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 Biology¹ 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 $\beta_5 S_4$ 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.

Vol. 50, 1984 NOTES 655

5	10 15	5 20	25 30	35
ATG AAT CCA AAT CAG AAA ATA ATA ATG AAT CCA AAT CAG AAA ATA ATA ATG AAT CCA AAT CAA AAG ACA ATA	ACC ATT GGA TCA ATC TGT CTG ACC ATT GGA TCA ATC TGT ATG ACA ATT GGC TCT GTC TCT CTC ACA ATT GGC TCT GTC TCT CTC	GTA GTC GGA CTA ATT AGC CTA A GCA NTC GGA ATA ATT AGT CTA A ACC ATT GCA ACA GTA TGC TTC C ACC ATT GCA ACA ATA TGC TTC C	TA TIG CAA ATA GGA AAT ATA ATC TCA ATA TGG AT TA TIG CAA ATA GGG AAT ATA ATC TCA ATA TGG AT TA TIG CAA ATA GGG AAT ATT ATC TCA ATA TGG AT TC ATG CAG ATT GCC ATC CTG GCA ACT ACT GTG AC TC ATG CAG ATT GCC ATC CTG GCA ACT ACT GTA AC TC ATG CAG ATT GCC ATC CAG GTA ACT ACT GTA AC	r agc a/pr/8 (N1) Tagc a/ussr (N1) A TTG a/ri/5-(N2)
CAT CAT TTT AAA CAA CAT GAG TGC GAC	TOC COC GOG AGC AAC CAA GTA	TCA ATT CAA ACC GGA AATT CA TCA ATT CAA ACT GGA AGT CA TCA ATT CAA ACT GGA AGT CA A ATG CCA TGT GAA CCA ATA ATA AT	60 AA AAC CAT ACT GGA ATA TGC AAC CAA GGC AGC ATA AAAC CAT ACT GGA ATA TGC AAC CAA AAC ATC ATC AA AAC CAT ACT GGA ATA TGC AAC CAA AAC ATC ATC AAA AAC CAT ACA GGA ATA TGC AAC CAA AAGA ATC ATC AT GAA AGG AAC ATA ACA GAG ATA GTG TAT TTG AC TA GAA AGG AAC ATA ACA GAG ATA GTG TAT TTG AC	r ACC A/PR/8 (N1) r ACC A/USSR (N1) r AAC A/RI/5-(N2)
TAT AAA — — — — — — — — — TAT AAA AAT AC ACC TGG GTA — TAT GAA AAT ACC ACC TGG GTA AAT ACC ACC ATA GAG AAA GAG ATT — ACC ACC ATTA GAG AAA GAG ATTA —	80 85	CANCE AND STEEL ST	95 AG GAC TCA ACT TCA GTG ATA TTA ACC GGC AAT TC RG GAC ACA ACT TCA GTG ATA TTA ACC GGC AAT TC RG GAC ACA ACT TCA <u>RTG IACA</u> TTA <u>GGC</u> GGC AAT TC RG GAC ACA ACT TCA <u>RTG IACA</u> TTA <u>GGC</u> GGC AAT TC GC GAA GTA GTG GAA TAC AGA AAT TGG TCA AA GC CCC AAA TTA GTG GAA TAC AGG AAT TGG TCA AA	G COCG A/RI/5-(N2)
CIT TGT COC ATC CGT GGG TGG GCT CTT TGT CCT ATC CGT GGG TGG GCT CAA TGT CAA ATT ACA GGA TTT GCA	ATA TAC AGC AAA GAC AAT AGC ATA TAC AGC AAA GAC AAC AGC OCT TIT TCT AAG GAC AAT TCA	125 ATA AGA ATT GGT TOC AAA GGA GG ATA AGA ATT GGT TOC AAA GGA GG ATA AGA ATT GGT TOC AAA GGA GG ATA AGA ATT GGT TOC AAA GGG GG ATC GGG GTT TOT GCT GGT GGG GG	130 