Solution Structure of a GA Mismatch DNA Sequence, d(CCATGAATGG)$_2$, Determined by 2D NMR and Structural Refinement Methods†

Karen L. Greene,‡ Robert L. Jones,§,‖ Ying Li,‡,⊥ Howard Robinson,# Andrew H.-J. Wang,# Gerald Zon,△ and W. David Wilson*,‡

Department of Chemistry and Laboratory for Chemical and Biological Sciences, Georgia State University, Atlanta, Georgia 30303, Department of Chemistry, Emory University, Atlanta, Georgia 30322, Lynx Therapeutics, Inc., Foster City, California 94404, and Department of Cell and Structural Biology, University of Illinois at Champaign–Urbana, Urbana, Illinois 61801

Received September 2, 1993; Revised Manuscript Received November 5, 1993®

ABSTRACT: GA mismatches in DNA have drawn attention because of their special repair mechanisms, stability, and variety of conformations. A symmetric 10-base oligodeoxyribonucleotide duplex, d(CCATGAATGG)$_2$, containing two GA mismatches has been investigated by one- and two-dimensional multinuclear NMR and molecular refinement procedures to ascertain the conformational details of the 5′-pyrimidine-GA-purine-3′ sequence. A molecular model established from the NMR results has a B-type right-handed helix with each of the bases retaining the normal anti-glycosidic torsional angles. Type I mismatched base pairs have GNH_2–AN7 and GN3–ANH_2 (edge-to-edge) hydrogen bonds, while type II base pairs have GN1H–AN1 and GO6–ANH_2 (face-to-face) bonds. The conformation at the GA mismatch site has type I GA base pairs and an unusual cross-strand stacking of the adjacent G5 and A6 bases, which causes significant overwinding of the helix at the mismatch site. Unusual shifts of the ^{31}P resonances suggest that the phosphate linkage between G5 and A6 is no longer in the low-energy B_I conformation. One-dimensional imino and phosphorus NMR studies were carried out on a number of DNA sequences containing adjacent 5′-GA-3′ mismatched base pairs to investigate the sequence dependence of the conformations and base-pairing types. Type I and type II conformations have very different imino proton and ^{31}P NMR spectral patterns that can be used to classify any sequence with adjacent GA mismatches by base-pairing and conformational type. The NMR results indicate that the conformation selected is dictated completely by the flanking sequence: 5′-pyrimidine-GA-purine-3′ sequences adopt the type I conformation, while 5′-purine-GA-pyrimidine-3′ sequences have the type II conformation.

As part of a long-term project to investigate the structure and thermodynamics of backbone-modified nucleic acids as antisense drugs, we discovered an oligodeoxyribonucleotide that exhibited an unexpected thermal transition (Wilson et al., 1988). Thermodynamic studies indicated that the oligomer formed a duplex, and base-pairing considerations suggested a purine-rich duplex with four GA mismatches out of 10 base pairs, as shown in **a-11** (Table 1). Replacement of the C in **a-11** by a G base caused a marked destabilization of the duplex, in agreement with the proposed base-pairing scheme. GA mismatch base pairs normally destabilize DNA duplexes relative to Watson–Crick base pairs (Aboul-ela et al., 1985), and the unusual stability of a duplex with four proposed GA mismatches was quite puzzling. We initiated a series of thermodynamic and structural studies on **a**, a 10-base analog of **a-11**, and several related sequences to determine the molecular basis for the unusual stability (Li et al., 1991a,b).

Initial NMR studies (Li et al., 1991b) indicated that the sequence does form a duplex with four GA mismatched base pairs. Each GA is in an anti–anti conformation with a previously unobserved base edge-to-base edge type of hydrogen-bonding (type I in Figure 1). The adjacent purine–purine mismatches stack over each other and fit quite well into a standard B-form duplex geometry. Single GA mismatches cannot form this type of stacked structure and, as expected from previous thermodynamic studies (Aboul-ela et al., 1985), are destabilizing (Li et al., 1991a). NMR investigations of other sequences subsequently have identified type I GA mismatches in RNA (Heus & Pardi, 1991) and DNA (Maskos et al., 1992; Chou et al., 1992). The RNA structure is quite different in that a single type 1 GA base pair helps close a tetraloop that is very stable and quite common in ribosomal

Table 1

label	sequence	base pair type
a-11	ATGAGCGAATA ATAAGCGAGTA	I
a	ATGAGCGAAT TAAGCGAGTA	I
a-AG	ATAGGCAGAT TAGACGGATA	
b	C_1 $C_2A_3T_4G_5A_6A_7T_8G_9G_{10}$ $G_{10}G_9T_8A_7A_6G_5T_4A_3C_2C_1$	I
c	CCAAGATTGG GGTTAGAACC	II
d	GCAGGC CGGACG	II
e	GCGAGC CGAGCG	I

† This work was supported by NIH Grants AI-27196 (W.D.W.) and GM-41612 (A.H.-J.W.).

‡ Georgia State University.
§ Emory University.
‖ Current address: Centers for Disease Control, Center for Environmental Health, Atlanta, GA 30341.
⊥ Current address: Hybridon, Inc., Worchester, MA.
University of Illinois at Champaign–Urbana.
△ Lynx Therapeutics, Inc.

® Abstract published in *Advance ACS Abstracts,* January 15, 1994.

0006-2960/94/0433-1053$04.50/0

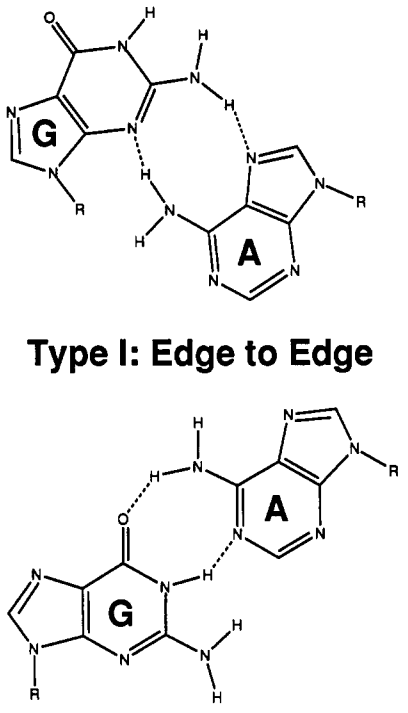

FIGURE 1: Schematic drawing of the hydrogen-bonding in G(anti)–A(anti) base pairs. The type I base pair involves hydrogen-bonding at the edges of the bases and is called an edge-to-edge base pair, while type II involves base face-to-face pairing (as in Watson–Crick base pairs).

RNA (Heus & Pardi, 1991). Morden and co-workers (Maskos et al., 1993) have found that adjacent GA and AA mismatches can form a stacked structure with type I edge-to-edge base pairs, and the overall structure is very similar to that obtained with two adjacent GA mismatches. The stability of the stacked purine–purine structure is, thus, not restricted to GA mismatched base pairs. Other studies on DNA oligomers have confirmed the primary features of our base-pairing and stacking model (Lane et al., 1992; Chou et al., 1992).

Thermodynamic studies on a number of sequences related to **a** (Table 1) have provided important information on the sequence dependence of the formation of the unusual adjacent-stacked GA mismatch structure (Li et al., 1991b; Ebel et al., 1992). Reversal of the GA (**a**) to an AG (**a-AG** in Table 1) sequence completely destabilizes the structure. Sequences 5′-XGAY-3′ where X is a pyrimidine and Y is a purine are very stable, but when X is the purine and Y is the pyrimidine, the stability decreases markedly.

NMR studies by Kan et al. (1983) and by Gorenstein and co-workers [cf. Nikonowicz and Gorenstein (1990)] demonstrated that the sequence **c** (Table 1) formed a duplex with adjacent anti–anti GA base pairs of type II (Figure 1), and Dickerson and co-workers (Prive et al., 1987) obtained an X-ray structure for **c** in which X is a purine and Y is a pyrimidine. The difference in base-pairing type for the adjacent GA mismatches in sequences **a** and **c** raises the interesting question of whether **c** would switch from type II to type I base-pairing for the GA mismatches if X were changed to a pyrimidine and Y to a purine (sequence **b** in Table 1). We have synthesized **b** and conducted a detailed NMR analysis of its structure to answer this question. The structure was determined by the method recently described by Robinson and Wang (1992).

