JOURNAL OF VIROLOGY, Sept. 1981, p. 663-672
0022-538X/81/090663-10$02.00/0

Vol. 39, No. 3

Conservation and Variation in the Hemagglutinins of Hong Kong Subtype Influenza Viruses During Antigenic Drift

GERALD W. BOTH* AND MERILYN J. SLEIGH

Commonwealth Scientific and Industrial Research Organization Molecular and Cellular Biology Unit, North Ryde, NSW, 2113, Australia

Received 10 March 1981/Accepted 18 May 1981

The nucleotide sequence was determined for the hemagglutinin gene of the Hong Kong subtype influenza strain A/Bangkok/1/79. The amino acid sequence predicted from these data shows a total of 36 amino acid changes as compared with hemagglutinin for a 1968 Hong Kong strain, 11 more than had occurred in a 1975 strain. The distribution of these changes confirmed that there are conserved and highly variable regions in hemagglutinin as the viral gene evolves during antigenic drift in the Hong Kong subtype. Of the four variable regions found in this study, only two have been seen previously. Correlation of highly variable areas in the hemagglutinins of Hong Kong subtype field strains with sites of amino acid changes in antigenically distinct influenza variants enabled us to predict likely antigenic regions of the protein. The results support and extend similar predictions made recently, based on the three-dimensional arrangement of hemagglutinin from a 1968 influenza strain.

Attempts to reduce outbreaks of influenza by vaccination have been frustrated because the virus undergoes continual alterations in its antigenic character. These antigenic changes result from changes in the primary structure of the two viral surface glycoproteins, hemagglutinin (HA) and neuraminidase (32). Antigenic shift, associated with the appearance of a new viral subtype, occurs when the virus acquires a new HA gene (and neuraminidase gene, in some cases) coding for a protein with entirely new antigenic characteristics. Within a viral subtype, the virus evolves under selective pressure supplied by host immunity. Strains able to grow and survive are those which have accumulated suitable mutations in the gene coding for HA, the most antigenically important of the new surface proteins. The resulting amino acid changes in HA are associated with progressive, small changes in viral antigenicity (antigenic drift) (28, 29, 34).

The Hong Kong subtype (with viral strains having type 3 HA and type 2 neuraminidase, i.e., H3N2) has been in circulation since 1968. In one study designed to analyze the progress of antigenic drift in this subtype, partial amino acid sequences for HA from different strains were compared (14). This study indicated that some regions in HA were more variable than others, with changes in the variable areas often accumulating in adjacent amino acids as the subtype evolved. However, this study provided information for only part of the HA protein sequence. With the advent of gene cloning and rapid DNA sequencing techniques, the analysis of antigenic drift has shifted to comparisons of HA gene nucleotide sequences for different strains (4, 10, 19, 22, 28, 29, 34). The most recent member of the Hong Kong subtype whose HA gene has been completely analyzed is A/Vic/3/75 (19).

We now report the HA gene sequence for the influenza strain A/Bangkok/1/79 (BK79). Assuming that between 1975 and 1979 the virus continued to accumulate HA amino acid changes at a rate similar to that seen earlier in subtype development (5, 29, 34), a comparison of the BK79 HA sequence with sequences from 1968 strains should show more clearly than does the comparison of the 1975 and 1968 strains (34) the distribution of constant and highly variable regions in the protein. In view of the selective pressure under which the virus is evolving, it might be expected that the areas of the protein in which most variation is seen would also be the most important antigenically. By correlating the variable areas with sites where amino acid changes are seen in antigenically distinct influenza variants isolated in laboratories (9, 16, 36), we were able to predict which regions of the protein are most likely to be involved in antibody binding. Similar predictions have been made by others, based on knowledge of the three-dimensional structure of HA of a 1968 Hong Kong strain (37, 38).

MATERIALS AND METHODS

Preparation of influenza viral RNA. The influenza strain BK79 (kindly provided by R. Webster, St. Jude's Hospital, Memphis, Tenn.) was grown in em-

bryonated chicken eggs (9). The virus was purified, and its RNA was extracted as described previously (26).

Preparation of DNA primers for copying of the HA gene RNA. Cloned DNA copies of the HA genes from influenza strains A/Mem/102/72 (Mem72) and A/NT/60/68/29C (29C) (4, 27, 28) were digested with restriction endonucleases (New England Biolabs). Reaction products were separated by polyacrylamide gel electrophoresis, and appropriate DNA fragments were excised and eluted from the gel as described previously (4, 5).

Determination of the HA gene nucleotide sequence. The influenza HA gene nucleotide sequence was determined by procedures described previously (3, 5). Total influenza genome RNA and a suitable DNA primer fragment were mixed, heat denatured (95°C, 1 min), and chilled on ice. The plus strand of the restriction fragment annealed during the incubation to a complementary region of the negative-stranded HA gene RNA segment and was then able to prime copying of the RNA into cDNA by reverse transcriptase. Dideoxynucleoside triphosphates were added to the reaction to generate partial cDNA copies (24). These copies were separated on a polyacrylamide gel so that the sequence of the cDNA could be determined (3, 5, 24). The cDNA copy has the same sense as does the mRNA of the HA gene; therefore, the corresponding HA amino acid sequence can be readily deduced. Sequence data were stored and analyzed with published computer programs (30, 31), and others were devised by A. Reisner and C. Bucholtz of this unit.

RESULTS AND DISCUSSION

Determination of the HA gene sequence for the Hong Kong influenza strain BK79. Most determinations of influenza HA gene nucleotide sequences have been made from cloned DNA copies of the appropriate segment of the negative-stranded RNA genome of the virus (4, 10, 19, 22, 28, 34). Recently, we abandoned this approach because it had become apparent that gene sequences determined from single cloned copies often differed from those of the predominant species present in the viral population (4, 29).

Our present approach (Fig. 1) makes use of the dideoxy chain termination method (24) to determine the sequence of single-stranded DNA copied from influenza genome RNA by reverse transcriptase. Primers for the reaction are provided by restriction endonuclease cleavage of cloned HA gene copies (4, 28). We have previously described the use of this technique to determine the nucleotide sequence of HA genes from several different influenza strains (4, 29). An example of the results obtained is shown in Fig. 2, which compares the sequences for BK79 and Mem72 through a region of the gene containing many differences between the two strains.

