VIROLOGY 148, 275–287 (1986)

The Antigenicity and Evolution of Influenza H1 Haemagglutinin, from 1950–1957 and 1977–1983: Two Pathways from One Gene

F. L. RAYMOND, A. J. CATON,[1] N. J. COX,*
A. P. KENDAL,* AND G. G. BROWNLEE[2]

*Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, United Kingdom, and *Center for Disease Control, Atlanta, Georgia 30333*

Received July 29, 1985; accepted September 25, 1985

Nucleotide sequence analysis of the region of the haemagglutinin gene coding for the HA1 domain of the protein was performed on 19 human influenza A strains of H1 subtype representative of the two epidemic periods from 1977–1983 and from 1950–1957. The amino acid changes relative to A/USSR/90/77 are summarised and are consistent with the view that variation in these field strains involves changes largely at the Sb and Ca antigenic sites previously characterised in laboratory mutants of the haemagglutinin of influenza A/PR/8/34. The Sa and Cb sites are less variant and are probably masked by carbohydrate side chains. We discuss the significance of other amino acid changes which do not correspond to previously defined antigenic sites. We also define the "mainstream" amino acid changes characteristic of the divergent evolutionary pathways of the 1950–1957 and 1977–1983 periods and note that the rate of evolution is faster in the earlier period.

INTRODUCTION

The haemagglutinin molecule is a remarkably polymorphic protein present on the surface of all influenza viruses. It can vary in amino acid sequence by up to 60% in different influenza A isolates (Winter *et al.*, 1981), yet still maintain its function in the life cycle of the virus (Lamb and Choppin, 1983; McCauley and Mahy, 1983). It is well known that this sequence variation contributes to antigenic variation and this property confers a selective advantage on the virus allowing it to escape the host immunity derived from previous infection or vaccination by influenza (for reviews, see Wiley *et al.*, 1981; Webster *et al.*, 1982). A better understanding of evolution and variation of the haemagglutinin is a prerequisite for improving the control of the virus through vaccination.

Our previous studies (Gerhard *et al.*, 1981; Caton *et al.*, 1982, 1983) of the antigenicity of the haemagglutinin of H1 subtype in influenza A/PR/8/34 indicated that, like the H3 subtype haemagglutinin (Wiley *et al.*, 1981), there are four regions on the globular head of this haemagglutinin trimer which form four more or less discrete antigenic areas. Two of these areas, Sa and Sb, are located at or very near the tip of the spike, whilst a third, the Ca site, is formed by the outer cleft between adjacent haemagglutinin monomers and lies on the side and approximately halfway down the globular head. The fourth site, Cb, occupies an area near the base of the globular head. All these sites are formed by the HA1 domain of the haemagglutinin. Because this model of antigenicity was based on the immune response of BALB/c mice and analysis of laboratory selected virus mutants, we have now investigated whether the antigenic residues purported to be part of the four immunodominant domains were residues which altered in human isolates of influenza of H1 subtype.

[1] Present address: The Wistar Institute, 36th Street at Spruce, Philadelphia, Pennsylvania 19104.

[2] Author to whom requests for reprints should be addressed.

0042-6822/86 $3.00

The H1 virus is uniquely suited to such a study, because analysis is possible not only of strains isolated from 1934 to 1957, but also of strains after 1977 when H1 virus reappeared in strains antigenically similar to a previous 1950 strain (Kendal *et al.*, 1978).

Therefore, we report here the sequence determination and analysis of HA1 regions for 19 strains of H1N1 subtype, 6 for the period 1950–1957 and 13 for the period 1977–1983.[3] The results obtained confirm that the sequence of the haemagglutinin of a 1977 reference strain is virtually identical to the 1950 reference strain. Thus in this study, we have followed the evolution of one gene on two separate occasions.

MATERIALS AND METHODS

Virus strains, virus growth, and RNA preparation. Table 1 lists the strains sequenced and an abbreviation for each strain used in subsequent tables and in the text. Table 2 shows the antigenic relationships of these strains to one another in a haemagglutination inhibition test (HI) with ferret sera. These have not been reported previously in a single simultaneous test but only as separate tests for the different eras (Kendal *et al.*, 1978, 1979). All strains were human isolates received by the WHO Influenza Centre, CDC, Atlanta, Georgia. Strains were selected from 1950–1957 to represent recognised antigenic variants from that period (Kendal *et al.*, 1978). For more recent strains where considerable heterogeneity has been recognised in antigenic tests with immune sera, strains were selected as either reference strains (USSR/77, BRZ/78, ENG/80, IND/80, CHL/83, DUN/27/83, VIC/83), or as antigenically similar in HI tests (Table 2) to reference strains (GA/114/83, TX/12/82, DUN/6/83, GA/79/83, HK/83), or as occasional variants (LACK/78, TX/29/82). Our analysis also included multiple isolates from the same epidemic wave (DUN/6/83 and DUN/27/83) as well as seven 1983 isolates from different parts of the world. The isolates were grown in the allantoic cavity of 11-day-old embryonated hen's eggs. Virus was purified by polyethylene glycol precipitation and sucrose density gradient centrifugation (Cox and Kendal, 1984) and total virion RNA purified as before (Raymond *et al.*, 1983), to give a preparation of RNA of approximately 1 mg/ml in 10 m*M* Tris-chloride, 0.1 m*M* EDTA, pH 7.4. Between 20 and 200 μg of RNA was prepared for most strains.

