Structural Analysis of Z−Z DNA Junctions with A:A and T:T Mismatched Base Pairs by NMR†

Xiang-Lei Yang and Andrew H.-J. Wang*

Center for Biophysics & Computational Biology and Department of Cell & Structural Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

Received December 2, 1996; Revised Manuscript Received February 4, 1997⊗

ABSTRACT: Left-handed Z-DNA structure is favored by the alternating $(dC\text{-}dG)_n$ sequence. Many Z-potentiating sequences are found in genome and they often do not have a perfect alternating $(dC\text{-}dG)_n$ sequence. When a single base pair is removed from the alternating $(dC\text{-}dG)_n$ sequence, a Z−Z junction is created. A Z−Z junction is energetically less favorable by 3.5 kcal/mol than a perfect Z-DNA sequence. We designed special sequences to probe the structural perturbation at the Z−Z junction. Four DNA oligomers, d(*CG*CGT*CG*CG) and d(*CG*CGA*CG*CG) (*C = C or br^5C), have been synthesized and analyzed by NMR. The two br^5C-modified DNA nonamers are in the Z-DNA conformation in 50% methanol solution. The T and the A nucleotides are in the *anti* and *syn* conformation, respectively, in the two br^5C nonamers. The NOE-restrained molecular dynamics refinement demonstrated that T−T and A−A bases in the two Z-DNA duplexes are dynamic and adopt a range of conformations. Mixing the br^5C-d(*CG*CGT*CG*CG) and br^5C-d(*CG*CGA*CG*CG) nonamers together converts a fraction of the two nonamer populations into a hetero duplex as evident from the presence of the imino proton (at 13.70 ppm) of an A:T base pair. A model has been generated for the br^5C-d(*CG*CGT*CG*CG)+br^5C-d(*CG*CGA*CG*CG) duplex which incorporates a Z−Z junction. Previous biophysical and biochemical data associated with the Z−Z junction are discussed in the context of the present model.

The left-handed Z-DNA structure is favored by the alternating $(dC\text{-}dG)_n$ sequence in which the dC and dG nucleotides are in the *anti* and *syn* glycosyl conformations, respectively (Wang et al., 1979; Rich et al., 1984). Interestingly the deoxyribose of the dC and dG residues points in the opposite direction (along the helix) within the (dC-dG) dinucleotide repeating unit of Z-DNA. When the sequence deviates from the perfect alternating dC-dG sequence, the formation of Z-DNA from B-DNA becomes less energetically favorable. For example it is significantly more difficult to convert the alternating pyrimidine−purine (dC-dA):(dT-dG) sequence to Z-DNA. The energetic penalty for various "defects" of Z-DNA has been measured through the use of supercoiled DNA plasmids containing various alternating $(C\text{-}G)_n$ inserts or their variants (Rich et al., 1984; Sinden, 1994). Those defects include the (A,T)-containing sequence (Ellison et al., 1986), the out-of-alternation sequence (Ellison et al., 1985), and the Z−Z junction (Johnston et al., 1991).

The Z−Z junction is created when a single base pair is removed from a perfect alternating dC-dG sequence. For example, a sequence like ...CGCGCGGCGCGCG... contains one Z−Z junction at the interface between the two adjacent underlined guanines. The thermodynamic and biochemical properties of Z−Z junctions have been studied (Johnston et al., 1991). It was found that the energy penalty of a GG-type of Z−Z junction is 3.5 kcal/mol. In addition the base pair at the junction appeared to be hypersensitive toward hydroxylamine. A computer model of the Z−Z junction structure was proposed using the chemical probing data obtained in that study (Johnston et al., 1991).

Inspecting the genomic DNA sequences reveals that there exist many Z-potentiating sequences (Schroth et al., 1992), many of which in fact contain imperfect DNA sequences. Whether those Z-DNA sequences adopt Z-DNA conformation *in vivo*, or they play any biological functions, have yet to be determined. However its role in regulating DNA supercoiling has been well-documented (Rich et al., 1984; Sinden, 1994). A recent study by Rich and colleagues has shown that the chicken double-stranded RNA adenosine deaminase has a strong Z-DNA binding property (Herbert et al., 1995; Herbert & Rich, 1996). It has been proposed that since this enzyme is known to work near the transcription apparatus where a high negative supercoiling density along the DNA chain exists in front of the polymerase-acting site (Sinden, 1994), Z-DNA may play some role for this enzyme.

In this paper we address the questions of the structural property of the Z−Z junction. We have designed two Z-DNA sequences which incorporate potential Z−Z junctions and studied them by NMR. We take advantage of the observation that $(dC\text{-}dG)_n$ oligomers under high alcohol or high salt concentrations adopt Z-conformation (Pohl & Jovin, 1972; Rich et al., 1984; Feigon et al., 1984). Moreover some chemical modifications such as the C5-methylation or C5-bromination of cytosine (br^5C) (Behe & Felsenfeld, 1981; Feigon et al., 1985; Orbons et al., 1986) or C8-bromination of guanine (br^8G) (Moller et al., 1984) have been shown to stabilize the Z conformation in linear DNA oligomers. More recently we have shown that C8-methyl guanine modification

† This work was supported by NIH Grant GM-41612 to A.H.-J.W.

* To whom correspondence should be addressed at Department of Cell & Structural Biology, B107 Chemical and Life Sciences Laboratory, University of Illinois at Urbana−Champaign, 601 S. Goodwin Ave., Urbana, IL 61801.

⊗ Abstract published in *Advance ACS Abstracts,* April 1, 1997.

S0006-2960(96)02937-6 CCC: $14.00

is a powerful modification in inducing many DNA sequences into Z conformation under physiological condition (Sugiyama et al., 1996). The sequences of d(br^{5}CGbr^{5}CGTbr^{5}CGbr^{5}CG) (denoted as Z(T−T)) and d(br^{5}CGbr^{5}CGAbr^{5}CGbr^{5}CG) (denoted as Z(A−A)) were synthesized and they have been converted into Z-DNA in 50% methanol for structural studies. In addition, the non-brominated d(CGCGTCGCG) (denoted as B(T−T)) and d(CGCGACGCG) (denoted as B(A−A)) sequences have been studied in parallel as a control. Our NMR structural studies provide a plausible model for the Z−Z junction embedded in a Z-DNA helix.

EXPERIMENTAL SECTION

The four DNA nonamers were synthesized on a DNA synthesizer in the Genetic Facility at UIUC. The NMR solutions of the four oligomers were prepared using the established procedure (Robinson & Wang, 1992). The two B-DNA oligomers, B(A−A) and B(T−T), were dissolved in D_2O with 0.05 M phosphate buffer at pH 7.0 and 0.15 M NaCl, and with 0.05 M phosphate buffer at pH 7.0 and 0.05 M NaCl, respectively. The duplex concentrations were 1.84 mM for B(A−A) and 1.72 mM for B(T−T). The two Z-DNA nonamers Z(A−A) and Z(T−T) were both dissolved in 50% CD_3OD and 50% D_2O with 0.02 M phosphate buffer at pH 7.0 with the duplex concentrations of 1.06 and 1.63 mM, respectively. NMR spectra were collected on a Varian VXR500 500 MHz spectrometer and processed with FELIX v1.1 (Hare Research, Woodinville, WA) on Silicon Graphics IRIS workstations. T_1 relaxation experiments (T_1IR) were carried out with the standard 180°-t-90° inversion-recovery sequence, and the average T_1 relaxation time is about 2.0 s for the B-form oligomers and 2.8 s for the Z-form oligomers.

