Crystal structure of a DNA Holliday junction

Miguel Ortiz-Lombardía[1,2], Ana González[3], Ramón Eritja[4], Joan Aymamí[1,5], Fernando Azorín[1] and Miquel Coll[1]

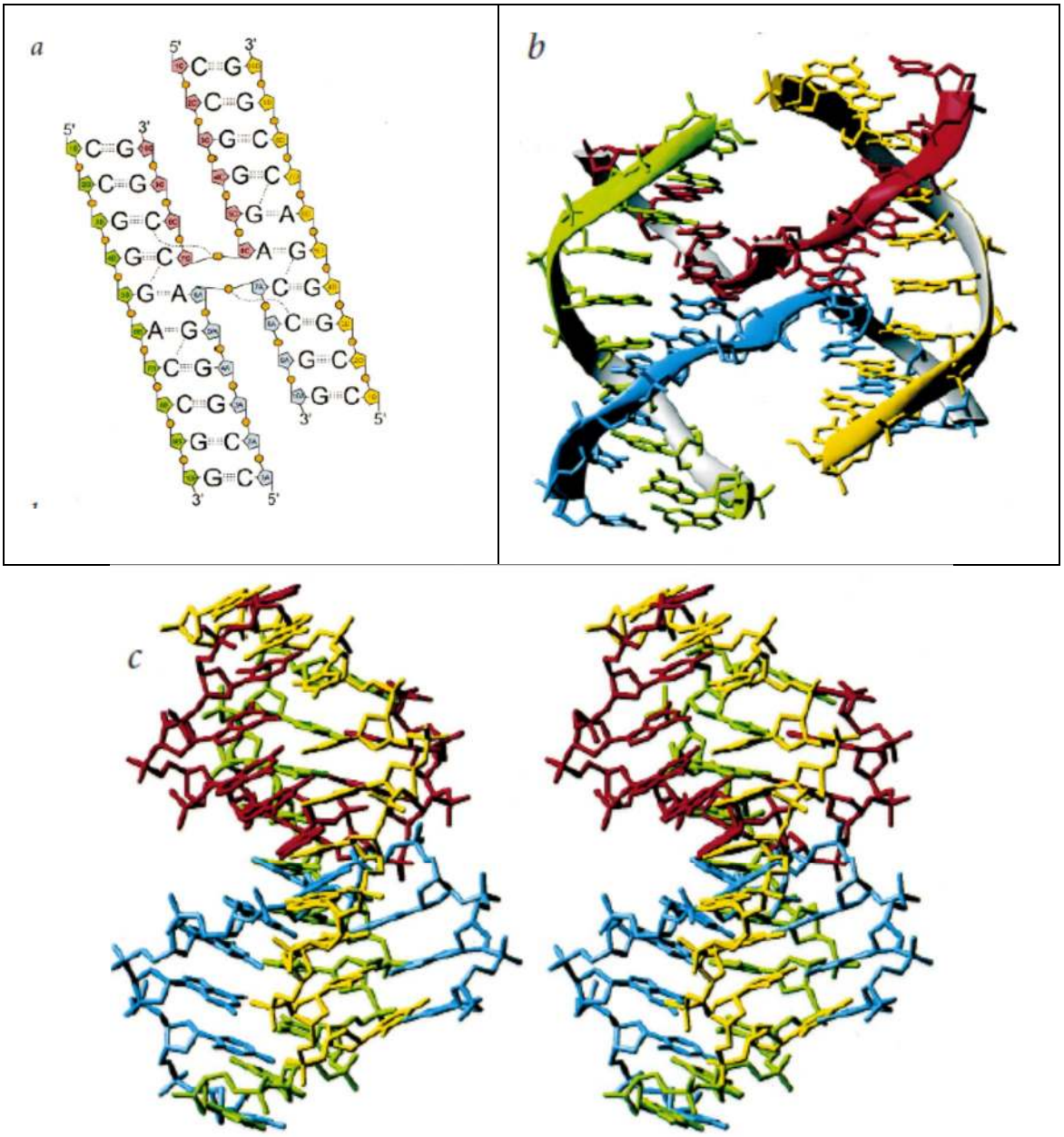
[1]Institut de Biologia Molecular de Barcelona, C.S.I.C., Jordi Girona 18, E- 08034 Barcelona, Spain. [2]Present address: Institut Pasteur, 28 rue du Dr. Roux, 75724 Paris Cedex 15, France. [3]EMBL, c/o DESY, Notkestraβe 85, D- 22603 Hamburg, Germany. [4]EMBL, Meyerhofstraβe 1, D-69117 Heidelberg, Germany. [5]Departament d'Enginyeria Química, Universitat Politècnica de Catalunya, Diagonal 647, E-08028 Barcelona, Spain.

