The Structure of Guanosine-Thymidine Mismatches in B-DNA at 2.5-Å Resolution*

William N. Hunter, Tom Browne‡, Geoff Kneale§, Naveen N. Anand, Dov Rabisovich, and Olga Kennard

From the University Chemical Laboratory, Cambridge CB2 1EW, United Kingdom

The structure of the deoxyoligomer d(C-G-C-G-A-A-T-T-T-G-C-G) was determined at 2.5-Å resolution by single crystal x-ray diffraction techniques. The final R factor is 18% with the location of 71 water molecules. The oligomer crystallizes in a B-DNA-type conformation, with two strands interacting to form a dodecamer duplex. The double helix consists of four A-T and six G-C Watson-Crick base pairs and two G-T mismatches. The G-T pairs adopt a "wobble" structure with the thymine projecting into the major groove and the guanine into the minor groove. The mismatches are accommodated in the normal double helix by small adjustments in the conformation of the sugar phosphate backbone.

A comparison with the isomorphous parent compound containing only Watson-Crick base pairs shows that any changes in the structure induced by the presence of G-T mismatches are highly localized. The global conformation of the duplex is conserved. The G-T mismatch has already been studied by x-ray techniques in A and Z helices where similar results were found. The geometry of the mismatch is essentially identical in all structures so far examined, irrespective of the DNA conformation. Hydration may be an important factor in stabilizing G-T mismatches.

A characteristic of Watson-Crick paired A-T and G-C bases is the pseudo 2-fold symmetry axis in the plane of the base pairs. The G-T wobble base pair is pronouncedly asymmetric. This asymmetry, coupled with the disposition of functional groups in the major and minor grooves, provides a number of features which may contribute to the recognition of the mismatch by repair enzymes.

Base pair mismatches occur when bases which are not complementary in the Watson-Crick sense form hydrogen-bonded pairs, for example, A-C or G-T instead of G-C. In theory a variety of hydrogen-bonded pairs, for example, A.C or G.T instead of G.C. In theory a variety of hydrogen-bonded pairs can be postulated between purine-pyrimidine, purine-purine, and pyrimidine-pyrimidine pairs (8, 9). Biosynthetic errors which occur during replication and genetic recombination or which are induced by chemical reactions (10, 11) may lead to the inclusion of mismatched base pairs in DNA. These errors reduce the fidelity of replication and may lead to a high rate of point mutations.

A complicated and as yet poorly understood recognition and repair system has developed to process mismatches and other structural anomalies. Enzymes, such as the DNA polymerases, are able to detect, excise, and correct errors in DNA (12). The repair efficiency of mispaired bases is strongly dependent on the type of mismatch (13-15). For example, it has been shown that in Escherichia coli, G-T mismatches are repaired with greater efficiency than A-A mismatches, which in turn are processed more efficiently than G-A pairs. The thermodynamic stability of the mismatches varies with the actual bases involved, their position in the double helix, and the neighboring nucleotide sequence (16-18). The thermodynamic properties strongly influence the frequency of occurrence of the different mismatches and may contribute to the efficiency of repair. The stereochemical nature of the mismatch is likely to be the primary determinant in the enzymatic recognition of mismatches.

We have characterized G-T mismatches in A- and Z-DNA (5, 6, 19), as well as C-A and G-A mismatches in B-DNA (20, 21). The single crystal x-ray study of the synthetic deoxyoligomer d(C-G-C-G-A-A-T-T-T-G-C-G) with two G-T mismatches in more than a full turn of a B-DNA helix, which is presented here, allows us to analyze the effect of this mispair on the local and global conformation of the double helix and to compare it to the effect of the C-A and G-A mismatches in identical sequence environments.

MATERIALS AND METHODS

Synthesis and Crystallization—The deoxyoligomer d(C-G-C-G-A-A-T-T-T-G-C-G) was synthesised by the solid-phase triester method (22) and was purified by ion-exchange and reverse-phase high pressure liquid chromatography. Crystals were grown in an aqueous solution containing approximately 0.5 mM DNA duplex, 8 mM sodium cacodylate buffer (pH 7.4), 25 mM magnesium acetate, 1.0 mM spermine, and 20% 2-methyl-2,4-pentanediol. The solution was maintained at 8 °C for a period of 1 month before crystals suitable for diffraction experiments were obtained.

