JOURNAL OF VIROLOGY, Feb. 1982, p. 730–734
0022-538X/82/020730-05$02.00/0

Vol. 41, No. 2

Complete Nucleotide Sequence of the Neuraminidase Gene of Human Influenza Virus A/WSN/33

ALAN L. HITI AND DEBI P. NAYAK*

Department of Microbiology and Immunology, University of California at Los Angeles School of Medicine, Los Angeles, California 90024

Received 22 July 1981/Accepted 8 September 1981

The complete nucleotide sequence of the neuraminidase (NA) gene of WSN/33 (H1N1) virus was determined. The entire sequence was derived from the insert of cDNA clones, except the last 20 nucleotides, which were determined by primer extension. The WSN NA gene contained 1,409 nucleotides beginning at the 5′ end (sense strand), with an untranslated region of 19 nucleotides followed by 1,359 nucleotides coding for 453 amino acids and finally ending with a 31-nucleotide sequence of untranslated region at the 3′ termini. The amino acid sequence of WSN NA, as deduced from the DNA sequence, showed the presence of a stretch of 29 amino acids (7 to 35) enriched in hydrophobic amino acids, which may anchor the protein into the viral or cellular membrane. When compared with the PR8 NA sequence, WSN NA appeared to possess a similar structure, including the identical location of all cysteine and proline residues. However, WSN NA contained only three of the five potential glycosylation sites present in PR8 NA. Additionally, WSN NA contained a substitution of a five-amino acid sequence for a six-amino acid sequence in PR8 NA. The possible significance of these sequence changes in the primary structure of WSN NA in the unique role of WSN NA as a virulence factor in mouse brain and MDBK cells is discussed.

Influenza is an important disease, against which there is as yet no effective prophylaxis. Because of the antigenic shift and drift of the two surface antigens, hemagglutinin (HA) and neuraminidase (NA), influenza virus can evade the immunity against the existing virus and cause frequent epidemics and infrequent worldwide pandemics. Recently, a great deal of attention has been paid to the study of HA. For example, the complete nucleotide sequences of the HA gene of each human subtype (H1, H2, and H3) (9, 11, 15, 25), as well as the complete amino acid sequence of the HA protein of the H3 subtype (22) and the partial amino acid sequence of the H2 subtype (23), have been determined. DNA clones of influenza HA have been expressed in both eucaryotic and procaryotic systems (3a, 4, 10, 13). Monoclonal antibody studies have identified the major antigenic epitopes involved in drift (8). Finally, a three-dimensional structure of HA has helped enormously in understanding the topological relationship among the different domains of HA protein involved in antigenic drift, receptor binding, and the infectious process (24).

However, a similar detailed understanding of the structure and function of NA, the other envelope protein which also undergoes antigenic drift and shift and plays a critical role in the infectious process, virulence, and cell-virus fusion (12, 18, 21) is lacking. Influenza NA is a glycoprotein with a molecular weight ranging from 48,000 to 63,000, including 20% carbohydrate, and is present as a homotetramer spike on the viral envelope (2). NA spikes constitute about 10% of the total envelope spikes. In an effort to contribute to the understanding of the nature of influenza NA, including its biological role and the nature of antigenic drift and shift, we cloned the WSN NA gene with recombinant DNA technology and determined its complete nucleic acid sequence. WSN NA was further considered important because of its contribution as a virulence factor to the growth of WSN virus in MDBK and neuroblastoma cells (16, 18) and neurovirulence in mice (21). In this report, we present the complete nucleotide sequence and the deduced amino acid sequence of WSN NA.

The procedures involved in cloning influenza viral RNA have been described previously (3). Briefly, viral RNA (vRNA) was isolated from gradient-purified virus particles. A synthetic dodecamer primer (5′ AGCAAAAGCAGG 3′) was used to make cDNA which was size selected and converted into double-stranded DNA with the Klenow fragment of *Escherichia coli* DNA polymerase. Double-stranded DNA was then treated with S1 nuclease, size selected in neutral agar-

ose gels, and cloned into the *Pst*I site of the pBR322 plasmid with G:C linkers. The clones were then characterized for insert size, cleavage by restriction enzymes, and hybridization to a specific vRNA probe. With this procedure, a number of clones were shown to contain a nearly full-size insert of the WSN NA gene. Two of these clones, 3-1 and 3-14, were subjected to detailed restriction analysis and sequencing with the Maxam and Gilbert procedure of DNA sequencing (14). Twenty nucleotides appeared to be missing from the 5′ termini of vRNA and were determined by the reverse transcription of WSN NA vRNA by extending a primer isolated from the WSN NA clone (11).

