Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 2402–2406, April 1986
Biochemistry

Molecular structure of the G·A base pair in DNA and its implications for the mechanism of transversion mutations

(x-ray diffraction/mutagenesis/deoxyolignonucleotides/base stacking/tautomer hypothesis)

TOM BROWN*, WILLIAM N. HUNTER, GEOFF KNEALE†, AND OLGA KENNARD‡

University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, United Kingdom

Communicated by Lord Todd, December 19, 1985

ABSTRACT **The synthetic deoxydodecamer d(C-G-C-G-A-A-T-T-A-G-C-G) was analyzed by x-ray diffraction methods, and the structure was refined to a residual error of R = 0.17 at 2.5-Å resolution (2 σ data) with 83 water molecules located. The sequence crystallizes as a full turn of a B-DNA helix and contains 2 purine·purine (G·A) base pairs and 10 Watson–Crick base pairs. The analysis shows conclusively that adenine is in the *syn* orientation with respect to the sugar moiety whereas guanine adopts the usual *trans* orientation. Nitrogen atoms of both bases are involved in hydrogen bonding with the N-1 of guanine 2.84 Å from the N-7 of adenine and the N-6 of adenine within 2.74 Å of the O-6 of guanine. The C-1′···C-1′ separation is 10.7 Å close to that for standard Watson–Crick base pairs. The incorporation of the purine·purine base pairs at two steps in the dodecamer causes little perturbation of either the local or the global conformation of the double helix. Comparison of the structural features with those of the G·T wobble pair and the standard G·C pair suggests a rationale for the differential enzymatic repair of the two types of base-pair mismatches.**

Investigations of synthetic DNA fragments by single crystal x-ray diffraction methods at near atomic resolution have brought about a reappraisal in our views of the conformational flexibility of DNA and the role of structural features in DNA–protein recognition (1). These studies have been extended to DNA fragments containing noncomplementary base pairs both in the solid state (2–5) and in solution (6, 7). The present paper gives an account of the first crystal structure determination of a full turn of a B-DNA helix incorporating two purine–purine base-pair mismatches.

The significance of noncomplementary base pairs (Fig. 1) and their potential involvement in spontaneous mutation was recognized by Watson and Crick in one of their early publications (8).

Noncomplementary bases may be incorporated into DNA during replication, during genetic recombination or, in the case of the G·T mispair, by deamination of 5-methyl cytosine (9). Such errors are generally detected and eliminated by proofreading and mismatch repair enzymes. Occasionally, however, the error is not corrected, leading to substitution mutations and possibly to evolutionary changes. Purine·purine base pairs specifically are involved in transversion mutations. The replacement of a purine·pyrimidine base pair by two large purine bases was thought to place considerable strain on the double helix (10), particularly if both bases are in the usual *anti* orientation with respect to the sugar-phosphate backbone, as illustrated in Fig. 1*f*. Base pairing between *N*-dimethyl guanine and adenine, with both bases in the *anti* orientation, was observed by several workers in crystal structures of tRNA (11, 12). G(*anti*)·A(*anti*) base pairs were also deduced from NMR solution studies of two synthetic deoxyoligonucleotides. Alternative base-pairing schemes between adenine and guanine have been proposed in which one of the bases is in the *anti* and one in the *syn* orientation so as to reduce the distance between the C-1′··· C-1′ atoms on opposite sides of the minor groove of the double helix. Topal and Fresco (13) postulated the involvement of a rare tautomer form and the formation of G(*syn*)·A(*imino*) and G(*enol, imino*)·A(*syn*) base pairs (Fig. 1 *c* and *d*) to account for the observed frequency of transversion mutations. The G(*anti*)·A(*syn*) (Fig. 1*b*) base pair, with both bases in the major tautomer form, was proposed by Traub and Sussman (14) for ribosomal RNAs. Theoretical investigations of the energetics of the different base-pairing schemes have been reported (15, 16).

We have analyzed the crystal structure of the synthetic deoxydodecamer d(C-G-C-G-A-A-T-T-A-G-C-G) to a resolution of 2.5 Å and have established that stable G(*anti*)·A(*syn*) base pairs are formed and are accommodated in the DNA double helix with little disruption of the local or global conformation.

