Reprinted from Nature, Vol. 286, No. 5775, pp. 771-776, August 21 1980

Antigenic drift between the haemagglutinin of the Hong Kong influenza strains A/Aichi/2/68 and A/Victoria/3/75

Martine Verhoeyen, Rongxiang Fang, Willy Min Jou, René Devos, Danny Huylebroeck, Eric Saman & Walter Fiers

Laboratory of Molecular Biology, State University of Ghent, Ledeganckstraat 35, B-9000 Ghent, Belgium

A DNA copy of the gene coding for the influenza A/Aichi/2/68 haemagglutinin protein was cloned in the plasmid pBR322 and the complete nucleotide sequence determined. Comparison of this primary structure and the deduced amino acid sequence with the haemagglutinin gene and protein of strains belonging to the same (H3) subtype and to different subtypes, of both human (H2) and avian (Hav1) origin, documents further at the molecular level the two independent modes of antigenic variation of the virus—drift and shift.

THE ability of type A influenza strains to change their antigenic properties is a major obstacle to controlling the viral disease by vaccination, and the virus continues to be an important cause of illness in the human population. Two types of change have been recognized: antigenic 'drift' seems to be the result of a certain number of local alterations in the viral surface proteins (thought to be selected under the pressure of the immune system), whereas antigenic 'shift' is a radical change of the antigenic properties of the surface protein, leading to the emergence of new pandemic strains. The origin of these pandemic strains is unknown, but it has been speculated that they result from a recombination event between a human virus and an influenza strain from an animal reservoir[1].

Type A influenza viruses have a negative-stranded RNA genome consisting of eight distinct RNA segments[2-4]. The two surface antigens, haemagglutinin and neuraminidase, are glycoproteins and have been assigned to viral RNAs 4 and 5 or 6 (depending on the strain), respectively[2,3]. We have recently determined the primary structure of the haemagglutinin gene from the relatively recent human influenza strain A/Victoria/3/75 (H3N2), starting from cloned DNA[5] and using fast DNA sequencing techniques. Compared with RNA and protein sequencing methods, this approach has the advantage of much greater speed and accuracy and avoids special problems, such as those encountered during the sequencing of hydrophobic regions of polypeptides.

We describe here the cloning and elucidation of the structure of the haemagglutinin gene of A/Aichi/2/68, which formed the beginning of the so-called Hong Kong (H3N2) period. The gene sequence and deduced amino acid sequence are compared with those of our previously reported Victoria strain[5], with an intermediate H3N2 isolate, A/Memphis/102/72 (this haemagglutinin was sequenced as cloned DNA[6] and except for a small hydrophobic region also as protein[7,8]), with amino acid data from monoclonal variants[9] and field strains[10], with the gene sequence of A/Japan/305/57 (H2N2)[11], which occurred at the start of the 1957–67 influenza wave, and with the fowl plague haemagglutinin gene (Hav1)[12].

Cloning the A/Aichi/2/68 haemagglutinin gene

The high yielder, recombinant strain X31, which carries the haemagglutinin and neuraminidase genes of A/Aichi/2/68, was used as the RNA source in our study. Double-stranded (ds) DNA copies of the haemagglutinin genetic information were synthesized and cloned essentially as described previously for the haemagglutinin gene of the A/Victoria/3/75 strain[5]. The major steps of the procedure are briefly summarized as follows: the mixture of eight negative-stranded X31 virion RNA segments was polyadenylated at the 3′ end[13], converted into dsDNA copies, sized on a 1.5% agarose gel and the material corresponding to full-length haemagglutinin dsDNA recovered form *Escherichia coli* K12 HB101 (refs 5, 14). Thirty-nine dTTP and annealed to oligo(dA)-tailed plasmid pBR322 in the *Pst*I restriction site, and the resulting hybrid was used to transform *Escherichia coli* K12 HB101 (refs 5, 14). Thirty-nine tetracycline-resistant colonies were obtained. When hybridized[15] to a ^{32}P-labelled gene 4 RNA probe, four of the clones showed positive hybridization. One plasmid (pXHA14) containing an insert of approximately 1,860 base pairs (including the A · T tails used for cloning) was chosen for further characterization.

