THE ROYAL SOCIETY

*doi* 10.1098/rstb.2001.1001

# The predicted antigenicity of the haemagglutinin of the 1918 Spanish influenza pandemic suggests an avian origin

**G. G. Brownlee*** **and E. Fodor**

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

In 1982 we characterized the antigenic sites of the haemagglutinin of influenza A/PR/8/34, which is an influenza strain of the H1 subtype that was isolated from humans in 1934, by studying mutants which escaped neutralization by antibody. Four antigenic sites, namely Cb, Sa, Sb and Ca, were found to be located near the tip of the trimeric haemagglutinin spike. Based on the sequence of the haemagglutinin of the 1918 Spanish influenza, we can now specify the extent of divergence of antigenic sites of the haemagglutinin during the antigenic drift of the virus between 1918 and 1934. This divergence was much more extensive (40%) than the divergence (20%) in predicted antigenic sites between the 1918 Spanish influenza and an avian H1 subtype consensus sequence. These results support the hypothesis that the human 1918 pandemic originated from an avian virus of the H1 subtype that crossed the species barrier from birds to humans and adapted to humans, presumably by mutation and/or reassortment, shortly before 1918.

**Keywords:** influenza; haemagglutinin; antigenicity; mutants; Spanish influenza

## 1. INTRODUCTION

The haemagglutinin of influenza virus is a variable, membrane-bound glycoprotein that is the major external structural protein of the virus (for reviews see Lamb & Krug 1996; Murphy & Webster 1996). New strains of influenza virus evolve every year or so by mutation in antigenic sites in the haemagglutinin. This antigenic drift allows the virus to evade the immune response in people with pre-existing immunity to influenza virus.

Antigenically unrelated haemagglutinin subtypes (antigenic shift) also appear or reappear in the human population, but much less frequently, every 15–30 years or so, almost certainly by influenza virus crossing the species barrier from an avian or animal host into humans. Influenza virus is widespread in the animal kingdom and infects birds and other mammals, e.g. pigs, horses and even seals, besides humans. Fifteen different distinct haemagglutinin subtypes (H1–H15) have been recognized (Webster 1999), yet only three, the H1–H3 subtypes, are known to have caused pandemics in humans (table 1). Direct evidence for avian to human transmission emerged in 1997 in Hong Kong when 18 people were known to have been infected and six died as a result of transmission of a lethal chicken influenza virus of H5 subtype to humans (Webster 1999). The segmented nature of the eight RNA gene segments of influenza virus (for a review, see Lamb & Krug 1996) favours the adaptation of avian strains to humans. Thus, genetic mixing or reassortment of the eight RNA gene segments can occur in a mixed infection by a human and an avian virus, thereby generating novel viruses.

The 1918–1919 Spanish influenza pandemic was one of most significant causes of death ever recorded for an epidemic and it has been estimated that up to 20 million people died throughout the world. A surprising feature of the 1918–1919 pandemic was the unusually high mortality in healthy adults in the age range of 15–35 years as compared to all other known influenza pandemics. In contrast, the age group over 60 years was no more at risk in 1918 than in later epidemics, e.g. in 1934 (Gale 1959). The reason for the unusual age-related mortality associated with the 1918–1919 pandemic has been the subject of speculation ever since. An obvious explanation is that there was a cohort of the population born between 1880 and 1900 who had no immunity to the 1918–1919 influenza strain, whereas people born before 1880 had some residual immunity that protected them. At first sight this hypothesis seems unlikely because influenza had been prevalent since the Russian influenza pandemic of 1889–1890 (figure 1). By 1918 there must have been significant population immunity to this virus or antigenic variants derived from it. However, it is possible that the 1889–1890 pandemic was caused by an influenza virus of different subtype (Murphy & Webster 1996), which may have given no significant cross-protective immunity to people infected with the 1918–1919 H1 subtype virus. An alternative explanation is that only a proportion of the population in 1918–1919 had any immunity to the pandemic originating in 1889–1890 or to subsequent epidemics occurring between 1890 and 1918. Those with immunity survived, whereas those with none succumbed. If it is assumed that the much earlier pandemic of 1840,

*Author for correspondence (george.brownlee@path.ox.ac.uk).

 

Table 1. Important influenza A virus pandemic strains.

