Volume 324, number 3, 296–300 FEBS 12600 June 1993

The catalytic domain of human hepatitis delta virus RNA

A proton nuclear magnetic resonance study

Bao-Shiang Lee[a], Huey-Nan Wu[b] and Tai-Huang Huang[a]

[a]*Institute of Biomedical Sciences* and [b]*Institute of Molecular Biology, Academia Sinica, Nankang, Taipei 11529, Taiwan, ROC*

Received 5 April 1993; revised version received 28 April 1993

We have obtained and analyzed the 600 MHz proton NMR spectra of a 74-mer RNA derived from the catalytic domain of hepatitis delta virus genomic RNA (HDV RNA) to determine its secondary structure. Deconvolution of the NMR spectrum obtained at 32°C indicates that part of the 74-mer RNA molecule may exist in multiple conformations in equilibrium. The major conformer contains two A–U base pairs and 14 ± 2 G-C base pairs. It appears to contain no standard G-U base pairs. Our NMR melting study suggests that this conformer has at least two stem-loop regions. One of the regions has been identified to be a tetra-loop. We have assigned five imino proton resonances of the tetra-loop stem. Our data is consistent with the pseudoknot model of Perrotta and Been.

Ribozyme; RNA; Structure; NMR; Hepatitis delta virus

1. INTRODUCTION

Hepatitis delta virus (HDV) is a 36-nm particle containing a single-stranded circular RNA genome of 1.7 kilobases [1]. The HDV RNAs of both genomic and antigenomic sense possess site-specific self-cleaving activity, yielding a 3′ cleavage product with a 5′ hydroxyl group and a 5′ cleavage product with a 2′3′-cyclic phosphoryl group [2]. Several secondary structures have been proposed for genomic HDV RNA, including pseudoknot [3] (Fig. 1), axehead [4], and the clover leaf model [2]. All three models support the presence of a stem region next to the cleavage site (Helix I in Fig. 1) and a stem-and-loop region. Unlike other ribozymes [5–9] however, the secondary structure of other regions of the HDV RNA catalytic domain is still uncertain and its detailed three-dimensional conformation is far from understood. Recent results of an extensive mutagenesis study have indicated that the HDV genomic catalytic domain contains a 6-base pair helix adjacent to the cleavage site and a stem-and-loop [10]. In this report we present the results of NMR studies of the secondary structure of a 74-mer RNA analogue (Fig. 1a). The specific features of this 74-mer are: (i) the hinge has been deleted to 5 nucleotides in length; (ii) the stem-and-loop has been replaced by a well-known super stable tetra-loop; (iii) the cleavage site has been changed from UG to UC with resulting loss of self-cleavage activity. The modification of the hinge and stem-and-loop did not alter the self-cleaving activity (Wu et al., unpublished observation). We assume that the UG-to-UC change at the cleavage site probably does not perturb the overall conformation of the 74-mer ribozyme molecule, although we have no evidence to support this assumption; however, since the change is at the end of the putative helix I region we believe that the secondary and tertiary structural features extracted from our NMR results are relevant.

Correspondence address T.-H. Huang, Institute of Biomedical Sciences and Academia Sinica, Nankang, Taipei 11529, Taiwan, ROC. Fax (886) (2) 785 3569.

2. MATERIALS AND METHODS

2.1. *RNA preparation*

The 74-mer RNA was synthesized by T7 RNA polymerase with a *Bam*HI-linearized plasmid as template, as described previously [11]. The isolated RNA was repurified by FPLC using a Mono-Q column (HR 5/5) (Pharmacia Co.) in 1 M NaCl and 10 mM sodium phosphate, pH 7.0. NMR samples were prepared by changing the sample buffer to the desired conditions (0.4 ml, 0.9 mM in 95% H_2O/5% D_2O, 100 mM NaCl, 0.1 mM sodium EDTA in 10 mM sodium phosphate at desired pHs) using a Centricon-10 ultrafilter (Millipore Co.). RNA in the NMR tube was heated to 95°C for 2 min and slowly cooled to room temperature.

