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

2 AUTHORS, INCLUDING:

Pui Shing Ho

Colorado State University

94 PUBLICATIONS   **4,317** CITATIONS

A rare nucleotide base tautomer in the structure of an asymmetric DNA junction

Patricia Khuu[1] and **P. Shing Ho**[1,2,*]

[1]Department of Biochemistry and Biophysics, ALS 2011, Oregon State University, Corvallis, OR 97331

[2]Department of Biochemistry and Molecular Biology, 1870 Campus Delivery, Colorado State University, Fort Collins, CO 80523

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 semi-continuous duplexes interrupted only by the crossing of the junction. These semi-continuous 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 differ from the more shallow ~40° for previous crystal structures of symmetric junctions that self-assemble from single identical inverted-repeat sequences. This supports the model that the unique set of intramolecular interactions at the trinucleotide core of the crossing strands, which are not present in the current asymmetric junction, affect 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 to study 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 involve symmetric sequences, asymmetric junctions can be assembled from four unique sequences to lock the position of DNA cross-over 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 near continuous duplexes that are related by an a 60° angle (the J_{twist} (11)) across the junction cross-over (Fig. 1A).

A number of single-crystal structures have now been reported of symmetric junctions that self-assembled from single inverted-repeat or near repeat sequences (12-16). Such symmetric junction are stabilized not by base complementarity, but by a set of sequence dependent

*To whom correspondences should be addressed Phone: 970-491-0569 FAX: 970-491-0494 E-Mail: Shing.Ho@ColoState.edu.

intramolecular interactions that lock the tight U-turn of the cross-over 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 the geometry is associated with the differences in the environment of the junction (solution *versus* crystal) or in the DNA constructs (asymmetric *versus* symmetric). The current study reports the 1.9 Å structure of an asymmetric junction constructed from four unique sequences (Fig. 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 the current 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 to create 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 guanines (G) and cytosine (C) bases (20). These hydrogen bonding patterns require that the A and C bases adopt the stable amino- while T and G adopt the keto-tautomer forms 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 to induce 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 adopting 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.

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 HPLC 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. Expedited by agitation of the setup, one to a few of these splinter-like slivers, depending on conditions, grew to become very stable, large diamond-shaped individual crystals with a concomitant loss the surrounding crystal balls as the DNA transferred 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 CuKα 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 *P1* 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 solved 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 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 insure that the asymmetric junction was properly oriented in the unit cell. Several constructs with various thymine bases replaced by 5-bromouracils 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 A2·T9 base pair was an initial concern and, therefore, the bases of these two base pairs were removed and the structure 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 only fit in the wobble configuration. We then forced the nucleotides in 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 showed 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 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 to S4, Fig. 1B), in contrast to the junctions that self-assemble from single inverted-repeat sequences as seen in previous crystal structures (12-16). The four sequences were designed to 1) fix the junction cross-overs at a specific point along the DNA strands through to base pair complementarity and 2) 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 four base pair arm stacked over a six base pair arms on either side of the junction cross-over, 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 *P1* space group with one four-stranded junction in the asymmetric unit (Table 1). The lattice, however, is similar to that of the *C2* 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 (Fig. 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 near continuous helices in the four-over-six base pair arrangement, as expected. The junction not only sits in the lattice in a single orientation, but there is a single conformation observed in which strands S1 and S3 wrap around the outside of the near 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 six-over-four stacking arrangement, with the junction cross-overs located two base pairs further down along the duplexes relative to the scheme in Fig. 1B. This was clearly not the case—there were no evidence from the electron density maps for 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 four-over-six stacking if the entire four-stranded complex were rotated 180° along the helices; however, this would have also 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 current crystal structure is different from the *iso*-II conformation that is seen in solution and in single molecule studies to be favored (~4:1) by the Junction 3 construct, from which the current asymmetric junction was based. Given the very small energy difference between the two isoforms (~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 is the case of the symmetric junctions. In the symmetric constructs, the junction cross-over is stabilized by a unique set of sequence-specific interactions at the $N_6N_7N_8$ positions of the CCnnn$N_6N_7N_8$GG inverted-repeat sequence motif (Fig. 2) (13,16). In particular, hydrogen bonds observed from N4 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 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 to an oxygen of the phosphate group linking C_6 to T_7 at the tight U-turn of strand 4 (Fig. 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 $J_{twist} = 56.5°$ (Table 2), an angle that defines a right-handed rotation relating the helical axes of the two near continuous stacked helices across the junction cross-over. 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 ~40° angle seen in structures of symmetric junctions (12-16). We can, therefore, directly attribute this more shallow J_{twist} of the previous crystal structures directly to the intramolecular electrostatic interactions in the symmetric junction, particularly the crucial hydrogen bonds that link the phosphates of the junction cross-over 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 helixes. The surprising observation was that the A2·T9 (across strands S1 to S2, Fig. 3) and C7·G4 base pairs (across strands S1 and S4) are highly sheared within the base planes, with the purine bases pushed ~1.4 Å towards the major groove of the DNA duplex (the average shear for all base pairs = -0.03 Å, SD = 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 N_1 imino nitrogen of A2 hydrogen bonding to the O_4 oxygen of T9 (< 2.9 Å) rather than to the N_3 nitrogen of the thymine ring (Fig. 4). In addition, the N_6 extracyclic amino group of the adenine base is too distant to be paired with the O_4 oxygen of the thymine. This wobble conformation is further supported by the angles relating the donor-acceptor groups (Fig. 4), which are more consistent with a single N_1 to O_4 hydrogen bond than the standard Watson-Crick hydrogen bonding pattern. Finally, the A2 adenine base is tipped 40.5° from the plane of the T9 thymine base, while the thymine remains sandwiched and essentially parallel to the flanking base pairs, resulting in a large propeller twist of the base pair (Fig. 5), and supporting the hypothesis that this is not a standard Watson-Crick base pair.