MATERIALS AND METHODS

Synthesis, Crystallization, and Data Collection. The deoxydodecamer d(C-G-C-G-A-A-T-T-A-G-C-G) was synthesized by the triester method using a solid-phase technique (17). Purification was by ion-exchange chromatography and by reverse phase HPLC. Crystals were grown from an aqueous solution containing 0.5 mM DNA duplex, 7.4 mM sodium cacodylate buffer (pH 7.4), 24 mM $Mg(OAc)_2$, 1.0 mM spermine, and 18% (vol/vol) 2-methyl-2,4-pentanediol. The deoxydodecamer crystallizes in the orthorhombic system with space group $P2_12_12_1$. A series of precession photographs suggested isomorphism with the deoxydodecamer d(C-G-C-G-A-A-T-T-C-G-C-G) analyzed by Dickerson and coworkers (18) and with the dodecamer d(C-G-C-G-A-A-T-T-T-G-C-G) containing two G·T mismatches (2, 3). This was confirmed by the unit cell dimensions of a = 25.69, b = 41.96, and c = 65.19 Å, as determined from diffractometer measurements. An assumed crystal density of 1.5 g·cm^{-3} indicates the presence of one double-stranded dodecamer per asymmetric unit and a DNA content of approximately 50% by weight in the unit cell.

Intensity data were collected at 5°C using a Syntex $P2_1$ diffractometer, equipped with a monochromator (λ = 1.5418 Å), a long arm and helium path. A crystal of dimensions 1.30 × 0.15 × 0.15 mm was mounted and sealed in a capillary. The dimensions of the specimen were not optimal, but no other crystal habit was available. Intensities were corrected for Lorentz and polarization factors, absorption, and time-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

*Present address: Department of Chemistry, University of Edinburgh, Edinburgh, United Kingdom.
†Present address: Biophysics Laboratory, Portsmouth Polytechnic, Portsmouth, United Kingdom.
‡To whom reprint requests should be addressed.

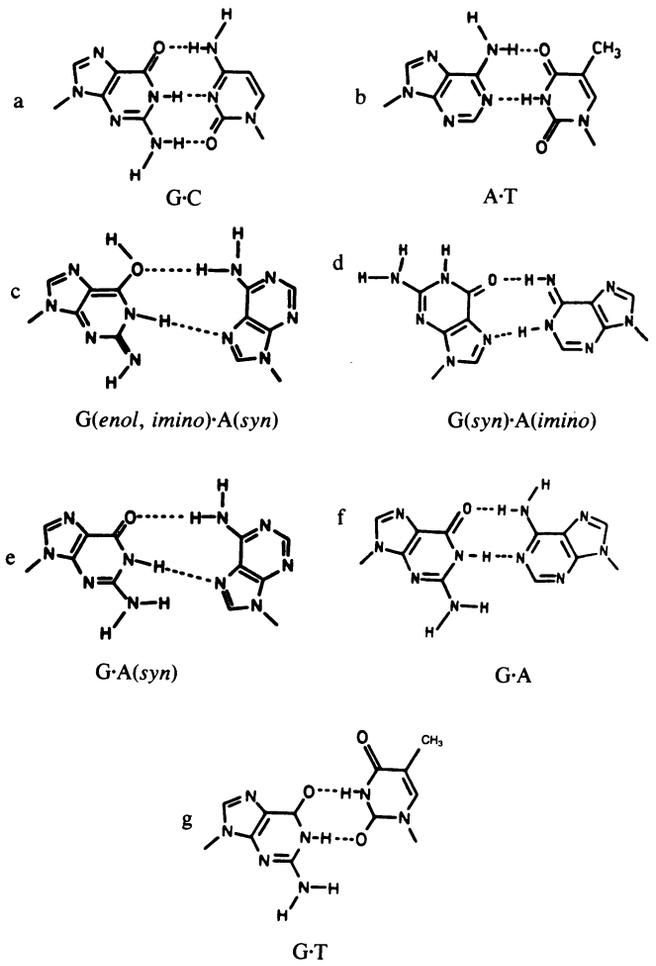

FIG. 1. Observed and postulated base-pairing schemes. Watson–Crick base pairs (*a, b*), tautomer pairs (*c, d*), two types of G·A pairs (*e, f*), G·T wobble base pair (*g*).

dependent decay. After averaging, 2707 independent reflexions were obtained to a resolution of 2.5 Å.

Structure Analysis. A molecular model of the dodecamer was derived from the parent sequence d(C-G-C-G-A-A-T-T-C-G-C-G) using B-DNA fiber coordinates (19). When constructing the model the bases were labeled C^1 through G^{12} in the 5′ to 3′ direction on strand 1, C^{13} through C^{24} on strand 2. By this convention $G^4{\cdot}A^{21}$ is referred to as the upper base-pair mismatch. Adenosine was substituted for cytosine at steps 9 and 21 and positioned by calculating the best molecular fit of the sugar and phosphate groups. The *anti* orientation was chosen for adenine since the G(*anti*)·A(*anti*) base pair was the one deduced from solid state (11, 12) and solution (6, 7) studies. We relied on refining this model against the experimental data and checking its validity by inspection of electron density maps calculated with refined phases.