AC GIT TIT GIC ATA AGA GAG CCT TIT ATT TCA TG AC GIT TIT GIC ATA AGA GAG CCC TIT ATT TCA TG AC GIT TIT GIC ATA AGA GAG CCC TIT ATT TCA TG AC ATT TGG GIG AGA AGA CCT TAT ATA TCA TG AC ATT TGG GIG AGA GAA CCT TAT GIG TCA ACT TGG GIG AGA GAA CCT TAT GIG TCA TG AC ATT TGG GIG AGA AGA GAA CCT TAT GIG TCA TG	140 F TCT A/WSN (N1) F TCT A/PR/8 (N1) F TCT A/USSR (N1) C GAC A/RI/5-(N2)
CAC TTG GAA TGC AGG ACC TTT TTT CAC TTG GAA TGC AGG ACC TTT TTT CAC TTG GAA TGC AGA ACC TTT TTT CCT GGC AAG TGT TAT CAA TTT GCA	150 1.55 CTG ACT CAA GGC GCC TITA CTG CTG ACC CAA GGT GCC TTA CTG CTG ACC CAA GGC GCT CTA TTA CTC GGG CAG GGG ACC ACA CTA	AAT GAC AAG CAT TCA AAG GGG A AAT GAC AAG CAT TCA AAT GGG AA AAT GAC AAG CAT TCA AAT GGG AA AAT GAC AAG CAT TCA AAT GGC AA	165 TITT AAG GAC AGA AGC COT TAT AGG GCC TTA ATT OC GTT AAG GAC AGA AGC COT TAT AGG GCC TTA ATT OC GTT AAG GAC AGA AGC COT TAT AGG GCC TTA ATT OC GTT AAG GAC AGA AGC COT TAT AGG GCC TTA ATT OC GTT AAG GAC AGA AGC COT TAT AGG GCC TTA ATT OC ATT AAG CAT GAT AGA ATC COT CAC CGA ACC CTA ATTA CAT GAT AGA ACC COT CAT CGA CCC ATTA TTA OA ATTA CAT GAT AGA ACC COT CAT CGA ACC CTA TTA OA ATTA CAT GAT AGA ACC COT CAT CGA ACC CTA TTA OA ATTA CAT GAT AGA ACC COT CAT CGA ACC CTA TTA OA ATTA CAT GAT AGA ACC COT CAT CGA ACC CTA TTA OA ATTA CAT GAT AGA ACC COT CAT CGA ACC CTA TTA OA ATTA CAT GAT AGA ACC COT CAT CGA ACC CTA TTA OA ATTA CAT GAT AGA ACC COT CAT CGA ACC CTA TTA OA ATTA CAT GAT AGA ACC COT CAT CGA ACC CTA TTA OA ATTA CAT GAT AGA ACC COT CAT CGA ACC CTA TTA OA ATTA CAT GAT AGA ACC COT CAT CGA ACC CTA TTA OA ATTA CAT GAT AGA ACC COT CAT CGA ACC CTA TTA OA ATTA CAT GAT AGA ACC COT CAT CGA ACC CTA TTA OA ATTA CAT GAT AGA ACC COT CAT CGA ACC CTA TTA OA ATTA CAT GAT AGA ACC COT CAT CGA ACC CTA TTA OA ATTA CAT GAT AGA ACC COT CAT CAT CGA ACC CTA TTA OA ATTA CAT GAT AGA ACC COT CAT CGA ACC CTA TTA OA ATTA CAT GAT AGA ACC COT CAT CGA ACC CTA TTA OA ATTA CAT GAT AGA ACC CTA TTA OA ATTA CAT GAT ACC CTA TTA OA ATTA CAT ACC CTA TTA OA ATTA CAT GAT ACC CTA TTA OA ATTA CAT ACC CTA TTA OA ATTA CAT GAT ACC CTA TTA OA ATTA CAT CAT ACC CTA TTA OA ATTA CAT ACC CTA TTA OA ATTA CAT CAT ACC CTA TTA OA ATTA CAT CAT CAT ACC	175 G AGC A/WSN (N1) G AGC A/PR/8 (N1) G AGC A/USSR (N1) A ATG A/RI/5-(N2)
