Site-Resolved Energetics in DNA Triple Helices Containing G·TA and T·CG Triads†

Daniel Coman and Irina M. Russu*

Department of Chemistry and Molecular Biophysics Program, Wesleyan University, Middletown, Connecticut 06459

Received September 24, 2001

ABSTRACT: Recognition of specific sites in double-helical DNA by triplex-forming oligonucleotides has been limited until recently to sites containing homopurine−homopyrimidine sequences. G·TA and T·CG triads, in which TA and CG base pairs are specifically recognized by guanine or by thymine, have now extended this recognition code to DNA target sites of mixed base sequences. In the present work, we have obtained a characterization of the stabilities of G·TA and T·CG triads, and of the effects of these triads upon canonical triads, in triple-helical DNA. The three DNA triplexes investigated are formed by the folding of the 31-mers d(GAAXAGGT$_5$CCTYTTCT$_5$CTTZTCC) with X = G, T, or C, Y = C, A, or G, and Z = C, G, or T. We have measured the exchange rates of imino protons in each triad of the three triplexes using nuclear magnetic resonance spectroscopy. The exchange rates are used to map the local free energy of structural stabilization in each triplex. The results indicate that the stability of Watson−Crick base pairs in the G·TA and T·CG triads is comparable to that of Watson−Crick base pairs in canonical triads. The presence of G·TA and T·CG triads, however, destabilizes neighboring canonical triads, two or three positions removed from the G·TA/T·CG site. Moreover, the long-range destabilizing effects induced by the T·CG triad are larger than those induced by the G·TA triad. These findings reveal the molecular basis for the lower overall stability of G·TA- and T·CG-containing triplexes.

Recognition of specific base sequences in double-helical DNA can be achieved by oligonucleotide-directed formation of triple-helical structures. Extensive experimental evidence has shown that formation of the triple-helical structure at the targeted sites inhibits the sequence-specific binding of proteins to DNA (*1*−*4*) and specifically alters gene expression (*5*−*8*). Moreover, triplex-forming oligonucleotides have been coupled with DNA cleaving agents, thus allowing site-specific cleavage of genomic DNA (*9*). Generally, the targeted sites consist of homopurine−homopyrimidine sequences. The bases in the third strand bind to the homopurine strand of the DNA double helix through specific Hoogsteen hydrogen bonds. Two families of structures exist depending upon the base composition and the orientation of the third strand (*10*). In pyrimidine−purine−pyrimidine (YRY)[1] triplexes the third strand consists of protonated cytosines (C^+) and thymines (T) and binds parallel to the purine strand of the double helix. The C^+ and T bases recognize respectively GC and AT base pairs by forming C^+·GC and T·AT triads. In purine−purine−pyrimidine (RRY) triplexes, the third strand is purine-rich and binds antiparallel to the purine strand in the DNA double helix. In this case, the GC and AT base pairs are recognized through the formation of A·AT, T·AT and G·GC triads. The practical applications of these two families of triple helices have been limited thus far by the fact that, in these systems, only GC and AT base pairs can be recognized by triplex formation. This limitation has prompted sustained efforts to expand the triplex recognition code to the other two Watson−Crick base pairs in duplex DNA (*11*, *12*). The two most promising base triad combinations identified thus far are the novel triads G·TA and T·CG. In the first triad, a guanine in the third strand recognizes a TA base pair in duplex DNA (*13*, *14*). In the second, a thymine in the third strand recognizes a CG base pair (*15*). Thus, these two triads open a route to achieving recognition of random base sequences in DNA through oligonucleotide-directed formation of a triple helix. However, at present, a limitation in this approach is that triple helices containing G·TA and T·CG triads are less stable than canonical YRY triple helices. This destabilization depends on the nature of the triad, being greater for triple helices containing T·CG triads than for those containing G·TA triads (*12*, *14*, *16*, *17*). New insights into the molecular origin of these destabilization effects have been gained from the three-dimensional structures of several G·TA- and T·CG-containing triple helices solved in solution state by nuclear magnetic resonance (NMR) spectroscopy (*18*−*22*). The structures have revealed that, in G·TA and T·CG triads, the base in the third strand is anchored to the duplex by only one hydrogen bond (Figure 1). In contrast, in canonical YRY triple helices, the third strand base forms two Hoogsteen hydrogen bonds to each Watson−Crick base pair. The structures have also shown that the presence of G·TA or T·CG triads induces large variations in the helical parameters. These structural variations distort the duplex part of the structure and allow novel

† Supported by a grant from the National Science Foundation (MCB-9723694).

* To whom correspondence should be addressed: Phone: (860) 685-2428. Fax: (860) 685-2211. E-mail: irussu@wesleyan.edu.

[1] Abbreviations: NMR, nuclear magnetic resonance spectroscopy; ppm, parts per million, A, adenine; G, guanine; C, cytosine; T, thymine; Y, either pyrimidine; R, either purine.

