Journal of Biomolecular NMR, **12:** 423–433, 1998.
KLUWER/ESCOM

On the NMR structure determination of a 44n RNA pseudoknot: Assignment strategies and derivation of torsion angle restraints

Michaël H. Kolk[a], Sybren S. Wijmenga[b], Hans A. Heus[a] and Cornelis W. Hilbers[a,*]
[a]*NSR Center for Molecular Structure, Design and Synthesis, Laboratory of Biophysical Chemistry, University of Nijmegen, Toernooiveld, 6525 ED Nijmegen, The Netherlands;* [b]*Present address: Department of Medical Biochemistry and Biophysics, University of Umeå, S901 87 Sweden*

Received 11 May 1998; Accepted 16 July 1998

Key words: assignment, pseudoknot, RNA, torsion angles

Abstract

The complete T- and pseudoknotted acceptor arm of the tRNA-like structure of turnip yellow mosaic virus (TYMV) genomic RNA has been studied by NMR spectroscopy. Resonance assignment and the gathering of restraints of the 44-mer are impeded by spectral complexity as well as by line broadening. The latter is caused by local dynamical effects in the pseudoknot domain in the molecule. These specific problems could be solved by using different field strengths and selectively $^{13}C/^{15}N$ labeled samples. Experiments for assigning the sugar spin systems were adjusted to satisfy the requirements of this system. Furthermore, the quality of the structure could be improved by determining the backbone torsion angles β, γ and ϵ, using new approaches that were tailored for use in large RNA molecules.

Introduction

RNA pseudoknots have been identified in many classes of RNA and have been shown to play important roles in replication initiation, frame-shifting, translational control and core formation in catalytic RNAs (Ten Dam et al., 1992; Pleij, 1994). This distinct RNA folding motif typically involves base-paring of residues in a hairpin loop with single-stranded nucleotides outside that loop. It was first discovered at the 3′-terminus of the genomic RNA of turnip yellow mosaic virus (TYMV) (Rietveld et al., 1982). Like in many other plant viruses containing positive strand RNA genomes, this segment contains a tRNA-like structure which is recognized by the viral replicase as well as by several tRNA-specific enzymes (Hall, 1979; Mans et al., 1991). On the basis of these functional properties and biochemical data obtained for the secondary structure, it was proposed that a pseudoknot plays a key role in folding the domain into an L-shaped molecule that closely resembles the overall shape of tRNA (Florentz et al., 1982; Rietveld et al., 1982).

Recently, we have determined the three-dimensional structure of the 44 nucleotides long pseudoknotted arm of the tRNA-like structure in TYMV (denoted as pk44, see Figure 1) by means of high resolution NMR spectroscopy (Kolk et al., 1998). The structure, which confirms the proposed secondary structure and provides a rationale for all available biochemical data, is highly resolved in all stem and loop regions, allowing a detailed description of a number of novel structural elements responsible for the stability and functionality of the pseudoknot. It also represents the first high resolution structure of an RNA pseudoknot determined to date, which allows for a considerable advance of our understanding of pseudoknots in general.

The system under study consists of three stem regions and comprises 44 nucleotides, which makes it a rather large RNA molecule to be investigated in detail by NMR. Moreover, it contains several regions of local mobility leading to line broadening and unfavorable spectral overlap. Although modern assignment strategies and heteronuclear techniques have greatly

*To whom correspondence should be addressed.

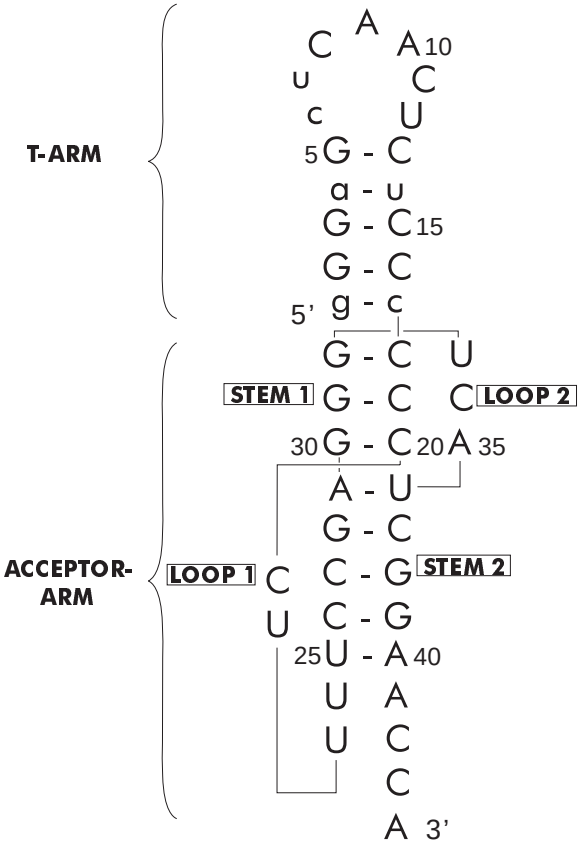

Figure 1. Secondary structure of the T-arm and pseudoknotted acceptor-stem of the tRNA-like structure of TYMV RNA. Non-wild-type residues are given in lowercase.

extended the size of RNA molecules apt for structural elucidation by NMR (Nikonowicz and Pardi, 1993; Dieckmann and Feigon, 1997), only part of these are applicable to the present system. In particular for resonances relating to the sugar moieties many of the conventional assignment procedures fell short due to unsurmountable problems of overlap in all spectral dimensions.

In this paper we present an overview of the approaches that we have undertaken to tackle these specific problems. They include the use of selective labeling of the molecule, employing different field strengths for the NMR experiments and the use of tailor-made experiments to obtain connectivities in the sugar regions. Furthermore, we describe several new applications of 2D and 3D heteronuclear NMR experiments that can be used to determine torsion angle restraints, at least in a qualitative way, for β, γ and ϵ, which otherwise have to be left mostly unconstrained in structure calculations of larger RNA molecules. All of these approaches are applicable to other systems of similar size and open the way towards the structure determination of even larger RNA molecules.

Materials and methods

Sample preparation

Four different RNA samples were prepared as described earlier (Kolk et al., 1998): an unlabeled sample, a sample that was uniformly labeled with $^{13}C/^{15}N$, and two samples that were selectively labeled at either the adenine or the uridine positions. Structure probing experiments revealed that the stability of the pseudoknot structure depends on the presence of Mg^{2+} (Van Belkum et al., 1989). All purified RNA samples were therefore dialyzed in a 10 mM $MgCl_2$ solution at pH 6.7 and subsequently concentrated with a Centricon microconcentrator to final RNA concentrations of 1 to 2.5 mM.

