Structure of influenza virus panhandle RNA studied by NMR spectroscopy and molecular modeling

Hae-Kap Cheong^{1,2}, Chaejoon Cheong^{1,*}, Yeon-Sook Lee³, Baik L. Seong⁴ and Byong-Seok Choi²

¹Magnetic Resonance Group, Korea Basic Science Institute, Eoun-dong 52, Yusung-gu, Taejon 305-333, Korea, ²Department of Chemistry, Korea Advanced Institute of Science and Technology, Kusung-dong 373-1, Yusung-gu, Taejon 305-701, Korea, ³Hanhyo Institute of Technology, Chonmin-dong 461-6, Yusung-gu, Taejon 305-390, Korea and ⁴Department of Biotechnology, College of Engineering and Bioproducts Research Center, Yonsei University, 134 Shinchon-dong, Seodaemun-gu, Seoul 120-749, Korea

Received October 12, 1998; Revised and Accepted January 7, 1999

ABSTRACT

The structure of a 34 nucleotide RNA molecule in solution, which contains the conserved panhandle sequences, was determined by NMR spectroscopy and molecular modeling. The partially double-stranded panhandle structure of the influenza virus RNA serves to regulate initiation and termination of viral transcription as well as polyadenylation. The panhandle RNA consists of internal loop flanked by short helices. The nucleotides at or near the internal loop are crucial for polymerase binding and transcriptional activity. They show more flexible conformational character than the Watson-Crick base-paired region, especially for the backbone torsion angles of α , γ and δ . Although residues A10 and A12 are stacked in the helix, the phosphodiester backbones are distorted. Residues A12, A13 and G25 show dynamic sugar conformations and the backbone conformations of these nucleotides are flexible. This backbone conformation and its associated flexibility may be important for protein-RNA interactions as well as base-specific interactions.

INTRODUCTION

The genome of influenza A viruses consists of eight segments of single-stranded, negative sense RNA (1). The viral RNAs (vRNAs) are transcribed by the three polymerase subunits (PA, PB1 and PB2) associated with the genome segments and are complexed with the polymerase subunits and nucleoprotein (NP; 2). Within infected cells, two classes of positive sense RNAs are made by the polymerase: (i) an incomplete, 3'-polyadenylated transcript which serves as an mRNA; and (ii) a complete, non-polyadenylated transcript which serves as a template for the synthesis of the vRNAs (2). During the initial phase of infection, the viral mRNAs are transcribed (3), while later in the infection process, the full-length transcripts (cRNA) are made. The switch between these two modes of RNA syntheses is thought to be

regulated by the intracellular concentration of NP and by conformational changes at the 3'- and 5'-ends of the vRNA (4).

The sequences of all eight genome segments are highly conserved at their 3'- and 5'-ends (5,6), and the sequences are partially complementary to each other. The two ends are held together by base pairs forming panhandle structures in the ribonucleoprotein (RNP) complex (7). The 3'- and 5'-end sequences of the genome segments contain all of the required signals for transcription, replication and genome packaging (8). The partially double-stranded panhandle structure serves to regulate initiation (9,10) and termination of viral transcription as well as polyadenylation (11,12). The influenza polymerase specifically binds to the conserved sequences located at the 3'- and 5'-ends of the vRNA, with a stronger affinity to the 5'-ends (13); however, the polymerase does not bind to a completely double-stranded RNA (13).

In our earlier studies, we demonstrated that a model vRNA formed a base-paired panhandle structure under protein-free condition (14). As shown in Figure 1A, the secondary structure of our model vRNA contained a G-U base pair and internal bulges. Moreover, an additional G-C base pair was found which was considered a part of an internal loop in the previous model. Here, we report the three-dimensional structure of the model RNA determined by NMR spectroscopy followed by structural calculations using restrained molecular dynamics.

MATERIALS AND METHODS

Synthesis and purification of model RNA

An RNA oligonucleotide, 5'-GGAGCAGAAACAAGGCUUC-GGCCUGCUUUUGCUC-3', was enzymatically synthesized using T7 RNA polymerase and synthetic DNA templates (15). The crude RNA was purified by using 15% polyacrylamide gel electrophoresis under denaturing conditions.