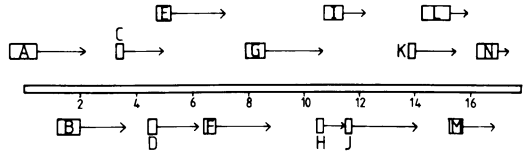


FIG. 1. *Strategy for obtaining the nucleotide sequence of the cDNA copies from BK79 genome RNA by using restriction fragments as primers. Arrows indicate the amount of sequence data obtained from each experiment. The numbers along the gene represent bases × 10^{-2}. The primers used were: (A) AluI-MboII (priming at base 44), (B) BstNI-AluI (200), (C) AvaII-HindIII (354), (D) DdeI-DdeI (469), (E) DdeI-AccI (524), (F) AvaII-HinfI (679), (G) HpaII-HhaI (859), (H) RsaI-HaeIII (1,066), (I) HaeIII-RsaI (1,128), (J) DdeI-BglI (1,172), (K) HinfI-HinfI (1,398), (L) AluI-HinfI (1,512), (M) HinfI-HpaII (1,571), and (N) BamHI-HaeIII (1,689). The MboII cut at base 44 is the first restriction site from the end of the gene and the first at which cDNA synthesis can begin. Therefore, the method does not permit the elucidation of part of the gene sequence coding for the signal peptide which extends between bases 30 and 77 (7). Mature HA1 begins at base 78, and mature HA2 begins at base 1,065.*

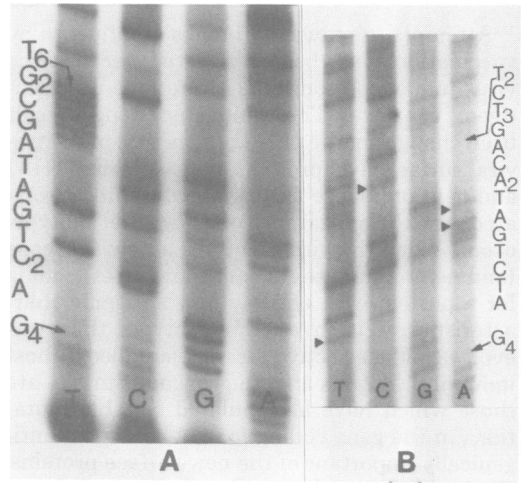


FIG. 2. *Base sequences of DNA copies of the HA genes from (A) Mem72 and (B) BK79 determined as described in text by using primer D (Fig. 1). The sequences are compared between bases 499 and 519, where there are four base differences (►) between the strains. An example of an artifact occasionally encountered with this method (24) is shown in base 509, where a band is seen in all four channels of the gel. This sometimes occurs when T is followed by a purine in the cDNA. The presence of a T can usually be determined from multiple sequencing experiments and, in this comparison, was supported by sequence data from a cloned copy of the HA gene (28).*

Figure 3 shows the complete HA gene sequence determined as described above for BK79, compared with the sequence for A/NT/60/68

