Journal of Biomolecular NMR, **14:** 157–167, 1999.
KLUWER/ESCOM

Structural studies of symmetric DNA undecamers containing non-symmetrical sheared (PuGAPu):(PyGAPy) motifs

Shan-Ho Chou[a,*], Yu-Yu Tseng[b], Yang-Ren Chen[a] & Jya-Wei Cheng[c]
[a]*Institute of Biochemistry and* [b]*Chemistry Department, National Chung-Hsing University, Taichung 402, Taiwan, Republic of China;* [c]*Division of Structural Biology and Biomedical Science, Department of Life Science, National Tsing Hua University, Hsinchu 300, Taiwan, Republic of China*

Received 25 January 1999; Accepted 15 April 1999

Key words: NMR structures, (PuGAPu)/(PyGAPy), (PyGAPu)2, sheared G•A pair, unusual DNA

Abstract

Interstrand purine-purine stacks originate from tandem sheared purine•purine pairing and represent one of the most important motifs in both DNA and RNA structures. Several RNA and DNA structures, solved recently in both solution and the solid state, contain these special motifs, which greatly increase the structural diversity of nucleic acid molecules. The direct evidence for the sheared purine-purine pairing at neutral pH in solution remains, however, elusive. In this manuscript, we have used high resolution NMR methods to study a series of symmetrical DNA duplexes containing two non-symmetrical 5′-(PuGAPu)/(PyGAPy)-3′ motifs. Many strong- and medium-strength NOEs across the G•A base pair were detected in the H_2O-NOESY spectra collected at a relatively low temperature (−5 °C). These NOEs, especially those from A-$^6NH_2$ to G-H1′, G-H4′, and G-$^2NH_2$, clearly define the proposed side-by-side sheared G•A pairing nature. Another interesting feature is the strong NOEs exhibited by the unpaired G-imino proton in the G•A pair to its own G-$^2NH_2$, which implies that G-$^2NH_2$ is involved in H-bonding with a base in the minor groove edge. The finding that non-symmetrical (PuGAPu):(PyGAPy) motif also form similarly stable structures loosens the requirement for a more restricted $(PyGAPu)_2$ motif in forming the interstrand purine-purine stacks.

Abbreviations: NOESY, NOE spectroscopy; 2D NMR, two-dimensional nuclear magnetic resonance; DG, distance geometry; EM, energy minimization; NOE, nuclear Overhauser enhancement; DQF-COSY, double-quantum filtered correlation spectroscopy; rmsd, root-mean-square deviation; ppm, parts per million; NH_{2b}, H-bonded amino proton; NH_{2n}, non-H-bonded amino proton.

Introduction

Complementary Watson–Crick G•C and A•T pairs are the foundation of double helical nucleic acid structures, and current antisense oligonucleotides designed for drug therapy also rely on this complementarity for the specific interaction between antisense and target molecules (Crook and Bennett, 1996). Several natural antisense RNA/target RNA duplexes studied recently have been found to contain non-canonical base pairs, the presence of which was suggested to add to the selectivity of antisense therapeutics and may even benefit oligonucleotide-based drug design (for a review, see Delihas et al., 1997).

Recent studies of several nucleic acid structures by both NMR (Chou et al., 1997; Dallas and Moore, 1997; Lin et al., 1998) and X-ray diffraction methods (Pley et al., 1994; Cate et al., 1996; Correll et al., 1997; Shepard et al., 1998) have identified novel structural features of *inter*-strand purine-purine stacks. These features originate from sheared purine•purine pairing and served as stable alternatives to the *intra*-strand base-base stacks commonly observed in a regular Watson–Crick G•C and A•T paired duplex. The

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

formation of these tandem G•A base pairs was also found to be quite common in an experiment of nucleic acid complementarity via the combinatorial library method when the studied tandem G•A base pair is preceded by a pyrimidine at the 5′-end (James and Ellington, 1997). These experiments indicated that interstrand purine-purine stacking motifs are no less stable than the now popular Watson–Crick A•T and G•C base-paired motifs. It is thus no wonder that nature has reserved this stable motif in forming several nucleic acid structures of biological importance, like those in the conserved core region of the hammerhead ribozyme (Pley et al., 1994), in the J4-J5 junction of the group 1 ribozyme (Cate et al., 1996), in the loop E–loop D region of *E. coli* 5S rRNA (Correll et al., 1997; Dallas and Moore, 1997), in the telomeres of the linear *Streptomyces* chromosome (Huang et al., 1998), in the replication origin of the circular ΦX174 virus (Shlomai and Kornberg, 1980), in the *Drosophila* centromeric dodeca-satellite repeats (Ferrer et al., 1995; Ortiz-Lombardia et al., 1998), and in the human centromeric tandem $(TGGAA)_n$ repeats (Chou et al., 1994b; Zhu et al., 1995). The wide occurrence of the stable tandem sheared G•A pairs and the resulting interstrand purine-purine stacks indicate that they are important motifs in nucleic acid secondary structures.

We have recently described the stable formation of interstrand purine-purine stacks in $(PyGAPu)_2$ motifs in DNA duplexes and found that dramatic structural changes can be induced by the flanking base-pair. Thus, when the preceding 5′-bases in the $d(PyGAPu)_2$ motifs were switched from a pyrimidine to a purine and the following 3′-bases from a purine to a pyrimidine, the resulting $d(PuGAPy)_2$ sequences become less stable. In the $d(GGAC)_2$-containing sequence, no discrete NMR resonance can be detected, while in the $d(AGAT)_2$-containing sequence, it adopts instead a head-to-head $G_{anti}•A_{anti}$ geometry, and only poor interstrand purine-purine stacking can be found. Interestingly, similar sequences in RNA behave quite different (Walter et al., 1994). For example, the adjacent G•A sequence in the RNA $(AGAU)_2$ is found to possibly adopt a sheared configuration, in contrast to the head-to-head $G_{anti}•A_{anti}$ geometry mentioned above in DNA. Besides, the RNA $(GGAC)_2$ sequence was found to be the most stable among all studied tandem G•A pairs and adopts the head-to-head $G_{anti}•A_{anti}$ geometry (Wu and Turner, 1996), although no stable species can be even detected by NMR methods in a similar DNA sequence. In addition, the RNA

