Structural Studies on the Hairpins at the 3′ Untranslated Region of an Anthrax Toxin Gene[†,‡]

Patrick R. Shiflett,[§] Kirsten J. Taylor-McCabe,[§] Ryszard Michalczyk,[||] Louis A. "Pete" Silks,[||] and Goutam Gupta*[,§]

McClintock Resource, Group B1, Bioscience Division, Mail Stop M888, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, and NIH National Stable Isotope Resource, Szilard Resource, Group B3, Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

Received January 22, 2003; Revised Manuscript Received March 27, 2003

ABSTRACT: Three proteins, namely, protective antigen (PA), edema factor (EF), and lethal factor (LF), encoded by the pX01 plasmid of *Bacillus anthracis* play a major role in the pathogenesis of target host cells. PA combines with EF and LF to form bipartite PA-EF and PA-LF toxins and facilitates intracellular delivery of EF and LF both of which cause cytotoxicity to the host. Since the level of PA is crucial to pathogenesis by anthrax toxins, it is important to understand how the host environment regulates the expression of the PA (or pagA) gene by utilizing the 5′ and 3′ untranslated regions (UTR). The 5′ UTR sequence determines the initiation of transcription, whereas the 3′ UTR sequence determines the efficient termination and stability of the transcript. Although, the role of the 5′UTR sequence of pagA has been investigated, little is known about the role of the 3′ UTR. Since hairpin formation at the 3′UTR of a gene is an established mechanism for efficient termination and stability of the transcript, we carried out structural studies, including gel electrophoresis, circular dichroism, and two-dimensional nuclear magnetic resonance spectroscopy, to determine whether the 3′ UTR sequences of pagA also form hairpin structures. Our results unequivocally demonstrate that both the coding and the noncoding 3′ UTR sequences form stable hairpin structures. It is quite likely that the hairpins at the 3′UTR may contribute to efficient termination and stability of the pagA transcript.

Three proteins of *Bacillus anthracis* play a major role in the pathogenesis of target host cells. The proteins, namely, protective antigen (PA, 83 kDa),[1] edema factor (EF, 89 kDa), and lethal factor (LF, 90 kDa), are all encoded by the pX01 plasmid and combine to form two separate toxins (*1*). These proteins act in a bipartite system where PA combines with EF to form edema toxin and separately with LF to form lethal toxin. PA, common to both the bipartite systems, initiates the interaction with the host cell by binding to a surface receptor (*2*). After PA binds the host cell surface receptor it is cleaved by the host furin protease to release a N-terminal 20 kD fragment and retain a 63 kDa fragment with an exposed EF or LF binding site. This 63 kDa fragment binds to EF or LF and aggregates to form a heptameric channel, which allows cytosolic delivery of EF and LF. EF is a calmodulin-dependent adenylate cyclase, and LF is a zinc dependent metallo-protease. These toxins act collectively on host target proteins to cause the onset of pathogenesis (*1*).

Since PA is crucial to toxin pathogenesis, it is important to understand how the host environment induces or regulates the expression of the PA (or pagA) gene. pagA, like any other bacterial gene, appears to be regulated by the promoter at the 5′ untranslated region (5′UTR) that initiates transcription and the signal at the 3′ untranslated region (3′UTR) that guarantees efficient transcription termination and mRNA stability. A systematic analysis of pagA promoter revealed the involvement of environmental factors such as CO_2 and the transcription activator protein, atxA, in the initiation of pagA transcription (*3*, *4*). However, very little is known about the role of the 3′UTR in the regulation of pagA. Fortunately, the complete sequencing of pX01 (*5*) gives us insight into the functional role of the 3′UTR of pagA. A 47 nucleotide stretch at the 3′UTR of pagA seems to be very unique to pX01 and appears to have several features, characteristic of a *rho*-independent (or intrinsic) transcription terminator (*6*−*10*), as summarized in Table 1 for various confirmed transcription terminators. First, it is proximal to the stop codon (i.e., 43 nucleotides downstream of pagA) like the majority of the established (or putative) intrinsic terminator shown in Table 1. Second, this sequence, TTAAAACATACTCTTTTTGTAAGAAATACAAGGAGAGTATGTTTTAA, as shown in Figure 1 tends to form a stable RNA or DNA hairpin with the underlined loop segment as determined by ΔG of hairpin formation. Third, the hairpin has a single-stranded poly-dT or poly-rU tail, which may act as an accessory factor in transcription termination. Finally, the noncoding strand (i.e., complementary to the mRNA strand) can also form a hairpin, although the structure/stability of

[†] This work was supported by the National Stable Isotope Resource (NIH/NCRR-RR02231) and DOE (KP-04-01-00-0) and LDRD-DOE grant on host−pathogen interactions.

[‡] Coordinates have been deposited at the Protein Data Bank (PDB) under 1NGO (110C) and 1NGU (110NC).

* To whom correspondence should be addressed. E-mail: gxg@lanl.gov. Tel.: (505) 663-5118. Fax: (505) 663-5150.

[§] McClintock Resource.

[||] NIH National Stable Isotope Resource.

[1] Abbreviations: PA, protective antigen; EF, edema factor; LF, lethal factor; UTR, untranslated region; DQF COSY, double quantum filtered correlation spectroscopy; NOESY, nuclear Overhauser effect spectroscopy; CD, cicrcular dichroism; T_m, melting temperature; kDa, kiloDalton; HMQC, hetero-nuclear multiple quantum coherence.

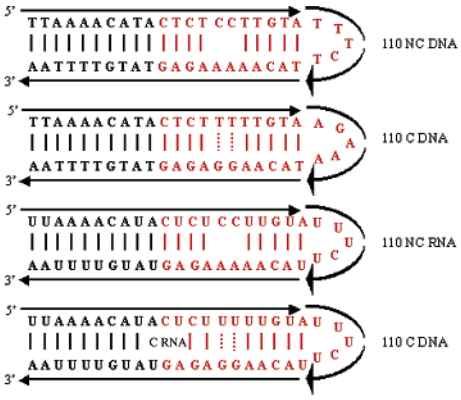
10.1021/bi034128f CCC: $25.00
Published on Web 05/03/2003

Table 1: Sequence and Structure of the 3′UTR Hairpins with Established Termination Activity

Organism	Gene	Location[a]	Hairpin	ΔG[b]	Ref.
GAS	aad9	0	A A A A A T T G A A A A A A GT G T G G A T T T T T A A C T T T T T T __ C A C A A	-14.8/-17.5 -14.1/-16.3	36
V. cholerae	crp	27	A A A A A T A G C C A C C C T T T T T T A T C G G T G G C G	-12.5/-17.2 -12.9/-17.1	37
V. cholerae	ctxB	5	A A A A A A G C C A C C T T T T T T T C G G T G A C	-10.5/-12.5 -10.4/-12.1	38
λ Phage	int	280	T T A G C G C A A G G T G A T T A A T C G C G T T C T T C T G T T	-9.7/-17.6 -9.8/-13.3	39
E. coli	crp	11	T G G C G C G T T A C C T A C C G C G C G A T G G	-10.5/-17.5 -8.8/-13.0	40
N gonorrhoeae	mtrR	17	C A A A C A T G C C G T C T G A A A A T A C G G C A A A C T T T C C C A A	-8.3/-12.2 -9.9/-18.3	41

[a] Location refers to the distance from the stop codon of the mentioned gene. Δ*G* values are computed for each sequence using *mfold* (*14*−*16*) algorithm. [b] The values on top refer to the DNA and RNA hairpins of the coding mRNA strand, whereas the values at the bottom refer to the DNA and RNA hairpins of the noncoding template strand.

A



B

Hairpin	ΔG (kcal/mol)
27mer	
pXO1-110 (DNA) Coding/Non-Coding	-4.2/-3.3
pXO1-110 (RNA) Coding/Non-Coding	-10.2/-6.4
47mer	
pXO1-110 (DNA) Coding/Non-Coding	-14.3/-14.2
pXO1-110 (RNA) Coding/Non-Coding	-23.6/-19.2

FIGURE 1: (A) Structures of DNA and RNA hairpins studied. (B) Free energy values. DNA or RNA *mfold* (Michael Zuker, Rensselaer Polytech Institute) and pattern recognition were performed using MOTIF and BLAST on *B. anthracis*: Sterne's strain (contains pXO1 plasmid).

this hairpin appears to be different from that of the coding strand (see Figure 1). The simultaneous formation of the hairpins on the mRNA and template strands may enhance the release of RNA polymerase facilitating efficient transcription termination. This prompted us to initiate structural studies on the 3′ UTR sequence of pagA in an attempt to understand how this sequence and structure may regulate the transcription of this gene.

In this article, we describe our studies on the 3′UTR of the pagA gene (ORF 110) in the pX01 plasmid of *B. anthracis*. We focus on the coding, referred to as 110C, and noncoding, referred to as 110NC, 27 nucleotide sequences colored red in Figure 1. More detailed studies are carried out on the 110C hairpin because the structure and stability of this hairpin may determine the efficient termination through the dissociation of the RNA polymerase complex and the mRNA stability, both of which may influence the regulation of the pagA gene. First, we predict using the computer algorithm *mfold* that a 47-nucleotide stretch at the 3′ UTR of pagA can form a stable hairpin. Second, we demonstrate by gel electrophoresis that the central 27 nucleotide-long DNA or RNA 3′ UTR sequence indeed folds into a hairpin. Third, we determine the stability of the hairpins by circular dichroism. Finally, we carry out structural studies by 2-D NMR on the coding and noncoding DNA strands to determine the conformation of the stem and loop of the hairpins.

