Nucleic Acids Research, 1998, Vol. 26, No. 21 **4925–4934**

NMR solution structure of a DNA dodecamer containing single G•T mismatches

Hatim T. Allawi[+] and John SantaLucia, Jr*

Department of Chemistry, Wayne State University, Detroit, MI 48202, USA

Received July 6, 1998; Revised and Accepted September 3, 1998

PDB accession no. 1bjd

ABSTRACT

The three-dimensional solution structure of the self-complementary DNA dodecamer $\underset{GCA\underline{T}TGCA\underline{G}TGC}{CGT\underline{G}ACGT\underline{T}ACG}$ which contains the thermodynamically destabilizing $\underset{A\underline{T}T}{T\underline{G}A}$ motif was determined using two-dimensional NMR spectroscopy and simulated annealing protocols. Relaxation matrix analysis methods were used to yield accurate NOE derived distance restraints. Scalar coupling constants for the sugar protons were determined by quantitative simulations of DQF-COSY cross-peaks and used to determine sugar pucker populations. Twenty refined structures starting from random geometries converged to an average pairwise root mean square deviation of 0.49 Å. Back calculated NOEs give R^c and R^x factors of 0.38 and 0.088, respectively. The final structure shows that each of the single G·T mismatches form a wobble pair with two hydrogen bonds where the guanine projects into the minor groove and the thymine projects into the major groove. The incorporation of the destabilizing $\underset{A\underline{T}T}{T\underline{G}A}$ motif has little effect on the backbone torsion angles and helical parameters compared to standard B-form duplexes, which may explain why G·T mismatches are among the most commonly observed in DNA. The structure shows that perturbations caused by a G·T mismatch extend only to its neighboring Watson–Crick base pair, thus providing a structural basis for the applicability of the nearest-neighbor model to the thermodynamics of internal G·T mismatches.

INTRODUCTION

The processes of cytosine methylation and spontaneous deamination continually create G·T mismatches in genomic DNA (1,2). In addition, G·T mismatches commonly occur during DNA replication but are efficiently repaired by DNA polymerase proofreading or by post-replication mismatch repair enzymes (3,4). To further aid in our understanding of the biological function of G·T mismatches, we are systematically studying their thermodynamic–structure relationships. We recently showed that internal G·T, G·G and G·A mismatches are the most stable mismatches found in DNA (5–8; N.Peyret, P.A.Seneviratne, H.T.Allawi and J.SantaLucia, unpublished results). Further, we showed that the nearest-neighbor model can be extended in its application to accurately predict the thermodynamics of internal G·T mismatches (6).

There have been a number of NMR studies performed on G·T mismatches in DNA, but none has resulted in a high resolution three-dimensional structure (9–12). To the best of our knowledge, there is only one X-ray crystallographic structure of a B-form DNA duplex containing single G·T mismatches obtained by Hunter and co-workers for the dodecamer d(CGC$\underline{G}$AATT$\underline{T}$GCG)$_2$ (13). Hunter and co-workers concluded that the G·T pair adopts a 'wobble' configuration with little and highly localized perturbations with respect to the overall double helix.

In our previous thermodynamic study on G·T mismatches (6), we showed that stability of internal G·T mismatches are strongly sequence dependent. The most stable trimer observed was $\underset{G\underline{T}G}{C\underline{G}C}$ (ΔG°_{37} = –1.05 kcal/mol of trimer) and the least stable trimers were $\underset{T\underline{T}T}{A\underline{G}A}$ and $\underset{A\underline{T}T}{T\underline{G}A}$ (ΔG°_{37} = 1.05 kcal/mol of trimer and 0.80 kcal/mol of trimer). As a first step towards understanding the relationship between thermodynamics and structure of mismatches in DNA, we decided to obtain the solution structure of one of the most unstable G·T mismatch containing trimer contexts (the trimer $\underset{A\underline{T}T}{T\underline{G}A}$) by solving its three-dimensional solution structure in the duplex d(CGT$\underline{G}$ACGT$\underline{T}$ACG)$_2$ using high-resolution NMR spectroscopy. This structure is used to provide a structural basis for the applicability of the nearest-neighbor model to the prediction of G·T mismatch thermodynamics (6). The biological implications of the G·T mismatch structure are also discussed.

MATERIALS AND METHODS

Sample synthesis and purification

dCGTGACGTTACG was synthesized on solid support with a Cruachem PS250 DNA/RNA synthesizer using standard phosphoramidite chemistry (14). Upon completion of synthesis, the DNA was deblocked, purified by thin layer chromatography, and desalted using a Sep-pak C-18 cartridge (Waters) as described (6). To further desalt the sample and prepare it for NMR studies, the oligonucleotide was dialyzed (Gibco BRL) twice against 1 l of double-distilled deionized water for 48 h. Residual divalent metals were removed by the addition and removal of several grains of chelex-100 resin. The oligonucleotide was then

*To whom correspondence should be addressed. Tel: +1 313 577 0101; Fax: +1 313 577 8822; Email: jsl@chem.wayne.edu

[+]Present address: Third Wave Technologies, Madison, WI 53719, USA

evaporated to dryness and dissolved in 0.33 ml of NMR buffer which consisted of 0.3 M NaCl, 10 mM sodium phosphate, 0.5 mM Na_2EDTA and 0.5 mM 3-(trimethylsilyl)propionate (TSP), pH 7.0. The sample was lyophilized twice from 99.96% D_2O and once from 99.996% D_2O. The sample was finally dissolved in 0.33 ml of 99.996% D_2O (or 90% H_2O and 10% D_2O for exchangeable proton NMR experiments) and placed in a microvolume NMR tube (Shigemi). The total strand concentration was 3 mM.

NMR experiments

All NMR spectra were recorded on a Varian Unity 500 MHz NMR spectrometer and analyzed on a Silicon Graphics Indigo^{2}Extreme workstation using VNMR-SGI (Varian) and FELIX95 (Biosym/MSI) software. All 2D NMR spectra were recorded in the phase-sensitive mode utilizing the States-TPPI method (15).

Exchangeable proton NMR spectra of the sample dissolved in 90% H_2O and 10% D_2O were recorded at 10°C using WATERGATE pulse sequence with a 'flip-back' pulse to suppress the water peak (16,17). Spectra were recorded with the transmitter frequency placed at the solvent frequency, a sweep width of 12 KHz, and gradient field strength of 10.0 G/cm and duration of 1 ms. 1D-NOE difference spectra were acquired with selective decoupling of individual resonances during the 1 s recycle delay as described (6). NOESY spectra in H_2O were recorded with mixing times of 100, 200 and 300 ms with a relaxation delay between scans of 5 s. A total of 600 complex FIDs were collected with 32 transients, 4096 complex points and a spectral width of 10 KHz in both dimensions. Natural abundance ^1H-^{15}N HMQC experiments were acquired in H_2O using jump and return proton pulses and *z*-axis gradient pulses to suppress the H_2O resonance (18). The spectral widths were 11 KHz in the proton dimension and 1500 Hz in the nitrogen dimension. A total of 80 FIDs with 1056 transients each and 4096 complex points were collected.

NOESY experiments in D_2O were recorded at 25°C with mixing times of 60, 100, 150, 300 and 500 ms using a relaxation delay of 7 s. The transmitter frequency was set at the residual HDO resonance, which was presaturated using low-power RF for 2 s during the recycle delay. For each FID, 32 transients were collected with 4096 complex points and a spectral width of 5000 Hz in both dimensions. A total of 450 complex FIDs were collected for each NOESY spectrum.

DQF-COSY spectra were recorded with 512 complex FIDs and spectral widths of 4000 Hz in both dimensions. A total of 64 transients with 2048 complex points were collected for each FID. A ^1H–^{31}P HETCOR spectrum (19) was acquired with spectral widths of 1650 Hz in the proton dimension and 1000 Hz in the phosphorus dimension. 256 complex FIDs were collected with 128 transients and 1024 complex points. TOCSY experiments were collected with a mixing time of 70 ms and spectral widths of 4000 Hz in both dimensions. A total of 850 FIDs were collected with 32 transients and 4096 complex points. Measurements of the spin-lattice relaxation times, T_1, were determined by the inversion–recovery method.

