Journal of Biomolecular NMR, **17:** 1–16, 2000.
KLUWER/ESCOM

Natural abundance heteronuclear NMR studies of the T_3 mini-loop hairpin in the terminal repeat of the adenoassociated virus 2

Shan-Ho Chou[a,b,*], Yu-Yu Tseng[b] & Bi-Yun Chu[a]
[a]*Institute of Biochemistry and* [b]*Chemistry Department, National Chung-Hsing University, Taichung 40227, Taiwan*

Received 10 January 2000; Accepted 13 March 2000

Key words: AAV2 hairpins, heteronuclear NMR, stem-loop interaction, stereospecific H5′ assignment, T_3 loop

Abstract

A DNA hairpin containing a T_3 loop, as occurs in the terminal repeat of a popular gene therapy vector (Adenoassociated Virus 2, AAV2), has been extensively studied using homo- and heteronuclear NMR experiments. Almost complete assignment of the proton and carbon resonances, including $H5'^{(Pro-S)}$ and $H5'^{(Pro-R)}$ protons, has been accomplished at natural abundance. NOESY spectra in H_2O and D_2O have revealed many unusual NOEs, which, when combined with the ϵ, β, γ, and χ torsion angles determined from heteronuclear ^{1}H-^{13}C, ^{1}H-^{31}P, and ^{13}C-^{31}P coupling constants, have allowed for a more detailed picture of the T_3 mini-loop hairpin. The three loop thymidines are all unpaired, yet are highly structured when bracketed by a 5′-GC...GC-3′ stem sequence. The structure determined in this manuscript is considerably different from several other structures reported so far. Contrary to an RNA oligomer with a central U_3 sequence that has the tendency to form a duplex with three U·U mismatches, the d(GAAGC-TTT-GCTTC) sequence exists mostly as a hairpin under millimolar NMR conditions. Since T_3 triloop was found to be an essential element for the site-specific non-homologous integration of the AAV2 virus, and modification of the T_3 loop residue abolishes such capability, the structure we report here may be of biological significance.

Introduction

Hairpins are the fundamental units for higher order nucleic acid structures. They can exist when intra-strand pairing occurs between inverted repeat sequences in double-stranded DNA (Dai et al., 1997, 1998). Such inverted repeats are widespread in the genomes of prokaryotes and eukaryotes, and have been suggested to play a role in the regulation of gene expression (McClellan et al., 1990). When present in the terminal end of linear virion DNAs, such inverted repeats have the potential to adopt three-way junction structures with hairpin formation. Adenoassociated virus 2 (AAV2) is one of such non-pathogenic viruses and has been the focus of recent efforts to develop a gene therapy vector (Muzyczka, 1992; Kotin, 1994), due to its ability to integrate into a specific site on human chromosome 19 (Kotin et al., 1992). The hairpin-like secondary structures at the ends have been found to serve both for its replication and site-specific non-homologous integration (Muzyczka, 1991), and were recently proposed to adopt a three-way junction with two hairpin arms, each with a 9-bp stem and a T_3 or an A_3 loop (Ren et al., 1999). Modification of the two loop-T residues within the C-TTT-G sequence was found to confer the largest effect on the specific binding of the AAV Rep protein to the terminal repeat (Ryan et al., 1996). It is thus of great interest to study the T_3 loop structure in greater detail to better understand AAV2 as an efficient vector for gene therapy.

While the A_3 loop structure has been well studied before (Chou et al., 1996a), and was found to contain a single A residue closed by a sheared A·A pair, its complementary T_3 loop structure still remains elusive, although several structural studies have been reported (Boulard et al., 1991; Baxter et al., 1993; Mooren et al., 1994; Kuklenyik et al., 1996). Boulard

*To whom correspondence should be addressed. E-mail: shchou@dragon.nchu.edu.tw

et al. (1991) have used model building and NMR/MD (molecular dynamics) to study the T_3 loop in a 21-mer hairpin and found no stacking of thymidine residues either over the 3′ or the 5′ end of the stem. Baxter et al. (1993) have studied a shorter d(CGATCG-TTT-CGATCG) hairpin, but found substantial flexibility within the loop segments, and suggested that it may not be possible to define a single hairpin structure. Mooren et al. (1994) have studied the hairpin formation of the d(TCTCTC-TTT-GAGAGA) oligomer, and found that two different loop conformations were in equilibrium. In one state, the first loop thymidine is folded into the minor groove, while in the second state, the second loop thymidine is folded into the minor groove. Kuklenyik et al. (1996) have studied the T_3 loop with a stem of alternating G·C base pairs (GCGC-TTT-GCGC), and were able to derive a structure with the first thymidine folded into the minor groove.

Such ambiguous results are most likely due to different stem sequences and insufficient experimental constraints. We have thus made several T_3 loop hairpins with different stem sequences and extensively collected homonuclear NOESY spectra in both H_2O and D_2O and heteronuclear ^{1}H-^{13}C HSQC (heteronuclear single quantum coherence) and HMBC (heteronuclear multiple bond coherence) experiments to extract as many distance and torsion angle constraints as possible. Our results do indicate that T_3 loop conformation is highly dependent upon the stem sequence; only when bracketed by a 5′-GC...GC-3′ sequence can it form a static conformer that allows detailed structural studies. After collecting abundant experimental constraints, we were able to define a well-converged T_3 loop hairpin structure with many unorthodox structural features. The three loop thymidines are all unpaired, yet are highly structured and form a stable mini-loop through several ζ and α torsion angle changes in the loop. The first loop T (T6) was found folded into the minor groove and forms three H-bonds with the juxtaposed stem G·C base pairs. The second T (T7) stacks directly upon the C5 of the closing base pair. The last loop T (T8) extrudes outward, with the T7-T8 backbone also partially stacking upon the G9 of the closing base pair. Thus, although pyrimidines are too small to engage in a sheared configuration to draw near the two strands of a hairpin, nature has used a completely different strategy to stabilize the pyrimidine-rich triloop hairpins. Since the TCC (Chou et al., 1999b) and TTC (Chou et al., unpublished results) triloops have also been found to adopt similar features, such structural characteristics may constitute a general principle for the formation of the pyrimidine-rich loops.

Materials and methods

Sample preparation

All DNA samples were synthesized on a 6 μmol scale on an Applied Biosystems 380B DNA synthesizer with the final 5′-DMT groups attached. The samples were purified and prepared for NMR studies as described previously (Chou and Tseng, 1999).

NMR experiments

All NMR experiments were obtained on a Varian Unity Inova 600 MHz spectrometer. One-dimensional imino proton spectra at 0 °C were acquired using the jump-return pulse sequence (Plateau and Gueron, 1982). The spectral width was 16 000 Hz with the carrier frequency set at the resonance of water. The maximum excitation was set at 12.5 ppm. For each experiment, 4K complex points were collected and 64 scans were averaged with a 2 s relaxation delay.

2D NOESY in 90% H_2O/10% D_2O was performed at −5 °C in a pH 6.8 low salt (20 mM) buffer with the following parameters: delay time 1 s, mixing time 0.12 s, spectral width 12195 Hz, complex points 2048, number of transients 128, and number of increments 490.

2D NOESY experiments in D_2O were carried out at 20 °C in the hypercomplex mode with a spectral width of 5006 Hz. Spectra were collected using three mixing times of 120, 240, and 360 ms with a relaxation delay of 1 s between each transient and with 2048 complex points in the t_2 and 450 complex points in the t_1 dimension. For each t_1 incrementation, 64 scans were averaged.

The 2D ^{1}H-^{13}C HSQC spectrum was acquired with broadband decoupling (Bax et al., 1983). The delay $1/(2^1J_{CH})$ was tuned to 3 ms for optimum excitation of sugar signals. Heteronuclear decoupling was achieved with the GARP-1 sequence (Shaka et al., 1985). 160 t_1 increments of 2K complex data points were collected. Each FID in the t_1 dimension was further linear-predicted to 320 data points. The repetition delay was 1 s, and 16 scans were averaged for each FID. The carrier was positioned at 4.6 ppm for protons and 90.8 ppm for carbons. The spectral width was 4669 Hz (7.8 ppm) in the proton dimension and 27183 Hz (180 ppm) in the carbon dimension. Total acquisition time was 1.5 h.

In the ^{1}H-^{13}C HSQC spectrum acquired without broadband decoupling, the delay was tuned to 2.5 ms for optimum excitation of base signals. 160 t_1 increments of 2K complex data points were collected with a repetition delay of 1.3 s and 32 scans were averaged for each FID. The carrier was positioned at 85.66 ppm for carbons. The spectral width was 4669 Hz (7.8 ppm) in the proton dimension and 2375 Hz (15.74 ppm) in the carbon dimension. Total acquisition time was about 11 h.

The high resolution ^{1}H-^{13}C HSQC spectrum was recorded with 800 t_1 increments of 2K complex data points and 32 scans each with broadband GARP-1 decoupling (Shaka et al., 1985). The carrier was positioned at 85.66 ppm for carbons. The spectral width was 4669 Hz (7.78 ppm) in the proton dimension and 2375 Hz (15.74 ppm) in the carbon dimension, leading to extensive folding. The total measuring time was 20 h.

The HMBC experiment was recorded with 400 t_1 increments of 2K complex data points and 128 scans each. The delay $1/(2^{1}J_{CH})$ was tuned to 2.5 ms and the mixing time to 62 ms for observing smaller long-range heteronuclear couplings. Each FID in the t_1 dimension was further linear-predicted to 512 data points. The carrier was positioned at 4.6 ppm for protons and 90.8 ppm for carbons. The spectral width was 4669 Hz (7.8 ppm) in the proton dimension and 27183 Hz (180 ppm) in the carbon dimension. The total measuring time was 49 h.

A proton-detected ^{31}P-^{1}H heteronuclear correlation spectrum (Sklenar et al., 1986) was collected in the hypercomplex mode with spectral widths of 5115 Hz in the ^{1}H dimension and 2000 Hz in the ^{31}P dimension. 1024 complex points in the t_2 (^{1}H) dimension and 128 t_1 increments in the t_1 (^{31}P) dimension were collected. Protons were presaturated for 1.0 s and 128 scans were accumulated for each t_1 incrementation. The total measuring time was 12 h.