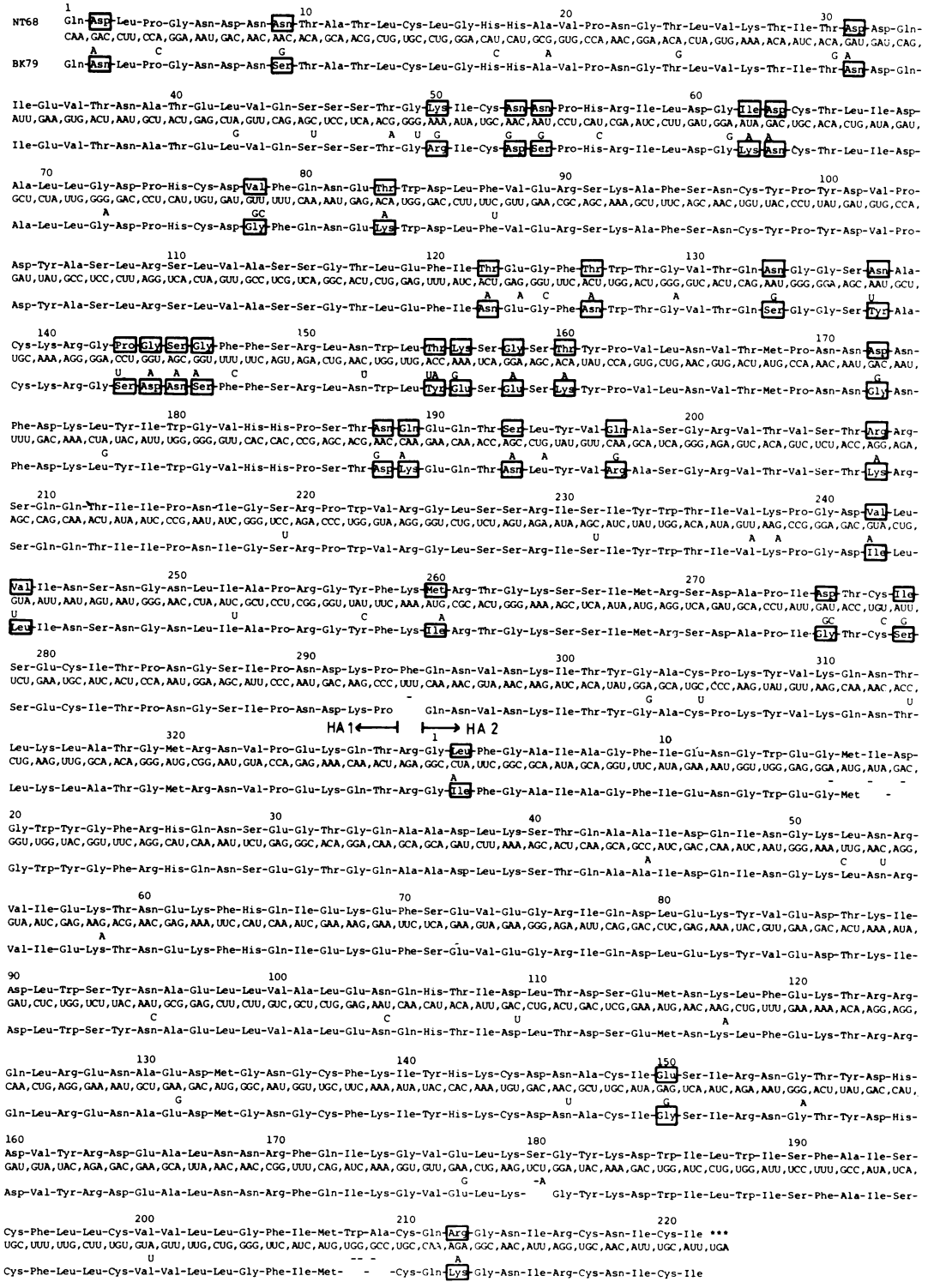


FIG. 3. *Comparison of cRNA and deduced amino acid sequences for that portion of the HA gene which codes for the mature HA protein from influenza strains NT68 and BK79. The NT68 sequence was derived as described in the text and in reference 29. Amino acid changes are boxed, and base changes in BK79 are indicated. Sequence data were stored and analyzed by using published programs (30, 31), and others were devised by A. Reisner and C. Bucholtz. We were unable to make a positive identification of the bases indicated by (–) in BK79 (see text).*

(NT68), an early isolate of the Hong Kong subtype (29). All base changes indicated were unambiguous. However, as shown in Fig. 3, in eight positions the base could not be unequivocally identified. This is a limitation of the technique and may be due to difficulties encountered by reverse transcriptase in copying regions of the template involved in the formation of secondary structure (4). Seven of the eight bases not positively identified for this strain fall in the region coding for HA2, the small subunit of the mature HA protein.

At three positions (no. 89, 383, and 923, all silent), the sequencing gels showed bands of approximately equal intensity in two channels. These mixtures always included the base present in NT68, and we have regarded them as conserved between the two strains. These results may reflect heterogeneity in the population of viral RNA molecules (20, 29), but we cannot eliminate the possibility that they are caused by problems with the sequencing method.

Conservation of structural features in the HA protein. In comparisons of the primary sequences of HAs from different viral subtypes, it was apparent that the number and relative positions of cysteine residues (4, 10, 14, 22, 35) and, to some extent, the proline residues (4) were conserved. This observation implies that, for some parts of the molecule, the shape is not permitted to vary extensively. Within the Hong Kong subtype, the same conservation is observed. In all of the members of the subtype for which partial or complete HA gene sequences have been determined (4, 14, 19, 28, 29, 34), the number and positions of the cysteine residues are constant. In addition, the proline residues are highly conserved. Only 1 of the 19 proline residues present in NT68 has disappeared in BK79, and no new proline codons are produced by mutation. Proline residues are also conserved in all other members of the subtype for which partial or complete HA sequences have been determined (14, 28, 29, 34; unpublished data). The importance of these residues in preserving the HA three-dimensional structure has been discussed previously (38).

Of the seven carbohydrate attachment sites in HA of NT68 and Aichi68 (35) characterized by the sequence Asn-X-Thr and Asn-X-Ser (21), six are preserved in all members of the subtype for which information is available. The site spanning resides 81 through 83 is lost in Eng69 (29), Vic375 (19), and BK79. New potential sites at residues 63 through 65 and 126 through 128 appear in several of the later strains of the subtype, but whether these are glycosylated is not known.

These results suggest that structural features impose constraints on changes which may occur in the HA protein during antigenic drift. In addition, highly hydrophobic areas within the protein in the N-terminal and C-terminal regions of HA2 retain their character within the subtype as well.

Conservation of HA2 in Hong Kong subtype viruses. The mature influenza HA protein consists of two disulfide-linked peptide chains, HA1 and HA2, produced from the primary HA translation product by proteolytic cleavage (18). There is considerable homology between HA2s from different viral subtypes (4, 10, 38). This conservation is even more striking within the Hong Kong subtype, with probably only three amino acid differences between HA2s from NT68 and BK79 (Fig. 3). This rate of change is much lower than that seen for the protein as a whole. However, in contrast to the conclusion drawn from a comparison of 1968 and 1975 strains (34), we found that the frequency of silent base mutation in the HA2 region of the gene is not significantly different from that in HA1 at the 5% level in a χ^2 analysis. It is now evident that the role of HA2 is a structural one, attaching HA to the viral membrane (25) and supporting the globular region of HA1 at the top of the HA spike (38). Conservation of amino acids within HA2 could reflect the necessity to maintain the extensive α-helical sections providing structural rigidity for the protein (38).

Conserved areas within HA1 of Hong Kong strains. HA is arranged on the viral surface in the form of closely packed trimeric spikes (11, 17). Regions at or near the tip of the spike are the most accessible to the exterior and contain the antigenically active region of HA (39) and presumably also areas involved in the protein functions of attaching virus particles to cells during infection (12) and virus penetration (23). Both the functional and the antigenic regions of HA are provided by the HA1 peptide (6, 7, 13). This finding is consistent with the proposed arrangement of HA on the virus surface whereby only residues from HA1 are found in the globular region of the protein near the top of the HA spike (38). Therefore, within a viral subtype, HA1 should contain areas where amino acids have changed to produce antigenic variation and areas which have been conserved to maintain protein function.

Regions of both of these types are found within HA1 of Hong Kong subtype viruses. This is apparent from the comparisons of BK79 and NT68 shown in Fig. 3, but can be seen more clearly in Fig. 4, where a comparison of the amino acid sequences for the HA1 regions of several Hong Kong isolates is shown. The HA1 region of BK79 contains two large blocks of

conserved amino acids (covering residues 84 through 121 and 279 through 328). No amino acid changes are seen in these regions in other members of the subtype (Fig. 4 and reference 14), although a third region conserved between NT68 and BK79, from residues 208 through 241, contains changes in some other strains (Fig. 4).

Interestingly, the number of silent base changes in the region of the gene (bases 327 through 440) corresponding to conserved amino acids 84 through 121 is unusually low. BK79 and Mem72 (28) each contain only one nucleotide different from NT68 in this area (bases 338 and 429, respectively), and no changes have been observed in any of the other Hong Kong strains examined (4, 19, 29, 34). This can be compared with a frequency of change for the whole of HA1 of BK79 of one silent change every 34 bases. The C-terminal conserved region of HA1 contains three silent base changes in the corresponding region of the gene for BK79 (bases 912 through 1,061), and changes in this area are also seen in other strains. It is important to consider that regions of apparent conservation in the HA protein may actually occur because of constraints on variation in the nucleotide sequence of the gene to conserve RNA secondary structure or specific sequences needed, for example, for protein binding. Some or all of the conservation of residues 84 through 121 may occur for this reason.

