Proton Exchange and Base Pair Opening in a DNA Triple Helix†

Stephen W. Powell,‡ Lihong Jiang,‡ and Irina M. Russu*,§

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

Received May 1, 2001; Revised Manuscript Received June 20, 2001

ABSTRACT: Nuclear magnetic resonance spectroscopy has been used to characterize opening reactions and stabilities of individual base pairs in two related DNA structures. The first is the triplex structure formed by the DNA 31-mer 5′-AGAGAGAACCCCTTCTCTCTTTTTCTCTCTT-3′. The structure belongs to the YRY (or parallel) family of triple helices. The second structure is the hairpin double helix formed by the DNA 20-mer 5′-AGAGAGAACCCCTTCTCTCT-3′ and corresponds to the duplex part of the YRY triplex. The rates of exchange of imino protons with solvent in the two structures have been measured by magnetization transfer from water and by real-time exchange at 10 °C in 100 mM NaCl and 5 mM $MgCl_2$ at pH 5.5 and in the presence of two exchange catalysts. The results indicate that the exchange of imino protons in protonated cytosines is most likely limited by the opening of Hoogsteen C^+G base pairs. The base pair opening parameters estimated from imino proton exchange rates suggest that the stability of individual Hoogsteen base pairs in the DNA triplex is comparable to that of Watson−Crick base pairs in double-helical DNA. In the triplex structure, the exchange rates of imino protons in Watson−Crick base pairs are up to 5000-fold lower than those in double-helical DNA. This result suggests that formation of the triplex structure enhances the stability of Watson−Crick base pairs by up to 5 kcal/mol. This stabilization depends on the specific location of each triad in the triplex structure.

Triple-helical nucleic acid structures are formed by binding of a third strand in the major groove of a double-helical nucleic acid. Formation of these structures generally requires a tract of purines in one strand of the double helix and a tract of pyrimidines in the other strand. The base composition of the third strand can be purine or pyrimidine rich. In YRY[1] triple helices, the third strand is pyrimidine rich and binds parallel to the homopurine strand of the double helix. In contrast, in RRY triple helices, the purine-rich third strand is oriented antiparallel to the homopurine strand of the double helix (*1*, *2*).

During the past decades there has been a resurgence of interest in triple-helical structures of nucleic acids due to their potential biological roles and applications. A biological role for these structures is suggested by the fact that long tracts of contiguous oligopurines occur in eukaryotic genomes at frequencies 3−6-fold higher than that predicted on a statistical basis (*3*). Many of these sequences are mirror repeats and can form triple-helical H-DNA structures in negatively supercoiled plasmids (*4*−*6*). Base sequences capable of forming H-DNA structures have been implicated in a variety of biological processes such as transcription control (*2*), lytic replication of the Epstein−Barr virus genome (*7*), and homologous recombination (*8*, *9*).

Applications of triple-helical structures in biotechnology and molecular medicine rely upon the ability of triplex-forming oligonucleotides to target base sequences in double-helical DNA with a degree of specificity comparable to, or even greater than, that of regulatory DNA-binding proteins. In the antigene strategy, the targeted sites are located within coding regions or upstream of the genes of interest. Formation of triple-helical structures at these sites inhibits transcription in vitro (*10*−*12*) and in vivo (*13*, *14*). Triplex-forming oligonucleotides can also be used as sequence-specific endonucleases by covalently linking them to DNA-cleaving agents such as Fe(II)-EDTA or copper(II) phenanthroline (*15*, *16*). Such artificial endonucleases can access rare sites on the genome with high specificity, thus having great value for physical mapping of chromosomes (*17*).

The understanding of the possible biological roles of nucleic acid triple helices, as well as the design and use of triplex-forming oligonucleotides, requires understanding how the structure and stability of triple helices depend on their base sequences and on solution conditions. High-resolution structures for several YRY and RRY DNA triple helices have been obtained by nuclear magnetic resonance (NMR) spectroscopy (*18*, *19*). The structures have revealed the base pairing schemes for canonical and noncanonical base triads, the orientation and the conformation of the third strand, and the conformational changes induced in the duplex by binding of the third strand in its major groove. The stability of triple helical structures has been extensively characterized by UV absorbance spectroscopy, calorimetric techniques, gel electrophoresis, and quantitative affinity cleavage titration (*20*−*24*). These studies have shown that the overall stability of DNA triple helices is strongly influenced by the base

† Supported by a grant from the National Science Foundation.

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

‡ Department of Molecular Biology and Biochemistry.

§ Department of Chemistry.

[1] Abbreviations: NMR, nuclear magnetic resonance; ppm, parts per million; Tris, tris(hydroxymethyl)aminomethane; A, adenine; G, guanine; C, cytosine; T, thymine; Y, either pyrimidine nucleotide; R, either purine nucleotide.

10.1021/bi010890a CCC: $20.00
Published on Web 08/24/2001

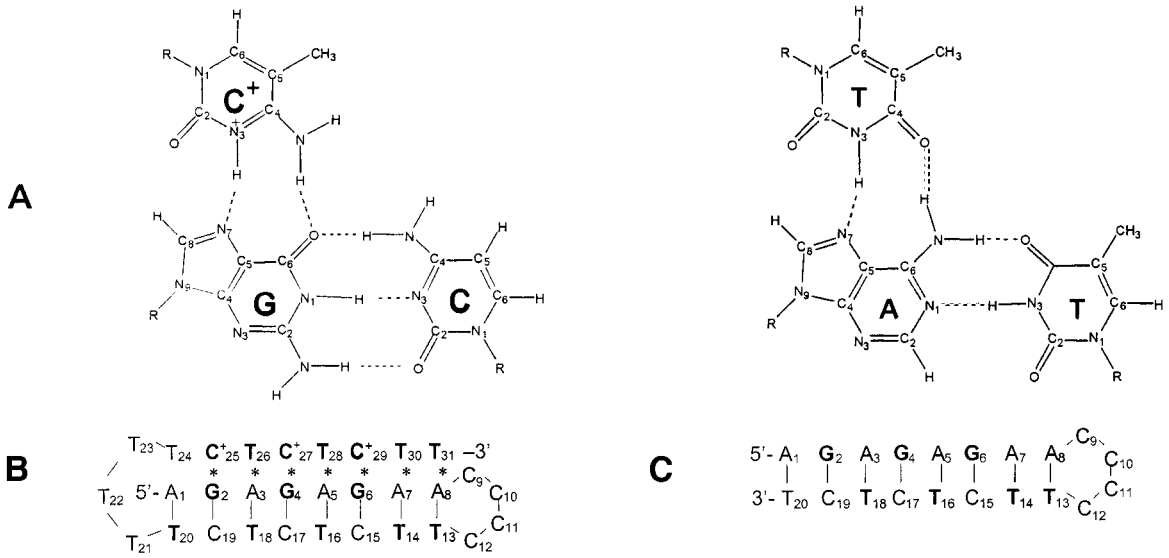

FIGURE 1: (A) Structures of C^+·GC and T·AT triads (*5*). (B) Base sequence and folded conformation of the DNA 31-mer triplex investigated (*25*). (C) Base sequence and folded conformation of the DNA 20-mer hairpin investigated. In panels B and C, Watson−Crick hydrogen bonding is indicated by vertical bars, and Hoogsteen hydrogen bonding is indicated by asterisks. The bases shown in bold contain hydrogen-bonded imino groups.