2.2. *NMR spectroscopy*

NMR data were collected on a Bruker AMX-600 spectrometer operating at 14.09 T (600.13 MHz for protons). Chemical shifts are reported relative to 2,2-dimethyl-2-silapentane-5-sulfonate (DSS). NMR spectra were collected with 4,096 complex data points over a spectrum width of 12.19 kHz using a binomial 1-1 pulse sequence centered at 10 or 12 ppm for water suppression. All spectra were processed with a 90°-shifted sine square window function before Fourier transformation. The 2D NOESY spectra were obtained in the phase-sensitive mode (TPPI) using a 1-1 pulse sequence for solvent suppression. Experimental conditions were: 250 ms mixing time, recycle delay time of 1.5 s, 400 scans with 2,048 complex data points in the F2 dimension and 320 complex free induction decays in the F1 dimension. The spectrum was zero-filled to 2K in the F1 dimension.

Published by Elsevier Science Publishers B.V.

Temperature was regulated with a Bruker variable temperature unit to within 0.1°C.

3. RESULTS AND DISCUSSION

2.1. *Determining the number of base pairs*

The proton NMR spectrum at 32°C of the 74-mer RNA fragment in the 9.0–15.0 ppm region in H_2O is shown in Fig. 1b. The sharp resonances between 12 and 15 ppm are due to base-paired imino protons, which are shifted to the low field region due to the combined effects of ring current shift and hydrogen bonding [12,13]. The imino protons not involved in hydrogen-bond formation, such as those in the loop region, resonate between 9 and 12 ppm and are broader due to exchange with solvent protons. Resonances at 9.8 and 14.28 ppm are two unique resonances which are well resolved and can be employed as intensity standards for determining the total number of base pairs in the RNA molecule, as discussed below.

To determine the number of base pairs present in the HDV RNA molecule, we have simulated the spectral region between 9 and 15 ppm using the spectra deconvolution sub-routine in the Bruker UXNMR package. The program accomplishes this by automatically assigning peaks and allowing the line width and intensity of each peak to vary until the best fit is obtained. As shown in Fig. 1c the simulation worked quite well. To quantify the resonances, the peaks at 9.8 and 14.28 ppm were taken to be one proton each and were used as intensity standards. The justification for using these two resonances is further discussed in the following paragraphs. The apparent higher intensity of the 9.8 ppm resonance than that of the 14.28 ppm resonance was due to non-uniform excitation of the 1–1 solvent suppression pulse when the carrier frequency was centered at 10 ppm for observing the aromatic resonances. When the center frequency was set at 12 ppm the intensities of the two peaks were similar. The result is shown in Fig. 1c. We found that the total intensity of resonances between 12 and 15 ppm corresponded to 16 ± 2 protons, suggesting the presence of 16 ± 2 base pairs. This agrees well with the number of discernible peaks in this region, however, although most of the resonances have intensities close to integer numbers of protons, the 12.99 and 13.09 ppm resonances have intensities corresponding to 0.2 and 0.4 protons, respectively. The presence of non-integer intensities could occur if the RNA molecule exists in multiple conformations or, perhaps as a result of partial melting at this temperature. The presence of multiple conformations is supported by the observation that: (i) the 14.28 ppm resonance was split into two peaks at a temperatures below 16°C, as discussed in the next section; and (ii) the total number of base-paired protons estimated from the 22°C spectrum was consistent with that estimated from the 32°C spectrum.

