JOURNAL OF VIROLOGY, Oct. 1983, p. 52–60
0022-538X/83/100052-09$02.00/0

Vol. 48, No. 1

Antigenic Drift in Influenza Virus H3 Hemagglutinin from 1968 to 1980: Multiple Evolutionary Pathways and Sequential Amino Acid Changes at Key Antigenic Sites

GERALD W. BOTH,[1]* MERILYN J. SLEIGH,[1] NANCY J. COX,[2] AND ALAN P. KENDAL[2]

Commonwealth Scientific and Industrial Research Organization, Molecular and Cellular Biology Unit, North Ryde, New South Wales, 2113, Australia,[1] and World Health Organization Collaborating Center for Influenza, Division of Viral Diseases, Center for Infectious Diseases, Centers for Disease Control, Atlanta, Georgia 30333[2]

Received 21 March 1983/Accepted 29 June 1983

Surveys of the antigenic properties of a wide range of variants of the H3N2 (Hong Kong) influenza virus subtype have revealed complex patterns of variants cocirculating during each of the main epidemic eras of the subtype. We determined hemagglutinin (HA) gene sequences for 14 isolates chosen to give the wildest possible spread of variant types. The addition of these data to existing HA gene sequence information for other variants provides a comprehensive picture of HA gene evolution during antigenic drift among H3N2 subtype viruses. The data reveal the existence of multiple evolutionary pathways during at least one period of development of the subtype and strikingly demonstrate that amino acid changes are limited to a small number of locations on the HA molecule during antigenic drift. The occurrence of sequential amino acid changes at key positions within these variable regions suggests that the HA structure has remained constant during subtype evolution so that only limited possibilities remain for further antigenic drift among H3N2 viruses.

Attempts to control influenza by vaccination have so far been of limited success and are hindered by continual changes in the major surface antigen of influenza viruses, the hemagglutinin (HA), against which neutralizing antibodies are primarily directed (3, 5, 8). In the last 25 years there have been several pandemics associated with a change in the hemagglutinin subtype of circulating virus (antigenic shift). Thus, Asian flu, which arose in 1957, contained H2 HA, the Hong Kong flu epidemic of 1968 was caused by viruses with H3 HA, and in 1977, a pandemic of Russian flu occurred in younger persons after the reappearance of viruses with H1 HA (reviewed in reference 22). H3N2 virus has now circulated in the human population for 14 years. Antigenic changes, detectable with whole sera, began to occur in HA only 3 to 4 years after the subtype was first isolated in 1968, although all isolates retain some antigenic properties common to the subtype (antigenic drift). Some of the variants were able to cause renewed epidemics (although of varying severity in different countries), presumably because changes accumulating in HA prevented neutralization by antibodies directed against previous epidemic strains. At the same time, these particular variants retained the ability to grow rapidly and spread widely in the community. Other variants with altered antigenic properties have also been detected, but many caused only localized outbreaks or isolated cases of influenza (9). It is certainly true that the relative abilities of different virus isolates to survive and achieve dominance might be influenced by changes other than those in HA, e.g., among a group of antigenically equivalent strains. However, it appears that the major force promoting influenza subtype evolution has been the need to overcome immunity, either natural or induced by vaccination, and that escape from immunity has been achieved by HA alterations. The most comprehensive studies of antigenic drift in HA have been carried out for the Hong Kong (H3N2) subtype of influenza virus, but our knowledge of changes occurring in the HA protein or in the gene coding for it is largely limited to a few key strains which have caused renewed epidemics since 1968. To understand the differences between successful and unsuccessful variants so that future evolutionary directions for the subtype might be predicted, information on a much wider range of variants is needed.

In the experiments reported in this paper we examined the HA gene sequence of influenza strains chosen to provide a much more comprehensive view of antigenic drift at the molecular level than has been available previously. The

study shows that there are significant restrictions operating on HA gene variation during antigenic drift; extrapolation of the results to other HA types may help to explain the limited lifespan of subtypes previously circulating in the population.

MATERIALS AND METHODS

Preparation of influenza virus RNA. Influenza viruses grown in embryonated chicken eggs were purified, and the RNA was extracted as previously described (17).

Determination of the HA gene nucleotide sequence. The methods for determining the sequence of cDNAs copied from influenza virus RNA have been described previously (2).

Antigenic analysis of strains. Viruses were characterized with postinfection ferret sera as described previously (9).

RESULTS AND DISCUSSION

Epidemiology of H3N2 subtype isolates. Tables 1 and 2 summarize epidemiological and antigenic data for H3N2 isolates spanning the period from 1968 to 1980. These isolates were chosen for detailed study after a preliminary survey of a much wider range of isolates (data not shown). Among the strains listed in Table 1 are all of the seven isolates for which HA gene sequence data have been reported previously (except for the X31 recombinant of Aichi/2/68 [20], which is closely related to NT68).