The acquired data were transferred to an IRIS 4D workstation and processed by the FELIX software (MSI Inc.) as described previously (Chou et al., 1994).

Structure determination

Three-dimensional structures of the 5′-GAAGC-TTT-GCTTC-3′ oligomer were generated by distance geometry and molecular dynamics calculations using distance and torsion angle constraints derived from NMR experiments. Most distance constraints from NOESY spectra in D_2O were classified as strong, medium, or weak based on their relative intensities at 120 ms mixing time and were given generous distance bounds of 2.0–4.0 Å, 3.0–5.0 Å, or 4.0–6.0 Å, respectively. Wider distance bounds of 5.0–10.0 Å were also applied to proton pairs that exhibited no NOE at 240 ms mixing time, but belonged to adjacent nucleotides. Canonical hydrogen-bond distances with bounds of 1.8–2.1 Å were assigned to Watson–Crick base pairs. A large number of distance constraints involving exchangeable protons (Figure 1) were also derived from H_2O/NOESY spectra and were given two distance bounds of either 2.0–4.0 Å or 3.0–6.0 Å.

Table 1. Carbon-phosphorus and proton-phosphorus scalar couplings (Hz) for the 5′-d(GAAGC-TTT-GCTTC)-3′ mini-loop hairpin

	$^{3}J_{C2'-P}$	$^{2}J_{C3'-P}$	$^{3}J_{H3'-P}$	$^{3}J_{C4'-P}$[a]	$^{4}J_{H4'-P}$[b]
G1	< 2	−4.6	2.9		0[d]
				8.3[c]	
A2	< 2	−4.1	2.0	9.2	2.6
				8.8	
A3	< 2	−4.2	2.4	10.3	4.0
				10.4	
G4	< 2	−3.3	4.0	9.3	2.6
				9.4	
C5	< 2	−3.8	4.0	9.4	2.4
				10.0	
T6	< 2	−6.2	9.2	7.5	2.6
				9.0	
T7	< 2	−3.2	4.6	8.7	2.9
				8.4	
T8	< 2	−4.7	4.3	7.0	2.3
				7.0	
G9	< 2	−4.0	2.8	9.3	1.0
				9.3	
C10	< 2	−4.7	2.0	9.6	3.2
				11.0	
T11	< 2	−5.1	2.0	10.8	2.4
				11.9	
T12	< 2	−3.7	3.0	11.9	2.0
				11.9	
C13	< 2	[e]	[e]		2.4
				9.5[f]	

[a] Each central residue has two $^{3}J_{C4'-P}$ values that are about equal, but could not be specifically identified.
[b] Sum of (n)H4′–(n)P and (n)H4′–(n+1)P couplings, but the latter value is usually zero.
[c] Three-bond (n)C4′–(n+1)P coupling constant.
[d] Four-bond (n)H4′–(n+1)P coupling vanishes when H4′–C4′–C3′–O3′–P atoms are not in the 'W' shape (Sarma et al., 1973; Chou et al., 1997).
[e] No phosphate at the 3′-end.
[f] Three-bond (n)C4′–(n)P coupling constant.

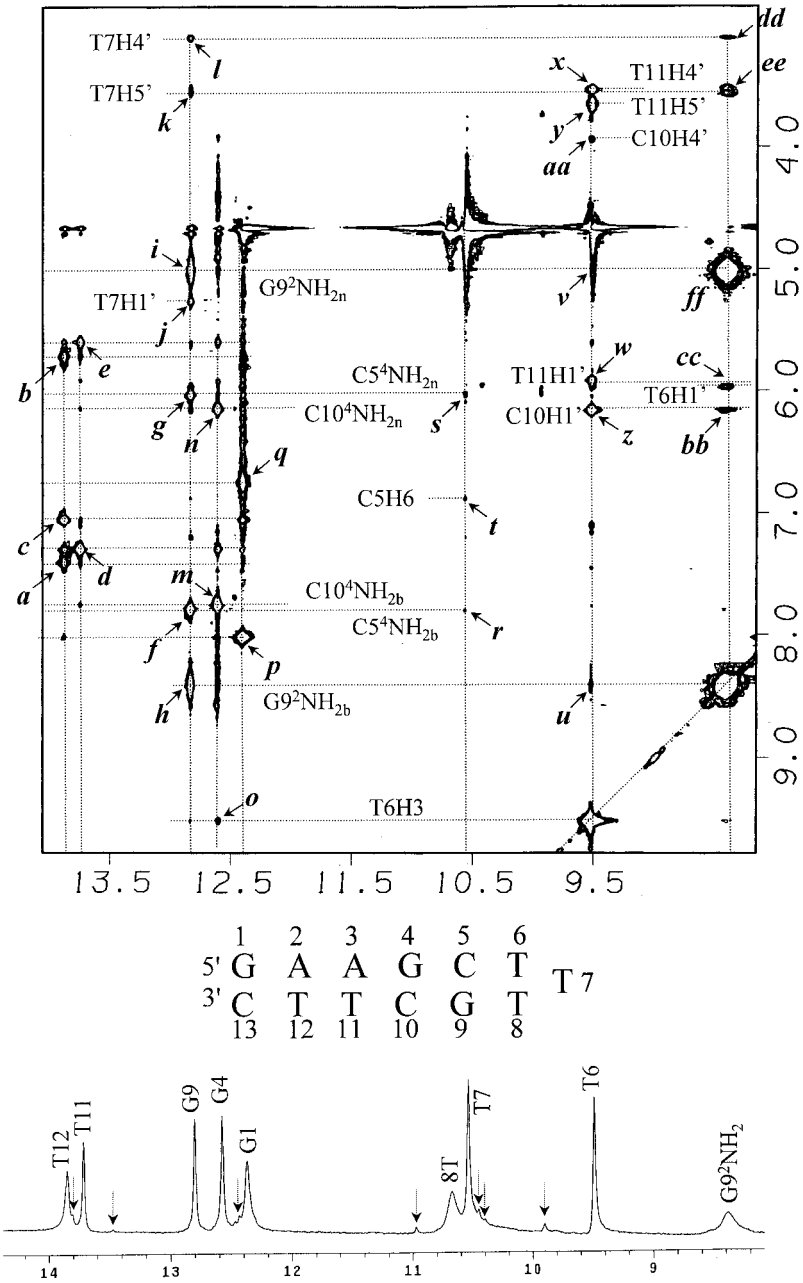

Figure 1. The expanded 2D-H_2O/NOESY contour plot (mixing time 0.12 s) and the one-dimensional imino proton spectrum (bottom) of the 5′-GAAGC-TTT-GCTTC-3′ sequence (T_3) at −5 °C. The assigned imino and G9 amino protons are labeled with residue numbers; several extra imino proton peaks, possibly arising from the duplex form, are also marked with arrows. These extra signals constitute less than 5% of the total signals and are not subject to further studies. The expanded 2D plot (top) covers NOE cross peaks from the imino to amino/base/H1′/H3′/H4′/H5′ protons. The readily observed NOEs exhibited by the loop T6H3 and T7H3 indicate that they are in slow exchange and are protected from exchange due to the special structural features adopted by the loop thymidines. The cross peaks were assigned as follows: a: T12H3–A2^{6}NH$_{2b}$; b: T12H3–A2^{6}NH$_{2n}$; c: T12H3–A2H2; d: T11H3–A3^{6}NH$_{2b}$; e: T11H3–A3^{6}NH$_{2n}$; f: G9H1–C5^{4}NH$_{2b}$; g: G9H1–C5^{4}NH$_{2n}$; h: G9H1–G9^{2}NH$_{2b}$; i: G9H1–G9^{2}NH$_{2n}$; j: G9H1–T7H1′; k: G9H1–T7H5′; l: G9H1–T7H4′; m: G4H1–C10^{4}NH$_{2b}$; n: G4H1–C10^{4}NH$_{2n}$; o: G4H1–T6H3; p: G1H1–C13^{4}NH$_{2b}$; q: G1H1–C13^{4}NH$_{2n}$; r: T7H3–C5^{4}NH$_{2b}$; s: T7H3–C5^{4}NH$_{2n}$; t: T7H3–C5H6; u: T6H3–G9^{2}NH$_{2b}$; v: T6H3–G9^{2}NH$_{2n}$; w: T6H3–T11H1′; x: T6H3–T11H4′; y: T6H3–T11H5′; z: T6H3–C10H1′; aa: T6H3–C10H4′; bb: G9^{2}NH$_{2b}$–C10H1′; cc: G9^{2}NH$_{2b}$–T6H1′; dd: G9^{2}NH$_{2b}$–T7H4′; ee: G9^{2}NH$_{2b}$–T7H5′; ff: G9^{2}NH$_{2b}$–G9^{2}NH$_{2n}$.

Table 2. Constraints used to determine the structure of the 5′-d(GAAGC-TTT-GCTTC)-3′ hairpin

Restraints	No. of constraints
Exchangeable NOEs	54
H-bonds (1.8 Å–2.1 Å)	13
2.0 Å–4.0 Å	5
3.0 Å–6.0 Å	36
Non-exchangeable NOEs	201
2.0 Å–4.0 Å	26
3.0 Å–5.0 Å	106
4.0 Å–6.0 Å	46
5 Å–10.0 Å	23
Total NOEs	255
Torsional angles	49
Backbone (β, γ, ϵ)	36
Glycosidic	13
NOEs per residue	19.6
NOEs and torsion angles per residue	23.4
Violations of experimental restraints	
Distance restraints (> 0.15 Å)	2[a]
Torsional angle restraints (> 3°)	0
Rmsd	0.46 ± 0.18 Å

[a]Five out of 20 final structures have two violations larger than 0.15 Å.

