duplex of d(CCGCTAGCGG) (Eichman *et al.*, 2000). The geometries of the unmodified DNA strands were optimized using the AMBER forcefield as implemented in the Discover module of the program InsightII (Molecular Simulation, Inc, San Diego, CA). One possible pathway for re-initiation of inactivated replication forks is shown and was adapted from (Cox *et al.*, 2000).

Bibliography

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. USA*, **97**, 8257-8262.

Baden, H.P., Parrington, J.M., Delhanty, J.D.A. and Pathak, M.A. (1972) DNA synthesis in normal and xeroderma pigmentosum fibroblasts following treatment with 8-methoxypsoralen and long wave ultraviolet light. *Biochim. Biophys. Acta*, **262**, 247-255.

Baikalov, I., Grzeskowiak, K., Yanagi, K., Quintana, J. and Dickerson, R.E. (1993) The crystal structure of the trigonal decamer C-G-A-T-C-G-6meA-T-C-G: a B-DNA helix with 10.6 base-pairs per turn. *J. Mol. Biol.*, **231**, 768-784.

Bennett, R.J. and West, S.C. (1995) Structural analysis of the RuvC-Holliday junction complex reveals an unfolded junction. *J. Mol. Biol.*, **252**, 213-226.

Benson, F.E., Stasiak, A. and West, S.C. (1994) Purifiacation and characterization of the human Rad51 protein, an analogue of *E. coli* RecA. *EMBO J.*, **13**, 5764-5771.

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

Berman, H.M. and Schneider, B. (1999) Nucleic acid hydration. In Neidle, S. (ed.) *Oxford Handbook of Nucleic Acid Structure*. Oxford University Press, New York, pp. 295-312.

Bessho, T., Mu, D. and Sancar, A. (1997) Initiation of DNA interstrand cross-link repair in humans: the nucleotide excision repair system makes dual incisions 5' to the cross-linked base and removes a 22- to 28-nucleotide-long damage-free strand. *Mol. Cell. Biol.*, **17**, 6822-6830.

Bethea, D., Fullmer, F., Syed, S., Seltzer, G., Tiano, J., Rischko, C., Gillespie, L., Brown, D. and Gasparro, F.P. (1999) Psoralen photobiology and photochemistry: 50 years of science and medicine. *J. Dermatol. Sci.*, **19**, 78-88.

Bhattacharyya, A., Murchie, A.I.H., von Kitzing, E., Diekmann, S., Kemper, B. and Lilley, D.M.J. (1991) Model for the interaction of DNA junctions and resolving enzymes. *J. Mol. Biol.*, **221**, 1191-1207.

Branlant, C., Krol, A. and Ebel, J.-P. (1981) The conformation of chicken, rat and human U1A RNAs in solution. *Nucleic Acids Res.*, **9**, 841-858.

Broker, T.R. and Lehman, I.R. (1971) Branched DNA molecules: intermediates in T4 recombination. *J. Mol. Biol.*, **60**, 131-149.

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

Brünger, A.T. (1992b) *X-PLOR version 3.1: a system for X-ray crystallography and NMR.* Yale University Press, New Haven, USA.

Carlström, G. and Chazin, W.J. (1996) Sequence dependence and direct measurement of crossover isomer distribution in model Holliday junctions using NMR spectroscopy. *Biochemistry*, **35**, 3534-3544.

Cassier, C., Chanet, R. and Moustacchi, E. (1984) Mutagenic and recombinagenic effects of DNA cross-links induced in yeast by 8-methoxypsoralen photoaddition. *Photochem. Photobiol.*, **39**, 799-803.

Chen, S.-m., Heffron, F. and Chazin, W.J. (1993) Two-dimensional 1H NMR studies of 32-base-pair synthetic immobile Holliday junctions: complete assignments of the labile protons and identification of the base-pairing scheme. *Biochemistry*, **32**, 319-326.

Chiu, T.K., Kaczor-Grzeskowiak, M. and Dickerson, R.E. (1999) Absence of minor groove monovalent cations in the crosslinked dodecamer C-G-C-G-A-A-T-T-C-G-C-G. *J. Mol. Biol.*, **292**, 589-608.