From the descriptions of strains in Table 1 it is apparent that virus isolates can be divided into four groups. The first group contains international reference strains for epidemic variants, e.g., Eng72, PC73, Vic75, Tex77, and BK179. The transition from one epidemic era to another has been less clear since 1975 because of the cocirculation and persistence of some variants. The second group is composed of lateral variants which, when tested with whole animal sera, are antigenically equivalent to epidemic strains but represent separate geographical isolates (e.g., NT68, Eng69, and Qu70, which are similar to HK68; Mem72, which is similar to Eng72; and Eng75, which is similar to Tex77). The third group contains terminal variants, which are inconsistently isolated in the midst of epidemics caused by viruses similar to reference strains. Examples include HK71, MC75, Tok75, Sin75, Vic76, and AC76. Although antigenically distinct from contemporary strains, these variants failed to spread beyond local or regional outbreaks and were normally not detected for more than a few months at most. The fourth group consists of variants which appear to derive from prototype strains and appear consistently with them in many parts of the world over a period of 1 or more years, i.e., distinct from local or regional isolates. Examples include BK279 and Sha80.

Since we could not hope to study every isolate from a particular influenza subtype, we examined, in detail, isolates from each of these groups to achieve the broadest possible view of HA variation during antigenic drift.

HA gene evolution since 1968. In previous studies of 1968, 1972, 1975, and 1979 H3N2 isolates, silent base differences between 1968 strains and later strains were found widely spread over the HA gene. However, base substitutions resulting in HA amino acid changes were almost entirely limited to that region of the gene coding for HA1, the major subunit of the mature HA protein (2, 20, 22). Since HA1 sequences form the antibody-binding and active sites on the HA molecule (23, 24), we concentrated on identifying changes in the region of the gene coding for HA1.

In Table 3 we report partial or complete HA1 gene sequences for all of the strains listed in Table 1, except for Eng69 and Qu70, whose sequences have been published previously (18). The results emphasize the progressive accumulation of mutations in the HA gene, with HA1 changes occurring at a constant rate of ca. 0.7% per year between 1968 and 1979.

Nucleotide changes shown in Table 3 are predominantly transitions, i.e., C-U or A-G changes. The ratio of transitions to transversions (54:17) (when BK179 is compared with NT68) shows a pattern of mutation similar to that reported for the RNA phages MS2 and R17 (10, 20) and is consistent with the introduction of errors during replication of the RNA genome by enzymes lacking editing functions.

The evolutionary relationships between different isolates can be assessed from the number of HA gene mutations separating them (Fig. 1). This genetic relatedness tree approximates an evolutionary tree for the subtype, although it ignores the possibility that the mutation rate is high enough to produce independent mutations at the same position in different strains or undetected reversions of previously acquired mutations. At least two examples of reversions are apparent at bases 692 and 726 (Table 3) in which 1979 strains now have a sequence similar to that of the original 1968 strains. An actual evolutionary tree for the subtype would no doubt be more complex than that shown in Fig. 1 since the strains examined in this study represent only a small proportion of the total number of H3N2 variants that have arisen since 1968.

Antigenic drift and conserved mutations. The sequence data summarized in Table 3 demonstrate that viruses present in the population at any one time constitute a heterogeneous array. The contemporary strains share some amino

TABLE 1. Summary of antigenic and epidemiological data on influenza A H3 subtype (Hong Kong) viruses isolated from 1968 to 1980

Strain	Strain characteristics	Epidemiological characteristics
A/Northern Territory/60/68 (NT68)	Identical to 1968 prototype strains	Typical of strains of the 1968 epidemic after the first appearance of Hong Kong subtype virus
A/England/878/69 (Eng69), A/Queensland/7/70 (Qu70)	Similar to 1968 isolates although distinguishable by monoclonal antibodies (12, 19)	Part of the spectrum of strains circulating during the first H3N2 epidemic era
A/Hong Kong/107/71 (HK71)	Exhibits major asymmetrical variation from 1968-like strains	Only isolated outbreaks caused by viruses of this type occurred in 1971 and 1972 (14, 15)
A/England/42/72 (Eng72)	Exhibits modest symmetrical variation from 1968 strains	Reference strain for variants which replaced HK68-like strains worldwide during 1972
A/Memphis/102/72 (Mem72)	Antigenically similar to Eng72	See Eng72
A/Port Chalmers/1/73 (PC73)	Poorly inhibited by antisera to previous epidemic strains	Displaced Eng42-like strains worldwide during 1973
A/Singapore/4/75 (Sin75)	Exhibits asymmetrical variation from A/Port Chalmers/73-like strains	Isolated from local outbreaks before the appearance of Eng75-like and Vic75-like strains in Singapore
A/Mayo Clinic/1/75 (MC75)	Exhibits major variation from all other strains	A unique isolate from one of the last PC73-like epidemics
A/Tokyo/1/75 (Tok75)	Exhibits major variation from all other strains, but closest to Vic75	Cocirculated with Vic75 in Japan only but disappeared in 1975 to 1976 (6)
A/Victoria/3/75 (Vic75)	Poorly inhibited by antisera to previous epidemic strains	Displaced PC73-like strains worldwide as predominant strain during 1975 to 1976 and caused severe epidemics in many countries; prevalence decreased in 1976 to 1977, after which Vic75-like viruses ceased to be isolated
A/England/864/75 (Eng75)	Poorly inhibited by antisera to previous epidemic strains	Similar strains were identified within a few months after recognition of Vic75-like strains; generally caused only sporadic cases and localized outbreaks in 1975 to 1976, but prevalence relative to Vic75-like viruses steadily increased; by 1977 was predominant strain (see Tex77); similar viruses still occasionally identified in 1982, in conjunction with variants emerging from 1979 on
A/Victoria/112/76 (Vic76)	Although this variant is poorly inhibited by antisera to previous or contemporary strains, antiserum to it reacts strongly with Eng75	Detected sporadically during Vic75-like outbreaks