Nucleotide sequence of pXHA14

The insert in plasmid pXHA14 was characterized by restriction mapping according to Smith and Birnstiel[16]. Subsequently, fragments were prepared by digestion with the appropriate restriction enzymes, 5′-terminally labelled, and sequenced according to the method of Maxam and Gilbert[17]. The nucleotide sequence is shown in Fig. 1 as the coding information or complement of the viral RNA. As the structure is highly homologous to the previously determined Victoria haemagglutinin gene (Fig. 1), we did not systematically attempt to read both strands. Nevertheless, more than half of the sequence was confirmed by analysis of both strands.

As expected from the cloning strategy used[5,14,18] (initiating second DNA strand synthesis starting from a terminal hairpin of the first DNA strand, followed by S_1 nuclease treatment), a small portion of the haemagglutinin genetic information corresponding to the 3′ end of the cDNA is missing from the clone. As we started from negative-stranded influenza RNA, this region corresponds to the 3′ end of the coding information (beyond the arrow in Fig. 1). Because two other strains belonging to the same subtype have been shown to share exactly the same 3′-noncoding sequence[5,6], we did not attempt to complete this sequence for the Aichi strain. Assuming identity in this region, 27 nucleotides would be missing from the pXHA14 plasmid, compared with 24 nucleotides missing from plasmid pVHA14, which contains the Victoria haemagglutinin information[5]. Taking into account an additional 27 base pairs at the 3′ end of the influenza-specific information in pXHA14, the total length of the Aichi haemagglutinin RNA is 1,765 nucleotides: 29 nucleotides precede the signal for the AUG translation initiation codon, the translated region extends over 1,698 nucleotides (specifying 566 amino acids), and the (presumed) 3′-terminal noncoding stretch comprises 38 nucleotides (including the signal for the UGA terminator).