Phase-sensitive 2D NOESY experiments for nonexchangeable proton spectra were performed at 2 °C for the two B molecules and at 25 °C for the two Z molecules. The NOESY spectra were collected by the States/TPPI technique (States et al., 1982) with 512 t_1 increments and 2048 t_2 complex points each the average of 24 transients. The recycle delay was 4.41 s for the two B oligomers and 7.0 s for the two Z oligomers, and the mixing time was 100 ms. Apodization of the data in the t_1 and t_2 dimensions consisted of 4 Hz exponential multiplication with one half of a sine-squared function for the last one-fourth of the data to reduce truncation artifacts. In addition to 2D NOESY, standard TOCSY spectra were collected to aid in the assignment process. 1D- and 2D-NOESY spectra of the exchangeable protons were collected in 90% H_2O/10% D_2O solution for B oligomers and in 50% H_2O/50% CD_3OD solution for Z oligomers, using the 1-1 pulse sequence (Hore, 1983) as the read sequence, with a mixing time of 80 ms, each data point the average of 24 transients. Radiation damping was avoided using the method described by Sklenar et al. (1987). Deuterium lock was put on the HDO signal rather than the CD_3OH signal to avoid a receiver overflow when the 2D-NOESY spectra were collected in 50% H_2O/50% CD_3OD solution for the Z oligomers.

Starting models were constructed by MidasPlus (University of California at San Francisco). Structural refinement of these molecules were carried out by the procedure SPEDREF (Robinson & Wang, 1992). The experimental 2D NOESY data sets were analyzed using MYLOR (Robinson & Wang, 1992) to define the line shapes and the chemical shift for each spin in the f_1 and f_2 dimensions. These line shapes were then used to determine the volumes of the NOESY cross peaks. The correlation time τ_c was determined to be 14, 12, 4, and 4 ns for B(A−A), B(T−T), Z(A−A) and Z(T−T), respectively. The long τ_c for B-DNA is likely related to the tendency of short B-DNA duplexes to form concatemers by the end-over-end stacking in solution. The program X-PLOR (Brunger, 1993) was used for the molecular dynamics simulation and energy minimization. For the first 40 refinement cycles, the molecular models with NOE restraints were slowly annealed from 300 to 40 K and then subjected to conjugate gradient minimization with NOE-restraints. For the next 20 refinement cycles, the low temperature annealing step was deleted from the procedure, and the refinement was performed with only NOE-restrained conjugated gradient minimization. The NMR *R*-factor is defined as $R = \Sigma|N_o - N_c|/\Sigma N_o$, where N_o and N_c are the experimental and calculated NOE cross peak intensity, respectively.

RESULTS AND DISCUSSION

Conversion between B- and Z-DNA. The non-brominated d(CGCGTCGCG) and d(CGCGACGCG) sequences in solution readily form B-DNA duplexes, B(T−T) and B(A−A), with T:T and A:A mismatched base pairs, respectively. In contrast, the brominated d(br^{5}CGbr^{5}CGTbr^{5}CGbr^{5}CG) and d(br^{5}CGbr^{5}CGAbr^{5}CGbr^{5}CG) sequences could be converted to Z-DNA, Z(T−T) and Z(A−A), in 50% methanol as shown by the 1D NMR spectra in the aromatic proton region (Figure 1). The assignment of the resonances in the spectra was carried out using 2D-NOESY and 2D-TOCSY data as described later. It is of interest to note that in the B(A−A) sequence, the A5H8 and A5H2 resonances are significantly broader than the remaining resonances. In the Z(A−A) sequence the A5H8 and G4H8 resonances are broad. A similar trend is also observed for the G4H8 proton in the B(T−T) to Z(T−T) transition. Such observations may be related to the different motional property of A5 and G4 residues. In going from B-DNA to Z-DNA the purine aromatic resonances generally move upfield while the pyrimidine aromatic resonances move downfield. Since there is only one resonance detected for a given proton, the duplex is presumed to be symmetrical, i.e., the A−A and T−T mismatches did not generate an apparent asymmetry in the molecule. A probable explanation is that the two homo bases switch rapidly (on the NMR time scale) between different mismatched base pair configurations, resulting in an equivalence between the two bases (due to averaging).

The proof that those four nonamer DNA molecules form duplex structures could be demonstrated by the presence of the imino proton resonances (Figure 2). All guanine imino protons (even for the terminal G9 residue) were clearly detected in all four molecules, including B(A−A), B(T−T), Z(A−A) and Z(T−T), and assigned by the 2D-NOESY data collected in 90%H_2O/10%D_2O or 50%H_2O/50%CD_3OD (*vide infra*). For B(T−T) one T-imino proton resonance at 10.65 ppm could be detected, suggesting that the two Ts are base paired and they flip-flop rapidly between two T:T base pair configurations. When the two molecules, B(A−A) and B(T−T), are mixed together, a new duplex was immediately seen as evident by the disappearance of the T5H3 resonance at 10.65 ppm and the emergence of the new T5H3 resonance

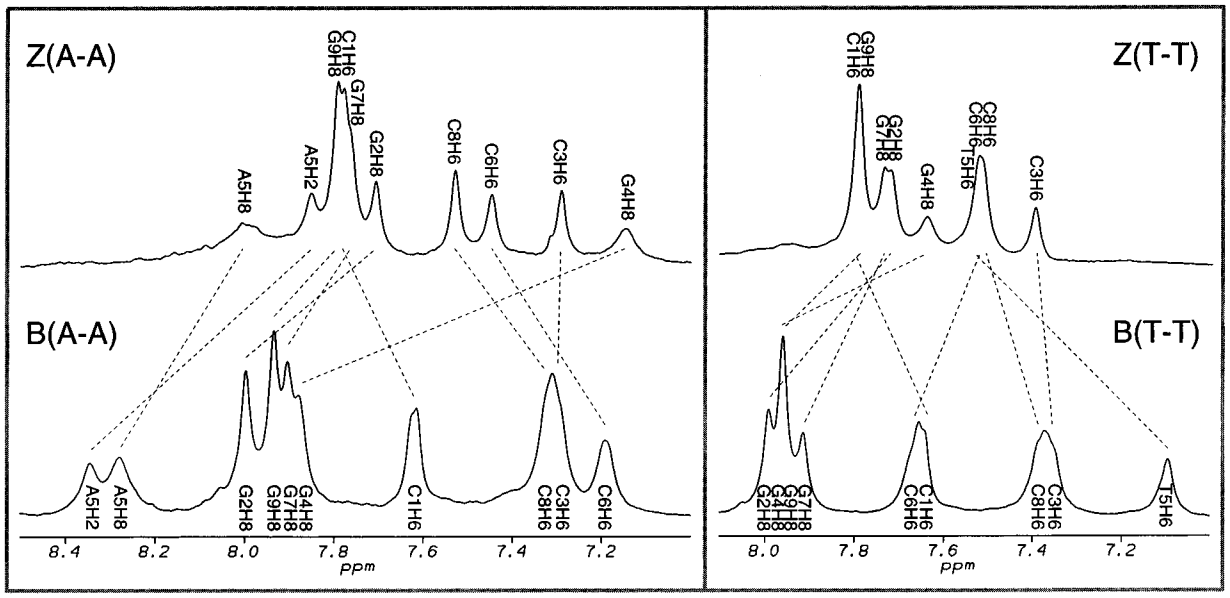

FIGURE 1: 1D NMR spectra of the aromatic proton regions of the two Z-DNA, Z(A−A) and Z(T−T), and two B-DNA, B(A−A) and B(T−T), molecules with A−A and T−T mismatch, respectively. The changes in the chemical shifts of the resonances associated with the B to Z transition are highlighted by the connecting dash lines. The spectra were collected in D_2O (B-DNA) or 50% D_2O/50% CD_3OD (Z-DNA) at 2 °C.