The fiber model was positioned in the unit cell by best molecular fit to the coordinates of the parent dodecamer obtained from Westhof and Prange (personal communication) based on data by Dickerson and coworkers (18).

A rigid body refinement was carried out at low resolution with the program SHELX (20). The resolution was gradually extended, and the model was refined by the constrained/restrained least-squares program CORELS (21). During these calculations the mismatched bases were allowed to move freely relative to each other, while the Watson–Crick base pairs were restrained to normal hydrogen bonding. The refinement converged at $R = 0.28$ for 1298 reflexions $I > 2\sigma(I)$ between 8 and 2.5 Å.

Various electron density maps were next prepared to check the atomic positions and to locate solvent molecules. We used, in addition to F_o-F_c and $2F_o-F_c$ maps, "fragment" F_o-F_c' electron density maps where in calculating F_c', specific regions (fragments) of the molecule were omitted. The maps were inspected on an Evans and Sutherland PS300 system. The G(*anti*)·A(*anti*) base pairs appeared, at this stage, to be acceptable, and the positions of a number of solvent molecules were clearly defined.

Refinement was continued with NUCLSQ (22), a modified version of the Konnert–Hendrickson PROLSQ program (23) for treatment of nucleotide structure. No restraints were placed on the conformations of the sugar moieties, the sugar-phosphate torsion angles, or, as mentioned above, the hydrogen bonding between the mismatched bases. Temperature factors for individual atoms were included as variables. Solvent positions were identified by using a peak search program and by inspection of maps on the graphics system and were gradually included in the calculations. Fragment maps were also used to inspect those regions of the sugar-phosphate backbone that were poorly defined due to high thermal vibration and/or disorder. After 50 solvent molecules had been located ($R = 0.20$), the Fourier syntheses and notably the fragment maps calculated with the mismatched base pairs omitted indicated significantly better electron density fit with the G(*anti*)·A(*syn*) than with the G(*anti*)·A(*anti*) base pair. The model was corrected, and the calculations were continued with the location of further solvent molecules. The refinement converged at $R = 0.17$ for the 2028 reflections between 8 and 2.5 Å ($Rw = 0.17$) with the inclusion of 83 solvent molecules, treated as oxygen atoms. There was no evidence, in the final electron density maps, of any disorder in the regions of the G(*anti*)·A(*syn*) base pairs (see Fig. 3).

RESULTS

The Structure of d(C-G-C-G-A-A-T-T-A-G-C-G). Two strands of the deoxydodecamer that are chemically equivalent constitute the asymmetric repeat unit of the crystal structure. They form an antiparallel double helix with 10 standard Watson–Crick base pairs and 2 mismatched base pairs at $G^4{\cdot}A^{21}$ and $A^9{\cdot}G^{16}$. The double helix, illustrated in Fig. 2, is of the B-DNA type.

In the crystal structure the double helices are inclined to the *c* crystallographic axis and staggered so that the minor grooves of adjacent helices are interlocked. The terminal G·C base pair of one helix is parallel to and in van der Waals contact with that of another helix related by the 2_1 operator parallel to the *c* axis.

The Geometry of the G·A(*syn*) Base Pairs. The geometry of the two G·A(*syn*) base pairs $G^4{\cdot}A^{21}$ and $G^{16}{\cdot}A^9$ is identical within the accuracy of the analysis. A view of the upper base pairs is given in Fig. 3, where it is superimposed on the electron density of a fragment map.

The mismatched base pair in this figure clearly has guanine in the *anti* and adenine in the *syn* orientation and accounts satisfactorily for all the electron density observed in this region. There is no residual electron density corresponding to the adenine in the *anti* orientation and thus no suggestion that the structure contains both G(*anti*)·A(*anti*) and G(*anti*)·A(*syn*) base pairs. The *syn* orientation of bases has so far only been observed in Z-DNA (24) where the regular alternation of *syn* and *anti* bases results in the double helix adopting the left-handed Z-conformation and in a DNA–triostin A complex (25), where adenine forms a Hoogstein base pair with thymine. In the present structure the handedness of the B-DNA helix is preserved.