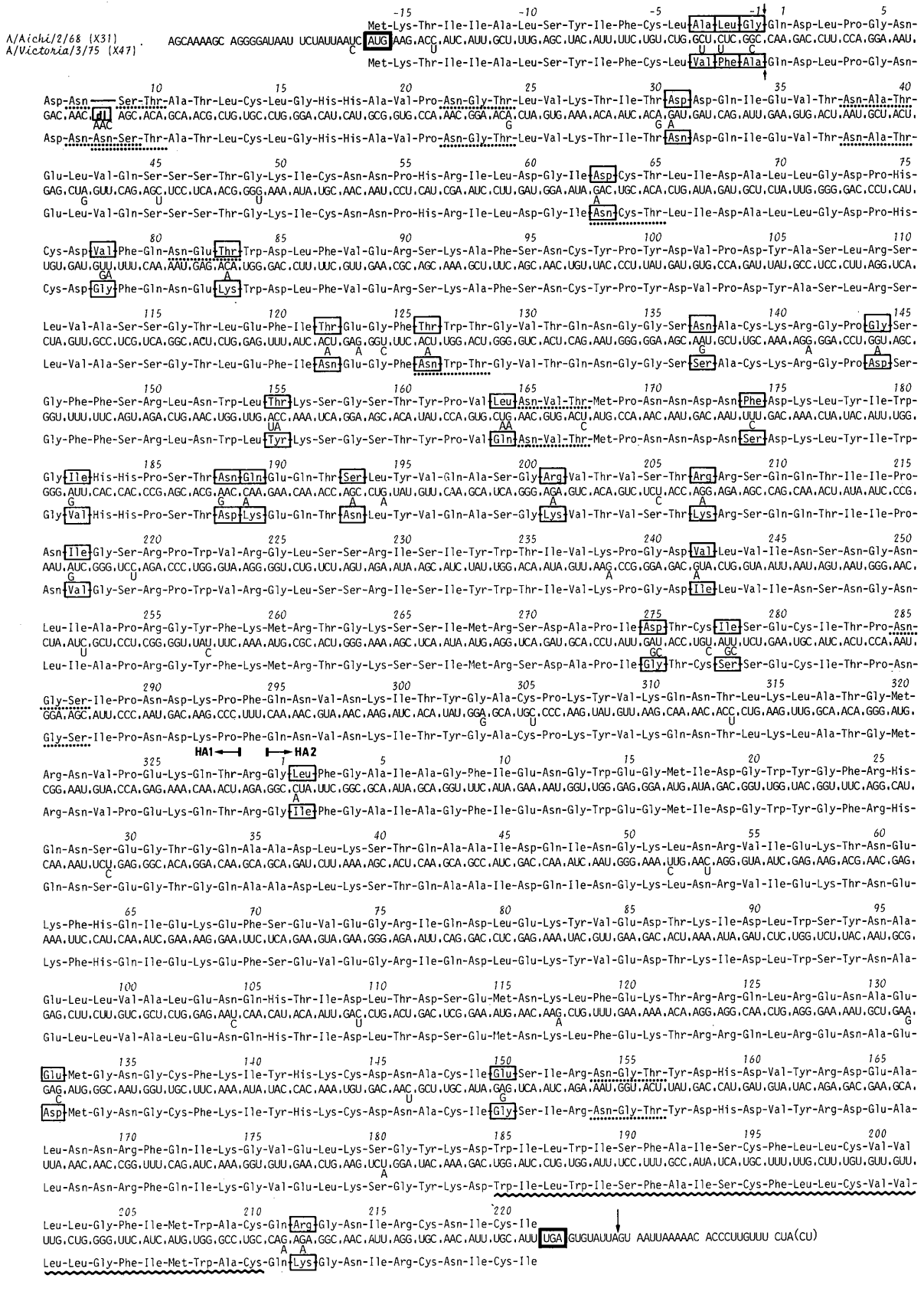

Fig. 1 Comparison of the haemagglutinin genes and deduced amino acid sequence of influenza strains A/Aichi/2/68 and A/Victoria/3/75. The nucleotide sequence of the Aichi gene is shown as the complementary viral RNA (coding strand); only the substitutions in the Victoria gene[5] are given. The Aichi nucleotide sequence is derived from cloned DNA; however, the sequence shown from the 3′ end up to the vertical arrow was missing from the cloned DNA and is assumed to be identical to the same region in Victoria (see text). The initiation and termination codons are shown in heavy boxes. Differences between amino acid sequences are shown in light boxes. Potential glycosylation sites are indicated by dotted lines. The wavy line indicates the hydrophobic region of HA2 that is embedded in the lipid membrane of the virion.

The detailed organization of the haemagglutinin RNAs belonging to the Hong Kong (H3) subtype (Victoria[5], Memphis[6–8], Aichi) seems to be identical except for the insertion of three nucleotides (specifying an asparagine residue at position 8′ of the HA1) in the relatively recent Victoria/3/75 strain (Figs 1, 2). The haemagglutinin RNA from another human subtype (H2)[11] and from an avian subtype (fowl plague)[12] are of comparable length.

The protein sequence

The haemagglutinin is synthesized as a single polypeptide chain which is subsequently modified by glycosylation and post-translational processing. The latter involves removal of an NH_2-terminal signal peptide[19] and proteolytic cleavage into HA1 and HA2 to yield mature haemagglutinin and infective virus[20,21]. Assuming that the Aichi sequence also starts with the NH_2-terminal blocking group pyroglutamic acid, as found in the Memphis H3 strain[22] (a blocked amino terminus of the HA1 is a general characteristic of H3 human strains and of two putative progenitor animal strains[23–25]), the signal sequence would be 16 amino acids long, as in Victoria[5] and Memphis[6,22]. Within the H3 subtype, the length and hydrophobic character of the signal peptide seem to be maintained, but the exact sequence is not (Figs 1, 2). There is no variation between the positions of cysteine residues in H3 strains. Also, the sequence of the Japan and fowl-plague strain can easily be aligned according to the cysteines, allowing for a minimum number of insertions/deletions between strains[5,26]. It is therefore to be expected that the disulphide bonding pattern is the same between these three subtypes and probably in all influenza A haemagglutinins, thus giving the haemagglutinin molecules a very similar three-dimensional conformation. The disulphide bonds have recently been determined for the A/Memphis/102/72 strain[27].