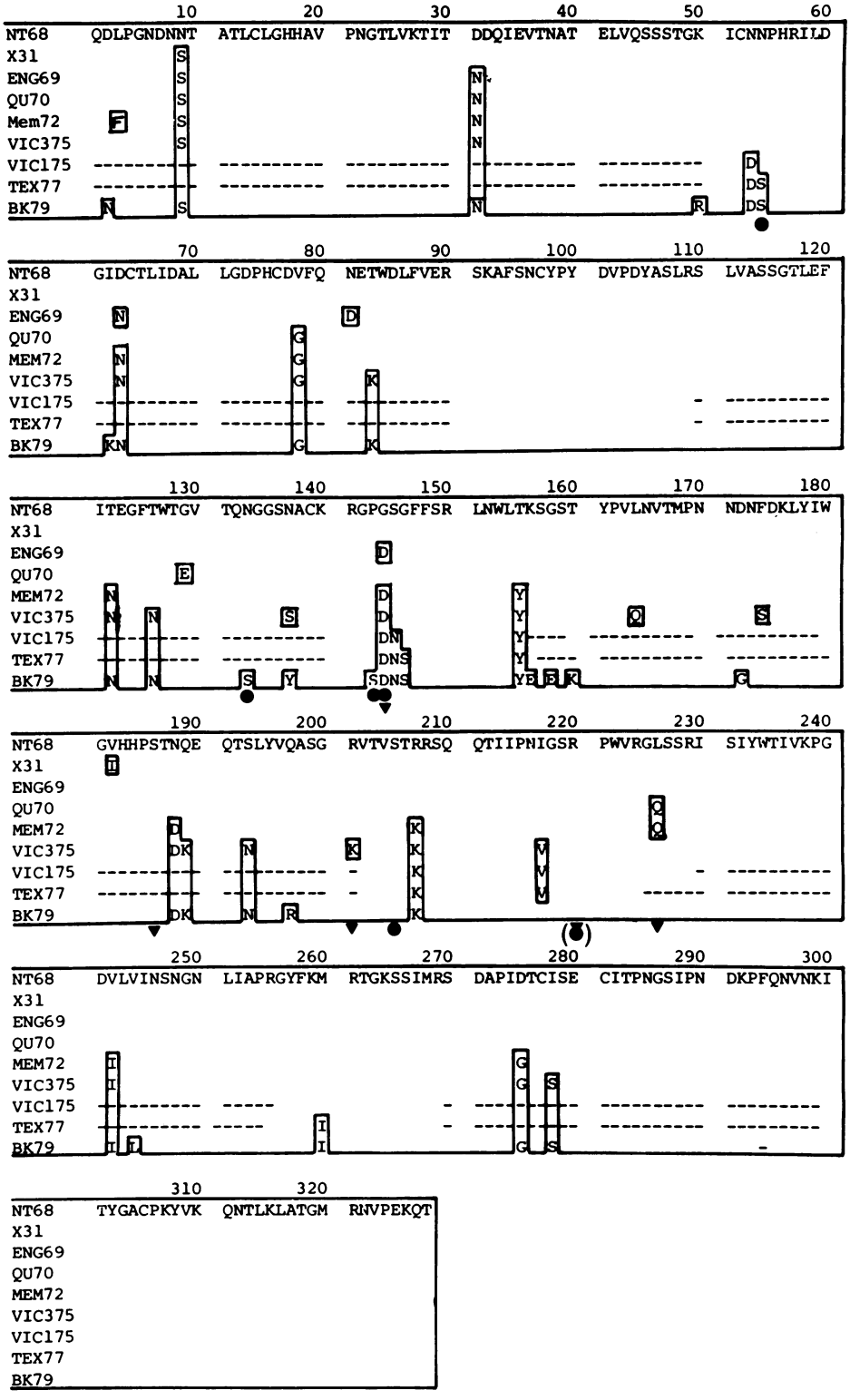
Consideration of the location of these conserved regions on the HA three-dimensional structure suggests that the reason for conservation of residues 279 through 329 of HA1 and of much of HA2 may be the same, since both areas of the protein are involved in the formation of the "stalk" connecting the globular region of HA1 at the top of the HA spike to the viral membrane (38). Residues 84 through 121 generally fall within the center of the globular region of HA1, whereas residues 208 through 241 mostly fall in the region of contact between the HA monomers making up the trimeric HA spike (38). It is possible that the conservation of such long blocks of amino acids is fortuitous; small stretches of amino acids conserved to preserve HA structure and function may be scattered on the HA primary structure, but brought together by protein folding.

Highly variable regions within HA1 of Hong Kong subtype viruses. A plot of the distribution of amino acid changes over the HA1 region of BK79 reveals not only conserved areas, but also some areas of high variability. This agrees with, but extends, earlier data (14, 15) which compared partial amino acid sequences for HA1 from several field strains of the Hong Kong subtype. A comparison with the complete HA1 sequences now available is shown in Fig. 4. As noted previously (14, 15), some of the amino acid changes occur in clusters; changes acquired in earlier strains of the subtype are retained and supplemented by changes in adjacent or nearby amino acids in later strains. This effect is particularly noticeable in variable areas covered by residues 50 through 63 and 122 through 160 (Fig. 4), but less so in the area between residues 188 and 197, which also contains several amino acid changes. Residue 217, which has changed from isoleucine to valine in Vic375, Vic175, and Tex77, has reverted to isoleucine in BK79. Other reversions have occurred at amino acids 3 and 201 (unpublished data). It is thought that these are due to independent mutations at the same nucleotide because the large number of amino acid and silent base changes common to Vic375 and BK79 suggest that they form part of a single evolutionary line (Both and Sleigh, manuscript in preparation).