NMR studies are carried out on the 110C DNA hairpin sequence and several of its analogues modified at specific nucleotides. DNA sequences were initially chosen for structure determination because of their inherently better spectral resolution over RNA. We are particularly interested in addressing two important questions: (i) whether the two successive G•T mispairs are integral to the stem of the 110 C hairpin and (ii) how the five purines are stacked in the loop. These two structural properties are also relevant for the RNA hairpin. In DNA, the sugar pucker is flexible and covers a range from C3′-endo to O1′-endo to C2′-endo with a bias toward the C2′-endo domain, whereas in RNA the sugar pucker is traditionally confined to the C3′-endo domain

(*11*, *12*). However, the transition from DNA to RNA sugar−phosphate backbone can be achieved without altering the general feature of the two successive G•T mispairs in the stem (i.e., internally stacked vs bulged out). The sugars of the nucleotides in the loop in both DNA and RNA hairpins tend to adopt C2′-endo pucker (*11*, *12*). Thus, the loop stacking in the 110C DNA hairpin may still be relevant in the RNA counterpart. The 110NC DNA hairpin structure in the noncoding or template strand is directly relevant to the release of RNA polymerase complex and efficient termination of transcription. Also, traditionally the difference in structure or stability between the hairpins formed by the coding and noncoding strands defines the directionality of transcriptional termination.

MATERIALS AND METHODS

NMR Samples. DNA was synthesized on a Perkin-Elmer ABI 394 DNA synthesizer using standard phosphoramidite chemistry on a 40 nM scale. Samples were deprotected and purified using reverse phase oligonucleotide purification columns (Annovis OPC catalog number 21-8550-86). After column purification, sample solvent was exchanged on a 3000 molecular weight cutoff microconcentrator (Millipore catalog number 42404 Microcon YM-3). NMR experiments were carried out in 10 mM phosphate buffer with 150 mM NaCl at pH 6.5. DNA concentrations were 1−2 mM. For experiments conducted in H_2O, 10% D_2O was added for the NMR lock signal. NMR samples in D_2O were prepared by several cycles of lyophilization in 99% D_2O in appropriate buffer and salt and finally dissolving the sample in 99.9% D_2O from Cambridge Isotope Laboratories. Site-specifically N^{15}/C^{13} labeled DNA at selected adenine residues was synthesized and purified in the manner described above. ^{13}C8, ^{15}N6, and ^{15}N9 of A12 and A15 of the 110C sequence were labeled, and the labeled phosphoramidites were prepared as described previously (*13*).

The following DNA sequences were examined by NMR: for the 3′-UTR of 110C, the coding strand of pagA (or gene 110 on pX01), 110C DNA 5′ CTCTTTTTGTAAGAAATACAAGGAGAG 3′, 110C RNA 5′ CUCUUUUUGUAAGAAAUACAAGGAGAG 3′, 110C T6C 5′ CTCTTCTTGTAAGAAATACAAGGAGAG 3′, 110C A12T 5′ CTCTTTTTGTATGAAATACAAGGAGAG 3′, and 110C A16C 5′ CTCTTTTTGTAAGAACTACAAGGAGAG 3′ and for the 3′-UTR sequences on 110NC, the noncoding strand of pagA (or gene 110 on pX01) 110NC DNA 5′ CTCTCCTTGTATTTCTTACAAAAAGAG 3′ and 110NC RNA 5′ CUCUCCUUGUAUUUCUUACAAAAAGAG 3′.

Computational Studies. DNA or RNA *mfold* (*14*−*16*) and pattern recognition were performed using MOTIF and BLAST on the 3′UTR of pagA on the pX01 plasmid of *B. anthracis* Sterne's strain. The sequence extending 150 base pairs downstream of pagA was modeled. Stem−loop structures were identified, and the ΔG for each structure was estimated by a folding algorithm *mfold* that uses free energy parameters from the laboratory of SantaLucia, Jr. (*16*). Parameters used as input for the folding were as follows: the DNA sequence is linear; folding temperature 37 °C; ionic conditions: $[Na^+]$ 1 M and $[Mg^{2+}]$ 0 M; correction type: oligomer; upper bound on the number of computed foldings: 50; window parameter: 0 for 27mer and 1 for 47-mer; and maximum distance between paired bases: no limit.

Electrophoresis Studies. DNA and RNA sequences were run on a nondenaturing 15% polyacrylamide gel (BioRad 30%Acrylamide: Bis 29:1 #161-0156). The following samples were run: 110NC DNA, 1.5 μL, 46 pmol, 110C DNA, 3.5 μL, 58 pmol, 110NC RNA, 1.2 μL, 212 pmol, 110C RNA, 1.2 μL, 260 pmol, 15 mM KCl in all samples, ladder = 10 bp, 1.5 μL, PROMEGA G447A, gel loading dye: 6X Blue/Orange, Promega #G190A. The gels were run at 50V in 0.5X TBE DEPC treated water using a BioRad Mini-Protean II cell and stained with ETBR (5 μL, 10 mg/mL stock).

Melting Studies by Circular Dichroism. Thermal stability of the hairpins was determined using circular dichroism experiments on a Jasco J700 attached to a temperature controlled water bath. 1 mM DNA samples were placed in a cell of 1 mm path length with 10 mM phosphate buffer and 20 mM $MgSO_4$ at pH 7.4. Multiple scans of each sample were run with 1 nm wavelength resolution and 200 mdeg sensitivity, 1 nm bandwidth, and 1 s response times.

NMR Experiments. All NMR experiments were performed on a 500 MHz Brüker Avance spectrometer equipped with a proton/carbon/nitrogen triple resonance probe with triple axis gradients and processed using Felix 98 (Biosym) on an SGI Oxygen workstation. All 2-D experiments were processed with an 80° shifted sinebell squared window function. No solvent suppression or baseline corrections were included in data processing. Experiments were performed in H_2O or D_2O solvents, as indicated. The 110NC strand experiments in D_2O were performed at 5 °C, and the 110C strand experiments in D_2O were performed at 27 °C. 110NC and 110C experiments in H_2O solvent were performed at 5 °C.

DQF COSY experiments were run with a 5000 Hz sweep width, taking 4096 points in t_2 and 1024 points in t_1. Water suppression was achieved through a 0.5 s presaturation at low RF field strength (γB_1 = 15 Hz), and the double quantum filter was achieved through phase cycle (*17*). Quadrature detection in T1 was achieved using the TPPI method (*18*). TOCSY experiments were run with a 5000 Hz sweep width, taking 4096 points in t_2 and 512 points in t_1. A 75 ms spin lock using a DIPSI2 sequence was utilized for Hartman Hahn transfer. Quadrature detection in t_1 was achieved using States-TPPI (*19*). NOESY experiments were run with a 5000 Hz sweep width, taking 4096 points in t_2 and 512 points in t_1. NOESY experiments in D_2O were conducted for mixing times 100, 200, and 300 ms. Water suppression was achieved through a 0.5 s presaturation pulse with weak RF field (γB_1 = 15 Hz), and quadrature detection in t_1 was achieved using the States-TPPI method (*19*). WATERGATE NOESY experiments in H_2O were run with a 12 000 Hz sweep width, taking 8192 points in t_2 and 512 points in t_1 with a mixing time of 150 ms. Water suppression was achieved using a (3-9-19)-pulse sequence with gradients (*20*, *21*). Data acquisition was done in phase sensitive mode using TPPI (*18*). HMQC ^{13}C-^{1}H experiments for 110C were conducted with the following parameters: ^{1}H dimension sweep width, 9.0 ppm, 2048 points and ^{13}C dimension sweep width, 3.0 ppm, 96 points. Carrier frequency was set at 138.7 ppm in the ^{13}C dimension and on water resonance in the ^{1}H dimension. Acquisition was in a phase sensitive mode using the TPPI method (*18*).

NMR Restraints and Structure Refinement. Distance restraints were generated from NOESY data of 110C and

110NC using three separate mixing times: 100, 200, and 300 ms. NOESY cross-peak volumes were measured as a function of mixing time and fit to a second-order polynomial curve to determine the initial rate of NOE buildup. The H5−H6 distance of 2.5 Å in cytosines was used a reference to estimate distances from the NOE intensities. The upper and lower bounds were set, respectively, at +10% or −10% of the estimated distances. NOE derived experimental distance restraints were subjected to force constants of 2 kcal/mol/Å^2, and hydrogen bond distance constraints were subjected to force constants of 50 kcal/mol/Å^2.

All modeling was done on an SGI Oxygen workstation. The initial models for both 110C and 110NC were generated in nucgen, a utility program included in Amber 4.1 (*22*). Nucgen was utilized to create canonical B-form DNA coordinates for the stems of 110C and 110NC. Loop regions were generated as double-stranded canonical B-form (for 110C) or A-form (for 110NC) DNA in nucgen, separated from their partner strand and then constrained on the O5′ and O3′, 16.5 Å apart and subjected to a minimization. All minimization and molecular dynamics were performed using SANDER module/Amber 4.1 (*22*). The minimized loop was positioned upon the stem and subjected to a restrained minimization to create the initial ($t = 0$) model for molecular dynamics. NOE and other restraints are outlined for 110C and 110NC in Tables 2 and 3. Minimization parameters include a dielectric constant of 80 to simulate bulk solvent, a 10 Å nonbonded interaction cutoff, scaling for 1−4 C interactions set at 1.2 (as per dat94 DNA parameter database), and termination when the energy gradient falls below 0.01. A minimized hairpin structure was used as $t = 0$ configuration in a 120 ps molecular dynamics at constant (300 K) temperature. All the constraints and other parameters were same as those used for initial energy minimization to obtain the $t = 0$ configuration. The first 20 ps segment was enough for thermal equilibration. 100 snapshots (1 after every ps) from the 21−120 ps trajectory were extracted from the trajectory using the CARNAL program in the Amber 4.1 package. The individual snapshots and the average of the 100 snapshots were subjected to the restrained minimization described above.

