Molecular cloning of a full-length cDNA encoding the hemagglutinin-neuraminidase glycoprotein of Sendai virus

Naoyuki Miura, Yoshihiro Nakatani, Masahiro Ishiura+, Tsuyoshi Uchida and Yoshio Okada

Institute for Molecular and Cellular Biology, Osaka University, 1-3 Yamada-Oka, Suita, Osaka 565, and *+National Institute for Basic Biology, Okazaki, Aichi 444, Japan*

Received 21 June 1985

We cloned a full-length complementary DNA for the hemagglutinin-neuraminidase (HN) mRNA of Sendai virus (HVJ) using a synthetic 27-mer as a probe. Nucleotide sequence analysis showed that there is a long open reading frame on the mRNA that encodes a protein of 575 amino acids. The deduced amino acid sequence indicated that only one hydrophobic region sufficiently long to anchor the protein in the membrane and located near the N-terminus (amino acids 35–60). It is suggested that HN protein is oriented with its N-terminus inside the membrane.

Recombinant DNA *cDNA library* *Oligodeoxynucleotide* *Sendai virus* *Hemagglutinin-neuraminidase protein* *Nucleotide sequence*

1. INTRODUCTION

The glycoproteins of Sendai virus, a prototype of the paramyxoviruses, consist of hemagglutinin-neuraminidase (HN) and F (fusion) proteins. These 2 proteins play important roles in viral infection and membrane fusion [1,2]. The F protein is involved in virus penetration, virus-induced cell-cell fusion and hemolysis. The HN protein has hemagglutination and neuraminidase activities and is responsible for adsorption of the virus to the receptor on host cells. Furthermore the HN protein is suggested to play an important role in the initial process of membrane fusion [3,4].

To investigate the role of HN protein at the molecular level, it is necessary to know the structure of the HN protein. However, the complementary DNA (cDNA) for Sendai virus HN protein has not yet been cloned and the primary structure of this protein is not known. Studies of the structure of the paramyxovirus HN protein may provide a means of elucidating the biochemical and biophysical events involved in the fusion process.

Here, a full-length cDNA clone encoding the HN protein was isolated from a cDNA library using as a probe a 27-mer synthetic oligodeoxynucleotide corresponding to a known nucleotide sequence of the Sendai virus HN genome. We present the complete nucleotide sequence of the Sendai virus HN mRNA and the predicted amino acid sequence of the HN protein. During preparation of this manuscript, another laboratory independently reported the nucleotide sequence (having a gap of 11 nucleotides) of the Sendai virus HN genome [5].

2. MATERIALS AND METHODS

2.1. *Oligodeoxynucleotide probe*

A 27-mer, d(TTTTGTAGTGCTACCACTAGGAGAGGT), was obtained by custom synthesis from Takara Shuzo (Kyoto, Japan). This oligonucleotide corresponds to the partially known nucleotide sequence (96–122th) of the Sendai virus HN genome [6].

2.2. *Construction of cDNA library and cloning*

RNA was purified from LLC-MK_2 cells, 20 h after infection with the Z strain of Sendai virus, by sedimentation through CsCl [7], and poly(A)$^+$ RNA was selected on columns of oligodeoxy-

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/85/$3.30

thymidylate-cellulose (type 3, Collaborative Research, MA). The cDNA library was constructed using the pCDV1 vector-primer and the pL1 linker fragment according to Okayama and Berg [8]. The cyclized cDNA preparation was used to transform into *E. coli* strain HB101 described by Hanahan [9]. Colony transfer to nitrocellulose filters was performed as described by Hanahan and Meselson [10]. The filters were prehybridized for 3 h at 65°C in 6 × NET/1 × Denhardt/0.1% SDS (1 × NET: 0.15 M NaCl/1 mM EDTA/0.03 M Tris-HCl, pH 8.0. 1 × Denhardt: 0.02% Ficoll/0.02% polyvinylpyrollidone/0.02% bovine serum albumin). Hybridization with (5′-^{32}P)-labeled oligonucleotide (3–5 × 10^5 cpm/ml) was done at 55°C for 16 h in 6 × NET/1 × Denhardt/0.1% SDS. The specific activity of the probe was 1–3 × 10^6 cpm/pmol. The filters were washed 4 times at room temperature and then once at 55°C for 1 min in 6 × SSC (1 × SSC: 0.15 M NaCl/15 mM sodium citrate, pH 7.0). The dried filters were exposed at −70°C for 12 h to Kodak XS-1 film with a Lightening Plus intensifying screen.

