(100% in signals, 85% in hydrophobic tails), which suggests a rather specific role for the terminal sequence in the fusion process.

There are two obvious difficulties in understanding the role of haemagglutinins in membrane fusion. First, the 135 Å length of the haemagglutinin seems to restrict membranes from any closer initial approach. Second, the HA_2 amino-terminus is ~100 Å from the distal tip and 35 Å from the membrane end of the molecule. For this region of the apolar terminal residues to approach either the virus or host-cell membrane a major conformational change would be required after receptor binding. Alternatively, some other region of the molecule significantly affected by the activation cleavage at the HA_2 amino terminus may participate.

Finally, antibodies that neutralize viral infectivity interact with the haemagglutinin. Identification of antibody binding sites and discussion of how their variation contributes to the recurrence of influenza disease in man is reported in the accompanying paper[28].

1. Laver, W. G. & Kilbourne, E. D. *Virology* **30**, 493–501 (1966).
2. Brand, C. M. & Skehel, J. J. *Nature new Biol.* **238**, 145–147 (1972).
3. Wiley, D. C. & Skehel, J. J. *Topics infect. Dis.* **3**, 135–138 (1978).
4. Skehel, J. J. & Waterfield, M. D. *Proc. natn. Acad. Sci. U.S.A.* **72**, 93–97 (1975).
5. Porter, A. G. *et al. Nature* **282**, 471–477 (1979).
6. Ward, C. W. & Dopheide, T. A. in *Structure and Variation in Influenza Virus* (eds Laver, W. G. & Air, G. M.) 27–38 (Elsevier, Amsterdam, 1980).
7. Dopheide, T. A. & Ward, C. W. in *Structure and Variation in Influenza Virus* (eds Laver, W. G. & Air, G. M.) 21–26 (Elsevier, Amsterdam, 1980).
8. Sleigh, M. J., Both, G. W., Brownlee, G. G., Bender, V. J. & Moss, B. A. in *Structure and Variation in Influenza Virus* (eds Laver, W. G. & Air, G. M.) 69–78 (Elsevier, Amsterdam, 1980).
9. Gething, M. J., Bye, J., Skehel, J. & Waterfield, M. D. in *Structure and Variation in Influenza Virus* (eds Laver, W. G. & Air, G. M.) 1–10 (Elsevier, Amsterdam, 1980).
10. Min Jou, W. *et al.* in *Structure and Variation in Influenza Virus* (eds Laver, W. G. & Air, G. M.) 63–68 (Elsevier, Amsterdam, 1980).
11. Verhoeyen, M. *et al. Nature* **286**, 771–776 (1980).
12. Threlfall, G., Barber, C., Carey, N. & Emtage, S. in *Structure and Variation in Influenza Virus* (eds Laver, W. G. & Air, G. M.) 51–61 (Elsevier, Amsterdam, 1980).
13. Min Jou, W. *et al. Cell* **19**, 683–696 (1980).
14. Ward, C. W. & Dopheide, T. A. *Biochem. J.* (in the press).
15. Springer, T. A. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* **73**, 2481–2485 (1976).
16. Coligan, J. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 3390–3394 (1978).
17. Tomita, M. & Marchesi, V. T. *Proc. natn. Acad. Sci. U.S.A.* **72**, 2964–2968 (1975).
18. Henderson, R. & Unwin, P. N. T. *Nature* **257**, 28–32 (1975).
19. Engelman, D. M., Henderson, R., McLachlan, A. D. & Wallace, B. A. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2023–2027 (1980).
20. Air, G. M. *Virology* **97**, 468–472 (1979).
21. McCauley, J. *et al. FEBS Lett.* **108**, 422–426 (1979).
22. Milstein, C., Brownlee, G. G., Harrison, T. M. & Mathews, M. B. *Nature new Biol.* **239**, 117–120 (1972):

23. Blobel, G. & Dobberstein, B. *J. Cell Biol.* **67**, 835–851 (1975).
24. Hirst, G. K. *J. exp. Med.* **75**, 49–64 (1942).
25. Lazarowitz, S. G. & Choppin, P. W. *Virology* **68**, 440–454 (1975).
26. Klenk, H. D., Rott, R., Orlich, M. & Blodorn, J. *Virology* **68**, 426–439 (1975).
27. Huang, R. T. C., Wahn, K., Klenk, H. D. & Rott, R. *Virology* **104**, 294–302 (1980).
28. Wiley, D. C., Wilson, I. A. & Skehel, J. J. *Nature* **289**, 373–378 (1981).
29. Wiley, D. C. & Skehel, J. J. *J. molec. Biol.* **112**, 343–347 (1977).
30. Monaco, H. L. thesis, Harvard Univ. (1978).
31. Harrison, S. C. *J. appl. Crystallogr.* **1**, 84–89 (1968).
32. Wiley, D. C., Skehel, J. J. & Waterfield, M. D. *Virology* **79**, 446–448 (1977).
33. Bricogne, G. *Acta crystallogr.* **A32**, 832–847 (1976).
34. Dopheide, T. A. & Ward, C. W. *J. gen. Virol.* (in the press).
35. Waterfield, M. D., Scrace, G. & Skehel, J. J. *Nature* **289**, 422–424 (1981).
36. Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, F. K. & Bricogne, G. *Nature* **276**, 368–373 (1978).
37. Abad-Zapatero, C. *et al. Nature* **286**, 33–39 (1980).
38. Richardson, J. *Adv. Protein Chem.* (in the press).
39. Wickner, W. *A. Rev. Biochem.* **48**, 23–45 (1979).
40. Air, G. M. in *Structure and Variation in Influenza Virus* (eds Laver, W. G. & Air, G. M.) 135–146 (Elsevier, Amsterdam, 1980).
41. Pauling, L. & Corey, R. B. *Nature* **171**, 59–61 (1953).
42. Cohen, C. & Holmes, K. C. *J. molec. Biol.* **6**, 423–432 (1963).
43. McLachlan, A. D. & Stewart, M. *J. molec. Biol.* **98**, 293–304 (1975).
44. Ward, C. W. & Dopheide, T. A. *Aust. J. biol. Sci.* **33**, 449–455 (1980).
45. Deisenhofer, J., Coleman, P. M., Epp, O. & Huber, R. *Hoppe-Seyler's Z. physiol. Chem.* **357**, 1421–1434 (1976).
46. Marquart, M., Deisenhofer, J., Huber, R. & Palm, W. in *Proc. 7th Aharon Kelzir-Katchalsky Conf. Ginossar* (ed. Balasan, N.) (in the press).
47. Laver, W. G., Air, G. M., Dopheide, T. A. & Ward, C. W. *Nature* **283**, 454–457 (1980).
48. Wright, C. S. *J. molec. Biol.* **141**, 267–291 (1980).
49. Gething, M. J., White, J. M. & Waterfield, M. D. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2737–2740 (1978).

Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation

D. C. Wiley & I. A. Wilson

Gibbs Laboratory, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138

J. J. Skehel

National Institute for Medical Research, Mill Hill, London NW7 1AA, UK

Four 'antigenic sites' on the three-dimensional structure of the influenza haemagglutinin are identified. At least one amino acid substitution in each site seems to be required for the production of new epidemic strains between 1968 and 1975.

THE three-dimensional structure of the major surface antigen, the haemagglutinin, of influenza virus was determined to advance our understanding of how a membrane antigen is recognized by the immune system[1]. Variation in the haemagglutinin's antigenic structure is associated with the familiar and uncontrolled recurrence of influenza epidemics in man[2]. The structural basis for this phenomenon is sought to answer the following questions: which regions of the molecule are immunogenic and dominant in humoral and cellular immune recognition? How do the observed mutations result in an escape from neutralization? Why are conserved surface regions of the haemagglutinin, such as those involved in its many activities[1], not involved in the induction of cross-reacting antibodies? What

are the requirements for the emergence of a novel epidemic strain?

This report presents an analysis of the three-dimensional structure of the 1968 Hong Kong haemagglutinin using the available primary sequence information on natural and laboratory-selected antigenic variants[3-16]. We propose initial answers to some of the above questions. Four antigenic sites are identified on the surface of the molecule. At least one amino acid substitution has been observed in each site of the haemagglutinins of recent epidemic strains and thus seems to be required for a new epidemic to occur. The probable stereochemical consequences of mutations at these sites and suggestions of how they might reduce antibody binding are discussed.
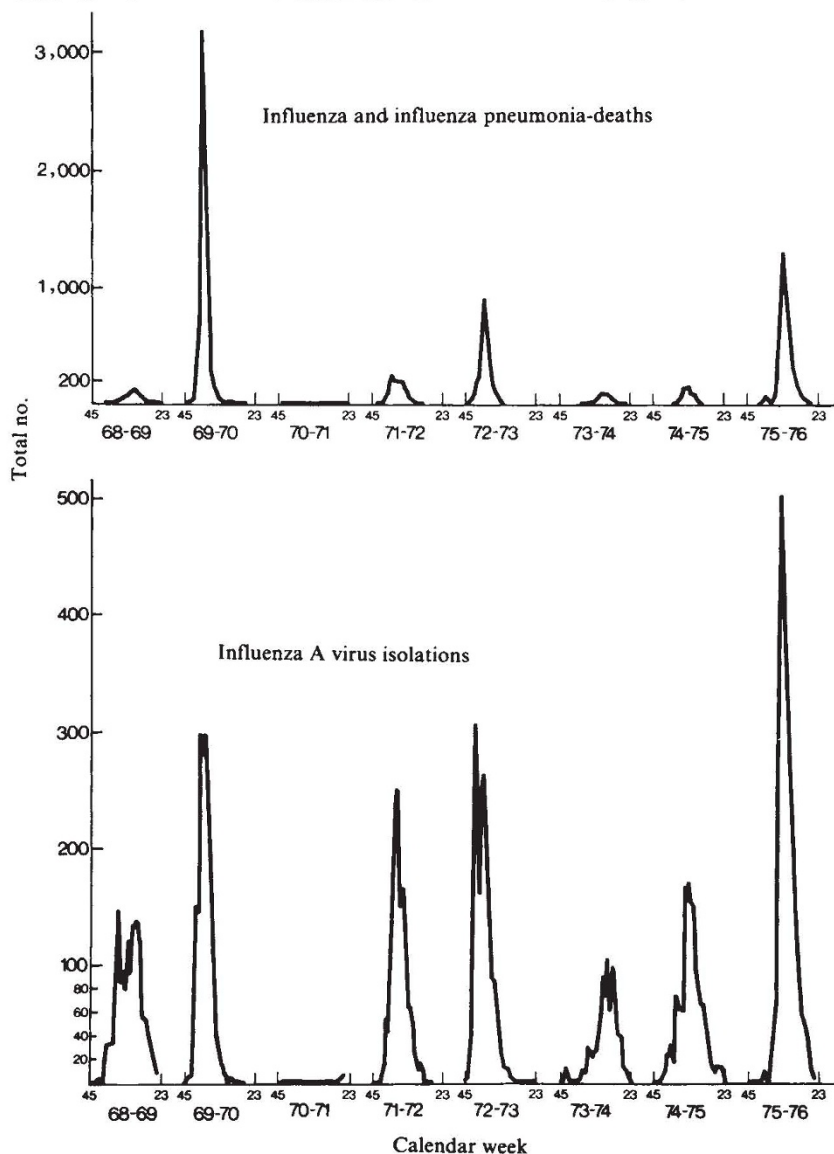
Fig. 1 The results of influenza surveillance in England and Wales for the years 1968–1976. The graphs were drawn from information provided by the Communicable Disease Surveillance Centre of the Public Health Laboratory Service, Colindale, London, part of which was published in detail in 1977 (ref. 29). Data of a similar nature for other countries up to 1973 are given in Assaad *et al.* (ref. 30).