Churchill, M.E.A., Tullius, T.D., Kallenbach, N.R. and Seeman, N.C. (1988) A Holliday recombination intermediate is twofold symmetric. *Proc. Natl. Acad. Sci. USA*, **85**, 4653-4656.

Cimino, G.D., Gamper, H.B., Isaacs, S.T. and Hearst, J.E. (1985) Psoralens as photoactive probes of nucleic acid structure and function: organic chemistry, photochemistry, and biochemistry. *Ann. Rev. Biochem.*, **54**, 1151-1193.

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

Clegg, R.M., Murchie, A.I.H., Zechel, A. and Lilley, D.M.J. (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.

Cole, R.S. (1973) Repair of DNA containing interstrand crosslinks in *Escherichia coli*: sequential excision and recombination. *Proc. Natl. Acad. Sci. USA*, **70**, 1064-1068.

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

Cooper, J.P. and Hagerman, P.J. (1989) Geometry of a branched DNA structure in solution. *Proc. Natl. Acad. Sci USA*, **86**, 7336-7340.

Coste, F., Malinge, J.M., Serre, L., Shepard, W., Roth, M., Leng, M. and Zelwer, C. (1999) Crystal structure of a double-stranded DNA containing a cisplatin interstrand cross-link at 1.63 Å resolution: hydration at the platinated site. *Nucleic Acids Res*, **27**, 1837-1846.

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.

de Massy, B., Weisberg, R.A. and Studier, F.W. (1987) Gene 3 endonuclease of bacteriophage T7 resolves conformationally branched structures in double-stranded DNA. *J. Mol. Biol.*, **193**, 359-376.

Dickerson, R.E. (1999) Helix structure and molecular recognition by B-DNA. In Neidle, S. (ed.) *Oxford Handbook of Nucleic Acid Structure*. Oxford University Press, New York, pp. 145-197.

Dickie, P., McFadden, G. and Morgan, A.R. (1987) The site-specific cleavage of synthetic Holliday junction analogs and related branched DNA structures by bacteriophage T7 endonuclease I. *J. Biol. Chem.*, **262**, 14826-14836.

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

Duckett, D.R. and Lilley, D.M.J. (1991) Effects of base mismatches on the structure of the four-way DNA junction. *J. Mol. Biol.*, **221**, 147-161.

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

Eichman, B.F., Mooers, B.H.M., Alberti, M., Hearst, J.E. and Ho, P.S. (200X) The crystal structures of psoralen cross-linked DNAs: drug dependent formation of Holliday junctions. *submitted for publication*.

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

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**, 11349-11358.

Gellert, M., Mizuuchi, K., H., O.D.M., Ohmori, H. and Tomizawa, J. (1979) DNA gyrase and DNA supercoiling. *Cold Spring Harb. Symp. Quant. Biol.*, **43**, 35-40.

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**, 1022-1029.

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.

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.

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

Grzeskowiak, K., Goodsell, D.S., Kaczor-Grzeskowiak, M., Cascio, D. and Dickerson, R.E. (1993) Crystallographic analysis of C-C-A-A-G-C-T-T-G-G and its implications for bending in B-DNA. *Biochemistry*, **32**, 8923-8931.

Hampel, A. and Tritz, R. (1989) RNA catalytic properties of the minimum (-)sTRSV sequence. *Biochemistry*, **28**, 4929-4933.

Han, G.W., Kopka, M.L., Cascio, D., Grzeskowiak, K. and Dickerson, R.E. (1997) Structure of a DNA analog of the primer for HIV-1 RT second strand synthesis. *J. Mol. Biol.*, **269**, 811-826.

Haran, T.E. and Crothers, D.M. (1988) Phased psoralen cross-links do not bend the DNA double-helix. *Biochemistry*, **27**, 6967-6971.

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 Å resolution. *Nat. Struct. Biol.*, **5**, 441-446.

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

Heinemann, U., Alings, C. and Bansal, M. (1992) Double helix conformation, groove dimensions and ligand binding potential of a G/C stretch in B-DNA. *EMBO J.*, **11**, 1931-1939.

Helm, T.N., Dijkstra, J.W.E., Ferrara, R.J., Jr. and Glanz, S. (1991) PUVA therapy. *Am. Fam. Physician*, **43**, 908-912.

