*Proc. Natl. Acad. Sci. USA*
Vol. 78, No. 12, pp. 7639–7643, December 1981
Evolution

# Sequence relationships among the hemagglutinin genes of 12 subtypes of influenza A virus

(antigenic shift/antigenic drift/dideoxy sequence analysis)

GILLIAN M. AIR

John Curtin School of Medical Research, Australian National University, P.O. Box 334, Canberra, A.C.T. 2601, Australia

*Communicated by Frank Fenner, June 3, 1981*

**ABSTRACT** **Nucleotide sequences of the 3′ 20% of the hemagglutinin gene of 32 influenza A virus strains from the 12 known hemagglutinin subtypes have been determined. Although the sequences of hemagglutinin genes and proteins of different subtypes differ greatly, cysteine and some other amino acid residues are totally conserved, presumably reflecting evolution of the 12 different hemagglutinins from a single gene. When viruses of one subtype, isolated over a period of time, are compared, the hemagglutinin gene and protein sequences show a slow accumulation of nucleotide changes and some amino acid changes. Since sequence data from the genes coding for the matrix and nonstructural proteins also show an accumulation of changes with time, it seems that antigenic selection (of the surface antigens) does not contribute significantly to the rate of change of influenza gene sequences. Although the rate of nucleotide change during drift is more than sufficient to account for the amino acid sequence differences observed in the 12 subtypes, there is a clear distinction, by antigenic as well as sequence analyses, between viruses of one subtype (0–9% amino acid variation) and viruses of other subtypes (20–74% amino acid variation). No virus has yet been found that is intermediate between subtypes.**

Influenza remains an uncontrolled disease in man, largely because of variations in the amino acid sequences of the two surface antigens, hemagglutinin (HA) and neuraminidase. Year by year, the influenza A virus becomes progressively more resistant to neutralization by antibody made against older HAs as antigenic changes accumulate. In addition, every so often, a "new" virus appears having such differences in the HA sequence and antigenic properties that it is clearly not derived by mutation from previously circulating viruses. The most recent examples are the appearance of "Asian" flu in 1957 (H2 subtype), "Hong Kong" flu in 1968 (H3 subtype) and "Russian" flu in 1977 (H1 subtype). It is possible that the new HA is normally derived from viruses circulating or sequestered in species other than man (1) and, for this reason, much work has been focused on characterizing influenza A viruses from birds and animals. Currently, 12 subtypes of HA are defined, designated H1–H12 (2), based on lack of serological crossreactivity.

HA contains two polypeptide chains, HA1 and HA2, which are derived by proteolysis of a single precursor molecule that includes loss of a signal peptide (3). Both polypeptides are glycosylated and contain hydrophobic regions (3, 4), making protein sequence analysis difficult. The amino acid sequence has been determined directly for HA1 and HA2 of only one strain of virus, of the Hong Kong (H3) subtype (4) and, in spite of much effort, part of the hydrophobic COOH-terminal sequence of HA2 could not be determined.

On the other hand, the segmented single-stranded genome is ideal for making a cDNA copy and inserting this into plasmids for amplification in *Escherichia coli* and nucleotide sequence analysis. Thus, the sequences of the fowl plague virus HA gene (H7) (5) and some human strains, an Asian (H2) type (6) and several Hong Kong (H3) strains (7–9), have been determined. From the nucleotide sequences, amino acid sequences were predicted using the genetic code.

When the sequences of these three subtypes of HA (H2, H3, and H7) are compared, at either the nucleotide or amino acid level, they are remarkably dissimilar. The complete lack of antigenic crossreactivity among them is well characterized but, because substitution of a single amino acid can completely destroy interaction between a particular antigenic determinant and monoclonal antibody (10) and because there is a limit to the number of independent determinants on the hemagglutinin (11, 12), the degree of divergence in sequence is remarkable.

This paper describes nucleotide sequence data from the HA genes of 32 virus strains. Most of the influenza sequences published have been derived from cloned cDNA copies (5–9), but there are possibilities of artifacts being introduced during the cloning procedure or cloning an unrepresentative molecule (5, 7). In this study, the sequences have been derived by dideoxy sequence analysis (13) of cDNA primed from the 3′ end of the viral RNA strand. The length of sequence obtained ranges up to 380 nucleotides and includes the 5′-noncoding region of the mRNA [except for the extra nucleotides derived from cell mRNA (14)], the signal peptide coding sequence, and the region that codes for up to a third of the HA1 polypeptide. The information for each HA is far from complete but is a significant proportion and of manageable size for detailed comparison.