As in the other H3 strains Victoria and Memphis and also in the H2 Japan strain, the region excised from between HA1 and HA2 consists of a single arginine residue; this situation contrasts with fowl plague haemagglutinin, where six residues are removed[12,28]. A second hydrophobic sequence, at the NH_2 terminus of the HA2, which may have a role in interaction with and penetration of the cell membrane during infection, is highly conserved in all influenza A strains examined so far, even in a single B-type strain, and also in the NH_2 terminus of the F1 subunit of Sendai virus[25]. Near the C terminus of HA2 there is a third hydrophobic region (amino acids 185–210 of HA2 in Fig. 1), which is involved in anchoring the haemagglutinin in the lipid viral envelope. The precise functioning of the signal interrupting the translocation of membrane proteins ('stop-transfer' sequence[29]) remains to be established. However, in contrast to the signal sequence, this stop-transfer sequence seems to be constant within the H3 subtype in at least three different field strains.

The percentage of homology in the HA1 portion is modest but usually in the same range between each two strains belonging to a different subtype (between 34 and 37%), whereas the homology is higher in HA2 (H3:H2, 50%; H3:Hav1, 66%; H2:Hav1, 52%). The fact that the HA2 is much more similar between H3 and Hav1 than between two human strains (H3 and H2) whereas the HA1 varies to approximately the same extent demonstrates how difficult it is to define the evolutionary relationship between different pandemic strains. The fraction conserved in all three subtypes (taking Aichi as the H3 prototype) is only 21% in HA1, but 42% in HA2. These values illustrate the importance of the HA1 portion to the antigenic variability of the influenza haemagglutinin.

Comparison of H3 strains: genetic drift

The nucleotide sequence of the haemagglutinin gene and the deduced amino acid sequence of the protein of the Aichi/68 and Victoria/75 strains are compared in Fig. 1. There are 67 nucleotide differences (3.8%): an insertion in the Victoria gene of three nucleotides close to the NH_2 terminus of the HA1 (coding for an extra asparagine residue) and 64 nucleotide substitutions. Of these latter changes, 63 occur in the coding region: 34 are silent nucleotide substitutions and the other 29 changes cause 28 amino acid differences (there is one case, at position 155 of the HA1, of two nucleotide changes resulting in a single amino acid change). Thus, in total there are 29 amino acid differences (including one insertion) or a 5.1% amino acid divergence accumulated over a 7-yr period (1968–75). By comparing maps of soluble tryptic peptides and amino acid composition data from nine strains isolated between 1968 and 1977 with the amino acid sequence of the A/Memphis/102/72 strain, Laver and collaborators[10] found a minimum (only soluble peptides were examined) of 20 amino acid changes occurring during that period. Another conclusion was that once a residue had changed, it did not change again in the later variants, except for a few apparent reversions to the original amino acid (apparent because there is a no reason to believe that prototypes of successive epidemics as isolated from nature form a direct genealogical lineage). The same observations are made when the complete amino acid sequence of the three H3 haemagglutinin molecules, Victoria/75 (ref. 5), Memphis/72 (refs 6–8) and Aichi/68, are compared (Fig. 2). Again, a likely explanation for the apparent reversions is that, for example, the Memphis strain is not the parent of the Victoria strain. Also, at the level of silent mutation, the same phenomenon of an occasional apparent reversion is observed (not shown).

