

A Rare Nucleotide Base Tautomer in the Structure of an Asymmetric DNA Junction[†]

Patricia Khuu[‡] and P. Shing Ho^{*,‡,§}

[‡]*Department of Biochemistry and Biophysics, ALS 2011, Oregon State University, Corvallis, Oregon 97331, and* [§]*Department of Biochemistry and Molecular Biology, 1870 Campus Delivery, Colorado State University, Fort Collins, Colorado 80523*

Received May 13, 2009; Revised Manuscript Received June 26, 2009

ABSTRACT: The single-crystal structure of a DNA Holliday junction assembled from four unique sequences shows a structure that conforms to the general features of models derived from similar constructs in solution. The structure is a compact stacked-X form junction with two sets of stacked B-DNA-type arms that coaxially stack to form semicontinuous duplexes interrupted only by the crossing of the junction. These semicontinuous helices are related by a right-handed rotation angle of 56.5°, which is nearly identical to the 60° angle in the solution model but differs from the more shallow value of ~40° for previous crystal structures of symmetric junctions that self-assemble from single identical inverted-repeat sequences. This supports the model in which the unique set of intramolecular interactions at the trinucleotide core of the crossing strands, which are not present in the current asymmetric junction, affects both the stability and geometry of the symmetric junctions. An unexpected result, however, is that a highly wobbled A·T base pair, which is ascribed here to a rare enol tautomer form of the thymine, was observed at the end of a CCCC/GGGG sequence within the stacked B-DNA arms of this 1.9 Å resolution structure. We suggest that the junction itself is not responsible for this unusual conformation but served as a vehicle for the study of this CG-rich sequence as a B-DNA duplex, mimicking the form that would be present in a replication complex. The existence of this unusual base lends credence to and defines a sequence context for the “rare tautomer hypothesis” as a mechanism for inducing transition mutations during DNA replication.

The four-stranded DNA complex known as the Holliday junction is the central intermediate in homologous recombination and recombination-mediated genetic mechanisms (1), including DNA repair and replication, resumption of stalled replication forks, and viral genome integration (2–7). Although homologous recombination by definition involves symmetric sequences, asymmetric junctions can be assembled from four unique sequences to lock the position of DNA crossover through base pair complementarity and, thereby, allow the structure and dynamics of junctions to be studied in solution (8) and in single molecules (9, 10). Structural models derived from solution studies show that DNA junctions under physiological salt conditions adopt a compact “stacked-X” conformation (8) in which the arms pair and coaxially stack to form two nearly continuous duplexes that are related by a 60° angle [the J_{twist} (11)] across the junction crossover (Figure 1A).

A number of single-crystal structures of symmetric junctions that self-assembled from single inverted-repeat or near-repeat sequences have now been reported (12–16). Such symmetric junctions are stabilized not by base complementarity but by a set of sequence-dependent intramolecular interactions that lock the tight U-turn of the crossover to prevent migration of the junction in the crystals (13) and in solution (17). The crystal structures recapitulate the general features of the stacked-X model (18), except that the J_{twist} is much shallower (~40°) (19). The question is whether this difference in geometry is associated

with the differences in the environment of the junction (solution vs crystal) or in the DNA constructs (asymmetric vs symmetric). This study reports the 1.9 Å structure of an asymmetric junction constructed from four unique sequences (Figure 1B), which serves to bridge the structural and dynamic properties of the junction in solution with the atomic details of the crystal structures.

An additional interesting feature of this structure is the observation that an A·T base pair within one of the B-DNA arms is wobbled and, thus, does not conform to the geometric requirements for Watson–Crick base pair complementarity. Base pair complementarity is the basis for accurate replication and transcription of the genetic information in DNA for creation of new daughter DNA and RNA molecules, respectively. Proper base pairing depends on patterns of hydrogen bond donors and acceptors that give rise to the now well-established Watson–Crick-type pairing of adenine (A) with thymine (T) and guanine (G) with cytosine (C) bases (20). These hydrogen bonding patterns require that the A and C bases adopt the stable amino tautomer form while T and G adopt the keto tautomer form that are now well recognized in biochemistry and molecular biology. The existence of rare tautomers of DNA bases, however, has been proposed as a mechanism for induction of transition mutations during replication; this is the “rare tautomer hypothesis” for transition mutations (21). We suggest that the wobbled A·T base pair seen in the current structure arises from having one of the bases adopt a rare tautomer that is stabilized by a CCCC·GGGG sequence within the DNA, and not by the junction itself. This model helps to explain the sequence context observed for rates of nucleotide misincorporation during DNA replication.

[†]This work was supported by grants from the Oregon Medical Research Foundation and from the National Institutes of Health (R01GM62957A), and from Colorado State University.

*To whom correspondences should be addressed. Phone: (970) 491-0569. Fax: (970) 491-0494. E-mail: Shing.Ho@ColoState.edu.