TEC COT GIC GOT GAA GOT COG TOC TEC COT GIC GOT GAA GOT COG TOC TOT COT ATA GOT GAA GOT COG TOC AAT GAG TIG GOT GIT — COA TIT	185 190 CCG TAC AAT TCA AGG TIT GAA CCG TAC AAT TCA AGA TIT GAA CCA TAC AAT TCA AGA TIT GAA CCA TTA GGA ACC AAA CAA GIG	195 A TOG GITT GCT TGG TCA GCA AGT GC A TGG GITT GCT TGG TCA GCA AGT GC A TGG GITT GCT TGG TCA GCA AGT GC A TGG GITT GCT TGG TCA GCA AGT GC G TGT GTA GCA TGG TCC AGC TCA AG	200 205 CA TGT CAT GAT GGA GTG GGC TGG CTA ACA ATC GG CA TGT CAT GAT GGC ATG GGC TGG CTA ACA ATC GG CA TGT TAT CAT GGC ATG GGC TGG CTA ACA ATC GG CA TGT TAT GAT GGC ATG GGC TGC CTA ACA ATC GG TT TGT CAC GAT GGA AAA GCA TGG TTG CAT GTT TG GT TGT TGC CAC GAT GGA AAA GCA TGG CTG CAT GTT TG	210 A ATT A/WSN (N1) A ATT A/PR/8 (N1) A ATT A/USSR (N1) F GTC A/RI/5-(N2)
VI 21.5 TCT GST CCA GAT GAT GCA GCA GTG TCA GGT CCA GAT AAT GCA GCA GTG TCT GST CCA GAT GAT GAT GCA GCA GTG TCT GST CCA GAT GAT AAT GCA GCA GTG ACT GSG GAT GAT ACA AAT GCG ACT	220 225 GCT GTA TTA AAA TAC AAC GGC GCT GTA TTA AAA TAC AAC GGC GCT GTA CTA AAA TAC AAC GGC GCT ACC TTC ATT TAT GAC GGG	230 ATA ATA ACT GAA ACC ATA AAA AC AGG CTT GTG GAC AGT ATT GGT TO	235 TTGG AGG AAG AAT ATA TTG AGG ACA CAA GAG TC FT TGG AGG AAG AAAT ATA TTG AGG ACA CAA GAG TC FT TGG AGG AAG AAAA ATA TTG AGA ACA CAA GAG TC FT TGG AGG AAG AAAA ATA TTG AGA ACA CAA GAG TC CA TGG TCT CAA AAT ATC CTC AGG ACC CAG GAG TC CA TGG TCT CAA AAT ATC CTC AGG ACC CAG GAG TG	245 GAA
TGT GCC TGT GTA AAC GGT TCA TGT TGT GTC TGT GTA AAC GGT TCA TGT TCC GTT TGT ATC AAT GGG ACT TGC TGC GTT TGT ATC AAT GGG ACT TGC TGC GTT TGT ATC AAT GGG ACT TGT	TIT ACT ATA ATG ACT GAT GGC TIT ACC ATA ATG ACC GAT GGC ACA GTA GTA ATG ACT GAT GGA	COCG AGT GAT GGG CTG GCC TCG TX COCG AGT GAT GGG CCG GCC TCG TX A AGT GCA TCA GGA AGA GCC GAT AC	270 AC AAA ATT TIC AAG ATC GAG AAG GGG AAG GTT AC AC AAA ATT TIC AAG ATC GAG AAG GGG AAG GTT AC AC AGA ATC TIC AAA ATC GAG AAG GGG AAG ATT AC ACT AGA ATC CTA TIC ATT AAA CAG GGG AAA ATT AC ATT AAA ATA CTA TIC ATT GAT GAG GAG GAG AAA ATT GT AAA ATA CTA TIC ATT GAT GAG GAG GAG AAA ATT GT	T AAA A/PR/8 (N1) T AAA A/USSR (N1) C CAT A/RI/5-(N2)
