

Structure of a small RNA hairpin

Peter W. Davis⁺, William Thurmes[§] and Ignacio Tinoco Jr^{*}

Department of Chemistry and Laboratory of Chemical Biodynamics, University of California, Berkeley, CA 94720, USA

Received October 19, 1992; Revised and Accepted December 18, 1992

ABSTRACT

The hairpin stem-loop form of the RNA oligonucleotide rCGC(UUU)GCG has been studied by NMR spectroscopy. In 10 mM phosphate buffer this RNA molecule forms a unimolecular hairpin with a stem of three base pairs and a loop of three uridines, as judged by both NMR and UV absorbance melting behavior. Distance and torsion angle restraints were determined using homonuclear proton-proton and heteronuclear proton-phosphorus 2-D NMR. These values were used in restrained molecular dynamics to determine the structure of the hairpin. The stem has characteristics of A-form geometry, although distortion from A-form occurs in the 3'-side of the stem, presumably to aid in accommodating the small loop. The loop nucleotides adopt C2'-endo conformations. NOE's strongly suggest stacking of the uracils with the stem, especially the first uracil on the 5'-side of the loop. The reversal of the chain direction in the loop seems to occur between U5 and U6. Loop structures produced by molecular dynamics simulations had a wide range of conformations and did not show stacking of the uracils. A flexible loop with significant dynamics is consistent with all the data.

INTRODUCTION

A variety of secondary and tertiary structures are required for the diverse biological functions of RNA (1). A hairpin loop is the functional element in a number of well characterized systems. The hairpin structure is required in the HIV TAR RNA for transactivation by the tat protein, and specific nucleotides within the loop are required for activity (2). R17 and Q β coat proteins bind to particular RNA hairpin-loop structures; the sequence is unimportant except for a few key unpaired nucleotides (3, 4). Extraordinarily stable tetra-loops of specific sequences occur much more often in ribosomal RNA than chance would dictate (5). These results suggest that sequence-dependent structural variation, rather than the primary sequence itself, may be important factors in RNA activity.

NMR studies in the past decade established that complementary RNA adopts a regular A-form structure. Haasnoot *et al.* (6)

showed that oligoribonucleotides in solution show characteristics of A-form structure. Happ *et al.* (7) developed a more detailed A-form model of r(CGAUGC)₂ with the combined use of NMR and restrained molecular dynamics. Studies on two hairpin stem-loops showed that the structures in solution were consistent with t-RNA crystal structures (8, 9), with a notable feature being single-stranded stacking by five loop nucleotides on the 3'-side of the stem. These models were based almost exclusively on proton NMR data. With this limited information, development of models required significant assumptions about phosphate-backbone conformations and base-stacking that might preclude the discovery of distinct differences among hairpin structures.

The recent advances in NMR data acquisition, processing, and interpretation have allowed more details to be determined experimentally, thus resulting in higher resolution structures. Studies on RNA tetra-loops show that these hairpins form structures with additional, unusual hydrogen-bonding and base-pairing interactions: a reverse-wobble G-U pair with the G in a syn conformation for the UUCG and UUUG loops (10), and an unusual G·A base pair in GCAA and GAAA loops (11). High-resolution studies of DNA hairpins have also uncovered sequence-dependent structural heterogeneity within the loops (12). NMR analysis of an RNA hairpin with two non-Watson-Crick base-pairs and a loop of three nucleotides (13) showed that these structural features are accommodated with apparently little distortion of the stem from A-form geometry. In addition, the results suggested that two of the three bases in the loop are involved in stacking on the 5'-side of the stem. It is clear that simple models are not sufficient to describe RNA hairpin stem-loop structure.

The present study was designed to determine the structure of a very small hairpin loop which is closed by Watson-Crick base-pairing. Although the previously mentioned tetra-loop hairpins contain loops of only two unpaired bases, this is accomplished through unusual base-pairing. The hairpin containing a three-base loop studied by Puglisi *et al.* (13) showed very little distortion from A-form in the stem region, however, it is unknown whether the unusual base pairs actually helped to accommodate the small loop. We chose to study the hairpin form of rCGC(UUU)GCG to determine how the steric constraints imposed by a small loop affect the overall structure of an RNA hairpin.

* To whom correspondence should be addressed

Present addresses: ⁺Isis Pharmaceuticals, 2292 Faraday Ave., Carlsbad, CA 92008 and [§]Displaytech, 2200 Central Ave, Boulder, CO 80301, USA

MATERIALS AND METHODS

Synthesis of RNA oligonucleotide

Preparation of monomers and supports. The RNA oligonucleotide was synthesized based on the strategy of Kierzek *et al.* (14), with some modifications as described. 5'-O-DMT-2-O-THP-N-amide-protected ribonucleoside monomers were prepared essentially as described by Markiewicz *et al.* (15). Exocyclic amino groups were protected with benzoyl groups for cytosine and isobutyryl for guanine. Nucleoside-bound supports were prepared as in (14), except for the use of 500 Å Long Chain Alkyl Amine (LCAA) support (Pierce) rather than 3-aminopropyl-derivatized silica. Typical loading was ~20 μ mole nucleoside per gram of support. Phosphoramidites were prepared as described for DNA by Atkinson and Smith (16) with the following modifications. Chloro-N,N-diisopropylamino-2-cyanoethoxyphosphine (ABN) was used instead of chloro-N,N-diisopropylaminomethoxyphosphine to achieve β -cyanoethyl- rather than methyl-protected phosphates (17). In addition the reaction time was extended to an hour or more to complete the reaction, and dry benzyl alcohol was used to quench excess phosphine. Fully protected ribonucleoside phosphoramidites were purified by silica gel chromatography (Merck, 230–400 mesh) with a gradient of hexane:acetone from 3:1 (v:v) to 2:3 with a constant 5% by volume of triethylamine. Fractions containing purified products were evaporated to a hard foam and crushed to a fine powder for use in oligonucleotide synthesis.

Oligonucleotide synthesis. Oligonucleotides were synthesized manually using a small vessel containing a sintered glass frit and a stopcock with a 14/20 joint on top to accommodate a rubber septum. In preparation for oligonucleotide synthesis, phosphoramidites and tetrazole were stored *in vacuo* over phosphorus pentoxide for at least 24 hours. Acetonitrile for washing and coupling reactions was stored over two different batches of dry 3 Å pore size molecular sieves for 24 hours each; preparing exceptionally dry acetonitrile and keeping it dry during oligonucleotide synthesis had the greatest single effect on the success of the synthesis. Reagents were introduced by syringe and all reactions were carried out under nitrogen.

