

# The three-dimensional structure and antigenic variation of the influenza virus haemagglutinin

D. C. WILEY<sup>1</sup> AND J. J. SKEHEL<sup>2</sup>

<sup>1</sup>*Fairchild Laboratory, Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA.*

<sup>2</sup>*Division of Virology, National Institute for Medical Research, Mill Hill, London NW7 1 AA, England*

## INTRODUCTION

The three-dimensional structure of the major glycoprotein antigen of influenza virus, the haemagglutinin (HA), was determined by high resolution protein X-ray crystallography to allow a more detailed study of the functions of the molecule (1). The HA is active in at least three stages of a virus infection: 1. The HA contains a binding site which recognizes and binds to a sialic-acid-containing receptor on a target cell to initiate an infection (2). 2. The HA undergoes a low-pH induced conformational change which activates a membrane-penetration activity that delivers the viral nucleocapsid into the target cell's cytoplasm (3). 3. The HA undergoes antigenic variation which is associated with the uncontrolled recurrence of influenza epidemics in man (4).

After a brief description of the three-dimensional structure of the HA from the 1968 Hong Kong influenza virus (A/AICHI/68), this review will focus on the antigenic structure of the HA.

## STRUCTURE DESCRIPTION

The HA is a trimeric glycoprotein of molecular weight 224,640 daltons which contains 19% carbohydrate (by weight). The HA monomer, although synthesized as a simple polypeptide chain, is normally post-translationally cleaved into two polypeptide chains HA<sub>1</sub> (328 AA) and HA<sub>2</sub> (221 AA), which remain covalently linked by a disulfide bond. For crystallization, the HA released from the virus membrane by removing a 5,406 dalton c-terminal, hydrophobic anchoring peptide from HA<sub>2</sub> by digestion with bromelain (4, 5).

The three-dimensional structure of the monomer is shown in *figure 1*. The N-Terminus of HA<sub>1</sub> is at the membrane end (bottom of *Fig. 1*) of the molecule. The HA<sub>1</sub> chain stretches up the structure 96 Å before folding into a Beta-sheet-rich (flat arrows, *Fig. 1*) compact globular domain at the top

of the structure. This globular domain contains the receptor binding site and some prominent loops and a helix at the distal end, which are antigenically important (6). The HA<sub>1</sub> chain then returns toward the membrane-end of the molecule where it composes part of a stem-like structure, which supports the distal globular region. The HA<sub>2</sub> polypeptide begins 22 Å away from the c-terminus of HA<sub>1</sub> (C1 of *Fig. 1*), indicating that a conformational change must have accompanied the

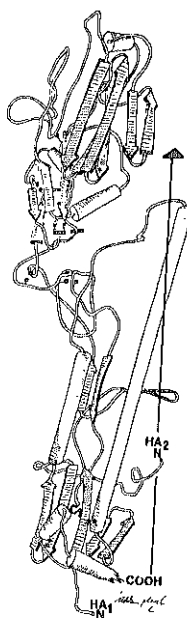


FIG. 1. Schematic drawing of the structure of the 1968 haemagglutinin monomer. Cylinders represent alpha-helices. Flat ribbons represent strands of beta sheet. The N and C termini of HA<sub>1</sub> are marked N and C<sub>1</sub>, the N and C termini of HA<sub>2</sub> are marked N and COOH. A solid line represents the position of the molecular three-fold symmetry axis that relates the three monomers of the HA trimer.

post-translational cleavage between HA<sub>1</sub> and HA<sub>2</sub>. The HA<sub>2</sub> chain includes two prominent alpha helices (cylinders in *Fig. 1*), the longest of which forms the backbone of the stem region of the structure.

The threefold molecular symmetry axis relating monomers within the trimer is shown as a solid vertical line. The trimer is formed primarily by the association of a 50 Å long hydrophobic face on the long alpha-helix with similar regions on three monomers. The result is a triple-stranded coil-coil of alpha-helices forming the core of the trimeric "stem". The amino-terminal of HA<sub>2</sub> is tucked into this hydrophobic core from which it emerges at low pH to take part in the membrane penetration activity (3).