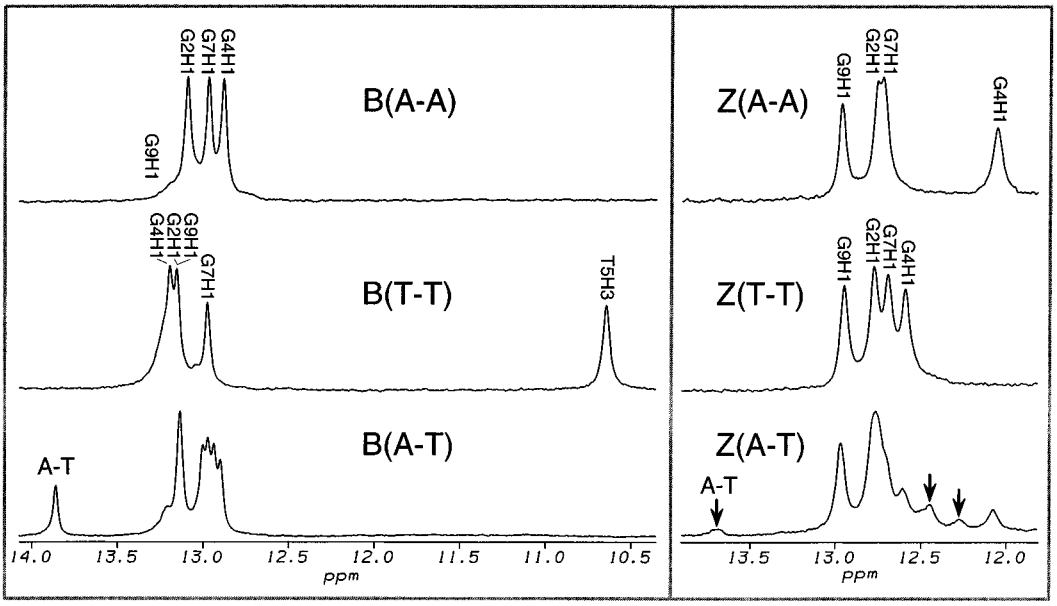

FIGURE 2: Exchangeable proton NMR spectra in H_2O (B-DNA) or 50% H_2O/50% CD_3OD (Z-DNA) at 2 °C. When the two Z-form molecules, Z(A−A) and Z(T−T), are mixed together, a new duplex with an A−T base pair is also formed as evident by the new T5H3 resonance at 13.70 ppm as well as other new resonances at 12.27 and 12.45 ppm (marked by arrows). However, the transition to the new helix is not complete since all resonances associated with the original Z(A−A) and Z(T−T) helixes remain.

at 13.80 ppm (which is consistent with an A−T Watson−Crick base pair).

For Z(A−A) the most upfield resonance is the G4H1 imino proton resonance. In Z(T−T) no T-imino proton resonance could be detected, suggesting that the two Ts are not base paired. It is remarkable to note that the imino proton resonance for the terminal G9 residue in both Z(A−A) and Z(T−T) duplexes shows up strongly. In fact both G9H1 resonances have a similar intensity to those of the remaining guanine H1 resonances. This is consistent with the conformational rigidity of Z-DNA (Kochoyan et al., 1990; Sugiyama et al., 1996) which restricts the motion (fraying) of the terminal G−C base pairs.

The cross peaks associated with the exchangeable protons in the 2D-NOESY spectra of Z(A−A) and Z(T−T) are again consistent with Z-DNA. For example, we note that the NH4a amino protons of the C1, C3, C6, and C8 cytosine bases have medium strong cross peaks to the H6 protons of the C8, C6, C3, and C1 residues, respectively. These corresponding cross peaks are not significantly observed in the spectra of B(A−A) and B(T−T). Such large cross peaks can only happen between the two interstrand cytosines in the CpG:CpG step of the Z-DNA due to its extreme sheared base pair stacking pattern. Similarly, the characteristic C1H1′−G2H1, C3H1′−G4H1, C6H1′−G7H1, and C8H1′−G9H1 cross peaks associated with Z-DNA are clearly detected in both Z(A−A) and Z(T−T) duplexes (data not shown). Our data unequivocally show that Z-DNA structure has been fully attained in both brominated molecules in 50% methanol solution.

Table 1: Chemical Shifts (ppm) for B(A−A) and B(T−T) at 2 °C

	H5/H2/Me	H6/H8	H1′	H2′/H2″	H3′	H4′	Hg′/H5″	H1/H3	H2a/4a[a]	H2b/4b[a]
C1	5.80	7.61	5.67	1.95/2.41	4.70	4.06	3.73/3.73		8.28	7.09
G2		7.99	5.87	2.71/2.71	4.99	4.36	4.07/3.99	13.09	na	na
C3	5.41	7.30	5.77	1.88/2.30	4.87	4.19	4.11/3.99		8.17	6.57
G4		7.88	5.77	2.54/2.54	4.93	4.22	4.07/3.95	12.88	na	na
A5	8.34	8.28	6.28	2.88/2.88	5.08	4.49	4.15/4.08		na	na
C6	5.22	7.19	5.54	1.82/2.24	4.80	4.13	4.31/4.14		8.34	6.57
G7		7.90	5.90	2.66/2.73	4.99	4.40	4.13/4.11	12.97	na	na
C8	5.44	7.31	5.67	1.89/2.31	4.82	4.13	4.19/4.10		8.49	6.73
G9		7.93	6.12	2.64/2.37	4.69	4.19	4.13/4.06	13.26	na	na
C1	5.85	7.64	5.71	2.00/2.43	4.71	4.07	3.73/3.73		8.23	7.14
G2		7.99	5.93	2.68/2.73	4.99	4.37	4.12/4.00	13.16	na	na
C3	5.39	7.36	5.64	2.07/2.43	4.86	4.22	4.08/4.08		8.39	6.59
G4		7.96	6.09	2.62/2.81	4.97	4.43	4.36/4.13	13.20	na	na
T5	1.37	7.10	5.91	1.81/2.32	4.81	4.13	3.99/4.00	10.65	na	na
C6	5.72	7.67	5.20	2.43/2.43	4.88	4.20	4.00/4.00		8.38	7.13
G7		7.91	5.94	2.61/2.73	5.01	4.36	4.13/4.13	12.98	na	na
C8	5.44	7.38	5.72	1.97/2.36	4.85	4.18	4.04/4.04		8.54	6.73
G9		7.96	6.15	2.65/2.37	4.69	4.20	4.08/4.08	13.16	na	na

[a] H2/4(a) are base-pair hydrogen bonded amino protons, H2/4(b) are not.

