JOURNAL OF VIROLOGY, May 1984, p. 654–656
0022-538X/84/050654-03$02.00/0
Copyright © 1984, American Society for Microbiology
Vol. 50, No. 2

Nucleotide Sequence of the Influenza Virus A/USSR/90/77 Neuraminidase Gene

PATRICK CONCANNON,[1*] CHRISTOPHER J. KWOLEK,[1†] AND WINSTON A. SALSER[1,2]

Departments of Biology[1] and Molecular Biology,[2] University of California at Los Angeles, Los Angeles, California 90024

Received 10 September 1983/Accepted 30 January 1984

The complete nucleotide sequence of the N1 neuraminidase gene of influenza virus A/USSR/90/77 was determined. Comparison of its predicted amino acid sequence with other N1 and N2 neuraminidases indicates that the N1 neuraminidases share most of the antigenic determinants mapped on the N2 neuraminidase but display at least one additional potentially antigenic region probably as a result of intersubtypic differences in glycosylation.

Influenza A virus displays two virus-specified glycoprotein antigens, hemagglutinin (HA), the major antigen, and neuraminidase (NA), a minor antigen. The capability of the virus to strongly vary these antigens by at least two general mechanisms allows the virus to continue to circulate in human and animal populations despite the ability of the hosts to mount immune responses against these viral antigens. In an attempt to better understand this variation, we undertook the determination of the nucleotide sequences of the genes encoding the surface antigens of influenza virus A/USSR/90/77.

We report here the nucleotide sequence of the A/USSR/90/77 NA gene. The coding sequence is 1,413 nucleotides in length, encoding 470 amino acids. This sequence was determined by the dideoxy chain termination method (8) from four cDNA clones generated by specific priming and reverse transcription of viral genomic RNA segments (3). All of these clones contained the entire NA-coding sequence but varied in the amount of the untranslated regions at the 5′ and 3′ ends that they contained. The sequence of the A/USSR/90/77 NA gene is shown in Fig. 1 aligned with other representative N1 and N2 sequences (4–7).

The amino acid sequence of the A/USSR/90/77 NA, translated from the nucleic acid sequence, displays less than 50% homology with representative sequences of the N2 subtype (Table 1). Although there is some intersubtypic conservation of structurally important amino acids, such as prolines and cysteines, and functionally important amino acids, such as those surrounding the proposed active site of the molecule, the low levels of homology reported in Table 1 could reflect a lack of similarity in the three-dimensional structures of the N1 and N2 NA molecules. We have addressed this question through comparison of the amino acid sequences of various N1 and N2 molecules.

Our previous work with H1 HA sequences, including that of the A/USSR/90/77 virus (2), suggested that there was a good correlation between clusters of amino acid substitutions found by comparing the sequences of different H1 HA molecules and antigenic sites identified by other criteria. The success of such correlations indicated that new antigenic sites could be identified in this way and further confirmed that the gross three-dimensional organization of the H1 HA molecule could be accurately deduced by comparison with the X-ray crystallographic data for the H3 HA molecule. We applied this same strategy in our analysis of the A/USSR/90/77 NA molecule to confirm that it is substantially similar in structure to the N2 NA molecule. In addition, we extended this analysis to predict a novel potential antigenic site not previously identified or characterized.

When amino acid differences in N2 field strains and in vitro-selected variants are superimposed on the three-dimensional structure of the neuraminidase molecule, they cluster into seven regions that have the potential of being involved in antibody interaction (1), as judged by their accessibility and variability. The regions have been numbered I through VII, and we have adopted this nomenclature in our sequence comparisons. These families of determinants defined for N2 are listed in Table 2 along with their positions in our sequence alignment in Fig. 1.

When the amino acid sequence of the A/USSR/90/77 NA is compared with those of early N1 strains, clusters of amino acid substitutions, diagnostic of antigenic determinants, appear. The locations of these clusters agree well with the locations of the antigenic determinant families defined for the N2 NA (1) (Table 2). For example, when the boundaries of the determinant families I through VII defined for the N2 NA were projected onto the N1 sequences, they contained 7.2% of the protein sequence but sustained 26% of the amino acid substitutions, corresponding to a 4.7-fold-greater rate of amino acid substitution than that observed for the rest of the molecule.

The two exceptions to this agreement are family VII, a single amino acid position at which we observed no changes in our sequence comparison, and family VIII, a new family we defined for the N1 NA (Table 2). No region of the N1 NA sequences displayed a greater concentration of amino acid substitutions than that which occurred in family VIII. Moreover, failure to observe such a cluster in N2 sequences can be explained by a difference in glycosylation between the subtypes, a phenomenon which we have previously observed in the HA molecule. The area of family VIII is primarily a surface β sheet, notated β_5S_4 in the N2 NA (9). In the assembled N2 NA tetramer, this region on each monomer is shielded by an oligosaccharide attached to position 216 (Fig. 1) of the neighboring monomer (1). The N1 NA sequences we compared lack this glycosylation site and expose this surface region to possible antibody interaction.