Spectroscopy

NMR experiments were performed on Varian Unity+ 500 and 750, and on Bruker AM 400 and AMX/DRX 600 spectrometers. Spectra were recorded at 15 °C for experiments relating to exchangeable proton resonances, and at 30 °C for all other experiments. No significant changes in chemical shifts are observed throughout the 5° to 40 °C range. Resonance assignment and restraint collection were established from experiments as described (Kolk et al., 1998). Additional experiments are discussed in the Results section.

Structure calculations

All structure calculations were performed on a Silicon Graphics O2 workstation using X-PLOR version 3.851 (Brünger, 1992). They include a torsion angle dynamics protocol (TAD, Stein et al., 1997) followed by rounds of simulated annealing (SA) and restrained conjugate gradient energy minimization (Kolk et al., 1998). Coordinates of the final ensemble of structures have been deposited in the Protein Data Bank under accession code 1A60.

Results

Two major obstacles were encountered in the spectral assignment of the pk44 molecule. The first has to do with the spectral complexity caused by the 44

nucleotides present in the system. The second relates to the line broadening that is observed for many residues in the pseudoknotted region (i.e. residues 18–40). Resonance identification for residues in the T-arm (residues 1–17) and the 3′-terminal tail (residues 41–44) only suffers from the first difficulty, as all nucleotides in this domain give rise to acceptable line-widths. Hence, standard methods (Nikonowicz and Pardi, 1993; Dieckmann and Feigon, 1997) provided a nearly complete assignment for all of these nucleotides.

For other parts of the molecule the spectral assignment was more troublesome, and it took several adjustments of existing NMR experiments to come to a full assignment. The severe overlap of the sugar carbons prompted for the development of HCCH experiments which combine the use of constant-time (CT) evolution in the ^{13}C-dimension with a high transfer efficiency (*vide infra*). Also, a 3D (^{31}P, ^{13}C, ^{1}H) correlation experiment was adapted so as to achieve optimal sensitivity for the present system. Finally, the structure could be improved appreciably by collecting torsion angle restraints using new approaches in experiments based on ^{31}P-spin-echo and on (^{1}H, ^{1}H) TOCSY transfer efficiencies. All of these experiments and their interpretation are discussed below.

Resonance assignments

T-arm and 3′-ACCA tail

Nearly complete assignment was possible for these regions using homonuclear 2D NOESY and ^{13}C-^{1}H 3D NOESY-HMQC/HMQC-NOESY (Ikura et al., 1990) experiments, all performed at 750 MHz. Sequential connectivities were established without discontinuities for all of these nucleotides following a standard anomeric to aromatic proton walk (Wijmenga et al., 1993). The only distinctive deviation from the A-helical pattern of NOE connectivities occurred around U7, which is involved in a well described U-turn, as found in the tRNA anticodon loop (Quigley and Rich, 1976). Confirmation of part of the assignments was achieved by the identification of A- and U-related resonances of the selectively labeled samples. Furthermore, the U14 imino proton could be connected to its H6 resonance via a 2D HNCCCH experiment (Simorre et al., 1995), and all adenine H2 and H8 resonances in the molecule could be interconnected from a 3D HCCH-TOCSY experiment (Legault et al., 1994).

Pseudoknotted region

The relative straightforwardness of the assignment is interrupted when entering the pseudoknot domain of the molecule. The advancement of the sequential walk is hampered by severe overlap of proton and carbon resonances, in particular for residues in stem 2. Moreover, many of the resonances are broadened due to local mobility around this part of the molecule. Evidently, the effects of these motions are not confined to the dynamic residues, but also determine the lineshapes of resonances belonging to nearby residues. This phenomenon accounts for the line broadening observed in all stem and loop regions of the pseudoknot domain, as is rationalized further in the Discussion section.

The effect of conformational averaging is most pronounced for residues at the interface of stem 1 and 2, as has been illustrated earlier (Kolk et al., 1998): residues A29 through G31, C20 and U36 all have resonances that are severely broadened at 750 MHz. Most of the assignments as well as the integration of the NOE cross-peaks belonging to these nucleotides were therefore carried out using spectra recorded at 400 MHz, in which the resonances narrow to acceptable line-widths. Line broadening of other residues in the pseudoknot domain is less prominent, but nevertheless contributes unfavorably to the assignment efficiency in ill-dispersed regions of the spectra. Therefore, the use of selectively labeled samples proved indispensable, in particular for the assignment of loop 1, which is composed primarily of uridines. Residues U22 through U25 could only be assigned from doubly X-filtered 2D NOESY (Folmer et al., 1995), 3D NOESY-HMQC and 3D PCH experiments performed on a ^{13}C/^{15}N U-enriched sample. The PCH experiment, which yielded P-C4′-H4′ connectivities, proceeded roughly analogous to the P(CC)H-TOCSY experiment (Wijmenga et al., 1995), but included gradient enhancement (Kay et al., 1992) and concatenated CT and INEPT steps for the ^{31}P and ^{13}C dimensions. Subsequent transfer of the C4′ to the C1′ and H1′ resonances, as in the P(CC)H-TOCSY, unfavorably affected the sensitivity of the experiment. Therefore, the ^{13}C-^{13}C DIPSI period was omitted and the H4′ and H1′ resonances were interconnected in a separate 2D CT (H)C(C)H-TOCSY (Bax et al., 1990), in which ^{1}H$\leftrightarrow$^{13}C coherence transfer was achieved using hetero-TOCSY steps.

Assignment of G- and C-residues in the pseudoknot domain, which are present almost exclusively in the stem regions and show considerable resonance

overlap, was done after all other resonances had been identified. Close comparison of homonuclear 2D NOESY and TOCSY spectra with ^{13}C-edited 3D NOESY and TOCSY data on the uniformly labeled pk44 yielded a complete sequential connectivity pattern which included all aromatic and H1′ resonances, and part of the H2′ and H3′ resonances. Subsequently, part of the remaining sugar spin systems could be assigned from 3D and 2D (CT) HCCH-TOCSY experiments. In many cases the latter approach failed, however, because of the poor dispersion in the ^{13}C dimension, which is of particular nuisance in 3D experiments with a limited number of data points for the indirect dimensions. Ideally, a 3D HCCH TOCSY experiment would be preferred with a constant-time ^{13}C-evolution. However, the lengthy CT delay of 25 ms in addition to the ^{13}C-^{13}C DIPSI period (about 20 ms) makes this experiment very insensitive and essentially useless for this system, even when gradient enhancement is used for the ^{13}C-dimension.