Table 2: ^{1}H and ^{31}P Chemical Shifts (ppm) of the $d(CCATGAATGG)_2$ Sequence at 30 °C

	nucleotide sequence									
signal	C1	C2	A3	T4	G5	A6	A7	T8	G9	G10
H1′	5.94	5.57	6.21	5.93	5.99	5.53	5.91	5.60	5.70	6.09
H2′	2.11	2.22	2.57	1.39	2.93	0.83	2.71	1.89	2.60	2.49
H2″	2.49	2.50	2.84	2.09	2.66	2.14	2.80	2.22	2.70	2.34
H3′	4.65	4.86	4.98	4.78	5.08	4.77	4.98	4.80	4.96	4.64
H4′	4.10	4.15	4.39	4.22	4.52	4.34	4.43	4.15	4.34	4.18
H5′	3.74	4.05	4.17	4.17	4.18	4.05	4.08	4.22	4.09	4.20
H5″	3.74	4.01	4.10	4.08	4.07	4.01	4.08	4.09	4.02	4.11
H2			7.78			7.82	7.52			
H5	5.90	5.69								
H6	7.74	7.60		6.87				7.12		
H8			8.29		8.17	7.39	8.34		7.77	7.76
CH_3				1.22				1.42		
^{31}P	−4.45	−4.25	−4.73	−5.42	−2.96	−4.65	−4.86	−4.25	−4.33	

Turner and co-workers (Santa Lucia et al., 1990) have found that the RNA sequence 5′-r(GCGAGCU) forms a stable duplex with a U overhang, but with no imino proton signal in the 12–14-ppm region for the G in the GA mismatch. The sequence 5′-r(GCAGGCG) forms a duplex with a G overhang and a more standard imino proton signal for the G in the central GA mismatched base pairs. To investigate whether related hexamer sequences as DNA exhibit similar behavior, and whether the observed differences could be explained by type I and II mismatched base pairs in the two RNA sequences, oligomers **d** and **e** (Table 1) were synthesized and investigated by ^{31}P and imino proton NMR methods. We find that **c** and **d**, where X is a purine and Y is a pyrimidine, have type II (face to face) GA mismatches, while **b** and **e**, where X is a pyrimidine and Y is a purine, have type I edge-to-edge base pairing for the mismatches. The overall conformation for the type I mismatch in **b** is similar to that originally proposed by Li et al. (1991a,b) for **a**.

EXPERIMENTAL PROCEDURES

Materials and Methods. The oligodeoxynucleotides were synthesized and purified as described previously (Li et al., 1991a). The concentrations of the oligonucleotides were determined optically using extinction coefficients per mole of strand at 260 nm determined by the nearest-neighbor procedure (Fasman, 1975). The NMR samples were prepared as described in Li et al. (1991a).

NMR Experiments. All of the 1D exchangeable proton spectra were collected on a Varian VXR 400 NMR spectrometer. The samples were in PIPES buffer (10 mM PIPES, 0.1 M NaCl, and 1 mM EDTA) at pH 7.0 and 0 °C. For all of the ^{1}H 1D NMR experiments, sodium (trimethylsilyl)-propionate-*2,2,3,3-d*$_4$ (TSP) was used as the reference at 0.0 ppm. The 1–1 solvent suppression pulse sequence (Hore, 1983) was used with the carrier frequency set at the water resonance. The sweep width was 10 000 Hz with a recycle delay of 1 s; 2-Hz line broadening was applied to the spectra. The ^{31}P NMR spectra were also collected at the same conditions as the 1D ^{1}H spectra. Samples in a 5-mm NMR tube were inserted into a 10-mm tube with PIPES buffer in D_2O. The spectra were obtained on a Jeol GX 270 under the following experimental conditions: 45° flip angle; 4-s pulse repetition, broad-band proton decoupling; 2000-Hz spectral width; 4-Hz line broadening applied before Fourier transformation. Trimethyl phosphate (TMP) was used as an external reference.

The 2D NMR spectra were taken on either a GE GN-600 or GE GN-500 spectrometer, and the data were processed on Silicon Graphics computers using the program FELIX (Hare

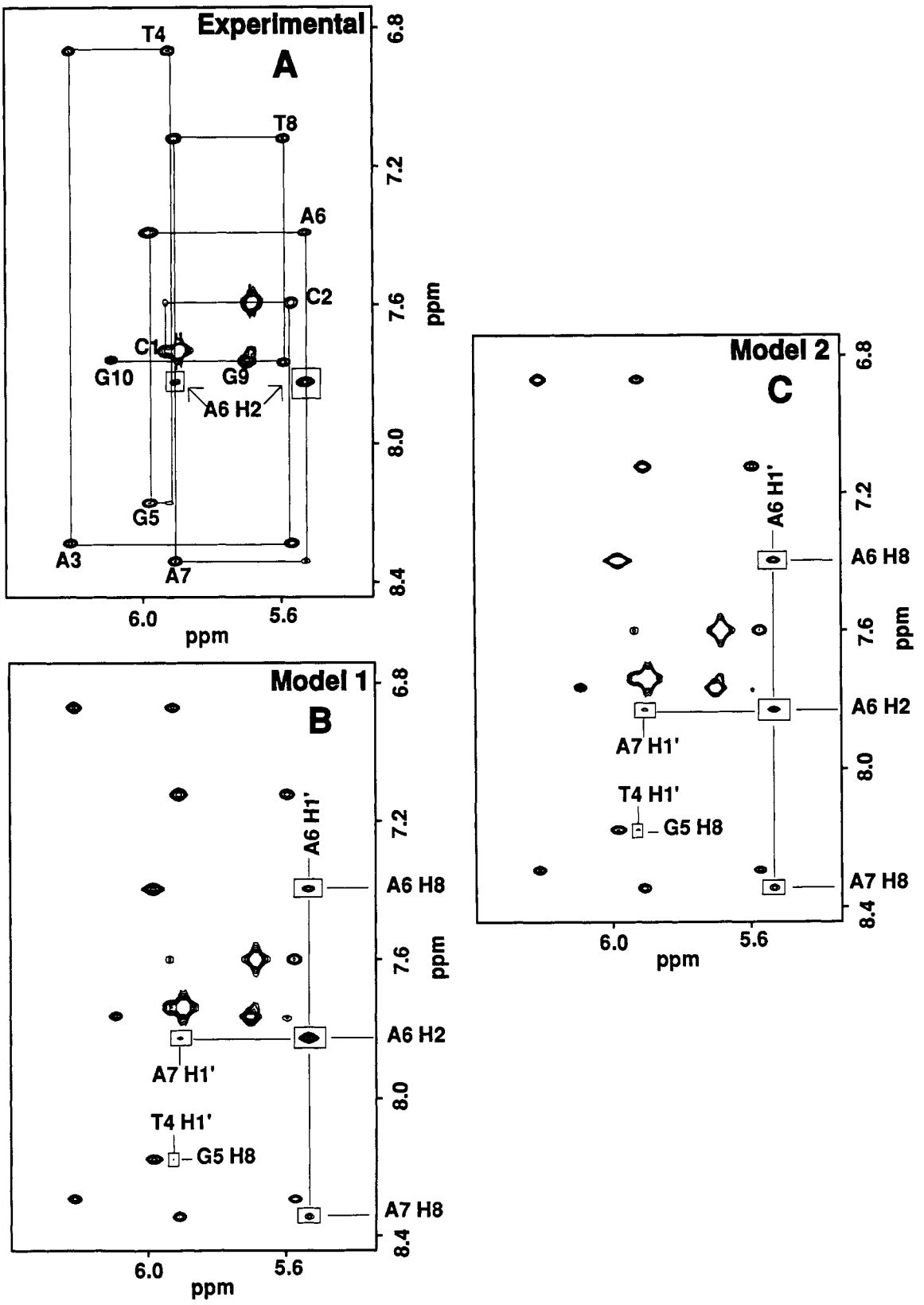

FIGURE 2: Experimental (A) and simulated spectra from the final structures of model 1 (B) and model 2 (C) of the H8/H6/H2 to H1′ region of the 2D NOESY spectrum at 200 ms of sequence **b**. The solid line in panel A follows an unbroken H6/H8 to H1′ network, indicating that the overall form of the structure is a B-form helix. As seen in panel A, the crosspeaks of A6H8–A6H1′, A7H8–A6H1′ and G5H8–T4H1′ are all of much lower intensity than other comparable crosspeaks. Also of note are the unusually strong A6H2 to A6H1′ and A7H1′ crosspeaks and the upfield shift of A6H8 (~1 ppm further upfield than A3H8 or A7H8). The 2D NOESY spectra in panels B and C both agree with the experimental spectrum, including reproduction of the A6H2 crosspeaks.