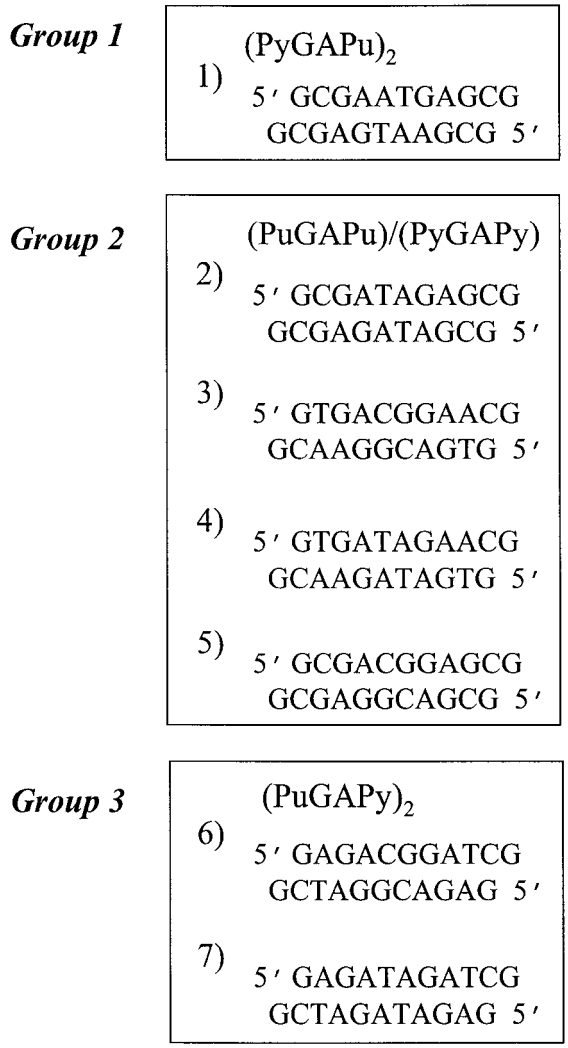

Figure 1. The DNA undecamers studied in this manuscript. All oligomers have symmetrical sequences, with group 2 sequences containing non-symmetrical 5′-(PuGAPu)/(PyGAPy)-3′ motifs.

$(GAGC)_2$ motif, which has never been observed to be stable in DNA, was found to be even more stable than the RNA $(CGAG)_2$ motif. These dramatic differences between the similar DNA and RNA sequences may be due to B-form versus A-form stems and the extra 2′-OH functional group in RNA.

It would therefore be of interest to know if non-symmetrical motifs in the (PuGAPu):(PyGAPy) sequences still allow the adjacent GA sequence to adopt a stable structure. We have thus synthesized a series of DNA undecamers (Figure 1) containing all kinds of combinations with two purines bracketing the G-A sequence in one strand and two pyrimidines bracket-

ing the G-A sequence in the other strand and used high resolution NMR methods to study their behavior. Our results indicate that they also form stable sheared G • A pairing. The previous elusive direct evidence for the sheared G•A pairing has been clearly identified in the H_2O-NOESY spectra. Many newly detected inter-strand NOE constraints originating from exchangeable G-imino and A-, G-amino protons of the sheared G•A pairs were also incorporated into distance bounds to derive the solution structure containing the 5′-(TGAC)/(GGAA)-3′ motif (sequence 3 in Figure 1).

Materials and methods

Sample preparation

All DNA samples were synthesized on a 6 μmol scale on an Applied Biosystems 380B DNA synthesizer (Chou et al., 1989) with the final 5′-DMT groups attached. After deprotection in concentrated NH_4OH at 55 °C overnight, the crude material was purified on a semi-preparative C-18 column and detritylated on column by a three-pump Varian DYNAMAX HPLC system. This oligonucleotide was further purified by anion-exchanger and gel filtration columns. Approximately 8 mg of purified lyophilized material was dissolved in 0.5 ml of 40 mM sodium phosphate buffer (pH 6.8) containing 200 mM NaCl and 0.5 mM EDTA. The samples were heated at 90 °C for 5 min and slowly annealed to room temperature. For non-exchangeable proton studies, the samples were lyophilized in 99.96% D_2O three times and finally dissolved in 0.5 ml of 99.996% D_2O.

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 16000 Hz with the carrier frequency set at the water resonance. The maximum excitation was set at 12.5 ppm. For each experiment, 16K complex points were collected and 64 scans were averaged with a 2 s relaxation delay. 1D-difference NOE spectra were collected as described previously (Chou et al., 1983).

2D NOESY in 90% H_2O/10% D_2O was performed at −5 °C in a pH 6.8 high salt (200 mM) buffer with the following parameters: delay time 1 s, mixing time 0.12 s, spectral width 11477 Hz, number of complex points 2048, number of transients 96, and number of increments 300, linear-predicted to 512.

NOESY experiments in D_2O were carried out at 20 °C in the hypercomplex mode with a spectral width of 4782 Hz. Spectra were collected using five mixing times of 60, 120, 240, and 360 ms with a relaxation delay of 2.0 s between transients and with 2048 complex points in the t_2 and 512 complex points in the t_1 dimension. For each t_1 incrementation, 32 scans were averaged.

A DQF-COSY spectrum was collected in the hypercomplex mode with a spectral width of 4782 Hz in both dimensions; 2048 complex points in the t_2 dimension and 400 complex points in the t_1 dimension were collected with a relaxation delay of 1 s, and 16 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 hypercomplex mode with a spectral width of 4782 Hz in the ^{1}H dimension and a spectral width of 2000 Hz in the ^{31}P dimension. A total of 1600 complex points in the t_2 (^{1}H) dimension and 128 complex points 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 acquired data were transferred to an SGI O2 workstation and processed by the software FELIX (MSI Inc.) as described previously (Chou et al., 1992).

Structural calculations

The solution structure of the d(5′-GTGACGGAACG-3′)$_2$ oligomer was determined by the combined use of distance geometry (DG) and molecular dynamics methods (DISCOVER) as described previously (Chou et al., 1992, 1994a). The 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 1.8–3.0 Å, 2.0–4.0 Å, or 3.0–6.0 Å, respectively. If a NOE cross peak was not detectable even at 360 ms mixing time, then a lower distance bound of 5.0 Å was applied for proton pairs between adjacent nucleotides. Canonical hydrogen-bond distances with bounds of 1.8–2.1 Å were assigned to Watson–Crick base pairs when their imino proton resonances were detected at the usual range between 12.5 and 14.5 ppm. A large number of distance constraints involving exchangeable protons (Figure 7) were also derived from the H_2O/NOESY spectra and only given two rather wide distance bounds of either 2.0–5.0 Å or 3.0–6.0 Å, due to the exchange phenom-