Nucleotide sequence analysis. RNA sequence analysis by the dideoxy chain termination method was performed essentially as described previously (Caton *et al.*, 1982; Raymond *et al.*, 1983). The six oligonucleotides synthesised (Sproat and Bannworth, 1983) were d(TAAAAACAACCAAAATG) primer I, d(AATCATGGTCCTACATTG) primer II, d(TTTTACAGAAATTTGCTA) primer III, d(CAATAATATTTGAGGCAA) primer IV,

TABLE 1

FULL IDENTIFICATION OF SEQUENCED STRAINS AND A SHORTHAND ABBREVIATION

Strain name	Abbreviation
A/Fort Warren/1/50	FW/50
A/England/1/51	ENG/51
A/Fort Leonard Wood/1/52	FLW/52
A/Queensland/34/54	QSL/54
A/Denver/1/57	DEN/57
A/USSR/90/77	USSR/77
A/Lackland/3/78	LACK/78
A/Brazil/11/78	BRZ/78
A/England/333/80	ENG/80
A/India/6263/80	IND/80
A/Texas/12/82	TX/12/82
A/Texas/29/82	TX/29/82
A/Georgia/79/83	GA/79/83
A/Georgia/114/83	GA/114/83
A/Hong Kong/32/83	HK/83
A/Chile/1/83	CHL/83
A/Dunedin/6/83	DUN/6/83
A/Dunedin/27/83	DUN/27/83
A/Victoria/7/83	VIC/83

[3] A report describing partial results, i.e., the sequence of 5 H1N1 isolates from 1977–1980 has been published previously (Raymond *et al.*, 1983).

TABLE 2

HAEMAGGLUTINATION INHIBITION REACTIONS OF INFLUENZA A (H1N1) VIRUSES ISOLATED FROM 1950 TO 1957 AND FROM 1977 TO 1983

	Ferret sera													
	"Old H1N1" strains					"New H1N1" strains								
Antigen	FW/50	ENG/51	FLW/52	QSL/54	DEN/57	USSR/77[a]	BRZ/78	LACK/78	ENG/80	IND/80	TX/29/82	DUN/27/83	CHL/83	VIC/83
"Old H1N1" strains														
FW/50	**320**	20	20	40	10	160	320	160	80	40	20	40	80	20
ENG/51	80	**640**	80	80	40	40	160	160	20	40	40	80	80	40
FLW/52	20	20	**1280**	320	40	40	20	10	<10	10	10	<10	<10	<10
QSL/54	40	10	80	**160**	40	10	80	80	20	80	40	80	80	40
DEN/57	10	<10	20	20	**2560**	<10	40	20	10	20	10	20	20	20
"New H1N1" strains[b]														
USSR/77	640	40	40	40	20	**320**	320	320	160	80	40	40	80	40
BRZ/78	320	40	40	80	40	40	**640**	320	320	80	40	80	160	40
LACK/78	320	20	40	80	40	80	160	**5120**	160	160	1280	320	320	80
ENG/80	160	20	40	40	20	80	320	640	**320**	320	160	80	320	80
GA/114/83	160	40	40	160	40	80	320	320	320	320	160	80	320	80
IND/80	80	10	80	80	80	80	160	320	160	**1280**	320	320	320	80
TX/12/82	80	10	40	80	40	80	160	160	160	640	320	160	320	80
TX/29/82	40	10	40	40	40	80	80	1280	80	160	**1280**	320	160	80
DUN/27/83	40	10	40	40	40	40	80	640	40	160	640	**640**	160	80
HK/83	40	10	40	40	40	20	80	640	40	80	640	160	80	40
CHL/83	40	<10	20	20	20	20	80	80	40	80	40	40	**160**	40
DUN/6/83	20	20	40	40	20	20	80	40	40	40	20	40	80	40
GA/79/83	40	<10	40	40	20	40	80	80	80	80	40	40	160	40
VIC/83	80	20	40	40	40	80	160	320	160	160	160	160	640	**640**

[a] Serum to recombinant with neuraminidase N7.

[b] Listed with antigenically related strains juxtaposed, i.e., GA/114/83 = ENG/80; TX/12/82 = IND/80; TX/29/82 and HK/83 = DUN/27/83; GA/79/83 and DUN/6/83 = CHL/83.

d(CTGGCTGACGGAGGCAAA) primer V, and d(GCCATGGTATGCTTTCGC) primer VI. Primers I–IV were complementary to nucleotide residues 19–35, 304–321, 513–530, and 799–816 of the haemagglutinin gene of influenza A/PR/8/34 (Winter *et al.*, 1981) and were used to sequence most viral strains with the exception that primer I failed to prime LACK/78 and FW/50 and this region of these strains remained unknown. Primer V was complementary to residues 530–547 of ENG/51 and replaced primer III for sequencing the corresponding region of ENG/51, FLW/52, QSL/54, and DEN/57. Primer VI was complementary to residues 833–850 of ENG/51 and replaced primer IV for sequencing the corresponding region of ENG/51. The nucleotide sequence of the HA1 coding region of USSR/77 was taken from Raymond *et al.* (1983) with the correction of residue 239 to a G and the assignment of residue 926 to an A from the results of Nakajima *et al.* (1983) and Concannon *et al.* (1984). Where there were technical difficulties in sequencing (i.e., a band of identical mobility present in all four tracks of the sequencing gel), we assumed that the sequence of the uncertain base was the same as in USSR/77. Because this assumption may not always be valid, we estimate that there may be one undetected sequence change in the 1950 and the 1978–83 strains, and up to three undetected changes in the 1951–1957 strains because overall they were more divergent in sequence from USSR/77 than the 1950 or 1978–1983 strains.

RESULTS

Table 3 shows the nucleotide sequence changes in the region of the haemagglutinin gene which encodes the HA1 domain of the protein in 19 different human influenza strains (see under Methods for selection of strains and their antigenicity), and Fig. 1 the deduced amino acid sequences. The HA1 domain of the haemagglutinin is 326 amino acid residues long (excluding the signal peptide) in all strains except DEN/57 which is one residue shorter due to a deletion at residue 130. Fifty-four positions occur, excluding changes in the signal peptide, where point mutations are observed relative to the sequence of USSR/77, which is taken as the parent strain. These alterations may be classified as sporadic, occurring in only one or at most a few usually unrelated strains, or more generalised, occurring in many strains in a systematic manner (Fig. 1). There are 10 hypervariable positions, i.e., residues 101, 132, 142, 143, 157, 163, 190, 225, 227, and 263, where two or more different amino acids occur—residue 142 being the most variable where four different amino acids occur. Amino acid changes can be classified as superficial or buried (Table 4) by comparison with the three-dimensional structure for a haemagglutinin of H3 subtype (Wilson *et al.*, 1981), assuming that the structures in H1 and H3 subtypes are similar, for which there is good circumstantial evidence (Winter *et al.*, 1981; Caton *et al.*, 1982; Both *et al.*, 1983). Forty-three changes are assigned to the surface of the trimer and 11 are buried. There is doubt about some residues classified as superficial which are close to positions of monomer-monomer contact. For example, residue 218 may be partly buried. Other superficial residues may be "masked" by carbohydrate side chains (see under Discussion). Superficial residues often cluster into discrete areas of the surface (see Fig. 2). The region exposed at the top of the molecule, forming an α helix, supports many changes, i.e., at residues 189–198. Similarly, the loop at the side of the molecule also supports several changes, at residues 140–143. Other smaller clusters occur (1) between residue 129–134, (2) around residue 225, and (3) around residue 219. Lower down in the hinge region of the haemagglutinin, there is another area of clustered amino acid changes including residues 273, 274, and 279 and 53, 54, and 56. There is a smaller cluster of changes at residues 79–82.