Antigenic variation

Since the first isolation of an influenza virus from humans in 1933 (ref. 17), two major pandemics have been recorded (in 1957 and 1968) and on both occasions, the viruses responsible were characterized by marked differences in the antigenic properties of their haemagglutinins. These viruses were designated prototype strains of distinct antigenic subtypes: H2 in the case of the Asian viruses prevalent between 1957 and 1968 and H3 for the Hong Kong viruses which appeared in 1968 and are still in circulation[18].

In addition to the pandemics which were observed initially within each period, epidemics were recorded which were thought to be caused by viruses which had undergone 'antigenic drift'[19] from the prototypes. We discuss the structural basis of this process as it occurred between 1968 and 1975.

Antigenic variation in the Hong Kong (H3) viruses. In 1968 a strain of influenza virus, A/Hong Kong/1/68, was isolated and found to be antigenically unrelated to the viruses of the Asian (H2) subtype which were prevalent during the previous decade. Worldwide epidemics of disease due to infection with this virus were reported in the winters of 1968/9 and 1969/70, during which time the results of international surveillance indicated little variation in the antigenic properties of the viruses isolated[20] (Fig. 1). In the winter of 1971/72, an antigenically distinct virus, A/England/42/72 (similar to A/Memphis/102/72 in the USA), was isolated—this caused widespread epidemics in 1972/73. Several antigenically related strains were reported in 1973/74 and in 1974/75, and there

were further worldwide epidemics in 1975/76 following the appearance of viruses antigenically similar to A/Victoria/3/75 (ref. 21). This type of information was available from several countries—for the UK (data shown in Fig. 1), the number of deaths due to influenza and influenza pneumonia increased in 1969, 1972 and 1975, coincident with increases in the number of virus isolations and with the appearance of antigenically distinct viruses.

The amino acid substitutions in the 1972 and 1975 disease-causing strains were located in the three-dimensional structure of the 1968 haemagglutinin (Fig. 2). The amino acid sequences of the haemagglutinins of representative viruses isolated in 1968, 1972 and 1975 were determined from a combination of protein and nucleic acid sequence analyses[4-6,10,13]. Fourteen amino acid substitutions were detected in the HA_1 polypeptide component of A/Memphis/102/72 relative to that of the A/Aichi/2/68 virus and a further 16 in HA_1 of A/Victoria/3/75. A summary of the structural locations and probable stereochemical consequences of these substitutions together with information from seven other partial sequences[9] and analyses of the haemagglutinins of laboratory-selected antigenic variants[15,16,22] is shown in Table 1.

Because the haemagglutinins of antigenic variants isolated in nature may also contain neutral substitutions, that is, mutations which do not affect the antigenic recognition by components of the immune system, two selection techniques have been used in attempts to identify those amino acid substitutions which influence antibody–antigen interaction. The work of Fazekas de

St. Groth and colleagues[15,16,23] has involved the selection of variants by growth of an H3 1968 virus (A/NT/60/68) and its derivatives in the presence of the most avid fractions of hyperimmune antisera and, similarly, Laver, Webster and colleagues[22,24,25] have selected variants of A/Memphis/1/71, a virus closely related to A/Hong Kong/1/68, by growth in the presence of monoclonal antibodies prepared against its haemagglutinin. After virus selection, amino acid sequence data were obtained for the haemagglutinins of the variants. Although only a partial sequence data are available, they indicate that variants selected using monoclonal antibodies contain single amino acid substitutions, which directly or indirectly alter an antigenic region of the haemagglutinin molecule sufficiently to prevent neutralization of the variant by the selecting antibody[22]. The locations of the amino acid substitutions in haemagglutinins of the laboratory-selected variants are summarized in Fig. 2c.

Suggested locations for the antigenic sites. An antigenic site is a region of a molecule involved in antibody binding[26]. Although there are no exhaustive immunochemical data such as those used in studies defining the antigenic sites on lysozyme and myoglobin[27,28], a combination of the amino acid sequence information and the three-dimensional structure of the 1968 haemagglutinin[1] allows identification of four regions of the Hong Kong haemagglutinin molecule in which amino acid substitutions seem to affect antibody binding.

Site A: an unusual protruding loop from amino acids 140 to 146, (see Fig. 2) which projects 8 Å from the local molecular surface, forms the centre of the most obvious antibody-binding site. Clearly the haemagglutinin of each antigenically distinct virus of epidemic significance has a mutation in this region (Fig. 2 and Table 1): A/Memphis/102/72: residue 144, Gly to Asp, and residue 122, Thr to Asn[4,6]; A/Victoria/1/75 and A/Victoria/3/75: residue 145, Ser to Asn, and residue 137, Asn to Ser respectively[9,12,13]. Furthermore, the haemagglutinins

of a group of monoclonal antibody-selected variants neutralization by single amino acid substitutions in this region: 143, Pro to His, Ser, Thr, Leu; 144, Gly to Asp, and 133, Asn to Lys[22]. Conserved amino acids Gly 134, Ala 138, Cys 139, Phe 147 and Phe (Tyr) 148 occurring at each end of the antigenically variable loop provide the structural foundation of the site in all 10 haemagglutinins examined (including one Avian[3] and one Asian[7] subtype).

