JMB

The Inherent Properties of DNA Four-way Junctions: Comparing the Crystal Structures of Holliday Junctions

Brandt F. Eichman[1], Miguel Ortiz-Lombardía[2], Joan Aymamí[3,4] Miquel Coll[3]* and Pui Shing Ho[1]*

[1]*Department of Biochemistry and Biophysics, ALS 2011 Oregon State University Corvallis, OR 97331-7305 USA*

[2]*Institut Pasteur, Unité de Biochimie Structurale, 25, rue du Dr. Roux, 75724 Paris Cedex 15, France*

[3]*Institut de Biologia Molecular de Barcelona, C.S.I.C. Jordi Girona 18, E-08034 Barcelona, Spain*

[4]*Department d'Enginyeria Quimica, Universitat Politècnica de Catalunya Diagonal 647, 08028 Barcelona Spain*

*Corresponding authors

Holliday junctions are four-stranded DNA complexes that are formed during recombination and related DNA repair events. Much work has focused on the overall structure and properties of four-way junctions in solution, but we are just now beginning to understand these complexes at the atomic level. The crystal structures of two all-DNA Holliday junctions have been determined recently from the sequences d(CCGGGACCGG) and d(CCGGTACCGG). A detailed comparison of the two structures helps to distinguish distortions of the DNA conformation that are inherent to the cross-overs of the junctions in this crystal system from those that are consequences of the mismatched dG·dA base-pair in the d(CCGGGACCGG) structure. This analysis shows that the junction itself perturbs the sequence-dependent conformational features of the *B*-DNA duplexes and the associated patterns of hydration in the major and minor grooves only minimally. This supports the idea that a DNA four-way junction can be assembled at relatively low energetic cost. Both structures show a concerted rotation of the adjacent duplex arms relative to *B*-DNA, and this is discussed in terms of the conserved interactions between the duplexes at the junctions and further down the helical arms. The interactions distant from the strand cross-overs of the junction appear to be significant in defining its macroscopic properties, including the angle relating the stacked duplexes across the junction.

© 2002 Elsevier Science Ltd. All rights reserved

Keywords: Holliday junctions; four-way junction; recombination; DNA structure

Introduction

Homologous recombination is recognized as an increasingly important process in DNA metabolism. It is involved in DNA repair, and for maintaining genomic integrity and genetic diversity.[1] For example, homologous recombination has been shown to be promoted by the product of the human breast cancer-associated BRCA2 gene,[2] while the BLM gene associated with Bloom's syndrome corresponds to an antirecombinase activity.[3] A four-stranded DNA junction was proposed by Holliday as the central intermediate in the mechanism of homologous recombination.[4] Holliday junctions are formed when strands cross-over and are shared between two different double-helical segments. The overall structure and properties of four-way junctions have been studied extensively in solution.[5] In the presence of divalent cations, these junctions exist predominantly as the stacked-X form in which the double-helical segments are coaxially stacked and twisted by 60° in a right-handed direction across the junction cross-over. In this structure, the stacked arms resemble two adjacent double-helices, but are linked at the junction by two common strands that cross-over between the duplexes. The overall features of several different types of four-way junctions from recent crystal structures[6–10] are in good agreement with the solution studies.

The first crystal structures of four-way junctions containing all deoxyribonucleotides were solved serendipitously in two different laboratories within

Present address: B. F. Eichman, Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, C2-226, Boston, MA 02115, USA.

E-mail addresses of the corresponding authors: mcccri@cid.csic.es; hops@ucs.orst.edu
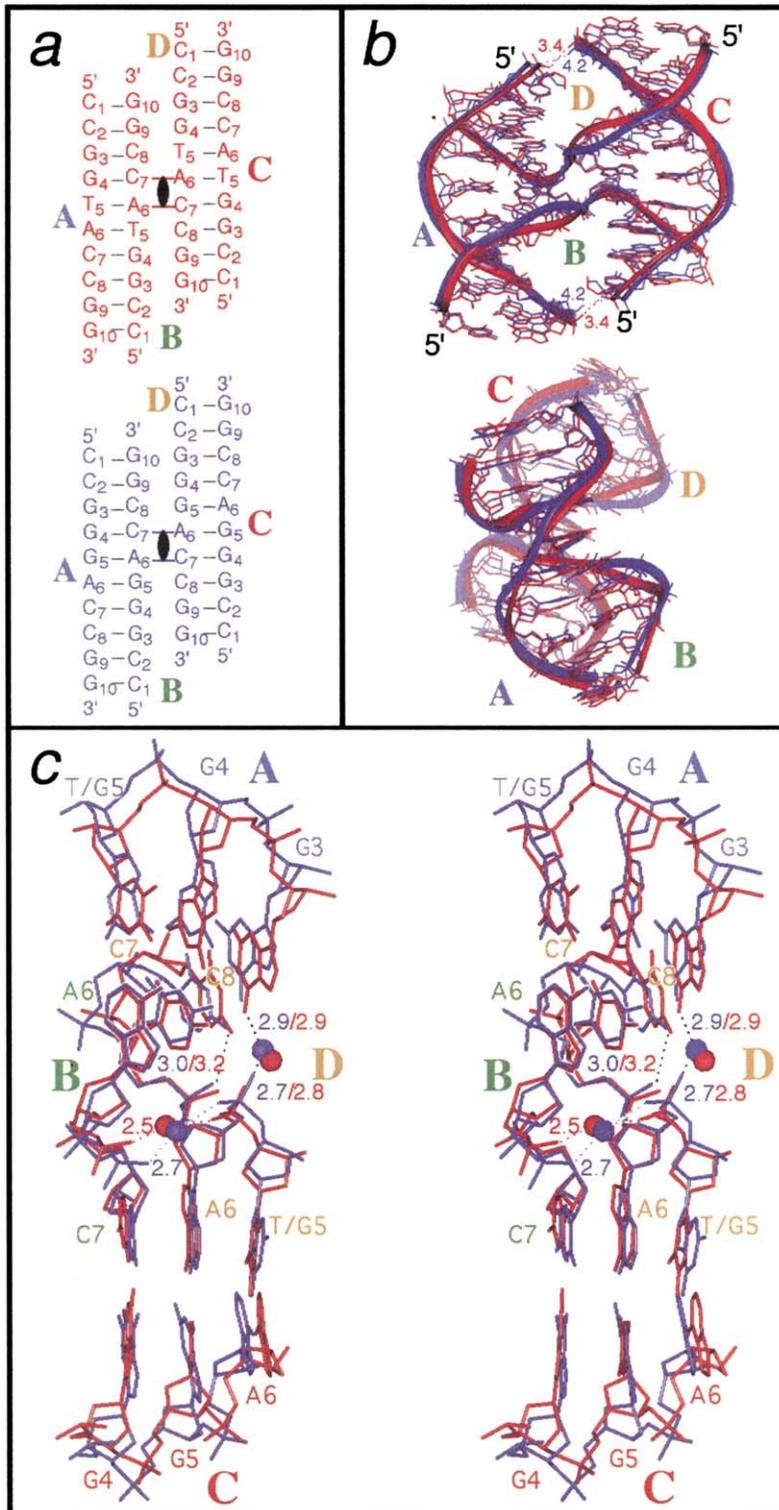
Figure 1. Comparison of the crystal structures of the TA junction from the sequence d(CCGGTACCGG) (red)[8] and the GA junction from d(CCGGGACCGG) (blue)[7]. All distances shown are in Å. (a) Schematics showing the association of four strands, A–D, into the stacked-X form of the four-way junctions. For simplicity, strands A–D of d(CCGGGACCGG) (bottom) were re-assigned relative to the published sequence in order to correspond to d(CCGGTACCGG) (top). (b) Overlay of the TA and GA structures, viewed into the major groove face (top) and along (bottom) the junction. All common atoms between the two overlaid structures show an rmsd of 1.15 Å. The phosphate backbones along each strand are traced with a ribbon. (c) Stereoview of atoms at the ACC core, showing the contacts formed between the bases and phosphate groups across the junction. Images are rotated 90° in the plane of the page with respect to the top structure in (b). Solvent molecules are rendered as spheres.

six months of one another. Ortiz-Lombardía *et al.* solved the first of these structures from the sequence d(CCGG**GA**CCGG) while studying the effects of tandem dG·dA mismatched base-pairs (underlined) on *B*-DNA.[7] The second all-DNA Holliday junction was solved from the sequence d(CCGGTACCGG), which was designed to study the effects of the photochemotheraputic drug psoralen on the DNA double-helix.[8] It should be noted that the two sequences (referred to as the

GA and TA sequences) are identical except for the nucleotide at position 5 in the DNA (Figure 1). Therefore, we can define d(CCGGNACCGG) as the sequence motif for the DNA Holliday junction in single crystals solved to date.