Research, Inc.). A 1D temperature study (supplementary material, Figure S1) showed 30 °C to be the optimal temperature for minimal chemical shift overlap. The 2D spectra were obtained with a sweep width of 5555 Hz in both dimensions for the GN-600 spectrometer and a sweep width of 5000 Hz on the GN-500. All processed spectra were zero-filled to a final dimension of 2K × 2K. Phase-sensitive NOESY (Jeener et al., 1979; Macura & Ernst, 1980) spectra

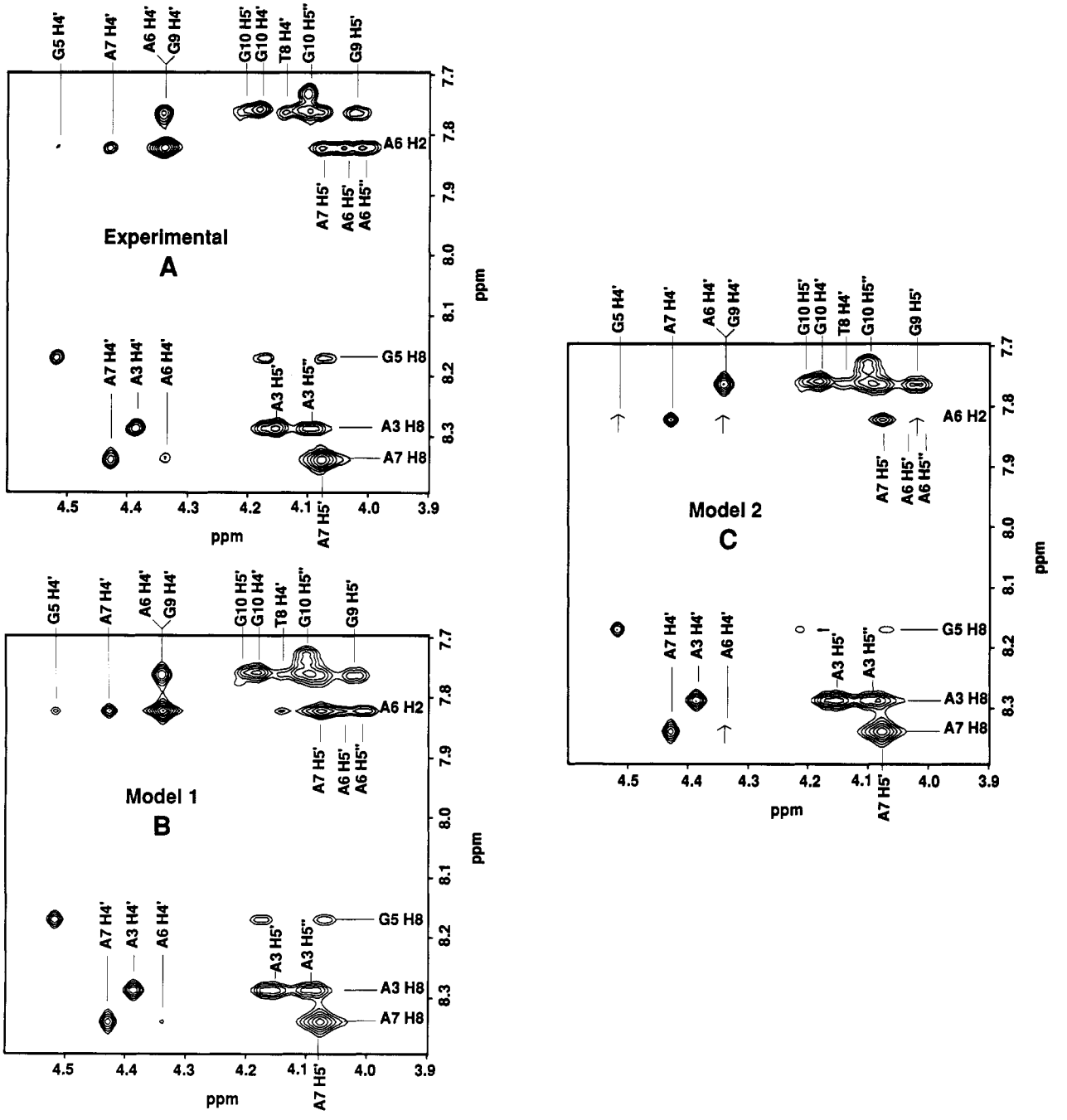

FIGURE 3: Expansion of the 2D NOESY spectrum at 200 ms of the H8/H6/H2 to H4′/H5′/H5″ region of sequence **b**. Shown here are the experimental spectrum (A) and the simulated spectra from the structures of model 1 (B) and model 2 (C). Of particular interest are the unique A6H2 crosspeaks. The A6H2 to A6H5′, A6H5″, A6H4′, A7H5′, A7H4′, and G5H4′ all appear in both the experimental spectrum (A) and the simulated spectrum of model 1 (B); however, only the crosspeaks of A6H2 to A7H5′ and A7H4′ are reproduced in the simulated spectra of model 2 (C). The A7H8–A6H4′ also appears in the spectra of panels A and B but not in that of panel C. These unusual A6H2 crosspeaks denote the type 1 interstrand cross-stacking of G5 and A6 as seen in Figure 1A. The type 2 base-pairing scheme of model 2 (Figure 1B) prevents cross-stacking, and these crucial crosspeaks do not appear.

were obtained at mixing times of 100 or 200 ms using the method of States (States et al., 1982). The NOESY spectrum used for refinement purposes was collected with a 200-ms mixing time, a 9-s relaxation delay, and a total of 2048 complex points in the t_2 dimension and 1024 points in the t_1 dimension with a total of 48 scans for each t_1 point. The NOESY spectra were processed in t_2 by an exponential multiplication of 4 and a fifth-order polynomial base-line correction and in t_1 by linear prediction of the first data point followed by an exponential multiplication of 4 and a fifth-order polynomial base-line correction. Special effort was given to the phasing of the spectral peaks and to the polynomial base-line corrections in both dimensions due to the sensitivity in the data refinement program to line shape.

The HOHAHA (Davis & Bax, 1985a,b) data set was collected as 1024 complex × 512 points in the t_2 and t_1 dimensions, respectively. Each t_1 point was the total of 32 scans. The data set was processed in t_2 by a sine-bell of 1024 points shifted by 90° and a fifth-order polynomial base-line correction and in t_1 by a sine-bell of 1024 points shifted by 90°. The phase-sensitive COSY (Aue et al., 1976; Nagayama et al., 1979) spectrum was collected with 2048 points in the t_2 dimension and 1024 points of 32 scans accumulated per point in the t_1 dimension and was processed in a manner similar to the HOHAHA data set.