The average B-factors (Fig. 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 the current structure at 1.9 Å resolution would not be able to resolve these alternative 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' carbon of the associated backbone atoms, which is not observed. The backbone atoms of A2·T9 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 bonds of the nucleotides, most likely rotational disorder 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 accounted for by the standard tautomer forms of the adenine and thymine bases, but could be with either an imino-adenine or 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 towards the major groove (Fig. 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 also 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 towards 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 wobbled in the opposite direction, with the thymine pushed towards the major groove to allow for the formation of two interstrand hydrogen bonds. The direction of this wobbled A·T 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(CGCGCGTTTTCGCGCG) (35); thus, thymine tautomers are possible and can be induced in order to facilitate base-base pairing. The current structure, however, is this first, to 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 semi-continuous 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 helps define the general sequence specific behavior of junctions regardless of the environment.

In comparison, previous 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 cross-over 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 (Fig 2). The specific inverted-repeat sequence, therefore, is shown to not only define the stability, but now 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 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 stabilizes 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·Gs 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 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 away from the junction cross-over, 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 four base pair arm of the current asymmetric junction, and it forms a canonical Watson-Crick base pair (Fig. 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 previous 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-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 (Fig. 7) because the replication machinery would recognize the associated wobble as a mismatched base pair. Either an enol-thymine or 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 misincorporation is dependent on the sequence of the flanking nucleotides, with misincorporation being one-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 the template nucleotides. The crystal structure of the high-fidelity DNA polymerase from *B. 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 stabilize a rare nucleotide tautomer, lending credence to the original "rare tautomer hypothesis" (21,48), that 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 polyG-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.

Acknowledgements

Abbreviations

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; mM, millimolar; R_{crys}, working crystallographic R-factor; R_{crys}, R-factor for cross-validation; kT, product of the Boltmann constant and absolute temperature.