*To whom correspondence should be addressed. Tel: +82 42 865 3431; Fax: +82 42 865 3419; Email: cheong@comp.kbsi.re.kr



Figure 1. (A) Secondary structure and numbering scheme of the influenza virus model RNA panhandle structure. Extra nucleotides introduced at the tetraloop and at both the 5'- and the 3'-ends and the U to C transition mutation within the duplex are shown in small characters. (B) *In vitro* transcription of the model RNA template. Lane 1, 34 nt model RNA T7 transcript as a size marker; lane 2, without added model RNA; lane 3, with the added model RNA.

In vitro transcription

Influenza virus polymerase was prepared according to Seong and Brownlee (16). Approximately 0.1 μ g micrococcal nuclease-treated influenza core protein was mixed with 60 ng of the model RNA and incubated at 30°C for 15 min. Transcription buffer comprised of the following was then added: 50 mM Tris–HCl (pH 7.8), 50 mM KCl, 10 mM NaCl, 1 mM DTT, 5 mM MgCl₂, 0.1% NP-40, 1 mM ATP, 0.5 mM GTP, 0.25 mM UTP, 10 μ M CTP, 50 μ M [α -³²P]CTP, 4 U RNasin (Promega) and 1 mM ApG as a primer, in a total reaction volume of 10 μ l. The reaction mixture was incubated at 30°C for 2 h. After extraction with phenol, the RNA was ethanol precipitated with 1 μ g yeast RNA as a carrier. The reaction product was analyzed using a 15% 8 M urea polyacrylamide gel.

NMR spectroscopy

The purified RNA sample was extensively dialyzed against 10 mM sodium phosphate + 0.01 mM EDTA (pH 6.5). It was then lyophilized and dissolved in 0.25 ml of a 9:1 H₂O/D₂O mixture for exchangeable proton experiments. For non-exchangeable proton experiments, the sample was lyophilized several times in 99.9% D_2O and then dissolved in 0.25 ml 99.96% D_2O (Aldrich). An NMR microtube from Shigemi was used to decrease the sample volume. All NMR experiments were carried out using a Bruker DMX600 NMR spectrometer operating at 600.13 MHz proton frequency. Exchangeable proton spectra were obtained using the 1-1 (jump-return) method (17). One-dimensional NOE experiments were performed with a pre-irradiation time of 600 ms. All two-dimensional NMR spectra were acquired in the phasesensitive mode using the TPPI method (18). The two-dimensional spectra were acquired with 400 or 512 FIDs of 2048 complex data points, using spectral widths of 3800 Hz for spectra acquired in D_2O and 12 000 Hz for spectra acquired in H_2O . The repetition delays were set to ~ 2 s, and 64 scans were averaged for each FID; the total acquisition time was <24 h. All NMR data were processed using FELIX (Biosym/MSI) and UXNMR (Bruker). The data were zero-filled to 2 K real points in t_1 and apodized by

using 45–60° phase-shifted squared sine bells in both dimensions. NOESY spectra acquired in H_2O with 150 and 400 ms mixing times were obtained using a jump–return pulse for solvent suppression (19). NOESY spectra acquired in D_2O were obtained with mixing times of 80, 150 and 400 ms. The residual HDO resonance was presaturated during the relaxation delay. DQF-COSY spectra were obtained using the standard pulse sequence (20).

