Volume 10 Number 16 1982 Nucleic Acids Research

Sequence of the N2 neuraminidase from influenza virus A/NT/60/68

D.R.Bentley and G.G.Brownlee

Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK

Received 11 July 1982; Accepted 22 July 1982

ABSTRACT

The complete sequence of the neuraminidase gene of influenza virus A/NT/60/68 (N2 subtype) was determined following cloning of full length complementary DNA into pBR322. Comparison of the predicted amino acid sequence with a closely related neuraminidase from A/Udorn/72 suggests that point mutations over an extensive region of the primary sequence can contribute to antigenic drift, although the region between amino acid residues 308 and 371 may be particularly significant.

INTRODUCTION

All influenza A viruses contain a neuraminidase surface glycoprotein in addition to their haemagglutinin. Like the haemagglutinin, the neuraminidase is well known to undergo antigenic variation by the accumulation of a small number of point mutations - antigenic drift, or by more extensive change through genetic reassortment - antigenic shift (1,2). "Shift" in the human influenza viruses occurred in 1957 when the H2N2 subtype replaced the H1N1 subtype with the emergence of Asian "flu". Since then, the H1N1 subtype has recycled as shown by the emergence of the Russian "flu" in 1977.

Although the haemagglutinin is the more significant antigen in the immune response of the host, the neuraminidase is also involved, as anti-neuraminidase antibodies can inhibit viral replication in animal experiments (3,4). It has been suggested in Man that anti-neuraminidase antibody limits the spread of the virus in the host, thus reducing the severity and extent of disease (5). The protection conferred by anti-neuraminidase antibodies may have been responsible (6) for the minor epidemic impact in the U.K. of the Hong Kong virus in 1968-69 (in contrast to 1969-70) despite the "shift" in haemagglutinin subtype.

In order to understand the variation of the neuraminidase at the molecular level, we previously used recombinant DNA methods to clone and sequence (7) the neuraminidase of human influenza A/PR/8/34, - a prototype

0305-1048/82/1016-5033$2.00/0

strain of the N1 subtype. Here we use related methods to sequence the neuraminidase of a human influenzal strain A/NT/60/68 as an example of an N2 subtype. We compare our results with the two presently known N1 (7,8) and a recently reported N2 sequence (9) in an attempt to define important antigenic regions of the molecule and a possible active site.

MATERIALS AND METHODS

Synthesis and cloning of double-stranded DNA by blunt-end ligation into phosphatase treated, PvuII cut, pBR322

Full-length double-stranded DNA copies of segment 6 of A/NT/60/68 virion RNA were prepared as previously described (10,11). Briefly, single-stranded cDNA was prepared from total virion RNA using the synthetic dodecanucleotide primer d(pA-G-C-A-A-A-A-G-C-A-G-G) (phosphorylated at its 5' terminus) which hybridises to the 3' end of all the viral RNA segments (12), and AMV reverse transcriptase (a gift from Dr J W Beard). Following purification of the band 6 cDNA from this mixture by electrophoresis using a denaturing 3% polyacrylamide gel, a second specific 13 long deoxyoligonucleotide primer, d(pA-G-T-A-G-A-A-A-C-A-A-G-G) (also phosphorylated at its 5' terminus) was used in the back-copy reaction with the "Klenow" subfragment of E.coli DNA polymerase I (Boehringer). We carried out blunt-end ligation of an aliquot of the double-stranded DNA with calf intestinal phosphatase-treated PvuII cut pBR322 (11), followed by transformation of competent E.coli X1974 cells. Transformants were screened by the Grunstein-Hogness procedure (13) using ^{32}P-labelled short-copy cDNA as a probe to identify thirty-one hybridisation-positive colonies. Eleven of these were selected and small-scale plasmid preparations carried out (14); when resolved by 1% agarose gel electrophoresis, five were longer than the pBR322 controls (results not shown). DNA from three of the five potential full-length clones was used to transform E.coli HB101 (15). One of the clones (NT60.6.7) was selected and plasmid DNA extracted and purified from a 1 ℓ culture by the alkali extraction procedure (16) followed by caesium chloride density gradient centrifugation.

Nucleotide sequence analysis of clone NT60.6.7 Sequence determination was performed by the chemical degradation method as previously described (17), except that the formic acid protocol was used for the "A+G" reaction. From the results, we established that both the oligonucleotide primers used in the cloning were present in the plasmid and that the clone NT60.6.7 was therefore full-length.