sequence, length of the strands, and presence of triplet mismatches and looped-out bases, as well as solution conditions such as temperature, pH, and salt concentration.

The present work extends these previous investigations by providing a characterization of the stability of a DNA triple helix at the level of individual base triads. The DNA triple helix investigated belongs to the YRY family (Figure 1B) and is the first intramolecular DNA triplex for which a model structure was derived from NMR data (*25*). Moreover, this triple helix has served as a reference for subsequent structural studies of triple helixes containing noncanonical triads (*26*) and changes in the intramolecular loops (*27*). To compare triple- and double-helical structures directly, we have also characterized the DNA hairpin corresponding to the duplex part of the triplex structure (Figure 1C).

MATERIALS AND METHODS

DNA Samples. The two DNA oligonucleotides were synthesized on an Applied Biosystems 381A automated DNA synthesizer using the solid-support phosphoramidite method. They were purified by reverse-phase HPLC on a PRP-1 column (Hamilton) in 50 mM triethylamine acetate buffer at pH 7 with a gradient of 10−20% acetonitrile in 32 min. The counterions were replaced with Na^+ ions by repeated centrifugation through Centricon YM-3 tubes (Amicon Inc.). Unless otherwise indicated, the final samples were in 100 mM NaCl and 5 mM $MgCl_2$ at pH 5.5 (measured at 10 °C). The sample of the triplex contained 230 OD_{260} units of DNA and that of hairpin 140 OD_{260} units.

NMR Experiments. The NMR experiments were performed at 10 °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 (*28*). Proton exchange rates were measured in experiments of transfer of magnetization from water and in real-time exchange experiments.

In transfer of magnetization experiments, the water proton resonance was selectively inverted using a Gaussian 180° pulse (6.4 ms). The intensity of imino proton resonances was measured as a function of the exchange delay, τ, following water inversion. A weak gradient (0.21 G/cm) was applied during the delay τ to prevent the effects of radiation damping upon the recovery of water magnetization to equilibrium. The observation was with the jump-and-return pulse sequence. The dependence of the intensity of an exchangeable proton resonance on the delay τ is described by (*29*)

$$I(\tau) = I^0 + [I(0) - I^0 - A]e^{-(R_1 + k_{ex})\tau} + Ae^{-R_w\tau} \quad (1)$$

with

$$A = (q - 1)\frac{k_{ex}}{R_1 + k_{ex} - R_w}I^0$$

where I^0 is the intensity at equilibrium, R_1 is the longitudinal relaxation rate, and k_{ex} is the exchange rate for the proton of interest; R_w is the longitudinal relaxation rate of water, and $q = I_w(0)/I^0_w$ expresses the efficiency of water inversion (e.g., $q = -1$ for perfect inversion). In each experiment, 27−30 values of the exchange delay were used. To avoid the effects of spin diffusion, the exchange delays ranged from 2 to 600 ms. The intensity of the exchangeable proton resonance of interest was fitted as a function of the exchange delay τ to eq 1, using a nonlinear least-squares program. The errors reported in the paper are those obtained from the fit. R_w and q were measured in separated experiments. The transfer of magnetization experiments allowed measurement of exchange rates faster than $\sim 0.2\ s^{-1}$. For slower exchange rates, the time dependence of the magnetization (eq 1) is dominated by the longitudinal relaxation of the imino proton and of water protons, and the exchange rates cannot be measured accurately.

In real-time exchange experiments, the exchange was initiated by adding D_2O to a concentrated DNA solution in H_2O, to a final volume fraction of D_2O of ∼80%. A total of 64 transients were accumulated for each spectrum with a total acquisition time of 4 min per spectrum. The intensity of each resonance was measured using standard deconvo-

lution software (Varian) and fitted as a function of the exchange time τ to the equation

$$I(\tau) = [I(0) - I(\infty)] \exp(-k_{ex}\tau) + I(\infty) \quad (2)$$

where $I(\infty)$ is the intensity in a fully exchanged sample. Due to the time elapsed between the initiation of exchange and the first NMR spectrum (7−12 min), the fastest exchange rate that can be measured in real-time exchange experiments is $\sim 10^{-3}$ s^{-1}.

The NOESY experiments used to assign the imino proton resonances of the DNA hairpin were performed with a mixing time of 150 ms and a regular NOESY pulse sequence in which each pulse was replaced by a jump-and return pulse (*30*).

Theory of Imino Proton Exchange. In nucleic acids, the exchange of imino protons with solvent protons occurs by transient opening of the base pairs. In the open state, the imino proton is accessible to proton acceptors, and its hydrogen bonds are broken. The open state is short-lived and thermodynamically unfavorable. Accordingly, the rate of exchange of a given proton is (*31*)

$$k_{ex} = \frac{k_{op} k_{ex,open}}{k_{cl} + k_{ex,open}} \quad (3)$$

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 $K_{op} = k_{op}/k_{cl}$ defines the free energy change for the opening reaction (*32*):

$$\Delta G_{op} = -RT \ln K_{op} \quad (4)$$

The exchange of the imino proton from the open state occurs by two mechanisms. In one, the exchange is catalyzed by external proton acceptors present in solution, e.g., OH^-, Tris base, and ammonia. Accordingly, the rate of exchange by external catalysis is proportional to the acceptor concentration (*31*, *32*):

$$k^{ext}_{ex,open} = k_A[\text{acceptor}] \quad (5)$$

The rate constant k_A depends on the pK values of the imino group and of the acceptor:

$$k_A = k_{coll} F \quad (6)$$

where k_{coll} is the rate of collision between the imino group and the acceptor and F is the fraction of productive transfers of the proton in the transient hydrogen-bonded complex between the imino group NH and the acceptor:

$$F = (1 + 10^{-\Delta pK})^{-1} \quad (7)$$

with $\Delta pK = pK(\text{acceptor}) - pK(\text{NH})$. The rate constant k_A is normally approximated as $k_A = \alpha k_A^0$, where k_A^0 is the rate constant for proton exchange in an isolated nucleoside. α is an empirical coefficient that accounts for any differences between external catalysis in the open state of the base pair and that in free nucleosides. Experimental evidence for various catalysts indicates that, in double-helical DNA, $\alpha \leq 1$ (*33*, *34*).

