Amino acid sequence of the Pronase-released heads of neuraminidase subtype N2 from the Asian strain A/Tokyo/3/67 of influenza virus

Colin W. WARD, Thomas C. ELLEMAN and Ahmed A. AZAD
C.S.I.R.O. Division of Protein Chemistry, 343 Royal Parade, Parkville, Vic. 3052, Australia

(Received 14 April 1982/Accepted 9 July 1982)

The amino acid sequence of the Pronase-released heads of neuraminidase subtype N2 from the A/Tokyo/3/67 strain of influenza virus was determined by a combination of peptide and nucleic acid sequence analysis. The results show that the Pronase-released heads contain 396 amino acid residues and extend from residue 74 in the original protein to the C-terminus at residue 469. The heads contain five potential glycosylation sites at asparagine residues 86, 146, 200, 234 and 402, but only the first four are glycosylated. The sequence homology with the corresponding region of the previously published sequence of the neuraminidase subtype N1 [Fields, Winter & Brownlee (1981) *Nature (London)* **290**, 213–217] is 45%. Detailed evidence for the sequence data has been deposited as Supplementary Publication SUP 50116 (14 pages) at the British Library Lending Division, Boston Spa, Wetherby, West Yorkshire LS23 7BQ, U.K., from whom copies may be obtained on the terms given in *Biochem. J.* (1981) **193**, 5.

New epidemics of influenza are associated with the appearance of viral strains with major (antigenic shift) or minor (antigenic drift) changes in the structure of one or both of the two surface glycoproteins. The structural basis of antigenic variation in the major coat protein (the haemagglutinin) has recently been established (for review see Ward, 1981). Similar investigations are now in progress in this laboratory to characterize antigenic shift and drift in the minor surface glycoprotein, the neuraminidase, by gene nucleotide sequence, protein amino acid sequence and X-ray-crystallographic analysis.

The neuraminidase is a tetramer (Colman & Laver, 1981; Blok *et al.*, 1982) with a box-like head and a centrally attached stalk by which it is embedded in the viral membrane (Laver & Valentine, 1969). Proteolytic digestion releases the soluble neuraminidase heads from the membrane-attached stalk (Laver & Valentine, 1969; Allen *et al.*, 1977).

In a previous paper (Blok *et al.*, 1982) we reported that the neuraminidase heads from the Asian strain A/Tokyo/3/67 of influenza virus had a monomer molecular weight of 50 000, and contained 5600 daltons (11.2%) of carbohydrate for an apoprotein molecular weight of 44 000. We also showed that protein translation starts at the first AUG triplet in the complementary gene sequence, that the 7000–12 000-*M_r* (Lazdins *et al.*, 1972; Allen *et al.*, 1977) membrane-embedded stalk region removed by Pronase digestion comes from the

N-terminal end of the molecule and that no modification of the C-terminal end occurs during Pronase digestion.

In the present paper the complete amino acid sequence and oligosaccharide distribution for the Pronase-released heads of the influenza-virus strain-A/Tokyo/3/67 neuraminidase subtype N2 is reported.

Materials and methods

The influenza virus was a recombinant containing the neuraminidase of strain A/Tokyo/3/67 and the haemagglutinin of strain A/NWS/33. The procedures employed were as described in the cited references: virus cultivation and purification (Laver, 1969); neuraminidase release by Pronase digestion and purification by sucrose-density-gradient centrifugation (Laver, 1978); S-carboxymethylation with iodo[2-¹⁴C]acetate, CNBr cleavage and peptide separation on Sephadex G-100 in 50% (v/v) formic acid (Dopheide & Ward, 1978); enzymic digestion, peptide fractionation, amino acid analysis and sequence determination (Ward & Dopheide, 1979, 1981). Nucleic acid sequence determination was by the dideoxy chain-termination method (Smith, 1980) on total viral RNA with the use of a restriction fragment (nucleotides 1125–1235) from complementary DNA for cloned strain-A/RI/5-57 neuraminidase as primer. Detailed descriptions of the experimental procedures are contained in Supplementary Publication SUP 50116.