RESULTS & DISCUSSION

Sequence of the neuraminidase gene of A/NT/60/68

The sequence of the neuraminidase gene is shown in the mRNA sense in Fig. 1. It contains one open reading frame (nucleotides 20-1426) coding for a protein of 469 amino acids with a calculated molecular weight of 52,084 daltons (excluding carbohydrate). The size of the protein is identical to A/Udorn/72 (9) but slightly longer than the 454 residues of the N1 neuraminidase of A/PR/8/34 (7) or the 453 residues of A/WSN/33 (8). The protein contains one markedly hydrophobic region, at residues 7-35 and the remaining sequence has eight potential glycosylation sites (see Fig. 1), which, by analogy with the haemagglutinin are probably glycosylated, although we do not expect both Asn 69 and Asn 70 to be modified.

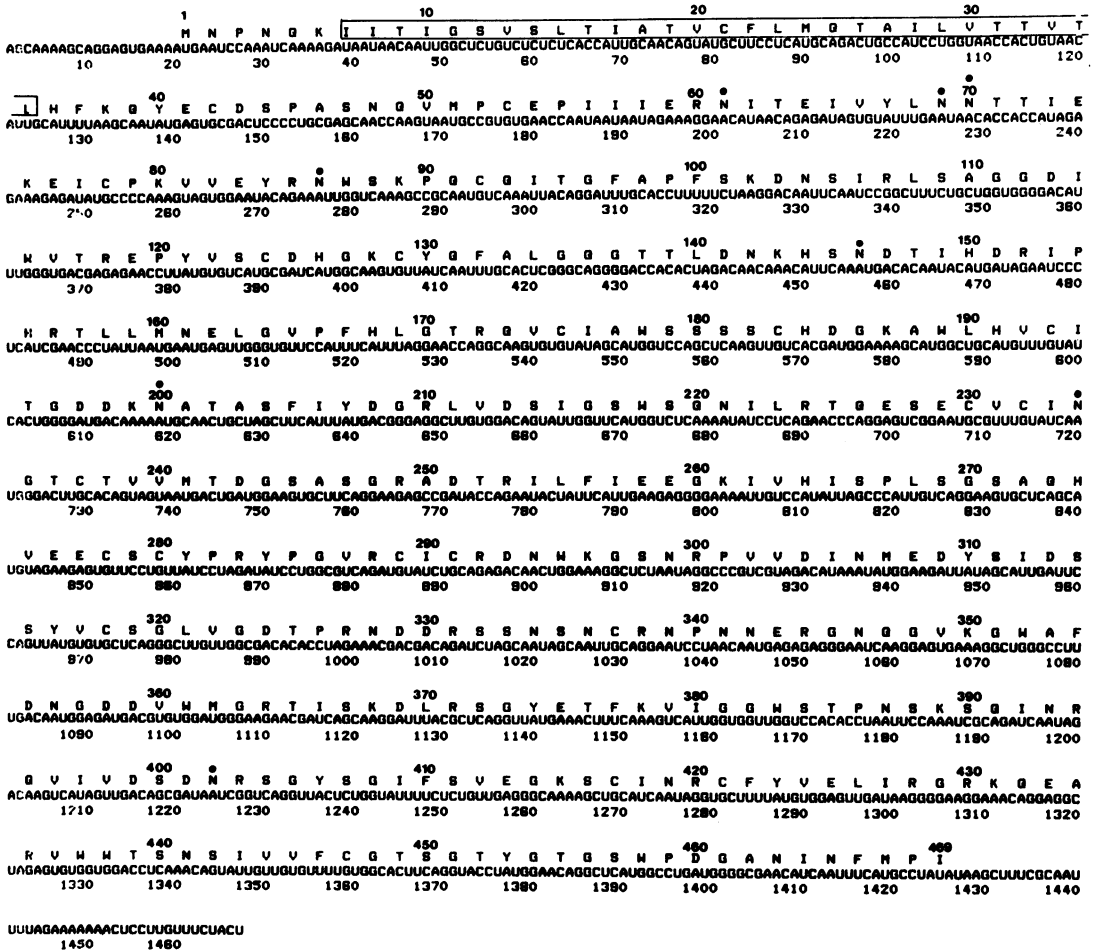
The amino acid composition of the N2 neuraminidase is shown in Table 1 and is very rich in cysteine residues (4.7% of the total compared with 3.2% for the haemagglutinin (18), and 1.0 and 1.3% for the internal virus polypeptides P1 protein (19) and the nucleoprotein (11) respectively). It also has a high content of serine and threonine (together 17%); this value is close to that for the HA1 subunit of the haemagglutinin (18% serine and threonine), but contrasts with the HA2 subunit (7.7%) and other influenza proteins (18).