References

1. Holliday R. A mechanism for gene conversion in fungi. Genet. Res 1964;5:282–304.

2. Nunes-Duby SE, Matsumoto L, Landy A. Site-specific recombination intermediates trapped with suicide substrates. Cell 1987;50:779–788. [PubMed: 3040260]

3. Cox MM, Goodman MF, Kreuzer KN, Sherratt DJ, Sandler SJ, Marians KJ. The importance of repairing stalled replication forks. Nature (London) 2000;404:37–41. [PubMed: 10716434]

4. Haber JE, Heyer WD. The fuss about Mus81. Cell 2001;107:551–554. [PubMed: 11733053]

5. Dickman MJ, Ingleston SM, Sedelnikova SE, Rafferty JB, Lloyd RG, Grasby JA, Hornby DP. The RuvABC resolvasome. Eur. J. Biochem 2002;269:5492–5501. [PubMed: 12423347]

6. Declais AC, Fogg JM, Freeman AD, Coste F, Hadden JM, Phillips SE, Lilley DM. The complex between a four-way DNA junction and T7 endonuclease I. EMBO J 2003;22:1398–1409. [PubMed: 12628932]

7. Subramaniam S, Tewari AK, Nunes-Duby SE, Foster MP. Dynamics and DNA substrate recognition by the catalytic domain of lambda integrase. J. Mol. Biol 2003;329:423–439. [PubMed: 12767827]

8. Lilley, DMJ. Structures and interactions of helical junctions in nucleic acids. In: Neidle, S., editor. Oxford Handbook of Nucleic Acid Structure. Oxford University Press; New York: 1999. p. 471-498.

9. McKinney SA, Declais AC, Lilley DM, Ha T. Structural dynamics of individual Holliday junctions. Nat. Struct. Biol 2003;10:93–97. [PubMed: 12496933]

10. McKinney SA, Freeman AD, Lilley DM, Ha T. Observing spontaneous branch migration of Holliday junctions one step at a time. Proc. Natl. Acad. Sci., USA 2005;102:5715–5720. [PubMed: 15824311]

11. Watson J, Hays FA, Ho PS. Definitions and analysis of DNA Holliday junction geometry. Nucleic Acids Res 2004;32:3017–3027. [PubMed: 15173384]

12. Ortiz-Lombardía M, González A, Eritja R, Aymamí J, Azorín F, Coll M. Crystal structure of a DNA Holliday junction. Nat. Struct. Biol 1999;6:913–917. [PubMed: 10504723]

13. Eichman BF, Vargason JM, Mooers BHM, Ho PS. The Holliday junction in an inverted repeat sequence: sequence effects on the structure of four-way junctions. Proc. Natl. Acad. Sci. USA 2000;97:3971–3976. [PubMed: 10760268]

14. Thorpe JH, Hobbs JR, Todd AK, Denny WA, Charlton P, Cardin CJ. Guanine specific binding at a DNA junction formed by d[CG(5-BrU)ACG](2) with a topoisomerase poison in the presence of Co (2+) ions. Biochemistry 2000;39:15055–15061. [PubMed: 11106483]

15. Thorpe JH, Gale BC, Teixeira SC, Cardin CJ. 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 2003;327:97–109. [PubMed: 12614611]

16. Hays FA, Teegarden A, Jones ZJ, Harms M, Raup D, Watson J, Cavaliere E, Ho PS. How sequence defines structure: a crystallographic map of DNA structure and conformation. Proc. Natl. Acad. Sci., USA 2005;102:7157–7162. [PubMed: 15870206]

17. Hays FA, Schirf V, Ho PS, Demeler B. Solution formation of Holliday junctions in inverted-repeat DNA sequences. Biochemistry 2006;45:2467–2471. [PubMed: 16489738]

18. Duckett DR, Murchie AIH, Diekmann S, von Kitzing E, Kemper B, Lilley DMJ. The structure of the Holliday junction, and its resolution. Cell 1988;55:79–89. [PubMed: 3167979]

19. Hays FA, Watson J, Ho PS. Caution! DNA crossing: crystal structures of Holliday junctions. J. Biol. Chem 2003;278:49663–49666. [PubMed: 14563836]

20. Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature (London) 1953;171:737–738. [PubMed: 13054692]

21. Watson JD, Crick FH. Genetical implications of the structure of deoxyribonucleic acid. Nature (London) 1953;171:964–967. [PubMed: 13063483]

22. Otwinowski Z, Minor W. Processing of x-ray diffraction data collected in oscillation mode. Methods Enzymol 1997;276:307–326.