It was found that the time required for complete removal of DMT groups with 0.3% DCA (14) varied significantly for the different bases. Guanosine detritylates completely in about 5 minutes, while uridine requires at least 10 minutes exposure for complete detritylation. Phosphoramidite for each coupling was dissolved directly prior to use, then mixed with tetrazole solution in the reaction vessel and allowed to react with the support for 15 minutes. After the final coupling reaction the last DMT was removed.

Deprotection and purification. Oligonucleotides were deprotected by heating the support in fresh concentrated ammonia at 55°C for 16 hours to remove benzoyl, isobutyryl and 2-cyanoethyl groups. The solution was then evaporated to a small volume and 2 ml of 0.01M aqueous HCl was added and the pH adjusted to 2.0 with 1.0 M HCl. The solution was stirred at room temperature for 20 hours with the pH kept at 2.0 to remove 2'-O-THP groups. For this last step and all subsequent manipulations, nuclease contamination was avoided by using latex gloves, sterile glass- and plastic ware, and doubly-distilled water.

Oligonucleotides were first purified by ion-exchange HPLC on DEAE with a gradient of 0.05 M to 1.5 M aqueous triethylammonium acetate at pH 6, with 20% acetonitrile. Fractions containing product were lyophilized repeatedly then

desalted through a column of Sephadex G-10 under moderate hydrostatic pressure. The product was dried and subsequently purified by reversed-phase HPLC with a gradient of 1 to 15% acetonitrile in 0.1 M ammonium acetate, pH 6. The product was again desalted on Sephadex G-10, and finally all counterions were replaced with sodium using a small (4×2 cm) column of Dowex 50-8 resin which had been properly conditioned. The final product was dried by lyophilization. Total yield for a 6 μ mole scale synthesis was 100 ODU (1.3 μ mole) of purified RNA.

Enzymatic digestion analysis. 0.1 ODU of purified RNA was dissolved in 25 μ l of 50 mM Tris, 25 mM MgCl₂, pH 7.5. 1 unit (1 μ l) calf intestinal phosphatase (CIP; Boehringer/Manheim) and 1 unit (1 μ l) of ribonuclease T2 (RNase T2; BRL) were added, and the solution was incubated at 37°C for 1 hour. A similar digestion was prepared in parallel but without the addition of RNase T2. The solution was then diluted to 75 μ l with water and ultra filtered (10,000 MW cutoff; Millipore) to remove protein. The filtered solutions were analyzed by reversed-phase HPLC at 260 nm (18) with a gradient of 0 to 20% acetonitrile in 100 mM ammonium acetate, pH 5.5. Peaks were integrated and converted to molar ratios with the proper extinction coefficients (19).

Thermodynamic analysis

Equilibrium melting curves were collected on a Gilford model 250 UV-Vis spectrophotometer with a Gilford model 2527 thermoprogrammer interfaced to an Apple IIe computer. RNA was dissolved in the standard buffer (10 mM sodium phosphate, 0.1 mM EDTA, pH 6.5) at strand concentrations of 24 μ M, 220 μ M, and 2.4 mM in 1 cm, 1 mm and 0.1 mm cells, respectively, to give an absorbance of ~1 at 280 nm. Absorbance data were collected at 280 nm and fitted to a two-state model.

NMR spectroscopy

Data acquisition. For all NMR measurements RNA was dissolved in 0.4 ml of the standard 10 mM phosphate buffer to a concentration of 2.4 mM, lyophilized repeatedly from 99.8%

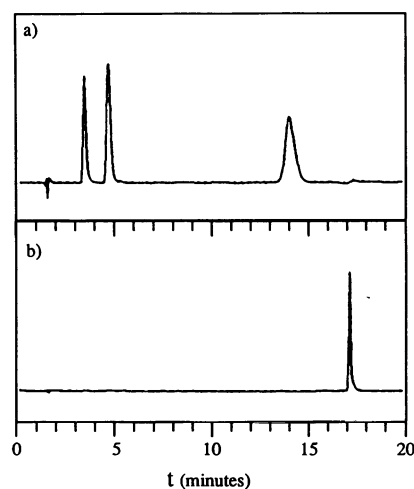


Figure 1. (a) HPLC trace of rCGCUUUGCG after 1 hour digestion with RNase T2 and CIP. The order of elution is rU (1.5 min), rC (4.8 min) and rG (14.1 min). (b) Trace of the same reaction mixture without T2, showing full-length oligonucleotide (17.1 min).

D₂O, then brought up into 0.4 ml 100.00% D₂O (Aldrich) under nitrogen. Spectra were taken on a General Electric GN 500 spectrometer. Unless otherwise noted data were collected at 25°C. 2-D NOESY, ³¹P-decoupled-2-quantum-filtered-COSY (2QF-COSY), double-quantum (DQ) and ¹H-³¹P heteronuclear correlated (HETCOR) experiments were performed and analyzed as described (10). Exchangeable protons were observed in 5% D₂O/95% H₂O using the 1-3-3-1 pulse sequence.

Analysis and interpretation. NMR data were used to create distance and angle restraints for use with the XPLOR (20) simulated annealing and molecular dynamics program. Approximate interproton distances were obtained from a 400 ms NOESY experiment using the two-spin approximation. Cross-peak volumes were integrated and used as estimates of the cross relaxation rate; the distance is calculated by comparison to the average pyrimidine H5-H6 cross-peak volume. Measured values were compared to expected values for both short (H5'-H5'') and medium (H1'-H2') distances and judged to be accurate to within 0.5 Å; significant spin diffusion was not a problem with this relatively small molecule. Distances were then classified as short, 1.5–3.5 Å; medium, 2.0–4.5 Å; or long, 2.5–6 Å, for use with XPLOR.

Proton-proton J-coupling constants were extracted from a high resolution (1 Hz/point in ω_2) ³¹P-decoupled-2QF-COSY and HETCOR experiments. Most couplings were measured directly as active couplings. In some cases passive couplings were used to estimate J, mostly for the H3'-H4' and H4'-H5'/H5'' couplings. Proton-phosphorus couplings were measured as active couplings from the HETCOR experiment.

Both NOE's and coupling constants were used to determine allowed dihedral angles. The angle γ (O5'-C5'-C4'-C3') was determined from H6/H5-H5'/H5'' NOE connectivities (21) and H4'-H5'/H5'' coupling constants (22). The angle δ (C4'-C3'-C2'-C1') was determined from the H1'-H2' coupling (22, 23). ϵ was determined from the H3'-³¹P coupling constant (22). Qualitative α and ζ (C5'-O5'-P-O3', O5'-P-O3'-C3') values (gauche or trans) were estimated from phosphorus chemical shift values.