RESULTS

A nucleic acid folding algorithm, *mfold*, (*14*−*16*) that predicts DNA secondary structure was used to identify possible stable hairpin formation in the 3′UTR of the pagA gene in *B. anthracis* Sterne's strain pXO1 plasmid. Stable stem−loops were identified for 110C and 110NC. Both the structures contained 42 base pairs in the stem and five bases in the loop (Figure 1). In this work, we have carried out structural studies on the central 27 nucleotide-long DNA hairpins (colored red in Figure 1). On a nondenaturing polyacrylamide gel, a single-stranded DNA hairpin with 2*N* nucleotides runs as the duplex with *N* base pairs plus migration effects from the loop (*23*, *24*). Figure 2 shows the mobility of 110C and 110NC DNA on a nondenaturing polyacrylamide gel relative to several duplex markers of defined lengths. 110C and 110NC DNA migrate around an 18−19 base pair duplex, suggesting that both of them adopt stable hairpin folds as shown in Figure 1. Figure 2 also shows the migration of 110C and 110NC RNA on the same nondenaturing gel. The migration of the two RNA sequences

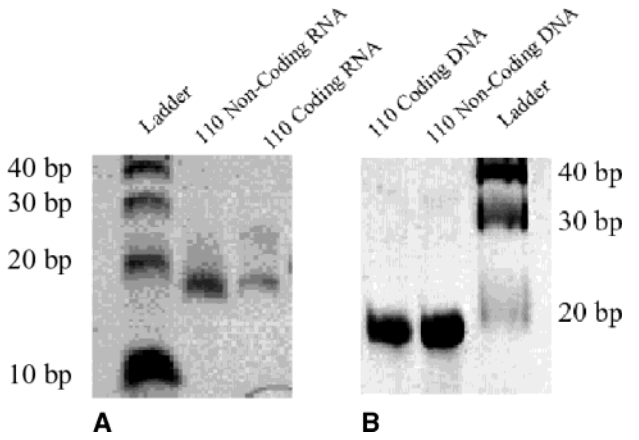


FIGURE 2: Nondenaturing gel (A) RNA and (B) DNA samples. Putative DNA stem−loop structures were run on 15% polyacrylamide gels and EtBr stained. 110 stem−loop structures were 27 nucleotides long and therefore should migrate as a shorter fragment when folded.

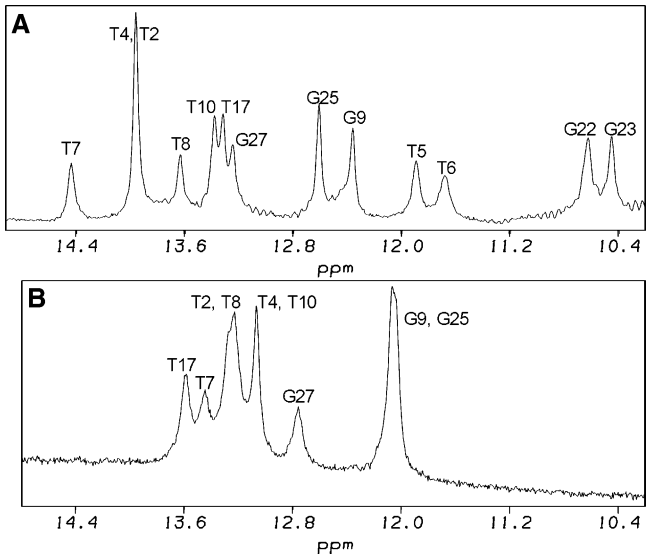


FIGURE 3: One-dimensional (1-D) NMR spectra of imino protons of 110C (A) and 110NC (B) DNA at 5 °C. Solution conditions: 2 mM DNA, 10 mM phosphate buffer, 150 mM NaCl, pH 6.5, 10% D_2O, 90% H_2O.

indicates a stable hairpin but migrate faster than their DNA counterparts. Thermal stability measurements by circular dichroism also confirm that both 110C and 110NC DNA are stable at room temperature. However, judging by the difference in their melting temperatures, T_m (i.e., 50 °C for 110C and 42 °C for 110NC), the 110C hairpin appears to be more stable than the 110NC hairpin. As discussed below, purine versus pyrimidine stacking in the loop and/or two H-bonded G•T versus single or no H-boned A•C mismatch in the stem (see Figure 1) may account for this difference in the stability of the two hairpins.

Figure 3A,B shows the 1-D NMR spectra of the exchangeable protons of 110C and 110NC DNA at 5 °C at pH 6.5. The presence of two G•T pairs is evidenced in 110C by the appearance of imino proton pairs in the 12−10 ppm range. Substitution of T6 by C6 causes the disappearance of the imino pair at (11.7, 10.7 ppm) and appearance of an imino signal at 12.6 ppm (see Figure 1s in the Supporting Information). These changes are consistent with a substitution of G22•T6 by G22•C6 because G-imino proton in Watson−

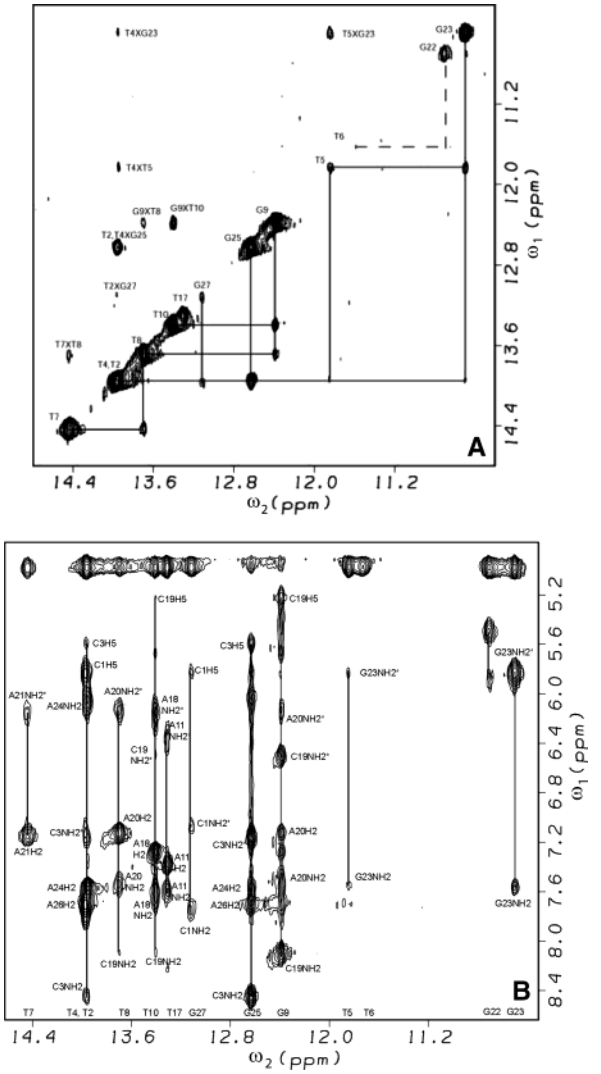


FIGURE 4: 2-D spectra depicting exchangeable proton correlations of the 110C sequence at 5 °C. The resonance assignments are marked. The spectrum is observed with a WATERGATE-NOESY pulse sequence with 150 ms mixing time. Solution conditions are detailed in Figure 3: (A) Exchangeable imino−imino correlations showing base pairing in the stem and sequential correlations in the strands. (B) Imino−amino, H2 (adenine), and H5 (cytosine) correlations.

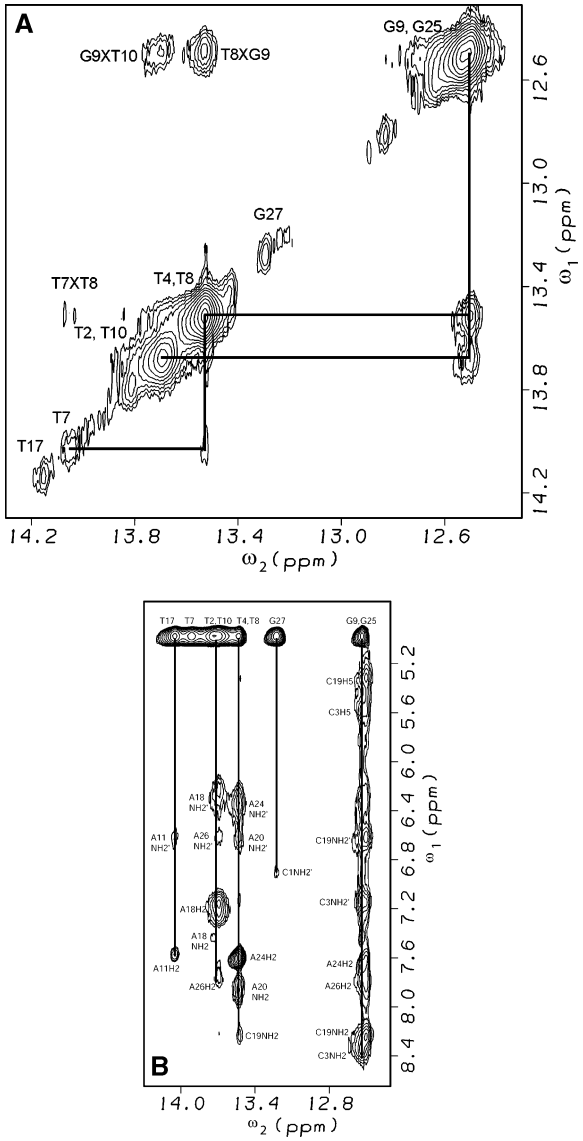


FIGURE 5: 2-D spectra depicting exchangeable proton correlations of the 110NC sequence at 5 °C. The resonance assignments are marked. The spectrum is observed with a WATERGATE-NOESY pulse sequence with 150 ms mixing time. Solution conditions are detailed in Figure 3: (A) Exchangeable imino−imino correlations showing base pairing in the stem and sequential correlations in the strands. (B) Imino−amino, H2 (adenine), and H5 (cytosine) correlations.