The second mechanism for exchange of imino protons involves proton acceptors in the same DNA molecule. The current model for this intrinsic catalysis (*35*) assumes that, in the open state, the two bases remain close to each other and form an outer-sphere complex with a water molecule. 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. The rate of exchange resulting from this intrinsic catalysis, $k^{int}_{ex,open}$, is also directly proportional to the fraction of productive transfers, F, from the NH group to the acceptors [eq 7 (*35*)]. Experimental evidence provided by Gueron and co-workers indicates that the same open state of the base pair is involved in both external and intrinsic catalysis (*35*). Accordingly, the total rate of exchange from the open state is

$$k_{ex,open} = k^{ext}_{ex,open} + k^{int}_{ex,open} \quad (8)$$

The base pair opening parameters that can be obtained from measurements of exchange rates depend on how the rate of proton exchange from the open state compares with the rate of base pair closing. According to eq 3, one distinguishes two limiting cases:

(i) When $k_{ex,open} \gg k_{cl}$ (EX1 regime)

$$k_{ex} = k_{op} \quad (9)$$

i.e., exchange is limited by the rate of base pair opening. In this case, the exchange rate provides directly the rate of opening.

(ii) When $k_{ex,open} \ll k_{cl}$ (EX2 regime)

$$k_{ex} = K_{op} k_{ex,open} \quad (10)$$

i.e., exchange is slowed relative to that in the open state by a factor equal to the equilibrium constant for opening, K_{op}.

RESULTS

The two DNA molecules investigated are shown in Figure 1. The DNA 31-mer has been first studied by Feigon and co-workers (*25*). These authors have shown that, in acidic solutions, the 31-mer folds into an intramolecular YRY triplex: the pyrimidine strand segment Y (bases 24−31) is located in the major groove of the hairpin duplex in an orientation parallel to the R strand (bases 1−8). The triplex contains seven canonical C^+·GC and T·AT triads (Figure 1). The NMR resonance of the imino proton in thymine T_{24} was not observed in the spectrum. This indicated that T_{24} does not form Hoogsteen hydrogen bonds with the A_1T_{20} base pair and, thus, is not part of a triad (*25*). The base sequence of the DNA 20-mer is the same as that in the duplex part of the triplex. As shown below, in solution, this 20-mer forms a double-helical hairpin structure.

The NMR resonances of interest to the present study are the proton resonances originating from imino groups in the two DNA structures. They are shown in Figure 2. In the DNA triplex, these resonances have been assigned to individual bases by Feigon and co-workers (*25*). The imino proton resonances of the hairpin occur in the same spectral region (i.e., 12.5−14.5 ppm), indicating that the 20-mer is in a double-helical hairpin conformation. We have assigned each resonance to a specific imino group in the hairpin using NOESY experiments (results not shown). The resonances

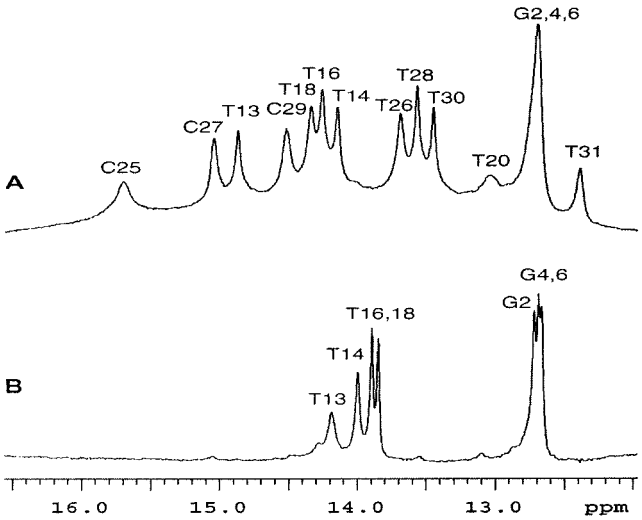

FIGURE 2: NMR resonances of imino protons in the DNA triplex (A) and the DNA hairpin (B) investigated in 100 mM NaCl and 5 mM $MgCl_2$ in 90% H_2O/10% D_2O at pH 5.5 and at 10 °C.

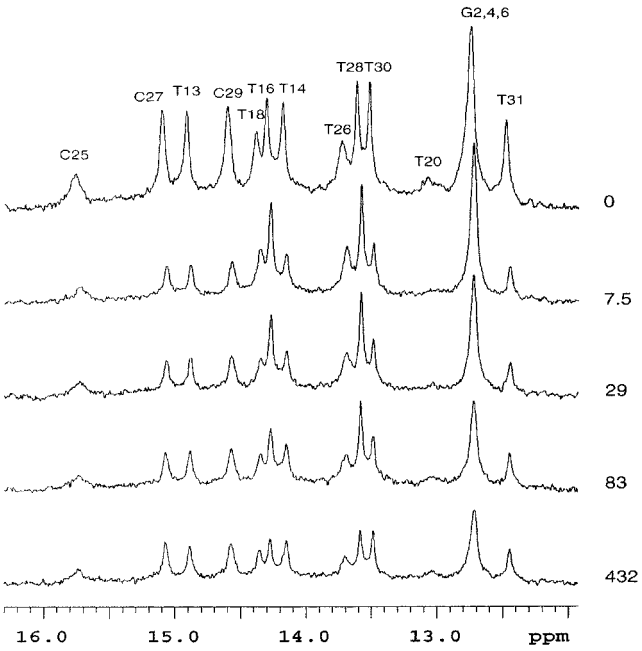

FIGURE 3: Imino proton NMR resonances of the DNA triplex during real-time exchange measurements. Experimental conditions are the same as in Figure 2. The exchange time (in minutes) is indicated for each spectrum.

of the bases located close to the ends of the duplex, i.e., G_2, T_{18}, and T_{13}, were identified by their gradual broadening upon increasing temperature.