## MATERIALS AND METHODS

Viruses were grown in embryonated eggs and purified by adsorption to and elution from chicken erythrocytes followed by sucrose density-gradient centrifugation (15).

Isolation of HA RNA and sequence analysis of the cDNA by using as primer a synthetic dodecanucleotide complementary to the 3′ 12 nucleotides common to all segments of influenza A viral RNA (16) (Collaborative Research, kindly given by C. J. Lai) and the dideoxy method (13, 17) were as described (18, 19). Sequences were obtained from HA genes of the following viruses: H1—A/NWS/33, A/PR/8/34 (Mt. Sinai), A/Bellamy/42, A/USSR/90/77, A/Fort Warren/1/50, A/Loyang/4/57, A/swine/Wisconsin/15/30, A/New Jersey/11/76, H2–A/RI/5⁻/57, A/Tokyo/3/67, A/Netherlands/68, A/Berkeley/68, A/duck/GDR/72, A/duck/Alberta/77/77, A/pintail/Alberta/293/77; H3—A/Memphis/1/71, A/black duck/Australia/702/78, A/duck/Ukraine/1/63; H4—A/duck/Alberta/28/76; H5—A/

Abbreviations: HA, hemagglutinin; HA1, large HA polypeptide; HA2, small HA polypeptide.

shearwater/Australia/75; H6—A/shearwater/Australia/72; H7—A/turkey/Oregon/71, A/equine/Prague/1/56; H8—A/turkey/Ontario/6118/68; H9—A/turkey/Wisconsin/1/66; H10—A/duck/Manitoba/53; H11—A/duck/England/56, A/duck/Ukraine/1/60, A/tern/Australia/75, A/duck/Memphis/546/76, A/duck/New York/12/78; H12—A/duck/Alberta/60/76.

The sequence data were stored and analyzed in a PDP-11 computer using programs adapted from those of Staden (20) and Gibbs and McIntyre (21). To enhance the accuracy of the data, every experiment was repeated, often many times with different virus preparations, and each region of sequence was read from several gels run for different times. The sequence gels of closely related strains were compared and differences were checked. It is, however, not possible to be certain that there is no error in the >10,000 nucleotides sequenced.

## RESULTS

Fig. 1 shows the nucleotide and predicted amino acid sequences obtained from HA genes of 12 viruses representing the 12 subtypes.