Crick G•C is shifted downfield by about 2 ppm from the G-imino proton in a wobble G•T pair (*25*−*28*). Thus, the imino pair at (11.7, 10.7 ppm) is assigned to G22•T6 of 110C, and by default the other imino pair at (11.9, 10.4 ppm) is assigned to G23•T5. Note that signals for G•T pairs are missing from the exchangeable protons of the 110NC 1-D spectrum (Figure 3B), where G•T mispairs are replaced by A•C mispairs devoid of any imino protons at neutral pH. Although A•C mispairs can form a single H-bonded pair involving amino protons (*29*−*31*), no direct evidence is available from our experiments to support the presence of such a pairing. However, as discussed later in the text, the two A•C mispairs seem to display time-averaged internally stacked conformations. In addition to the lack of two base pairs in the stem, 110NC also has pyrimidine stacking in the loop as opposed to purine stacking in 110C (compare the hairpin structures in Figure 1). It has been reported that substitution of the A-rich loop by a T-rich loop in a hairpin does not cause significant difference in stability (*32*). However, several reports indicate that a G•T (or G•U) pair is more stable than an A•C pair in a DNA (or RNA) duplex or in the stem of a hairpin (*33*, *34*). Therefore, it is not surprising that we observe less stability in the 110NC hairpin than the 110C hairpin. Also, note that the imino protons of the 110NC hairpin show broad peaks as well as a chemical shift rearrangement relative to the imino protons of the 110C hairpin (compare hairpins in Figure 1A,B).

A 2-D NOESY cross-section (Figure 4A) shows the connectivity of the imino protons in 110C. Note that imino protons from all base pairs in the stem display NOESY connectivities excluding G22•T6, which appears to be more labile at 5 °C and results in weak imino connectivity between G22•T6 and A21•T7. However, as discussed later, H1′ to H6/H8 and H2″ to H6/H8 inter-nucleotide connectivities indicate that all base pairs, including G22•T6 and G23•T5, are stably stacked in the stem of 110C (see Figure 1). Figure

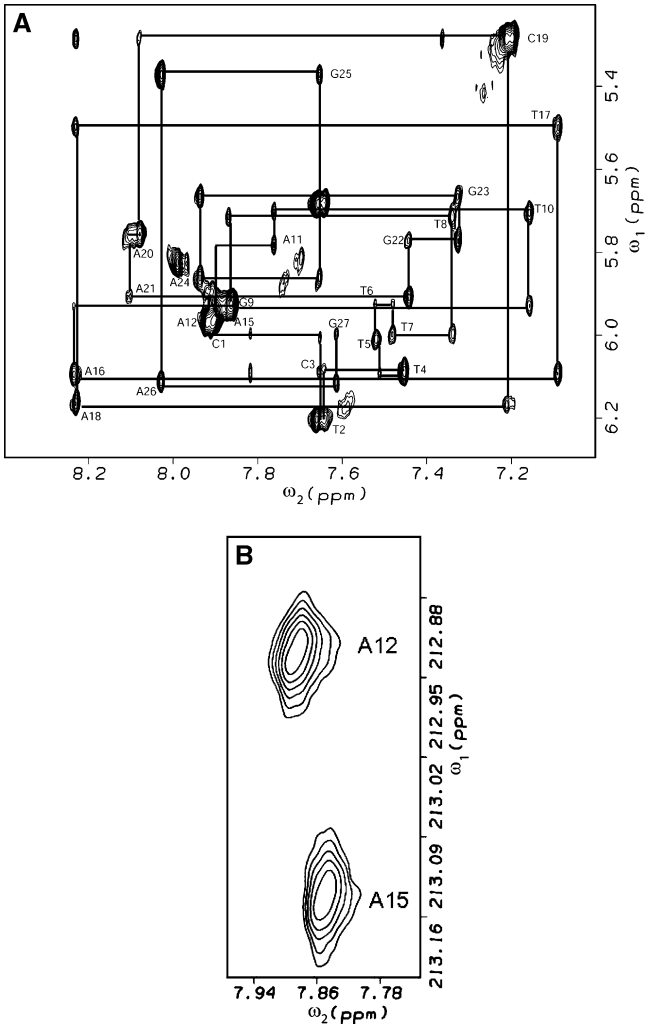


FIGURE 6: (A) 2-D NOESY (200 ms mixing time) depicting correlations between base H6/H8 to sugar H1′ protons for the 110C sequence at 27 °C. Internal correlations and nearest neighbors show the sequential assignment of the 110C sequence. Solution conditions are detailed in Figure 3 with the exception for all D_2O solvent. (B) HSQC spectrum depicting the H8 correlation to the ^{13}C labeled C8 at 27 °C. The A12 and A15 residues of the 110C sequence were labeled to confirm the H8 proton chemical shifts.

4B shows the assignment and NOESY connectivities between the H-bonded imino proton and their in-plane and out-of-plane neighbors. The T-N3H proton of a Watson−Crick A•T pair shows in-plane connectivities with A-H2 and A-N6H2/N6H2′. The G-N1H proton of a Watson−Crick G•C pair shows in-plane connectivities with C-N4H2/N4H2′ and C-H5. The T-N3H proton of a G•T mispair shows in-plane connectivities with G-N1H and G-N2H2/N2H2′. Several out-of-plane NOEs are observed in addition to the in-plane NOEs. N3H protons of T2 and T4 from the Watson−Crick A26•T2 and A24•T4 pairs show NOEs with N4H and H5 of C1 and C3 from the neighboring Watson−Crick G27•C1 and G25•C3 pairs. G-N1H of the Watson−Crick G25•C3 pairs shows NOEs with the A-H2 from the Watson−Crick A24•TT4 and A26•T2 pairs.

Figure 5A shows the NOESY cross-section depicting the connectivities between imino protons from the neighboring Watson−Crick base pairs in the 110NC hairpin at 5 °C. The NOESY connectivities involving imino protons are present on either side of the two A•C mispairs (i.e., G27•C1 to A26•T2 to G25•C3 to A24•T4 and A21•T7 to A20•T8 to

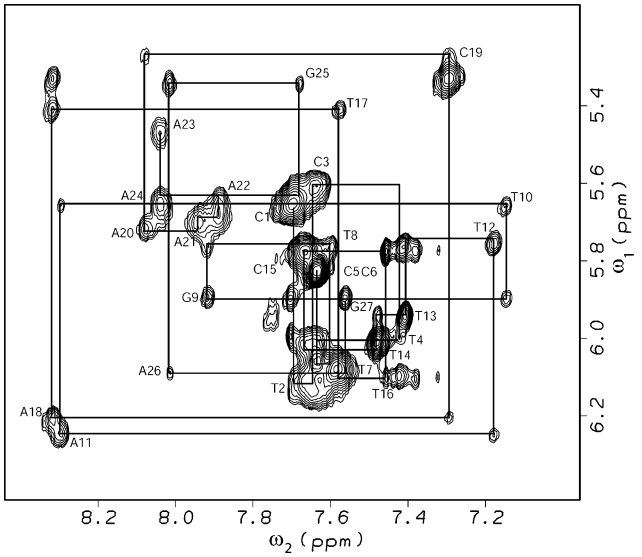


FIGURE 7: 2-D NOESY (200 ms mixing time) depicting correlations between base H6/H8 to sugar proton H1′ protons for the 110NC sequence at 5 °C. Internal correlations and nearest neighbors show the sequential assignment of the 110NC sequence. Solution conditions are detailed in Figure 3 with the exception of 100% D_2O solvent.

G9•C19 to T10•A18 to A11•T17). Note that NOEs involving the imino protons of the terminal G27•C1 and A11•T17 are weak and not seen at the contour level of Figure 5A. However, as discussed later, H1′ to H6/H8 and H2″ to H6/H8 internucleotide connectivities indicate that all the base pairs, including the A22•C6 and A23•C5 mispairs, are stably stacked in the stem of 110NC (see Figure 1). Figure 5B shows the assignment and NOESY connectivities between the H-bonded imino protons and their in-plane and out-of-plane neighbors. Note that only a small number of out-of-plane NOEs, as compared to 110C (Figure 4B), can be discerned because of resonance overlap and line broadening. The imino protons belonging to the core of Watson−Crick paired segments of the stem, such as A26•T2, G9•C19, A18•T10, A20•T10, and G25•C3, show out-of-plane NOEs. However, in this cross-section, out-of-plane connectivities are missing from A24•T4 and A21•T7 neighboring A•C mispairs and from A11•T17 adjacent to the pyrimidine loop. The lack of out-of-plane NOEs from these three base pairs may be attributed to their enhanced flexibility because they are located at the boundary of dynamic (i.e., the two A•C mispairs and the pyrimidine loop) segments of the hairpin structure.