The exchange rates of imino protons in both DNA structures were measured using transfer of magnetization from water and real-time exchange experiments. A representative series of spectra during real-time exchange is shown in Figure 3. Spectra from transfer of magnetization experiments are provided as Supporting Information. The exchange rates are summarized in Table 1. Depending on the exchange rate three classes of protons can be distinguished. In class I, the exchange rates range from 0.4 to 42 s^{-1}. This class contains the N3H protons of the protonated cytosines (i.e., C_{25}^+, C_{27}^+, and C_{29}^+) and of the terminal thymines (i.e., T_{13}, T_{20}, and T_{31}) in the triplex and G-N1H and T-N3H protons of the Watson−Crick base pairs in the hairpin. In class II,

Table 1: Exchange Rates of Imino Protons in the DNA Triplex and Hairpin Investigated in 100 mM NaCl and 5 mM $MgCl_2$ at pH 5.5 and at 10 °C

	DNA triplex		DNA hairpin	
base	base pairing	k_{ex} (s^{-1})	base pairing	k_{ex} (s^{-1})
G_2	Watson−Crick	$(4.4 \pm 0.1) \times 10^{-4}$ [a]	Watson−Crick	1.9 ± 0.1
G_4	Watson−Crick	$(4.4 \pm 0.1) \times 10^{-4}$ [a]	Watson−Crick	0.45 ± 0.05
G_6	Watson−Crick	$(4.4 \pm 0.1) \times 10^{-4}$ [a]	Watson−Crick	0.43 ± 0.07
T_{13}	Watson−Crick	2.7 ± 0.2	Watson−Crick	7.0 ± 0.7
T_{14}	Watson−Crick	*b*	Watson−Crick	1.20 ± 0.07
T_{16}	Watson−Crick	$(2.2 \pm 0.1) \times 10^{-4}$	Watson−Crick	1.10 ± 0.06
T_{18}	Watson−Crick	$(1.2 \pm 0.1) \times 10^{-3}$	Watson−Crick	1.62 ± 0.08
T_{20}	Watson−Crick	10 ± 2	Watson−Crick	*c*
C_{25}^+	Hoogsteen	42 ± 4		
T_{26}	Hoogsteen	$(2.6 \pm 0.1) \times 10^{-4}$		
C_{27}^+	Hoogsteen	1.8 ± 0.1		
T_{28}	Hoogsteen	$(1.30 \pm 0.06) \times 10^{-4}$		
C_{29}^+	Hoogsteen	6.6 ± 0.6		
T_{30}	Hoogsteen	*b*		
T_{31}	Hoogsteen	25 ± 1		

[a] In the DNA triplex, the imino proton resonances of G_2, G_4, and G_6 overlap (Figure 2). The value of k_{ex} given is obtained from the overlapped resonance. [b] The exchange rate is too slow to be measured by transfer of magnetization from water ($k_{ex} < \sim 0.2\ s^{-1}$) and too fast to be measured by real-time exchange ($k_{ex} > \sim 10^{-3}\ s^{-1}$). [c] The resonance is not observable due to exchange broadening by fraying at the ends of the duplex.

the exchange rates are $\sim 10^{-3}\ s^{-1}$ or slower. This class contains only triplex protons: G-N1H and T-N3H protons of Watson−Crick base pairs (i.e., G_2, G_4, G_6, T_{16}, and T_{18}) and T-N3H protons of Hoogsteen base pairs (i.e., T_{26} and T_{28}). The exchange rates of the imino protons in Watson−Crick base pairs of the triplex are lower than those in the hairpin by factors of up to 5000. In class III, the exchange is too slow to be measured in experiments of transfer of magnetization from water ($k_{ex} < \sim 0.2\ s^{-1}$) and too fast to be detectable in real-time exchange measurements ($k_{ex} > \sim 10^{-3}\ s^{-1}$). The imino protons of T_{14} and T_{30} in the triplex fall in this class.

To elucidate the mechanism of exchange of imino protons in the DNA triplex, we have also investigated the effects of external catalysis by added proton acceptors. The catalyst investigated was acetate. The choice of acetate was dictated by its low p*K* value [p*K* = 4.76 at 10 °C (*36*)], which allows larger concentrations of acetate base to be obtained in the pH range in which the triplex structure is stable. The concentration of acetate was varied from 0 to 0.45 M at pH 5.5 (corresponding range of the concentration of acetate base 0−0.38 M) in the presence of 100 mM NaCl and 5 mM $MgCl_2$. Catalysis by acetate was observed for N3H protons of the thymines in the third strand, except T_{31}. An example of acetate catalysis is shown in Figure 4 for T_{28}. No catalysis by acetate was observed for imino protons of protonated cytosines or for those of Watson−Crick base pairs in the DNA triplex. This result is illustrated in Figure 4, which shows that the exchange rates of cytosine N3H protons are constant upon increasing the acetate base concentration to 0.34 M.

The exchange of imino protons in the DNA hairpin was further characterized in order to obtain the kinetic parameters for base pair opening in this structure. The exchange rates were measured in transfer of magnetization experiments at pH 8.26 using Tris as an external catalyst. The rationale for choosing these experimental conditions was 2-fold. First, the p*K* value of Tris [8.61 at 10 °C (*37*)] is close to the p*K*'s of

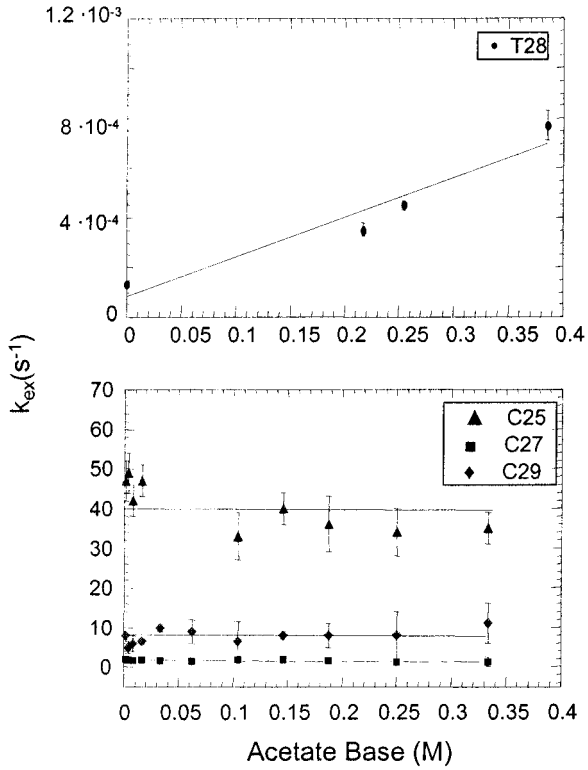

FIGURE 4: Dependence of the exchange rates of N3H protons of T_{28} (upper graph) and of protonated cytosines (lower graph) in the DNA triplex on the concentration of acetate base at 10 °C.