As expected, most amino acid changes were found in the HA1 part of the protein (Fig. 2, Table 1); indeed, the antigenic character is thought to reside largely, if not entirely, in this section[30–32]. Over the 1968–75 period of antigenic drift, 22 (6.7%) of the amino acids have changed in HA1 whereas only 4 (1.8%) have been replaced in HA2. Comparing all three H3 subtypes (Fig. 2), 27 variable residues are observed in HA1 and 5 in HA2. Note that there is some clustering of the amino acid changes in the middle third (amino acids 110–226) of the HA1 in

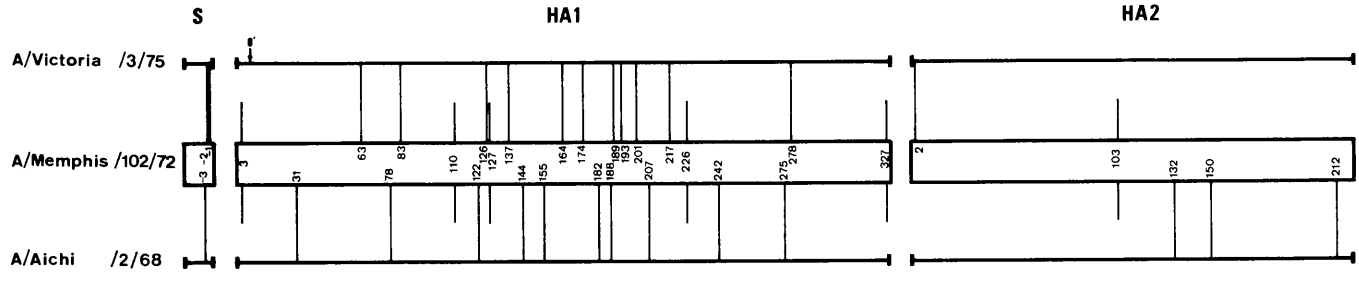

Fig. 2 Differences between the amino acid sequence of three H3 haemagglutinins derived from the virus strains A/Aichi/2/68 (the present paper), A/Memphis/102/72 (ref. 6) and A/Victoria/3/75 (ref. 5). The Memphis sequence used here is the complete one obtained by Sleigh *et al.*[6], rather than the one obtained by Ward and Dopheide[7,8], which lacks the hydrophobic region close to the C-terminus of the HA2. Apart from this, both sequences are essentially the same. Each difference is indicated by its number in the signal sequence (S), HA1 and HA2 portions of the molecule, respectively, and by a line connecting the strains involved. Two small lines at the same position in the Memphis strain indicate non-identity with either Aichi or Victoria (apparent reversions; see text).

Table 1 Variation observed among three H3 haemagglutinins

Comparison	Nucleotide changes leading to an amino acid change			Silent nucleotide changes		Total of nucleotide changes	
	No. of changes	% Nucleotide changes	% Amino acid changes	No. of changes	% Nucleotide changes	No. of changes	% Nucleotide changes
Aichi-Memphis	16 (4)	1.6 (0.6)	4.6 (1.8)	14 (9)	1.4 (1.4)	30 (13)	3.0 (2.0)
Memphis-Victoria	16 (2)	1.6 (0.3)	5.2 (0.9)	13 (3)	1.3 (0.4)	32 (5)	3.2 (0.8)
Aichi-Victoria	22 (4)	2.2 (0.6)	6.7 (1.8)	23 (10)	2.3 (1.5)	48 (14)	4.9 (2.1)
Total variable sites	27 (5)	2.7 (0.8)	8.2 (2.3)	24 (11)	2.4 (1.7)	54 (16)	5.5 (2.4)

The HA1 and HA2 regions are considered separately; the values for HA2 are given in parentheses. For the HA1 region the total number of nucleotides and amino acids is 984 and 328, respectively. The AAC insertion (asparagine residue 8') in the Victoria strain has been taken into account only in columns 3 (% amino acid changes) and 6 and 7 (total of nucleotide changes). The HA2 region comprises 663 nucleotides and 221 amino acids in all strains.

the three strains examined. A similar differential variability between both subchains of the haemagglutinin molecule has been observed by Laver and colleagues[10]. Variants have been isolated which are resistant to monoclonal antiserum, and the haemagglutinins of these have probably changed at only a single amino acid position[9]. As these variants are still neutralized by total antiserum against the parent virus, it seems very likely that a more profound change (of at least two amino acids) would be required to generate a new field strain capable of overcoming the immunity of the population.