2.3. *Sequence analysis*

DNA fragments were subcloned in M13 vector mp10 and mp11. The nucleotide sequences were determined by the dideoxy method [11].

3. RESULTS

3.1. *Screening of Sendai HN cDNA*

mRNA from infected cells was used as a template and a cDNA library was constructed as described in section 2. We screened 2000 ampicillin-resistant transformants with (5′-^{32}P)-labeled probe and obtained one clone that had the longest insert, pCD-HN12. This clone was analyzed further.

3.2. *Nucleotide sequence of HN cDNA*

The strategy of nucleotide sequences analysis is shown in fig.1. Fig.2 shows the nucleotide sequence of cloned cDNA for HN mRNA. The open reading frame starting from a methionine at the 57th nucleotide from the 5′-terminus codes a peptide consisting of 575 amino acids of M_r 63 433. This value is consistent with the unglycosylated HN protein in vivo (63 kDa) [12]. The sequence 5′-AGGGTGAAAG-3′ at the 5′-terminus of our HN cDNA corresponds to the 10 base long consensus start sequence of the HN gene, 3′-UCCCACUUUC-5′,reported by Gupta and Kingsbury [6]. On the other hand, the sequence 5′-ATTAAG-3′ next to the poly(A) sequence corresponds to the end sequence of the HN gene, 3′-UAAUUCUUUUU-5′ [6]. Thus pCD-NH12 is found to be a full-length cDNA for HN mRNA and there are untranslated regions of 56 and 102 nucleotides at the 5′- and 3′-termini, respectively, of the mRNA.

Blumberg et al. [5] reported the nucleotide sequence of the Sendai virus HN genome. We find close agreement between the sequence in fig.2 and that published by Blumberg et al., but we note the following differences: (i) our cDNA clone is full-length, whereas their overlapping clones have a