The nonexchangeable protons (i.e., base protons (H6/H8, H2/H5/CH3) and sugar protons (H1′, H2′/H2″, H3′, H4′, H5′/H5″)) are sequentially assigned using TOCSY, DQF-COSY, and NOESY data following the methodology described previously for other DNA hairpins (*23*). Briefly, the sugar spin system (H1′, H2′/H2″, H3′, H4′, H5′/H5″) and base spin system (H6, H5/CH3) are identified from TOCSY or DQF-COSY experiments. The spin systems are sequentially assigned with the aid of NOESY data at various mixing times. H1′(i) to H6/H8(i + 1) NOESY connectivities are most useful in the sequential assignment of the spin systems. Figure 6A shows the H1′ versus H6/H8 NOESY cross-section of 110C in D2O at 27 °C for a mixing time of 200 ms. A continuous H1′(i) → H6/H8 (i + 1) connectivity is observed from C1 through A12 and A15 through G27. The

Table 2: Conformational Parameters of 110C Structure and Its Agreement with NMR Data[a]

	$\alpha^b \pm$ SD	$\beta^c \pm$ SD	$\gamma^d \pm$ SD	$\delta^e \pm$ SD	$\epsilon^f \pm$ SD	$\zeta^g \pm$ SD	$\chi^h \pm$ SD
C1			185.4 0.6	143.6 8.5	181.1 0.66	5.3 0.8	237.8 3.5
T2	292.7 1.4	180.6 1.1	58.9 0.9	133.6 1.3	177.5 1.3	265.9 2.0	250.6 2.1
C3	292.8 2.9	176.2 2.4	57.9 2.2	125.1 4.4	179.1 2.3	264.3 1.1	237.6 3.7
T4	295.7 2.3	178.6 2.6	58.0 1.6	115.6 5.2	179.0 3.1	267.6 1.9	230.4 2.0
T5	295.2 1.7	174.2 2.7	58.9 1.4	123.5 4.6	176.9 1.6	264.8 1.2	244.6 3.3
T6	297.2 3.5	172.7 2.6	60.8 1.8	116.7 4.1	174.4 3.0	265.9 1.5	235.7 2.3
T7	297.2 4.1	175.8 3.4	60.2 3.8	114.4 4.8	173.9 6.1	265.9 2.0	231.7 3.0
T8	287.1 3.0	179.3 4.5	58.8 2.4	126.3 7.6	184.2 2.2	263.0 3.5	239.5 4.0
G9	294.4 2.3	180.0 2.6	53.1 2.1	124.7 4.5	176.7 2.2	261.6 1.7	249.4 2.5
T10	285.8 3.9	177.1 2.0	59.2 1.4	115.2 6.3	194.1 3.5	273.5 2.2	239.1 5.2
A11	302.3 4.5	171.8 1.0	52.3 1.3	138.0 1.5	195.8 2.2	94.9 1.5	281.5 5.0
A12	284.1 4.5	267.9 4.0	82.7 4.8	118.1 3.1	263.6 3.7	290.0 4.2	163.1 80.4
G13	275.0 2.3	156.3 8.1	50.4 3.1	104.0 24.2	185.5 4.0	277.2 6.7	315.9 12.3
A14	299.1 1.8	167.3 2.4	45.4 0.9	118.0 7.7	189.1 16.2	275.2 5.9	332.7 3.3
A15	274.7 4.4	179.9 2.8	59.0 1.5	140.6 2.5	197.9 4.9	276.3 2.8	260.0 13.0
A16	295.3 2.0	175.8 2.2	45.5 0.9	138.9 1.3	175.9 18.4	264.3 2.8	279.1 7.9
T17	290.4 1.5	174.1 8.6	64.0 1.6	134.1 1.7	177.3 1.2	264.0 1.2	250.4 2.9
A18	297.3 1.4	179.2 1.2	53.1 1.5	120.2 3.6	173.4 1.1	268.8 0.5	251.6 1.9
C19	292.9 7.3	174.2 1.3	59.7 1.7	105.3 7.2	186.0 27.0	252.3 28.3	226.6 5.3
A20	294.3 2.9	176.9 8.1	56.5 5.1	132.2 4.0	176.0 3.6	265.4 2.5	251.3 3.4
A21	292.8 3.1	175.8 1.8	58.5 2.6	119.9 6.2	177.1 2.7	260.8 4.1	237.2 4.3
G22	289.4 2.0	182.5 1.9	55.1 3.1	131.8 4.6	180.3 1.4	263.6 1.7	238.2 2.4
G23	291.5 2.2	185.5 2.8	52.6 2.7	129.7 3.7	178.1 2.3	265.2 0.9	247.5 3.5
A24	294.3 1.7	179.1 1.8	55.2 1.5	126.4 3.4	175.9 1.6	266.0 1.0	244.9 2.9
G25	294.4 2.1	179.4 2.8	55.4 2.9	131.4 2.1	175.0 1.6	258.8 1.7	245.4 2.7
A26	291.1 1.0	185.7 2.1	55.2 2.3	134.5 2.8	178.1 1.9	267.9 1.8	245.2 2.1
G27		180.7 3.0	52.9 0.9	130.3 2.1			246.2 2.0

NOE violations		av. min.	20	51	85
RMSD (Å) from the average minimized structure			3.8	3.2	3.5
violations >1.0 Å		:0	0	0	0
violations >0.5 Å and <1.0 Å		:7	8	9	6
violations 0.5 Å		:27	26	29	38
number of NOEs between upper and lower bound		:157	157	153	147
total number of NOEs		:191			
NOEs involving exchangeable protons in the stem	25				
NOEs involving nonexchangeable protons in the stem	153				
NOEs involving nonexchangeable protons in the loop	13				
distance restraints involving hydrogen bonds	50				
dihedral restraints involving δ torsion angles	22				
dihedral restraints involving (C4′ C3′ C2′ C1′) angles	22				

[a] Table gives angle and then standard deviation. Average values and standard deviation of torsion angles for the 100 minimized 110C hairpin structures from the 21−120 ps MD trajectory. Three minimized snapshots (at 22, 51, and 85 ps on the MD trajectory) are further analyzed. Not only are these structures different from the average minimized structure, but also they are different from each other as revealed by the (3 × 3) RMSD matrix below.

	22	51	85
22	1.83	2.52	
51	1.83	1.84	
85	2.52	1.84	

[b] α: C5′ O5′ P O3′. [c] β: C4′ C5′ O5′ P. [d] γ: C3′ C4′ C5′ O5′. [e] δ: O3′ C3′ C4′ C5′. [f] ϵ: P O3′ C3′ C4′. [g] ζ: O5′ P O3′ C3′. [h] χ: O4′ C1′ N1 C2 or O4′ C1′ N9 C4.

H8 chemical shift values of A12 and A15 are confirmed by ^{1}H-^{13}C-HSQC experiments in D_2O on the 110C hairpin with ^{13}C8-labeled A12 and A15 (Figure 6B). NMR spectroscopy of a single nucleotide substitution (A12T) in the 110C hairpin facilitated the identification of G13 and A14 in the loop. This substitution extends the stem through A16•T12 and resolves the overlapping H1′-H8 cross-peak at (5.8, 8.0 ppm) into two cross-peaks. One cross-peak (5.2, 8.0 ppm) is assigned to G13 because it connects to T12 (the substituted nucleotide), whereas the other (5.85, 8.05 ppm) is assigned to A14 because it connects to previously assigned A15 and newly assigned G13 (see Figure 2s in the Supporting Information). Once (H1′, H6/H8) all the nucleotides are assigned, the other sugar protons H2′/H2″, H3′, H4′, and H5′/H5″ are assigned by linking them with the corresponding H1′ using TOCSY (H1′ vs H2′/H2″), (H3′ vs H2′/H2″), and (H3′ vs H4′, H5′/H5″) connections. These assignments are further confirmed using intra- and inter-nucleotide H2′/H2″ to H6/H8 or H3′ to H6/H8 connectivities in NOESY spectra. C-H5s and T-CH3s protons are assigned using H6/H8 versus H5/CH3 DQF-COSY, TOCSY, or NOESY cross-peaks. A-H2 protons are assigned using the NOESY in-plane and out-of-plane cross-peaks involving T-N3Hs of A•T pairs. Figure 7 shows the H1′ versus H6/H8 NOESY cross-section of 110NC in D_2O at 5 °C for a mixing time of 200 ms. Although there are a few overlaps, a continuous H1′ to H6/H8 NOESY connectivity can be obtained for all 27 nucleotides; therefore, all 27 (H1′, H6/H8) can be sequentially

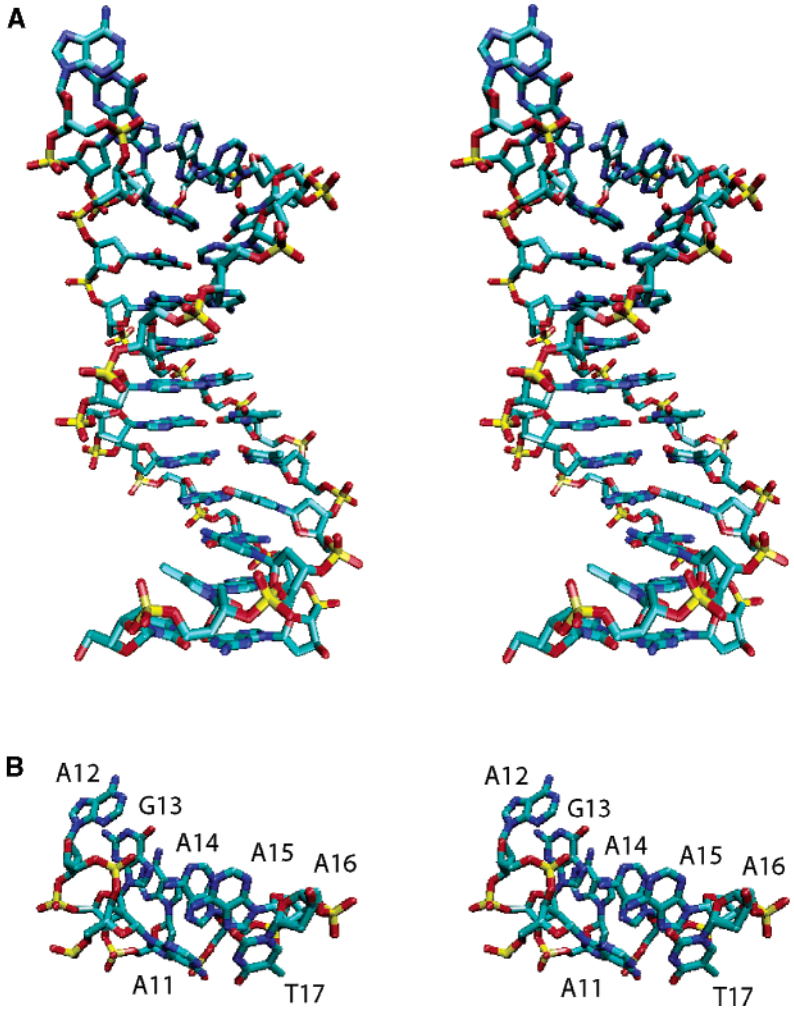


FIGURE 8: (A) Minimized average structure from the molecular dynamics (MD) of the 110C sequence. (B) Loop stacking in the average minimized structure viewed relative to the loop closing A11•T17 pair.

assigned. The rest of the protons in the sugar and base spin systems are obtained by following the procedure as described above for 110C.