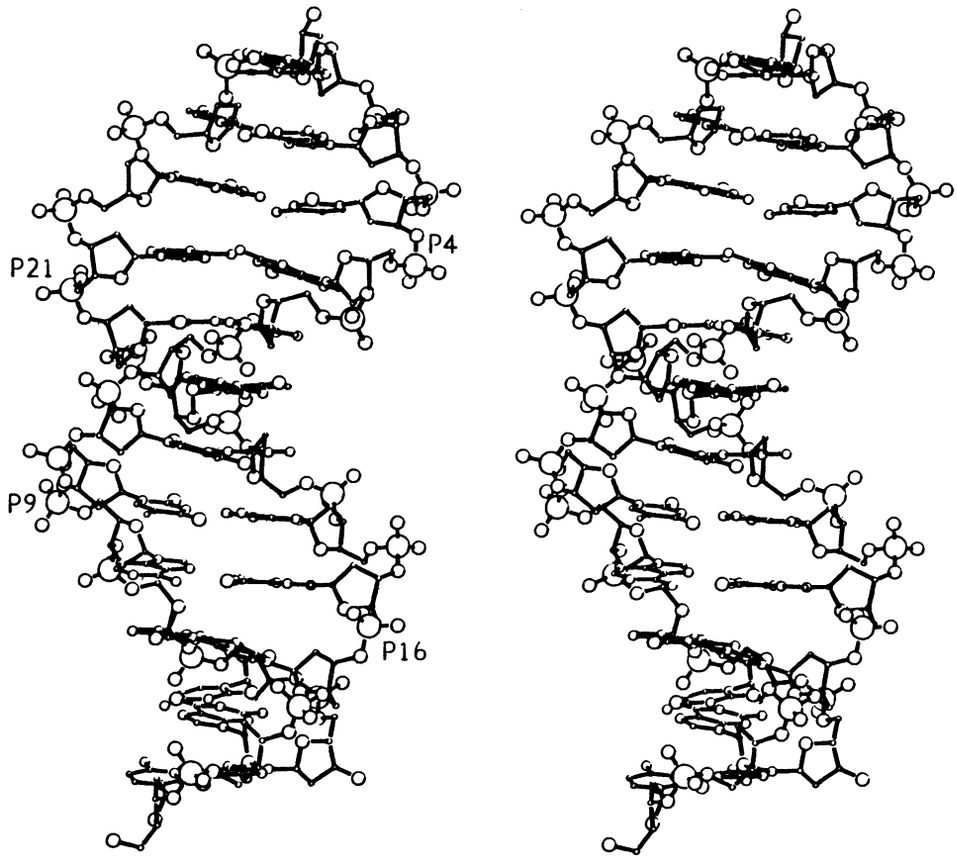

FIG. 2. Stereo diagram of the mismatch dodecamer d(C-G-C-G-A-A-T-T-A-G-C-G). The G·A(*syn*) base pairs are at positions G^4·A^{21} and A^6·G^{19}. The phosphorous atoms of these residues are labeled. The double helix is viewed from the side, so as to highlight the major and minor grooves. Note the large propellor twists that are characteristic of B-DNA helices observed in crystal structures and are not specifically related to the presence of the two G·A(*syn*) mismatches.

The two purine bases are joined by hydrogen bonds between N-1 of guanine and N-7 of adenine and O-6 of guanine and N-6 of adenine. The distances for the two base pairs are 2.85 and 2.82 Å for NH···N and 2.74 and 2.75 Å for the NH···O separation corresponding to normal values. The mismatched bases are tilted with respect to each other and have propellor twists of 17.5° and 11.8° at G^4·A^{21} and A^9·G^{16}, respectively. The local helical twists are 27.6° and 29.4°. These values fall well within the range of those for the Watson–Crick base pairs in the rest of the dodecamer (Table 1).

Table 1 also contains values for the C-1′···C-1′ separations. The values for the two G·A(*syn*) base pairs of 10.81 and 10.56 Å are similar to the corresponding values for the 10 Watson–Crick base pairs in the rest of the dodecamer and with values of 10.44 Å for A·T (26) and 10.72 Å for the G·C (27) obtained in small molecule structures. The angle between the glycosyl bonds (C-1′···N-1 in pyrimidines and C-1′···N-9 in purines) is also listed in this table and will be discussed below.

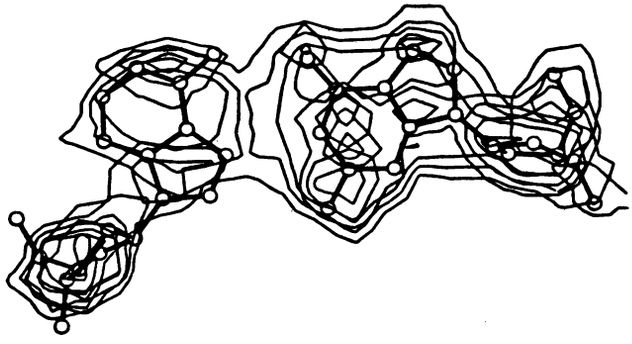

FIG. 3. The base-pair mismatch G^4·A^{21}(*syn*) superimposed on a composite section of the "fragment" electron density map, $F_o - F_c'$. All atoms shown in the diagram were omitted from the structure factor (F_c') calculations.