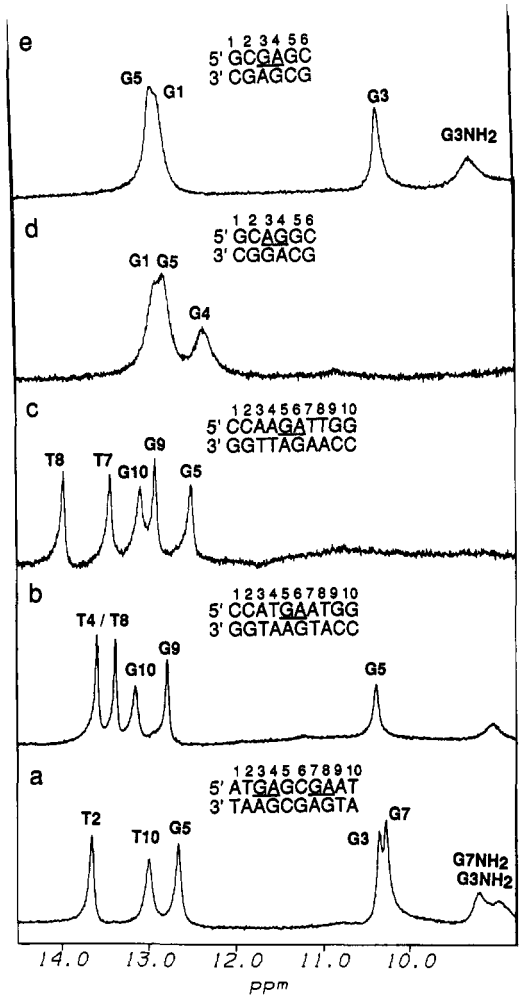

FIGURE 4: Exchangeable proton NMR spectra of sequences **a**–**e**. The characteristic pattern for duplexes with 5'-pyrimidine-GA-purine-3' is easily distinguished from the pattern for duplexes with 5'-purine-GA-pyrimidine-3' or 5'-AG-3'. The guanine imino proton associated with the mismatch has been shifted almost 2 ppm upfield of normal guanine imino resonances for sequences **a**, **b**, and **e**. All three of these sequences are of the form 5'-pyrimidine-GA-purine-3', and **a** and **b** have a type I base pair in which the imino proton is not hydrogen-bonded. The imino shifts of sequences **c** and **d** remain in the spectral range for Watson–Crick base pairs. The exchangeable proton peaks at about 9.5 ppm in **a** and **e** belong to the amino protons of the mismatched guanines.

A phase-sensitive ^1H–^{31}P heteronuclear reverse chemical shift correlation (Sklenar & Bax, 1987; Sklenar et al., 1986) spectrum was collected with 1024 complex data points in the t_2 dimension and 80 scans accumulated per each of 128 t_1 increments. The spectral width was optimized to include the H3', H4', and H5'/H5'' regions.

Structural Refinement. All models of sequence **b** (Table 1) were created with the SYBYL software package (Tripos, Inc.) as starting structures for the cyclic refinement program, SPEDREF (Robinson & Wang, 1992). The starting structures utilized the two separate hydrogen-bonding schemes, types I and II (Figure 1), for the GA mismatch site. For each of the type I models, the 16 flanking Watson–Crick base-paired nucleotides (Table 1, **b**) were built using the B-form coordinates of Arnott (Chandrasekaran & Arnott, 1989). These flanking nucleotides were docked onto a previously built GA mismatch region having type I base pairing (Li et al., 1991b). The starting model with type II base pairing was initially from coordinates obtained from the Protein Data Bank (Bernstein et al., 1977) for the Dickerson GA mismatch (Table 1, **c**) (Prive et al., 1987). The base pairs flanking the mismatch site were then mutated in SYBYL to give sequence **b** (Table 1). Initial minimization of all of the models was carried out to an RMS gradient of 0.08 using the Kollman all-atom force field and a distance-dependent dielectric function of 4 as previously described (Veal & Wilson, 1991). For comparison purposes, two other structures were built and minimized in SYBYL using the same parameters as above. One structure was a similar B-form duplex having only Watson–Crick base pairs. The second structure was a mutated version of the duplex with type I base pairs in which the pyrimidine and purine bases flanking the GA mismatch site have been reversed to give a 5'-purine-GA-pyrimidine-3' sequence. Analysis of the helical parameters of the refined models was carried out using NEWHEL92 (provided by R. E. Dickerson).

RESULTS

Nonexchangeable Proton Assignments for d(CCAT-GAATGG)$_2$. The nonexchangeable chemical shifts of **b** (Table 1 with nucleotide numbering) were obtained through the standard sequential assignment methods with NOESY and COSY 2D spectra (Feigon et al., 1982; Scheek et al., 1983; Chary et al., 1987; Neuhaus & Williamson, 1989; Wuthrich, 1986). The aromatic to H1' region (Figure 2A) was assigned via the NOE crosspeaks seen between each of the base H8/H6 protons to its own H1' proton and to the H1' proton in the 3' direction. The ability to "walk through" the base aromatic to H1' protons identifies the duplex as a B-form right-handed helix. Each of the H8/H6 to H1' crosspeaks is of an intensity that is significantly less than that for the CH5 to CH6 crosspeaks, and this result shows that all bases, including G5 and A6, have the anti-glycosidic torsional angles.

Several of the signals in this region are worth noting for their importance in establishing structural details (Figure 2A). The crosspeak between A6H8 and A6H1' is of unusually low intensity compared to the A6H8–G5H1' crosspeak, and the G5H8–T4H1' and A7H8–A6H1' signals are unusually small. Further, the A6H8 resonance itself is shifted much farther upfield, between 0.8 and 1 ppm, than the adenine H8 signals not involved in mismatched base pairing, indicating that A6H8 has been shifted into a stronger ring current region than either A3 or A7. While the expected sequential connectivities are seen, the unusual intensities of these crosspeaks argue that there is some unique deformation of the helix within the GA mismatch site. Each of these unique features is also seen in the spectra simulated from the refined models (Figure 2B,C).

The AH2 crosspeaks also play an important role in delineating the cross-strand stacking at the mismatch site. Adenosine H2 protons have relatively long T1 relaxation values (2.8 s); therefore, the NOESY spectrum for refinement was collected with a 9-s delay. Unusual A6H2–A6H1' and A6H2–A7H1' crosspeaks are seen (Figure 2A). The A6H2–A6H1' crosspeak is identified as an A6H2 interaction with the opposite strand A6H1' on the basis of the work of Li et al. (1991) with sequence **a** (Table 1). The intensity of this crosspeak is quite strong and is reproduced in the spectra simulated from both models 1 and 2.

The crosspeaks identified in the H8/H6/H2 to H4'/H5'/H5'' region (Figure 3) are also valuable in defining the structure [the H5' and H5'' signals are assigned as in Robinson and Wang (1992)]. Again, a number of the A6H2 crosspeaks were instrumental in identifying the cross-strand stacking nature of the mismatch site. Interstrand crosspeaks between A6H2 and A6H5' and between A6H5'' and A6H4' were all cleanly resolved in the spectrum (Figure 3A), as were intrastrand A6H2 to A7H5' and G5H4' signals.