The following sets of constraints are obtained from the analysesof NMR data. *Base Pairing in the Stem.* NOESY experiments in water as described in Figures 4 and 5 provide information about the base pairing pattern in the stem of 110C and 110NC. The presence of G-N1H above 12 ppm and in-plane G-N1H to C-N4H2/2′ NOE indicates the presence of a Watson−Crick G•C pair. The presence of T-N3H above 13 ppm and in-plane T-N3H to A-H2 NOE indicates the presence of a Watson−Crick A•T pair. The presence of an imino (T-N3H, G-N1H) within 12−10 ppm and in-plane T-N3H to G-N1H and T-N3H to G-N2H2/2′ NOEs suggest the presence of a two H-bonded wobble G•T pair. The H-bond lengths N···O and N···N are set within 2.7−3.2 Å, and the H-bond angles N−H···O and N−H···N are set within 120−180°. Our data did not provide any direct evidence of H-bonding in the A•C mispairs of 110NC, although they appear to be internally stacked. The atomic sites with like charges, such as A-N6 and C-N4, in A•C mispairs were constrained to be far (>5 Å) from each other.

Nucleotide Geometry. Intranucleotide NOEs for H1′···H6/H8, H2′···H6/H8, and H3′···H6/H8 indicated that all the 27 nucleotides in both the 110C and the 110NC hairpins predominantly displayed (*C2′*-endo, anti) conformations with torsion angles $\delta = 110-160°$ and $\chi = 210-270°$. Intranucleotide NOEs for H1′···H6/H8, H2′···H6/H8, and H3′···H6/H8 monitored as a function of mixing times 100, 200, and 300 ms showed that in general, for the 27 nucleotides, the corresponding distances were 2−2.7, 3−3.8, and 3.7−4.5 Å that are consistent with (*C2′*-endo, anti) conformations and the ranges of (δ, χ) as shown above. Wherever (H1′, H2′/H2″) spin systems could be resolved in the DQF-COSY spectrum, we observed large $J_{H2'\cdots H1'}$ (>10 Hz) and small $J_{H1'\cdots H2''}$ (∼3 Hz) couplings that are also consistent with *C2′*-endo sugar puckers.

Base Stacking at the Stem. Out-of-plane NOEs involving G-N1H, T-N3H, A-C2H, A-N6H2/2′, C-αN4H2/2′ from the NOESY experiments in water (Figures 4 and 5) and internucleotide H1′(*i*)···H6/H8(*i* + 1), H2″(*i*)···H6/H8(*i* + 1), H8/H6(*i*)···H5C/CH3T(*i* + 1), etc. indicated that base pairs are stacked in a B-DNA geometry with the stem helix-axis lying close to the center of the base pairs. Among all, the two internucleotide distances, ΔH2″(*i*)···H6/H8(*i* + 1) within 2.0−2.7 Å and H1′(*i*)···H6/H8(*i* + 1) within 3.0−4.0 Å, are most critical for the assignment of the stem stacking to B-DNA geometry.

As previously described (*35*, *36*), NOESY cross-peak intensities at mixing times 100, 200, and 300 ms are used to obtain distance constraints (i.e., upper and lower limit) for pairwise interactions involving nonexchangeable protons. To

Table 3: Conformational Parameters of 110NC Structure and Its Agreement with NMR Data[a]

	$\alpha^b \pm$ SD	$\beta^c \pm$ SD	$\gamma^d \pm$ SD	$\delta^e \pm$ SD	$\epsilon^f \pm$ SD	$\zeta^g \pm$ SD	$\chi^h \pm$ SD
C1			54.2 0.4	135.3 0.6	181.3 0.9	268.7 1.1	244.1 1.2
T2	295.3 1.1	176.2 1.0	54.7 1.1	129.8 2.2	176.2 1.1	260.6 2.6	248.2 3.2
C3	297.8 2.9	181.9 2.4	55.7 1.2	135.0 3.7	174.5 1.8	263.9 1.6	250.5 2.9
T4	301.2 9.5	177.3 3.1	60.4 3.7	126.3 4.2	183.6 37.4	241.8 42.1	228.4 7.2
C5	291.8 1.7	184.4 15.7	59.6 3.9	137.9 2.2	182.5 3.6	268.8 3.5	245.9 5.0
C6	285.8 2.7	175.9 3.6	51.6 2.1	126.4 6.7	274.6 2.5	129.9 5.7	255.8 12.4
T7	292.5 2.8	125.8 4.3	59.3 2.1	127.9 4.8	185.3 2.9	260.6 2.6	234.5 5.0
T8	268.0 8.6	174.4 2.8	54.3 3.3	144.3 3.0	275.7 23.0	140.4 20.1	264.5 11.6
G9	298.1 2.0	164.4 4.2	47.4 5.1	139.9 2.9	172.1 2.1	266.7 1.2	246.5 3.0
T10	282.0 6.2	185.5 2.0	55.7 1.7	129.9 4.6	188.1 4.5	274.3 3.3	243.2 3.5
A11	154.3 14.7	177.5 2.6	44.7 2.4	126.2 6.9	208.3 6.8	287.8 4.9	265.0 4.5
T12	293.8 0.8	161.1 8.2	53.7 5.3	145.7 2.2	175.9 1.6	268.3 2.7	260.3 3.4
T13	280.1 6.8	171.1 2.4	60.8 3.1	123.6 5.5	262.4 16.4	284.8 6.8	240.2 3.7
T14	258.6 34.3	148.5 5.2	47.4 3.9	141.0 3.7	225.7 45.6	182.5 47.9	305.0 7.2
C15	292.1 3.8	114.8 26.9	50.8 4.2	140.6 3.0	243.0 38.2	215.1 45.6	348.1 8.1
T16	107.5 38.2	154.1 22.5	62.0 4.6	140.6 2.0	182.5 6.6	252.6 13.9	246.6 6.0
T17	288.6 2.8	173.5 4.3	227.1 42.3	148.8 3.2	181.8 3.3	266.3 3.2	253.0 12.0
A18	295.4 1.3	183.9 4.0	48.4 2.2	138.1 2.4	175.5 2.2	256.2 2.9	263.2 5.0
C19	293.9 1.4	182.4 2.4	55.0 1.2	138.5 1.5	178.3 1.3	267.0 1.0	258.2 2.8
A20	279.5 2.9	176.6 0.8	55.3 1.0	139.7 1.7	268.8 3.8	150.1 8.7	265.3 2.2
A21	293.0 2.4	173.3 2.5	42.1 2.9	145.8 1.7	175.6 3.0	262.3 2.6	292.7 8.0
A22	297.8 1.5	166.2 3.6	60.4 2.5	136.8 3.7	166.8 4.6	259.9 7.8	276.0 5.2
A23	297.9 1.9	178.9 3.6	58.6 1.7	136.9 3.6	172.9 1.9	266.3 2.2	250.2 3.4
A24	293.1 3.6	181.0 3.6	58.1 1.7	137.5 3.2	177.2 2.6	261.3 2.8	247.5 3.6
G25	297.4 1.9	184.8 3.6	52.9 2.7	137.8 3.0	173.3 2.3	260.0 7.8	251.0 2.6
A26	290.5 1.9	188.7 2.3	52.2 2.7	141.8 1.6	175.0 2.5	268.1 2.2	254.7 2.4
G27		184.0 2.8	47.7 1.6	138.0 1.3			253.2 3.8

NOE violations		av. min.	63	100	116
RMSD (Å) from the average minimized structure			4.4	4.2	4.6
violations >1.0 Å		:0	0	0	0
violations >0.5 Å and <1.0 Å		:20	23	19	22
violations <0.5 Å		:41	31	35	31
number of NOEs between upper and lower bound		:145	152	152	153
total number of NOEs		:206			
NOEs involving exchangeable protons in the stem	23				
NOEs involving nonexchangeable protons in the stem	131				
NOEs involving nonexchangeable protons in the loop	52				
distance restraints involving hydrogen bonds	46				
dihedral restraints involving δ torsion angles	22				
dihedral restraints involving (C4′ C3′ C2′ C1′) angles	22				

[a] Table gives angle and then standard deviation. Average values and standard deviation of torsion angles for the 100 minimized 110NC hairpin structures from the 21−120 ps MD trajectory. Three minimized snapshots (at 63, 100, and 116 ps on the MD trajectory) are further analyzed. Not only are these structures different from the average minimized structure, but also they are different from each other as revealed by (3 × 3) RMSD matrix below.