Results

CNBr cleavage of *S*-carboxymethylated neuraminidase heads (400 nmol) gave five peptides, CN1 (residues 74–160), CN2 (residues 161–307), CN3 (residues 308–362), CN4 (residues 363–467) and CN5 (residues 468–469), which could be partially resolved by gel filtration (Fig. 1). End-group analysis and the isolated peptides show that CNBr did not cleave the Met-241–Thr-242 bond. The peptide CN2/CN4 mixture (tubes 17–21), the peptide CN1/CN2 mixture (tubes 27–32) and two-thirds of the peptide CN3 sample (tubes 33–41) were each digested with trypsin, and the peptides were purified as indicated in the Materials and methods section. The large soluble tryptic peptide (residues 225–249) from peptide CN2 and the unpurified mixtures of large insoluble tryptic peptides (residues 188–210 and 211–224) from the digest of the peptide CN1/CN2 mixture and those (residues 188–210, 211–224 and 436–437) from the digest of the peptide CN2/CN4 mixture were not sequenced but were further digested with thermolysin. The large soluble tryptic peptide (residues 308–327) from peptide CN3 was not sequenced but was further digested with chymotrypsin. The mixture of peptides CN1, CN2 and CN4 in fractions 22–26 (Fig. 1) was digested with chymotrypsin, and the remaining one-third of the peptide CN3 sample was digested with thermolysin. The peptides isolated from these different digests are summarized in Fig. 2. Most peptides were purified to homogeneity and had analyses and electrophoretic mobilities consistent with their final sequence.

Fig. 1. Separation of CNBr-cleavage peptides CN1–CN5 from *S*-[2-¹⁴C]carboxymethylated influenza-virus strain-A/Tokyo/3/67 neuraminidase heads

The digest was fractionated on a column (150 cm × 1.0 cm) of Sephadex G-100 in 50% (v/v) formic acid. Peptides were identified by A_{280} (—) and radioactivity in 10 µl samples (----). The fraction size was 2.1 ml and the flow rate 4.2 ml/h. Further experimental details were as indicated in the Materials and methods section.

Fig. 2. Summary of the data used to determine the amino acid sequence of influenza-virus strain-A/Tokyo/3/67 neuraminidase heads. The peptide regions sequenced completely are shown by continuous lines. Broken lines denote peptides not sequenced but digested further with thermolysin or chymotrypsin, and those portions of peptides not sequenced. Details of the purification, amino acid composition, electrophoretic mobility and determination of amino acid sequence for each of these peptides are given in Supplementary Publication SUP 50116.

The peptide sequences collectively accounted for all of the structure except for a seven-residue section (residues 436–442) at the *N*-terminus of the large insoluble peptide from peptide CN4. They also provided unambiguous evidence for 33 of the 52 overlaps needed to align the tryptic peptides and the thermolysin or chymotryptic fragments of those five tryptic peptides digested further. The sequence of the missing seven residues was obtained indirectly by sequencing strain-A/Tokyo/67 viral RNA with the use of an internal restriction fragment (nucleotides 1125–1235, codons 373–405) from the cloned neuraminidase N2 gene of the closely related strain A/RI/5-/57 as primer. The sequence of the complementary DNA for this region was:

AAA-CAG-GAG-ACT-AGA-GTA-TGG-TGG-ACC-TCA-AAC-AGT-ATT-GTT-GTG-TTT-TGT-GGC-ACT
Lys-Gln-Glu-Thr-Arg-Val-Trp-Trp-Thr-Ser-Asn-Ser-Ile-Val-Val-Phe-Cys-Gly-Thr
431 435 440 445

which bridges the gap between residue 435 and 443. The alignment of the strain-A/Tokyo/3/67 neuraminidase peptides was completed by using the sequence data for the neuraminidase gene of strain A/RI/5-/57 (T. C. Elleman, A. A. Azad & C. W. Ward, unpublished work). The sequence homology between these two N2 neuraminidases is 93%.

