Neuraminidase gene from the early Asian strain of human influenza virus, A/RI/57/57 (H2N2)

T.C.Elleman, A.A.Azad and C.W.Ward

CSIRO, Division of Protein Chemistry, Parkville, Victoria 3052, Australia

Received 27 July 1982; Revised and Accepted 8 September 1982

ABSTRACT

The complete structure of the neuraminidase gene from the A/RI/57/57 strain of influenza virus has been determined. It is 1467 nucleotides long and codes for a protein of 469 amino acid residues. Comparison with the gene sequence for the N1 strains A/WSN/33 and A/PR/8/34, the N2 strain A/Udorn/72 and the protein sequence for the N2 strain A/Tokyo/3/67 shows the amino acid sequence changes that have occurred during antigenic shift (60%) and drift (7-9%).

INTRODUCTION

The influenza A virus is an enveloped particle containing eight separate gene segments of single-stranded, negative sense RNA which code for seven viral and three non-structural proteins. The protein/gene segment assignments have been established (see refs. 1-3 for reviews) and complete sequences for representatives of these eight gene segments are now available (see refs. 4 & 5 for further references).

Only two of the viral proteins, the haemagglutinin and neuraminidase, are found on the surface of the virion. Both are anchored in the viral membrane and project out from the virus surface, both are glycosylated and both exhibit major (shift) and minor (drift) changes in antigenicity during the evolution of new virus strains⁶. The structural basis of antigenic variation in the major surface antigen, the haemagglutinin has recently been established from protein sequence, gene sequence and X-ray crystallographic studies of natural field strains and variants selected with monoclonal antibodies or hyperimmune sera (see 7 for review). Similar studies are now in progress to characterise antigenic variation in the neuraminidase.

Peptide maps^{8,9} and nucleotide sequences (150-300 residues long) from the 3'-end of the neuraminidase gene^{10,11} have been used to partially characterise antigenic drift and shift but whole sequences are required if the primary and tertiary structural changes associated with antigenic variat-

ion are to be established. Complete sequences have been published to date only for the two closely related N1 variants A/WSN/33¹² and A/PR/8/34¹³ and an N2 variant A/Udorn/72⁴⁵ from the Hong Kong subtype. These studies showed the extensive sequence differences between N1 and N2 variants associated with the antigenic shift that occurred in 1957.

In this paper we report the complete structure of the neuraminidase gene from the early Asian influenza virus A/RI/5⁻/57. This sequence is required to establish the extent of the nucleotide and amino acid substitutions that have accumulated by antigenic drift in the N2 neuraminidase since its appearance in 1957.

EXPERIMENTAL PROCEDURES

Preparation of influenza viral RNA

The virus used was a recombinant containing the neuraminidase of A/RI/5⁻/57 prepared as described by Webster¹⁴. Virus was grown in the allantoic sac of 11-day chick embryos and purified as described by Laver¹⁵. Viral RNA was prepared from viral suspension by pronase digestion and phenol extraction. Total yield was ca. 1500 µg RNA from 1500 eggs.

Cloning and Sequence Analysis

The neuraminidase gene structure was obtained by a combination of cloning in pBR322, dideoxy sequencing on viral RNA (vRNA) and dideoxy sequencing on a single-stranded cDNA copy of the 5' end of the neuraminidase gene. For cloning, cDNA transcripts of vRNA segments were prepared by reverse transcription of unfractionated vRNA using the commercially-available, synthetic, dodecanucleotide primer, d(pAGCAAAAGCAGG) complementary to the 3' end of all eight influenza virus gene segments¹⁶. These cDNA transcripts were made double-stranded by 3'-end self-priming, and treated with S1 nuclease¹⁷ to render them suitable for cloning into the PstI site of pBR322 by G/C tailing¹⁸. Clones containing copies of the neuraminidase gene were identified by probing with neuraminidase vRNA isolated from agarose gel and 5' end labelled with [γ -32P] ATP.

