Journal of Biomolecular NMR, **21:** 307–319, 2001.
KLUWER/ESCOM

Novel cross-strand three-purine stack of the highly conserved 5′-GA/AAG-5′ internal loop at the 3′-end termini of *Parvovirus Genomes*

Shan-Ho Chou* & Ko-Hsin Chin
Institute of Biochemistry, National Chung-Hsing University, Taichung, 40227, Taiwan

Received 7 June 2001; Accepted 18 October 2001

Key words: asymmetric internal loop, cross-strand stack, parvovirus genome, sheared GA pairing, three-purine stack

Abstract

We have used two-dimensional nuclear magnetic resonance (2D-NMR), distance geometry (DG) and molecular dynamics / energy minimization (MD/EM) methods to study a 2 × 3 asymmetric internal loop structure of the highly conserved '5′-(GA)/(AAG)-5′ bubble' present at the 3′-end hairpin of the single-stranded DNA genome of parvoviruses. This motif contains an unpaired adenosine stacked between two bracketed sheared G•A pairs. However, the phenomenal cross-strand G-G and A-A stacking in the tandem sheared G•A pairs has undergone considerable change. A novel three-purine stacking pattern is observed instead; the inserted A18 base is completely un-stacked from its neighboring G17 and A19 bases, but well stacked with the cross-strand A4 and G3 bases to form a novel A4/A18/G3 stack that is different from the double G/G, A/A or quadruple G/G/G/G stack present in the 5′-(GA)/(AG)-5′ or 5′-(GGA)/(AGG)-5′ motifs. Unlike the bulged purine residue that usually causes about 20 degree kink in the helical axis of the parent helix when bracketed by canonical G•C or A•T base pairs, no significant kink is observed in the present helix containing a bulged-adenine that is bracketed by sheared G •A pairs. The phosphodiesters connecting G3-A4 and G17-A18 residues adopt unusual ζ torsional angles close to the *trans* domain, yet that connecting A18-A19 residues resumes the normal ζ(g^-) value. The well structured '5′-(GAA)/(AG)-5′' internal loop in the parvovirus genomes explains its resistance to single-strand specific endonuclease susceptibility.

Introduction

Parvoviruses are unique viruses of linear single-stranded DNA genomes with complex Y-shaped 3′-termini structure that contain a highly conserved asymmetric 2 × 3 '5′-(GA)/(AAG)-5′ bubble' close to its replication origin (Astell et al., 1979). These viruses replicate by the so-called 'rolling hairpin' mechanism (Astell et al., 1985), in which the replication fork flipped back and forth along the linear genome by the sequential synthesis and rearrangement of the palindromic viral termini into hairpin structures (Willwand et al., 1998). Interestingly, while the 5′-end telomere was found to incorporate heterogeneous sequences, the 3′-end telomere was instead found to incorporate a unique sequence containing a 5′-(GA)/(AAG)-5′ 'bubble' (Astell et al., 1985). This non-symmetrical synthesis of the 3′-end telomere was presumably due to the asymmetric resolution of the virus dimeric replication form caused by the crucial selection of a single cutting site of the virally coded NS1 protein (Cotmore et al., 1993; Cotmore and Tattersall, 1994; Liu et al., 1994). Since the 5′-(GA)/(AAG)-5′ internal loop sequence is close to the 3′-hairpin that acts as a primer to initiate the genome synthesis, has been found to be highly conserved, and is resistant to single-strand specific endonuclease, the unusual structure it adopts may play an important role in the life cycle of parvoviruses. We are thus endeavored to solve its structure using high-resolution NMR technique.

Cross-stand base stacking has been found to be a major factor in stabilizing unusual nucleic acid structures. In RNA, cross-strand stacks resulting from

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

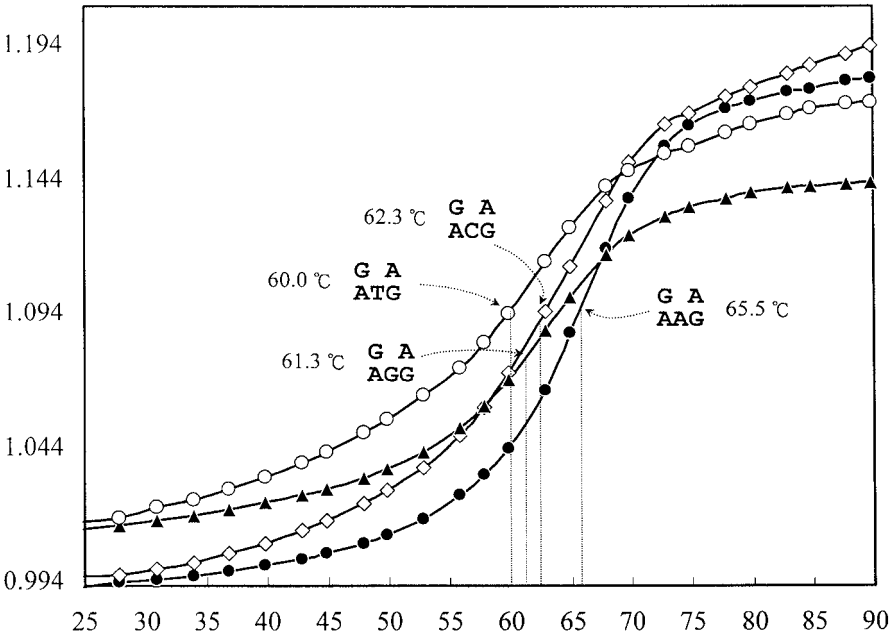

Figure 1. The absorbance (OD) versus temperature melting curves for the d(GC**GA**GTCG*CTTG*CGAC**GAA**GC) hairpin (abbreviated as GA/AAG thereafter), d(GC**GA**GTCG*CTTG*CGAC**GGA**GC) hairpin (GA/AGG), d(GC**GA**GTCG*CTTG*CGAC**GCA**GC) hairpin (GA/ACG), and d(GC**GA**GTCG*CTTG*CGAC**GTA**GC) hairpin (GA/ATG). The T_m values were determined from the maximum of the first differential curves.

paired bases have been observed in several crystal or solution structures (Cate et al., 1996a,b; Correll et al., 1997; Moore, 1999; Pley et al., 1994; Zimmermann et al., 1997). While in DNA, cross-strand base stacking can result from either paired or unpaired purine bases and has been repetitively observed in several purine-rich sequences (Chou et al., 1997; Shepard et al., 1998; Spackova et al., 2000). In the paired case, for examples, the tandem sheared $(GA)_2$ motif is engaged in phenomenal cross-strand G/G and A/A stacking in the 5′-(PyGAPu)/(PuAGPy)-5′ or 5′-(PyGAPy)/(PuAGPu)-5′ segment (Chou et al., 1994; Chou et al., 1999a; Gao et al., 2000). Similarly, a mixed cross-strand purine-purine G/A and purine-pyrimidine A/C stacks were observed in the 5′ (PyGCPu)/(PuAAPy)-5′ segment (Chou and Tseng, 1999). In the unpaired case, such cross-strand stacking is even more prominent. Thus in the highly tandem $(TGGAA)_n$ repeats in the human centromere, the two unpaired central guanines in the $(GGA)_2$ motif are engaged in cross-strand stacking to form a G/G/G/G stack with the bracketed sheared G•A base pairs (Chou et al., 1994b; Zhu et al., 1995). Quadruple intercalation motif with four inter-strand guanine stacking has also recently been discovered in the $(GGGA)_2$ motif embedded in a hairpin stem to form a G/G/G/G/G/G stack (Chou and Chin, unpublished result). Interestingly, the canonical paired G•C or A•T bases can even be split apart to form cross-strand stack when bracketed by sheared G•A pairs (Chou and Chin, 2001). The stable formation of such motifs do point out that cross-strand stacking plays a significant role in stabilizing double helical nucleic acid structures of unusual nature.