(Hong Kong (bird) influenza only infected a few people in Hong Kong.)

| name | year first seen | haemagglutinin subtype |
|---|---|---|
| Spanish influenza | 1918 | H1 |
| Asian influenza | 1957 | H2 |
| Hong Kong influenza | 1968 | H3 |
| Russian influenza | 1977 | H1 |
| Hong Kong (bird) influenza | 1997 | H5 |

which petered out around 1860 (figure 1), was caused by an H1 subtype virus, then people born between 1840 and 1860 would have acquired some resistance to an H1 subtype virus. Thus, people over 60 years old in 1918 might have been no more at risk from influenza in 1918 simply because of immunity acquired from a previous epidemic originating in 1840.

The historical epidemiological records of deaths from influenza in Victorian England and Wales (Gale 1959) give us another important clue that might explain the severity of the 1918–1919 pandemic. Deaths from influenza between 1861 and 1889 were insignificant in England and Wales (figure 1). 'In 1889 this country had been free from pandemic influenza for more years than in any previous epoch since the middle of the seventeenth century' (quoted in Gale 1959, p. 47). As a consequence, in 1918 there would have been little population immunity to influenza in people born between 1860 and 1918 other than to the 1889 pandemic or to subsequent epidemics between then and 1918.

Despite the fact that the absence of pre-existing immunity must have played an important part in the susceptibility of people to infection by the 1918–1919 virus, an explanation for its unusually high mortality may still be required. Most commentators have suggested that the 1918 virus had some special virulence features. That is why it was so lethal. However, the alternative argument, that the virus lacked any specific virulence features, cannot be excluded. Thus, it is possible that the unusually high death rate in 1918–1919 was simply a lack of population immunity. Under this hypothesis, the 1918–1919 pandemic was caused by a new virus, which had recently crossed the species barrier from birds to humans and adapted to humans by mutation and/or reassortment. Only the elderly had any specific immunity, which was acquired through infection from the earlier 1840 pandemic, thus protecting them and not the younger age groups from excess deaths.

The sequences of three of the gene segments of the 1918 influenza virus, namely the haemagglutinin, neuraminidase and segment 8 (the NS1 and nuclear export protein coding sequence), have recently been deduced from reverse transcriptase polymerase chain reaction analysis of frozen or formalin-preserved lung tissues from people who died of influenza in 1918 (Reid *et al*. 1999, 2000; Basler *et al*. 2001). The 1918 virus was, as expected, an H1N1 subtype virus. However, the haemagglutinin and neuraminidase gene sequences revealed no unusual sequence features that could explain their virulence in 1918. Furthermore, novel viruses that were constructed from the 1918 NS1 and nuclear export protein genes in a background of genes from a 1933 influenza virus (A/WSN/33) by using the recent plasmid-based rescue methods (Fodor *et al*. 1999; Neumann *et al*. 1999) showed no specific virulence in mice. However, the virulence of the 1918 haemagglutinin and neuraminidase gene segments remains to be tested experimentally.

There is little doubt that the antigenicity of the haemagglutinin must have been a primary factor in the 1918–1919 pandemic, irrespective of whether other specific virulence features were present in the 1918 virus and were encoded by genes other than the haemagglutinin. The antigenicities of the 1968 (H3 subtype) and 1934 (H1 subtype) viruses were studied extensively in the 1980s (Caton *et al*. 1982). Four main antigenic sites were identified near the tip of the globular head of the haemagglutinin in both subtypes (figure 2). Despite this common feature there were differences in the exact positions of two of the four antigenic sites because of the presence of N-linked carbohydrate, which masked potential antigenic sites. Now that the sequence of the haemagglutinin of the 1918 influenza virus that is also of the H1 subtype, is known (Reid *et al*. 1999) it was of interest to compare its predicted antigenic sites with those previously studied in the 1934 virus. Furthermore, a comparison of the antigenic sites of the 1918 virus with those of the classical swine 1930 virus and an avian H1 subtype consensus should test the hypothesis that the 1918 virus was antigenically derived from an avian precursor.

Our results suggested that the 1918 virus is more closely related to the avian H1 consensus and the 1930 swine virus than to other known 1934 human viruses. This study thus confirms previous antigenic analysis that suggested that the antigenic sites of the 1918 virus resembled the avian consensus closely. However, here we examine the data in much more detail than before (Reid *et al*. 1999).