* Corresponding author.
† Present address: School of Medicine, University of California, San Francisco, CA 94143.

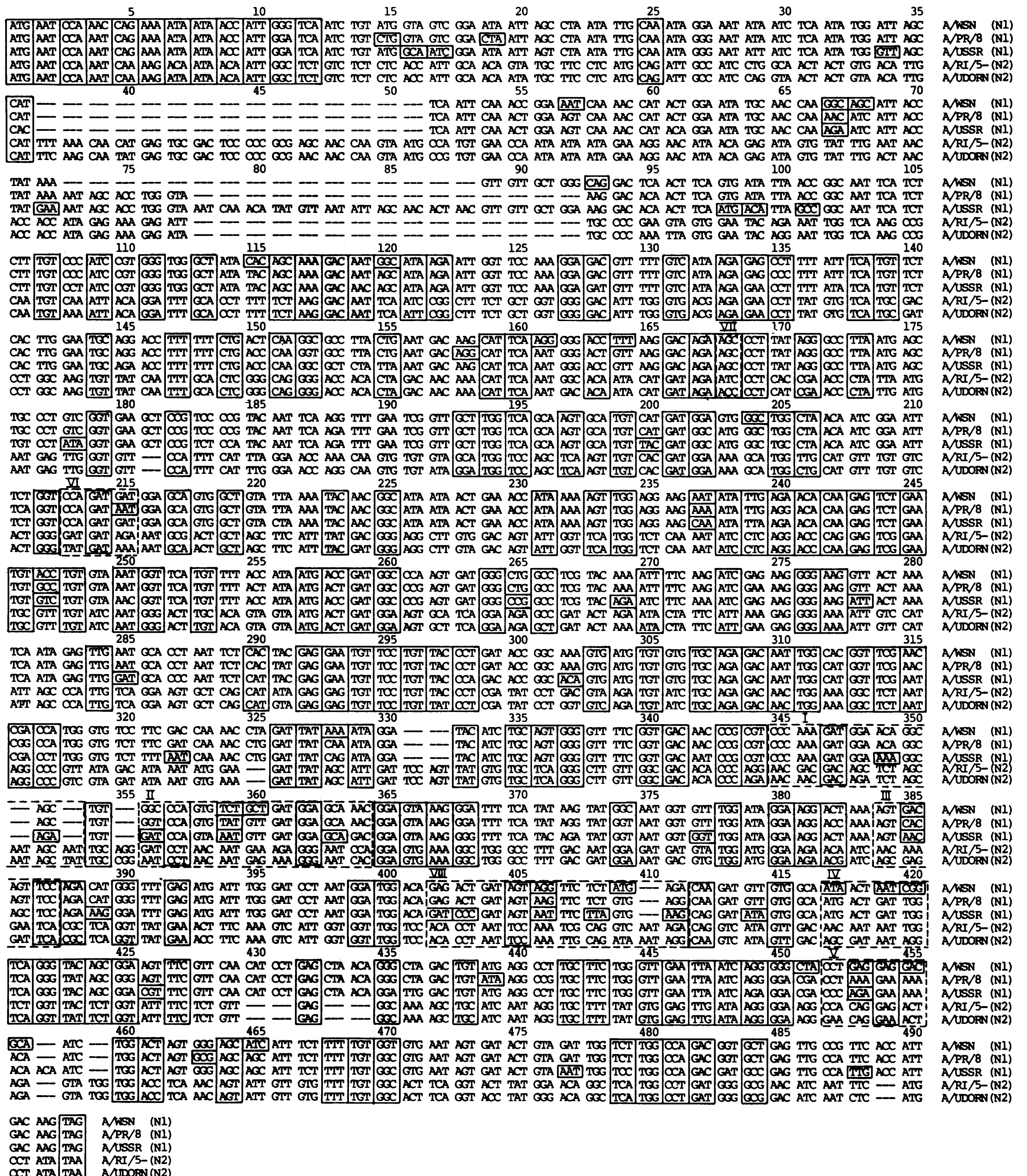

FIG. 1. Nucleotide sequence of the A/USSR/90/77 NA gene. The nucleotides are arranged in triplets corresponding to the correct translational reading frame and aligned on the basis of amino acid homology. Solid-line boxes surrounding single triplets indicate amino acid differences relative to the other N1 sequences. Solid-line boxes enclosing a triplet in all five sequences indicate an amino acid homology at that position in all five sequences. Dashed-line boxes identified by roman numerals enclose the proposed antigenic determinants as described in the text. These sites were identified both by reference to X-ray crystallographic data for the N2 NA and a moving average analysis of sequence homology among N1 NA amino acid sequences which identifies regions that display clustering of amino acid substitutions at levels significantly higher than background. The highest signals that we ignored in this analysis were changes at single amino acid positions 221 and 231, which are separated by a large region of complete intersubtypic homology.