The evolutionary relationships (Fig. 3) between different virus isolates can be deduced from the number of base changes (Table 3) separating them (Both *et al.*,

TABLE 3

NUCLEOTIDE CHANGES IN THE HA1 DOMAIN OF 19 H1 SUBTYPE STRAINS FROM 1950-1983[a]

BASE NO.	DEN/57	QSL/54	FLM/52	ENG/51	FW/50	USSR/77	LACK/78	BRZ/78	ENG/80	GA/114/83	IND/80	TX/12/82	TX29/82	HK/83	GA/79/83	CHL/83	DUN/6/83	DUN/27/83	VIC/83	A.A.No.
43			G		A															-4
50					G									A						
51	A	A	A	A	G															-1
59		G			A															
66					C									T						5
68					T											C	C	C		
69					T									A						6
71	C				A															
110	A				G															
116	T		T		C															
125					C				T		T	T								
139					T									C						29
146	G				A															
152	T	T	T	T	C															
173	A	A	A	A	C															
182	C	C	C	C	T															
185	A	A	A	A	G															
190					G										A	A	A	A	A	46
191	C	C	C	C	T															
197	T	T	T		C															
200	G	G	G	G	A											G	G	G		
204	T	T	T	T	C															
211					G										A	A	A	A	A	53
213	T	T		T	C															
217					A												T			54
223			A		T															56
227					C											G	G	G		
242					G									A						
245	C				A															63
248	T	T	T	T	C															
250					A		G	G	G	G	G	G	G	G	G	G	G	G		65
264	G	G	G		A															70
269				G	A															
284					C				T		T	T								
291	T	T	T	T	C															
294	C	C	C	C	T															79
296	C			C	T															
299	C	C	C		T															
302	T	T	T	T	G															81
304	G	G	G	G	A															82
338		T	T		T	C														
347	G	G	G		A															
349		T			C															96
353	C				T															
362	G				A			G	G	G	G	G	G	G	G	G	G			
363	G	G	G	G	T															101
364				T	A															101
374				T	C															
380					G												A	A	A	
384	C	C	C	C	C	T	C	C	C	C	C	C	C	C	C	C	C	C		
392					G										A	A	A	A	A	
401					C	T	T	T	T	T	T	T	T	T	T	T	T	T		
413					A				T											
446				G	A			T	T	T	T	T	T	T	T	T	T	T		125a
449					A						G	G								
458	C				A											C	C	C		129
465	A	A	A	A	G		A													132
466	C	C	C		T															132
469		T			C															133
470					C									T						
472					G									A	A	A	A	A	A	134
476	A	A	A	A	A	C	A	A	A	A	A	A	A	A	A	A	A	A		
479	G	G	G		A															
486	G	G	G	G	G	T									G					138
491		T			C															
492	C				T															140
497			C		T															
498	G	G	G	G	A															142
499	C	C		C	A															142
500					G							T	T							142
501	A	A	A	A	G															143
502				A	G															143
520	A	A	A	A	G															149
524					T		C													

BASE NO.	DEN/57	QSL/54	FLM/52	ENG/51	FW/50	USSR/77	LACK/78	BRZ/78	ENG/80	GA/114/83	IND/80	TX/12/82	TX29/82	HK/83	GA/79/83	CHL/83	DUN/6/83	DUN/27/83	VIC/83	A.A.No.
527		A				G														
528	G	G	G			C														152
530	C	C	C	C		A														
543					A	G		A	A	A	A	A	A	A	A	A	A	A	A	157
544	C	C	C	C		A														157
554	A	A	A	A		G														
561			G			A														163
562			C			A		G												163
563						T												G		163
571	G					A														166
587	T	T	T			C														
588	C	C	C			A														172
602						C													T	
605						T													C	
608	G	G	G	G		A	C													
617	A					T														
632		C				T														
639						G	A								A	A			A	189
642						G						A	A							190
644	A	A	A	T		C														190
649			G			A												G	G	192
651	G					A														193
653	A					C														
654	C	C	C	C		A														194
661			A			G														196
662						G		T	T	T	T	T	T	T	T	T	T	T		
665						A									C					197
668	T	T	T			A														198
686				G		A														
696						A		C	C	C	C	C	C	C	C	C	C	C		208
726						G			T											218
729	A	A	A			G		A	A	A	A	A	A	A	A	A	A	A		219
732						A			C		C	C								
734	G	G	G			A														
747						G									A	A	A			225
748	A	A	A			G	A			A	A	A	A	A	A	A	A	A		225
753	T	C	C			G														227
754						C		A	A	A	A	A	A	A	A	A	A	A		227
758						G									A					
764		G	G			T														230
770	T	T		T		C														
780	T	T	T	T		C														
785	A	A	A	A		G														
791					T	C														
794	A		A	A		G														
799				A		C														242
817	C	C	C			A														248
824				C		T														
833	A	A	A			G														
840	T	T	T	T		C	T	T	T	T	T	T	T	T	T	T	T	T		256
856	G	G	G	G	G	A	G	G	G	G	G	G	G	G	G	G	G	G		260a
864	C	C	C	C		T														263
865	C	C	C			T														263
869	A	A	A	A		G														
875	G				G	A														
878	A					C														
894	C	C	C			T														273
896	A	A	A	A		G					A									
897	C					A														274
912						A		G	G	G	G	G	G	G	G	G	G	G		279
914	A					G														
953	C	C	C	C	C	T														
959	T	T	T	T		C										T	T	T	T	
965				C		T														
966						A	G	G	G	G	G	G	G	G	G	G	G	G		297
974	T					A														
986	A	A	A	A	A	G	A	A	A	A	A	A	A	A	A	A	A	A		
998						A					G		G	G	G	G	G	G		
1001				T		C							T	T						
1013	G			C		A														
1019						G	A	A	A	A	A	A	A	A	A	A	A	A		
1022		A				G														
1053	G					A														326
1064	G	G	G			A														
1068						C			T	T	T	T	T	T	T	T	T	T		