Diffraction Measurements—Precise photography indicated isomorphism with the native deoxyoligomer d(C-G-C-G-A-A-T-T-T-G-C-G) analyzed by Dickerson and co-workers (1). A single, well-formed crystal (1.0 × 0.2 × 0.3 mm) was mounted and sealed in a glass capillary and proved to be of sufficient quality to yield data to a resolution of 2.5 Å. The unit cell dimensions, obtained from diffractometer measurements, are a = 25.53, b = 41.22, c = 65.65 Å. The space group is P2₁2₁2₁. Diffraction intensities were recorded in ω-scan mode on a Syntex P2₁ diffractometer equipped with a long arm helium path and using graphite monochromated CuKα radiation. The crystal was maintained at a constant temperature of 6 °C throughout the data collection. Corrections were applied for crystal decomposition, Lorentz polarization, and absorption effects. Of the 2497 inde-
Mispairs in B-DNA

FIG. 1. Stereoview of mispair T9-G16 superimposed on fragment electron density map. All atoms of the nucleotides were given a low occupancy so they do not affect the structure factor calculation. The electron density obtained is an unbiased representation of where the atoms should be positioned. The map is contoured between levels ±0.7 and ±1.8 eÅ⁻³ at intervals of 0.2 eÅ⁻³. At this level of contouring, only positive density is observed. This is represented with continuous lines.

Structure Refinement—The starting model for the analysis was obtained from the coordinates of the native dodecamer d(C-G-C-G-A-A-T-T-C-G-C-G) which had been deposited with the Brookhaven Protein Data Bank (1). However, the atoms of the bases at positions 4, 9, 16, and 21 corresponding to the positions of the mismatched bases were given low occupancy in the early stages of the refinement. As a result, they made virtually no contribution to structure factor calculations. The starting model was refined against the experimental data as a rigid body using a modified version of SHELX (23). Initially, only reflections in the range of 10 to 7 Å were used in the calculations. The resolution was extended in four steps of 1 Å. The rigid body refinement converged with a crystallographic residual, R, of 0.40 for the 1404 reflections with I > 3σ(I) in the 10 to 3 Å resolution range. Next, the restrained least-squares method of Hendrickson and Konnert (24) was applied using NUCLSQ, a program modified for nucleic acid components (25). The resolution was gradually extended to 2.5 Å. When the refinement converged electron density (2F₀-Fᵢ) and difference (Fᵢ-F₀) maps were calculated and displayed on an Evans & Sutherland PS300 computer graphics system using the program FRODO (26). The atoms of the G'T and T.G base pairs were fitted to the maps, and their coordinates were substituted for the atoms of the G.C and C.G base pairs of the starting model at positions 4 and 21 and 9 and 16. Both of the mispaired bases were well defined as wobble base pairs. The atoms of these bases were now given full occupancy in the refinement.

Subsequent refinement of the structure was carried out by a combination of restrained least-squares calculations and manipulation of the atomic positions in various regions of the dodecamer on the graphics system to improve the fit with the electron density. To avoid introducing bias in the conformations of the deoxyribose moieties, their constituent atoms were allowed to move freely during graphic manipulation. Any regularization of coordinates of the furanose ring atoms would have produced a predetermined sugar conformation. In particular, tying the sugars to the C₂'endo conformation (as normally associated with B-DNA) would have been incorrect since a variety of sugar conformations were observed in the native dodecamer (1) as well as in the dodecamers with C.A and G.A mismatches (20,21).

Refinement was continued with the inclusion of solvent molecules. Positive peaks, greater than 2.5 standard deviations in height, were located in difference syntheses using the automated search program PKSER. These peaks were then inspected on the graphics to ascertain whether they could represent solvent molecules. Our criteria for identification of solvent required that a well-shaped peak be within 2.2-3.3 Å of plausible hydrogen-bonding partners and that in the subsequent refinement an acceptable geometry and thermal parameter were obtained. Solvent molecules which appeared dubious were discarded prior to calculation of new Fourier syntheses from which more solvent molecules could be identified. Several solvent molecules refined to positions not meeting our criteria but were retained due to their well-defined presence in the maps. Despite careful scrutiny of the various Fourier syntheses, we have not been able to identify any of the solvent peaks unambiguously as cations. It is likely that they are disordered and that some of the solvent positions identified are either water molecules or cations. Calculations based on a probable crystal density of 1.5 gcm⁻³ and unit cell volume of approximately 69,000 Å³ suggest that there can be about 400 solvent molecules in the asymmetric unit. That we observe less than 20% of these supports the notion of highly disordered bulk solvent.

Occupancy and thermal parameters (B) of solvent sites are corre-

The following convention is used in this paper: the nucleotides are numbered C1 to G12 in the 0-5' to 0-3' direction on stand 1 and C13 to G24 on stand 2.

1 D. Rabinovich, personal communication.
lated. In crystals which diffract to high resolution (1.0-1.5 Å), the number of experimental observations is sufficiently large to allow refinement of both positional occupancy and $B$ values. In the present case, at 2.5-Å resolution, such a rigorous treatment is not possible, and a choice must be made between refinement of occupancy while maintaining a fixed $B$ value or keeping the occupancy fixed and refining $B$. We adopted the latter strategy; and thus, all solvent molecules were treated as oxygen atoms with a fixed occupancy of 1.0.