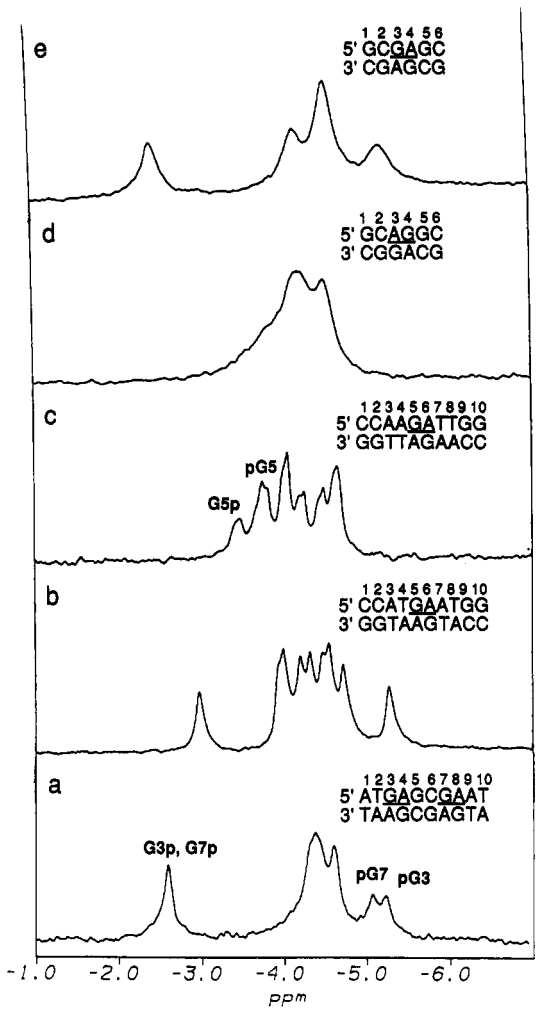

FIGURE 5: ^{31}P NMR spectra of sequences **a**–**e** showing the characteristic ^{31}P NMR pattern for duplexes with 5′-pyrimidine-GA-purine-3′ is easily seen and distinguished from the patterns for other 5′-GA-3′ or 5′-AG-3′ mismatched sequences. Sequences **a**, **b**, and **e** all have phosphate resonances shifted downfield of the main band by ~1.5 ppm and a resonance shifted upfield by ~0.5–1 ppm. The downfield-shifted phosphate resonances of **a** and **b** have been assigned as belonging to the GpA steps, while the upfield peaks belong to the preceding 5′-TpG step. In contrast, the ApG and GpA steps of sequence **c** are clustered within a main band of phosphate resonances. The clear differences in the 1D ^{31}P spectra can then be used to predict that the GpA phosphate of sequence **d** remains in a normal B_I form, while that of sequence **e** does have an unusual phosphate conformation at the GpA step.

The resonances of the H2′ and H2″ signals are unusually shifted for the G5 and A6 bases (supplementary material, Figure S2A). As has been observed by others (Li et al., 1991b; Lane et al., 1992; Chou et al., 1992), in a type I GA mismatch, the G5H2′ is shifted downfield of the H2″, while A6H2′ is far upfield of normal B-form H2′ signals. The H2′ is generally shifted downfield of the H2″ only for the H3′ terminal bases, and this result indicates that the G5 base of **b** is shifted out of the ring current field of its 3′ neighbor, A6. Conversely, the upfield shifts of A6H2′ and A6H8 both indicate that A6 has moved further into the ring current field of its neighboring base. Also, the A6H8 to A6H2′ and H2″ crosspeaks are much less intense than comparable crosspeaks and the A6H8 to G5H2′ and H2″ signals are almost nonexistent, whereas G5H8 has a strong NOE to the T4H1′, H2′, and H2″ signals. All of these crosspeaks are consistent with deviations in the G5 and A6 bases from a standard B-form helix conformation and with the cross-stacking as proposed by Li et al. (1991b).

Model Refinement for **b**. Imino proton results [see below and Li et al. (1991a,b)] clearly demonstrated that **b** formed a duplex with two mismatched and eight Watson–Crick base pairs. Two families of models were constructed for the duplex, as described in the Experimental Procedures section, to investigate convergence to the NMR results by the SPEDREF routine. One family had type I GA mismatched base pairs, while the other had type II GA base pairs. In both families, the mismatched base pairs were flanked by normal Watson–Crick base pairs with a B-form type structure. From both the SYBYL and SPEDREF results, it was quickly apparent that the models with type II GA base pairs have higher energy and do not fit the experimental NMR spectrum as well as the type I models. For the type I family, five separate starting models, with significantly different ϵ and ζ angles in the GA mismatch site, were refined to assist in finding the global minimum.

The rotational correlation time was determined in SPEDREF for each of the refined models through comparison of the R factor and error values to τ_c. Final minimizations were carried out at the optimized τ_c value (~2.9 ± 0.1 ns). Each of the five models with the type I GA mismatches minimized to similar R factors (<16.5%) and error values. Comparisons of the simulated spectra to the original NOESY spectrum [cf. Figures 2, 3, and S2 (supplementary material)] clearly show that model 1 with type I mismatches produces a simulated spectrum in excellent agreement with the experimental data set, while the spectrum of model 2 with type II mismatches is missing many of the NOEs unique to the GA mismatch site.

The spectrum simulated from model 1 (Figure 3B) closely matches the spectrum derived from the experimental data. As seen in Figure S3 of the supplementary material, the cross-strand stacking of A6 and G5 is unambiguous. The placement of the adenosines over one another brings the A6H2 of one strand into close contact with the A6H5′, A6H5″, and A6H4′ of the cross-strand. The A6H2 is now quite close to the cross-strand G5H4′ and to the H4′ of the neighboring A7 as well. The hydrogen bonding in model 2 (Figure 3C), however, prevents cross-stacking of the G5 and A6 bases, and the A6H2 to A6H5′, A6H5″, A6H4′, and G5H4′ crosspeaks are missing entirely from the simulated spectrum. Moreover, although the protons of G5H8 are close enough to those of T4H5′ and T4H5″ to produce NOEs, they are much weaker crosspeaks than those in the experimental spectrum. Finally, while the crosspeak between A7H8 and A6H4′ is small in the experimental spectrum and in the spectrum of model 1, it is totally missing from the spectrum of model 2. It is clear that, although the SPEDREF routine can force model 2 into a structure that matches many of the experimental signals, it cannot overcome the incorrect base pairing at the mismatch site.

There is excellent agreement between the experimental and simulated spectra of model 1 for the aromatic to H2′/H2″ region (supplementary material, Figure S2A,B). Again, this region contains crosspeaks that are particular to the nature of the cross-strand stacking of GA mismatches, such as the far upfield shifted A6H2′ signal and the G5H2″ crosspeak, which is moved upfield of the G5H2′ peak. The twisting of the G5 and A6 bases causes the A6H8 to G5H2′ and G5H2″ crosspeaks to be much smaller in intensity than standard (n)-H8 to (n – 1)H2′/H2″ signals. This is reflected in the longer distances between the A6H8 to G5H2′ and H2″ as measured in model 1 and, hence, the much smaller crosspeaks in the spectrum. The position of G5 has also brought the G5 and T4 bases close enough for there to be a tiny but unusual crosspeak between G5H8 and T4CH_3.

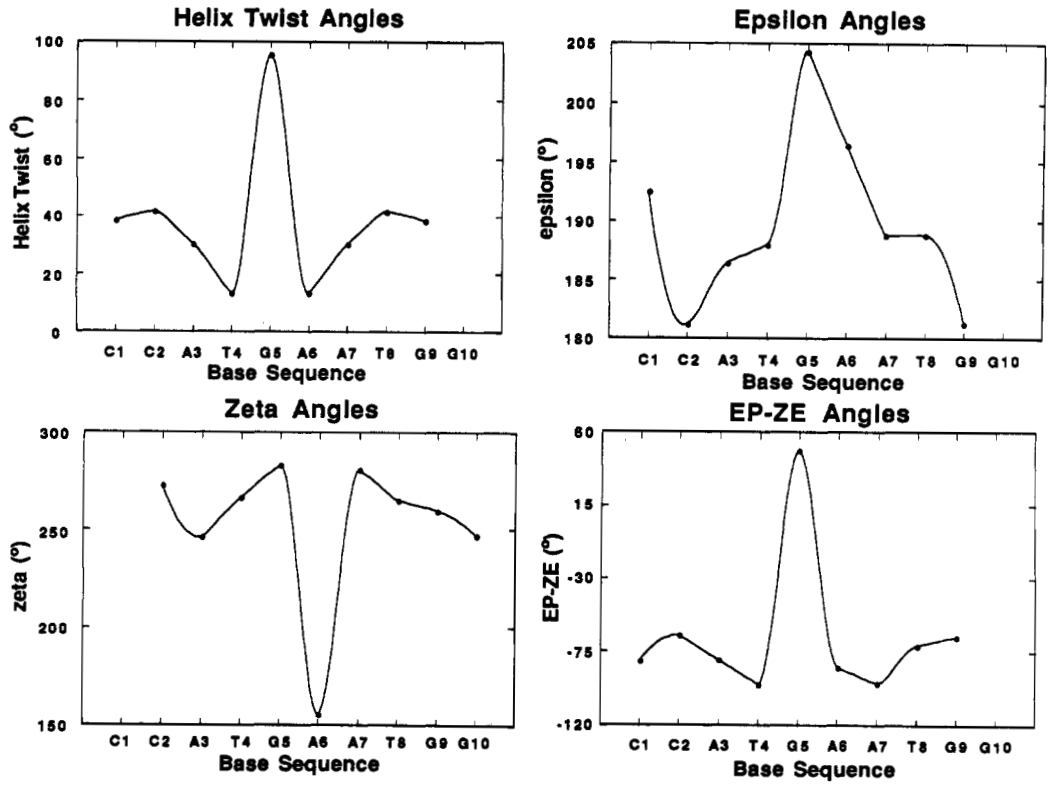

FIGURE 6: Plot of the backbone angles EP-ZE (A, bottom right), ϵ (B, top right), ζ (C, bottom left), and the helix twist angles (D, top left) as calculated by the NEWHEL92 program (R. E. Dickerson) from the structure of model 1. In each of the plots, the individual values for G5 and A6 deviate greatly from those of their neighbors, indicating a distortion in the phosphate backbone at the GA mismatch site. In particular, the ϵ and ζ values of G5 are ~155 and ca. −155°, respectively, which denotes a lack of BI conformation. The helix twist angles are overwound for G5 at ~100° and are underwound for the neighboring T4 and A6 bases at ~5° each.