Here, we compare and contrast the details of the two structures, focusing on the differences and similarities between the structures of the inverted-repeat TA and mismatched GA Holliday junctions in order to determine which structural

Table 1. Crystallographic data for the Holliday junction structures of d(CCGGGACGG)[7] and d(CCGGTACCGG)[10]

	CCGGGACCGG	C^{5Br}CGGGACCGG	CCGGTACCGG
A. *Data collection*			
Space group	C2	C222$_1$	C2
Unit cell parameters			
a, b, c (Å)	64.20, 23.74, 38.30	23.75, 63.90, 71.40	66.45, 23.50, 76.94
β (deg.)	112.43	90	114.83
Volume (Å^3)	53,957	108,358	109,040
DNA strands/asymmetric unit	2	2	4
Resolution range (Å)	19.39–2.16	19.07–2.71	30.16–2.10
R$_{merge}$	0.069 (0.137)	0.099 (0.293)	0.045 (21.8)
B. *Refinement*			
Resolution range	19.39–2.16	–	8.00–2.10
DNA atoms (solvent molecules)	408	–	808
Water molecules	30	–	92
R (R$_{free}$)	0.240 (0.288)	–	0.230 (0.318)
r.m.s.d. bond lengths (Å)	0.008	–	0.017
r.m.s.d. bond angles (deg.)	1.7	–	1.9

Statistics for d(C^{5Br}CGGGACCGG) are for data collected at the peak wavelength (0.9224 Å). Numbers in parentheses refer to data in the highest-resolution shell.

properties are truly inherent to the DNA junction in the crystal structures, and which are effects of the base mismatches. We find that the DNA structure across the stacked arms at the junction is influenced by base-stacking, and that the interactions at the ACC-core sequence and at the ends of the duplex arms are conserved in both structures. These results indicate a strong nucleotide sequence-dependence in the crystal structures of Holliday junctions. The interactions between adjacent duplex arms impose slight distortions to the helical twist at the base steps flanking the junction, showing how the interactions removed from the junction influence the overall geometry of the four-stranded complex.

Results

Comparisons of the crystallography

The two junction sequences were crystallized under very similar conditions. The GA sequence formed thin, plate-like crystals from a solution containing 0.33 mM DNA, 133.3 mM MgCl$_2$, 6.7% (v/v) 2-methyl-2,4-pentanediol (MPD), and which was equilibrated against 45% MPD. Thin plates were obtained from the TA sequence using 0.25 mM DNA, 75 mM sodium cacodylate buffer (pH 7), 15 mM CaCl$_2$, 2.5% MPD, and equilibrated against 30% MPD. Diffraction quality crystals of the TA sequence could be obtained with 50 to 150 mM MgCl$_2$ in place of CaCl$_2$ in the crystallization solutions.

Both GA and TA native crystals belong to the monoclinic space group C2. The Br-cytosine derivative of the GA crystal (d(C^{5Br}CGGGACCGG), referred to as the Br-GA sequence from this point on), used for multiwavelength anomalous diffraction (MAD) phasing, belongs to the orthorombic space group C222$_1$ (Table 1). The unit cell dimensions of both the GA and Br-GA crystals accommo-

date two DNA strands in the asymmetric unit, while the asymmetric unit of the TA crystal consists of four DNA strands. Therefore, the mismatch junction is formed in both the GA and Br-GA crystals from two duplexes related by 2-fold crystallographic symmetry, while the junction in the TA sequence is composed of four structurally unique DNA strands. This loss of crystallographic symmetry in the TA sequence is a result of a ∼2 Å shift in the DNA along the a-axis (away from the c-axis). Even with this lattice shift, the DNA crystal packing is identical between all three structures. The TA junction has been crystallized recently in the smaller GA-type lattice,[11] and shows essentially the same conformation as the structure in the larger unit cell.

The structures were solved and refined using different X-ray diffraction and refinement methods. The Br-GA structure was solved with experimental phases obtained from a bromine MAD experiment, although the structure could be solved as well by molecular replacement, using a double-helical starting model based on the mismatch-containing decamer d(CCAAGATTGG).[12] The GA structure was solved by molecular replacement from the resulting Br-GA model, and was refined to 2.16 Å using REFMAC.[13] In contrast, the TA structure was solved using molecular replacement with idealized B-DNA helices as search models, and was refined to 2.10 Å with X-PLOR 3.851.[14]

The overall structures of the DNA Holliday junctions

The structures formed from the TA junction[8] and GA junction sequences[7] are stacked-X type Holliday junctions, and, overall, are virtually identical (Figure 1). The root-mean-square deviation (r.m.s.d.) between the junction structures of the GA and TA sequences is 1.15 Å for all atoms except

Table 2. Backbone torsion angles (degrees) for nucleotides at the junction cross-overs in d(CCGGGACCGG)[7] and d(CCGGTACCGG)[8]

Torsions		d(CCGGGACCGG)		d(CCGGTACCGG)		Canonical *B*-DNA	
Angle	Nucleotide	Junction	Arms	Junction	Arms	Dodecamer[36]	Decamer[15]
δ	A6	137.1	148 ± 12	147.1 (1.5)	143 ± 9	123 ± 21	133 ± 19
ε	A6	**− 73.6**	− 132 ± 29	**− 890 (0.9)**	− 125 ± 34	− 169 ± 25	− 151 ± 34
ζ	A6	− 96.6	− 156 ± 51	− 77.8 (1.2)	− 165 ± 42	− 108 ± 34	− 130 ± 52
α	C7	− 47.1	− 62 ± 17	− 70.1 (7.0)	− 55 ± 21	− 63 ± 8	− 68 ± 5
β	C7	**− 177.1**	153 ± 16	**− 156.8 (1.4)**	145 ± 18	171 ± 14	162 ± 14
γ	C7	47.5	42 ± 9	57.8 (5.4)	42 ± 11	54 ± 8	50 ± 6
χ	C7	**− 152.1**	− 83 ± 16	**− 147.3 (0.9)**	− 82 ± 14	− 117 ± 14	− 96 ± 18

Torsion angles[38] are defined as P-(α)-O5′-(β)-C5′-(γ)-C4′-(δ)-C3′-(ε)-O3′-(ζ)-P. Angles at the junctions are from the specified nucleotides that flank the cross-over point, and angles from the arms were averaged across all remaining nucleotides (± one standard deviation). In the d(CCGGTACCGG) structure, junction angles were averaged across the specified nucleotide in strands B and D (values in parentheses show the spread between the highest and lowest values). Canonical *B*-DNA torsion angles are reported for the Drew dodecamer d(CGCGAATTCGCG)[37] and the 0.74 Å structure of d(CCAGTACTGG),[15] and are averages for all nucleotides in those structures except cytosine C1. Values in bold-face differ from the corresponding angle in the Drew dodecamer by greater than two standard deviations.

nucleotides G5 and T5. They are composed of four strands of DNA, with two strands (labeled A and C) wrapping around the outside of the duplexes and thereby tracing a pathway nearly identical with that expected for two resolved *B*-DNA duplexes. The two inner strands (B and D) cross-over between the duplexes, making an abrupt turn as the strands exchange between the duplexes. This therefore defines the structures as antiparallel four-way junctions, where the orientations of each strand alternate in directions (5′ to 3′, followed by 3′ to 5′, and so forth). DNA strands B and D each cross-over between A6 and C7 and base-pair with complementary strands A and C to form two pseudo-continuous *B*-DNA helices, each composed of a 6 bp arm stacked on a 4 bp arm (Figure 1(a) and (b)). These stacked helices are related in a right-handed sense by ∼ 41° in both structures.

The interactions between the adjacent duplexes, those at the junction A6–C7–C8 triplet sequence and between the ends of the 4 bp and 6 bp arms are conserved in both structures. The hydrogen-bonding network and van der Waals interactions at the ACC junction core[7,8] are evident in both structures, showing that these are defining characteristics of the two junctions (Figure 1(c)). In addition, there are contacts between adjacent duplex arms that are further removed from the ACC core. The phosphate oxygen atoms of C2 on the 4 bp arms and G10 on the 6 bp arms are within 3.4 Å and 4.2 Å in the TA and GA structures, respectively (Figure 1(b)). Although these interactions are distant from the junction core, they appear to be equally important in defining the overall geometry of the DNA junctions. Therefore, the overall structures are extremely similar and have the same sequence-dependent stabilization of the junction.

A more detailed analysis of the helical parameters of both structures shows that at the backbone and base-pair levels, there are minor distortions from *B*-DNA at the junction. The distortions that are common to both structures are

inherent properties of these junctions, and differences between the two structures highlight the effects of the mismatched d(G·A) base pairs.

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

The Holliday junction does not dramatically affect the *B*-DNA nature of the helical arms in either the GA or the TA structures. The phosphoribose backbone is surprisingly unperturbed, even with the sharp re-direction of the strands at the cross-overs. The phosphate group positions in both structures are nearly identical with the positions expected in two adjacent, resolved duplexes. Aside from the deviation in backbone trajectory imposed by the mismatched d(G·A) base-pairs (described in detail below), the only differences in backbone torsion angles from canonical *B*-DNA occur as a result of the sharp direction change in the backbone at strands B and D of the junction. This direction change can be described primarily by rotations around χ (glycosidic bond), ε (C4′–C3′–O3′–P), and β (P–O5′–C5′–C4′) of nucleotides A6 and C7 (Table 2).