The complete amino acid sequence of strain-A/Tokyo/3/67 neuraminidase heads is shown in Fig. 3. The peptide data and automated sequenator analysis of intact heads showed that Pronase cut the N2 neuraminidase in at least two positions to give two overlapping *N*-terminal sequences:

Glu-Lys-Glu-Ile-Cys-Pro-Lys-Val-Val-Glu-Tyr-Arg-Asn-Trp-Ser
Ile-Cys-Pro-Lys-Val-Val-Glu-Tyr-Arg-Asn-Trp-Ser

Nucleotide sequence data for the neuraminidase gene of strain A/RI/5-/57 (T. C. Elleman, A. A. Azad & C. W. Ward, unpublished work) show that these *N*-terminal glutamic acid and isoleucine residues occur at positions 74 and 77 in the original molecule. The sequence extends 396 residues to the C-terminal isoleucine residue at position 469. It contains five methionine, twelve tryptophan, nineteen half-cystine and four glycosylated asparagine residues (positions 86, 146, 200 and 234). There are five potential Asn-Xaa-Thr glycosylation sites in the sequence, but the asparagine residue at position 402 is not glycosylated.

Discussion

Antigenic analysis has shown that there are nine distinct neuraminidase subtypes found in the influ-

enza A viruses (Schild *et al.*, 1980), only two of which, N1 and N2, have been found in human strains.

Partial nucleotide sequence data (Blok, 1981) have shown that the *N*-terminal stalk region of the different neuraminidase subtypes showed no sequence homology apart from the first twelve residues. This lack of homology extends to residue 91 (Fig. 4), suggesting that residues 74–91 in the Pronase-released neuraminidase represent the last vestiges of the stalk. The data in Fig. 4 do show, however, that the enzymically active heads from the neuraminidase N1 (Fields *et al.*, 1981) and N2 subtypes are homologous. Sixteen of the half-cystine residues occur in identical positions in both

sequences, with additional cysteine residues occurring at positions 78, 175 and 193 (N2) and at position 160 (N1). Three of the glycosylation sites are found at identical (residues 146 and 234) or closely adjacent positions (residue 86/88) on the two proteins. The fourth oligosaccharide unit on the Asn-200 residue in the N2 protein is not present in the N1 neuraminidase.

The overall sequence homology between the heads of neuraminidase subtypes N1 and N2 is 45%, which is comparable with that found between the different haemagglutinin subtypes (for review see Ward, 1981). To maximize these structural rela-

tionships, gaps have been inserted in the aligned sequences. Blok (1981) showed there was a major deletion of 15 amino acid residues in N1 strains isolated between 1933 and 1935. In strain PR/8/34 this deletion involves residues 63–77 and occurs in the stalk region that contains the Pronase-cleavage site for the N2-subtype molecule (Fig. 4).

In addition to establishing the major structural changes associated with the antigenic shift from subtype N1 to subtype N2, the amino acid sequence of the strain-A/Tokyo/3/67 neuraminidase will be required for the interpretation of the electron-density map being derived from X-ray-crystallographic analysis of crystals of the strain-A/Tokyo/3/67 neuraminidase (Colman & Lever, 1981) and for the comparative structural analyses of other N2-subtype field strains and antigenic variants selected with monoclonal antibodies to the neuraminidase of this strain of influenza virus (Webster *et al.*, 1982).