## 2. METHODS

Influenza haemagglutinin sequences were accessed from the Los Alamos influenza sequence database (www-flu.lanl.gov/). A consensus of the amino acid sequences of the HA1 domain of seven avian haemagglutinin sequences of the H1 subtype, which were isolated from 1976 to 1990, was constructed (data not shown) by multiple alignment of the following sequences: A/duck/Alberta /35/76 (D10477), A/duck/WI/259/80 (407004), A/turkey/Minnesota/1661/81 (AAD25305), A/duck/WI/1938/80 (407002), A/chicken/Hong Kong/14/76 (U46782), A/duck/Bavaria/1/77 (AF091313) and A/turkey/Germany/2482/90 (U96766). The consensus antigenic sites (which are named 'duck' in figure 3) have identical amino acid sequences to those in A/duck/WI/1938/80 and A/turkey/Minnesota/1661/81. The other sequences used were A/swine/Iowa/15/30, (AF091308), A/PR/8/34 (J 02143) and A/South Carolina/1/18 (AF117241).

The antigenic Cb site (Caton *et al*. 1982) was originally defined by amino acid residues 78, 79 and 81–83 of the haemagglutinin using the Winter *et al*. (1981) numbering system, which is based on the alignment of the haemagglutinin amino acid sequence of A/PR/8/34 (H1 subtype) with A/Aichi/2/68 (H3 subtype). We widened the definition here in order to include residue 80 because A/USSR/90/77 diverged from A/PR/8/34 at this position (Raymond *et al*. 1983).

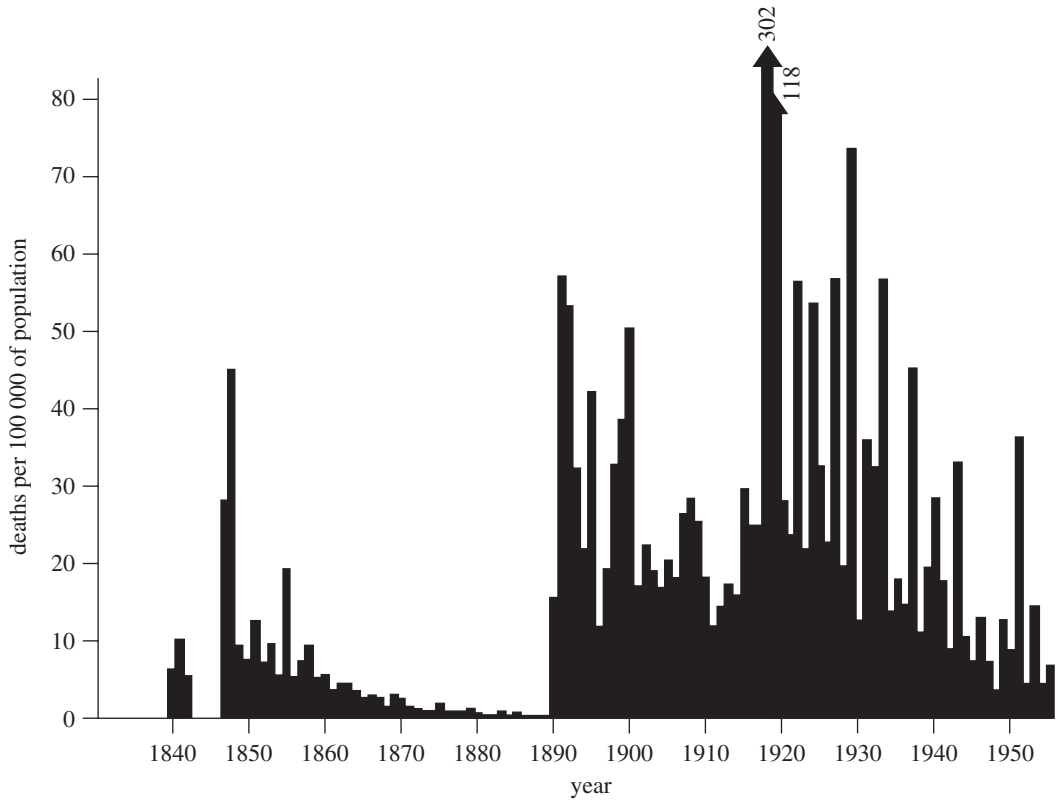


Figure 1. Deaths from influenza in England and Wales per 100 000 of the population, 1840–1955 (from Gale 1959, p. 48, reproduced by permission of Penguin Books Ltd.).