Antigenic significance of the changes in HA1. The main selective pressure on influenza virus strains emerging during development of a subtype is that supplied by the antibodies of the host. Therefore, successive strains would be expected to acquire amino acid changes which alter viral antigenicity without unacceptably deleterious effects on growth potential. Immunological studies have suggested that successive members of a subtype are able to overcome neutralization by antibodies against earlier members of that subtype (33); i.e., antigenically favorable changes are retained and supplemented in later strains. Where successive amino acids of the protein chain form part of a single antigenic site (1), progressive evolution at this site would be expected to produce clustering of amino acid changes as described above. If, on the other hand, amino acids contributing to an antigenic site are not arranged continuously on the HA1 polypeptide, but come from different regions brought together by protein folding (2), then amino acid changes in different parts of the molecule could affect antibody binding to a common site. A third kind of antigenically important amino acid change would be one affecting the conformation of an antigenic site, although not located within it.

Antigenically important amino acid changes in HA1 have been identified for some early strains of the Hong Kong subtype by measuring the effects of these changes on the binding of monoclonal antibodies (20, 29). However, such an approach is useful only in comparing strains with few amino acid differences.

An alternative approach to the identification of antigenically important amino acid changes is the selection of influenza variants able to grow


10 20 30 40 50 60
NT68 QDLPGNDNNT ATLCLGHHAV PNGTLVKTIT DDQIEVTNAT ELVQSSSTGK ICNNPHRILD
X31
ENG69
QU70
Mem72
VIC375
VIC175
TEX77
BK79
70 80 90 100 110 120
NT68 GIDCTLIDAL LGDPHCDVFQ NETWDLFVER SKAFSNCYPY DVPDYASLRS LVASSGTLEF
X31
ENG69
QU70
MEM72
VIC375
VIC175
TEX77
BK79
130 140 150 160 170 180
NT68 ITEGFTWTGV TQNGGSNACK RGPGSGFFSR LNWLTKSGST YPVLNVTMPN NDNFDKLYIW
X31
ENG69
QU70
MEM72
VIC375
VIC175
TEX77
BK79
190 200 210 220 230 240
NT68 GVHHPSTNQE QTSLYVQASG RVTVSTRRSQ QTIIPNIGSR PWVRGLSSRI SIYWTIVKPG
X31
ENG69
QU70
MEM72
VIC375
VIC175
TEX77
BK79
250 260 270 280 290 300
NT68 DVLVINSNGN LIAPRGYFKM RTGKSSIMRS DAPIDTCISE CITPNGSIPN DKPFQNVNKI
X31
ENG69
QU70
MEM72
VIC375
VIC175
TEX77
BK79
310 320
NT68 TYGACPKYVK QNTLKLATGM RNVPEKQT
X31
ENG69
QU70
MEM72
VIC375
VIC175
TEX77
BK79

in the presence of monoclonal antibodies (16, 36) or of a subfraction of whole antiserum (9). Most of these variants have only single amino acid changes whose effect on antigenicity can readily be determined (20, 25, 29). The locations of these amino acid changes are shown in Fig. 4. Such data have been used in conjunction with the location of residues on the HA three-dimensional structure to predict possible locations for antigenically important regions (37). We have found that it is possible to arrive at very similar predictions by using the same data in conjunction with the locations of highly variable regions in the HA of the influenza field strain BK79.

The most obvious cluster of altered amino acids in BK79 covers residues 143 through 146. Residue 144 in this region is changed both in monoclonal variants (36) and in several variants selected with whole antiserum (9, 20). The same change (Gly → Asp) at residue 144 is responsible for part of the altered antigenicity of Eng69 (29, 36) and appears in several of the more recent field isolates. The adjacent residue (Pro 143) was altered in several of the monoclonal variants (16, 36), and a change in this residue (to Ser) is now seen (in BK79) for the first time in a field strain. Included in the same antigenic region as the residue 143 through 146 cluster must be residue 133, since antibodies responding to the change at residue 143 were affected by a change at this residue (Asn → Lys) in one of the monoclonal variants (36). Residue 133 has changed from Asn to Ser in BK79. Nearby residue 137 (Asn) is changed to Ser and Tyr by independent adjacent mutations in later field strains (Fig. 3) and may also be included in this antigenic region.

Another monoclonal variant derived from Mem71 (16) shows a change at residue 54 (Asn → Lys), which falls within a second variable region in BK79 (residues 50 through 54). Three residues within this cluster are altered in BK79. A third cluster of altered amino acids in BK79 covers residues 155 through 160. No laboratory-isolated variants have been detected with changes in this region, and most of the changes seen in BK79 are not present in earlier field strains (Fig. 4). Therefore, this region has not previously been recognized as highly variable. Residue 160, altered from Thr to Lys in BK79, is Ala in PC73 (unpublished data). This is the second of two examples (see also residue 137) of mutations in adjoining bases producing separate changes at the same amino acid residue.

The remaining variable region in the HA1 of BK79 covers residues 186 through 189. Adjacent residues 188 and 189 are altered in some of the later field strains (Fig. 4), and one laboratory variant shows a Ser → Ile change at residue 186 (20). This observation suggests that this group of amino acids could also form a variable cluster of the type described in other regions.

Therefore, a consideration of our results for BK79 in conjunction with information available on changes in antigenically distinct influenza variants has suggested the existence of four clusters of variable amino acids in HA1 (50 through 54, 143 through 146, 186 through 189, and 155 through 160). At least the first three of these are likely to be antigenically significant. Fifty percent of the amino acid changes seen in BK79 fall within these clusters or occur in associated amino acids (residues 133 and 137). The remaining changes are scattered in HA1, but at least some may be antigenically significant since, in laboratory variants, changes outside the variable clusters are also seen at residues 205, 220, 201 (the latter is altered in Vic375 [Fig. 3]), and 226 (altered in Qu70 and Mem72 [Fig. 4]) (5, 16, 20, 28, 29).