Recently, the use of constant-time evoluted HMQC experiments was reported during the preparation of this manuscript (Marino et al., 1997). Owing to the favorable relaxation properties of multiple-quantum coherences, this approach can significantly enhance sensitivity in some heteronuclear experiments, provided that proton-proton coupling, which then becomes active during the CT delay, does not cancel this effect. An additional caveat concerning this method is that, for dynamical molecules such as pk44, the effect of proton exchange-broadening may exceed that of relaxation, since the former affects multiple-quantum $H_{x,y}C_{x,y}$ coherences, but not $H_zC_{x,y}$ antiphase coherences as in a 'regular' CT-HSQC experiment. It therefore remains to be established whether the application of this method in the afore-mentioned 3D CT HCCH-TOCSY could make this experiment work for large RNA molecules.

Here, we have utilized the high magnetization transfer efficiency of HCCH-COSY type of experiments (Pardi and Nikonowicz, 1992), which relieves much of the sensitivity problem. For the present purposes, the 3D HCCH correlation experiment (Kay et al., 1990) was carried out with CT ^{13}C-evolution (in a way similar to that described in Ikura et al. (1991), see Figure 2). Coherence transfer to H1′ proceeds from H2′, or can be relayed from H3′ with the incorporation of a relay step (Figure A). Carbon evolution can be executed at the C2′/C3′ or C1′ antiphase coherence level, the latter approach allowing for gradient sensitivity enhancement in the carbon dimension. Using the experiment of Figure A, the H3′ resonances could be identified starting from the well-dispersed H1′-C1′ crosspeaks, after which the exact C3′ chemical shifts could be determined from a separate 2D CT (H)C(C)H-TOCSY experiment. H2′ resonances were assigned from a similar experiment in which the relay step was omitted and ^{13}C frequency labeling was performed prior to coherence transfer to the C1′s, which connects the desired H2′ and C2′ chemical shifts to the previously assigned H1′s. It has the concomitant

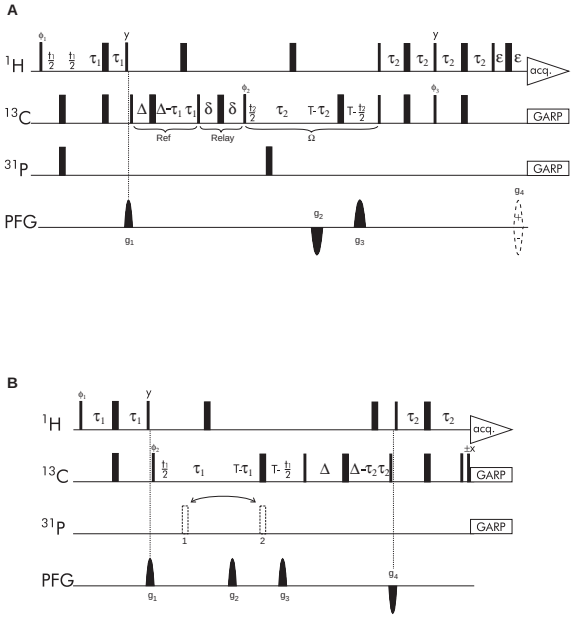

Figure 2. Pulse sequences for the 3D CT-HC(C)H correlation (A) and the ^{31}P spin-echo difference CT-HCCH experiment (B). Narrow and wide bars denote 90° and 180° pulses, respectively. All pulses are along the x-axis, except when indicated otherwise. ^{13}C offsets were at 77.2 ppm. A. The sequence as depicted correlates H3′, C1′ and H1′. The 'Relay' step can be omitted so as to establish vicinal connectivities. Proton refocusing element 'Ref' and frequency labeling period Ω may be interchanged (see text), provided that the Kay-enhancement sequence (Kay et al., 1992) is replaced by a refocusing INEPT step. This corresponds to the ^{31}P-decoupled 3D version of experiment B. $\phi_1 = x, x, -x, -x$, $\phi_2 = x, -x$, acq. $= x, -x, -x, x$. For each increment ϕ_3, which is along the y-axis, was inverted simultaneously with the sign of the final gradient in order to achieve N- and P-type coherence selection. Transfer delays: $\tau_1 = 1.64$ ms, $\Delta = 2.9$ ms, $\delta = 5.5$ ms, $\tau_2 = 1.48$ ms, T = 16.5 ms and $\varepsilon = 1$ ms. Gradients g_1 through g_4, all 500 μs long, had values of 48, 48, 48 and 24.6 G/cm, respectively. B. Two experiments were recorded in an interleaved manner with the ^{31}P 180° pulse at position (1) or (2), corresponding to the decoupled and coupled spectrum, respectively. Phase cycling was identical to the previous sequence. The transfer delays, optimized for H2′→ C2′→ C1′ → H1′, were: $\tau_1 = 1.65$ ms, T = 14.5 ms, $\Delta = 5.8$ ms and $\tau_2 = 1.48$ ms. Gradients g_1 through g_4 all had durations of 500 μs at 48 G/cm. Quadrature detection in t_1 was achieved using the States-TPPI method (Marion et al., 1989).

advantage of identifying all H1′-C2′ crosspeaks that are of use in determining the torsion angle ϵ, which is described in the following section.

A similar approach was employed using a 3D CT HCCH correlation experiment for the assignment of the H5′ and H5″ resonances, starting from H4′-C4′ crosspeaks. The latter are remarkably well dispersed and could therefore largely be identified from intra-residue NOEs to H8/H6/H1′ resonances in 3D NOESY-HMQC/HMQC-NOESY experiments. The assignments of the H4′ and C4′ resonances were verified in 2D and 3D HCCH-TOCSY experiments.

Torsion angle restraints

Large difficulties arise when trying to assess the backbone torsion angles in large RNA molecules. Conformationally relevant J-couplings can be accurately determined from direct measurement of resonance splittings in E.COSY or P.E.COSY type of experiments (Wijmenga et al., 1993), but this method relies heavily on the dispersion and line-widths of the resonances involved. For molecules larger than about 25 nucleotides, both resonance overlap and unfavorable line shapes hamper the determination of J-couplings using these methods. For the pk44 molecule the measurement of resonance splittings is altogether impossible and other approaches had to be taken for the determination of the sugar pucker and the backbone torsion angles β, γ and ϵ.

Sugar puckers

For the sugar pucker conformation, the H1′ to H2′ J-coupling serves as the most direct determinant (Wijmenga et al., 1993). All ^{3}J(H1′,H2′)s could be evaluated in a qualitative way from ^{1}H-^{1}H TOCSY-type of experiments (2D TOCSY, TOCSY-HMQC, TOCSY-CT-HSQC). Very weak or strong TOCSY H1′-H2′ crosspeaks correspond to pure N- or S-type conformations, respectively. Intermediate intensities indicate an equilibrium between these states which can not be entered as a restraint in any conventional structure calculation protocol. Time-averaged protocols (Yao et al., 1997) are better equipped for this, and should be the method of choice if the overall structure depends significantly on the sugar conformation of an individual residue. Such is clearly not the case for pk44, as was ascertained by executing additional simulated annealing (SA) runs for each residue with intermediate TOCSY H1′-H2′ crosspeaks in which the respective sugar puckers were set to either N- or S-type. This did not noticeably alter the overall conformation of the molecule.

