Volume 15 Number 16 1987 Nucleic Acids Research

Structural features and hydration of a dodecamer duplex containing two C.A mispairs

William N.Hunter, Tom Brown[1] and Olga Kennard*

University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW and [1]Department of Chemistry, University of Edinburgh, West Mains Road, Edinburgh, UK

Received May 19, 1987; Revised and Accepted July 8, 1987

ABSTRACT

X-ray diffraction techniques have been used to characterise the crystal and molecular structure of the deoxyoligomer d(C-G-C-A-A-A-T-T-C-G-C-G) at 2.5Å resolution. The final R factor is 0.19 with the location of 78 solvent molecules. The oligomer crystallises in a B-DNA type conformation with two strands coiled about each other to produce a duplex. This double helix consists of four A.T and six G.C Watson-Crick base pairs and two C.A mispairs. The mismatched base pairs adopt a "wobble" type structure with the cytosine displaced laterally into the major groove, the adenine into the minor groove. We have proposed that the two close contacts observed in the C.A pairing represent two hydrogen bonds one of which results from protonation of adenine. The mispairs are accommodated in the double helix with small adjustments in the conformation of the sugar-phosphate backbone. Details of the backbone conformation, base stacking interactions, thermal parameters and the hydration are now presented and compared with those of the native oligomer d(C-G-C-G-A-A-T-T-C-G-C-G) and with variations of this sequence containing G.T and G.A mispairs.

INTRODUCTION

The accuracy of DNA replication may be reduced by the inclusion of mismatched base pairs in an otherwise standard Watson-Crick double helix (1). The manner in which non-complementary bases can form hydrogen bonded pairs has been the subject of a number of theoretical studies (2,3). More recently, experimental information pertaining to the thermodynamic stability (4-6) and precise geometry of certain mispair combinations has become available (7-12). The combination of this information with biochemical investigations of the recognition and repair of mispairs (13-18) permits a limited understanding of the complex systems that have evolved to control mutation rates.

The molecular structure of the synthetic dodecamer d(C-G-C-A-A-A-T-T-C-G-C-G) was determined at 2.5Å resolution by single crystal X-ray diffraction techniques. In a previous paper we presented the structure of the C.A mispair and the implications of the observed molecular geometry for

the enzymic recognition of certain mismatches (12). This work is a detailed analysis of the crystal and molecular structure of the dodecamer and a comparison with the structure of the parent compound d(C-G-C-G-A-A-T-T-C-G-C-G) determined by Dickerson and colleagues (19-21). Comparisons will also be made with the dodecamer structures d(C-G-C-G-A-A-T-T-T-G-C-G) and d(C-G-C-G-A-A-T-T-A-C-G-C) which contain G.T and G.A mismatches (9,10).

RESULTS AND DISCUSSION

Crystal Data

The oligomer d(C-G-C-A-A-A-T-T-C-G-C-G) crystallised in the orthorhombic space group $P2_12_12_1$ with two independent strands in the asymmetric unit. The unit cell dimensions (a=25.37, b=41.44, c=65.20 Å) indicated isomorphism with the native dodecamer (19). The structure was refined to a final R-factor of 0.19 at a resolution limit of 2.5Å. The analysis included the location of 78 solvent molecules all assigned as oxygen atoms. Experimental details have been reported (12) and coordinates will be deposited with the Brookhaven Protein Databank.

Duplex Structure and Crystal Packing

The dodecamer adopts a double helical structure in the crystal lattice which consists of ten Watson-Crick base pairs and two C.A mispairs. The duplex, shown in Figure 1, is a member of the B-DNA family. There are about ten base pairs per turn with each base pair located astride the helix axis. The bases are labelled C1 to G12 in the 5' to 3' direction on strand one, C13 to G24 in the 5' to 3' direction on strand two. Solvent molecules are labelled W1 through to W78.

The helix axis is directed along c and there are intermolecular interactions between the upper part of one duplex with the lower part of another related by the 2_1 axis parallel to c. There are five intermolecular hydrogen bonds, which link functional groups in the minor grooves of symmetry related duplexes and one of the terminal O-3' atoms. Several solvent molecules act as intermolecular bridges. There are also van der Waals type interactions involving the stacking of a base in one duplex on the furanose ring of a symmetry related duplex. This type of intermolecular interaction occurs between G12 and G24 and the deoxyribose moieties of C3 and C15 of the symmetry related duplex.

The C.A mispairs

Figure 2 shows several possible hydrogen bonding schemes that can be constructed for a C.A base pair with bases in major and minor tautomer forms.

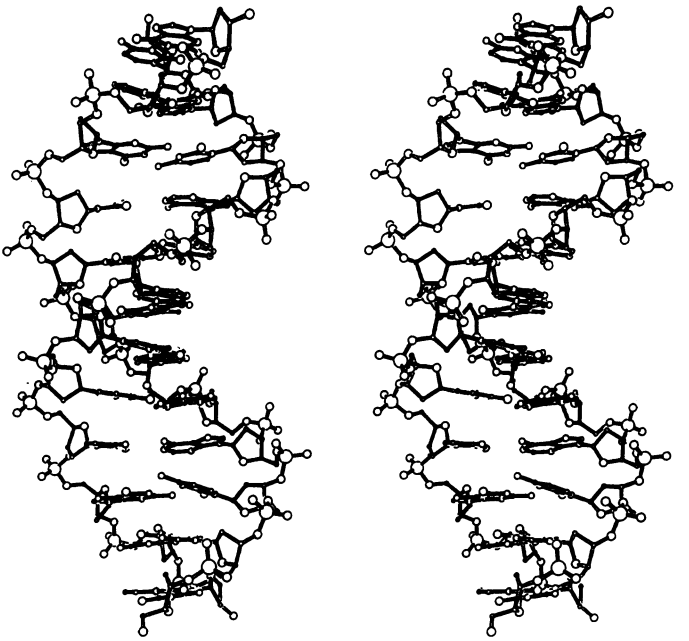

Figure 1
Stereoview of the duplex d(C-G-C-A-A-A-T-T-C-G-C-G) showing the minor groove at the bottom, major groove at the top. The bonds of atoms in the C.A mispairs are filled in to highlight these positions.