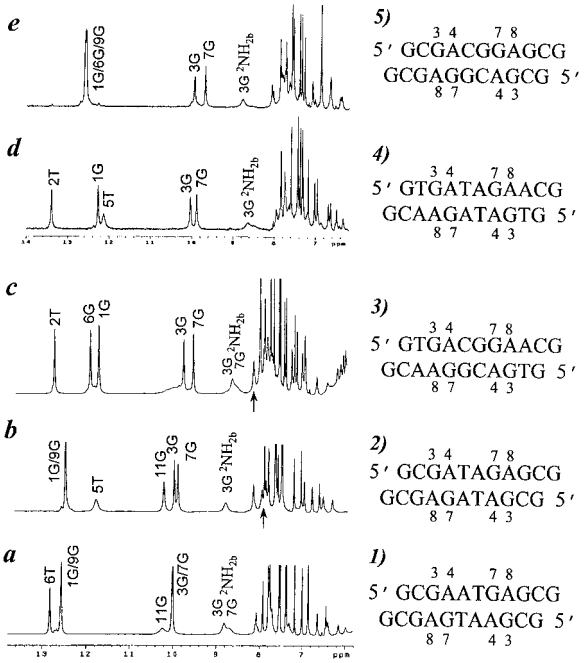

Figure 2. The one-dimensional 600 MHz imino, amino and aromatic proton NMR spectra of the group 1 and 2 sequences at 0 °C. Samples (ca. 8 mg each) were dissolved in 0.5 ml of 40 mM phosphate buffer (pH 6.8) containing 200 mM NaCl, 0.2 mM EDTA. The spectra were assigned either by 1D-NOE or 2D H_2O/NOESY methods. The two peaks at ca. 10 ppm are the resonance from the unpaired G-imino protons in the 5′-(PuGAPu)/(PyGAPy)-3′ motifs. Note the upfield shifting of the 5T imino proton resonance in both sequences 2 and 4, which are also broader in lineshape. $8A^6$-NH_{2b} protons are labeled with arrows.

enon. The dihedral angle constraints were primarily determined semi-quantitatively from the ^{31}P-^{1}H heteronuclear correlation data (Chou et al., 1996). The β and γ dihedral angles were constrained 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 dihedral angles were constrained to the *trans* (180° ± 30°) and *gauche*$^+$ (60° ± 30°) domains, respectively. Otherwise they were left unconstrained. The ϵ dihedral angle only adopts either the *trans* or *gauche*$^-$ conformation (Altona, 1982). The *gauche*$^+$ conformation is not sterically allowed. Based on the absence of the long-range $^4J_{H2'-P}$ coupling, all ϵ torsion angles could be constrained to the *trans* domain (180° ± 30°) (Mooren et al., 1994) except for the tandem G-A nucleotides which were left unconstrained. 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 aromatic-anomeric cross peaks of comparable intensity to the C-H5/C-H6 cross peaks were detected (Dallas and Moore, 1997). These NOE distance (438 in total) and dihedral angle (78 in total) constraints were used to generate initial structures using the DGII program (MSI, Inc.). The abundant NOE constraints, especially those around the tandem GA sequence, although quite wide, have allowed us to efficiently generate DG structures with correct folding at a quite good success rate (∼70%). The initial structure was further refined by restrained molecular dynamics using the program DISCOVER (MSI, Inc.). Well-converged final structures with pairwise rmsd values of ca. 0.5 Å were obtained after molecular dynamics calculations.

Results

Sequence studied

We originally studied the symmetrical undecamer duplexes $d(ATGAGCGAATA)_2$ (Chou et al., 1992) and $d(GCGAATGAGCG)_2$ (Chou et al., 1994a) (sequence 1 in Figure 1), that contain two symmetrical $(PyGAPu)_2$ motifs with preceding 5′-pyrimidines and following 3′-purines. In this paper, we have studied symmetrical sequences containing two non-symmetrical (PuGAPu):(PyGAPy) motifs with purines bracketing the GA sequence in one strand and pyrimidines bracketing the GA sequence in the other strand (group 2 sequences in Figure 1) to see if they can still form stable motifs. Furthermore, two sequences containing two symmetrical $(PuGAPy)_2$ motifs with preceding 5′-purines and following 3′-pyrimidines were also prepared and studied (group 3 in Figure 1). Unlike those for RNA (Walter et al., 1994), no discrete NMR signals for DNA containing two $(AGAT)_2$ sequences (sequence 7 in Figure 1) can be observed, although a decamer or an octamer containing only one such motif did exhibit an imino proton resonance at ca. 12.4 ppm that is characteristic of the G-imino proton resonance involved in the G_{anti}•A_{anti} pairing (Cheng et al., 1992). This indicates that the head-to-head G_{anti}•A_{anti} motif de-stabilizes the parent DNA duplex and only one such mismatch is allowed at neutral pH for a short oligomer less than 12 nucleotides. Similarly, we could not detect any discrete NMR signal for the 5′-(AGAC)/(GGAT)-3′-containing sequence (sequence 6 in Figure 1). On the contrary, all group 2 sequences containing two non-symmetrical (PuGAPu):(PyGAPy) motifs form stable structures and are suitable for NMR studies. We will

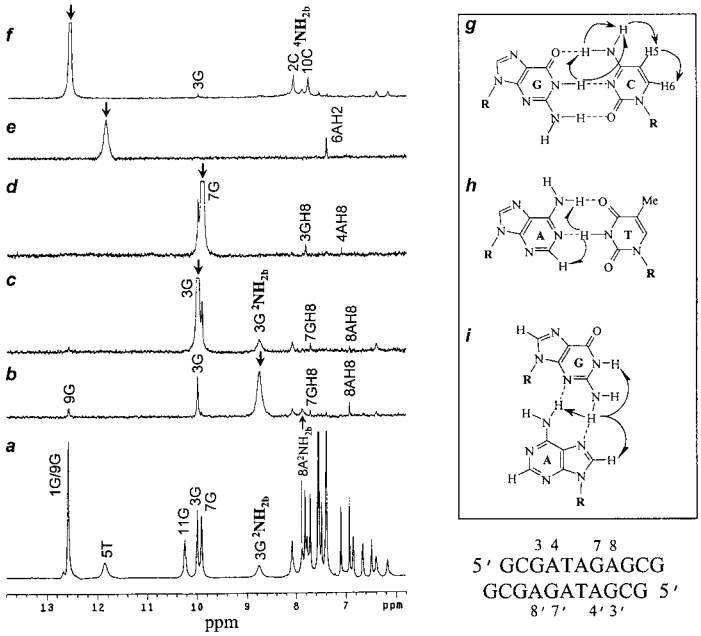

Figure 3. The 1D-NOE spectra for sequence 2 (b–f) and the expected coherence transfer pathways for a regular G•C pair (g), an A•T pair (h), and a sheared G•A pair (i). The irradiated peaks were labeled with thick arrows, and the NOE peaks of interest were labeled with corresponding text.

thus only focus on the NMR data of the group 2 sequences in the following description.