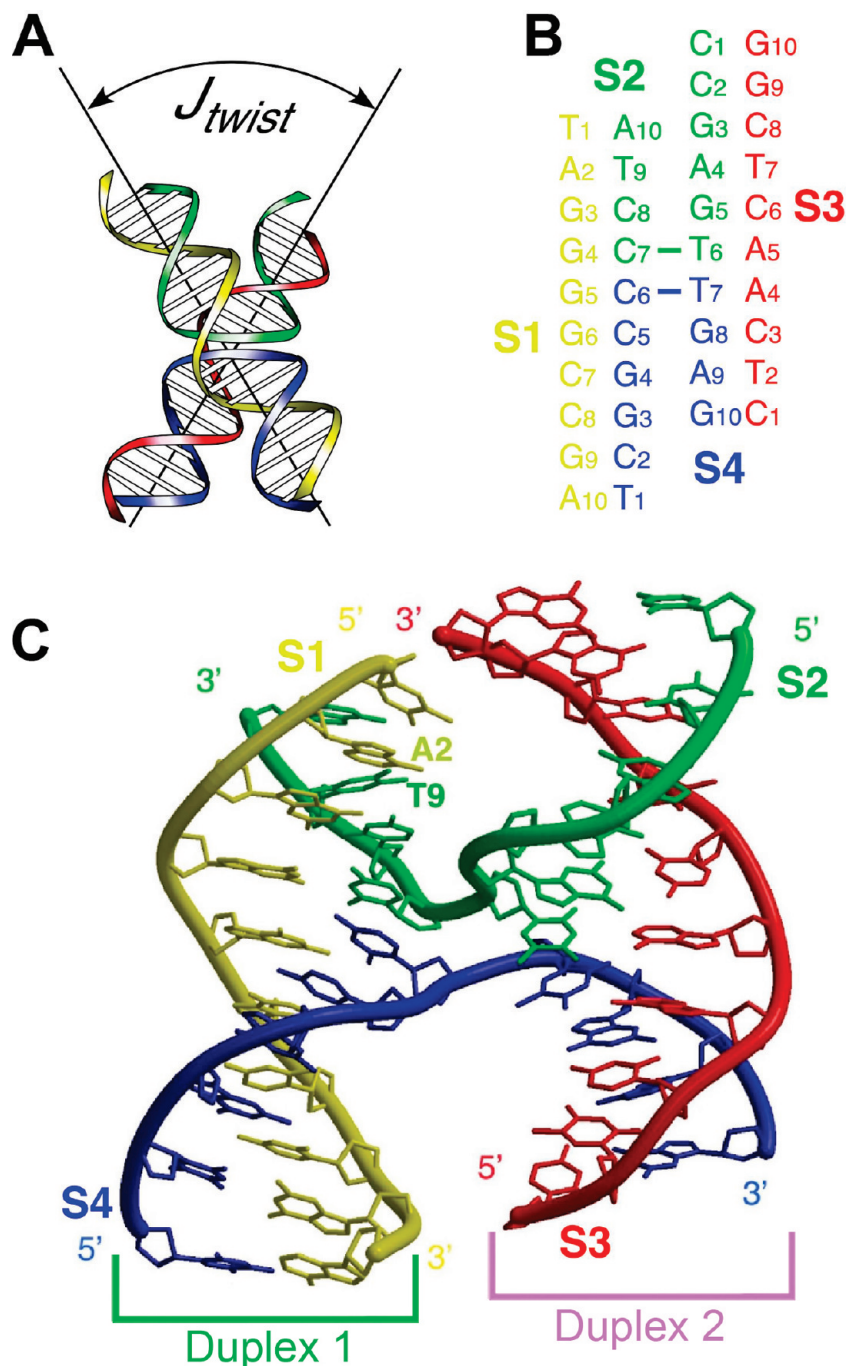


FIGURE 1: Construct of the asymmetric DNA junction. (A) Solution model for the four-stranded DNA Holliday junction in the stacked-X form (8), with the angle relating the helical axes of the stacked duplex arms defined as J_{twist} (11). (B) The four DNA sequences, TAGGGGCCGA (strand S1, yellow), CCGAGTCCTA (strand S2, green), CTCAACTCGG (strand S3, red), and TCGGCCTGAG (strand S4, blue), were annealed to assemble the asymmetric junction. (C) Crystal structure of the asymmetric junction assembled from four sequences, with the phosphodeoxyribose backbone represented as ribbons and colored according to the strands in panel B.

MATERIALS AND METHODS

DNA Synthesis and Crystallization. The four DNA oligonucleotide sequences used to assemble the junction (TAGGGGCCGA, TCGGCCTGAG, CCGAGTCCTA, and CTCAACTCGG) were synthesized by Midland Oligos with the trityl protecting group attached and subsequently purified

by HPLC¹ followed by size exclusion chromatography on a Sephadex G-25 column after detritylation with 3% acetic acid. The DNA sequences were first annealed and then crystallized by the sitting-drop vapor diffusion method from solutions of 0.8 mM DNA and 25 mM sodium cacodylate (pH 7.0) buffer with 100 mM calcium chloride, and equilibrated against a reservoir of 35% aqueous 2-methyl-2,4-dimethylpentanediol. The first phase of growth involved formation of many “crystalline balls” in a droplet, with each “ball” consisting of highly dense, very fine crystal slivers growing from a central point. With the process expedited by agitation of the setup, one to a few of these splinterlike slivers, depending on conditions, grew to become

¹Abbreviations: HPLC, high-performance liquid chromatography; EPMR, molecular replacement by evolutionary search; IDA, interduplex angle for junctions; J_{roll} , junction roll angle; J_{slide} , junction slide translation; J_{twist} , junction twist angle; R_{work} , working crystallographic R-factor; R_{crys} , R-factor for cross validation; kT , product of the Boltzmann constant and absolute temperature.

Table 1: Data Collection and Refinement Statistics

space group	<i>P</i> 1
unit cell constants	$a = 22.344 \text{ \AA}$, $b = 34.554 \text{ \AA}$, $c = 37.631 \text{ \AA}$, $\alpha = 109.921^\circ$, $\beta = 90.028^\circ$, $\gamma = 109.016^\circ$
wavelength (\AA)	1.542
resolution (\AA)	50–1.90
total (unique) no. of reflections	6797 (3048)
R_{sym}^a	6.8 (28.9)
completeness (%) ^a	89.9 (62.9)
$\langle I/\sigma I \rangle^a$	15.7 (2.8)
refinement	
R_{cryst}	22.43
R_{free}^b	26.48
no. of DNA atoms	809
no. of solvent atoms	114

^a The value in parentheses refers to that of the highest-resolution shell.

^b R_{free} is calculated as R_{cryst} for 10% of the reflections not used in refinement (49).

very stable, large diamond-shaped individual crystals with a concomitant loss the surrounding crystal balls as the DNA passed from one crystalline form to another.

Data Collection. X-ray diffraction data were collected from one of the single crystals at liquid nitrogen temperatures using Cu K α radiation from a Rigaku (Tokyo, Japan) RU-H3R rotating anode generator with a RAXIS-IV image plate detector. Diffraction images were processed and reflections merged and scaled using DENZO and SCALEPACK from the HKL suite of programs (22). The crystals were indexed in the *P*1 space group, with the volume of the unit cell consistent with the four unique strands of one Holliday junction defining the asymmetric unit of the crystal (Table 1).

Structure Solution. The structure was determined by molecular replacement using EPMR (23) with a complete four-stranded assembly constructed from the previous symmetric junction d(CCGGTACCGG) as the starting model. A solution with correlation coefficients of 79.8% and an R_{cryst} of 39.8% was obtained with a complete four-stranded junction in the asymmetric unit. The initial model was subsequently refined in CNS (24), applying rigid body refinement, simulated annealing, positional refinement, and addition of solvent molecules. At each stage of refinement, the electron density associated with the DNA was scrutinized to ensure that the asymmetric junction was properly oriented in the unit cell. Several constructs with various thymine bases replaced with 5-bromouracil were used to further confirm that there was only a single orientation for the junction in the asymmetric unit of the unit cell.

The surprising observation of a wobble configuration at the A2·T9 base pair was an initial concern; therefore, the bases of these two base pairs were removed, and the structure was subjected to simulated annealing, followed by several rounds of refinement. When the bases were added back to the backbone according to the residual densities for the bases, they could fit only in the wobble configuration. We then forced the nucleotides into positions that placed their bases in the standard Watson–Crick paired positions, refined the structure, and, again, saw that the bases returned to their wobbled positions. Furthermore, refinement of the DNA with the bases disconnected from the backbone always placed them in the wobble position, even when the starting positions were the standard Watson–Crick paired positions. Finally, a lower-resolution brominated version of the DNA construct exhibited the identical wobble at the same

position. We were, thus, convinced that the data supported a highly wobbled A2·T9 base pair and, at that point, completed the refinement of the structure, resulting in a final model with an R_{cryst} of 22.43% and an R_{free} of 26.48% (Table 1). The values are typical of reliable DNA structures at this resolution.