80 90
 Glu-Lys-Glu-Ile-Cys-Pro-Lys-Val-Val-Glu-Tyr-Arg-Asn-Trp-Ser-Lys-Pro-
 100 110 120
 Gln-Cys-Gln-Ile-Thr-Gly-Phe-Ala-Pro-Phe-Ser-Lys-Asp-Asn-Ser-Ile-Arg-Leu-Ser-Ala-Gly-Gly-Asp-Ile-Trp-Val-Thr-Arg-Glu-Pro-
 130 140 150
 Tyr-Val-Ser-Cys-Asp-Pro-Val-Lys-Cys-Tyr-Gln-Phe-Ala-Leu-Gly-Gln-Gly-Thr-Thr-Leu-Asp-Asn-Lys-Ser-Asn-Asp-Thr-Val-His-
 160 170 180
 Asp-Arg-Ile-Pro-His-Arg-Thr-Leu-Leu-Met-Asn-Glu-Leu-Gly-Val-Pro-Phe-His-Leu-Gly-Thr-Arg-Gln-Val-Cys-Ile-Ala-Trp-Ser-Ser-
 190 200 210
 Ser-Ser-Cys-His-Asp-Gly-Lys-Ala-Trp-Leu-His-Val-Cys-Ile-Thr-Gly-Asp-Asp-Asn-Ala-Thr-Ala-Ser-Phe-Ile-Tyr-Asp-Gly-Arg-
 220 230 240
 Leu-Val-Asp-Ser-Ile-Gly-Ser-Trp-Ser-Gln-Asn-Ile-Leu-Arg-Thr-Gln-Glu-Ser-Glu-Cys-Val-Cys-Ile-Asn-Gly-Thr-Cys-Thr-Val-Val-
 250 260 270
 Met-Thr-Asp-Gly-Ser-Ala-Ser-Gly-Arg-Ala-Asp-Thr-Arg-Ile-Leu-Phe-Ile-Glu-Glu-Gly-Lys-Ile-Val-His-Ile-Ser-Pro-Leu-Ala-Gly-
 280 290 300
 Ser-Ala-Gln-His-Val-Glu-Cys-Ser-Cys-Tyr-Pro-Arg-Tyr-Pro-Gly-Val-Arg-Cys-Ile-Cys-Arg-Asp-Asn-Trp-Lys-Gly-Ser-Asn-Arg-
 310 320 330
 Pro-Val-Val-Asp-Ile-Asn-Met-Glu-Asp-Tyr-Ser-Ile-Asp-Ser-Ser-Tyr-Val-Cys-Ser-Gly-Leu-Val-Gly-Asp-Thr-Pro-Arg-Asn-Asp-Asp-
 340 350 360
 Arg-Ser-Ser-Asn-Ser-Asn-Cys-Arg-Asn-Pro-Asn-Asn-Glu-Arg-Gly-Thr-Gln-Gly-Val-Lys-Gly-Trp-Ala-Phe-Asp-Asn-Gly-Asn-Asp-Leu-
 370 380 390
 Trp-Met-Gly-Arg-Thr-Ile-Ser-Lys-Asp-Leu-Arg-Ser-Gly-Tyr-Glu-Thr-Phe-Lys-Val-Ile-Gly-Gly-Trp-Ser-Thr-Pro-Asn-Ser-Lys-Ser-
 400 410 420
 Gln-Ile-Asn-Arg-Gln-Val-Ile-Val-Asp-Ser-Asp-Asn-Arg-Ser-Gly-Tyr-Ser-Gly-Ile-Phe-Ser-Val-Glu-Gly-Lys-Ser-Cys-Ile-Asn-Arg-
 430 440 450
 Cys-Phe-Tyr-Val-Glu-Leu-Ile-Arg-Gly-Arg-Lys-Gln-Glu-Thr-Arg-Val-Trp-Trp-Thr-Ser-Asn-Ser-Ile-Val-Val-Phe-Cys-Gly-Thr-Ser-
 460
 Gly-Thr-Tyr-Gly-Thr-Trp-Pro-Asp-Gly-Ala-Asn-Ile-Asn-Phe-Met-Pro-Ile

Fig. 3. Amino acid sequence of influenza-virus strain A/Tokyo/3/67 neuraminidase heads

The numbers refer to the positions in the intact neuraminidase, Pronase digestion removing 73–76 amino acid residues from the N-terminal end. The glycosylated asparagine residues are underlined. The fifth site, at Asn-402, does not carry carbohydrate.