Figure 1 shows the important restriction sites and the strategy used to determine the sequence of the WSN NA insert. We sequenced through all restriction sites. Nearly the entire sequence was confirmed by overlapping gels, and more than 50% was confirmed by sequencing both DNA strands. The insert was checked for methylated cytosine with *Bst*N1, an isochizomer of *Eco*RII, and was found to be negative. Thus, no ambiguity in the sequence was left unresolved. Clone 3-1 was missing an A residue around positions 35 to 39. The correct sequence was obtained by sequencing clone 3-14 since the loss of an A residue would switch the coding potential to a different reading frame with numerous stop codons and would be completely different from that of PR8 NA. We have also observed an occasional omission of an A residue in some of the DNA clones obtained from defective interfering vRNA segments (D. P. Nayak, unpublished data).

Figure 2 shows the complete nucleotide sequence of cDNA and the deduced amino acid sequence of WSN NA. The amino acid sequence

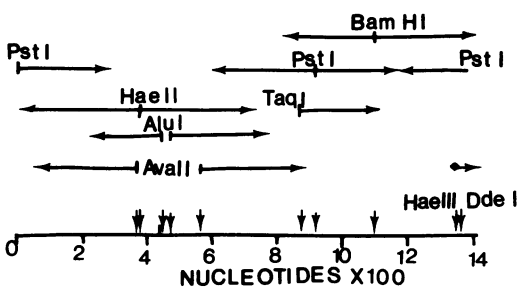

FIG. 1. Strategy used in sequencing cloned WSN NA. DNA with the 5′ end of the sense strand is on the left. Arrows (← →) show composite sequence information obtained from multiple gels. The restriction sites (↓) used for sequencing are *Pst*I (5′ and 3′ ends and nucleotide 924), *Bam*HI (nucleotide 1099), *Hae*II (nucleotide 383), *Taq*I (nucleotide 869), *Alu*I (nucleotides 452 and 469), and *Ava*II (nucleotide 363). The primer (◆) from *Hae*III (nucleotide 1344) to *Dde*I (nucleotide 1356) was used for extension.

of PR8 NA recently determined from the DNA sequence (5) is also included for comparison. The noncoding region at the 5′ end of the sense-strand DNA consisted of 19 nucleotides compared with 20 nucleotides in PR8 NA (5) and was identical to that of NWS/33 (1). A single change at the common 5′-end sequence (position 4, G → A) was observed and was due to the primer used in cDNA synthesis. The open reading frame started at positions 20 to 22 (ATG) and terminated at positions 1382 to 1384 (TAG). This was followed by a sequence of 28 nucleotides at the 3′ end of cRNA, which also contained a polyadenylic acid addition site (AAAAAA) at positions 1392 to 1397 (17). This 3′ untranslated sequence was different by a single nucleotide from that of PR8 NA at position 1386 (T → C). The coding region consisted of an open frame of 1,359 nucleotides (positions 20 to 1381) coding for 453 amino acids, compared with 1,362 nucleotides coding for 454 amino acids in PR8 NA.

WSN NA contained only three potential glycosylation sites, compared with five potential glycosylation sites in PR8 NA (Fig. 3). Of these, the sites beginning at amino acids 44, 72, and 220 were identical. The loss of the glycosylation site at amino acid 131 was due to the conversion of Asn → Arg, involving a single-base change (G → A) in position 2 of the codon. The loss of the glycosylation site at amino acid 58 was due to a change in this part of the molecule, involving a stretch of 15 nucleotides in WSN NA in place of 18 nucleotides in PR8 NA, beginning at nucleotide 191 (see below).

Like PR8 NA, WSN NA contained one stretch of a major hydrophobic region consisting of 29 amino acids (amino acids 7 to 35) that are likely to be embedded in the viral (or cellular) lipid bilayer (5). Two changes in this region (Leu → Met, Leu → Ile) did not alter the hydrophobicity of WSN NA. The first six amino acids (amino acids 1 to 6) that are believed to be involved in an interaction with a hydrophilic milieu inside the membrane were identical to those of PR8 NA (5). The other major uncharged stretch of the molecule consisted of 16 amino acids (amino acids 420 to 435) but contained fewer hydrophobic amino acids and, therefore, is unlikely to be anchored in the membrane.