Site B: this comprises the external residues 187–196 of an α helix, and adjacent residues along the upper edge of a pocket tentatively implicated in virus receptor binding[1] (see Fig. 2). Two substitutions in the haemagglutinin of the 1972 virus, A/Memphis/102/72: 188, Asn to Asp, and 155, Thr to Tyr[4,6]; and two more in that of the 1975 virus, A/Victoria/3/75: 189, Gln to Lys and 193, Ser to Asn[9,12,13] characterize the site, but the lack of sequence information for this region of the haemagglutinins of several variants leaves the comparison incomplete[9]. The substitution 186, Ser to Ile in the haemagglutinin of the 30D virus[16], selected using 'avid-fraction' antisera, may also support the site designation. In addition, the existence of this structure in other haemagglutinins is suggested by the observed conservation of neighbouring residues. On the inward face of the helix residues 190 Glu, 191 Gln, 194 Leu and 195 Tyr are conserved in all the known sequences, as are 153 Trp and 154 Leu. However, because no amino acid substitutions in the haemagglutinins of variants selected by monoclonal antibodies comparable with those which support the delineation of site A have yet been reported, more direct evidence of antibody binding to this region is not available.

Site C: a bulge in the tertiary structure at the disulphide bond between Cys 52 and Cys 277, 60 Å from the distal tip of the molecule, comprises another antibody-binding site (see Fig. 2). The haemagglutinins of the viruses from both epidemic periods have substitutions clustered in this region:
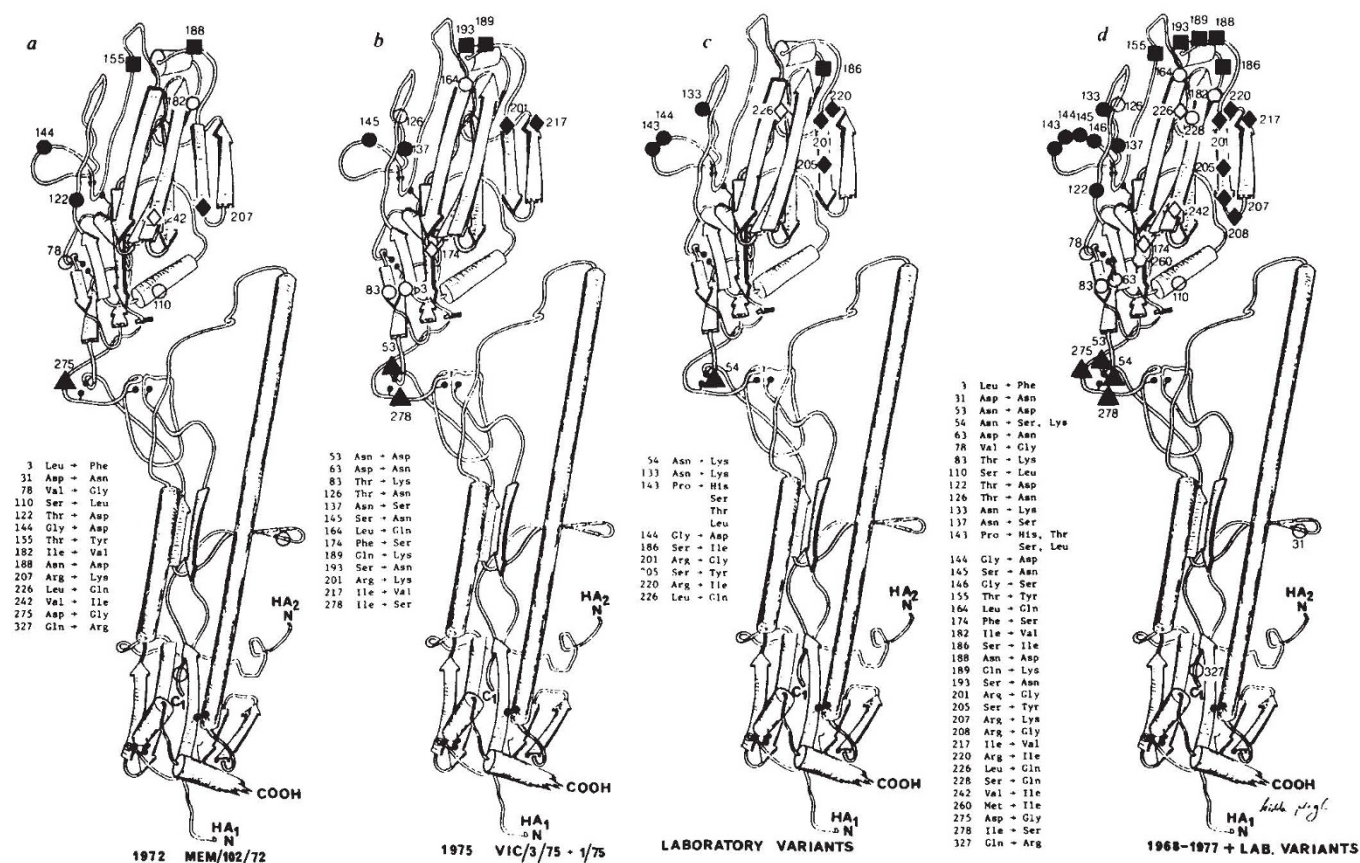

Fig. 2 Schematic drawing of a monomer of the AICHI-1968 haemagglutinin showing locations of amino acid substitutions on HA₁. ●, Site A; ■, site B; ▲, site C; ◆, site D (subunit interface); ◇, surface; ○, neutral. *a*, Amino acid changes from AICHI/2/68 to MEM/102/72 (refs 4–6, 10, 11). *b*, Amino acid changes from MEM/102/72 to VIC/3/75 (refs 12, 13). *c*, Laboratory-selected antigenic variants from growth in monoclonal or most-avid-fraction antibody[9,15,16,22]. (See Table 1 for details.) *d*, Summary of all known alterations from 1968 to 1977 including laboratory-selected variants[3–16,22]. A three-dimensional (stereo) drawing of the same view of the molecule is found in Fig. 3 of ref. 1.

