Amino acid sequence and oligosaccharide distribution of the haemagglutinin from an early Hong Kong influenza virus variant A/Aichi/2/68 (X-31)

Colin W. WARD and Theo A. DOPHEIDE

CSIRO, Division of Protein Chemistry, 343 Royal Parade, Parkville, Victoria 3052, Australia

(Received 8 July 1980/Accepted 19 November 1980)

The amino acid sequence and oligosaccharide distribution for the haemagglutinin from the early Hong Kong influenza virus A/Aichi/2/68 (X-31) was investigated. The two polypeptide chains, HA1 and HA2, were fragmented by CNBr and enzymic digestion, and the amino acid sequence of each small peptide was deduced by comparing its chromatographic behaviour, electrophoretic mobility, amino acid composition and *N*-terminus with that of the corresponding peptide of the haemagglutinin of known structure from the influenza-virus variant A/Memphis/102/72. Those peptides in which changes were detected were sequenced fully. The complete amino acid sequence of the haemagglutinin HA1 chain (328 residues) and 188 of the 221 residues of the HA2 chain were established by this approach, and revealed only twelve differences between the amino acid sequences of variant-A/Aichi/68 and -A/Memphis/72 haemagglutinins. These occurred at positions 2, 3, 122, 144, 155, 158, 188, 207, 242 and 275 in the HA1 chain and 150 and 216 in the HA2 chain. The highly aggregated hydrophobic region (residues 180–121) near the *C*-terminal end of the HA2 chain was not resolved by peptide sequencing. The oligosaccharide distribution in variant-A/Aichi/68 haemagglutinin was identical with that found in that of A/Memphis/72, with sugar units attached at asparagine residues 8, 22, 38, 81, 165 and 285 in the HA1 chain and 154 on the HA2 chain. The monosaccharide compositions of the individual carbohydrate units on variant-A/Aichi/68 haemagglutinin differed from those of the corresponding units in variant-A/Memphis/72 haemagglutinin, and evidence was found for heterogeneity in the oligosaccharide units attached at single glycosylation sites.

Influenza remains a poorly controlled infection that frequently affects a large proportion of the population, largely because of the ability of the virus to undergo frequent and often dramatic antigenic change. Although both viral coat proteins (haemagglutinin and neuraminidase) can change, the haemagglutinin is considered the more important antigen because of its central role in the infection process (Drzeniek, *et al.*, 1966; Laver & Kilbourne, 1966; Klenk *et al.*, 1975; Lazarowitz & Choppin, 1975).

Two kinds of antigenic variation have been observed in influenza virus: major changes, termed 'antigenic shift', and minor changes, termed 'antigenic drift'. Since the first influenza virus was isolated from man in 1933 (Smith *et al.*, 1933), antigenic shifts have occurred in 1957 (Asian influenza), 1968 (Hong Kong influenza) and 1977 (Russian influenza). Since the appearance of the Hong Kong sub-type of influenza type A virus in man in 1968, rapid antigenic drift has occurred,

resulting in the appearance of new variants each year (Schild *et al.*, 1974; Pereira, 1976). Peptide 'maps' have shown that antigenic shift and drift are associated with major and minor differences respectively in the amino acid sequences of the haemagglutinin (Laver & Webster, 1968, 1972).

Previous reports have described the amino acid sequence of the haemagglutinin from the Hong Kong variant A/Memphis/102/72 as determined by protein (Dopheide & Ward, 1978, 1979, 1980; Ward & Dopheide, 1979, 1980; Ward *et al.*, 1980a) and nucleic acid (Sleigh *et al.*, 1980) procedures, and these data have been used to determine the partial amino acid sequences of several natural Hong Kong field strains (Laver *et al.*, 1980), and variants selected with monoclonal hybridoma antibodies (Laver *et al.*, 1979). In the present paper we describe our investigations of the amino acid sequence of the haemagglutinin from X-31, a recombinant virus containing the haemagglutinin from the early Hong Kong strain A/Aichi/68. We have determined the

structure to assist the characterization of the extent of chemical and antigenic change that has occurred in the Hong Kong sub-type during antigenic drift, and to provide the sequence data required for the interpretation of the 0.3 nm (3 Å)-resolution electron-density map derived from X-ray-crystallographic analysis of X-31 haemagglutinin crystals (Wiley & Skehel, 1977; Wilson *et al.*, 1980).