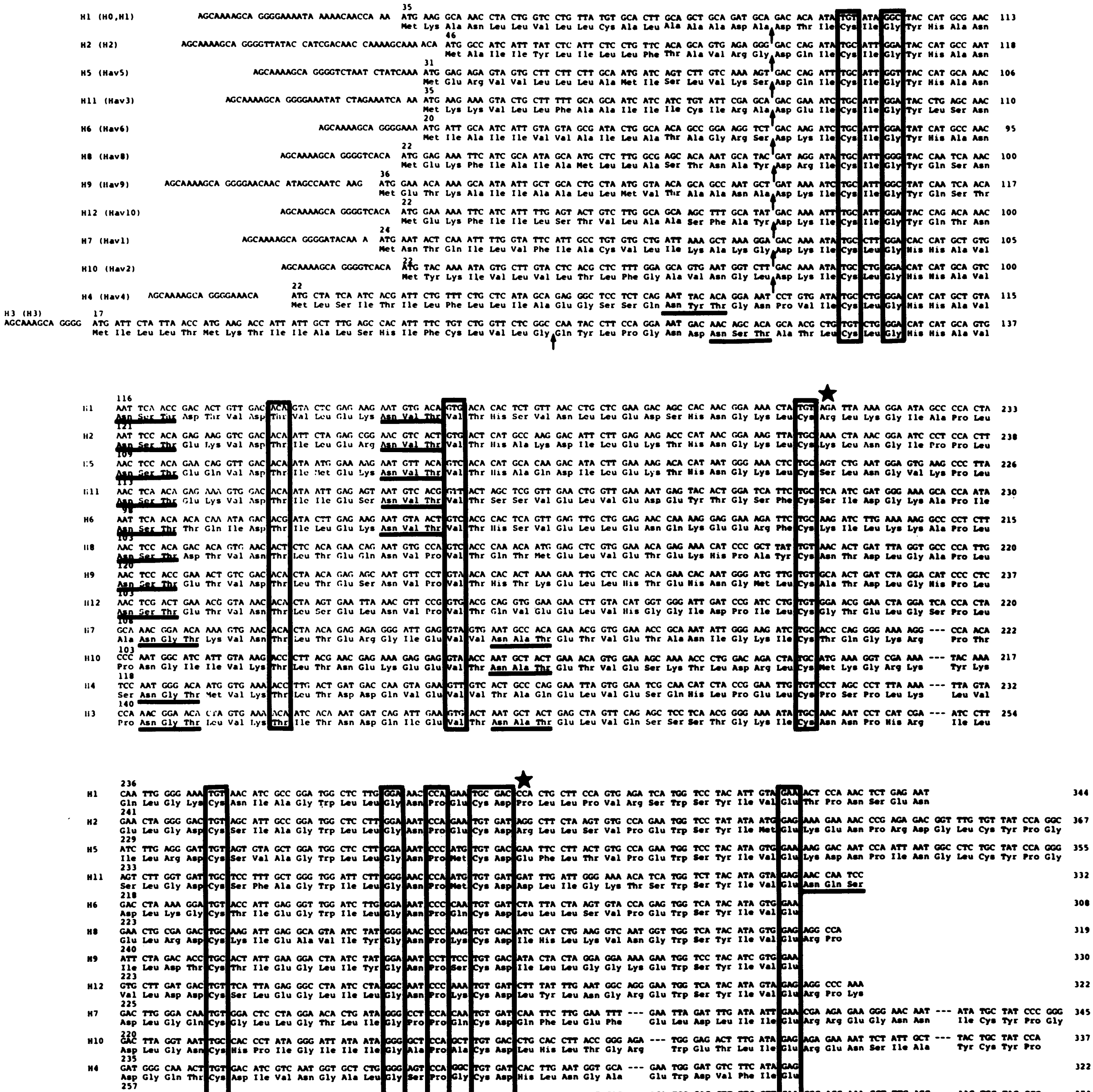

FIG. 1. Sequences of cDNA transcribed from the 3′ end of HA genes of viruses representing the 12 HA subtypes and predicted amino acid sequences of the precursor HA. The $NH_2$-terminal amino acid of HA1 (actual or presumed) is indicated by an arrow, and potential glycosylation sites are underlined. Boxes indicate amino acid residues conserved through all subtypes; ★, positions at which most variation is seen. The sequences are aligned at cysteine residues. Those shown are H1, A/PR/8/34; H2, A/RI/5⁻/57; H3, A/Memphis/1/71; H4, A/duck/Alberta/28/76; H5, A/shearwater/Australia/75; H6, A/shearwater/Australia/72; H7, A/turkey/Oregon/71; H8, A/turkey/Ontario/6118/68; H9, A/turkey/Wisconsin/1/66; H10, A/duck/Manitoba/53; H11, A/duck/Memphis/546/76; H12, A/duck/Alberta/60/76.

Table 1. Percentage amino acid sequence identity among the $NH_2$-terminal regions of HA1 of viruses representing each of the 12 HA subtypes

| | H1 | H2 | H5 | H11 | H6 | H8 | H9 | H12 | H7 | H10 | H4 | H3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| H1 | | 69 | 66 | 57 | 62 | 51 | 56 | 49 | 32 | 33 | 29 | 26 |
| H2 | | | 80 | 54 | 61 | 44 | 56 | 47 | 39 | 35 | 33 | 28 |
| H5 | | | | 56 | 61 | 46 | 53 | 49 | 38 | 36 | 33 | 32 |
| H11 | | | | | 61 | 51 | 54 | 49 | 33 | 30 | 29 | 29 |
| H6 | | | | | | 52 | 56 | 52 | 37 | 35 | 35 | 30 |
| H8 | | | | | | | 65 | 61 | 34 | 31 | 33 | 28 |
| H9 | | | | | | | | 67 | 35 | 32 | 35 | 26 |
| H12 | | | | | | | | | 33 | 34 | 33 | 30 |
| H7 | | | | | | | | | | 49 | 42 | 47 |
| H10 | | | | | | | | | | | 44 | 42 |
| H4 | | | | | | | | | | | | 47 |