The total numbers of distance constraints for the unpaired T6, T7, and T8 residues are 25, 30, and 24, respectively. Some of the critical constraints are schemed in Figure 6. The β and γ torsion angle constraints were determined primarily semi-quantitatively from the ^{31}P-^{1}H heteronuclear correlation data (Chou et al., 1996a) using the in-plane 'W' rule (Sarma et al., 1973). If long-range (n)P ↔ (n)H4′ four-bond couplings were detected, then the β and γ torsion angles were constrained to the *trans* (180° ± 30°) and *gauche*$^{+}$ (60° ± 30°) domains, respectively. The β torsion angle constraints could further be checked by the heteronuclear $^{3}J_{C4'-P}$ coupling constants (Table 1). It is important to note that, while the value for the 5′-end G1 residue can correspond to the ϵ torsion angle and the value for the 3′-end C13 residue to the β torsion angle, the two $^{3}J_{C4'-P}$ values for each non-terminal phosphorus atom could not be assigned to specific torsion angles (Figure 3). However, these two values are approximately equal for each residue and most are larger than 7 Hz, which indicates that all β torsion angles are located in the *trans* domain (Marino et al., 1999). The ϵ torsion angle can only be located in either the *trans* or the *gauche*$^{-}$ domain (Altona, 1982). The *gauche*$^{+}$ conformation is not sterically allowed. The combination of $^{3}J_{H3'-P}$, $^{3}J_{C4'-P}$, and $^{3}J_{C2'-P}$ also allows one to determine the ϵ torsion angle (Marino et al., 1999). The ζ and α dihedral angles could not be determined from any heteronuclear coupling until now and were all left unconstrained. The χ dihedral angles were constrained to −100° ± 30° (ideal B-DNA values) when no aromatic–anomeric cross peaks of comparable intensity to the CH5/CH6 cross peaks were detected (Dallas and Moore, 1997). This is consistent with the data (see Supplementary material) that all $^{3}J_{C8/C6-H1'}$ coupling constants are larger than the corresponding $^{3}J_{C4/C2-H1'}$ coupling constants (Schmieder et al., 1992). The combined NOE distance (255 in total) and torsion angle (49 in total) constraints were used to generate initial structures using the DGII program (MSI, Inc.). The initial structures were further refined by restrained molecular dynamics using the program DISCOVER (MSI, Inc.). A 10 ps dynamics run was performed at 300 K with a step size of 1.0 fs, which was followed by a conjugate gradient minimization of 2000 iterations looped 10 times. Thirty well-converged final structures with pair-wise rmsd values of less than 0.5 Å were obtained after these calculations.

Results

Exchangeable proton studies

Figure 1 shows the one-dimensional (bottom) and the expanded two-dimensional (top) imino/G9NH$_2$–H6/H8/H1′/H4′/H5′ proton spectra of the d(GAAGC-TTT-GCTTC) oligomer. Spectra were collected at −5 °C in 90% H_2O/10% D_2O containing 20 mM NaCl. These imino and amino protons were systematically assigned using the standard procedure as described (Tseng and Chou, 1999). The two resonances at a lower chemical shift (13.6–14 ppm) and the three at a higher chemical shift (12–13 ppm) are characteristic of the imino protons participating in a regular Watson-Crick A·T and G·C base pair. These imino protons were assigned through the G1-T12-T11-G4-G9 imino proton NOE connectivity of the H_2O/NOESY spectrum (data not shown). The resonances at 9.4 ppm, 10.6 ppm, and 10.7 ppm are, on the other hand, characteristic of imino protons not involved in classical Watson–Crick H-bonding. These resonances possibly belong to the imino protons of the three loop thymidines. The sharp imino proton resonances at 9.4 ppm and 10.6 ppm indicate that they are protected from solvent exchange, while the broader

imino proton peak at 10.7 ppm indicates that it is exposed to the solvent and exchanges faster with H_2O. Several smaller peaks, marked by arrows in the bottom figure, were also detected for this sequence. These resonances possibly belong to the duplex form, which disappeared when the sample was diluted twice.

The assignments of the unpaired loop T-imino protons are shown in the expansion spectrum of the imino to NH_2/base/H1′/H3′/H4′/H5′ proton region of the 2D H_2O/NOESY spectrum (top Figure 1) collected at $-5\,^{\circ}$C. The T12, T11, G9, G4, and G1 imino protons are all involved in regular Watson–Crick base-pairing as described above, and exhibit classical NOEs to their base-paired and neighbored A^6NH_2, AH2, and C^4NH_2 protons (Tseng and Chou, 1999). However, several rather unusual NOEs exhibited by the G9-imino proton were observed; the NOEs to the T7H1′ (cross peak j), T7H4′ (l), and T7H5′ (k) protons were detectable. More notably, G9-imino also exhibits strong NOEs to its H-bonded $^2NH_{2b}$ (cross peak h) and non-H-bonded $^2NH_{2n}$ (cross peak i) protons, which is confirmed by the strong geminal G9$^2NH_{2b}$ ↔ G9$^2NH_{2n}$ cross peak (ff). This is quite unusual, as the G^2NH_2 protons are normally located in the minor groove and exhibit no NOE to the G-imino proton in a regular G·C base pair. This phenomenon has been attributed to the faster rotation rate of the G^2C-G^2N bond (Mueller et al., 1995). A similar situation was also detected for the G^2NH_2 protons participating in the tandem sheared G·A base pairs, in which an adenosine pairs with a guanosine through the minor groove edge (Chou et al., 1999a). This indicates that the G9$^2NH_{2b}$ amino proton in this hairpin possibly experiences protection from solvent exchange in the minor groove by the folded-in loop thymidine.

The unpaired T-imino proton at 9.4 ppm was assigned to T6H3, since it exhibits strong NOEs to the C10H1′ (cross peak z), T11H1′ (w), T11H4′ (x), and T11H5′/H5″ protons (y), which were assigned by the D_2O/NOESY spectrum described later; similar NOEs were also observed for the loop T6 imino proton in the d(GAAGC-TCC-GCTTC) hairpin (Chou et al., 1999a). The imino proton at 10.6 ppm was assigned as the T7H3 by the observation that it exhibits weak NOEs to the C5H6 (t), C5$^4NH_{2b}$ (r) and C5$^4NH_{2n}$ (s) protons. These NOEs are consistent with the non-exchangeable proton NOEs between the T7CH3↔C5H5 and T7CH3↔C5H1′ protons (see Supplementary material), which together suggest that the T7 base stacks upon the C5 nucleotide. The broader imino proton at 10.7 ppm could now be assigned to T8H3, which exhibits no NOEs to the surrounding protons, suggesting that it is in fast exchange with the solvent. The three critical unpaired loop TH3 protons were thus confidentially assigned. Finally, one weak NOE between the T6 and G4 imino protons (cross peak o) was also detected, reinforcing the idea that T6 folds into the minor groove and interacts with the stem G4·C10 base pair.

2D D_2O/NOESY studies

The evidence that T7 stacks upon C5 further comes from the non-exchangeable proton NOESY spectrum (see Supplementary material). Except for the regular NOEs commonly detected in this region, several unexpected NOEs were also observed. The two T7CH$_3$ ↔ C5H5 and T7CH$_3$ ↔ C5H1′ NOEs, as well as the T7CH$_3$ ↔ C5H3′ and T7CH$_3$ ↔ C5H2′ NOEs, clearly confirm that T7 stacks upon C5. This is also consistent with the T7H3 ↔ C5H6 (cross peak t in Figure 1) and T7H3 ↔ C5$^4NH_2$ (cross peaks r and s in Figure 1) NOEs described above.

Stereospecific assignment of H5′/H5″ protons

The assignment of H5′/H5″ using 2D-NOESY has been difficult, due to the overlap of the H5′/H5″ signals with the H4′ signals and the weak H6/H8 to H5′/H5″ NOE intensities. The stereospecific proR or proS assignment of the H5′ proton is even more difficult without an isotope enriched sample (Hines et al., 1994; Marino et al., 1996). Besides, the C5′ signals are also quite crowed in a regular ^{1}H-^{13}C HSQC experiment. Although they are more dispersed in the high resolution ^{1}H-^{13}C HSQC experiment (see Supplementary material), they still cannot be assigned without prior knowledge of the proton assignment. The stereospecific H5′$^{(pro-R)}$ and H5′$^{(pro-S)}$ assignment has been previously achieved through the 3D-HMQC-TOCSY or 3D-HMQC-NOESY (Hines et al., 1994) or directed HCC-TOCSY-CCH-E.COSY and selective H5′/C5′ HSQC experiments using a ^{13}C labeled molecule (Marino et al., 1996, 1999). We have, however, found an alternative way to simultaneously assign the H5′$^{(pro-S)}$, H5′$^{(pro-R)}$ and C5′ atoms through the more resolved C3′/C4′–H5′ and H3′/H4′–C5′ correlations by simply running ^{1}H-^{13}C HMBC and HSQC correlation experiments at natural abundance. This is shown in Figure 2a, in which the two spectra are superimposed with the folded-in HSQC spectrum displayed above (Figure 2b) to allow for easier assignments. The single-bond ^{1}H-^{13}C correlation cross peaks in the HSQC spectrum are drawn in hollow and labeled with