Figure 3 presents one of the mispairs observed in this study. The two C.A mispairs are, within the accuracy of this analysis, identical. The N-3 acceptor groups of the cytosines are 2.8 and 3.0Å from the hydrogen donating N-6 groups of the adenines in the A4.C21 and A16.C9 mispairs respectively. This contact clearly represents a standard hydrogen bond. There is another close contact between the mispaired bases. The electron acceptor groups O-2 of cytosine and N-1 of adenine are 2.8 and 2.7Å apart in the mispairs. At 2.5Å resolution we cannot observe hydrogen atoms but have proposed that this close contact represents a hydrogen bond stabilising the C.A mispair (12). All base pairs characterised so far utilise a minimum of two inter-base hydrogen bonds. These closed circular bonds are known to confer additional stability due to cooperativity. The question then arises as to how the O-2 and N-1 atoms can form a hydrogen bond.

Various hydrogen bonding schemes can be deduced that invoke various tautomeric forms of the bases, examples of which are given in Figure 2.

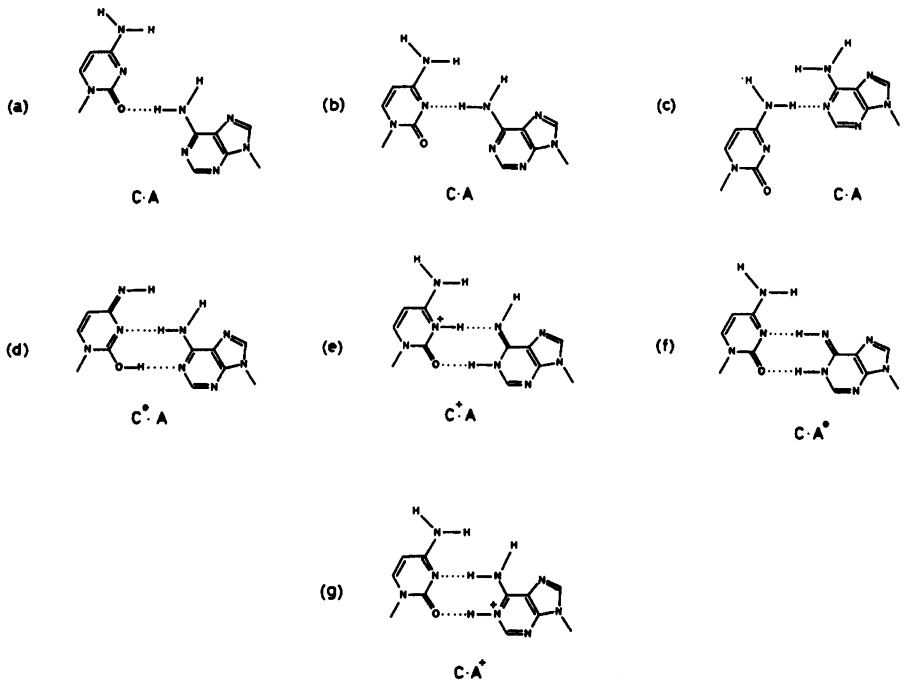

Figure 2
Some hydrogen bonding schemes for the C.A pairing. (a), (b) and (c) involve major tautomer forms of the bases and only one hydrogen bond. (d) and (f) involve rare tautomer forms of C and A respectively with two inter-base hydrogen bonds. (e) shows a possible pairing with cytosine N-3 protonated. In this case however the adenine would adopt a less likely tautomeric form. In (g) the adenine N-1 is protonated and there is no requirement for unusual tautomers, it is this scheme which we propose for the C.A pairing observed in this work.

Although we cannot unambiguously discount hydrogen bonding schemes utilising rare tautomers we believe that they would be unlikely given the calculated high energy of tautomerization (3,22). A more probable explanation involves protonation. The pKa (association constant) value of adenine N-1 is approximately 4.0 (23). However, the pKa value in a base pair could be expected to differ from that of the free base. Crystals were obtained from conditions buffered at pH 7.4. We suggest that duplex formation may have initiated a C.A mispair with a single hydrogen bond between N-3 and N-6. This could have produced variations in the pKa values of the two bases. The abstraction of a proton from bulk solvent could lead to protonation of N-1 of the adenines, with a subsequent delocalisation of the positive charge, and the formation of

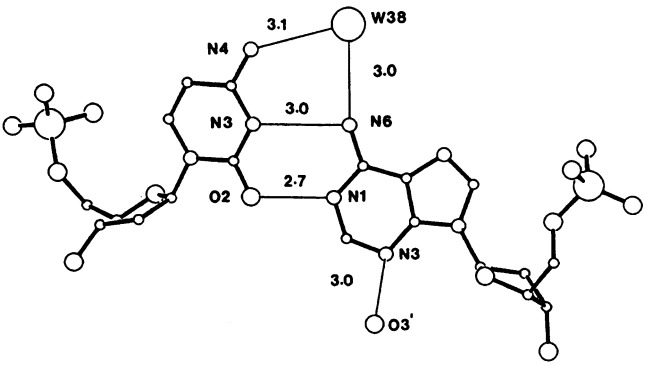

Figure 3
The C9.A16 mispair observed in this study with the associated water molecule (W38) in the major groove and the O-3'(G24) of a symmetry related molecule in the minor groove. Thin lines are used to represent hydrogen bonds. Atoms are shown in the decreasing order of size O(water)>P>O>N>C. Atoms involved in hydrogen bonds are labelled. Distances are given in Å.

the second hydrogen bond between the bases. Protonated forms of adenine have been observed in single crystal structures at atomic resolution (24) and proposed in relation to fibre diffraction studies (25).