When compared with PR8 NA, WSN NA contained a total of 32 amino acid changes (excluding the 5 amino acid changes beginning at amino acid 58) scattered throughout the molecule. As in PR8 NA, these changes did not greatly alter the property of the molecule. For example, 21 amino acids were replaced by amino acids with similar properties without affecting the net charge. Six neutral or basic amino acids were replaced by acidic amino acids, and five neutral or acidic amino acids were changed to

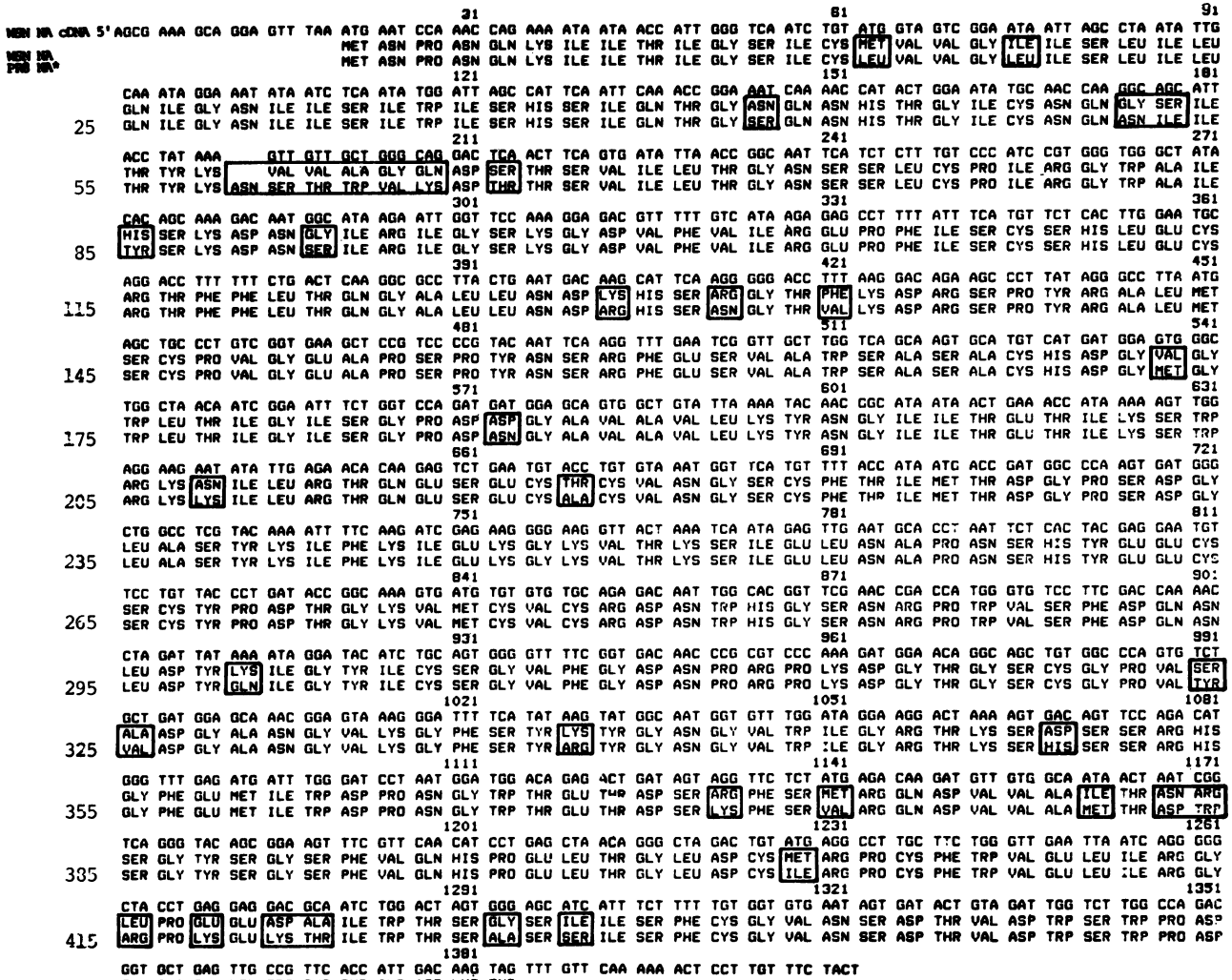

FIG. 2. Nucleotide sequence of the sense strand of WSN NA DNA and the deduced amino acid sequence. The amino acid sequence of PR8 NA (5) is also included for comparison. The boxed amino acid residues show the differences between WSN NA and PR8 NA.