The angles β and ϵ

Backbone torsion angles β and ϵ can in principle be derived from both ^{3}J(H,P) and ^{3}J(C,P) coupling constants. Experiments yielding proton-phosphorus scalar couplings have been amply used to assess these angles before the current age of isotopically labeled RNA (Varani and Tinoco, 1991; Wijmenga et al., 1993; Blommers et al., 1994). For reasons explained earlier, this is an impractical method for large RNA molecules. Introduction of ^{13}C-nuclei in the ribose moieties can be used to disperse the multiplets along the carbon axis, as is discussed in Schwalbe et al. (1994) and Varani et al. (1995). Moreover, ^{3}J(C,P) values can now be obtained from P-FIDS-CT-HSQC (Schwalbe et al., 1994) or ^{31}P spin-echo difference CT-HSQC (Vuister et al., 1993; Legault et al., 1995). Practically, the latter method is preferred for larger molecules, since it does not rely on proton or carbon line-widths for an accurate measurement of the J-couplings, as long as the crosspeaks are reasonably well resolved. Evidently, it only yields the ^{3}J(C,P) coupling constants, and a somewhat higher precision in the determination of β and ϵ may be obtained for smaller molecules using the P-FIDS-CT-HSQC experiment, which produces both (H,P) and (C,P) scalar couplings.

As has been described by Legault et al. (1995), the spin-echo difference CT-HSQC allows for an accurate measurement of ^{3}J(C2′,P) and ^{3}J(C4′,P). Experiments were recorded with constant time delays equaling $1/J_{CC}$ and $2/J_{CC}$, the latter proving less informative due to a poor signal to noise ratio. The former experiment typically produced the coupling constants with an accuracy of about 0.5 Hz. There are, however, two problems concerning this method. Firstly, the measured value for ^{3}J(C4′,P) is in fact the combined effect of the $^{31}P_{3'}$ and the $^{31}P_{5'}$ scalar coupling to each C4′. Secondly, the determination of ^{3}J(C2′,P) is hampered by the severe overlap of the H2′-C2′ crosspeaks. As a consequence, only a small fraction of these couplings could be evaluated in a regular spin-echo difference CT-HSQC experiment.

The latter problem can be solved via an extension to the CT-HSQC experiment. ^{31}P-modulated antiphase C2′ coherence can be transferred to C1′ in a spin echo difference CT-HCCH correlation experiment, as is depicted in Figure 2B. Analogously to the procedure in the spin-echo CT-HSQC experiment, ^{31}P-modulation

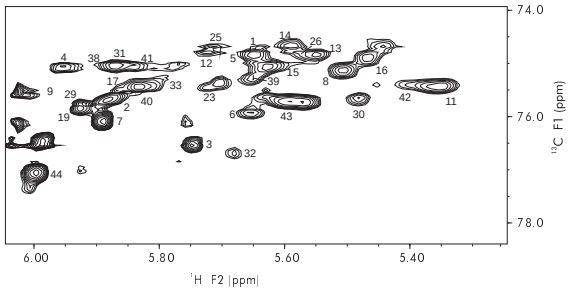

Figure 3. Contour plot of the uncoupled ^{31}P spin-echo difference CT-HCCH spectrum of uniformly ^{13}C/^{15}N-labeled pk44, recorded with the pulse sequence given in Figure 2B. Water suppression was achieved by low-power saturation of the residual HDO peak during the relaxation delay of 1 s. H1′-C2′ crosspeaks are labeled by their corresponding residue number. Unlabeled peaks are caused by degradation products. The spectrum was acquired at 600 MHz with spectral widths of 5285 and 6000 Hz for the ω_1 and ω_2 dimension, respectively. For each of the 148 increments 240 scans were recorded.

can then be measured from H1′-C2′ crosspeaks which are far better resolved than the H2′-C2′s (Figure 3). The concatenation of constant time and refocusing delays makes this experiment sensitive enough to be also recorded as a 3D experiment, in which the H1′-C2′ cross peaks may be dispersed along the H2′ or C1′ dimension.

Using these different spin-echo difference methods on uniformly and selectively labeled samples, the ^{3}J(C2′,P) coupling constants could be determined for most residues, with values ranging from 0 to 5 Hz. In principle, the non-bijective nature of the Karplus curve (Wijmenga et al., 1993) prevents a definitive determination of the corresponding ϵ-angles from these data alone. Fortunately, ϵ has a rather restricted habitat within the conformational wheel, and is in practice confined to the 180–300° range (Saenger, 1984). This leaves a nearly one-to-one correlation between the ^{3}J(C2′,P) value and the torsion angle ϵ, which could therefore be established for most of the residues in the molecule (see Table 1).

The evaluation of ϵ offers a possibility to circumvent the ambiguity problem concerning the determination of the ^{3}J(C4′,P) coupling constant. Knowing ϵ, the value for ^{3}J(C4′,P$_{3'}$) can be calculated from the corresponding Karplus equation (Wijmenga et al., 1993). Subsequently, a correction factor CF$_{3'}$ can be computed:

$$CF_{3'} = 1/\cos(\pi \times {}^{3}J_{C4'P_{3'}} \times T)$$

Table 1. Experimental backbone torsion angle restraints used in the structure calculation of the TYMV pseudoknot. The values for β, γ, the sugar pucker P and ϵ were determined as described in the text. The angles α and ζ were set to exclude the *trans* conformation whenever the ^{31}P chemical shift resided in the regular A-helical region