Location of BK79 amino acid changes on HA three-dimensional structure. Now that the three-dimensional structure of the HA protein has been determined by X-ray crystallographic techniques (38), it is possible to locate the variable regions in BK79 on the HA spike. Figure 5 shows the three-dimensional structure of HA from the 1968 influenza strain X31 (38) with the amino acids altered in the BK79 strain (indicated by filled circles). The concentration of amino acid changes near the tip of the spike is striking. The variable areas 155 through 160 and 186 through 189 are adjacent and together

FIG. 4. *Comparison of amino acid sequences in the HA1 region of HAs from strains of the Hong Kong subtype. HA1 amino acid sequence data for strains NT68, Eng69, and Qu70 were obtained from HA gene sequences determined previously (29). The data for Mem72, the X31 recombinant of Aichi68 and Vic375, are from nucleotide sequences of cloned HA gene copies (9, 28, 34). Note that bases 110, 127, and 327 of Mem72 differ from the sequence reported previously (28). The extra Asn in Vic375 between amino acids 8 and 9 (19) has been omitted for clarity. The partial sequences for Vic75 and Tex77 were obtained by amino acid sequencing and comparative peptide mapping of HA proteins (14). Residues altered in later strains compared with the NT68 strain are boxed. Blank areas indicate conserved sequences. Sites where changes have been seen in variant strains selected in the presence of monoclonal antibodies (16, 36) or a fraction of whole antiserum (5, 9, 20) are indicated by ● and ▼, respectively. The bracketed circle was not precisely mapped and is deduced. The one-letter code for amino acids is: A, Ala; R, Arg; N, Asn; D, Asp; C, Cys; Q, Gln; E, Glu; G, Gly; H, His; I, Ile; L, Leu; K, Lys; M, Met; F, Phe; P, Pro; S, Ser; T, Thr; W, Trp; Y, Tyr; and V, Val.*

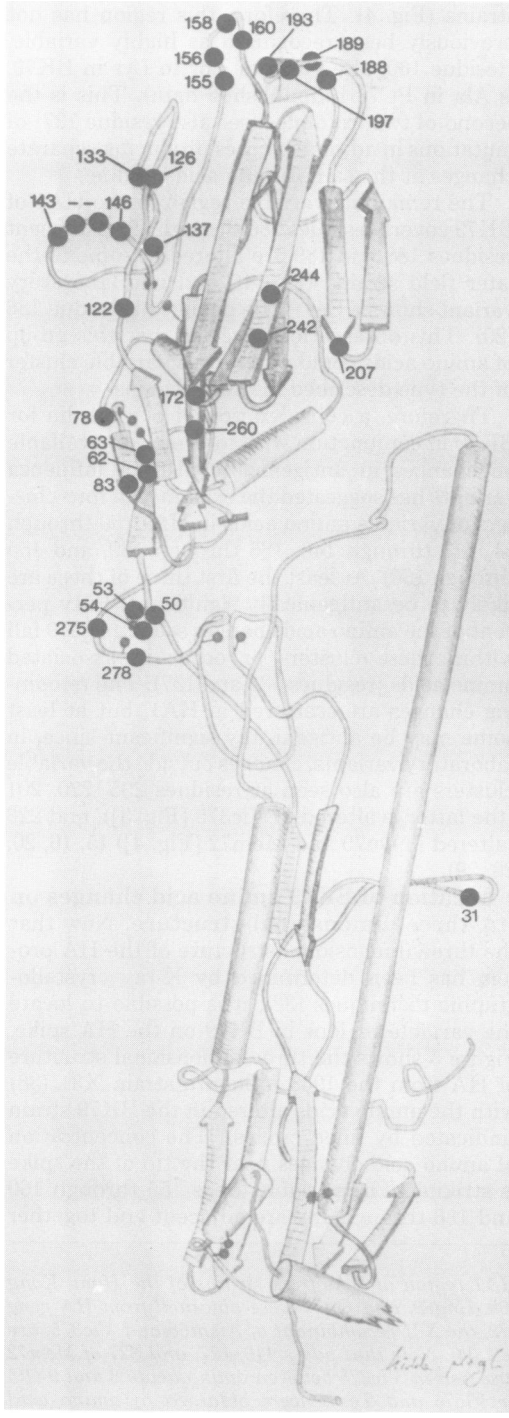


FIG. 5. *Schematic drawing of a monomer of HA from Aichi68 showing the locations of the amino acid changes in HA1 for BK79. The monomer was drawn by Hidde Pleogh, and the picture was supplied by Don Wiley (37, 38).*

probably form the proposed antigenic site B (37). Amino acid 129, a change which appeared to be antigenically important in Qu70 (29), may also contribute to this site. Residues 143 through 146 are located near the tip on the prominent loop projecting out from the molecule. Amino acids 133 and 137, both altered in BK79, are found in the same region and probably also form part of antigenic site A. Residues 122 and 126, also altered in BK79, are nearby, but whether they influence antibody binding at site A is not clear.

The last variable cluster (50 through 54) predicted from the BK79 sequence is found lower down on the spike (38) in association with residues 275 and 278. These are also altered in BK79, giving a total of five amino acid changes in the proposed antigenic site C (37). Of these, residue 50 was conserved in earlier members of the subtype and was proposed as an "anchor" residue, maintaining the structure of antigenic site C (38). The change at residue 50 (Arg → Lys) is a conservative one, whereas most of the changes proposed to be antigenically important are nonconservative, an observation consistent with the idea that residue 50 may play a role in maintaining structure rather than in antibody binding.

Therefore, each of the four variable regions found in BK79 seems likely, as predicted, to be of antigenic importance. From the location of amino acid changes on HA in BK79, it can be estimated that up to 65% of the amino acid changes are associated with these antigenically important sites. The remaining changes are scattered, but generally fall within the upper (globular) part of the HA molecule. Whether some form part of an antigenic site D is not yet certain (37). Influenza variants selected with the "avid" fraction of whole antiserum (9) contain amino acid changes which profoundly affect antigenicity (5, 20, 29). Three of these are located outside the proposed sites A, B, and C, at residues 201, 220, and 226, whereas variants selected with monoclonal antibodies have amino acid changes at position 205 and in a peptide comprising residues 217 through 224 (possibly Arg 220) (16). It has been suggested that some of these changes may be able to disturb the interface between the globular portions of HA1 at the top of the trimeric spike (37). Thus, they may disturb antibody binding in the interface region, or indirectly, possibly altering the conformation of more than one antigenic region. This would be consistent with the escape of the avid fraction variants from neutralization during selection by what was essentially a panel of monoclonal antibodies assumed to be binding to more than one antigenic determinant (9).