23. Kissinger CR, Gehlhaar DK, Fogel DB. Rapid automated molecular replacement by evolutionary search. Acta Crystallogr. D Biol. Crystallogr 1999;55(Pt 2):484–491. [PubMed: 10089360]

24. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr 1998;54(Pt 5):905–921. [PubMed: 9757107]

25. Hays FA, Vargason JM, Ho PS. Effect of sequence on the conformation of DNA holliday junctions. Biochemistry 2003;42:9586–9597. [PubMed: 12911300]

26. Churchill MEA, Tullius TD, Kallenbach NR, Seeman NC. A Holliday recombination intermediate is twofold symmetric. Proc. Natl. Acad. Sci. USA 1988;85:4653–4656. [PubMed: 3387432]

27. Murchie AIH, Clegg RM, von Kitzing E, Duckett DR, Diekmann S, Lilley DMJ. Fluorescence energy transfer shows that the four-way DNA junction is a right-handed cross of antiparallel molecules. Nature (London) 1989;341:763–766. [PubMed: 2797209]

28. Clegg RM, Murchie AIH, Zechel A, Lilley DMJ. The solution structure of the four-way DNA junction at low-salt conditions: a fluorescence resonance energy transfer analysis. Biophys. J 1994;66:99–109. [PubMed: 8130350]

29. Egli M, Usman N, Zhang SG, Rich A. Crystal structure of an Okazaki fragment at 2-A resolution. Proc. Natl. Acad. Sci., USA 1992;89:534–538. [PubMed: 1370582]

30. Kielkopf CL, Ding S, Kuhn P, Rees DC. Conformational flexibility of B-DNA at 0.74 Å Resolution: d(CCAGTACTGG)$_2$. J. Mol. Biol 2000;296:787–801. [PubMed: 10677281]

31. Zamora F, Kunsman M, Sabat M, Lippert B. Metal-stabilized rare tautomers of nucleobases. 6-Imino tautomer of adenine in a mixed-nucleobase complex with mercury (II). Inorg. Chem 1997;36:1583–1587. [PubMed: 11669746]

32. Thiyagarajan S, Rajan SS, Gautham N. Cobalt hexammine induced tautomeric shift in Z-DNA: the structure of d(CGCGCA)*d(TGCGCG) in two crystal forms. Nucleic Acids Res 2004;32:5945–5953. [PubMed: 15534365]

33. Ho PS, Frederick CA, Quigley GJ, van der Marel GA, van Boom JH, Wang AH, Rich A. G.T wobble base-pairing in Z-DNA at 1.0 A atomic resolution: the crystal structure of d(CGCGTG). EMBO J 1985;4:3617–3623. [PubMed: 4092690]

34. Hunter WN, Brown T, Kneale G, Anand NN, Rabinovich D, Kennard O. The structure of guanosine-thymidine mismatches in B-DNA at 2.5-A resolution. J. Biol. Chem 1987;262:9962–9970. [PubMed: 3611072]

35. Chattopadhyaya R, Grzeskowiak K, Dickerson RE. Structure of a T4 hairpin loop on a Z-DNA stem and comparison with A-RNA and B-DNA loops. J. Mol. Biol 1990;211:189–210. [PubMed: 2299669]

36. Kryachko E, Sabin JR. 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 2003;91:695–710.

37. Biertumpfel C, Yang W, Suck D. Crystal structure of T4 endonuclease VII resolving a Holliday junction. Nature (London) 2007;449:616–620. [PubMed: 17873859]

38. Hadden JM, Declais AC, Carr SB, Lilley DM, Phillips SE. The structural basis of Holliday junction resolution by T7 endonuclease I. Nature (London) 2007;449:621–624. [PubMed: 17873858]

39. Khuu PA, Voth AR, Hays FA, Ho PS. The stacked-X DNA Holliday junction and protein recognition. J. Mol. Recognit 2006;19:1–9. [PubMed: 16180204]

40. Tsai AG, Engelhart AE, Hatmal MM, Houston SI, Hud NV, Haworth IS, Lieber MR. Conformational variants of duplex DNA correlated with cytosine-rich chromosomal fragile sites. J. Biol. Chem 2009;284:7157–7164. [PubMed: 19106104]