10.1021/bi011832s CCC: $22.00
Published on Web 03/06/2002

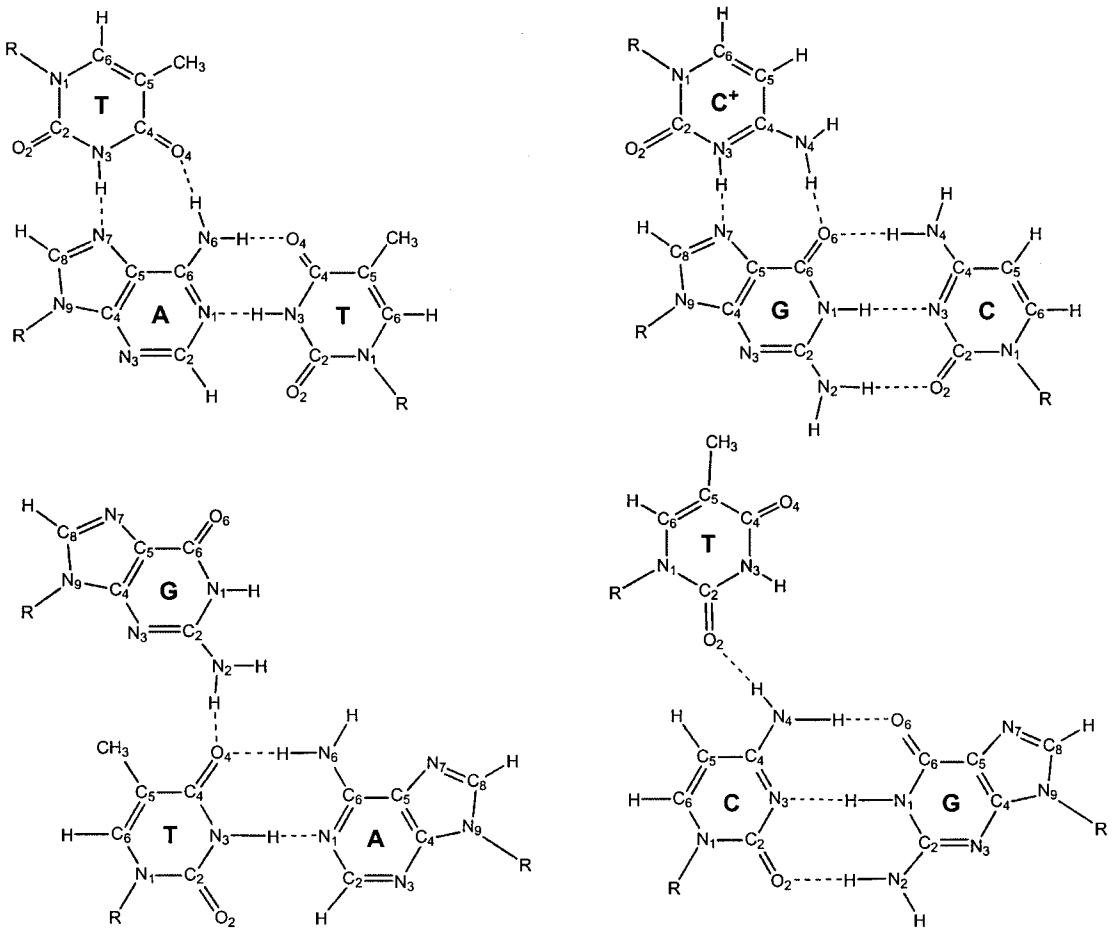

FIGURE 1: Structures of the canonical T·AT and C^+·GC triads (*10*) and of the novel G·TA and T·CG triads (*18*−*20*).

stacking interactions between the base in the third strand and neighboring bases (*18*−*20*).

Complete elucidation of the molecular basis of the stability of G·TA- and T·CG-containing triple helices requires understanding how these perturbations in conformation affect the stability of each triad in these structures. One experimental approach that is uniquely suited for obtaining this information is proton exchange (*23*). This method has been extensively used to define the energetic effects of structural changes in proteins, induced by single amino acid substitutions or by allosteric transitions (*24*). We have also used proton exchange to characterize the energetic effects of structural modifications in DNA double helices, for example, base pair mismatches (*25*). In the present work, we have extended these previous investigations by a study of two DNA triple helices containing G·TA and T·CG triads, using NMR spectroscopy and proton exchange. The DNA triple helices investigated are shown in Figure 2. One triple helix (henceforth abbreviated G·TA triplex) incorporates a G·TA triad into an otherwise YRY triplex. The solution structure of this triple helix has been solved by NMR methods by Patel and co-workers (*19*). The second triple helix (henceforth abbreviated T·CG triplex) contains a T·CG triad, and its solution structure has also been solved by Patel and co-workers (*20*). Except for the replaced triad, the two triple helices are identical. Thus, they permit a comparison of the energetic effects of G·TA and T·CG triads within the same base sequence context. As a reference for the structural energetics in these two triple helices, we have also constructed and studied a homologous triplex in which a canonical C^+·GC triad is present at the same site as the G·TA and T·CG triads in the other two triplexes (henceforth abbreviated C^+·GC triplex).

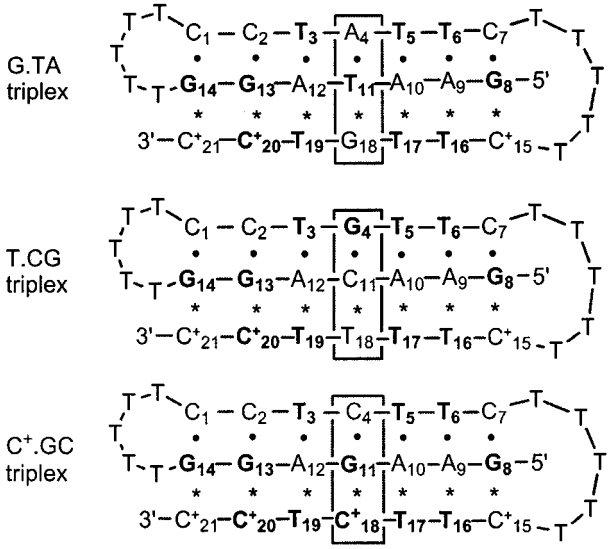

FIGURE 2: Base sequences and folded conformation of the DNA triplexes investigated. Watson−Crick hydrogen bonds are indicated by dots, and Hoogsteen hydrogen bonds are indicated by asterisks. The imino proton resonances for the bases shown in bold are observable in the NMR spectra. The numbering of the bases is the same as that used in the determination of the structures of the G·TA and T·CG triplexes.

MATERIALS AND METHODS

DNA Samples. The DNA oligonucleotides were synthesized on an automated DNA synthesizer (Applied Biosystems 381A) using phosphonate chemistry for the G·TA and T·CG triplexes and phosphoramidite chemistry for the C^+·GC triplex. They were purified by reverse-phase HPLC on a PRP-1 column (Hamilton) in 50 mM triethylammonium acetate buffer at pH 7 (with a gradient of 5−32% acetonitrile in 39 min). The counterions were replaced with Na^+ ions by repeated centrifugation through Centricon YM-3 tubes (Amicon Inc.). The final samples were in 100 mM NaCl and 5 mM $MgCl_2$ at pH 4.6 (measured at 5 °C). The samples contained 240 OD_{260} units for the G·TA triplex, 140 OD_{260} units for the T·CG triplex, and 230 OD_{260} units for C^+·GC triplex.