No full-length cloned copies of the neuraminidase gene were obtained, the longest cloned fragment being 84% of the gene. The sequence was determined for this fragment which represents the first 1235 nucleotides of the A/RI/5⁻/57 neuraminidase gene. Restriction digests of whole chimeric plasmid, or restriction fragments of the insert isolated from acrylamide gels, were cloned in various vectors of the M13mp2 series^{19,20} (Fig. 1). Sequences of these M13-cloned fragments were determined by the chain-termination method²¹

Fig. 1: Summary of sequencing strategy. The length of the cloned copy of the neuraminidase gene together with the sequence directly determined from this double-stranded copy using primer d(pAGCAAAAGCAGG) are shown relative to the virion RNA molecule. From this clone, HpaII restriction fragments were sub-cloned in M13mp7¹⁹, PstI fragments in M13mp2/PstI²⁰, and AluI fragments into the SmaI site of M13mp8 (Messing, unpublished). The further 5'-terminal vRNA sequence was determined by copying vRNA using a HinfI restriction fragment comprising nucleotides 1055-1124 or an exonucleaseIII-treated HinfI/PstI fragment comprising nucleotides 1125-1235.

using universal primers^{22,23} or internal restriction fragments isolated from the cloned neuraminidase gene as primers.

Discrepancies between this longest pBR322-cloned fragment and a second shorter one were found at nucleotide positions 253, 478 and 669 of the first 894 positions. RNA sequence analysis using internal restriction fragments as primers showed these discrepancies all corresponded with the incorrect pairing of thymidine with viral guanosine. In view of the estimated error frequency (10^{-3}) for reverse transcriptase^{24,25,26}, these discrepancies are not unexpected. RNA sequence analysis confirmed the remainder of the cloned sequence.

To extend the sequence towards the 5' end of the vRNA two restriction fragments derived from the cloned insert (nucleotides 1055-1124 and 1125-1235) after exonuclease III treatment were used as specific primers in the chain-termination method using reverse transcriptase²⁷ and unfractionated vRNA as template. This extended sequence was confirmed in the opposite direction by first producing a single stranded cDNA copy of the neuraminidase gene from whole vRNA with reverse transcriptase using a PstI/Bam HI restriction fragment (nucleotides 894-1033). The sequence of this cDNA was determined by the chain-termination method using the commercially-available dodecanucleotide primer d(pAGTAGAAACAAG). This primer is equivalent to the constant 5'-

end of all influenza vRNA genes¹⁶ and so is complementary to the 3'-end of cDNA molecules that extend to the end of the neuraminidase gene. This back sequencing was done in the usual manner with [α -32P] dATP²¹, and also with 5'-end labelled primer²⁸ and unlabelled deoxyribonucleoside triphosphates to determine the two nucleotides, 1454 and 1455, directly adjacent to the primer. These approaches yielded the complete sequence of the A/RI/5⁻/57 neuraminidase gene and 94% of the sequence (nucleotides 34-1400), was determined in both directions.

RESULTS

Nucleotide Sequence of the Neuraminidase Gene

The sequence of the A/RI/5⁻/57 N2 neuraminidase gene, expressed as the complementary DNA, is shown in Fig. 2. The sequence contains a total of 1467 nucleotides. There is only one long reading frame in the cDNA sequence and it codes for a primary gene product of 469 amino acid residues. The five alternative reading frames are blocked by numerous termination codons. Nineteen bases precede the first ATG codon in the cDNA sequence; 1407 nucleotides code without interruption for the neuraminidase protein, and there is a single stop codon TAA followed by 38 bases to the end of the sequence.

Previous sequence studies of the neuraminidase gene from A/RI/5⁻/57 and A/Tokyo/67 strains have been limited to partial sequences of up to 260 nucleotides from the 3'-end of the vRNA^{10,11}. Our data differs from the published 151 nucleotide sequence of A/RI/5⁻/57 at nucleotides 103,106 and 108, and from the published A/Tokyo/67 sequence which has a frameshift error after nucleotide 158. Our nucleotide assignments at these positions are confirmed by our primed RNA sequence analysis, by the amino acid sequence of the first 31 amino acids at the N-terminus of the closely related neuraminidase from the recombinant strain X-7.F1²⁹, and by the amino acid sequence of the neuraminidase heads from A/Tokyo/67³⁰.