An alternative scheme could be constructed with the cytosine rather than the adenine being protonated. There is evidence for such protonation from spectroscopic and diffraction experiments (26,27). The complex of d(G-C-G-T-A-C-G-C) and the bis intercalator triostin A (27) contains C.G(syn) base pairs with the cytosines N-3 being protonated. If in the C.A mispair the cytosine was protonated then the adenine would have to be in a rare tautomeric state (Figure 2e).

Subsequent to our report of the C.A mispair in d(C-G-C-A-A-A-T-T-C-G-C-G) (12) an nmr study by Sowers et al., (28) investigated the relative concentrations of tautomeric and protonated forms of bases with particular emphasis on the C.A pairing. Their results concur with our proposal of the protonated adenine and presence of two hydrogen bonds between bases in the C.A mispair.

The A4.C21 mispair has two closely associated solvent molecules, W35 and W36. One of these, W35 lies in the major groove and forms a well defined hydrogen bond to N-6(A4) of length 2.8Å. This solvent molecule is 3.5Å from N-4(C21), a distance which may be taken to represent a hydrogen bond since the solvent atom has a high thermal parameter (71Å^2) and the positional error

is probably between 0.2-0.5Å. In the minor groove W36 forms a hydrogen bond to N-3(A4).

The A16.C9 mispair has an intermolecular contact (of length 3.0Å) in the minor groove between N-3(A16) and O-3'(G24). This contact replaces the W36...N-3(A4) interaction observed for A4.C21. In the major groove a well ordered solvent molecule, W38 with a thermal parameter of 20Å^2, forms a bridge between the N-4(C9) and N-6(A16) groups. This bridging interaction is reminiscent of that observed in G.T mismatches in A, B and Z-DNA (8, 29-31) where the guanine O-6 and thymine O-4 atoms interact with a solvent molecule.

The sugar-phosphate backbone

Distances between adjacent phosphorus atoms in each strand are given in Table 1; they range from 5.9 to 7.3Å with an average of 6.6Å. In the parent compound the observed range is 6.2 to 7.1Å with an average of 6.7Å (20). The P...P distances between residues 8 and 9, 20 and 21, that is around the C9 and C21 nucleotides which are involved in the mispairs, are 0.8 and 0.1Å shorter than the corresponding distances in the native oligomers. Around A4 there is an increase in P...P separation of 0.3Å compared to the parent compound whilst around A16 the P...P separation does not change upon introduction of the C.A mispair.

DNA groove widths can be defined in terms of the separation of the phosphorus atoms across the grooves (32). In the mismatch structure the minor groove width varies from 13.8Å (P(A5)...P(G24)) to 9.2Å (P(T8)...P(C21)). There is a pronounced narrowing of the minor groove at the A-A-T-T region, as observed in the native structure, but there is no significant change at the mismatch sites. The major groove widths are more uniform and range from 18.5Å (P(T7)...P(C15)) to 17.0Å (P(A4)...P(A18)). These values show good agreement with those observed in the native structure and in closely related structures (20,21,30).

The backbone torsion angles and torsion angles around the glycosyl bond are presented in Table 1 together with average values and, for comparison, the average values of the parent oligomer (calculated from coordinates deposited with the Brookhaven Protein Databank). The range of individual torsion angles is wider in the mismatch than in the native structure but the average values agree to within 17°. The sugar-phosphate backbone is clearly flexible enough to accommodate the C.A mispairs with only minor adjustments to local conformations.

The glycosidic torsion angle χ ranges from -83 to -151° with an average

TABLE 1
Torsion angles (^{o}) and distances between adjacent phosphorous atoms (Å)

Residue	χ	α	β	γ	δ	ε	ζ	P_i-P_{i+1}
C1	-107	---	---	-111	133	-246	-21	---
G2	-88	-111	236	43	152	-178	-131	6.0
C3	-151	-49	139	65	75	-164	-91	6.5
A4	-109	-35	156	58	128	-165	-120	7.1
A5	-105	-31	155	34	105	-228	-45	6.9
A6	-83	-93	238	43	141	-200	-75	6.3
T7	-115	-68	195	52	111	-186	-88	6.7
T8	-111	-64	174	62	125	-159	-127	7.1
C9	-125	-23	125	39	74	-156	-97	5.9
G10	-97	-41	146	61	140	-132	-168	7.0
C11	-115	-31	111	54	111	-170	-73	6.8
G12	-93	-93	204	49	85	---	---	---
C13	-87	---	---	177	154	-197	-102	---
G14	-100	-79	179	65	142	-154	-163	6.6
C15	-144	-21	96	66	63	-186	-72	6.3
A16	-98	-74	198	66	132	-228	-60	7.1
A17	-85	-89	230	56	126	-241	-28	6.6
A18	-92	-93	239	18	133	-224	-51	6.5
T19	-123	-94	196	63	87	-204	-53	6.4
T20	-105	-53	193	44	128	-188	-64	7.0
C21	-97	-82	179	58	105	-138	-137	6.1
G22	-101	-42	127	65	128	-151	-159	7.3
C23	-120	-60	140	60	58	-146	-67	6.0
G24	-144	-52	167	50	82	---	---	---
Average	-108	-63	174	53[1]	113	-184	-91	6.6
Average values for d(C-G-C-G-A-A-T-T-C-G-C-G)								
	-117	-63	171	54	123	-169	-108	6.7

1 omits γ for C1 and C13

Main-chain torsion angles are defined by P $\overset{\alpha}{-}$ O-5 $\overset{\beta}{-}$ C-5' $\overset{\gamma}{-}$ C4' $\overset{\delta}{-}$ C3' $\overset{\varepsilon}{-}$ O-3' $\overset{\zeta}{-}$ P
The glycosyl torsion angle (χ) is defined by O-4' - C-1' - N-1 - C-2 for pyrimidines O-4' - C-1' - N-9 - C-4 for purines

value of -108°. In the parent compound χ varies from -88 to -135°, average of -117°. These ranges correspond to -synclinal through to -anticlinal orientation as defined by IUPAC-IUB nomenclature (33). We note that, as described previously (34), χ falls into two distinct categories. For purines, χ ranges from -83° to -144° (average -100°) and for pyrimidines the range is -87° to -151° (average -117°). This difference may reflect steric effects of having either a five or six membered ring adjoining the furanose moieties.