Table 1 Antigenic drift residues

Strain	Residue	Replacement (from)	Replacement (to)	Probable antigenic site	Structural location and probable effects of replacement
MEM/102/72*	144	Gly	Asp	A	External loop. Substitution is both larger and charged. Change from glycine may restrict ϕ, ψ angles of main chain, altering the loop shape.
	122	Thr	Asn	A?	External extended chain (β). Substitution is larger, 18 Å from residue 144, and may effect the edge of a contact region centred at the 144 loop.
	155	Thr	Tyr	B	4 Å from residue 193 (VIC) and the C-terminal end of the 187–196 α-helix. Substitution is considerably larger and would project over the 'pocket'.
	188	Asn	Asp	B	N-terminal exterior face of α helix 187–196, 18 Å from residue 155 altered charge.
	275	Asp	Gly	C	Near 52–77 disulphide bond bulge. Asp 275 was H-bonded to Asn 53, an interaction that would be lost by the Gly substitution.
	207	Arg	Lys	D?	Subunit interface near residues 99 and 100 of another subunit. Substitution is conservative but shorter and unbranched.
	242	Val	Ile	D??	At the surface but near residues 205 and 207 of same subunit and 221 and 222 of an adjacent subunit, also near the oligosaccharide chain attached at 165. Substitution is larger.
	226	Leu	Gln	?	External on the front of the 'pocket', 8 Å from residue 137 (A), 8 Å from 186 (B?), 16 Å from 242 (D?). Substitution increases size and changes polarity. Gln could reach Ser 136 or Asn 38 for H-bonds. Found in 29C 'avid-fraction' variant. Reverts in VIC/3/75.
	3	Leu	Phe	None	Near membrane, probably inaccessible to antibodies. Reverts in VIC/3/75.
	327	Gln	Arg	None	100 Å from distal tip. Reverts in VIC/3/75.
	31	Asp	Asn	None	82 Å from distal tip.
	78	Val	Gly	None (E?)	Probably buried by oligosaccharide at 81.
	182	Ile	Val	None	Buried residue near 220.
	110	Ser	Leu	None	Near Arg 261, buried. Reverts in VIC/3/75.
VIC/3/75† (VIC/1/75)‡	(145)	Ser	Asn	A	External loop. Substitution is larger and has more potential for H-bonding. If the Asn interacts with the 135/136 region the loop conformation might be altered (VIC/1/75).
	137	Asn	Ser	A?	External 7 Å from residue 145. Substitution is smaller, possibly near an edge of the A site.
	189	Gln	Lys	B	External face of 187–196 α helix. Substitution is larger and charged.
	193	Ser	Asn	B	External face of 187–196 α helix, 4 Å from 189 and 4 Å from 155. Substitution is larger with more H-bond potential. Possible new interaction with Tyr 155.
	278	Ile	Ser	C	Surface residue in 52–277 disulphide bulge. Substitution is smaller and polar and could H-bond to Asn 54.
	(53)	Asn	Asp	C	External in 52–277 disulphide bulge, next to 275. Introduces charge (VIC/1/75).
	201	Arg	Lys	D?	Packed in the subunit interface against 217 of the other subunit, below 188. Substitution is smaller and less bulky and could effect the fit between subunits or the site 'B' helix or be recognized directly if the subunit interface were exposed by a conformational charge.
	217	Ile	Val	D?	Packed in the subunit interface against 201 of the other subunit. Similar to the 201 change. Under 189 (B).
	174	Phe	Ser	D??	External after surface chain, 169 to 173: Pro, Asn, Asn, Asp, Asn. 14 Å from 242. 169 is next to 242.
	164	Leu	Gln	None	Buried next to 127. Substitution gets larger and polar.
	126	Thr	Asn	None	New oligosaccharide attachment site. Carbohydrate would cover a smooth face. Remote possibility that it could reach 133 or 122 and affect site A.
	327				⎫
	3				⎬ Same as 1968 AICHI strain.
	226				⎪
	110				⎭
	63	Asp	Asn	None (E?)	New oligosaccharide attachment site, very near old site at 81.
	83	Thr	Lys	None (E?)	Destroys oligosaccharide attachment site at 81.
TEX/1/77§	146	Gly	Ser	A	External loop. Substitution increases size and potential for H-bonding to 135, 136 region, also limits Φ, Ψ angles of main chain (see MEM/102/72 residue 144).
	54	Asn	Ser	C	External as 53 above, next to 278 which is not yet sequenced in TEX/1/77.
	260	Met	Ile	None	Buried next to 177. Close to site D?.
ENG/187/70§	228	Ser	Gln	None	In pocket near 226. Substitution is much larger, and does not reappear after this strain.
ENG/42/72§	208	Arg	Gly	D?	Near subunit interface, residue 207. Substitution smaller and lacks charge.
Monoclonally selected variant‖	143	Pro	His, Ser Thr, Leu	A	Loop. All these residue 143 variations were selected with a single antibody. Selection with other antibodies has produced identical sequence changes.
	144	Gly	Asp	A	Loop. (See MEM/102/72 residue 144).
	133	Asn	Lys	A	Selected with the same antibody that 143 Pro–Ser was selected.
	54	Asn	Lys	C	External 52–277 disulphide bulge. Substitution is larger and charged.
	205	Ser	Tyr	D	Buried in the subunit interface near 221 and 220 (See VIC/3/75 residue 201).
Avid-fraction variant¶					
30D	186	Ser	Ile	B or D	Below 189 in B-site helix. Ser probably H-bonded to carbohydrate attached at 165. Ile may move carbohydrate affecting site D or directly affect site B.
30D	220	Arg	Ile	D	Subunit interface residue near 203 and 205 of an adjacent subunit 217, of the same subunit. Arg 220 probably H-bonds to Gln 210 of another subunit—the Ile substitution could not make such an interaction.
34D	201	Arg	Gly	D?	Refer to VIC/3/65 residue 201.
29C	226	Leu	Gln	?	Refer to MEM/102/72 residue 226.