DISCUSSION

Structure of the Neuraminidase Protein

The protein sequence encoded by the neuraminidase gene is shown in Fig. 2. Amino acid sequence analysis of intact neuraminidase has shown that protein synthesis does start at the first AUG triplet in the complementary gene sequence, and that there is no proteolytic processing at the N-terminal end²⁹. Neither the initiating methionine residue, nor a hydrophobic signal peptide is removed from the neuraminidase during biosynthesis. The gene

Fig. 2: Sequence of the neuraminidase gene in complementary RNA sense. The predicted protein sequence is shown above, using the single-letter code. Potential glycosylation sites are indicated by an asterisk. The N-terminal hydrophobic region is shown underlined, with the further possible extension of the stalk indicated by a broken line. Arrows indicate pronase-cleavage sites in the neuraminidase. The sequences of the twelve nucleotides at each end of the molecule are those of the commercially-available primers.

sequence in conjunction with the protein data^{29,30} shows that there is also no post-translational processing at the C-terminal end. The protein contains 469 amino acid residues for an apoprotein MW of 52,006 (Table 1).

Molecular weight estimates^{29,31,32,33} and the emerging three-dimensional crystal structure³⁴ show that the enzyme as it exists in the virus is a tetramer. It appears in electron micrographs^{35,36} to have a box-like head, 80x80x40 Å, made up of four co-planar, roughly spherical sub-units, and a

Table 1. Amino Acid Compositions of the Neuraminidases from A/PR/8/34 (N1)¹³ and A/RI/5⁻/57

Amino Acid	Residues/mole		Amino Acid	Residues/mole	
	A/PR/8/34	A/RI/5 ⁻ /57		A/PR/8/34	A/RI/5 ⁻ /57
Asn (N)	25	34	Val (V)	28	34
Asp (D)	25	27	Met (M)	7	8
Thr (T)	29	33	Ile (I)	41	39
Ser (S)	51	47	Leu (L)	21	19
Gln (Q)	11	15	Tyr (Y)	14	13
Glu (E)	17	22	Phe (F)	16	14
Pro (P)	21	22	Trp (W)	16	13
Gly (G)	44	39	His (H)	10	10
Ala (A)	16	16	Lys (K)	23	17
Cys (C)	19	22	Arg (R)	20	25
Total Residues				454	469
Apoprotein MW				50,087	52,006

narrow, centrally attached stalk (15Å x 100Å) with a small, highly aggregating knob at its base. This knob region is very hydrophobic and serves to anchor the neuraminidase in the viral membrane. Comparative peptide maps of N2, N5 and N8 strains²⁹ and the complete amino acid sequence of the A/Tokyo/67 N2 neuraminidase heads³⁰ have shown that this membrane-embedded stalk region comes from the N-terminal end of the protein.

As found for the N1 neuraminidase¹³, there is a hydrophobic sequence from residues 7 to 35 (Fig. 2) which probably serves to anchor the neuraminidase in the viral membrane. It is of similar length to the membrane-embedded regions of some other membrane proteins¹³ and is flanked by charged residues at positions 6 and 36 that could interact with phosphate groups on the surfaces of the lipid bilayer³⁷. Chou-Fasman structural predictions³⁸ show that the residues that constitute the stalk-region of the neuraminidase include a potential 27 residue β -strand (residues 47-73). This strand from each monomer may be arranged in the tetramer as a four chain β -barrel similar to the structure found in ovomucoid crystals³⁹ and the parallel, triple-stranded β -structure in tomato bushy stunt virus⁴⁰. The globular head region is predicted to be rich in potential β -sheet and β -bend structures.