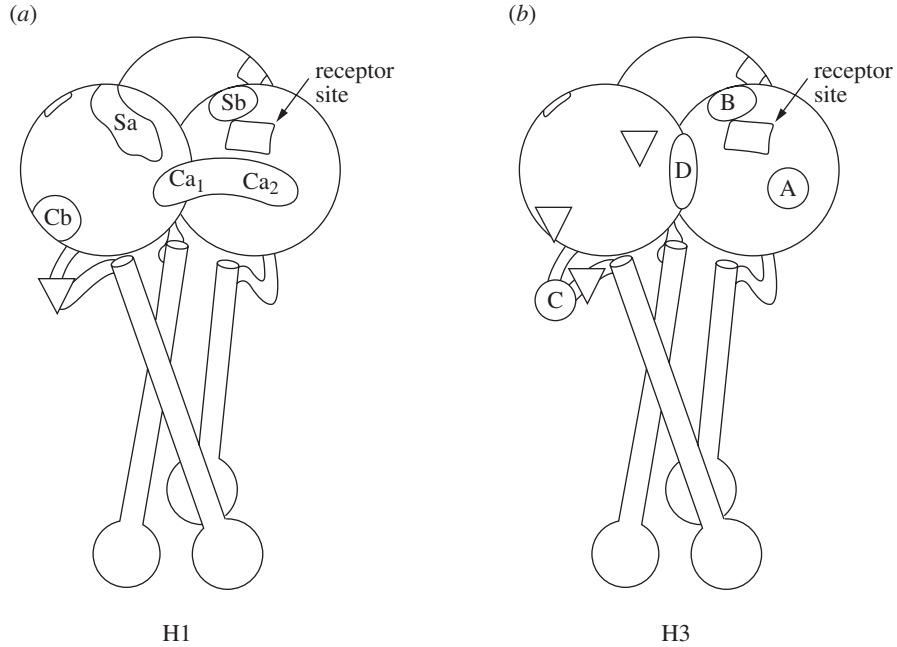


Figure 2. Antigenic sites on the surface of the haemagglutinin trimer. (*a*) The H1 subtype virus A/PR/8/34 showing the approximate location of the Cb, Sa, Sb and Ca antigenic sites on the globular head. (*b*) H3 subtype viruses showing the location of antigenic sites A–D. The inverted open triangles show the carbohydrate attachment sites on the globular head; those elsewhere are omitted. The receptor binding sites are shown. Note that antigenic sites are present on each monomer, but are only shown on one monomer for simplicity (from Caton *et al*. 1982).

The Sa site was originally defined by amino acid residues 128, 129, 158, 160, 162, 163 and 165–167 (Caton *et al*. 1982). We extended this site to include residues 156 and 159, even though they had been formally assigned to the Sb site by Caton *et al.* (1982), as they were immediately adjacent to Sa site residues. Residues 157 and 164 were included because they were bracketed by antigenic residues. Residue 161 was excluded because it was conserved in different haemagglutinin subtypes (Winter *et al*. 1981).

If we exclude residues 156 and 159, which for convenience have been included in the Sa site here (see above), then the