### ANTIGENIC VARIATION

Since the first isolation of an influenza virus in humans in 1933 (7), two human pandemics have been recorded (1957 and 1968) which were associated with viruses with antigenically distinct HAs. The viral HAs from these pandemics were designated H2 in the case of the Asian strains prevalent from 1957-1968 and H3 for the Hong Kong virus strains which appeared in humans in 1968 and are still in circulation (8). H1 containing strains, which circulated between 1933-1957, have reappeared recently (1977) and now co-circulate with the H3 type strains. Amino-acid sequences of HA from each pandemic era show only 40-60% sequence homology, indicating that HAs from novel viral subtypes have entered the human population at the start of each pandemic (for review, see 10, 11).

Within each pandemic era, periodic epidemics have been recorded which are associated with virus strains, which have undergone 'antigenic-drift' (12) from the prototype strains. Amino acid sequence homologies among 'drifted' strains are high, 80-90% (10, 11).

Since neutralizing antibodies against influenza virus are directed at the HA glycoprotein (13), studies of the amino acid sequences of drifted strains (for review, see 11) and analyses of antigenic variants selected by growing virus in the presence of monoclonal antibodies against the HA (14, 15, 16), have been combined with knowledge of the HA three-dimensional structure to begin to provide a molecular explanation for the epidemiology of influenza infection (6).

### ANTIGENIC VARIATION IN THE HONG KONG (H3) VIRUSES

Since the initial pandemic of 1968 caused by the

influenza virus strains similar to A/Hong Kong/1/68, epidemics have occurred in 1972/73, caused by strains similar to A/Memphis/102/72, 1975/76 caused by strains similar to A/Victoria/3/75, and 1979/80 caused by strains similar to A/Bangkok/1/79. The amino acid substitutions in the 1972, 1975, and 1979 disease causing strains were located in the three-dimensional structure of the HA (*Fig. 2a, b, d*). The amino acid sequences of these HAs have been determined by protein and nucleic acid sequencing techniques (10, 17, 18, 19, 20). Fourteen amino acid substitutions were detected in the HA<sub>1</sub> polypeptide chain of A/Memphis/102/72 relative to that of the A/AICHI/2/68 virus, a further 13 in HA<sub>1</sub> of A/Victoria/1/75, and a further 14 in HA<sub>1</sub> of A/Bangkok/1/79. A detailed summary of the structural locations and possible stereochemical consequences of these substitutions is presented elsewhere (6, 20, 21).

Further definition of the regions of the 1968 HA, which may be involved in antibody binding, is provided by studies of antigenic variants selected by growth in monoclonal antisera against the HA (14, 15, 21, 22, 23). *Figure 2c* shows the location of single amino acid positions from 8 variants of the A/Memphis/1/71 virus strain (6). *Figure 2e* shows the location of the single amino acid position from 12 antigenic variants of the A/AICHI/1/68 HA from the X:31 strain of the 1968 Hong Kong influenza virus (15, 21).

### SUGGESTED LOCATION OF ANTIGENIC SITES

All of this data is summarized in *figure 2f*, where it is evident that amino acid substitutions on a large fraction of the surface of the globular domain of HA<sub>1</sub> appear to affect antibody binding. On the basis of small cross reactivity among classes of monoclonal antibodies with the antigenic variants and on the basis of the clustering of amino acid substitutions in the natural strains, it is possible to separate the antigenic surface into regions labeled by five different symbols in *figure 2f* (6, 21). Each one of these proposed antigenic sites is identified by a class of monoclonal antibody selected variants. A (closed spheres in *Fig. 2f*), B (squares), and E (inverted triangles) are defined by monoclonal antibodies directed against the A/AICHI/1/68 HA (15, 21), while A, C (triangles) and D (diamonds) were defined by monoclonal antibodies directed against A/MEM/1/71 (6, 14). Furthermore, each site AE, has accumulated at least one amino acid substitution between each reoccurrence of influenza epidemics in the human population since 1968 (see *Fig. 2a, b, d*).