NMR Experiments. The NMR experiments were performed at 5 °C on a Varian INOVA 500 spectrometer operating at 11.75 T. One-dimensional (1D) NMR spectra were obtained using the jump-and-return pulse sequence (*26*). For the measurements of proton exchange rates, we have used two methods: transfer of magnetization from water and real-time exchange. The choice of the method was dictated by the values of the exchange rates to be measured; namely, transfer of magnetization experiments were used to measure exchange rates higher than 0.3 s^{-1}, and real-time exchange experiments were used to measure exchange rates lower than 4×10^{-3} s^{-1}.

In transfer of magnetization experiments, the exchange was initiated by inverting selectively the water proton resonance using a Gaussian 180° pulse (5.7−5.9 ms). A weak gradient (0.21 G/cm) was applied during the exchange delay following water inversion to prevent the effects of radiation damping upon the recovery of water magnetization to equilibrium. At the end of the exchange delay, a second Gaussian pulse (1.7−1.8 ms) was applied to bring the water magnetization back to the oz axis. The observation was with the jump-and-return pulse sequence. Twenty-two values of the exchange delay in the range from 2 to 600 ms were used in each experiment. The relaxation delay between successive scans was 8 s. The longitudinal relaxation rate of water (0.5−0.6 s^{-1}) was measured in separate experiments. The exchange rates were calculated from the dependence of the intensity of the proton resonance of interest on the exchange delay as we have previously described (*27*−*29*). Due to effects of the longitudinal relaxation of the protons of interest and of water protons, the lowest exchange rate that can be measured reliably using the transfer of magnetization method is ∼0.3 s^{-1} at 5 °C.

In real-time exchange experiments, the DNA samples in water were dried down to 60 ± 30 μL by flushing with argon. The exchange was initiated by adding 540 μL of D_2O such that the final volume fraction of D_2O was between 86% and 95%. The time elapsed between the initiation of exchange and the first NMR spectrum varied between 4 and 6 min. A total of 96 transients were accumulated for each spectrum with a total acquisition time of 3.85 min per spectrum. The intensity of each resonance was fitted as a function of the exchange delay t to the equation:

$$I(t) = [I(0) - I(\infty)] \exp(-k_{ex}t) + I(\infty) \quad (1)$$

where $I(\infty)$ is the intensity in a fully exchanged sample. Due to the time elapsed between the initiation of the exchange and the first NMR spectrum, the fastest exchange rate that could be measured accurately in these experiments is 4×10^{-3} s^{-1}.

The 2D NOESY experiments were carried out for the C^+·GC triplex using the water flip-back WATERGATE NOESY pulse sequence (*30*). A total of 512 increments were used in the second dimension at a spectral resolution of 5 Hz/point. The mixing time was 100 ms.

Theory of Imino Proton Exchange. The exchange of imino protons with solvent protons in nucleic acids occurs by opening of the base pairs. During opening, the hydrogen bond holding the imino proton breaks, and the proton is moved into a state where it is accessible to proton acceptors. Accordingly, the rate of exchange of a given proton is (*31*)

$$k_{ex} = \frac{k_{op}k_{ex,open}}{k_{cl} + k_{ex,open}} \quad (2)$$

where k_{op} is the rate of opening of the base pair, k_{cl} is the rate of closing, and $k_{ex,open}$ is the rate of proton exchange from the open state. The equilibrium constant $K_{op} = k_{op}/k_{cl}$ is related to the free energy change for the opening reaction by (*23*)

$$\Delta G_{op} = -RT \ln K_{op} \quad (3)$$

In the open state of the base pair, the exchange of the imino proton occurs by two mechanisms. In one, the exchange is catalyzed by external proton acceptors present in solution, e.g., water, OH^-, and buffers. In the other, the proton acceptors are the nitrogens in the other base of the open base pair; for example, for an open Watson−Crick AT base pair, the acceptor of the N3H proton of thymine is the N1 group in adenine. Accordingly, the rate of exchange from the open state is (*23*, *32*)

$$k_{ex,open} = k_A[\text{acceptor}] + k^{int}_{ex,open} \quad (4)$$

where k_A is the rate constant for proton transfer to an external acceptor and $k^{int}_{ex,open}$ is the rate of internal catalysis. The efficiency of proton transfer to external and internal acceptors is determined by the fraction of productive transfers, F, in the transient hydrogen-bonded complex between the imino group NH and the acceptor. The fraction F depends on the pK values of the imino group and of the acceptor:

$$F = (1 + 10^{-\Delta pK})^{-1} \quad (5)$$

with $\Delta pK = pK(\text{acceptor}) - pK(\text{NH})$.

Two regimes for proton exchange in nucleic acids have been observed (*23*, *33*): (i) In the EX1 regime, when $k_{ex,open} \gg k_{cl}$:

$$k_{ex} = k_{op} \quad (6)$$

and the exchange is rate-limited by the opening of the base pair. (ii) In the EX2 regime, when $k_{ex,open} \ll k_{cl}$:

$$k_{ex} = K_{op}k_{ex,open} \quad (7)$$

and the measured exchange rate is proportional to the

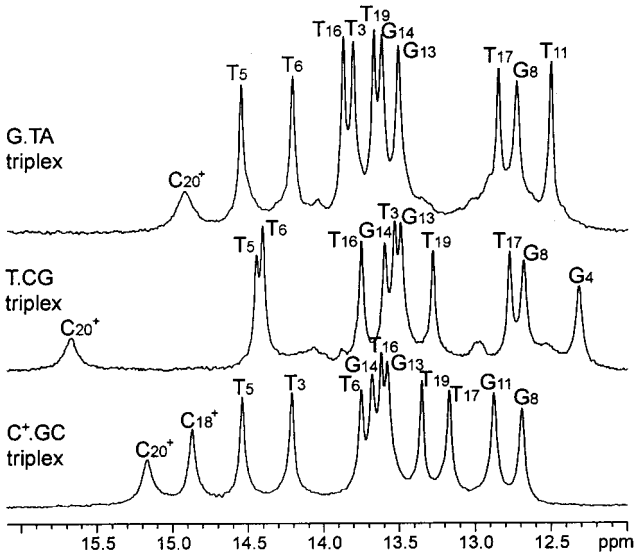

FIGURE 3: NMR resonances of imino protons in the DNA triplexes investigated in 100 mM NaCl and 5 mM $MgCl_2$ at pH 4.6 and at 5 °C. The resonance assignments for the G·TA and T·CG triplexes are from Patel and co-workers (*19*, *20*).

equilibrium constant for opening of each base pair and to the rate of proton transfer from the open state.