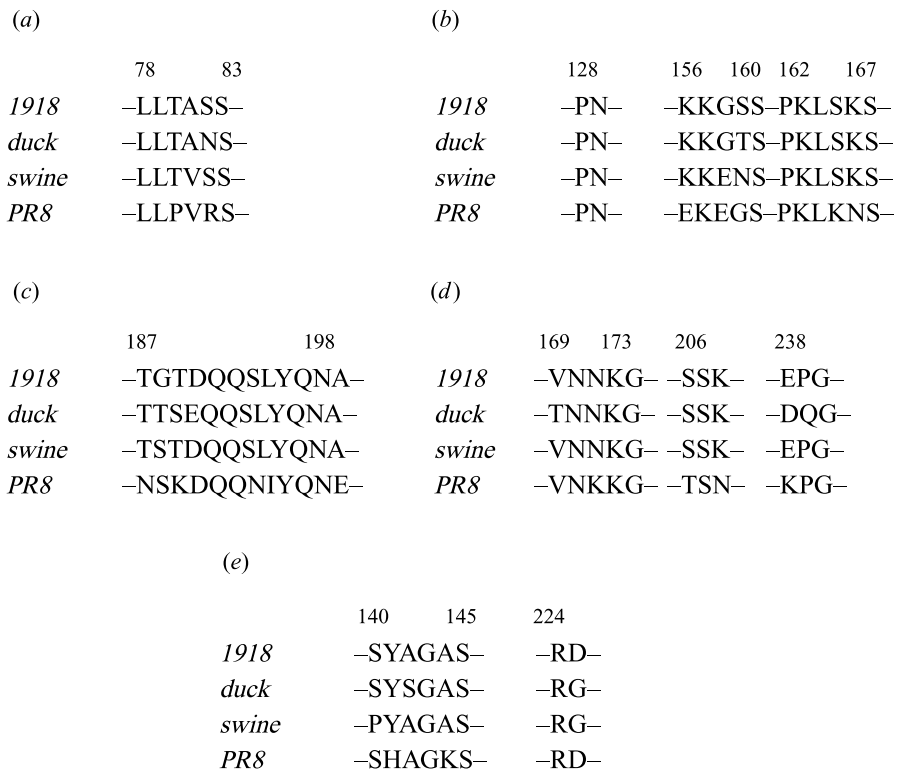


Figure 3. Amino acid sequence alignment of antigenic sites (*a*) Cb, (*b*) Sa and (*c*) Sb and the (*d*) $Ca_1$ and (*e*) $Ca_2$ subsites of A/South Carolina/1/18 (1918), an A/avian H1 consensus (duck), A/swine/Iowa/15/30 (swine) and A/PR/8/34 (PR8). See the text for further details.

antigenic Sb site was originally defined by residues 192, 193, 196 and 198. We initially extended this site to include all residues from 190 to 198, either because of the immediate proximity of the additional residues to known antigenic ones or because of known antigenic variation in field strains (Raymond *et al.* 1986). Subsequently, residues 187–189 were also included because they are highly variable in the alignment of seven avian H1 subtype sequences (see above). Residues 187–189 are also variable in the sequence comparison shown in figure 3 and are potentially involved in antigenic variation.

The antigenic Ca site is formed from two subsites—$Ca_1$ and $Ca_2$. These subsites, although on opposite sides of the haemagglutinin monomer, are actually close to one another, forming a single antigenic site in a cleft between two adjacent monomers (figure 2). The $Ca_1$ subsite was originally defined by residues 169, 173, 207 and 240 on the basis of the isolation of laboratory escape mutants (Caton *et al.* 1982). Residues 172 and 208 were subsequently included since they varied in the evolution of H1 field strains (Raymond *et al.* 1986). In addition, we included residues 170, 171, 206, 207 and 239 as they were bracketed by or adjacent to antigenic residues. Residue 238 was included because it was a nearby surface residue that varied between A/USSR/90/77 and A/PR/8/34 (Raymond *et al.* 1983). The $Ca_2$ subsite was originally defined by residues 140, 143 and 145 and also included residues 224 and 225 (Caton *et al.* 1982). Residue 142 was included here because it varied in some field strains (Raymond *et al.* 1986). Residues 141 and 144 were included because they were bracketed between two adjacent antigenic residues.

## 3. RESULTS

Figure 2 shows the location of antigenic sites that were previously located (Caton *et al.* 1982) on the globular head of the haemagglutinin trimer (H1 subtype) of influenza A/PR/8/34 (figure 2*a*). For comparison the antigenic sites for influenza strains of the H3 subtype are also shown (figure 2*b*). There are four antigenic sites in both subtypes. Two are located in roughly similar positions (Sb is equivalent to B and Ca is equivalent to D) and the others are essentially in different locations due to the masking effect of N-linked carbohydrate.

The antigenic sites Sa and Sb are close to one another near the receptor-binding site on the tip of each monomer in the case of influenza A/PR/8/34. However, site Cb lies somewhat below the tip of the globular head whilst the Ca site is formed from two subsites, i.e. $Ca_1$ and $Ca_2$. These subsites are on opposite sides of the monomer, but are actually one site formed in a cleft between two adjacent monomers (figure 2*a*). It is presumed that antibody binding to these antigenic sites blocks the receptor-binding pocket on the tip of the haemagglutinin monomer by steric hindrance. This inhibits viral infection by preventing the attachment of virus to sialic acid receptors on host cells.