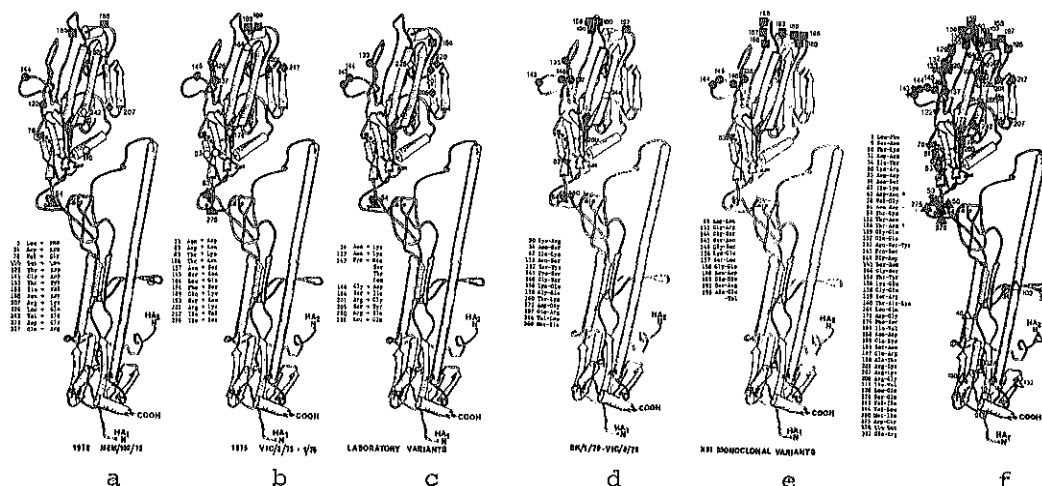


Fig. 2. Schematic drawing of the monomer of the 1968 haemagglutinin showing locations of amino acid substitutions on HA<sub>1</sub>, ●, site A; ■, site B; ▲, site C; ◇, site D; ▽, site E.

- (a) Amino acid changes from AICHI/2/68 to MEM/102/72 (17, 18, 10).  
 (b) Amino acid changes from MEM/102/72 to VIC/3/75 (19).  
 (c) Laboratory selected antigenic variants of MEM/1/71 (14).  
 (d) Amino acid changes from VIC/3/75 to Bangkok/1/79 (20).  
 (e) Laboratory selected antigenic variants of X:31 (AICHI/2/68-HA) from growth in monoclonal antisera (15, 21).  
 (f) Summary of all known variation from 1968 to 1979, including laboratory selected variants.

A three-dimensional (stereo) drawing of the same view of the molecule is found in Figure 3 of reference 1.

Site A includes part of an unusual protruding loop of amino acids (140-146) and the adjacent surface of HA<sub>1</sub>. The amino acids ala 138, cys 139, phe 147 and phe(tyr) 148 occurring at each end of this loop are conserved in all thirty HA sequences available, which include HAs from all three pandemic eras as well as avian and equine viruses. Thus the structural foundation of this antigenic site is conserved in known HA<sub>2</sub>.

Site B is centered on a loop of amino acids 155-160 and the region 188-198 which includes an  $\alpha$  helix, both located at the distal tip of the HA<sub>1</sub> domain. This site was originally defined on the basis amino acid changes in the natural epidemic strains since 1968 (6). However, more recent data from antigenic variants selected by monoclonal antibodies indicate that it can be split into two partially overlapping sites B1 and B2, which comprise opposite sides of the region (21). Like site A, the structural potential for site B is conserved in other haemagglutinins (6).

Site C is a bulge in the tertiary structure of the HA, 60 Å from the distal tip in a region containing a highly conserved disulfide bond. Although a few natural substitutions have been observed in this region, its definition relies heavily on the partial

sequence of a single variant selected with monoclonal antisera against A/MEM/1/71 (14, 6).

Site D: In first three sites, the amino acids which are suggested to cause antigenic changes are external and, in principle, directly recognizable by molecules of the immune system. Parts of site D depart from this apparently single situation. Again as in site C, a single monoclonally selected variant defines this region, along with a loose set of naturally occurring substitutions (14, 6). The only amino acid substitution reported for the monoclonal variant is at serine 205, which is found in the trimer interface between adjacent globular HA<sub>1</sub> domains. Whether this substitution is recognized directly as a result of a relative movement of the globular domains (a possibility that cannot accurately be accessed at this time but which appears plausible based on the design of the protein) or indirectly at a nearby surface is unknown.

Site E: On the basis of natural variation, site E was not convincingly defined, although single substitution have occurred in that region in each epidemic strain since 1968 (6). The subsequent discovery of an antigenic variant selected by a new class of monoclonal antibody, which resulted in the single amino acid substitution ASP-63 to ASN-

63, further characterized this site (21) and led to an interesting observation. The amino acid substitution found in the antigenic variant creates a new oligosaccharide attachment site ASN-63, CYS-64, thr-65, which is glycosylated. Immunoprecipitation experiments with extracts from variant virus-infected cells prepared in the presence or absence of tunicamycin, which inhibits glycosylation of the HA, demonstrate that the addition of the new oligosaccharide chain is required to escape reaction with the monoclonal antisera (24). Thus the ASP to ASN change is not recognized directly, but instead the new oligosaccharide chain on the variant HA must interfere with antibody binding to a nearby region of the HA surface. This is a direct indication that carbohydrate can alter the antigenic structure of the HA by effectively covering some region of the HA surface, an idea that became apparent when the positions of oligosaccharide at-

tachment sites from fowl plague virus (25) and the asian strain A/JAPAN/305/57 (26) HA sequences were noted to occur in the regions A, B, and D described as antigenic sites in the 1968 structure (6).