1D Exchangeable and Phosphorus Spectra. The imino and amino assignments of five sequences (Table 1) were obtained through standard one-dimensional NOE methods in H_2O (Li et al., 1991a,b; Cheng et al., 1992). The 1D exchangeable proton and 1D phosphorus spectra were collected to further study the role of sequence in the determination of type I or type II base pairing in duplexes with GA mismatch structures. As seen in Figure 4, the pattern of imino chemical shifts is quite distinct for the five sequences shown and depends on the type of bases flanking the GA mismatch site. For sequence **b**, whose structure has been discussed above, the eight Watson–Crick-type base-paired imino proton resonances are located in the normal region (12.5–14 ppm); only four peaks are observed because the sequence is self-complementary (Figure 4). The imino proton resonance of the mismatched G is at 10.5 ppm, and the peak at about 9.2 ppm is identified as the resonance of the mismatched G amino group. This pattern is the same as for sequences **a** and **e**. Such an upfield shift for G5 indicates that the imino proton is not involved in a hydrogen bond as expected for type I base pairing. The imino spectra of **c** and **d** are consistent with each other and with a standard base-pairing interaction as in the type II model.

Pattern groupings are also seen in the phosphorus spectra of sequences **a**–**e** (Figure 5). The ^{31}P NMR spectrum of **b** resembles the spectra of **a** and **e**, with similar upfield and downfield ^{31}P NMR signals clearly shifted from the main band and much more widely dispersed than the ^{31}P signals of **c** and **e** (Nikonowicz & Gorenstein, 1990). The clustering of phosphorus signals in the spectra of **c** and **e** is consistent with a general B-form backbone conformation. The downfield shifts of the mismatched guanine phosphorus signals in sequences **a**, **b**, and **e** indicate that the phosphate backbone deviates from a standard B-type structure (Gorenstein, 1984; Gorenstein et al., 1988; Nikonowicz & Gorenstein, 1990; Roongta et al., 1990). The remarkable similarity of the imino proton and phosphorus NMR spectra of **a**, **b**, and **e** strongly indicates that the mismatches in **e** adopt type I GA pairing and, hence, have a conformation similar to that of **a** and **b**. Each of the type I base-paired structures has a 5′-pyrimidine-GA-purine-3′ sequence in common.

A heteronuclear ^{1}H–^{31}P reverse chemical shift correlation spectrum was used to confirm the phosphorus and the H3′/H4′ proton resonance assignments in **b**. In such an experiment, the $(n)^{31}P$ is correlated to the $(n-1)$H3′ and to the (n)H4′ and H5′/H5″protons. As seen in Figure S4 of the supplementary material, there are nine reasonably well-resolved signals in the H3′ proton region between 5.2 and 4.6 ppm. The H4′ and H5′/H5″ proton region, between 4.5 and 3.8 ppm, is more densely overlapped. Phosphorus assignments were made by comparison with the known proton assignments. As in sequence **a**, the phosphorus resonances flanking the mismatched guanosine are shifted either downfield or upfield of the main band of phosphorus shifts.

DISCUSSION

All of the ^{1}H and ^{31}P results indicate that oligodeoxynucleotides with a 5′-purine-GA-pyrimidine-3′ sequence have type II face-to-face hydrogen bonding (Table 1), while those GA mismatches with reversed flanking bases, 5′-pyrimidine-GA-purine-3′, have type I edge-to-edge hydrogen bonds (Table 1). Oligodeoxynucleotide **b**, $d(CCATGAATGG)_2$, with a 5′-pyrimidine-GA-purine-3′ sequence, was synthesized for in-depth NMR analysis to investigate the hypothesis that the sequence would switch to type I instead of the type II hydrogen-bonding pattern observed for sequence **c** (Table 1). The structure resulting from this study proves this to be the case. A comparison of the spectra derived from the refined structures