Table 1

Amino acid composition of neuraminidase of A/NT/60/68

Amino Acid	Molar %	Amino Acid	Molar %
Cys (C)	4.7	His (H)	2.1
Gly (G)	8.3	Lys (K)	3.8
Pro (P)	4.1	Arg (R)	5.8
Ser (S)	10.0	Asp (D)	6.0
Thr (T)	6.8	Glu (E)	4.5
Ala (A)	3.4	Asn (N)	6.8
Val (V)	7.3	Gln (Q)	3.4
Leu (L)	4.3	Phe (F)	3.0
Ile (I)	8.5	Tyr (Y)	3.0
Met (M)	3.4	Trp (W)	2.6

Total number of amino acids = 469



Comparison of N1 and N2 neuraminidases - antigenic shift

In Figure 2, the amino acid sequence of the neuraminidase of A/NT/60/68 (N2 subtype) is aligned with that of A/PR/8/34 (7), as an example of a prototype N1 subtype, to maximise their overall homology. Various blocks of amino acids, from one to twelve residues in length, are conserved throughout the molecule, together accounting for a 39% homology. In detail, sixteen out of the total of twenty-two cysteine residues are conserved in position; glycine and proline residues are also highly conserved (72% and 47% respectively). As these residues confer constraints upon the conformation of the polypeptide, this suggests that the different neuraminidases may share a greater three-dimensional structural similarity than is evident from the 39% homology of their primary sequences. Further, because most of the cysteine residues are clustered (note especially the Cys-X-Cys sequences at 231-233, 279-281 and 290-292, Fig. 2), parts of the globular head of the protein are likely to be structurally rigid.

The overall primary sequence homology 39% is similar to the value (41%) when the haemagglutinin molecules of these two viruses (H3 and H1) are compared; and markedly differs from the values derived by comparison of the nucleoprotein (94%) (11) and the P1 protein (95%) (20) of the same two strains. The antigenic "shift" of the neuraminidase therefore appears to be as extensive as that of the haemagglutinin. Similar conclusions can be drawn if the A/NT/60/68 neuraminidase is compared with the A/WSN/33 sequence (8), as would be expected considering that there are only 33 amino acid differences between the A/PR/8/34 and A/WSN/33 sequences (8).

The markedly hydrophobic region of the neuraminidase of A/NT/60/68 (Fig. 1) is exactly colinear with the hydrophobic region of the N1 neuraminidase (at residues 7-35). Potential glycosylation sites of both neuraminidases (see arrows in Fig. 2) are mostly in the N-terminal half of the molecule, and are particularly clustered in the region of amino acids 40-90, although the precise positions of these carbohydrate residues is not conserved.

A Chou-Fasman (20) and Robson (21) analysis of secondary structure predicts that both neuraminidases are rich in β-sheet and have little α-helix. In detail, the neuraminidase of A/NT/60/68 contains 16 regions of

Figure 1. The nucleotide sequence of the neuraminidase of A/NT/60/68 in the mRNA sense. The amino acid sequence is listed in one-letter code (see Table 1); the large box delineates the hydrophobic N-terminal region, and the solid circles (●) indicate potential glycosylation sites.