1D NOE experiments on the (PuGAPu)/(PyGAPy) motifs

The one-dimensional imino and aromatic proton spectra for sequences 1 to 5 in Figure 1 at a neutral (pH 6.8) buffer condition at 40 °C are shown in Figure 2. All imino protons and amino protons were either assigned by 1D-NOE or by 2D-H_2O/NOESY methods (see later description). Compared with sequence 1, which has been studied before (Chou et al., 1994a), all other four sequences have two separated imino proton resonances at around 10 ppm, due to the non-symmetrical nature of the tandem 5′-(PuGAPu)/(PyGAPy)-3′ motifs. These sharp resonances are characteristic of the unpaired guanosine imino protons participating in the sheared G•A pairing (Li et al., 1991a; Cheng et al., 1992; Chou et al., 1997). 3G-$^2NH_{2b}$ (b stands for H-bonded and n for non-H-bonded) and/or 7G-$^2NH_{2b}$ and 8A-$^6NH_{2b}$ for sequences 2 and 3 (assigned by 2D-H_2O/NOESY, see later) also move down to ca. 8.8 and 8.0 ppm. This is another indication that the G- and A-amino protons in the $(GA)_2$ motifs are involved in H-bonding. Another interesting point is that the 5T-imino proton resonances of the Watson–Crick 5T•6A base pair in sequences 2 and 4 exhibit rather broad lineshapes and shift upfield even beyond other G-imino protons participating in the Watson–Crick G•C base pairs (Figures 2b and 2d). This unusual phenomenon is confirmed by the 1D NOE method as shown in Figure 3. It is clear from the spectrum that the broader imino proton resonance at 11.92 ppm is indeed an H-bonded T-imino proton, as revealed by its strong NOE to 6A-H2 (Figure 3e). Figure 3b also shows the important NOEs exhibited by the 3G-$^2NH_{2b}$ to its own imino proton and to its paired 8A-$^6NH_{2b}$ and 8A-H8. Similarly, the 3G-imino proton can also exhibit NOEs to its own $^2NH_{2b}$ and paired 8A-H8 by spin diffusion via its own $^2NH_2$ (Figure 3c). In the same way, the 7G-imino proton can also exhibit NOEs to its paired 4AH8 via its own $^2NH_2$ protons (Figure 3d). The cross-strand NOEs from 3G-imino to 7′G-H8 and from 7G-imino to 3′G-H8 are also obvious from Figures 3c and 3d, which indicate that 3G stacks on the 7′G and vice versa. These assignments were further confirmed by the 2D-H_2O/NOESY experiments (data not shown).

2D H_2O/NOESY experiments of the (GGAA)/(TGAC) motifs

The expanded two-dimensional H_2O/NOESY spectrum of d(GTGACGGAACG)$_2$ is shown in Figures 4, 5, and 6, respectively. Figure 4 shows the imino-proton related NOE cross peak region. As shown in Figure 3h,

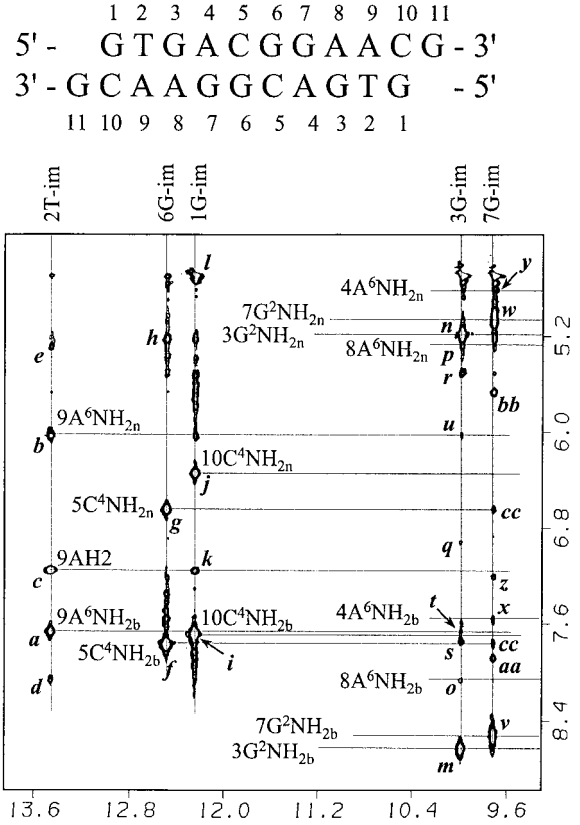

Figure 4. The expanded NOESY (mixing time 0.12 s) contour plot of sequence 3 in 90% H_2O/10% D_2O (200 mM NaCl, high salt, pH 6.8 buffer) at −5 °C. This expanded region covers NOE cross peaks from imino to amino/base protons. The cross peaks are labeled from a to cc as follows: (a) 2T-imino–9A^{6}NH$_{2b}$; (b) 2T-imino–9A^{6}NH$_{2n}$; (c) 2T-imino–9AH2; (d) 2T-imino–8A^{6}NH$_{2b}$; (e) 2T-imino–8A^{6}NH$_{2n}$; (f) 6G-imino–5C^{4}NH$_{2b}$; (g) 6G-imino–5C^{4}NH$_{2n}$; (h) 6G-imino–5CH5; (i) 1G-imino–10C^{4}NH$_{2b}$; (j) 1G-imino–10C^{4}NH$_{2n}$; (k) 1G-imino-9AH2; (l) 1G-imino–10CH5; (m) 3G-imino–3G^{2}NH$_{2b}$; (n) 3G-imino–3G^{2}NH$_{2n}$; (o) 3G-imino–8A^{6}NH$_{2b}$; (p) 3G-imino–8A^{6}NH$_{2n}$; (q) 3G-imino–8AH8; (r) 3G-imino–7GH1′; (s) 3G-imino–7GH8; (t) 3G-imino–9A^{6}NH$_{2b}$; (u) 3G-imino–9A^{6}NH$_{2n}$; (v) 7G-imino–7G^{2}NH$_{2b}$; (w) 7G-imino–7G^{2}NH$_{2n}$; (x) 7G-imino–4A^{6}NH$_{2b}$; (y) 7G-imino–4A^{6}NH$_{2n}$; (z) 7G-imino–4AH8; (aa) 7G-imino–3GH8; (bb) 7G-imino–3GH1′; (cc) 7G-imino–5C^{4}NH$_{2n}$.