	α	β	γ	P	ϵ	ζ
G1			54±30			0±120
G2	0±120	180±30	54±30	N	230±20	0±120
G3	0±120		54±30	N	225±20	0±120
A4	0±120	180±30	54±30	N	215±20	0±120
G5	0±120		54±30	N	220±20	0±120
C6	0±120	220±30	54±30	N	215±20	0±120
U7	0±120	180±30	54±30	N	240±20	
C8		215±30	54±30		215±20	0±120
A9	0±120	180±30	54±30	N	220±20	0±120
A10	0±120	180±30	54±30	N	235±20	0±120
C11	0±120	180±40	54±30	N	205±20	0±120
U12	0±120	150±30	54±30	N	235±20	0±120
C13	0±120	165±30	54±30	N	220±20	0±120
U14	0±120	180±20	54±30	N	220±20	0±120
C15	0±120		54±30	N	205±20	0±120
C16	0±120		54±30	N	205±20	0±120
C17	0±120		54±30	N		0±120
C18	0±120		54±30	N		0±120
C19	0±120		54±30	N		0±120
C20	0±120		54±30	N		0±120
C21	0±120			N		0±120
U22	0±120	230±30	180±30	S	230±30	0±120
U23	0±120	180±30	54±30	N	255±20	0±120
U24	0±120	145±30	54±30	N	240±20	0±120
U25	0±120	180±30	54±30	N	230±20	0±120
C26	0±120		54±30	N		0±120
C27	0±120			N		0±120
G28	0±120		54±30	N	220±30	0±120
A29	0±120	150±30	54±30	N	200±20	0±120
G30	0±120		54±30	N	205±20	0±120
G31	0±120		54±30	N	205±20	0±120
G32	0±120		54±30	N	205±20	0±120
U33	0±120	180±30	54±30	S	245±20	0±120
C34	0±120			N	205±20	0±120
A35	0±120	150±30	54±30	S	235±20	0±120
U36	0±120	120±30	54±30	N	230±20	0±120
C37	0±120		54±30	N		0±120
G38	0±120		54±30	N		0±120
G39	0±120		54±30	N		0±120
A40	0±120	180±30	54±30	N	210±20	0±120
A41	0±120	180±30	54±30	N	240±20	0±120
C42	0±120	180±30	54±30	N	225±20	0±120
C43	0±120	180±30	54±30	N	245±20	
A44						

Table 2. Chemical shifts of proton and carbon resonances relative to TSP in the TYMV pseudoknot

	H8/H6	H2/H5	H1′	H2′	H3′	H4′	H5′/H5′′	im/am
G1	8.03	n.a.[a]	5.65	4.50	4.12	4.23	4.07/3.99	12.56
G2	7.49	n.a.	5.88	4.60	4.50	4.31	4.20/4.15	12.64
G3	7.24	n.a.	5.77	4.50	4.09	4.19	4.10/4.08	12.22
A4	7.68	7.52	5.96	4.57	4.53	4.54	4.28/4.11	
G5	7.00	n.a.	5.63	4.43	4.40	4.18	4.18/4.07	13.33
C6	7.48	5.11	5.64	4.17	4.42	4.32	4.32/4.17	
U7	7.89	5.82	5.88	4.35	4.65	4.38	4.14/4.13	
C8	7.74	5.76	5.50	4.43	4.30	4.29	4.08/3.98	
A9	8.05	7.86	6.02	4.67	4.87	4.39	4.30/4.26	
A10	7.66	8.04	5.36	4.35	4.52	4.49	4.35/4.33	
C11	7.29	5.43	5.36	4.14	4.33	4.35	4.11/4.06	
U12	7.78	5.70	5.74	4.38	4.65	4.52	4.17/4.10	
C13	8.01	5.82	5.56	4.37	4.65	4.45	4.18/4.12	8.44/7.15
U14	8.13	5.50	5.59	4.46	4.65	4.48	4.59/4.18	14.32
C15	8.01	5.65	5.64	4.29	4.55	4.44	4.18/4.11	8.39/6.93
C16	7.88	5.45	5.47	4.44	4.47	4.43	4.12/4.09	8.49/6.85
C17	7.80	5.45	5.85	4.55	4.45	4.42	4.20/4.08	8.64/6.84
C18	7.59	5.84	5.59	4.93	4.38	4.39	4.16/4.06	8.84/6.85
C19	8.02	5.92	5.92	4.37	4.50	4.37	4.17/4.13	8.01/6.97
C20	7.43	5.23	5.52	4.11	4.43	4.37	4.19/4.10	8.68/6.84
C21	7.86	5.61	5.81	3.99	4.22	4.34[b]	4.12/4.09[b]	
U22	7.58	5.54	5.92	5.05	4.95	4.44	4.23/4.14	
U23	7.95	5.79	5.74	4.46	4.62	4.53	4.37/4.16	
U24	8.04	5.95	6.03	4.44	4.55	4.58	4.31/4.25	
U25	7.76	5.79	5.69	4.55	4.71	4.49	4.33/4.16	13.93
C26	7.96	5.57	5.56	4.46	4.30	4.41	4.12/4.12[b]	8.29/6.86
C27	7.77	5.91	5.93	4.43[b]	4.21	4.36[b]	4.10/4.10[b]	8.33/6.92
G28	7.70	n.a.	5.64	4.50	4.50	4.43	4.12/4.12[b]	11.99
A29	8.23	7.64	5.97	4.66	4.46	4.45	4.44/4.44	
G30	6.92	n.a.	5.45	4.33	3.99	4.42	4.38/4.32	12.60
G31	6.83	n.a.	5.85	4.41	4.55	4.34	4.03/3.90	13.36
G32	7.64	n.a.	5.71	4.04	4.18	4.34	4.11/4.11	12.35
U33	7.76	5.47	5.79	4.29	4.60	4.39	4.38/4.23	
C34	7.94	6.14	6.20	4.42	4.28	4.65	4.22/4.17	
A35	8.54	8.38	5.99	5.08	4.86	4.47	3.97/3.87	
U36	7.35	6.29	5.83	4.67	4.49	4.47	3.95/3.95	14.59
C37	7.69	5.63	5.64	4.81	n.d.[c]	4.35	n.d./n.d.	8.00/6.49
G38	7.63	n.a.	5.88	4.81	4.59	4.25	4.16/4.09	12.44
G39	7.30	n.a.	5.65	4.48[b]	4.27[b]	4.40	4.13/4.06	12.40
A40	7.65	7.27	5.84	4.48	4.58	4.47	4.09/4.07	
A41	7.57	7.74	5.83	4.27	4.42	4.41	4.08/4.03	
C42	7.32	5.25	5.42	4.08	4.24	4.23	4.08/4.01	
C43	7.55	5.59	5.60	4.47	4.40	4.14	4.03/3.95	
A44	8.31	8.07	5.99	4.51	4.41	4.27	4.16/4.05	