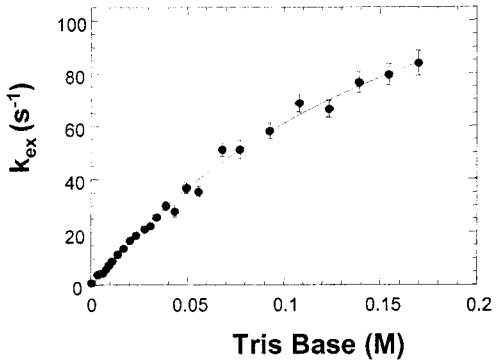

FIGURE 5: Dependence of the exchange rate of N3H proton of T_{16} in the DNA hairpin on the concentration of Tris base. Experimental conditions: 100 mM NaCl and 5 mM $MgCl_2$ at pH 8.26 and at 10 °C. The curve corresponds to the nonlinear least-squares fit to eq 3, which yields $k_{op} = 182 \pm 21$ s^{-1} and $K_{op}k_A = 793 \pm 80$ M^{-1} s^{-1}.

G-NH1 and T-N3H groups [10.1 for thymidine and 9.6 for guanosine at 10 °C (*38*)]. Hence, Tris is an efficient catalyst for the exchange of these imino protons (eq 7). Second, by using a higher pH value, the concentration of Tris base could be increased in the range in which the exchange of these protons approaches the EX1 regime. An example of the dependence of the exchange rate on the concentration of Tris base is shown in Figure 5 for the imino proton of T_{16}. The exchange rates were fitted as a function of base catalyst concentration to eq 3. The fits provided directly the values of the opening rates k_{op}: 182 ± 21 s^{-1} for the A_5T_{16} base pair and 5 ± 1 s^{-1} for the G_4C_{17} and G_6C_{15} base pairs. The opening rates of the A_3T_{18} and A_7T_{14} base pairs could not be determined accurately because the imino proton resonances of these base pairs overlapped at higher Tris concentrations. The equilibrium constants for opening K_{op} were calculated from the fits to eq 3 assuming that the rate constants k_A for proton transfer to Tris base are 2×10^7 M^{-1} s^{-1} for T-N3H and 5×10^7 M^{-1} s^{-1} for G-N1H (*39*). The obtained K_{op} values are 4.2×10^{-5} for the A_5T_{16} base pair and 4.3×10^{-7} for the G_4C_{17} and G_6C_{15} base pairs. These values are typical for Watson−Crick base pairs in double-helical DNA (*39*−*41*).

DISCUSSION

The exchange of imino protons investigated monitors distinct fluctuations of the DNA triplex structure. On one hand, the exchange of T-N3H and G-N1H in the duplex part of the structure occurs through opening of AT and GC Watson−Crick base pairs. On the other hand, the exchange of T-N3H and C^+-N3H in the third strand reflects the opening of TA and C^+G Hoogsteen base pairs. Since the exchange properties of imino protons in Watson−Crick base pairs are different from those of imino protons in Hoogsteen base pairs, we discuss each of these types of protons separately.

Imino Proton Exchange and Opening of Watson−Crick Base Pairs in the DNA Triplex. As shown in Table 1, the exchange rates of T-N3H and G-N1H in the duplex part of the triplex structure are in the range from 2×10^{-4} to 10 s^{-1}. Comparable values have been found by Cain and Glick (*42*) for an intramolecular YRY DNA triplex with and without a disulfide bridge cross-linking the pyrimidine strand to the duplex. Under the experimental conditions used in the present work (i.e., 100 mM NaCl and 5 mM $MgCl_2$ at pH 5.5), the exchange of imino protons occurs in the absence of external catalysts. As explained in the Theory of Imino Proton Exchange section, under these conditions, the exchange is catalyzed by proton acceptors in the same DNA molecule, namely, the C-N3 group for the exchange of G-N1H and the A-N1 group for the exchange of T-N3H (*35*). The exchange rates observed experimentally are slower than the exchange rates resulting from this intrinsic catalysis by a factor corresponding to the equilibrium constant for opening K_{op} (eq 10). Therefore, determination of the equilibrium constants for opening requires knowledge of the rates of intrinsic catalysis, $k_{ex,open}^{int}$. We have calculated these rates on the basis of our present data on the DNA hairpin as follows.

As described in the Results section, for the DNA hairpin, the K_{op} values were determined from the dependence of the exchange rates on Tris concentration at pH 8.26 (namely, K_{op} values are 4.2×10^{-5} for A_5T_{16} and 4.3×10^{-7} for G_4C_{17} and G_6C_{15} base pairs). The rates of intrinsic catalysis at pH 5.5 were calculated from eq 10, using the determined K_{op} values and the exchange rates given in Table 1. This calculation assumes that the K_{op} values in the DNA hairpin at pH 5.5 and in the absence of added catalyst are the same as those in the presence of Tris at pH 8.26. This assumption is justified by the following: (i) in double-helical DNA, external and intrinsic catalysis occurs from the same open state of the base pair (*35*), (ii) in the DNA hairpin, the exchange rates of imino protons in A_5T_{16}, G_4C_{17}, and G_6C_{15} base pairs in the absence of Tris at pH 8.26 are the same as those at pH 5.5, and (iii) in general, for DNA duplexes, the equilibrium constants for opening determined from the Tris dependence of the exchange rates are similar to those

Table 2: Estimates of the Equilibrium Constants and Free Energies for Opening of Individual Base Pairs in the DNA Triplex Investigated at 10 °C

triad	base pair	K_{op}	ΔG_{op}[a]
$C_{25}^+ \cdot G_2C_{19}$	$C_{25}^+G_2$	4×10^{-6} [b]	7.0
	G_2C_{19}	4×10^{-10} [c]	12.1
$T_{26} \cdot A_3T_{18}$	$T_{26}A_3$	9×10^{-6}	6.5
	A_3T_{18}	4×10^{-8}	9.5
$C_{27}^+ \cdot G_4C_{17}$	$C_{27}^+G_4$	2×10^{-7} [b]	8.6
	G_4C_{17}	4×10^{-10} [c]	12.1
$T_{28} \cdot A_5T_{16}$	$T_{28}A_5$	4×10^{-6}	7.0
	A_5T_{16}	7×10^{-9}	10.5
$C_{29}^+ \cdot G_6C_{15}$	$C_{29}^+G_6$	7×10^{-7} [b]	7.9
	G_6C_{15}	4×10^{-10} [c]	12.1
$T_{30} \cdot A_7T_{14}$	$T_{30}A_7$	n/a	n/a
	A_7T_{14}	n/a	n/a
$T_{31} \cdot A_8T_{13}$	$T_{31}A_8$	~1	~0
	A_8T_{13}	9×10^{-5}	5.2
none	A_1T_{20}	3×10^{-4}	4.5