TABLE 1—*Continued*

Strain	Strain characteristics	Epidemiological characteristics
A/Allegheny County/29/76 (AC76)	Quite similar to Vic75 except that antiserum to this variant reacts strongly with HK68 and Eng42 variants	Detected sporadically during Vic75-like outbreaks
A/Texas/1/77 (Tex77)	Interchangeable with Eng75 in hemagglutination inhibition tests with animal sera	Chosen as reference strain for viruses which displaced Vic75 as predominant epidemic virus in 1977 (see Eng75)
A/Bangkok/1/79 (BK179)	Poorly inhibited by antisera to previous epidemic strains, but closest to Eng75-like viruses	During 1979 the predominant H3N2 viruses worldwide exhibited good reactivity with BK179, although many were, in addition, still also highly reactive with Eng75 and Tex77; thus, BK179 exhibited slightly more pronounced drift than predominant epidemic strains
A/Bangkok/2/79 (BK279)	Exhibits major antigenic drift from all other strains, although closest to Vic112, Eng75, and BK179	Isolated occasionally worldwide, usually from same outbreaks as BK179-like viruses
A/Shanghai/31/80 (Sha80)	Minor variant of BK179 which exhibits decreased inhibition compared to BK179 with Eng75 or Tex77 animal sera	About 33% of 1981 isolates, particularly from Asian and Western Pacific regions were of this type; they were isolated from the same outbreaks as BK179- and BK279-like strains during this period

acid differences from earlier strains but also carry changes characteristic of each strain. Presumably the combination of changes most favorable to the virus is the one that is seen to be preserved during evolution because successful strains dominate and so become progenitors for future generations. Thus, the most favorable combinations of changes might be identified by their transmission through the viral evolutionary line. To highlight the mutations associated with the evolution of new variants while minimizing the importance of background noise changes, which are not in the vertical evolutionary path of the virus, we illustrated the HA amino acid changes which have accumulated and been conserved in the major epidemic strains from 1968 to 1980 (Fig. 2). The changes are distributed over the proposed antigenic sites, A, B, and C (23) (Fig. 3). If putative site D (23) is extended to include residues such as 172, 174, 207, 242, and 244 (3a), then some D site changes are also conserved and transmitted during the evolutionary process (Fig. 2).

The results shown in Fig. 1 and 2 suggest that in about 1975 viral evolution resulted in two divergent pathways. One led to viruses with sequences most closely related to Vic75. This path subsequently died out after 1 or 2 years. The other led, via Eng75-like strains, to viruses of the type still in circulation. Analysis of the sequences implies that divergence occurred from a hypothetical variant which existed during the last period of circulation of PC73-like viruses (such as Sin75), but in addition, the variant acquired changes at amino acids 189 and 217 seen in all subsequent epidemic strains.

Selective pressures conserving sequences in some areas of the gene but favoring changes in others. From our analysis of the HA gene sequence for the strain BK179 it was apparent that HA amino acid changes in this successful strain were found clustered in proposed antigenic regions of the molecule, with other regions containing few, if any, changes (2). It is evident that when a larger number of variants is studied the changes are still restricted to these same variable regions even in relatively unsuccessful strains. This is illustrated in Fig. 3, which shows the locations of HA amino acid changes in all of the strains described in this study. It appears

TABLE 2. Hemagglutination inhibition reactions of influenza A (H3N2) virus isolates, from 1968 to 1980, for which HA1 sequences have been determined

Virus	Hemagglutination inhibition reactions[a] with the following ferret sera:													
	A/Hong Kong/8/68[b]	A/Hong Kong/107/71	A/England/42/72[b]	A/Port Chalmers/1/73[b]	A/Mayo Clinic/4/75	A/Singapore/4/75	A/Victoria/3/75[b]	A/Tokyo/1/75	A/England/864/75	A/Allegheny County/29/76	A/Victoria/112/76	A/Bangkok/1/79[b]	A/Bangkok/2/79	A/Shanghai/31/80
A/Hong Kong/8/68[c]	**1,280**	320	960	80	20		30			480		20		
A/Hong Kong/107/71		**480**												
A/England/42/72	640	80	**1,280**	640	60	120	80			640	80	20		
A/Port Chalmers/1/73		40	40	**320**	20	160	40			20	60			
A/Mayo Clinic/4/75				80	**320**	20								
A/Singapore/4/75		60	20	100	50	**960**	60		80	60	120		20	30
A/Victoria/3/75	30		40	20	240	40	**640**	20	20	480	640			
A/Tokyo/1/75							80	**320**			120	20		20
A/England/864/75		40		80	30	160	80		**1,920**	120	1,920	240	160	320
A/Allegheny County/29/76	40	20	80	20	20	40	240			**640**	240		30	
A/Victoria/112/76						20	80		120	80	**2,560**	20	20	40
A/Bangkok/1/79									160		160	**960**	240	1,280
A/Bangkok/2/79									80		320	120	**2,560**	40
A/Shanghai/31/80									40		80	320	160	**1,280**