Taking into account the number of different natural variants that have been isolated and the number of epidemics reported during the period considered[33,34], it is reasonable to assume that a not unimportant fraction of the 22 differences in HA1 observed between Aichi/68 and Victoria/75 (probably at least one-third and possibly up to two-thirds or even more, that is, 7–15 amino acids or more) have played a part in the antigenic variability. Also, if one considers the relative abundance of differences between HA1 and HA2 to reflect the antigenic variability within HA1 above the background of tolerable mutation rates, a value of 16 residues is obtained. However, these changes may be clustered in several 'antigenic sites' (each comprising a limited number of amino acids involved in the reaction with the antibody[9,35]) or they may influence the way the antigenic site is presented to the antibody by conformational changes. Thus, the analysis of amino acid sequences in a limited number of natural strains is not in itself sufficient for a delineation of antigenic sites.

Another approach is the analysis of variants selected with monoclonal antibodies. The Hong Kong variants studied by Laver and colleagues[19] have amino acid changes at positions 54, 143, 205 and a fourth one in the region 217–224 of the HA1 (the latter two changes were observed in the variants selected with the same monoclonal antibody). There are no amino acid changes around residue 54 in the three strains compared in Fig. 2, but a change of both the adjacent residue 53 and amino acid 54 (but to a different amino acid from that in the monoclonal variant) have been described[10]. Again, the amino acids next to residue 143, residues 144 (Fig. 2) and 145 (ref. 10), have been found to change in field strains. Natural changes around the third variant (205) are at positions 207 (Fig. 2 and ref. 10) and 208 (ref. 10). Thus, from all cases where a direct comparison is possible, in only one instance was an amino acid change found in the same position in a monoclonal antibody-selected variant and in a field strain (residue 54), and even then the residue by which it was replaced was completely different (to a lysine and a serine, respectively), suggesting that the immunological selection in nature works differently from that in the laboratory model experiments. The reason for this (systematic) discrepancy is unknown, although it is quite clear that the immunological pressure in an infected immune individual is very different from the selection pressure working in a murine monoclonal hybridoma to select laboratory variants.

Most of the amino acid changes in these H3 haemagglutinins have been acquired by single base substitutions and are relatively frequently observed interchanges between proteins in general[36]. Apart from this relationship directly determined by the genetic code, chemically, the substitutions are often of a conservative nature: Gly↔Ala; Ile↔Val; Leu↔Phe; Leu↔Ile; Glu↔Asp; Arg↔Lys. There is one amino acid change (Thr→Tyr) which is both infrequently observed and of a chemically drastic nature. Whether this latter change (at position 155 of HA1) is antigenically important is not known. Not surprisingly, it is the only change caused by two nucleotide substitutions.

Glycosylation sites

All potential glycosylation sites (known to be at Asn-X-Ser or Asn-X-Thr sequences) of the Aichi haemagglutinin are identical to the sites in A/Memphis/102/72, and these glycosylation sites have recently been directly confirmed[7,8]. Thus, it is probable that the glycosylation pattern has remained constant during the period 1968–72. However, one site (position 81 in HA1) disappears and two potential new sites (at positions 63 and 126) are created in Victoria relative to the 1968 prototype (and Memphis/72). Thus, the pattern of glycosylation in the HA1 is not invariable within the H3 subtype. The effect (if any) of this variation on the antigenic properties remains to be established. Note that the positions of the glycosylation sites are rather variable between different subtypes, except at position 154 of the HA2 chain, which has been conserved in all influenza A strains examined so far[5–8,11,12].