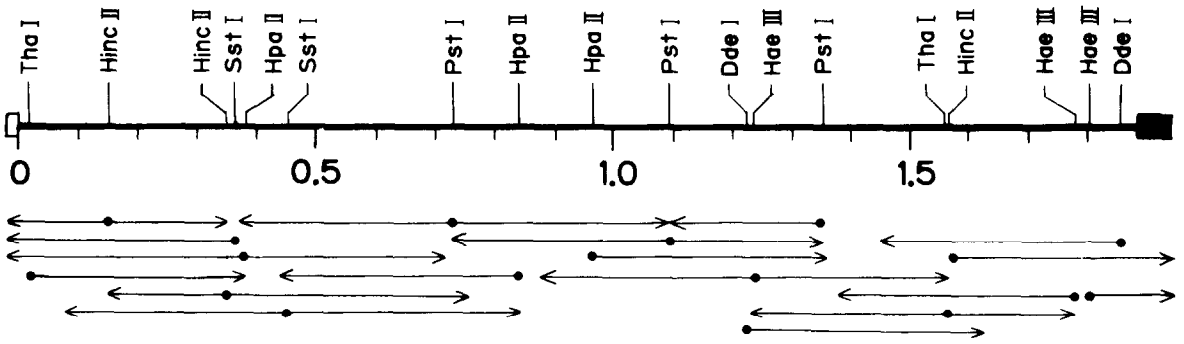

Fig.1. Restriction cleavage maps and sequencing strategy for pCD-HN12. Index numbers represent kilobases from the 5′-end of the gene (mRNA). The direction of each sequence determination is shown by horizontal arrows, starting at appropriate cleavage sistes. (□) Poly(dG) tail and (■) poly(A) sequence.

```
1-AGGGTGAAAGTGAGGTCGCGCGGTACTTTAGCTTTCACCTCAAACAAGCACAGATC ATG GAT GGT GAT AGG GGC AAA CGT GAC TCG TAC TGG TCT
                                                           met asp gly asp arg gly lys arg asp ser tyr trp ser -13

 96- ACT TCT CCT AGT GGT AGC ACT ACA AAA TTA GCA TCA GGT TGG GAG AGG TCA AGT AAA GTT GAC ACA TGG TTG CTG ATT CTC TCA
     thr ser pro ser gly ser thr thr lys leu ala ser gly trp glu arg ser ser lys val asp thr trp leu leu ile leu ser -41

180- TTC ACC CAG TGG GCT TTG TCA ATT GCC ACA GTG ATC ATC TGT ATC ATA ATT TCT GCT AGA CAA GGG TAT AGT ATG AAA GAG TAC
     phe thr gln trp ala leu ser ile ala thr val ile ile cys ile ile ile ser ala arg gln gly tyr ser met lys glu tyr -69

264- TCA ATG ACT GTA GAG GCA TTG AAC ATG AGC AGC AGG GAG GTG AAA GAG TCA CTT ACC AGT CTA ATA AGG CAA GAG GTT ATA GCA
     ser met thr val glu ala leu asn met ser ser arg glu val lys glu ser leu thr ser leu ile arg gln glu val ile ala -97

348- AGG GCT GTC AAC ATT CAG AGC TCT GTG CAA ACC GGA ATC CCA GTC TTG TTG AAC AAA AAC AGC AGG GAT GTC ATC CAG ATG ATT
     arg ala val asn ile gln ser ser val gln thr gly ile pro val leu leu asn lys asn ser arg asp val ile gln met ile -125

432- GAT AAG TCG TGC AGC AGA CAA GAG CTC ACT CAG CAC TGT GAG AGT ACG ATC GCA GTC CAC CAT GCC GAG GGA ATT GCC CCA CTT
     asp lys ser cys ser arg gln glu leu thr gln his cys glu ser thr ile ala val his his ala glu gly ile ala pro leu -153

516- GAG CCA CAT AGT TTC TGG AGA TGC CCT GTC GGA GAA CCG TAT CTT AGC TCA GAT CCT GAA ATC TCA TTG CTG CCT GGT CCG AGC
     glu pro his ser phe trp arg cys pro val gly glu pro tyr leu ser ser asp pro glu ile ser leu leu pro gly pro ser -181

600- TTG TTA TCT GGT TCT ACA ACG ATC TCT GGA TGT GTT AGG CTC CCT TCA CTC TCA ATT GGC GAG GCA ATC TAT GCC TAT TCA TCA
     leu leu ser gly ser thr thr ile ser gly cys val arg leu pro ser leu ser ile gly glu ala ile tyr ala tyr ser ser -209

684- AAT CTC ATT ACA CAA GGT TGT GCT GAC ATA GGG AAA TCA TAT CAG GTC CTG CAG CTA GGG TAC ATA TCA CTC AAT TCA GAT ATG
     asn leu ile thr gln gly cys ala asp ile gly lys ser tyr gln val leu gln leu gly tyr ile ser leu asn ser asp met -237

768- ATC CCT GAT CTT AAC CCC GTA GTG TCC CAC ACT TAT GAC ATC AAC GAC AAT CGG AAA TCA TGC TCT GTG GTG GCA ACC GGG ACT
     ile pro asp leu asn pro val val ser his thr tyr asp ile asn asp asn arg lys ser cys ser val val ala thr gly thr -265

852- AGG GGT TAT CAG CTT TGC TCC ATG CCG ACT GTA GAC GAA AGA ACC GAC TAC TCT AGT GAT GGT ATC GAG GAT CTG GTC CTT GAT
     arg gly tyr gln leu cys ser met pro thr val asp glu arg thr asp tyr ser ser asp gly ile glu asp leu val leu asp -293

936- GTC CTG GAT CTC AAA GGG AGA ACT AAG TCT CAC CGG TAT CGC AAC AGC GAG GTA GAT CTT GAT CAC CCG TTC TCT GCA CTA TAC
     val leu asp leu lys gly arg thr lys ser his arg tyr arg asn ser glu val asp leu asp his pro phe ser ala leu tyr -321

1020- CCC AGT GTA GGC AAC GGC ATT GCA ACA GAA GGC TCA TTG ATA TTT CTT GGG TAT GGT GGA CTA ACC ACC CCT CTG CAG GGT GAT
      pro ser val gly asn gly ile ala thr glu gly ser leu ile phe leu gly tyr gly gly leu thr thr pro leu gln gly asp -349

1104- ACA AAA TGT AGG ACC CAA GGA TGC CAA CAG GTG TCG CAA GAC ACA TGC AAT GAG GCT CTG AAA ATT ACA TGG CTA GGA GGG AAA
      thr lys cys arg thr gln gly cys gln gln val ser gln asp thr cys asn glu ala leu lys ile thr trp leu gly gly lys -377

1188- CAG GTG GTC AGC GTG ATC ATC CAG GTC AAT GAC TAT CTC TCA GAG AGG CCA AAG ATA AGA GTC ACA ACC ATT CCA ATC ACT CAA
      gln val val ser val ile ile gln val asn asp thr leu ser glu arg pro lys ile arg val thr thr ile pro ile thr gln -405

1272- AAC TAT CTC GGG GCG GAA GGT AGA TTA TTA AAA TTG GGT GAT CGG GTG TAC ATC TAT ACA AGA TCA TCA GGC TGG CAC TCT CAA
      asn tyr leu gly ala glu gly arg leu leu lys leu gly asp arg val tyr ile tyr thr arg ser ser gly trp his ser gln -433

1356- CTG CAG ATA GGA GTA CTT GAT GTC AGC CAC CCT TTG ACT ATC AAC TGG ACA CCT CAT GAA GCC TTG TCT AGA CCA GGA AAT AAA
      leu gln ile gly val leu asp val ser his pro leu thr ile asn trp thr pro his glu ala leu ser arg pro gly asn lys -461

1440- GAG TGC AAT TGG TAC AAT AAG TGT CCG AAG GAA TGC ATA TCA GGC GTA TAC ACT GAT GCT TAT CCA TTG TCC CCT GAT GCA GCT
      glu cys asn trp tyr asn lys cys pro lys glu cys ile ser gly val tyr thr asp ala tyr pro leu ser pro asp ala ala -489

1524- AAC GTC GCT ACC GTC ACG CTA TAT GCC AAT ACA TCG CGT GTC AAC CCA ACA ATC ATG TAT TCT AAC ACT ACT AAC ATT ATA AAT
      asn val ala thr val thr leu tyr ala asn thr ser arg val asn pro thr ile met tyr ser asn thr thr asn ile ile asn -517

1608- ATG TTA AGG ATA AAG GAT GTT CAA TTA GAG GCT GCA TAT ACC ACG ACA TCG TGT ATC ACG CAT TTT GGT AAA GGC TAC TGC TTT
      met leu arg ile lys asp val gln leu glu ala ala tyr thr thr thr ser cys ile thr his phe gly lys gly tyr cys phe -545

1692- CAC ATC ATC GAG ATC AAT CAG AAG AGC CTG AAT ACC TTA CAG CCG ATG CTC TTT AAG ACT AGC ATC CCT AAA TTA TGC AAG GCC
      his ile ile glu ile asn gln lys ser leu asn thr leu gln pro met leu phe lys thr ser ile pro lys leu cys lys ala -573

1776- GAG TCT TAAATTTAACTGACTAGCAGGCTTGTCGGCCTTGCTGACACTAGAGTCATCTCCGAACATCCACAATATCTCTCAGTCTCTTACGTCTCTCACAGTATTAAG -1883
      glu ser ***
```