TCA ATA GAG TTG AAT GCA CCT AAT TCA ATA GAG TTG GAT GCA CCC AAT ATT AGC CCA TTG TCA GGA AGT GCT	TCT CAC TAT GAG GAA TGT TCC TCT CAT TAC GAG GAA TGT TCC CAG CAT ATA GAG GAG TGT TCC	TIGT TAC OCT GAT ACC GGC AAA G. TIGT TAC OCT GAT ACC GGC AAA G. TIGT TAC OCA GAC ACC GGC ACA TIGT TAC OCA GAC ACC GGC ACA TIGT TAC OCT GAC TAT OCT GAC GC	305 ITS ATC TOT GTC TOC ACA GAC AAT TOC CAC GOT TO ITS ATC TOT GTC TOC ACA GAC AAT TOC CAC GOT TO ITS ATC TOT GTC TOC ACA GAC AAT TOC CAT GOT TO ITS ATC TOT GTC TOC ACA GAC AAT TOC CAT GOT TO ITS ACA TOT ATC TOC ACA GAC AAC TOC AAA GOC TO ITS ACA TOT ATC TOC ACA GAC AAC TOC AAA GOC TO ITS ACA TOT ATC TOC ACA GAC AAC TOC AAA GOC TO	GAAC A/PR/8 (N1) GAAT A/USSR (N1)
CGA CCA TOG GTG TCC TTC GAC CAA CGG CCA TGG GTG TCT TTC GAT CAA CGA CCT TGG GTG TCT TTT TAMT CAA AGG CCC GTT ATA GAC ATA AAT ATG AGG CCC GTT ATA GAC ATA AAT ATG AGG CCC GTC GTA GAT ATA AAT GTG	325 AAC CTA GAT TAT AAA ATA GGA AAC CTG GAT TAT CAA ATA GGA AAC CTG GAT TAT CAG ATA GGA AAC CTG GAT TAT ACC ATT GAT AAA — GAT TAT ACC ATT GAT AAA — GAT TAT ACC ATT GAT	335	340 THE THE GOT GAC AAC COG COT COC AAA GAT GGA AC THE THE GOT GAC AAC COG COT COC AAA GAT GGA AC THE THE GOT GAC AAT COG COT COC AAA GAT GGA AC THE THE GOT GAC AAT COG COT COC AAA GAT GGA AC THE GOT GC GAC ACA COC AGGA ACA CAC CAC THE GOT GC GAC ACA COC AGA AAC AAC CAC CAC AGA ACC THE GOT GOC GAC ACA COC AGA AAC AAC GAC CAC AGA TO	350 \\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\
AGC TGT GGT CCA GTG AGA TGT GAT CCA GTA	TAT GIT GAT GGA GCA AAC GGA AAT GIT GAT GGA GCA GAC GGA AAT GAA AGA GGG AAT CCA GGA AAT GAG AAA GGG AAT CAC GGA	GTA AAG GGA TITT TCA TAT AAG TI GTA AAG GGA TITT TCA TAT AGG TI GTA AAG GGG TTT TCA TAC AGA TI	375 AT GGC AAT GGT GTT TGG ATA GGA AGG ACT AAA AG AT GGT AAT GGT GTT TGG ATA GGA AGG ACC AAA AG AT GGT AAT GGT GGT TGG ATA GGA AGG ACC AAA AG AT GGT AAT GGT GGT TGG ATG GGA AGA ACA ACA ATG GGA GGT GGT GGT TGG ATG GGA AGA ACA ATG GGA GGT GGT TGG ATG GGA AGA ACA ATG GGA AGA ACC ATC AG	CAC A/PR/8 (N1)
AGT TCC AGA CAT GGG TTT GAG ATG AGC TCC AGA AAG GGA TTT GAG ATG GAA TCA CCC TCA GGT TAT GAA ACT GAT TCA CCC TCA GGT TAT GAA ACC	ATT TGG GAT CCT AAT GGA TGG ATT TGG GAT CCT AAT GGA TGG TTC AAA GTC ATT GGT GGT TGG TTC AAA GTC ATT GGT GGT TGG	ACA GAT COC GAT AGT AAG TIC TO ACA GAT COC GAT AGT AAT TIC TO TOC ACA CCT AAT TOC AAA TOG C	410 CT MTG — AGA (CAA GAT GTT) GTC CCA