Materials and methods

Virus

The primary inoculum of the X-31 virus was kindly supplied by Dr. W. G. Laver, Australian National University, Canberra, Australia. Virus was grown in the allantoic sac of 11 day-old-chick embryos and purified as described by Laver (1969). A total of 4000 chick embryos yielded 600 mg of virus.

Haemagglutinin isolation and peptide fragmentation

The procedures for haemagglutinin isolation by electrophoresis on cellulose acetate blocks, the separation of the heavy (HA1) and light (HA2) chains by centrifugation on guanidine hydrochloride density gradients, reduction and *S*-carboxymethylation with iodo[2-¹⁴C]acetate, CNBr cleavage and the separation of the CNBr peptides have been fully described, as have the procedures for pyroglutamyl aminopeptidase (EC 3.4.11.8), trypsin (EC 3.4.21.4), thermolysin (EC 3.4.24.4), pepsin (EC 3.4.23.1) and *Staphylococcus aureus* proteinase (EC 3.4.21.19) digestion (Dopheide & Ward, 1978, 1979, 1980; Ward & Dopheide, 1979, 1980; Ward *et al.*, 1980a). The digestion of peptide HA2 CN1 (for nomenclature, see below) with trypsin (1% w/w) was extended for 48 h at 37°C in 0.1 M-*N*-ethylmorpholine/acetate, pH 8.0, with stirring; a further addition of enzyme was made after 24 h. Soluble tryptic peptides were initially fractionated on two coupled Sephadex G-50 (Fine grade) columns (150 cm × 0.9 cm each, in series) in 0.01 M-NH₄HCO₃/10% (v/v) propan-2-ol. Pooled fractions were further purified by high-voltage electrophoresis and, where necessary, chromatography in butanol/acetic acid/water/pyridine (15:3:12:10, by vol.) on Whatman 3 MM paper.

Analytical

Amino acid analyses were performed after hydrolysis in 5.7 M-HCl/0.004 M-thioglycolic acid at 108°C *in vacuo* for 24 h. Tryptophan-containing peptides were identified by staining side strips with Ehrlich reagent (Easley, 1965), and were hydrolysed in 4 M-methanesulphonic acid containing 0.2% tryptamine at 115°C for 24 h (Simpson *et al.*, 1976).

C-Terminal amino acids were detected by selective ³H-labelling (Holcomb *et al.*, 1968). Automated sequence degradation of the deblocked HA1 chain was performed with a protein program (Inglis *et al.*, 1979), and manual dansyl (5-dimethylaminonaphthalene-1-sulphonyl)-Edman degradations were performed as described previously (Ward & Dopheide, 1979). Glucosamine was determined on an amino acid analyser after hydrolysis in 3 M-*p*-toluenesulphonic acid at 100°C for 24 h (Allen & Neuberger, 1975). Neutral sugars were determined as alditol acetates by g.l.c. (Albersheim *et al.*, 1967). All samples (12–70 nmol) were hydrolysed in 2 ml of 2.5 M-trifluoroacetic acid at 100°C for 2 h in a sealed tube under N₂. After hydrolysis, the acid was removed by rotary evaporation and the hydrolyses reduced with NaBH₄ and acetylated (Albersheim *et al.*, 1967). The resulting alditol acetate derivatives were separated by g.l.c. on a column (1.5 m × 2 mm) of 3% SDP 2340 on Supelcoport 100/120.

Peptide nomenclature

Peptides have been assigned a prefix according to the digest from which they were recovered (CN, CNBr; T, tryptic; C, chymotryptic; Th, thermolytic; P, peptic; S.a., *Staphylococcus aureus* proteinase) and a number (1, 2, 3) according to the order in which they occur in the final sequence. Peptides resulting from partial cleavages, sharing common regions of sequence, are denoted 1a, 1b, 1c etc.

Results

In the present study we have fragmented the two chains (HA1 and HA2) of variant-X-31 haemagglutinin by CNBr cleavage and enzymic digestion and deduced the amino acid sequence of each peptide by comparing its chromatographic behaviour, electrophoretic mobility, amino acid composition and *N*-terminus with that of the corresponding peptide from the haemagglutinin of variant A/Memphis/102/72 (A/Mem/72), the structure of which is already known (see the introduction for references). For those peptides in which changes were detected, the amino acid sequence was determined directly.