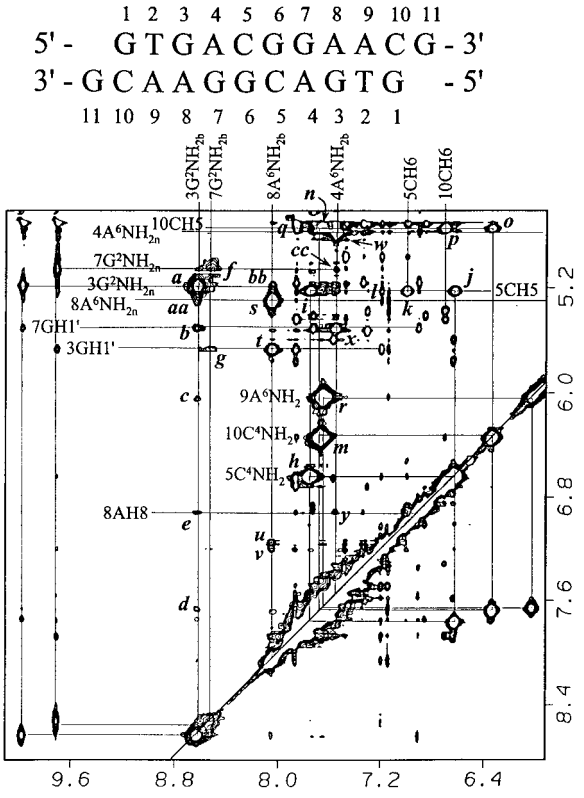

Figure 5. The expanded H_2O/NOESY contour plot for the amino proton region of sequence 3. The cross peaks are labeled from a to cc as follows: (a) 3G^{2}NH$_{2b}$–3G^{2}NH$_{2n}$; (b) 3G^{2}NH$_{2b}$–7GH1′; (c) 3G^{2}NH$_{2b}$–9A^{6}NH$_{2n}$; (d) 3G^{2}NH$_{2b}$–9A^{6}NH$_{2b}$; (e) 3G^{2}NH$_{2b}$–8AH8; (f) 7G^{2}NH$_{2b}$–7G^{2}NH$_{2n}$; (g) 7G^{2}NH$_{2b}$–3GH1′; (h) 5C^{4}NH$_{2b}$–5C^{4}NH$_{2n}$; (i) 5C^{4}NH$_{2b}$–5CH5; (j) 5C^{4}NH$_{2n}$–5CH5; (k) 5CH5–5CH6; (l) 5CH5–4AH8; (m) 10C^{4}NH$_{2b}$–10C^{4}NH$_{2n}$; (n) 10C^{4}NH$_{2b}$–10CH5; (o) 10C^{4}NH$_{2n}$–10CH5; (p) 10CH5–10CH6; (q) 10CH5–9AH8; (r) 9A^{6}NH$_{2b}$–9A^{6}NH$_{2n}$; (s) 8A^{6}NH$_{2b}$–8A^{6}NH$_{2n}$; (t) 8A^{6}NH$_{2b}$–3GH1′; (u) 8A^{6}NH$_{2b}$–9AH2; (v) 8A^{6}NH$_{2b}$–4AH8; (w) 4A^{6}NH$_{2b}$–4A^{6}NH$_{2n}$; (x) 4A^{6}NH$_{2b}$–7GH1′; (y) 4A^{6}NH$_{2b}$–8AH8; (aa) 3G^{2}NH$_{2b}$–8A^{6}NH$_{2n}$; (bb) 8A^{6}NH$_{2b}$–3G^{2}NH$_{2n}$; (cc) 4A^{6}NH$_{2b}$–7G^{2}NH$_{2n}$.

an H-bonded T-imino proton usually exhibits three strong NOEs to its paired A-^{6}NH$_{2b}$, A-^{6}NH$_{2n}$, and A-H2 protons. These corresponding NOEs are clearly depicted in Figure 4 (cross peaks a, b, and c). On the contrary, an H-bonded G-imino proton only exhibits strong NOEs to its base-paired C-^{4}NH$_{2b}$ and C-^{4}NH$_{2n}$ located in the major groove; the NOEs to its own G-^{2}NH$_{2b}$ located in the minor groove are usually not detected even at a very low temperature or under acidic conditions in a regular Watson–Crick base-paired duplex (Chou et al., unpublished results; Rajagopal and Feigon, 1989; Radhakrishnan et al., 1991; Mueller et al., 1995). This phenomenon is possible due to the efficient relaxation transfer through the H_2O molecules well aligned in the minor groove, as observed in a DNA crystal structure (Han et al., 1997), or caused by rotation of the G-amino group around the C-N bond (Mueller et al., 1995). The strong NOE cross peaks from 6G- and 1G-imino protons to their corresponding 5C-^{4}NH$_{2b}$/5C-^{4}NH$_{2n}$ (cross peaks f and g) and 10C-^{4}NH$_{2b}$/10C-^{4}NH$_{2n}$ NOEs (cross peaks i and j) can therefore be assigned with ease. However, contrary to what occurred in a Watson–Crick G•C pair, strong NOE cross peaks can be detected be-

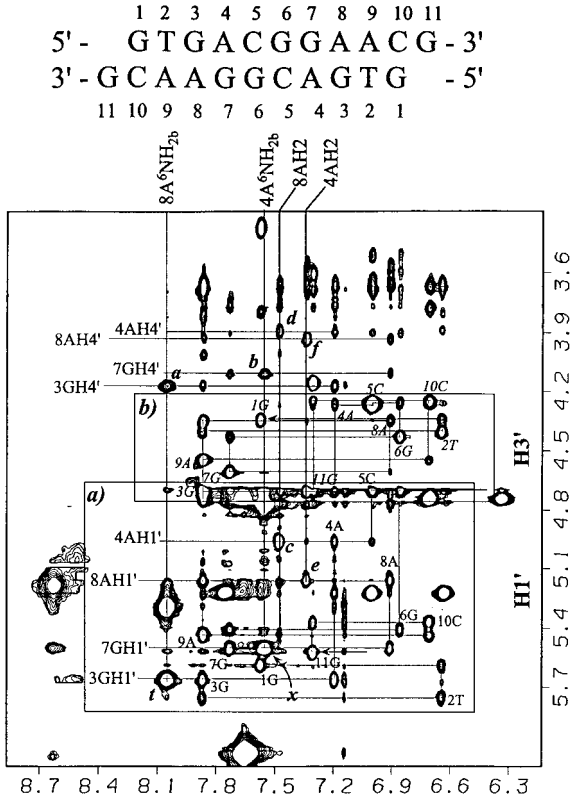

Figure 6. The expanded H_2O/NOESY contour plot for the aromatic proton to H1′ (boxed in a) and aromatic proton to H3′ (boxed in b) regions of sequence 3. Only intra-nucleotide aromatic–H1′/H3′ are labeled with the nucleotide numbers. Other cross peaks of interest are labeled as follows: (a) 8A^{6}NH$_{2b}$–3GH4′; (b) 4A^{6}NH$_{2b}$–7GH4′; (c) 8AH2–4AH1′; (d) 8AH2–4AH4′; (e) 4AH2–8AH1′; (f) 4AH2–8AH4′; (t) 8A^{6}NH$_{2b}$–3GH1′; (x) 4A^{6}NH$_{2b}$–7GH1′.

tween the 3G- and 7G-imino protons and their own ^{2}NH$_{2b}$/^{2}NH$_{2n}$ protons (cross peaks m, n and v, w). This unusual phenomenon is supportive of the notion that in a sheared G•A pair, the guanosine pairs with the adenosine through its minor groove edge (Figure 3i) to repel the H_2O molecules possibly located in the minor groove, which prevents the relaxation transfer through the H_2O molecules. In this figure, many weaker NOE cross peaks exhibited by 3G- and 7G-imino protons were also detected across the strand. For example, those between 3G-imino and 7G-H1′/7G-H8 (peaks r and s) and those between 7G-imino and 3G-H1′/3G-H8 (peaks bb and aa) are further evidence that residue 3G cross-stacks with residue 7G in the (TGAC):(GGAA) motif.