MTG ACT MAG CT GTG — AGG (CAA GAT GTT) GTG CCA ATTA ACT GAT TA GTG — AAG (CAG GAT ATTA GTG CCA ATG ACT GAT AG GTC AAT AGA (CAG GTC ATTA GTT) GAC AAC AAT AAT AG ATA AAT AGG (CAG GTC ATTA GTT) GAC AAC AAT AAT AG ATA AAT AGG (CAG GTC ATTA GTT) GAC AAC AAT AAT AG ATA AAT AGG (CAA GTC ATTA GTT) GAC AGC GAT AAT	TGG A/PR/8 (N1) TGG A/USSR (N1) TGG A/RI/5-(N2) TAGG A/UDORN(N2)
TCA GGG TAC AGC GGA CGT TTC GTT TCT GGT TAC TCT GGT ATT TTC TCT TCA GGT TAT TCT GGT ATT TTC TCT	CAA CAT CCT GAG CTA ACA GGA GTT GAG GGC GTT GAG GGC	TTG GAC TGT ATG AGG CCT TGC TT AAA AGC TGC ATC AAT AGG TGC TT AAA AGC TGC ATC AAT AGG TGC TT	445 450 TT CAA TTA ATC LAGG GGG CTA COT AAA GA GA CTA CT AAA TTA ATC LAGG GGG GGA CGA CCT AAA GA CT GGG GTT GAA TTA ATC LAGG GGA CGA CCA CCC AAA GA CTA ATC LAGG GGA CGA CCC AAA CTA ATC LAGG GGA CGA CCC AAA CTA ATC LAGG GGA CGA CCC AAA CTA CTA TTA ATC LAGG GGA AGG CCA CGG CA CT TATA GGG GGA CGG CAA CCAG CAA CAG CGA CGC CAC CCC CAC CA	AAA A/USSR (N1) ACT A/RI/5-(N2)
ACA ACA ATC TGG ACT AGT GGG AGA GTA TGG TGG ACC TCA AAC	AGC AGC ATT TCT TTT TGT GGC AGT ATT GTT GTG TTT TGT GGC	GIG AAT AGT GAT ACT GTA GAT TO GIG AAT AGT GAT ACT GTA GAT TO GIG AAT AGT GAT ACT GTA (AAT) TO ACT TOA GGT ACT TAT GGA ACA GG	480 35 TOT TOS COA GAC GOT GCT GAG TTG COA TTC ACC 36 TOT TOS COA GAC GOT GCT GAG TTG COA TTC ACC 36 TOT TOS COA GAC GAT GCT GAG TTG COA TTC ACC 36 TOC TOS COA GAC GAT GCC GAG TTG COA TTG ACC 37 TOA TOS COT GAT GOS GCS AAC ATC AAT CTC 37 TOA TOS COT GAT GOS GCS GAC ATC AAT CTC 38 TOA TOS COT GAT GOS GCS GAC ATC AAT CTC	ATT A/USSR (N1)
GAC AAG TAG A/NSN (N1) GAC AAG TAG A/PR/8 (N1) GAC AAG TAG A/VR/7S (N1) CCT ATA TAA A/VR/7S (N2) CCT ATA TAA A/VLORN (N2)				

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.

656 NOTES J. VIROL.

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

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	

[&]quot;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	F "1"		
	A/WSN/33	A/PR/8/34	A/USSR/90/77	Family"
349	Thr	Thr	Lys	Ī
352	Ser	Ser	Arg	I
356	Gly	Gly	Asp	П
359	Ser	Tyr	Asn	H
360	Ala	Val	Val	H
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

[&]quot; 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

- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.