2D-NOE intensity analysis and distance restraints

Non-exchangeable interproton distances were derived using the program MARDIGRAS (20) from NOESY spectra with mixing times of 60, 100 and 150 ms. The starting model for the MARDIGRAS calculations was a standard B-form DNA duplex constructed using InsightII (Biosym/MSI). Prior to MARDIGRAS calculations, the B-form starting duplex was submitted to a 1000 steps of energy minimization using XPLOR (21). To account for peak integration errors, each of the three intensity sets was assigned an error of 25% of the smallest intensity value in that set. A three-site jump model was used to treat the thymine methyl protons (22). All MARDIGRAS calculations were carried out assuming a single isotropic correlation time (τ_c) of 4.5 ± 0.5 ns. The number of experimental cross-peak intensities were 119, 150 and 186 for the 60, 100 and 150 ms spectra, respectively. Calculations were carried out using the 'RANDMARDI' option of MARDIGRAS to provide error bounds (23). A total of 30 RANDMARDI runs were performed on the three intensity sets. The resultant distances were averaged and their standard deviations computed. 0.20 Å was added to the standard deviations and used as a conservative estimate of the lower and upper bounds of the distance restraints input into restrained molecular dynamics (rMD) calculations. Eleven restraints were not included in the rMD calculations since the peak intensities were inaccurate due to either spectral overlap or low signal-to-noise ratios. A total of 175 non-exchangeable interproton distance restraints per strand were obtained.

Exchangeable interproton distance restraints were computed from the 100 and 200 ms NOESY in H_2O spectra using the isolated spin pair approximation and scaling the observed intensities by the H5–H6 of cytosine as a fixed distance of 2.46 Å as described (24). MARDIGRAS calculations were not performed on exchangeable protons since the exchange rate with solvent is unknown (25). Exchangeable distance restraints were classified as strong (1.5–3.0 Å), medium (3.0–4.5 Å) and weak (4.5–6.0 Å). A total of 32 exchangeable interproton distance restraints per strand were obtained. A total of 26 hydrogen bonding distance restraints between G · C and A · T pairs of 1.90 ± 0.2 Å were also included in the rMD calculations. Two imino–imino distance restraints (1.5–3.0 Å) for the H1 guanine to H3 thymine of the G · T pairs were also included.

Structural modeling

All molecular modeling experiments were carried out on a Silicon Graphics Indigo^{2}Extreme computer running InsightII (Biosym/MSI) and XPLOR (21). rMD and restrained energy minimizations (rEM) were performed using the all-atom CHARMM force field and charges with modified bond lengths and angles appropriate for DNA (26). NOE and dihedral angle restraints were included in the force field as a quadratic pseudopotential with a flat well within the upper and lower bounds of the restraints. Force constants used for the NOE and dihedral angle potentials were set to 50 kcal/(mol·Å^2) and 50 kcal/(mol·rad^2), respectively. The DNA in this study forms a self-complementary duplex, and therefore, the same NOE and dihedral angle restraints were applied to both strands. In addition, 2-fold symmetry was imposed (consistent with the observation of one resonance from symmetry related protons in both strands of the duplex), using non-crystallographic symmetry (NCS) as a term in the XPLOR force field, and was included in the final stages of the refinement (24).

A total of 32 structures with random backbone dihedrals were generated using XPLOR and used as starting coordinates for the rMD and rEM protocols. These protocols were performed *in vacuo* in a simulated annealing stage and a refinement stage as

Table 1. Proton and phosphorous chemical shifts assignments (p.p.m.) for d(CGTGACGTTACG)$_2$ at 25°C[a]

	Proton Assignments										
Base	H8/H6	H5/H2	Methyl	H1'	H2'	H2"	H3'	H4'	H5'/H5"[b]	P	imino/amino[c]
C1	7.61	5.81	na[d]	5.71	2.00	2.40	4.69	4.07	3.71	na	6.99
G2	7.99	na	na	6.00	2.70	2.82	4.98	4.38	4.10/4.04	-4.59	12.83
T3	7.29	na	1.55	5.84	2.22	2.57	4.89	4.21	4.29/4.18	-4.70	13.51
G4	7.81	na	na	5.75	2.55	2.80	5.00	4.37	4.24/4.20	-4.63	10.17
A5	8.12	7.78	na	6.12	2.58	2.85	5.01	4.44	4.25/4.13	-4.83	
C6	7.19	5.18	na	5.53	1.99	2.32	4.82	4.14	4.19	-4.71	8.07/6.49
G7	7.82	na	na	5.95	2.59	2.76	4.94	4.36	4.23/4.13	-4.58	12.76
T8	7.20	na	1.37	6.02	2.03	2.56	4.84	4.25	4.16	-4.74	13.88
T9	7.47	na	1.78	5.55	2.17	2.32	4.87	4.06	4.13	-4.77	12.01
A10	8.34	7.52	na	6.19	2.74	2.84	5.04	4.44	4.02/4.15	-4.50	
C11	7.27	5.35	na	5.62	1.86	2.25	4.78	4.14	4.22/4.13	-4.70	8.26/6.74
G12	7.87	na	na	6.09	2.59	2.38	4.66	4.18	4.15/4.06	-4.57	13.00

[a]Chemical shift are relative to TSP for proton and to external TMP for phosphorous resonances.
[b]The H5′ and H5″ assignments are not stereospecific.
[c]Exchangeable proton chemical shifts are at 10°C.
[d]Not applicable.

described (24,27). During the simulated annealing calculations, dihedral restraints and electrostatic terms were turned off and van der Waals repulsive interactions were turned on. Upon completion of the simulated annealing stage, the resulting structures were examined to test for convergence. The criterion for convergence was the proper formation of Watson–Crick base pairs. The converged structures were used in the refinement protocol as described (24).

RESULTS

Residues of the dodecamer duplex in this study were labeled and numbered according to:

1	2	3	4	5	6	7	8	9	10	11	12
C	G	T	G	A	C	G	T	T	A	C	G
G	C	A	T	T	G	C	A	G	T	G	C
12	11	10	9	8	7	6	5	4	3	2	1

Note that G·T mismatches are underlined. This residue numbering format will be referred to through out the text.

Non-exchangeable proton and phosphorus assignments

NMR resonances of d(CGTGACGTTACG)$_2$ were assigned in a sequential manner using well described procedures for DNA duplexes (28). Table 1 summarizes all proton and phosphorus chemical shift assignments at 25°C. The base (H8/H6/H2) to the H1′/H5 region of the 300 ms NOESY (Fig. 1) was used to assign base (H8/H6/H2) and H1′ protons in a sequential manner. As expected, A5 and A10 H2 protons exhibited the longest T_1 times from inversion-recovery experiments (3.6 and 3.2 s, respectively) while the rest of the base protons had shorter T_1 times (1.7–2.2 s).

Cross-peaks in the thymine H6 to methyl region of the 60 ms NOESY spectrum were used to unambiguously assign T3, T8 and T9 methyl protons. The 60 ms NOESY spectrum was also used to stereospecifically assign all H2′ and H2″ resonances since the distance H1′-H2″ is smaller than H1′–H2′ for all sugar conformations (28). H2′ and H2″ assignments were further confirmed from analysis of the H1′–H2′ and H1′–H2″ DQF-COSY cross-peak patterns (Fig. 2) (28).