RESULTS

The DNA junction presented in this study was constructed by assembling four unique DNA sequences [labeled S1–S4 (Figure 1B)], in contrast to the junctions that self-assemble from single inverted-repeat sequences as seen in previously published crystal structures (12–16). The four sequences were designed (1) to fix the junction crossovers at a specific point along the DNA strands through to base pair complementarity and (2) to favor one particular isoform (iso-I) of the stacked-X conformation through base stacking of purines along the outside strands, based on the junction 3 construct studied in solution (18). For this study, however, the construct was designed to have a 4 bp arm stacked over 6 bp arms on either side of the junction crossover, an arrangement that is observed in all the current crystals of symmetric junctions and, therefore, was predicted to facilitate the crystallization of an asymmetric junction in a similar lattice. The asymmetric junction crystallized in a *P*1 space group with one four-stranded junction in the asymmetric unit (Table 1). The lattice, however, is similar to that of the *C*2 forms of the symmetric junctions, even though the asymmetric unit of the latter form can be either all four strands or, more commonly, two strands, with the other two strands generated in the unit cell by crystallographic symmetry. Each DNA duplex arm is coaxially stacked end to end against another duplex arm from an adjacent junction. Furthermore, this lower-symmetry space group indicates that the asymmetric junction is truly asymmetric not only in sequence but also in structure; i.e., the coaxially stacked duplexes on either side of the junction are not equivalent, nor are they conformationally averaged to be pseudosymmetric.

Overall Structure of the Asymmetric DNA Junction. The structure of this asymmetric junction (Figure 1C) is in the antiparallel stacked-X conformation as predicted from solution studies (18) and as seen previously in crystal structures of symmetric junctions (12–16). The arms of the DNA clearly adopt B-type double-helical conformations, as with all previous crystal structures, and are stacked to form nearly continuous helices in the 4 bp-over-6 bp arrangement, as expected. The junction sits in the lattice in a single orientation; there also is a single conformation observed in which strands S1 and S3 wrap around the outside of the nearly continuous helices, and S2 and S4 make sharp U-turns to cross over between duplexes. This is the so-called iso-I form of the junction. One can imagine that this particular conformation could isomerize to place S1 and S3 on the inside and the S2 and S4 strands on the outside of the complex (the alternative iso-II form). This alternative conformational isomer would have generated a 6 bp-over-4 bp stacking arrangement, with the junction crossovers located 2 bp farther down along the duplexes relative to the scheme in Figure 1B. This was clearly not the case; there was no evidence from the electron density maps of any crossover phosphate groups at these alternative positions within the crystal unit cell to suggest that this second isomer was present either partially or as a whole. This alternative conformation could have been accommodated in a 4 bp-over-6 bp stacking if the entire four-stranded complex were rotated 180° along the helices; however, this would have also

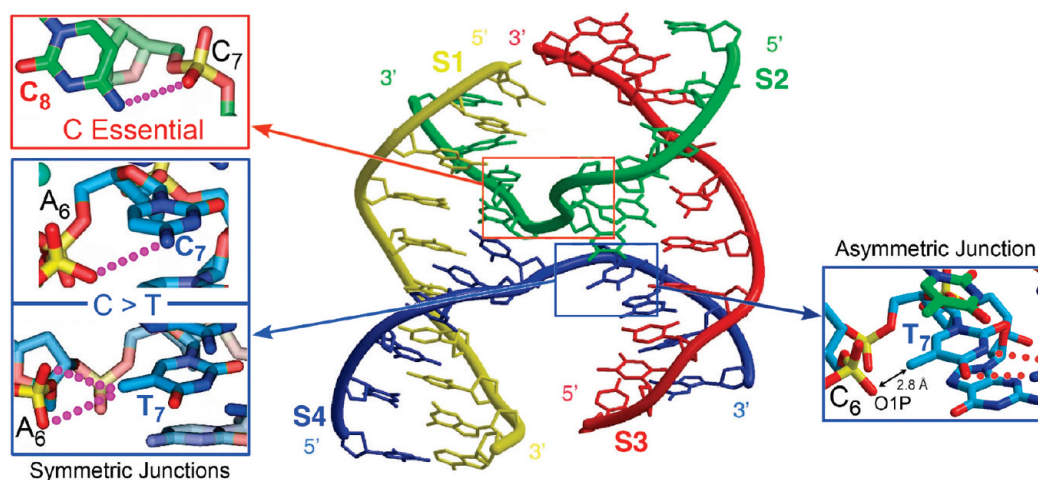


FIGURE 2: Intramolecular interactions at the U-turn of symmetric and asymmetric Holliday junctions. Shown on the left are the stabilizing hydrogen bonds from the cytosine bases to the phosphate oxygens at nucleotide positions C_7 and C_8 of the inverted-repeat sequence motif CCnnn $N_6N_7N_8$ GG. These help to fix the junction crossover in symmetric junctions, preventing the junction from migrating and resolving to separate DNA duplexes. The C_8 interaction is considered to be essential in defining the ACC trinucleotide core at the N_6 , N_7 , and N_8 positions (13, 16). In the asymmetric junction, a short distance interaction from the C5-methyl of thymine T_7 to the phosphate oxygen is observed (right) which is analogous to that observed in the junction formed by the sequence with ATC at the $N_6N_7N_8$ trinucleotide core (25).

rotated the major and minor grooves of the respective helices so that they would face in opposite directions, and there was no evidence to suggest that was the case. Therefore, it is clear that there was only a single stacking conformation observed for this assembly of four sequences. It is interesting, however, that the iso-I conformation seen in the current crystal structure is different from the iso-II conformation that is seen in solution and in single-molecule studies to be favored ($\sim 4:1$) by the junction 3 construct, on which the current asymmetric junction was based. Given the very small energy difference between the two isoforms ($\sim kT$), it is not surprising that the crystal lattice as it was designed could affect the conformational preference seen in solution.

The asymmetric junction is sequence locked, and therefore, its position along the DNA strands is not dependent on any particular set of intramolecular interactions, as in the case of the symmetric junctions. In the symmetric constructs, the junction crossover is stabilized by a unique set of sequence-specific interactions at the N_6 , N_7 , and N_8 positions of the CCnnn $N_6N_7N_8$ GG inverted-repeat sequence motif (Figure 2) (13, 16). In particular, hydrogen bonds observed from the N_4 amino group of cytosines C_7 and C_8 to oxygens of the phosphate groups at the U-turn of the junction specify the ACC trinucleotide as a predominantly junction forming sequence in this motif. The C_7 hydrogen bond can be replaced by an electrostatic interaction with the methyl group of a thymine base (in the junction formed by the ATC trinucleotide variant) or by a bromine halogen bond (in the ABrUC variant) to this phosphate group (25). In the current asymmetric junction, a single short contact was observed between the methyl group of thymine T_7 and an oxygen of the phosphate group linking C_6 to T_7 at the tight U-turn of strand 4 (Figure 2). In symmetric junctions, however, the C_7 interaction cannot be replaced by a T to stabilize the tight U-turn; therefore, we would not consider this to be a particularly strong interaction.

One prediction from the design of this study is that the lack of specific intramolecular interactions in this asymmetric junction would result in a geometry that more closely mirrors the solution model over that of the crystal structures of symmetric junctions. The current asymmetric junction has a J_{twist} of 56.5° (Table 2), an angle that defines a right-handed rotation relating the helical axes

of the two nearly continuous stacked helices across the junction crossover. This nearly is identical to the 60° angle that has been measured for asymmetric junction constructs by anomalous gel migration (18, 26) and FRET studies (27, 28) and significantly different from the more shallow angle of $\sim 40^\circ$ seen in structures of symmetric junctions (12–16). We can, therefore, directly attribute this more shallow J_{twist} of the previously published crystal structures directly to the intramolecular electrostatic interactions in the symmetric junction, particularly the crucial hydrogen bonds that link the phosphates of the junction crossover to the bases of the stacked duplex arms at an essential ACC trinucleotide core (13, 16). In the absence of these sequence-specific interactions, as in asymmetric constructs of the current structure and studied in solution, the junction apparently assumes a more “relaxed” conformation.