In this respect, we now report a novel cross-stand three-purine stack in the highly conserved asymmetric 2 × 3 '5′-(G_3A_4)/$(A_{19}A_{18}G_{17})$-5′ bubble' at the 3′-end termini of the parvovirus genomes. Unlike the cross-stand purine stacks encountered in the $(GA)_2$ or $(GGA)_2$ motifs that exhibit excellent G-2, A-2, or G-4 stack, an odd three-purine stacking motif is observed instead, with the bulged A18 residue completely unstacked from its neighboring G17 and A19 bases, but well stacked with the cross-strand A4 and G3 bases to form a mixed purine G3/A18/A4 stack. Such compact 5′-(GA)/(AAG)-5′ internal loop structure in the parvovirus genomes can well explain its resistance to single-strand specific endonuclease susceptibility.

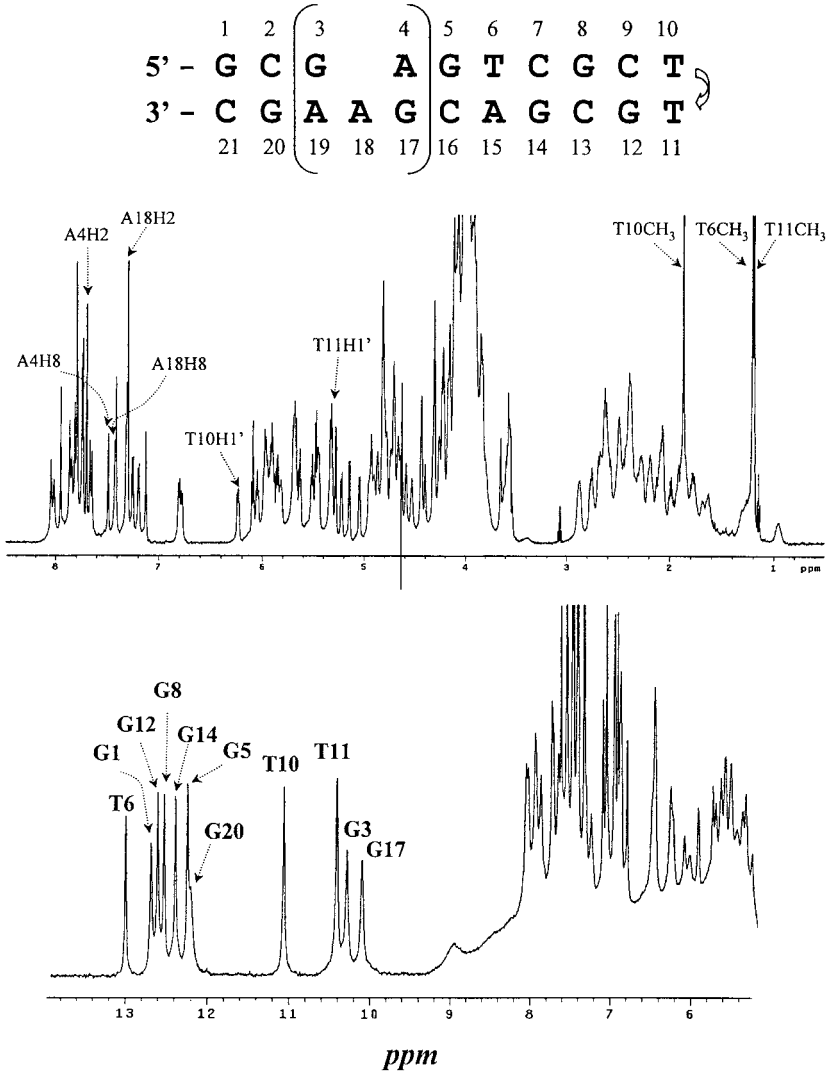

Figure 2. The one-dimensional exchangeable and non-exchangeable proton NMR spectrum of the GA/AAG hairpins at 600 MHz. Assignments of the exchangeable protons were made from NOESY experiments in 90% H_2O/10% D_2Osolution at 0 °C while that of the non-exchangeable protons from NOESY experiments in 100% D_2O solution at 25 °C. Several proton signals of particular interest were labeled and marked by arrows.

Materials and methods

Sample preparation

All DNA samples were synthesized in 3 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 before (Chou and Tseng, 1999).

UV melting studies

Absorbance (OD) versus temperature profile was obtained at 260 nm with a Cary 100 photospectrometer equipped with a temperature controller. A temperature probe was placed inside the UV chamber to monitor the cell temperature. The temperature in each run was increased from 25 °C to 90 °C at a rate of 0.5 °C/min. The melting temperature was calculated from the first differential of the melting curves using the program supplied by the vendor.

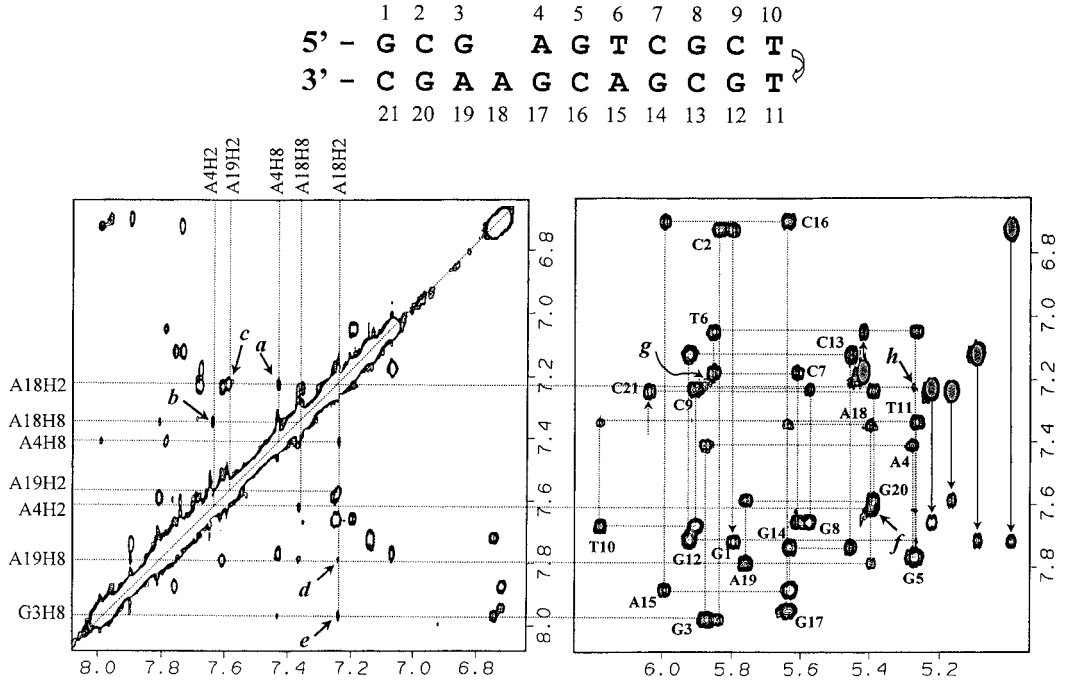

Figure 3. The expanded 2D-NOESY in D_2O of the GA/AAG hairpin at 25 °C and 600 ms mixing time. The base-base region was shown in left while base-H1′/H5 region in right. Some critical base-base NOEs in the internal loop region were marked by arrows in left, i.e., (a) A4H8-A18H2; (b) A4H2-A18H8; (c) A19H2-A18H2; (d) A18H2-A19H8; (e) A18H2-G3H8. The sequential base-H1′ connectivity were followed in the right figure with the intra-residue H6/H8-H1′ NOEs labeled with residue numbers. Several unusual NOEs were also marked, namely, (f) A4H2-A18H1′; (g) A18H2-G3H1′; (h) A18H2-A4H1′.