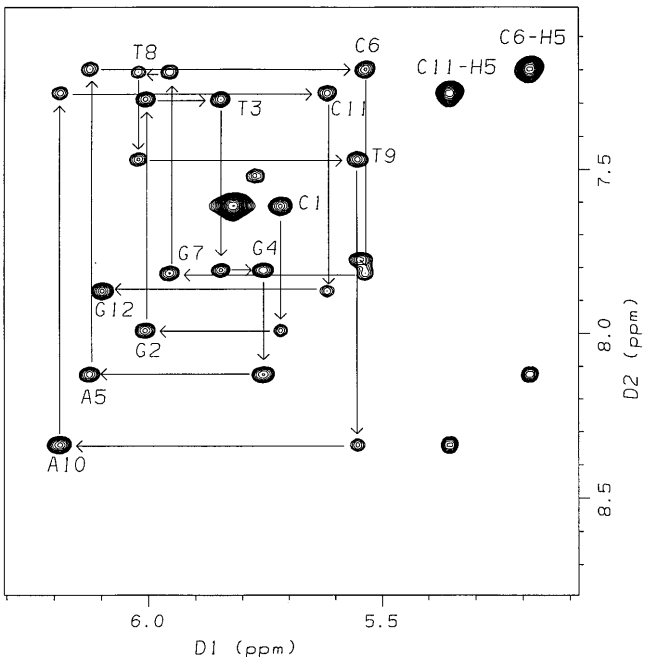

Figure 1. H8/H6/H2–H1′/H5 region of the 300 ms NOESY of d(CGTGACGTTACG)$_2$ at 25°C in 100% D_2O. Sequential H8/H6(n)–H1′(n)–H8/H6(n + 1) connectivities are shown in solid lines. Resonances at 7.78 and 7.52 are from A5H2 and A10H2.

Assignments of H3′ and H4′ were obtained from analysis of TOCSY spectra (Supplementary Material). Non-stereospecific assignments of H5′/H5″ resonances were made from analysis of DQF-COSY, TOCSY, ^{1}H-^{31}P HETCOR and NOESY experiments. Phosphorus assignments were obtained by analysis of the ^{1}H–^{31}P HETCOR spectrum (Supplementary Material). In the ^{1}H–^{31}P

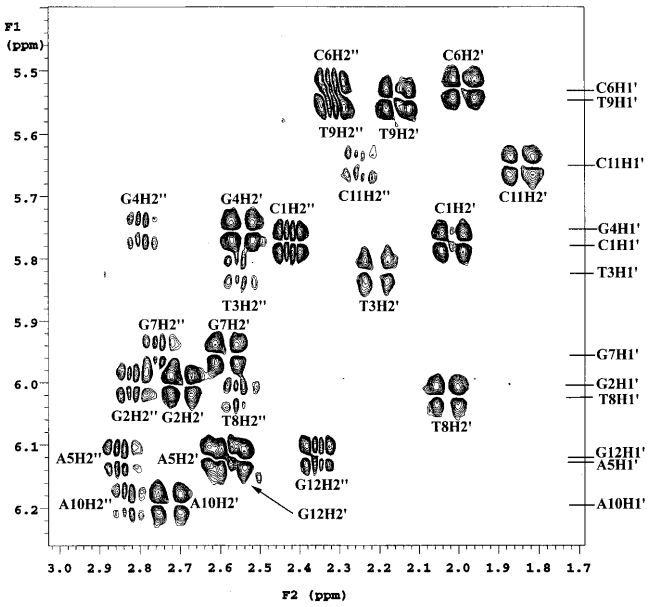

Figure 2. H1′–H2′/H2″ region of the DQF-COSY of d(CGTGACGTTACG)$_2$ at 25°C in 100% D_2O. Assignments for the H1′ protons are on the right axis. Assignments for H2′ and H2″ are adjacent to their cross peaks.

HETCOR experiment, each phosphorus resonance of residue (*n*) exhibits cross peaks between H3′ of residue (*n* – 1) and H4′, H5′/H5″ of residue (*n*).

Exchangeable proton assignments

The imino protons of d(CGTGACGTTACG)$_2$ were assigned using 1D-NOE difference spectra and 300 ms 2D-NOESY spectrum recorded in H_2O (Supplementary Material). The resonances at 12.0 and 10.2 p.p.m. were assigned to the guanine (G4) and thymine (T9) imino protons of the G·T mismatch. Since G4 and T9 imino resonances exhibited the same NOEs, we used natural abundance ^{1}H-^{15}N HMQC spectroscopy (18) to obtain specific assignments (Supplementary Material). Exchangeable amino protons were assigned from the 300 ms H_2O 2D-NOESY experiment (Supplementary Material). C6 hydrogen-bonded and non-hydrogen-bonded amino resonances were assigned from their NOE cross peaks with the imino protons of G7 and T8. Similarly, C11 amino resonances where assigned from their cross peaks with the imino resonances of G2 and T3. Imino to amino NOEs for the C1 amino resonances were not observed due to the rapid exchange of the G12 imino proton with solvent.

Coupling constants and sugar pucker estimates

J-coupling constants were obtained from simulations of DQF-COSY cross-peaks using SPHINX and LISNHA (29). Digital resolution, apodization functions and acquisition times were the same as in the experimental spectra. Values of –14.0 ± 0.5 Hz were used for all $J_{H2'–H2''}$. All other coupling constants were determined in an iterative fashion as described by Schmitz and James (30) by changing their values by ± 0.2 Hz and comparing the resultant cross-peak patterns with their experimental counterpart. Figure 3 shows the experimental versus simulated H1′–H2′ and H1′–H2″ cross-peak patterns obtained for G4 and A10. A table of *J*-coupling constants obtained for all residues is provided in Supplementary Material. All $J_{H1'–H2'}$ values ranged from a minimum of 8.0 Hz to maximum of 9.8 Hz. This range is what is expected for a C2′-*endo* sugar geometry ($J_{H1'–H2'}$ ~ 10.0 Hz for pure C2′-*endo* conformer and ~ 1.0 Hz for pure C3′-*endo* conformer; 31). No obvious changes in the *J*- coupling constants for the G·T mismatches and adjacent residues are observed.

The sums of the H1′ *J*-coupling constants, Σ1′, were used to calculate fractions of south (C2′-*endo*), f_S, conformers in each sugar residue by assuming a two-state dynamic equilibrium and using the equation (28):

$$f_S = (\Sigma 1' - 9.8)/5.9 \quad \mathbf{1}$$

where $\Sigma 1' = J_{H1'–H2'} + J_{H1'–H2''}$. With the exception of the terminal C1 and G12 residues, f_S values indicate that the majority of sugar populations (>88%) are in the S-type conformer (Supplementary Material).

Backbone torsion angles

Backbone dihedral angles for ε were estimated based on measurements of *J*-coupling constants of H3′–P ($J_{H3'–P}$) obtained from the ^{1}H–^{31}P HETCOR as described (32). Values of $J_{H3'–P}$ ranged from 2.5 to 4 Hz, consistent with previous observations for B-form DNA (9), and are given in the Supplementary Material. These coupling constants were converted to ε angles using the relationship (9):

$$J_{H3'–P} = 15.3 \cos^2(\theta) – 6.1 \cos(\theta) + 1.6 \quad \mathbf{2}$$

where ε = –θ – 120. The two solutions obtained using equation **2** (using the observed *J*-coupling with the upper and lower error estimates) were used as lower and upper bounds for ε angles (Supplementary Material).

In the ^{1}H–^{31}P HETCOR spectrum, weak H5′–P and H5″–P cross-peaks were observed. However, strong four bond coupling between H4′ and phosphates were detected for all residues indicating that the atoms P–O5′–C5′–C4′–H4′ lie in the same plane forming a 'W-shaped' conformation (32). This situation occurs if β and γ are in the *trans* and the g^+ conformations, respectively (28). Therefore, torsion angle values for β and γ were restricted to be in ranges of 180 ± 30° and 60 ± 30°, respectively.

Sugar pucker estimates (see above) indicate that the majority of all residues are in the S-type conformation and, therefore, values for δ were restricted to represent S-type sugars with values ranging from 137 to 172° (28). In this study, a small range of ^{31}P chemical shifts was observed (^{31}P chemical shifts varied over a range of 0.33 p.p.m.) indicating that there are no major distortions resulting in abnormal α and ζ present in the studied duplex (33). Nonetheless, since there were no direct measurements of α and ζ torsion angles in this study, no restraints were imposed on their values. A total of 46 torsion angle restraints per strand were obtained.