DNA recombination is a universal biological event responsible both for the generation of genetic diversity and for the maintenance of genome integrity. A four-way DNA junction, also termed Holliday junction, is the key intermediate in nearly all recombination processes. This junction is the substrate of recombination enzymes that promote branch migration or catalyze its resolution. We have determined the crystal structure of a four way DNA junction by multiwavelength anomalous diffraction, and refined it to 2.16 Å resolution. The structure has two-fold symmetry, with pairwise stacking of the double-helical arms, which form two continuous B-DNA helices that run antiparallel, cross in a right-handed way, and contain two G-A mismatches. The exchanging backbones form a compact structure with strong van der Waals contacts and hydrogen bonds, implying that a conformational change must occur for the junction to branch-migrate or isomerize. At the branch point, two phosphate groups from one helix occupy the major groove of the other one, establishing sequence-specific hydrogen bonds. These interactions, together with different stacking energies and steric hindrances, explain the preference for a particular junction stacked conformer.

Since a four-way DNA junction was first proposed in 1964 as the central intermediate in homologous genetic recombination[1], a large amount of data, obtained with a range of techniques including electron microscopy[2], gel electrophoresis[3,4], fluorescence resonance energy transfer[5,6], NMR[7,8], chemical or enzyme probing[3,9ま11] and molecular modeling[12ま16], has revealed many of the characteristics of the junction (for reviews see refs 17,18). For example, in the absence of cations the junction is believed to be extended, with its arms unstacked, and to have a four-fold planar conformation. However, in the presence of metal ions (as occurs in physiological conditions), the arms stack in pairs and form a two-fold nonplanar junction, known as the X-stacked model. Recently, three proteinまDNA complex structures have provided an image of the four-way DNA junction when bound to two recombination proteins[19ま21]. The junction conformations of these two complexes differ, having a four-fold symmetry in the RuvA complex and a two-fold symmetry in the Cre complex. However, in both structures, the arms are unstacked and lie almost in the same plane. Despite the efforts of several laboratories, no unliganded four-way DNA junction has been crystallized, and NMR data have proved insufficient to determine its three-dimensional structure fully. Consequently, the precise details of this biologically significant DNA structure remain unknown. To date, the only crystallographic report concerns an 82-nucleotide RNAまDNA complex containing an RNAま

DNA four-way junction[22]. Yet, in that hybrid structure, two of the stems are in the A conformation and one of these is limited to one base pair and a loop. We have crystallized and determined the structure of a four-way junction at high resolution, using the DNA decamer d(CCGGGACCGG) (Figs 1*a*, 2). The 5-Br-cytosine derivative d(C^{Br}CGGGACCGG) has been crystallized in a different space group, yet it forms a similar four-way junction (Fig. 2*a*).

Fig. 1 Global structure of the Holliday junction. ***a***, Scheme of the junction with nucleotide numbering. Phosphate groups are represented as orange circles. Hydrogen bonds are indicated by dashed lines. Nonexchanging strands are in green and yellow, and exchanging strands are in blue and red. The four strands have been labeled A, B, C and D for clarity, although the asymmetric unit in the crystal contains only strand A and strand B, C and D being related by symmetry to them. ***b***, Ribbon diagram of the Holliday junction, as seen from its major groove side. ***c***, Stereo lateral view. The strand color code in (*b*) and (*c*) is the same as in (*a*).