Conclusions. With information now available on the changes in influenza HA covering the period 1968 to 1979, an assessment of the extent of and limitations on change in viral HA can be made. From the HA gene sequence for BK79, together with information on earlier isolates, it is apparent that the HA2 region is almost completely conserved within the subtype. Conservation in this area among subtypes is also high (4, 10). Within HA1, features needed to maintain three-dimensional structure (cysteines, prolines, and glycosylation sites) are highly conserved. There are also two extensive regions of amino acid conservation in HA1 within the subtype, although in one of these, the number of silent base changes in the corresponding region of the gene is unusually low, suggesting that constraints on evolution may be imposed at the nucleic acid and the protein levels. The importance of these extensive conserved regions is not clear, since it seems likely that functional domains of the protein are constructed of short runs of amino acids brought into proximity by protein folding (37, 38).

In addition to conserved areas, the HA1 of BK79 contains three major variable regions (residues 50 through 54, 143 through 146, and 155 through 160). The first two of these contain residues also changed in laboratory-isolated antigenic variant strains, and one of the laboratory strains points to a fourth potential cluster of variable amino acids in field strains at residues 186 through 189. A consideration of the three-dimensional arrangement of HA suggests that all of these variable regions are likely to be important antigenically.

Of the HA1 amino acids within the regions thought to be involved in antibody binding (37), a large proportion has changed between 1968 and 1979. So far we have seen no examples of successive changes at a single amino acid, a mechanism once proposed to explain the progressive nature of antigenic drift (8). On the other hand, we have reported at least three examples of amino acids which have changed and then apparently reverted (residues 3, 201, and 217), suggesting that the mutation frequency is high enough to permit multiple changes at a single site.

Within the proposed antigenic sites, there are several amino acids not regarded as necessary to maintain structure (37) which have not altered between 1968 and 1979. These seem to provide some further scope for evolution in the Hong Kong subtype, although it is possible that changes that have already occurred in nearby residues may preclude changes at these sites (for functional reasons) or reduce their importance in antibody binding. Thus, if progressive changes at single amino acids do not occur, the possibilities for evolution in the Hong Kong subtype beyond 1979 may be limited.

ACKNOWLEDGMENTS

We thank J. Harrison and J. Rawlinson for growing and purifying virus; V. Bender for purifying RNA; R. Webster for providing BK79 virus stocks; A. Reisner and C. Bucholtz for adapting and writing computer programs for sequence data analysis; and E. Hamilton for excellent technical assistance. We are grateful to P. A. Underwood, R. Whittaker, and B. Moss for valuable discussions; G. Grigg, P. A. Underwood, V. Bender, and B. Moss for critical reading of the manuscript; D. Wiley for informative discussion and for providing Hidde Ploegh's drawing of the structure in Fig. 5.; and J. W. Beard for continuing supplies of reverse transcriptase.

ADDENDUM IN PROOF

Our conclusion that amino acid residue 201 (changed from Arg → Lys in Vic375) had reverted to Arg in BK79 was based on the unpublished nucleotide sequence data for A/Tex/1/77 of B. A. Moss and G. G. Brownlee. In this sequence, residue 201 was also Lys. However, the sequence has now been revised and shows this residue as Arg. We now conclude that the Arg → Lys change at residue 201 is unique to Vic375.

LITERATURE CITED

1. **Atassi, M. Z.** 1977. The complete structure of myoglobin: approaches and conclusions for antigenic structures of proteins, Article title, p . 77-176. *In* M. Z. Atassi (ed.), Immunochemistry of proteins, vol. 2. Plenum Publishing Corp., New York.
2. **Atassi, M. Z.** 1978. Precise determination of the entire antigenic structure of lysozyme. Immunochemistry **15:** 909-936.
3. **Both, G. W., and G. M. Air.** 1979. Nucleotide sequence coding for the N-terminal region of the matrix protein of influenza virus. Eur. J. Biochem. **96:**363-372.
4. **Both, G. W., and M. J. Sleigh.** 1980. Complete nucleotide sequence of the haemagglutinin gene from a human influenza virus of the Hong Kong subtype. Nucleic Acids Res. **8:**2561-2575.
5. **Both, G. W., M. J. Sleigh, V. J. Bender, and B. A. Moss.** 1980. A comparison of antigenic variation in Hong Kong influenza virus haemagglutinins at the nucleic acid level, p. 81-89. *In* W. G. Laver and G. M. Air (ed.), Structure and variation in influenza virus. Elsevier/North-Holland Publishing Co., New York.
6. **Brand, C. M., and J. J. Skehel.** 1972. Crystalline antigen from the influenza virus envelope. Nature (London) New Biol. **238:**145-147.
7. **Eckert, E. A.** 1973. Properties of an antigenic glycoprotein isolated from influenza virus hemagglutinin. J. Virol. **11:**183-192.
8. **Fazekas de St. Groth, S.** 1975. The phylogeny of influenza, p. 741-754. *In* B. W. J. Mahy and R. D. Barry (ed.), Negative strand viruses. Academic Press, Inc., London.
9. **Fazekas de St. Groth, S., and C. Hannoun.** 1973. Selection par pression immunologique de mutants, dominants du virus de la grippe A (Hong Kong). C.R. Acad. Sci. Ser. D **276:**1917-1920.
10. **Gething, M. J., J. Bye, J. Skehel, and M. Waterfield.** 1980. Cloning and DNA sequence of double stranded copies of haemagglutinin genes from H2 and H3 strains elucidates antigenic shift and drift in human influenza virus. Nature (London) **287:**301-306.
11. **Griffiths, I. P.** 1975. The fine structure of influenza virus, p. 121-132. *In* B. W. J. Mahy and R. D. Barry (ed.),