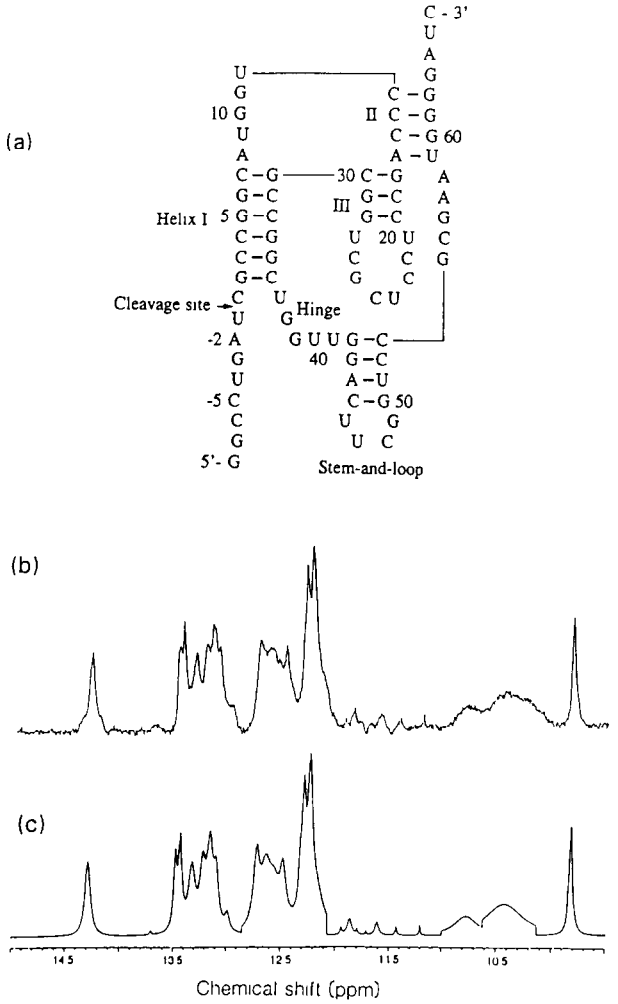

Fig. 1. (a) The proposed secondary structure of the 74-mer RNA fragment of Perrotta and Been [3]. The cleavage site is indicated. A long stem-loop IV has been replaced by a tetra-loop. The nucleotides are numbered with respect to the cleavage site. (b) 600 MHz proton NMR spectrum of the 74-mer HDV RNA in H_2O at 32°C, pH 6.5. (c) Simulated spectrum.

3.2. *Determining the type of base pairs*

To identify the type of base pairs we relied on two pieces of information: chemical shift and NOE pattern [12–14]. In a standard Watson–Crick base pair, the imino proton of an A–U base pair resonates between 13 and 15 ppm, and that of a G–C base pair resonates between 11 and 13 ppm. One can not assign the type of base-pairs based on chemical shift information alone, however, especially for resonances near 13 ppm, or for protons in regions of unusual structures. Thus, NOE patterns are very important pieces of additional information for identifying the type of base pairs. In A–U base pairs, the imino proton of U is 2.8 Å from the aromatic proton H2 of A (around 7.2 ppm), thus giving a strong NOE cross-peak in the imino proton–aromatic proton region. For a G–C base pair, usually two strong NOE peaks in this region can be detected; one of the

NOE peaks is due to the direct effect of the imino proton of G and the hydrogen bonded amino proton of C (around 6.7 ppm), which is 2.6 Å away from the imino proton. The other cross-peak is due to the non-hydrogen bonded exocyclic amino proton of C (around 8.2 ppm), which is partially saturated due to the presence of fast rotation of the amino group about the C–N bond. Fig. 2 shows the 2D NOESY spectrum in the amino-to-imino proton region which displays the connectivities between imino and amino protons. Thirteen imino proton peaks are clearly discernable. Apparently three base-paired imino protons at 12.24, 12.71, and 12.99 ppm are exchanging with solvent too fast to exert NOE with adjacent protons. With the exception of the 13.47 and 14.28 ppm resonances, each imino proton resonance shows two major cross-peaks with the amino protons at around 6.7 and 8.2 ppm, as expected for G–C base pairs. Thus, these resonances are assigned to G–C base pairs. For the 13.47 and 14.28 ppm resonances, a very strong cross-peak at 7.15 and 7.75 ppm, respectively, was observed, presumably between the imino proton of an A–U base pair and the AH2 proton. Therefore, the 13.47 and the 14.28 ppm resonances are assigned to the imino protons of an A–U base pair. Apparently during the long mixing time used for the NOE experiment, inter-base pair NOE peaks such as those at (14.28, 6.75), (14.28, 8.42), (13.43, 7.8) and (12.29, 7.8) are also observable. This result suggests that the three base pairs resonating at 14.28, 13.43 and 12.29 ppm are adjacent base pairs. This is confirmed by additional 1- and 2D NMR experiments which suggests that these base pairs belong to a tetra-loop. From chemical shift alone, the three resonances at 12.24, 12.70 and 12.99 ppm, which give no NOE peaks in the 2D NOESY spectrum, are also likely to be G–C base pairs.