	63	100	116
63		1.17	1.89
100	1.17		1.55
116	1.89	1.55	

[b] α: C5′ O5′ P O3′. [c] β: C4′ C5′ O5′ P. [d] γ: C3′ C4′ C5′ O5′. [e] δ: O3′ C3′ C4′ C5′. [f] ϵ: P O3′ C3′ C4′. [g] ζ: O5′ P O3′ C3′. [h] χ: O4′ C1′ N1 C2 or O4′ C1′ N9 C4

this constraint list, appropriate H-bonding constraints are added based upon the WATERGATE-NOESY data described in Figures 4 and 5 and the other torsion and distance constraints as described above. An ensemble of hairpin structures are then obtained for the coding or noncoding sequence using a methodology previously described by us (*35*, *36*). Briefly, the methodology involves the following steps. First, a starting structure is obtained by constrained minimization. Second, the minimized structure is used as $t = 0$ configuration in a 300 K (constant temperature) 120-ps MD simulation. Third, the 100 snapshots (1 after every ps) from last 100-ps of the MD trajectory (i.e., after 20-ps of equilibration) are isolated. Fourth, these 100 snapshots are minimized with constraints to a low gradient. Note that the 100 minimized structures capture the conformational variants of the hairpin that are in agreement with the NMR data. Fifth, an average structure is obtained from the 100 MD snapshot and minimized with constraints. Finally, conformations of the 100 ensemble member and the ensemble average are analyzed as well as their agreement with the NMR data.

110C Hairpin. Figure 8A shows the structure of the 110C DNA hairpin. This structure is obtained from an average of 100 snapshots from the 21−120 ps MD trajectory and then subjecting the average structure to constrained minimization. Note that all the 11 base pairs in the stem are stably stacked including the two G•T mispairs in the middle of the stem. Also, five purines (A12-G13-A14-A15-A16) in the loop are split into two stacked arrangements. The A12-G13-A14 segment forms an independent stack, whereas the A15-A16 segment is stacked on top of T17 of the loop closing

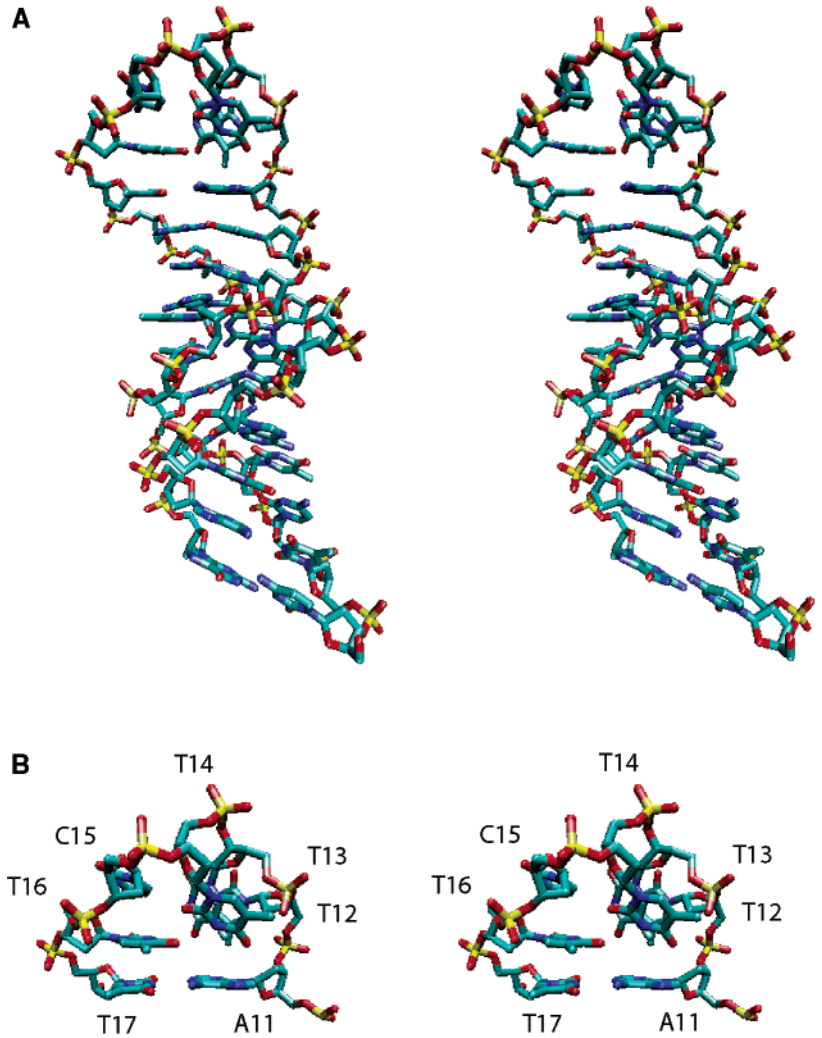


FIGURE 9: (A) Minimized average structure from the molecular dynamics (MD) of the 110NC sequence. (B) Loop stacking in the average minimized structure viewed relative to the loop closing A11•T17 pair.

T17•A11 pair. Figure 8B shows the close-up of the loop stacking with reference to the loop-closing A11•T17. A12 in the loop is quite flexible. It samples two glycosyl torsions: one around syn and the other around low anti. Table 2 shows the torsion angles and their deviations for 27 nucleotides in the 110C DNA hairpin. The sampling using MD and energy minimization is performed to capture conformational variants of the 110C hairpin that show agreement with the NMR data. That different conformational variants can satisfy the same NMR data is illustrated in Table 2 by describing the agreement of three different structures with respect to the NMR data. These three structures are different from each other (RMSD = 1.9−2.5 Å). These minimized snapshots are obtained from time points 22, 51, and 85 ps of the 120-ps MD trajectory. Even though these structures are quite different from the average minimized structure (RMSD > 3Å), they show quite similar agreement with the NMR data. Note that the conformation variation is somewhat localized at the end of the stem and at the loop (especially at A12).

Table 2 shows that the torsion angles of the sugar−phosphate backbone and the glycosyl torsions of the stem nucleotides reflect a canonical B-DNA conformation (i.e., C2′-endo sugar pucker, anti glycosyl torsion, gauche+ C4′-C5′ torsion, trans C5′-O5′ and C3′-O3′ torsions, (gauche−, gauche−) for the phosphodiester bonds). Interestingly, all the loop nucleotides except A12 also display torsion angles similar to the B-DNA sugar−phosphate backbone and glycosyl torsions. A12 shows quite a departure from the rest of the loop nucleotides. The sugar−phosphate backbone is quite different and so is the glycosyl torsion, which as stated above, samples syn and low anti conformations. Because of the paucity of internucleotide constraints in the loop, it is difficult to determine whether the observed peculiarity of A12 is due to actual structural constraint of loop closure or a modeling artifact.

110NC Hairpin. Figure 9A shows the structure of the 110NC DNA hairpin. This structure is obtained from an average of 100 snapshots from the 21−120 ps MD trajectory and then subjecting the average structure to constrained minimization. Note that all the 11 base pairs in the stem are stably stacked. Note that in all these structures, although the two A•C mispairs do not break the continuity of the stem, they introduce a great deal of flexibility in and around them. This feature may originate from the fact that the A•C mispairs lack H-bond or at most have only one at neutral pH. The torsion angles of 110NC DNA hairpin as shown in Table 3 also document flexibility and departure from the canonical B-DNA conformation for some of the torsions belonging to the (T4-C5−C6-T7) stretch; note that C5 and

C6 are in involved in A•C mispairs. The rest of the stem nucleotides display canonical B-DNA conformation. The five pyrimidines (T12-T13-T14-C15-T16) in the loop show considerable conformational variations. T13, T14, and C15 appear to be most flexible as evidenced by larger deviations around some torsion angles in the T13-T14-C15 stretch. Also, the phophodiester torsions linking C15 and T16 show a guache+, trans conformation, which is a departure from the canonical B-DNA. Again, because of the paucity of internucleotide constraints in the loop, it is difficult to determine whether the observed peculiarity of the pyrimidine-rich loop is really an actual structural feature of the loop or a modeling artifact. Figure 9B shows the close-up of the loop stacking with reference to the loop-closing A11•T17 base pair. It appears that T12-T13-T14 and C15-T16 stacks are present in the average structure of the loop. Table 3 shows the torsion angles and their deviations for 27 nucleotides in the 110NC DNA hairpin. The conformational variation of the 110NC hairpin is pronounced at and around the two A•C mispairs, at the end of the stem, and at the loop. The agreement of different structures with respect to the NMR data is described in Table 3 using three conformational variants of the 110NC hairpin. These three structures are different from each other (RMSD = 1.2−1.9 Å). These minimized snapshots are obtained from time points 63, 100, and 116 ps of the 120-ps MD trajectory. Even though these structures are quite different from the average minimized structure (RMSD > 3Å), they show quite similar agreement with the NMR data.

DISCUSSION

Intrinsic or rho-independent transcription terminators are defined as the 3′UTR gene sequences that are capable of forming stable hairpins. Detailed computer analysis reveals that the terminator hairpins at 3′UTRs fall into a wide range of categories (*37*−*42*). They appear with or without the 3′ poly-dT or poly-rU tail. Two hairpins may occur in succession to form an L-shaped structure at the 3′UTR of the transcript, or two hairpins may form on complementary strands and face each other. These stem−loop structures are called factor-independent or intrinsic terminators because their presence alone enables transcription termination in vitro, and their efficiencies are generally enhanced in vivo by the presence of effector proteins that specifically bind to the hairpins (*43*). DNA hairpins at the transcription termination sites may also be utilized for pause or arrest of replication that is initiated at a distal site (*44*). Note that the replication origin in pX01 is distant from pagA.