Figure 3 shows the amino acid sequence alignment of the antigenic sites Cb, Sa and Sb and the $Ca_1$ and $Ca_2$ subsites of A/South Carolina/1/18 (1918), an A/avian H1 consensus (see § 2) (duck), A/swine/Iowa/15/30 (swine) and A/PR/8/34 (PR8). The purpose of this alignment was first to compare the predicted antigenic sites of the haemagglutinin of 1918 influenza with the prototype human influenza virus of the H1 subtype A/PR/8/34 and, second, to compare them with A/swine/Iowa/30 which is the oldest known classical swine influenza virus of the H1 subtype. Avian consensus antigenic sites derived from seven viruses of the H1 subtype (see § 2) that were isolated

Table 2. Number of amino acid differences in the antigenic sites of haemagglutinin.

| site | pairwise comparisons | total numbers | number of differences |
|---|---|---|---|
| Cb site | 1918 versus duck | 6 | 1 |
| | 1918 versus swine | 6 | 1 |
| | 1918 versus PR8 | 6 | 3 |
| | duck versus swine | 6 | 2 |
| | duck versus PR8 | 6 | 3 |
| | swine versus PR8 | 6 | 2 |
| Sa site | 1918 versus duck | 13 | 1 |
| | 1918 versus swine | 13 | 2 |
| | 1918 versus PR8 | 13 | 5 |
| | duck versus swine | 13 | 2 |
| | duck versus PR8 | 13 | 5 |
| | swine versus PR8 | 13 | 4 |
| Sb site | 1918 versus duck | 12 | 3 |
| | 1918 versus swine | 12 | 1 |
| | 1918 versus PR8 | 12 | 6 |
| | duck versus swine | 12 | 3 |
| | duck versus PR8 | 12 | 7 |
| | swine versus PR8 | 12 | 5 |
| $Ca_1$ subsite | 1918 versus duck | 11 | 3 |
| | 1918 versus swine | 11 | 0 |
| | 1918 versus PR8 | 11 | 4 |
| | duck versus swine | 11 | 3 |
| | duck versus PR8 | 11 | 6 |
| | swine versus PR8 | 11 | 4 |
| $Ca_2$ subsite | 1918 versus duck | 8 | 2 |
| | 1918 versus swine | 8 | 2 |
| | 1918 versus PR8 | 8 | 2 |
| | duck versus swine | 8 | 2 |
| | duck versus PR8 | 8 | 4 |
| | swine versus PR8 | 8 | 4 |

between 1976 and 1990 were also included in this comparison because avian strains are thought to be in a state of evolutionary stasis and believed not to undergo significant antigenic drift (Murphy & Webster 1996).

Table 2 lists the number of amino acid differences for each of the six pairwise comparisons at each antigenic site or subsite. Thus, the Cb site of the 1918 influenza differs from swine/30 at only one of the six positions (residue 81) and from the avian consensus (duck) at another single residue (residue 82) (figure 3). In contrast, the 1918 virus differs from the 1934 A/PR/8/34 virus in three of the six positions (residues 80–82). Thus, the 1918 virus is closer to the avian consensus and swine sequence than to one of the earliest isolated human viruses when considering the Cb site. The other Cb site pairwise comparisons (duck versus swine, duck versus PR8 and swine versus PR8) give results (table 2) consistent with the interpretation that the avian and swine sequences are closest to the 1918 influenza in antigenicity. Is this pattern of difference valid for all antigenic sites? The data in table 2 suggest that the pattern of differences noted for the Cb site is also valid for the Sa and Sb sites, but not for the $Ca_1$ and $Ca_2$ subsites.

Table 3 summarizes the number of amino acid differences obtained by summing the amino acid differences for all four antigenic sites, including the $Ca_1$ and $Ca_2$ subsites. This 'averaged' result confirms the detailed picture

Table 3. Total number of amino acid differences in all of the haemagglutinin antigenic sites.

| strains | differences | % difference |
|---|---|---|
| 1918 versus duck | 10/50 | 20 |
| 1918 versus swine | 6/50 | 12 |
| 1918 versus PR8 | 20/50 | 40 |
| duck versus swine | 12/50 | 24 |
| duck versus PR8 | 25/50 | 50 |
| swine verus PR8 | 19/50 | 38 |

derived from the Cb site alone. The 1918 virus is closest in its haemagglutinin antigenic sequences to the swine 1930 virus (six differences) and reasonably close to the avian consensus (10 differences). On the other hand, there are 20 antigenic differences between the haemagglutinin of the 1918 and A/PR/8/34 viruses. This confirms the idea that the haemagglutinin of the 1918 virus is antigenically more distant from the A/PR/8/34 virus than either the swine 1930 or the avian H1 consensus.