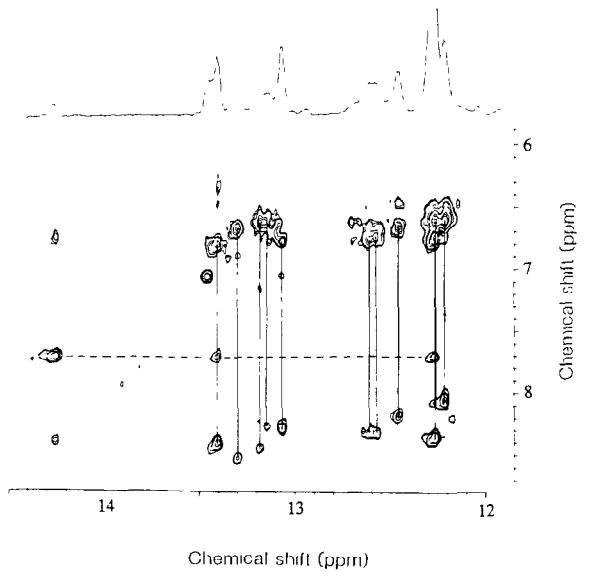

Fig. 2. Contour plot of the imino-aromatic proton region of the 2D NOESY spectrum of the 74-mer HDV RNA in H_2O at 32°C, pH 6.5. Cross-peaks belonging to the same aromatic groups are connected by vertical solid lines. The horizontal dashed line connects the inter-base-paired NOE cross-peaks

3.3. *Melting of the 74-mer RNA fragment*

A set of NMR spectra of the 74-mer RNA obtained at various temperatures is shown in Fig. 3. At lower temperatures the general pattern of the base-paired imino proton resonances between 12 and 15 ppm did not change, except for an expected general broadening of all resonances; however, the loop region proton resonances between 9 and 12 ppm sharpened substantially at low temperatures, consistent with the fact that disappearance of these resonances at higher temperatures is due to rapid exchanging with solvent. An exception to this is the resonance at 9.82 ppm which broadened and disappeared at a higher temperature, similar to the behavior of the resonance at 14.28 ppm. We propose that these two resonances belong to two base pairs in a stable tetra-loop, as further explained in the next paragraph. Interestingly, the A–U imino proton resonance at 14.28 ppm split into two resonances at temperatures between 12 and 16°C, indicating the presence of two conformations in nearly equal populations which undergo slow exchange at low temperatures. At higher temperatures only one resonance was observed, presumably due to the increased exchange rate between the two conformations. Furthermore, at 62°C five prominent resonances at 14.28, 13.43, 13.15, 12.28 and 9.82 ppm were clearly identifiable. Additional NOE results discussed below