Negative strand viruses. Academic Press, Inc., London.
12. **Hirst, G. K.** 1942. The quantitative determination of influenza virus and antibodies by means of red cell agglutination. J. Exp. Med. **75**:49-64.
13. **Jackson, D. C., T. A. Dopheide, R. J. Russell, D. O. White, and C. W. Ward.** 1979. Antigenic determinants of influenza virus haemagglutinin. II. Antigenic reactivity of the isolated N-terminal cyanogen bromide peptide of A/Memphis/72 haemagglutinin heavy chain. Virology **93**:458-465.
14. **Laver, W. G., G. M. Air, T. A. Dopheide, and C. W. Ward.** 1980. Amino acid sequence changes in the haemagglutinin of A/Hong Kong (H3/N2) influenza virus during the period 1968-1977. Nature (London) **283**:454-457.
15. **Laver, W. G., G. M. Air, R. G. Webster, W. Gerhard, C. W. Ward, and T. A. Dopheide.** 1980. The antigenic sites on influenza virus haemagglutinin. Studies on their structure and variation, p. 295-306. *In* W. G. Laver and G. M. Air (ed.), Structure and variation in influenza virus. Elsevier/North-Holland Publishing Co., Amsterdam.
16. **Laver, W. G., G. M. Air, R. G. Webster, W. Gerhard, C. W. Ward, and T. A. Dopheide.** 1979. Antigenic drift in type A influenza virus. Sequence differences in the haemagglutinin of Hong Kong (H3N2) variants selected with monoclonal hybridoma antibodies. Virology **98**:226-237.
17. **Laver, W. G., and R. D. Valentine.** 1969. Morphology of the isolated haemagglutinin and neuraminidase subunits of influenza virus. Virology **38**:105-119.
18. **Lazarowitz, S. G., R. W. Compans, and P. W. Choppin.** 1971. Influenza virus structural and nonstructural proteins in infected cells and their plasma membranes. Virology **46**:830-843.
19. **Min Jou, W., M. Verhoeyen, R. Devos, E. Saman, R. Fang, D. Huylebroeck, and W. Fiers.** 1980. Complete structure of the haemagglutinin gene from the human influenza A/Victoria/3/75 (H3N2) strain as determined from cloned DNA. Cell **19**:683-696.
20. **Moss, B. A., P. A. Underwood, V. J. Bender, and R. G. Whittaker.** 1980. Antigenic drift in the haemagglutinin from various strains of influenza virus A/Hong Kong 68 (H3N2) p. 329-338. *In* W. G. Laver and G. M. Air (ed.), Structure and variation in influenza virus. Elsevier/North-Holland Publishing Co., New York.
21. **Neuberger, A., A. Gottschalk, R. D. Marshall, and R. G. Spiro.** 1972. Carbohydrate-peptide linkages in glycoproteins and methods for their elucidation, p. 450-490. *In* A. Gottschalk (ed.), Glycoproteins. Elsevier/North-Holland Publishing Co., New York.
22. **Porter, A. G., C. Barber, N. H. Carey, R. A. Hallewell, G. Threlfall, and J. S. Emtage.** 1979. Complete nucleotide sequence of an influenza virus haemagglutinin gene from cloned DNA. Nature (London) **282**:471-477.
23. **Rott, R., H. D. Klenk, and C. Scholtissek.** 1978. Activation of influenza virus infectivity by proteolytic cleavage of the haemagglutinin, p. 69-81. *In* W. G. Laver, H. Bachmayer, and C. Weil (ed.), The influenza virus haemagglutinin. Springer-Verlag New York, Inc., New York.
24. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74**:5463-5467.
25. **Skehel, J., and M. D. Waterfield.** 1975. Studies on the primary structure of the influenza virus haemagglutinin. Proc. Natl. Acad. Sci. U.S.A. **72**:93-97.
26. **Sleigh, M. J., G. W. Both, and G. G. Brownlee.** 1979. A new method for the size estimation of the RNA genome segments of influenza virus. Nucleic Acids Res. **6**:1309-1321.
27. **Sleigh, M. J., G. W. Both, and G. G. Brownlee.** 1979. The influenza virus haemagglutinin gene: cloning and characterisation of a double-stranded DNA copy. Nucleic Acids Res. **7**:879-893.
28. **Sleigh, M. J., G. W. Both, G. G. Brownlee, V. J. Bender, and B. A. Moss.** 1980. The haemagglutinin gene of influenza A virus: nucleotide sequence analysis of cloned DNA copies, p. 69-78. *In* W. G. Laver and G. M. Air (ed.), Structure and variation in influenza virus. Elsevier/North-Holland Publishing Co., New York.
29. **Sleigh, M. J., G. W. Both, P. A. Underwood, and V. J. Bender.** 1981. Antigenic drift in the hemagglutinin of the Hong Kong influenza subtype: correlation of amino acid changes with alterations in viral antigenicity. J. Virol. **37**:845-853.
30. **Staden, R.** 1977. Sequence data handling by computer. Nucleic Acids Res. **4**:4037-4051.
31. **Staden, R.** 1977. A strategy of DNA sequencing employing computer programs. Nucleic Acids Res. **6**:2601-2610.
32. **Stuart-Harris, C. H., and C. H. Schild.** 1976. Influenza: the viruses and the disease, p. 57-68. Edward Arnold, London.
33. **Underwood, P. A.** 1980. Serology and energetics of cross-reactions among the H3 antigens of influenza viruses. Infect. Immun. **27**:397-404.
34. **Verhoeyen, M., R. Fang, W. Min Jou, R. Devos, D. Huylebroeck, E. Saman, and W. Fiers.** 1980. Antigenic drift between the haemagglutinin of the Hong Kong influenza strains A/Aichi/2/68 and A/Victoria/3/75. Nature (London) **286**:771-776.
35. **Ward, C. W., and T. A. Dopheide.** 1980. The Hong Kong (H3) haemagglutinin: complete amino acid sequence and oligosaccharide distribution for the heavy chain of A/Memphis/102/72, p. 27-38. *In* W. G. Laver and G. M. Air (ed.), Structure and variation in influenza virus. Elsevier/North-Holland Publishing Co., New York.
36. **Webster, R. G., and W. G. Laver.** 1980. Determination of the number of nonoverlapping antigenic areas on Hong Kong (H3N2) influenza virus haemagglutinin with monoclonal antibodies and the selection of variants with potential epidemiological significance. Virology **104**:139-148.
37. **Wiley, D. C., I. A. Wilson, and J. J. Skehel.** 1981. Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation. Nature (London) **289**:373-378.
38. **Wilson, I. A., J. J. Skehel, and D. C. Wiley.** 1981. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3Å resolution. Nature (London) **289**:366-373.
39. **Wrigley, N. G., W. G. Laver, and J. C. Downie.** 1977. Binding of antibodies to isolated haemagglutinin and neuraminidase molecules of influenza virus observed in the electron microscope. J. Mol. Biol. **109**:405-421.