Energy minimization and restrained molecular dynamics

Structures based on NMR data were created with XPLOR using a series of three simulated annealing cycles (24, 25). The annealing cycle involved 1) energy minimization; 2) restrained

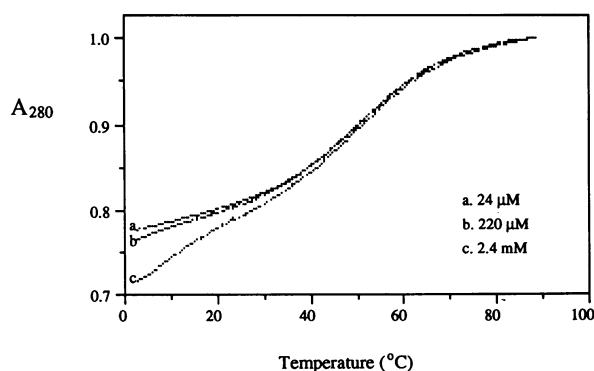


Figure 2. Normalized absorbance melting curves for rCGCUUUGCG in standard buffer: 10 mM sodium phosphate, 0.1 mM EDTA, pH 6.5. The lack of concentration dependence of the absorbance indicates that the melting is unimolecular; the transition is from hairpin to coil.

molecular dynamics (RMD) at 1000° K; 3) RMD while incrementing the relevant force constants to their final values; 4) RMD during which the temperature is decremented to 300° K; and 5) energy minimization.

Seven starting structures with randomized torsion angles were created with XPLOR. The first annealing step allowed the random coil to fold into the correct general shape using only NOE-derived distance constraints and base-pairing and base-pair stacking distance constraints. This initial annealing was followed by refinement annealing during which the torsion angle restraints were introduced. Refinement was done twice, first with the β , δ , γ and ϵ angle restraints, then again including α and ζ restraints. This stepwise introduction of restraints discouraged trapping of high-energy structures. Of the seven starting structures, only one failed to converge to a reasonable energy during the first annealing cycle. The remaining six converged to very similar final energies during each of the three annealing cycles.

RESULTS

RNA synthesis

Both enzymatic digestion and NMR indicate the synthesis of the desired oligonucleotide. The HPLC chromatogram of the RNase T2 digest in Figure 1 shows only peaks corresponding to cytidine, uridine and guanosine; the relative molar ratios were 0.96 : 0.96 : 1.08 for rC : rU : rG, which is close to the expected ratio of 1 : 1 : 1. Since RNase T2 is specific for RNA 3'-5' linkages (26), the oligonucleotide contains only the proper 3'-5' linkages. The limit of detection for undigestible 2'-5' linkages is estimated to be <0.2% per site, or <2% per molecule. This analysis also indicates that there was no significant base modification during the synthesis and that the bases are completely deprotected (18). In addition, the sequential assignment of resonances in the NMR spectra were consistent with an oligonucleotide of nine residues with the proper sequence.

The yield of purified oligonucleotide from a 6 μ mole-scale synthesis was ca. 100 A₂₆₀ units, or 1.3 μ mole, for a 22% yield. Thus the stepwise yield was at least 83% per coupling, and likely greater than that, since there was a significant loss of material during the purification steps. Qualitative observation of trityl cation suggested greater than 90% efficiency per coupling.

Thermodynamics

The equilibrium melting curves were obtained over a 100-fold range of RNA concentration to determine both the concentration dependence and thermodynamic parameters for the hairpin-to-random coil transition. Figure 2 shows that the curves are nearly superimposable, with the exception of the lower baseline of the most highly concentrated sample. The very similar T_m's strongly indicate the presence of a monomolecular hairpin stem-loop structure. The sloping low baseline of the 2.4 mM sample

Table 1. Thermodynamic parameters for hairpin formation obtained from the melting curves shown in Figure 2.

Strand Concentration (mM)	ΔH° (kcal/mol)	ΔS° (e.u.)	T _m (°C)	ΔG° (kcal/mol)
2.4	-34.9	-105.7	57.0	-3.4
0.22	-24.7	-75.5	54.0	-2.2
0.024	-23.3	-71.3	53.6	-2.0

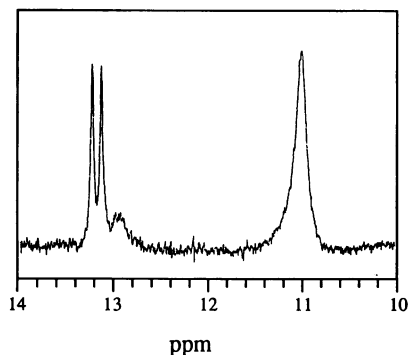


Figure 3. Downfield region of the NMR spectrum of rCGCUUGCG in standard buffer with 95% H₂O/5% D₂O at 22°C. The three hydrogen-bonded guanine H1 protons resonate near 13 ppm, while the three non-hydrogen-bonded uracil H3 protons are at 11 ppm. The two sets of resonances integrate differently due to the uneven excitation profile of the 1-3-3-1 pulse.

suggests the formation of a duplex structure at low temperature, so the data from this sample were not considered in the analysis.

The results of the thermodynamic analysis are shown in Table 1. There is good agreement between two samples of the lowest concentration, with an average ΔH° of -24 kcal/mole. The calculated ΔH° in 1M NaCl (27) for a molecule containing this sequence is -22.7 kcal/mole; this value should be very similar to that obtained in a solution 10 mM in sodium ions (28). Thus the experimental value is consistent with an A-form stem with 3 G-C base-pairs.

NMR spectroscopy

Initial experiments showed that the hairpin form of the oligonucleotide was favored by both higher temperature and very low salt concentration. Incomplete de-salting or addition of sodium chloride resulted in a second set of resonances in the NMR spectrum; lower temperature also favored the second set of resonances. These resonances were therefore assigned as duplex RNA. A salt concentration of 10 mM and moderate temperature (25°C) resulted in predominantly hairpin form of the RNA, with barely detectable amounts of duplex present in the NMR spectrum. This observation is consistent with the UV absorbance melting data.