The refinement was terminated after the location of 71 solvent molecules. The final residual was 0.18 for 2004 reflections with $I > \sigma(I)$ in the range 8 to 2.5 Å. The highest peak in the final difference map was 0.3 eÅ$^{-3}$. This peak lies close to the sugar phosphate backbone at G24 and may represent some disorder at this part of the structure. The refined coordinates will be deposited with the Brookhaven Protein Data Bank.

RESULTS AND DISCUSSION

**G·T Mismatch**—Fig. 1 shows a stereoview of the mismatch T9·G16 superimposed on a fragment electron density map, $F_c$-$F'_c$. In calculating the structure factors ($F'_c$) for this map, the atoms of nucleotides T9 and G16 and solvent molecules nearby were given a low occupancy factor. These atoms do not contribute to the phasing, and the electron density observed is an unbiased representation of where the atoms should be positioned. The positions of the bases are extremely well defined. At this high level of contouring (see figure legend), the peaks attributable to solvent molecules are not observed. Fig. 2 is a stereoview of the same mispair superimposed on a different fragment map. Now the nucleotide atoms have been given full occupancy, but solvent molecules have been left out in calculating $F'_c$. These solvent molecules, W1, W43, and W48, are well defined in this map.

Fig. 3 shows the mismatch T9·G16 and associated solvent molecules. As can be seen, the G·T mismatch adopts a wobble-type structure, with both bases in the major tautomer form. The thymine 0-4-keto group is displaced into the major groove, and the guanine amino group protrudes into the minor groove compared to a normal Watson-Crick pair. The hydrogen-bonding distances between the guanine N-1 and 0-6 and thymine 0-2 and N-3 groups, respectively, are in the range of 2.7-2.8 Å in the two mismatches. The wobble base pair is very similar to the G·T pairs observed in A- and Z-DNA fragments.

The disposition of the bases allows water molecules to form hydrogen-bonded bridges between the thymine 0-4 and 0-2 atoms and guanine 0-6 and N-2 atoms, respectively. Similar bridges were observed around the G·T pairs both at atomic resolution in the Z-DNA hexamer d(C-G-C-G-T-G) (7) and at 2.1-Å resolution in the A-DNA octamer d(G-G-G-T-C-C-C-C) (5). In the present work, at 2.5-Å resolution, well-defined solvent molecules are located at these bridging positions around the T9·G16 pair. The other mismatch, G4·T21 has a water bridging the 0-6(G4) and 0-4(T21) atoms in the major groove. There is a weak smear of positive electron density between the N-2(G4) and 0-2(T21) atoms in the minor groove which may be attributed to disordered solvent. Attempts to refine a water molecule in this position produced unacceptably high thermal parameters. A similar observation applies to a small positive peak of electron density approximately 3.2 Å from N-7(G4) in the major groove. This position is well ordered at N-7(G16), where the water molecule is also able to

![Fig. 2. Stereoview of T9·G16 mispair on fragment map showing peaks corresponding to solvent positions associated with this base pair. The solvent molecules W1, W43, and W48 (with $B$ values of 29 43 and 50 Å$^2$) have been omitted from the structure factor calculations. Their presence is well defined. The map is contoured between ±0.2 and ±0.5 eÅ$^{-3}$ at intervals of 0.07 eÅ$^{-3}$. Negative density is depicted with dashed lines; positive with continuous lines.](image-url)
FIG. 3. Guanine-thymine mismatch (G16-T9) and three associated solvent molecules. The hydrogen bonds are depicted as dashed lines with distances given in angstroms. Atoms are shown as spheres of decreasing radii in the order $P > O$ (water) $> O > N > C$. Atoms involved in hydrogen bonds are labeled.