Fig.2. Nucleotide sequence of the cDNA for HN mRNA and the predicted amino acid sequence. The N-terminal hydrophobic domain is doubly underlined. (□) Positions of potential asparagine-linked acceptor sites for carbohydrate.

gap of 11 nucleotides (45–55th nucleotides); (ii) 20 nucleotide substitutions, at positions 40, 45–47, 98, 161, 209, 253, 344, 573, 589, 768, 956, 1079, 1094, 1198, 1210, 1459, 1639 and 1825; (iii) between nucleotides 1653 and 1654, they reported a 3-nucleotide insertion, 5′-TTT, which results in an increase in the total nucleotide length of the HN mRNA and the predicted amino acid number of the HN protein. We found a total of 1883 nucleotides (excluding the poly(A) sequence) and

575 amino acids, whereas they reported a total of 1886 nucleotides and 576 amino acids.

Incidentally, in the sequence (96–122th) that corresponds to the 27-mer probe, there is one base substitution (98th nucleotide is T instead of C). this mismatch did not inhibit screening the colonies by use of this probe.

4. DISCUSSION

The amino acid sequence deduced from the nucleotide sequence is shown in fig.2.

Glycosylation of the HN protein is inhibited in the presence of tunicamycin, indicating that carbohydrate chains are attached through N-linkage to asparagine residues. The amino acid sequence shown in fig.2 contains 5 potential acceptor sites for N-linked carbohydrate (Asn-X-Thr/Ser).

The biosynthesis and membrane orientation of glycoprotein are mediated by a characteristic hydrophobic 'signal' sequence in the polypeptide chain [12]. To search for possible hydrophobic signals in the sequence of HN protein, a plot of local hydrophilicity vs sequence position was prepared (fig.3). The N-terminus of HN protein (amino acids 35–60) is the only strongly hydrophobic region that is sufficiently long for attachment of the protein to the membrane. It seems likely that HN contains an extended signal sequence which both transfers the protein across the membrane and remains in the bilayer to anchor the protein with the first 34 N-terminal amino acids protruding on the cytoplasmic side of the membrane as a hydrophilic tail. N-terminal membrane anchor structures are also exhibited in HN protein of Simian virus 5 [14], NA protein of influenza virus [15] and transferrin receptor [16].