RESULTS

Solution Structures and Assignments of Imino Proton Resonances in the DNA Triple Helices Investigated. The solution structures of the G·TA and T·CG triplexes have been solved by Patel and co-workers using NMR spectroscopy (*19*, *20*). Both triplexes belong to the YRY family and contain, in addition to the G·TA or the T·CG triad, three canonical C^+·GC and three canonical T·AT triads (Figure 2). The imino proton resonances of these two triplexes, and their assignments, are shown in Figure 3. The imino proton resonances from the Hoogsteen bases of the G·TA and T·CG triads (i.e., G_{18} and T_{18}, respectively) occur at 11.0 and 11.1 ppm, respectively, upfield from the spectral region shown in Figure 3, where they overlap with resonances of the imino protons of thymines in the two loops. The resonances of imino protons from the terminal C_{21}^+ and C_{15}^+ bases are not observed. For these protons the exchange with solvent is very fast, and their resonances are broadened beyond detection.

For the C^+·GC triplex we have characterized its solution structure using 2D ^{1}H NOESY experiments. Formation of the triple-helical structure was indicated by the appearance of imino proton resonances from Hoogsteen base pairs and by the characteristic downfield shift of amino proton resonances of protonated cytosines. The imino proton resonances of the C^+·GC triplex are shown in Figure 3. Expanded regions of the ^{1}H−^{1}H NOESY spectra for the same triplex are shown in Figure 4. Sequential connectivities between imino protons are observed for the central bases in the Hoogsteen strand (i.e., T_{19}-C_{18}^+-T_{17}-T_{16}) and in the duplex part of the structure (i.e., G_{13}-T_3-G_{11}-T_5-T_6-G_8). For the latter, the connectivity between imino protons in G_{13} and G_{14} is not resolved because the chemical shifts of these resonances are close to each other (Figure 3). Formation of the C^+·GC triplex was confirmed using all of the other NOESY signatures of triple-helical structures (*34*). For example, as shown in Figure 4B, the imino protons in Hoogsteen T·A

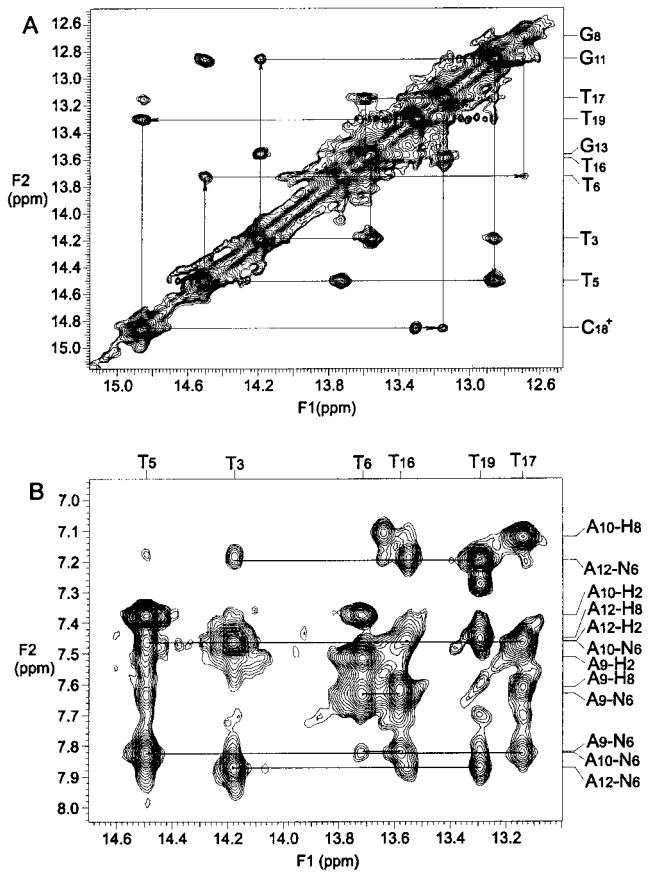

FIGURE 4: Expanded regions of the ^{1}H−^{1}H NOESY spectrum of the C^+·GC triplex in 90% H_2O/10% D_2O at pH 4.6 and at 5 °C. (A) Sequential connectivities between imino protons in Watson−Crick base pairs (G_{13}-T_3-G_{11}-T_5-T_6-G_8) and in Hoogsteen base pairs (T_{19}-C_{18}^+-T_{17}-T_{16}). (B) Connectivities between imino and amino protons and between imino and aromatic protons in T_{19}·$A_{12}T_3$, T_{17}·$A_{10}T_5$, and T_{16}·A_9T_6 triads. The horizontal lines indicate thymine imino protons that have connectivities to the same adenine amino protons.

base pairs have NOESY connectivities to A-C8H protons in the same T·AT triad, namely, T_{17}-A_{10}, T_{16}-A_9, and T_{19}-A_{12}. Moreover, the thymine imino protons in Watson−Crick and Hoogsteen base pairs also show connectivities to both amino protons of the adenine in the same triad, i.e., T_3 and T_{19} to A_{12}, T_5 and T_{17} to A_{10}, and T_6 and T_{16} to A_9 (Figure 4B). These results, and the connectivities between exchangeable and nonexchangeable protons (not shown), confirm that the C^+·GC oligonucleotide folds in solution into a YRY triple-helical structure.

Proton Exchange in the DNA Triple Helices Investigated. We have measured the rates of exchange with solvent for the imino protons in the three triple-helical structures. Representative examples from proton exchange measurements are shown in Figure 5. The upper panel in the figure illustrates results from transfer of magnetization experiments by showing the dependence of the intensity of the imino proton resonance in G_8 on the exchange delay. For this proton the exchange rates are, within experimental errors, the same in the three triple helices. The lower panel in the figure illustrates results from real-time exchange experiments. After initiation of the exchange by addition of D_2O to the sample, the intensity of the imino proton resonance from T_{17} decreases with time. The exchange in the canonical C^+·GC triplex is slower than that in the G·TA triplex. For the same proton in the T·CG triplex the exchange is too fast to be