The virus strains used are those shown in Fig. 1. Because the signal peptide sequences are so variable, comparisons have been made from the aspartic acid residue at the $NH_2$-terminus of H1, H2, and H7 (3, 25) and of H5, H11, H6, H8, H9, H10, and H12 (by homology) or from the corresponding amino acid where the $NH_2$-terminus is longer [H3 (4) and probably H4]. The sequences were compared pairwise by diagonal analysis (21) and percent identity was counted, deletions or insertions being counted as mismatches. A similar analysis of published complete sequences shows that the overall homology between H2 and H7 is 43%, that between H7 and H3 is 49%, and that between H2 and H3 is 42% (5–7).

**The Signal Peptides.** The amino acid sequences shown in Fig. 1 are those predicted if translation begins at the first ATG of the cDNA sequence (22). Only for the H2 subtype has this ATG been demonstrated to be the start of a signal peptide (23), which is cleaved off at or in the cell membrane during virus maturation. *In vivo*, the mRNA extends beyond the 3′ end of the viral strand RNA, having 10–15 nucleotides including the cap spliced on from nonviral mRNAs (14). The sequences of the extra 10–15 nucleotides vary (24) and it is not known what would happen if they contained an AUG sequence. Two inphase ATG sequences are present in several of the cDNA sequences shown in Fig. 1, including one near the start of the H3 (A/Mem/1/71) sequence. In other H3 sequences, only the second ATG is present (7–9), but it is possible that in A/Mem/1/71 the signal peptide is five amino acids longer.

The signal peptide sequences are almost wholly hydrophobic, but little homology is apparent in either nucleotide or amino acid sequences among the different subtypes. There are some amino acid similarities between H5 and H11 and between H8 and H12. The length is variable, as is the position of the first ATG in the cDNA, ranging from nucleotides 15–18 in A/Mem/1/71 (H3) to 44–46 in A/RI/5$^-$/57 (H2). There is no set pattern of charged residues before or after the central hydrophobic portion, which is 11–14 residues long.

**The $NH_2$ Terminus of HA1.** Viruses of subtypes H1, H2, and H7 have been shown to have aspartic acid at the $NH_2$ terminus of HA1 (3, 25) and, because all viruses of all subtypes except H3 and H4 have this amino acid at the same position, I have treated this as the $NH_2$ terminus of the mature HA1. The H3 $NH_2$ terminus is a cyclized glutamine in A/Mem/102/72 (4), and this is coded by all other H3 strains sequenced, giving an extra 10 amino acids in these strains. The only subtype in which the $NH_2$ terminus is entirely unknown is H4. Because attempts to directly identify the $NH_2$-terminal amino acid in this HA1 failed (W. G. Laver, personal communication), a candidate is the glutamine at amino acid 17.

It is clear from Fig. 1 that, although the cysteine residues can be aligned in all subtypes, very few other amino acids are totally conserved. At each of two positions, there are nine different amino acid residues in the 12 sequences, both positions being adjacent to conserved residues. One group of conserved amino acids, Gly-X-Pro-Y-Cys-Asp (Fig. 1), corresponds to a rather tight turn in the three-dimensional structure of Hong Kong (H3) HA derived by Wilson *et al.* (26), and the conservation of these residues supports their hypothesis that the basic structure of the HA molecule is the same in all subtypes.

Some subtypes are clearly more similar than others; the percentage amino acid identities is shown in Table 1. The relationships are shown as the dendrogram in Fig. 2. The hemagglutinins fall into two main families, the Hong Kong group (H3, H4, H7, and H10) being very different from the rest. The closest pair are H2 and H5 (80% amino acid sequence homology), but no antigenic crossreactivity has been detected between these. One of the three or four putative antigenic areas of the Hong Kong HA1 (12, 26) is present in the 30% of HA1 sequenced and, although what happens in the rest of the molecule does not affect the arguments presented here, it is possible that the partial sequences in Fig. 1 are indeed representative of the whole. Comparisons of complete sequences of H2, H7, and H3 (5–7) show the same relationships as found in the $NH_2$-terminal regions (Table 1 and Fig. 2).