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 jump-return pulse sequence (Plateau and Gueron, 1982). The spectral width was 12 000 Hz with the carrier frequency set at the resonance of water. The maximum excitation was set at 12.5 p.p.m. For each experiment, 12K 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 0 °C in a pH 6.8 low salt (10 mM) buffer with the following parameters; delay time 1 s, mixing time 0.12 s, spectra width 11 000 Hz, complex points 2048, number of transients 112, and number of increments 300.

NOESY experiments in D_2O were carried out at 25 °C in the hypercomplex mode with a spectral width of 4750 Hz. Spectra were collected using three mixing times of 100, 300, and 600 ms with a relaxation delay of 1 s between each transient and with 2048 complex points in the t_2 and 225 complex points in the t_1 dimension. For each t_1 incrementation, 40 scans were averaged.

A DQF-COSY spectrum was collected in the TPPI mode with a spectral width of 4750 Hz in both dimensions; 2048 complex points in the t_2 dimension and 350 (real) points in the t_1 dimension were collected with a relaxation delay of 0.8 seconds, and 40 scans were averaged for each t_1 incrementation.

A proton-detected ^{31}P-^{1}H heteronuclear correlation spectrum (Sklenar et al., 1986) was collected in the TPPI mode with a spectral width of 4750 Hz in the ^{1}H dimension and a spectral width of 1200 Hz in the ^{31}P dimension. 1024 complex points in the t_2 (^{1}H) dimension and 100 complex points in the t_1 (^{31}P) dimension were collected. Protons were presaturated for 0.9 s and 128 scans were accumulated for each t_1 incrementation.

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

Structure determination

Three-dimensional structures of the d(GCGAGTCG CTTGCGACGAAGC) hairpin were generated by distance geometry and molecular dynamics calculations using distance and torsional angle constraints derived from NMR experiments. Most distance constraints derived from NOESY spectra in D_2O were classified as strong, medium, or weak based on their relative intensities at 100 and 300 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. 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 were also derived from H_2O/NOESY spectra and were given only two wide distance bounds of either 2.0–5.0 Å or 3.0–6.0 Å, due to the exchange phenomena. The β and γ torsional angle constraints were determined primarily semi-quantitatively from the ^{31}P-^{1}H heteronuclear correlation data (Chou et al., 1996) using the in-plane 'W rule' (Sarma et al., 1973). If the long-range (n)P ↔ (n)H4′ four-bond couplings were detected, then the β and γ torsional angles were constrained to the *trans* (180° ± 30°) and *gauche*$^+$ (60° ± 30°) domains, respectively. Otherwise they were left unconstrained. The ε torsional angle can only locate in either the *trans* or *gauche*$^-$ domain (Altona, 1982). The *gauche*$^+$ conformation is not sterically allowed. Based on the absence of long-range $^4J_{H2'-P}$ coupling, all ε torsion angles were constrained to the trans domain (180° ± 30°). The ζ and α dihedral angles were all left unconstrained. The χ dihedral angles were constrained to −100° (ideal B-DNA values) ± 30° when no base-H1′ cross-peaks of comparable intensity to the CH5/CH6 cross-peaks was detected. These NOE distance (242 in total) and torsional angle (112 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.). Well-converged final structures with pair-wise r.m.s.d. value of 1.35 ± 0.26 Å were obtained after molecular dynamics calculations.

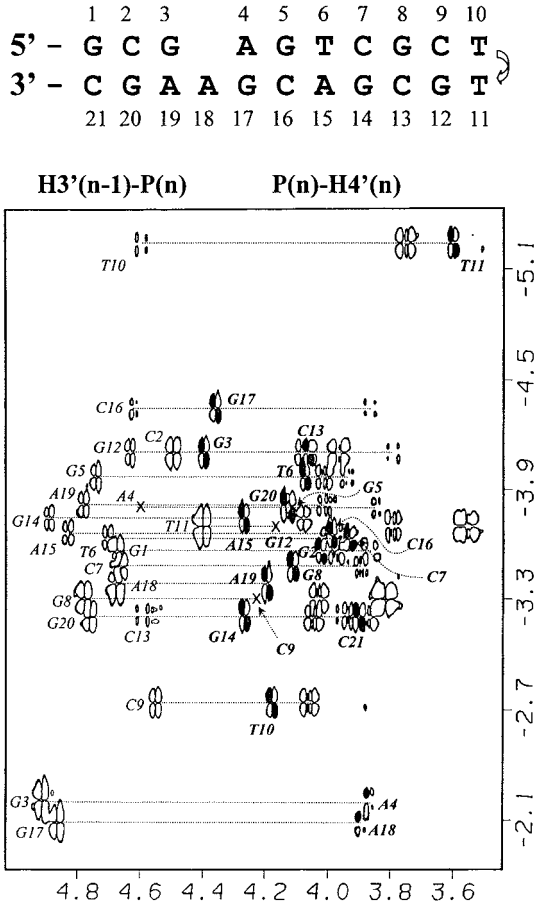

Figure 4. The proton-detected ^{1}H-^{31}P heteronuclear correlation spectrum of the GA/AAG 21-mer hairpin. Nineteen (n−1)H3′–(n)P cross peaks (A4 cross peak is missing because its chemical shift is coincident with that of water) and eighteen (n)P-(n)H4′ cross peaks are observable (C9 and G12 cross peaks are missing due to the different torsional angles in the CTTG loop region). Two upfield phosphodiester signals (G3-A4 and G17-A18) at approximately −2.1 ppm are observed. Interestingly, the A18-A19 phosphodiester signal has resumed normal ^{31}P chemical shift value.

Results

Thermodynamic studies

Figure 1 shows the UV melting curves from 25 to 90 °C under a low salt condition (20 mM NaCl and 10 mM pH 6.8 sodium phosphate buffer) for the four 5′-GC***GA***GTCG*CTTG*CGAC***GXA***GC-3′ oligonucleotides containing the internal loop motifs of 5′-(GA)/(A**A**G)-5′, 5′-(GA)/(A**G**G)-5′, 5′-(GA)/(A**C**G)-5′, and 5′-(GA)/(A**T**G)-5′ sequences respectively in the stem region. All such oligomers exhibit well-behaved transition curves with melting temperatures of 65.5 °C, 61.3 °C, 62.3 °C, and 60.0 °C,

respectively, indicating good structural formation of these asymmetrical internal loops. The two strands are connected by a stable CTTG tetraloop (Ippel et al., 1995), and the hairpin formation of such oligomers is demonstrated by the concentration-independence of the melting temperatures; no change in the melting temperature was observed when the sample concentration was diluted up to fifty times. NMR data also reveal the characteristic NOEs expected for the CTTG tetraloop (see later section).