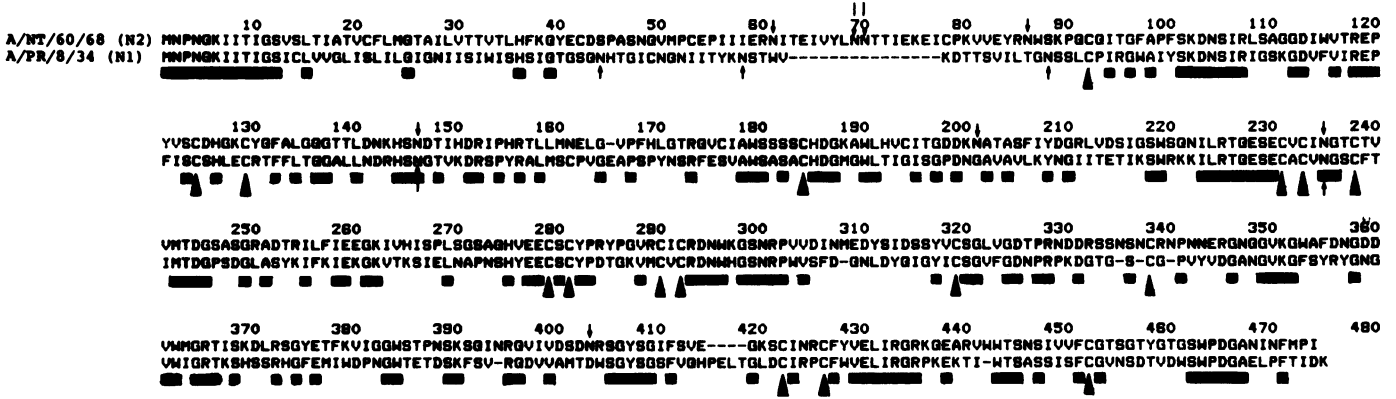


Figure 2. Comparison of the amino acid sequences of A/NT/60/68 (N2 subtype) and A/PR/8/34 (N1 subtype). The alignment required the occasional insertion of gaps in both sequences, and the inclusion of a space of fifteen gaps immediately after residue 62 of the A/PR/8/34 sequence. Amino acids conserved in both sequences are underlined (▬) except for conserved cysteines, which are indicated by triangles (▲). Potential glycosylation sites in both sequences are shown by arrows (↓).

predicted β-sheet and only one of α-helix (results not shown), all of which lie in the globular head domain. Thus the secondary structure of the neuraminidase would appear to resemble the HA1 component of the haemagglutinin (22).

The domain organisation and a possible active site of the neuraminidase

From our results and others (9,23), we define four domains of the molecule (see Fig. 3). We have previously postulated that the membrane attachment site of the neuraminidase is near the N-terminus (domain "B" in Fig. 3) (7). Recent analysis (24) of tryptic peptides of the neuraminidase of A/RI/5^{+}/57 (N2 subtype) has provided direct confirmatory evidence.

The neuraminidase and the HA1 component of the haemagglutinin both bind terminal sialic acid residues of carbohydrate chains and it is possible that they have similar substrate binding sites. We have therefore compared the regions of conserved neuraminidase sequence with those of all the influenza haemagglutinins for which the sequence is known. We find that the tripeptide cysteine-tyrosine-proline occurs in all neuraminidases of both subtypes so far sequenced at positions 281-283 (see Fig. 2), and is also present in all the haemagglutinins of all influenza "A" viruses so far sequenced at position 97-99 (22). [Influenza "B" strains may differ in that this tyrosine residue is replaced in B/Lee/40 by a phenylalanine residue (25).] Following the postulate (22) that the tyrosine of this tripeptide is part of the receptor binding site of the haemagglutinin, we propose that it could be part of the active site of the neuraminidase.

Comparison of A/NT/60/68 and A/Udorn/72 neuraminidases - antigenic drift

A comparison of the amino acid sequences of these two neuraminidases (sequence reported here and ref. 9) shows 28 amino acid point mutations (Table 2). Among these 28 must be some which significantly alter antigenicity and others with lesser or no antigenic significance. We can reasonably exclude changes at amino acids 20, 26 and 29 and those at amino acids 47, 69, 81 and 93 as being within or very close to the hydrophobic membrane attachment domain or "stalk" domain of the molecule. This leaves 21 point mutations, of which those at amino acids 195, 254 and 345 represent conservative amino acid changes (I → V, R → K) which are unlikely to be important antigenically.

The remaining 18 changes span residues 126-471 in the primary sequence but it is noticeable that the greatest density of change (9 changes) occur within 70 residues (i.e. between 308 and 371) and there are no changes

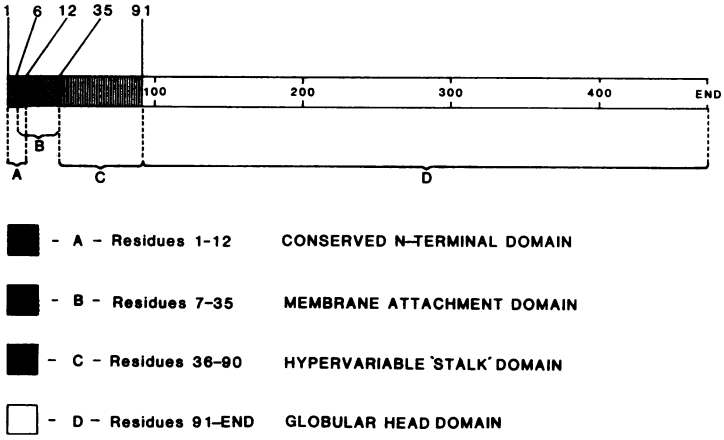


Figure 3. Domain organisation of the neuraminidase. Note that domain "A" and "B" overlap, as residues 7-12 (the overlapping region) have characteristics of both domains.