TABLE I

<table>
<thead>
<tr>
<th>Residue</th>
<th>$\chi$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\gamma$</th>
<th>$\delta$</th>
<th>$\epsilon$</th>
<th>$\zeta$</th>
<th>$P_1 - P_{11}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>-120</td>
<td>-73</td>
<td>191</td>
<td>39</td>
<td>149</td>
<td>-179</td>
<td>-74</td>
<td>6.6</td>
</tr>
<tr>
<td>G2</td>
<td>-97</td>
<td>-73</td>
<td>191</td>
<td>39</td>
<td>149</td>
<td>-179</td>
<td>-74</td>
<td>6.6</td>
</tr>
<tr>
<td>C3</td>
<td>-147</td>
<td>-45</td>
<td>121</td>
<td>62</td>
<td>76</td>
<td>-157</td>
<td>-106</td>
<td>6.4</td>
</tr>
<tr>
<td>G4</td>
<td>-115</td>
<td>-32</td>
<td>136</td>
<td>64</td>
<td>114</td>
<td>-176</td>
<td>-112</td>
<td>6.9</td>
</tr>
<tr>
<td>A5</td>
<td>-104</td>
<td>-21</td>
<td>155</td>
<td>32</td>
<td>119</td>
<td>-195</td>
<td>-70</td>
<td>6.8</td>
</tr>
<tr>
<td>A6</td>
<td>-89</td>
<td>-21</td>
<td>155</td>
<td>32</td>
<td>119</td>
<td>-195</td>
<td>-70</td>
<td>6.8</td>
</tr>
<tr>
<td>T7</td>
<td>-107</td>
<td>-68</td>
<td>198</td>
<td>48</td>
<td>124</td>
<td>-178</td>
<td>-101</td>
<td>6.7</td>
</tr>
<tr>
<td>T8</td>
<td>-115</td>
<td>-71</td>
<td>168</td>
<td>75</td>
<td>115</td>
<td>-174</td>
<td>-84</td>
<td>7.1</td>
</tr>
<tr>
<td>T9</td>
<td>-127</td>
<td>-73</td>
<td>173</td>
<td>58</td>
<td>86</td>
<td>-150</td>
<td>-106</td>
<td>6.0</td>
</tr>
<tr>
<td>G10</td>
<td>-82</td>
<td>-46</td>
<td>165</td>
<td>37</td>
<td>155</td>
<td>-67</td>
<td>-333</td>
<td>6.9</td>
</tr>
<tr>
<td>C11</td>
<td>-101</td>
<td>-78</td>
<td>171</td>
<td>30</td>
<td>144</td>
<td>-196</td>
<td>-72</td>
<td>7.0</td>
</tr>
<tr>
<td>G12</td>
<td>-79</td>
<td>-72</td>
<td>234</td>
<td>18</td>
<td>132</td>
<td>-208</td>
<td>-69</td>
<td>6.2</td>
</tr>
<tr>
<td>G13</td>
<td>-109</td>
<td>-72</td>
<td>234</td>
<td>18</td>
<td>132</td>
<td>-208</td>
<td>-69</td>
<td>6.2</td>
</tr>
<tr>
<td>G14</td>
<td>-101</td>
<td>-95</td>
<td>219</td>
<td>55</td>
<td>128</td>
<td>-182</td>
<td>-104</td>
<td>6.2</td>
</tr>
<tr>
<td>C15</td>
<td>-141</td>
<td>-61</td>
<td>155</td>
<td>65</td>
<td>79</td>
<td>-175</td>
<td>-72</td>
<td>6.4</td>
</tr>
<tr>
<td>G16</td>
<td>-99</td>
<td>-65</td>
<td>184</td>
<td>65</td>
<td>142</td>
<td>-200</td>
<td>75</td>
<td>7.2</td>
</tr>
<tr>
<td>A17</td>
<td>-95</td>
<td>-79</td>
<td>220</td>
<td>57</td>
<td>132</td>
<td>-204</td>
<td>-61</td>
<td>6.6</td>
</tr>
<tr>
<td>A18</td>
<td>-100</td>
<td>-72</td>
<td>207</td>
<td>42</td>
<td>131</td>
<td>-175</td>
<td>-104</td>
<td>6.5</td>
</tr>
<tr>
<td>T19</td>
<td>-133</td>
<td>-60</td>
<td>157</td>
<td>62</td>
<td>87</td>
<td>-186</td>
<td>-65</td>
<td>6.7</td>
</tr>
<tr>
<td>T20</td>
<td>-100</td>
<td>-64</td>
<td>198</td>
<td>41</td>
<td>138</td>
<td>-177</td>
<td>-85</td>
<td>6.8</td>
</tr>
<tr>
<td>T21</td>
<td>-114</td>
<td>-67</td>
<td>174</td>
<td>55</td>
<td>83</td>
<td>-164</td>
<td>-85</td>
<td>6.0</td>
</tr>
<tr>
<td>G22</td>
<td>-81</td>
<td>-55</td>
<td>182</td>
<td>50</td>
<td>150</td>
<td>-115</td>
<td>-181</td>
<td>6.8</td>
</tr>
<tr>
<td>G23</td>
<td>-135</td>
<td>-71</td>
<td>138</td>
<td>44</td>
<td>51</td>
<td>-160</td>
<td>-86</td>
<td>6.5</td>
</tr>
<tr>
<td>G24</td>
<td>-136</td>
<td>-63</td>
<td>173</td>
<td>53</td>
<td>94</td>
<td>-196</td>
<td>-91</td>
<td>6.7</td>
</tr>
</tbody>
</table>

Average values for $d$(C-G-C-G-A-A-T-T-C-G-C-G)

Average values for $d$(C-G-C-G-A-A-T-T-C-G-C-G)

Main chain torsion angles defined by

$$P^0 - O(5')^2 - C(5')^2 - C(4')^2 - C(3')^2 - O(5')^2 - P$$

Glycosyl torsion angle ($\chi$) defined by

$$O(4') - C(1') - N(1) - C(2)$$

for pyrimidines

$$O(4') - C(1') - N(9) - C(4)$$

for purines

* Excludes Cl.

form a hydrogen bond to the water bridging 0-6(G16) and 0-4(T9).