[a] Blank indicates titer of <20. Results are means of duplicate tests.
[b] Serum to recombinant with neuraminidase N7.
[c] Sequences have been determined for antigenically similar strains A/Aichi/2/68 and A/Northern Territory/68.

that the sites where amino acid changes can alter antigenicity are so limited that in some strains key amino acids in these sites change for a second time as evolution proceeds (Fig. 3 and Table 3). The occurrence of these sequential changes at residues 145, 188, 189, and 193 suggests that the HA three-dimensional structure remains constant during antigenic drift, presumably so that the biological function of HA can be maintained.

The considerable conservation of core residues both within the upper globular section of HA and in the supporting stalk region noted earlier (23, 24) now extends across four influenza subtypes, H1, H2, H3, and H7 (1, 4, 7, 11, 13, 21, 25), suggesting that the overall conformation and the locations of biologically active and antigenic regions are fundamentally similar for HAs of all subtypes. It has been suggested that some residues may be conserved in different subtypes to act as anchors that maintain equivalent loops of amino acids at the HA surface (23, 24). In the B antigenic region (Fig. 2), anchor residues corresponding to amino acids 184, 185, 190, 191, 194, and 195 of H3 HA are conserved both within and among HA types (2, 4, 7, 13, 20, 25). Of the surface residues in this region, 188, 189, and 193 are highly variable among H3 strains and among subtypes as well. On the other hand, the adjacent residues 187, 192, and 196 vary among subtypes but have been conserved among all H3 strains examined so far. It is possible that changes in these residues (and in analogous residues in other sites, e.g., 140 to 142 in the A site) may alter the nature of the antigenic loops sufficiently to account for the antigenic differences among subtypes (16).

Generation of antigenic diversity among 1975 to 1976 isolates by sequential amino acid changes. Analysis of strains from the 1975 to 1976 and 1979 to 1981 periods has demonstrated that at these times a variety of distinctive viruses were isolated (3b, 26). Four variants from the 1975 to 1976 period (MC75, Tok75, Vic76, and AC76) were also the first variants identified to carry sequential amino acid changes of the type described above. Amino acid 193, which changed to asparagine in PC73, changed again to lysine in MC75 and to aspartic acid in Tok75, AC76, and Vic76. AC76 also has a second base change within the codon for residue 189, producing arginine in AC76 compared to lysine in contemporary strains. (In contrast, Sin75 appears to have mutated at residue 189 directly from glutamine to arginine.) AC76 and Vic76 differ from each other by changes at residues 157, 189, 209, and 240, and of these, the changes at 157 and 189 (both part of the B region of Fig. 3) are most

TABLE 3. Summary of HA1 nucleotide sequence changes from A/Northern Territory/60/68 (18) for 15 H3N2 influenza isolates[a]

BASE NO.	NT68	HK71	ENG72	MEM72	PC73	SIN75	MC75	TOK75	VIC75	AC76	VIC76	ENG75	TEX77	BK179	BK279	SHA80	AA NO.	AA CHANGE
59	U													C				
64	G	A															-5	C→Y
67	U												A				-4	L→Q
70	C		U	U	U			U	U	U				U	U		-3	A→V
72	C							U	U	U				U	U		-2	L→F
75	G	A															-1	G→S
76	G							C	G	-		C	C				-1	G→A
76	G													U			-1	G→V
81	G												A	A	A		2	D→N
83	C		U															
84	C			U	U			C	C	C		C	C				3	L→F
89	A													C	C			
103	A	G	G	G	G			G	G	G		G	G	G			9	N→S
104	C																	
106	C	A															10	T→K
113	G								U	U								
128	U													C	C			
134	G													A	A			
149	A							G	G	G		G	G	G				
167	A							G	G	G		G	G	G				
168	G		A	A	A			A	A	A		A	A	A			31	D→N
178	U	C															34	I→T
188	U									C								
200	G								A	A						=		
203	A		G	G	G		=	G	G	G	=	G	G	G	G			
212	C			U	U	=	U	U	U	U	U	U	U	U		U		
221	G										A	A	A	A	A			
224	G					U		U	U	U	U	U	U	U	U			
226	A										G	G	G	G	G	50		K→R
234	A								G	G	A	G	G	G	G	53		N→D
238	A											G	G	G	G	54		N→S
245	U												C	C	C			
260	A												G	G	G	G		
262	U												A	A	A	A	62	I→K
264	G				A	A	A		A	A	A	A	A	A	A	A	63	D→N
275	G												A					
287	A			G														
293	G													A	A	A		
310	U		G	G	G	G	G		G	G	G	G	G	G	G	G	78	V→G
311	A		C	C	C	C	C		A	C	C	C	C	C	C	C		
321	G												A				82	E→K
325	C				A	A		A	A	A	A	A	A	A	A	A	83	T→K
338	C						=				U	U	U	U	=			
404	G	A													=			
428	U													C	C			
429	C			U														
440	C	U																
442	C		A	A	A	A	A	A	A	A	A	A	A	A	A	A	122	T→N
446	G			A	A	A	A	A	A	A	A	A	A	A	A	A		
449	U			C	C	C	C	C	C	C	C	C	C	C	C	C		
454	C						U										126	T→I
454	C				A	A		A	A	A	A	A	A	A	A	A	126	T→N
463	G	A															129	G→E
464	G												A					
471	C	G															132	Q→E
475	A													G	G	G	133	N→S
486	A											U	U	U	U	U	137	N→Y
487	A								G	G	G						137	N→S
500	G								A									
504	C													U	U	U	143	P→S
508	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	144	G→D
510	A							C									145	S→R
511	G								A	A	A	A	A	A	A	A	145	S→N
511	G						U									A	145	S→I
512	C															A	145	N→K*
513	G													A	A	A	146	G→S
518	U												C	C	C	C		
533	C													U	U	=		
540	A		U	U	U	=	U	U	U	U	U	U	U	U			}	
541	C		A	A	A		A	A	A	A	A	A	A	A			155	T→Y