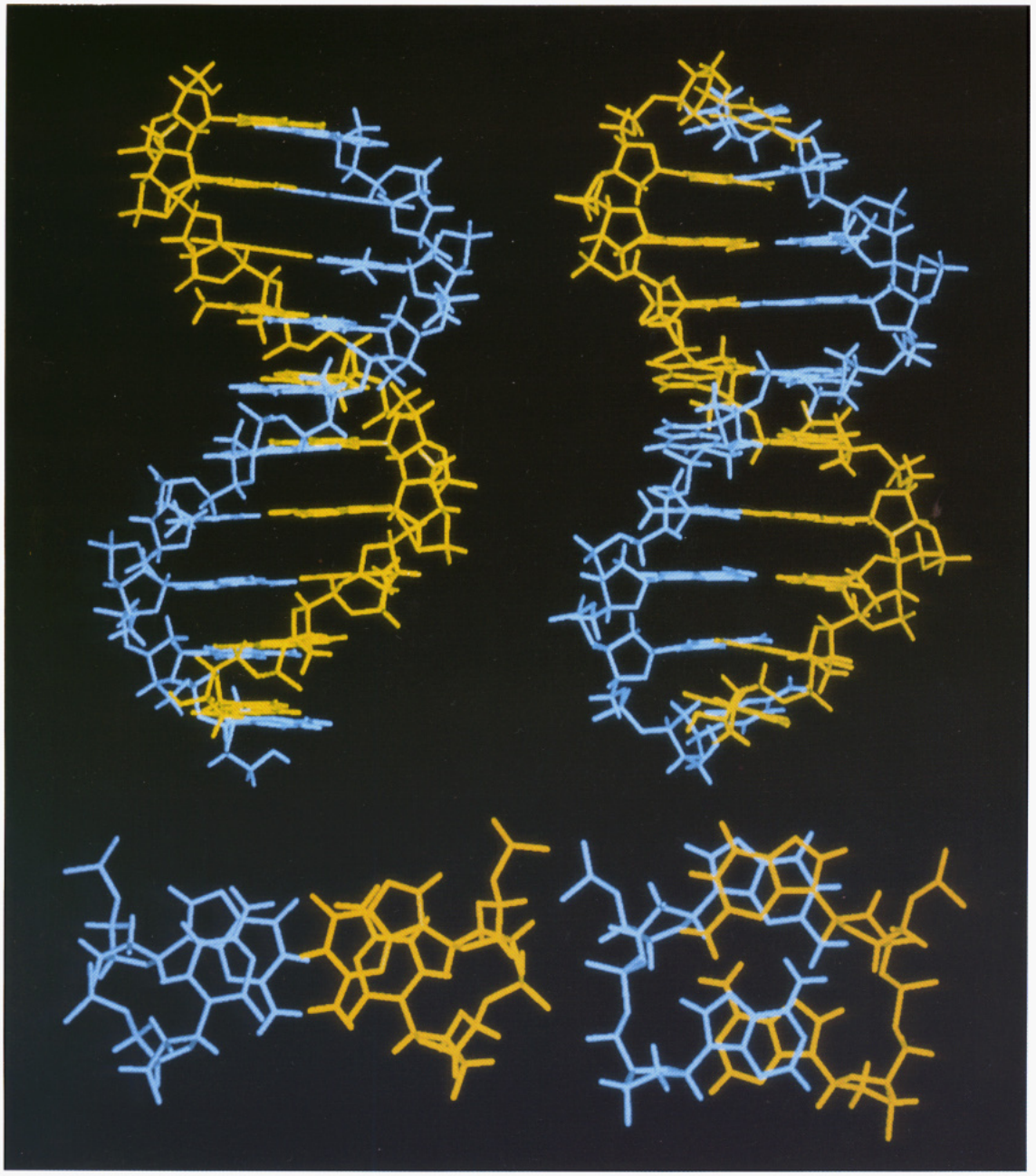

FIGURE 7: Panels A (top) and B (bottom) both show three-dimensional views of sequence **b**, d(CCATGAATGG)$_2$ (right), and a fully minimized B-form helix of similar sequence, d(CCATGCATGG)$_2$ (left). Panel A shows two full-length views of the sequences. The cross-strand stacking at the G5–A6 mismatch site of sequence **b** (panel A, right) is evident, while a comparison of the two sequences shows that sequence **b** still fits into an overall B-form helix. Panel B (right) shows the stacking pattern of the mismatched base pairs G5 and A6 of sequence **b** (panel B, right) and the Watson–Crick base pairs G5 and C6 of the B-form duplex (panel B, left). The interstrand stacking of sequence **b** is seen in this view, as well as the excellent stacking of the mismatched base pairs. Also note the unusual conformation of the phosphate backbone between the G5 and A6 bases of sequence **b**.

based on model 2 (type II) and model 1 (type I) with the experimental 2D NOESY spectrum shows conclusively that model 1 has excellent agreement with the experimental 2D NMR results, while the spectra of model 2 fail to agree with the experimental data for those NOEs unique to the GA mismatch site.

A complete structural characterization of the GA mismatch sequence, d(CCATGAATGG)$_2$, can be derived from model 1. The structure is, as a whole, a B-form helix, and the conformations of all the bases, including those of G5 and A6, are anti. However, a closer look at the backbone angles (Figure 6) in the oligonucleotide sequence clearly marks deviations from B-form helicity, most notably in the GA mismatch region. These plots unequivocally indicate that the four bases involved in the mismatch are in a unique conformation, but that the conformation is one that fits surprisingly well into the overall structure of a B-form helix.

The plots in Figure 6 are similar in that the backbone angles at G5 and A6 are dramatically different from the rest of the B-form backbone. The base pairs are quite overwound at the G5/A6 step with a helical twist value of $\sim$98°, while the flanking steps of A4/G5 and A6/A7 are subsequently underwound with a twist of only $\sim$16° (relative to 36° for a B-form duplex). This distortion yields the cross-strand stacking that characterizes this structure. The unusual base pairs in the structure have $\sim$25° propeller twist values, which are larger than those for the surrounding bases. Other unusual values at this site are those of the λ angles, the angle between the glycosidic bonds and the C1′–C1′ vectors. In a Watson–Crick base pair, these values are symmetric and are in the

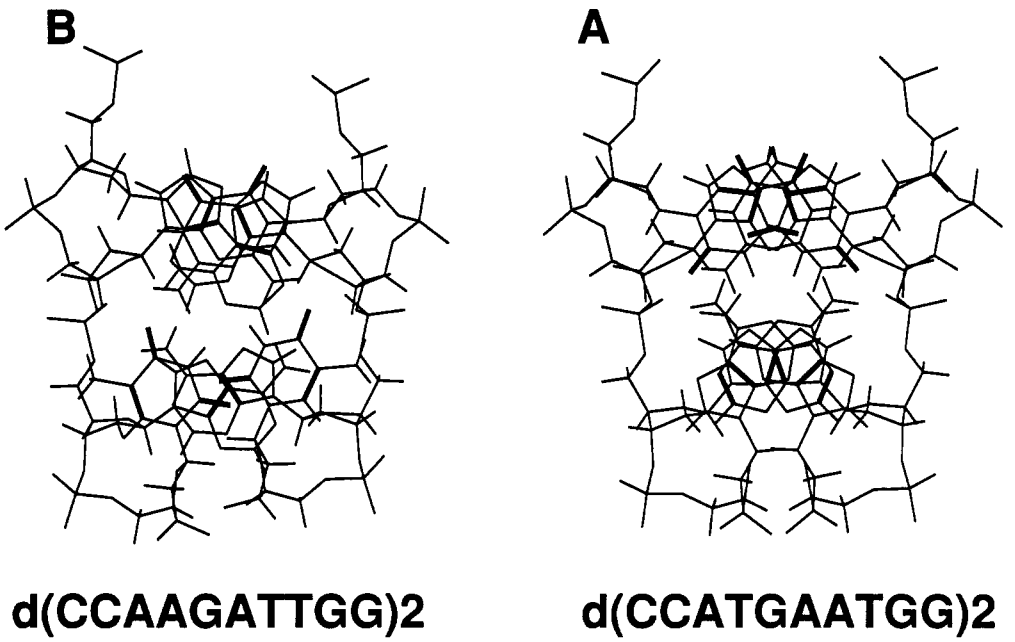

FIGURE 8: Views of the stacking interactions of sequence **b**, d(CCATGAATGG)$_2$ (A), and the mutated sequence, d(CCAAGATTGG)$_2$ (B), which was minimized in SYBYL. Only nucleotides 4–7 of each sequence are shown. The excellent stacking of sequence **b** is again evident in this view, while the view of the mutated sequence (B) shows that the stacking of the GA mismatched base pairs is far less compact.

range of 50–55° for AT and GC base pairs (Seeman, 1976). The λ values for the model 1 GA base pairs are highly asymmetric at approximately 98° and 6° and this again indicates an enormous offset in the base-pair stacking. Such deviations from symmetry are similar to those observed by Morden et al. (1992) in an oligomer with type I GA and AA mismatches.

The ε and ζ values for model 1 were measured to be ca. −155° and +155°, respectively, at the GA mismatch step, which would put the phosphate group conformation closer to ε(*t*),ζ(*t*) rather than the well-characterized B_I (t,g$^-$) or B_{II} (g$^-$,t) conformations. It should be noted that all of the models that were constructed with type I hydrogen bonds refine with their ε and ζ values in the ε(*t*),ζ(*t*) range. The large upfield and downfield shifts of the G5pA6 and T4pG5 phosphorus signals also indicate a deviation from the B_I backbone conformation. Li et al. (1991b) found that the phosphate at the GA site in sequence **a** is not in a B_I conformation, and Chou et al. (1992) have suggested that **a** is in a B_{II} conformation. It is clear from all of these results that the GA phosphate conformation is not B_I, but the DNA backbone conformation at phosphate groups is underdetermined by proton NMR methods. It should be emphasized, however, that the B_I and B_{II} conformations were discovered in fairly traditional duplexes with Watson–Crick base pairs, and alternative structures, especially when base-pair mismatches, base loops, or bulges are involved, are likely to be found.

The experimental NMR data presented above and depicted by model 1 all lead to the conclusion that the GA mismatched bases in sequence **b** (Table 1) undergo cross-stand stacking, i.e., the G5 of one strand stacks on the corresponding G5 of the second strand rather than on the A6 of its own chain as occurs in a B-form helix. The same interstrand stacking is true, of course, for the mismatched A6 bases. The ability of the type I GA mismatches to form cross-strand purine stacking that fits well into the overall B-form helix greatly enhances the stability of these structures. This unique interstrand purine stacking feature is dramatically illustrated in the views of Figure 7. Close examination of the GA mismatch site (Figure 7A) clearly shows the alignment of G5 and A6 stacked upon their cross-strand G5 and A6 partners; also delineated in this view is the manner in which a general B-form structure is retained for the flanking sequences. This is especially apparent when the full GA sequence is compared side by side to a similar sequence with Watson–Crick base pairs and a standard B-form helix (Figure 7A). The excellent nonsequential stacking of the GA mismatch bases can be appreciated from Figure 7B, where the two adenosines and the two guanosines are shown to fit completely over one another. The extension of the phosphate backbone between G5 and A6 that must occur to accommodate the interstrand stacking is visible in this view as well. Again, a visual comparison of the mismatch base-pairing with an analogous B-form structure (Figure 7B) illustrates the deviation of the backbone from the normal B-conformation. Just as this DNA sequence has unusual base stacking, it also has additional backbone conformations that do not correspond to the low-energy states of B_I and B_{II}, but it instead has moved into a conformation designed to accommodate the altered base stacking at the GA mismatch site.