[a] Only positions varying from USSR/77 are shown. Where a gap is left, the sequence is identical to USSR/77, except for residues 43-321 of FW/50 and LACK/78 which were not determined. Residues 459-461 are deleted in DEN/57 causing the loss of a single amino acid residue corresponding to amino acid 130 of USSR/77. The position of amino acid changes is identified in the last column and the nature of the change is shown in Fig. 1.

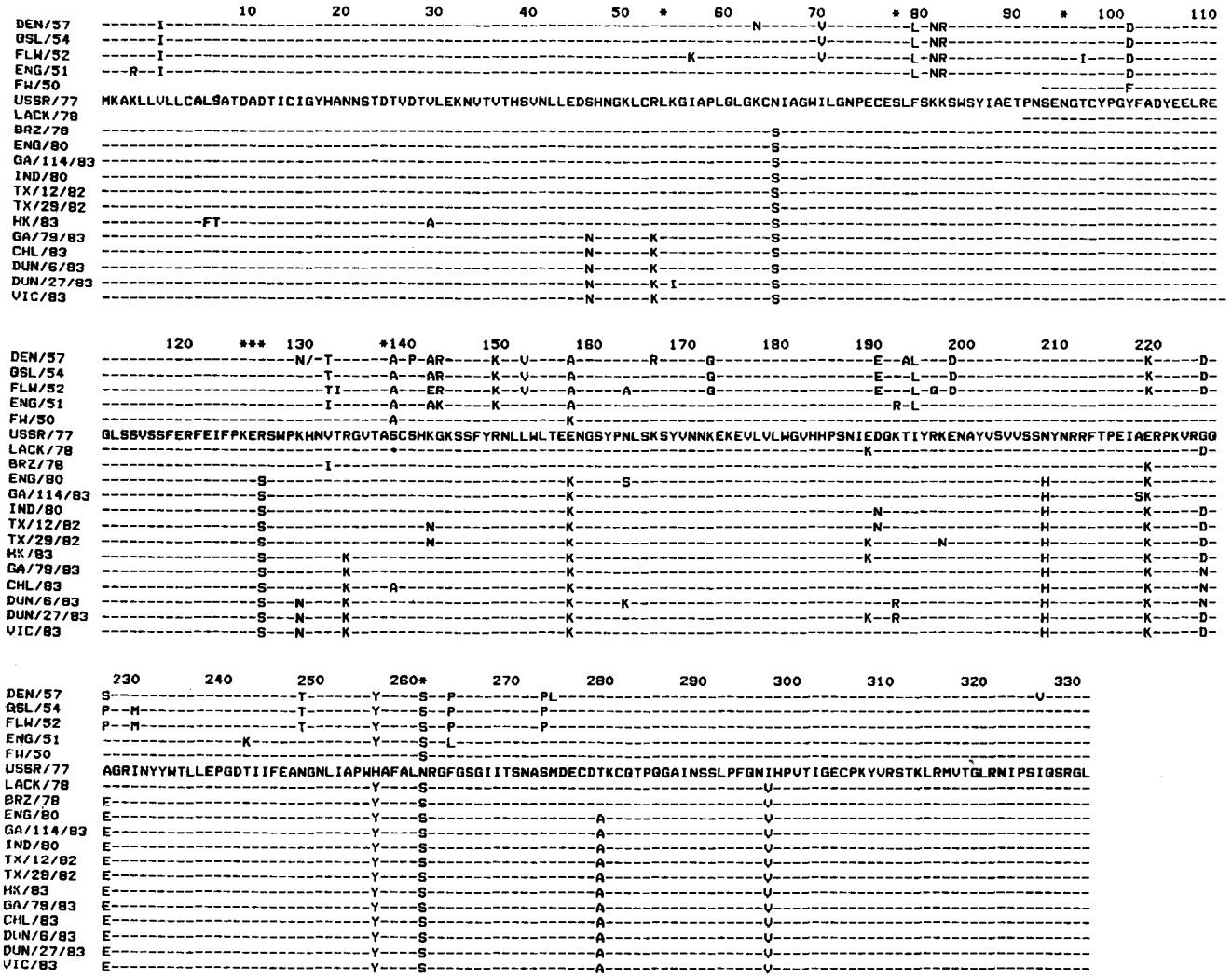
10 20 30 40 50 * 60 70 * 80 90 * 100 110
DEN/57
GSL/54
FLW/52
ENG/51
FW/50
USSR/77 MKAKLLVLLCALSATDADTICIGYHANNSTDTVDTVLEKNVTVTHSVNLLEDSHNGKLCRLKGIAPLQLGKCNIAGWILGNPECESLFSKKSWSYIAETPNSENGTCYPGYFADYEELRE
LACK/78
BRZ/78
ENG/80
GA/114/83
IND/80
TX/12/82
TX/29/82
HK/83
GA/79/83
CHL/83
DUN/6/83
DUN/27/83
VIC/83
120 *** 130 *140 150 160 170 180 190 200 210 220
DEN/57
GSL/54
FLW/52
ENG/51
FW/50
USSR/77 GLSSVSSFERFEIFPKERSWPKHNVTRGVTASCSHKGKSSFYRNLLWLTEENGSYPNLSKSYVNNKEKEVLVLWGVHHPSNIEDQKTIYRKENAYVSVVSSNYNRRFTPEIAERPKVRGQ
LACK/78
BRZ/78
ENG/80
GA/114/83
IND/80
TX/12/82
TX/29/82
HK/83
GA/79/83
CHL/83
DUN/6/83
DUN/27/83
VIC/83
230 240 250 260* 270 280 290 300 310 320 330
DEN/57
GSL/54
FLW/52
ENG/51
FW/50
USSR/77 AGRINYYWTLLEPGDTIIFEANGNLIAPWHAFALNRGFGSGIITSNASMDECDTKCQTPQGAINSSLPFGNIHPVTIGECPKYVRSTKLRMVTGLRNIPSIQSRGL
LACK/78
BRZ/78
ENG/80
GA/114/83
IND/80
TX/12/82
TX/29/82
HK/83
GA/79/83
CHL/83
DUN/6/83
DUN/27/83
VIC/83