The largest distortions to base-pair stacking in the GA structure occur at the junction and the immediately adjacent base steps (Figure 2). There is no analogous d(GpG/ApC) base step in a regular *B*-DNA double helix to compare to the GA junction, but the helical twist (27°) is very low for a standard *B*-DNA duplex. The helical twist at the d(G4pT5/A6∗C7) dinucleotide step of the TA junction (34°), where the asterisk (∗) refers to the strand cross-over, however, is nearly identical with that in the analogous dinucleotide of the d(CCAGTACTGG) *B*-DNA structure (Figure 2(a) and (b)).[15] The low twist at the GA junction is thus not an intrinsic characteristic of the four-way junction, but is expected to show the sequence-dependent variations seen in the base-pair stacking of *B*-DNA duplexes (Figure 3).
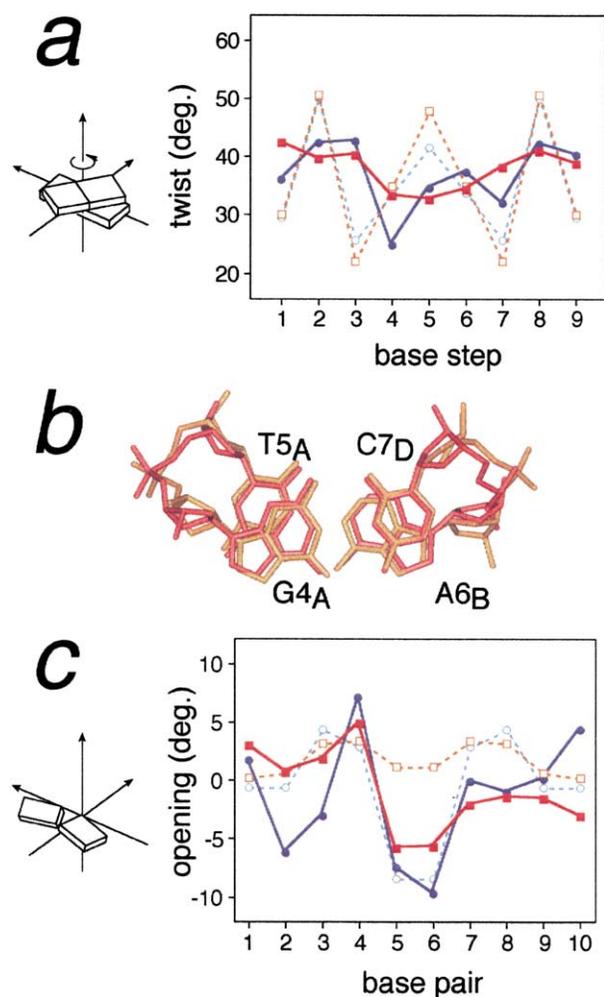
Figure 2. Similarity in base-stacking (helical twist between base-pairs and opening of base-pairs) between the DNA junctions and corresponding *B*-DNA sequences. Filled symbols and continuous lines correspond to the junction structures, while open symbols and broken lines correspond to analogous *B*-DNA duplex structures with central d(TpA) and d(GpA) steps. Data are shown for the TA junction of d(CCGGTA-CCGG) (red squares), to GA junction of d(CCGGGA-CCGG) (dark blue circles), the *B*-DNA duplex of d(CCA-GTACTGG) (orange squares)[15], and the *B*-DNA duplex of d(CCAAGATTGG) (light blue circles)[12]. Parameters were calculated using CURVES 5.2[17]. (a) Twist angles for the nine dinucleotide steps, numbered 1–9 from the 5′-end of strand A. The strand cross-overs of the junctions occur at base step 4. (b) Sequence-dependence in the base-stacking at the d(G4pT5/A6∗C7) base step across the junction in d(CCGGTACCGG) (red) and in d(CCAGTACTGG) (orange), where the asterisk (∗) refers to the position of the strand cross-over. Subscripts refer to the DNA strands A, B, and D. (c) Opening angle for the ten base-pairs, numbered as in (a).

Interestingly, the base steps flanking the junction along the 4 bp and 6 bp arms are overwound and underwound, respectively (Figures 2(a) and 3). The twist angles at the 4 bp arm d(G3pG4)/C7pC8) steps are overwound by ∼5° compared to the analogous d(GG/CC) step in the *B*-DNA struc-

ture of d(CCAGGCCTGG).[16] Conversely, the twist angles at the central d(TpA) and d(GpA) base steps, along the 6 bp arms, are lower than the corresponding steps in regular *B*-DNA structures (Figure 2(a)). On average, the helical twist of the 4 bp arms of the GA-junction is 1.2° (calculated by CURVES[17]) to 2.1° (from 3DNA[18]) larger compared to the 6 bp arms. In the TA junction, the shorter arms are 2.4° (CURVES[17]) to 1.8° (3DNA[18]) overwound relative to the longer arms. Thus, the overwinding on one side of the junction is compensated for by an underwinding on the other side, and can be explained by the interactions between the adjacent arms that are distant from the cross-overs of the junctions.

In addition to the perturbations seen in the base stacking, the geometry within the d(T·A) base-pairs of the TA junction appear to be affected by the strand cross-overs. For example, the opening angles of these base-pairs are lower than expected in a *B*-DNA duplex (Figure 2(c)). The mismatched d(G·A) base-pairs exhibit the same negative opening, but those distortions are inherent to d(G·A) mismatches, since they are observed also in the *B*-DNA structure of d(CCAAGATTGG).[12] The negative opening angle therefore may not be attributed to the junction directly but, instead, may be associated with the flexible nature of d(T·A) base-pairs in general.[19] In summary, the largest distortions of the *B*-DNA arms that are common to both the GA and TA junctions are seen in the backbone torsion angles between A6 and C7 of the cross-over strands, the low helical twist angles between bases that span the point of cross-over, and the compensating overwinding and underwinding of the 4 bp and 6 bp arms, respectively (Table 2).

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

The most obvious difference between the TA and GA Holliday junction crystal structures is seen as a distortion of the backbone near the strand cross-overs (Figure 1(b)). A comparison of the local helix axes in both Holliday structures (Figure 4) indicates that the duplexes are slightly bent (10°–15°), with a change in their direction localized at the junction base step. In the TA sequence, the axis trajectory is smoother than in the GA sequence, where the mismatches clearly enhance a zig-zag effect in the axis trajectory.

The positions of the phosphate groups overlay precisely between the two structures, except at phosphate group 5, where the strands cross-over and, to a lesser degree, between phosphate groups 4 and 7 on the non-cross-over strands. It is not surprising that the main differences in the backbone trace between the two structures occur at the uncommon nucleotide. Superimposition of crystal structures of *B*-DNA decanucleotides

5' 3'
C₁ᵦ G₁₀ₐ

C₂ᵦ G₉ₐ

G₃ᵦ C₈ₐ

G₄ᵦ C₇ₐ

T₅ᵦ A₆ₐ

A₆ᵦ T₅ₐ

C₇ᴅ G₄ₐ

C₈ᴅ G₃ₐ

G₉ᴅ C₂ₐ

G₁₀ᴅ C₁ₐ
3' 5'