The unusually effective stacking of the GA mismatched bases is apparently made possible by the flanking bases, as can be seen by comparison of the structures in Figure 8A, d(CCATGAATGG)$_2$, and Figure 8B, d(CCAAGATTGG)$_2$. The order of flanking nucleotides in sequence **b**, 5′-pyrimidine-GA-purine-3′, was changed to 5′-purine-GA-pyrimidine-3′, and the structure was minimized to assess sequence dependence on the formation of either type I or type II (Figure 1) base pairs in a model 1 structural context. As discussed above, the G and A nucleotides are fully stacked over one another in sequence **b**. However, when the identity of the flanking bases is changed to that of 5′-purine-GA-pyrimidine-3′, the purine base to the 5′ side prevents the complete, and therefore more energy efficient, stacking of the mismatched base pairs. Rather than the compact stacking that is seen in Figure 8A for the 5′-pyrimidine-GA-purine-3′ structure, the bulkier purine base interferes with the cross-strand stacking of the bases and causes the nucleotides to be stacked spirally rather than stacked atop one another. The decrease in stacking causes the energy of this conformation to be raised significantly. The

identity, then, of the base 5′ to the GA mismatch serves to act as a conformational switch, causing type I base pairing with a 5′-pyrimidine and shifting the conformation to type II hydrogen bonding when the 5′-base is a purine.

The deoxyoligonucleotide sequence, d(CCATGAATGG)$_2$, forms a type I GA mismatch base pair. In such a structure, the imino signal belonging to the guanine is unpaired and is therefore shifted upfield of the normal Watson–Crick base-paired imino signals, while the phosphorus linking the mismatched GA bases has a signal shifted downfield of normal phosphorus resonances. Type I hydrogen bonding results in a structure where the G5 and A6 stack not on their sequential neighbors, but on the G5 and A6, respectively, of the opposite strand. The cross-strand stacking region still fits easily within an overall B-form helix, although the phosphate backbone within the mismatch region deviates from the B_I conformation of a classic B-form DNA. The impact of the overwound helix within this region extends only to the bases on either side of the mismatch, thereby minimizing the effect of the distorted backbone and leaving intact the stability, as a whole, of the GA mismatch sequence.

SUPPLEMENTARY MATERIAL AVAILABLE

Figures showing the 1D NMR, NOESY, and COSY spectra of sequence **b** and a stereoview of model 1 (6 pages). Ordering information is given on any current masthead page.

REFERENCES

Aboul-ela, F., Koh, D., Martin, F. H., & Tinoco, I. (1985) *Nucleic Acids Res. 13*, 4811–4824.

Aue, W. P., Bartholdi, E., & Ernst, R. R. (1976) *J. Chem. Phys. 64*, 2229–2246.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977) *J. Mol. Biol. 112*, 535–542.

Borden, K. L. B., Jenkins, T. C., Skelly, J. V., Brown, T., & Lane, A. N. (1992) *Biochemistry 31*, 5411–5422.

Chandrasekaran, R., & Arnott, S., (1989) in *Landolt-Bornstein, Nucleic Acids, VII/1b* (Saenger, W., Ed.) pp 55, Springer-Verlag, Berlin,

Chary, K. V. R., Hosur, R. V., Govil, G., Zu-Kun, T., & Miles, H. T. (1987) *Biochemistry 26*, 1315–1322.

Cheng, J.-W., Chou, S.-H., & Reid, B. R. (1992) *J. Mol. Biol. 228*, 1037–1041.

Chou, S.-H., Cheng, J.-W., & Reid, B. R. (1992) *J. Mol. Biol. 228*, 138–155.

Davis, D. G., & Bax, A. (1985a) *J. Am. Chem. Soc. 107*, 2820.

Davis, D. G., & Bax, A. (1985b) *J. Am. Chem. Soc. 107*, 7197.

Ebel, S., Lane, A. N., & Brown, T. (1992) *Biochemistry 31*, 12083–12086.

Fasman, G. D., Ed. (1975) *Handbook of Biochemistry and Molecular Biology: Nucleic Acids*, 3rd ed., Vol. I, CRC Press, Cleveland, OH.

Feigon, J., Wright, J. M., Leupin, W., Denny, W. A., & Kearns, D. R. (1982) *J. Am. Chem. Soc. 104*, 5540–5541.

Gorenstein, D. G. (1984) in *Phosphorus-31 NMR: Principles and Applications* (Gorenstein, D. G., Ed.) Academic Press, New York.

Gorenstein, D. G., Schroeder, S. A., Fu, J. M., Metz, J. T., Roongta, V., & Jones, C. R. (1988) *Biochemistry 27*, 7223–7237.

Heus, H. A., & Pardi, A. (1991a) *Science 253*, 191–194.

Hore, P. J. (1983) *J. Magn. Reson. 55*, 283–300.

Jeener, J., Meier, B. H., Bachman, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546–4553.

Kan, L.-S., Chandrasegaran, S., Pulford, S. M., & Miller, P. S. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 4263–4265.

Lane, A., Martin, S. R., Ebel, S., & Brown, T. (1992) *Biochemistry 31*, 12087–12095.

Li, Y., Zon, G., & Wilson, W. D. (1991a) *Biochemistry 30*, 7566–7572.

Li, Y., Zon, G., & Wilson, W. D. (1991b) *Proc. Natl. Acad. Sci. U.S.A. 88*, 26–30.

Macura, S., & Ernst, R. R. (1980) *Mol. Phys. 41*, 95–110.

Maskos, K., Gunn, B. M., LeBlanc, D. A., & Morden, K. M. (1993) *Biochemistry 32*, 3583–3595.

Nagayama, K., Kumar, A., Wuthrich, K., & Ernst, R. R. (1979) *J. Magn. Reson. 40*, 321–334.

Neuhaus, D., & Williamson, M. (1989) in *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, VCH, New York.

Nikonowicz, E. P., & Gorenstein, D. G. (1990) *Biochemistry 29*, 8845–8858.

Orbons, L. P. M., Marel, G. A. v. d., Boom, J. H. v., & Altona, C. (1987) *Eur. J. Biochem. 170*, 225–239.

Patel, D. J., Kozlowski, S. A., Ikuta, S., & Itakura, K. (1984) *Biochemistry 23*, 3207–3217.

Prive, G. G., Heinemann, U., Chandrasegaran, S., Kan, L.-S., Kopka, M. L., & Dickerson, R. E. (1987) *Science 238*, 498–504.

Robinson, H., & Wang, A. H.-J. (1992) *Biochemistry 31*, 3524–3533.

Roongta, V. A., Jones, C. R., & Gorenstein, D. G. (1990) *Biochemistry 29*, 5245–5258.

Santa Lucia, J. J., Kierzek, R., & Turner, D. H. (1990) *Biochemistry 29*, 8813–8819.

Scheek, R. M., Russo, N., Boelens, R., & Kaptein, R. (1983) *J. Am. Chem. Soc. 105*, 2914–2916.

Seeman, N. C., Rosenberg, J. M., & Rich, A. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 804–808.

Sklenar, V., & Bax, A. (1987) *J. Am. Chem. Soc. 109*, 7525–7526.

Sklenar, V., Miyashiro, H., Zon, G., Miles, H. J., & Bax, A. (1986) *FEBS Lett. 208*, 94.

States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286–292.

Veal, J., & Wilson, W. D. (1991) *J. Biomol. Struct. Dyn. 8*, 1119–1145.

Wilson, W. D., Dotrong, M.-H., Zou, E. T., & Zon, G. (1988) *Nucleic Acids Res. 16*, 5137–5151.

Wuthrich, K. (1986) in *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, Inc., New York.