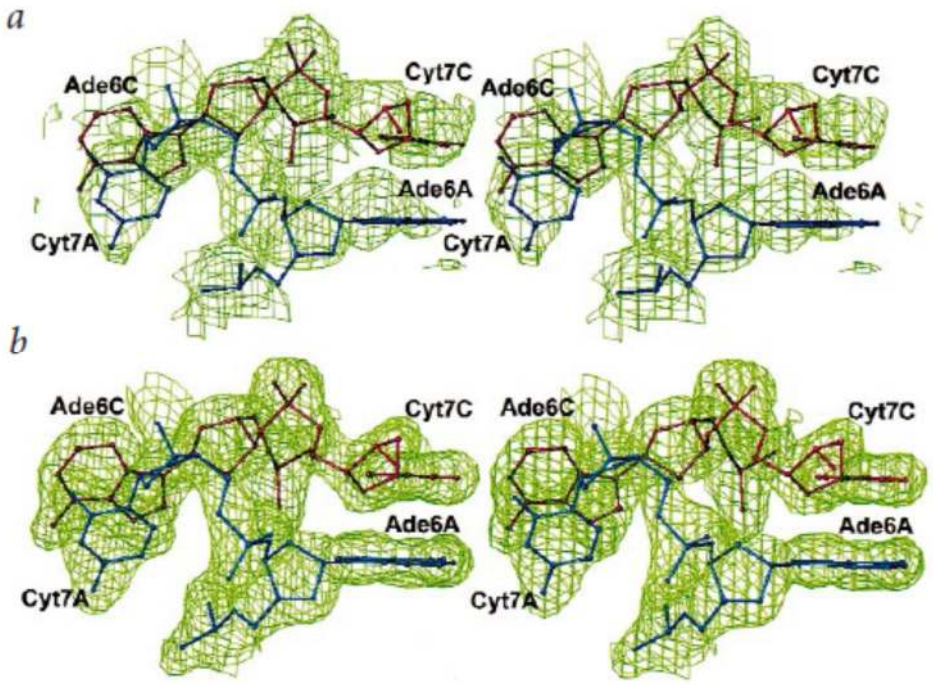


Fig. 2 Electron density maps. Stereo diagrams showing the electron density maps at the exchanging point of the junction and the final atomic model with the two exchanging strands colored as in Fig. 1. ***a***, Initial experimental MAD electron density map calculated at 2.7 Å resolution. ***b***, Final σA-weighted 2Fo - Fc map after refinement at 2.16 Å resolution. Both maps have been contoured at the 1σ level.

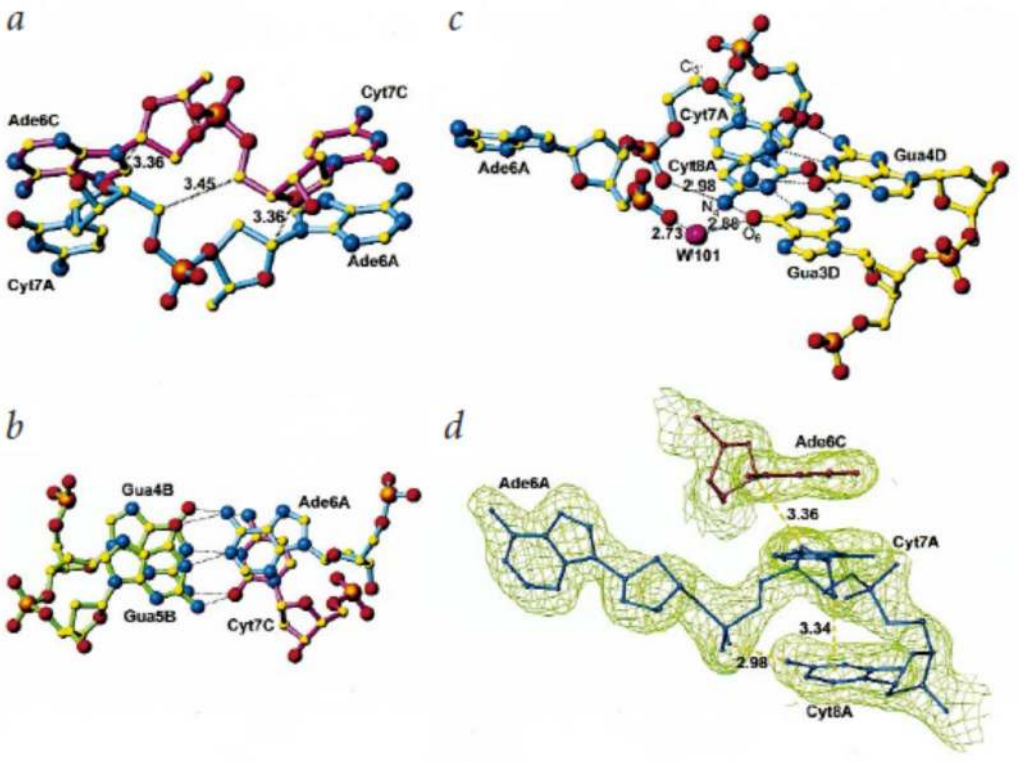


Fig. 3 Structural details of the four-way junction. ***a***, Detail of the exchanging backbones at the junction. One C5'-C5' van der Waals contact and the two O4'-C1' hydrogen bonds are shown with dotted lines, and the distances are indicated. ***b***, Base pair overlap at the stacked arms. The twist is 10° lower than in regular DNA allowing an extensive stacking of the two base pairs. ***c***,