TABLE 1. Percent homology shared among the NA gene of A/USSR/90/77 and the NA genes of other influenza viruses[a]

Virus	Amino acid (%)	Nucleic acid (%)
A/USSR/90/77 (N1)	100	100
A/PR/8/34 (N1)	89.0	87.5
A/WSN/33 (N1)	86.5	85.8
A/RI/5⁻/57 (N2)	43.8	39.8
A/UDORN/72 (N2)	47.6	41.6

[a] Percent homology has been calculated as the total number of identical positions in the aligned sequences divided by the total number of comparable positions. Positions that contain deletions are omitted from the calculation.

TABLE 2. Positions of amino acid substitutions arranged into families of potential determinants

Position	Amino acid substitution for virus:			Family[a]
	A/WSN/33	A/PR/8/34	A/USSR/90/77	
349	Thr	Thr	Lys	I
352	Ser	Ser	Arg	I
356	Gly	Gly	Asp	II
359	Ser	Tyr	Asn	II
360	Ala	Val	Val	II
364	Asn	Asn	Asp	II
385	Asp	His	Asn	III
417	Ile	Met	Met	IV
419	Asn	Asp	Asp	IV
420	Arg	Trp	Trp	IV
453	Glu	Lys	Arg	V
455	Asp	Lys	Lys	V
215	Asp	Asn	Asp	VI
402	Glu	Glu	Asp	VIII
403	Thr	Thr	Pro	VIII
406	Arg	Lys	Asn	VIII
408	Ser	Ser	Leu	VIII
409	Met	Val	Val	VIII
411	Arg	Arg	Lys	VIII

[a] Families I through VI correspond exactly to those previously defined for the N2 neuraminidase. Family VII contains only a single position at which no changes were observed in the N1 field strains. Family VIII is a novel family of determinants defined in the text for the N1 neuraminidase.

The amino acid substitutions we observed at this site strongly suggest that such interaction has occurred during circulation in vivo to select variants at this position.

Our comparison of N1 NA sequences yields two significant conclusions. First, the finding that differences in glycosylation may result in differences in the organization of antigenic sites in NA molecules of different subtypes confirms previous observations made for the HA molecule, indicating that this may be a general mechanism employed by influenza A viruses. Second, the comparison of the locations of the highly variable regions of the N1 and N2 NA molecules is a stringent test of whether the gross three-dimensional structures of the molecules are similar, since these areas are those most likely to display differences between subtypes. The striking similarity in the locations of these sites is a strong argument that the N1 and N2 NA molecules share similar structures.

We thank Gary Wilcox and the members of his laboratory, particularly Patrick Clarke, for assisting us in establishing the dideoxy sequencing technique in our laboratory and Susan Salser for expert assistance in preparing this manuscript.

This work was supported in part by Public Health Service grant GM-18586 from the National Institutes of Health.

LITERATURE CITED

1. **Colman, P. M., J. N. Varghese, and W. G. Laver.** 1983. Structure of the catalytic and antigenic sites in influenza virus neuraminidase. Nature (London) **303**:41–44.
2. **Concannon, P., I. W. Cummings, and W. A. Salser.** 1984. Nucleotide sequence of the influenza virus A/USSR/90/77 hemagglutinin gene. J. Virol. **49**:276–278.
3. **Cummings, I., and W. A. Salser.** 1980. The synthesis and cloning of large influenza A cDNAs using synthetic DNA primers, p. 147–155. *In* G. Laver and G. Air (ed.), Structure and variation in influenza virus. Elsevier/North-Holland Publishing Co., New York.
4. **Elleman, T. C., A. A. Azad, and C. W. Ward.** 1982. Neuraminidase gene from the early Asian strain of human influenza virus A/RI/5⁻/57 (H2N2). Nucleic Acids Res. **10**:7005–7015.
5. **Fields, S., G. Winter, and G. G. Brownlee.** 1981. Structure of the neuraminidase gene in human influenza virus AA/PR/8/34. Nature (London) **290**:213–217.
6. **Hiti, A. L., and D. P. Nayak.** 1982. Complete nucleotide sequence of the neuraminidase gene of human influenza virus A/WSN/33. J. Virol. **41**:730–734.
7. **Markoff, L., and C.-J. Lai.** 1982. Sequence of the influenza A/UDORN/72 (H3N2) virus neuraminidase gene as determined from full-length DNA. Virology **119**:288–297.
8. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74**:5463–5467.
9. **Varghese, J. N., W. G. Laver, and P. M. Colman.** 1983. Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 angstrom resolution. Nature (London) **303**:35–40.