NMR studies

The one-dimensional exchangeable and non-exchangeable proton spectra at a neutral (pH 6.8) low salt buffer condition at 0 °C for the 5′-(GA)/(A**A**G)-5′ 21mer were shown in Figure 2. The 5′-(GA)/(AAG)-5′ internal loop was bracketed by a pair of G•C base pair to increase its stability, although good NMR spectra was also obtained when it is bracketed by a pair of A•T base pair as in the original sequence (Astell et al., 1979). The exchangeable proton spectrum was assigned by 2D-NOESY in 90% H_2O/10% D_2O as previously described (Tseng and Chou, 1999), while that of non-exchangeable protons by 2D-NOESY in 100% D_2O using the standard sequential assignment procedure (Hare et al., 1983). Except for the seven expected imino proton signals participating in the canonical base pairing in the 12–13 ppm region, four imino proton signals situating in the unpaired 10–11 ppm region were clearly observed. The two imino proton signals at approximately 10.2 ppm are typical of the unpaired G imino protons belonging to the two bracketed sheared G•A base pairs (Cheng et al., 1992). The right one of the two was assigned as due to residue G17 from the observation of two medium strength NOEs to the $C16NH_2$ protons, while the left one to residue G3 from the observation of two medium strength NOEs to the $C2NH_2$ protons (data not shown). The two imino proton signals at approximately 10.4 and 11.1 ppm belong, on the other hand, to the two unpaired thymine residues in the CTTG loop (Chou and Chin, unpublished result). Although unpaired (as demonstrated by their characteristic chemical shifts), they do exhibit sharp resonance peaks, possibly due to the protection from solvent exchange through the folding of the first thymine into the minor groove and the stacking of the second thymine upon stem residue C9. Such features have been repetitively observed in several NMR spectra of pyrimidine-rich loop sequences, i.e., the TCC (Chou et al., 1999b), TTT (Chou et al., 2000), TTTG (Chou et al., 2001), and CTTG (Chou and Chin, unpublished result) sequences etc. The formation of CTTG loop is also corroborated by the observation of consistently downfield shifting of the looped-out T10H1, T10H1′, and $T10CH_3$ protons as marked in Figure 2 (Chou and Chin, unpublished result). The rather narrow linewidth of the A4H8/H2 and A18H8/H2 protons (marked in Figure 2) and the abundant NOEs they exhibit (shown in Figure 3) indicate that they are not subject to local motion possibly present in such kind of internal loop. These data, along with the excellent chemical shift dispersion, and the observation that only one set of signal is present, indicating that the present 2×3 internal loop motif is stable in NMR time scale and suitable for NMR studies.

Figure 4 shows the expanded base-base and base-H1′/H5 D_2O/NOESY at 25 °C for the GA/AAG 21-mer. The spectrum was collected at a mixing time of 600 ms to reveal weaker AH2-related NOEs. However, the distances were judged from the 100 ms and 300 ms NOESY to prevent possible spin diffusion error. The sequential base-H1′ connectivity (right figure) were followed by the well-established procedure without much difficulty (Hare et al., 1983). The most important feature of the spectrum is a strong NOE cross peak from A4H2 to A18H1′ (cross peak f). This, along with the weak cross peaks from A18H2 to A4H1′ (cross peak h) and A18H2 to G3H1′ (cross peak g), strongly suggest that residue A18 is stacked between residues G3 and A4. This conclusion is further corroborated by the observation of weak NOEs in the base proton - base proton region (left figure), i.e., those between A4H8-A18H2 (cross peak a), A4H2-A18H8 (cross peak b), A18H2-A19H2 (cross peak c), A18H2-A19H8 (cross peak d), and A18H2-G3H8 (cross peak e) protons.

Anti glycosidic torsional angle and C2′-endo sugar conformation were adopted for all residues, including those in the GA/AAG internal loop, as demonstrated by the relatively weak intra-residue H6/H8-H1′ NOEs in the NOESY and by the strong *J*-coupled H1′-H2′/H2″ cross peaks, absent H2″-H3′ cross peaks, and weak H3′-H4′ cross peaks in the DQ-COSY spectrum (data not shown). This is different from the sugar puckers adopted for the unpaired guanine residues in the $(GGA)_2$ motif, in which C3′-endo conformation are adopted to extend the backbone for intercalation (Chou et al., 1994b).

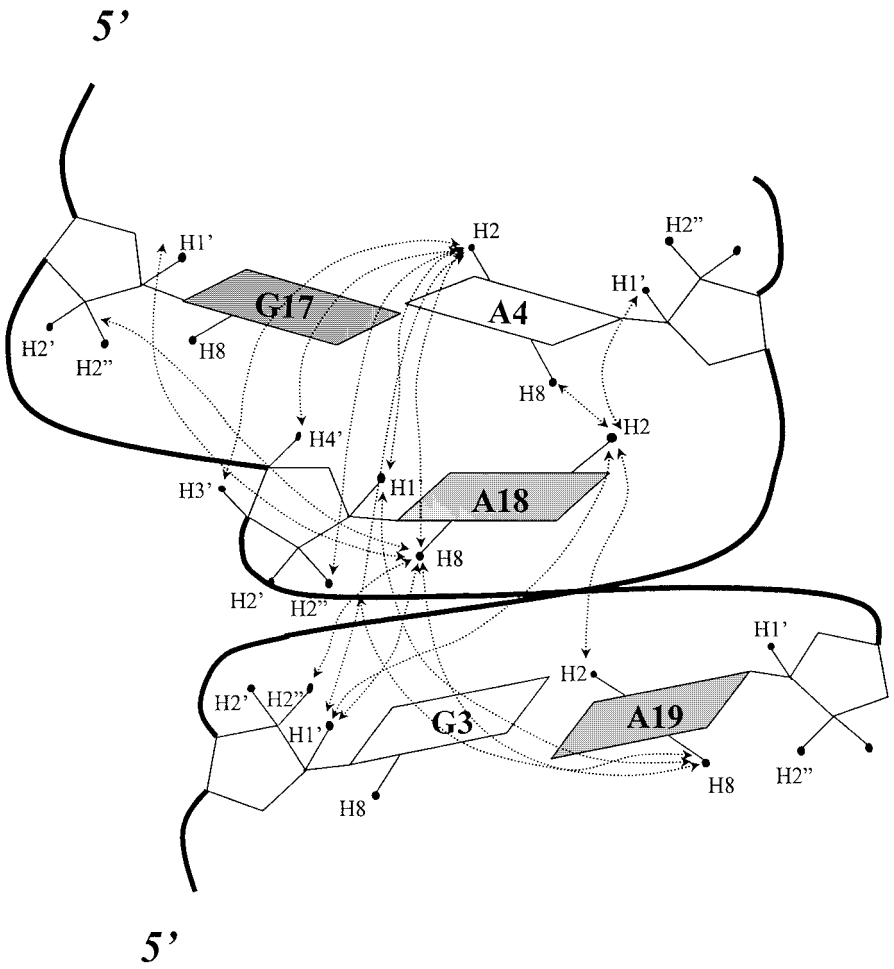

Figure 5. The schematic idiosyncratic NOEs in the GA/AAG internal loop. The intercalation of unpaired A18 residue into G3-A4 step is clearly revealed by the detection of cross-strand NOEs between the A18H2-A4H8, A18H2-A4H1′, A4H2-A18H1′, A4H2-A18H8, A18H2-G3H1′, A18H8-G3H1′, and A18H8-G3H2′ proton pairs.