5' 3'
C₁ᵦ G₁₀ₐ

C₂ᵦ G₉ₐ

G₃ᵦ C₈ₐ

G₄ᵦ C₇ₐ

G₅ᵦ   A₆ₐ

A₆ᵦ G₅ₐ

C₇ᴅ G₄ₐ

C₈ᴅ G₃ₐ

G₉ᴅ C₂ₐ

G₁₀ᴅ C₁ₐ
3' 5'
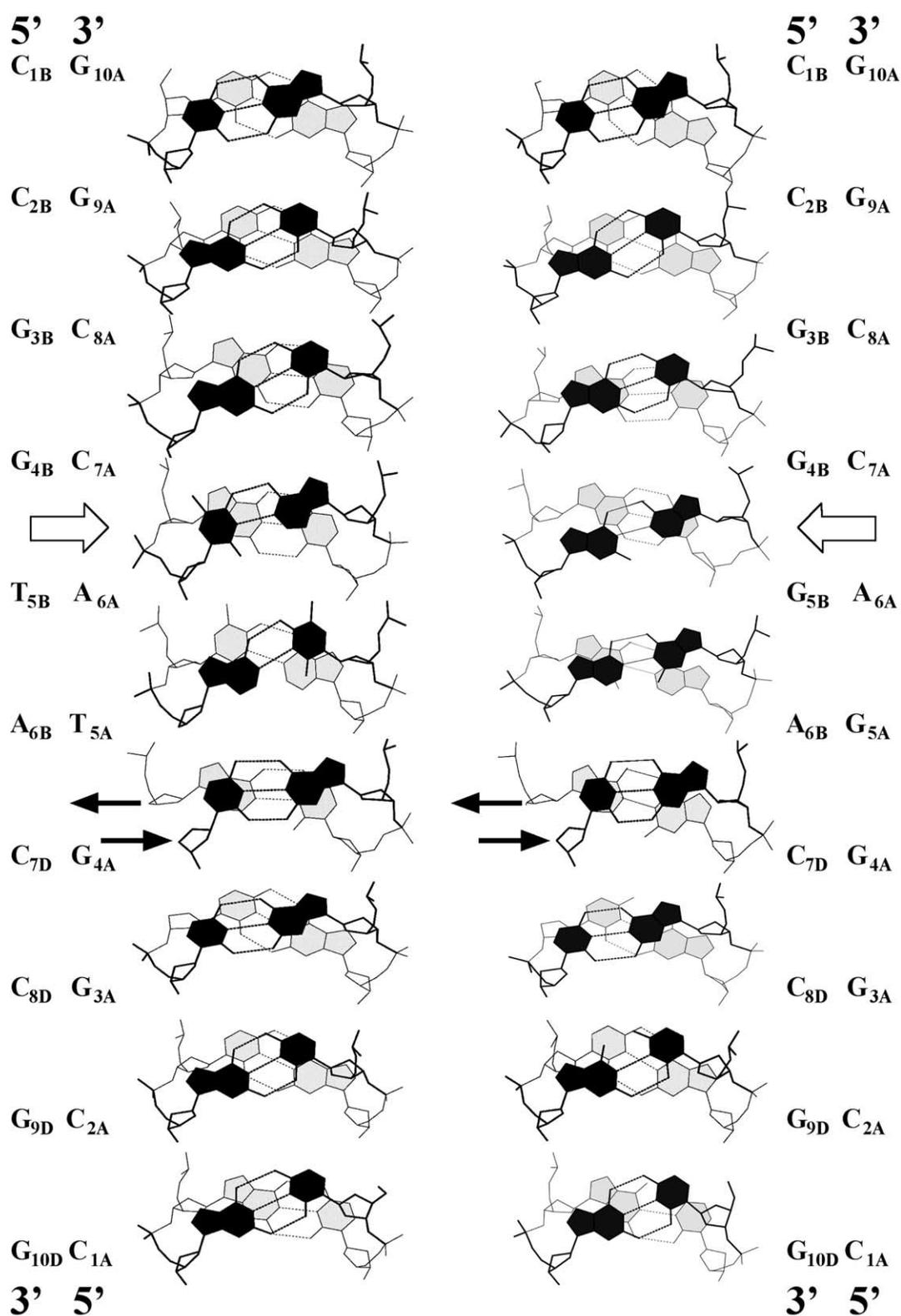
Figure 3. Base-pair stacking along the stacked arms of the d(CCGGTACCGG) (left) and d(CCGGGACCGG) (right) structures. The filled arrows indicate the cross-over point where strands of the junction exchange to and from the adjacent stacked arms. Open arrows indicate the equivalent sequence at a non-crossing step. The plots were generated with 3DNA.[18]

Table 3. Angles between *B*-DNA duplexes packed into X-type crystal lattices

Length (bp)	Sequence	Space group	Angle (deg.)	NDB code	PDB code	References
12	ACCGGCGCCACA/TGTGGCGCCGGT	*R*3	74	BD0022	1QP5	[39,40]
	ACCGCCGGCGCC/GGCGCCGGCGGT	*R*3	74	BDL035	330D	[39]
10	CCGGCGCCGG	*R*3	77	BDJ039	1CGC	[41]
	CCGCCGGCGG	*R*3	77	BD0015	1QC1	[28]
	CGATCG^{6m}ATCG	*P*3$_2$21	60	BDJB48	1DA3	[42]
	CCAACITTGG	*P*3$_2$21	60	BDJB43	1D60	[43]
	CCACTAGTGG	*P*3$_2$21	60	BDJ061	–	[44]
	CCATTAATGG	*P*3$_2$21	60	BDJ055	167D	[45]
	CAAAGAAAAG/CTTTTCTTTG	*C*2	51	BDJ081	307D	[46]
	CGCAATTGCG	*C*2	51	BDJ069	252D	[47]
	CCGCTAGCGG	*C*2	44	BD0028	1DCV	[8]
	CTCTCGAGAG	*C*2	43	BDJ060	196D	[48]
6	GGCGCC	*P*4$_1$2$_1$2	90	BD0040	1F6C	[49]

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

containing central d(TpA) and mismatched d(GpA) base steps onto the two stacked arms of the Holliday structures show that the backbone traces are strongly determined by the nucleotide sequence (Figure 5). The differences seen in the backbone traces between the two Holliday structures can be directly attributed to the mismatched d(G·A) base-pairs. However, it is clear that at phosphate group 6, the backbone trajectory in the TA Holliday structure is similar to that in the GA Holliday structure, but different from regular *B*-DNA, indicating that the conformation at this point is defined by the junction, not the sequence.

In addition to the backbone trajectories, other differences in the base-pairs between the two Holliday structures are a result of the mismatches in the GA structure, and not an inherent feature of DNA junctions. Inspection of base-pair parameters stretch, stagger, *y*-displacement, propeller twist, and inclination in the two Holliday structures and in two analogous *B*-DNA sequences shows that the central base-pairs are different between the TA and GA structures, and that these differences are sequence-dependent (Figure 6). Distortions at the mismatched d(G·A) base-pairs are identical with those observed in the mismatched d(G·A) *B*-DNA structure.[12] For example, the large negative propeller twist at the mismatch is associated with the extra hydrogen bonding across the stacked arms between N2 of guanine G5 (strand A) and O2 of C7 (strand D). In contrast, the base-pair parameters in the TA junction structure are remarkably similar to a d(T·A)-containing *B*-DNA structure,[15] again illustrating the fact that the junction does not significantly distort the *B*-DNA nature of the duplexes.

In conclusion, the differences in backbone and axis trajectories and base-pair geometries between the TA and GA junctions structures are a direct result of the d(G·A) mismatched base-pairs, and cannot be attributed to the junction itself.

Solvent structure of the Holliday junction

So far, the similarities and differences between the TA and GA junction structures have been described at the DNA level. How does the surface of a four-stranded DNA Holliday junction complex appear to the surrounding solvent? The solvent-accessible surface (SAS) shows that the major and minor grooves are as they appear in standard *B*-DNA double-helices, with the minor grooves on one face of the complex separated by the raised phosphoribose backbone ridge at the junction (Figure 7(a)), and the major groove surfaces connected smoothly by the junction on the other face of the complex (Figure 7(c)). Two cavities are formed on either side of the strand cross-overs by the interactions between two adjacent arms (Figure 7(a)–(b)). The cavity floor is formed from the cross-over strands at the junction, the ceiling from the interaction between phosphate groups at the ends of the arms, and the walls from the major groove of the 4 bp arm facing the minor groove of the 6 bp arm. This cavity is the site of six out of the 13 conserved water molecules between the TA and GA structures (Figure 7(b)), including an essential sequence-dependent bridging water molecule at the ACC core.[7,8]