Torsion angle δ is correlated with the conformation of the furanose ring (35); in the present structure it varies from 58° to 154° (+synclinal to

TABLE 2
Analysis of sugar puckers in d(C-G-C-A-A-A-T-T-C-G-C-G)

Residue	Amplitude of pucker	Pseudorotation phase angle	Conformation	Conformation of parent structure
C1	31.5	200	C3'-exo	C2'-exo
G2	39.7	180	C2'-endo	C1'-exo
C3	35.2	34	C3'-endo	O4'-exo
A4	32.7	162	C2'-endo	C2'-endo
A5	35.9	154	C1'-exo	C1'-exo
A6	38.0	193	C3'-exo	C1'-exo
T7	29.1	130	C1'-exo	O4'-endo
T8	35.1	165	C2'-endo	C1'-exo
C9	25.7	44	C3'-endo	C1'-exo
G10	43.7	164	C2'-endo	C2'-endo
C11	21.1	164	C2'-endo	C2'-endo
G12	13.9	73	O4'-endo	C1'-exo
C13	43.9	225	C4'-endo	C2'-endo
G14	42.4	158	C2'-endo	C1'-exo
C15	49.9	50	C4'-exo	O4'-endo
A16	35.2	183	C2'-endo	C2'-endo
A17	32.0	186	C2'-endo	C2'-endo
A18	23.9	205	C3'-exo	C2'-endo
T19	30.9	80	O4'-endo	C1'-exo
T20	30.2	195	C3'-exo	C1'-exo
C21	37.5	135	C1'-exo	C1'-exo
G22	41.4	163	C2'-endo	C2'-endo
C23	39.8	172	C3'-endo	C1'-exo
G24	31.6	18	C3'-endo	C3'-endo

+antiperiplanar) and the apparent sugar conformations vary from C3'-endo to C4'-endo. Similar variation is found in the native dodecamer structure structure at a slightly higher resolution (20). It is interesting, however, that in the phosphorothiolate hexamer d(Gp(S)CpGp(S)CpGp(S)C) which also crystallises in a B-DNA type helix (36) the sugar conformation is consistently C2'-endo.

The furanose rings of C3 and C15 have quite small δ values. As mentioned previously these sugar moieties are involved in van der Waals type packing with symmetry related duplexes and their conformations may be affected by the inter-duplex packing in the unit cell.

Sugar conformations may also be analysed with the pseudorotation concept (37,38). In Table 2 we give the psuedorotation phase angles and amplitudes of pucker together with inferred furanose conformations.

Conformations of the remaining backbone torsion angles may be summarised as follows. Torsion angle α ranges from -synperiplanar to -anticlinal, β varies from -anticlinal to +anticlinal while γ is restricted to +synperiplanar and +synclinal when γ C1 and γC13 are ignored due to end

effects. Torsion angles ε and ζ fall in the range -anticlinal to +anticlinal and -synperiplanar to +anticlinal respectively.

The Bases

The vertical stacking of bases (23,39) is an important factor in stabilising the double helix. Disruption of stacking interactions induced by the presence of mispairs could destabilise the duplex and may contribute to mechanisms which maintain the accuracy of DNA replication. In d(C-G-C-A-A-A-T-T-C-G-C-G) there are eleven base pair steps. Seven steps involve only Watson-Crick base pairs and the base stacking interactions are very similar to those observed at corresponding steps in the native dodecamer. These are steps 1 and 11 of type CpG(=CpG), 2 and 10 of type GpC(=GpC), 5 and 7 of type ApA(=TpT) and 6 of type ApT(=ApT). The stacking interactions may be summarised as follows. There is considerable variation in base pair overlap from one step to another. Purine-pyrimidine steps (2, 6 and 10) show greater overlap of bases than pyrimidine-purine steps (1 and 11). The ApA(=TpT) steps show good base overlaps with the six membered rings of the purine bases interacting with each other (20).

Steps 3 and 9, of type CpA(=CpG) and 4 and 8, of type ApA(=TpC) involve a C.A mismatch. The same base stacking interactions are observed at each mispair; therefore only one set,(steps 9 and 8), is illustrated and compared with the corresponding set in the native structure (Figure 4). In the parent compound step 9, CpG(=CpG) shows only a slight overlap of C15 on G16 and G10 on C9. The corresponding step in the mismatch structure shows that the pyrimidine purine interaction, C15 on A16 overlap, is very similar. On the other strand the lateral displacement of C9 into the major groove leads to a decrease in the intra-strand purine pyrimidine overlap. In both the native and mismatch dodecamer structures a small degree of inter-strand purine purine overlap is observed, although the nature of this overlap differs.

Step 8 in the parent structure, GpA(=TpC) shows moderate overlap between G16 on A17 and between C9 and T8. This is conserved at the same step in the mismatch structure, A16 on A17 while the pyrimidine pyrimidine overlap, C9 on T8 is actually improved due to the displacement of C9 into the major groove. It should be noted that the relative importance of the various forces contributing to base stacking is not fully understood. Hence, although the hydrophobic interactions in the pyrimidine pyrimidine overlap in the mismatch seem favourable, dipole-dipole interactions may be unfavourable due to the close proximity of the electronegative O and N atoms due to the CC stacking. In the native structure, dipole-dipole interactions seem more favourable in