Hoess, R., Wierzbicki, A. and Abremski, K. (1987) Isolation and characterization of intermediates in site-specific recombination. *Proc. Natl. Acad. Sci. USA*, **84**, 6840-6844.

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

Hwang, G.-S., Kim, J.-K. and Choi, B.-S. (1996) The solution structure of a psoralen cross-linked DNA duplex by NMR and relaxation matrix refinement. *Biochem. Biophys. Res. Comm.*, **219**, 191-197.

Isaacs, S.T., Shen, C.-K.J., Hearst, J.E. and Rapoport, H. (1977) Synthesis and characterization of new psoralen derivatives with superior photoreactivity with DNA and RNA. *Biochemistry*, **16**, 1058-1064.

Kanne, D., Straub, K., Hearst, J.E. and Rapoport, H. (1982a) Isolation and characterization of pyrimidine-psoralen-pyrimidine photoadducts from DNA. *J. Am. Chem. Soc.*, **104**, 6754-6764.

Kanne, D., Straub, K., Rapoport, H. and Hearst, J.E. (1982b) Psoralen-deoxyribonucleic acid photoreaction. Characterization of the monoaddition products from 8-methoxypsoralen and 4,5',8-trimethylpsoralen. *Biochemistry*, **21**, 861-871.

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

Kissinger, C.R., Gehlhaar, D.K. and Fogel, D.B. (1999) Rapid automated molecular replacement by evolutionary search. *Acta Crystallogr.*, **D 55**, 484-491.

Kitts, P.A. and Nash, H.A. (1987) Homology-dependent interactions in phage λ site-specific recombination. *Nature*, **329**, 346-348.

Kowalczykowski, S.C. (2000) Initiation of genetic recombination and recombination-dependent replication. *Trends Biochem. Sci.*, **292**, 156-165.

Kowalczykowski, S.C., Dixon, D.A., Eggleston, A.K., Lauder, S.D. and Rehrauer, W.M. (1994) Biochemistry of homologous recombination in *Escherichia coli. Microbiol. Rev.*, **58**, 401-465.

Lanskin, J.D. (1994) Cellular and molecular mechanisms in photochemical sensitization: studies on the mechanism of action of psoralens. *Food Chem. Toxicol.*, **32**, 119-127.

Lanskin, J.D. and Lee, E. (1991) Psoralen binding and inhibition of epidermal growth factor binding by psoralen/ultraviolet light (PUVA) in human epithelial cells. *Biochem. Pharmacol.*, **41**, 125-132.

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

Lilley, D.M. (1980) The inverted repeat as a recognizable structural feature in supercoiled DNA molecules. *Proc. Natl. Acad. Sci. U S A*, **77**, 6468-6472.

Lilley, D.M.J. (1999) Structures and interactions of helical junctions in nucleic acids. In Neidle, S. (ed.) *Oxford Handbook of Nucleic Acid Structure*. Oxford University Press, New York, pp. 471-498.

Lipanov, A., Kopka, M.L., Kaczor-Grzeskowiak, M., Quintana, J. and Dickerson, R.E. (1993) Structure of the B-DNA decamer C-C-A-A-C-I-T-T-G-G in two different space groups: conformational flexibility of B-DNA. *Biochemistry*, **32**, 1373-1389.

Lüftl, M., Röcken, M., Plewig, G. and Degitz, K. (1998) PUVA inhibits DNA replication, but not gene transcription at nonlethal dosages. *J. Invest. Dermatol.*, **111**, 399-405.

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.

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. USA*, **94**, 9080-9084.

Møllegaard, N.E., Murchie, A.I.H., Lilley, D.M.J. and Nielsen, P.E. (1994) Uranyl photoprobing of a four-way DNA junction: evidence for specific metal ion binding. *EMBO J.*, **13**, 1508-1513.

Mu, D., Bessho, T., Nechev, L.V., Chen, D.J., Harris, T.M., Hearst, J.E. and Sancar, A. (2000) DNA interstrand cross-links induce futile repair synthesis in mammalian cell extracts. *Mol. Cell. Biol.*, **20**, 2446-2454.

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

Murchie, A.I.H., Thomson, J.B., Walter, F. and Lilley, D.M.J. (1998) Folding of the hairpin ribozyme in its natural conformation achieves close physical proximity of the loops. *Mol. Cell*, **1**, 873-881.