Structure calculation

Interproton distances between non-exchangeable protons were estimated by integrating the cross-peak intensities of NOESY spectra at 80, 150 and 400 ms mixing times. The pyrimidine H5–H6 distance (2.45 Å) was used as a reference. The cross-peak intensities were classified as strong, medium, weak or very weak, and converted into upper bounds of distance restraints of 2.5, 3.5, 5.0 or 7.0 Å, respectively. Proton–proton coupling constants were measured in a high-resolution (1 Hz/point) DQF-COSY spectrum. Six distance constraints per base pair were added to maintain hydrogen-bonded Watson-Crick and G-U base pairs and to make the base planes coplanar. Distance constraints involving exchangeable protons were set to 2-5 Å when NOEs were seen at 150 ms mixing time. Sugar conformations were determined from J_{H1'-H2'} values. No nucleotides, with the exception of G20, had strong intranucleotide H1'-H6/H8 NOEs at short mixing times (80 ms), thus, all the glycosidic angles χ except G20 were constrained to be anti $(-120 \sim -90^{\circ})$. The amino groups were constrained to be in the plane of the base using torsion angle restraint. Backbone torsion angles (α , β , γ , ϵ and ζ) for the G2~A8, U28~C34, G14~G15 and C22~C23 regions were constrained to the standard A-form. The backbone conformation of the tetra loop region was constrained to the structure obtained by Varani et al. (21). Structure computation and graphic display were achieved using the AMBER force field and InsightII, NMRChitech and Discover software packages running on a Silicon Graphics workstation. Initial structures were generated by distance geometry methods using the isolated two-spin approximation. The best structures, which exhibited the smallest violation of the constraints, were subjected to a simulated annealing procedure with NMRChitech. The structures were first minimized using 2000 cycles of restrained energy minimization. After minimization, a restrained molecular dynamics simulation was initiated at 1000 K with a step size of 1.0 fs for a period of 20 ps. Then, the temperature was gradually lowered to 300 K over 20 ps followed by 2000 cycles of energy minimization.

RESULTS AND DISCUSSION

RNA transcription from the model panhandle RNA

The template activity of the model panhandle RNA was studied in an *in vitro* transcription assay using the micrococcal nuclease treated influenza core as polymerase (22). As shown in Figure 1B, full-length RNA transcript was observed (lane 3) with similar size of the template RNA used (lane 1), and the transcription was specific for the added template RNA (compare with the control lane 2 without RNA). The transcript showed some degree of microheterogeneity in size. Similar observations were made previously with panhandle (22) or 3'-RNA templates (16). This is probably due to heterogeneity of T7 RNA transcripts since similar reactions with pure synthetic RNAs gave full-length transcripts of predominantly one size (16).



Figure 2. (A) Section of a NOESY spectrum in water (9:1 H_2O/D_2O) at 150 ms mixing time for the panhandle RNA in 10 mM sodium phosphate, 0.01 mM EDTA at pH 6.5 and 14°C. The G25 imino to C11 H5 and amino (exposed He and bonded Hb) NOE cross peaks indicate a Watson–Crick base pair between the nucleotides. (B) Section of a DQF-COSY spectrum in D_2O at 30°C showing the H1′–H2′ couplings.

NMR analysis

Exchangeable and non-exchangeable proton resonances were assigned and reported in our previous paper (14). The sequential NOE connectivities of imino-imino and aromatic-anomeric protons suggest that G2~A9 and A13~C16 form double helical stem regions by base pairing with U27~C34 and G21~U24, respectively. The stem contains a G-U pair with little distortion from A-type helical geometry. NOEs observed between the imino proton of G25 and the H5 and amino protons of C11 suggest that the G25 be involved in Watson-Crick base paring with the C11 (Fig. 2A). At pH 6.5, there is no evidence of a protonated A⁺-C pair between A10 and C26. NOE between the H2 proton of A10 and the H1' proton of U27 suggests that A10 is stacked within the helix. A12 exhibits NOE between the A12, H2 proton and the G25, H1' proton suggesting that A12 is stacked within the helix. Figure 2B shows the H1'-H2' region of a DQF-COSY spectrum of the RNA. The observed strong coupling between H1' and H2' protons of U18 and C19 is indicative of the C2'-endo sugar conformation of those nucleotides. Likewise, the medium couplings of the nucleotides A12, A13 and G25 at and near the internal loop region as well as G2 and C34 at the ends are indicative of the flexible sugar conformations of the nucleotides. We introduced a UUCG tetraloop to ensure the stability of the stem and to provide a starting point for NMR assignment. The tetraloop shows the same chemical shift patterns and NOE connectivities as were seen earlier by Varani et al. (23).

Description of the RNA structure

Sixteen refined structures were obtained by the combined use of distance geometry and molecular dynamics. The statistics and

energy analysis of the final structures is given in Table 1. The root mean square deviation (r.m.s.d.) of the converged structures excluding G1, shown superimposed in Figure 3, was 1.18 ± 0.29 Å for all atoms. The G-U base pair between G7 and U29 was not deviated far from the A-form helical structure, but it was associated with partial opening compared to other base pairs.