Further considerations

Not only nucleotide substitutions leading to an amino acid change but also silent mutations occur somewhat more frequently in the sequence coding for HA1 (2.3%) than in that coding for HA2 (1.5%) (Table 1). This enhanced value is statistically significant at the 5% level as tested by χ^2 analysis. If this trend is consolidated by more data, the question arises as to the meaning of this higher variability in HA1 independent of selection pressure at the protein level. It could perhaps be interpreted as base changes selected as a function of the preservation of a defined secondary structure in the viral RNA that has been more disturbed by nucleotide changes causing amino acid replacements in HA1 than HA2. A well defined secondary (and tertiary) structure may be important in the formation of the nucleoprotein complex or for packaging the eight RNAs in a molar ratio in the particle. Another case where the importance of secondary structure in the genome of RNA viruses has been well documented involves the RNA phages, where it has been shown that the number of base changes between RNA phages is 2.5 times lower in double-stranded regions than in single-stranded regions[37].

If one considers the actual number of silent and non-silent single base changes in the 61 codons for the 20 amino acids in the HA1 and the HA2 coding sequence and compares these values with those that would be expected assuming that single base replacements are acquired on a random basis, the differences are

quite striking. According to this assumption, 25.5% of the substitutions would be expected to be neutral and 74.5% would cause an amino acid change[38]. In reality, one finds in the HA1: silent changes, 23; non-silent changes, 20; non-silent changes expected taking the silent ones as the 25.5% value, 67. In the HA2 region these values are: silent changes, 10; non-silent changes, 4; non-silent changes expected, 29. This means that in HA1, 70% of the potentially non-silent mutations would be selected against; this percentage is even higher (86%) in HA2. A similar comparison between RNA phages leads to a value of 93% (ref. 37). All these deviations from randomness are statistically meaningful at the 0.1% level (χ^2 test). Although these estimates are not very accurate (it is not valid to assume an equal probability for a nucleotide to change in any direction), they are, nevertheless, a relative measure of the selection pressure to maintain the amino acid sequence in a molecule or its subregions.

Considering the nature of the nucleotide substitutions, transitions are much more frequent than transversions: of the 64 single-base substitutions between Aichi and Victoria, 49 are transitions and only 15 are transversions. Similarly, comparison of nucleotide sequences from two RNA bacteriophages, MS2 and R17, reveals 29 transitions and 5 transversions[37]. Even in the (at least at first approximation) neutral third-letter position of fourfold degenerate codons, 12 out of 15 changes are transitions between the influenza strains, and all 11 changes are transitions in the comparison between the two phages. It is quite possible that these transitions arise during replication by the formation of G · U instead of the normal G · C or A · U interactions[37].

Finally, it is too early to understand the basis of drift from the sequence information available for the haemagglutinins of natural H3N2 strains and of variants selected by monoclonal antibodies. The three-dimensional structure of the X31 haemagglutinin will probably soon be solved[39], especially as the primary amino acid sequence is now available for reference. This may then, in combination with the available and forthcoming sequence information, provide a solid framework for understanding the variability of the virus within a subtype, the mechanism of antigenic drift and the number and distribution of antigenic sites on the haemagglutinin molecule.

We thank Dr J. Skehel for a gift of X31 RNA and W. Kuziel for help with the manuscript. This research was supported by grants from the Fonds voor Kollektief Fundamenteel Onderzoek and the Gekoncerteerde Akties of the Belgian Ministry of Science. D. H. thanks the IWONL and M.V. thanks the NFWO for fellowships.

Received 16 April; accepted 17 June 1980.