Murshudov, G.N., Vagin, A.A., Lebedev, A., Wilson, K.S. and Dodson, E.J. (1999) Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr. D*, **55**, 247-255.

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**, 151-156.

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

Orr-Weaver, T.L., Szostak, J.W. and Rothstein, R.J. (1981) Yeast transformation: a model system for the study of recombination. *Proc. Natl. Acad. Sci. USA*, **78**, 6354-6358.

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

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

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**, 1819-1826.

Panyutin, I.G. and Hsieh, P. (1994) The kinetics of spontaneous DNA branch migration. *Proc. Natl. Acad. Sci. USA*, **91**, 2021-2025.

Peckler, S., Graves, B., Kanne, D., Rapoport, H., Hearst, J.E. and Kim, S.-H. (1982) Structure of a psoralen-thymine monoadduct formed in photoreaction with DNA. *J. Mol. Biol.*, **162**, 157-172.

Pflugrath, J.W. (1999) The finer things in X-ray diffraction data collection. *Acta Crystallogr.*, **D 55**, 1718-1725.

Pikkemaat, J.A., van den Elst, H., van Boom, J.H. and Altona, C. (1994) NMR studies and conformational analysis of a DNA four-way junction formed in a linear synthetic oligonucleotide. *Biochemistry*, **33**, 14896-14907.

Potter, H. and Dressler, D. (1976) On the mechanism of genetic recombination: electron microscopic observation of recombination intermediates. *Proc. Natl. Acad. Sci. USA*, **73**, 3000-3004.

Potter, H. and Dressler, D. (1978) *In vitro* system from *Escherichia coli* that catalyses generalised genetic recombination. *Proc. Natl. Acad. Sci. USA*, **75**, 3698-3702.

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

Privé, 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**, 177-199.

Raaijmakers, H., Vix, O., Törõ, 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**, 1447-1458.

Robinson, B.H. and Seeman, N.C. (1987) Simulation of double-stranded branch point migration. *Biophys J.*, **51**, 611-626.

Roe, S.M., Barlow, T., Brown, T., Oram, M., Keeley, A., Tsaneva, I.R. and Pearl, L.H. (1998) Crystal structure of an octameric RuvA-Holliday junction complex. *Cell*, **2**, 361-372.

Saenger, W. (1984) *Principles of Nucleic Acid Structure*. Springer-Verlag, New York.

Sastry, S.S. and Hearst, J.E. (1991a) Studies on the interaction of T7 RNA polymerase with a DNA template containing a site-specifically placed psoralen cross-Link I. Characterization of elongation complexes. *J. Mol. Biol.*, **221**, 1091-1110.

Sastry, S.S. and Hearst, J.E. (1991b) Studies on the interaction of T7 RNA polymerase with a DNA template containing a site-specifically placed psoralen cross-Link II. Stability and some properties of elongation Complexes. *J. Mol. Biol.*, **221**, 1111-1125.

Shakked, Z., Guzikevich-Guerstein, G., Frolow, F., Rabbinovich, D., Joachimiak, A. and Sigler, P.B. (1994) Determinants of repressor/operator recognition from the structure of the trp operator binding site. *Nature*, **368**, 469-473.

Shi, Y.-B., Gamper, H., Van Houten, B. and Hearst, J.E. (1988a) Interaction of *E. coli* RNA polymerase with DNA in an elongation complex arrested at a specific psoralen cross-link site. *J. Mol. Biol.*, **199**, 277-293.

Shi, Y.-B., Griffith, J., Gamper, H. and Hearst, J.E. (1988b) Evidence for structural deformation of the DNA helix by a psoralen diadduct but not by a monoadduct. *Nucleic Acids Res.*, **16**, 8945-8953.

Shinohara, A., Ogawa, H. and Ogawa, T. (1992) Rad51 protein involved in repair and recombination in *S. cerevisiae* is a RecA-like protein. *Cell*, **69**, 457-470.

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**, 8341-8355.

Sigal, N. and Alberts, B. (1972) Genetic recombination: the nature of a crossed strand-exchange between two homologous DNA molecules. *J. Mol. Biol.*, **71**, 789-793.