* Refs 4, 5, 6.
† Refs 8, 9.
‡ Refs 12, 13.
§ Ref. 9.
‖ Refs 23, 24.
¶ Refs 19, 20.

A/Memphis/102/72: 275, Asp to Gly[4-6]; A/Victoria/1/75 and A/Victoria/3/75: 53, Asn to Asp and 278, Ile to Ser respectively[9,12,13]. The absence of amino acid sequences for the 275 to 278 region of HA$_1$ from a number of partial sequences may also mean that this list is incomplete[9]. A substitution at position 54, Asn to Lys, in a variant selected by monoclonal antibody prevents its neutralization[22] and variant 30D also fails to react with the same monoclonal antibody[16]. Again the variable amino acids seem to be anchored in a matrix of conserved residues: 50 Lys, 51 Ile (Leu), 52 Cys, 277 Cys, 281 Cys and 286 Gly. The variable residues at all four positions are in direct contact. In the structure of the 1968 haemaglutinin, Asp 275 is hydrogen-
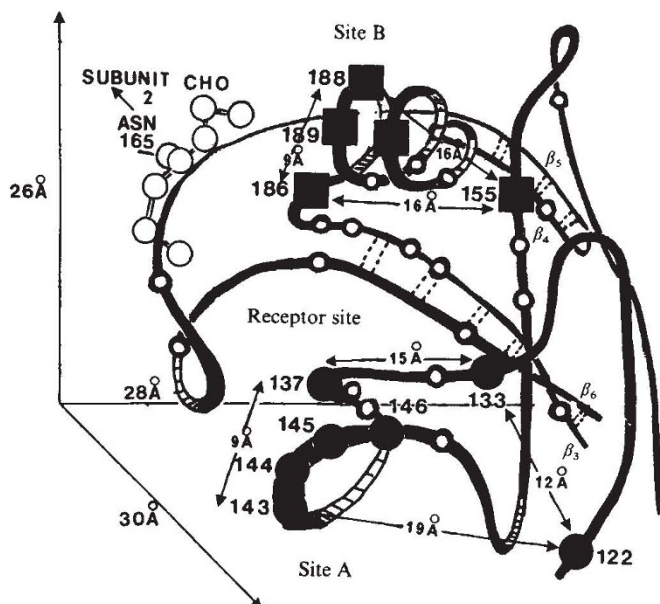
Fig. 3 Schematic drawing showing the size and shape of two of the proposed antigenic sites: ●, A and ■, B (see Fig. 2 and text). Their relationship to the pocket of conserved residues tentatively identified as the host sialic-acid-containing receptor binding site is shown.

bonded to Asn 53, Ile 278 is packed against Asn 54, and Cys 52 bonds to Cys 277. Changes in the types of contact between these residues (for details see Table 1) may be the structural alterations to which antibody binding is sensitive.

Site D: in the first three sites, the amino acid substitutions suggested as causing antigenic variation are external and, in principle, directly recognizable by molecules of the immune system. Site D departs from this apparently simple situation (site C may also). Several amino acid substitutions in the haemagglutinins of both natural and laboratory-selected antigenic variants occur in the interface region between subunits in the haemagglutinin trimer: A/Memphis/102/72: 207, Arg to Lys[4,6]; A/Victoria/3/75: 201, Arg to Lys and 217, Ile to Val[9,12,13]; monoclonal variant: 205, Ser to Tyr[22], and 'avid-fraction' variants 34C: 201, Arg to Gly, and 30D: 220, Arg to Ile[15,16] (see Figs 2, 4). These amino acids may be recognized directly as a result of a relative movement of the globular domains of HA₁ to expose the interface regions. However, it is also possible that the actual antibody binding site is remote from these amino acids but affected by the exact fit at the interface, which might be disturbed by the substitutions listed. A site which spans the monomer–monomer boundary or one near the oligosaccharide chain attached at residue 165, which overlaps that

boundary, could be particularly sensitive to alterations in the subunit contact region. Residues 242, Val to Ile and 226, Leu to Gln in A/Memphis/102/72[4,5]; 174, Phe to Ser in A/Victoria/3/75[9,12,13,25] and 226, Leu to Gln (also near sites A and B) in laboratory variant 29C (refs 15, 16) are likely components of this area (see Fig. 2). The amino acid sequences of peptides in the 174 and 242 regions are absent from many partial sequences[9].