Detail of phosphates 6 and 7 of one helix that occupy the major groove of the adjacent helix and interact with the second base pair after the junction by a direct hydrogen bond and a water bridge. ***d***, Final σA-weighted 2Fo 未Fc electron density map, contoured at 1σ, with the final atomic model showing one of the exchanging backbones as it breaks the helical path (represented here by Ade6C, Cyt7A, Cyt8A) and crosses to the adjacent double helix. The unique orientation of deoxyribose 7A is visible, with the O4'7A-C1'6C hydrogen bond indicated as a dashed line. The interaction of the C2' group with cytosine 8A and the hydrogen bond between the oxygen atom of phosphate of Cyt7A and the N4 of Cyt8A are also indicated. Strand color codes are as in Fig. 1.

A planar scheme of the structure (Fig. 1*a*), with its base pairing arrangement and numbering, reveals that the four-way junction is formed by four chemically identical copies of the d(CCGGGACCGG) oligonucleotide. In the plane, they outline an H-shaped junction, in which two of the four arms are six base pairs long and the other two are four base pairs long. As can be seen in diagrams of the three-dimensional structure of the junction (Fig. 1*b*,*c*), the junction has two-fold, rather than four-fold symmetry, with a two-fold crystallographic axis running through the exchanging point, perpendicular to the plane of the paper (Fig. 1*b*). The structure defines two kinds of DNA strands: two (red and blue) are exchanging strands, and two (green and yellow) are nonexchanging strands. The four double-stranded arms of the junction are stacked pairwise, forming two continuous helices, and therefore the structure closely corresponds to the X-stacked model3,9. The two double helices are not in plane but form an angle of 40°, which is slightly smaller than the 60° predicted by the model. The helices cross each other in a righthanded way and are almost antiparallel (Fig. 1*c*).

The two DNA double helices are in the B-form and straight, with a calculated curvature of only 3°. It is remarkable that the distortion from regular B-DNA caused by the junction affects only the four base pairs involved (two on each side) and does not extend any farther. The distortion consists mainly in a reduced helical twist of 25° and a peculiar orientation of the sugar moieties of residues 7A and 7C (Fig. 1*a*), whose phosphate groups lie at the branch point (Fig. 3*a*). The glycosyl χ torsion angle is -152° at these residues, and the orientation of the sugar is such that the C5' atom points to the outside of the helix cylinder, allowing the backbone to protrude from the helical path (Fig. 3*d*). Therefore, the critical conformational change responsible for the sharp switch in the direction of the sugar-phosphate backbone is the sugar-to-base orientation, rather than the sugar pucker, which remains in the B-DNA C2'-endo conformation. The sugar-phosphate backbone torsion angles also differ from those in canonical B-DNA at the exchanging strands, notably ε6 = -73.6° (~180° in BI-DNA), in agreement with earlier preditions13.

The area of exchange is reduced to two sugars and two phosphate groups of residues 7A and 7C. These two exchanging backbone segments run tightly antiparallel at the junction, with strong van der Waals contacts between them at the C5' sugar carbon (Fig. 3*a*). The unique orientation of the deoxyriboses of 7A and 7C places their O4' atom at a distance of 3.36 Å from the C1' atom of sugars 6C and 6A, respectively (Fig. 3*a*,*d*). These two interactions can be

considered O···H-C hydrogen bonds23. The orientation of sugars 7A and 7C allows further interaction of their C2' hydrogens with the heterocyclic ring of cytosines 8A and 8C (Fig. 3*d*).

The two phosphates of the junction, 7A and 7C, are oriented away from each other (Fig. 3*a*), avoiding their electrostatic repulsion, as predicted by the von Kitzing model13. The distance between those phosphates is 6.8 Å. They occupy the major groove of the adjacent double strand, and interact with the second base pair beyond the exchanging point (Figs 1*a*, 3*c*). The interaction consists of a hydrogen bond with the N4 group of Cyt8A and Cyt8C, respectively. This interaction is base specific, as it can only occur with the N4 group of cytosines or the N6 group of adenines. Phosphates 6A and 6C also occupy the major grooves of the adjacent double strands. In this case no direct hydrogen bonds with the bases are observed. Instead, a water bridge connects these phosphate groups with the O6 atoms of guanines 3D and 3B, the two bases paired with cytosines 8A and 8C, respectively.