Exchangeable protons. Figure 3 shows the down field region of the NMR spectrum taken in 95% H₂O at 22°C to observe the imino protons of rCGCU₃GCG. At ~ 13 ppm there are three resonances, corresponding to three hydrogen-bonded imino protons within G-C base-pairs (13, 29). Two of them are quite sharp, while the third is rather broad. This third resonance likely belongs to the terminal base pair, which is more accessible to solvent and thus is broadened by exchange at this relatively high temperature (29, 30). The broad peak at ~ 11 ppm corresponds to the three uridines in the loop. Non-hydrogen-bonded iminos resonate in this region; they are rather broad and indistinct, due to their moderately fast exchange rate. Thus these spectra further indicate that the molecule adopts a hairpin with three G-C base-pairs and a loop of three uridines.

Non-exchangeable protons. The non-exchangeable proton spectrum was assigned (Table 2) by the usual sequential-assignment method. Strong NOE's were observed between H2' and H3' protons and the 3'-neighboring H6 or H8 proton within the proposed stem region; they are consistent with an A-form

helix. In addition, a pair of NOE contacts from C₃H2' and C₃H3' to U₄H6, and single contacts from U₄H2' to U₅H6 and U₆H3' to G₇H8 were observed. No NOE's were observed between U₅ and U₆.

The assignments were greatly facilitated by the results of the DQ and HETCOR experiments. The lack of diagonal peaks in the DQ experiment allowed identification of resonances which were so close in chemical shift that they were obscured by diagonal peaks. Even so, G₇, C₈ and G₉ H5' and H5'' protons appeared to have nearly identical chemical shifts, resulting in no measurable J-coupling (31). The HETCOR experiment confirmed the identity of several questionable H3' and H5'/H5'' resonances, as well as verifying proper sequence assignments of resonances belonging to the three uridines.

Quantitative analysis of the 2-D spectra revealed some interesting features about the conformations of the nucleosides. NOE's showed that the bases all adopted the usual anti conformation about the glycosidic bond. Both NOESY and COSY data, however, indicated that only three of the nine ribose rings adopt a nearly pure N-type (e.g. C3'-endo) conformations. Coupling constants listed in Table 2 were used to determine relative populations of S- and N-type sugars from the H1'-H2' coupling, and the approximate pseudorotation phase angle (P°) from the additional H2'-H3' and H3'-H4' couplings (23, 32). The results in Table 2 show that C1, G2, and C8 are $\geq 85\%$ N-type conformation from the H1'-H2' coupling, and the additional couplings suggest a strong preference for a particular conformer close to classical C3'-endo. Likewise, U5 and U6 are $\geq 85\%$ S-type, and they display couplings suggesting a preference for a particular conformer as well, close to C2'-endo. The remaining nucleotides exhibit behavior unlike C2'-endo or C3'-endo sugars. The data support an equilibrium between a significant population of N and S conformations, with a preference for N-type in the case of C₃, and S-type for U₄ and G₇. The P° range consistent with each of the residues was converted to a range of δ value restraints in the RMD; the uncertainties in P° were reflected in appropriately broad δ ranges.

NOE's and coupling constants indicated that in most cases γ is in the range of gauche⁺. The exceptions are G₇ and C₈, which both have 8 Hz coupling between H4' and H5'/H5'. For these two nucleotides γ was allowed to be in the range of gauche⁻ or trans, -30° to -210° , for the RMD.

Phosphorus NMR. The ³¹P NMR spectrum displays a moderate range of chemical shifts between -2.9 and -4.2 ppm (Figure 4). Two resonances belonging to the U₅/U₆ phosphate, P₅, and the U₅/U₆ phosphate, P₆, are shifted noticeably down field from the rest, suggesting a trans conformation about the α or ζ P-O bond (30). This chemical shift dispersion is similar to that seen in the ³¹P spectra of Z-RNA (32), though more pronounced chemical shift dispersions were observed in hairpin structures containing unusual structural features (10, 13). For these two phosphates α and ζ are likely within the range of gauche⁻ to gauche⁺, -100° to 100° .

Strong couplings between phosphorus and H5'/H5'' protons suggest unusual dihedral angles for β about the O5'-C5' bonds. There is one large coupling of 16 Hz between P₄ and U₅, indicating gauche⁻ or gauche⁺ conformation. Smaller couplings ~ 9 Hz for P₅-U₆, P₆-G₇, and P₇-C₈ do not limit the dihedral angle to a specific range. Data presented by Lankhorst *et al.* (33) show that couplings of this magnitude could be consistent with deviation from the three low-energy rotamers, as well as a dynamic exchange between them; therefore no conclusion was

Table 2a. Chemical shifts* for the proton and phosphorus resonances.

	H6/H8	H5	H1'	H2'	H3'	H4'	H5'/H5''	P
C ₁	8.09	6.03	5.61	4.54	4.58	4.36	3.96/4.07	-3.82
G ₂	7.86	—	5.82	4.49	4.75	4.25	4.20/4.54	-4.12
C ₃	7.57	5.41	5.81	4.25	4.58	4.45	4.12/4.51	-3.55
U ₄	7.90	5.89	5.90	4.40	4.49	4.44	4.10 4.19	-3.94
U ₅	7.68	5.72	4.18	4.18	4.55	4.11	4.10/4.48	-3.17
U ₆	7.70	5.84	5.84	4.27	4.69	4.29	3.78/3.88	-3.52
G ₇	8.12	—	5.69	4.85	4.70	4.55	4.21/4.21	-2.92
C ₈	7.69	5.40	5.73	4.44	4.42	4.54	4.43/4.43	-3.89
G ₉	7.75	—	5.92	4.20	4.37	4.29	4.10/4.10	—

* Proton shifts are measured relative to internal TSP; phosphorus shifts are relative to external TMP. Only U₄ H5' and H5'' are stereospecifically assigned.

Table 2b. Scalar-coupling values (Hz) for proton-proton and proton-phosphorus couplings.*

	H1'-H2'	H2'-H3'	H3'-H4'	H4'-H5'/H5''	H3'-P	P-H5'/H5''	%N #
C ₁	2	6	9	3/5	7	n.a.	85
G ₂	1.5	4	9	<3/3	7	<4	95
C ₃	3.0	5.3	7	<3/<3	7	<4	70
U ₄	5.8	5.0	≤3	~4 ~3	9	<4	30
U ₅	7.6	5.0	≤3	<3/<3	7	16	5
U ₆	6.8	5.5	<3	~5/~3	~7	(~9/~4)	15
G ₇	6.1	5.6	~4	8	~8	9	25
C ₈	1.5	~6	~8	~8	8	9	95
G ₉	2.5	5	7	(Σ≤7)	—	<4	80

* Values given to 2 significant digits are estimated to be accurate to ±0.5 Hz; values given to 1 significant digit are accurate to ±1 Hz. Approximate values (e.g. ~8) are estimated to have an error of about 2–3 Hz. Values in parentheses are estimates with high error due to peak overlap or weak intensity. Couplings too small to yield a cross-peak are indicated as less than an estimated threshold value.