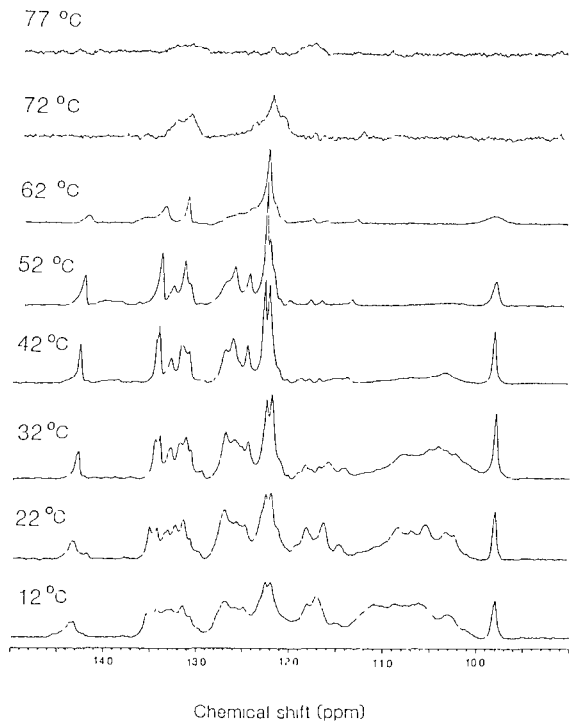

Fig. 3. Temperature variation of the 600 MHz proton NMR spectra of the 74-mer HDV RNA at pH 6.

suggest that four of these resonances (at 14.28, 13.43, 12.28 and 9.82 ppm) are likely to originate from the same tetra-loop, as reported by Varani et al. [15]. The fifth resonance, at 13.15 ppm, plus additional broad resonances centered around 12.7 and 13.6 ppm, must originate from another stem with a high melting temperature. Thus, these data indicate that the 74-mer RNA consists of at least two stem-and-loops with melting temperatures above 70°C.

3.4. *Assignment of the tetra-loop imino proton resonances*

A tetra-loop sequence 5'GGACUUCGGUCC-3' [15] (sequence 42–53 in Fig. 1a) was built into the 74-mer RNA sequence to replace a longer stem-loop region, and caused no change in self-cleavage activity (Wu, unpublished observation). Thus it is likely that the tetra-loop structure is retained in the overall folding of the 74-mer. The NMR spectrum of this loop alone contains 7 distinct imino proton resonances, four from the four stem base pairs (14.52, 13.63, 12.93 and 12.68 ppm), two from the loop region, bases, U46 and U47 (11.99 and 11.36 ppm), and one from the wobble base pair, G49–U46 (9.96 ppm). The NMR spectrum shown in Fig. 1 clearly shows the presence of the two most distinct resonances at 14.28 and 9.82 ppm. The melting study discussed above further supports this assignment. Thus, the 14.28 ppm resonance is assigned to the A44–U51 base pair and 9.82 ppm resonance is assigned to the imino proton of G49. To identify other resonances of the tetra-loop we obtained a 2D NOESY spectrum in the imino–imino region (Fig. 4). The 14.28 ppm resonance shows two NOE peaks at 13.43 and 12.28 ppm. To observe smaller NOE peaks we have obtained a series of 1D NOE difference spectra at 400 ms presaturation time (spectra not shown). These data show a closed NOE pattern among the three resonances at 14.28, 13.43 and 12.29 ppm, but not extending to other peaks. From the chemical shift we assign the 13.43 ppm resonance to the G43–C52 base pair. The two end G–C base pair imino protons are both believed to resonate at 12.29 ppm, which has an intensity of two protons. These assignments are consistent with the melting behavior of the four resonances. Surprisingly, the two loop region uridine imino protons are not observable, presumably due to fast exchange. Thus, the tetra-loop structure observed in the 74-mer is not completely identical to that observed in the isolated sequence.

In conclusion, our study shows that the 74-mer RNA may exist in multiple conformations. The major conformer contains two A–U base pairs and 14 ± 2 G–C base pairs. Our melting study and NOESY results show that there are at least two stem-loop regions with melting points greater than 65°C. One of the regions has been identified to be an unusually stable tetra-loop. Four resonances belonging to the tetra-loop have been identified. In discriminating among the three proposed