TABLE 4

POSITION OF AMINO ACID MUTATIONS IN THE HAEMAGGLUTININ TRIMER AND THEIR RELATIONSHIP TO PREVIOUSLY NOTED ANTIGENIC SITES[a]

Buried residues	Surface residues	Antigenic site[b]	Buried residues	Surface residues	Antigenic site
5				163[c]	Sa
6				166	Sa
29				172	Ca?
	46			189	
	53			190	
	54			192	Sb
	56			193	Sb
	63			194	Sb?
65				196	Sb
70				197	Sb?
	79	Cb		198	Sb
	81	Cb		208	Ca?
	82	Cb		218	
	96[c]			219	
	101			225	Ca
	125c			227	
	129	Sa	230		
	132			242	
	133[c]		248		
	134		256		
	138			260a	
	140	Ca		263	
	142	Ca?		273[c]	
	143	Ca		274	
	149			279	
152			297		
	157	Sa	326		

[a] USSR/77 was taken as the parent strain.

[b] If the position is at, or very close (indicated by a ?) to a previously defined antigenic site, this site is marked.

[c] Causes loss of carbohydrate.

1983). There are two distinct pathways, each characterised by a distinctive set of "mainstream" mutations, which become "fixed" in subsequent strains, and "side-stream" mutations, which are usually specific to that strain. Considering the later

FIG. 1. The deduced amino acid sequence of the HA1 region of the haemagglutinin of 19 H1 subtype influenza strains. The amino acid sequence is written in full for USSR/77 using the single-letter amino acid code. Where the amino acid sequences of the other 18 strains are identical to that of USSR/77, they are marked by a dash (—) and where different, the altered amino acids are indicated. A slash (/) at residue 130 for DEN/57 indicates the amino acid deletion at this residue. The blank region at the N terminus of the sequence of FW/50 and LACK/78 was not determined. The amino acid sequences are numbered according to the alignment of Winter *et al.*, 1981. Additional amino acid residues present in the H1 subtype sequence but absent in the H3 subtype strains are marked by an asterisk (*). These additional residues are numbered by reference to the previous numbered residue followed by a postscript, e.g., 53a, 77a, etc.

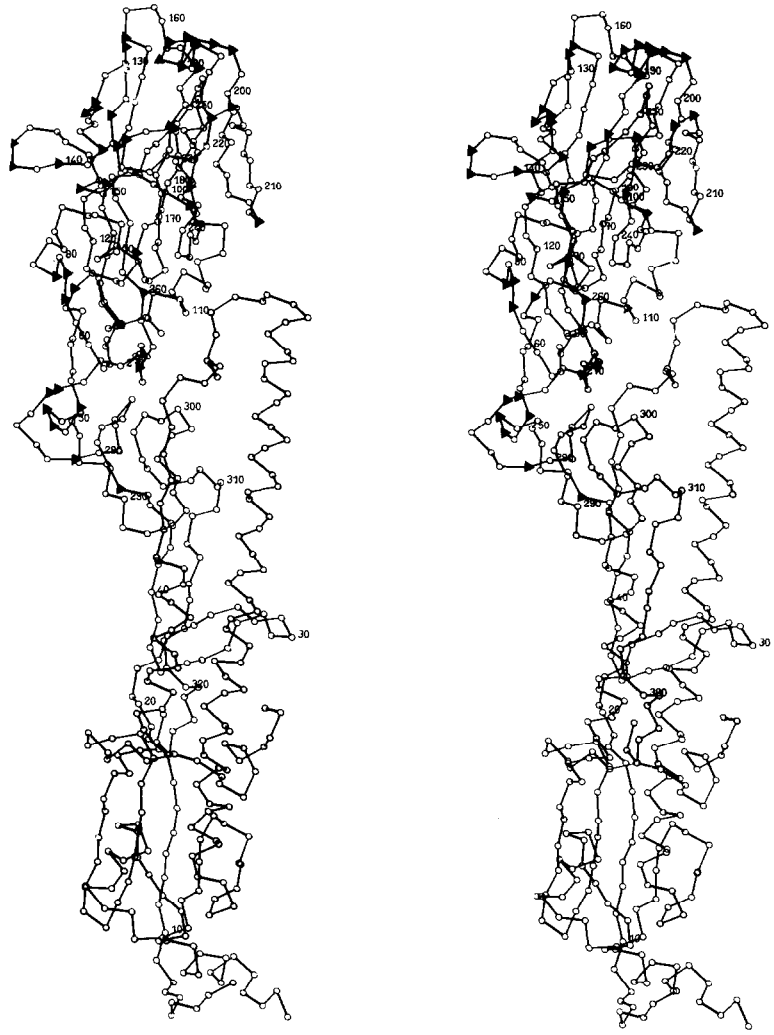

FIG. 2. Stereodiagram showing the position of superficial amino acid changes in the evolution of the HA1 domain region of the haemagglutinin of H1 viruses in the period 1950–1983 relative to strain USSR/77. These changes are marked on a side view of the complete haemagglutinin molecule beside the α carbon positions of the respective amino acids by the symbol ▶ and tend to concentrate into discrete areas (see text). The lower part of the figure is formed from the N-terminal region of the HA1 domain along with the HA2 domain of the protein.