Backbone torsional angles

Figure 4 shows the ^{31}P-^{1}H heteronuclear correlation spectrum that provides important information on the backbone conformation. The horizontal dotted lines connect the (n−1)H3′-(n)P cross peak to the (n)P-(n)H4′ cross peak. One very downfield (−5.3 ppm) and two very upfield ^{31}P signals are detected. The very downfield one was assigned to be the T10-T11 backbone phosphodiester signal. Such phenomenon has also been observed in a d(CTTG) loop hairpin (Chou et al., unpublished result). Similar to the tandem sheared $(GA)_2$ motif, two characteristic upfield ^{31}P signals belonging to the G3-A4 and G17-A18 backbone phosphodiesters were observed (Chou et al., 1992a). However, ^{31}P signal belonging to the A18-A19 phosphodiester has resumed the normal chemical shift. In this respect, the 5′-(GA)/(A**A**G)-5′ motif is more similar to the 5′-(GA)/(AG)-5′ motif than to the 5′-(GGA)/(AGG)-5′ motif, in that two upfield phosphorous signals were also observed for the 5′-(GA)/(AG)-5′ motif (Chou et al., 1992a; Chou et al., 1999a) while only regular phosphorous signals were observed for the 5′-(GGA)/(AGG)-5′ motif (Chou et al., 1994b). However, the upfield shifting of the phosphorous signals in this 2 × 3 internal loop (approximately 1.5 ppm) is not as large as that in the 2 × 2 internal loop of the tandem sheared $(GA)_2$ motif (approximately 2 ppm), possibly due to the minor backbone adjustment necessary to accommodate this asymmetric internal loop.

Several cross peaks are absent in the spectrum. The A4H3′-P5 cross peak is not observable because

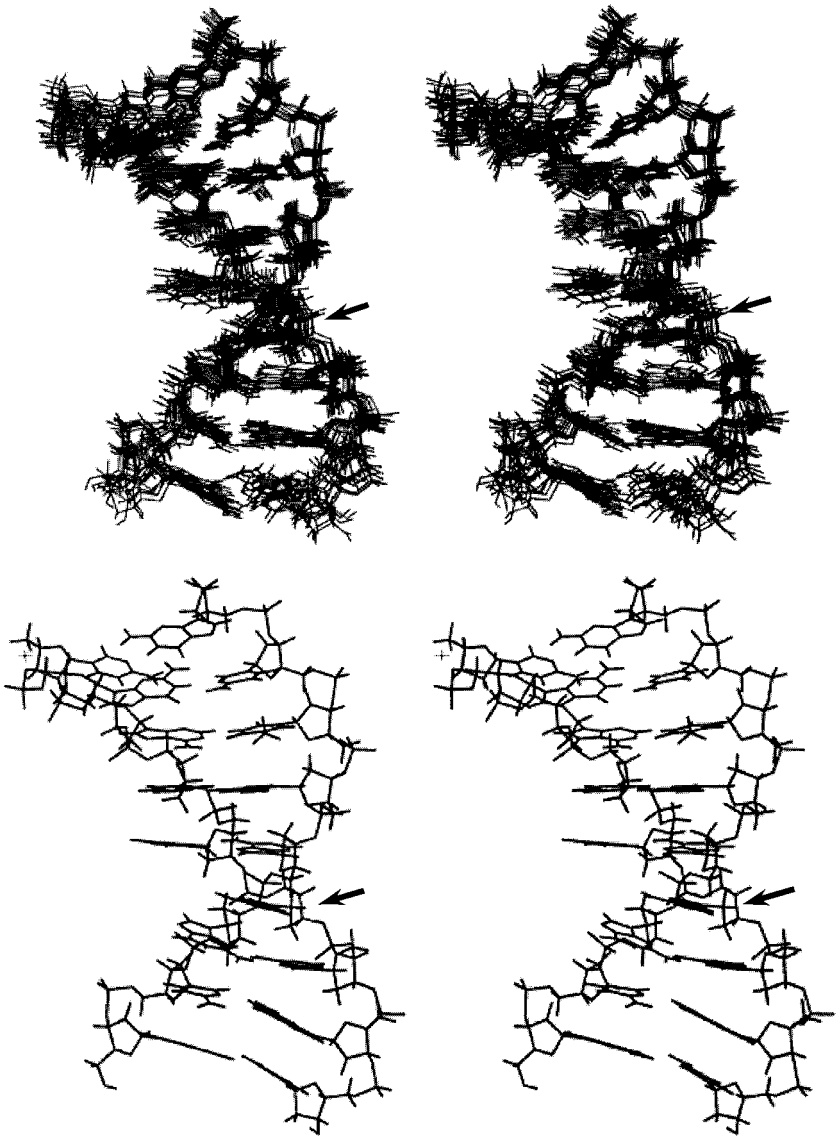

Figure 6. Wide-eye stereo views of the 15 superimposed structures of DNA oligomers containing the asymmetric GA/AAG internal loop before (top) and after (bottom) molecular dynamics calculation (the top CTTG loop were excluded for comparison). Only one of the final structures was shown after MD calculation. Before MD calculation, r.m.s.d. value of 1.35 ± 0.26 Å was observed, indicating good superimposition. Interestingly, the bulged A18 residue (indicated by an arrow) doesn't create large kink in the helical axis; only a small kink of less than five degree is observed in the present GA/AAG internal loop, possibly due to the different stacking pattern with the bracketed sheared G•A base pairs.

the chemical shift of A4H3′ is coincident with that of water, and weak power was applied to saturate the water signal. The G12H4′-P12 cross peak is not observable because the γ torsional angle of G12 in the mini-loop has been altered from the *gauche*$^+$ domain to the *trans* domain to accommodate the sharp turn necessary to connect the two strands. The five P12-G12O5′-G12C5′-G12C4′-G12H4′ atoms are therefore not situated in a plane to form a 'W' shape to exhibit detectable $^4J_{\mathrm{P-H4'}}$ coupling (Altona, 1982; Sarma et al., 1973). The C9H4′-P9 cross peak is also not observed, possible due to the same reason. The β and γ torsional angles for these two residues are therefore left unconstrained.

Using the information from through-space NOE connectivity and through-bond *J*-coupling connectiv-

Table 1. Structural Statistics for the d(GC**GA**GTCG*CTTG*CGAC**GAA**GC) hairpin

Restraints	Numbers
Exchangeable NOEs	
H-bonds (1.8–2.1 Å)	24
2.0–5.0 Å	2
3.0–6.0 Å	13
Non-exchangeable NOEs	
2.0–4.0 Å	31
3.0–5.0 Å	91
4.0–6.0 Å	73
> 5 Å	8
Total NOEs	242
Torsional Angles	112
Backbone (β, γ, ε)	91
Glycosidic	21
NOEs per residue	12
NOEs and torsion angles per residue	17
Violations of experimental restraints	
Distance restraints (> 0.15)	0
Torsional angles restraints (> 3°)	0
r.m.s.d.	1.35 ± 0.26 Å

ity, all exchangeable protons, non-exchangeable protons (H5′/H5′ protons excluded), and phosphorous atoms of the 5′-d(GC***GA***GTCG*CTTG*CGAC***GAA***GC)-3′ hairpin were unambiguously assigned with their chemical shifts listed in the supplementary Table 1.