Sinden, R.R. and Hagerman, P.J. (1984) Interstrand psoralen cross-links do not introduce appreciable bends in DNA. *Biochemistry*, **23**, 6299-6303.

Spielmann, H.P., Dwyer, T.J., Hearst, J.E. and Wemmer, D.E. (1995) Solution structures of psoralen monoadducted and cross-linked DNA oligomers by NMR spectroscopy and restrained molecular dynamics. *Biochemistry*, **34**, 12937-12953.

Spielmann, H.P., Sastry, S.S. and Hearst, J.E. (1992) Methods for the large-scale synthesis of psoralen furan-side monoadducts and diadducts. *Proc. Natl. Acad. Sci. USA*, **89**, 4514-4518.

Srinivasan, A.R. and Olson, W.K. (1994) Computer models of DNA four-way junctions. *Biochemistry*, **33**, 9389-9404.

Straub, K., Kanne, D., Hearst, J.E. and Rapoport, H. (1981) Isolation and characterization of pyrimidine-psoralen photoadducts from DNA. *J. Am. Chem. Soc.*, **103**, 2347-2355.

Sun, W., Mao, C., Liu, F. and Seeman, N.C. (1998) Sequence dependence of branch migratory minima. *J. Mol. Biol.*, **282**, 59-70.

Taylor, R. and Kennard, O. (1982) Crystallographic evidence for the existence of C-H···O, C-H···N, and C-H···Cl hydrogen bonds. *J. Am. Chem. Soc.*, **104**, 5063-5070.

Thompson, B.J., Camien, M.N. and Warner, R.C. (1976) Kinetics of branch migration in double-stranded DNA. *Proc. Natl. Acad. Sci. USA*, **73**, 2299-2303.

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

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

Timsit, Y., Vilbois, E. and Moras, D. (1991) Base-pairing shift in the major groove of (CA)n tracts by B-DNA crystal structures. *Nature*, **354**, 167-170.

Timsit, Y., Westhof, E., Fuchs, R.P.P. and Moras, D. (1989) Unusual helical packing in crystals of DNA bearing a mutation hot spot. *Nature*, **341**, 459-462.

Tomic, M.T., Wemmer, D.E. and Kim, S.-H. (1987) Structure of a psoralen cross-linked DNA in solution by nuclear magnetic resonance. *Science*, **238**, 1722-1725.

Trosko, J.E. and Isoun, M. (1971) Photosensitizing effect of trisoralen on DNA synthesis in human cells grown *in vitro*. *Int. J. Radiat. Biol.*, **19**, 87-92.

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

Voet, D. and Voet, J.G. (1995) *Biochemistry*. John Wiley & Sons, Inc., New York.

von Kitzing, E., Lilley, D.M.J. and Diekmann, S. (1990) The stereochemistry of a four-way DNA junction: a theoretical study. *Nucleic Acids Res.*, **18**, 2671-2683.

Vos, J.-M.H. and Hanawalt, P.C. (1989) DNA interstrand cross-links promote chromosomal integration of a selected gene in human cells. *Mol. Cell. Biol.*, **9**, 2897-2905.

West, S.C. (1997) Processing of recombination intermediates by the RuvABC proteins. *Annu. Rev. Genet.*, **31**, 213-244.