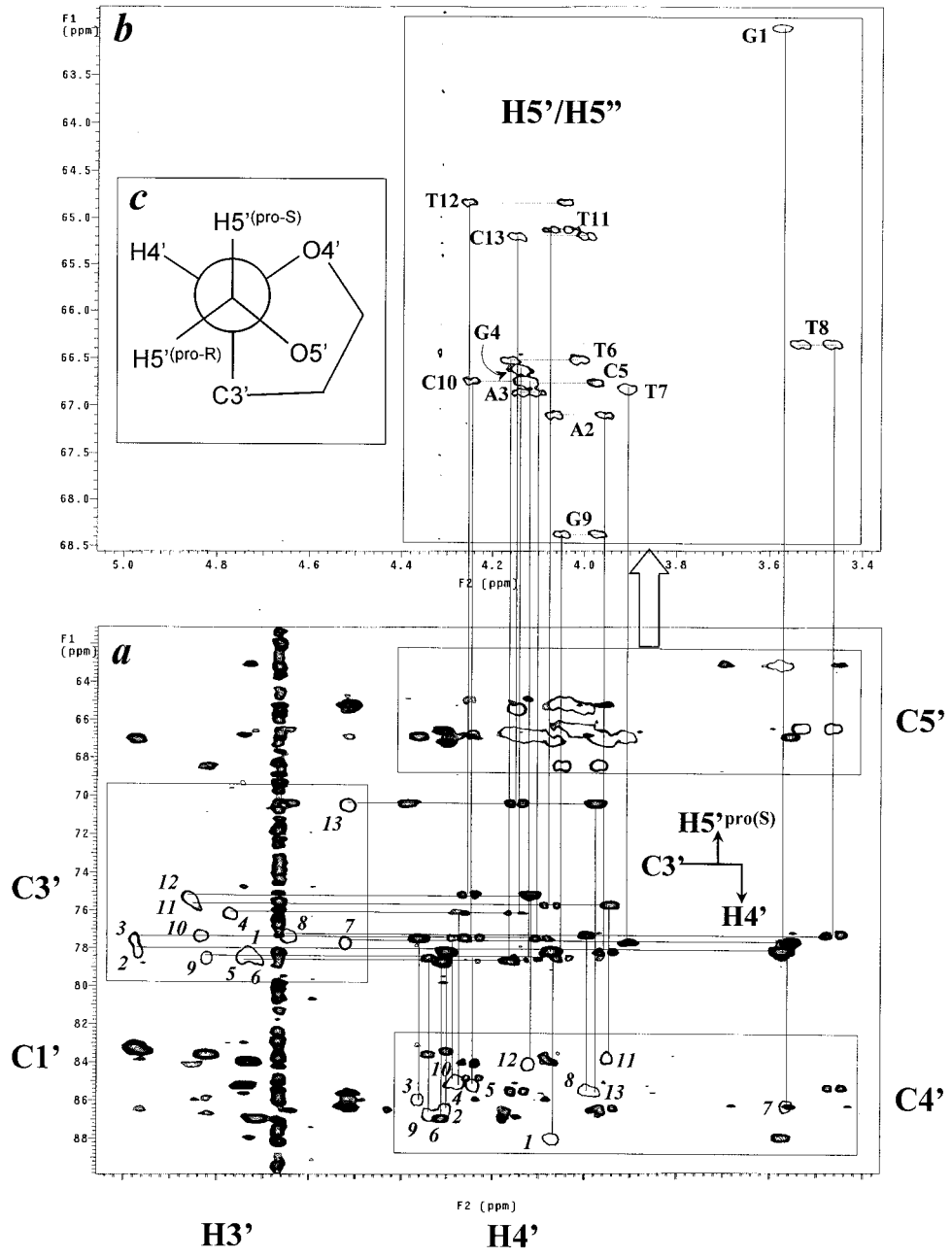

Figure 2. The superimposed ^{1}H-^{13}C HSQC (plotted in hollow) and HMBC (plotted in solid) spectra (box a). The crowded H5′-C5′ correlation region was further expanded above by the high-resolution HSQC spectrum (box b) for easier H5′/H5″ identification. All assigned HSQC cross peaks were labeled with residue numbers. Box c shows the cartoon of the rotamer when the γ torsion angle is in the *gauche*$^{+}$ conformation. In such a conformation, the H5′$^{(pro-S)}$–C5′–C4′–C3′ atoms are in the *trans* domain, and exhibit a larger $^{3}J_{H5'-C3'}$ coupling, while the H5′$^{(pro-R)}$–C5′–C4′–C3′ atoms are in the *gauche*$^{-}$ domain, and should exhibit smaller or non-detectable $^{3}J_{H5'-C3'}$ coupling. Each C3′ atom thus only reveals two cross peaks in this region; one to the H4′ atom that is connected downward, and the other to the H5′$^{(pro-S)}$ that is connected upward. The stereospecific H5′$^{(pro-S)}$ and H5′$^{(pro-R)}$ protons could thus be assigned. Some more H3′/H4′–C5′ correlation peaks in the upper-left corner and H3′–C1′ cross peaks in the lower-left corner were obvious in box a. These were not labeled for clarity reasons.

the residue numbers, while those from the HMBC spectrum are drawn in solid. Although the C5′ signals in Figure 2b are well resolved, their assignment still requires prior knowledge of the H5′/H5″ assignment, which is difficult to achieve based only upon the homonuclear NOESY experiment. However, using the easily assigned C3′ resonances and the C3′–H5′ (upward) and C3′–H4′ (downward) correlations, one can then stereospecifically assign H5′ protons. As shown in the C3′–H4′/H5′ region in the center of Figure 2a, we observed two cross peaks for each C3′ signal. One is from the C3′–H5′ correlation (connected upward to the $H5'^{(pro-S)}$) that is split by the geminal H5′/H5″ coupling and appeared as a doublet, while the other is from the C3′–H4′ correlation (connected downward). The C3′–$H5'^{(pro-S)}$ and C3′–H4′ correlations can thus be easily identified. Except for the weak C5C3′–C5H4′ peak, all other pairs of cross peaks were clearly detected. Interestingly, only one correlation peak for each C3′–H5′/H5″ pair was observed, which can be explained by the insert shown in Figure 2c. When the γ torsion angle is located in the *gauche*$^+$ domain (as present in most nucleotides in either DNA or RNA, and can be confirmed by the observation of the $^4J_{H4'-P}$ coupling (Sarma et al., 1973; Chou et al., 1996a, 1997)), the $H5'^{(pro-S)}$–C5′–C4′–C3′ atoms are in the *trans* conformation, and a stronger $^3J_{H5'-C3'}$ coupling constant is expected. In contrast, the $H5'^{(pro-R)}$–C5′–C4′–C3′ atoms are in the *gauche* domain in a similar situation, and only weaker or non-observable $^3J_{H5'-C3'}$ coupling constants can be detected. The observed C3′–H5′ cross peak is thus due to the C3′–$H5'^{(pro-S)}$ correlation, and the other H5′ signal must belong to the $H5'^{(pro-R)}$. The C5′ signals could also be assigned through the help of the H3′/H4′–C5′ cross peaks (upper-left corner of Figure 2a). Although the idea of stereospecific $H5'^{(pro-S)}$ and $H5'^{(pro-R)}$ assignment based on the $^3J_{H5'-C3'}$ and $^3J_{H5''-C3'}$ couplings as well as the γ torsion angle has been proposed (Hines et al., 1994; Varani et al., 1996), this is the first time to our knowledge that such an idea is put into practice. Absolute confirmation of such an approach would require preparation of labeled oligomer with selective H5′-deuteration, which is being pursued in our laboratory (data not published). The chemical shifts for all assigned protons and carbons are available as Supplementary material.

Determination of heteronuclear coupling constants

There is no doubt that ^{1}H-^{1}H, ^{1}H-^{13}C, ^{1}H-^{31}P, and ^{13}C-^{31}P couplings have great potential in determining nucleic acid structure, as has been well discussed in several papers (Varani and Tinoco, 1991a; Hines et al., 1994; Marino et al., 1999). Ironically, the homonuclear ^{1}H-^{1}H coupling is the most difficult to measure, due to the severe overlapping of the proton signals. However, such a ^{1}H-^{1}H coupling could be determined through the passive coupling of the HCCH-E.COSY or directed HCC-TOCSY-CCH-E.COSY experiment that requires preparation of ^{13}C-labeled samples (Marino et al., 1999). Fortunately, many critical backbone torsion angles can now be determined solely from the high resolution ^{1}H-^{13}C HSQC experiments at natural abundance, as has been demonstrated by several groups (Schmieder et al., 1992; Avizonis and Kearn, 1995).

Of the many heteronuclear couplings, the one-bond ^{1}H-^{13}C scalar couplings are the easiest to measure and could be determined from the fully coupled high-resolution HSQC experiment. Since the spectrum is very well resolved (data not shown), such a measurement is straightforward. All the measured $^1J_{13C-H}$ values are listed in the Supplementary material.

On the other hand, the long-range ^{1}H-^{31}P and ^{13}C-^{31}P coupling constants have to be extracted from the high-resolution ^{1}H-^{13}C HSQC experiment, as previously demonstrated (Schmieder et al., 1992). Figure 3a shows the cross peaks correlating the H3′ and C3′ of residues T7 and T8. The 'E.COSY' pattern of the cross peak is due to the passive $^3J_{H3'-P}$ coupling in the proton dimension and the passive $^2J_{C3'-P}$ coupling in the carbon dimension. Except for the C13 residue that has no attached phosphate, every other residue exhibits such a typical tilted cross peak, and the respective coupling constants could be measured as the displacement of the multiplet components, which are listed in Table 1. While utilization of the size and relative sign of the two-bond $^2J_{C3'-P}$ coupling constants in structural determination remains to be explored, the three-bond $^3J_{H3'-P}$ couplings are useful in locating the ϵ torsion angle through the modified Karplus equation (Varani et al., 1996; Marino et al., 1999). However, since a single $^3J_{H3'-P}$ coupling constant value can correspond to up to four different ϵ values, it is necessary to further determine the $^3J_{C4'-P}$ and $^3J_{C2'-P}$ couplings to restrict the ϵ value to a single domain. Such couplings could also be determined from the high-resolution ^{1}H-^{13}C HSQC experiment, with the determination of $^3J_{C4'-P}$ values shown in Figure 3b and 3c. Figure 3b shows the H4′-C4′ coupling pattern of the 5′-end G1 residue, which is split in the carbon dimension by the $^3J_{(n)C4'-(n+1)P}$ coupling. However,