Structural modeling

To help in the convergence process, five additional distance restraints specified as >6 Å between protons from non-consecutive residues were added to the restraint data set. There were no observed NOEs for these additional distances, even in the 500 ms mixing time NOESY experiment, thus justifying their addition. After the simulated annealing stage of the rMD protocol, 20 out of the 32 random starting structures formed proper Watson–Crick base

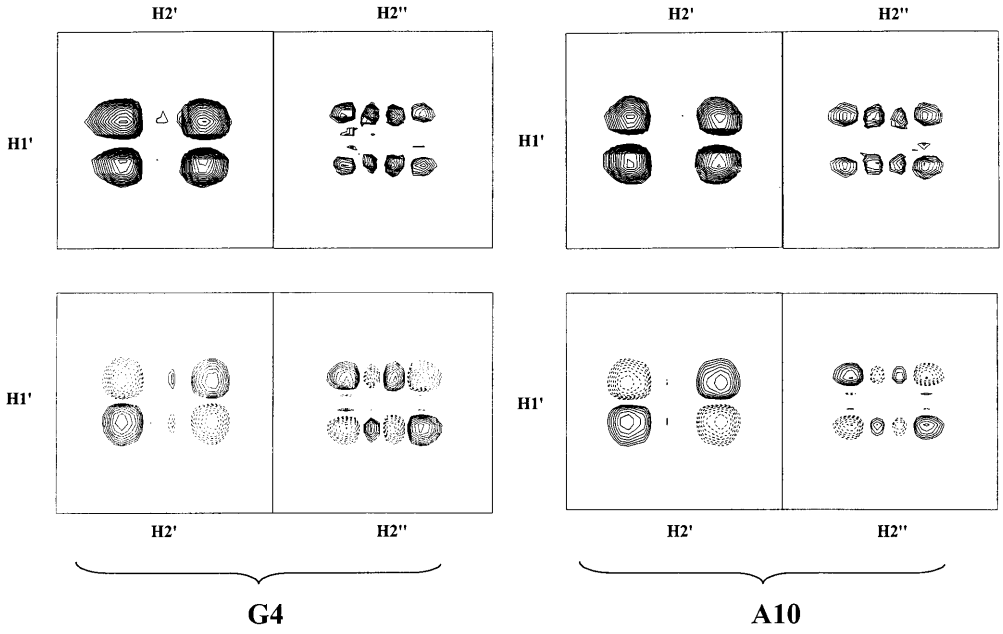

Figure 3. Expansion (70 × 70 Hz) of the DQF-COSY cross-peaks (top) of G4 (right) and A10 (left) H1′(ω1)–H2′(ω2) and H1′(ω2)–H2″(ω1) cross peaks at 25°C and their corresponding simulations (bottom) obtained from SPHINX and LINSHA (29). Negative contours are shown in broken lines in the simulated cross-peaks. The following coupling constants (Hz) produced the best fits for G4: $J_{H1'–H2'}$ = 9.2, $J_{H1'–H2''}$ = 6.2, $J_{H2'–H3'}$ = 6.2, $J_{H2''–H3'}$ = 2.5 and $J_{H2'–H2''}$ = –14.0 and for A10: $J_{H1'–H2'}$ = 9.8, $J_{H1'–H2''}$ = 5.6, $J_{H2'–H3'}$ = 6.2, $J_{H2''–H3'}$ = 2.3 and $J_{H2'–H2''}$ = –14.0.

pairs and thus were considered to converge. These 20 structures were used as input for the refinement stage in which dihedral angle and electrostatic terms were added to the force field as described previously (24). The resulting 20 'refined' structures converged to satisfy all distance and dihedral angles restraints within 0.1 Å and 1°, respectively. The total NOE and dihedral energy violations for all final structures were <20 and 30 kcal/mol, respectively.

Superposition of the resulting 20 structures (Fig. 4) reveals that the overall structure as well as local features such as base stacking and sugar conformations are well determined. Terminal residues did not superimpose as well as the internal residues since fewer restraints are imposed on them and because of accumulation of errors along the helix. The average root mean square deviation (RMSD) for the all-atom pairwise superposition of these 20 structures was 0.49 Å (Fig. 4A). An average RMSD of 0.25 Å was obtained for the superposition of the $\frac{TGA}{ATT}$ trimers of all 20 structures (Fig. 4B).

The accuracy of the final structures can be judged by comparing their calculated NOEs with the experimental 2D NOE data (34). We calculated theoretical NOEs for the final structures using the program CORMA (35). We used the *R* factors below to compare the theoretical NOE intensities from the final structures with their experimental counterpart (36):

$$R^c = \frac{\sum_i |a_o(i) - a_c(i)|}{\sum_i [a_o(i)]} \qquad 3$$

$$R^x = \frac{\sum_i \left|a_o(i)^{\frac{1}{6}} - a_c(i)^{\frac{1}{6}}\right|}{\sum_i \left[a_o(i)^{\frac{1}{6}}\right]} \qquad 4$$

where R^c is the crystallographic-equivalent *R*-factor. a_o and a_c are the observed and calculated NOE intensities for the *i*th cross-peak, and R^x is the sixth-root residual index. R^c and R^x values were obtained for the NOESY experiments (mixing times of 60, 100 and 150 ms) and averaged. The average R^c and R^x value of the 20 final structures were 0.38 and 0.088, respectively. These *R*-factor values indicate that the final structures are in good agreement with the measured NOEs (36). For comparison, the initial energy minimized B-DNA model used in the MARDIGRAS calculations fits the experimental NOEs with residual R^c and R^x factors of 1.10 and 0.160, respectively.

Dihedral angles analysis and helical parameters of final structures

Torsion angles and helical parameters for the final refined structure of d(CGTGACGTTACG)$_2$ obtained using the programs CURVES 5.3 (37) and NEWHEL93 (38,39) are listed in Tables 2 and 3. The dihedral angle results show that the duplex has an overall geometry of a B-form DNA with slight distortions near the G·T mismatches and terminal residues. Only minor distortions in the backbone geometry are observed as a result of the G·T mismatch and are as follows: the α torsion angles of T3, T8 and T9 (α values of –69.2, –66.8 and –69.3°, respectively), the ε torsion angles of T3 and T9 (ε values of –178.9 and 178.3°,