1977–1983 pathway first (Figs. 1 and 3), we see at what point the hypervariable amino acids (163, 189, 190, and 225) arose. The 163, 189, and 190 mutations are sporadic and are unrelated to the mainstream, whilst residue 225 has the unique property of being both a mainstream and sidestream mutation. This arises because it mutated from G → D (codons GGT → GAT) early in the 1977–1983 pathway in a minor variant LACK/78, whilst the mainstream of viruses until 1980 had 225G, until a mutation occurred to D just before the emergence of IND/80. Subsequently 225 has mutated again from D → N (codons GAT → AAT) once at a branch leading to

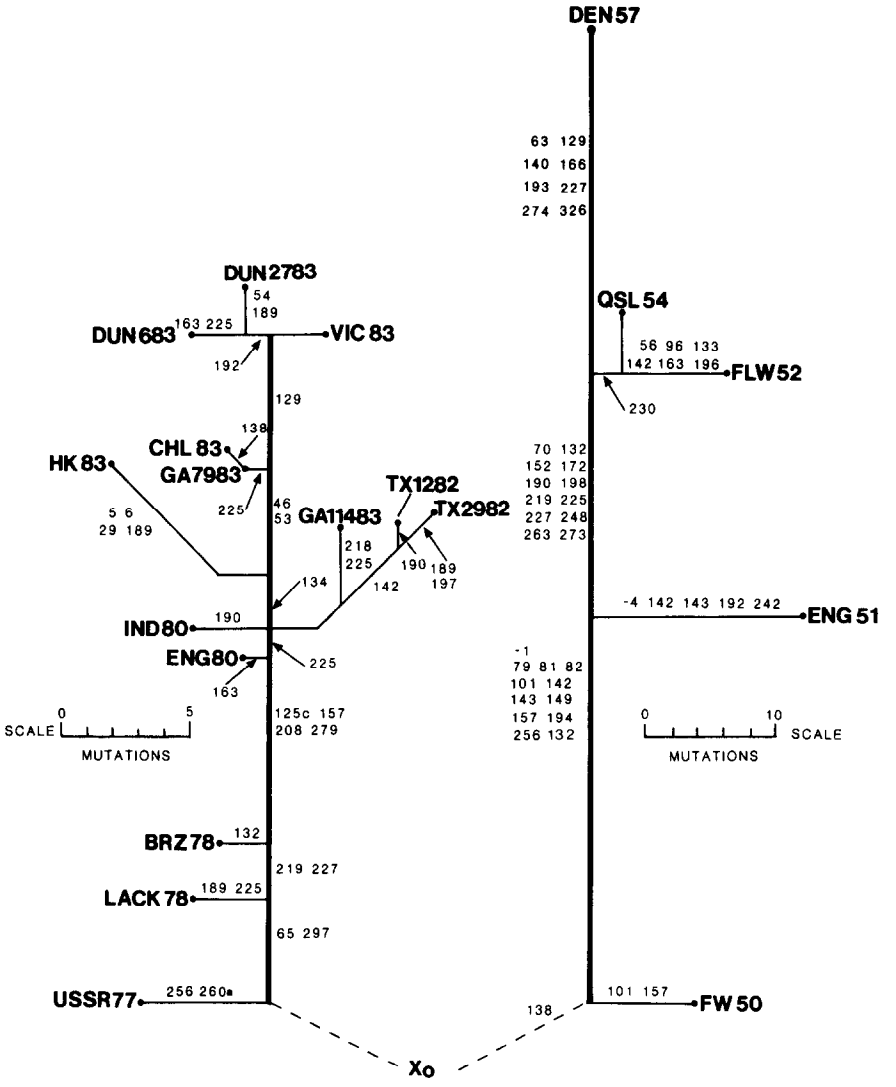

FIG. 3. The divergent evolutionary pathway of the 1950-1957 and 1977-1983 strains deduced from the nucleotide changes, encoding the HA1 domain of the haemagglutinin. The two main branches are drawn to two different scales and derive from the common precursor X_0. Distances between strains are the minimal mutational distances, including both coding and silent changes. Numbers refer to the amino acid residues which become "fixed" in all subsequent strains when these residues are on the mainstream (bold line); or to strain-specific amino acid residues on the branches. Numbers mentioned twice are amino acids which undergo independent (e.g., 163) or sequential (e.g., 225) changes.

GA/79/83 and CHL/83 and a second time in DUN/6/83 (see Figs. 1 and 3). It has also back mutated from AAT → GAT in GA/114/83. One 1983 strain, GA/114/83, seems to be a relic of the 1980-1982 era deriving from an earlier branch point leading to the two 1982 strains (Fig. 3). HK/83 is also representative of earlier 1983 strains and has diverged more than any other 1983 strain from the main evolutionary line.

The analysis of the earlier 1950–1957 pathway confirms the close relationship between USSR/77 and FW/50 (Young *et al.*, 1979), which differ only at amino acids 101, 138, 157, and 260a. The five 1950–1957 strains are very obviously different from one another in many positions and only QSL/54 and FLW/52 show any reasonably close relationship (see Fig. 3). This is consistent with the antigenic analysis (Table 2) with ferret sera. Indeed, there are 33 amino acid changes between FW/50 and the last member of the series—DEN/57. The rapid rate of evolution in the 8-year period 1950–1957 (1.2% amino acid change/HA1 domain/year) contrasts with the slower rate in the 7-year period 1977–1983. Here, only 18 amino acid changes separate the two strains, USSR/77 and DUN/27/83, which are furthest apart (0.8% amino acid change/HA1 domain/year). For comparison, Both *et al.* (1983) report 0.7% amino acid change/HA1 domain/year in H3 strains between 1968 and 1979.

DISCUSSION

It is commonly supposed that new influenza outbreaks arise by antigenic drift through natural selection imposed by population immunity. Our previous view of the antigenicity of the haemagglutinin of influenza strains of H1 subtype was derived from the study of such antigenically altered laboratory mutants of A/PR/8/34. This can now be related to the sequence of the HA1 domain of the haemagglutinin of 19 field strains of H1 subtype which were isolated from clinical cases between 1950 and 1983. The fact that 11 of the 54 amino acid changes (taking USSR/77 as the standard sequence) were in "buried" residues in the haemagglutinin trimer and in positions unavailable to direct antibody selection pressure, suggests that not all mutations are related to antigenicity. The mainstream amino acid mutations plot the evolutionary mainline, without interference from the specific adaptative features or "noise" in any one strain, and approximately 50% of changes are in positions other than recognised antigenic sites in both pathways (see Fig. 3 and Table 4). These additional residues may either represent antigenic regions of the molecule which differed from or were not detected in the previous antigenic analysis of influenza A/PR/8/34, or may represent neutral mutations, or may be advantageous mutations for functions of the virus unrelated to its antigenicity (see for example Daniels *et al.*, 1985).