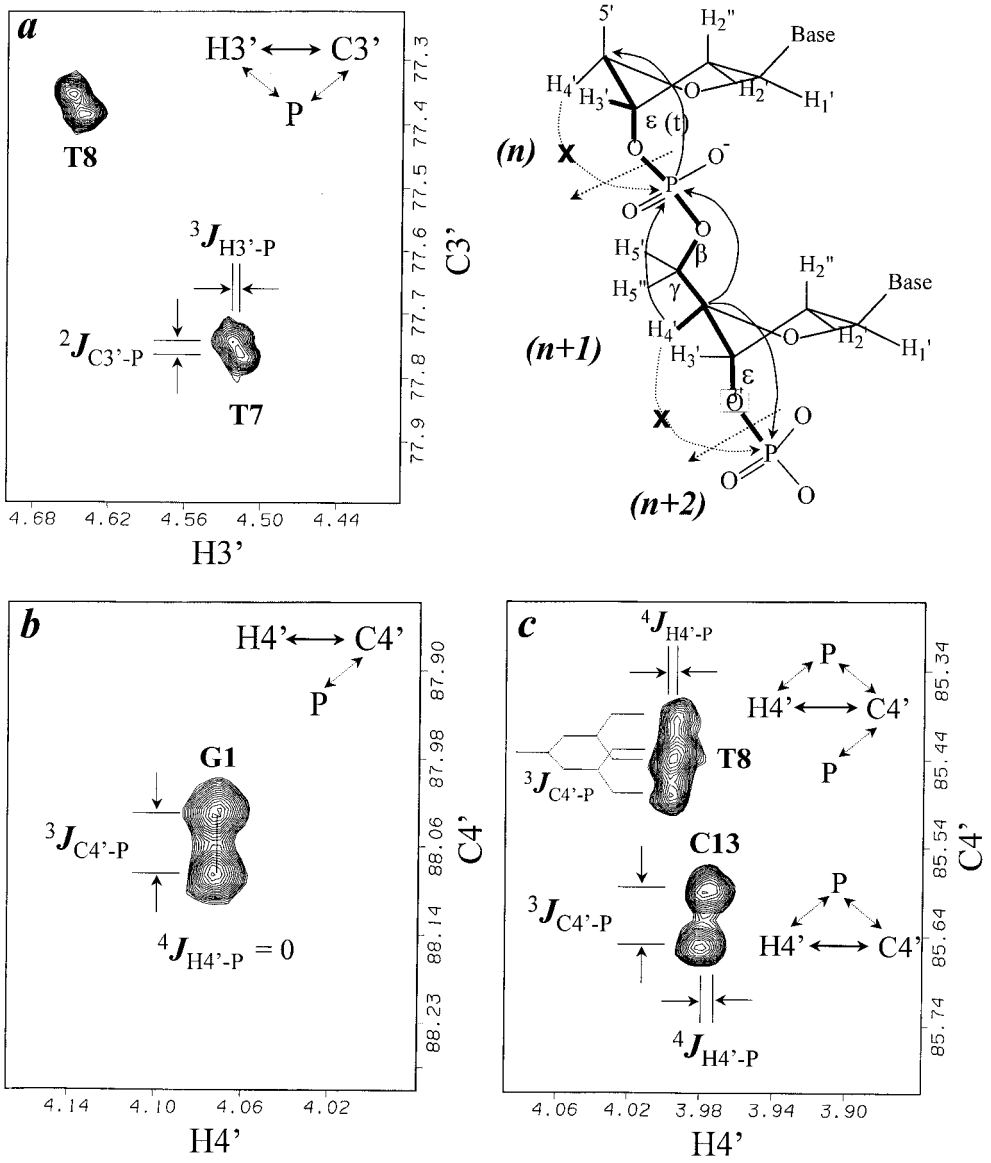

Figure 3. The expanded high-resolution HSQC contour plots of the H3′–C3′ (box a), G1H4′–G1C4′ (box b), and C13H4′–C13C4′ and T8H4′–T8C4′ (box c) regions. The cartoon displays the backbone of a dinucleotide and is plotted in the upper-right corner, with the atom pairs exhibiting four-bond scalar couplings connected by curved arrows and the atom pairs exhibiting no such couplings connected by curved arrows marked with an 'x'. In box a, the H3′–C3′ correlation peak is tilted in the carbon dimension by the passive two-bond $^2J_{C3'-P}$ coupling, and tilted in the proton dimension by the passive three-bond $^3J_{H3'-P}$ coupling. In box b, the H4′–C4′ cross peak of the 5′-end G1 residue is only tilted in the carbon dimension by the $^3J_{C4'-P}$ coupling, and not tilted in the proton dimension, due to the absence of the $^4J_{H4'-P}$ coupling. In box c, the H4′–C4′ cross peak of the 3′-end C13 residue is tilted in both dimensions by the corresponding coupling. Every central residue is characterized by a tilted quartet pattern split in the carbon dimensions by two $^3J_{C4'-P}$ couplings and only split once in the proton dimension by one $^4J_{H4'-P}$ coupling, as typified by the T8 residue.

the proton dimension is not split, due to the absence of long-range four-bond $^4J_{(n)H4'-(n+1)P}$ couplings. Such couplings only occur when the atoms connecting the four bonds are coplanar and the four bonds form a 'W' shape (Sarma et al., 1973; Chou et al., 1997), as happens in the (n)P–(n)H4′ coupling. One example of such coupling can be seen in the 3′-end C13 residue (Figure 3c). In this H4′-C4′ cross peak, the C4′ and H4′ exhibit three-bond $^3J_{(n+1)C4'-(n+1)P}$ and four-bond $^4J_{(n+1)H4'-(n+1)P}$ couplings to the 5′-end phosphorus atom, respectively, and can be examined by the tilting of the cross peaks. It is worthy to note that the $^3J_{(n)C4'-(n+1)P}$ coupling for the 5′-end G1 residue is characteristic of the ϵ torsion angle, while the three-bond $^3J_{(n+1)C4'-(n+1)P}$ coupling for the 3′-end C13 residue is characteristic of the β torsion angle. Both values are around 9 Hz and correspond to the *trans* domain for these two torsion angles. On the other hand, the carbon dimension for any internal residue is split twice by the $^3J_{(n)C4'-(n+1)P}$ and $^3J_{(n+1)C4'-(n+1)P}$ couplings into a quartet, as shown in the figure for the T8 residue. However, these two values could not be assigned to a specific coupling. The proton dimension of the internal residue is still only split once by the $^4J_{(n+1)H4'-(n+1)P}$ coupling, due to the absence of the $^4J_{(n)H4'-(n+1)P}$ coupling.

The three-bond $^3J_{(n)C2'-(n+1)P}$ couplings could be determined in the same way from the tilting of the H2′–C2′ correlation cross peaks (data not shown) and were found to be less than 2 Hz in most cases. All measured coupling values are listed in Table 1.

^{31}P-^1H HETCOR experiment

Stereospecific $H5'^{(pro-S)}$ and $H5'^{(pro-R)}$ assignment needs prior knowledge of the γ torsion angle, which is difficult to obtain without measuring the small $^3J_{H4'-H5'}$ coupling constants. Measurement of such couplings, however, requires tedious preparation of ^{13}C-labeled samples (Hines et al., 1994; Marino et al., 1996; Varani et al., 1996). Fortunately, by running the heteronuclear ^{1}H-^{31}P correlation spectrum at natural abundance (Sklenar et al., 1986), one can still obtain the required γ torsion angle data, at least in a semi-quantitative way (Sarma et al., 1973; Blommers et al., 1989; Chou et al., 1996a, 1997). Except for the G9 phosphorus atom at the loop-turning point that reveals no detectable (n)P-(n)H4′ correlation, all other phosphorus atoms exhibit approximately equal intensity correlation peaks to the 3′-end H4′ (see Supplementary material), indicating that most backbone torsion angles are located in the regular $\beta(t)\gamma(g^+)$ domain (Sarma et al., 1973; Blommers et al., 1989; Chou et al., 1996a). Such results are consistent with the $^3J_{C4'-P}$ couplings measured from the high-resolution HSQC. Although the G9H4′–G9P cross peak could not be observed, the β torsion angle could still be constrained in the *trans* domain, due to the large G9 $^3J_{C4'-P}$ coupling value (9.3 Hz, Table 1). The absence of the G9 $^4J_{H4'-P}$ coupling is thus due to the distortion of the γ torsion angle from the required *gauche*$^+$ domain.

Structural features

The distance and torsion angle constraints used to determine the T_3 mini-loop hairpin structure are listed in Table 2. By increasing the number of constraints, we were able to get a well-converged family of 30 final structures from 50 random embeds with rmsd values of 0.46 ± 0.18 Å (compared with the average structure). The superposition of the 20 final structures with the lowest energy is shown in Figure 4 from two different views, with their measured torsion angles listed in Table 3. As is the case with most other DNA sugars, all nucleotides in this T_3 mini-loop hairpin, including the three T_3 unpaired nucleotides, adopt the regular C2′-endo pucker. The glycosidic angle is also in the *anti* domain for all nucleotides. The superimposed pictures clearly reveal the unusual arrangement of the bases in the loop. While the T6 nucleotide folds into the minor groove, the T7 base stacks well with the C5 of the closing C5·G9 base pair, with the T8 base also pointing toward the major groove. The folding of the T6 base into the minor groove makes it possible for T6 to form three H-bonds with the G4N2, G9N2, and C10O2 atoms of the stem G·C base pairs to stabilize this unusual structure, as occurred in the TCC mini-loop (Chou et al., 1999b). Unlike other base pairs in the stem region that use the Watson–Crick edges to participate in H-bonding, those of residues T7 and T8 face instead toward the major groove, and are available for protein binding. Such a feature might be responsible for the specific binding of the AAV Rep protein to the terminal repeat of the AAV2 virus (Ryan et al., 1996). Even with such dramatic base rearrangement, the backbone around the loop still looks smooth and appears as an 'S' shape (Figure 5). Except for the $\gamma(g^+)$ to $\gamma(t)$ torsion angle change at the loop-turn nucleotide (G9), the backbone is adjusted mainly through the ζ and α torsion angles. Most β, γ, and ϵ torsion angles remain unchanged and are located in the regular $(t)(g^+)(t)$ domain, although the ϵ torsion angles of residues T7 and T8 have adjusted somewhat toward the *gauche*$^-$ domain (Table 3). The corresponding

Table 3. Backbone torsion angles (deg) in the T_3 mini-loop hairpin

Residue	α	β	γ	δ	ϵ	ζ	χ
1G				132 ± 8	-175 ± 2	-93 ± 3	-131 ± 3
2A	-74 ± 3	178 ± 23	60 ± 2	118 ± 9	-178 ± 3	-96 ± 7	-126 ± 7
3A	-69 ± 3	176 ± 3	59 ± 2	115 ± 10	-173 ± 2	-92 ± 5	-122 ± 7
4G	-74 ± 2	167 ± 3	61 ± 2	101 ± 7	180 ± 2	-91 ± 3	-132 ± 3
5C	-66 ± 2	170 ± 2	66 ± 2	128 ± 2	-177 ± 2	-103 ± 2	-126 ± 2
6T	$\boldsymbol{-162 \pm 2}$	175 ± 2	65 ± 2	144 ± 2	-140 ± 3	-78 ± 2	-136 ± 2
7T	-83 ± 2	-175 ± 3	68 ± 2	133 ± 3	$\boldsymbol{-93 \pm 2}$	$\boldsymbol{162 \pm 2}$	-135 ± 2
8T	-74 ± 2	170 ± 3	58 ± 2	137 ± 2	$\boldsymbol{-96 \pm 2}$	$\boldsymbol{50 \pm 2}$	-129 ± 3
9G	$\boldsymbol{107 \pm 3}$	-178 ± 4	$\boldsymbol{-170 \pm 2}$	136 ± 10	-176 ± 2	-97 ± 5	-123 ± 6
10C	-73 ± 4	176 ± 4	62 ± 2	115 ± 5	-174 ± 2	-88 ± 2	-124 ± 6
11T	-71 ± 4	170 ± 2	61 ± 2	116 ± 10	178 ± 2	-102 ± 7	-124 ± 6
12T	-66 ± 2	175 ± 2	63 ± 2	134 ± 5	-178 ± 2	-98 ± 6	-115 ± 5
13C	-74 ± 2	180 ± 3	56 ± 2	131 ± 5			-118 ± 4

Parameters were calculated based on the values of 20 final structures. Values are expressed as mean values ± standard deviations. Special torsional angles are listed in italic-bold.

torsion angle changes at the loop-turn have been observed many times in single-nucleotide loops closed by sheared G·A (Chou et al., 1996b; Zhu et al., 1996), A·A (Chou et al., 1996a), or A·C (Chou et al., 1999c) base pairs.