It is possible to gain further insight into the nature of four-way junctions by comparing the solvent patterns observed in the crystal structures. Most importantly, given the similarity of the helical arms to *B*-DNA, we can raise the question of whether the hydration of the stacked arms are what would be expected for standard *B*-DNA, or whether the junctions have unique patterns of hydration. A direct comparison between the two junction structures is slightly hampered by the asymmetric unit of the TA structure consisting of four unique DNA strands, while that of the GA consists of two strands, with the full four-stranded junction being generated by crystallographic symmetry. Thus, each duplex in the TA structure
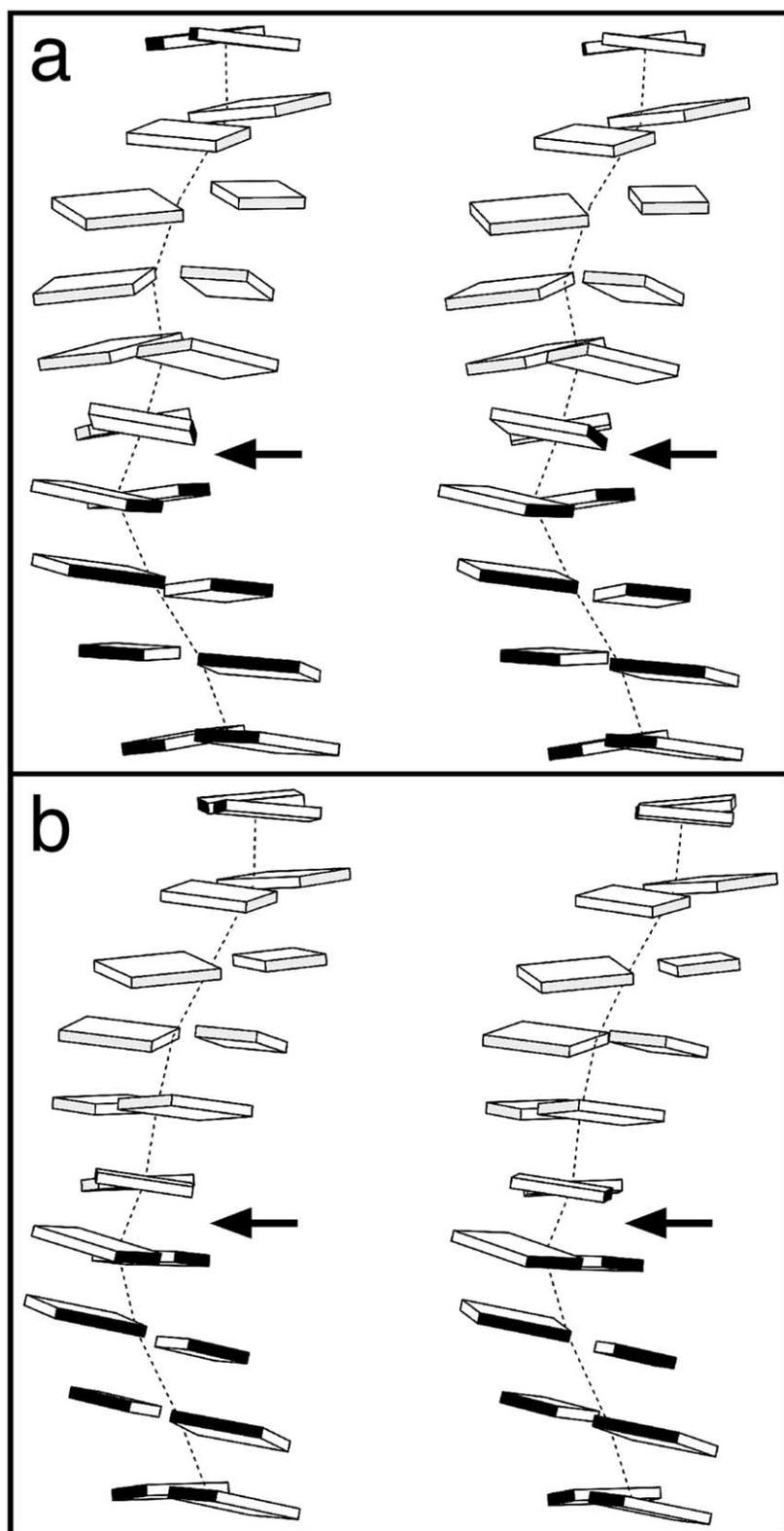
Figure 4. Block-base stereo representations of the pseudocontinous helices formed by the stacked long and short arms of the TA (left) and GA (right) junctions. The local helix axes are represented by broken lines. The arrows indicate the cross-over steps. The plot was generated with the program 3DNA.[18]

shows unique sets of associated water molecules. Therefore, to compare the water molecules between the two structures accurately, one complete solvation pattern for the four-stranded TA complex was generated by applying 2-fold symmetry to map the water molecules from one duplex onto the other.

DNA structure is highly dependent on the detailed interactions of solvent with each base-pair, and each particular DNA conformation has been shown to have a unique pattern of hydration.[20] In *B*-DNA, a characteristic "spine of hydration" has been observed as a hydrogen-bonded network of water molecules in the minor groove with
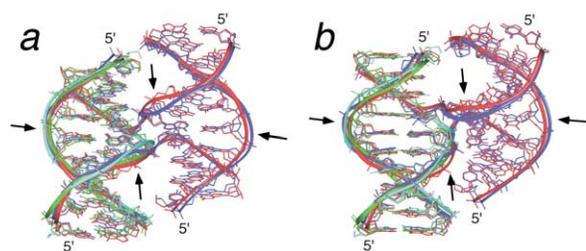
Figure 5. Comparison of the phosphoribose backbone trajectories between the TA and GA junction structures and analogous *B*-DNA duplexes. The TA and GA junction structures were superimposed and are colored as in Figure 1. Central d(T·A)- and d(G·A)-containing *B*-DNA duplexes d(CCGCTAGCGG) (green)[8] and d(CCAAGA-TTGG) (light blue)[12] are superimposed on one (left) stacked duplex of the junction structures. Phosphate group positions along each DNA strand are traced with a ribbon. Arrows indicate the positions of the d(G·A) mispairs. Two views are shown, and are related by a 20° rotation along the junction.

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

With the exception of a few solvent molecules at the strand cross-overs, all of the first-shell water molecules at the arms of the TA and GA junction structures are indicative of *B*-DNA. However, not all of the DNA hydrogen-bond donor and acceptor sites are occupied with water molecules. For example, the minor groove spine of hydration typical of *B*-DNA is not fully extended along the length of the duplex arms. The water molecules seem to be clustered at the ends of the arms and absent from the central base-pairs, especially in the minor grooves of the mismatch junction structure (Figure 8(a)). This apparent lack of water molecules at the central base steps is most likely a function of the resolution of the two structures, and not a result of the junction. The total number of water molecules observed in the two junction structures is not unusual for structures at this resolution (Figure 9). Thus, higher-resolution structures will be required to assess the effect of the junction on the hydration pattern at the base-pairs across the junction.

The four-stranded complexes of both structures contain 49 first-shell water molecules that were observed in the electron density maps, 13 of which (six per duplex plus one at the center of the junction) have identical positions between the two structures. Almost all of these conserved water

molecules lie in the cavity formed at the interface of adjacent duplexes (Figures 7(b) and 8(b)). In the minor groove, there are four conserved interactions. The water molecules at G3 (strand A) and G9 (strand D), at the ends of the 4 bp arms, are the only two that do not face the adjacent duplex across the junctions. The remaining two are located at the 6 bp arms, one at G9 (strand A), and the other at the d(T·A) and d(G·A) base-pairs, which is hydrogen-bonded to the O2 keto oxygen atom of T5 (strand B) and the corresponding N3 imino nitrogen atom of G5 (strand B). In the major groove, the conserved water molecules are located on the 4 bp arms at C2 and G3. The water at G3 bridges the O6 oxygen atom of this guanine base with the O1P oxygen atom of A6 at the strand cross-overs, apparently contributing to the stability of the junction.[7,8] Thus, the majority of the water molecules that occupy common positions in the two structures are lining the duplex interface cavity.

The conformations of four-way DNA junctions have been shown to be dependent on the particular cations present; the stacked-X form is stabilized by polyvalent cations.[5] As expected, the junction cross-overs, where four phosphate groups are packed closely within 6.5 Å of each other (Figure 7(a) and (b)), show very highly negative electrostatic potential surfaces. Although crystals of both sequences were grown in the presence of divalent metals (Mg^{2+} or Ca^{2+}), none was observed in either structure, even at the junction.[7,8] A single solvent molecule was observed bridging the O2P oxygen atoms of A6 at the junction (strands B and D) in both structures. This electron density peak, present in both structures (in the GA structure sitting on a 2-fold crystallographic axis) can be assigned as a sodium ion, although further high-resolution studies are required to elucidate the true identity of this solvent molecule. This cation would help to counter the negatively charged phosphate groups at the junction, and it is buried completely beneath the SAS (Figure 7(d)). The inaccessibility of this ion suggests that it was bound initially to a less compact form of the junction that had a more open cation-binding pocket and, perhaps, facilitated collapse to the fully compact stacked-X junction through favorable electrostatic interactions.