Conformation of the Double Helix. The substitution of the two G·A(*syn*) base pairs has surprisingly little effect on the global conformation of the double helix. The average global twist angle is 36.0° compared with 37.3° in the native dodecamer and 36.0° in the B-DNA fiber model. The average rise per residue is 3.4 Å (3.3 and 3.4 Å for native and fiber helices, respectively), and the roll and tilt angles are characteristic of B-DNA. Values at individual steps are also comparable to those reported for the native structure (28) with the exception of the C^3-G^4 and A^9-G^{10} steps where the rise per residue is significantly larger and at the G^2-C^3 and

Table 1. Characteristics of base-pairs in the mismatch dodecamer d(C-G-C-G-A-A-T-T-A-G-C-G)

Base pair	Propellor twist, degrees	C-1′···C-1′, Å	λ_1, degrees	λ_2, degrees
C^1·G^{24}	11.2	10.60	53.6	53.6
G^2·C^{23}	16.9	10.35	53.0	59.8
C^3·G^{22}	16.2	10.60	52.3	55.2
G^4·A^{21}	17.5	10.81	55.2	40.8
A^5·T^{20}	11.7	10.22	56.2	59.1
A^6·T^{19}	18.8	10.05	55.2	61.4
T^7·A^{18}	15.1	10.10	60.8	55.7
T^8·A^{17}	13.0	10.26	57.8	48.6
A^9·G^{16}	11.8	10.56	39.5	62.3
G^{10}·C^{15}	11.2	10.62	58.6	50.7
C^{11}·G^{14}	18.2	10.67	51.7	48.4
G^{12}·C^{13}	10.2	10.72	51.1	52.5

λ_1 is the angle between glycosidic bond C-1′–N-1(9) and the C-1′···C-1′ vector of a base on strand 1, and λ_2 is the angle between glycosidic bond C-1′–N-1(9) and the C-1′···C-1′ vector of a base on strand 2.

G^{10}-C^{11} steps where it is smaller than in the native structure. This compression and extension may be related to changes in base stacking. The presence of the two purine·purine base pairs does not appear to perturb the conformation of the sugar-phosphate backbone significantly. This is indicated in Table 2 where the average torsion angles of the sugar-phosphate backbone of the mismatch dodecamer are compared with average values for the native dodecamer and for a B-DNA fiber. The values for the mismatch dodecamer are intermediate between the two.

The conformations of the deoxyribose rings in crystal structures of B-type dodecamers appear to be much less uniform than in crystal structures of A-type helices (29). In the present structure, as in the native dodecamer (28), they range between C-2′ (*endo*) and C-3′ (*endo*) with δ values between 70.5° and 155°. A detailed description of the local conformational features, base stacking, and hydration pattern will be reported elsewhere.

Comparison with Related Studies of G·A Base-Pair Mismatches. X-ray studies have been reported on single crystals of yeast phenylalanine tRNA (11, 12) where a mismatched base pair, m^2G^{26}·A^{44}, with both bases in the *anti* orientation, was found in the D loop. In that structure the N-1(G)···N-1(A) and O-6(G)···N-6(A) distances are 3.3 and 3.1 Å (30), respectively, and the propellor twist between the bases is large (25°) to relieve steric clash. The tRNA and the present DNA studies are not directly comparable, partly because of the presence of a methyl substituent on the guanine and partly because of the very different local environments in the two structures. Traub and Sussman (14) have proposed, from an analysis of the secondary structure of ribosomal RNA, that the G(*anti*)·A(*syn*) base pair may be a better model than the G(*anti*)·A(*anti*) pair to fit in the end of an undistorted RNA double helix. Our observation of the G(*anti*)·A(*syn*) base pair in double helical DNA lends support to this suggestion.

NMR investigations on G·A mismatches have been reported for d(C-G-A-G-A-A-T-T-C-G-C-G) (6) and d(C-C-A-A-G-A-T-T-G-G) (7). In both studies the spectra have been interpreted to represent the presence of G(*anti*)·A(*anti*) pairs. Theoretical studies indicate that the G(*anti*)·A(*anti*) and the G(*anti*)·A(*syn*) base pairs are energetically near equivalent (15, 16). It may well be that both types of mismatches are formed in DNA (and RNA) helices, dependent on the local sequence environment and helix conformation.