ACKNOWLEDGEMENT

This work was supported by a research grant from the Ministry of Education of Japan.

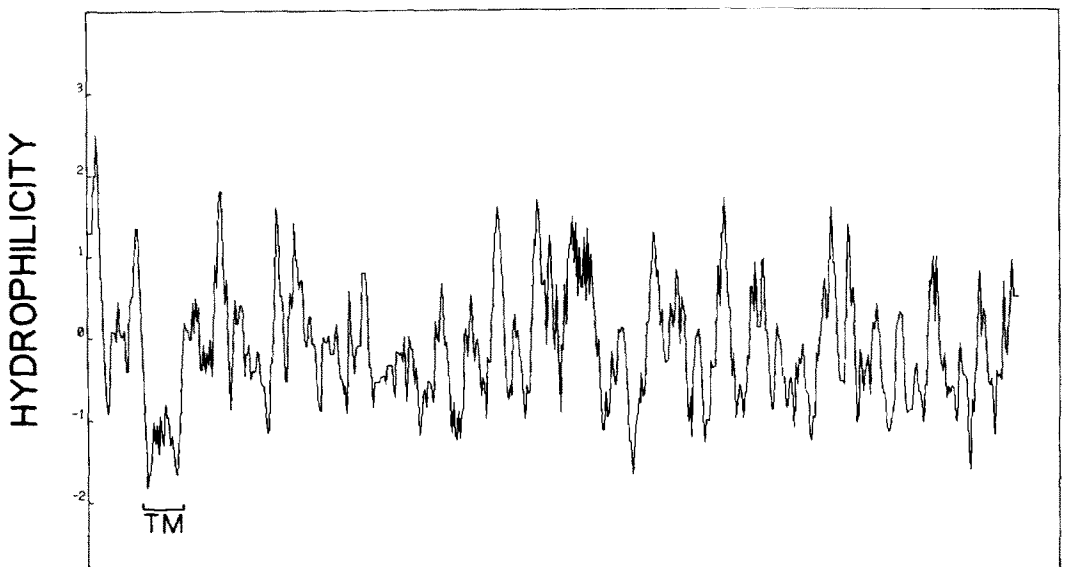

Fig.3. Hydrophilicity profile of the unprocessed HN protein. A window of 6 amino acids was used to calculate local hydrophilicity of each position according to Hopp and Woods [17]. Numbers under the horizontal axis are positions from N-terminal methionine. TM, the proposed site of the transmembrane anchor.

REFERENCES

[1] Choppin, P.W. and Scheid, A. (1980) Rev. Infect. Dis. 2, 40–61.
[2] Homma, M. and Ohuchi, M. (1973) J. Virol. 12, 1457–1465.
[3] Miura, N., Uchida, T. and Okada, Y. (1982) Exp. Cell Res. 141, 409–420.
[4] Uchida, T., Ueda, S., Nakanishi, M., Miura, N. and Okada, Y. (1984) Exp. Cell Res. 152, 313–321.
[5] Blumberg, B., Giorgi, C., Roux, L., Rjau, R., Dowling, P., Chollet, A. and Kolakofsky, D. (1985) Cell 41, 269–278.
[6] Gupta, K.C. and Kingsbury, D.W. (1984) Nucleic Acids Res. 12, 3829–3841.
[7] Chirgwin, J.M., Przybyra, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294–5299.
[8] Okayama, H. and Berg, P. (1983) Mol. Cell. Biol. 3, 280–289.
[9] Hanahan, D. (1983) J. Mol. Biol. 166, 557–580.
[10] Hanahan, D. and Meselson, M. (1983) Methods Enzymol. 100, 333–342.
[11] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[12] Nakamura, K., Homma, M. and Compans, R.W. (1982) Virology 119, 474–487.
[13] Blobel, G. (1980) Proc. Natl. Acad. Sci. USA 77, 1496–1500.
[14] Hiebert, S.W., Paterson, R.G. and Lamb, R.A. (1985) J. Virol. 54, 1–6.
[15] Fields, S., Winter, G. and Brownlee, G.G. (1981) Nature 290, 213–217.
[16] Schneider, C., Owen, M.J., Banville, D. and Williams, J.G. (1984) Nature 311, 675–678.
[17] Hopp, T.P. and Woods, K.R. (1981) Proc. Natl. Acad. Sci. USA 78, 3824–3828.