Table 2. Continued

	C8/C6	C2/C5	C1′	C2′	C3′	C4′	C5′
G1	142.1	n.a.[a]	92.7	74.9	72.0	82.5	65.4
G2	141.4	n.a.	91.9	75.7	72.4	82.3	66.1
G3	141.2	n.a.	92.2	76.1	72.3	84.7	66.0
A4	139.1	153.7	92.2	75.0	72.0	81.2	65.8
G5	136.1	n.a.	92.2	74.8	71.8	85.7	65.4
C6	141.7	96.5	93.1	76.0	71.8	82.2	64.0
U7	143.7	104.5	92.4	75.8	73.5	83.1	65.5
C8	143.1	97.7	93.4	75.1	72.5	82.3	64.2
A9	141.1	154.9	91.4	75.4	73.3	82.4	66.0
A10	140.0	154.9	91.5	75.0	73.6	82.9	66.2
C11	140.8	97.0	93.0	75.5	72.3	81.9	62.2
U12	142.8	105.0	92.1	74.7	75.0	84.1	66.6
C13	142.1	97.7	94.1	74.9	71.7	81.9	65.2
U14	143.1	103.1	93.3	74.8	71.6	81.6	63.9
C15	142.1	96.7	93.1	75.1	71.8	81.3	65.4
C16	141.8	97.2	93.4	74.9	72.2	81.3	64.0
C17	141.3	97.2	93.4	75.5	71.7	81.3	63.8
C18	144.9	98.5	94.1	75.7	70.8	84.0	66.9
C19	142.1	98.5	92.3	75.9	72.0	86.3	66.9
C20	141.4	97.1	94.3	74.3	71.5	83.6	64.2
C21	141.8	96.7	91.7	75.4	72.2	82.7[b]	65.4[b]
U22	145.7	104.7	94.3	83.0	79.1	86.0	67.0
U23	144.1	105.2	93.3	74.8	72.4	82.0	63.3
U24	143.8	104.5	90.2	75.8	72.3	85.1	67.0
U25	143.6	105.0	92.7	74.8	73.3	83.0	65.3
C26	142.4	96.6	93.4	74.8	72.8	81.3	65.2[b]
C27	142.3	98.5	95.8	78.9[b]	72.0	83.9[b]	66.9[b]
G28	143.0	n.a.	93.3[b]	74.9	72.0	81.2	65.3[b]
A29	141.1	153.9	90.5	75.5	72.0	83.4	66.0
G30	136.2	n.a.	92.2	75.4	75.8	83.0	64.8
G31	135.7	n.a.	91.9[b]	75.1	72.3	85.8	64.4[b]
G32	139.4	n.a.	85.7	76.7	71.5	84.0	65.0
U33	143.1	103.4	90.6	75.3	75.6	84.0	66.5
C34	142.1	99.5	90.0	75.9	72.1[b]	84.6	67.6
A35	142.1	157.6	90.2	79.7	75.1	83.4	65.9
U36	145.5	107.1	88.7	77.4	74.9	83.4	67.8
C37	143.6	97.7	93.2	75.5[b]	n.d.	86.8	n.d.
G38	142.4	n.a.	91.9	75.2[b]	71.8	84.2	66.0
G39	136.1	n.a.	92.3	75.4[b]	71.9[b]	81.3	64.3[b]
A40	139.2	153.5	92.0	75.0	72.1	81.4	64.5
A41	139.0	154.4	91.7	75.5	72.4	81.8	64.4
C42	141.4	97.3	92.9	75.5	72.7	82.5	64.5
C43	142.1	97.7	91.7	75.7	71.6	82.4	65.4
A44	142.0	155.7	89.5	76.7	72.0	85.3	66.8

[a] Not applicable.
[b] Tentative assignment.
[c] Not determined.

where T is the total constant time period. Then:

$$^{3}J_{C4'P_{5'}} = \frac{\cos^{-1}(R \times CF_{3'})}{\pi \times T}$$

in which R denotes the ratio of intensities of a particular H4′-C4′ crosspeak in the coupled and uncoupled spectrum, respectively. Most of the values for R could be obtained from spin-echo difference CT-HSQC experiments owing to the favorable resolution of the H4′-C4′ crosspeaks. For some helical residues of which these peaks are overlapping, the ratio could be obtained from H1′-C4′ crosspeaks in spin-echo difference 2D CT-HCCH-DIPSI spectra. This experiment, in which the homonuclear DIPSI period was optimized for C4′ to C1′ coherence transfer, proceeds analogously to the ^{31}P-modulated HCCH correlation experiment.

The ^{3}J(C4′,P$_{5'}$) coupling constants thus calculated ranged from 2 to 12 Hz. This covers the entire area of 90 to 270° that is typically occupied by the angle β (Saenger, 1984). However, the majority of determined scalar couplings amount to about 11 Hz corresponding to the β^t regime, which reflects the presence of an A-helical rotameric state. Values for ^{3}J(C4′,P$_{5'}$) deviating from this class indicated the presence of a $\pm ac$ β angle. This ambiguity could be solved by iteratively entering the possible solutions into the structure calculation protocol in order to make out which value best fitted all other experimental data. If no discernible preference for either of the two possibilities was observed, the particular angle was eventually left unrestrained.

The angle γ

Highly degenerate proton and carbon chemical shifts and unfavorable line-widths obstruct the measurement of the γ-related ^{3}J(H4′,H5′) and ^{3}J(H4′,H5″) coupling constants from an HCCH-E.COSY (Griesinger and Eggenberger, 1992) or any other COSY-type experiment. However, an estimate of these couplings can also be derived from (^{1}H,^{1}H) TOCSY-based experiments, as has been shown for a homonuclear TOCSY-NOESY (Wijmenga et al., 1994). For our purposes, a TOCSY-CT-HSQC was executed so as to profit from the C4′ dispersion. The advantage of this method is that it produces in-phase H4′-H5′/H5″ crosspeaks whose intensities are dependent on γ. At a mixing time of 40 ms the coherence transfer is poor for γ^+ but rather efficient for γ^t conformations (and for γ^-, which in practice does not occur (Saenger, 1984)), irrespective of the sugar pucker (Wijmenga et al.,

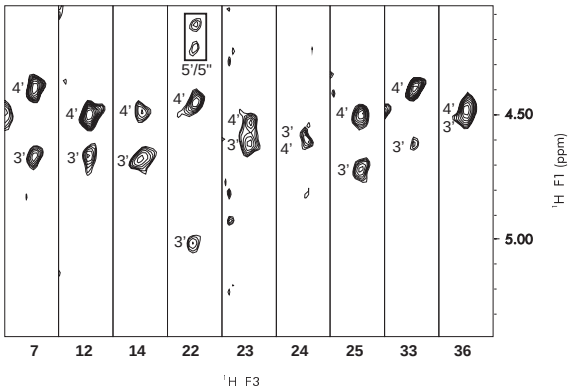

Figure 4. Strips taken from a 600 MHz 3D TOCSY-HSQC, using constant time evolution and gradient enhancement for the ^{13}C-dimension, performed on the U-labeled pk44. Strips were cut at the appropriate C4′ plane (F_2) and H4′ chemical shift (F_3) of the uridine residues, as numbered at the bottom. Crosspeaks to the H3′s are strong in case of N-type sugar puckers, and weak for S-type puckers. Transfer to the H5′/H5′′ protons (boxed), indicative of a γ^t conformation at the employed ^{13}C-^{13}C DIPSI mixing time of 40 ms, is observed for U22. Spectral widths of 969 and 4045 Hz were used for F_1 and F_2, recording 60 and 92 complex increments, respectively, to a final experiment time of 30 h. States-TPPI phase-cycling was used in F_1.