These values are the percent of N conformer (3'-endo) for each ribonucleoside estimated from the coupling constants.

made as to the conformation about these particular bonds. Non-existent P-H5'/H5'' cross-peaks for four of the stem nucleotides were interpreted as β-trans and assigned values of 180° ± 60°.

Energy minimization and restrained molecular dynamics

The ΔH° and imino proton chemical shifts indicated the formation of three C·G hydrogen-bonded base-pairs that stack upon each other in a regular A-form double-helix. Thus for the initial structural refinement six intermolecular H-N, N-N and N-O distance constraints were included to insure Watson–Crick base-pairing, as well as two guanine imino H1-H1 distance constraints to promote stacking. In these structures the base pairs tended to be buckled, suggesting that either insufficient data were available or the data lacked the precision to limit the conformation.

For the structures presented here additional constraints were used. Inter- and intraresidue distance constraints and dihedral angles derived from canonical A-form helical parameters (21) were introduced for the C₁-G₉ and G₂-C₈ base pairs. This resulted in a more standard A-form appearance of the first two base pairs, yet there was little effect on the structure of the loop and loop closing base-pair.

Comparison of input restraints versus distances and angles in the final six structures revealed some violations of the restraints. The distance violations were small, usually on the order of 0.1 to 0.2 Å outside the input range. Nearly all of them were in the first two base pairs, where the additional, tight (±0.5 Å) constraints had been introduced. The dihedral violations were also small, usually on the order of 5° outside the input range.

Like the distance violations, most of the dihedral violations were in the stem portion, where A-form constraints (±5–10°) were used.

Although there is considerable heterogeneity among the resulting structures, they appear to group into two different types. The two types, shown in Figure 5, are characterized by the position of U₄ relative to C₃, which in turn affects orientation of the loop with respect to the stem. In three structures (Fig. 5a) U₄ tends toward the minor groove and the loop crosses the top of the stem along the G₇-C₃ base-pair long axis. In the other structures (Fig. 5b) the loop crosses the closing base pair in a zig-zag fashion, with U₄ directed toward the major groove and U₆ toward the minor groove. There seems to be less variation between the loop conformations of structures in Figure 5a than in Figure 5b. The second set of structures shows significant buckling of the loop-closing base pair. Also, despite loose angle restraints in the loop, there were several small, but consistent, angle violations in the loop nucleotides of these three structures.

DISCUSSION

Synthesis of RNA

There has been a controversy as to whether or not DMT and THP groups are compatible in RNA synthesis. Christodoulou *et al.* (34) claim that loss of THP groups during the acidic conditions required for DMT removal cause migration of the phosphate from the 3' to the 2' position, thus compromising the authenticity of the RNA. On the other hand, Tanimura *et al.* (35)

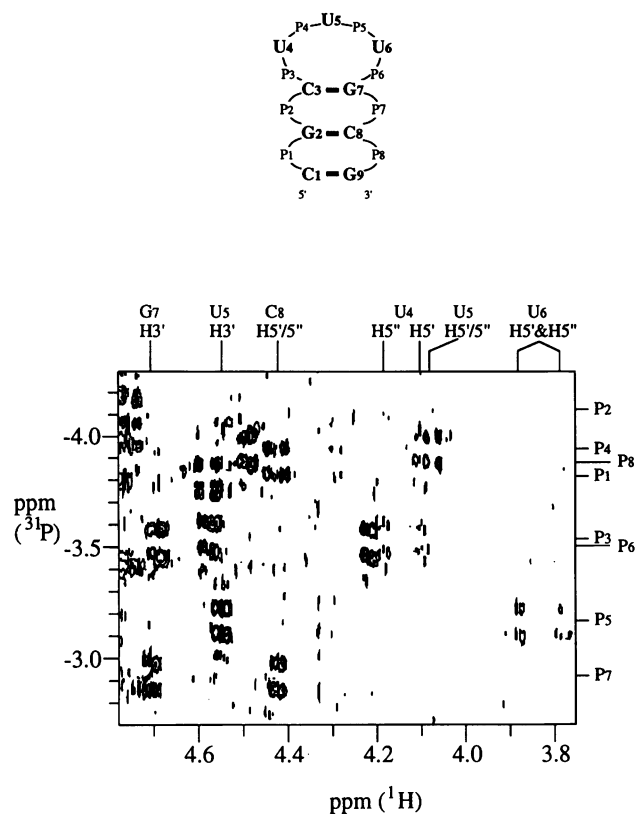


Figure 4. A two-dimensional ^1H - ^{31}P heteronuclear COSY (HETCOR) experiment with relevant assignments. The secondary structure of rCGCUUUGCG is shown above.

claim that loss of THP groups leads only to chain degradation in concentrated ammonia. This group synthesized oligonucleotides up to 13 residues in length using a combination of 5'-O-pixyl and -moxyl groups in conjunction with the 2'-O-THP. Kierzek *et al.* (14) were able to selectively remove the DMT group by treating the support-bound oligomer with 0.3% dichloroacetic acid (DCA) in CH_2Cl_2 for about 7 minutes. Typical deblocking conditions for DNA synthesis are 3% DCA or 3% trichloroacetic acid (TCA) in CH_2Cl_2 for about 1 minute.

The advantage of THP over silyl groups is two-fold. First, THP can be selectively introduced at the 2'-hydroxyl with the use of tetraisopropylidisiloxane-1,3-diyl group (15), simplifying the preparation of the proper monomer building block. The preparation of 2'-O-silyl monomers requires separating nearly equimolar amounts of 2'-O- and 3'-O-silyl nucleosides, and then identifying the proper isomer. In addition, one is faced with at least 50% loss of yield at that step. The second advantage is in the final deprotection and purification steps. The THP group is stable to the aqueous ammonia used for removal of acyl groups on the bases, while the stability of silyl groups under these conditions is questionable (36).