There is base pair overlap of the stacked arms (Fig. 3*b*). Full stacking is observed between the two bases of the exchanging strands, adenine 6A and cytosine 7C, indicating that this interaction is essential to the stability of the junction structure.

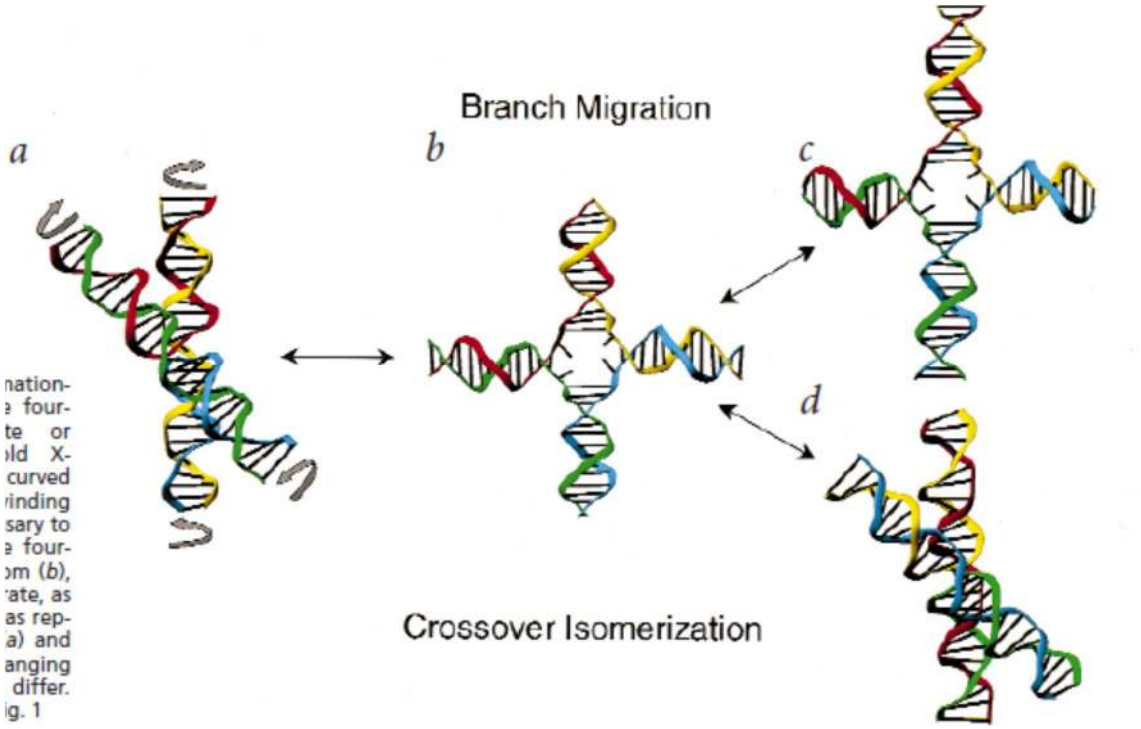


Fig. 4 Scheme of the conformational changes required for the fourway junction to migrate or isomerize. ***a***, The two-fold Xstacked structure with curved arrows indicating the unwinding movement of the arms necessary to open the junction to ***b***, the fourfold unstacked structure. From (*b*), the junction can branch-migrate, as indicated in ***c***, or isomerize, as represented in ***d***. Structurally (*a*) and (*d*) are identical, but the exchanging and nonexchanging strands differ.Strand color codes are as in Fig. 1

The junction does not cause a widening of the minor groove, as predicted in some models13. On the contrary, the minor groove is slightly narrower than in regular B-DNA immediately beyond the junction, with a distance between phosphates 8A and 8B of 10.9 Å. Nevertheless, the major groove widens out at the junction, with a distance between phosphates 6A and 2B of 19.1 Å. It is in precisely that area of the major groove that the phosphates 6C and 7C of the adjacent helix are located.

There are several well-defined water molecules in both the minor and major grooves. However, no metal ion could be clearly assigned from the electron density maps, even though the X-stacked junction is known to be stabilized by divalent cations24 and the crystals analyzed here were obtained in the presence of Mg2+. A possible ion location coordinating phosphates 6A and 6C was investigated; however, the electron density corresponds better to a water molecule, and no other ligands for a typical Mg2+ octahedral coordination are present.