Structural studies

Figure 5 shows some of the idiosyncratic NOEs present in this unusual motif. The statistics of the constraints used to determine its solution structure was listed in Table 1. Due to the abundant NOEs in this 2×3 internal loop region, a well-converged family of the final structures was obtained. Figure 6 shows the overlapping of fifteen final structures from thirty initial structures before molecular dynamics (top) and the one selected final structure after molecular dynamics calculation (bottom). It is interesting to note that insertion of an extrahelical adenine (indicated by an arrow) into the tandem sheared G•A base pairs causes only a minor kink in the helical axis in the present case. This is different from the common belief that a significant kink of approximately 20 deg in the helical axis usually results when an extrahelical purine is inserted into the canonical G•C or A•T base pairs of a duplex (Hare et al., 1986; Rosen et al., 1992a). Such unusual bulge-without-kink behavior of the 5′-(GA)/(AAG)-5′ internal loop is possibly due to the novel three-purine cross-strand stack as shown in Figures 7c and 7d. Without the bulged A18 residue, the sheared G17•A4 and G3•A19 base pairs form excellent cross-strand G3/G17 and A4/A19 stacks similar to those shown in Figure 7a (Chou et al., 1994a). Incorporation of an unpaired adenine into the tandem sheared G•A pairs has, however, considerably changed the stacking pattern, as clearly revealed from the down-the-helix view of the 5′-(GA)/(AAG)-5′ motif in Figure 7c. Residues G17 and A19 have been completely dislocated and un-stacked from the core of internal loop stacking. Instead, bulged residue A18 intercalates well into the internal loop and forms good cross-strand stacking with residues G3 and A4 of the opposite strand. Such intercalation results in a novel mixed three-purine G3/A18/A4 stack that is very different from the double-guanine or double-adenine stack in the tandem sheared $(GA)_2$ motif (Figure 7a) (Chou et al., 1994a) or the quadruple-guanine stack in the $(GGA)_2$ motif (Figure 7b) (Chou et al., 1994b). However, the displaced G17 and A19 residues do not enter freely into the solvent, but stack well with the neighboring C16 and G20 residues respectively as shown in Figure 7d. The overall stacking in the 5′-(CGAG)/(GAAGC)-5′ motif therefore undergoes considerable change, with the intra-strand G17-A18 and A18-A19 stacking completely disrupted, but replaced by a novel three-purine cross-strand A4-A18-G3 stacking that is accompanied by two excellent intra-strand G17-C16 and A19-G20 stacking.

All measured torsional angles were listed in Table 2 of the supplementary material.

Discussions

Parvoviruses are special in that they replicate through rolling hairpin mechanism, in which the replication starts from the 3′-hydroxyl group of the 3′-termini and reaches to the 5′-end (Astell et al., 1985). Importantly, the structures of the 3′-termini of the four different parvovirus genomes were shown to exist in a Y-shaped hairpin structure, in which a highly conserved 2×3 internal loop of the 5′-(GA)/(AAG)-5′ sequence close to the origins of DNA replication was detected. Although DNA structures with one (Bhattacharyya and Lilley, 1989; Hare et al., 1986; Joshua-Tor et al., 1992), two (Stassinopoulos et al., 1996), three (Aboul-ela et al., 1993; Rosen et al., 1992a,b), or five bulged

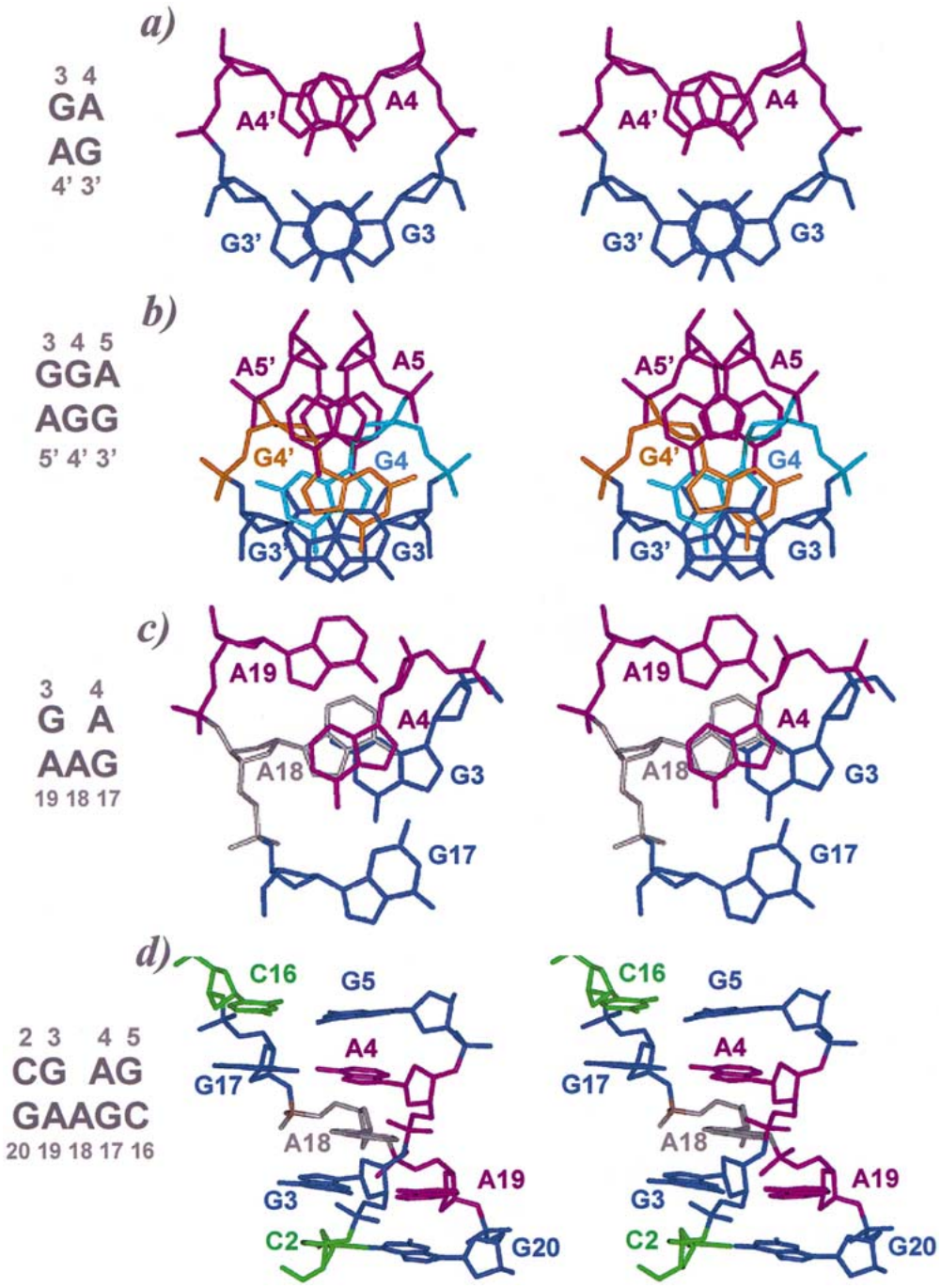

Figure 7. The stereo pictures from the view down-to-the-helical axis of the 5′-(GA)/(AG)-5′ (a), 5′-(GGA)/(AGG)-5′ (b), and 5′-(GA)/(AAG)-5′ (c) motifs, and from the view perpendicular to the helical axis of the 5′-(CGAG)/(GAAGC)-5′ (d) motif. Excellent cross-strand G/G and A/A stacks are obvious for the 5′-(GA)/(AG)-5′ motif (Figure 7a), while excellent cross-strand G/G/G/G base-base stacking and G4(deoxyribose)/A5′(base) and G4′(deoxyribose)/A5(base) stacking are observed for the 5′-(GGA)/(AGG)-5′ motifs (b; residues G4 and G4′ were colored blue and brown respectively). Intercalation of residue A18 (colored gray) into the tandem sheared A4•G17 and G3•A19 base pairs has caused considerable change in the stacking pattern. The A4-A19 and G3-G17 bases are now completely un-stacked. Instead, a novel cross-strand three-purine A4/A18/G3 stacking is observed, which is also revealed from the view perpendicular to the helical axis in (d). The displaced G17 base does not enter into the solvent freely, but stacks very well with the intra-strand C16 base. Similarly, the displaced A19 also stacks well with the intra-strand G20 base to create a novel three-purine stack for this unusual 2 × 3 internal loop motif.

residues (Dornberger et al., 1999) have been well studied before, all such structures are bracketed by conventional canonical G•C and/or A•T base pairs. No paper dealing with bulged DNA structure bracketed by non-canonical base pairs has been published as far as we know until to date. Since such structure may play important roles in the life cycle of these viruses, we are therefore endeavored to determine its three-dimensional structure using high-resolution NMR technique.