**Sequences Within a Subtype: Genetic (Including Antigenic) Drift.** Nucleic acid and peptide analyses have shown that rather large differences in antigenic properties of human viruses within the H3 subtype occur with only about 20 amino acid substitutions, mostly in HA1 (9, 29). Analysis of the cDNA sequences complementary to the 3′ end of the HA genes of the H1 subtype of viruses isolated between 1933 and 1957 showed an accumulation of nucleotide changes but very few amino acid

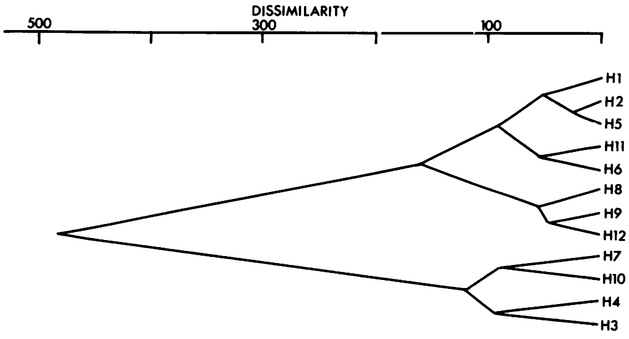

FIG. 2. Sequence relationships among the 12 HA subtypes. "Dissimilarity" was calculated from the amino acid identities shown in Table 1 by the method of Dayhoff (27), which allows for reversions, and then the dendrogram was calculated from these relationships by using the program of Lance and Williams (28) (Euclidean distance, Burr's strategy).

```
                                                                        35                                                                                                                122
Duck/Eng/56 (Viola)                 AGCAAAAGCA GGGGATCTAT CAAGAAGTCG AA ATG GAG AAA ATC CTG CTA TTT GCA GCT ATT TTC CTT TGT GTG AAA GCA GAT GAG ATC TGT ATC GGG TAT TTA AGC AAC AAC TCG ACA GAC
                                                                        Met Glu Lys Ile Leu Leu Phe Ala Ala Ile Phe Leu Cys Val Lys Ala Asp Glu Ile Cys Ile Gly Tyr Leu Ser Asn Asn Ser Thr Asp
                                                                                                                                        ↑
Duck/Ukraine/1/60 (Cumin)                                                           ACA                                                              ATT GGC
                                                                                    Thr

tern/Aust/75 (G55C)                                   A         A  A                CTC CTT             ACT ATC ATC     TGC         GCG                  ATT GGC                 AAT
                                                                                    Leu                 Thr     Ile

Duck/Mem/576/76 (Russ)                               AA    T    A  A        AAG     GTA     CTT         GCA ATC ATC ATC     ATT CGA     GAC GAA     TGC ATT GGA TAC CTG             TCA     GAG
                                                                            Lys     Val                         Ile Ile     Ile Arg                                                         Glu

Duck/NY/12/78 (Buzz)                                            A  A  G     AAG                     ACA GTT         TTA TAT GCA             GAA     TGC     GGT
                                                                            Lys                     Thr Val             Tyr Ala


                                                                                                                                                                                            242
Viola                           AAA GTT GAC ACA ATA ATT GAG AAC AAT GTC ACG GTC ACT AGC TCA GTG GAA CTG GTT GAG ACA GAA CAC ACT GGA TCA TTC TGT TCA ATC AAT GGA AAA CAA CCA ATC AGC CTT GGA GAT
                                Lys Val Asp Thr Ile Ile Glu Asn Asn Val Thr Val Thr Ser Ser Val Glu Leu Val Glu Thr Glu His Thr Gly Ser Phe Cys Ser Ile Asn Gly Lys Gln Pro Ile Ser Leu Gly Asp

Cumin                                                                                                                                                       GGG                             GAC

G55C                                                                    ACA                         TTG TTG                                                 GGG             ACA         AGA GAC
                                                                                                        Leu                                                                 Thr         Arg

Russ                                GTG                     AGT             GTT         TCG GTT             GAA AAT GAG TAC                 TGC         GAT GGG     GCA         AGT     GGT
                                                            Ser                                                 Asn     Tyr                             Asp         Ala

Buzz                                            ACG     ATC GAA AGC         ACA             TCG             GTG                                                 GGG AAG             AGT
                                                            Ser


                                                                                281
Viola                           TGT TCA TTT GCT GGA TGG ATA TTA GGC AAC CCT ATG TGT|
                                Cys Ser Phe Ala Gly Trp Ile Leu Gly Asn Pro Met Cys|
                                                                                                                                                    332
Cumin                                                           GGA AAT             GAT GAC CTA ATT GGA AAG AAT TCA TGG TCT TAC ATA GTG GAA AAC CAA TCT
                                                                                    Asp Asp Leu Ile Gly Lys Asn Ser Trp Ser Tyr Ile Val Glu Asn Gln Ser

G55C                            TGC TCC                             AAT CCC CAA    |
                                                                            Gln    |

Russ                            TGC TCC         GGG     ATT CTT GGG     CCA             GAT TTG     GGG AAA ACA                     GTA GAG
                                                                                                            Thr

Buzz                                TCC                         GGA     CCC             GAT     ATA GGG     ACT         TCA     ATT     GAG
                                                                                                            Thr
```