Table 1 Data collection, phasing and refinement statistics

Data collection and phasing statistics					
	Native	^{5Br}C2 derivative			
λ (Å)	1.5418	0.9220	1.2500	0.9224	0.8856
Space group	C2	$C222_1$			
Cell parameters	64.20 Å	23.75 Å			
	23.74 Å	63.90 Å			
	38.30 Å	71.40 Å			
	β = 112.43°				
Resolution range (Å)	19.39–2.16	19.07–2.71	23.84–2.68	19.07–2.71	19.07–2.71
Unique reflections	2,786	1,403	1,449	1,421	1,423
Completeness (%)[1]	92.0 (75.5)	87.7 (72.5)	87.7 (82.0)	89.4 (80.7)	90.0 (82.0)
Multiplicity[1]	2.1 (1.7)	3.6 (3.3)	3.7 (4.0)	5.7 (5.1)	6.5 (5.9)
$\langle I\rangle/\langle\sigma I\rangle$[1]	10.4 (5.2)	6.3 (2.5)	6.4 (3.1)	5.9 (2.5)	5.9 (2.6)
R_{merge}[1,2]	0.069 (0.137)	0.090 (0.285)	0.088 (0.245)	0.099 (0.293)	0.101 (0.280)
R_{merge}anomalous[3]		0.055	0.051	0.051	0.053
Anom. data comp. (%)[4]		87.8	89.4	91.1	90.6
R_{cullis} (acentric/centric)[5]		0.88/0.84	0.80/0.74	–	0.86/0.81
Phasing power (acentric/centric)[6]		0.79/0.64	1.27/0.92	–	1.02/0.76
Refinement statistics					
Reflections	2,786				
DNA atoms	408				
Water molecules	30				
R factor[7]	0.240				
R_{free}[8]	0.288				
R.m.s.d. bond lengths (Å)	0.008				
R.m.s.d. bond angles (°)	1.7				

[1]Values in parentheses correspond to the outermost resolution shell
[2]$R_{merge} = \Sigma_{hkl}\Sigma_i|I_i(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl}\Sigma_i I_i(hkl)$, calculated for the whole data set.
[3]R_{merge}anomalous $= \Sigma_{hkl}\Sigma_i|I_i(hkl\pm) - \langle I(hkl)\rangle| / \Sigma_{hkl}\Sigma_i I_i(hkl\pm)$.
[4]Percentage of reflections with a Bijvoet pair.
[5]R_{cullis} is the mean residual lack of closure error divided by the dispersive difference.
[6]Phasing power = r.m.s.($|F_H|$ / E), where F_H is the heavy atom amplitude and E is the residual lack of closure error.
[7]R factor $= \Sigma_{hkl}||F_o(hkl)| - k|F_c(hkl)|| / \Sigma_{hkl}|F_o(hkl)|$.
[8]R factor for 10% of reflections not used in the refinement.

The exchanging region is flanked, in two of the four arms, by four G-A mismatches (Fig. 1). DNA duplexes can easily accommodate G-A mismatches with only minimal distortions, which is the case in the present structure. The conformation of the mispairs (both in the *anti* orientation) and their hydrogen bonding scheme are identical to those observed in similar sequences when found in double-helical structures[25], including a high propeller twist and three-centered hydrogen bonds. Two of these three centered hydrogen bonds occur at the

stacked arms and seem to stabilize the junction, but their contribution to the formation or immobilization of the junction is difficult to assess since a similar interaction would occur if this junction were to be resolved in two duplexes. It is notable, however, that previous crystallization attempts with designed immobile junctions having all Watson–Crick base pairs were unsuccessful. It is believed that non-Watson–Crick base pairing occurs when the recombining sequences are not strictly homologous. Our results confirm that G-A mismatches, and likely other types of base mispairs[26], are compatible with the four-way junction structure.

Three functional consequences derive from the structure of area is very compact, with strong van der Waals, hydrogen bonding and stacking interactions, branch migration, which implies the breakage of the Watson–Crick hydrogen bonding and a concomitant swing movement of two base pairs, is not possible unless the structure unfolds. As other evidence has shown[24], the junction must undergo a movement to an unstacked structure, close to a four-fold type, in order to migrate (Fig. 4). Proteins that promote branch migration must do so primarily by opening the X-stacked structure, as shown in the RuvA–DNA complexes[19,20], where the junction is planar, unstacked and has four-fold symmetry.