Sugar Phosphate Backbone—The distances between adjacent phosphorus atoms in each strand range from 6.0 to 7.2 Å, with an average of 6.6 Å. In the native structure, the range is 6.2–7.1 Å, with an average of 6.7 Å (2). The P...P distances between residues 8 and 9 and 20 and 21, i.e., around the thymines T9 and T21 of the mismatch, are 0.7 and 0.2 Å shorter than in the native oligomer. The P...P distances involving the guanine of the G·T mismatches are equal to within 0.2 Å in the two structures. Interestingly, the presence of the two mismatches has little effect on the groove widths, which are defined in terms of the separation of the phosphorus atoms across the grooves with corrections for van der Waals radii (28). In the mismatch structure, the width of the minor groove varies from 8.8 Å (P(A5)...P(G24)) to 3.5 Å (P(T9)...P(T20)).
The major groove is more uniform, with values ranging from 12.5 Å (P(T7)...P(C15)) to 10.4 Å (P(A5)...P(A17)). Similar ranges were obtained in the native structure (4).

The backbone torsion angles and the torsion angles around the glycosyl bond are presented in Table I, which also lists average values and, for comparison, corresponding average values of the native dodecamer (2). The range of individual torsion angles is wider in the mismatch than in the native dodecamer, but the average values agree to within 10°. Evidently, the sugar phosphate backbone is sufficiently flexible to accommodate the G-T wobble pairs by small local adjustments while preserving the overall backbone conformation.

The values of $\chi$ fall into two groups depending on whether a purine or pyrimidine base is involved. Those for purines, including G4 and G16 which are involved in mismatches, range from -79° to -136° (average -98°); and those for pyrimidines, including T9 and T21 also involved in mismatches, range from -101° to -147° (average -121°). These values may represent steric differences between five- and six-membered rings adjoining the sugar moieties (3). The overall range of $\chi$ values for all bases is -81° to -147° in the mismatch and -88° to -135° in the native structure and corresponds to $\psi$-synclinal through to $\psi$-anticlinal orientation as defined by IUPAC-IUB Nomenclature (29).
Torsion angle $\delta$ ranges from 76° to 155° (+syndclinal to +antiprerpilanoal). This torsion angle is correlated with the conformation of the furanose ring (30) and, in the present structure, corresponds to conformations C-3'-endo through C-3'-exo; it is also correlated with the glycosidic bond torsion angle $\chi$. As noted above, $\chi$ has a bimodal distribution depending on whether the base is a purine or pyrimidine. Lower $\delta$ values are observed for pyrimidines (average = 105°) than for purines (average = 132°). This tendency was attributed by Dickerson (32) to cross-chain steric hindrance between purine bases. The lowest $\delta$ values are observed for the furanose rings of C3 and C15: 76° and 79° respectively. These sugar moieties are involved in van der Waals-type packing with symmetry-related duplexes, and their conformations may be affected by this interaction.

The conformations of the other sugar phosphate torsion angles may be summarized as follows. Torsion angle $\alpha$ ranges from -synderiplanar to -antiplanar, $\beta$ varies from -synderiplanar to -antiplanar, whereas $\gamma$ is restricted to +synderiplanar when $\gamma$ C1 is ignored. The remaining torsion angles $\epsilon$ and $\zeta$ fall in the range -synderiplanar to +synderiplanar. The largest values of $\epsilon$ are observed at G10 (-233°) and G22 (-181°), but they appear to be a characteristic of the dodecamer structure since in the native duplex the corresponding values are -210° and -188°. There is no indication, from an examination of the intermolecular contacts in this area, that this variation is related to crystal packing effects.

**Base Stacking Interactions**—Perturbations of base stacking interactions due to the presence of mismatch base pairs may play a role in the mechanism by which fidelity of replication is maintained. In d(C-G-C-G-A-A-T-T-T-G-C-G), there are 11 base pair steps of which seven involve only Watson-Crick base pairs. These steps are steps 1 and 11 of type CpG(=CpG), steps 2 and 10 of type GpC(=GpC), steps 5 and 7 of type ApA(=TpT), and step 6 of type ApT(=ApT). These base pair steps all occur in the parent dodecamer. The base stacking interactions at these steps are similar to those found in the native dodecamer. There is considerable variation in base pair overlap from one step to another. Purine-pyrimidine steps (steps 2, 6, and 10) display a greater degree of overlap than the pyrimidine-purine steps (steps 1 and 11). The ApA(=TpT) steps show good base overlaps, with the six-membered rings of the purine bases interacting strongly.