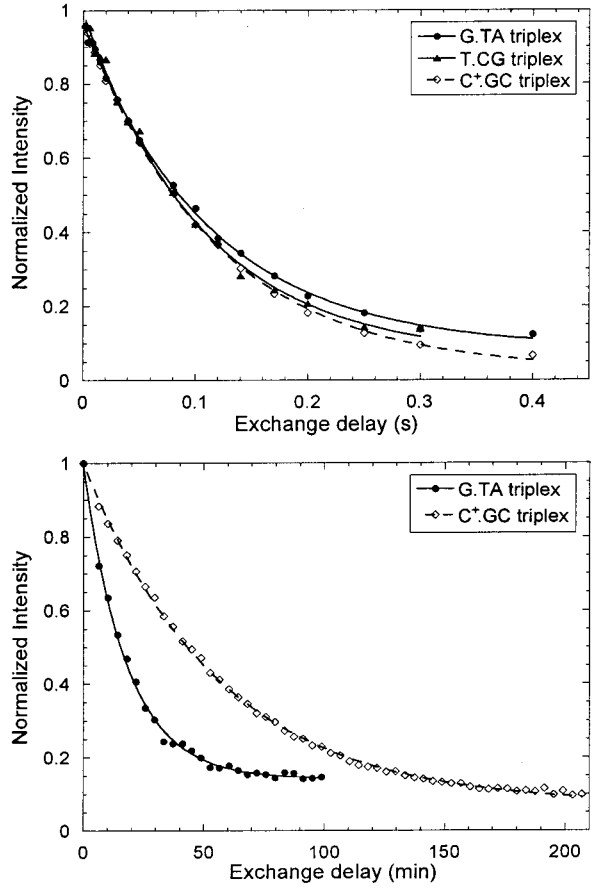

FIGURE 5: Examples of the dependence of the intensity of imino proton resonances on the exchange delay. Upper panel: exchange of the imino proton in G_8 measured in transfer of magnetization experiments. Lower panel: exchange of the imino proton in T_{17} measured in real-time exchange experiments.

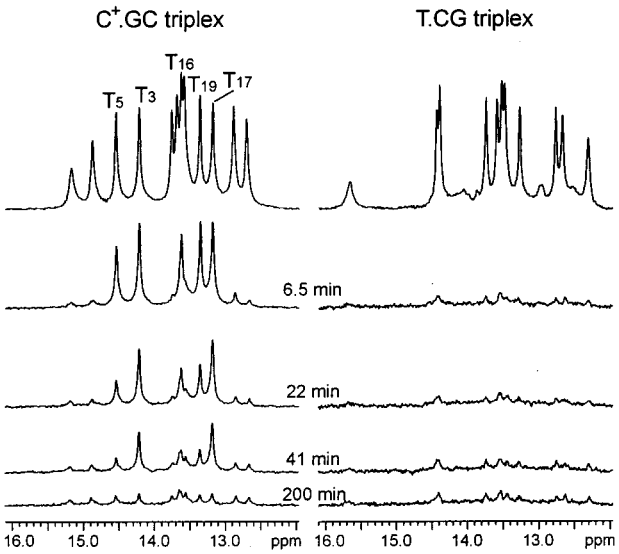

FIGURE 6: Selected imino proton spectra of the C+·GC and T·CG triplexes during real-time exchange measurements at 5 °C. The exchange time (minutes) is given for each spectrum. The spectra at the top of the figure are control spectra taken in 90% H_2O/10% D_2O. The imino proton resonances of the C+·GC triplex which show changes in intensity during real-time exchange are indicated.

measurable in real-time exchange experiments. This result is shown in Figure 6. As one can see, in the first spectrum acquired after the initiation of exchange (i.e., exchange delay of 6.5 min), the imino protons in this triplex have already exchanged with deuterium, indicating that their exchange rates are faster than $\sim 4 \times 10^{-3}$ s^{-1}. In contrast, for the canonical triplex, the exchange of several imino protons can be monitored in these experiments.

The exchange rates of all imino protons in the three DNA triplexes are summarized in Table 1. As the data show, the range of values spans 5 orders of magnitude, from 100 to 4×10^{-4} s^{-1}. Within this range, the exchange of several imino protons is too fast to be measured in real-time exchange experiments (i.e., $k_{ex} > 4 \times 10^{-3}$ s^{-1}) and too slow to be measured in transfer of magnetization experiments (i.e., $k_{ex} < 0.3$ s^{-1}). For these protons we have tried to define the exact value of the exchange rate by measurements at different temperatures. The range of temperatures that could be used was limited by the stability of the triple-helical structure to temperatures lower than 15 °C. We have found that, in this temperature range, the exchange rates of the imino protons of interest remained in the range from 4×10^{-3} to 0.3 s^{-1}, which is not accessible to our NMR measurements.

DISCUSSION

Imino Proton Exchange in the Canonical C+·GC Triple Helix. The exchange of imino protons in DNA triple helices results from the fluctuations in the structure that yield open, solvent-accessible states for Hoogsteen and Watson−Crick T, G, and C+ bases. The wide range of values of the exchange rates reflects (i) variations in the rates of opening and closing of individual bases throughout the structure (eqs 2 and 7) and (ii) different efficiencies of proton transfer from the open state, which depend on the chemical nature of the imino group and of the acceptor (eq 5). We have previously shown that, in YRY DNA triple helices and under experimental conditions similar to those used in the present work, the imino proton in the open state is transferred to other functional groups in the same DNA molecule or to water (*28*). Internal catalysis dominates the exchange of imino protons in Hoogsteen C+·G and in Watson−Crick base pairs. For protonated cytosines, the acceptor of the N3H proton is the N7 group of the guanine in the same C+·G base pair (Figure 1). The pK value of this acceptor group [pK = 2.0 (*35*)] is only 2.4 pH units lower than that of the cytosine imino group [pK = 4.4 (*35*)]. Therefore, the efficiency of internal catalysis is high, e.g., $k_{ex,open}^{int} \approx 6 \times 10^8$ s^{-1} at 10 °C, and the exchange is in, or close to, the EX1 regime (*28*). The high efficiency of internal catalysis also explains why the rates of exchange of cytosine imino protons are large (Table 1). For Watson−Crick base pairs, the pK values of the imino groups [9.6 and 10.1 for G-N1H and T-N3H, respectively (*35*)] are much higher than those of the acceptors [3.7 and 4.4 for A-N1 and C-N3, respectively (*35*)]. As a result, internal catalysis is less efficient, i.e., $k_{ex,open}^{int} \approx 10^4$−$10^6$ s^{-1} at 10 °C (*28*). For Watson−Crick base pairs located in the center of the triple-helical structure (e.g., $A_{12}T_3$, $G_{11}C_4$, and $A_{10}T_5$) the exchange rates are further decreased by their lower equilibrium constants for opening. For base pairs located close to the ends of the structure (e.g., $G_{14}C_1$ and G_8C_7) the low efficiency of internal catalysis is compensated, in part, by the larger equilibrium constants for opening of these base pairs, and the exchange rates are higher.