Table 2: Chemical Shifts (ppm) for Z(A−A) and Z(T−T) in 50% Methanol at 25 °C

	H2/Me	H6/H8	H1′	H2′/H2″	H3′	H4′	H5′/H5″	H1/H3	H2a/4a[a]	H2b/4b[a]
C1		7.78	5.73	1.53/2.38	4.55	3.66	2.48/3.16		8.95	6.97
G2		7.68	6.10	2.61/2.61	5.06	4.03	4.08/4.02	12.74	na	na
C3		7.31	5.38	1.60/2.58	4.67	3.58	2.37/3.70		8.99	6.31
G4		7.16	5.64	2.68/2.04	4.60	4.11	4.16/3.87	12.06	na	na
A5	7.85	8.01	6.01	2.64/2.19	5.00	4.11	4.05/4.02		na	na
C6		7.46	5.52	1.61/2.57	4.78	3.88	2.64/3.72		8.56	6.42
G7		7.74	6.16	2.81/2.68	5.02	4.07	4.12/4.12	12.70	na	na
C8		7.54	5.79	1.72/2.65	4.87	3.94	2.66/3.90		9.15	6.48
G9		7.77	6.20	3.14/2.33	4.69	4.12	4.34/4.03	12.93	na	na
C1		7.79	5.78	1.54/2.37	4.56	3.69	2.47/3.17		8.99	6.98
G2		7.70	6.12	2.66/2.66	5.07	4.05	4.09/4.05	12.76	na	na
C3		7.41	5.56	1.65/2.63	4.79	3.80	2.53/3.84		9.17	6.22
G4		7.66	6.10	3.35/2.46	4.85	4.30	4.02/3.95	12.59	na	na
T5	1.69	7.58	6.18	2.18/2.21	4.83	4.09	4.05/4.11	na	na	na
C6		7.53	5.70	1.69/2.46	4.68	3.87	2.52/3.65		8.93	6.30
G7		7.71	6.14	2.80/2.66	5.06	4.06	4.10/4.05	12.67	na	na
C8		7.52	5.81	1.73/2.66	4.88	3.95	2.66/3.91		9.17	6.45
G9		7.77	6.20	3.14/2.34	4.70	4.12	4.33/4.04	12.92	na	na

[a] H2/4(a) are base-pair hydrogen bonded amino protons, H2/4(b) are not.

When the two molecules, Z(A−A) and Z(T−T), are mixed together, a new duplex with an A−T base pair was formed as evident by the new T5H3 resonance at 13.70 ppm as well as other new resonances at 12.27 and 12.45 ppm (marked by arrows in Figure 2). However, the transition to the new helix is not complete since all resonances associated with the original Z(A−A) and Z(T−T) helixes remain, albeit with diminished intensities. To ensure that the Z(A−A) and Z(T−T) helixes were not trapped in their local energy minima, the mixed solution was heated to 70 °C and then slowly cooled back down to 2 °C. The same spectrum was obtained, indicating that an equilibrium had been reached with the three different helixes, Z(A−A), Z(T−T) and Z(A−T), coexisting and in slow exchange.

NMR Refinement of Z-DNA. In order to obtain more definitive structural information of the Z−Z junction, NOE-restrained refinement has been performed on the Z(A−A) and Z(T−T) duplexes. In addition, parallel refinements were also carried out on their B-DNA counterparts, B(A−A) and B(T−T). 2D-NOESY and TOCSY in D_2O were used to assign the resonances of all nonexchangeable protons. The chemical shifts of all resonances are tabulated in Tables 1 and 2.

For B(A−A) and B(T−T), the usual sequential assignment procedure was used (data not shown). Inspection of the NOE data suggested that both A−A and T−T bases are intrahelical, not unlike other A−A and T−T mismatched base pair found in other molecules (Lian et al., 1996; Gervais et al., 1995; Stolarski et al., 1987; Arnold et al., 1987). Two B-DNA duplexes incorporating an A−A and a T−T mismatched base pair, respectively, were constructed using MidasPlus (University of California at San Francisco), and these models were subjected to a combined SPEDREF (Robinson & Wang, 1992) and NOE-constrained refinement (Brünger, 1993). We measured 942 and 881 NOE integrals, respectively, as the input for the NOE-restrained refinement. The refined structures (not shown), with NMR *R*-factor of 16.8% and 17.8%, respectively, are not unusual compared to other B-DNA containing A−A and T−T mismatched base pair studied previously (Lian et al., 1996; Gervais et al., 1995; Stolarski et al., 1987; Arnold et al., 1987). The refinement statistics of the four structures are listed in the Table 3.

For the Z(A−A) and Z(T−T) duplexes, the usual sequential assignment procedure would not be applicable. For example, the aromatic-H1′ cross peak region of the 2D-NOESY spectrum of Z(A−A) and Z(T−T) (Figures 3 and

Table 3: NMR Refinement Statistics for Z(A−A), Z(T−T), B(A−A), and B(T−T)

	Z(A−A)	Z(T−T)	B(A−A)	B(T−T)
number of NOE restraints	593	452	942	881
NMR *R* factor ($S\|N_o - Nc\|/SN_c$), %	19.9	18.6	16.8	17.8
structure statistics (rmsd)				
NOE distance deviations (pairwise distance deviations between refined structure and NOE-derived values), Å	0.303	0.237	0.294	0.323
bond distance deviations from ideal values, Å	0.012	0.009	0.012	0.014
bond angle deviations from ideal values, deg	3.5	3.3	3.3	3.7
temperature of data collection (°C)	25	25	2	2

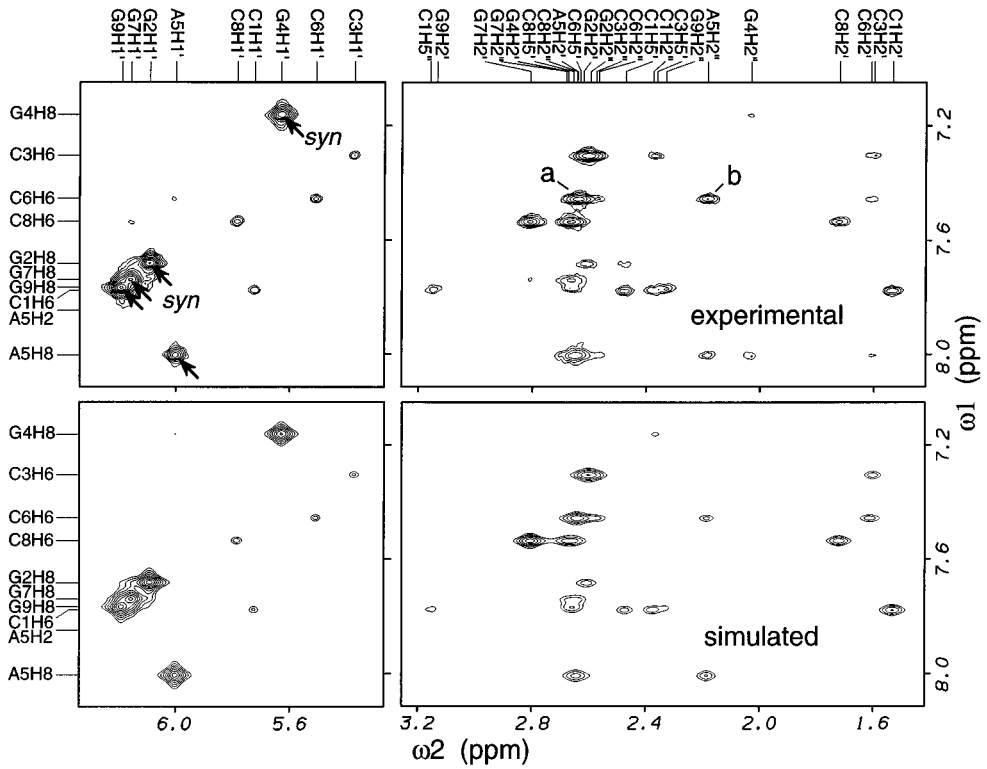

FIGURE 3: Experimental and simulated 2D-NOESY spectra of the aromatic to H1′/H2′/H2″ region of Z(A−A). The strong H1′−H8 cross peaks associated with the *syn* glycosyl conformations of all purine nucleotides including the adenosine in the A−A mismatch are marked by arrows. Cross peaks between A5H2′/H2″ and C6H6 are labeled a and b, which indicate that the H2′/H2″s of A are pointing toward the 3′-direction.