Although the role of 3′UTR hairpin structures is experimentally verified (see Table 1), there is a debate as to the necessity of the 3′ poly-dT or poly-rU tail for efficient termination (*42*). The argument favoring the necessity of the 3′ poly-dT or poly-rU tail is based upon the unusual structure of the poly-rU•poly-dA hetero-duplex and its poor stability, which causes RNA polymerase complex dissociation and subsequent termination of transcription right before the poly-rU tail. In fact, for some terminators the poly-dT or poly-rU tail has been proven to be necessary. However, there are instances as well where their role is far from being clear. Experimental data on all confirmed terminators indicate that the hairpin structure is critically important in termination and mRNA stability of the associated gene. The hairpin also dictates the regulation of the downstream gene. Determination of the hairpin structures for the coding and noncoding strands serves various purposes. It helps us to determine the changes in structure and stability caused by large changes in sequence simultaneously in the stem and in the loop. It enables us to examine whether the hairpin formation is directional in terms of stability. It also offers some clues whether formation of hairpin on both the mRNA and the template facilitate more efficient dissociation of the RNA polymerase complex.

In this paper, our focus has been solely on the 3′UTR hairpin structure of pagA. Although the hairpin at the 3′UTR of pagA can be predicted by various pattern recognition algorithms or on the basis of stability (ΔG) criteria, experimental determination or verification of the actual hairpin structure helps to suggest that it may play the role of a transcription terminator (*37*−*42*). As discussed in the introductory paragraphs, the 47 nucleotide-long stem−loop located in the 3′UTR of the pagA gene satisfies the general features of the terminators listed in Table 1, but it also possess several unique characteristics. It has a long (22 base pair) stem with five purines in the loop. It has two consecutive G•T mispairs in the stem. As discussed in the previous section, we focused on the central 27 nucleotide-long hairpin with a shorter 11 base pair stem and the same five nucleotide loop. We believe this hairpin forms the core, and the longer hairpin is merely an extension of the stem with all Watson−Crick pairs. For the RNA hairpins, we have proven that it has the gross morphology of a hairpin by gel electrophoresis study (*36*). For the 110C DNA hairpin, in addition to gel electrophoretic studies, we have performed 2-D NMR studies on the original sequence and three analogues modified at single sites. These experiments unequivocally demonstrate that the 3′UTR 110C DNA sequence indeed forms a hairpin with two G•T mispairs in the stem and five purines in the loop. We also observe that the two G•T mispairs are internally stacked as opposed to being bulged out. Note that bulged bases are more susceptible to the endonucleolytic activity of nucleases. The purines, AGAAA, in the loop also show stable stacking. We have also completed gel electrophoresis and 2-D NMR studies on the 27 nucleotide-long hairpin formed by the noncoding strand, 110NC (see Figure 1). In this hairpin, the G•T mispairs are replaced by A•C mispairs, and the purine-rich loop is replaced by the pyrimidine-rich loop. The combined results from the 110C and 110NC structure analysis indicate that the vast differences in the loop and mismatches in the stem cause small but finite changes in overall stability.

In summary, the structural properties of the hairpin at the 3′UTR of pagA reported in this article may be useful in designing experiments that will provide biological proof that the hairpin indeed acts as a transcription terminator.

ACKNOWLEDGMENT

We would like to thank Linda S. Thompson for supplying the unlabeled DNA sequences used in this study.

SUPPORTING INFORMATION AVAILABLE

Two 2-D spectra. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Lacy, D. B., and Collier, R. J. (2002) *Curr. Top. Microbiol. Immunol. 271*, 61−85.

2. Bradley, K. A., Mogridge, J., Mourez, M., Collier, R. J., and Young, J. A. T. (2001) *Nature 414*, 225−9.
3. Dai, Z. H., Sirard, J. C., Mock, M., and Koehler, T. M. (1995) *Mol. Microbiol. 16*, 1171−81.
4. Koehler, T. M., Dai, Z. H., and Kaufman-Yarbray, M. (1994) *J. Bacteriol. 176*, 586−95.
5. Okinaka, R. T., Cloud, K., Hampton, O., Hoffmaster, A. R., Hill, K. K., Keim, P., Koehler, T. M., Lamke, G., Kumano, S., Mahillon, J., Manter, D., Martinez, Y., Ricke, D., Svensson, R., and Jackson, P. J. (1999) *J. Bacteriol. 181*, 6509−15.
6. Henkin, T. M., and Yanofsky, C. (2002) *Bioessays 24*, 700−7.
7. Henkin, T. M. (2000) *Curr. Opin. Microbiol. 3*, 149−53.
8. Richardson, J. P. (1993) *Crit. Rev. Biochem. Mol. Biol. 28*, 1−30.
9. Holmes, W. M., Platt, T., and Rosenberg, M. (1983) *Cell 32*, 1029−32.
10. von Hippel, P. H., Rees, W. A., and Wilson, K. S. (1995) *Nucleic Acids Symp. Ser. 33*, 1−4.
11. Zidek, L., Stefl, R., and Sklenar, V. (2001) *Curr. Opin. Struct. Biol. 11*, 275−81.
12. Shen, L. X., Cai, Z. P., and Tinoco, I., Jr. (1995) *FASEB J. 9*, 1023−33.
13. Orji, C. C., Michalczyk, R., and Silks, L. A., III. (1999) *J. Org. Chem. 64*, 4685−9.
14. Mathews, D. H., and Turner, D. H. (1999) *Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide.*, Vol. 11−43, Kluwer Academic Publishers, New York.
15. Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999) *J. Mol. Biol. 288*, 911−40.
16. SantaLucia, J., Jr. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 1460−5.
17. Derome, A. E., and Williamson, M. R. (1990) *J. Magn. Reson. 88*, 177−85.
18. Marion, D., and Wuthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967−74.
19. Marion, D., Ikura, M., Tschudin, R., and Bax, A. (1989) *J. Magn. Reson. 85*, 393−9.
20. Sklenar, V., Piotto, M., Leppik, R., and Saudek, V. (1993) *J. Magn. Reson. Ser. A 102*, 241−5.
21. Piotto, M., Saudek, V., and Sklenar, V. (1992) *J. Biomolec. NMR 2*, 661−6.
22. Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., Ferguson, D. M., Seibel, G. L., Chandra Singh, U., Weiner, P. K., and Kollman, P. A. (1995) University of California, San Francisco, San Francisco.
23. Mariappan, S. V. S., Garcia, A. E., and Gupta, G. (1996) *Nucleic Acids Res. 24*, 775−83.
24. Mariappan, S. V. S., Catasti, P., Chen, X., Ratliff, R., Moyzis, R. K., Bradbury, E. M., and Gupta, G. (1996) *Nucleic Acids Res. 24*, 784−92.
25. Patel, D. J., Kozlowski, S. A., Ikuta, S., and Itakura, K. (1984) *Fed. Proc. 43*, 2663−70.
26. Allawi, H. T., and SantaLucia, J., Jr. (1997) *Biochemistry 36*, 10581−94.
27. Allawi, H. T., and SantaLucia, J., Jr. (1998) *Nucleic Acids Res. 26*, 4925−34.
28. McDowell, J. A., and Turner, D. H. (1996) *Biochemistry 35*, 14077−89.
29. Sarma, M. H., Gupta, G., and Sarma, R. H. (1987) *Biochemistry 26*, 7707−7714.
30. Gao, X. L., and Patel, D. J. (1987) *J. Biol. Chem. 262*, 16973−84.
31. Boulard, Y., Cognet, J. A., Gabarro-Arpa, J., Le Bret, M., Carbonnaux, C., and Fazakerley, G. V. (1995) *J. Mol. Biol. 246*, 194−208.
32. Chou, S. H., Tseng, Y. Y., and Chu, B. Y. (1999) *J. Mol. Biol. 292*, 309−20.
33. Roongta, V. A., Jones, C. R., and Gorenstein, D. G. (1990) *Biochemistry 29*, 5245−58.
34. SantaLucia, J., Jr., Kierzek, R., and Turner, D. H. (1991) *Biochemistry 30*, 8242−51.
35. Catasti, P., Chen, X., Deaven, L. L., Moyzis, R. K., Bradbury, E. M., and Gupta, G. (1997) *J. Mol. Biol. 272*, 369−82.
36. Mariappan, S. V., Catasti, P., Silks, L. A., III, Bradbury, E. M., and Gupta, G. (1999) *J. Mol. Biol. 285*, 2035−52.
37. Rudd, K. E. (1999) *Res. Microbiol. 150*, 653−64.
38. Lesnik, E. A., Sampath, R., Levene, H. B., Henderson, T. J., McNeil, J. A., and Ecker, D. J. (2001) *Nucleic Acids Res. 29*, 3583−94.
39. Ermolaeva, M. D., Khalak, H. G., White, O., Smith, H. O., and Salzberg, S. L. (2000) *J. Mol. Biol. 301*, 27−33.
40. Bachellier, S., Clement, J. M., and Hofnung, M. (1999) *Res. Microbiol. 150*, 627−39.
41. d'Aubenton Carafa, Y. D., Brody, E., and Thermes, C. (1990) *J. Mol. Biol. 216*, 835−58.
42. Unniraman, S., Prakash, R., and Nagaraja, V. (2002) *Nucleic Acids Res. 30*, 675−84.
43. Bermudez-Cruz, R. M., Chamberlin, M. J., and Montanez, C. (1999) *Biochimie 81*, 757−64.
44. Hyrien, O. (2000) *Biochimie 82*, 5−17.

BI034128F