Table 1. Statistics and energy analysis of the final structures

Number of distance restraints	
Inter residue	111
Intra residue	87
Hydrogen bonds	77
Energy type	Energy value (kcal/mol)
Total energy	-139.5 ± 4.27
Bond energy	22.1 ± 0.23
Valence angle energy	156.2 ± 1.81
Dihedral angle energy	408.5 ± 2.99
Hydrogen bond energy	-25.9 ± 0.51
Non-bond energy	-280.0 ± 3.68
Non-bond repulsion energy	1123.8 ± 5.54
Non-bond dispersion energy	-1403.8 ± 7.27
Coulomb energy	-420.7 ± 2.25

Energy values are expressed as mean values \pm standard deviations.

Figure 4 shows plots of the backbone torsion angles and sugar puckers for the final structures. Base stacking of the bulged nucleotides, A10 and A12 continues at the 5'-side of the panhandle. The C11–A12 step is underwound (helical twist = 8°), while the A10-C11 step is overwound (45°). Distortions caused by the bulged nucleotides force the backbone conformations of A10, A12, A13 and G14 to deviate from the A-helical structure. The backbones of A10 and A12 are in conformational equilibrium between $\alpha^t \gamma^t$ and $\alpha^- \gamma^+$. The dihedral angles of A13 are α^t and γ^t , while those of G14 are α^+ and γ^t . It is assumed that the regular A-form has α^- and γ^+ dihedral angles. In contrast to A10 and A12, the bulged nucleotide C26 is not stacked within the helix. Modeling results suggest that C26 is shifted to the minor groove, perpendicular to the neighboring bases, and stabilized by hydrogen bonds between an amino proton of C26 and N1 of A12 and between an amino proton of G25 and O2 of C26. However, we could not find direct NMR evidence for these hydrogen bonds. Perpendicular orientation of a bulged base in the minor groove has been also observed in other RNA structures such as the HIV-2 TAR-argininamide complex (24). C11 and G25, which have been regarded as a part of internal loop in the previous studies, form a Watson-Crick base pair. This kind of base pair in the internal loop region is not uncommon. An NMR study of IRE RNA has shown a similar result (25). At the protein (IRP) contact site, a G-C Watson-Crick base pair in the internal loop/bulge region was observed by NMR, which has not been suggested by other experimental methods or folding programs (25).

The central region of the panhandle is shown from the perspective of the minor groove in Figure 5. The H1'-H2' coupling constants of the riboses of A12, A13 and G25 are 4~6 Hz (Fig. 2B). Among the final structures, some of A12 adopted C3'-endo and the rest adopted C2'-endo sugar conformations,



Figure 3. Superposition of the 16 converged structures of the panhandle RNA. With the exception of nucleotide G1, the average r.m.s.d. value was 1.18 ± 0.29 Å. The hydrogen atoms and the dangling G1 nucleotide are not shown for clarity.

whereas A13 preferred the C2'-endo sugar conformation. The ribose of G25 was equilibrated between O4'-endo and C3'-exo conformations. This sugar conformation facilitated for G25 and C11 to form a Watson–Crick base pair since the modeling results showed that with C3'-endo conformation, Watson–Crick base pair between G25 and C11 was not possible without a severe structural distortion. G25 is not stacked well on U24 because the bulged nucleotide A12 is partially stacked on G25.

Relation of structure and function

The panhandle structure of influenza virus RNA plays an important role in the regulation of transcription and replication and in the packaging into the virion of the vRNA. Many detailed mutational analyses of various sequences of the panhandle RNAs have been performed in order to better understand the functional importance of specific sequences for both replication and transcription. Accumulated data suggest the importance of base pairs in the panhandle which corresponds to nucleotides 13–16 and 21–24 in the model RNA (Fig. 1A) (10,26,27). In the distal part of the promoter, corresponding to nucleotides 3–12 and 25–23, the nature of nucleotides rather than ability to form base pairs was important (10,26). However, systematic analysis *in vivo* is still lacking especially on the distal part of the promoter, rendering functional implication of the structure difficult.