4) showed only strong intranucleotide GH1′−GH8 and A5H1′−A5H8 cross peaks, indicative of the *syn* conformation of guanine and adenine residues and the *anti* conformation of cytosine and thymine residues. As has been noted before (Feigon et al., 1984, 1985; Sugiyama et al., 1996), there is no internucleotide connectivity in Z-DNA, in contrast to that in the right-handed B-DNA. The assignment was subsequently extended to the aromatic-H2′/H2″ region and finally to all regions of the spectrum. The TOCSY data supported the assignment (data not shown). Note that all deoxycytidine H2′ and H5′ resonances in Z-DNA are unusually upfield (∼1.5−1.7 ppm and ∼2.4−2.7 ppm, respectively), analogous to those seen before (Feigon et al., 1984, 1985; Sugiyama et al., 1996). The upfield shifts are due to the orientation of the sugar moiety of dC nucleotide in Z-DNA which places the H2′ and H5′ protons directly under the ring current of the neighboring 5′- and 3′-dG guanine bases, respectively. Note that in Z-DNA for the pyrimidine nucleotide H2′ resonances are more upfield than H2″, but for the purine nucleotide H2′ resonances are generally more downfield than H2″ (see Table 2).

The important internucleotide NOE cross peaks for Z(A−A) are summarized in Figure 5. Of particular interest are those associated with the A5 nucleotide. It was noted that A5 is in the *syn* conformation although the A5H8−A5H1′ cross peak intensity appears somewhat weaker than those from the G nucleotides. Furthermore there is a quite large A5H8−A5H3′ NOE cross peak (data not shown) which is not usually observed for a *syn* nucleotide since the distance between the two protons is ∼5 Å. One possible explanation is that the A5 nucleotide at the Z−Z junction may flip-flop between the *syn* and *anti* conformations, although the *syn* conformation predominates. The A5 nucleotide has extensive cross peaks both to G4 and C6 bases (Figure 5), suggesting that A5 is likely to be intrahelical and sandwiched between the G4 and C6 nucleotides. An important piece of information is that the H2′/H2″ of A5 have strong cross peaks to H2′/H2″/H6 of C6, indicating that the deoxyriboses of those two nucleotides are oriented with their O4′ pointing in the opposite direction.

We constructed a model of the Z(A−A) duplex and subjected it to a combined SPEDREF (Robinson & Wang, 1992) and NOE-constrained refinement (Brünger, 1993). We measured 593 NOE integrals as the input for the NOE-restrained refinement. The refined structure, which has an NMR *R*-factor of 19.9%, is shown in Figure 6A. The

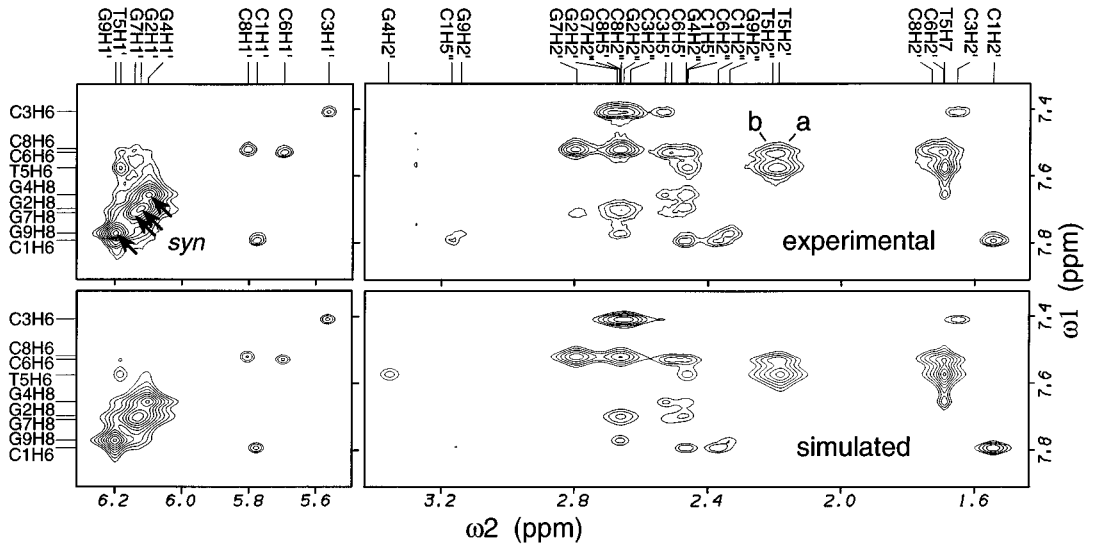

FIGURE 4: Experimental and simulated 2D-NOESY spectra of the aromatic to H1′/H2′/H2′ region of Z(T−T). Again, the strong H1′−H8 cross peaks associated with the *syn* glycosyl conformations of all purine nucleotides are marked by arrows. Cross peaks between T5H2′/H2′ and C6H6 are labeled a and b, which indicate that the H2′/H2″s of T nucleotide are pointing toward the 3′-direction.

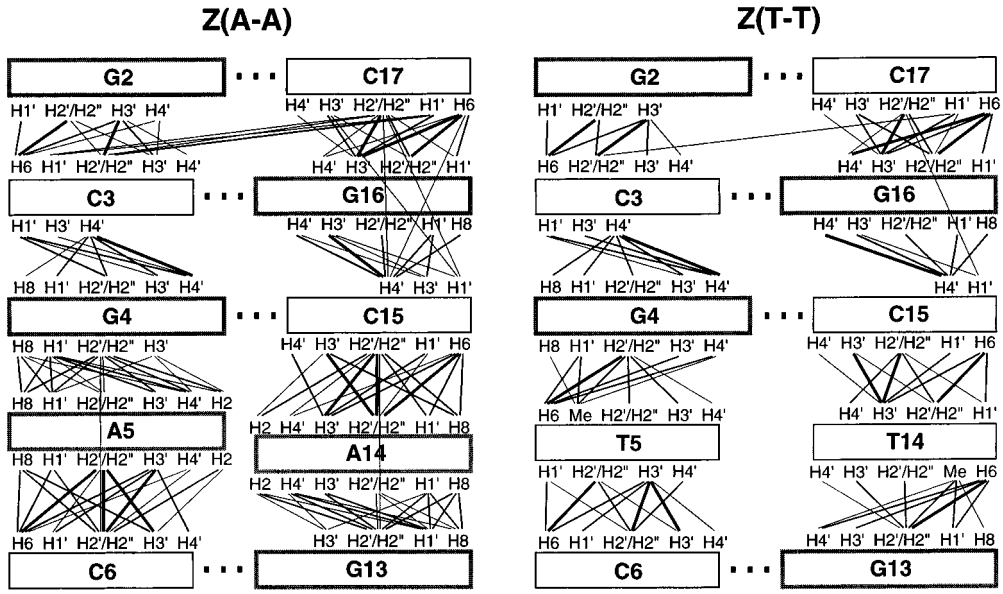

FIGURE 5: Schematic diagram showing important internucleotide cross peaks observed in the experimental 2D-NOESY spectra of Z(A−A) (left panel) and Z(T−T) (right panel). The nucleotides are depicted as rectangular boxes, with the *syn* purine nucleotides shown as boxes with dark borders. The dotted lines between bases indicate that the two bases are paired with their imino protons observed by NMR. The NOE intensities are represented by straight lines of varying thickness, denoting strong, medium, and weak NOE cross peaks.

simulated NOESY spectrum is in good agreement with the observed spectrum (Figure 3). The two equivalent A5 bases are slightly non-coplanar and no apparent hydrogen bond is found between them. In the NOE-refined structure the A5 base is stacked on the G4 base on the one side and at the same time lies above the C2′ of the C6 residue on the other side. The CGCG segments at both ends are quite similar to the $d(CGCGCG)_2$ Z-DNA structure determined by X-ray crystallography (Wang et al., 1979). The root mean square deviation between the two Z-structures, one from Z(A−A) and the other from the crystal structure, is 2.14 Å.