BASE NO.	NT68	HK71	ENG72	MEM72	PC73	SIN75	MC75	TOK75	VIC75	AC76	VIC76	ENG75	TEX77	BK179	BK279	SHA80	AA NO.	AA CHANGE
543	A							C									156	K→Q
543	A													G	G		156	K→E
547	C										U						157	S→L
550	G						A					A	A	A	A		158	G→E
554	C	A															159	S→R
555	A				G												160	T→A
555	A							U									160	T→S
556	C													A	A		160	T→K
568	U								A	A	A	A					164	L→Q
569	G									A								
572	C			U														
578	U								C	C	C	C						
584	A										G							
592	A												G	G	G		172	D→G
596	U									C	C							
598	U								C	C	C	C					174	F→S
608	A												G	G	G			
620	G					=							A		=			
638	G														A			
639	A	G		G	G	G	G	G	G	G	G	G	G	G		G	188	N→D
639	A														U		188	D→Y*
642	C									A	A	A	A	A	A	A	189	Q→K
643	A					G											189	Q→R*
643	A										G						189	K→R*
654	A								G		G	G					193	N→D*
655	G				A	A	A	A	A	A	A	A	A	A	A	A	193	S→N
656	C						A										193	N→K*
656	C														G		193	N→K*
659	G		A	A	A	A	A	A	A	A	A	A	A	A	A	A		
667	A													G	G	G	197	Q→R
669	G	A			A											=	198	A→T
679	G								A	A	A	A					201	R→K
692	U				C	C	C	C	C	C	C	C	C	U	U			
695	C				A													
697	G		A	A	A	A	A	A	A	A	A	A	A	A			207	R→K
699	A		G														208	R→G
703	G										A					=	209	S→N
714	A									G	G						213	I→V
726	A						G	G	G	G	G	G	G	A	A	A	217	I→V
731	G	A													A			
732	U															C	219	S→P
734	C		U	U	U	U	U	U	U	U	U	U	U	U	U	U		
754	U			A		=		=								=	226	L→Q
762	A												U				229	R→G
765	A										G	G					230	I→V
770	C													U	U			
788	U												A	A	A			
791	G		A	A	A		A		A	A	A	A	A	A				
795	G										A						240	G→R
801	G		A	A	A		A		A	A	A	A	A	A			242	V→I
807	G													U	U	U	244	V→L
824	G										A	A						
833	C		U	U	U		U		U	U	U	U	U	U				
848	U				C		C		C	C	C	C	C	C				
857	G											A	A	A	A		260	M→I
896	U					=							C					
901	A		G	G	G	G	G		G	G	G	G	G	G	G		275	D→G
902	U			C	C	C	C		C	C	C	C	C	C	C			
908	U			C	C	C	C		C	C	C	C	C	C	C			
910	U					G	G		G	G	G	G	G				278	I→S
911	U										C	C	C					
950	C	U																
953	G		A															
977	C		U				=											
986	A		G	G	G	G			G	G	G	G	G	G				
992	C		U	U	U	U			U	U	U	U	U	U				
998	G	A				=												
999	U											A						
1016	C				U				U	U		U	U	U				
1034	G	A																

[a] Blank areas indicate sequence identity with NT68. Amino acids (AA) are numbered from the N-terminal residue of HA1. Amino acid changes relative to NT68 are given in the single-letter code (2). Symbols: *, Amino acid residues which have undergone a sequential change; =, a partial sequence was determined. The nucleotide sequences not determined are as follows: Sin75, nucleotides 59 to 212, 540 to 620, 754 to 896, and 998 to 1,061; MC75, nucleotides 59 to 203 and 977 to 1,061; Tok75, nucleotides 59 to 338 and 754 to 1,061; Eng75, nucleotides 59 to 203; and Sha80, nucleotides 59 to 200, 338 to 404, 533 to 620, 669 to 703, and 754 to 1,061.