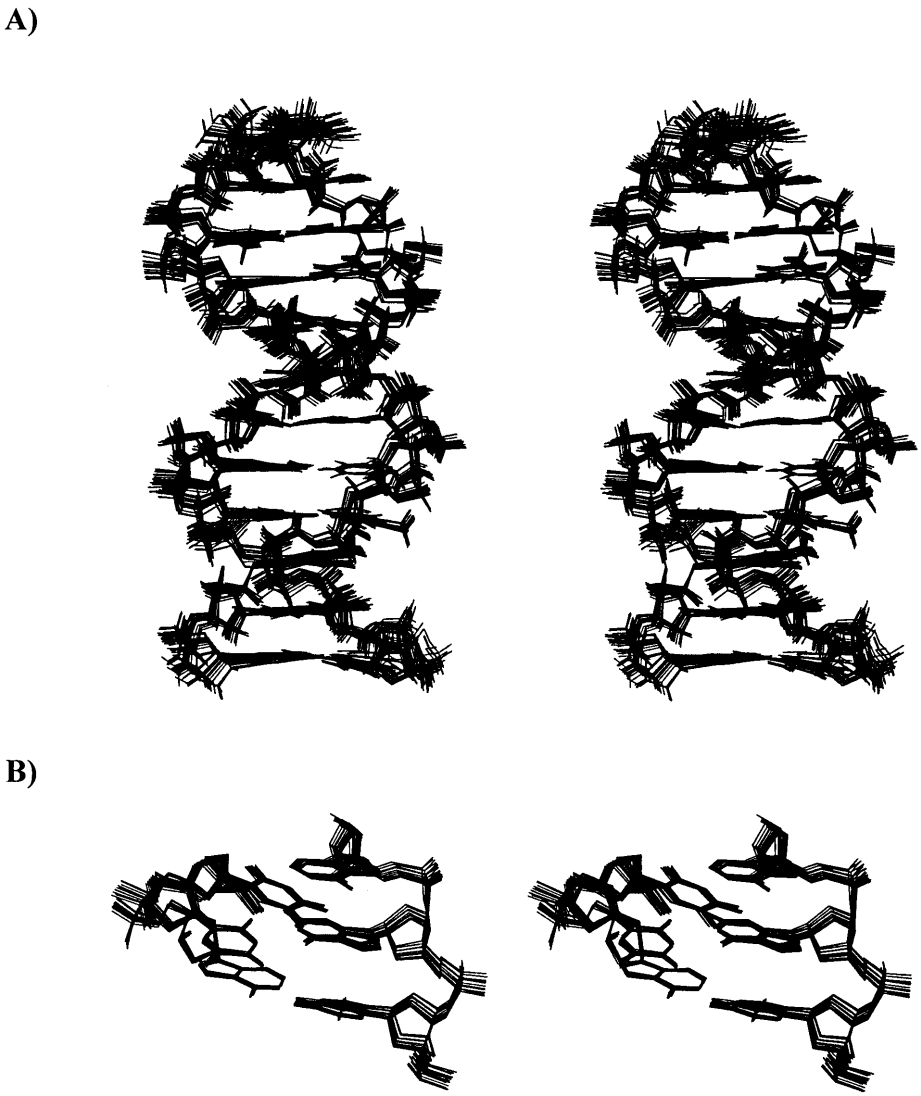

Figure 4. (**A**) Superposition (RMSD 0.49) of 20 overlapped final structures of d(CGTGACGTTACG)$_2$. (**B**) View of the minor groove of the superposition of the heavy atoms of the TGA/ATT trimer (RMSD 0.25) of the 20 final structures.

respectively) and the ζ torsion angles of T3 (ζ = –104.3°) and T9 (ζ = –107.1°). These torsion angle variations are within 7, 4 and 12° from their average value for α, ε and ζ, respectively (Table 2). These results indicate that the G·T mismatch is accommodated in the double helix with only minor changes to backbone torsion angles of the mismatched pair and their neighboring Watson–Crick residues.

Helical parameters for the final structure also indicate that no major perturbation on either the base steps or on the base pair parameters (Table 3 and Supplementary Material). The only major distortions are noted in λ_1 and λ_2 of the G·T mismatch. λ_1 and λ_2 are defined as the angles between the N9/N1–C1′ bond and the C1′–C1′ vector of each base pair. The subscripts 1 and 2 designate the right and left bases. For canonical Watson–Crick pairs in B-DNA, both λ_1 and λ_2 are ~56° (38). In this study, we find that for Watson–Crick base pairs values for λ_1 and λ_2 fall in that range. However, λ_1 and λ_2 of guanine and thymine of the G·T mismatch (G4 and T9) are 40.8 and 69.5°, respectively, indicating that mismatch is asymmetric and that the guanine is shifted towards the minor groove and the thymine towards the major groove as expected for a wobble pair.

DISCUSSION

Final structures

The precision of the final structures is evaluated from the convergence of all 20 random structures to a single well-determined final structure. The RMSD deviation of 0.49 Å for the superposition of all 20 final structures indicates good convergence (Fig. 4). A value of 0.49 Å is in the range that is expected considering thermal fluctuations and NOE and dihedral experimental error (40–44). The fact that there were no distance or dihedral angle restraint violations >0.1 Å and 1° indicates that the final structures satisfy the imposed experimental restraints. An RMSD value of 0.25 for the $\frac{TGA}{ATT}$ trimers for the superposition of all 20 structures in this

Table 2. Dihedral angles (°) for d(CGTGACGTTACG)$_2$ [a]

Residue	α	β	γ	δ	ε	ζ	χ	P[b]
C1	na	na	55.8 ± 0.3	137.8 ± 0.3	-171.6 ± 0.2	-84.6 ± 0.5	-149.8 ± 2.8	197.4 ± 2.1
G2	-81.3 ± 1.4	-161.2 ± 1.0	52.6 ± 0.3	138.8 ± 0.2	172.2 ± 0.4	-88.1 ± 0.5	-105.7 ± 0.6	179.3 ± 0.9
T3	-69.2 ± 0.4	-177.5 ± 0.3	54.3 ± 0.2	135.2 ± 0.2	-178.9 ± 0.2	-104.3 ± 1.6	-108.5 ± 0.5	151.3 ± 0.4
G4	-78.3 ± 1.3	-171.9 ± 1.0	53.5 ± 0.4	138.3 ± 0.3	170.6 ± 0.3	-96.2 ± 0.6	-114.2 ± 0.8	158.8 ± 0.8
A5	-72.7 ± 0.5	-167.6 ± 0.9	55.6 ± 0.2	138.6 ± 0.2	174.1 ± 0.2	-89.0 ± 1.5	-112.4 ± 0.2	179.2 ± 1.4
C6	-73.2 ± 1.6	-174.0 ± 0.9	55.6 ± 0.3	135.9 ± 0.2	177.1 ± 0.2	-93.0 ± 1.7	-108.0 ± 0.9	161.6 ± 1.0
G7	-75.2 ± 2.1	-165.4 ± 1.2	48.7 ± 1.0	136.3 ± 0.5	173.0 ± 0.7	-93.4 ± 0.8	-108.7 ± 0.4	162.6 ± 2.8
T8	-66.8 ± 0.5	-177.4 ± 0.3	54.5 ± 0.2	134.6 ± 0.2	178.0 ± 0.5	-91.0 ± 1.1	-111.0 ± 0.8	152.2 ± 0.5
T9	-69.3 ± 0.7	-178.0 ± 0.3	57.1 ± 0.2	137.0 ± 0.3	178.3 ± 0.2	-107.1 ± 1.0	-103.6 ± 0.2	167.4 ± 2.0
A10	-72.2 ± 1.3	-168.1 ± 1.3	50.6 ± 0.3	138.8 ± 0.2	173.3 ± 0.2	-89.0 ± 0.4	-105.7 ± 0.6	186.7 ± 0.6
C11	-72.2 ± 0.4	-177.4 ± 0.3	55.2 ± 0.2	136.1 ± 0.3	-179.0 ± 0.3	-110.1 ± 0.4	-110.0 ± 0.2	154.5 ± 0.4
G12	-73.3 ± 0.4	-170.8 ± 0.7	54.1 ± 0.2	139.7 ± 0.2	na	na	-110.5 ± 0.4	182.7 ± 0.7
Average	-73.0 ±1.0	-171.8 ± 0.8	54.0 ± 0.3	137.3 ± 0.3	177.0 ± 0.3	-95.1 ± 0.9	-112.3 ± 0.7	169.5 ± 1.1
B-DNA[c]	-46	-147	36	157	155	-96	-98	162
B-DNA[d]	-65	-193	51	129	203	-120	-103	147

[a]Dihedral angles were calculated using the program CURVES 5.3 (37). Reported values are the averages of the 20 final structures. Error bars are the standard deviations among the 20 final structures and represent the precision of the final structure. See text for discussion of accuracy of the structure. Extra significant figures are given to allow for the reproduction of the structure.
[b]Pseudorotation angle.
[c]B-DNA values from ref. 26.
[d]B-DNA values from ref. 38.