Table 2

Amino acid changes between A/NT/60/68 and A/Udorn/72

Position (Fig. 2)	Mutation	Position (Fig. 2)	Mutation
20	V ⟶ I	309	E ⟶ K
26	T ⟶ I	330	D ⟶ N
29	L ⟶ Q	337	N ⟶ Y
47	S ⟶ N	345	R ⟶ K
69	N ⟶ T	348	Q ⟶ H
81	V ⟶ L	357	N ⟶ D
93	Q ⟶ K	359	D ⟶ N
126	H ⟶ P	369	K ⟶ E
153	I ⟶ T	371	L ⟶ S
178	A ⟶ G	391	S ⟶ L
195	I ⟶ V	436	K ⟶ E
198	D ⟶ Y	439	A ⟶ T
254	R ⟶ K	468	N ⟶ D
308	M ⟶ V	471	F ⟶ L

(excluding the conservative change at amino acid 254) between residues 198 and 307. Within this high density region we note that (i) changes occur at the adjacent residues 308 and 309 and that this is the longest variable region when comparing N1 and N2 sequences (Fig. 2), (ii) residues 330 and 337 are also in a variable region when comparing N1 and N2 sequences in the sense that three deletions had to be placed in the A/PR/8/34 sequence in this region to maintain alignment.

We favour the hypothesis that some, at least, of the point mutations between residues 308 and 371 will be important antigenically. Formally, however, we must conclude from our sequence comparisons that amino acid change over a wide region of the primary sequence of the 'globular' domain could be important antigenically. A knowledge of the three dimensional structure of the neuraminidase, now in progress for A/Tokyo/67 (26), as well as the isolation and location of many single amino acid variants in the laboratory, will be required before a definitive statement of the number and location of antigenic sites is possible.

ACKNOWLEDGEMENTS

We thank Mrs Joyce Huddleston for assistance in cloning, and the Medical Research Council for financial support.

REFERENCES

1. Paniker, C.K.J. (1968) J.Gen.Virol. 2, 385-394.
2. Schulman, J.L. and Kilbourne, E.D. (1969) Proc.Natl.Acad.Sci.USA 63, 326-333.
3. Schulman, J.L., Khakpour, M. and Kilbourne, E.D. (1968) J.Virol. 2, 778-786.
4. Rott, R., Becht, H. and Orlich, M. (1974) J.Gen.Virol. 22, 35-41.
5. Murphy, B.R., Kasel, J.A. and Channock, R.M. (1972) New England J.of Med. 286, 1329-1332.
6. Stuart-Harris, C. (1979) Br.Med.Bull. 35, 3-8.
7. Fields, S., Winter, G. and Brownlee, G.G. (1981) Nature 290, 213-217.
8. Hiti, A.L. and Nayak, D.P. (1982) J.Virol. 41, 730-734.
9. Markoff, L. and Lai, C.-J. (1982) Virology 119, 288-297.
10. Winter, G., Fields, S., Gait, M.J. and Brownlee, G.G. (1981) Nucl.Acids Res. 9, 237-245.
11. Huddleston, J.A. and Brownlee, G.G. (1982) Nucl.Acids Res. 10, 1029-1038.
12. Skehel, J.J. and Hay, A.J. (1978) Nucl.Acids Res. 5, 1207-1219.
13. Rothstein, R.J., Lall, L.F., Bahl, C.P., Narang, S.A. and Wu, R. (1979) Methods in Enzymol. 68, 101-110.
14. Holmes, D.S. and Quigley, M. (1981) Analytical Biochem. 114, 193-197.
15. Boyer, H. and Rouland-Dossoix, D. (1969) J.Mol.Biol. 41, 459-472.
16. Birnboim, H.C. and Doly, J. (1979) Nucl.Acids Res. 7, 1513-1523.
17. Maxam, A.M. and Gilbert, W. (1980) Methods in Enzymol. 65, 499-560.
18. Winter, G., Fields, S. and Brownlee, G.G. (1981) Nature 292, 72-75.

19. Bishop, D.H.L., Huddleston, J.A. and Brownlee, G.G. (1982) Nucl.Acids Res. 10, 1335-1343.
20. Chou, P. and Fasman, G.D. (1974) Biochemistry 13, 212-245.
21. Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) J.Mol.Biol. 120, 97-120.
22. Wilson, I.A., Skehel, J.J. and Wiley, D.C. (1981) Nature 289, 366-373.
23. Blok, J. (1982) ICN-UCLA Symp.Molec.Cell Biol. 22, 45-54.
24. 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.
25. Elliot, R.M., personal communication.
26. Colman, P.M., Tulloch, P.A. and Laver, W.G. (1981) in Structure and Variation in Influenza Virus, Laver, W.G. and Air, G.M., Eds., 351-356, Elsevier, New York.