Similar immunoprecipitation experiments with the site E monoclonal antisera on two viruses in the 1968 pandemic era A/England/878/69 and A/Vic/3/75, which are glycosylated at position 63 also indicate that the oligosaccharide is required to prevent reaction with the antisera to the 1968 strain site E, providing evidence for the epidemiological significance of carbohydrate-mediated modifications of HA antigenicity.

To establish the relative importance of the antigenic sites described, a definition is needed of the spectrum of antibody molecules in postinfection human sera with regard to amount, avidity, and site specificity.

#### OTHER ANTIGENIC SUBTYPES

Figure 3 shows a comparison of the antigenic structure of the 1968 H3 HA discussed above with the antigenic structure of the H1 strain A/PR/8/34 as proposed from partial sequencing of a series of antigenic variants of that strain selected with monoclonal antibodies (16) and displayed on the 1968 structure (6). A similar array of antigenic regions A, B1, B2, E and D can be seen, although differences in detail are apparent (16). No evidence for a site equivalent to C is available. One interesting feature of the comparison is that antigenic variation has been selected at position 165, 166, and 167 in the H1 strain, while no variation in that area is seen in either the natural or monoclonal-antibody selected H3 viruses. That region is covered by an oligosaccharide chain attached at 165 in the H3 strain, which would block access to that region.

Although no monoclonal antibody selection experiments have been reported for type B influenza, a sequence comparison of three naturally occurring strains B/LEE/40, B/MD/59, and B/HK/73 using the 1968 HA structure model, indicate that similar regions on the surface of the HA<sub>1</sub> domain have altered sequences, which parallel alterations in antigenicity (27).

#### ACKNOWLEDGMENTS

We thank William Willard for artwork, Rose Gonsalves and Gary White for assistance, and acknowledge support from the National Institutes of Health, Grant No. AI-13654, and from Nato Research Grant No. 0215/83. We thank Eugene Brown, Genetics Institute, Boston, Massachusetts, for the synthesis of an oligonucleotide primer.

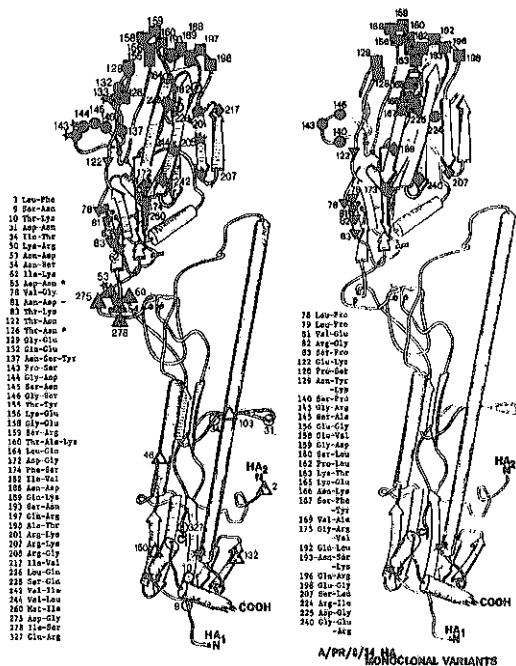


FIG. 3. A comparison of the amino acid variation observed in both natural and monoclonal antibody selected antigenic variants of the H3 HA subtype (A) (see Fig. 2 for references) with amino acid variation observed in monoclonal antibody selected variants of the H1 HA from the A/PR/8/34 virus (B) (6, 21). Of particular interest is the observation that an area covered by an oligosaccharide chain (165-168) of the 1968 HA shows no variation, while the equivalent unglycosylated region in the H1 HA shows variation at 165, 166, 167.