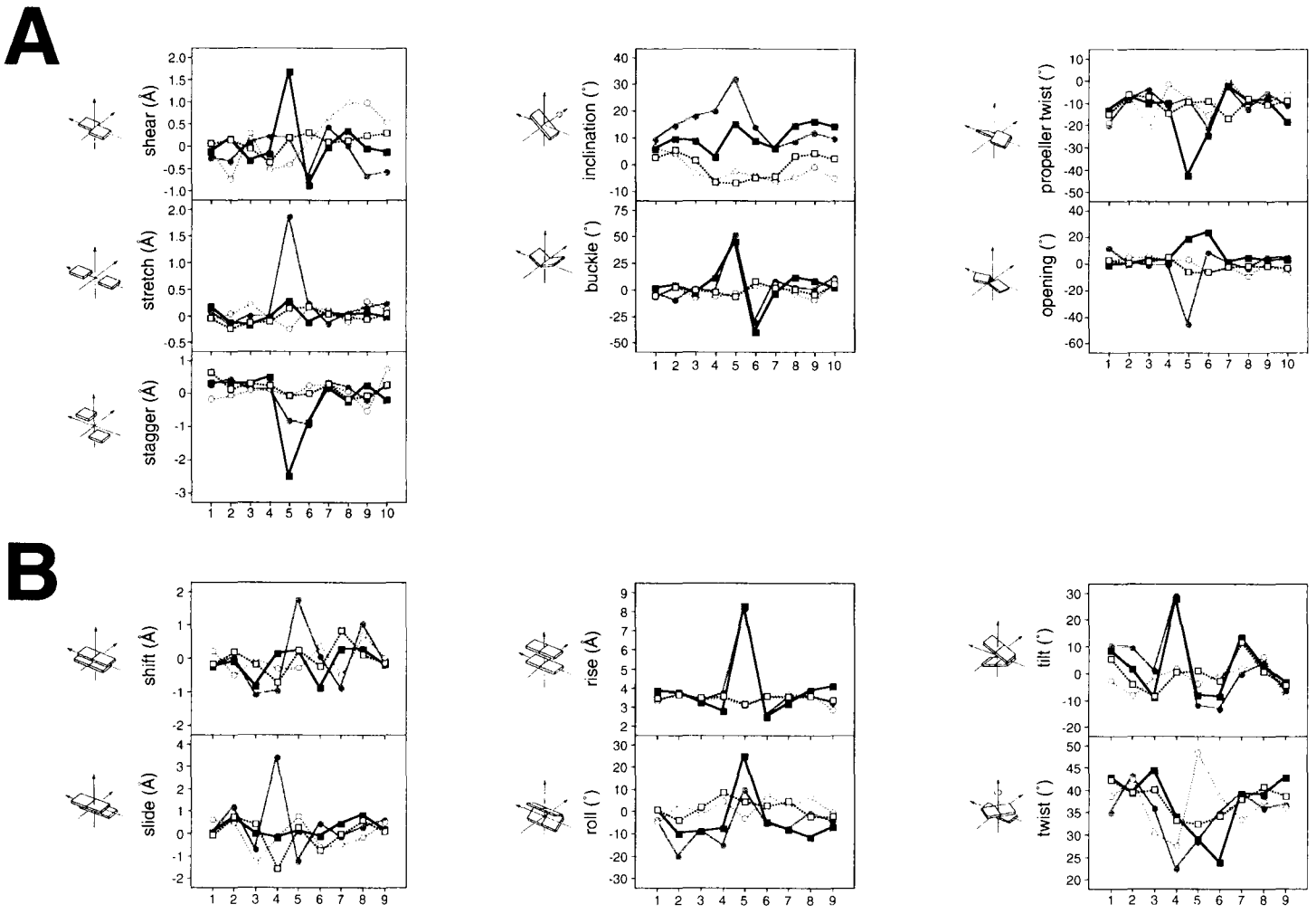
White, M.F., Giraud-Panis, M.-J.E., Pöhler, J.R.G. and Lilley, D.M.J. (1997) Recognition and manipulation of branched DNA structure by junction-resolving enzymes. *J. Mol. Biol.*, **269**, 647-664.

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**, 827-841.

Zhang, S. and Seeman, N.C. (1994) Symmetric Holliday junction crossover isomers. *J. Mol. Biol.*, **238**, 658-668.

APPENDIX

Appendix. Effect of psoralen cross-links on the strand cross-over and B-DNA arms of Holliday junctions. Base pair (A) and base step (B) parameters are shown for cross-linked (filled symbols) and native (open symbols) DNA sequences d(CCGGTACCGG) (squares) and d(CCGCTAGCGG) (circles). Strands are numbered from the 5'-end of the pyrone strand. The strand cross-overs to the junction occur at between base pairs 4 and 5, and at base step 4. Data were calculated using the program CURVES 5.2 (Lavery and Sklenar, 1989).


A
shear (Å)
stretch (Å)
stagger (Å)
inclination (°)
buckle (°)
propeller twist (°)
opening (°)
B
shift (Å)
slide (Å)
rise (Å)
roll (°)
tilt (°)
twist (°)

Appendix