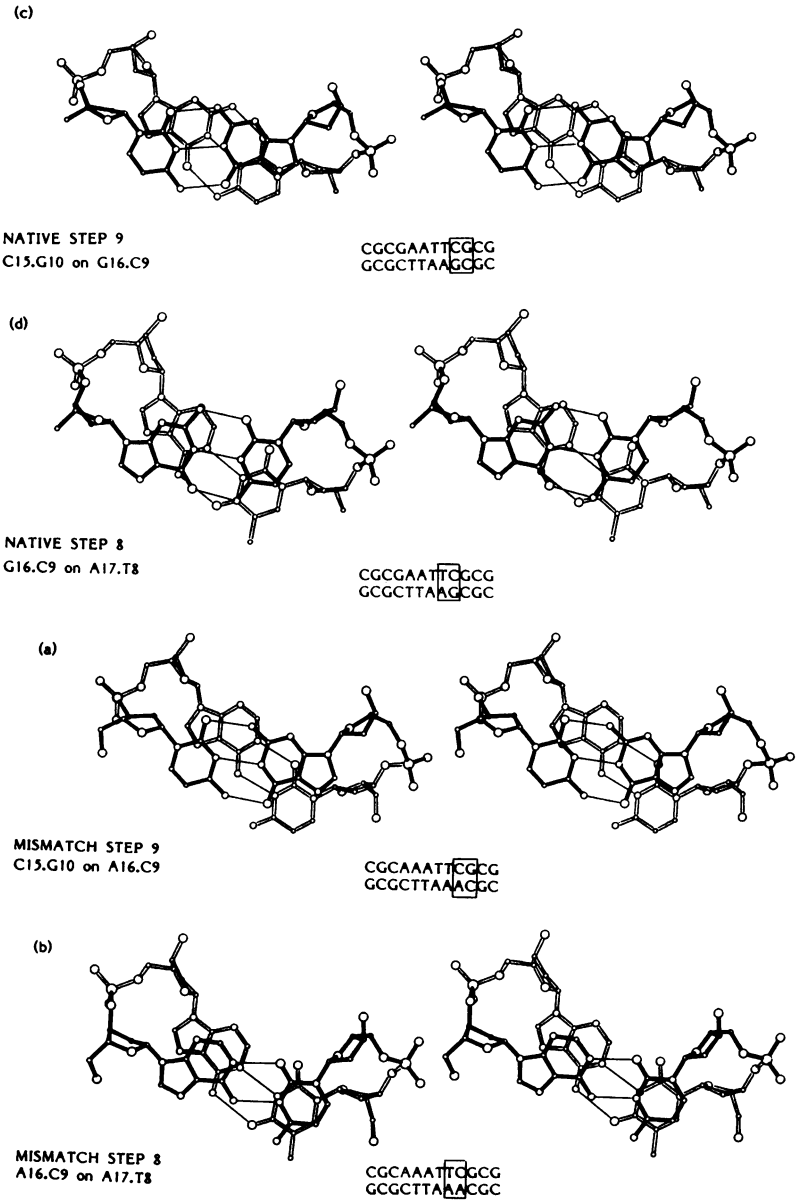

Figure 4
Stereoviews of base stacking interactions at steps involving the C9.A16 mispair, (a) and (b) and corresponding steps, (c) and (d) in the native dodecamer. Hydrogen bonds are drawn as thin lines and covalent bonds of the upper residue are filled in. Atoms are depicted as spheres of decreasing radii in the order P>O>N>C.

TABLE 3
Geometrical properties of base pair steps and base pairs in d(C-G-C-A-A-A-T-T-C-G-C-G)

Base Pair	Step	Roll (°)	Twist (°)	Rise (Å)	Propellor twist (°)	λ_1(°)	λ_2(°)	C1'...C1' (Å)
C1.G24					9	52	54	10.8
	1	0	37	3.1				
G2.C23					12	52	56	10.6
	2	-7	37	2.7				
C3.G22					18	51	55	10.6
	3	5	35	4.1				
A4.C21					13	46	64	10.4
	4	7	30	3.5				
A5.T20					17	58	59	10.1
	5	4	40	3.1				
A6.T19					17	53	61	10.1
	6	-7	30	3.2				
T7.A18					14	57	60	10.2
	7	1	36	3.2				
T8.A17					17	60	54	10.3
	8	-6	36	3.1				
C9.A16					15	72	49	10.2
	9	10	36	4.3				
G10.C15					7	57	51	10.7
	10	-10	38	2.7				
C11.G14					5	56	51	10.5
	11	-4	37	3.6				
G12.C13					5	55	54	10.6

the CT stacking steps. The electron rich N-4 of C9 overlaps the electron deficient C-4 of T8 and the electron rich N-3 of T8 overlaps the electron deficient C-4 of C9. Similarly there is good overlap between N-1 of G16 and C-6 of A17, C-6 of G16 and N-6 of A17 in the native structure. In the mismatch, the overlaps occur at the mid-point in the C-N and C-O bonds and hence will be less favourable. Thus, from a qualitative view of dipole-dipole interactions, the native base stacking interactions may be more favourable.

Stacking interactions can be described in a more quantitative fashion using parameters derived by Dickerson, Calladine and Drew (20,40). Table 3 gives selected geometrical properties of base pair steps and base pairs in d(C-G-C-A-A-A-T-T-C-G-C-G).

Twist is the rotation from one base pair to the next with respect to the overall helix axis. The average twist value for both the mismatch and native structures is 36° and corresponds to ten base pairs per turn of helix. There is an interesting change at steps involving the mispairs. Steps 3 and 9,

type CpA(=CpG) are overwound by 9^{o} and 6^{o} compared to the corresponding steps in the native dodecamer. On the other side of the mispairs, steps 4 and 8, type ApA(=TpC) are underwound by 6^{o} and 5^{o}.

Roll refers to the angle by which adjacent base pairs open up to a helix groove. In the parent compound, purine pyrimidine steps, GpC(=GpC) and ApT(=ApT), open out towards the major groove at steps 2, 6 and 10 with roll values of -3^{o}, -6^{o} and -10^{o}. Steps 1, 3 and 9 CpG(=CpG) open towards the minor groove with rolls of 7^{o}, 5^{o} and 2^{o} respectively. This directionality which has been attributed to the avoidance of steric clash between purines from opposite strands (39) is maintained in the mismatch dodecamer.

The rise per base step shows little difference between the mismatch and parent compounds, values within 0.3Å being observed for ten out of eleven corresponding steps. The one marked exception is step 4 where an increase in rise of 0.5Å is observed upon incorporation of a C.A mispair. This may well be a consequence of the unfavourable electrostatic interactions discussed earlier. Rise per base pair will be more fully discussed further on.