Changes in base stacking are localized at the mismatch sites. There are four base pair steps which involve G-T mispairs. These are steps 3 and 9 of type CpG(=CpG) and steps 4 and 8 of type GpA(=TpT). We observe the same stacking interactions around both mismatch sites; and therefore, only one set, that involving G4-T21, is illustrated (Fig. 4) and compared with the equivalent set in the native structure. In the parent compound, step 3, CpG(=CpG), shows a moderate interaction of C21 with G22 but a weak interaction between G4 and C3. The corresponding step in the mismatch dodecamer shows that the G4...C3 interaction remains conserved. However, on the other strand, there is a deterioration in the pyrimidine...purine overlap as T21 is displaced into the major groove and G4 is displaced into the minor groove. Step 4 in the parent compound, GpA(=TpT), shows good overlap between the purines. The six-membered ring of G4 interacts with both the five- and six-membered rings of A5. The degree of pyrimidine...pyrimidine overlap is quite small. The good overlap between purines is retained at step 4 in the mismatch structure. Moreover, the pyrimidine...pyrimidine overlap between T20 and T21 is greatly enhanced in the mismatch structure.

The parameters of Dickerson and Drew (Ref. 2; see also Refs. 31 and 32) allow further analysis of base pair interactions. In Table II, we present some of the geometrical properties of base pair steps and base pairs in d(C-G-C-G-A-A-T-T-T-G-C-G).

**Table II**

<table>
<thead>
<tr>
<th>Base pair</th>
<th>Step</th>
<th>Roll (°)</th>
<th>Twist (°)</th>
<th>Rise Propeller</th>
<th>$\lambda_1$ (°)</th>
<th>$\lambda_2$ (°)</th>
<th>C-1'...C-1'</th>
</tr>
</thead>
<tbody>
<tr>
<td>A6-T19</td>
<td>6</td>
<td>-5.8</td>
<td>33.1</td>
<td>3.0</td>
<td>13.4</td>
<td>7.2</td>
<td>44.6</td>
</tr>
<tr>
<td>T7-A18</td>
<td>7</td>
<td>0.3</td>
<td>34.6</td>
<td>3.2</td>
<td>17.8</td>
<td>59.6</td>
<td>55.2</td>
</tr>
<tr>
<td>T8-A17</td>
<td>8</td>
<td>0.9</td>
<td>34.5</td>
<td>3.2</td>
<td>13.4</td>
<td>7.2</td>
<td>44.6</td>
</tr>
<tr>
<td>T9-G16</td>
<td>9</td>
<td>4.6</td>
<td>33.5</td>
<td>4.3</td>
<td>11.5</td>
<td>5.6</td>
<td>55.7</td>
</tr>
<tr>
<td>G10-C15</td>
<td>10</td>
<td>-7.2</td>
<td>39.7</td>
<td>2.8</td>
<td>17.0</td>
<td>58.1</td>
<td>52.8</td>
</tr>
<tr>
<td>C11-G14</td>
<td>11</td>
<td>-1.0</td>
<td>45.1</td>
<td>3.5</td>
<td>2.3</td>
<td>56.0</td>
<td>51.2</td>
</tr>
<tr>
<td>G12-C13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Twist** is the rotation from one base pair to the next with respect to the helix axis. The average twist values for the mismatch and native structures are 36.2° and 35.8°, respectively, and correspond to 10 base pairs/turn of helix. The largest difference in twist between the two structures (+8.1°) is observed at step 11. At steps involving the G-T mispairs (steps 3, 4, 8, and 9), the differences are +2.4°, -3.2°, -6.4°, and +3.1°.
Rise per base pair step shows a decrease of 0.6 and 0.4 Å at steps 2 and 10 and an increase of 1.0 and 0.7 Å at steps 3 and 9 when compared to the native structure. The remaining steps all agree to within 0.3 Å in the two structures. The increase in rise per base pair at steps 3 and 9, CpG(=TpG), is associated with a disruption of the base stacking interactions as mentioned earlier. This increase in rise on the 5' side of the G-T mismatches is compensated for by a compression in the adjacent base pair steps (steps 2 and 10).

The angle by which adjacent base pairs open up to the helix grooves is termed the roll. In the parental compound, purine-pyrimidine base steps, GpC(=GpC) and ApT(=ApT), open out toward the major groove at steps 2, 6, and 10 with roll values of -3.2°, -6.0° and -9.8°. Pyrimidine-purine base steps open the other way. Steps 1, 3, and 9 have roll angles of 7.0°, 5.0°, and 1.7°, respectively (2, 4). The same trend is observed in the mismatch dodecamer (see Table II) and has been explained on the basis of avoidance of steric clashes between purines from opposite strands (32).

Propeller twist is the dihedral angle between individual base planes viewed along the long axis of a base pair. In the native structure, the range is 4.9–17.8°; in the mismatch dodecamer, it is 2.3–19.1°. We observe no major differences in propeller twist consequent to the introduction of two G-T mismatches.