DISCUSSION

Fidelity of replication depends on the frequency of misinsertion of noncomplementary bases, on the efficiency of proofreading, and on postreplicative mismatch repair. *In vitro* studies with *Escherichia coli* DNA polymerase III show that G·A mismatches are less frequently inserted (1×10^{-5}) than G·T mismatches (8×10^{-5}) (31). Lower misinsertion frequencies might well arise if insertion of G·A base pairs depends upon adenine adopting the *syn* orientation. The effect would be most marked when the template residue is adenine, since the normal orientation of the bases in DNA is *anti* and rotation about the glycosyl bond is likely to be hindered by base stacking interactions. This restriction is less relevant when the template residue is guanine since it is the incoming ATP that needs to be in the *syn* orientation. These considerations do not apply to the formation of G·T and A·C pairs since in these mismatches both bases are *anti* isomers.

NMR experiments on synthetic deoxyoligonucleotides indicate that, when incorporated in a B-DNA helix, G·A and G·T mismatches have similar stabilities (6). Their behavior toward DNA polymerase during proofreading, however, is strikingly different.

Experiments with DNA polymerase III show that about 10% of all G·A mismatches avoid detection by proofreading, whereas only 0.5% of G·T mismatches escape (31). It seems that discrimination does not depend entirely on energetics; enzymic recognition of structural features is likely to be important in maintaining fidelity.

DNA polymerase is thought to have one active site for synthesis. It does not distinguish between the four combinations of Watson–Crick base pairs—A·T, T·A, G·C, and C·G—but does recognize the overall shape of a normal base pair. When a mispaired base is inserted, it is normally removed from the primer terminus by the 3′–5′ exonuclease (proofreading) activity of the enzyme. What, then, is the basis of the discrimination? Watson–Crick base pairs have a common structural feature. They are pseudosymmetric about the glycosyl bond so that the angle between these bonds and the C-1′···C-1′ vector falls within a narrow range of 52° to 62° (26, 27). The G·A(*syn*) base pair found in the present study is less symmetric, with glycosyl bond angles of 40° for adenine and 58° for guanine. The G·T wobble base pair (Fig. 1*g*) studied in several crystal structures (2–5) is, however, much more asymmetric with values of 40° and 70° for the two bases guanine and thymine, respectively. The C-1′···C-1′ separation, on the other hand, for both the G·A(*syn*) and the G·T wobble base pairs falls in the range of 10.2 to 10.8 Å as for Watson–Crick base pairs. Thus it seems that, at least for the G·A(*syn*) and G·T wobble mispairs, the degree of asymmetry of the glycosyl bonds is an important structural feature, which follows the same trend as the experimentally determined DNA polymerase proofreading efficiencies.

Enzymic recognition of noncomplementary bases may involve steric interactions in the major or minor grooves. It is thus of interest to compare the disposition of mismatched base pairs with Watson–Crick pairs. Fig. 4*a* shows the G·A(*syn*) base pair as determined in the present study, and Fig. 4*b* the same base pair superimposed on the G·C base pair at the same position in the native dodecamer. In Fig. 4*c* the G·T wobble pair from d(C-G-C-G-A-A-T-T-T-G-C-G) (ref. 3 and unpublished data) is superimposed on the same G·C base pair. Both figures were constructed by calculating the best molecular fit between the guanines so as to emphasize the displacement of the mispaired bases.

For the G·T wobble base pair there are substantial differences both in the major and in the minor grooves, whereas for the G·A(*syn*) base pair these differences are mostly in the major groove. The precise disposition of the mismatched base pairs in different DNA sequences (and different conformations) will, however, depend on the local environment.

Mismatches that escape proofreading may still be corrected by postreplicative mismatch repair enzymes. Some of these enzymes are apparently unable to recognize G·A mismatches, whereas G·T mismatches are repaired with

Table 2. Average torsion angles of the sugar-phosphate backbone

Model	Average torsion angles, °						
	α	β	γ	δ	ε	ξ	X
Mismatch dodecamer	−62	154	54*	112	159	−94	−107†
B-DNA fiber model‡	−41	136	38	139	133	−157	−102
Native dodecamer§	−63	171	54	123	169	−108	−117

Main chain torsion angles are defined by:
P–$\underline{\alpha}$–O-5′–$\underline{\beta}$–C-5′–$\underline{\gamma}$–C-4′–$\underline{\delta}$–C-3′–$\underline{\varepsilon}$–O-3′–$\underline{\xi}$–P.
The glycosyl torsion angle (X) is defined by:
O-4′—C-1′—N-1—C-2 for pyrimidines and O-4′—C-1′—N-9—C-4 for purines.
*Excludes C^1.
†Excludes A^9 and A^{21} that are in the *syn* orientation with X values of 67° and 56°, respectively.
‡Ref. 19.
§Ref. 28.