Discussion

From the correlation between each junction structure and similar *B*-DNA sequences, we can discern what distortions of the phosphoribose backbone and the base-pairs across the junction are inherent to the four-stranded complex in the crystal structures and which are sequence effects common to both junctions and their resolved *B*-DNA duplexes. In particular, this comparison distinguishes distortions that are caused by the mismatch in the GA sequence. Most of the conformational features of the two DNA junction structures appear to be characteristic of the sequence-
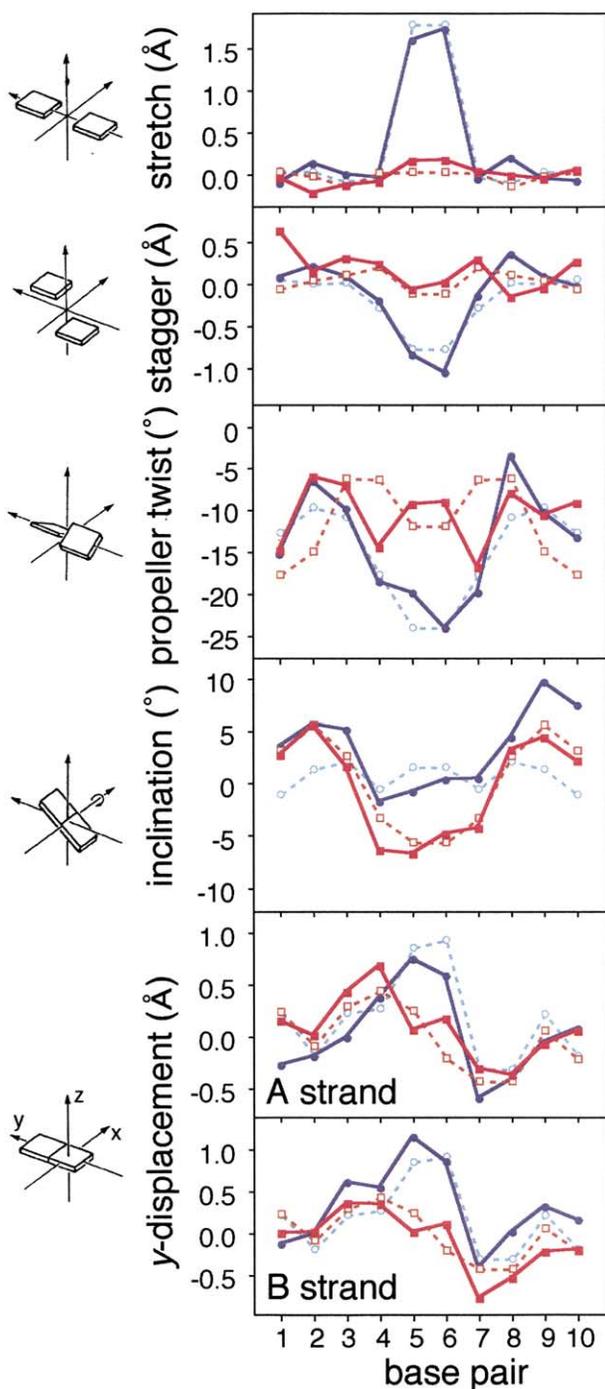
Figure 6. Comparison of local base-pair parameters for the GA and TA junctions, and analogous *B*-DNA duplexes. Stretch, stagger, propeller twist, inclination, and *y*-displacement are shown for the same structures and colored as in Figure 2, with the continuous lines representing the TA (red) and GA (blue) junctions structures, and the broken lines the analogous *B*-DNA duplexes of d(CCAGTACTGG) (orange squares)[15] and d(CCAAGATTGG) (light blue circles).[12] Data were calculated using CURVES 5.2.[17] For the plots of *y*-displacement, the A-strand panel shows the structural parameters for the DNA strand that wraps around the outside of the stacked duplexes, while the B-strand panel is for the strand that crosses over in the junction (Figure 1). For the TA-junction structure, which contains four unique DNA strands, the A-strand parameters are

dependent perturbations of *B*-DNA duplexes. The conserved interactions at the ACC core of the junction and between the ends of the arms suggest strongly that these are the defining characteristics of the overall four-stranded complex, and thus are responsible for fixing the Holliday junction at a particular point within the d(CCGGNACCGG) motif.

The crystal structure of the GA sequence shows that the distortion of the DNA induced by the mismatches (i.e. a high propeller twist) is compatible with the formation of a stacked junction.[7] It has been shown by comparative gel electrophoresis that stacked-X junctions in solution are capable of accommodating d(G·A) mismatches at the crossover, and that in some cases mismatched sequences require increased concentrations of magnesium ion to form the stacked junction.[21] Indeed, the GA crystals were grown in a ninefold greater concentration of divalent cation than the TA crystals, suggesting that formation of a mismatched junction in the crystal is less favorable than a junction consisting of all standard Watson–Crick base-pairs.

The reduced helical twist and strong stacking of the d(GpT/A∗C) or d(GpG/A∗C) base steps (where the asterisk indicates the position of the strand cross-overs) and the unique hydrogen-bonding at the junction and adjacent bases are the dominant interactions that define the DNA structure at the junction. These interactions, as well as the network of hydrophobic contacts at the ACC core, provide an explanation for the ability of these sequences to crystallize as immobile junctions. Panyutin *et al*. have observed that conditions that favor the stacked-X form of the junction decrease the rate of branch migration dramatically.[22] This has led to the theory that disruption to the base-pairing and stacking at the junction *via* the open-X form is necessary for each step of branch migration. In the crystal structures, migration of the junction along the duplex was most likely arrested by the unique ACC sequence. It is interesting to speculate that prior to crystallization, all oligonucleotide sequences at such high concentrations are capable of forming four-stranded structures, and eventually form resolved duplexes due to branch migration past the end in those sequences that lack junction-stabilizing interactions. Indeed, junctions have been observed to form at high concentrations in the PCR products of the p53 gene.[23]

The stacked-X junctions seen in the crystal structures of d(CCGGGACCGG) and d(CCGGTACCGG) have an interduplex angle of ~41°, which is shallower than had been estimated by gel electrophoresis, fluorescence resonance energy transfer (FRET), and atomic force microscopy.[24–26]

averaged between the two outside (A and C) strands of the structure and the B-strand parameters averaged for strands B and D. For the GA junction, strand A is crystallographically identical with C and B is crystallographically identical with D.

Although there are some uncertainties in the estimates from the gel and spectroscopic methods, this discrepancy has led to some discussion as to whether the crystal structures say anything about the geometry of the junction in solution.

The 41° interhelical angle that relates the two stacked duplex arms across both junctions is not a consequence of the crystal lattice. Identical packing has been observed in several crystal structures of DNA duplexes of different lengths and in different space groups, showing that this is a common feature of the DNA molecule (Table 3). The crystal packing in these *B*-DNA duplex and Holliday junction structures involves pseudo-continuous stacked helices that span the length of the crystal and that cross each other in an X, with the backbone of one duplex sitting in the major groove of the adjacent duplex. Timsit *et al.* have shown that this backbone-groove packing is sequence-dependent.[27,28] The interhelical angles of these duplex crystal structures range from 42° to 90° (Table 3). In addition, the brominated derivative of the GA sequence used in the structure determination of the mismatched junction crystallizes in a different space group, and has a geometry that is identical with that of the native TA and GA structures. Finally, the recently determined structure of the TA junction in the smaller GA-crystal lattice showed that the overall geometries are not determined by whether the structures are absolutely symmetric across the junction.[29] This further suggests that the DNA sequence, and not the crystal lattice, determines the overall structure of the four-stranded complex. Finally, the interhelical angle of constructs that contain the d(CCGGTA-CCGG) sequence[30] has been determined recently by atomic force microscopy to be ∼43°, as opposed to the larger 60° determined by the same method for non-symmetric sequences.[26] In addition, the studies show that the junction cross-overs are located at the same sequence position as seen in the crystal structures. This emphasizes the sequence-dependence in the geometry as well as the detailed conformational features of Holliday junctions.

It now appears that interactions at the ends of the arms and at the ACC core define the 41° orientation of adjacent helices. A 60° angle measured in solution[24–26] would be too wide to accommodate the specific interduplex contacts observed in the DNA crystal structures, given the specific helical parameters of the sequences. The overwinding of the 4 bp arms and underwinding of the 6 bp arms at the base steps flanking the junction, the largest perturbations of the DNA common to both structures, act in concert to preserve the interhelical contacts. A clockwise rotation of one arm results in a counter-clockwise rotation of the other, much like two connected gears rotating in opposite directions. Thus, the interactions between the arms are strong enough to effect the rotation of one helix from the other. In addition, it is evident that elimination of one of the arms results in increased variability in the conformation of the junction.

When one of the arms consists of only one base-pair, as in the crystal structures of four-way junctions formed from RNA and DNA strands,[6,9] the angle between the arms can vary from 55° to −80°. Thus, the sequence-specific interactions between the arms that are further removed from the junction seem to play a special role in defining the geometric features of the junctions in the crystal. We suggested previously that the 60° angle can be accommodated by changing the average twist of the *B*-DNA arms from ∼10 bp/turn in the crystals to ∼10.5 bp/turn typically seen in solution.[31]

It is surprising that no polyvalent cation was specifically located in either structure. The floor of the interduplex cavity directly at the strand cross-overs is the most likely binding site for cations because of the high charge potential and proximity to potential ligands. Most of the conserved water molecules reside in this cavity, and are more than likely partially occupied cations. A single well-ordered Co^{3+} was observed in the crystal structure of the 55° RNA−DNA junction, in a position identical with that of the bridging water molecule observed between G3 and the phosphate group of A6 in the DNA junctions.[6] Presumably, the RNA−DNA junction is able to accommodate the larger cobalt ion because of the wider angle between the adjacent arms, although the Co^{3+}−DNA coordination bonds in that structure were longer than the expected 2 Å. Crystals of d(CCGGTACCGG) could not be obtained in the presence of Co^{3+}, and soaking very small concentrations of Co^{3+} into existing crystals resulted in very fragile or cracked crystals (B.F.E., unpublished results), suggesting that the cobalt distorted the DNA structure severely. Therefore, the compact junction formed as a result of the interactions between the arms does not seem to provide an optimum binding site for polyvalent cations.