Figure 5 shows the amino-proton-related NOE cross peak region. Five strong cross peaks (a, h, m, r, and s) can be readily detected. Four of them (a, h, m, and r) are the geminal NH$_{2b}$ ↔ NH$_{2n}$ cross peaks belonging to residues 3G, 5C, 10C, and 9A. The presence of these strong cross peaks confirms the imino-NH$_2$ assignments shown in Figure 4. The other strong cross peak (s), however, cannot be assigned by this pathway. It was identified as the geminal 8A-^{6}NH$_{2b}$ ↔ 8A-^{6}NH$_{2n}$ cross peak by the mutual NOEs (Figure 7b) between the 8A-^{6}NH$_{2b}$ ↔ 3G-^{2}NH$_{2n}$ (cross peak bb) and between the 8A-^{6}NH$_{2n}$ ↔ 3G-^{2}NH$_{2b}$ (cross peak aa) proton pairs. The strong cross-strand NOEs to 3G-H1′ (t) and medium strength NOEs to 3G-H4′ (cross peak a in Figure 6) also confirm the 8A-^{6}NH$_{2b}$ assignment. The geminal 7G-^{2}NH$_{2b}$ ↔ 7G-^{2}NH$_{2n}$ cross peak (cross peak f) is much weaker, due to the rather broad linewidth of the 7G amino protons (Figure 2c). But its presence can be confirmed by the rather strong 7G-imino ↔ 7G-^{6}NH$_2$ NOE cross peaks v and w as revealed in Figure 4. A weak cross-strand NOE from 7G-^{2}NH$_{2b}$ to 3G-H1′ (cross peak g) was also detectable. The reason for the broader ^{2}NH$_2$ linewidth for residue 7G compared to 3G is not clear at the moment. The geminal 4A-^{6}NH$_{2b}$ ↔ 4A-^{6}NH$_{2n}$ cross peak (w) is close to the water ridge and more difficult to assign, but the similar cross-strand NOE pattern exhibited by 4A-^{6}NH$_{2b}$ to 7G-H1′ (cross peak x), 7G-H4′ (cross peak b in Figure 6), and 7G-^{2}NH$_{2n}$ (cross peak cc) as those shown by 8A-^{6}NH$_{2b}$ confirms its assignment. The capability to confidently assign the 3G, 7G, 4A, and 8A amino protons and to detect the many related NOEs exhibited by these protons unambiguously defines the sheared 3G•8A and 7G•4A base pairing nature at neutral pH. The quite large chemical shift differences between the two geminal G and A amino protons (> 2.5 ppm) also indicate that one of them participates in H-bonding while the other does not.

In the regular Watson–Crick G•C base pairs of this oligomer, further NOE cross peaks from 5C- and 10C-^{4}NH$_{2b}$ to their corresponding C-H5 were also detected (i and n), due to the efficient spin-diffusion from ^{4}NH$_{2b}$ to ^{4}NH$_{2n}$ (see Figure 3g). Besides, one can also detect the direct NOE cross peaks from 5C- and 10C-^{4}NH$_{2n}$ to their corresponding C-H5 protons (j and o). The assignments can even be extended from C-H5 to C-H6 (cross peaks p and k) to connect the exchangeable proton connectivity to the non-exchangeable proton connectivity (see Figure 6).

Figure 6 shows the non-exchangeable NOE connectivities for the aromatic to H1′ and H3′ regions of the H_2O/NOESY spectrum of sequence 3. Even

though many exchangeable proton-related NOE cross peaks are present, the aromatic to H1′ (box a) and aromatic to H3′ (box b) NOE cross peaks are still well resolved, and can be connected using the well-established assignment procedure developed for D_2O/NOESY spectra (Hare et al., 1983). Similar results were obtained for the D_2O/NOESY spectrum under identical conditions (data not shown). The quite detectable cross-strand NOEs from 4A-H2 to 8A-H1′ and 8A-H4′ (peaks e and f) and from 8A-H2 to 4A-H1′ and 4A-H4′ (peaks c and d) are obvious, and have been detected before in the $(PyGAPu)_2$ motifs (Chou et al., 1992, 1994a).

Structure determination of the 5′-(GGAA)/(TGAC)-3′ motifs

Three-dimensional structures for sequence 3 containing the 5′-(GGAA)/(TGAC)-3′ motifs were generated by distance geometry and molecular dynamics calculations using distance and dihedral angle constraints derived from NMR experiments. The combined exchangeable and non-exchangeable idiosyncratic NOEs around the sheared G•A pair region are summarized in Figure 7, and the constraints used to determine the solution structure are listed in Table 1. The dihedral angle constraints were derived from the two-dimensional ^{1}H-^{31}P correlation spectrum. Ten approximately equal strength $^{4}J_{H4'-P}$ cross peaks were detected (data not shown), indicating that all four consecutive bonds in the H4′–C4′–C5′–O5′–P backbone linkages lie in the same plane to adopt the regular $\beta(t)\gamma(g^+)$ conformation (Sarma et al., 1973; Chou et al., 1996).

After initial distance geometry calculation, 15 out of 20 embedded structures with the lowest energies were further submitted to molecular dynamics calculations with 10000 cycles of dynamics run at 300 K and 6000 cycles of conjugate gradient energy minimization in the AMBER force field. The 15 structures all exhibit correct folding and overlap well with the average structure (rmsd values of 0.5 Å). No distance violation of larger than 0.15 Å and dihedral angle violation larger than 3° was observed for all final structures (structural statistics are listed in Table 1). The overlapped final 15 structures are shown in Figure 8, and its overall structure is very similar to the one containing two symmetrical $(PyGAPu)_2$ [5′-(TGAG)/(CGAA)-3′] motifs we have determined previously (Chou et al., 1994a). Large propeller-twisting (ca. 25°) is observed for the sheared G•A base pairs. All χ glycosidic bonds are in the anti domain. Most backbone torsion angles are also within the regular B-DNA domain, i.e. in the $\alpha(g^-)\beta(t)\gamma(g^+)\epsilon(t)\zeta(g^-)$ ranges, except for the ϵ and ζ angles connecting G-A nucleotides in the tandem G•A pairs, which are closer to the *gauche*$^-$ and *trans* domains respectively. α Torsion angles connecting the tandem G•A pairs also deviate from the *gauche*$^-$ domain and are located at around $-120°$.