As shown in Fig. 2 there are twenty two half-cystine residues in the

N2 neuraminidase, including one (residue 21) in the proposed membrane attachment sequence and three (residues 42,53 and 78) in the stalk region. There are eight potential glycosylation sites in the N2 protein, three of which occur in the stalk region, at residues 61,69 and 70, and five in the heads at residues 86,146,200,234 and 402. The protein sequence data for the heads of A/Tokyo/67 showed only four of the five potential sites are glycosylated. These are asparagine residues 86,146,200,234³⁰. Nothing is known about the glycosylation status of the three sites in the stalk. In the B/Lee/40 strain of influenza virus the stalk region is known to be rich in carbohydrate⁴¹.

Sequence Changes and Antigenic Shift

Comparison of complete N1¹³ and N2 nucleotide sequences (Fig. 3) shows that antigenic shift in the neuraminidase, like that in the haemagglutinin⁷, does not involve an extension of the point mutation process occurring during antigenic drift, but involves extensive changes in sequence consistent with genetic reassortment⁴². The non-translated region at the 5'-end of the cDNA contains 20 nucleotides in A/PR/8/34 (N1) and 19 in A/RI/5⁻/57 (N2), the coding regions contain 1362 and 1407 bases respectively, and the number of nucleotides at the 3'-end of the cDNA are 31 in A/PR/8/34 and 41 in A/RI/5⁻/57. There is a 45 nucleotide deletion in A/PR/8/34 which, by comparison with the partial sequences of other N1 neuraminidase genes¹¹, involves nucleotides 207-251. Other N1 strains isolated between 1933 and 1935 showed similar deletions which differ in size and position⁴⁶.

The nucleotide sequences of A/PR/8/34 and A/RI/5⁻/57 differ by approximately 50% with the expected¹⁶ high conservation of sequence at the 5'-non coding region of the cDNA and at the 3'-non coding end (21 identical residues). Other regions of sequence that are highly conserved between A/PR/8/34 (N1) and A/RI/5⁻/57 (N2) are nucleotides 20-54, 449-459, 738-750, 845-865, 890-904 and to a lesser extent 689-708.

The overall amino acid sequence homology between the N1 and N2 proteins is 40%. As shown in Fig. 3 the sequence homology is not distributed evenly over the length of the molecule. As previously reported^{11,37} the first six amino acids are conserved in all sub-types and the next six residues conserved in most sub-types. From then on there are major differences in sequence. As shown in Fig. 3 there is little homology in the hydrophobic membrane attachment sequence and no homology in the region (residues 37-73) which comprises the narrow-stalk region of the protein. In this stalk region neither the half-cystine residues nor the potential glycosylation

with the protein sequence³⁰ of the heads of the N2 variant A/Tokyo/67 and the proposed sequence of A/Udorn/72 shows that antigenic drift in the N2 neuraminidase involves an accumulation of point mutations at a rate similar to that found in the Hong Kong haemagglutinin. Within the period of the Asian influenza epidemics 1957-1967, twenty-seven amino acid substitutions have accumulated in the heads of the neuraminidase. This represents a mutation rate of 0.68 residues/year/100 amino acid residues which is comparable to the rate of 0.66 found for the whole haemagglutinin in the period 1968 to 1979.

The Hong Kong influenza sub-type which appeared in man in 1968 retained the Asian N2 neuraminidase, and this has continued to undergo antigenic drift. Comparison of the A/Tokyo/67 and A/Udorn/72 heads shows that an additional 19 changes in amino acid sequence have occurred during this period. Three occur at positions already changed in A/Tokyo/67 and seven are apparent reversions, reflecting the fact that strains isolated in different parts of the world in chronological order do not necessarily represent a direct genealogical lineage, and that as found in the Hong Kong haemagglutinin some sequence changes were found to be restricted to individual variants (see ref. 7 for review).