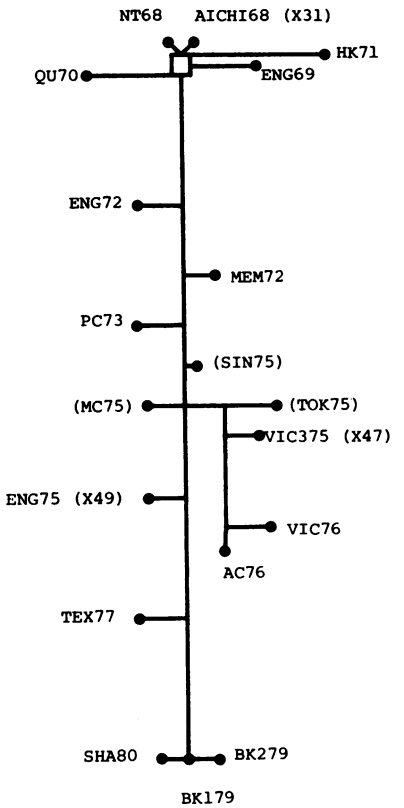

FIG. 1. Interrelationship of Hong Kong subtype strains according to the minimum mutational distances between their HA1 coding regions. The figure is drawn to scale, and the number of mutations separating different strains can be calculated relative to the NT68-BK179 distance (73 mutations). Strains in parentheses have been only partially sequenced. Three mutations at bases 84, 149, and 167 occurred at some point between 1973 and 1975 but cannot be positioned exactly. The full names of the virus strains examined are listed in Table 1. The Aichi68 and Eng75 strains sequenced were the recombinants X31 and X47, respectively. HA gene sequences have been published previously for strains NT68 (18), Eng69 (18), Qu70 (18), Mem72 (18), Vic75 (11), and BK179 (2).

likely to produce the considerable antigenic differences between these strains. Similarly, the contemporary strains BK179 and BK279, antigenically quite distinct from one another (Table 2), differ only by second changes in amino acids 188 and 193.

Thus, it is clear that the sequential change of key amino acids within antigenic regions generates new antigenic variants. It is a curious observation that, so far, no virus carrying a sequential amino acid change has given rise to a new variant having major epidemiological impact. Despite the large number of strains with sequential changes arising during the 1975 to 1976 period, the successful variants which emerged from this period carried changes of a different type, e.g., at residues 133 and 143 in the A region and residues 156 and 197 in the B region (Fig. 2 and Table 3). This suggests that not all changes which appear to be antigenically favorable are optimal for virus survival.

Evolutionary future of the H3 subtype. One of the most striking observations to emerge from our analysis of H3 HA variation is that the orientation of amino acids involved in antibody binding appears to remain constant as evolution proceeds. Amino acids such as 145, 188, 189, and 193, which occupy key positions in potential antigenic sites in 1968 strains (24), change early

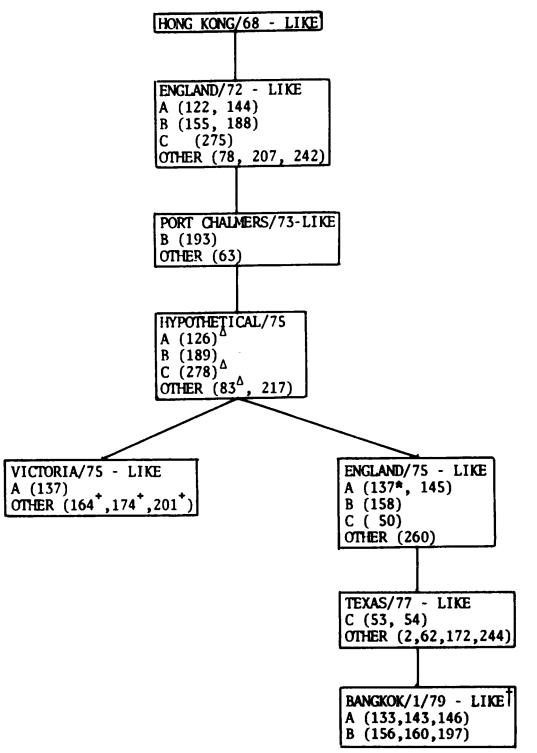

FIG. 2. Evolution of fixed amino acid changes in the HA1 region of H3 influenza HA. Only the changes forming the mainstream of evolution are included (see text), with viruses grouped according to their antigenic and epidemiological characteristics listed in Table 1. Changes are shown for putative antigenic sites A through C or elsewhere on the molecule (Other). Symbols: +, lacking in Tex77 and successive viruses, consistent with two independent evolutionary paths; *, different from Vic75-like viruses, consistent with two independent evolutionary paths; †, change at amino acid 217 has reverted in these strains; Δ, these changes are seen in Sin75, except that amino acid 189 changed to arginine instead of lysine, which was found in subsequent epidemic strains.