The same argument is valid for the conformational change known as crossover isomerization (Fig. 4). In the two possible conformers, the exchanging and nonexchanging strands differ, and going from one conformer to the other involves complete rearrangement of van der Waals contacts, hydrogen bonds and stacking interactions at the branch point. The flattening or opening of the junction is concomitant with the unstacking of the arms. From the present structure, the most likely movement leading to the opening of the junction is the rotation of the stacked arms in opposite directions, with further unwinding of the continuous helix at the stacking point, where the helix is already 10° unwound (Fig. 4*a*).

Second, the local DNA sequence at the junction determines which of the two conformers is more stable[6,10], and this might determine which strand is to be exchanged[21]. The four base pairs at the branch point are clearly critical in that their stacking differs in each conformer. At the same time, the second base pair beyond that point is also important because of the specific phosphate–base hydrogen bonds described above. The observed isomer is not the one that maximizes the number of purines in the exchanging strands, which is in disagreement with other observations[13,27]. In the present sequence the alternative conformer (Fig. 4*d*) would have a guanine residue at the positions now occupied by Cyt8, and therefore the direct hydrogen bond from N4 to phosphate 7 would not be feasible, since, in the guanine, the spatially corresponding group would be O6. Other sequence effects can be inferred from the present structure. For example, an adenine at position 8 would allow the observed hydrogen bond from its N6 nitrogen to phosphate 7. On the other hand, a thymine at position 8 would tend to destabilize the structure because steric clashes would occur between the C5 methyl group of the base and the phosphate group of residue 7. Likewise, a thymine at position 7 would place its C5

methyl carbon at 2.7 Å from phosphate 8. Obviously, small rearrangements might take place, which would avoid these clashes, and additional structural analysis with other sequences would further clarify this issue.

Third, the structure of the four-way DNA junction defines two very different faces, as predicted by the X-stacked model. This can be clearly seen in the lateral view shown in Fig. 1*c*, where the left surface or major groove side differs markedly from the right surface or minor groove side. Proteins involved in recombination and interacting with the junction must recognize either one of the faces[28] or the two faces with diverse recognition elements[21].

Methods

Crystallization and data collection. Thin, platelike crystals of the synthetic d(CCGGGACCGG) oligonucleotide were obtained by the hanging-drop vapor diffusion method by mixing 1 μl of a 1mM DNA solution in water with 2 μl of precipitant solution, containing 10% 2-methyl-2,4-pentanediol (MPD) and 0.2 M $MgCl_2$, and equilibrating against a reservoir of 0.5 ml of 45% MPD. Similar conditions were used for crystallizing the d(CBrCGGGACCGG) derivative. The native crystals belong to space group C2, while the Br derivative crystals belong to space group C2221 (Table 1). In both cases the DNA contents of the asymmetric unit consist of two strands. Native diffraction data were collected using a Rigaku RU200 X-ray generator with a Cu rotating anode and a MarResearch image plate detector. These data were processed with DENZO[29] and SCALEPACK[29] (Table 1). Anomalous diffraction data from the Br derivative were collected at four different wavelengths on the tunable beamline X31 at the EMBL outstation at the Deutsches Elektronensynchrotron (Hamburg). The data were indexed and integrated with MOSFLM[30] and scaled with SCALA[31].

Phasing and refinement. The structure was solved by the multiwavelength anomalous diffraction (MAD) method with a MIR-like approach. The data set collected at a wavelength of 0.9224 Å was used as a pseudo-native. Heavy atom refinement and phasing were carried out with MLPHARE[31] (Table 1). Density modification procedures, namely solvent flattening and histogram matching, were applied with DM31. The experimental MAD electron density map, calculated at 2.7 Å resolution, showed unambiguously the four-way junction arrangement of the DNA backbone and a crystallographic two-fold axis running through the junction (Fig. 2*a*). An atomic model was readily traceable, and after a few refinement cycles with REFMAC[31] this model was used to solve the structure of the native crystals by molecular replacement with AMORE[32]. Refinement of the native structure continued with REFMAC and manual

model building using σ_A-weighted $2F_o$ - F_c and F_o - F_c electron density Fourier maps (Fig. 2*b*). The refinement was monitored by the R_{free} behavior, and all reflections were used in the process, with no low-resolution or σ cutoffs. Bulk solvent correction was applied. The solvent structure was built applying strict criteria, so water molecules were only added when clear peaks appeared in both $2F_o$ - F_c and F_o - F_c maps, had hydrogen bond interactions with the DNA molecule and lowered the R_{free}. The final model, at 2.16 Å resolution, has excellent geometry with no strain. No disordered areas were detected. Table 1 shows the final refinement statistics.

Coordinates. The atomic coordinates have been deposited with the Nucleic Acids Data Base (entry code UD0006).