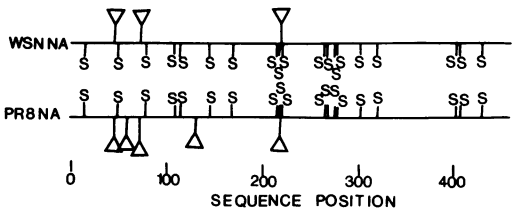

FIG. 3. Positions of 19 cysteine residues (S) and the potential glycosylation sites (△) of WSN NA and PR8 NA.

basic amino acids. Again, none of these changes were clustered; instead, they were scattered throughout the molecule. Therefore, the basic structures of WSN NA and PR8 NA are essentially the same. This was further confirmed by the fact that no change was observed in any of the 19 cysteine or 21 proline residues which are believed to be important in maintaining the structure of a protein molecule. In fact, the overall amino acid sequence, as well as the nucleotide sequence of both WSN NA and PR8 NA (both belonging to the N1 subtype), is very similar despite the fact that these viruses were independently isolated nearly 50 years ago and since then may have been subjected to different growth conditions and selective pressure. Only 65 nucleotides (4.7%) and 32 amino acid residues (7.1%) were different in WSN NA and PR8 NA. Nearly half of the base substitution (30 of 65) did not cause any amino acid alteration. The majority of amino acid changes (29 of 32, or 91%) were due to a single-base substitution, and only three (9%) required a two-base substitution.

PR8/34 and WS/33 (the parent virus of WSN) are both the H1N1 subtype, isolated from human influenza by the inoculation of ferrets and possessing similar biological properties, including growth behavior and host range (6, 19). However, WSN virus was deliberately selected by passaging WS virus many times in mouse brain, whereas similar attempts to select neurovirulent strains of PR8 virus have failed (7). The WSN virus selected by this procedure acquired new properties, despite retaining the similar antigenic determinants of the parent virus. For example, WSN is the only virus that can grow in MDBK cells as well as cause neurovirulence in mice (18, 21). WSN NA has been shown to play an important role in the virulence of WSN virus in both of these host systems (16, 18, 21). A comparison of the primary structure of WSN NA and PR8 NA showed an important difference at amino acids 58 to 63. In this region, there was a substitution of six amino acid residues in PR8 NA by a completely different five-amino acid residue in WSN NA. These amino acid changes cannot be explained by the usual single-base, or even the infrequent two-base, mutation observed in the rest of the WSN NA gene, but may possibly require substitution of a stretch of the nucleic acid involving deletion and insertion. However, we have no data as yet for either the source of this sequence or the mechanism involved in substitution. Alternatively, this site may be a "hot spot" of mutation which can only be ascertained by determining the sequence of a number of NA genes. Furthermore, we cannot rule out the possibility that this sequence is present in the parent WS virus (19). A comparative sequence determination of WS virus as well as NWS, another neurotropic variant of WS virus (20), may help to determine whether this sequence was acquired later.

The data presented here show that the primary structures of WSN NA and PR8 NA are essentially the same except for a major difference near amino acids 58 to 63. In the absence of a three-dimensional structure of NA, it is difficult to predict the role of these residues in the specificity of WSN NA or to define the structural domains and antigenic determinants of NA. However, since it is now possible to express the cloned influenza HA gene in eucaryotic systems (10, 13), similar experiments, including a specific base substitution in this region, may elucidate the role of these and other sequences in the function and specificity of influenza NA.

We thank Keiichi Itakura for providing us with the synthetic primers and Alan R. Davis for his invaluable aid and expertise.
This work was supported by a grant from Genentech, Inc.

LITERATURE CITED

1. **Blok, J., and G. M. Air.** 1980. Comparative nucleotide sequences at the 3′ end of neuraminidase gene of eleven influenza type A viruses. Virology **107**:50–60.
2. **Bucher, D., and P. Palese.** 1975. The biologically active proteins of influenza virus: neuraminidase, p. 83–123. *In* E. Kilbourne (ed.), The influenza viruses and influenza. Academic Press, Inc., New York.
3. **Davis, A. R., A. L. Hiti, and D. P. Nayak.** 1980. Construction and characterization of a bacterial clone containing the hemagglutinin gene of the WSN strain (H0N1) of influenza virus. Gene **10**:205–218.

3a. **Davis, A. R., D. P. Nayak, M. Ueda, A. L. Hiti, D. Dowbenko, and G. D. Kleid.** 1981. Expression of antigenic determinants of the hemagglutinin gene of a human influenza virus in *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. **78**:5376–5380.