Table 3. Helical parameters[a] for base pairs in the final structure of d(CGTGACGTTACG)$_2$

Base-pairs	propeller twist (°)	inclination (°)	buckle (°)	X-DISP (Å)	interstrand C1′-C1′ distance (Å)	λ_1[b] (°)	λ_2[b] (°)
C1-G12	3.5 ± 2.4	-3.4 ± 3.3	14.1 ± 2.0	0.6 ± 0.2	10.4 ± 0.1	57.5 ± 0.1	50.1 ± 0.2
G2-C11	-9.5 ± 0.5	1.7 ± 1.2	2.5 ± 0.8	0.7 ± 0.2	10.4 ± 0.1	51.2 ± 0.1	58.5 ± 0.2
T3-A10	-13.3 ± 0.3	2.6 ± 1.4	-4.2 ± 0.9	0.7 ± 0.1	10.3 ± 0.1	56.2 ± 0.2	56.0 ± 0.1
G4-T9	-11.0 ± 0.5	3.8 ± 2.3	4.3 ± 0.6	0.7 ± 0.1	10.3 ± 0.1	40.8 ± 0.1	69.5 ± 0.1
A5-T8	-16.8 ± 0.8	3.7 ± 2.1	-1.6 ± 1.0	0.4 ± 0.3	10.6 ± 0.1	52.7 ± 0.1	53.2 ± 0.1
C6-G7	-15.5 ± 0.6	2.5 ± 1.5	-6.3 ± 0.6	0.3 ± 0.4	10.4 ± 0.1	57.2 ± 0.3	50.9 ± 0.2
G7-C6	-15.5 ± 0.6	2.5 ± 1.5	6.3 ± 0.6	0.3 ± 0.4	10.4 ± 0.1	50.9 ± 0.2	57.2 ± 0.3
T8-A5	-16.8 ± 0.8	3.7 ± 2.1	1.6 ± 1.0	0.4 ± 0.3	10.6 ± 0.1	53.2 ± 0.1	52.7 ± 0.1
T9-G4	-11.0 ± 0.5	3.8 ± 2.3	-4.3 ± 0.6	0.7 ± 0.1	10.3 ± 0.1	69.5 ± 0.1	40.8 ± 0.1
A10-T3	-13.3 ± 0.3	2.6 ± 1.4	4.2 ± 0.9	0.7 ± 0.1	10.3 ± 0.1	56.0 ± 0.1	56.2 ± 0.2
C11-G2	-9.5 ± 0.5	1.7 ± 1.2	-2.5 ± 0.8	0.7 ± 0.2	10.4 ± 0.1	58.5 ± 0.2	51.2 ± 0.1
G12-C1	3.5 ± 2.4	-3.4 ± 3.3	-14.1 ± 2.0	0.6 ± 0.2	10.4 ± 0.1	50.1 ± 0.2	57.5 ± 0.1
average	-10.4 ± 0.8	1.8 ± 1.9	0.0 ± 1.0	0.6 ± 0.2	10.4 ± 0.1	54.5 ± 0.2	54.5 ± 0.2
B-DNA[c]	-11.1	2.4	-0.2	0.8	10.9	55.9	55.9

[a]Values were obtained using the program NEWHEL93 (39). Error bars are the standard deviations among the 20 final structures.
[b]λ_1 and λ_2 are the angles between N9/N1–C1′ bonds and the vector for each base pair, where subscripts 1 and 2 represent the left and right base in column 1, respectively.
[c]B-DNA values are from refs 26 and 38.

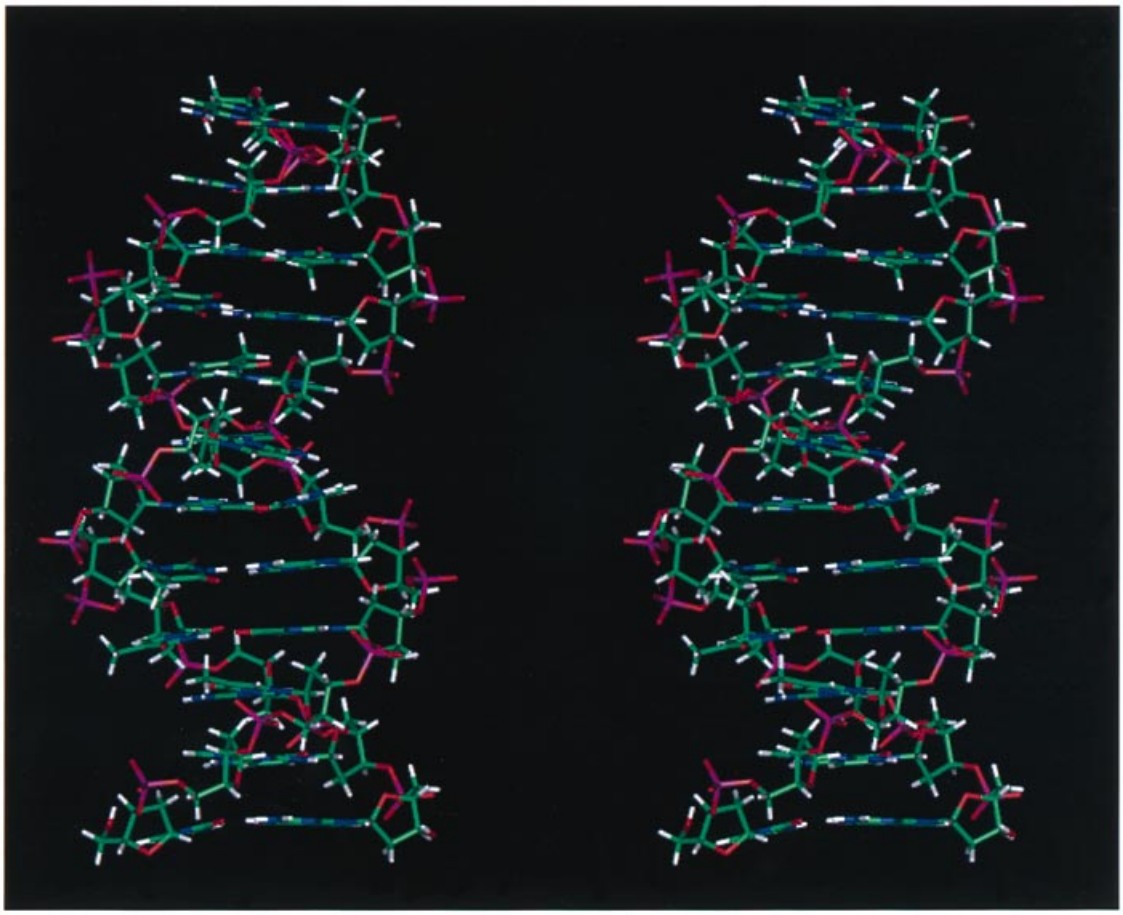

Figure 5. Stereoview of 1 of the 20 final structures of d(CGTGACGTTACG)$_2$.

study indicates precise structural determination in the vicinity of the mismatched pair (Fig. 4B). The small standard deviations in the dihedral angles (Table 2) also indicate precise structure determination.

The accuracy of the final structures is more difficult to test than their precision (45). This is due to the approximations used in generating the final structures (e.g. *in vacuo* simulations, force field parameters, isotropic motion, etc.) and accumulation of errors along the helix. For example, the sugar puckers are known to be dynamic and we observe 75% C2′-*endo* (for C1) but this is not reflected in the precision in δ reported in Table 2. Nonetheless, the quality of the calculated residual R^c and R^x factors indicate that the final structures are in excellent agreement with the experimental NOESY data.

G•T mismatches and the nearest-neighbor model

Close inspection of the 20 final structures reveals that the G·T mismatch is incorporated into the double helix without causing major perturbations that extend beyond the mismatched residues and the neighboring Watson–Crick base pairs (Fig. 5). In fact, the torsion angle and helical parameters indicate there are no major perturbations caused by the G·T mismatch in the double helix that extend beyond its neighboring Watson–Crick residues. This is remarkable considering that the G·T mismatch trimer sequence in this duplex is among the most thermodynamically destabilizing for G·T mismatches (6). This result is, however, consistent with the applicability of the nearest-neighbor model to G·T mismatch thermodynamics in DNA.