[a] In kcal/mol. [b] Calculated assuming a rate of closing of 10^7 s^{-1}. [c] Calculated from the overlapped resonance (see footnote *a* in Table 1).

determined using other catalysts (*33*). Using this approach, we have found that the values of $k_{ex,open}^{int}$ in the DNA hairpin at 10 °C are 3×10^4 s^{-1} for the AT base pair and 1×10^6 s^{-1} for the GC base pairs. These values are comparable to those previously found for an intermolecular DNA duplex at 15 °C (*35*). Moreover, the values are in good agreement with the predictions made on the basis of the p*K* values of the groups involved in proton transfer (eq 7). For example, for an open AT base pair, the p*K* of the proton acceptor [p*K*(A-N1) = 3.7 in adenosine (*38*)] is 6.4 units lower than the p*K* of the proton donor [p*K*(T-N3H) = 10.1 (*38*)]. Therefore, the fraction of productive proton transfers between these two groups (*F*) is $\sim 4 \times 10^{-7}$ (eq 7). For an open GC base pair, intrinsic catalysis should be more efficient since the p*K* values of the two groups are closer to each other [namely, p*K*(G-N1H) = 9.6 in guanosine and p*K*(C-N3H) = 4.4 in cytidine (*38*)]. Consequently, in these base pairs, the fraction *F* should increase to 7×10^{-6}, and the rate of intrinsic catalysis should be enhanced at least 1 order of magnitude, in accordance to our findings.

The equilibrium constants for opening of Watson−Crick base pairs in the DNA triplex were estimated on the basis of eq 10 from the exchange rates measured experimentally (Table 1) and the rates of intrinsic catalysis determined as described above. The values are 4×10^{-8}, 7×10^{-9}, and 9×10^{-5} for A_3T_{18}, A_5T_{16}, and A_8T_{13} base pairs, respectively; the K_{op} value obtained from the overlapped resonance of the three GN1-H imino protons is 4×10^{-10} (Table 2).

Exchange of Thymine Imino Protons and Opening of TA Hoogsteen Base Pairs in the DNA Triplex. The exchange of imino protons in Hoogsteen TA base pairs may be expected to occur by transfer of the T-N3H proton to the N7 group of adenine in the open TA base pair. However, this mechanism is unlikely since, in free adenine, the N7 group does not protonate (*38*, *43*). Therefore, in the absence of added catalysts, it is more likely that water itself is the acceptor of these protons. In free thymine, the rate of water-catalyzed exchange is ~30 s^{-1} at 10 °C and at pH <5.5 (*44*). This rate should be comparable to the rate of exchange of the T-N3H proton in the open state of a Hoogsteen TA base pair, $k_{ex,open}^{ext}$. Hence, on the basis of eq 10, the equilibrium constants for opening of $T_{26}A_3$ and $T_{28}A_5$ base pairs in the DNA triplex can be estimated as 9×10^{-6} and 4×10^{-6}, respectively (Table 2).

The different mechanisms for exchange of T-N3H protons in Hoogsteen TA and Watson−Crick AT base pairs explain why these two types of protons exhibit different responses to catalysis by acetate (Results section). The efficiency of acetate in catalyzing the exchange of the imino proton in free thymine is relatively low because acetate's p*K* value [4.76 (*36*)] is much lower than the p*K* of T-N3H [10.1 at 10 °C (*38*)]. Thus, the rate constant for acetate-catalyzed exchange of T-N3H protons should be $\sim 3 \times 10^3$ M^{-1} s^{-1} [eqs 6 and 7 with $k_{coll} = 6 \times 10^8$ M^{-1} s^{-1} at 10 °C (*45*)], and for the concentrations of the acetate base investigated, $k_{ex,open}^{ext}$ for acetate catalysis should be less than 1×10^3 s^{-1}. For thymine imino protons in Watson−Crick AT base pairs, this rate is significantly lower than the rate of intrinsic catalysis (3×10^4 s^{-1}), explaining why the effect of acetate upon the exchange rates is not detectable. In contrast, for thymine imino protons in Hoogsteen TA base pairs, the rate for acetate catalysis exceeds that for water catalysis and the effect of acetate is observed (Figure 4).

Exchange of Imino Protons in Protonated Cytosines and Opening of C^+G Hoogsteen Base Pairs in the DNA Triplex. The exchange of imino protons in protonated cytosines is most likely dominated by intrinsic catalysis (*46*), in which the acceptor of the cytosine imino proton is the G-N7 group. The efficiency of intrinsic catalysis in this case can be estimated by comparing a Hoogsteen C^+G base pair to, for example, a Watson−Crick GC base pair. For the C^+G Hoogsteen base pair, the p*K* values of C^+-N3H and G-N7 groups (i.e., 4.4 and 2.0, respectively) differ by only 2.4 pH units ($\Delta pK = -2.4$). In contrast, for a Watson−Crick GC base pair the difference in p*K*'s between G-N1H and C-N3 is 5.2 units (vide supra). Therefore, on the basis of eq 7, the efficiency of intrinsic catalysis in the exchange of the C^+-N3H proton should be $\sim 6 \times 10^2$-fold higher than that for a G-N1H proton. Using our determination of the rate of intrinsic catalysis for G-N1H protons (namely, $k_{ex,open}^{int} = 1 \times 10^6$ s^{-1}), this predicts a rate of intrinsic catalysis, $k_{ex,open}^{int}$, for the C^+-N3H proton of 6×10^8 s^{-1}.

A dominant contribution from intrinsic catalysis to the exchange of imino protons in protonated cytosines is also indicated by our experimental observation that the exchange rates of these protons are independent of the concentration of the acetate base (Figure 4). Acetate is expected to be a good catalyst of the exchange of these protons since its p*K* value (p*K* = 4.76) is similar to that of the C^+-N3H group (p*K* = 4.4). The predicted rate constant for acetate catalysis is 4×10^8 M^{-1} s^{-1} [for $k_{coll} = 6 \times 10^8$ M^{-1} s^{-1} (*45*)]. Hence, for the concentrations of acetate base investigated (i.e., 0−0.34 M), the rate of acetate-catalyzed exchange, $k_{ex,open}^{ext}$, should reach a value of 1.4×10^8 s^{-1}. The lack of an observable effect of acetate upon the exchange rates therefore implies that the rate of intrinsic catalysis, $k_{ex,open}^{int}$, is larger than 1.4×10^8 s^{-1}, in accordance to the prediction made above based on p*K* values.