Highly Sheared and Wobbled Base Pairs in the B-DNA Arms. For the most part, the base pairs along the arms can be classified as standard Watson–Crick base pairs in B-type helices. The surprising observation was that A2·T9 [across strands S1 and S2 (Figure 3)] and C7·G4 (across strands S1 and S4) base pairs are highly sheared within the base planes, with the purine bases pushed ~ 1.4 Å toward the major groove of the DNA duplex (the average shear for all base pairs = -0.03 Å; standard deviation = 0.62). The C7 and G4 bases could be interpreted as being paired through a set of shared bifurcated hydrogen bonds (29). The A2·T9 base pair, however, is so sheared that it is best interpreted as a “wobbled” base pair, with the imino N_1 atom of A2 hydrogen bonding to O_4 of T9 (< 2.9 Å) rather than to N_3 of the thymine ring (Figure 4). In addition, the N_6 extracyclic amino group of the adenine base is too distant to be paired with O_4 of the thymine. This wobble conformation is further supported by the angles relating the donor–acceptor groups (Figure 4), which are more consistent with a single N_1 – O_4 hydrogen bond than the standard Watson–Crick hydrogen bonding pattern. Finally, the A2 base is tipped 40.5° from the plane of the T9 base, while the thymine remains sandwiched and essentially parallel to the flanking base pairs, resulting in a large propeller twist of the base pair (Figure 5), and supporting the hypothesis that this is not a standard Watson–Crick base pair.

Table 2: Geometry of the Asymmetric Junction Assembled from Four Unique Sequences Compared to Symmetric Junctions Formed by Single Inverted-Repeat Sequences^a

structure	sequence	PDB accession number	IDA (deg)	J_{twist} (deg)	J_{slide} (Å)	J_{roll} (deg)
Asymmetric Junctions						
solution model ^b (18)	various	NA	60	60°	NA	NA
crystal structure	see Figure 1	3IGT	70.4	56.5	0.8	134.5
Symmetric Junctions (see ref 11 for nomenclature)						
ACC (13)	d(CCGGTACCGG)	1DCW	65.9	37.8	0.2	139.9
tACC-4Ca (15)	d(TCGGTACCGA)	1L4J	66.6	39.3	0.3	142.1
tACC-2Sr1 (15)	d(TCGGTACCGA)	1M6G	70.2	44.5	0.6	150.1
tACC-2Ca (15)	d(TCGGTACCGA)	1NQS	67.4	40.6	1.1	146.6
GCC(25)	d(CCGGCGCCGG)	1P4Y	62.0	37.0	-0.04	143.6
CCC(16)	d(CCGGGCCCGG)	1ZF2	67.4	43.1	1.03	147.0
ATC(16)	d(CCGATATCGG)	1ZF4	70.0	43.9	0.90	148.4
average symmetric junctions			67.1	40.9	0.58	145.4
standard deviation			2.8	3.0	0.44	3.6

^a The geometries are defined by interduplex angles relating the ends of the stacked arms (IDA), the angle between the helical axes of the stacked arms (J_{twist}), the displacement of the helical axes from the bisecting plane (J_{slide}), and the rotation of the helical arms from the center of the junction (J_{roll}) (11). The geometries were calculated for the single-crystal DNA-only Holliday junction structures listed with global helix axes defined by CURVES (50). ^b J_{twist} approaches the IDA in junctions with long duplex arms such as those used to study the structure in solution.

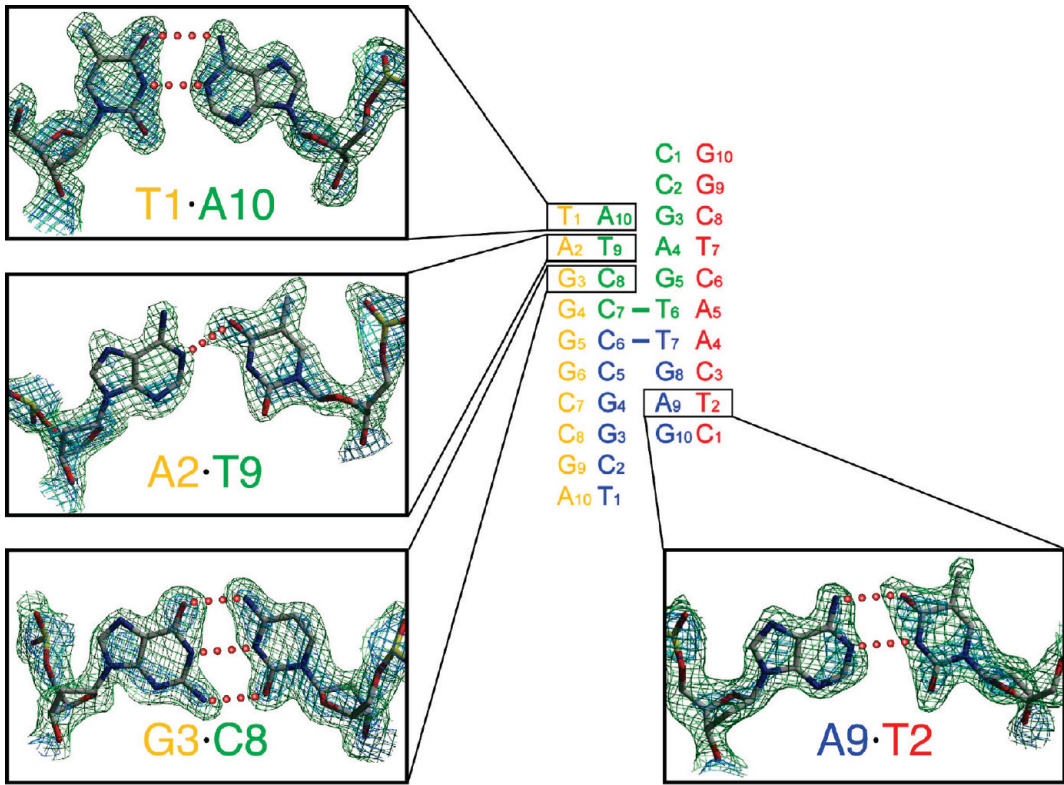


FIGURE 3: Electron density ($2F_o - F_c$) maps for the T1·A10, A2·T9, and G3·C8 base pairs of the S1·S2 arm and the A9·T2 base pair (equivalent to the A2·T9 pair) of the S3·S4 arm of the asymmetric junction. The strands are labeled and colored as in Figure 1. Atoms with the shortest approach distances that, therefore, would be considered to be hydrogen bonded are connected with dots. The map is contoured at 2σ (blue mesh) and 1σ (green mesh) contour levels.

The average B -factors (Figure 6) for the atoms in the bases and in the backbone of each nucleotide along the four chains show that the bases at the wobbled A2·T9 positions are not as well structured as the bases of other nucleotides, particularly compared to those at the junction. This is consistent with this A2·T9 base pair having only a single hydrogen bond and the A2 base being displaced from the stacked bases of the flanking base pairs.