The dihedral angle between individual base planes when viewed along the base pair long axis is the propellor twist. In the native structure the range is 5^{o} to 18^{o}, in the mispair structure it is 5^{o} to 25^{o}. There is only one base pair which shows any large difference compared to the native structure. This is C3.G22, adjacent to a C.A mispair. The twist increases from 6^{o} to 18^{o} when the mispair is present.

Table 3 also lists values for and the C-1'...C-1' distances for each base pair. We use λ to represent the angle between the glycosyl bond (C1'-N) and the C1'...C1' vector for any given base pair; for example λ_1 is the N-1(C1)...C-1'(C1)...C-1'(G24) angle, λ_2 the complementary angle subtended at C-1'(G24). Watson-Crick base pairs display a relatively narrow range of λ values. In the present structure they range from 51^{o} to 60^{o}. Note, however, that the C.A mispairs have a wider range of values 46^{o} to 64^{o} in A4.C21 and 49^{o} to 72^{o} in C9.A16. This pronounced asymmetry of the mispair may be an important structural feature in enzymic recognition of mismatches (9, 12). The C-1'...C-1' distances seem to be unaffected by the incorporation of mispairs.

Thermal parameters

Isotropic thermal parameters for individual atoms were refined against the diffraction data. These parameters ($B=8\pi^{2}\bar{u}^{2}$Å^{2}) are linked to molecular motion dynamics (41) but also contain contributions from errors in the model, the diffraction data and the quality of the fit of atomic positions to the

electron density. Final B values in $Å^2$ ranged from 17 to 54 (average 35) for the bases; 37 to 67 (average 48) for the sugars and 44 to 69 for the phosphates (average 57). These values agree with those observed in the native structure where the ranges were 20 to 44 (average 28) for bases; 32 to 51 (average 42) for sugars and 38 to 61 (average 50) for the phosphates (these values were calculated from coordinates deposited with the Brookhaven Protein Databank). There are no indications of disorder or instability at the mismatch sites when compared to other parts of the duplex or to the native structure. The average B values decrease from the phosphates to the furanose rings to the bases in the mismatch structure. This trend has been observed in all oligonucleotide structures determined to date by single crystal diffraction techniques and represents the inherent flexibility of the phosphate-sugar backbone in comparison to the bases. The thermal parameters observed for solvent molecules range from 17 to 83$Å^2$ with an average of 53$Å^2$.

Hydration

Crystallographic techniques can only identify the locally stable solvent positions. Oligonucleotide crystals contain approximately 50% w/v of DNA, the remainder being bulk solvent. With the exception of Z-DNA crystals, which diffract to a very high resolution (42), the quality of the experimental data does not allow identification of more than about 30% of the solvent molecules or of cations. Our criteria for identification of solvent were that a well shaped peak in difference maps with a minimum value of 2.5 standard deviations in height was within 2.2 to 3.3Å of plausible hydrogen bonding partners and that in the subsequent refinement acceptable hydrogen bonding contacts and thermal parameters were obtained. This approach, together with rigorous examination of electron density and difference maps, was used to reduce the chances of introducing artefacts into the refinement. Several solvent molecules refined to positions outside our criteria but were nevertheless retained due to a well defined presence in the various maps. We have not been able to identify cations unambiguously and hence solvent molecules were refined as oxygen atoms of water molecules with a fixed occupancy. In the first report of this structure 81 solvent molecules were included in the refinement (12). Subsequent refinement has lead to the deletion of three of these atoms.

Of the 78 independent solvent molecules identified, 60 have contacts with one or more atoms of the DNA and 43 of these have additional contacts to one or more other solvent molecules. The sugar-phosphate backbone interacts with 46 of the solvent molecules, the bases with 18 solvents.

In the parent dodecamer structure a total of 72 solvent molecules were identified, 50 of which form contacts to the DNA (43). In the structural analysis of a modification, sequence d(C-G-C-G-A-A-T-T-^{Br}C-C-G-C-G) crystallised in 60% (v/v) 2-methyl-2,4-pentanediol and analysed at 7^{o}C (termed MPD7), 114 solvent molecules were located. The marked increase in the number of solvent molecules identified was attributed to immobilization of water due to the high MPD concentrations (44). The hydration pattern is nevertheless similar to that of the parent structure. It is worth noting that in the MPD7 structure the hydration of the phosphate backbone was more ordered than in the parent compound. We now give details of the hydration observed in the present study and where appropriate draw comparisons with the native and MPD7 structures.

In the minor groove of the mismatch dodecamer the bases provide 6 hydrogen bond donors (N-2 of guanine) and 24 hydrogen bond acceptors (O-2 of cytosine and thymine, N-3 of guanine and adenine). Inter-molecular hydrogen bonding interactions prevent 9 of these groups from being hydrated. Ten of the remaining groups are hydrated, 8 of these in the narrow A-A-T-T region of the duplex where they form a hydration pattern similar to that observed previously (36, 37).

In the major groove there are 36 polar groups; 14 hydrogen bond donors (N-4 of cytosine, N-6 of adenine) and 22 hydrogen bond acceptors (O-4 of thymine, N-7 of adenine) and 22 hydrogen bond donors (O-4 of thymine, N-7 of adenine and guanine, O-6 of guanine). In total, 15 of these groups (4 donors, 11 acceptors) are hydrated.