In the present study we found that deblocking with 0.3% DCA was satisfactory. Despite the long detritylation times, loss of THP groups was sufficiently small to afford a good yield of oligonucleotide. In addition no trace of improper 2'-5' linkages was detected in the RNase T2 digestions. The overall yield of purified oligonucleotide was about 22%, yielding a sufficient amount for NMR studies. However, the method could be

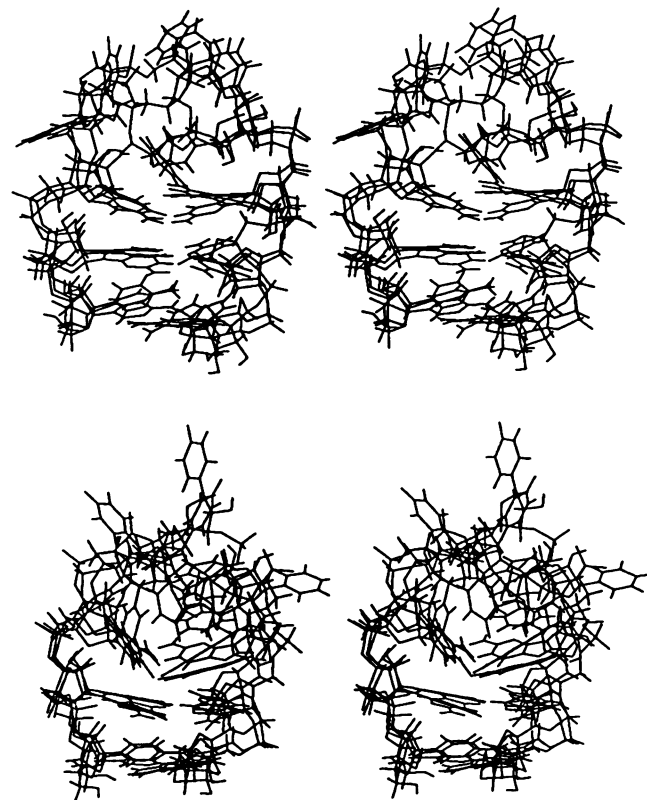


Figure 5. Stereoviews of rCGCUUUGCG of two different types of structural models obtained from restrained molecular dynamics. Two sets of three superimposed structures each are shown.

improved, and perhaps extended to the synthesis of longer oligonucleotides, by using more labile 5'-O-protecting groups on the pyrimidines (35) in order to shorten the exposure to acidic condition during the deblocking step.

Hairpin structure

Stem-loop structure from NMR. The NMR data are consistent with A-form geometry for the first two G·C base pairs in the stem. All of the measurable torsion angles and inter- and intrasidue NOE's are characteristic of A-form nucleotide conformation and stacking interactions. This is not the case for the C₃-G₇ pair, however. Although the NOE's are basically characteristic of A-form, several of the G₇ and one of the C₈ backbone torsion angles are not. G₇ exhibits significant S-type sugar pucker rather than the typical N-type; G₇ and C₈ both have γ values other than γ^+ ; and G₇ seems to have some trans-character in either the α or ζ torsion angle.

Qualitative analysis of the NMR indicates considerable organization within the loop, with a continuation of stacking from the stem into the loop. U₄ has NOE's from both H2' and H3' to the neighboring C₃H6 with magnitudes similar to that seen in the stem region. J-coupling and ^{31}P chemical shifts indicate that the backbone angles for U₄ are in the realm expected for A-form, with the exception being a preference for N-type sugar conformation. Weak NOE's are also observed between U₄H2'-U₅H6 and U₆H3'-G₇H6 protons, suggesting additional stacking on the 3' and 5'-sides of the stem, respectively. It is apparent that the stacking must be strongly distorted from A-

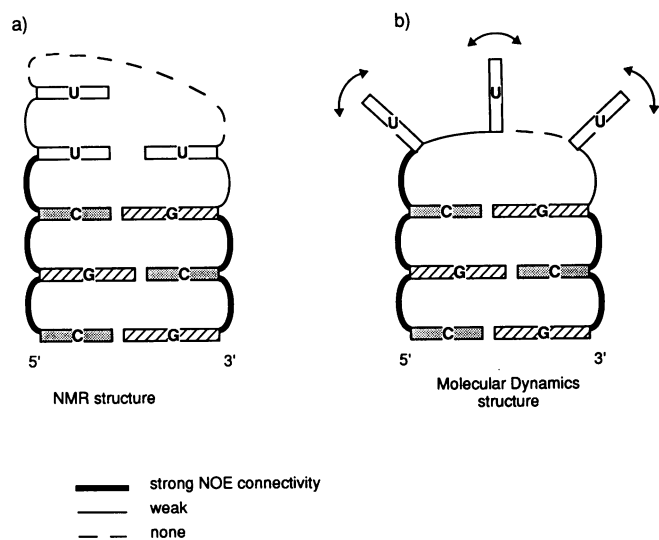


Figure 6. Schematic diagram of the hairpin structure implied by (a) NMR NOE connectivity data, which suggests some stacking within the loop, and (b) molecular dynamics structures, which do not indicate stacking of the loop nucleotides.

form in order to close the loop. Many of the loop torsion angles display unusual behavior, such as S-type sugars, a gauche β angle for U₅, and the possibility of trans α or z angles. There are no observable NOE's between U₅ and U₆, which is consistent with this being the point between the 3'- and 5'-sides of the stem where the strand changes direction. A schematic model interpreted from the NMR data is shown in Figure 6a.

The results are similar in some respects to those obtained with the DNA hairpin dCGCG(TTGTT)CGCG in the seminal work of Williamson and Boxer (30), who considered hairpin dynamics as well as structure. In this study the authors determined that although there is greater flexibility in the loop and loop-stem junction than in the stem region, there is still considerable organization within the loop. They found NOE connectivities continuing from the stem into the loop, and from these data concluded that significant stacking of the unpaired bases occurred on both sides of the stem. In addition, they found that one phosphorus had an unusual downfield chemical shift which they assigned to the (C₁₀)p(G₁₁) phosphate within the stem on the 3'-side. They suggested that the conformation and dynamics of this phosphate may be important in the interplay between maximizing stacking interactions and minimizing steric constraints within the loop.

Stem-loop structure from restrained molecular dynamics. The RMD structures indicate that the stem is able to deviate from A-form to accommodate the loop while retaining good stacking of all three base-pairs. Analysis of the backbone angles of the six structures showed strong preferences for non-A-form backbone angles in the 3'-side of the stem, but not on the 5'-side, as indicated by the NMR data. Unlike the structure implied by NMR, however, there was no evidence for stacking within the loop for the six RMD structures.