Structure of the HA1 chain

CNBr cleavage of *S*-carboxymethylated HA1 (700 nmol) gave five peptides, CN1–CN5, which could be readily separated by gel filtration (Fig. 1). The large fragments CN1 (residues 1–168), CN2 (residues 169–260), and CN3 (residues 269–320) were digested with trypsin and the digests fractionated as shown in Figs. 2(a), 2(b) and 2(c). The large soluble tryptic peptides, CN1.T1b, CN1.T1c, CN1.T3 and CN3.T2, as well as the large insoluble

Fig. 1. Separation of CNBr peptides CN1–CN5 from *S*-[2- ^{14}C]carboxymethylated variant-X-31 haemagglutinin chain HA1

The digest was fractionated on a column (150 cm $\times$ 1.0 cm) of Sephadex G-100 in 50% (v/v) formic acid. Peptides were identified by A_{280} (—) and radioactivity in 10 μl samples (----). The fraction size was 2.1 ml, and the flow rate 4.2 ml/h.

tryptic peptides, CN1.T5, CN2.T2a and CN2.T10, were further digested with either thermolysin, *S. aureus* proteinase or pepsin.

The complete sequence of X-31 HA1 is shown in Fig. 3, along with the known sequence of A/Mem/72 HA1 and the location of the peptides used to deduce the sequence. Only ten sequence changes were found between X-31 and A/Mem/72 HA1 polypeptides. These occurred at positions 2, 3, 122, 144, 155 and 158 in peptide CN1, 188, 207 and 242 in peptide CN2 and 275 in peptide CN3. The amino acid compositions, electrophoretic mobilities and amino acid sequences for the peptides containing these changes are shown in Table 1. Direct sequenator analysis (40 cycles) of carboxymethylated HA1 (150 nmol), after the *N*-terminal pyroglutamic acid blocking group has been removed with calf-liver pyroglutamyl aminopeptidase, confirmed the suspected sequence changes at residues 2 and 3 in the blocked peptide CN1.T1b.Th1a (see Fig. 3).

The nucleic acid sequence for the haemagglutinin gene from A/Mem/72 (Sleigh *et al.*, 1980) showed that the 'Asx' residues at positions 137 and 250 in HA1 were asparagine not aspartic acid as found in the protein sequence (Ward & Dopheide, 1980; Ward *et al.*, 1980a) and that the sequence at residues 290–293 was Asn-Asp-Lys-Pro not Lys-Pro-Asp-Asp (Dopheide & Ward, 1978). As shown in Fig. 3, the X-31 sequence agrees with the A/Mem/72 nucleic-acid-sequence data. Peptides CN1.T5.Th5, CN2.T10.P3 and CN3.T2.Th4 were each recovered as both neutral and acidic (deamidated) versions, supporting the assignment of residues 137, 250 and 290 or 291 as asparagine. Manual sequence analysis

Fig. 2. Chromatography of tryptic digests of CNBr peptides (0.6 μmol , 18 h digestion) from variant-X-31 haemagglutinin chain HA1

Two columns (each 150 cm $\times$ 0.9 cm) of Sephadex G-50 (Fine grade) connected in series were used, and the columns eluted with 0.01 M-NH $_4$ HCO $_3$ /10% (v/v) propan-2-ol at a flow rate 4.4 ml/h and monitored by A_{230} (—) and radioactivity in 10 or 20 μl samples (----). Fractions (2.2 ml) were pooled as shown, and the peptides present at each fraction are indicated. (a) Tryptic peptides from *S*-[2- ^{14}C]carboxymethylated peptide CN1; (b) tryptic peptides from peptide CN2; (c) tryptic peptides from *S*-[2- ^{14}C]carboxymethylated peptide CN3.

of peptide CN3.T2.Th4 (100 mol) confirmed the sequence of residues 290–293 as Asn-Asp-Lys-Pro. All other peptides isolated from X-31 HA1 and indicated in Fig. 2 were identical with the corresponding peptides from A/Mem/72 in end group, composition and chromatographic behaviour (see Dopheide & Ward, 1978; Ward & Dopheide, 1980; Ward *et al.*, 1980a). Oligosaccharide units were

Fig. 3. Amino acid sequence of variant-X-31 haemagglutinin chain HA1

The alignment of the tryptic and appropriate thermolytic, peptic and *S. aureus*-proteinase peptides used to derive the sequences is shown and is based on comparison with the known amino acid sequence of A/Mem/102/72 haemagglutinin. The N-terminal 'Glx' residue is pyroglutamic acid; methionine residues occur in both strains at positions 168, 260, 268 and 320. Carbohydrate groups are attached in both strains at asparagine residues 8, 22, 35, 81, 165 and 285. The regions sequenced directly are underlined with dots. The residues that differ between the two haemagglutinins are enclosed in boxes.