41. McCall M, Brown T, Kennard O. The crystal structure of d(G-G-G-G-C-C-C-C). A model for poly (dG).poly(dC). J. Mol. Biol 1985;183:385–396. [PubMed: 4020865]

42. Gao YG, Robinson H, Wang AH. High-resolution A-DNA crystal structures of d(AGGGGCCCCT). An A-DNA model of poly(dG) x poly(dC). Eur. J. Biochem 1999;261:413–420. [PubMed: 10215851]

43. Basham B, Schroth GP, Ho PS. An A-DNA triplet code: thermodynamic rules for predicting A- and B-DNA. Proc. Natl. Acad. Sci. USA 1995;92:6464–6468. [PubMed: 7604014]

44. Topal MD, DiGuiseppi SR, Sinha NK. 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 1980;255:11717–11724. [PubMed: 7002928]

45. Watanabe SM, Goodman MF. On the molecular basis of transition mutations: frequencies of forming 2-aminopurine.cytosine and adenine.cytosine base mispairs in vitro. Proc. Natl. Acad. Sci., USA 1981;78:2864–2868. [PubMed: 6942407]

46. Singer B, Chavez F, Goodman MF, Essigmann JM, Dosanjh MK. 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., USA 1989;86:8271–8274. [PubMed: 2682644]

47. Warren JJ, Forsberg LJ, Beese LS. The structural basis for the mutagenicity of O(6)-methyl-guanine lesions. Proc. Natl. Acad. Sci., USA 2006;103:19701–19706. [PubMed: 17179038]

48. Topal MD, Fresco JR. Complementary base pairing and the origin of substitution mutations. Nature (London) 1976;263:285–289. [PubMed: 958482]

49. Brünger AT. Free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature (London) 1992;355:472–475. [PubMed: 18481394]

50. Lavery R, Sklenar H. Defining the structure of irregular nucleic acids: conventions and principles. J. Biomol. Struct. Dyn 1989;6:655–667. [PubMed: 2619933]