The important internucleotide NOE cross peaks for Z(T−T) are summarized in Figure 5. Unlike the B(T−T) duplex in which a clear base-paired T-imino proton resonance was detected (at 10.65 ppm, see Figure 2), no such resonance could be seen for the Z(T−T) duplex. The NOE cross peak pattern associated with the T5 residue is somewhat abnormal. The methyl group of T5 has many cross peaks to the G4 nucleotide, while the deoxyribose is oriented so that the H2′/H2″ are pointing toward the C6 residue. Since the T5 is clearly in the *anti* conformation (evident by the weak intranucleotide H6−H1′ cross peak), it is difficult to place the T5 nucleotide in a regular intrahelical arrangement with the T5 base stacked between G4 and C6 bases. Indeed this observation was borne out by the molecular dynamics simulated annealing refinement.

We constructed a model of the Z(T−T) duplex with the initial position of the two T5 nucleotides in the stacked conformation and subjected it to a combined SPEDREF (Robinson & Wang, 1992) and NOE-constrained simulated annealing refinement (Brünger, 1993). We measured 452 NOE integrals as the input for the NOE-restrained refinement. It should be noted that there are significantly fewer NOE cross peaks between the T5 nucleotide and the neighboring G4 and C6 nucleotides when compared with the number of A5 to G4 and C6 in Z(A−A) or the number

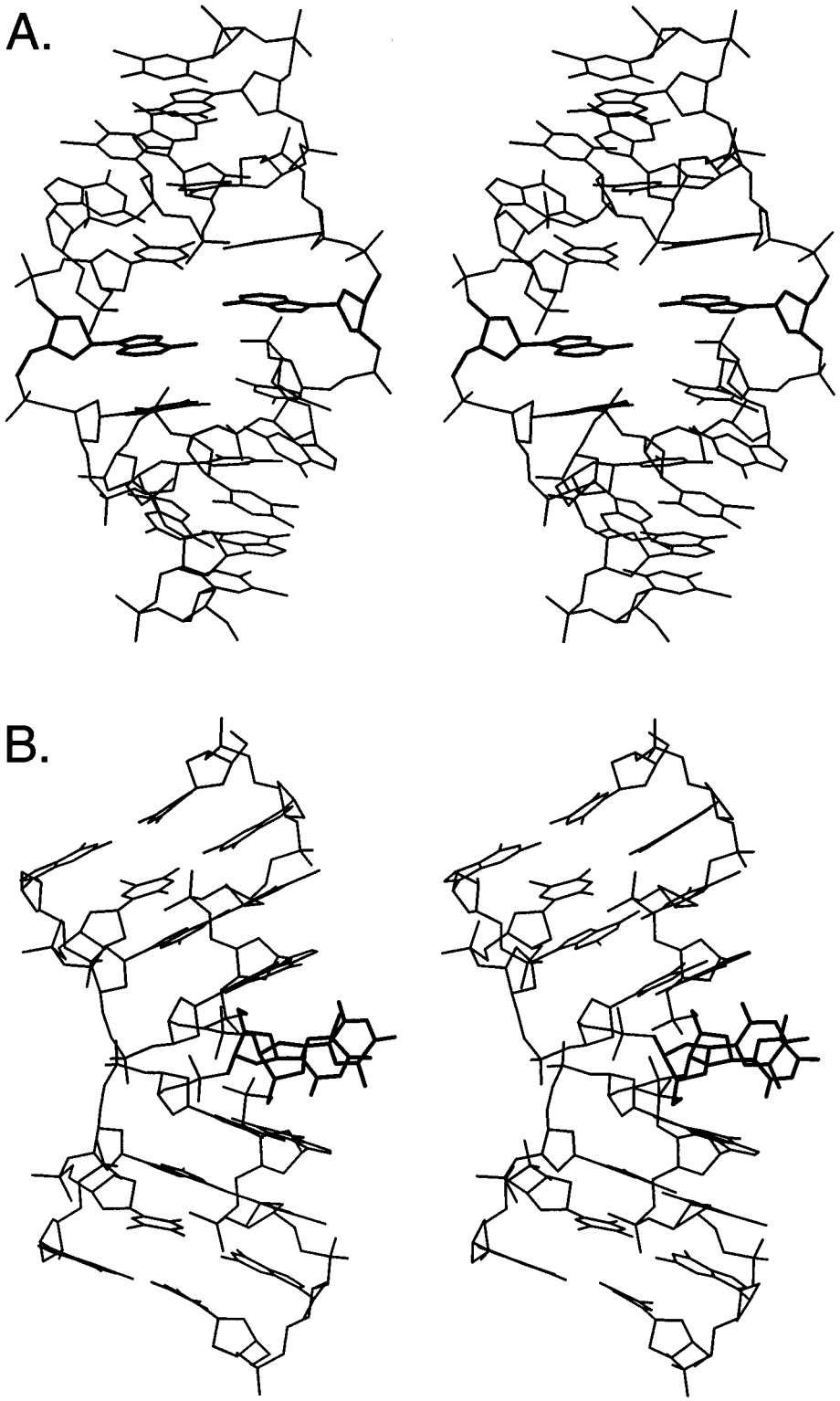

FIGURE 6: Stereoscopic drawings of the refined structures of (A) Z(A−A) and (B) Z(T−T) duplexes. A−A and T−T mismatches are highlighted by thick lines.

of T5 to G4 and C6 in B(T−T). This may suggest that T5 is not making as many contacts to its neighbors. The refined structure, which has an NMR *R*-factor of 18.6%, is shown in Figure 6B. The simulated NOESY spectrum is again in good agreement with the observed spectrum (Figure 4). It is interesting to note that the final conformation of the two equivalent T5 bases always end up in an extrahelical position. The extrusion of the T5 nucleotide is toward the major groove direction. The chemical shifts of T5H6 (7.58 ppm) and T5Me (1.69 ppm) in Z(T−T) are significantly more downfield than the corresponding protons (7.10 ppm and 1.37 ppm, respectively) in B(T−T). This is also consistent with the model of extrahelical T bases in Z(T−T) duplex since the extrahelical T bases no longer are under the ring current influence, therefore causing their protons to be more downfield shifted.

In fact the extrusion of the central T5:T5 nucleotides leaves a gap between the two equivalent G4:C6 base pairs, and the two four-base pair Z-DNA segments have a varying kink angle of up to 30°. The two equivalent G4:C6 base pairs

are not directly stacked over each other, supported by the observation that there is no NOE cross peaks between them (Figure 5). However, the two $(CGCG)_2$ segments in the duplex still adopt an intact Z-DNA structures despite the opening between the two segments resulting from the two unpaired T:T bases. The root mean square deviation between the two Z-structures, one from Z(T−T) and the other from the crystal structure, is 1.56 Å. The intact nature of the Z segments is reflected in the observation that even though the G4:C6 base pair in this model is partially exposed to solvent on one side, the G4H1 imino proton is still readily observable, similar to other G imino protons (Figure 2).