Figure 5 shows the stacking of the 5′-CG-TTT-CG-3′ motif. While the G4–C5–T7 bases have excellent intra-strand stacking at one side, those of G9 and C10 also stack well at the other side. The T6 base does not involve in any stacking, but instead participates in H-bonding with the functional groups of the C5–G9 and G4–C10 base pairs at the minor groove edge. Figure 6 shows the stereoview of this motif from the major groove (bottom) and minor groove (top) sides of one of the final structures. The idiosyncratic protons that are critical in determining this unusual T_3 mini-loop structure are marked by white circles and connected by pink arrows. The G9H8–T8H1′, G9H8–T8H2′, and G9H8–T8H2″ proton pairs that do not exhibit NOEs, even at 240 ms mixing time, are connected by dotted cyan arrows. The $\beta(t)$ and $\gamma(t)$ torsion angles of the turning G9 residue are evident in the bottom figure and are marked by white and yellow arrows, respectively. From this figure, the upfield shifting for the T7H1′, T7H4′, and T8H5′ protons is evident, as they are directly above the G9 base. The abundant experimental constraints around this mini-loop have led to reliable determination of this unusual structure, which is also consistent with the chemical shift changes.

Discussion

Structural study of nucleic acid molecules through isotope-enriched samples has been a subject of extensive studies (Varani and Tinoco, 1991b; Nikonowicz et al., 1992; Pardi, 1995; Zimmer and Crothers, 1995; Patel et al., 1997; Kolk et al., 1998; Butcher et al., 1999). However, isotope enrichment of the RNA or DNA samples by the in vitro transcription/replication method is still tedious and too expensive to be of general application, and preparation of nucleic acids containing unusual nucleotides can only be carried out by the chemical synthesis method (Jones, 1994). We have therefore tried to look for alternate ways to solve nucleic acid structure at natural abundance and to push the limits for such approaches.

As a first step toward this goal, we have used HMBC in combination with high-resolution HSQC experiments to stereospecifically assign the crowded $H5'^{(pro-S)}$, $H5'^{(pro-R)}$ and C5′ atoms, which is usually the most difficult task in DNA or RNA structural determination. Even though the H2′/H3′/H4′/H5′/H5″ signals of an RNA molecule are usually overlapped in the homonuclear NOESY spectrum, the carbon signals are, on the contrary, very well resolved in the carbon dimension of the heteronuclear ^{1}H-^{13}C HSQC spectrum (Varani et al., 1996). Thus, by running an HMBC experiment, one may be able to assign the crowded H2′/H3′/H4′/H5′/H5″ signals through their attached carbons by observing the multiple-bond C1′–H3′, H1′–C3′, C2′–H3′, C4′–H2′, and C3′–H5′ correlations, etc. Most of these cross peaks were in-

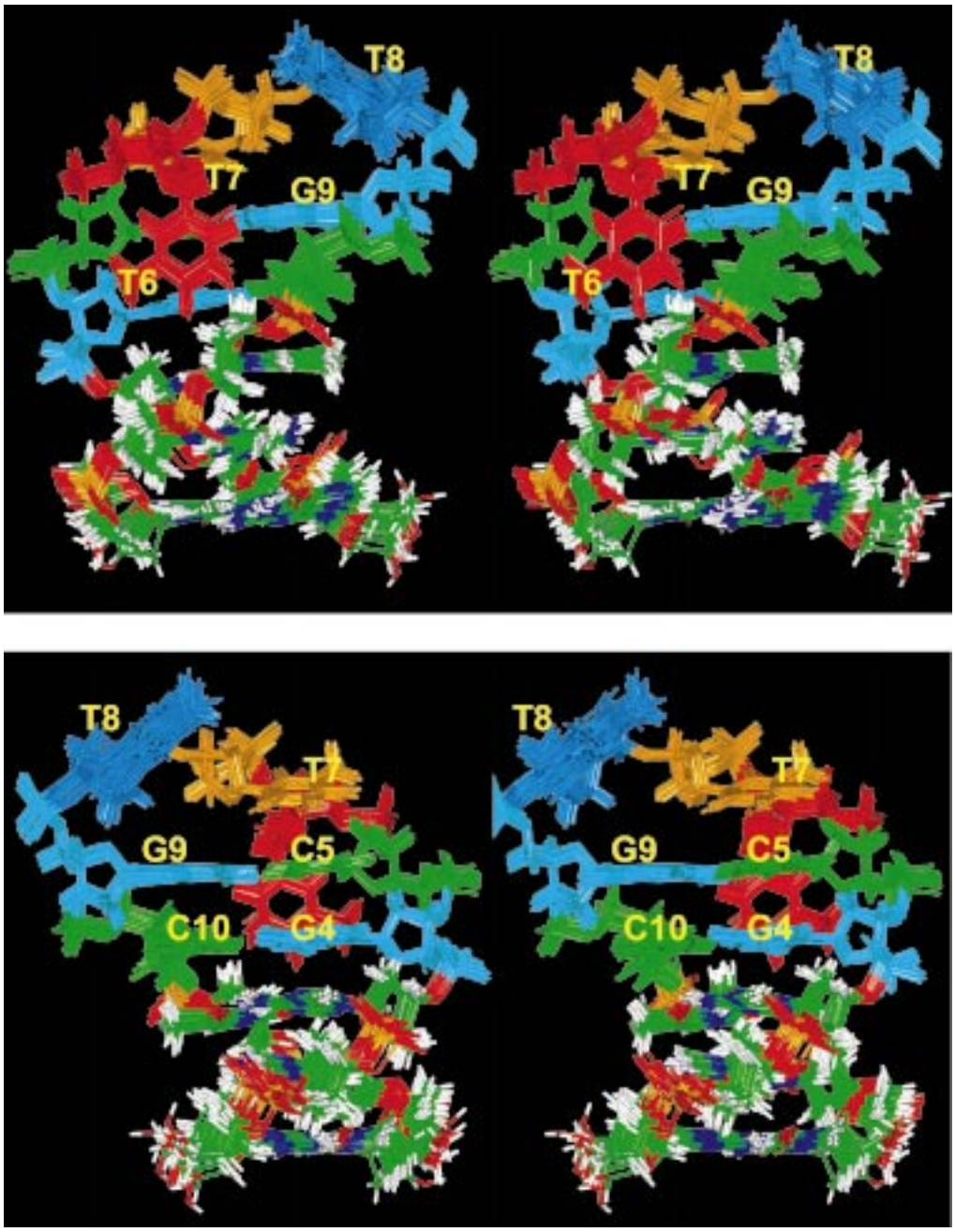

Figure 4. Superimposed wide-eye stereoviews of the 20 final structures of the T_3 mini-loop hairpin in two different orientations. These final structures overlap well with the average structure with rmsd values of 0.46 ± 0.18 Å.

deed present in our current study (data not shown). Of course, conformation-independent sequential assignment of the fingerprint H6/H8–H1′ region is a prerequisite for this approach to be feasible. The aromatic and H1′ proton assignments can be individually accomplished by the through-bond 2D or 3D hetero-TOCSY-NOESY H1′–^{31}P and H6/H8–^{31}P spectra (Kellog et al., 1992; Kellogg and Schweitzer, 1993). We are currently testing this idea using a synthetic RNA sequence with tandem sheared G·A base pairs and a DNA sequence that may form a structure

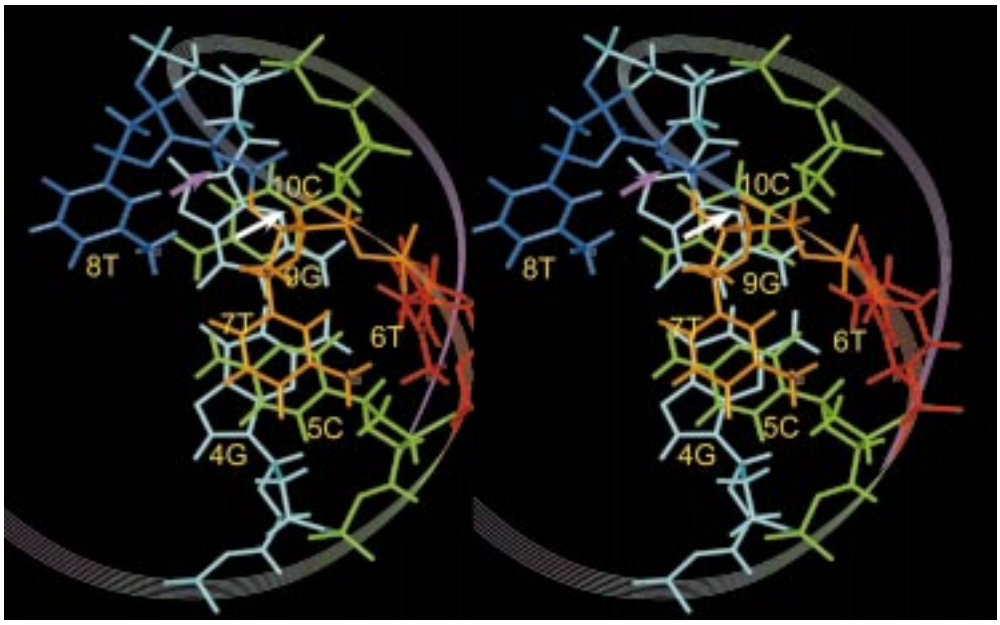

Figure 5. Stereoview of the stacking pattern in the 5′-GC-TTT-CG-3′ motif. Excellent intra-strand stacking of the G4–C5–T7 bases on one side and the G9–C10 on the other side is clear. The backbone between T7–T8 also partially stacks upon the G9 base, causing upfield shifting of the T7H4′ (indicated by a white arrow) and T8H5′ (indicated by a pink arrow) protons. The T6 base folds into the minor groove and forms three H-bonds with the stem G4·C10 and C5·G9 base pairs to stabilize this unusual T_3 mini-loop structure. Both Watson–Crick binding sites of the loop residues T7 and T8, on the contrary, point in parallel toward the major groove and are available for protein binding.