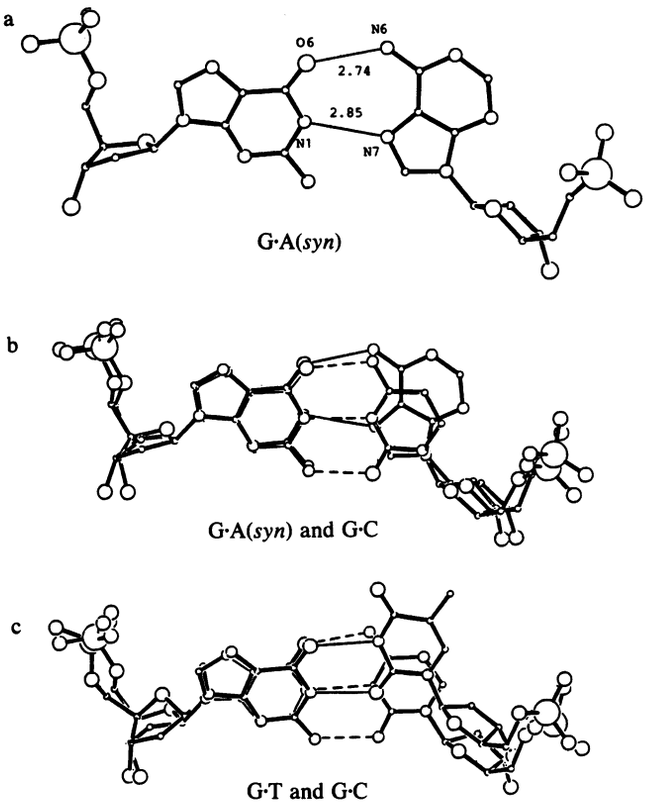

FIG. 4. Comparison of the G·A(*syn*) base pair and the G·T wobble base pair with a Watson–Crick base pair. (*a*) The G·A(*syn*) base pair as found in the crystal structure of the mismatch dodecamer at position $G^4 \cdot A^{21}$. Distances marked are in Å units. (*b*) Superposition of the G·A(*syn*) base pair on the G·C base pair at the same position in the native dodecamer. (*c*) Superposition of the G·T wobble base pair obtained from a mismatch dodecamer and the G·C pair in the same position in the native dodecamer. In preparing *b* and *c* the superpositions were obtained by calculating the best fit between the guanine bases, to emphasize the change in orientation of the glycosyl bond direction. The coordinates used for the native dodecamer were obtained from the Protein Databank and for the G·T mismatch dodecamer (2, 3) from unpublished data.

reasonably high efficiency (32–35). Mismatch repair systems, however, are complex and as yet not fully understood.

The G·A(*syn*) base pair as characterized in the present study has important implications for the tautomer hypothesis (13), which postulates that substitution mutations arise principally from mispairs in which one base is present in a rare tautomeric form and where the orientation of the glycosyl bond angles and the glycosyl bond separation are close to those in Watson–Crick pairs. It is implicitly assumed in this model that such base pairs are not structurally distinguished from Watson–Crick pairs.

In support of this theory it has been suggested that it is the G(*enol, imino*)·A(*syn*) (Fig. 1*c*) and the G(*syn*)·A(*imino*) base pairs (Fig. 1*d*) that give rise to adenine to guanine and guanine to adenine transversions. The G·A(*syn*) base pair, directly observed for the first time in the present study, is essentially isostructural with both tautomer pairs. Additionally, the hydrogen bonding between the bases involves the same heteroatoms as those in the G(*enol, imino*)·A(*syn*) pair. It is thus not necessary to evoke a minor tautomer to postulate a G·A base pair that mimics a Watson–Crick base pair to account for the frequency of transversion mutations involving such G·A base pairs.

We thank Dr. P. R. Evans for discussions and help in using the graphics facility at the Medical Research Council Laboratory for Molecular Biology, Prof. D. Rabinovich for the peak search program, and Drs. E. Westhof and T. Prange for the unpublished coordinates of the native dodecamer. We are grateful to Prof. R. A. Raphael for his help and advice. Financial support for this work was provided by the Medical Research Council under a Program Grant (to O.K.).