The dynamic properties associated with the A:A and T:T base pairs in the Z(A−A) and Z(T−T) duplexes are reflected in the T_1 relaxation inversion recovery time (T_1IR) measurments. Z-DNA is more rigid than B-DNA (Kochoyan et al., 1990; Sugiyama et al., 1996). The averaged T_1IR of 3.0 s and 2.8 s for the C/G nucleotides in Z(A:A) and Z(T:T), and 1.9 s and 2.1 s for B(A:A) and B(T:T), respectively, support this notion. Interestingly the A and T nucleotides in the Z(A:A) and Z(T:T) duplexes have lower respective T_1IR values of 2.7 s and 2.4 s, indicating that the mismatched base pair region is more flexible than the remaining C/G nucleotides.

A Model for the Z−Z Junction. The analyses of the two Z-DNA structures with the A−A and T−T bases located at the junction provide us with a starting point to construct a Z−Z junction with a normal A−T base pair. Our data showed that the A nucleotide is in the *syn* conformation with the H2′/H2″ pointing toward the 3′-direction, whereas the T nucleotide is in the *anti* conformation with the H2′/H2″ also pointing toward the 3′-direction. Furthermore when the two molecules, Z(A−A) and Z(T−T), are mixed together, a clear T5H3 imino proton resonance from an A−T base pair was detected (Figure 2). This suggests that a fully base-paired Z helix, Z(A−T), was formed, although such a helix apparently has a similar stability as the mismatched Z(A−A) and Z(T−T) duplexes. Using those available structural information, a plausible arrangement for the Z−Z junction is schematically summarized in Figure 7.

An important question in the Z−Z junction model is the type of base pair between A and T. We have constructed a model in which the A−T base pair is in the reverse Watson−Crick base pair conformation (Figure 8A). Such a base pair satisfies all available structural information. Other attempts to use either the normal Watson−Crick, the Hoogsteen, or the reverse Hoogsteen base pairs always resulted in implausible models. The model with the reversed Watson−Crick base pair has been energy-minimized and depicted in Figure 8A. Note that a reversed Watson−Crick G−C base pair can be substituted for the A−T base pair without an extensive conformational rearrangement. The three consecutive base pairs at the junction, i.e., G4-A5-C6, show excellent stacking interactions. In order to maintain a good stacking pattern, the three base pairs adopt unusual backbone conformation so that they have very small helical twist angles at those two steps (i.e., G4pA5 and A5pC6). Certain phosphate groups (e.g., those from T14 and C15) become close to each other which may destabilize the structure due to charge repulsion.

The model presented above is different from that has been proposed by Johnston et al. (1991) in which the A(*syn*) and T(*anti*) bases form a Watson−Crick base pair. We have built

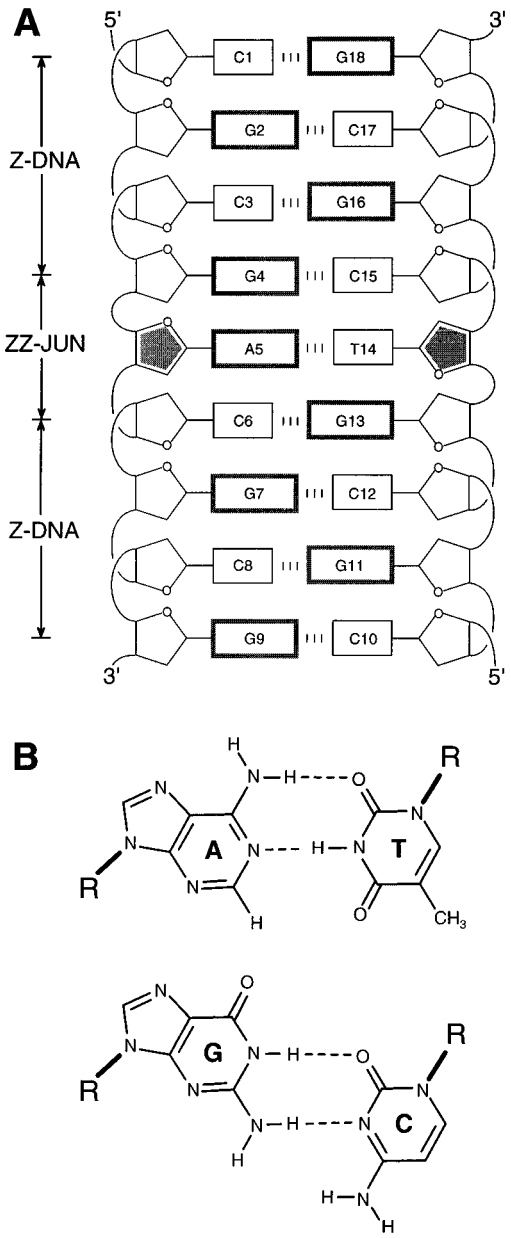

FIGURE 7: (A) Schematic diagram showing the structural characteristics of the Z−Z junction. The purine nucleotides (large rectangular boxes) are in the *syn* conformation, and the pyrimidine nucleotides (small rectangular boxes) are in the *anti* conformation. The polarity of the deoxyribose ring for the A5 nucleotide is the same as that of the G4 nucleotide, but the polarity of the deoxyribose ring for the T14 is opposite to that of the C15 nucleotide. (B) The reverse Watson−Crick A−T and G−C base pairs.

a similar model (Figure 8B) in order to compare with the model incorporating a reverse-Watson−Crick A−T base pair (Figure 8A). The principal difference is that in the latter model the A5 nucleotide has its H2′/H2″ pointing in the 3′-direction, whereas in the model of Johnston et al. the A5 nucleotide has its H2′/H2″ pointing in the 5′-direction.

However it should be pointed out that the actual Z−Z junction, while showing a compact structure, may be a dynamic, instead of the static, structure shown in Figure 8A. It is likely that the A:T base pair can transiently open up, causing the T nucleotide to become extrahelical, exposing it for chemical reaction. This scenario is fully consistent with the chemical probing data by Johnston et al. (1991), which showed that the pyrimidine nucleotides become