The high efficiency of intrinsic catalysis suggests that the exchange of imino protons in C^+G base pairs is in, or close to, the opening-limited regime. For this to be true the rate of exchange from the open state must be larger than the rate

of base pair closing (eqs 3 and 9). The rates of closing of C^+G base pairs are expected to be similar to those in double-helical DNA [i.e., 10^7-10^8 s $^{-1}$ (*40*, *41*)]. Hence, it is likely that $k_{ex,open} \geq k_{cl}$ and the exchange rates of C^+-N3H protons measured experimentally (Table 1) are close to the rates of opening C^+G base pairs in C^+·GC triads. On the basis of these arguments, we have estimated the equilibrium constants for opening of C^+G base pairs by assuming a value of 10^7 s $^{-1}$ for the rates of closing of these base pairs. The values of K_{op} and opening free energies obtained with this assumption are shown in Table 2. Higher values of the closing rates would result in lower K_{op} values; hence, our estimates most likely provide upper limits for the equilibrium constants for opening for these bases.

Site-Resolved Structural Energetics in the DNA Triplex. The results obtained in the present work allow us to map the stability of the DNA triplex structure at the level of individual base pairs and distinct base pairing interactions. The exchange of imino protons reflects those conformational fluctuations in the DNA triplex which yield an open, exchange-competent state of the imino group. Accordingly, the local stabilization energies derived from exchange rates represent the free energy changes associated with these conformational fluctuations (ΔG_{op}) (*32*).

One important result of the present investigation is that the equilibrium constants for opening of individual base pairs in the duplex part of the triplex structure are much lower than those in a DNA double helix. In the double-helical hairpin investigated, the equilibrium constants for opening are 4.2×10^{-5} for A_5T_{16} and 4.3×10^{-7} for G_4C_{17} and G_6C_{15} (Results section). Formation of the triple-helical structure reduces these equilibrium constants up to 6000-fold (Table 2). Hence, relative to DNA double helices, the Watson−Crick base pairs in the triplex are further stabilized by extra free energies of up to 5 kcal/mol (eq 4). This stabilization is significant only for Watson−Crick base pairs located in the central part of the triplex structure. For terminal base pairs, such as T_{20} and T_{13}, the exchange rates and the opening equilibrium constants are comparable to those in double-helical DNA (Tables 1 and 2). For the AT base pairs located in the central part of the triplex the exchange rates also depend on the exact location of the base pair in the structure. As shown in Table 1, the exchange rate for T_{18} is five times larger than that for T_{16}. This difference clearly reflects a destabilization of the A_3T_{18} base pair relative to the A_5T_{16} base pair (Table 2). This effect may be related to the location of the A_3T_{18} base pair toward the 5′-end of the triplex structure (this directionality is defined relative to that of the purine Watson−Crick strand). As shown by Feigon and co-workers (*25*), in this part of the structure the $T_{24}A_1T_{20}$ triad does not form. Hence, these observations suggest that the 5′-end part of the triplex structure is less stable than the rest, possibly due to conformational constraints imposed by the short, three-base loop connecting the two pyrimidine strands.

In the DNA triplex, the energetic cost of the structural fluctuations which bring the imino protons of Watson−Crick base pairs into the open, solvent-accessible state is expected to be higher than that in double-helical DNA. For example, formation of the open state for the guanine in a C^+·GC triad should perturb the hydrogen bonds involved in both Watson−Crick and Hoogsteen base pairing (Figure 1). Thus, one expects the free energy change for opening these bases to be higher than in double-helical DNA. However, it is interesting to note that, in Watson−Crick AT base pairs of T·AT triads, the hydrogen-bonding pattern for thymines is the same as that in a DNA double helix. Moreover, analysis of NMR-derived triplex structures (*19*) shows that placing the third strand in the major groove of the double helix does not introduce significant steric hindrance for opening of thymines in either the major or the minor groove. Despite these structural characteristics, the stabilization free energies for AT base pairs in the triplex are significantly higher than those in double-helical DNA (Table 2). This suggests that the opening transitions probed by the exchange of thymine imino protons in Watson−Crick base pairs may involve larger perturbations of the triplex structure at or near the exchanging proton.

Our present results also provide new insights into the opening kinetics and stabilization free energies of Hoogsteen base pairs in the DNA triplex. As shown in Table 2, the equilibrium constants for opening of Hoogsteen base pairs in the central part of the triplex range from 2×10^{-7} to 9×10^{-6}. These values are comparable to those normally found for Watson−Crick base pairs in double-helical DNA [4.2×10^{-5} and 4.3×10^{-7} in the DNA hairpin investigated and other results (*40*, *41*)]. This finding implies that, in a canonical YRY triplex such as the one investigated here, the stability of Hoogsteen base pairs is comparable to that of Watson−Crick base pairs in duplex DNA. As in duplex DNA, the base pairs formed by Hoogsteen interactions appear to open one at the time, independently of each other. The exchange rates of Hoogsteen imino protons, however, depend on the location of the base pairs in the structure. As in the case of double-helical DNA, the exchange of the last two bases, T_{30} and T_{31}, is faster, being affected by fraying at the end of the strand. Moreover, for Hoogsteen imino protons located in the 5′-end part of the structure the exchange rates are higher than those in central triads. For instance, the exchange rate of T_{26} is approximately 2-fold higher than that of T_{28}. Similarly, among the three C^+·GC triads present in the structure, the highest exchange rate is observed for C^+_{25} (Table 1). These observations provide additional support to the suggestion that the triads located at the 5′-end of the structure are less stable than the other central triads in the triplex.

In summary, the results presented in this paper demonstrate that, in a YRY triplex, the binding of a third strand in the major groove of double-helical DNA significantly stabilizes the Watson−Crick base pairs against the structural fluctuations responsible for imino proton exchange. This stabilization appears to be dependent on the location of the base triad relative to the ends of the triplex and on steric constraints imposed by the loop connecting the third strand to the rest of the structure. Similar dependencies are observed for the stabilization free energies of Hoogsteen base pairs. The DNA investigated in the present work is one of only a few DNA canonical triplexes for which proton exchange has been characterized at the level of individual triads (*42*). Thus, no inferences can be drawn yet on the dependence of the local stabilization energy on base sequence. Characterization of nucleic acid triple helices containing different base sequences and selected noncanonical base triads will provide this important information.