Figure 6 shows the stacking interactions in the studied structure (i.e. $\begin{smallmatrix}5'-\underline{T}A-3'\\3'-\underline{G}T-5'\end{smallmatrix}$ and $\begin{smallmatrix}5'-\underline{G}A-3'\\3'-\underline{T}T-5'\end{smallmatrix}$, Fig. 6B and E) and compares them to Watson–Crick interactions when the G·T mismatch is replaced by a G·C pair [i.e. $\begin{smallmatrix}5'-CA-3'\\3'-GT-5'\end{smallmatrix}$ (46) and $\begin{smallmatrix}5'-GA-3'\\3'-CT-5'\end{smallmatrix}$ (47), Fig. 6A and D]. The results show that in the $\begin{smallmatrix}5'-\underline{T}A-3'\\3'-\underline{G}T-5'\end{smallmatrix}$ step (Fig. 6B) there is only slight interstrand and intrastrand overlap between the mismatched guanine and the A·T pair below it. The mismatched thymine does not overlap with the A·T bases below it. In contrast, the step $\begin{smallmatrix}5'-\underline{G}A-3'\\3'-\underline{T}T-5'\end{smallmatrix}$ (Fig. 6E) has considerably more interstrand stacking but essentially no intrastrand stacking interactions between the G·T mismatch and the A·T pair below it. Interestingly, the stability of the nearest-neighbor dimer $\begin{smallmatrix}5'-\underline{T}A-3'\\3'-\underline{G}T-5'\end{smallmatrix}$ is approximately equal to $\begin{smallmatrix}5'-\underline{G}A-3'\\3'-\underline{T}T-5'\end{smallmatrix}$ (ΔG°_{37} of 0.43 and 0.34 kcal/mol of dimer, respectively) (6). Our structural results, in terms of overlap and stacking interactions of $\begin{smallmatrix}5'-\underline{T}A-3'\\3'-\underline{G}T-5'\end{smallmatrix}$ versus $\begin{smallmatrix}5'-\underline{G}A-3'\\3'-\underline{T}T-5'\end{smallmatrix}$, rationalize this thermodynamic trend. Comparison of our G·T mismatch containing dimer $\begin{smallmatrix}5'-\underline{G}A-3'\\3'-\underline{T}T-5'\end{smallmatrix}$ (Fig. 6E) with that determined by Hunter and co-workers (13) (Fig. 6F) shows

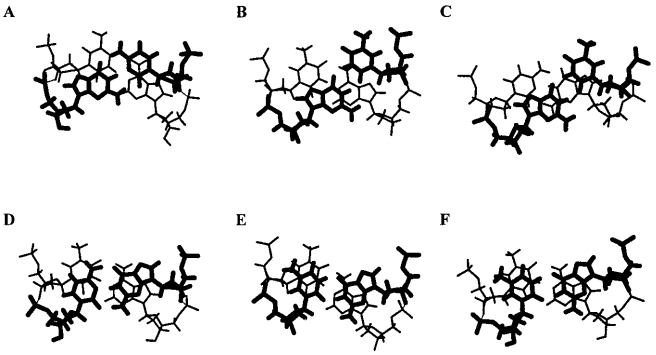

Figure 6. Base overlap showing the stacking interactions between A = CA/GT (46), B = TA/GT, C = TG/GC (13), D = GA/CT (47), E = GA/TT and F = GA/TT (13).

similar stacking interactions between the G·T mismatch and the A·T pair below. This result indicates that the stacking of $^{5'-\underline{GA}-3'}_{3'-\underline{TT}-5'}$ does not depend on the identity of the base pair on the other side of the G·T mismatch since they are different in both studies ($^{5'-\underline{TG}-3'}_{3'-A\underline{T}-5'}$ in this work and $^{5'-\underline{CG}-3'}_{3'-G\underline{T}-5'}$ in the Hunter *et al.* structure; 13). This is also consistent with the nearest-neighbor model.

Comparison of the G·T mismatch containing dimers with the corresponding dimers with G·C pairs (Fig. 6A and D versus B and E), reveals that the guanine residue does not undergo major changes in its stacking orientation with respect to the bases below it upon changing from a G·C pair to a G·T mismatch. However, this is not the case when comparing stacking interactions between A·T versus G·T pairs (Supplementary Material). That is, the thymine residue adopts totally different stacking geometries going from an A·T base pair to a G·T mismatch. This result may explain our previous observations of the strong linear correlation (R^2 of 0.97) observed between the nearest-neighbor thermodynamic parameters of G·T versus G·C pairs (6) and the weak correlation (R^2 of 0.56) when comparing the thermodynamics of G·T versus A·T pairs.

Comparison with previous structural studies on G•T mismatches

Previous one- and two-dimensional NMR studies on G·T mismatch containing duplexes have concluded that the G·T mismatch does not cause major perturbations that extend beyond the neighboring base pairs (9–12). This is confirmed in our solution structure of the duplex d(CGT<u>G</u>ACGT<u>T</u>ACG)$_2$. A complete three-dimensional structure of a G·T mismatch containing B-DNA at a resolution of 2.5 Å of the dodecamer duplex d(CGC<u>G</u>AATT<u>T</u>GCG)$_2$ has also been reported by Hunter and co-workers (13). The duplex in our study and the Hunter *et al.* duplex are similar in terms of their length and the position of the G·T mismatches. The Hunter *et al.* duplex, however, contains a G·T timer motif, $^{5'-\underline{CG}A-3'}_{3'-G\underline{T}T-5'}$, that is more stable than the motif in our duplex, $^{5'-\underline{TG}A-3'}_{3'-A\underline{T}T-5'}$ (ΔG°_{37} of –0.13 and +0.77 kcal/mol of trimer, respectively). Nonetheless, the 3′ neighboring Watson–Crick pair is the same in both studies. Since stacking geometry is not very different between the $^{CG}_{G\underline{T}}$ and $^{TG}_{A\underline{T}}$ dimers (Fig. 6), we conclude that most of the 0.90 kcal/mol difference in stability is due to the extra hydrogen bond in $^{5'-\underline{CG}A-3'}_{3'-G\underline{T}T-5'}$ versus $^{5'-\underline{TG}A-3'}_{3'-A\underline{T}T-5'}$. Hunter *et al.* concluded that the G·T mismatch forms a stable Wobble base pair with minor perturbations of the overall duplex (13). A comparison of stacking interactions between the two studies (Fig. 6B and E versus C and F) reveals that the geometry of stacking interactions in G·T mismatch containing dimer $^{5'-\underline{GA}-3'}_{3'-\underline{TT}-5'}$ that is present in both duplexes are in good agreement.

Biological implications

G·T mismatches, along with G·A and A·C mismatches, are among the most commonly observed mismatches in genomic DNA (48). While thermodynamics may play a major role in the frequency of the occurrences of different mismatches, it is more likely that the enzymatic recognition and repair of mismatches is influenced by the geometry and three-dimensional structure of the mismatch (13,49–51). We have previously shown that G·T, G·A and A^+·C mismatches are among the most stable mismatches in DNA which, in part, explains their common incorporation into DNA during its replication (5,6). However, repair of mismatches in DNA does not seem to follow the observed thermodynamic trend. One might expect that the more stable the mismatch, the less efficient is its repair. However, the stable G·T mismatch is among the most efficiently repaired mismatches in DNA, which may suggest that its repair is based on recognition of structure (52–54).

In this study, we find that while the G·T mismatch is stable compared to other mismatches, its structure has small and highly localized perturbations. Repair enzymes that recognize and excise G·T mismatches may recognize subtle backbone perturbations such as in the torsion angles perturbations in α, ε and ζ that result from the G·T mismatch. A more likely possibility, however, is that mismatch repair enzymes directly recognize the base pair parameters λ_1 and λ_2, which are approximately the same in canonical G·C and A·T pairs (Table 3) but are highly asymmetric for G·T mismatches (λ_1 = 40.8° and λ_2 = 69.5°). This mismatch asymmetry could serve as a recognition element that directs repair enzymes to correct mispairs (55).

ACKNOWLEDGEMENTS

We thank Dr Pete Spielmann for suggestions about MARDIGRAS calculations. We thank Shikha Varma, Rostem Irani and Nicolas Peyret for stimulating conversations. We also thank Wayne State University for funding this research.

SUPPLEMENTARY MATERIAL

Three tables showing the *J*-coupling constants obtained from SPHINX/LINSHA and fraction of S-conformers (Table S1), the *J*-coupling constants and ε values obtained from analysis of the ^{1}H–^{31}P HETCOR (Table S2) and helical parameters for base steps in the final structure (Table S3). Seven figures showing the H1′–H3′ region of the TOCSY spectrum (Fig. S1), the H3′–H4′ region of the TOCSY spectrum (Fig. S2), the ^{1}H–^{31}P HETCOR spectrum (Fig. S3), the imino proton region of the 1D-NOE-difference spectra (Fig. S4), the imino region of the natural abundance ^{1}H–^{15}N HMQC spectrum (Fig. S5), the amino/H8/H6/H1′/H5–imino region of the 300 ms H_2O NOESY (Fig. S6) and the base overlap showing the stacking interactions between $^{TA}_{AT}$, $^{TA}_{G\underline{T}}$, $^{AA}_{TT}$ and $^{GA}_{\underline{TT}}$ (Fig. S7).