For technical reasons (see above), some modifications were introduced in the panhandle RNA. These are two additional G

residues at the 5'-end and one C residue at the 3'-end and the U5 to C5 transition. The added nucleotides at the extreme ends and the U to C single base change within the structure may increase the stability of the panhandle. However, the model panhandle promoter was active *in vitro* (Fig. 1). In other studies, a significant level of transcription was observed even when 20–30 nt were added to the 3'-end of the panhandle RNA (28). The effect of the U5 to C5 transition, as introduced in our model RNA template, has not been studied before, although the mutation to A5 significantly inhibited transcription (26).

Individual nucleotides C11 and G25 are crucial for polymerase binding and/or transcriptional activity (9,10,26), although double mutants that restore the Watson-Crick base pair have never been tested before. It is possible that the Watson-Crick base pair formed between C11 and G25 plays an important role for recognition and binding of the RNA polymerase. Modification of A10 nucleotide by diethyl pyrocarbonate or by base substitution mutation decreased polymerase binding and transcriptional activity (26,29). The three-dimensional structure showed that the backbone of A10 was distorted from A-helical geometry (Fig. 4). A hydrogen bond between A12 and C26 was seen in the calculated structures, although this could not be observed directly from the NMR data. The potential importance of this hydrogen bonding was neither tested nor predicted in previous models for influenza transcription, including the RNA-fork model (10,26) or the corkscrew model (4). Although this hydrogen bonding may not be crucial for transcriptional activity (16,26), since it causes



Figure 4. Plots of dihedral angles and sugar puckers for the final structures of the panhandle RNA. Standard deviations from the mean are represented by error bars. Angles of the standard A-form RNA are indicated by horizontal lines.



Figure 5. Central region of the panhandle RNA from the perspective of the minor groove.

distortions of the backbone, the conformation of the backbone might be an important factor for polymerase–promoter interaction in the influenza RNP complex. Watson–Crick base pair was assigned between the A13 and U24. Previous *in vivo* promoter analysis did suggest a Watson–Crick base pair at this position, but with weak stability (27). This may have an implication for the melting of the RNA duplex in the promoter region during transcription, especially in transition of initiation into elongation of RNA synthesis.

It should be remembered, however, that the structure presented here is that of a naked RNA. Since additional interactions would be involved in the promoter–polymerase complex, it is uncertain whether the same structure would be maintained in the RNP complex. Biochemical studies showed that the binding of the polymerase increased the stability of the panhandle (30).

It is largely unknown whether the same polymerase complex without any modification is used for both transcription (mRNA synthesis) and replication (cRNA synthesis). However, the same vRNA promoter is used for the two different modes of RNA synthesis (31–33). It is conceivable that the interaction of polymerase would affect the conformation as well as the stability of the promoter. If so, slight structural change in the promoter and differential interaction between the promoter and polymerase would affect the two closely related, but distinct modes of initiation of RNA synthesis. In this regard, it is worth noting a unique variation of promoter sequence observed in natural influenza isolates, that is U30 to C30 transition at the 3'-end (5,6). It has been shown that this nucleotide is involved in temporal control of the ratio of transcription and replication (34).

Perhaps the availability of sufficient quantities of the three polymerase subunits (PB1, PB2 and PB3) (33) and subsequent structural information of the RNA promoter in the RNP complex would be required for detailed information on structure–function relationships of the promoter. Similar panhandle conformation of RNA promoter is expected to be shared among related viruses, such as Bunyavirus, Hantavirus and a group of viruses in the Arenaviridae (35). As the first three-dimensional structure of RNA promoter ever reported, the present information will be a useful platform for similar studies on viral promoters in these medically important viruses.

ACKNOWLEDGEMENTS

This work was supported by the Korea Science and Engineering foundation through the Center for Molecular Catalysis at Seoul National University (to B.-S.C.) and by the Kuk-je-kong-dong Program of the Ministry of Science and Technology, the Republic of Korea (to C.C.).

REFERENCES

- 1 Palese, P. and Kingsbury, D.W. (eds) (1983) *Genetics of Influenza Virus*. Springer, Vienna.
- 2 Seong, B.L. (1993) Infect. Agents Dis., 2, 17-24.
- 3 Krug,R.M., Alonso-Caplan,F.V., Julkunen,I. and Katz,M.G. (1989) *The Influenza Viruses*. Plenum Press, NY, pp. 89–152.