FIG. 3. Genetic drift in the HA gene of the H11 (Hav3) subtype of influenza virus. The nucleotide sequence and predicted amino acid sequence are given for cDNA transcribed from the 3′ end of RNA segment 4 of A/duck/England/56. For the other strains, where there is a nucleotide change, the entire codon is shown; if this also involves an amino acid change, the new amino acid is shown. Blank areas therefore indicate regions in which nucleotide and amino acid sequences are identical. Vertical lines indicate the end of data for the strain, and the arrow shows the start of mature HA1.

changes in HA1. Two "swine" strains of this subtype—A/swine/Wisconsin/15/30 and A/New Jersey/11/76—are also highly conserved (30).

Similar sequence analyses of the HA gene of H2 (Asian) strains, isolated from humans in 1957 and at the end of the H2 era (1968), and of three Asian-type strains isolated later from birds (1972 and 1977) showed the same pattern (31). There were many nucleotide changes, but only 6 amino acids changed in the $NH_2$-terminal 77 amino acids of HA1 through this time. Peptides from HA1 of the late human H2 strains—A/Netherlands/68 and A/Berkeley/68—have been analyzed, and show a similar proportion of changes through the rest of the HA1 polypeptide (unpublished results). The same pattern is seen in subtypes occurring only in birds, such as in H11, shown in Fig. 3. Because these are sporadic isolates from different bird species in different parts of the world, they cannot be considered as an evolutionary series, but the earlier strains are more similar to each other than to later strains and vice versa. There is again high conservation of the amino acid sequence, although there are many nucleotide changes.

Although the numbers of nucleotide changes within a subtype increase with time (30, 31), the vast majority of these are "silent" and do not alter the protein sequence except in the apparently less constrained signal peptides.

## DISCUSSION

In the matrix and nonstructural protein genes from human influenza viruses isolated during 1933–1977, we have observed a gradual accumulation of nucleotide (and some amino acid) changes, the rate of change being much the same from 1934 to 1957 (H1) as from 1957 to 1977 (H2 and H3) (32). Genetic drift also appears to occur at a constant rate in the various subtypes of HA. Of particular note is that the H2 (Asian) HA, which underwent antigenic drift in humans from 1957 to 1968, continued to drift in avian hosts at least up to 1977 (31, 33). The rate of amino acid change within a subtype is not significantly greater than that observed in the neuraminidase gene (19) or in the nonselected genes coding for the matrix and nonstructural proteins, which have also changed with time (32). As new variants of HA appear, the older viruses disappear, but isolations of an old virus are sometimes reported [e.g., Hong Kong/68 from South Australia in 1979 (34)]. In the best documented and most widespread case of reappearance, a 1950-type of H1N1 virus was isolated in 1977 (35). This virus was a relatively late member of the H1 era in humans and, although the subtype disappeared from humans in 1957, the 1977 virus caused major epidemics following its reappearance and has since drifted significantly (36). The close relationship of the A/USSR/90/77 strain to A/FW/1/50 (2 nucleotide differences in 335—adenosine at position 31 and thymidine at 182; see ref. 30) indicates that this virus was not replicating during its disappearance.