51. Lu XJ, Olson WK. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. Nucleic Acids Res 2003;31:5108–5121. [PubMed: 12930962]
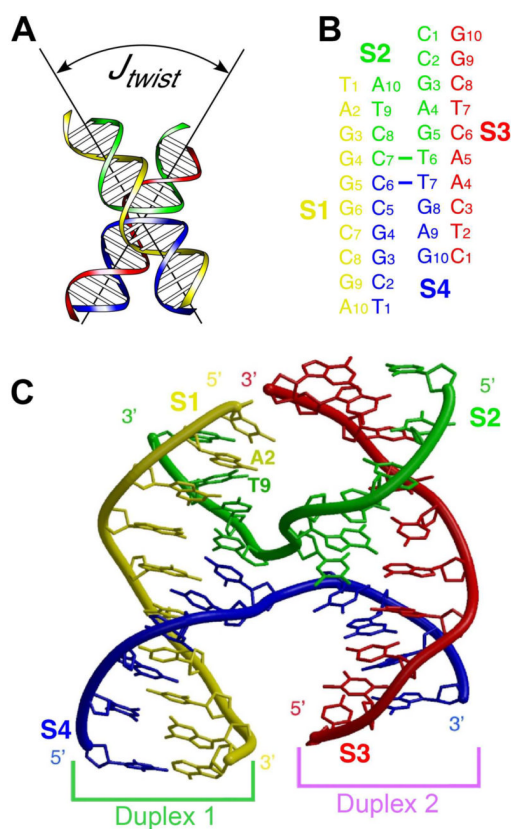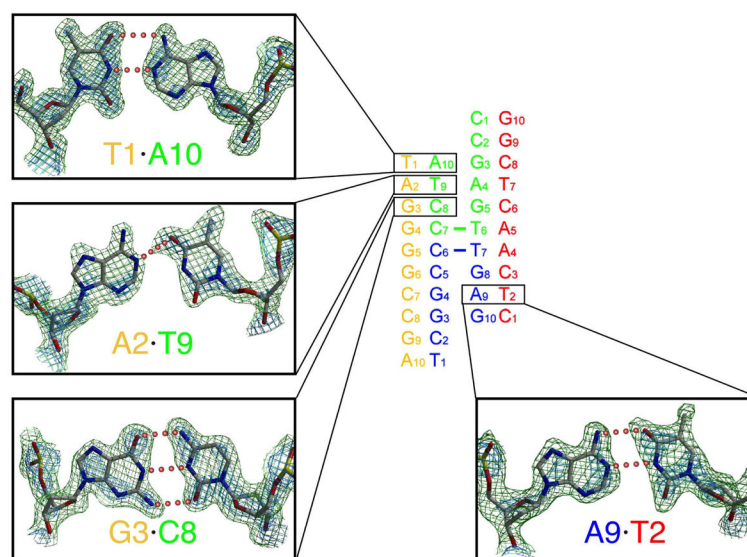
Figure 1.
Construct of asymmetric DNA junction. **A**. Solution model for the four-stranded DNA Holliday junction in the stacked-X form (8), with the angle relating the helical axes of the stacked duplex arms defined as J_{twist} (11). **B**. The four DNA sequences TAGGGGCCGA (Strand S1, yellow), CCGAGTCCTA (strand S2, green), CTCAACTCGG (strand C3, 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 **B**.
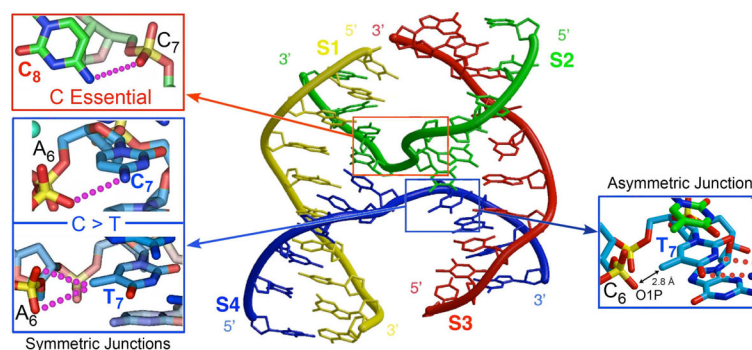
Figure 2.
Intramolecular interactions at the U-turn of symmetric and asymmetric Holliday junctions. Shown on the left are the stabilizing hydrogen bonds from the cytosine bases to the phosphate oxygens at nucleotide positions C_7 and C_8 of the inverted-repeat sequence motif CCnnn$N_6N_7N_8$GG. These help to fix the junction cross-over 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 $N_6N_7N_8$ positions (13,16). In the asymmetric junction, a short distance interaction from the C5-methyl of thymine T7 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).