Table 4 shows the percentage amino acid identity summed for all four antigenic sites in the various pairwise comparisons. This confirms the similarity of the predicted antigenic sites of the haemagglutinin of the 1918 virus with that of the swine/30 virus (88% identity) and the avian consensus sequence (80% identity). In contrast, there is only 60% identity between the 1918 haemagglutinin and the haemagglutinin of the A/PR/8/34 virus.

## 4. DISCUSSION

The purpose of this paper was to compare the predicted antigenic sites of the haemagglutinin of the 1918 Spanish influenza with the known sites previously studied in influenza A/PR/8/34 (Caton *et al.* 1982). Both viruses are of the H1 subtype. Although antigenic sites have been reasonably well defined for influenza A/PR/8/34, it is still difficult to obtain a complete picture of the antigenicity of the haemagglutinin experimentally. However, to a good approximation (see Caton *et al.* (1982) for a fuller discussion of the limitations of the experimental definition of antigenic sites) there are four antigenic sites, namely Cb, Sa, Sb and Ca, near the tip of the globular head of the trimeric haemagglutinin spike.

We have also compared the amino acid sequences of the predicted antigenic sites of the haemagglutinin of the 1918 virus with those of the classic H1 subtype swine influenza of 1930 and an avian H1 subtype consensus sequence derived from strains that were isolated between 1976 and 1990. Our purpose was to test the hypothesis that the 1918 virus was derived from an avian H1 subtype strain. We also included a comparison with the classic swine influenza of 1930 because serological evidence had suggested that they were antigenically closely related.

The decision as to which amino acids should be included in the antigenic sites was somewhat subjective. It was based primarily on amino acid residues that were identified as being antigenic because of the isolation of mutants that escaped neutralization by antibody (Caton *et al.* 1982). However, this would have been too restrictive a definition of antigenic sites because the experiments for

Table 4. Amino acid identity (%) of all of the haemagglutinin antigenic sites in various pairwise comparisons.

| | 1918 | duck | swine | PR8 |
|---|---|---|---|---|
| 1918 | 100 | 80 | 88 | 60 |
| duck | — | 100 | 76 | 50 |
| swine | — | — | 100 | 62 |
| PR8 | — | — | — | 100 |

isolating escape mutants could not possibly have defined all antigenic residues (for further discussion see Caton *et al*. 1982). Instead, we chose a broader definition of antigenic sites, thereby widening the definition in order to include nearby amino acid residues. Thus, adjacent residues that are known to be involved in the evolution of H1 strains and, therefore, implicated in antigenic variation were included (Raymond *et al*. 1983, 1986). Other highly variable residues adjacent or very close to known antigenic sites were also included. Thus, the antigenic Cb site was originally defined by amino acid residues 78, 79 and 81–83 of the haemagglutinin by isolation of escape mutants at these positions. We widened the definition here to include residue 80 because there was a potential antigenic variant in influenza A/USSR/90/77 at this position (Raymond *et al*. 1983). Details of the reasons for including the particular residues chosen for the remaining antigenic sites are given in § 2. Overall our definition of antigenic sites includes 50 amino acid residues in four antigenic sites.

The location of N-linked carbohydrate in the HA1 domain of haemagglutinin is important since carbohydrate can mask antigenic sites (Caton *et al*. 1982) and might invalidate the sequence comparisons of potential antigenic sites in tables 2–4. Both the 1918 and 1934 (A/PR/8/34 Mount Sinai strain) HA1 domains of haemagglutinin have four predicted glycosylation sites. However, these are not located in identical positions, since A/PR/8/34 lacks one site present in the 1918 virus (residue 94) and has gained a new site (residue 271) that is absent in the 1918 virus (Reid *et al*. 1999). Since both these 'swapped' carbohydrate moieties are distant from the antigenic sites it is unlikely they will affect antigenicity. Interestingly, the pattern of carbohydrate residues in the HA1 domain of the haemagglutinin of the 1918 virus, the avian consensus sequence and the swine 1930 virus are identical to one another, thereby emphasizing their close relationship.