Fig. 1 shows that the nucleotide sequence changes between even the closest subtypes are much greater than any changes within subtypes (refs. 30 and 31; Fig. 3). At the amino acid sequence level, the relationships shown in Fig. 2 are clearly seen, but there is still a marked gap between the most dissimilar strains within a subtype (A/duck/England/56 vs. A/duck/Memphis/546/76, H11; Fig. 3) and the most similar pair of subtypes (H2 and H5). In most subtypes in which several strains have been examined [H1 (30), H2 (31), and H3 (9)], there are so few amino acid differences that it is impossible to tell whether the subtype is drifting toward any other. In H11 (Fig. 3), which shows rather more variation, the other viruses of the subtype are no more similar to any other subtype than the example used in Fig. 2.

Three conclusions can be drawn from the sequence data.

(*i*) Drift within HA subtypes has been continuous at an overall rate of 5% nucleotide change per 20 years, with very few observed reversions of nucleotides for at least 27 years (human H1 strains), 45 years (swine H1), 21 years (H2), and 22 years (H11). The H11 subtype shows drift, even though viruses of different bird species from different parts of the world (England, Ukraine, USA, and Australia) were compared (Fig. 3). The H2 subtype drifted in human hosts for 12 years and, although slightly diversified at the end of its human era (A/Tokyo/3/67, A/Berkeley/68 and A/Netherlands/68), a remarkably similar virus appeared in bird species and continued to drift for at least 9 years more (31). Some host specificity is seen. The H1 subtype in humans drifted between 1933 and 1956. The H1 swine virus, a probable cause of the 1918 human epidemic, is very similar in HA sequence to the human H1 viruses (30), but there are characteristic differences suggesting that it evolved independently of the human strains between 1930 and 1976. However, in the H3 subtype, it is not clear from sequence data whether a 1978 avian virus (A/black duck/Australia/702/78) is derived from duck/Ukraine/1/63 or from later human H3 strains (31).

(*ii*) The rate of sequence change in the HA gene is not significantly greater than that in the antigenically unselected genes.

(*iii*) Drift in any HA subtype sequence has not been observed to tend toward any other subtype.

There is a clear distinction between the extent of amino acid variation within subtypes (0–9%) and that between subtypes (20–74%). Although it is obvious that, if drift continued over a long period of time, the sequence variation would easily exceed that between subtypes, but there is currently no evidence that drift can proceed indefinitely; no viruses have been found that are intermediate between subtypes.

The different subtypes of HA are not confined to different species; all subtypes have been grown in embryonated chicken eggs, indicating that HA subtypes are not species specific, although transmission to other species may be relatively rare. Drift occurs in human, bird, and animal hosts, in HA and other genes. As drift proceeds, the earlier strains disappear, except for scattered but perhaps significant reappearances (34). The data presented here give no hint of the mechanisms underlying two extremes of influenza virus evolution—i.e., how a 1950 H1N1 virus reappeared almost unchanged in 1977 or how 12 subtypes that have little recognizable nucleotide sequence homology but various degrees of amino acid sequence homology have evolved.

I thank Dr. Adrian Gibbs for the computer analysis to generate Table 1 and Fig. 2 and Anne Mackenzie, Sally Campbell, and Kathy Fry for expert technical assistance. This work was supported in part by Grant AI 15343 from the National Institutes of Health, and by the Australian Government Poultry Industry Trust Fund.