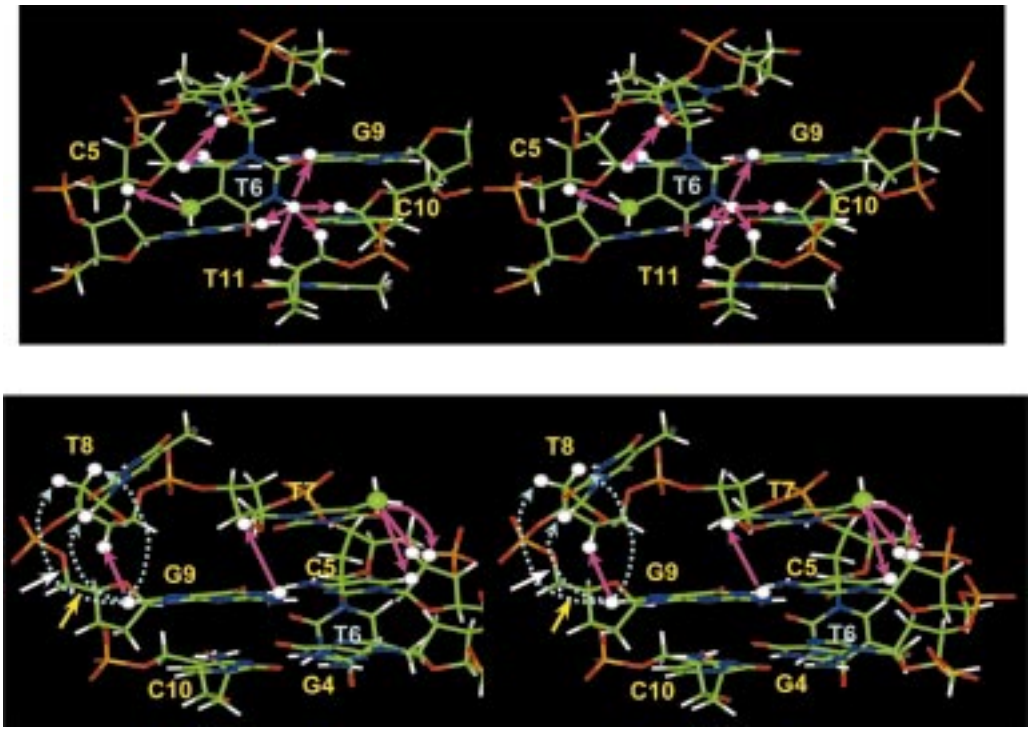

Figure 6. Wide-eye stereoviews of one of the selected T_3 mini-loop structures from the minor groove (top) and major groove (bottom) direction. The protons exhibiting idiosyncratic NOEs are shown as white circles and are connected by pink arrows, while proton pairs exhibiting no NOE even at 240 ms mixing time are connected by dotted blue arrows. The unusual $\epsilon(t)\gamma(t)$ torsion angles adopted by the turning G9 residue are indicated by white and yellow arrows, respectively.

containing no canonical G·C and A·T base pair (data not shown).

It is also interesting to note that almost all backbone ϵ, β, and γ torsion angles could be conveniently extracted by the combination of high-resolution ^{1}H-^{13}C HSQC and ^{1}H-^{31}P correlation experiments without resorting to isotope-enriched samples (Figure 3). The glycosidic χ torsion angle could also be determined either qualitatively by inspecting the intensity difference between the H1′–C8/C6 and H1′–C4/C2 cross peaks (Schmieder et al., 1992) or quantitatively by curve fitting to simulate the HMBC peak patterns (Keeler et al., 1988).

Several other groups are also working on the methodology for solving nucleic acid structures at natural abundance. For example, a ^{1}H-^{13}C-^{31}P triple-resonance 3D NMR experiment is currently available for the H3′/C3′, H4′/C4′, and phosphorus assignments (Krishnamurthy, 1995). Furthermore, a ^{1}H-^{15}N HMQC spectrum at natural abundance has also been reported (Live et al., 1991), which could lead to the imino or amino nitrogen assignment that could provide important information about the base pairing schemes in an unusual nucleic acid structure. At higher magnetic field (> 800 MHz), it may be feasible to run even a 3D ^{1}H-^{13}C HSQC-NOESY at natural abundance. It is therefore possible to study nucleic acid molecules of larger molecular weight at natural abundance.

Both the carbon chemical shifts and one-bond ^{1}H-^{13}C couplings have been found to be sensitive to the ribose conformation and have potential in the structural analysis of nucleic acids (Varani et al., 1996). The chemical shifts of C1′ were found to shift upfield from 92.5 ± 0.9 ppm to 88.2 ppm, while those of C3′, C4′, and C5′ shifted downfield from 72.0 ± 1.4 ppm, 81.8 ± 0.7 ppm, and 64.3 ± 0.9 ppm to 78.3 ppm, 84.9 ppm, and 67.0 ppm, respectively when the sugar puckers are switched from C3′-endo to C2′-endo (Varani et al., 1996). Such significant chemical shift changes of −4.3 ppm, +6.3 ppm, +3.8 ppm, and +2.7 ppm for the C1′, C3′, C4′, and C5′ signals are useful for identifying unusual sugar puckers. Interestingly, the chemical shifts of C3′, C4′, and C5′ of all sugars (including the three unpaired loop thymidines) in our present DNA T_3 mini-loop hairpin have values of 77.0 ± 2.2 ppm, 85.8 ± 1.2 ppm, and 66.4 ± 1.3 ppm, respectively (see Supplementary material). The chemical shifts of C3′, C4′, and C5′ are thus in fact identical for both DNA and RNA C2′-endo sugar puckers. The C1′ chemical shifts of RNA are, however, about 3.3 ppm downfield from those of the C1′ of DNA (88.2 ppm vs. 84.9 ppm), which is most likely due to the inductive effect of the O2′ atom in RNA, since the chemical shifts of the C2′ of RNA are about 35 ppm downfield to those of the C2′ of DNA.

The one-bond sugar $^{1}J_{HC}$ coupling is also sensitive to the sugar pucker (Varani et al., 1996). The $^{1}J_{H1'C1'}$ coupling constants were found to decrease from 176 ± 4 Hz to 169 Hz, while the $^{1}J_{H3'C3'}$ coupling constants increase from 145 ± 3 Hz to 156 Hz and the $^{1}J_{H4'C4'}$ coupling constants from 146 ± 1 Hz to 152 Hz when the RNA sugar puckers are switched from C3′-endo to C2′-endo (Varani and Tinoco, 1991a). However, the corresponding $^{1}J_{H1'C1'}$, $^{1}J_{H3'C3'}$, and $^{1}J_{H4'C4'}$ coupling constants in our DNA T_3 mini-loop hairpin have values of 163 ± 4 Hz, 155 ± 2 Hz, and 147 ± 2 Hz, respectively. Except for the $^{1}J_{H3'C3'}$, both the values of $^{1}J_{H1'C1'}$ and $^{1}J_{H4'C4'}$ coupling constants in the DNA C2′-endo sugars are about 5 Hz smaller than those in the RNA C2′-endo sugars. This difference is not unreasonable, since the RNA C2′-endo sugar has an extra O2′ atom, which may change the scalar interaction between the neighboring H1′–C1′ and H4′–C4′ atoms. However, since the $^{1}J_{H3'C3'}$ values are the same for both DNA and RNA C2′-endo pucker, it can therefore be used as another criterion for identifying the unusual C3′-endo or C2′-endo sugars in the DNA or RNA sequences; a C2′-endo sugar will have a $^{1}J_{H3'C3'}$ value of 155 ± 2 Hz, while a C3′-endo sugar will have a value of 145 ± 3 Hz, about 10 Hz smaller, irrespective of DNA or RNA.