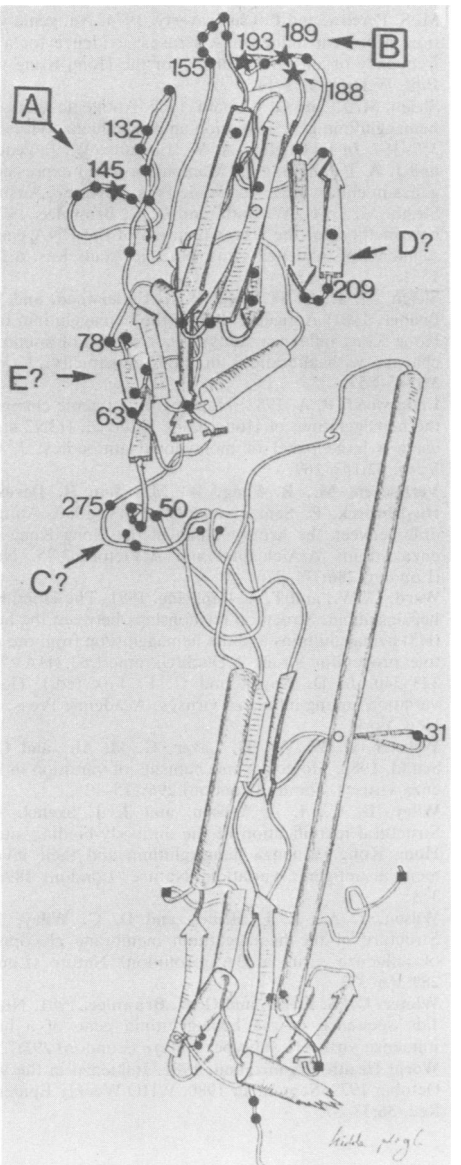

FIG. 3. Three-dimensional structure of HA for a 1968 strain. Symbols: ●, location of HA1 amino acid changes since 1968 for all of the strains shown in Table 1; ★, amino acids undergoing sequential change during H3 evolution; ■, HA2 changes in strains Mem72, Vic75, and BK179. Antigenic sites marked A through E are those proposed in reference 24.

during subtype development and are also the sites for sequential changes. As discussed above, no changes have been seen in anchor residues for the antigenic loops or in nearby residues (e.g., 140 to 142 in the A site and 187, 192, and 196 in the B site) which might influence loop conformation. If no such changes occur as H3 evolution proceeds, then pressure will continue for change in the same small groups of residues most closely involved in antibody attachment. Thus, there may be a limited number of variants which can be produced during antigenic drift within an influenza virus subtype. Of these, only a few may be epidemiologically successful. There are already some indications that not all substitutions at particular residues are favorable for subsequent survival of the virus. There is also evidence from field isolates for a strong bias in favor of particular types of changes in different antigenic regions (to charged or larger residues or both in the A and B regions and to smaller residues in the C region; Table 3). Since such substitutions in the protein result from single base changes in the HA gene, the options for further evolution at particular sites are restricted. For example, the change at residue 155 from threonine to tyrosine may represent an evolutionary endpoint, and at residue 193, already changed from serine to asparagine to lysine, only a further change to glutamic acid may be effective. An interesting possibility resulting from this type of analysis is the synthesis of peptides representing predicted end points of antigenic site evolution for evaluation as prospective vaccines.

The usefulness of this approach in analyzing constraints and predicting future directions for HA evolution will become apparent as new H3 variants arise or with examination of the completed evolutionary process in earlier influenza subtypes, such as H1N1, which circulated in humans between 1918 and 1957.

ACKNOWLEDGMENTS

We are most grateful to E. Hamilton, L. Siegman, C. A. Stocksdale, J. G. Galphin, and R. Black for expert technical assistance, to V. Bender, B. Moss, and P. Underwood for helpful discussions, to J. Drummond and B. Lawrence for typing the manuscript, and to C. Bucholtz and A. Reisner for assistance with computing. We also thank D. Wiley for the sketch of the HA structure by Hidde Ploegh.

LITERATURE CITED

1. **Both, G. W., and M. J. Sleigh.** 1980. Complete nucleotide sequence of the hemagglutinin gene from a human influenza virus of the Hong Kong subtype. Nucleic Acids Res. **8**:2561–2575.
2. **Both, G. W., and M. J. Sleigh.** 1981. Conservation and variation in the hemagglutinins of Hong Kong subtype influenza viruses during antigenic drift. J. Virol. **39**:663–672.
3. **Brand, C. M., and J. J. Skehel.** 1972. Crystalline antigen from the influenza virus envelope. Nature (London) New Biol. **238**:145–147.

3a. **Caton, A. J., G. G. Brownlee, J. W. Yewdell, and W. Gerhard.** 1982. The antigenic structure of the influenza virus A/PR/8/34 haemagglutinin (H1 subtype). Cell **31**:417–427.

3b. **Cox, N. J., Z. S. Bai, and A. P. Kendal.** 1983. Laboratory-based surveillance of influenza A(H1N1) and (H3N2) viruses in 1980–81: antigenic and genomic analyses. Bull. W.H.O. **61**:143–152.