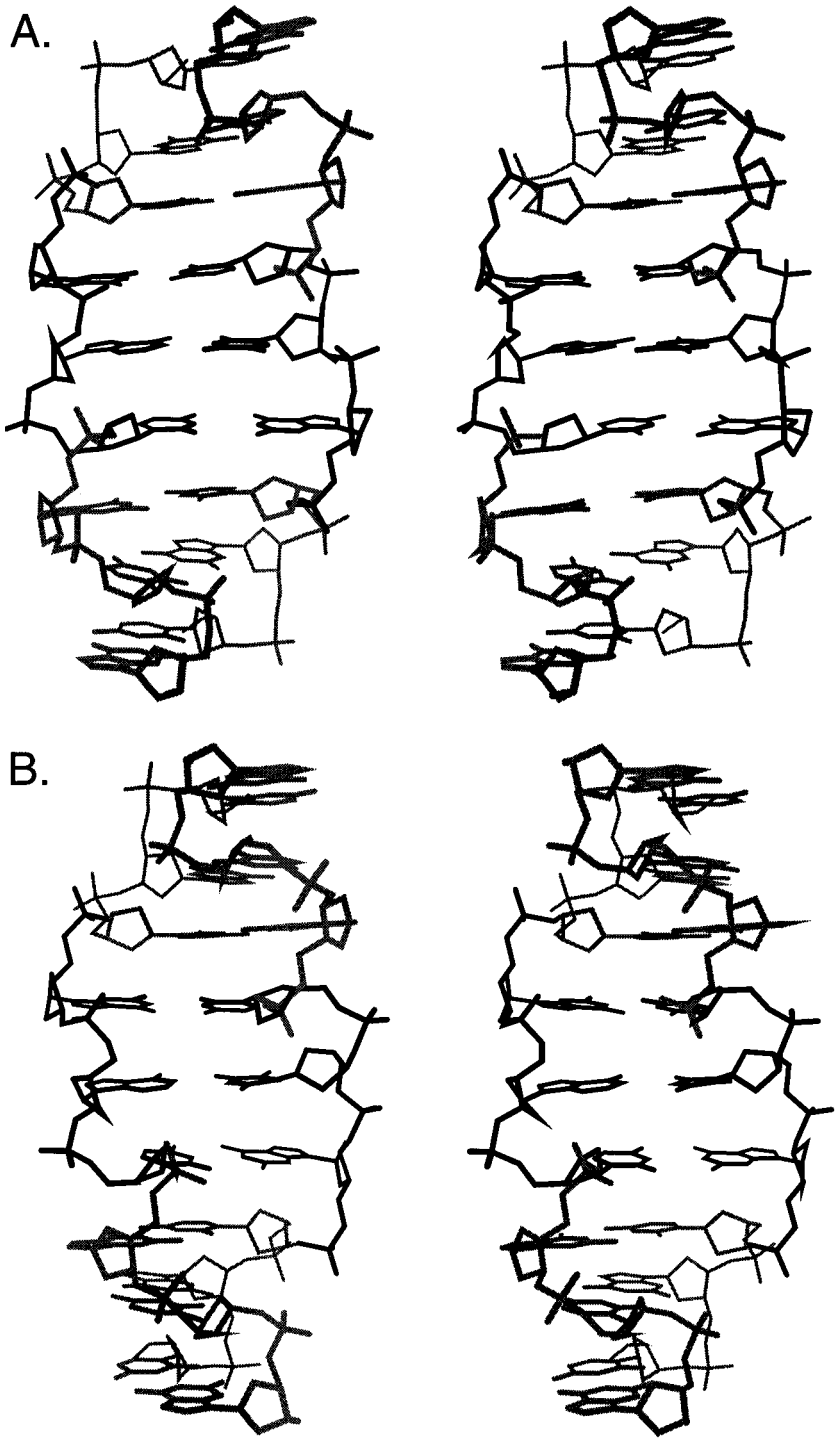

FIGURE 8: (A) Stereoscopic drawing of the proposed Z−Z junction model constructed using MidasPlus and energy-minimized by conjugated gradient minimization. The A−T base pair at the junction is in the reverse Watson−Crick base pair conformation. The DNA backbone conformation at the G4pA5 step has a right-handed helical twist which causes the two $(CGCG)_2$ Z-DNA segments having a very small helix twist angle. (B) An alternative model similar to that of Johnston et al. (1991) in which A−T Watson−Crick base pair is used.

hypersensitive at the Z−Z junction toward the reaction of hydroxylamine and osmium tetraoxide. The dynamic property of the Z−Z junction provides a mechanism for the pyrimidine base to be thrust into the groove, exposing them toward those chemical agents.

CONCLUSION

In this work we have used two Z-DNA-forming molecules, Z(A−A) and Z(T−T), to obtain critical information regarding the conformation of A and T at the junction of two Z-DNA segments. A plausible model has been proposed and presented which explains the previous biochemical data satisfactorily. Since different Z-DNA sequences, including the Z−Z junction, are found *in vivo* (Schroth et al., 1992), it is important to obtain a more complete understanding of the structure, dynamics, and energetics of Z-DNA related motifs. Recently we have shown that the substitution of a methyl group at the guanine C8 position dramatically stabilizes the Z conformation of short oligonucleotides of a variety of base sequences (Sugiyama et al., 1996). Some of those m^8G-modified oligomers exist as a stable Z form at physiological salt conditions without added organic solvent

or divalent metal. In the future we should be able to design unique sequences using m^8G-modified DNA oligomers (e.g., C*GC*GACACG + C*GT*GTC*GCG where *G = m^8G) such that a complete conversion to Z-conformation with a single Z−Z junction is possible. Such a system may provide a better opportunity for the quantitative structural analysis of a Z−Z junction.

REFERENCES

Arnold, F. H., Wolk, S., Cruz, P., & Tinoco, I. Jr. (1987) *Biochemistry 26*, 4068−4075.

Behe, M., & Felsenfeld, G. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 1619−1623.

Brünger, A. X-PLOR (version 3.1) (1992), The Howard Hughes Medical Institute and Yale University, New Haven, CT.

Ellison, M. J., Feigon, J., Kelleher, R. J., Wang, A. H.-J., Habener, J. F., & Rich, A. (1986) *Biochemistry 25*, 3648−3655.

Ellison, M. J., Kelleher, R. J., Wang, A. H.-J., Habener, J. F., & Rich, A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 8320−8324.

Feigon, J., Wang, A. H.-J., van der Marel, G. A., van Boom, J. H., & Rich, A. (1984) *Nucleic Acids Res. 12*, 1243−1263.

Feigon, J., Wang, A. H.-J., van der Marel, G. A., van Boom, J. H., & Rich, A. (1985) *Science 230*, 82−84.

Gervais, V., Cognet, J. A., Le Bret M., Sowers, L. C., & Fazakerley G. V. (1995) *Eur. J. Biochem. 228*, 279−290.

Herbert, A., Lowenhaupt, K., Spitzer, J., & Rich, A. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 7550−7554.

Herbert, A., & Rich, A. (1996) *J. Biol. Chem. 271*, 11595−11598.

Hore, P. J. (1983) *J. Magn. Reson. 54*, 539−542.

Johnston, B. H., Quigley, G. J., Ellison, M. J., & Rich, A. (1991) *Biochemistry 30*, 5257−5263.

Kochoyan, M., Leroy, J. L., & Gueron, M. (1990) *Biochemistry 29*, 4799−4805.

Lian, C., Robinson, H. and Wang, A. H.-J. (1996) *J. Am. Chem. Soc. 118*, 8791−8801.

Moller, A., Nordheim, A., Kozlowski, S. A., Patel, D. J., & Rich, A. (1984) *Biochemistry 23*, 54−62.

Orbons, L. P., van der Marel, G. A., van Boom, J. H., & Altona, C. (1986) *Eur. J. Biochem. 160*, 131−139.

Pohl, F. M., & Jovin, T. M. (1972) *J. Mol. Biol. 67*, 375−396.

Rich, A., Nordheim, A., & Wang, A. H.-J. (1984) *Annu. Rev. Biochem. 53*, 791−846.

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

Schroth, G. P., Chou, P.-J., & Ho, P. S. (1992) *J. Biol. Chem. 267*, 11846−11855.

Sinden, R. R. (1994) *DNA Structure and Function*, Academic Press: New York.

Sklenar, V., Brooks, B. R., Zon, G., & Bax, A. (1987) *J. Magn. Reson. 75*, 352−357.

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

Stolarski, R., Buck, F., Fera, B., & Ruterjans, H. (1987) *Eur. J. Biochem. 169*, 603−609.

Sugiyama, H., Kawai, K., Matsunaga, A., Fujimoto, K., Saito, I., Rosinson, H., & Wang, A. H.-J. (1996) *Nucleic Acids Res. 24*, 1272−1278.

Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G. A., & Rich, A. (1979) *Nature 282*, 680−682.

BI962937B