Growth of influenza virus in chick eggs can introduce artefactual mutations probably in surface positions of the haemagglutinin, which affect binding to antibody and to cell receptors (Schild *et al.*, 1983; Robertson, personal communication). Such artefact mutations could affect the antigenicity in laboratory tests (Table 2) and could therefore artificially link strains which in fact are not closely related. Thus it seems possible that the similarity in haemagglutinin tests with ferret sera of HK/83, TX/29/82, and DUN/27/83 is related to an E → K mutation at residue 189 common to these three evolutionarily distinct (see Fig. 3) strains. The same rationale applies to residue 190 in IND/80 and TX/12/82 which are evolutionarily distinct but antigenically related. The artefactual nature of such changes as well as those at residue 163 explains their sporadic occurrence, but does not affect the validity of earlier conclusions based on mainstream mutations. Indeed, changes at 190 may not always be artefactual as this position mutates in the mainstream in the 1950–1957 pathway. Nakajima *et al.* (1983) also noted that laboratory mutants of USSR/77 at positions 189 and 190 were antigenically easily distinguishable from the parent USSR/77 virus.

The relative significance of the four antigenic sites defined in A/PR/8/34 probably differs in the recent field strains. The evidence for variation at the Sb site at the top of the molecule is particularly clear in the 1950–1957 pathway where, for example, residues 194 and 198 change and become fixed. Surprisingly, fixation of changes at

this site was not observed in the 1977–1983 pathway. On the other hand, the significance of the Sa site, residues 129 and 160–167, in recent field strains seems uncertain. Most have a potential carbohydrate attachment site at residue 163 which is lacking in A/PR/8/34 and this could mask some or all of the Sa antigenic site, explaining the relative lack of variation here. A similar situation exists in the haemagglutinin of H3 subtype where carbohydrate is attached to residue 165, effectively masking this region of molecule (Caton *et al.*, 1983; Both *et al.*, 1983). Evidence for the significance of the Cb antigenic site in the recent field strains is also uncertain. Although the Cb site residues 79, 81, and 82 differ in the four 1950–1957 strains (FW/50 was not sequenced in this region) from USSR/77, all four strains are identical in this region (see Fig. 1) and all have a potential carbohydrate attachment site at N(81). Carbohydrate, therefore, could well mask this site also. In addition, no mutations occur in this site in the 1977–1983 pathway. We conclude that if the Cb site has antigenic significance in the evolution of viruses in the field, variation at this position was only significant in the very early evolution of strains in the period from 1934 to 1950.

The Ca site of A/PR/8/34, which included the residues in the loop between 140 and 143 as well as those around residues 207, 225, 173, and 240 forming an antigenic cleft between subunits, clearly changes. Variation between residues 140 and 143 has occurred in both main branches of evolution but results in fixation of changes only in the 1950–1957 pathway. Variation is also seen at the remainder of the Ca site positions and 208 and 225 become fixed in the mainstream of the 1977–1983 pathway. Residue 225 changes in both evolutionary pathways, whereas 172 and 208 change in only one pathway.

The significance of the remaining superficial changes (Table 4) is uncertain but it seems reasonable to assign residues which are quite close to previously assigned antigenic sites, e.g., 138 (near Ca), and 227 (near Ca). More speculatively, we think that 218 and 219 might also be part of the Ca site. This leaves the changes at residues 132–134 and the hinge region (53–63 and 273–279) unaccounted for in terms of previously defined antigenic sites. Most field strains have carbohydrate attached at 131 and at 271 which could mask these regions from antibody depending on the exact disposition of the carbohydrate side chain. Therefore, for this region and the hinge region, we cannot distinguish whether these are new antigenic domains which were absent or not detected in the A/PR/8/34 laboratory studies, or whether they represent neutral mutations. Interestingly, a similar problem exists for the hinge region of the H3 subtype virus where field strains commonly show amino acid alterations (Both *et al.*, 1983), but only one laboratory antigenic variant at residue 54 is known (Laver *et al.*, 1979; Underwood *et al.*, 1984).

Many features of the evolution of the haemagglutinin of H1 viruses described here are similar to those described by Both *et al.* (1983) for viruses of H3 subtype. For example, both proceed by "fixation" of mainstream changes, although a major branch point occurs. In both there is some restriction in the repertoire of amino acids which can change, as exemplified by amino acids undergoing independent or sequential changes. The most obvious differences are in the rate of evolution of 1950–1957 strains, which is greater than that of either the 1977–1983 H1 strains or the 1968–1979 H3 strains (Both *et al.*, 1983). We suggest that this more rapid evolution in 1950–1957 correlates with circulation of the virus in the presence of a higher overall level of population immunity at that time, than in the subsequent 1977-present era, where virus circulated in younger people, many of whom had no preexisting antibody. High antibody pressure could have resulted in the emergence of highly specialised strains with many unique features, like DEN/57, which lacked the potential for further evolutionary change. It is unlikely that "laboratory drift" (change in sequence due to

passaging of viral stocks) is responsible for this higher rate of evolution of the 1950–1957 strains, as estimates derived from the influenza PB2 gene (Jones *et al.*, 1983) suggest these contribute less than 0.05% amino acid change per year.

In conclusion, the sequence of the haemagglutinin of the recent field strains of H1 subtype confirms the importance of two main antigenic areas—the Sb site, at the tip of the haemagglutinin spike, and the Ca site, somewhat lower down, previously characterised by studying laboratory mutants of A/PR/8/34. However, the other previously defined antigenic areas Sa and Cb are probably masked by carbohydrate, or do not vary sufficiently for there to be convincing evidence that they are major antigenic sites of general significance in the recent field strains. The mutations characteristic of the mainstream of the evolutionary pathway are not subject to individual features or "noise" present in any one strain, and differ (see Fig. 3) in the two main branches of H1 subtype viruses described here, except at two residues, 219 and 225, as well as from the mainstream mutations in H3 subtype viruses (Both *et al.*, 1983). These observations yet again demonstrate the flexibility and evolutionary potential of the influenza viral haemagglutinin, making it difficult to predict the existence or molecular nature of future epidemic strains.