**Thermal Parameters**—The isotropic thermal parameters for individual atoms were refined against the diffraction data together with the coordinates. Final thermal parameters (B = 8π2uÅ2) ranged from 12 to 42 (average = 22) for the bases, from 24 to 48 (average = 35) for the sugars, and from 34 to 57 (average = 43) for the phosphates. These values (which were observed at 6°C) are slightly lower than those observed in the native structure at room temperature when the ranges were 20–44 (average = 28) for bases, 32–51 (average = 42) for the sugars, and 38–61 (average = 50) for the phosphates (calculated from coordinates deposited with the Brookhaven Protein Data Bank). The thermal parameters, averaged for each base, sugar, and phosphate moiety and presented in Table III, do not give any indication of disorder or instability at the mismatch sites when compared both to the native structure and to those sections of the mismatch dodecamer which contain Watson-Crick base pairs. The solvent molecules have B values ranging from 25 to 78 Å2 with an average of 56 Å2. The B values associated with all solvent positions are also presented in Table III. Those solvent molecules with high thermal parameters can be expected to represent ordered or partially occupied sites. As discussed previously, there are not enough data at a high resolution to allow for occupancy refinement. It is worth noting that the three solvents, W1, W43, and W48, associated with the T9-G16 mispair have B values of 29, 43, and 50 Å2, respectively. These solvent molecules are very well defined in the electron density maps; an example is presented in Fig. 2.

**Hydration**—Oligonucleotide crystals are heavily hydrated, with unit cells containing approximately 50% w/v bulk solvent. In one sense, this is an advantage since the crystal structure is likely to be a good model for DNA in vivo. Crystallographic techniques, when applied to proteins and oligonucleotides, are only able to locate the most static or locally stable solvent positions. The assignment of solvent molecules can at times be subjective. We believe that the rather conservative approach to solvent assignment, discussed under "Materials and Methods," may omit some solvent positions but was required to reduce chances of introducing artifacts into the refinement.

There are three observations which support the relevance of deduction based on solvent molecule identification at the medium resolution of 2.5 Å of the present analysis. First, a recent crystal structural study of d(G-G-BrU-A-BrU A-C-C) (where BrU represents 5-bromouridine) (33) extended the resolution from about 2.2 to 1.7 Å with the use of a high intensity synchrotron source for data collection. Increasing the resolution led to the identification of only three additional solvent molecules and the omission of only a few solvent positions in the outer shell of hydration. The overall pattern of hydration remained conserved. Second, the pattern of hydration observed in several variants of the dodecamer structure is similar, although each refinement process was carried out in different laboratories and with different refinement procedures (34–36). This would suggest that any trends observed are real and not biased toward any particular criteria for identification of solvent or affected by the choice of refinement techniques. Third, the hydration of the T9-G16
Fig. 5. Stereoview of d(C-G-C-A-A-T-T-T-G-C-G) with solvent molecules located in course of analysis. The top left base is C1. The well-ordered solvent molecules in the narrow A-A-T-T region of the duplex, as observed in the native structure (35, 36), are evident.

Mispairs in B-DNA

mispair in this work is almost identical to the hydration of G-T mismatches in A-DNA at 2.2 Å (5) and Z-DNA at near-atomic resolution of 1.2 Å (7). Our experience of not being able to unambiguously identify cations is common to most oligonucleotide structures so far studied. The only exception is Z-DNA, crystals of which diffract x-rays to atomic resolution (33).

Fig. 5 shows the mismatch dodecamer together with the 71 independent solvent positions located in the analysis. All of the solvent molecules identified have contacts in the range of 2.5-3.4 Å. Of the 71 independent solvent molecules identified, 62 have contacts with one or more residues of the DNA, and 45 of these have additional contacts to one or more other solvent molecules. The remaining nine solvent molecules have contacts only to solvent. The deoxyribose phosphate backbone interacts with 44 of the solvent molecules; the bases interact with 21. In the major groove, the dodecamer base pairs provide 10 hydrogen bond donors (N-4 and N-6 of cytosine and adenine, respectively) and 26 hydrogen bond acceptor groups (N-7 of adenine, O-6 and N-7 of guanine, and O-4 of thymine). Of these 36 groups, 19 are observed to be hydrated, 7 donors, and 12 acceptors.

On the minor groove side of the base pairs, there are 32 functional groups, 8 hydrogen bond donors (N-2 of guanine), and 24 hydrogen bond acceptors (O-2 of cytosine and thymine and N-3 of adenine and guanine). Solvent is found to be associated with only seven of these groups, mainly in the narrow A-A-T-T center of the duplex where a hydration pattern similar to that observed in the native structure is formed (35). There are a further 10 functional groups in the minor groove which participate in interduplex hydrogen bonds and are thus blocked to solvent molecules.

Along the sugar phosphate backbone, the O-1(P) atoms are directed toward the major groove, and the O-2(P) atoms are orientated away from the double helix toward bulk solvent. The phosphate groups are fully hydrated both in the present and native structures (35, 36). We observe no preference for hydration of O-1(P) or O-2(P). Of the 44 free phosphate oxygen atoms, 28 are associated with solvent molecules, 14 O-1(P) and 14 O-2(P). The solvent molecules tend to form localized clusters around individual phosphate groups, which seems to be a characteristic of the mixed sequence B-DNA conformation (37). Of the 48 O-3' and O-5' atoms, only 11 are hydrated. The furanose oxygen atoms also tend to be sparsely hydrated; only 8 out of 24 O-4' atoms are seen to be associated with well-defined solvent molecules.