An alternative interpretation may be that the bases are statically disordered and represents an average of two or more conformations of the nucleotides; this could also account for the higher B -factors for the A2·T9 bases. Multiple conformations of bases and deoxyribose backbones have been seen in ultra-high-resolution structures of DNA duplexes (30), and our structure at 1.9 Å resolution would not be able to resolve these alternative

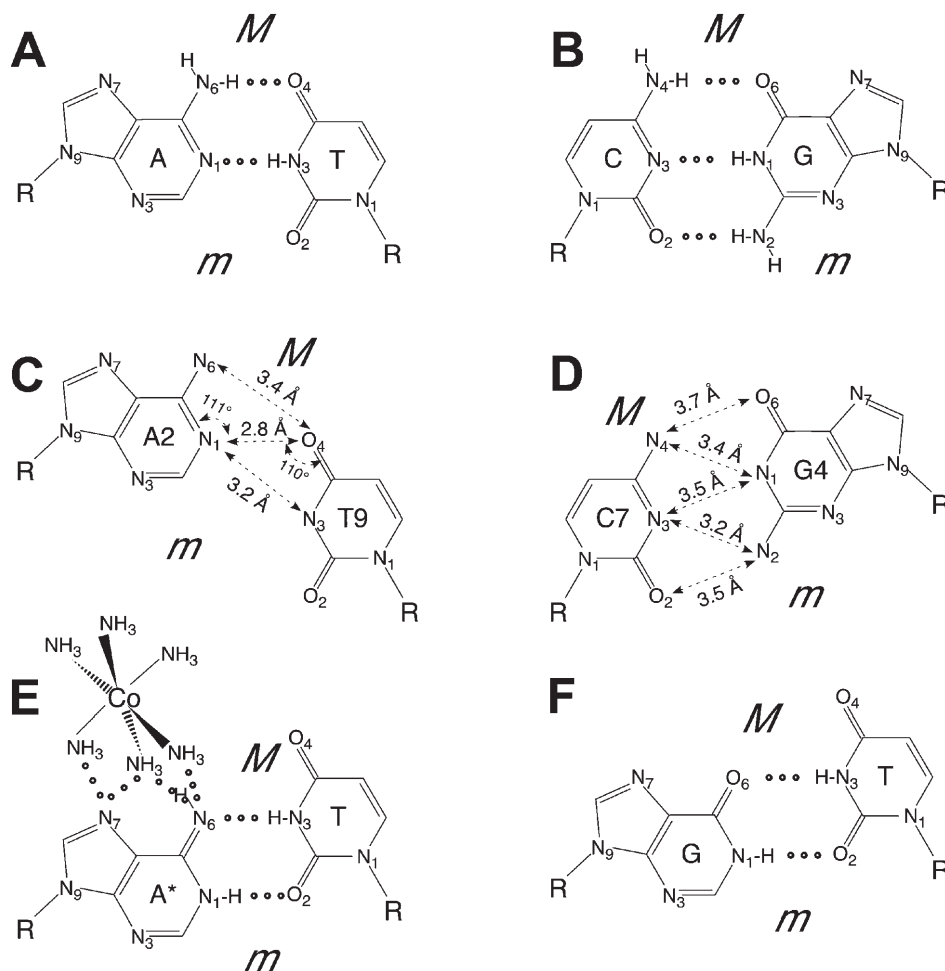


FIGURE 4: Structure of Watson–Crick, wobbled rare tautomers and wobbled mismatched base pairs. Hydrogen-bonded atoms are connected with dots, and the major and minor grooves are labeled M and m, respectively. A·T (A) and C·G (B) base pairs as formed by the bases in their stable tautomer forms (amino forms of A and C and keto forms of T and G). Geometries of A2·T9 (C) and C7·G4 (D) base pairs observed in the B-DNA arms of an asymmetric DNA Holliday junction. The distances and angle relating potential hydrogen bond donor and acceptor groups are labeled for each. (E) Wobble base pair observed in the cobalt(III) hexamine-induced imino tautomer of adenine (A*) with T (32). (F) Structure of a wobble G·T mismatched base pair (33, 34).

positions. However, positional disorder along the base planes that would result in the large shear displacements of the bases would also be reflected in correlated disorder and, consequently, higher *B*-factors for at least the C1' atom of the associated backbone atoms, which is not observed. The backbone atoms of the A2·T9 base pair are as well structured as those of other nucleotides in the structure, indicating that the conformation and positions of the phosphodeoxyribose atoms are very well-defined. The disorder in the bases is, therefore, most likely rotational disorder around the glycolytic bond of the out-of-plane adenine A2 base. Such rotational disorder could not reposition the hydrogen bond donor and acceptor groups of the two bases into the proper geometry to effect Watson–Crick pairing but would leave them sheared and in a wobbled configuration. Thus, the trajectory from the corresponding deoxyribose sugars constrains the positions of the A2·T9 nucleobases to only their wobbled positions and, thus, further supports the interpretation that the A2·T9 base pair is indeed highly wobbled.

The A2·T9 wobbled base pair cannot be ascribed to the standard tautomer forms of the adenine and thymine bases but could be with either an imino adenine or an enol thymine. The current structure in itself cannot distinguish between the two possibilities, but the direction of the base shearing may. Wobbled

base pairs have previously been observed to be induced by Hg(II) in an isolated A·U base pair (31) or by cobalt hexamine at the terminal A·T base pair in a Z-DNA crystal (32), and both are similar to that of G·T mismatches (33, 34) with the pyrimidine sheared toward the major groove (Figure 4). In these metal-induced wobbles, the adenine was assumed to adopt the imino tautomeric form, since it would still allow two hydrogen bonds to form between the A and T bases, as with the canonical A·T base pair. In addition, *ab initio* calculations on isolated bases suggest that the imino tautomer of adenine is as much as 14 kcal/mol more stable than an enol thymine (36). The wobbled A2·T9 base pair in the current structure shears the adenine toward the major groove and has a single N1...O4 hydrogen bond. If this were an imino adenine tautomer, we would have expected the base pair to wobble in the opposite direction, with the thymine pushed toward the major groove to allow for the formation of two inter-strand hydrogen bonds. The direction of this wobbled A·T base pair would, therefore, suggest that, in contrast to the metal-induced tautomers, the thymine adopts the rare enol form. An unusual tautomer form of an extrahelical thymine base has been invoked previously to explain a T·T pairing between the loops of the hairpins in the structure of d(CGCGGTTTTCGCGCG) (35); thus, thymine tautomers are possible and can be induced to facilitate base–base pairing. The current