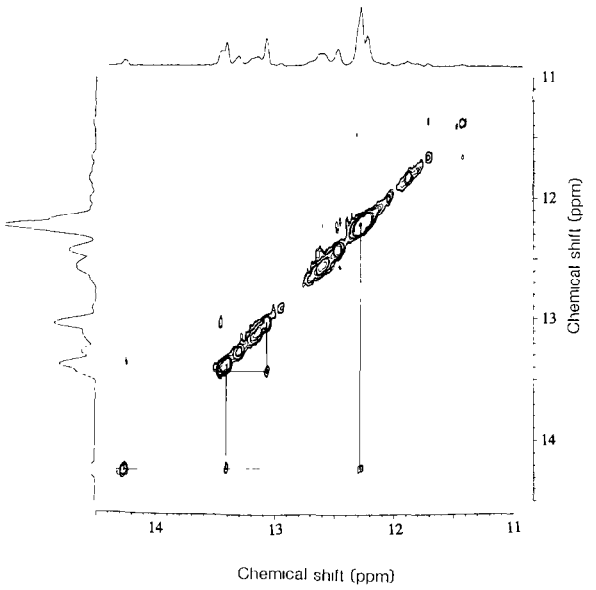

Fig. 4. Contour plot of the imino–imino proton region of the 2D NOESY spectrum of the 74-mer HDV RNA in H_2O at 32°C, pH 6.5. The cross-peaks are connected by solid lines.

models regarding to the secondary structure of HDV RNA we found that the axehead model [4] which contains 11 G-C pairs, one A–U pair, and two G–U pairs (based on our sequence) does not agree with the present NMR results, whilst both the pseudoknot model [3] (Fig. 1), which contains 15 G–C pairs and two A–U pairs, and the clover leaf model [2], which contains 13 G–C pairs and 2 A–U pairs, are consistent with our NMR data. Additional biochemical evidences led us to favor the pseudoknot model (Wu et al., unpublished observation). Thus, we tentatively assign the 13.47 ppm resonance to the A16–U59 imino proton.

Acknowledgements We thank Prof. B.C. Wang for the gift of T7 RNA polymerase, Ms. Y.C. Lai and H.W. Hwang for assistance with sample preparation, and Ms. L.H. Tseng for assistance with NMR experiments. This work was supported by a grant from National Science Council of the Republic of China.

REFERENCES

[1] Chen, P.J., Kapana, G., Goldberg, J., Mason, W., Werner, B., Gerin, J.L. and Taylor, J. (1986) Proc. Natl. Acad. Sci. USA 83, 8774–8778.
[2] Wu, H.-N., Wang, Y.-J., Hung, C.-F., Lee, H.-J. and Lai, M.M.C. (1992) J. Mol. Biol. 223, 233–245.
[3] Perrotta, A.T. and Been, M.D. (1991) Nature 350, 434–436.
[4] Branch, A.D. and Robertson, H.D. (1991) Proc. Natl. Acad. Sci. USA 88, 10163–10167
[5] Foster, A.C. and Symons, R.H. (1987) Cell 49, 211–220.
[6] Uhlenbeck, O.C. (1987) Nature 328, 596–600.
[7] Haseloff, J. and Gerlach, W.L. (1988) Nature 334, 585–591.
[8] Felstein, P.A., Buzayan, J.M. and Breuning, G. (1989) Gene 82, 53–61.
[9] Hampel, A., Tritz, R., Hicks, M. and Cruz, P. (1990) Nucleic Acids Res. 18, 299–204.

[10] Wu, H.-N., Lin, Y.-L., Lin, F.-P., Makino, S., Chang, M.F. and Lai, M.M.C. (1989) Proc. Natl. Acad. Sci. USA 86, 1831–1835.
[11] Milligan, J.F., Groebe, D.R., Witherell, G.W. and Uhlenbeck, O.C. (1987) Nucleic Acids Res. 15, 8783–8798.
[12] Kearns, D.R. (1976) Prog. Nucleic Acids Res. Mol. Biol. 18, 91–149.
[13] Robillard, G.T. and Reid, B.R. (1979) in: Biological Applications of Magnetic Resonance (Shulman, R.G., ed.) pp. 45–112, Academic Press, New York.
[14] Heus, H.A. and Pardi, A. (1991) J. Mol. Biol. 217, 113–124.
[15] Varani, G., Cheong, C. and Tinoco Jr. I. (1991) Biochemistry 30, 3280–3289.