In contrast to Watson−Crick and Hoogsteen C+·G base pairs, the exchange of imino protons in Hoogsteen T·A base

Table 1: Imino Proton Exchange Rates (s^{-1}) at 5 °C: C^+·GC Triplex (Upper Row), G·TA Triplex (Middle Row), and T·CG Triplex (Lower Row)

C1	C2	T3	C4/A4/G4	T5	T6	C7
		$(4.1 \pm 0.3) \times 10^{-4}$		$(1.2 \pm 0.1) \times 10^{-3}$	$4 \times 10^{-3} < k_{ex} < 0.3$	
		$(2.1 \pm 0.2) \times 10^{-3}$		$4 \times 10^{-3} < k_{ex} < 0.3$	$4 \times 10^{-3} < k_{ex} < 0.3$	
		$4 \times 10^{-3} < k_{ex} < 0.3$	1.3 ± 0.1	$4 \times 10^{-3} < k_{ex} < 0.3$	$4 \times 10^{-3} < k_{ex} < 0.3$	
G14	G13	A12	G11/T11/C11	A10	A9	G8-5′
1.69 ± 0.06	0.43 ± 0.01		$4 \times 10^{-3} < k_{ex} < 0.3$			4.2 ± 0.1
1.91 ± 0.04	0.39 ± 0.01		$(1.3 \pm 0.1) \times 10^{-3}$			4.1 ± 0.1
2.10 ± 0.08	0.41 ± 0.02					4.6 ± 0.3
3′-C^+21	C^+20	T19	C^+18/G18/T18	T17	T16	C^+15
a	45 ± 2	$(9.9 \pm 0.9) \times 10^{-4}$	2.4 ± 0.1	$(3.0 \pm 0.3) \times 10^{-4}$	$(0.9 \pm 0.1) \times 10^{-3}$	*a*
a	100 ± 12	$(2.5 \pm 0.3) \times 10^{-3}$	*b*	$(9.2 \pm 0.9) \times 10^{-4}$	$(1.0 \pm 0.1) \times 10^{-3}$	*a*
a	65 ± 3	0.35 ± 0.01	*b*	$4 \times 10^{-3} < k_{ex} < 0.3$	0.40 ± 0.01	*a*

[a] Imino proton resonance is not observable. [b] Imino proton resonance overlaps the T-N3H resonances from the loops.

pairs most likely occurs by direct transfer of the proton from the open state to water. The rate of this transfer is small, i.e., $k_{ex,open} \approx 30\ s^{-1}$(*36*), explaining why the imino proton exchange for thymines in the Hoogsteen strand is very slow (Table 1).

Imino Proton Exchange and Stabilities of the G·TA and T·CG Triads. The exchange rates obtained in the present work allow us to characterize the stabilities of the Watson−Crick base pairs in the G·TA and T·CG triads. This characterization is based on the exchange rates of imino protons in T_{11} in the G·TA triplex and in G_4 in the T·CG triplex. The exchange rate of the T_{11} imino proton in the G·TA triplex ($k_{ex} = 1.3 \times 10^{-3}\ s^{-1}$, Table 1) is similar to those of imino protons in Watson−Crick AT base pairs in the canonical triplex, e.g., $1.2 \times 10^{-3}\ s^{-1}$ for the $A_{10}T_5$ base pair. This result indicates that the stability of the $T_{11}A_4$ base pair in the G·TA triad is comparable to that of AT base pairs in canonical T·AT triads. Similarly, for the imino proton of the $C_{11}G_4$ base pair in the T·CG triplex, the exchange rate ($1.3\ s^{-1}$) is of the same order of magnitude as those of imino protons in Watson−Crick GC base pairs in the canonical triplex, i.e., $G_{13}C_2$, $G_{14}C_1$, and G_8C_7 (Table 1). This suggests that the stability of the CG base pair in the T·CG triad is also comparable to that of GC base pairs in canonical triads. One should note, however, that the Watson−Crick GC base pairs used for this comparison are all located within two positions from the ends of the triplex. Due to this location, their stabilities are probably lower than that of a central GC base pair. A more informative comparison would be to the $G_{11}C_4$ base pair in the canonical triplex. Unfortunately, for this base pair, the value of the exchange rate falls into the range that is not accessible to NMR measurements.

Our present results do not allow a characterization of the stabilities of the Hoogsteen bases in the G·TA and T·CG triads for two reasons. First, the imino proton resonances of G_{18} and T_{18} in the two substituted triplexes overlap the imino proton resonances from the loops (see Results section). Thus, their individual exchange rates cannot be measured. Second, according to the structures of G·TA and T·CG triplexes (*19*, *20*), formation of G·TA and T·CG triads involves a single Hoogsteen hydrogen bond from the amino groups of G_{18} and C_{11}, respectively (Figure 1). This hydrogen-bonding pattern does not involve the imino protons of G_{18} and T_{18} and leaves these protons accessible to solvent. Therefore, were the resonances of these imino protons resolved in the spectra, their exchange rates would not have been indicative of the stability of the Hoogsteen base in G·TA and T·CG triads.