Our overall results (table 4) show that the antigenic sites of the haemagglutinin of the 1918 influenza virus were 80% and 88% identical to the H1 avian consensus and swine 1930 sequences, respectively. On the other hand, there is only 60% identity between the predicted antigenic sites of the haemagglutinin of the 1918 virus and the A/PR/8/34 virus. It is no surprise that the 1918 virus is antigenically closely related to the 1930 swine influenza (tables 3 and 4) since these two viruses are known to be antigenically related from serological studies. Moreover, current opinion favours the hypothesis that the 1918 human virus was transmitted to pigs shortly after 1918 (Webster 1999). The fact that the antigenic sites of the haemagglutinin of the 1918 virus are closely related to an avian virus H1 subtype consensus is striking. It supports previous conclusions, which are also based on antigenicity, that the 1918 influenza virus was of avian origin (Reid *et al*. 1999; Webster 1999). Perhaps the most remarkable finding is how little the avian H1 consensus haemagglutinin sequence has diverged from the 1918 human influenza, despite the fact that the avian sequence was derived from avian strains that were isolated over 50 years later. This observation is consistent with the proposal that the avian haemagglutinin is in evolutionary stasis in birds (Murphy & Webster 1996).

This work was supported by a Medical Research Council programme grant (G9523972) to G.G.B.

## REFERENCES

Basler, C. F. (and 11 others) 2001 Sequence of the 1918 pandemic influenza virus nonstructural gene (NS) segment and characterization of recombinant viruses bearing the 1918 NS genes. *Proc. Natl Acad. Sci. USA* **98**, 2746–2751.

Caton, A. J., Brownlee, G. G., Yewdell, J. W. & Gerhard, W. 1982 The antigenic structure of the influenza virus A/PR/8/34 hemagglutinin (H1 subtype). *Cell* **31**, 417–427.

Fodor, E., Devenish, L., Engelhardt, O. G., Palese, P., Brownlee, G. G. & García-Sastre, A. 1999 Rescue of influenza A virus from recombinant DNA. *J. Virol.* **73**, 9679–9682.

Gale, A. H. 1959 *Epidemic diseases*. London: Penguin.

Lamb, R. A. & Krug, R. M. 1996 Orthomyxoviridae: the viruses and their replication. In *Fields virology*, 3rd edn (ed. B. N. Fields, D. M. Knipe, P. M. Howley, R. M. Chanock, J. L. Melnick, T. P. Monath, B. Roizman & S. E. Straus), pp. 1353–1395. Philadelphia: Lippincott-Raven.

Murphy, B. R. & Webster, R. G. 1996 Orthomyxoviridae. In *Fields virology*, 3rd edn (ed. B. N. Fields, D. M. Knipe, P. M. Howley, R. M. Chanock, J. L. Melnick, T. P. Monath, B. Roizman & S. E. Straus), pp. 1397–1445. Philadelphia: Lippincott-Raven.

Neumann, G. (and 11 others) 1999 Generation of influenza A viruses entirely from cloned cDNAs. *Proc. Natl Acad. Sci. USA* **96**, 9345–9350.

Raymond, F. L., Caton, A. J., Cox, N. J., Kendal, A. P. & Brownlee, G. G. 1983 Antigenicity and evolution amongst recent influenza viruses of H1N1 subtype. *Nucl. Acids Res.* **11**, 7191–7203.

Raymond, F. L., Caton, A. J., Cox, N. J., Kendal, A. P. & Brownlee, G. G. 1986 The antigenicity and evolution of influenza H1 haemagglutinin, from 1950–1957 and 1977–1983: two pathways from one gene. *Virology* **148**, 275–298.

Reid, A. H., Fanning, T. G., Hultin, J. V. & Taubenberger, J. K. 1999 Origin and evolution of the 1918 'Spanish' influenza virus hemagglutinin gene. *Proc. Natl Acad. Sci. USA* **96**, 1651–1656.

Reid, A. H., Fanning, T. G., Janczewski, T. A. & Taubenberger, J. K. 2000 Characterization of the 1918 'Spanish' influenza virus neuraminidase gene. *Proc. Natl Acad. Sci. USA* **97**, 6785–6790.

Webster, R. G. 1999 1918 Spanish influenza: the secrets remain elusive. *Proc. Natl Acad. Sci. USA* **96**, 1164–1166.

Winter, G., Fields, S. & Brownlee, G. G. 1981 Nucleotide sequence of the haemagglutinin gene of a human influenza virus H1 subtype. *Nature* **292**, 72–75.