- 4 Flick, R., Newmann, G., Hoffmann, E., Newmeier, E. and Hobom, G. (1996) *RNA*, **2**, 1046–1057.
- 5 Desselberger, U., Racaniello, V.R., Zazra, J.J. and Palese, P. (1980) Gene, 8, 315–328.
- 6 Robertson, J.S. (1979) Nucleic Acids Res., 6, 3745–3757.
- 7 Hsu,M.T., Parvin,J.D., Gupta,S., Krystal,M. and Palese,P. (1987) Proc. Natl Acad. Sci. USA, 84, 8140–8144.
- 8 Luytjes, W., Krystal, M., Enami, M., Parvin, J.D. and Palese, P. (1989) Cell, 59, 1107–1113.
- 9 Fodor, E., Seong, B.L. and Brownlee, G.G. (1993) J. Gen. Virol., 74, 1327–1333.
- 10 Fodor, E., Pritlove, D.C. and Brownlee, G.G. (1994) J. Virol., 68, 4092–4096.
- 11 Luo, G., Luytjes, W., Enami, M. and Palese, P. (1991) J. Virol., 65, 2861–2867.
- 12 Li,X. and Palese,P. (1994) J. Virol., 68, 1245–1249.
- 13 Tiley,L.S., Hagen,M., Matthews,J.T. and Krystal,M. (1994) J. Virol., 68, 5108–5116.
- 14 Cheong,H.K., Cheong,C. and Choi,B.S. (1996) Nucleic Acids Res., 24, 4197–4201.
- 15 Milligan, J.F., Groebe, D.R., Witherell, G.W. and Uhlenbeck, O.C. (1987) Nucleic Acids Res., 15, 8783–8798.
- 16 Seong, B.L. and Brownlee, G.G. (1992) J. Gen. Virol., 111, 3115-3124.
- 17 Hore, P.J. (1983) J. Magn. Reson., 55, 283-300.
- 18 Marion, D. and Wuthrich, K. (1983) Biochem. Biophys. Res. Commun., 113, 967–974.
- 19 Sklenar, V. and Bax, A. (1987) J. Magn. Reson., 75, 378-383.
- 20 Muller, N., Ernst, R.R. and Wuthrich, K. (1986) J. Am. Chem. Soc., 108, 6482–6492.
- 21 Allain, F.H. and Varani, G. (1995) J. Mol. Biol., 250, 333–353.
- 22 Seong, B.L. and Brownlee, G.G. (1992) Virology, 186, 247-260.
- 23 Varani, G., Cheong, C. and Tinoco, I., Jr (1991) Biochemistry, 30, 3280-3289.
- 24 Brodsky, A.S. and Williamson, J.R. (1997) J. Mol. Biol., 267, 624-639.
- 25 Gdaniec, Z., Sierzputowska-Gracz, H. and Theil, E.C. (1998) Biochemistry, 37, 1505–1512.
- 26 Fodor, E., Pritlove, D.C. and Brownlee, G.G. (1995) J. Virol., 69, 4012–4019.
- 27 Kim,H.J., Fodor,E., Brownlee,G.G. and Seong,B.L. (1997) J. Gen. Virol., 78, 353–357.
- 28 Yamanaka, K., Ogasawara, N., Yoshikawa, H., Ishihama, A. and Nagata, K. (1991) Proc. Natl Acad. Sci. USA, 88, 5369–5373.
- 29 Baudin, F., Bach, C., Cusack, S. and Ruigrok, R.W.H. (1994) *EMBO J.*, 13, 3158–3165.
- 30 Klumpp, K., Ruigrok, R.W.H. and Baudin, F. (1997) EMBO J., 16, 1248–1257.
- 31 McCanley, J.W. and Maky, B.W.J. (1983) *Biochem. J.*, **211**, 281–294.
- 32 Lamb, R.A. and Choppin, P.W. (1983) Annu. Rev. Biochem., 52, 461-506.
- 33 Ishihama, A. and Nagata, K. (1988) CRC Crit. Rev. Biochem., 23, 27-76.
- 34 Lee,K.H. and Seong,B.L. (1998) J. Gen. Virol., **78**, 1923–1934.
- 35 Schmaljohn,C.S. (1996) In Fields,B.N., Knipe,D.M and Howley,P.M. (eds), *Fundamental Virology* (3rd ed.). Ravan Press, New York, pp. 675–690.