Effects of G·TA and T·CG Triads upon the Stability of Canonical Triads. The exchange rates of imino protons in canonical triads can be used to map the energetic effects that the G·TA or the T·CG triad has upon the rest of the triple-helical structure. According to the local opening model for proton exchange (*23*), the free energy change in the opening reaction, ΔG_{op} (eq 3), is a measure of the structural stability at the site containing the imino proton. Therefore, any structural perturbation that affects that stability at that site should result in a change in the rate of exchange of the imino proton contained in that site. This change in stabilization free energy is related to the observed change in exchange rate by

$$\delta\Delta G_{op} = -RT \ln\frac{K_{op}}{K'_{op}} = -RT \ln\frac{k_{ex}}{k'_{ex}} \qquad (8)$$

where the primed symbols refer to the perturbed site.[2]

The data presented in Table 1 indicate that the exchange rates of imino protons in triads located in the vicinity of the G·TA/T·CG site are different from the corresponding ones in the canonical triplex. One example is the T_{19}·$A_{12}T_3$ triad, located on the 3′-side of the G·TA/T·CG site (this directionality is defined relative to the purine strand in the triplexes). In this triad, the imino proton exchange rates in both Watson−Crick and Hoogsteen bases are increased in the substituted triplexes as compared to the canonical triplex. Interestingly, these increases are larger in the T·CG triplex than in the G·TA triplex. For example, relative to the canonical triplex, the exchange rate of the imino proton in the Hoogsteen T_{19}·A_{12} base pair increases 2.5-fold in the G·TA triplex and 350-fold in the T·CG triplex. These increases correspond to unfavorable changes in the stabilization free energy of the T·A Hoogsteen base pair of 0.5 and 3.2 kcal/mol, respectively (at 5 °C). A similar pattern is observed for the T_{17}·$A_{10}T_5$ triad, which is located on the 5′-side of the G·TA/T·CG site. The exchange rates of the imino proton in the Watson−Crick base pair of this triad are increased at least 3.3-fold in the two substituted triplexes. For the Hoogsteen base pair in the same triad the exchange rate increases 3-fold in the G·TA triplex and more than 10-fold

[2] This equation is valid only in the EX2 regime of exchange (eq 7). For protons whose exchange is in the EX1 regime (eq 6), the equation is valid only if the structural perturbation does not affect the rate of base pair closing. Moreover, in using this equation to calculate the changes $\delta\Delta G_{op}$, we make the assumption that the structural perturbation does not affect the rate of proton exchange from the open state (eq 7).

in the T·CG triplex. Therefore, the T_{17}·$A_{10}T_5$ triad is also destabilized relative to the canonical triplex by 0.6 to, at least, 1.3 kcal/mol (at 5 °C).

The energetic effects of the G·TA and T·CG triads extend to canonical triads at positions removed from the site of the substitution. Examples of these long-range effects are the Hoogsteen T_{16}·A_9 and C_{20}^+·G_{13} base pairs and the Watson−Crick G_{14}·C_1 base pair (Table 1). As for the other triads, these effects are larger in the T·CG than in the G·TA triplex. For instance, the stability of the Hoogsteen T_{16}·A_9 base pair is the same in the C^+·GC and G·TA triplexes but is decreased by 3.3 kcal/mol in the T·CG triplex. The only exception to this trend is the Hoogsteen C_{20}^+·G_{13} base pair for which the exchange rate suggests that the stability in the G·TA triplex is lower than that in the T·CG triplex.

The energetic effects observed in the present work for the G·TA triplex are very similar to the effects that we have recently reported for the triple helix formed by the 32-mer oligonucleotide d(AGATAGAACCCCTTCTATCTTATA-TCTGTCTT) (*29*). This oligonucleotide also forms a YRY triple helix with a central G·TA triad (*18*). However, this triple helix is one triad longer than the G·TA triplex investigated in the present work, and the base sequence is different. As in the triplex investigated here, the stability of the TA base pair in the G·TA triad in the longer triplex was found to be comparable to that of AT base pairs in the duplex part of the structure. Moreover, destabilization effects were observed for triads two positions removed from the G·TA site. This similarity between the two triple helices suggests that the effects of the G·TA triad do not depend strongly on base sequence context.

The availability of solution structures for two of the DNA triple helices investigated allows an unique opportunity to compare structural variations to the dynamic changes probed by proton exchange. For the G·TA triplex the structure reveals perturbations in both the duplex part and in the third strand (*19*). The duplex part of the structure is generally underwound, with an average helical twist of 31° (as compared to 33° and 36° in A-form and B-form, respectively). The only exception to this trend is at the A_4-T_5 step where the helical twist increases to 37°. As a result, the intrastrand stacking at the (A_4-T_5)·(A_{10}-T_{11}) step is enhanced. For the third strand, the largest perturbations in the structure occur for the trinucleotide segment T_{17}-G_{18}-T_{19} and for the bases at the 3′-end. At the T_{17}-G_{18} step the helical twist increases to 72°, and this results in a novel stacking interaction between T_{17} and G_{18}. The opposite effect is observed at the G_{18}-T_{19} step where the helical twist decreases to −20° and the stacking between G_{18}-T_{19} is greatly reduced. At the 3′-end of the strand, the axial rise at the C_{20}^+−C_{21}^+ step increases from 3.4 to ∼4.5 Å. The presence of these perturbations in the structure certainly predicts changes in the stability of individual triads. However, the correlation between the nature of the structural variations and the changes in proton exchange rates that we have observed is not immediate. For example, on the basis of the structure, one may assign the destabilization of Watson−Crick base pairs in the G·TA triplex to the fact that the double helix is underwound. However, this mechanism would not explain the higher exchange rates for the bases within the regions with increased helical twist, namely, A_4-T_5 and T_{17}-G_{18}. Similarly, variations in the stacking interactions do not account for the observed changes in proton exchange rates. For example, at the trinucleotide segment T_{17}-G_{18}-T_{19} in the third strand, both 5′- and 3′-bases are destabilized although on the 5′-side the stacking is enhanced and on the 3′-side the stacking is poor. The complementarity between the structures of the two triplexes and their dynamics probed by proton exchange is even more evident for the T·CG triplex. The structure of this triplex shows that the magnitude and the pattern of the perturbations induced by the T·CG triad are very similar to those observed in the G·TA triplex (*20*). In contrast, our results indicate that the destabilization of canonical triads in the T·CG triplex is larger than that in the G·TA triplex.