1994). This procedure therefore serves as a monitor for a qualitative discrimination of γ^+ and γ^t oriented residues (see figure). Interestingly, two nucleotides in the pk44 molecule were found to have the rather exotic γ^t conformation: U22, which is discussed below, and A44, which is very mobile and apparently adopts different orientations that are in rapid equilibrium.

Discussion

Structural aspects

The presented procedure for resonance assignment and restraint collection permits the determination of a solution structure that is resolved at an atomic level for large RNA molecules. An overview of the calculated structure of the TYMV pseudoknot has been given previously (Kolk et al., 1998). Due to the elongated shape of the molecule the 2.1 Å overall rmsd found for the molecule obscures the local structural resolution that was obtained for its subdomains. Local rmsd's ranged from 0.7 Å for the T-loop to 1.4 Å for loop 1.

The described methods for the determination of torsion angles provide a substantial contribution to the set of restraints derived from standard methods. Moreover, they deliver direct experimental evidence

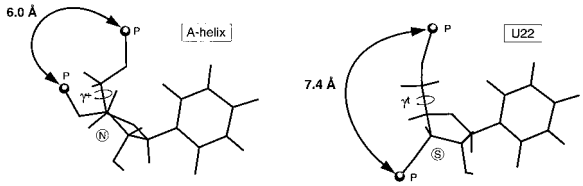

Figure 5. Representation of the interphosphate distances observed in helical RNA (left) and for residue U22 (right). The backbone around this nucleotide is fully extended through an S-type sugar conformation and a γ-angle in the *trans* region.

for some structural properties that follow from the structure calculations. As a clear example hereof serve the deviating torsion angles that were determined for residue U22. This residue is located in the major groove of stem 2 and spans most of the distance between C20 and U23 (see Figure 1). For this purpose U22 has to be fully extended, thereby maximizing the distance between the phosphates at both sides of the nucleotide. This distance is generally about 6 Å for residues in an A-type helix, and about 7 Å in case of an S-type sugar pucker. U22 attains a value as high as 7.4 Å via a combination of an S-type sugar pucker and an unusual γ-*trans* conformation (see Figure 5), both of which were confirmed experimentally.

In retrospect, the major obstacles for the structure calculation of pk44 relate more to the problems of resonance broadening than to its mere size. The effects of conformational averaging are mostly centered around the junction site of the pseudoknot. This particular location constitutes the nexus of the pseudoknot's helical realm to both loop 1 and loop 2. This is reflected by the helical twist between the A29-U36 and G30-C20 base-pairs, which is increased by 20° to an angle of 52° in order to accommodate the negatively charged C20-**p**-C21 and A35-**p**-U36 phosphate groups. As a consequence, the base moieties of C20 and U36 are no longer in a stacked orientation, as is the case in a regular A-helix.

As has been argued by Wijmenga et al. (1997), ring-current effects are by far the most important determinants of chemical shifts in nucleic acids, which accounts, for instance, for the unusually high value of 6.3 ppm for the U36 H5 resonance. By the same argument, even minute alterations in the relative orientations of stems 1 and 2 will influence the line shapes of related resonances in a manner depending on the time scale at which they take place. As has been pointed out, resonance line-widths of A29 through G31, C20 and U36 correlate strongly to the field strength in the 400 to 750 MHz range. This is a

clear indication of mutual flexibility of stem 1 and 2, taking place at a time scale of the order of milliseconds, which appears to be a rather common exchange rate in RNA molecules.

The amplitudes of these internal motions are difficult to determine exactly, due to the limited simulation times of the TAD and SA protocols, which are in the picosecond rather than the millisecond range. Hence, some of the displacements may not show up in the final ensemble of structures, and the local rmsd of 0.3 Å that is observed for the residues at the junction (i.e. C20, A29, G30 and U36) may underestimate the true distribution of conformations in that area. Nevertheless, high-amplitude motions are improbable considering the intrinsically consistent set of NOEs that was collected for this region. Furthermore, it could be verified that even small dislocations of the bases can have significant effects on the relating chemical shifts using the in-house written program NUCHEMICS (Wijmenga et al., 1997). This chemical shift calculation software was originally written for DNA but was reparametrized for RNA. Chemical shifts were calculated for the final ensemble of 10 structures including only the helical residues near the junction site. Consequently, the chemical shift differences induced by proximal single-stranded residues were disregarded so as to monitor exclusively the mutual effects of the stem regions. The results indicated a spread in chemical shifts of 0.2–0.3 ppm for the relating resonances within the set of structures (data not shown). At a millisecond exchange rate this chemical shift difference convincingly explains the line broadening observed in the NMR spectra.

Evaluation of torsion angles

The described approach for the determination of torsion angles β, γ and ϵ relies on a few assumptions. Most importantly, it is postulated that these angles are always within the mentioned ranges. For ϵ, it is generally accepted that the *gauche*$^+$ region is stereochemically forbidden, but conformations somewhat beyond the peripheries of the canonical domain of 180°–270° cannot be ruled out conclusively. The same holds true for γ-angles in the g^- region. Such conformations are very unorthodox, however, and are rarely, if ever, observed in RNA. Furthermore, they are expected to reveal themselves in deviating chemical shifts and NOE connectivities. Therefore, these angles were only used as a supplement to the NOE restraints dominating the structure, and were found to refine rather than distort the 'NOE-only' structure of the molecule.

A similar argument is valid for the error bounds that were taken into account. Obviously, the angle restraints are most influential for residues that are least confined in motion by distance restraints, such as U22. A very limited number of NOEs is observed for this residue, and one might question the validity of the experimentally derived torsion angles for a mobile residue. However, the exchange rate for this residue is rather slow, considering its broadened resonances, and the observed deviating sugar pucker (almost entirely of the S-type) and chemical shifts (in particular for sugar ^{1}H and ^{13}C resonances) indicate the presence of one major conformation and do not support a large conformational freedom for this residue. It is therefore concluded that the applied 20–30° bounds for the determined backbone torsion angles, which account for experimental errors, also allow for sufficient conformational uncertainty to account for the mobility of this and similar residues. Clearly, this approach is not valid for distinctly flexible nucleotides such as A44, whose small line-widths and inconsistent NOEs indicate an averaging of a wide range of conformations. Consequently, the torsion angles for this residue were left unconstrained.