Interestingly, the highly conserved motif of the nature 5′-(GA)/(A**A**G)-5′ sequence happens to be more stable than the 5′-(GA)/(A**G**G)-5′ sequence by approximately 4 °C. This is different to what observed in the double-intercalation $(GPuA)_2$ motif (Chou et al., 1994b) or the quadruple-intercalation $(GPuPuA)_2$ motif (Chou and Chin, 2001), in which the $(GGA)_2$ and $(GGGA)_2$ motifs are more stable than the $(GAA)_2$ and $(GAAA)_2$ motifs, due to the formation of cross-strand H-bonds of the central unpaired guanines with the opposite backbone phosphodiesters. The present melting data do hint that the unpaired guanine in the 5′-(GA)/(A**G**G)-5′ sequence is not situated in a position to form cross-strand H-bond formation to stabilize this internal loop motif. This can be revealed from Figure 7c, in which when the unpaired A18 residue is substituted into G18, its 2′-NH_2 is too far away from the cross-strand A4 phosphodiester oxygen atom to form a H-bond. This is unlike the unpaired GNH_2 in the $(GGA)_2$ motif, which does form good H-bond with the cross-strand phosphodiester oxygen atom, namely, the G4′NH_2 with the A5 phosphodiester and the G4NH_2 with the A5′ phosphodiester (Figure 7b). Although the unpaired guanosine is possibly not involved in H-bonding in the (GA)/(AGG) internal loop, the reason why it is less stable than (GA)/(AAG) is still not clear (Figure 1), but is likely due to the guanine-guanine base repulsion between the G18 and G3 residues.

While tandem $(GA)_2$ sequence (which can be considered as a 2 × 2 internal loop) in the $(PyGAPu)_2$ (Chou et al., 1994a) or (PuGAPu)/(PyGAPy) motif (Chou et al., 1999a) is characterized by excellent cross-strand G/G and A/A stacking, the double-intercalation G2 motif of the $(GGA)_2$ sequence (which can be considered as a 3 × 3 internal loop) is characterized instead by the intercalation of two unpaired guanines to exhibit excellent cross-strand G/G stacking. Furthermore, each unpaired guanine base also displays partial intra-strand G/G stacking with the guanine base of the bracketed sheared G•A pairs, and each deoxyriboses of the unpaired guanosines also stacks with the adenines of the bracketed sheared G•A pairs, i.e., deoxyribose(G4′) over A5 and deoxyribose(G4) over A5′ as shown in Figure 7b (Chou et al., 1994b; Zhu et al., 1995). So when examined from the down-the-helix view of the $(GGA)_2$ motif, the G3′ residue is situated on top of G3 residue and A5 on top of A5′ residue, as that happened in the $(GA)_2$ motif, even though two unpaired guanosine residues are intercalated between the two sheared G•A base pairs. However, the stacking of asymmetric 2 × 3 $(G_3A_4)/(A_{19}A_{18}\ G_{17})$ internal loop is considerably different to those of the $(GA)_2$ and $(GGA)_2$ motifs. With the incorporation of an unpaired A18 residue, A4 base is now completely un-stacked from the A19 base and G17 base completely un-stacked from the G3 base (Figure 7c). Instead, A18 base now stacks well with the cross-strand A4 base and partially with the cross-strand G3 base to form a mixed three-purine A4-A18-G3 stack. Such novel stacked structure can explain its resistance to single-stranded specific endonuclease susceptibility.

Interestingly, the backbone conformation for the asymmetric 2 × 3 $(G_3A_4)/(A_{19}A_{18}\ G_{17})$ internal loop is more similar to that of $(GA)_2$ than to $(GGA)_2$ motif, i.e., two very upfield ^{31}P signals were also observed in the asymmetric internal loop motif. However, it is the first G17-A18 phosphodiester backbone, not the second one of A18-A19, that is transformed into $\zeta(t)$ domain to extend the backbone for dislocating base A18 for cross-strand stacking. The ζ torsion angle of the A18-A19 backbone has in fact resumed the normal *gauche*$^-$ values. Furthermore, no unusual C3′-endo sugar conformation is found in this 2 × 3 asymmetric internal loop; for all measurable residues, the sugars existed predominantly in the C2′-endo conformation, as readily determined from the sum of the $^3J_{1'-2'}$ and $^3J_{1'-2''}$ coupling constants (Altona and Sundaralingam, 1972), which were in all cases greater than 15 Hz (data not shown), although the sugar pucker conformation of the central guanosine residues in the $(GGA)_2$ motif are transformed from C2′-endo to C3′-endo to extend the backbone for base intercalation. The unpaired A18 residue still adopts the regular C2′-endo conformation, possibly due to the different stacking profile.

Another interesting point deserves noting is that in the present case, the bulged pyrimidine also adopts similar stacking-in conformation as bulged purine when bracketed by sheared G•A pairs. This is different from the common belief that extra-helical pyrim-

idine has the tendency to be looped out while extrahelical purine to be stacked inside a B-DNA duplex (Hare et al., 1986; Rosen et al., 1992a). This is most likely due to the strong conformational constraint imposed by the tandem sheared G•A base pair that does not allow an unpaired residue to be looped out. This unexpected finding is confirmed through the detection of several crucial NOEs in the 2D-NOESY of the 5′-(GA)/(A**T**G)-5′ motif, i.e., strong NOEs between the T18CH_3-G17H1′ protons and weak NOEs between the T18CH_3-G17H8 and T18CH_3-A4H1′ protons (data not shown). The lower melting temperature of the 5′-(GA)/(A**T**G)-5′ and 5′-(GA)/(A**C**G)-5′ motifs is thus due to the weaker stacking force between the unpaired pyrimidine base and the bracketed purine bases. Nevertheless, the unpaired thymine in the 5′-(GA)/(A**T**G)-5′ motif does exhibit some local dynamics, as demonstrated by the absence of several NOE cross peaks expected in this region (data not shown). Due to this reason, the 5′-(GA)/(A**T**G)-5′ internal loop is not subject to further structural studies.