Other amino acid substitutions in HA₁ and HA₂ found in nature are considered to be antigenically neutral for reasons detailed in Table 1. For example, the substitution at residue 78, Val to Gly, in the A/Memphis/102/72 haemagglutinin, is buried by the oligosaccharide at position 81. Similarly, in A/Victoria/3/75 at residue 83, the Thr to Lys substitution destroys the oligosaccharide attachment site at residue 81, but the Asp to Asn change at 63 creates a new site—this means that the carbohydrate could occupy almost the same position.

The characterization of new variants may provide evidence for additional sites (for example, 122 and 126, site E?) or extend the size of those proposed.

Antibody interactions and neutralization. This analysis has not defined the extent of each site but only amino acid positions within sites that seem sensitive to substitution when examined serologically, Nevertheless, areas of the surface remain, not in the interfaces nor covered by oligosaccharide chains, that seem serologically unimportant. It is not clear why these regions are not immunogenic, although, for example, the concave shape of the pocket of conserved amino acids tentatively identified with receptor binding[1] probably prevents the production of avid antibodies against that region. The proximity of sites A, B and D to this tentative receptor site (Fig. 3) may be important in the mechanism of neutralization.

Table 1 summarizes changes in the size and/or change of the observed substitutions that would result in a decrease in the binding constant of antibody at a site as a result of steric repulsion or unfavourable charge interactions (the burying of an unpaired charge in the antibody–antigen complex). In addition to these mechanisms, the glycine-rich loop in site A protrudes in such a way as to make so few contacts with the rest of the structure that the effect of mutations in it may be amplified by altering the conformation of the entire loop. The more restricted range of torsion angles possible for amino acids replacing glycine and the potential for hydrogen and ionic bonds from the loop to adjacent regions of the molecule possible with polar or charged substitutions (Table 1) are candidates for such an alteration. (X-ray data on isomorphous crystals of singly substituted, monoclonally selected variant haemagglutinins prepared in collaboration with R. Webster are being collected to allow a direct visualization of these altered sites.)

Antigenic variation and disease incidence (1968–1975)

It seems from the above discussion that the haemagglutinins of both major disease-causing viruses of this period had at least one
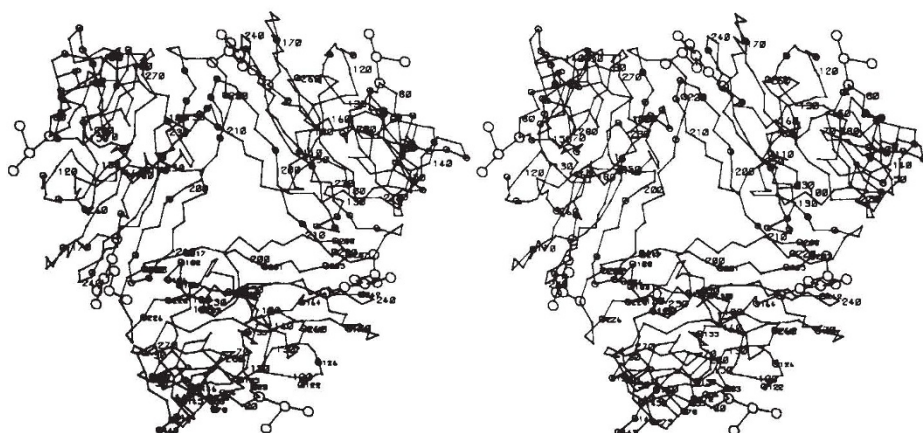

Fig. 4 Stereo drawing of the α-carbon tracing of residues 48–280 (HA₁) of the haemagglutinin trimer. Circled residues are natural and laboratory variations[3–16,22] (see Fig. 2d). Tentative (see ref. 1) carbohydrate positions are shown as large open circles for the attachment sites at Asn 81 and Asn 165. The view is down the trimer axis, viewed from the outside towards the membrane.

mutation in each of the four antigenic sites described, and may have required these to avoid virus neutralization. Thus, the viruses acquired the antigenic potential to cause a recurrence of disease in individuals previously infected with viruses of the H3 subtype. Because other characteristics of viruses such as transmissibility and virulence are important in the timing and impact of epidemics, antigenic potential (the recreation of a threshold of susceptible hosts) may, in general, only be a necessary condition for, but not capable of, the generation of epidemics.

The possibility has also been raised[25] that one locus may be more immunogenic than others and therefore, when altered, might be of greater advantage to the virus. Site A, due to its accessibility, seems to be ideally suited both for avid antibody binding and for accepting amino acid substitutions without affecting the main framework of the molecule. To establish the relative importance of the four antigenic sites, a definition will be needed of the spectrum of antibody molecules in human sera with regard to amount, avidity and site specificity.

The character and location of amino acids on the 1968 haemagglutinin structure conserved in three subtypes[3-14] indicates that the basic framework of the molecule is preserved[1]. Because strictly conserved amino acids are found flanking each of the proposed antigenic sites, these structures are also probably present in haemagglutinins from other subtypes. Whether or not the same sites are involved in the interaction of these other haemagglutinins with antibodies is unknown. In the haemagglutinin of other subtypes, these sites have not only varied considerably in amino acid sequence but probable points of oligosaccharide chain attachment are sometimes found within or near regions of the molecule sequentially equivalent to those proposed here as antigenic sites in the Hong Kong haemagglutinins: for example, in fowl plague virus[3] (Hav1) at 133 (site A), 158 (site B) and 240 (possibly site D), and in A/Japan/305/57 (H2) (refs 7, 14) at 169 (site D). It is therefore possible that the immune recognition of such sites in the haemagglutinins of other subtypes is sterically prevented by oligosaccharide chains.