When the nucleotide sequence of the neuraminidase gene of the Asian strain A/RI/57/57 is compared with that of the Hong Kong variant A/Udorn/72 the sequence homology is 93%. Ninety-one nucleotide differences occur, 87 of these within the coding region. Thirty-nine of these changes are silent and five are second mutations within one codon. Of the 43 amino acid differences (91% homology), 36 occur in the head region (residues 74-469) while seven occur in the stalk, representing an overall rate of substitution in the whole molecule of 0.61 amino acid residues/year/100 amino acids. This rate from 1957-1972 does not significantly differ from the rate within the Asian era.

As shown in Fig. 3, 18 of the 43 amino acid sequence changes between the A/RI/57/57 and A/Udorn/72 neuraminidase involve radical differences in charge. In particular the regions 197-199, 344-347, 367-370, 400-403 and residue 431 in which sequential substitutions occurred at the same or adjacent positions, are strong contenders for involvement in the antigenic determinants.

Concluding Remarks

This laboratory is investigating the molecular basis of antigenic variation in the influenza virus neuraminidase by protein sequence, gene sequence and X-ray crystallographic analyses. The gene sequence of the

neuraminidase from A/RI/5⁻/57 was determined to establish the extent of structural change that accompanied the N1 to N2 shift in 1957 and to provide the original structure from which antigenic drift in the N2 neuraminidase proceeded. The protein sequence of the A/Tokyo/67 neuraminidase heads³⁰ was determined to aid the resolution of the X-ray crystal structure³⁴ and to check the authenticity of the sequence of the cloned gene. It was also expected to have accumulated a sufficient number of amino acid substitutions to allow the antigenic regions of the neuraminidase to be identified on the three-dimensional structure as was done with the Hong Kong haemagglutinin⁴³. Serological analysis of antigenic variants of A/Tokyo/67 selected with anti-N2 monoclonal antibodies provided evidence for at least three non-overlapping antigenic regions on the N2 tetramer⁴⁴. The structures reported here and in the previous paper³⁰, together with the comparative sequences of other N2 field strains, and variants selected with monoclonal antibodies will allow the number of antigenic determinants and the way these change during the evolution of new influenza strains of epidemic importance, to be characterised.

We thank Dr. W.G. Laver for the primary A/RI/5⁻/57 inoculum and Susan Barrett for expert technical assistance.

REFERENCES

1. Scholtissek, C. (1978) *Curr. Top. Microbiol. Immunol.* 80, 139-169.
2. Webster, R.G., Laver, W.G., Air, G.M. and Schild, G.C. (1982) *Nature* 296, 115-121.
3. Palese, P. (1977) *Cell* 10, 1-10.
4. Bishop, D.H.L., Huddleston, J.A. and Brownlee, G.G. (1982) *Nucleic Acids Res.* 10, 1335-1343.
5. Winter, G. and Fields, S. (1982) *Nucleic Acids Res.* 10, 2135-2143.
6. Schild, G.C., Newman, R.W., Webster, R.G., Major, D. and Hinshaw, V.S. in *Structure and Variation in Influenza Virus*. Laver, W.G. and Air, G.M. Eds., pp.373-383 Elsevier, New York.
7. Ward, C.W. (1981) *Curr. Top. Microbiol. Immunol.* 94/95, 1-74.
8. Kendal, A.P. and Kiley, M.P. (1973) *J. Virol.* 12, 1482-1490.
9. Laver, W.G. (1978) *Virology* 86, 78-87.
10. Blok, J. and Air, G.M. (1980) *Virology* 107, 50-60.
11. Blok, J. (1981) *ICN-UCLA Symp. Molec. Cell Biol.* 21, 45-54.
12. Hiti, A.L. and Nayak, D.P. (1982) *J. Virol.* 41, 730-734.
13. Fields, S., Winter, G. and Brownlee, G.G. (1981) *Nature* 290, 213-217.
14. Webster, R.G. (1970) *Virology* 42, 633-642.
15. Laver, W.G. (1969) in *Fundamental Technique of Virology*, Habel, K. and Salzman, N.P. Eds., pp.82-86 Academic Press, New York.
16. Robertson, J.S. (1979) *Nucleic Acids Res.* 6, 3745-3757.
17. Higuchi, R., Paddock, G.V., Wall, R. and Salser, W. (1976) *Proc. Nat. Acad. Sci. USA.* 73, 3146-3150.
18. Otsuka, A. (1981) *Gene* 13, 339-346.
19. Messing, J., Crea, R. and Seeburg, P.H. (1981) *Nucleic Acids Res.* 9, 309-321.