20 40 60
 N1 MNPNQKIITIGSICLVVGLI SLILQIGNIISIWISHSIQT GSNHTGICNQNIITYKNST LW-----KDT TSVILTCNSSLCPIRGMAIY
 N2 EKEICPK VVEYRNWSKPOQITGFAPP 80 100
 N1 SONSTRIGSKGVNIREP FISCCHLECRITFFETQALL NDRSHGTRKRSFYRA-ILM SCPVGEARSPYNRSFESVAN SASACHDGGWLTIGISPD
 N2 SONSIRLSACGDIWTRP YVSGDPVKCYQFALCGQTIL DNKESNDIVHRIPIHRTLLM NE-LG-VPHLCTRGVCIAW SSSSCHDGKAMLHVCLTDD 160 178 198
 N1 NCATAVLEKNGIITETIKM RKKILRTQESCEACVNGSCF TIMDGPDSGLASVYKFKIE KGVTKSIELNAPNSHYEEC SCYPTDCKFMVCGRNNHGS
 N2 DNATASFIDGRLVDSIGM SONILRTQESCEVINGTCT VVMTDSASERADTRILFIE ECKIVHISPLAGSAQRVEEC SCYTPYPCVRICRDNKNGS 258 278 298
 N1 NEPWFSDONLDYQIG--YI CSGVFCDNPRKDTG-G-S-C -GPVVYDGAANGVKGFSYRYG NGVILGRKTKSHSRHGFEM WDPNGMTETDTSKFSV-RQDV
 N2 NEPVDVINE-DYSDSSV CSGLVGDTPNDRRSSNMC RNPNNERGTQGVKGWAFDNG NDLNGRITISKDLRSGYETTF KVICGNSTPNKSKSQINRQVI 337 357 377 397
 N1 VAMTDHSCYSSCFVQHPILT GLDCIRPCFWELINGRPE KTI-MTSASSISFCGVNSDT VDSNPDDGAELEPFTIDK
 N2 VDSNRSCTSGTIF--SVE-- GKSCINRCFWELINGRKE TRVMWTSNSIVVFCGTSGTY GTCSNPDDGANINEMPI 413 433 453 469

Fig. 4. Comparative amino acid sequences for neuraminidases from influenza virus strains A/PR/8/34 subtype N1 and A/Tokyo/3/67 subtype N2

The data for the sequence of the N1 enzyme are from Fields *et al.* (1981). Regions of identity are shaded. For the key to the one-letter notation for the amino acid residues see *Biochem. J.* (1969) 113, 1–4.

We thank Dr. W. G. Laver for supplying the neuraminidase material under contract, Dr. A. S. Inglis for the automated sequenator run, Dr. C. M. Roxburgh for the amino acid analyses and Mr. Neil Cook and Ms. Susan Barrett for technical assistance.

References

- Allen, A. K., Skehel, J. J. & Yuferov, V. (1977) *J. Gen. Virol.* **37**, 625–628
- Blok, J. (1981) in *Genetic Variation among Influenza Viruses* (Nayak, D. P., ed.), pp. 45–54, Academic Press, New York
- Blok, J., Air, G. M., Laver, W. G., Ward, C. W., Lilley, G. G., Woods, E. F., Roxburgh, C. M. & Inglis, A. S. (1982) *Virology* **119**, 109–121
- Colman, P. M. & Laver, W. G. (1981) in *Structural Aspects of Recognition and Assembly in Biological Macromolecules* (Balaban, M., Sussman, J. L., Traub, W. & Yonath, A., eds.), pp. 869–872, Balaban ISS, Rehovot and Philadelphia
- Dopheide, T. A. & Ward, C. W. (1978) *Eur. J. Biochem.* **85**, 393–398
- Fields, S., Winter, G. & Brownlee, G. (1981) *Nature (London)* **290**, 213–217
- Laver, W. G. (1969) in *Fundamental Techniques in Virology* (Habel, K. & Salzman, N. P., eds.), pp. 82–86, Academic Press, New York
- Laver, W. G. (1978) *Virology* **86**, 78–87
- Laver, W. G. & Valentine, R. C. (1969) *Virology* **38**, 105–119
- Lazdins, I., Haslam, E. A. & White, D. O. (1972) *Virology* **49**, 758–765
- Schild, G. C., Newman, R. W., Webster, R. G., Major, D. & Hinshaw, V. S. (1980) in *Structure and Variation in Influenza Virus* (Laver, W. G. & Air, G. M., eds.), pp. 373–383, Elsevier, New York
- Smith, A. J. H. (1980) *Methods Enzymol.* **65**, 560–580
- Ward, C. W. (1981) *Curr. Top. Microbiol. Immunol.* **94/95**, 1–74
- Ward, C. W. & Dopheide, T. A. (1979) *Virology* **95**, 107–118
- Ward, C. W. & Dopheide, T. A. (1981) *Biochem. J.* **193**, 953–962
- Webster, R. G., Hinshaw, V. S. & Laver, W. G. (1982) *Virology* **117**, 93–104