4. **Davis, A. R., D. P. Nayak, M. Veda, A. L. Hiti, D. G. Kleid, and D. Dowbenko.** 1981. Structure of the hemagglutinin gene of H0N1 strain of human influenza virus and its expression in *E. coli*, p. 159–167. *In* D. Nayak and C. F. Fox (ed.), Genetic variation among influenza viruses. Academic Press, Inc., New York.
5. **Eckert, E. A.** 1973. Properties of an antigenic glycoprotein isolated from influenza virus hemagglutinin. J. Virol. **11**:183–192.
6. **Fukumi, H., and J. Takeuchi.** 1977. Antigenic variation of influenza viruses isolated from the recent epidemics. Dev. Biol. Stand. **39**:417–420.
7. **Gething, M.-J., J. Bye, J. Skehel, and M. Waterfield.** 1980. Cloning and DNA sequence of double stranded copies of hemagglutinin genes from H2 and H3 strains elucidates antigenic shift and drift in human influenza virus. Nature (London) **287**:301–306.
8. **Jackson, D. C., T. A. Dopheide, R. J. Russell, D. O. White, and C. W. Ward.** 1979. Antigenic determinants of influenza virus hemagglutinin. II. Antigenic reactivity of the isolated N-terminal cyanogen bromide peptide of A/Memphis/72 hemagglutinin heavy chain. Virology **93**:458–465.
9. **Kendal, A. P., J. M. Joseph, G. Kobayashi, D. Nelson, C. R. Reyes, M. R. Ross, J. L. Sarandria, R. White, D. F. Woodall, G. R. Noble, and W. R. Dowdle.** 1979. Laboratory-based surveillance of influenza virus in the United States during the winter of 1977–1978. Am. J. Epidemiol. **110**:449–461.
10. **Min Jou, W., and W. Fiers.** 1976. Studies on the bacteriophage MS2: comparison of the nucleotide sequences in related bacteriophage RNAs. J. Mol. Biol. **106**:1047–1060.
11. **Min Jou, W., M. Verhoeyen, R. Devos, E. Saman, R. Fang, D. Huylebroeck, and W. Fiers.** 1980. Complete structure of the hemagglutinin gene from the human influenza A/Victoria/3/75 (H3N2) strain as determined from cloned DNA. Cell **19**:683–696.
12. **Moss, B. A., P. A. Underwood, V. J. Bender, and R. G. Whittaker.** 1980. Antigenic drift in the hemagglutinin 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.
13. **Porter, A. G., C. Barber, N. J. Carey, R. A. Hallewell, G. Threlfall, and J. S. Emtage.** 1979. Complete nucleotide sequence of an influenza virus hemagglutinin gene from cloned DNA. Nature (London) **282**:471–477.
14. **Schild, G. C., M. Henry-Aymard, M. S. Pereira, P. Chakraverty, W. Dowdle, M. Coleman, and W. K. Chang.** 1973. Antigenic variation in current human type A influenza viruses: antigenic characteristics of the variants and their geographic distribution. Bull W.H.O. **48**:269–278.
15. **Schild, G. C., J. S. Oxford, W. R. Dowdle, M. Coleman, M. S. Pereira, and P. Chakraverty.** 1974. Antigenic variation in current influenza A viruses: evidence for a high frequency of antigenic "drift" for the Hong Kong virus. Bull. W.H.O. **51**:1–11.
16. **Sleigh, M. J., and G. W. Both.** 1982. Antigenic change and hemagglutinin gene evolution among influenza viruses, p. 357–364. *In* P. Nagley, A. W. Linnane, W. J. Peacock, and J. A. Pateman (ed.), Manipulation and expression of genes in eukaryotes. Academic Press, Sydney, Australia.
17. **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.
18. **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.
19. **Underwood, P. A.** 1982. Mapping of antigenic changes in the hemagglutinin of Hong Kong influenza H3N2 strains using a large panel of monoclonal antibodies. J. Gen. Virol. **62**:153–169.
20. **Verhoeyen, M., R. Fang, W. Min Jou, R. Devos, D. Huylebroeck, E. Saman, and W. Fiers.** 1980. Antigenic drift between the hemagglutinin of the Hong Kong influenza strains A/Aichi/2/68 and A/Victoria/3/75. Nature (London) **286**:771–776.
21. **Ward, C. W., and T. A. Dopheide.** 1981. The Hong Kong hemagglutinin. Structural relationships between the human (H3) hemagglutinins and the hemagglutinin from the putative progenitor strain A/Duck/Ukraine/1/63 (HAV7), p. 323–340. *In* D. Nayak and C. F. Fox (ed.), Genetic variation among influenza viruses. Academic Press, Inc., New York.
22. **Webster, R. W., W. G. Laver, G. M. Air, and G. C. Schild.** 1982. Molecular mechanisms of variation in influenza viruses. Nature (London) **296**:115–121.
23. **Wiley, D. C., I. A. Wilson, and J. J. Skehel.** 1981. Structural identification of the antibody-binding sites of Hong Kong influenza hemagglutinin and their involvement in antigenic variation. Nature (London) **289**:373–378.
24. **Wilson, I. A., J. J. Skehel, and D. C. Wiley.** 1981. Structure of the hemagglutinin membrane glycoprotein of influenza virus at 3Å resolution. Nature (London) **289**:366–373.
25. **Winter, G., S. Fields, and G. G. Brownlee.** 1981. Nucleotide sequence of the hemagglutinin gene of a human influenza virus H1 subtype. Nature (London) **292**:72–75.
26. **World Health Organization.** 1981. Influenza in the world October 1979–September 1980. WHO Weekly Epidemiol. Rec. **56**:33–37.