An oligonucleotide that is capable of forming a hairpin structure can also form a duplex structure, albeit with possible mismatched base pairs or bulges. Indeed, an RNA 11-mer r(GGCG-UUU-CGCC) at NMR conditions (1.35 mM RNA, 10 mM phosphate, pH 6, 20 °C) was found to adopt mostly a duplex structure with three consecutive U·U base pairs (Sich et al., 1997). The monomer hairpin was only predominant when the stem length was increased to eight base pairs. We thus carefully analyzed our data to see if the stable species we observed is a hairpin with a T_3 loop or a duplex with three consecutive T·T base pairs or two TTT bulges. Several pieces of evidence indicate that the stable species is indeed a hairpin, not a duplex. First, the one-dimensional imino proton spectrum has two sharp (6T and 7T) and one broad (8T) imino proton resonances. Two-dimensional H_2O/NOESY contains many NOEs resulting from the 6T imino proton. On the contrary, only weak NOEs from 7T-imino to 5CNH_2/5CH6 were observed, and no NOEs from 8T-imino could be detected at all. The

unequal linewidths and different NOEs for these three T-imino protons are thus consistent with a hairpin structure with a T_3 mini-loop. 6T is folded into the minor groove with its imino proton forming an H-bond with the C10O2 atom. It thus has the slowest exchange rate and exhibits stronger NOEs to its surrounding protons. 7T stacks upon 5C and also has a slow exchange rate. However, its imino proton is exposed to the solvent and is only close to the $5CNH_2$ and 5CH5/H6 protons. Therefore only weaker NOEs to these nearby protons were observed. On the contrary, the 8T imino proton is quite broad and did not exhibit any NOE at all. This is consistent with the fact that the 8T base extrudes into the major groove and exchanges fast with the H_2O solvent. The unequal linewidths and different NOEs for these T-imino protons thus argue against a duplex with three consecutive T·T mismatches; three equal sharp resonances would be observed in such a case (Sich et al., 1997). Second, it is impossible for a duplex with either three consecutive T·T mismatches or two TTT bulges to exhibit the special tertiary NOE constraints observed for the 6T-imino protons. The 6T-imino to 4G-imino, 10CH1′/H2′, or 11TH1′/H4′/H5′/H5″ NOEs are only possible when 6T folds into the minor groove of a hairpin structure. A duplex with T·T mismatches or TTT bulges containing such unusual tertiary constraints would result in a highly distorted structure. Finally, no new imino proton peaks appeared when the sample concentration was diluted two or three times, which indicates that the species we observed is already a monomeric hairpin, not a dimeric duplex. Since the RNA hairpin studied by Sich et al. (1997) contains a 5′-CG-UUU-CG-3′ motif and adopted completely different loop conformations, it is unclear if an RNA 5′-GC-UUU-GC-3′ loop motif will form a similar folded-in structure as in DNA. But the RNA 5′-GC-UU-GC-3′ motif published by Jucker and Pardi (1995) did contain a similar loop conformation as those in DNA. It is also interesting to note that a purine-rich sequence $d(A_2G_2T_4A_2G_2)$ also contains a T_3 loop structure that is almost identical to what we publish here (Kuryavyl et al., 2000).

Conclusions

Detailed heteronuclear NMR studies of the T_3 mini-loop hairpin of the AAV2 terminal repeat have been performed at natural abundance. Combination of ^{1}H-^{13}C HSQC, HMBC, and ^{1}H-^{31}P correlation spectra has allowed for measurement of many parameters that are critical in determining the three-dimensional nucleic acid structure. The stereospecific $H5'^{(pro-S)}$ and $H5'^{(pro-R)}$ assignment solves one of the major obstacles in determining nucleic acid structures using an NMR approach. Continuing development in this area may allow for determination of oligomers of larger molecular weight without resorting to the preparation of isotope-enriched samples.

Acknowledgements

We thank the National Science Council and the Chung-Zhen Agricultural Foundation Society of Taiwan, ROC for the instrumentation grants. This work was supported by the NSC grant 88-2113-M-005-016 to S.H.C.

References

Altona, C. (1982) *Recl. Trav. Chim. Pays-Bas*, **101**, 413–433.
Avizonis, D.Z. and Kearn, D.R. (1995) *Nucleic Acids Res.*, **23**, 1260–1268.
Bax, A., Griffey, R.H. and Hawkins, B.L. (1983) *J. Magn. Reson.*, **55**, 301–315.
Baxter, S.M., Greizerstein, M.B., Kushlan, D.M. and Ashley, G.W. (1993) *Biochemistry*, **32**, 8702–8711.
Blommers, M.J.J., Walters, J.A.L.I., Haasnoot, C.A.G., Aelen, J.M.A., van der Marel, G.A., van Boom, J.H. and Hilbers, C.W. (1989) *Biochemistry*, **28**, 7491–7498.
Boulard, Y., Gabarro-Arpa, J., Cognet, J.A.H., Bret, M.L., Guy, A., Teoule, R., Guschlbauer, W. and Fazakerley, G.V. (1991) *Nucleic Acids Res.*, **19**, 5159–5167.
Butcher, S.E., Allain, F.H.-T. and Feigon, J. (1999) *Nat. Struct. Biol.*, **6**, 212–216.
Chou, S.-H., Cheng, J.-W., Fedoroff, O. and Reid, B.R. (1994) *J. Mol. Biol.*, **241**, 467–479.
Chou, S.-H. and Tseng, Y.-Y. (1999) *J. Mol. Biol.*, **285**, 41–48.
Chou, S.-H., Tseng, Y.-Y., Chen, Y.-R. and Cheng, J.-W. (1999a) *J. Biomol. NMR*, **14**, 157–167.
Chou, S.-H., Tseng, Y.-Y. and Chu, B.-Y. (1999b) *J. Mol. Biol.*, **292**, 309–320.
Chou, S.-H., Tseng, Y.-Y. and Wang, S.-W. (1999c) *J. Mol. Biol.*, **287**, 301–313.
Chou, S.-H., Zhu, L., Gao, Z., Cheng, J.-W. and Reid, B.R. (1996a) *J. Mol. Biol.*, **264**, 981–1001.
Chou, S.-H., Zhu, L. and Reid, B.R. (1997) *J. Mol. Biol.*, **267**, 1055–1067.
Chou, S.-H., Zhu, L. and Reid, R.R. (1996b) *J. Mol. Biol.*, **259**, 445–457.
Dai, X., Greizerstein, M.B., Nadas-Chinni, K. and Rothman-Denes, L.B. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 2174–2179.
Dai, X., Kloster, M. and Rothman-Denes, L.B. (1998) *J. Mol. Biol.*, **283**, 43–58.
Dallas, A. and Moore, P.B. (1997) *Structure*, **5**, 1639–1653.
Hines, J.V., Landry, S.M., Varani, G. and Tinoco Jr., I. (1994) *J. Am. Chem. Soc.*, **116**, 5823–5831.
Jones, R.A. (1994) *Methods Mol. Biol.*, **26**, 207–231.

Jucker, F.M. and Pardi, A. (1995) *Biochemistry*, **34**, 14416–14427.
Keeler, J., Neuhaus, D. and Titman, J.J. (1988) *Chem. Phys. Lett.*, **146**, 545–548.
Kellogg, G.W. and Schweitzer, B.I. (1993) *J. Biomol. NMR*, **3**, 577–595.
Kellogg, G.W., Szewczak, A.A. and Moore, P.B. (1992) *J. Am. Chem. Soc.*, **114**, 2727–2728.
Kolk, M.H., Wijmenga, S.S., Heus, H.A. and Hilbers, C.W. (1998) *J. Biomol. NMR*, **12**, 423–433.
Kotin, R.M. (1994) *Human Gene Therapy*, **5**, 793–801.
Kotin, R.M., Linden, R.M. and Berns, K.I. (1992) *EMBO J.*, **11**, 5071–5078.
Krishnamurthy, V.V. (1995) *J. Magn. Reson.*, **B108**, 117–120.
Kuklenyik, Z., Yao, S. and Marzilli, L.G. (1996) *Eur. J. Biochem.*, **236**, 960–969.
Kuryavyl, V., Kettani, A., Wang, W., Jones, R. and Patel, D.J. (2000) *J. Mol. Biol.*, **295**, 456–469.
Live, D.H., Radhakrishnan, I., Misra, V. and Patel, D.J. (1991) *J. Am. Chem. Soc.*, **113**, 4687–4688.
Marino, J.P., Schwalbe, H., Glaser, S.J. and Griesinger, C. (1996) *J. Am. Chem. Soc.*, **118**, 4388–4395.
Marino, J.P., Schwalbe, H. and Griesinger, C. (1999) *Acc. Chem. Res.*, **32**, 614–623.
McClellan, J.A., Bouoblikova, P., Palecek, E. and Lilley, D.M.J. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 8373–8377.
Mooren, M.M.W., Pulleyblank, D.E., Wijmenga, S.S., van de Ven, F.J.M. and Hilbers, C.W. (1994) *Biochemistry*, **33**, 7315–7325.
Mueller, L., Legault, P. and Pardi, A. (1995) *J. Am. Chem. Soc.*, **117**, 11043–11048.
Muzyczka, N. (1991) *Semin. Virol.*, **2**, 281–290.
Muzyczka, N. (1992) *Curr. Top. Microbiol. Immunol.*, **158**, 97–129.
Nikonowicz, E.P., Sirr, A., Legault, P., Jucker, F.M., Baer, L.M. and Pardi, A. (1992) *Nucleic Acids Res.*, **20**, 4507–4513.
Pardi, A. (1995) *Methods Enzymol.*, **261**, 350–380.
Patel, D.J., Suri, A.K., Jiang, F., Jiang, L., Fan, P., Kumar, R.A. and Nonin, S. (1997) *J. Mol. Biol.*, **272**, 645–664.
Plateau, P. and Gueron, M. (1982) *J. Am. Chem. Soc.*, **104**, 7310–7311.
Ren, J., Qu, X., Chaires, J.B., Trempe, J.P., Dignam, S.S. and Dignam, J.D. (1999) *Nucleic Acids Res.*, **27**, 1985–1990.
Ryan, J.H., Zolotukhin, S. and Muzyczka, N. (1996) *J. Virol.*, **70**, 1542–1553.
Sarma, R.H., Mynott, R.J., Wood, D.J. and Hruska, F.E. (1973) *J. Am. Chem. Soc.*, **95**, 6457–6459.
Schmieder, P., Ippel, J.H., van den Elst, H., van der Marel, G.A., van Boom, J.H., Altona, C. and Kessler, H. (1992) *Nucleic Acids Res.*, **20**, 4747–4751.
Shaka, A.J., Barker, P.B. and Freeman, R.J. (1985) *J. Magn. Reson.*, **64**, 547–552.
Sich, C., Ohlenschlager, O., Ramachandran, R., Gorlach, M. and Brown, L.R. (1997) *Biochemistry*, **36**, 13989–14002.
Sklenar, V., Miyashiro, H., Zon, G., Miles, H.T. and Bax, A. (1986) *FEBS Lett.*, **208**, 94–98.
Tseng, Y.-Y. and Chou, S.-H. (1999) *J. Chin. Chem. Soc.*, **46**, 699–706.
Varani, G., Aboul-ela, F. and Allain, F.H.-T. (1996) *Prog. NMR Spectrosc.*, **29**, 51–127.
Varani, G. and Tinoco Jr., I. (1991a) *J. Am. Chem. Soc.*, **113**, 9349–9354.
Varani, G. and Tinoco Jr., I. (1991b) *Quart. Rev. Biophys.*, **24**, 479–532.
Zhu, L., Chou, S.-H. and Reid, B.R. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 12159–12164.
Zimmer, D.P. and Crothers, D.M. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 3091–3095.