Acknowledgements

This work was supported by grants from the Ministerio de Educación y Cultura of Spain, the Generalitat de Catalunya, the Centre de Referència en Biotecnologia and the European Union. Synchrotron data collection was supported by the EU program for access to large installations.

Correspondence and requests for materials should be addressed to M.C. *email: mcccri@cid.csic.es*.

1. Holliday, R. *Genet. Res.* **5**, 282–304 (1964).
2. Broker, T.R. *J. Mol. Biol.* **81**, 1–6 (1973).
3. Duckett, D.R. *et al. Cell* **55**, 79–89 (1988).
4. Cooper, J.P. & Hagerman, P.J. *J. Mol. Biol.* **198**, 711–719 (1987).
5. Murchie, A.I.H. *et al. Nature* **341**, 763–766 (1989).
6. Miick, S.M., Fee, R.S., Millar, D.P. & Chazin, W.J. *Proc. Natl. Acad. Sci. USA* **94**, 9080–9084 (1997).
7. Chen, S.M. & Chazin, W.J. *Biochemistry* **33**, 11453–11459 (1994).
8. Overmars, F.J. & Altona, C. *J. Mol. Biol.* **273**, 519–524 (1997).
9. Churchill, M.E.A., Tullius, T.D., Kallenbach, N.R. & Seeman, N.C. *Proc. Natl. Acad. Sci. USA* **85**, 4653–4656 (1988).
10. Grainger, R.J., Murchie, A.I.H. & Lilley, D.M.J. *Biochemistry* **37**, 23–32 (1998).
11. Panyutin, I.G., Biswas, I. & Hsieh, P. *EMBO J.* **14**, 1819–826 (1995).
12. Sigal, N. & Alberts, B. *J. Mol. Biol.* **71**, 789–793 (1972).
13. von Kitzing, E., Lilley, D.M.J. & Dikmann, S. *Nucleic Acids Res.* **18**, 2671–2683 (1990).

14. Timsit, Y. & Moras, D. *J. Mol. Biol.* **221**, 919ま940 (1991).
15. Goodsell, D.S., Grzezkowiak, K. & Dickerson, R.E. *Biochemistry* **34**, 1022ま1029 (1995).
16. Wood, A.A., Nunn, C.M., Trent, J.O. & Neidle, S. *J. Mol. Biol.* **269**, 827ま841 (1997).
17. Seeman, N.C. & Kallenbach, N.R. *Annu. Rev. Biophys. Biomol. Struct.* **23**, 53ま86 (1994).
18. Lilley, D.M.J. In *Nucleic acid structure* (ed. Neidle, S.) 471ま498 (Oxford University Press, Oxford; 1999).
19. Hargreaves, D. *et al*. *Nature Struct. Biol.* **5**, 441ま446 (1998).
20. Roe, S.M. *et al*. *Mol. Cell* **2**, 361ま372 (1998).
21. Gopaul, D.N., Guo, F. & Van Duyne, G.D. *EMBO J.* **17**, 4175ま4187 (1998).
22. Nowakowski, J., Shim, P.J., Prasad, G.S., Stout, C.D. & Joyce, G.F. *Nature Struct. Biol.* **6**, 151ま156 (1999).
23. Taylor, R. & Kennard, O. *J. Am. Chem. Soc.* **104**, 5063ま5070 (1982).
24. Duckett, D.R., Murchie, A.I.H. & Lilley, D.M.J. *EMBO J.* **9**, 583ま590 (1990).
25. Privé, G.G. *et al*. *Science* **238**, 498ま504 (1987).
26. Duckett, D.R. & Lilley, D.M.J. *J. Mol. Biol.* **221**, 147ま161 (1991).
27. Azaro, M.A. & Landy, A. *EMBO J.* **16**, 3744ま3755 (1997).
28. Raaijmakers, H. *et al*. *EMBO J.* **18**, 1447ま1458 (1999).
29. Otwinowski, Z. & Minor, W. *Methods Enzymol.* **276**, 307ま326 (1997).
30. Leslie, A.G.W. In *Crystallographic computing 5: from chemistry to biology* (eds Moras, D., Podjanry, A.D. & Thierry, J.C.) 27ま38 (Oxford University Press, Oxford, United Kingdom; 1991). 31. Collaborative Computational Project, Number 4. *Acta Crystallogr. D* **50**, 760ま763 (1994).
32. Navaza, J. *Acta Crystallogr. A* **50**, 157ま163 (1994).