1. Laver, W. G. & Webster, R. G. (1979) *Br. Med. Bull.* **35**, 29–34.
2. World Health Organization (1980) *Memo. Bull. W. H. O.* **58**, 585–591.
3. Waterfield, M. D., Espelie, K., Elder, K. & Skehel, J. J. (1979) *Br. Med. Bull.* **35**, 57–64.
4. Ward, C. W. & Dopheide, T. A. (1980) in *Structure and Variation in Influenza Virus*, eds. Laver, W. G. & Air, G. M. (Elsevier, Amsterdam), pp. 27–38.
5. Porter, A. G., Barber, C., Carey, N. H., Hallewell, R. A., Threlfall, G. & Emtage, J. S. (1979) *Nature (London)* **282**, 471–477.
6. Gething, M. J., Bye, J., Skehel, J. & Waterfield, M. (1980) *Nature (London)* **287**, 301–306.
7. Sleigh, M. J., Both, G. W., Brownlee, G. G., Bender, V. J. & Moss, B. A. (1980) in *Structure and Variation in Influenza Virus*, eds. Laver, W. G. & Air, G. M. (Elsevier, Amsterdam), pp. 69–80.
8. Min Jou, W., Verhoeyan, M., Devos, R., Saman, E., Fang, R., Huylebroeck, D., Fiers, W., Threlfall, G., Barber, C., Carey, N. & Emtage, S. (1980) *Cell* **19**, 683–696.
9. Verhoeyan, M., Fang, R., Min Jou, W., Devos, R., Huylebroeck, D., Saman, E. & Fiers, W. (1980) *Nature (London)* **286**, 771–776.
10. Laver, W. G., Air, G. M., Webster, R. G., Gerhard, W., Ward, C. W. & Dopheide, T. A. (1979) *Virology* **98**, 226–237.
11. Yewdell, J., Webster, R. G. & Gerhard, W. (1979) *Nature (London)* **279**, 246–248.
12. Webster, R. G. & Laver, W. G. (1980) *Virology* **104**, 139–148.
13. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
14. Plotch, S. J., Bouloy, M. & Krug, R. M. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 1618–1622.
15. Laver, W. G. (1969) in *Fundamental Techniques in Virology*, eds. Habel, K. & Salzman, N. P. (Academic, New York), pp. 82–86.
16. Skehel, J. J. & Hay, A. J. (1978) *Nucleic Acids Res.* **5**, 1207–1219.
17. Sanger, F. & Coulson, A. R. (1978) *FEBS Lett.* **87**, 107–110.
18. Air, G. M. (1979) *Virology* **97**, 468–472.
19. Blok, J. & Air, G. M. (1980) *Virology* **107**, 50–60.
20. Staden, R. (1979) *Nucleic Acids Res.* **6**, 2601–2610.
21. Gibbs, A. J. & McIntyre, G. A. (1970) *Eur. J. Biochem.* **16**, 1–11.
22. Kozak, M. (1978) *Cell* **15**, 1109–1123.
23. Elder, K. T., Bye, J. M., Skehel, J. J., Waterfield, M. D. & Smith, A. E. (1979) *Virology* **95**, 343–350.
24. Lai, C. J., Markoff, L. J., Sveda, M., Dhar, R. & Chanock, R. M. (1980) in *Structure and Variation in Influenza Virus*, eds. Laver, W. G. & Air, G. M. (Elsevier, Amsterdam), pp. 115–123.
25. Bucher, D. J., Li, S. S. L., Kehoe, J. M. & Kilbourne, E. D. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 238–242.
26. Wilson, I. A., Skehel, J. J. & Wiley, D. C. (1981) *Nature (London)* **289**, 366–372.
27. Dayhoff, M. O. (1972) *Atlas of Protein Sequence and Structure*, (Natl. Biomed. Res. Found., Silver Spring, MD), Vol. 5.
28. Lance, G. N. & Williams, W. T. (1967) *Aust. Comput. J.* **1**, 15–20.
29. Laver, W. G., Air, G. M., Dopheide, T. A. & Ward, C. W. (1980) *Nature (London)* **283**, 454–457.
30. Air, G. M., Blok, J. & Hall, R. M. (1981) in *Replication of Negative Strand Viruses*, eds. Bishop, D. H. L. & Compans, R. W. (Elsevier, Amsterdam), 225–239.
31. Air, G. M. & Hall, R. M. (1981) in *ICN-UCLA Symposia on Molecular and Cellular Biology*, eds. Nayak, D. & Fox, C. F. (Academic, NY), Vol. 22, in press.
32. Hall, R. M. & Air, G. M. (1981) *J. Virol.* **38**, 1–7.
33. Hinshaw, V. S., Webster, R. G., Bean, W. J. & Sriram, G. (1980) *Comp. Immun. Microbiol. Infect. Dis.* **3**, 155–164.
34. Moore, B. W., Webster, R. G., Bean, W. J., van Wyke, K. L., Laver, W. G., Evered, M. G. & Downie, J. C. (1981) *Virology* **109**, 219–222.
35. Nakajima, D., Desselberger, U. & Palese, P. (1978) *Nature (London)* **274**, 334–339.
36. Young, J. F., Desselberger, W. & Palese, P. (1979) *Cell* **18**, 73–83.