It is interesting that, in the absence of Co^{3+}, no Mg^{2+} was observed in the RNA−DNA structure.[6] These data suggest that magnesium and cobalt have different binding sites at the junction. Indeed, a magnesium ion was bound in the major groove of the stacked base-pairs at the strand cross-over in the −80° DNA−RNA junction structure. Very few solvent molecules were observed in this region of the DNA junctions, suggesting that a diffuse, poorly ordered cloud of counterions could be present to help counterbalance the negative electrostatic potentials and thereby stabilize these structures. An accumulation of cations at or near the junction is expected to be higher than one would expect for an equivalent set of phosphate charges for *B*-DNA duplexes, which would help to stabilize the four-stranded complex.[32] This, however, may not result in an ordered array of complexes that can be distinguished in the current crystal structures. On the other hand, the solvent molecules that have been identified in the minor groove of *B*-DNA duplexes at medium resolution have recently been reassigned as monovalent
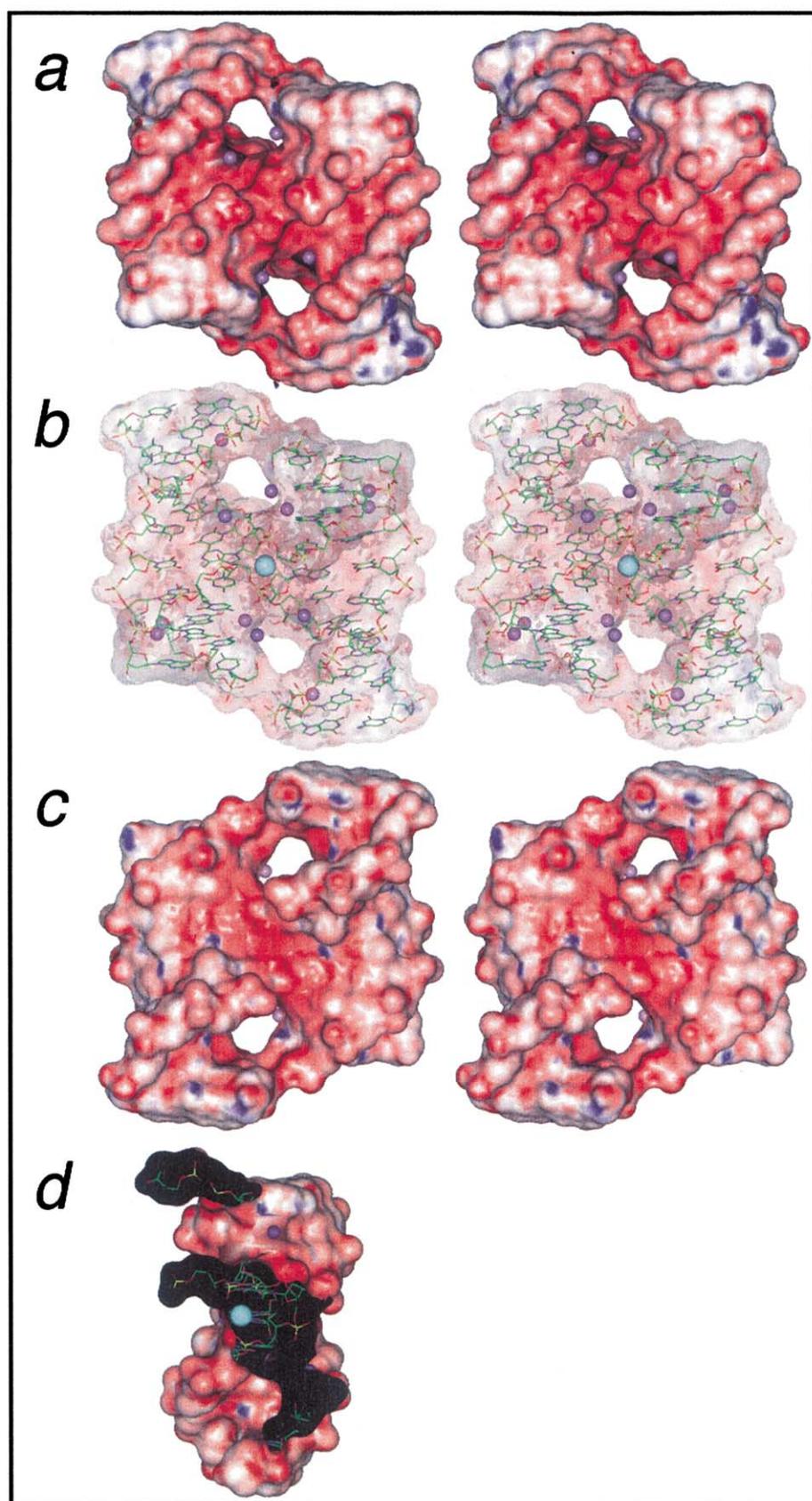
Figure 7. Solvent-accessible surface and electrostatic potential of d(CCGGTACCGG). Surfaces and potentials (red negative and blue positive) were calculated using the Delphi module of the program InsightII (Molecular Simulation, Inc., San Diego, CA). Conserved water molecules (purple spheres) between the two junction structures and the sodium ion (blue sphere) at the center of the junction are shown. (a) Stereoview into the minor groove face showing the cavity formed between the minor groove of the 6 bp arm and the major groove of the 4 bp arm. (b) Location of solvent on the
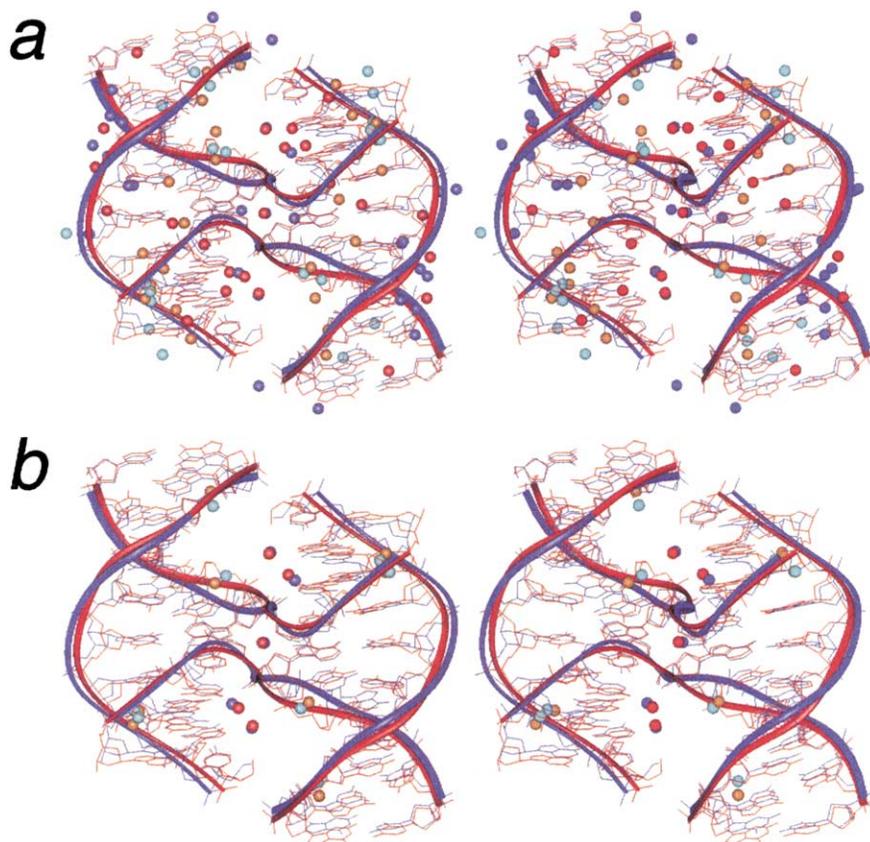
Figure 8. Hydration patterns in d(CCGGTACCGG) (red atoms) and d(CCGGGACCGG) (blue atoms). (a) First-shell water molecules (spheres) are colored according to the DNA structure and to their positions in the major and minor groove. Water molecules in the major (red) and minor (orange) grooves of d(CCGGTACCGG) are compared to those in the major (dark blue) and minor (light blue) grooves of d(CCGGGACCGG). Phosphate group positions along each DNA strand are traced with a ribbon. (b) Water molecules that are common to both the TA and GA junctions. Water molecules that correspond with each other (within 1 Å) between the two structures are shown, and colored as in (a).
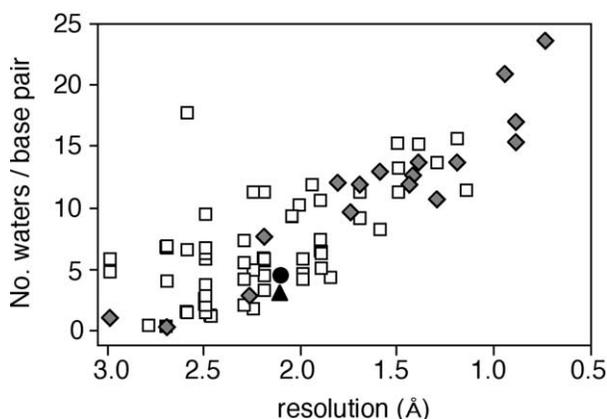
Figure 9. Number of water molecules located in the TA and GA junction and in *B*-DNA single crystal structures as a function of resolution. Data are shown for all *B*-DNA structures in the Nucleic Acid Database,[35] and were normalized as the number of water molecule s per base-pair in each structure. Data for the d(CCGGTA-CCGG) (filled circle) and d(CCGGGACCGG) (filled triangle) junctions are compared with all *B*-DNA duplex structures (squares). Shaded diamond symbols correspond to *B*-DNA structures cross-validated with the R_{free} calculation,[36] while those assigned without cross-validation are shown as open squares.

cations in high-resolution structures.[33] For the stacked-X junction in the crystals, the gap between adjacent neighboring arms is expected to have very highly negative electrostatic potentials and, thus, many of the solvent molecules modeled in this space may ultimately prove to be cations. Indeed, a hexaaquo-calcium(II) complex was observed in the minor groove, but diametrically opposed to the junction cross-over in the recently determined high-resolution (1.5 Å) structure of the methylated sequence d(CCGGTACm5CGG).[34]

surface of the junctions viewed into the minor groove, as in (a). The transparent surface is overlaid over the DNA atoms and solvent molecules. (c) Stereoview into the major groove face of the junction. (d) Cross-section at the junction, revealing the sodium ion buried beneath the solvent-accessible surface. The molecule is rotated 90° with respect to the views in (a)–(c), and cut-away to show only half the junction.