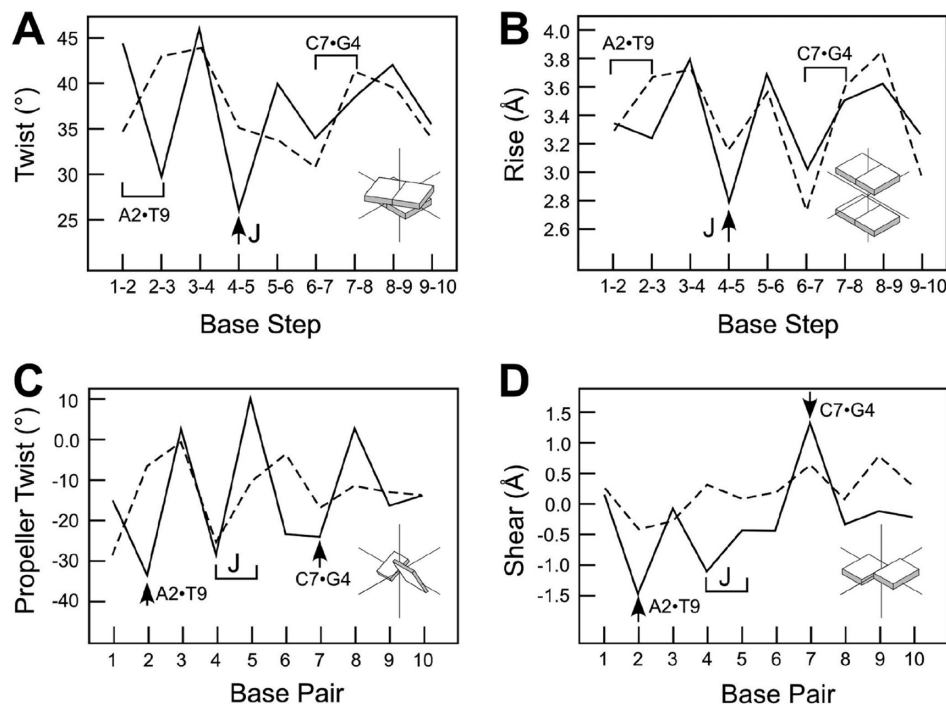


FIGURE 5: Helical parameters for dinucleotides and base pairs within the asymmetric junction. The helical rise (A) and twist (B) between dinucleotides and propeller twist (C) and shear for base pairs (D) were calculated using 3DNA (57). The semicontinuous double helices (duplex 1 and duplex 2) of the junction are as labeled in Figure 1. The location of the highly sheared A2·T9 and C7·G4 base pairs in duplex 1 and the junction crossover are as indicated.

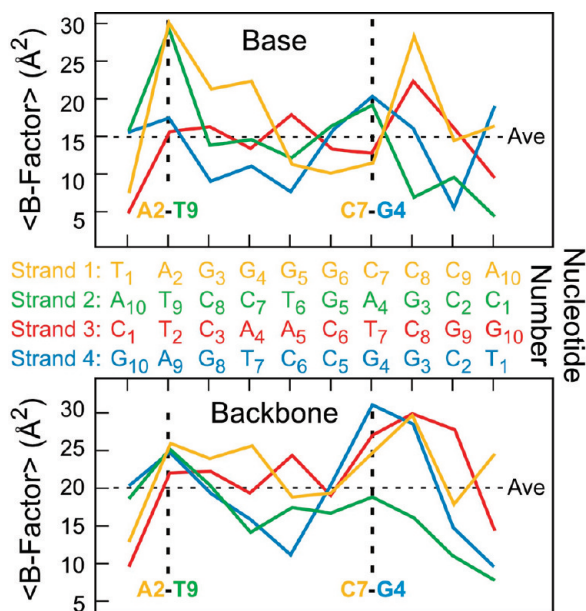


FIGURE 6: *B*-Factors averaged across the base (A) and phosphodeoxyribose (B) atoms for each nucleotide along the chains of the asymmetric junction formed by assembly of the sequences TAGGGGCCGA (yellow), CCGAGTCCTA (green), CTC AAC-TCGG (red), and TCGGCCTGAG (blue), colored according to strands in Figure 1.

structure, however, is the first, to the best of our knowledge, of such a tautomer in a B-type DNA double helix.

DISCUSSION

The first conclusion drawn from the single-crystal structure of an asymmetric junction presented here is that the original solution model (18) was remarkably accurate in describing the

geometry of the complex. The current four-stranded complex assembled from four unique sequences recapitulates the geometry of the solution model in terms of the antiparallel orientations of the strands, the stacking of the duplex arms to generate a sequence-specific conformation, and the relative rotation of the two semicontinuous double helices formed by the stacked duplex arms. This structure, therefore, serves to bridge the solution properties with the crystal structures of DNA junctions, indicating that the atomic details are compatible with and help to define the general sequence-specific behavior of junctions regardless of the environment.

In comparison, previously published crystal structures of symmetric junctions are more compact (reflected in the much shallower J_{twist} of 40°) compared to the structure of asymmetric junctions in the solution state and in crystals. We can now directly attribute this difference in geometry between the two classes of structures to the sequence-specific intramolecular interactions at the core trinucleotide of the junction crossover that is seen only in the symmetric junctions. Specifically, a pair of stabilizing cytosine to phosphate hydrogen bonds defines the sequence preference for this trinucleotide core as NCC (where N = A > G > C) within an inverted-repeat motif (Figure 2). The specific inverted-repeat sequence, therefore, is shown to define not only the stability but also the detailed geometry of the four-stranded junction. Thus, the sequence specificity of junction-resolving enzymes that recognize and bind to the Holliday junction in the stacked-X form, such as T4 endonuclease VII (37) and T7 endonuclease I (38), can be linked to the geometry as defined by these unique sequence-dependent molecular interactions (39).

The other interesting conclusion from this structure is that that highly sheared base pairs, particularly an A·T base pair induced to a wobble configuration, are stabilized by the sequences in this junction. We propose that the A2·T9 wobble base pair in this structure can be attributed to sequence effects that stabilize a rare

enol tautomer of the thymine base or, less likely, the imino tautomer of an adenine along a B-DNA duplex. The A·T wobble sits at the end of four contiguous C·G base pairs (CCCC/GGGG), and this is the only crystal structure of a B-type helix for this sequence motif. The observation that the C7·G4 base pair at the opposite end of this stretch is also highly sheared indicates that both ends of the polyC·polyG B helix induce similar distortions to the flanking base pairs.

Contiguous C·G base pairs are particularly prone to bisulfite deamination (40), suggesting that the sequence motif has unique structural properties. Sequences with long contiguous stretches of C·G base pairs tend to adopt the alternative A-conformation (16, 41–43). We had previously observed that the sequence d(CCGGGCCCGG) will crystallize either as an A-form duplex or as a junction, but in the four-stranded complex, the arms all adopt the B-type conformation (16). The junction, therefore, appears to allow sequences that would normally form A-DNA in double helices to be studied as B-form double helices, similar to the conformation of the DNA in protein complexes.

The wobbled A2·T9 base pair, and the presumed tautomerization of the thymine, is not directly induced by the structure of the Holliday junction. The A2·T9 base pair is two steps from the junction crossover, where one would expect the structure to be most perturbed. In addition, there is a T·A base pair at the equivalent position of the opposite 4 bp arm of the current asymmetric junction, and it forms a canonical Watson–Crick base pair (Figure 3). Furthermore, the A2·T9 position is not in contact with the opposing S3·S4 arm of the junction. Finally, no such wobbled A·T or C·G base pair has been reported in any previously published crystal structure of symmetric junctions. Thus, the DNA junction itself is not the cause of the wobbled base pairs but serves as a vehicle by which this stretch of contiguous C·G base pairs, which would normally crystallize as an A-type duplex, can be studied in the standard B-form; DNA junctions apparently do not accommodate A-type helices, even for G·C base pair-rich sequences that strongly favor A-DNA duplexes (16).