1. Laver, W. G. & Webster, R. G. *Br. med. Bull.* **35,** 29–33 (1979).
2. Palese, P. & Schulman, J. L. *Proc. natn. Acad. Sci. U.S.A.* **73,** 2142–2146 (1976).
3. Scholtissek, C., Harms, E., Rohde, W., Orlich, M. & Rott, R. *Virology* **74,** 332–344 (1976).
4. McGeoch, D. J., Fellner, P. & Newton, C. *Proc. natn. Acad. Sci. U.S.A.* **73,** 3045–3049 (1976).
5. Min Jou, W. *et al. Cell* **19,** 683–696 (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.) 69–79 (Elsevier, Amsterdam, 1980).
7. 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).
8. 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).
9. Laver, W. G. *et al. Virology* **98,** 226–237 (1979).
10. Laver, W. G., Air, G. M., Dopheide, T. A. & Ward, C. W. *Nature* **283,** 454–457 (1980).
11. 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).
12. Porter, A. G. *et al. Nature* **282,** 471–477 (1979).
13. Devos, R., Gillis, E. & Fiers, W. *Eur. J. Biochem.* **62,** 401–410 (1976).
14. Devos, R., van Emmelo, J., Contreras, R. & Fiers, W. *J. molec. Biol.* **128,** 595–619 (1979).
15. Grunstein, M. & Hogness, D. S. *Proc. natn. Acad. Sci. U.S.A.* **72,** 3961–3965 (1975).
16. Smith, H. O. & Birnstiel, M. L. *Nucleic Acids Res.* **3,** 2387–2398 (1976).
17. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74,** 560–564 (1977).
18. Efstratiadis, A., Kafatos, F. C. & Maniatis, T. *Cell* **10,** 571–585 (1977).
19. McCauley, J. *et al. FEBS Lett.* **108,** 422–426 (1979).
20. Lazarowitz, S. G. & Choppin, P. W. *Virology* **68,** 440–454 (1975).
21. Klenk, H.-D., Rott, R., Orlich, M. & Blödorn, J. *Virology* **68,** 426–439 (1975).
22. Ward, C. W. & Dopheide, T. A. *Virology* **103,** 37–53 (1980).
23. Laver, W. G. & Webster, R. G. *Virology* **81,** 482–485 (1977).
24. Ward, C. W. & Dopheide, T. A. *Br. med. Bull.* **35,** 51–56 (1979).
25. Waterfield, M. D., Espelic, K., Elder, K. & Skehel, J. J. *Br. med. Bull.* **35,** 57–63 (1979).
26. Brownlee, G. G. in *Structure and Variation in Influenza Virus* (eds Laver, W. G. & Air, G. M.) 385–390 (Elsevier, Amsterdam, 1980).
27. Dopheide, T. A. & Ward, C. W. *FEBS Lett.* **110,** 181–183 (1980).
28. Klenk, H.-D. in *Structure and Variation in Influenza Virus* (eds Laver, W. G. & Air, G. M.) 213–222 (Elsevier, Amsterdam, 1980).
29. Blobel, G. *et al. Soc. exp. Biol. Symp.* **33,** 9–36 (1979).
30. Jackson, D. C., Dopheide, T. A., Russell, R. J., White, D. O. & Ward, C. W. *Virology* **93,** 458–465 (1979).
31. Jackson, D. C., Russell, R. J., Ward, C. W. & Dopheide, T. A. *Virology* **89,** 199–205 (1978).
32. Potter, C. W. & Oxford, J. S. *Br. med. Bull.* **35,** 69–75 (1979).
33. Pereira, M. S. in *Influenza Virus, Vaccines and Strategy* (ed. Selby, P.) 25–31 (Academic, London, 1976).
34. Pereira, M. S. *Br. med. Bull.* **35,** 9–14 (1979).
35. Atassi, M. Z. & Smith, J. A. *Immunochemistry* **15,** 609–610 (1978).
36. Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C. in *Atlas of Protein Sequence and Structure* Vol. 5, Suppl. 3 (ed. Dayhoff, M. O.) 345–352 (The National Biomedical Research Foundation, Silver Springs, Maryland, 1978).
37. Min Jou, W. & Fiers, W. *J. molec. Biol.* **106,** 1047–1060 (1976).
38. Jukes, T. H. & King, J. L. *Nature* **281,** 605–606 (1979).
39. Wilson, I. A., Skehel, J. J. & Wiley, D. C. in *Structure and Variation in Influenza Virus* (eds Laver, W. G. & Air, G. M.) 339–348 (Elsevier, Amsterdam, 1980).