1. Saenger, W. (1984) *Principles of Nucleic Acid Structure* (Springer, New York).
2. Brown, T., Kennard, O., Kneale, G. & Rabinovich, D. (1985) *Nature (London)* **315**, 604–606.
3. Kennard, O. (1985) *J. Biomol. Struct. Dyn.* **3**, 205–225.
4. Brown, T., Kneale, G., Hunter, W. N. & Kennard, O. (1986) *Nucleic Acids Res.*, in press.
5. Kneale, G., Brown, T., Kennard, O. & Rabinovich, D. (1985) *J. Mol. Biol.* **186**, 805–814.
6. Patel, D. J., Kozlowski, S. A., Ikuta, S. & Itakura, K. (1984) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **43**, 2663–2670.
7. Kan, L. S., Chandrasegaran, S., Pulford, S. M. & Miller, P. S. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 4263–4265.
8. Watson, J. D. & Crick, F. H. C. (1953) *Nature (London)* **171**, 964–967.
9. Radding, C. M. (1978) *Annu. Rev. Biochem.* **47**, 847–880.
10. Crick, F. H. C. (1976) *J. Mol. Biol.* **19**, 548–555.
11. Jack, A., Ladner, J. E. & Klug, A. (1976) *J. Mol. Biol.* **108**, 619–649.
12. Rich, A. (1977) *Acc. Chem. Res.* **10**, 338–402.
13. Topal, M. D. & Fresco, J. R. (1976) *Nature (London)* **263**, 285–293.
14. Traub, W. & Sussman, J. L. (1982) *Nucleic Acids Res.* **19**, 2701.
15. Churprina, V. P. & Poltev, V. I. (1983) *Nucleic Acids Res.* **11**, 5205–5222.
16. Rein, R., Shibata, M., Gardino-Juarez, R. & Keiber-Emmons, T. (1983) in *Structure and Dynamics of Nucleic Acids and Proteins*, eds. Clementi, E. & Sarma, R. (Adenine Press, Guilderland, NY), pp. 269–288.
17. Gait, M. J., Matthes, H. W. D., Singh, M., Sproat, B. S. & Titmus, R. C. (1982) *Nucleic Acids Res.* **10**, 6243–6248.
18. Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K. & Dickerson, R. E. (1980) *Nature (London)* **287**, 755–758.
19. Arnott, S. & Hukins, D. W. I. (1972) *Biochem. Biophys. Res. Commun.* **47**, 1504–1509.
20. Sheldrick, G. M. (1976) *SHELX 76 System of Computing Programs* (Univ. of Cambridge, Cambridge, England).
21. Sussman, J. L., Holbrook, S. R., Church, G. M. & Kim, S.-H. (1977) *Acta Crystallogr. Sect. A* **33**, 800–804.
22. Westhof, E., Dumas, P. & Moras, D. (1985) *J. Mol. Biol.* **184**, 119–145.
23. Hendrickson, W. A. & Konnert, J. H. (1981) in *Biomolecular Structure, Conformation, Function and Evolution*, ed. Srinivasan, R. (Pergamon, Oxford), Vol. 1, pp. 43–57.
24. Wang, A. H. J., Quigley, G. J., Kolpack, F. J., Crawford, L., van Boom, J. H., van der Marel, G. & Rich, A. (1979) *Nature (London)* **282**, 680–686.
25. Wang, A. H. J., Ughetto, G., Quigley, G. J., Hakoshima, T., van der Marel, G. A., van Boom, J. M. & Rich, A. (1984) *Science* **225**, 115–121.
26. Seeman, N. C., Rosenberg, J. M., Suddath, F. L., Kim, J. J. P. & Rich, A. (1976) *J. Mol. Biol.* **104**, 109–144.
27. Rosenberg, J. M., Seeman, N. C., Day, R. O. & Rich, A. (1976) *J. Mol. Biol.* **104**, 145–167.
28. Dickerson, R. E. & Drew, H. R. (1981) *J. Mol. Biol.* **149**, 761–786.
29. Shakked, Z. & Kennard, O. (1985) in *Biological Macromolecules and Assemblies*, eds. McPherson, A. & Jurnak, F. (Wiley, New York), pp. 1–36.
30. Ladner, J. E., Jack, A., Robertus, J. D., Brown, R. S., Rhodes, D., Clark, B. F. C. & Klug, A. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 4414–4418.
31. Fersht, A. R., Knill-Jones, J. W. & Tsui, W. C. (1982) *J. Mol. Biol.* **156**, 37–51.
32. Kramer, B., Kramer, W. & Fritz, H. J. (1984) *Cell* **38**, 878–887.
33. Lu, A. L., Welsh, K., Clark, S., Su, S. S. & Moldrich, P. (1984) *Cold Spring Harbor Symp. Quant. Biol.* **49**, 589–596.
34. Claverys, J. P., Mejean, V., Gasc, A. M. & Sicard, A. M. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 5956–5960.
35. Dohet, C., Wagner, R. & Radman, M. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 503–505.