This model suggests that an A·T base pair abutted to the ends of contiguous C·G sequences as B-DNA would be highly wobbled as a result of one base (most likely the thymine) adopting a rare tautomer form. Such a tautomerization would be highly mutagenic (Figure 7) because the replication machinery would recognize the associated wobble as a mismatched base pair. Either an enol thymine or an imino adenine, upon replication, would result in a transition mutation from an A·T to G·C base pair, one of the more common mutations observed in nature. It has been shown that A·C base pair misincorporation is dependent on the sequence of the flanking nucleotides, with misincorporation being 1 order of magnitude higher with a polyG as compared to a poly(A/T) template or primer (44, 45), while the rate of misincorporation of a T for a C has been shown to be highly sequence dependent (46). These dramatic sequence context-dependent rates of misincorporation have previously been attributed to base stacking effects at the DNA ends but may in fact arise from sequence effects on tautomer preferences of the incoming or template nucleotides. The crystal structure of the high-fidelity DNA polymerase from *Bacillus stearothermophilus* shows a rare protonated tautomer of cytosine paired to a mutagenic O6-methylguanine base (47), indicating that rare tautomers can be accommodated in a replication complex. Our results demonstrate that the DNA sequence could itself

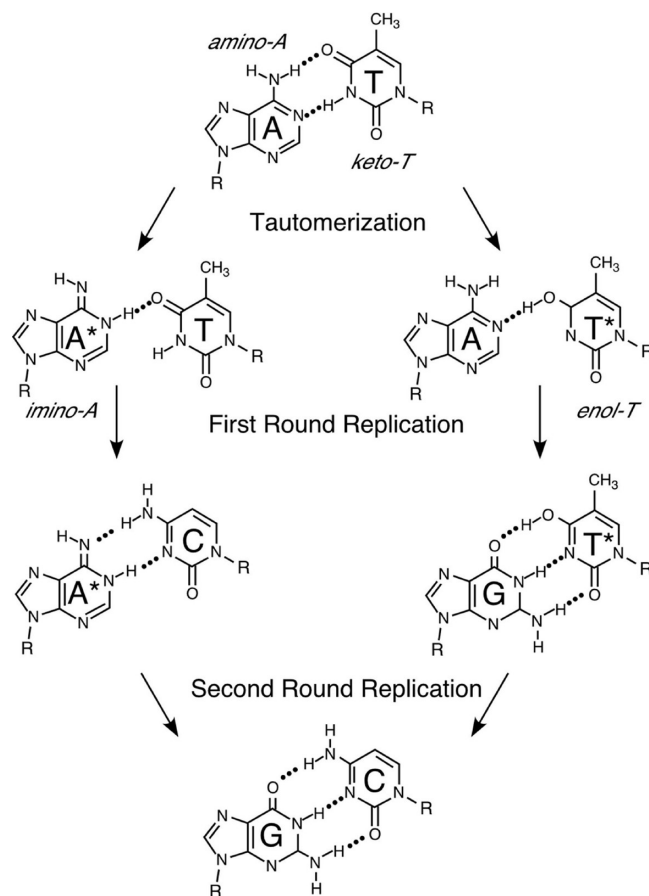


FIGURE 7: Potential role of "rare" tautomers in A·T to G·C base pair transition mutations. A normal Watson–Crick A·T base pair, with the adenine in its normal amino form and thymine in the keto form, would become "wobbled" if the adenine were to adopt the rare imino tautomer (A*) or the thymine adopt the rare enol tautomer (T*). Such tautomers would be recognized as a base pair mismatch, which, after the first round of replication, would result in either an A*·C or G·T* base pair that would appear "normal" in terms of the hydrogen bonding pattern and the displacement of bases toward the major and minor grooves. In the second round, the transition mutation would be fixed, with either the A*·C or G·T* base pair resulting in a G·C base pair in one set of daughter DNA duplexes.

stabilize a rare nucleotide tautomer, lending credence to the original rare tautomer hypothesis (21, 48), in which genomic mutations could result from polymerases misincorporating nucleotides that are misread because of the uncommon hydrogen bonding patterns associated with these base tautomers. In this case, we suggest that poly-G-poly-C sequences in the context of a B-DNA duplex, as would be seen in replication complexes, help stabilize either an enol thymine or imino adenine tautomer, which could contribute to the relatively common occurrence of transition mutations from A·T to G·C base pairs in this sequence context.

ACKNOWLEDGMENT

We thank Prof. P. Andrew Karplus for helpful discussion during this work.

REFERENCES

- Holliday, R. (1964) A mechanism for gene conversion in fungi. *Genet. Res.* 5, 282–304.
- Nunes-Duby, S. E., Matsumoto, L., and Landy, A. (1987) Site-specific recombination intermediates trapped with suicide substrates. *Cell* 50, 779–788.