Figure 3.
Electron density (2Fo-Fc) 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 Fig. 1. Atoms with the shortest approach distances and, therefore, that 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.
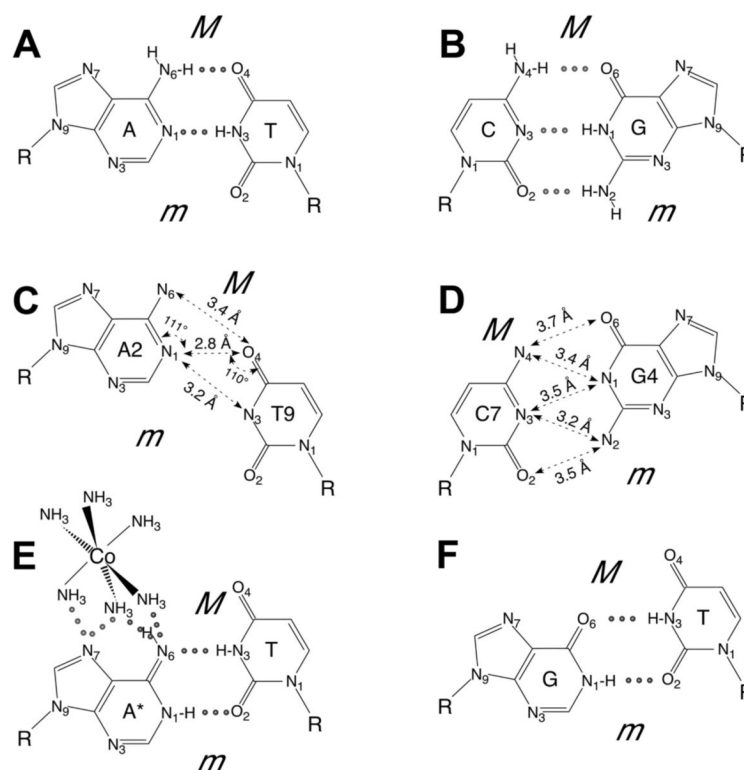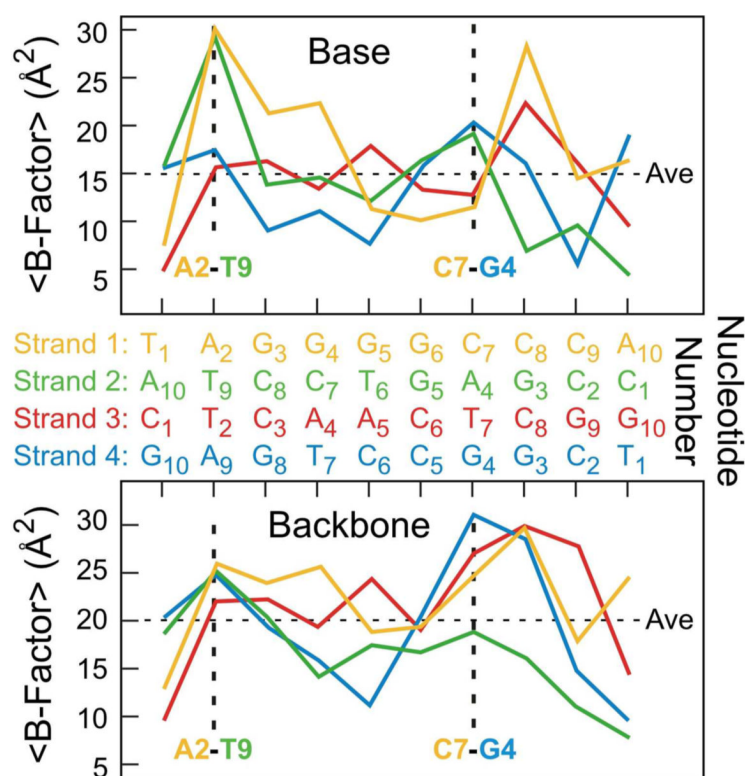
Figure 4.
Structure of Watson-Crick, wobbled rare tautomers, and wobbled mismatched base pairs. Hydrogen bonded atoms are connected with dots, and the major and minor grooves labeled as M and m, respectively. Base pairs of A·T (**A**) and C·G (**B**) as formed by the bases in their stable tautomer forms (amino-forms of A and C, and keto forms of T and G). The 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**. The 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).
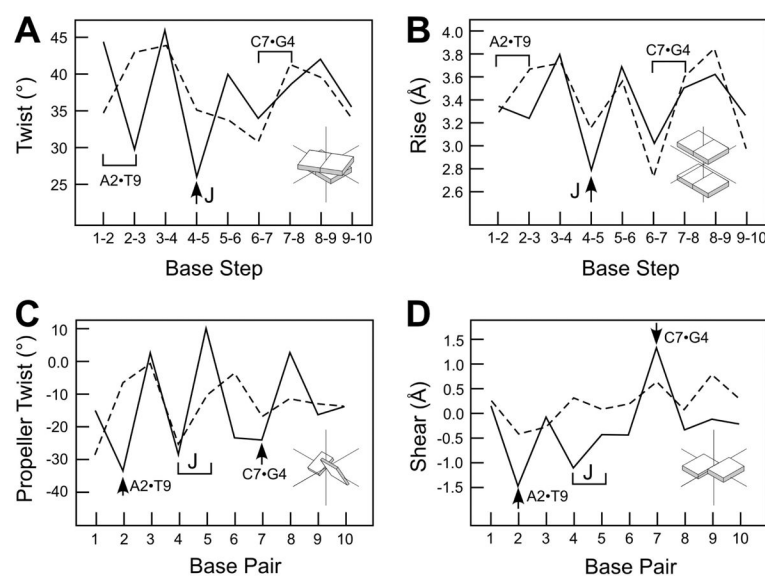
Figure 5.
Helical parameters for dinucleotides and base pairs within the asymmetric junction. The helical rise (**A**) and twist (**B**) between dinucleotides, and propeller twist (**C**) and shear for base pairs (**D**)were calculated using the 3DNA software (51). The semi-continuous double helices (Duplex 1 and Duplex 2) of the junction are as labeled in Fig. 1. The location of the highly sheared A2·T9 and C7·G4 base pairs in Duplex1, along with 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 assembling the sequences TAGGGGCCGA (yellow), CCGAGTCCTA (green), CTCAACTCGG (red), TCGGCCTGAG (blue), colored according to strands in Fig. 1.
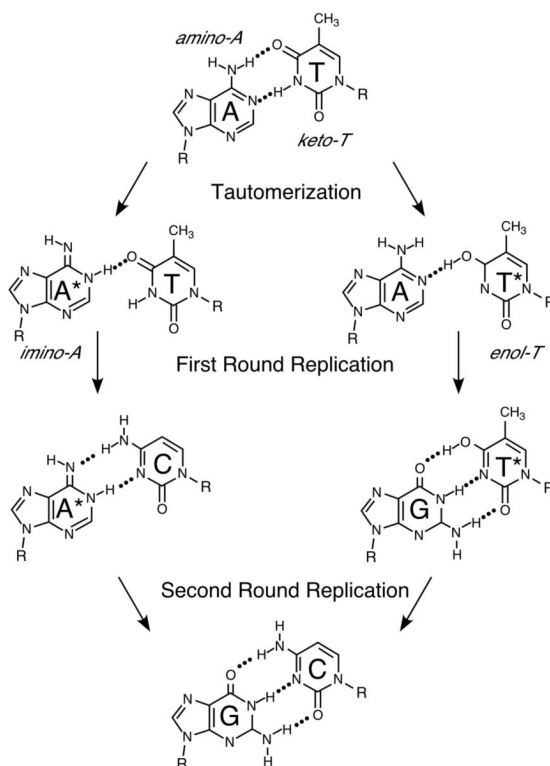
Figure 7.