Table 1. Structural statistics for the d(GTGACGGAACG) duplex

Restraints	Numbers
Exchangeable NOEs	
H-bonds (1.8–2.1 Å)	24
2.0–5.0 Å	10
3.0–6.0 Å	66
Non-exchangeable NOEs	
1.8–3.0 Å	74
2.0–4.0 Å	104
3.0–6.0 Å	134
> 5 Å	26
Total NOEs	438
Torsion angles	78
Backbone (β, γ, ϵ)	56
Glycosidic	22
NOEs per residue	20
NOEs and torsion angles per residue	23.5
Violations of experimental restraints	
Distance restraints (> 0.15 Å)	0
Torsional angles restraints (>3°)	0
rmsd	0.5 ± 0.1 Å

Discussion

The sheared G•A pair is an important motif in nucleic acid structure. It not only occurs in nature (Shlomai and Kornberg, 1980; Pley et al., 1994; Ferrer et al., 1995; Cate et al., 1996; Correll et al., 1997; Huang et al., 1998; Ortiz-Lombardia et al., 1998), but also in the nucleic acid aptamers screened by in vitro selection. For example, a large DNA hairpin loop is zippered up to form an amino acid binding pocket involving three sheared G•A pairings (Lin et al., 1998). However, the direct evidence for this important pairing geometry at a physiological pH still remains elusive. Maskos et al. (1993) were able to show some NMR evidence for the protonated sheared A•A pairing in the 5′-(CGAA)/(TAAG)-3′ motif at a rather acidic pH (4.23). However, upon raising the pH to 4.72, one

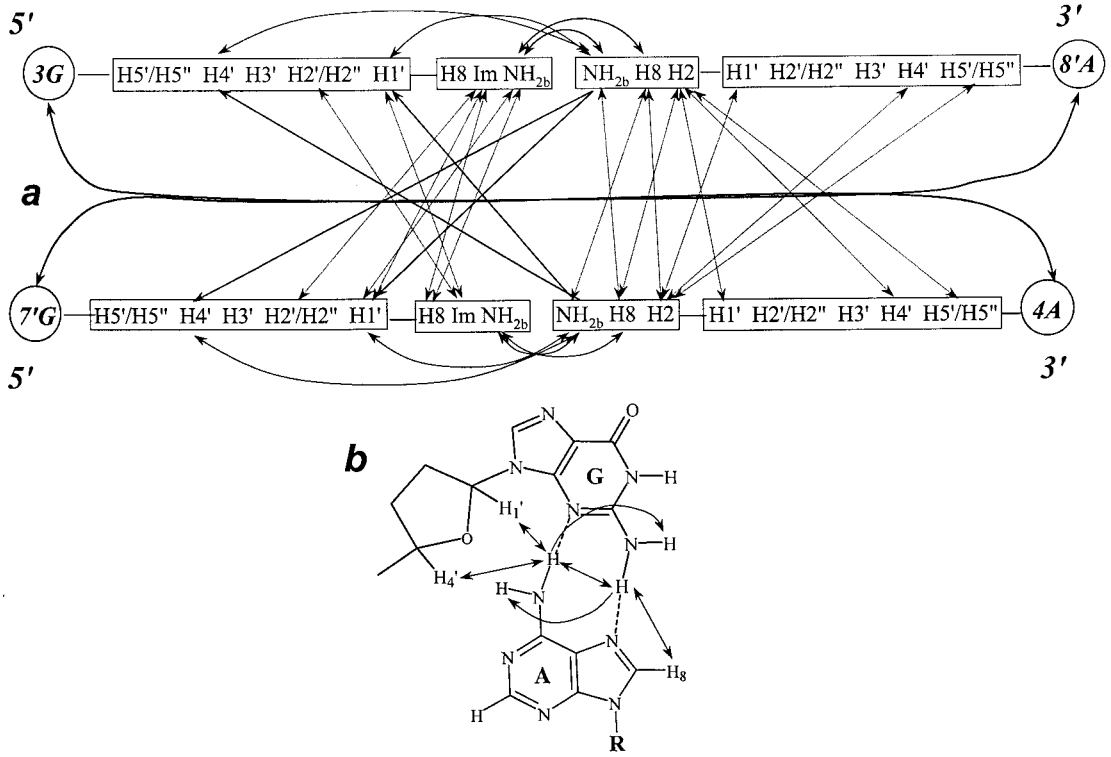

Figure 7. The scheme of the NOE restraints of the exchangeable and non-exchangeable protons around the tandem G•A base pairs (a) and the critical NOEs detected for a sheared G•A base pair (b).

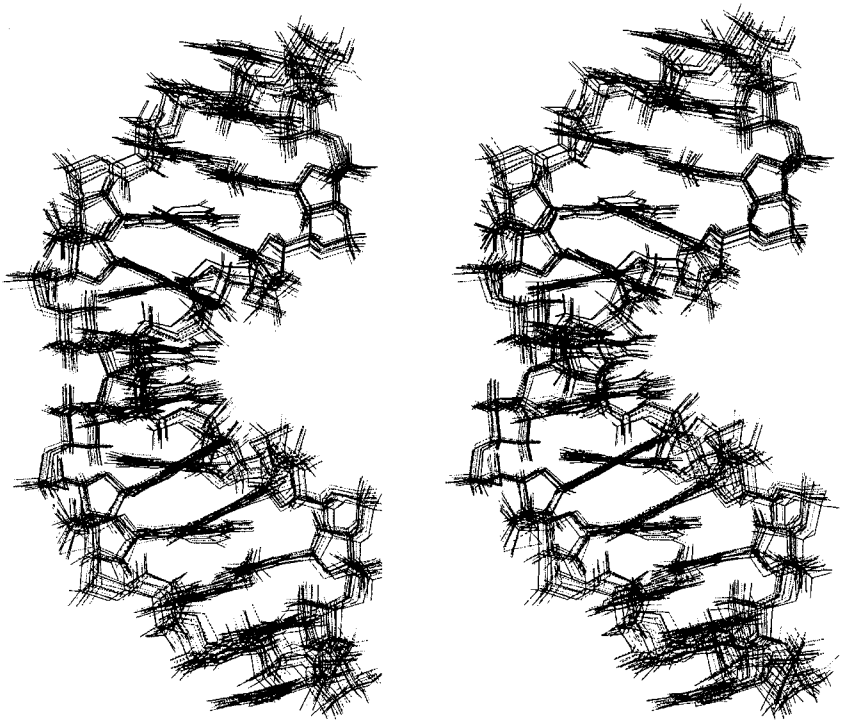

Figure 8. Superimposed wide-eye stereo view of the final 15 structures of the 5′-(GTGACGGAACG)$_2$-3′ duplex. These final structures overlap well with the average structure with rmsd values of 0.5 ± 0.1 Å.

of the A^6NH_2 NMR signals quickly disappeared. At pH around 6, almost no such signal could be detected (Maskos et al., 1993). We are able, for the first time, to observe very clear NOEs for all G- and A-NH_2 protons of the sheared G•A pair in the 5′-(G**GA**A)/(T**GA**C)-3′ motif at neutral pH.