## REFERENCES

1. Wilson, I. A., J. J. Skehel and D. C. Wiley. 1981. *Nature* 289:366-373.
2. Rogers, B. N., J. C. Paulson, R. S. Daniels, J. J. Skehel, I. A. Wilson and D. C. Wiley. 1983. *Nature* 304:76-78.
3. Skehel, J. J., P. M. Bayley, E. B. Brown, S. R. Martin, M. D. Waterfield, J. M. White, I. A. Wilson and D. C. Wiley. 1982. *PNAS* 79:968-972.
4. Braud, C. M. and J. J. Skehel. 1972. *Nature New Biol.* 238:145-147.
5. Wiley, D. C. and J. J. Skehel. 1977. *J. Mol. Biol.* 112:343-347.
6. Wiley, D. C., I. A. Wilson and J. J. Skehel. 1981. *Nature* 289:373-378.
7. Smith, W., G. H. Andrews and P. P. Laidlaw. 1933. *Lancet* ii:66-68.
8. World Health Organization. 1971. *Bull. Wod Hlth. Org.* 45:119-124.
9. Kendal, A. P., G. R. Noble, J. J. Skehel and W. R. Dowdle. 1978. *Virology* 89:632-636.
10. Ward, C. W. and T. A. Dopheide. 1981. *Biochem. J.* 195:337-340.
11. Skehel, J. J., A. R. Douglas, I. A. Wilson and D. C. Wiley. 1982. In *Virus Persistence Symposium 33*. Soc. Gen. Microbiol. (ed. Mahy, B. W. J., Minson, A. C., and Darby, G. K. Cambridge University Press).
12. Burnet, F. M. 1955. *Principles of Animal Virology* 382ff.
13. Webster, R. G. and W. G. Laver. 1975. In *Influenza Viruses and Influenza* (ed. Kilbourne, E. D.), pp. 270-314. Academic Press. New York
14. Laver, W. G., G. M. Air, C. W. Ward and T. A. Dopheide. 1980. *Nature* 283:454-457.
15. Daniels, R. S., A. R. Douglas, J. J. Skehel and D. C. Wiley. 1983. *J. Gen. Virol.* 64:1657-1662.
16. Caton, A. J., G. G. Brownlee, J. W. Kendell and W. Gerhard. 1982. *Cell* 31:417-427.
17. Sleight, M. J., G. W. Both, G. G. Brownlee, V. J. Bender and B. A. Moss, B. A. 1980. In *Structure and Variation in Influenza Virus* (eds. Laver, W. G., and Air, G. M.) 69-78. Elsevier, Amsterdam.
18. Verhoeyen, M., R. Fang, W. Min Jou, R. Devos, D. Huylebroeck, E. Saman and W. Fiers. 1980. *Nature* 286:771-776.
19. Min Jou, W., M. Verhoeyen, R. Devos, E. Saman, R. Fang, D. Huylebroeck, W. Fiers, G. Threfall, D. Backer, N. Carey and S. Emtage. 1980. *Cell* 19:683-696.
20. Both, G. W. and M. S. Sleight. 1980. *Nucl. Acid. Res.* 8:2561-2575.
21. Daniels, R. S., A. R. Douglas, F. Gonsalves-Scarano, G. Palu, J. J. Skehel, E. Brown, M. Knossow, I. A. Wilson and D. C. Wiley. 1983. In *Antigenic Structure of Influenza Virus Haemagglutinin* (eds. Laver, W. G. and Air, G. M.). Elsevier-North Holland, New York, in press.
22. Moss, B. A., P. A. Underwood, V. J. Bender and R. G. Whitaker. 1980. In *Structure and Variation in Influenza Virus* (eds. Laver, W. G. and Air, G. M.), 329-338, Elsevier, Amsterdam.
23. Bot, G. W., M. J. Sleight, V. J. Bender and B. A. Moss. 1980. In *Structure and Variation in Influenza Virus* (eds. Laver, W. G., and Air, G. M.), 81-89, Elsevier, Amsterdam.
24. Skehel, J. J., D. J. Stevens, R. S. Daniels, A. R. Douglas, M. Knossow, I. A. Wilson and D. C. Wiley. 1983. *PNAS* (submitted).
25. Porter, A. G., C. Barber, N. H. Casey, R. A. Hollowell, G. Threfall and J. S. Emtage. 1979. *Nature* 282:471-477.
26. Gething, M. J., J. Bye, J. J. Skehel and M. Waterfield. 1980. *Nature* 287:301-306.
27. Krystal, M., J. F. Young, P. Palase, I. A. Wilson, J. J. Skehel and D. C. Wiley. 1983. *Proc. Nat. Acad. Sci. U.S.A.* 80:4527-4531.