Individual torsion angles for each of the structures were analyzed to search for consistent trends or correlations. Despite the apparent heterogeneity of the structures, several notable features emerged. As expected, the first two base pairs exhibited basically A-form angles from C₁ to angle ζ of C₃, and from

angle δ of C₈ through G₉. On the other hand, the loop nucleotides displayed variability, with several torsion angles showing no preference for gauche⁺, gauche⁻ or trans conformations.

All of the structures deviate from A-form in a similar fashion on the 3'-side of the stem. There is a strong preference for trans γ in both G₇ and C₈, despite the fact that gauche⁻ and trans γ were allowed for these two nucleotides during the RMD. Values for the angles G₇ through C₈ were added for each structure as a measure of the twist in that portion of the helix. The result was an average value of $70^\circ \pm 56^\circ$, while the sum of the same angles in an A-form dinucleotide step is 186° . (Note that a single structure skews the experimental result; the average for five of the six sums is $90^\circ \pm 26^\circ$). This shows that the helix is significantly underwound at these two nucleotides relative to A-form.

The particular conformation of nucleotide G₇ appears to be important. In addition to predominantly S-type sugar pucker and trans γ , the guanine ring prefers a high-anti conformation ($\chi = 95 \pm 10^\circ$) about the glycosidic bond. These features coupled with the general unwinding of the helix allows closure of the small loop with minimal distortion of the base-pairing or base-stacking.

Analysis of angles for the sum of C₃ ϵ through U₄ ζ on the 5'-side of the stem reveals the underlying cause of the difference between the two families of structures. The average sum for the first three structures is $117^\circ \pm 20^\circ$, and for the second it is $207^\circ \pm 50^\circ$. The analogous sum for canonical A-form is 158° . Thus in one family the backbone is directed 45° toward, and in the other it is directed 45° away from, the center of the helix relative to the A-form geometry.

The results underscore the importance that the restraints imposed by the stereo chemistry of chemical bonds plays in the conformation of small hairpin loops. The distance between phosphates on opposite strands of an A-form helix is about 18 Å. One way the tri-loop can cross this distance is by changing from the usual N-type sugars to S-type; the distance between successive phosphates in C3'-endo sugars is 5.9 Å, while in C2'-endo it is 7.0 Å (37). Therefore the net distance that can be spanned by three nucleotides is about 21 Å, exceeding the cross-strand distance by 3 Å. However, bridging a double-helix requires not only physically spanning the distance, but also orienting the chemical bonds properly. Despite an increase in loop length afforded by the sugar conformations, there still seems to be insufficient conformational freedom for correct bond orientation. Thus the stem must deviate from A-form in order to allow the bonds to align properly. Changing the 3'-side of the stem, even as far as the second residue, apparently has a lesser impact on the global stem structure and stability than changing the 5'-side would have.

CONCLUSION

The RNA nonanucleotide rCGC(UUU)GCG forms a stable monomolecular hairpin in 10 mM phosphate solution. The nucleotides in the loop adopt S-type (C2'-endo) conformations, and there is distortion from A-form near the loop on the 3'-side of the stem. These distortions are not reflected in the value determined for ΔH° , which is close to the calculated value for an RNA helix with three C·G base-pairs.

The structure deduced directly from NMR differs from the structures calculated using restrained molecular dynamics. One explanation is that the loop is not constrained to a particular

conformation. The loop nucleotides exhibit NOE's that are characteristic of stacking, but they are weak, suggesting that the bases are only spending part of the time in a stacked conformation. Several of the proton-phosphorus coupling constants have values intermediate between those expected for any particular conformation, suggesting again that they are an average of the time spent by the atoms in two or more different torsion angles.

A second possibility is that the data are not sufficient to describe the structure adequately. The extensive set of quantitative interatom distances and torsion angles alone were insufficient to constrain the stem portion. One likely factor was the conservative interpretation of both the NOE's and J-coupling-constants. This approach reduced their impact on the structures produced. Another factor was a lack of NOE data for exchangeable protons. These deficiencies were accommodated by adding stacking and hydrogen-bonding restraints based on thermodynamic and imino-proton NMR data, resulting in the formation of a double-helical, A-type stem. No such structural assumptions could be made about the loop nucleotides, however, and there was also less interatom distance and torsion information for this portion of the molecule. We found no evidence of structural features such as unusual hydrogen-bonding or nucleotide conformations which would strongly influence the structure. There were only a few NOE's among the three uridines, and none between U₅ and U₆. Despite assignment of the ³¹P spectrum and measurement of J-coupling constants, values for β and ϵ could not be determined precisely, and in many cases not at all. Values for α and ζ were estimated from chemical shift, and not determined quantitatively. Thus the calculated loop structures were mainly dependent on the bond length, bond angle and van der Waals interaction energies used in XPLOR. The fact that a 'stacked' structure, such as shown in Fig. 6a, was not found in the molecular dynamics calculations is probably caused by a combination of too little data for the loop, too few structures calculated, and approximations present in the modeling process.

To develop a more rigorous model of molecules such as the one presented here, a greater variety of complementary NMR experiments will be necessary. NOE connectivities to exchangeable protons have been critical in structure determinations (10). ¹H-¹³C and ¹³C-³¹P NMR experiments would be helpful in the determination of more accurate torsion angles, and would also extend the number which are measurable. Also, ¹³C relaxation behavior could give insight to the dynamics of the structure, and thus determine whether the molecule displays significant conformational lability (38). There may in fact be a unique structure for the UUU tri-loop in solution which cannot be determined solely from the methods used in this study.

It is clear that the structure of RNA hairpin loops is an integral part of their function. The classic example is the anticodon loop of tRNA, in which single-stranded A-type stacking of the anticodon likely facilitates formation of a stable codon-anticodon mini-helix. Hairpins containing UNCG and GNRA tetra-loops have the ability to form an extra non-Watson-Crick base-pair which gives them extraordinary stability. Tri-loops are fairly common in a variety of naturally occurring RNAs indicating that they may possess some particularly useful characteristic. Although bacterial 16s rRNAs strongly favor the highly stable tetra-loops, the UUU tri-loop is the most common replacement (5). Wolters (39) showed that eukaryotic 16s-like RNAs display an even greater abundance of UUU loops, with nearly equal frequency as UUCG and GCAA families. He suggested that

evolutionary development of RNA-protein interaction might have reduced the need for highly stable loops. Loops of three adenosines or uridines are about 1 kcal/mole less stable than loops of four (28). Perhaps then the important structural feature of tri-loops is the accessibility of the loop nucleotides for recognition by proteins. A lack of defined structure, or an easily perturbed one, may be important in the biological role of some small hairpin loops.