Conclusions—The detailed analysis of the dodecamer d(C-G-C-G-A-A-T-T-T-G-C-G) confirms a number of observations based on single crystal x-ray studies of DNA fragments with mismatched base pairs (5-7, 19-21). Noncomplementary bases are able to form stable base pairs linked by two hydrogen bonds. The global conformation of the double helix remains conserved, with any changes in the sugar phosphate backbone highly localized. The base overlap between the mismatch and adjacent base pairs, although altered in comparison to Watson-Crick structures, is still extensive and may contribute to the stabilization of mismatches in these structures.

The hydration of polar groups makes an important contribution to thermodynamic stability (38). Potential hydrogen bond donors and acceptors will always be satisfied when accessible to solvent. The hydrogen bonds formed by the free functional groups of mismatched base pairs with the solvent are necessarily different from those which can be formed by Watson-Crick base pairs. These hydrogen bonds will contribute to mismatch stability and may be a significant factor in determining thermodynamic properties of the different mispair combinations.

The change in base pair geometry upon misinsertion provides a number of ways in which an enzyme could, by highly specific hydrogen-bonding interactions, recognize a mismatch. A prominent feature of Watson-Crick base pairs is a pseudodyad relating the glycosidic bonds of the bases. This symmetry may be important for the initial recognition of Watson-Crick base pairs, and further discrimination could then be provided by directed hydrogen bonds to exposed functional groups (27). In Table II, we present values for λ, the angle formed between glycosyl bonds and the C-1'-...C-1' vector. For instance, λ1 is the N-1(C1)...C-1'(C1)...C-1'(G24) angle. The complementary angle, λS, is N-9(G24)...C-1'(G24)...C-1'(C1). In Watson-Crick base pairs, these values are similar whether subtended at a purine or a pyrimidine base. In d(C-G-C-G-A-A-T-T-T-G-C-G), the largest differ-
ence between $\lambda_1$ and $\lambda_2$ for the Watson-Crick base pairs is 5.8° at G-C.22. The G-T mismatches, however, display a much wider difference of 27.4° in both mismatches. This combination of bases leads to the formation of an asymmetric base pair.

The C-A mismatches in d(C-G-C-A-A-A-T-T-C-G-C-G) are also asymmetric with values differing by 13.3° and 23.0° for the two mismatches (20). The two G-A mismatches in d(C-G-C-G-A-A-T-T-C-G-C-G) have differences of 14.4° and 22.8° in the $\lambda$ values. These three mismatches may be arranged in the order G-T > C-A > G-A, depicting the degree of asymmetry compared to Watson-Crick base pairs. This is precisely the order of efficiency, observed in a number of studies, in which each type of mismatch is recognized and repaired (13-15). It is possible that mismatch enzymes are able to recognize a non-Watson-Crick base pair on the basis of a deviation from pseudosymmetry.

In addition to the changes in $\lambda$ values, we note perturbations in the rise per base pair at steps involving G-T, G-A, and C-A mismatches when compared to the native structure. Each mismatch dodecamer contains two mispair sites involving residues 4 and 21 and 9 and 16. Steps 3 and 9 are referred to as being on the 5' side of the mispair, and steps 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 per base pair at steps 3 and 9 is 3.6 and 3.3 Å. For the G-T mismatch, both 5' steps have rise values of 4.3 Å and for the G-A mismatch, rise values of 4-2; whereas in the C-A structure, values of 4.1 and 4.3 Å are observed.

Steps 4 and 8 have rise values of 3.1 and 3.3 Å in the native structure. The G-T structure gives rise values of 2.7 and 3.3 Å at corresponding steps, and the C-A structure gives values of 3.5 and 3.1 Å. The G-A structure has values of 3.1 Å at both steps.

The best repaired mismatch, G-T, is observed to produce the largest perturbation in rise per base pair in this dodecamer sequence. C-A and G-A produce changes very similar to one another. We would expect that rise per base pair which is correlated with helical twist and base stacking could be a parameter strongly influenced by the nature of adjoining nucleotides. It may well be a factor with the observed variation of mismatch repair with base sequence (16). Further studies of mismatch structure and sequence effects are required to aid the understanding of the many complex processes involved in mismatch recognition and repair.

Acknowledgments—We thank the Medical Research Council Laboratory of Molecular Biology for the use of their Evans & Sutherland graphics system, Drs. M. J. McCall, P. R. Evans, and Z. Shakked for discussions, and H. Belchamber for preparation of the manuscript. Our final thanks go to Dr. Richard Brennan for constructive and thorough criticism.

REFERENCES

23. Sheldrick, G. M. (1976) SHELX 76 System of Computer Programs, University of Cambridge, Cambridge, United Kingdom