1. Wilson, I. A., Skehel, J. J. & Wiley, D. C. *Nature* **289**, 366–373 (1981).
2. Webster, R. G. & Laver, W. G. in *The Influenza Viruses and Influenza* (ed. Kilbourne, E. D.) 269–314 (Academic, New York, 1975).
3. Porter, A. G. *et al. Nature* **282**, 471–477 (1979).
4. Ward, C. W. & Dopheide, T. A. in *Structure and Variation in Influenza Virus* (eds Laver, W. G. & Air, G. M.) 27–38 (Elsevier, Amsterdam, 1980).
5. Dopheide, T. A. & Ward, C. W. in *Structure and Variation in Influenza Virus* (eds Laver, W. G. & Air, G. M.) 21–26 (Elsevier, Amsterdam, 1980).
6. Sleigh, M. J., Both, G. W., Brownlee, G. G., Bender, V. J. & Moss, B. A. in *Structure and Variation in Influenza Virus* (eds Laver, W. G. & Air, G. M.) 69–78 (Elsevier, Amsterdam, 1980).
7. Gething, M. J., Bye, J., Skehel, J. & Waterfield, M. in *Structure and Variation in Influenza Virus* (eds Laver, W. G. & Air, G. M.) 1–10 (Elsevier, Amsterdam, 1980).
8. Min Jou, W. *et al.* in *Structure and Variation in Influenza Virus* (eds Laver, W. G. & Air, G. M.) 63–68 (Elsevier, Amsterdam, 1980).
10. Verhoeyen, M. *et al. Nature* **286**, 771–776 (1980).
11. Ward, C. W. & Dopheide, T. A. *Biochem. J.* (submitted).
12. Threlfall, G., Barber, C., Carey, N. & Emtage, S. in *Structure and Variation in Influenza Virus* (eds Laver, W. G. & Air, G. M.) 51–61 (Elsevier, Amsterdam, 1980).
13. Min Jou, W. *et al. Cell* **19**, 683–696 (1980).
14. Gething, M.-J., Bye, J., Skehel, J. & Waterfield, M. *Nature* **287**, 301–306 (1980).
15. Both, G. W., Sleigh, M. J., Bender, V. J. & Moss, B. A. in *Structure and Variation in Influenza Virus* (eds Laver, W. G. & Air, G. M.) 81–89 (Elsevier, Amsterdam, 1980).
16. Moss, B. A., Underwood, P. A., Bender, V. J. & Whittaker, R. G. in *Structure and Variation in Influenza Virus* (eds Laver, W. G. & Air, G. M.) 329–338 (Elsevier, Amsterdam, 1980).
17. Smith, W., Andrewes, G. H. & Laidlaw, P. P. *Lancet* ii, 66–68 (1933).
18. World Health Organization (1971) *Bull. Wld Hlth Org.* **45**, 119–124 (1971).
19. Burnett, F. M. *Principles of Animal Virology*, 382 (1955).
20. Pereira, M. S. *Br. med. Bull.* **35**, 9–14 (1979).
21. Stewart-Harris, C. *Br. med. Bull.* **35**, 3–8 (1979).
22. Laver, W. G. *et al.* in *Structure and Variation in Influenza Virus* (eds Laver, W. G. & Air, G. M.) 295–306 (Elsevier, Amsterdam, 1980).
23. Fazekes de St. Groth, S. *Topics Infect. Dis.* **3**, 25–48 (1978).
24. Laver, W. G. *et al. Virology* **98**, 226–237 (1979).
25. Webster, R. G. & Laver, W. G. *Virology* **104**, 139–148 (1980).
26. Atassi, M. Z. & Smith, J. A. *Immunochemistry* **15**, 609–610 (1978).
27. Atassi, M. Z. *Immunochemistry* **15**, 909–936 (1978).
28. Atassi, M. Z. *Immunochemistry* **12**, 423–438 (1975).
29. Public Health Laboratory Service Standing Advisory Committee on Influenza *J. Hyg., Camb.* **78**, 223–233 (1977).
30. Assad, F. A., Cockburn, W. C. & Sundaresan, T. K. *Bull. Wld Hlth Org.* **49**, 219–233 (1973).

A gene chimaera of SV40 and mouse β-globin is transcribed and properly spliced

Gilbert Chu & Phillip A. Sharp

Center for Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

A gene chimaera of the first exon from simian virus 40 (SV40) sequences coding for T antigen placed upstream from the third exon from mouse β-globin, and separated by the resultant new chimaeric intron, was cloned into a bacterial plasmid. When transfected into monkey cells, the gene chimaera was transcribed, polyadenylated and spliced using the donor splice site from SV40 and the acceptor splice site from mouse β-globin. This result suggests that a donor site from one gene can be spliced to an acceptor site from another gene.

THE finding of intervening sequences within mammalian genes has led to speculation that they take part in both phylogeny and ontogeny by facilitating the rearrangement of genes. These ideas are based on the assumption that the cell can properly splice together RNA from any two exons. Thus, evolution and differentiation may proceed in quantum jumps as a new gene is built by rearrangement of exons coding for functional protein domains from different pre-existing genes[1]. Differentiation of the immunoglobulin gene proceeds in such a manner[2]. Similarly, the regulation of a gene may be radically altered if a promoter sequence and short exon are introduced upstream from a series of exons coding for a complete protein. Insertion of this transposable promoter could occur thousands of nucleotides from the protein-coding exons.

If exons can be shuffled arbitrarily and the RNA is properly spliced, then there should be similarity among all donor sites and among all acceptor sites. In fact, the boundaries of intervening sequences are characterized by two consensus sequences, one for donor sites and one for acceptor sites[3]. Furthermore, there are examples of splicing pattern in which two or more donor sites