ACKNOWLEDGEMENTS

We wish to thank Dr B. Wimberly for assistance in NMR spectroscopy and molecular dynamics simulations; Dr J. Wyatt for advice in handling and analyzing RNA; Drs J. Puglisi and G. Varani for useful discussions on RNA structure and NMR spectroscopy; Drs J. Jaeger and M. Chastain for computer assistance; and Drs R. Kierzek, S. Dreiker and D. Turner for information on RNA synthesis methods. We also wish to thank Dr Puglisi and Dr Wyatt for critical reading of this manuscript. This research was supported in part by National Institutes of Health Grant GM 10840, by the department of Energy, Office of Energy Research, Office of Health and Environmental Research under Grant DE-FG03-86ER60406 and through instrumentation grants from the Department of Energy, DE-FG05-86ER75281 and from the National Science Foundation, DMB 86-09305 and BBS 87-20134.

REFERENCES

- Chastain, M.A. and Tinoco, I., Jr. (1991) *Prog. Nucleic Acid Res. Mol. Biol.* **41**, 131-177
- Feng, S. and Holland, E.C. (1990) *Nature* **334**, 165
- Romaniuk, P.J., Lowary, P., Wu, H.-N., Stormo, G. and Uhlenbeck, O.C. (1987) *Biochemistry* **26**, 1563-1568
- Witherell, G.W. and Uhlenbeck, O.C. (1989) *Biochemistry* **28**, 71-76
- Woese, C.R., Winker, S. and Gutell, R.R. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8467-8471
- Haasnoot, C.A.G., Westerink, H.P., van der Marel, G.A. and van Boom, J.H. (1984) *J. Biomol. Struct. Dyn.* **2**, 345-360
- Happ, C.S., Happ, E., Nilges, M., Gronenborn, A.M. and Clore, G.M. (1988) *Biochemistry* **27**, 1735-1743
- Clore, G.M., Gronenborn, A.M., Piper, E.A., McLaughlin, L.W., Graeser, E. and van Boom, J.H. (1984) *Biochem. J.* **221**, 737-751
- Haasnoot, C.A.G., Hilbers, C.W., van der Marel, G.A., van Boom, J.H., Singh, U.C., Pattabiraman, N. and Kollman, P.A. (1986) *J. Biomol. Struct. Dyn.* **3** (5), 843-857
- Varani, G., Cheong, C. and Tinoco, I., Jr. (1991) *Biochemistry* **30**, 3280-3289
- Heus, H.A. and Pardi, A. (1991) *Science* **253**, 191-194
- Blommers, M.J.J., van de Ven, F.J.M., van der Marel, G.A., van Boom, J.H. and Hilbers, C.W. (1991) *Eur. J. Biochem.* **201**, 33-51
- Puglisi, J.D., Wyatt, J.R. and Tinoco, I., Jr. (1990) *Biochemistry* **29**, 4215-4226
- Kierzek, R., Caruthers, M.H., Longfellow, C.E., Swinton, D., Turner, D.H. and Freier, S.M. (1986) *Biochemistry* **25**, 7840-7846
- Markiewicz, W.T., Biala, E. and Kierzek, R. (1984) *Bulletin Pol. Acad. Sci.* **32**, 433-451
- Atkinson, T. and Smith, M. (1984) In Gait, M.J. (ed.) *Oligonucleotide synthesis—a practical approach* (IRL Press, Oxford, England)
- Sinha, N., Biernat, J., McManus, J. and Köster, H. (1984) *Nucleic Acids Res.* **12**, 4539-4557
- Eadie, J.S., McBride, L.J., Efcavitch, J.W., Hoff, L.B. and Cathcart, R. *Analytical Biochemistry* **165** (1987) 442-447
- Puglisi, J.D. and Tinoco, I., Jr. (1989) *Methods Enzymol.* **180**, 304-325
- Brünger, A.T. (1990) XPLOR-Version 2.1 User Manual. Yale University, New Haven, CT
- Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*. John Wiley & Sons, New York
- Altona, C. (1982) *Recl. Trav. Chim. Pays. Bas.* **101**, 413-433

23. de Leeuw, F.A.A.M. and Altona, C. (1982) *J. Chem. Soc. Perkin Trans. 2*, 375–384.
24. Kraulis, P.J., Clore, G.M., Nilges, M., Jones, T.A., Pettersson, G., Knowles, J. and Gronenborn, A.M. (1989) *Biochemistry* **28**, 7241–7257
25. Wimberly, B. (1992) *Ph.D. Dissertation, U. C. Berkeley Department of Chemistry*
26. Wang, Y.Y., Lyttle, M.H. and Borer, P.N. (1990) *Nucleic Acids Res.* **18** (11) 3347–3352
27. Freier, S.M., Kierzek, R., Jaeger, J.J., Sugimoto, N., Caruthers, M.H., Neilson, T. and Turner, D.H. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 9373–9377
28. Groebe, D.R. and Uhlenbeck, O.C. (1988) *Nucleic Acids Res.* **16** (24), 11725–11767
29. Haasnoot, C.A.G., de Bruin, S.H., Berendsen, R.G., Janssen, H.G.J.M., Binnendijk, T.J.J., van der Marel, G.A. and van Boom, J.H. (1983) *J. Biomol. Struct. Dyn.* **1**, 115–129
30. Williamson, J.R. and Boxer, S.G. (1989) *Biochemistry* **28**, 2819–2831
31. Becker, E.D. (1980) *High Resolution NMR*. Academic Press, New York, p. 92
32. Davis, P.W., Adamiak, R.W. and Tinoco, I., Jr. (1990) *Biopolymers* **29**, 109–122
33. Lankhorst, P.P., Haasnoot, C.A.G., Erkelens, C. and Altona, C. (1984) *J. Biomol. Struct. Dyn.* **1**, 1387–1405
34. Christodoulou, C., Sudhir, A., Gait, M.J. (1986) *Tetrahedron Lett.* **27**, 1986
35. Tanimura, H., Fukazawa, T., Sekine, M., Hata, T., Efcavitch, J.W. and Zon, G. (1988) *Tett. Let.* **29**, 577–578
36. Scaringe, S.A., Francklyn, C. and Usman, N. (1990) *Nucleic Acids Res.* **18**, 5433–5441
37. Saenger, W. (1984) *Principles of Nucleic Acid Structure*. Springer-Verlag, New York
38. Williamson, J.R. and Boxer, S.G. (1988) *Nucleic Acids Res.* **16**, 1529–1540
39. Wolters, J. (1992) *Nucleic Acids Res.* **20**, 1843–1850