On the backbone the O-1P atoms are directed into the major groove, the O-2P atoms project away from the duplex into bulk solvent. Of the 44 free phosphate oxygen atoms, 26 (15 O-2P, 11 O-1P) are associated with solvent molecules. There is one water molecule, which forms a bidentate intraphosphate link between O-1P and O-2P. In B-DNA the sugar-phosphate backbone is extended relative to A-DNA (45). The later conformation contains a large number of solvent mediated interphosphate bridging interactions as evidenced in several crystal structures (45). In B-DNA the free oxygen atoms of adjacent phosphates are in general about 6.6Å apart, too far for a single solvent bridge and so each phosphate group tends to be individually hydrated. The extension in B-DNA backbone makes the O-3' and O-5' atoms more accessible to hydration when compared to A-DNA (46). In this study 7 O-3' and 5 O-5' atoms are hydrated and there are 6 bidentate intraphosphate interactions between free phosphate oxygens and phosphodiester oxygens. In the MPD7

structure there are 7 solvents which make this type of interaction (44). The furanose oxygen atoms are sparsely hydrated with only 6 associated with well defined solvent molecules.

CONCLUSIONS

The analysis of the structural features and hydration of the mismatch dodecamer d(C-G-C-A-A-A-T-T-C-G-C-G) substantiates previous observations based on single crystal X-ray studies of DNA fragments with G.A and G.T mispairs (7-10, 12, 29-31). Non-complementary bases are able to form stable base pairs linked by two interbase hydrogen bonds. The mismatch base pairs characterised so far do not cause large disruptions to either the sugar-phosphate backbone or to base stacking interactions. Any small changes in conformation and stacking are localised.

Mismatch recognition and repair mechanisms are highly selective and efficient. They must be able to detect deviations from Watson-Crick complementarity and alterations in base pair geometry induced by mispair formation provides a number of ways by which enzymes could interact with mismatches. Specific hydrogen bonding interactions could help to identify mispairs or distinguish between the differant combinations of bases. We note that mispairs are pronouncedly asymmetric with respect to the approximate twofold axis inherent in Watson-Crick base pairs. The asymmetry, measured by λ as discussed previously, shows that the degree of asymmetry is G.T>C.A$\geq$G.A (12). This is the order of efficiency, observed in a number of studies, in which these mispairs are recognised and repaired (13-16).

In addition to changes in the λ values we also observe perturbations in the rise per base pair at steps involving G.T, C.A and G.A mispairs compared to the native structure. Each mismatch dodecamer contains two mispair sites involving residues 4 and 21, 9 and 16. Steps 3 and 9 are referred to as being on the 5' side of the mispair, 4 and 8 on the 3' side. When these mismatches are present the rise per base pair increases on the 5' side. In the native structure the rise at steps 3 and 9 is 3.6 and 3.3Å. For the G.T mismatch both 5' steps have rise values of 4.3Å, for the G.A 4.2Å. The C.A mispair produces rise values of 4.1 and 4.3Å at steps 3 and 9. Steps 4 and 8 have rise values of 3.1 and 3.3Å in the native structure. The G.T mispair shows rise values of 2.7 and 3.3Å at the corresponding steps, the C.A structure values of 3.5 and 3.1Å. The G.A has values of 3.1Å at both of these steps. The best repaired mismatch (G.T) is observed to produce the largest perturbations in rise per base pair in this dodecamer sequence. The

C.A and G.A mismatches produce changes very similar to each other. Rise per base pair is correlated with helical twist and base stacking and will be strongly influenced by the nature of adjoining nucleotides. It may well be a contributing factor to the observed variation of mismatch repair with base sequence (5). In summary, we think that the polymerase enzymes have a capacity to recognise differences in structure between normal and mismatched base pairs and eliminate the errors accordingly.

Thermodynamic properties of mismatch base pairs greatly affect the frequency of mispair formation and the processes involved in repair (13). We note that in the investigation of the thermodynamics of various mispairs little or no attention was paid to the role that hydration of polar groups could play (4-6, 47, 48). Potential hydrogen bond donors and acceptors will be satisfied when accessible to solvent. This hydration of polar groups contributes to thermodynamic stability (49). The disposition of bases in the G.T and C.A mispairs places functional groups in a position to allow bridging interactions by solvent molecules. The G.T mispair can have bridging solvent molecules in both the major and minor grooves but the C.A mispair can only have such an interaction in the major groove. These additional hydrogen bonds must contribute to mismatch stability. The presence of an additional bridging solvent molecule in the G.T. mispair might well help explain the greater thermodynamic stability of G.T compared to C.A mispairs as observed experimentally (4). Work is currently underway to investigate the thermodynamic stability of the sequences discussed in this paper by UV melting techniques (Brown, private communication).

ACKNOWLEDGEMENTS

We are grateful to the Medical Research Council Laboratory of Molecular Biology Cambridge, U.K for allowing access to their Evans and Sutherland graphics system, to Dr P R Evans for all his help, Professor A Eschenmoser for discussions and to Mrs H. Belchamber for preparation of this manuscript. Special thanks go to a referee for many helpful comments. This work was supported by an MRC program grant to O.K.

*To whom correspondence should be addressed

REFERENCES

1 Watson, J.D. & Crick, F.H.C. (1953), Nature, 171, 964-967
2 Topal, M.D. & Fresco, J.R. (1976), Nature, 263, 285-293
3 Rein, R., Shibata, M., Gardino-Juarez, R. & Keiber-Emmons, T. (1983),