Potential role of "rare" tautomers in A·T to G·C transition mutations. A normal Watson-Crick A·T base pair, with the adenine in its normal amino and thymine in the keto-forms, would become "wobbled" if the adenine were to adopt the rare imino- (A*) or the thymine adopt the rare enol-tautomers (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* that would appear "normal" in terms of the hydrogen bonding pattern and the displacement of bases towards the major and minor grooves. In the second round, the transition mutation would be fixed, with either the A*·C or G·T* resulting in a G·C base pair in one set of daughter DNA duplexes.

Table 1

Data collection and refinement statistics

Space group	*P1*
Unit cell constants (Å′)	$a = 22.344$ $b = 34.554$ $c = 37.631$ $\alpha = 109.921$ $\beta = 90.028$ $\gamma = 109.016$
Wavelength (Å′)	1.542
Resolution (Å′))	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
Number of DNA atoms	809
Number of solvent atoms	120

[a]Value in parentheses refer to the highest resolution shell.

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

Table 2

Geometry of the asymmetric junction assembled from four unique sequences compared to symmetric junctions formed by single inverted-repeat sequences. The geometries are defined by interduplex angle 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 the program CURVES (50).

Structure	Sequences	PDB Accession Number	IDA (°)	J_{twist} (°)	J_{slide} (Å)	J_{roll} (°)
Asymmetric Junctions						
Solution Model[1] 18	Various	NA	60°	60°	NA	NA
3IGT Crystal Structure	see Fig. 1		70.4	56.5	0.8	134.5
Symmetric Junctions (see Watson, *et al.* 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*

[1]J_{twist} approaches the IDA in junctions with long duplex arms such as those used to study the structure in solution.