2 × 3 asymmetric internal loops also exist abundantly in RNA of important biological functions (Baumstark and Reisner, 1995; Fodor et al., 1995; Gutell, 1994; Jang and Wimmer, 1990; Mathews et al., 1997; Powers and Noller, 1995). In this respect, it is interesting to note that a thermodynamics and NMR study of a similar RNA internal loop motif of the 5′-(GA)/(AAG)-5′ sequence has recently been reported (Schroeder and Turner, 2000). This oligomer also exhibits a decent melting temperature of approximately 52 °C with well-resolved NMR imino proton signals, although those characteristic guanosine imino protons involved in the sheared G•A pairs are rather broad. It remains to be seen if such RNA internal loops adopt a similar structure as that of DNA, since DNA and RNA of the same sequence may adopt quite different structures under identical condition (Wu and Turner, 1996, Cheng et al., 1992).

Acknowledgements

We thank the National Science Council and the Chung-Zhen Agricultural Foundation Society of Taiwan, ROC for the instrumentation grants and Dr Larvery for the kind gift of CURVE program. SHC is a recipient of the Outstanding Research Award from the National Science Council. This work was supported by the NSC grants 89-2113-M-005-034 to S.H.C.

References

Aboul-ela, F., Murchie, A.I.H., Homans, S.W. and Lilley, D.M. (1993) *J. Mol. Biol.*, **229**, 173–188.
Altona, C. (1982) *Recl. Trav. Chim. Pays-Bas.*, **101**, 413–433.
Altona, C. and Sundaralingam, M. (1972) *J. Am. Chem. Soc.* **94**, 8205–8212.
Astell, C.R., Chow, M. B. and Ward, D.C. (1985) *J. Virol.*, **54**, 171–177.
Astell, C.R., Smith, M., Chow, M.B. and Ward, D.C. (1979) *Cell*, **17**, 691–703.
Baumstark, T. and Reisner, D. (1995) *Nucl. Acids Res.*, **23**, 4246–4254.
Bhattacharyya, A. and Lilley, D.M. (1989) *Nucl. Acids Res.*, **17**, 6821–6840.
Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Kundrot, C.E., Cech, T.R. and Doudna, J.A. (1996a) *Science*, **273**, 1678–1685.
Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Szewczak, A.A., Kundrot, C.E., Cech, T.R. and Doudna, J.A. (1996b) *Science*, **273**, 1696–1699.
Cheng, J.-W., Chou, S.-H. and Reid, B.R. (1992) *J. Mol. Biol.*, **228**, 1037–1041.
Chou, S.-H. and Chin, K.-H. (2001) *J. Mol. Biol.*, in press.
Chou, S.-H. and Tseng, Y.-Y. (1999) *J. Mol. Biol.*, **285**, 41–48.
Chou, S.-H., Chin, K.-H. and Chen, C.W. (2001) *J. Biomol. NMR*, **19**, 33–48.
Chou, S.-H., Cheng, J.-W., Fedoroff, O.Y., Chuprina, V.P. and Reid, B.R. (1992) *J. Am. Chem. Soc.*, **114**, 3114–3115.
Chou, S.-H., Cheng, J.-W., Fedoroff, O. and Reid, B.R. (1994a) *J. Mol. Biol.*, **241**, 467–479.
Chou, S.-H., Zhu, L. and Reid, B.R. (1994b) *J. Mol. Biol.*, **244**, 259–268.
Chou, S.-H., Zhu, L., Gao, Z., Cheng, J.-W. and Reid, B.R. (1996) *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., 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 Chu, B.-Y. (2000) *J. Biomol. NMR*, **17**, 1–16.
Correll, C.C., Freeborn, B., Moore, P.B. and Steitz, T.A. (1997) *Cell*, **91**, 705–712.
Cotmore, S.F. and Tattersall, P. (1994) *EMBO J.*, **13**, 4145–4152.
Cotmore, S.F., Nuesch, J.P.F. and Tattersall, P. (1993) *J. Virol.*, **67**, 1579–1589.
Dornberger, U., Hillisch, A., Gollmick, F.A., Fritzsche, H. and Diekmann, S. (1999) *Biochemistry*, **38**, 12860–12868.
Fodor, E., Pritlove, D.C. and Brownlee, G.G. (1995) *J. Virol.*, **69**, 4012–4019.
Gao, Y.-G., Robinson, H., Sanishvili, R., Joachimiak, A. and Wang, A.H.-J. (2000) *Biochemistry*, **38**, 16452–16460.
Gutell, R.R. (1994) *Nucl. Acids Res.*, **22**, 3502–3507.
Hare, D., Shapiro, L. and Patel, D.J. (1986) *Biochemistry*, **25**, 7456–7464.
Hare, D.R., Wemmer, D. E., Chou, S.-H., Drobny, G. and Reid, B.R. (1983) *J. Mol. Biol.*, **171**, 319–336.
Ippel, J.H., Lanzotti, V., Galeone, A., Mayol, L., van den Boogaart, J.E., Pikkemaat, J.A. and Altona, C. (1995) *J. Biomol. NMR*, **6**, 403–422.
Jang, S.K. and Wimmer, E. (1990) *Genes Dev.*, **4**, 1560–1572.

Joshua-Tor, L., Frolow, F., Appella, E., Hope, H., Rabinovich, D. and Sussman, J.L. (1992) *J. Mol. Biol.*, **225**, 397–431.
Liu, Q., Yong, C.B. and Astell, C.R. (1994) *Virology*, **201**, 251–262.
Mathews, D.H., Banerjee, A. R., Luan, D.D., Eickbush, T.H. and Turner, D.H. (1997) *RNA*, **3**, 1–16.
Moore, P.B. (1999) *Annu. Rev. Biochem.*, **67**, 287–300.
Plateau, P. and Gueron, M. (1982) *J. Amer. Chem. Soc.*, **104**, 7310–7311.
Pley, H.W., Flaherty, K.M. and McKay, D.B. (1994) *Nature*, **372**, 68–74.
Powers, T. and Noller, H.F. (1995) *RNA*, **1**, 194–209.
Rosen, M.A., Live, D. and Patel, D.J. (1992a) *Biochemistry*, **31**, 4004–4014.
Rosen, M.A., Shapiro, L. and Patel, D.J. (1992b) *Biochemistry*, **31**, 4015–4026.
Sarma, R.H., Mynott, R.J., Wood, D.J. and Hruska, F.E. (1973) *J. Am. Chem. Soc.*, **95**, 6457–6459.
Schroeder, S.J. and Turner, D.H. (2000) *Biochemistry*, **39**, 9257–9274.
Shepard, W., Cruse, W.B.T., Fourme, R., Fortelle, E.D.L. and Prange, T. (1998) *Structure*, **6**, 849–861.
Sklenar, V., Miyashiro, H., Zon, G., Miles, H.T. and Bax, A. (1986) *FEBS Lett.*, **208**, 94–98.
Spackova, N., Berger, I. and Sponer, J. (2000) *J. Am. Chem. Soc.*, **122**, 7564–7572.
Stassinopoulos, A., Ji, J., Gao, X. and Goldberg, I.H. (1996) *Science*, **272**, 1943–1946.
Tseng, Y.-Y. and Chou, S.-H. (1999) *J. Chin. Chem. Soc.*, **46**, 699–706.
Willwand, K., Mumtsidu, E., Kuntz-Simon, G. and Rommelaere, J. (1998) *J. Biol. Chem.*, **273**, 1165–1174.
Wu, M. and Turner, D.H. (1996) *Biochemistry*, **35**, 9677–9689.
Zhu, L., Chou, S.-H. and Reid, R.B. (1995) *J. Mol. Biol.*, **254**, 623–637.
Zimmermann, G.R., Jenison, R.D., Wick, C.L., Simorre, J.-P. and Pardi, A. (1997) *Nat. Struct. Biol.*, **4**, 644–649.