Thermodynamics studies have shown that non-symmetrical 5′-(PuGAPu)/(PyGAPy)-3′ motifs are no less stable than their symmetrical counterparts (Li et al., 1991b; Li and Agrawal, 1995). Although nucleic acid structures containing such symmetrical 5′-$(PyGAPu)_2$-3′ motifs have been well studied before [for a review, see Chou et al. (1997)], few NMR studies are available for non-symmetrical 5′-(Pu**GA**Pu)/(Py**GA**Py)-3′ motifs (Katahira et al., 1993; Li and Agrawal, 1995). This is the first detailed structural study for such motifs. Our extensive NMR data for both 5′-(GGAG)/(CGAC)-3′ (this manuscript) and 5′-$(CGAG)_2$-3′ (Cheng et al., 1992) sequences have indicated that they form a similar sheared G•A motif, which is, however, not detected for the 5′-$(GGAC)_2$-3′ sequence. The reason why the polarity of the Watson–Crick C•G pair preceding or following the sheared G•A pair has such a dramatic effect on the stability of the parent oligomer is still not clear. But the finding that such non-symmetrical 5′-(PuGAPu)/(PyGAPy)-3′ sequences can also form stable motifs adds to the repertoire for unusual nucleic acid structural formations.

The NOE between a G-imino proton and its own G-$^2NH_2$ amino proton in a regular Watson–Crick G•C paired duplex has rarely been observed. This phenomenon has been attributed to the fast rotation of the ^2C-^2N single bond (Mueller et al., 1995). However, we have detected pretty strong (n)G-imino proton to (n)G-$^2NH_2$ amino proton NOEs (cross peaks m, n and v, w in Figure 4). It is not clear if the G-$^2NH_2$–A^7N H-bond in a sheared G•A pair is stronger than the G-$^2NH_2$–C^2O H-bond in a regular G•C pair, but another explanation is that the adenosine pairs with the guanosine through the minor groove edge to repel the H_2O molecules possibly located in the minor groove. This can decrease the relaxation transfer through these H_2O molecules and retains the strong NOEs to the G-$^2NH_2$ protons.

Conclusions

As we (Chou and Tseng, 1999; Chou et al., 1999) and others (Mueller et al., 1995) have shown, inclusion of NH_2-related constraints can greatly increase the accuracy of the unusual nucleic acid structure determination, even though only loose constraints are utilized. Direct pieces of evidence for the sheared G•A pairs have been obtained at neutral pH at a relatively low temperature (−5 °C). Our results show that for some unusual nucleic acid structures, NH_2-related NOE data can be obtained without the need to prepare ^{15}N isotope-enriched samples (Mueller et al., 1995). Only a high quality H_2O/NOESY spectrum at a lower temperature would be sufficient. The strong NOEs exhibited by the unpaired G-imino proton in the sheared G•A pair to its own G-$^2NH_2$ are quite unusual, and imply that G-$^2NH_2$ is involved in H-bonding with a base in the minor groove edge.

Acknowledgements

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

References

Altona, C. (1982) *Recl. Trav. Chim. Pays-Bas*, **101**, 413–433.
Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Kundrot, C.E., Cech, T.R. and Doudna, J.A. (1996) *Science*, **273**, 1678–1685.
Cheng, J.-W., Chou, S.-H. and Reid, B.R. (1992) *J. Mol. Biol.*, **228**, 1037–1041.
Chou, S.-H., Cheng, J.-W., Fedoroff, O. and Reid, B.R. (1994a) *J. Mol. Biol.*, **241**, 467–479.
Chou, S.-H., Cheng, J.-W. and Reid, B. (1992) *J. Mol. Biol.*, **228**, 138–155.
Chou, S.-H., Flynn, P. and Reid, B. (1989) *Biochemistry*, **28**, 2422–2435.
Chou, S.-H., Hare, D.R., Wemmer, D.E. and Reid, B.R. (1983) *Biochemistry*, **22**, 3037–3041.
Chou, S.-H. and Tseng, Y.-Y. (1999) *J. Mol. Biol.*, **285**, 41–48.
Chou, S.-H., Tseng, Y.-Y. and Wang, S.-W. (1999) *J. Mol. Biol.*, in press.
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. (1994b) *J. Mol. Biol.*, **244**, 259–268.
Chou, S.-H., Zhu, L. and Reid, B.R. (1997) *J. Mol. Biol.*, **267**, 1055–1067.
Correll, C.C., Freeborn, B., Moore, P.B. and Steitz, T.A. (1997) *Cell*, **91**, 705–712.
Crook, S.T. and Bennett, C.T. (1996) *Annu. Rev. Pharmacol. Toxicol.*, **36**, 107–129.
Dallas, A. and Moore, P.B. (1997) *Structure*, **5**, 1639–1653.

Delihas, N., Rokita, S.E. and Zheng, P. (1997) *Nat. Biotechnol.*, **15**, 751–753.
Ferrer, N., Azorin, F., Villasante, A., Gutierrez, C. and Abad, J.P. (1995) *J. Mol. Biol.*, **245**, 8–21.
Han, G.W., Kopka, M.L., Cascio, D., Grzeskowiak, K. and Dickerson, R.E. (1997) *J. Mol. Biol.*, **269**, 811–826.
Hare, D.R., Wemmer, D.E., Chou, S.-H., Drobny, G. and Reid, B.R. (1983) *J. Mol. Biol.*, **171**, 319–336.
Huang, C.-H., Lin, Y.-S., Yang, Y.-L., Huang, S.-w. and Chen, C.W. (1998) *Mol. Microbiol.*, **28**, 905–916.
James, K.D. and Ellington, A.D. (1997) *Chem. Biol.*, **4**, 595–605.
Katahira, M.H.S., Mishima, K., Uesugi, S. and Fujii, S. (1993) *Nucleic Acids Res.*, **21**, 5418–5424.
Li, Y. and Agrawal, S. (1995) *Biochemistry*, **34**, 10056–10062.
Li, Y., Zon, G. and Wilson, W.D. (1991a) *Proc. Natl. Acad. Sci. USA*, **88**, 26–30.
Li, Y., Zon, G. and Wilson, W.D. (1991b) *Biochemistry*, **30**, 7566–7572.
Lin, C.-H., Wang, W., Jones, R.A. and Patel, D.J. (1998) *Chem. Biol.*, **5**, 555–572.
Maskos, K., Gunn, B.M., LeBlanc, D.A. and Morden, K. M. (1993) *Biochemistry*, **32**, 3583–3595.
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.
Ortiz-Lombardia, M., Cortes, A., Huertas, D., Eritia, R. and Azorin, F. (1998) *J. Mol. Biol.*, **277**, 757–762.
Pley, H.W., Flaherty, K.M. and McKay, D.B. (1994) *Nature*, **372**, 68–74.
Radhakrishnan, I., Gao, X., de los Santos, C., Live, D. and Patel, D.J. (1991) *Biochemistry*, **30**, 9022–9030.
Rajagopal, P. and Feigon, J. (1989) *Nature*, **339**, 637–640.
Sarma, R.H., Mynott, R.J., Wood, D.J. and Hruska, F.E. (1973) *J. Am. Chem. Soc.*, **95**, 6457–6459.
Shepard, W., Cruse, W.B.T., Fourme, R., Fortelle, E.d.l. and Prange, T. (1998) *Structure*, **6**, 849–861.
Shlomai, J. and Kornberg, A. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 799–803.
Walter, A.E., Wu, M. and Turner, D.H. (1994) *Biochemistry*, **33**, 11349–11354.
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.