3. 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.
4. Haber, J. E., and Heyer, W. D. (2001) The fuss about Mus81. *Cell* 107, 551–554.
5. Dickman, M. J., Ingleston, S. M., Sedelnikova, S. E., Rafferty, J. B., Lloyd, R. G., Grasby, J. A., and Hornby, D. P. (2002) The RuvABC resolvosome. *Eur. J. Biochem.* 269, 5492–5501.
6. Declais, A. C., Fogg, J. M., Freeman, A. D., Coste, F., Hadden, J. M., Phillips, S. E., and Lilley, D. M. (2003) The complex between a four-way DNA junction and T7 endonuclease I. *EMBO J.* 22, 1398–1409.
7. Subramaniam, S., Tewari, A. K., Nunes-Duby, S. E., and Foster, M. P. (2003) Dynamics and DNA substrate recognition by the catalytic domain of lambda integrase. *J. Mol. Biol.* 329, 423–439.
8. Lilley, D. M. J. (1999) Structures and interactions of helical junctions in nucleic acids. In *Oxford Handbook of Nucleic Acid Structure* (Neidle, S., Ed.) pp 471–498, Oxford University Press, New York.
9. McKinney, S. A., Declais, A. C., Lilley, D. M., and Ha, T. (2003) Structural dynamics of individual Holliday junctions. *Nat. Struct. Biol.* 10, 93–97.
10. McKinney, S. A., Freeman, A. D., Lilley, D. M., and Ha, T. (2005) Observing spontaneous branch migration of Holliday junctions one step at a time. *Proc. Natl. Acad. Sci. U.S.A.* 102, 5715–5720.
11. Watson, J., Hays, F. A., and Ho, P. S. (2004) Definitions and analysis of DNA Holliday junction geometry. *Nucleic Acids Res.* 32, 3017–3027.
12. 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.
13. 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. U.S.A.* 97, 3971–3976.
14. Thorpe, J. H., Hobbs, J. R., Todd, A. K., Denny, W. A., Charlton, P., and Cardin, C. J. (2000) Guanine specific binding at a DNA junction formed by d[CG(5-BrU)ACG]₂ with a topoisomerase poison in the presence of Co²⁺ ions. *Biochemistry* 39, 15055–15061.
15. Thorpe, J. H., Gale, B. C., Teixeira, S. C., and Cardin, C. J. (2003) Conformational and Hydration Effects of Site-selective Sodium, Calcium and Strontium Ion Binding to the DNA Holliday Junction Structure d(TCGGTACCGA)₄. *J. Mol. Biol.* 327, 97–109.
16. Hays, F. A., Teegarden, A., Jones, Z. J., Harms, M., Raup, D., Watson, J., Cavaliere, E., and Ho, P. S. (2005) How sequence defines structure: A crystallographic map of DNA structure and conformation. *Proc. Natl. Acad. Sci. U.S.A.* 102, 7157–7162.
17. Hays, F. A., Schirf, V., Ho, P. S., and Demeler, B. (2006) Solution formation of Holliday junctions in inverted-repeat DNA sequences. *Biochemistry* 45, 2467–2471.
18. 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.
19. Hays, F. A., Watson, J., and Ho, P. S. (2003) Caution! DNA crossing: Crystal structures of Holliday junctions. *J. Biol. Chem.* 278, 49663–49666.
20. Watson, J. D., and Crick, F. H. (1953) Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature* 171, 737–738.
21. Watson, J. D., and Crick, F. H. (1953) Genetical implications of the structure of deoxyribonucleic acid. *Nature* 171, 964–967.
22. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276, 307–326.
23. Kissinger, C. R., Gehlhaar, D. K., and Fogel, D. B. (1999) Rapid automated molecular replacement by evolutionary search. *Acta Crystallogr. D* 55 (Part 2), 484–491.
24. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* 54 (Part 5), 905–921.
25. Hays, F. A., Vargason, J. M., and Ho, P. S. (2003) Effect of sequence on the conformation of DNA holliday junctions. *Biochemistry* 42, 9586–9597.
26. 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. U.S.A.* 85, 4653–4656.
27. 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.
28. 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. *Bioophys. J.* 66, 99–109.
29. Egli, M., Usman, N., Zhang, S. G., and Rich, A. (1992) Crystal structure of an Okazaki fragment at 2-Å resolution. *Proc. Natl. Acad. Sci. U.S.A.* 89, 534–538.
30. 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.
31. Zamora, F., Kunsman, M., Sabat, M., and Lippert, B. (1997) Metal-stabilized rare tautomers of nucleobases. 6-Imino tautomer of adenine in a mixed-nucleobase complex with mercury(II). *Inorg. Chem.* 36, 1583–1587.
32. Thiagarajan, S., Rajan, S. S., and Gautham, N. (2004) Cobalt hexamine induced tautomeric shift in Z-DNA: The structure of d(CGCGCA)*d(TGCGCG) in two crystal forms. *Nucleic Acids Res.* 32, 5945–5953.
33. Ho, P. S., Frederick, C. A., Quigley, G. J., van der Marel, G. A., van Boom, J. H., Wang, A. H., and Rich, A. (1985) G·T wobble base-pairing in Z-DNA at 1.0 Å atomic resolution: The crystal structure of d(CGCGTG). *EMBO J.* 4, 3617–3623.
34. Hunter, W. N., Brown, T., Kneale, G., Anand, N. N., Rabinovich, D., and Kennard, O. (1987) The structure of guanosine-thymidine mismatches in B-DNA at 2.5-Å resolution. *J. Biol. Chem.* 262, 9962–9970.
35. Chattopadhyaya, R., Grzeskowiak, K., and Dickerson, R. E. (1990) Structure of a T4 hairpin loop on a Z-DNA stem and comparison with A-RNA and B-DNA loops. *J. Mol. Biol.* 211, 189–210.
36. Kryachko, E., and Sabin, J. R. (2003) Quantum chemical study of the hydrogen-bonded patterns in A·T base pair of DNA: Origins of tautomeric mispairs, base flipping, and Watson-Crick to Hoogsteen conversion. *Int. J. Quantum Chem.* 91, 695–710.
37. Biertumpfel, C., Yang, W., and Suck, D. (2007) Crystal structure of T4 endonuclease VII resolving a Holliday junction. *Nature* 449, 616–620.
38. Hadden, J. M., Declais, A. C., Carr, S. B., Lilley, D. M., and Phillips, S. E. (2007) The structural basis of Holliday junction resolution by T7 endonuclease I. *Nature* 449, 621–624.
39. Khuu, P. A., Voth, A. R., Hays, F. A., and Ho, P. S. (2006) The stacked-X DNA Holliday junction and protein recognition. *J. Mol. Recognit.* 19, 1–9.
40. Tsai, A. G., Engelhart, A. E., Hatmal, M. M., Houston, S. I., Hud, N. V., Haworth, I. S., and Lieber, M. R. (2009) Conformational variants of duplex DNA correlated with cytosine-rich chromosomal fragile sites. *J. Biol. Chem.* 284, 7157–7164.
41. McCall, M., Brown, T., and Kennard, O. (1985) The crystal structure of d(G-G-G-G-C-C-C-C). A model for poly(dG)·poly(dC). *J. Mol. Biol.* 183, 385–396.
42. Gao, Y. G., Robinson, H., and Wang, A. H. (1999) High-resolution A-DNA crystal structures of d(AGGGGCCCT). An A-DNA model of poly(dG) × poly(dC). *Eur. J. Biochem.* 261, 413–420.
43. Basham, B., Schroth, G. P., and Ho, P. S. (1995) An A-DNA triplet code: Thermodynamic rules for predicting A- and B-DNA. *Proc. Natl. Acad. Sci. U.S.A.* 92, 6464–6468.
44. Topal, M. D., DiGiuseppi, S. R., and Sinha, N. K. (1980) Molecular basis for substitution mutations. Effect of primer terminal and template residues on nucleotide selection by phage T4 DNA polymerase in vitro. *J. Biol. Chem.* 255, 11717–11724.
45. Watanabe, S. M., and Goodman, M. F. (1981) On the molecular basis of transition mutations: Frequencies of forming 2-aminopurine·cytosine and adenine·cytosine base mispairs in vitro. *Proc. Natl. Acad. Sci. U.S.A.* 78, 2864–2868.
46. Singer, B., Chavez, F., Goodman, M. F., Essigmann, J. M., and Dosanjh, M. K. (1989) Effect of 3' flanking neighbors on kinetics of pairing of dCTP or dTTP opposite O6-methylguanine in a defined primed oligonucleotide when *Escherichia coli* DNA polymerase I is used. *Proc. Natl. Acad. Sci. U.S.A.* 86, 8271–8274.
47. Warren, J. J., Forsberg, L. J., and Beese, L. S. (2006) The structural basis for the mutagenicity of O(6)-methyl-guanine lesions. *Proc. Natl. Acad. Sci. U.S.A.* 103, 19701–19706.
48. Topal, M. D., and Fresco, J. R. (1976) Complementary base pairing and the origin of substitution mutations. *Nature* 263, 285–289.
49. Brunger, A. T. (1992) Free R value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355, 472–475.
50. Lavery, R., and Sklenar, H. (1989) Defining the structure of irregular nucleic acids: Conventions and principles. *J. Biomol. Struct. Dyn.* 6, 655–667.
51. Lu, X. J., and Olson, W. K. (2003) 3DNA: A software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* 31, 5108–5121.