See supplementary material available in NAR Online.

REFERENCES

1 Lindhal,T. (1982) *Annu. Rev. Biochem.*, **51**, 61–68.
2 Wang,R.Y., Kuo,K.C., Gehrke,C.W., Huang,L.H. and Ehrlich,M. (1982) *Biochem. Biophys. Acta.*, **69**, 371–377.
3 Johnson,K.A. (1993) *Annu. Rev. Biochem.*, **62**, 685–713.
4 Modrich,P. and Lahue,R. (1996) *Annu. Rev. Biochem.*, **65**, 101–133.
5 Allawi,H.T. and SantaLucia,J., Jr (1998) *Biochemistry*, **37**, 2170–2179.
6 Allawi,H.T. and SantaLucia,J., Jr (1997) *Biochemistry*, **36**, 10581–10594.
7 Allawi,H.T. and SantaLucia,J., Jr (1998) *Biochemistry*, **37**, 9435–9444.
8 Allawi,H.T. and SantaLucia,J., Jr (1998) *Nucleic Acids Res.*, **26**, 2694–2701.
9 Roongta,V., Jones,C.R. and Gorenstein,D.G. (1990) *Biochemistry*, **29**, 5245–5258.
10 Patel,D.J., Kozlowski,S.A., Ikuta,S. and Itakura,K. (1984) *Fed. Proc. Fed. Amer. Soc. Exp. Biol.*, **43**, 2663–2670.
11 Hare,D., Shapiro,L. and Patel,D.J. (1986) *Biochemistry*, **25**, 7445–7456.
12 Kalnik,M.W., Kouchakdjian,M., Li,B.F., Swann,P.F. and Patel,D.J. (1988) *Biochemistry*, **27**, 108–115.
13 Hunter,W.N., Brown,T., Kneale,G., Anand,N.N., Rabinovich,D. and Kennard,O. (1987) *J. Biol. Chem.*, **262**, 9962–9970.
14 Brown,T. and Brown,D.J.S. (1991) In Eckstein,F. (ed.), *Oligonucleotides and Analogues*. IRL Press, NY, pp. 1–24.
15 Marion,D., Ikura,M., Tschudin,R. and Bax,A. (1989) *J. Magn. Reson.*, **85**, 393–399.
16 Piotto,M., Saudek,V. and Sklenar,V. (1992) *J. Biomol. NMR*, **2**, 661–665.
17 Lippens,G., Dhalluin,C. and Wieruszeski,J.-M. (1995) *J. Biomol. NMR*, **5**, 327–331.
18 Szewczak,A., Kellogg,G. and Moore,P. (1993) *FEBS Lett.*, **327**, 261–264.
19 Sklenar,V., Miyashiro,H., Zon,G., Miles,H.T. and Bax,A. (1986) *FEBS Lett.*, **208**, 94–98.
20 Borgias,B.A. and James,T.L. (1990) *J. Magn. Reson.*, **87**, 475–487.
21 Brunger,A.T. (1992) *X-PLOR: A System for Crystallography and NMR*. Yale University Press, New Haven, CT.
22 Liu,H., Thomas,P.D. and James,T.L. (1992) *J. Magn. Reson.*, **98**, 163–175.
23 Liu,H., Speilmann,H.P., Ulyanov,N.B., Wemmer,D.E. and James,T.L. (1995) *J. Biomol. NMR*, **6**, 390–402.
24 SantaLucia,J.J. and Turner,D.H. (1993) *Biochemistry*, **32**, 12612–12613.
25 Lukavsky,P., Billeci,T.M., James,T.L. and Schmitz,U. (1997) In Leontis,N.B. and SantaLucia,J., Jr (eds), *Molecuar Modeling of Nucleic Acids*. ACS symposium series. Washington, DC, Vol., **682**, pp. 122–149.
26 Saenger,W. (1984) *Principles of Nucleic Acid Structure*. Springer Verlag, NY, pp. 51–158.
27 Wimberly,B., Varani,G. and Tinoco,I., Jr (1992) *Biochemistry*, **32**, 1078–1087.
28 Wijmenga,S.S., Mooren,M.W. and Hilbers,C.W. (1993) In Roberts,G.C.K. (ed.), *NMR of Macromolecules*. Oxford University Press, NY, pp. 217–288.
29 Widmer,H. and Wuthrich,K. (1986) *J. Magn. Reson.*, **70**, 270–279.
30 Schmitz,U. and James,T.L. (1995) *Methods Enzymol.*, **259**, 3–44.
31 Altona,C. and Sundaralingam,M. (1972) *J. Am. Chem. Soc.*, **94**, 8205–8212.
32 Varani,G. and Tinoco,I., Jr (1991) *Q. Rev. Biophys.*, **24**, 479–532.
33 Gorenstein,D.G. (1992) *Methods Enzymol.*, **211**, 254–286.
34 James,T.J. (1991) *Curr. Opin. Struct. Biol.*, **1**, 1042–1053.
35 Keepers,J.W. and James,T.L. (1984) *J. Magn. Reson.*, **57**, 404–426.
36 Thomas,P., Basus,V.J. and James,T.L. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 1237–1241.
37 Lavery,R. and Skelnar,H. (1989) *J. Biomol. Struct. Dyn.*, **6**, 655–667.
38 Dickerson,R.E. (1992) *Methods Enzymol.*, **211**, 67–111.
39 Dickerson,R.E. (1989) *EMBO J.*, **8**, 1–4.
40 Wu,M. and Turner,D.H. (1996) *Biochemistry*, **35**, 9677–9689.
41 Wu,M., SantaLucia,J., Jr and Turner,D.H. (1997) *Biochemistry*, **36**, 4449–4460.
42 Ulyanov,N.B., Schmitz,U. and James,T.L. (1993) *J. Biomol. NMR*, **3**, 547–568.
43 Jaishree,T.N., van der Marel,G.A., van Boom,J.H. and Wang,A.H. (1993) *Biochemistry*, **32**, 4903–4911.
44 Lane,A.N. (1997) In Leontis,N.B. and SantaLucia,J., Jr (eds), *Molecular Modeling of Nucleic Acids*. ACS symposium series, Washington, DC, Vol. 682, pp. 106–121.
45 Coppel,Y., Berthet,N., Coulombeau,C., Coulombeau,C., Garcia,J. and Lohmme,J. (1997) *Biochemistry*, **36**, 4817–4830.
46 Goodsell,D.S., Kopka,M.L., Cascio,D. and Dickerson,R.E. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 2930–2934.
47 Drew,H.R., Wing,R.M., Takano,T., Broka,C., Tanaka,S., Itakura,K. and Dickerson,R.E. (1981) *Proc. Natl Acad. Sci. USA*, **78**, 2179–2183.
48 Sloane,D.L., Goodman,M.F. and Echols,H. (1998) *Nucleic Acids Res.*, **16**, 6465–6475.
49 Moran,S., Ren,R.X., Sheils,C.J., Rumney,S. and Kool,E.T. (1996) *Nucleic Acids Res.*, **24**, 2044–2052.
50 Hunter,W.N. (1992) *Methods Enzymol.*, **211**, 221–231.
51 Goodman,M.F. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 10493–10495.
52 Kramer,B., Kramer,W. and Fritz,H.-J. (1984) *Cell*, **38**, 879–887.
53 Modrich,P. (1987) *Annu. Rev. Biochem.*, **56**, 435–466.
54 Dohet,C., Wagner,R. and Radman,M. (1985) *Proc. Natl Acad. Sci. USA*, **82**, 503–505.
55 Goodman,M.F., Creighton,S., Bloom,L.B. and Petruska,J. (1993) *Crit. Rev. Biochem. Mol. Biol.*, **28**, 83–126.