References

1. Lilley, D. M. J. (2000). Structures of helical junctions in nucleic acids. *Quart. Rev. Biophys.* **33**, 109−159.
2. Xia, F., Taghian, D. G., DeFrank, J. S., Zeng, Z.-C., Willers, H., Iliakis, G. & Powell, S. N. (2001). Deficiency of human BRCA2 leads to impaired homologous recombination but maintains normal nonhomologous end joining. *Proc. Natl Acad. Sci. USA*, **98**, 8644−8649.
3. Karow, J. K., Constantinou, A., Li, J.-L., West, S. C. & Hickson, I. D. (2000). The Bloom's syndrome gene product promotes branch migration of Holliday junctions. *Proc. Natl Acad. Sci. USA*, **97**, 6504−6508.
4. Holliday, R. (1974). Molecular aspects of genetic exchange and gene conversion. *Genetics*, **78**, 273−287.
5. Lilley, D. M. J. (1999). *Oxford Handbook of Nucleic Acid Structure* (Neidle, S., ed.), pp. 471−498, Oxford University Press, New York.
6. Nowakowski, J., Shim, P. J., Prasad, G. S., Stout, C. D. & Joyce, G. F. (1999). Crystal structure of an 82-nucleotide RNA−DNA complex formed by the 10-23 DNA enzyme. *Nature Struct. Biol.* **6**, 151−156.
7. Ortiz-Lombardía, M., González, A., Eritja, R., Aymamí, J., Azorín, F. & Coll, M. (1999). Crystal structure of a DNA Holliday junction. *Nature Struct. Biol.* **6**, 913−917.
8. Eichman, B. F., Vargason, J. M., Mooers, B. H. M. & Ho, P. S. (2000). The Holliday junction in an inverted repeat sequence: sequence effects on the structure of four-way junctions. *Proc. Natl Acad. Sci. USA*, **97**, 3971−3976.
9. Nowakowski, J., Shim, P. J., Stout, C. D. & Joyce, G. F. (2000). Alternative conformations of a nucleic acid four-way junction. *J. Mol. Biol.* **300**, 93−102.
10. Eichman, B. F., Mooers, B. H. M., Alberti, M., Hearst, J. E. & Ho, P. S. (2001). The crystal structures of psoralen cross-linked DNAs: drug dependent formation of Holliday junctions. *J. Mol. Biol.* **308**, 15−26.
11. Thorpe, J. H., Teixeira, S. C. M., Gale, B. C. & Cardin, C. J. (2002). Structural characterization of a new crystal form of the four-way Holliday junction formed by the DNA sequence d(CCGGTACCGG)₂: sequence versus lattice? *Acta Crystallog. sect. D*, **58**, 567−569.
12. Privé, G. G., Heinemann, U., Chandrasegaran, S., Kan, L.-S., Kopka, M. L. & Dickerson, R. E. (1987). Helix geometry, hydration, and G·A mismatch in a B-DNA decamer. *Science*, **238**, 498−504.
13. Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallog. sect. D*, **55**, 247−255.
14. Brünger, A. T. (1992). *X-PLOR Version 3.1: A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT.
15. Kielkopf, C. L., Ding, S., Kuhn, P. & Rees, D. C. (2000). Conformational flexibility of B-DNA at 0.74 Å Resolution: d(CCAGTACTGG)₂. *J. Mol. Biol.* **296**, 787−801.
16. Heinemann, U. & Alings, C. (1989). Crystallographic study of one turn of G/C-rich B-DNA. *J. Mol. Biol.* **210**, 369−381.
17. Lavery, R. & Sklenar, H. (1989). Defining the structure of irregular nucleic acids: conventions and principles. *J. Biomol. Struct. Dynam.* **6**, 655−667.
18. Lu, X.-J., Shakked, Z. & Olson, W. K. (2000). A-form conventional motifs in ligand-bound DNA structures. *J. Mol. Biol.* **300**, 819−840.
19. Dickerson, R. E. (1999). *Oxford Handbook of Nucleic Acid Structure* (Neidle, S., ed.), pp. 145−197, Oxford University Press, New York.
20. Berman, H. M. & Schneider, B. (1999). *Oxford Handbook of Nucleic Acid Structure* (Neidle, S., ed.), pp. 295−312, Oxford University Press, New York.
21. Duckett, D. R. & Lilley, D. M. J. (1991). Effects of base mismatches on the structure of the four-way DNA junction. *J. Mol. Biol.* **221**, 147−161.
22. Panyutin, I. G., Biswas, I. & Hsieh, P. (1995). A pivotal role for the structure of the Holliday junction in DNA branch migration. *EMBO J.* **14**, 1819−1826.
23. Yakubovskaya, M., Neschastnova, A., Humphrey, K., Babon, J., Popenko, V., Smith, M. Lambrinakos, A., Lipatova, Z., Dobrovolskaia, M., Cappai, R., Masters, C., Belitsky, G., & Cotton, R. (2001). Interaction of linear homologous DNA duplexes via Holliday junction formation. *Eur. J. Biochem.* **268**, 7−14.
24. Murchie, A. I. H., Clegg, R. M., von Kitzing, E., Duckett, D. R., Diekmann, S. & 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.
25. Cooper, J. P. & Hagerman, P. J. (1989). Geometry of a branched DNA structure in solution. *Proc. Natl Acad. Sci. USA*, **86**, 7336−7340.
26. Mao, C., Sun, W. & Seeman, N. C. (1999). Designed two-dimensional DNA Holliday junction arrays visualized by atomic force microscopy. *J. Am. Chem. Soc.* **121**, 5437−5443.
27. Timsit, Y. & Moras, D. (1991). Groove-backbone interaction in B-DNA. Implication for DNA condensation and recombination. *J. Mol. Biol.* **221**, 919−940.
28. Timsit, Y. & Moras, D. (1994). DNA self-fitting: the double helix directs the geometry of its supramolecular assembly. *EMBO J.* **13**, 2737−2746.
29. Thorpe, J. H., Teixeira, S. C., Gale, B. C. & Cardin, C. J. (2002). Structural characterization of a new crystal form of the four-way Holliday junction formed by the DNA sequence d(CCGGTACCGG)₂: sequence versus lattice? *Acta Crystallog. sect. D*, **58**, 567−569.
30. Sha, R., Liu, F. & Seeman, N. C. (2002). Atomic force microscopic measurement of the interdomain angle in symmetric Holliday junctions. *Biochemistry*, **41**, 5950−5955.
31. Ho, P. S. & Eichman, B. F. (2001). The crystal structures of DNA Holliday junctions. *Curr. Opin. Struct. Biol.* **11**, 302−308.
32. Olmsted, M. C. & Hagerman, P. J. (1994). Excess conterion accumulation around branched nucleic acids. *J. Mol. Biol.* **243**, 919−929.
33. Shui, X., McFail-Isom, L., Hu, G. G. & Williams, L. D. (1998). The B-DNA dodecamer at high resolution reveals a spine of water on sodium. *Biochemistry*, **37**, 8341−8355.
34. Vargason, J. M. & Ho, P. S. (2002). The effect of cytosine methylation on the structure and geometry of the Holliday junction. The structure of d(CCGGTACm5CGG) at 1.5 Å resolution. *J. Biol. Chem.* **277**, 21041−21049.
35. Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T. Hsieh, S.-H., Srinivasan, A. R. & Schneider, B. (1992). The nucleic acid database. A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.* **63**, 751−759.

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

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

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

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

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

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

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

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

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

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

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

47. Wood, A. A., Nunn, C. M., Trent, J. O. & Neidle, S. (1997). Sequence-dependent crossed helix packing in the crystal structure of a B-DNA decamer yields a detailed model for the Holliday junction. *J. Mol. Biol.* **269**, 827–841.

48. Goodsell, D. S., Grzeskowiak, K. & Dickerson, R. E. (1995). Crystal structure of C-T-C-T-C-G-A-G-A-G. Implications for the structure of the Holliday junction. *Biochemistry*, **34**, 1022–1029.

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

Edited by P. J. Hagerman