ACKNOWLEDGMENT

We thank Dr. Ryszard Michalczyk for help with the DNA synthesis.

SUPPORTING INFORMATION AVAILABLE

Figure 1S showing representative spectra from experiments of transfer of magnetization from water on the DNA triplex investigated. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Frank-Kamenetskii, M. D., and Mirkin, S. M. (1995) *Annu. Rev. Biochem. 64*, 65−95.
2. Soyfer, V. N., and Potaman, V. N. (1995) *Triple-Helical Nucleic Acids*, Springer, New York.
3. Behe, M. J. (1995) *Nucleic Acids Res. 23*, 689−695.
4. Mirkin, S. M., Lyamichev, V. I., Drushlyak, K. N., Dobrynin, V. N., Filippov, S. A., and Frank-Kamenetskii, M. D. (1987) *Nature (London) 330*, 495−497.
5. Mirkin, S. M., and Frank-Kamenetskii, M. D. (1994) *Annu. Rev. Biophys. Biomol. Struct. 23*, 541−576.
6. Schroth, G. P., and Ho, P. S. (1995) *Nucleic Acids Res. 23*, 1977−1983.
7. Portes-Sentis, S., Sergeant, A., and Gruffat, H. (1997) *Nucleic Acids Res. 25*, 1347−1354.
8. Kohwi, Y., and Panchenko, Y. (1993) *Genes Dev. 7*, 1766−1778.
9. Rooney, S. M., and Moore, P. D. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 2141−2144.
10. Cooney, M., Czernuszewicz, G., Postel, E. H., Flint, S. J., and Hogan, M. E. (1988) *Science 241*, 456−459.
11. Duval-Valentin, G., Thuong, N. T., and Helene, C. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 504−508.
12. Maher, L. J., Dervan, P. B., and Wold, B. (1992) *Biochemistry 31*, 70−81.
13. Orson, F. M., Thomas, D. W., McShan, W. M., Kessler, D. J., and Hogan, M. E. (1991) *Nucleic Acids Res. 19*, 3435−3441.
14. Postel, E. H., Flint, S. J., Kessler, D. J., and Hogan, M. E. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 8227−8231.
15. Francois, J. C., Saison-Behmoaras, T., Barbier, C., Chassignol, M., Thuong, N. T., and Helene, C. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 9702−9706.
16. Moser, H. E., and Dervan, P. B. (1987) *Science 238*, 645−650.
17. Strobel, S. A., Doucette-Stamm, L. A., Riba, L., Housman, D. E., and Dervan, P. B. (1991) *Science 254*, 1639−1642.
18. Radhakrishnan, I., and Patel, D. J. (1993) *Structure 1*, 135−152.
19. Wang, E., and Feigon, J. (1999) in *Oxford Handbook of Nucleic Acid Structure* (Neidle, S., Ed.) pp 355−388, Oxford University Press, New York.
20. Plum, G. E., Park, Y. W., Singleton, S. F., Dervan, P. B., and Breslauer, K. J. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 9436−9440.
21. Plum, G. E., Pilch, D. S., Singleton, S. F., and Breslauer, K. J. (1995) *Annu. Rev. Biophys. Biomol. Struct. 24*, 319−350.
22. Plum, G. E. (1997) *Biopolymers 44*, 241−256.
23. Roberts, R. W., and Crothers, D. M. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 9397−9401.
24. Singleton, S. F., and Dervan, P. B. (1992) *J. Am. Chem. Soc. 114*, 6957−6965.
25. Macaya, R., Wang, E., Schultze, P., Sklenar, V., and Feigon, J. (1992) *J. Mol. Biol. 225*, 755−773.
26. Wang, E., Malek, S., and Feigon, J. (1992) *Biochemistry 31*, 4838−4846.
27. Tarkoy, M., Phipps, A. K., Schultze, P., and Feigon, J. (1998) *Biochemistry 37*, 5810−5819.
28. Plateau, P., and Gueron, M. (1982) *J. Am. Chem. Soc. 104*, 7310−7311.
29. Ernst, R. R., Bodenhausen, G., and Wokaun, A. (1987) *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Clarendon Press, Oxford.
30. Michalczyk, R., and Russu, I. M. (1993) *FEBS Lett. 331*, 217−222.
31. Hvidt, A., and Nielsen, S. O. (1966) *Adv. Protein Chem. 21*, 287−386.
32. Englander, S. W., and Kallenbach, N. R. (1984) *Q. Rev. Biophys. 16*, 521−655.
33. Gueron, M., Charretier, E., Hagerhorst, J., Kochoyan, M., Leroy, J. L., and Moraillon, A. (1990) in *Structure & Methods* (Sarma, R. H., and Sarma, M. H., Eds.) Vol. 3, pp 113−137, Adenine Press, New York.
34. Nonin, S., Leroy, J.-L., and Gueron, M. (1995) *Biochemistry 34*, 10652−10659.
35. Gueron, M., Kochoyan, M., and Leroy, J. L. (1987) *Nature (London) 328*, 89−92.
36. Weast, R. C. (1987) in *CRC Handbook of Chemistry and Physics*, CRC Press, Boca Raton, FL.
37. Gueffroy, D. E. (1975) in *CALBIOCHEM: Buffers*, Calbiochem-Novabiochem Co., San Diego, CA.
38. Ts'o, P. O. P. (1974) *Basic Principles in Nucleic Acid Chemistry*, Academic Press, New York.
39. Moe, J. G., and Russu, I. M. (1992) *Biochemistry 31*, 8421−8428.
40. Folta-Stogniew, E. J., and Russu, I. M. (1994) *Biochemistry 33*, 11016−11024.
41. Gueron, M., and Leroy, J. L. (1995) *Methods Enzymol. 261*, 383−413.
42. Cain, R. J., and Glick, G. D. (1998) *Biochemistry 37*, 1456−1464.
43. Buchner, P., Maurer, W., and Ruterjans, H. (1978) *J. Magn. Reson. 29*, 45−63.
44. Nonin, S., Leroy, J.-L., and Gueron, M. (1996) *Nucleic Acids Res. 24*, 586−595.
45. Benight, A. S., Schurr, J. M., Flynn, P. F., Reid, B. R., and Wemmer, D. E. (1988) *J. Mol. Biol. 200*, 377−399.
46. Leroy, J. L., Gehring, K., Kettani, A., and Gueron, M. (1993) *Biochemistry 32*, 6019−6031.

BI010890A