In summary, the procedure that we have currently exploited to derive backbone torsion angle restraints is able to deal with most of the problems invariably connected to large RNA molecules: a considerable degree of line broadening and resonance overlap. It does not aim to reach a very high precision, as do other methods that have been previously reported (Schwalbe et al., 1994; Sich et al., 1997), but rather at feasibility for systems with complexities that are well beyond the level of an RNA hairpin. A similar semi-quantitative approach for the determination of backbone torsion-angles, using ^{13}C-edited ^{31}P-^{1}H HETCOR and hetero-TOCSY experiments, has been described earlier (Varani et al., 1995; Varani et al., 1996), but these methods turned out less useful for the present system, mainly due to problems of sensitivity. Finally, it is argued that a very precise determination of torsion angles is often unrealistic in regions for which few NOEs have been collected, since this is likely to be the result of dynamical effects.

Acknowledgements

The NMR experiments were performed at the SON Large Scale NMR facility (Nijmegen and Utrecht, The Netherlands). This research was supported by the Netherlands Foundation of Chemical Research (SON) with financial aid from the Netherlands Organization of Scientific Research (NWO). H.A.H. was supported by a grant from the Royal Netherlands Academy of Arts and Sciences (KNAW).

References

Bax, A., Clore, G.M. and Gronenborn, A.M. (1990) *J. Magn. Reson.*, **88**, 425–431.

Blommers, M.J., Nanz, D. and Zerbe, O. (1994) *J. Biomol. NMR*, **4**, 595–601.

Brünger, A.T. (1992) *X-PLOR. A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT.

Dieckmann, T. and Feigon, J. (1997) *J. Biomol. NMR*, **9**, 259–272.

Florentz, C., Briand, J.P., Romby, P., Hirth, L., Ebel, J.P. and Giegé, R. (1982) *EMBO J.*, **2**, 269–276.

Folmer, R.H.A., Hilbers, C.W., Konings, R.N.H. and Hallenga, K. (1995) *J. Biomol. NMR*, **5**, 427–432.

Griesinger, C. and Eggenberger, U. (1992) *J. Magn. Reson.*, **97**, 426–434.

Hall, T.C. (1979) *Int. Rev. Cytol.*, **60**, 1–26.

Ikura, M., Kay, L.E. and Bax, A. (1991) *J. Biomol. NMR*, **1**, 299–304.

Ikura, M., Kay, L.E., Tschudin, R. and Bax, A. (1990) *J. Magn. Reson.*, **86**, 204–209.

Kay, L.E., Ikura, M. and Bax, A. (1990) *J. Am. Chem. Soc.*, **112**, 888–889.

Kay, L.E., Keifer, P. and Saarinen, T. (1992) *J. Am. Chem. Soc.*, **114**, 10663–10664.

Kolk, M.H., Van der Graaf, M., Wijmenga, S.S., Pleij, C.W.A., Heus, H.A. and Hilbers, C.W. (1998) *Science*, **280**, 434–438.

Legault, P., Farmer, B.T., Mueller, L. and Pardi, A. (1994) *J. Am. Chem. Soc.*, **116**, 2203–2204.

Legault, P., Jucker, F.M. and Pardi, A. (1995) *FEBS Lett.*, **362**, 156–160.

Mans, R.M., Pleij, C.W.A. and Bosch, L. (1991) *Eur. J. Biochem.*, **201**, 303–324.

Marino, J.P., Diener, J.L., Moore, P.B. and Griesinger, C. (1997) *J. Am. Chem. Soc.*, **119**, 7361–7366.

Marion, D., Ikura, M., Tschudin, R. and Bax, A. (1989) *J. Magn. Reson.*, **85**, 393–399.

Nikonowicz, E.P. and Pardi, A. (1993) *J. Mol. Biol.*, **232**, 1141–1156.

Pardi, A. and Nikonowicz, E.P. (1992) *J. Am. Chem. Soc.*, **114**, 9292–9203.

Pleij, C.W.A. (1994) *Curr. Opin. Struct. Biol.*, **4**, 337–344.

Quigley, G.J. and Rich, A. (1976) *Science*, **194**, 796–806.

Rietveld, K., Van Poelgeest, R., Pleij, C.W.A., Van Boom, J.H. and Bosch, L. (1982) *Nucleic Acids Res.*, **10**, 1929–1946.

Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, NY.

Schwalbe, H., Marino, J.P., King, G.C., Wechselberger, R., Bermel, W. and Griesinger, C. (1994) *J. Biomol. NMR*, **4**, 631–644.

Sich, C., Ohlenschlager, O., Ramachandran, R., Gorlach, M. and Brown, L.R. (1997) *Biochemistry*, **36**, 13989–14002.

Simorre, J.P., Zimmermann, G.R., Pardi, A., Farmer, B.T. and Mueller, L. (1995) *J. Biomol. NMR*, **6**, 427–432.

Stein, E.G., Rice, L.M. and Brünger, A.T. (1997) *J. Magn. Reson.*, **124**, 154–164.

Ten Dam, E., Pleij, C.W.A. and Draper, D. (1992) *Biochemistry*, **31**, 11665–11676.

Van Belkum, A., Wiersema, P.J., Joordens, J., Pleij, C.W.A., Hilbers, C.W. and Bosch, L. (1989) *Eur. J. Biochem.*, **183**, 591–601.

Varani, G., Aboul-ela, F. and Allain, F. H.-T. (1996) *Prog. NMR Spectrosc.*, **29**, 51–127.

Varani, G., Aboul-ela, F., Allain, F. H.-T. and Gubser, C.C. (1995) *J. Biomol. NMR*, **5**, 315–320.

Varani, G. and Tinoco, I.J. (1991) *Q. Rev. Biophys.*, **24**, 479–532.

Vuister, G.W., Wang, A.C. and Bax, A. (1993) *J. Am. Chem. Soc.*, **115**, 5334–5335.

Wijmenga, S.S., Heus, H.A., Leeuw, H.A.E., Hoppe, H., van der Graaf, M. and Hilbers, C.W. (1995) *J. Biomol. NMR*, **5**, 82–86.

Wijmenga, S.S., Heus, H.A., Werten, B., van der Marel, G., van Boom, J.H. and Hilbers, C.W. (1994) *J. Magn. Reson. B*, **103**, 134–141.

Wijmenga, S.S., Kruithof, M. and Hilbers, C.W. (1997) *J. Biomol. NMR*, **10**, 337–350.

Wijmenga, S.S., Mooren, M.W. and Hilbers, C.W. (1993) In *NMR of macromolecules. A practical approach* (Ed., Roberts, G.C.K.), Oxford University Press, New York, NY, pp. 217–288.

Yao, L.J., James, T.L., Kealey, J.T., Santi, D.V. and Schmitz, U. (1997) *J. Biomol. NMR*, **9**, 229–244.