"Structure and Dynamics of Nucleic Acids and Proteins" (eds., Clementi, E. and Sarma, R.) Adenine Press, New York, 269-288
4 Aboul-ela, F., Koh, D., Tinoco, Jr. F. & Marvin, F.H. (1985), Nucleic Acids Res., 13, 4811-4824
5 Tibanyenda, N., De Bruin, S.H., Haasnoot, C.A., van Marel, G.A., van Boom, J.H. & Hilbers, C.W. (1984), Eur. J. Biochem., 139, 19-25
6 Patten, J.E., So, A.G. & Downey, K.M. (1984), Biochem., 23, 1613-1618
7 Brown, T., Kennard, O., Kneale, G. & Rabinovich, D. (1985), Nature, 315, 604-606
8 Kneale, G., Brown, T., Kennard, O. & Rabinovich, D. (1985), J. Mol. Biol., 186, 805-814
9 Brown, T., Hunter, W.N., Kneale, G. & Kennard, O. (1986), Proc. Nat. Acad.Sci. (USA), 83, 2402-2406
10 Kennard, O. (1985), J. Biomol. Struct. Dynam., 3, 205-226
11 Patel, D.J., Kozlowski, S.A., Ikuta, S. & Itakura, K. (1984), Biochemistry, 42, 2663-2670
12 Hunter, W.N., Brown, T., Anand, N.N. & Kennard, O. (1986), Nature, 320, 552-555
13 Kramer, B., Kramer, W. & Fritz, H.J. (1985), Cell, 38, 879-887
14 Fersht, A.R., Knill-Jones, J.W. & Tsui, W.C. (1982), J. Mol. Biol., 156, 37-51
15 Lu, A.L., Welsh, K., Clark, S., Su, S.S. & Moldrich, P. (1984) Cold Spring Harbour Symp. Quant. Biol., 49, 589-596
16 Muster-Nassal, C. & Kolodner, R. (1986), Proc. Nat. Acad. Sci. (USA), 83, 7618-7622
17 Laengle-Rouault, F., Maenhaut-Michel, G. & Radman, M. (1986), EMBO Journal, 5, 2009-2013
18 Accuracy in Molecular Processes. (1986) University Press, Cambridge Edited by Kirkwood, T.B.L., Rosenberger, R.F., and Galas, D.J.
19 Wing, R., Drew, H.R., Takano, T., Broka, C., Tanaka, S., Itakura, K. & Dickerson, R.E. (1980), Nature, 287, 755-758
20 Dickerson, R.E. & Drew, H.R. (1981), J. Mol. Biol., 149, 761-786
21 Dickerson, R.E., Kopka, M.L. & Drew, H.R. (1983) "Structure and Dynamics of Nucleic Acids and Proteins" (eds., Clementi, E. & Sarma, R.) Adenine Press, New York, 149-179
22 Churpina, V.P. & Poltev, V.I. (1985), Nucleic Acids Res., 13, 141-152
23 Saenger, W. Principles of Nucleic Acid Structure. (1984) Springer-Verlag
24 Taylor, R. & Kennard, O. (1982), J. Molec. Struct., 78, 1028-1032
25 Rich, A., Davies, D.R., Crick, F.H.C. & Watson, J.D. (1961), J. Mol. Biol., 3, 71-86
26 Sarocchi, M-T. & Guschlbauer, W. (1972), Eur.J.Biochem. 34, 232-240
27 Quigly, G.J., Ughetto, G., van der Marel, G.A., van Boom, J.H., Wang, A.H-J. & Rich, A. (1986), Science, 232, 1255-1258.
28 Sowers, L.C., Fazakerly, G.V., Kim, H., Dalton, L. & Goodman, M.F. (1986), Biochemistry, 25, 3983-3988
29 Hunter, W.N., Kneale, G., Brown, T., Rabinovich, D. & Kennard, O. (1986), J. Mol. Biol., 190, 605-618
30 Hunter, W.N., Brown, T., Kneale, G., Anand, N.N., Rabinovich, D. & Kennard, O. (1986), J. Biol. Chem. (in press)
31 Ho, P.S., Frederick, C.A., Quigley, G., van der Marel, G.A., van Boom, J.H., Wang, A.H-J. & Rich, A. (1985), EMBO.J. 4, 3617-3623.
32 Arnott, S. (1976) "Organisation and Expression of Chromosomes" (eds., Allfrey, V.G., Bantz, E.K.F., McCarthy, B.J., Schimke, R.T. & Tissieres, A.) Dahlem Conference, Berlin, 209-222
33 IUPAC-IUB Joint Commission on Biochemical Nomenclature (1983), Eur. J. Biochem., 131, 9-15

34 Drew, H.R., Wing, R.M., Takano, T., Broka, C., Tanaka, S., Itakura, K. & Dickerson, R.E. (1981), Proc. Nat. Acad. Sci. (USA), 78, 2179-2183

35 Levitt, M. & Warshel, A. (1978), J. Amer. Chem. Soc., 100, 2607-2613

36 Cruse, W.B.T., Salisbury, S.A., Brown, T., Eckstein, F., Costick, R. & Kennard, O. (1986), J.Mol.Biol. 192, 891-905.

37 Altona, C. & Sundaralingam, M. (1972), J. Amer. Chem. Soc., 94, 8205-8212

38 Rao, S.T., Westhof, E. & Sundaralingam, S. (1981), Acta Cryst Sect. A, 36, 421-425

39 Dickerson, R.E. (1983), J. Mol. Biol., 166, 419-441

40 Calladine, C.R. & Drew, H.R. (1984), J. Mol. Biol., 179, 773-782

41 Ringe, D. & Petsko, G.A. (1985), Prog. Biophys. Molec. Biol., 45, 197-235

42 Wang, A.H-J., Quigley, G.J., Kolpack, F.J., Crawford, L., van Boom, J.H., van der Marel, G. & Rich, A. (1979), Nature 282, 680-686

43 Drew, H.R. & Dickerson, R.E. (1981), J. Mol. Biol., 151, 535-556

44 Kopka, M.L., Fratini, A.V., Drew, H.R. & Dickerson, R.E. (1983), J. Mol. Biol., 163, 129-146

45 Saenger, W., Hunter, W.N. & Kennard, O. (1986), Nature, 324, 385-388

46 Kennard, O., Cruse, W.B.T., Nachman, T., Prange, T., Shakked, Z. & Rabinovich, D. (1986), J. Biomolec. Stereodynamics. 4, 623-647

47 Jivicny, J., Wood, S.G., Martin, D. & Ubasawa, A. (1986), Nucleic Acids Res., 16, 6579-6590

48 Wood, S.G., Ubasawa, A., Martin, D. & Jiricny, (1986), Nucleic Acids Res. 14, 16, 6591-6602

49 Fersht, A.R., Shi, J-P., Knill-Jones, J., Lowe, D.M., Wilkinson, A.J., Blow, D.M., Brick, P., Carter, P., Waye, M.M.Y. & Winter, G. (1985) Nature, 314, 235-238