REFERENCES

BOTH, G. W., SLEIGH, M. J., COX, N. J., and KENDAL, A. P. (1983). Antigenic drift in influenza virus H3 hemagglutinin from 1968 to 1980: Multiple evolutionary pathways and sequential amino acid changes at key antigenic sites. *J. Virol.* **48**, 52–60.

CATON, A., BROWNLEE, G. G., YEWDELL, J. W., and GERHARD, W. (1982). The antigenic structure of the influenza virus A/PR/8/34 hemagglutinin (H1 subtype). *Cell* **31**, 417–427.

CATON, A. J., RAYMOND, F. L., BROWNLEE, G. G., YEWDELL, J. W., and GERHARD, W. (1983). Antigenic variation in influenza virus. *Biochem. Soc. Trans.* **11**, 43–49.

CONCANNON, P., CUMMINGS, I. W., and SALSER, W. A. (1984). Nucleotide sequence of the influenza virus A/USSR/90/77 hemagglutinin gene. *J. Virol.* **49**, 276–278.

COX, N. J., and KENDAL, A. P. (1984). Genetic stability of A/Ann Arbor/6/60 cold-mutant (temperature-sensitive) live influenza virus genes: Analysis of oligonucleotide mapping of recombinant vaccine strains before and after replication in volunteers. *J. Infect. Dis.* **149**, 194–200.

DANIELS, R. S., DOWNIE, J. C., HAY, A. J., KNOSSOW, M., SKEHEL, J. J., WANT, M. L., and WILEY, D. C. (1985). Fusion mutants of the influenza virus hemagglutinin glycoprotein. *Cell* **40**, 431–439.

GERHARD, W., YEWDELL, J., FRANKEL, M. E., and WEBSTER, R. (1981). Antigenic structure of influenza virus haemagglutinin defined by hybridoma antibodies. *Nature (London)* **290**, 713–717.

JONES, K. L., HUDDLESTON, J. A., and BROWNLEE, G. G. (1983). The sequence of RNA segment 1 of influenza virus A/NT/60/68 and its comparison with the corresponding segment of strains A/PR/8/34 and A/WSN/33. *Nucleic Acids Res.* **11**, 1555–1566.

KENDAL, A. P., JOSEPH, J. M., KOBAYASHI, G., NELSON, D., REYES, C. R., ROSS, M. R., SARANDRIA, J. L., WHITE, R., WOODALL, D. F., NOBLE, G. R., and DOWDLE, W. R. (1979). Laboratory-based surveillance of influenza virus in the United States during the winter of 1977–78 I. *Amer. J. Epidemiol.* **110**, 449–461.

KENDAL, A. P., NOBLE, G. R., SKEHEL, J. J., and DOWDLE, W. R. (1978). Antigenic similarity of influenza A (H1N1) viruses from epidemics in 1977–1978 to "Scandinavian" strains isolated in epidemics of 1950–1951. *Virology* **89**, 632–636.

LAMB, R. A., and CHOPPIN, P. W. (1983). The gene structure and replication of influenza virus. *Annu. Rev. Biochem.* **52**, 467–506.

LAVER, W. G., AIR, G. M., WEBSTER, R. G., GERHARD, W., WARD, C. W., and DOPHEIDE, T. A. (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.

MCCAULEY, J. W., and MAHY, B. W. J. (1983). Structure and function of the influenza virus genome. *Biochem. J.* **211**, 281–294.

NAKAJIMA, S., NAKAJIMA, K., and KENDAL, A. P. (1983). Identification of the binding sites to monoclonal antibodies on A/USSR/90/77 (H1N1) hemagglutinin and their involvement in antigenic drift in H1N1 influenza viruses. *Virology* **131**, 116–127.

RAYMOND, F. L., CATON, A. J., COX, N. J., KENDAL, A. P., and BROWNLEE, G. G. (1983). Antigenicity and evolution amongst recent influenza viruses of H1N1 subtypes. *Nucleic Acids Res.* **11**, 7191–7203.

SCHILD, G. C., OXFORD, J. S., DE JONG, J. C., and WEBS-

TER, R. G. (1983). Evidence for host-cell selection of influenza virus antigenic variants. *Nature (London)* **303,** 706-709.

SPROAT, B. S., and BANNWORTH, W. (1983). Improved synthesis of oligodeoxynucleotides on controlled pore glass using phosphotriester chemistry and a flow system. *Tetrahedron Lett.* **24,** 5771-5774.

UNDERWOOD, P. A. *et al.* (1984). An antigenic map of the haemagglutinin of the influenza Hong Kong subtype (H3N2) constructed using mouse monoclonal antibodies. *Mol. Immunol.* **21,** 663-671.

WEBSTER, R. G., LAVER, W. G., AIR, G. M., and SCHILD, G. C. (1982). Molecular mechanisms of variation in influenza viruses. *Nature (London)* **296,** 115-121.

WILEY, D. C., WILSON, I. A., and SKEHEL, J. J. (1981). Structural identification of the antibody-binding sites of Hong Kong influenza hemagglutinin and their involvement in antigenic variation. *Nature (London)* **289,** 373-378.

WILSON, I. A., SKEHEL, J. J., and WILEY, D. C. (1981). Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3Å resolution. *Nature (London)* **289,** 366-373.

WINTER, G., FIELDS, S., and BROWNLEE, G. G. (1981). Nucleotide sequence of the haemagglutinin gene of a human influenza virus H1 subtype. *Nature (London)* **292,** 72-75.

YOUNG, J. F., DESSELBERGER, U., and PALESE, P. (1979). Evolution of human influenza A viruses in nature: Sequential mutations in the genomes of new H1N1 isolates. *Cell* **18,** 73-83.