4. **Davis, A. R., D. P. Nayak, M. Ueda, A. L. Hiti, D. G. Kleid, and D. Dowbenko.** 1981. Structure of hemagglutinin gene of H0N1 strain of human influenza virus and its expression in *Escherichia coli*. *In* D. P. Nayak (ed.), Genetic variation among influenza viruses. Academic Press, Inc., New York.
5. **Fields, S., G. Winter, and G. G. Brownlee.** 1981. Structure of the neuraminidase gene from human influenza virus A/PR/8/34. Nature (London) **290**:213–217.
6. **Francis, T.** 1934. Transmission of influenza by filtrable virus. Science **80**:457–459.
7. **Francis, T., and A. E. Moore.** 1940. A study of the neurotropic tendency in strains of the virus of epidemic influenza. J. Exp. Med. **72**:717–728.
8. **Gerhard, W., J. Yewdell, M. E. Frankel, and R. Webster.**

1981. Antigenic structure of influenza virus hemagglutinin defined by hybridoma antibodies. Nature (London) **290**:713–717.

9. **Gething, M. J., J. Bye, J. Skehel, and M. Waterfield.** 1980. Cloning and DNA sequence of double-stranded copies of hemagglutinin genes from H2 and H3 strains elucidates antigenic shift and drift in human influenza virus. Nature (London) **287**:301–306.
10. **Hartman, J., D. P. Nayak, and G. Fareed.** 1981. Expression of the human influenza hemagglutinin using simian virus 40 vectors. *In* D. P. Nayak (ed.), Genetic variation among influenza viruses. Academic Press, Inc., New York.
11. **Hiti, A. L., A. R. Davis, and D. P. Nayak.** 1981. Complete sequence analysis shows that hemagglutinins of the H0 and H2 subtypes of human influenza viruses are closely related. Virology **111**:113–124.
12. **Huang, R. T. C., R. Rott, K. Wahn, H.-D. Klenk, and T. Kohama.** 1980. The function of the neuraminidase in membrane fusion induced by myxoviruses. Virology **107**:313–319.
13. **Lai, C.-J., M. Sveda, L. J. Markoff, B.-C. Lin, R. A. Lamb, and J. A. Berndt.** 1981. Nonviral sequences in complete influenza viral DNA clones and functional expression of cloned DNA coding for the hemagglutinin. *In* D. P. Nayak (ed.), Genetic variation among influenza viruses. Academic Press, Inc., New York.
14. **Maxam, A. M., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. U.S.A. **74**:560–564.
15. **Min Jou, W., M. Verhoeyen, R. Devos, E. Saman, R. Fang, D. Huylebroeck, W. Fiers, G. Threlfall, C. Barber, N. Carey, and S. Emtage.** 1980. Complete structure of the hemagglutinin gene from the human influenza A/Victoria/3/75 (H3N2) strain as determined from cloned DNA. Cell **19**:683–696.
16. **Nakajima, S., and A. Sugiura.** 1980. Neurovirulence of influenza virus in mice. II. Mechanism of virulence as studied in a neuroblastoma cell line. Virology **101**:450–457.
17. **Robertson, J. S., M. Schubert, and R. A. Lazzarini.** 1981. Polyadenylation sites for influenza virus mRNA. J. Virol. **38**:157–163.
18. **Schulman, J. L., and P. Palese.** 1977. Virulence factors of influenza A viruses: WSN virus neuraminidase required for plaque production in MDBK cells. J. Virol. **24**:170–176.
19. **Smith, W., C. H. Andrews, and P. P. Laidlaw.** 1933. A virus obtained from influenza patients. Lancet **i**:66–68.
20. **Stuart-Harris, C. H.** 1939. A neurotropic strain of human influenza virus. Lancet **i**:497–499.
21. **Sugiura, A., and M. Ueda.** 1980. Neurovirulence of influenza viruses in mice. I. Neurovirulence of recombinants between virulent and avirulent virus strains. Virology **101**:440–449.
22. **Ward, C. W., and T. A. Dopheide.** 1980. The Hong Kong (H3) hemagglutinin. Complete amino acid sequence and oligosaccharide distribution for the heavy chain of A/Memphis/102/72, p. 27–38. *In* G. Laver and G. Air (ed.), Structure and variation of influenza viruses. Elsevier/North-Holland Publishing Co., Amsterdam.
23. **Waterfield, M. D., K. Espelie, K. Elder, and J. J. Skehel.** 1979. Structure of the haemagglutinin of influenza virus. Br. Med. Bull. **35**:57–63.
24. **Wilson, I. A., J. J. Skehel, and D. C. Wiley.** 1981. The hemagglutinin membrane glycoprotein of influenza virus: structure at 3A resolution. Nature (London) **289**:366–373.
25. **Winter, G., S. Fields, and G. G. Brownlee.** 1981. Nucleotide sequence of the hemagglutinin gene of a human influenza virus H1 subtype. Nature (London) **292**:72–75.