The global stability of G·TA- and T·CG-containing triple helices has been previously measured by UV melting (*16*, *37*), quantitative affinity cleavage (*14*), and gel electrophoresis (*15*). The results have shown that T·CG-containing triple helices are less stable than G·TA-containing triple helices. For example, Dervan and co-workers have measured the association equilibrium constants for formation of triple-helical complexes between a 15-base oligonucleotide and a 242 base pair DNA fragment using quantitative affinity cleavage titration (*14*). Substitution of a canonical triad by the G·TA triad was found to decrease the overall stability of the triplex by 1.5−1.7 kcal/mol (at 22 °C). In the presence of the T·CG triad, the stability of the triplex was found to be even lower, namely, by 2.1−2.3 kcal/mol. These changes most likely contain energetic contributions from many more interactions than those probed in the present work by the exchange of imino protons. Therefore, a quantitative comparison between the two sets of results is not possible. Nevertheless, our present findings provide several new insights into the molecular mechanisms that are responsible for the differences in the free energy of triplex formation observed by Dervan and co-workers (*14*). First, our results suggest that the lower overall stability of G·TA- and T·CG-containing triplexes is not due to an intrinsically lower stability of these two triads. Instead, they show that the stability of Watson−Crick base pairs in these two triads approaches that in canonical triads. Second, a significant contribution to the lower stabilities of G·TA- and T·CG-containing triplexes most likely results from the destabilization effects that these triads have upon neighboring, canonical triads. According to our present results, as well as other recent results from our laboratory (*29*), these destabilization effects are long range and can extend to triads two or three positions removed from the G·TA/T·CG site. These long-range effects also explain why T·CG-containing triplexes have a lower overall stability than G·TA-containing triplexes. As our results show, the T·CG triad affects a larger number of neighboring triads than the G·TA triad does. Moreover, the magnitude of the long-range effects induced by the T·CG triad is also larger than that of the effects induced by the G·TA triad. The general validity of these suggestions could be established by mapping the stability of individual base pairs in DNA triplexes using proton exchange, over a range of solvent conditions, such as pH and ionic strength, known to affect the free energy of formation of these structures. This work is currently in progress in our laboratory.

REFERENCES

1. Cooney, M., Czernuszewicz, G., Postel, E. H., Flint, S. J., and Hogan, M. E. (1988) *Science 241*, 456−459.
2. Postel, E. H., Flint, S. J., Kessler, D. J., and Hogan, M. E. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 8227−8231.
3. Orson, F. M., Thomas, D. W., McShan, W. M., Kessler, D. J., and Hogan, M. E. (1991) *Nucleic Acids Res. 19*, 3435−3441.
4. Vasquez, K. M., and Wilson, J. H. (1998) *Trends Biochem. Sci. 23*, 4−9.
5. Young, S. L., Krawczyk, S. H., Matteucci, M. D., and Toole, J. J. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 10023−10026.
6. Maher, L. J., Dervan, P. B., and Wold, B. (1992) *Biochemistry 31*, 70−81.
7. Duval-Valentin, G., Thuong, N. T., and Helene, C. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 504−508.
8. Maher, L. J. I. (1992) *Biochemistry 31*, 7587−7594
9. Strobel, S. A., Doucette-Stamm, L. A., Riba, L., Housman, D. E., and Dervan, P. B. (1991) *Science 254*, 1639−1642.
10. Frank-Kamenetskii, M. D., and Mirkin, S. M. (1995) *Annu. Rev. Biochem. 64*, 65−95.
11. Soyfer, V. N., and Potaman, V. N. (1995) *Triple-Helical Nucleic Acids*, Springer, New York.
12. Gowers, D. M., and Fox, K. R. (1999) *Nucleic Acids Res. 27*, 1568−1577.
13. Griffin, L. C., and Dervan, P. B. (1989) *Science 245*, 967−971.
14. Best, G. C., and Dervan, P. B. (1995) *J. Am. Chem. Soc. 117*, 1187−1193.
15. Yoon, K., Hobbs, C. A., Koch, J., Sardaro, M., Kutny, R., and Weis, A. L. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 3840−3844.
16. Mergny, J.-L., Sun, J.-S., Rougee, M., Montenay-Garestier, T., Barcelo, F., Chomilier, J., and Helene, C. (1991) *Biochemistry 30*, 9791−9798.
17. Kiessling, L. L., Griffin, L. C., and Dervan, P. B. (1992) *Biochemistry 31*, 2829−2834.
18. Wang, E., Malek, S., and Feigon, J. (1992) *Biochemistry 31*, 4838−4846.
19. Radhakrishnan, I., and Patel, D. J. (1994) *Structure 2*, 17−32.
20. Radhakrishnan, I., and Patel, D. J. (1994) *J. Mol. Biol. 241*, 600−619.
21. Radhakrishnan, I., and Patel, D. J. (1994) *Biochemistry 33*, 11405−11416.
22. Wang, E., and Feigon, J. (1999) in *Oxford Handbook of Nucleic Acid Structure* (Neidle, S., Ed.) pp 355−388, Oxford University Press, New York.
23. Englander, S. W., and Kallenbach, N. R. (1984) *Q. Rev. Biophys. 16*, 521−655.
24. Englander, S. W., Sosnick, T. R., Englander, J. J., and Mayne, L. (1996) *Curr. Opin. Struct. Biol. 6*, 18−23.
25. Moe, J. G., and Russu, I. M. (1992) *Biochemistry 31*, 8421−8428.
26. Plateau, P., and Gueron, M. (1982) *J. Am. Chem. Soc. 104*, 7310−7311.
27. Mihailescu, M. R., and Russu, I. M. (2001) *Proc. Natl. Acad. Sci. U.S.A. 98*, 3773−3777.
28. Powell, S., Jiang, L., and Russu, I. M. (2001) *Biochemistry 40*, 11065−11072.
29. Jiang, L., and Russu, I. M. (2001) *Nucleic Acids Res. 29*, 4231−4237.
30. Lippens, G., Dhalluin, C., and Wieruszeski, J.-M. (1995) *J. Biomol. NMR 5*, 327−331.
31. Hvidt, A., and Nielsen, S. O. (1966) *Adv. Protein Chem. 21*, 287−386.
32. Gueron, M., Kochoyan, M., and Leroy, J. L. (1987) *Nature (London) 328*, 89−92.
33. Gueron, M., and Leroy, J. L. (1995) *Methods Enzymol. 261*, 383−413.
34. Macaya, R., Wang, E., Schultze, P., Sklenar, V., and Feigon, J. (1992) *J. Mol. Biol. 225*, 755−773.
35. Ts'o, P. O. P. (1974) *Basic Principles in Nucleic Acid Chemistry*, Vol. I, Academic Press, New York.
36. Nonin, S., Leroy, J.-L., and Gueron, M. (1996) *Nucleic Acids Res. 24*, 586−595.
37. Roberts, R. W., and Crothers, D. M. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 9397−9401.

BI011832S