20. Bentley, D.L. and Rabbitts, T.H. (1980) *Nature* 288, 730-733.
21. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Nat. Acad. Sci. USA* 74, 5463-5467.
22. Anderson, S., Gait, M.J., Mayol, L. and Young, I.G. (1980) *Nucleic Acids Res.* 8, 1731-1745.
23. Duckworth, M.L., Gait, M.J., Goellet, P., Hong, G.F., Singh, M. and Titmas, R.C. (1981) *Nucleic Acids Res.* 9, 1691-1706.
24. Battula, N. and Loeb, L. (1974) *J. Biol. Chem.* 249, 4086-4093.
25. Gopinathan, K.P., Weymouth, L.A., Kunkel, T.A. and Loeb, L.A. (1979) *Nature* 278, 857-859.
26. Fields, S. and Winter, G. (1981) *Gene* 15, 207-214.
27. Hamlyn, P., Brownlee, G.G., Cheng, C.C., Gait, M.J. and Milstein, C. (1978) *Cell* 15, 1067-1075.
28. Maxam, A. and Gilbert, W. (1980) in *Methods in Enzymology*. Grossman, L. and Moldave, K. Eds., Vol. 65, pp.499-560. Academic Press, New York.
29. Blok, J., Air, G.M., Laver, W.G., Ward, C.W., Lilley, G.G., Woods, E.F., Roxburgh, C.M. and Inglis, A.S. (1982) *Virology* 119, 109-121.
30. Ward, C.W., Elleman, T.C. and Azad, A.A. (1982) *Biochem. J.* 207, 91-95.
31. Bucher, D.J. and Kilbourne, E.D. (1972) *J. Virol.* 10, 60-66.
32. Lazdins, I., Haslam, E.A. and White, D.O. (1972) *Virology* 49, 758-765.
33. Wrigley, N.G., Skehel, J.J., Charlwood, P.A. and Brand, C.M. (1973) *Virology* 51, 525-529.
34. Colman, P.M. and Laver, W.G. (1981) in *Structural Aspects of Recognition and Assembly in Biological Macromolecules*. Balaban, M., Sussman, J.L., Traub, W. and Yonath, A. Eds., pp.869-872. Balaban ISS, Rehovot/Philadelphia.
35. Laver, W.G. and Valentine, R.C. (1969) *Virology* 38, 105-119.
36. Wrigley, N.G. (1974) *Brit. Med. Bull.* 35, 35-38.
37. Blok, J. and Air, G.M. (1982) *Biochemistry*. In Press.
38. Ward, C.W. and Dopheide, T.A. (1980) *Aust. J. Biol. Sci.* 33, 449-455.
39. Weber, E., Papamokos, E., Bode, W., Huber, R., Kato, I. and Laskowski, M. (1981) *J. Mol. Biol.* 149, 109-123.
40. Harrison, S.C., Olson, A.J., Schutt, C.E., Winkler, F.K. and Bricogne, G. (1978) *Nature* 276, 368-373.
41. Allen, A.K., Skehel, J.J. and Yuferov, V. (1977) *J. Gen. Virol.* 37, 625-628.
42. Scholtissek, C., Rohde, W., von Hoyningen, V. and Rott, R. (1978) *Virology* 87, 13-20.
43. Wiley, D.C., Wilson, I.A. and Skehel, J.J. (1981) *Nature* 298, 373-378.
44. Webster, R.G., Hinshaw, V.S. and Laver, W.G. (1982) *Virology* 117, 93-104.
45. Markoff, L. and Lai, C.J. (1982) *Virology* 119, 288-297.
46. Blok, J. and Air, G.M. (1982) *Virology* 118, 229-234.