Table 1. Amino acid composition and sequence of variant-X-31 haemagglutinin peptides that differed from the corresponding peptides in variant A/Mem/72

For the location of these peptides and their comparison with A/Mem/72, see Figs. 3 and 6. Abbreviations used: <Glu, pyroglutamic acid; Hse, homoserine; CmCys, carboxymethylcysteine.

Chain	Peptide	N-terminus	Mobility	Composition	Sequence	Position
HA1	CN1.T1b.Th1a	<Glu	−0.30	Asp _{2.7} Thr _{1.1} Ser _{0.8} Glu _{1.1} Pro _{1.1} Gly _{1.1} Leu _{0.9} Tyr _{0.8} GlcNAc _{2.1}	Glx-Tyr-Leu-Pro-Gly-Asn- Asp-Asn	1–10
	CN1.T5.Th3b	Ile		Thr _{0.9} Glu _{1.1} Gly _{1.1} Ile _{0.9}	Ile-Thr-Glu-Gly	121–124
	CN1.T6b	Gly	+0.32	Ser _{1.7} Pro _{1.0} Gly _{2.9} Phe _{2.2} Arg _{1.0}	Gly-Pro-Gly-Ser-Gly-Phe- Phe-Ser-Arg	142–150
	CN1.T7	Leu	+0.32	Asp _{1.0} Thr _{1.0} Leu _{2.0} Lys _{1.1} Trp	Leu-Asn-Trp-Leu-Thr-Lys	151–156
	CN1.T8	Ser	−0.13	Asp _{1.2} Thr _{2.1} Ser _{1.8} Glu _{1.0} Hse _{0.9} Pro _{1.1} Val _{1.9} Leu _{1.1} Tyr _{1.0} GlcNAc _{1.2}	Ser-Glu-Ser-Thr-Tyr-Pro-Val- Leu-Asn-Val-Thr-Met	157–168
	CN2.T2.Th3	Val	+0.22	Asp _{1.1} Thr _{2.0} Ser _{2.0} Glu _{3.2} Pro _{1.1} Val _{0.7} His _{1.7}	Val-His-His-Pro-Ser-Thr-Asn- Gln-Glu-Gln-Thr-Ser	182–183
	CN2.T5a	Val	+0.39	Thr _{2.0} Ser _{1.0} Val _{1.9} Arg _{1.1}	Val-Thr-Val-Ser-Thr-Arg	202–207
	CN2.T10.P3	Ile	0	Asp _{1.1} Pro _{1.1} Gly _{1.0} Val _{1.8} Ile _{0.8} Leu _{1.2} Lys _{0.9}	Ile-Val-Lys-Pro-Gly-Asp-Val- Leu	236–243
	CN3.T2-Th2	Ile	−0.89	CmCys _{1.0} Asp _{1.1} Thr _{0.9} Ile _{1.0}	Ile-Asp-Thr-Cys	274–277
HA2	CN1.T3	CmCys	−0.6	CmCys _{1.2} Asp _{2.0} Ser _{0.9} Glu _{1.1} Ala _{1.0} Ile _{1.8} Arg _{0.9}	Cys-Asp-Asn-Ala-Cys-Ile-Glu- Ser-Ile-Arg	144–153
	CN1.T8	Gly	+0.45	Asp _{1.0} Gly _{1.0} Ile _{0.8} Arg _{0.9}	Gly-Asn-Ile-Arg	213–216

Fig. 4. Separation of CNBr peptides CN1–CN4 from S-[2-¹⁴C]carboxymethylated haemagglutinin chain HA2 of variant X-31

For details, see the legend to Fig. 1.

found at asparagine residues 8, 22, 38, 81, 165 and 285.

Structure of HA2

CNBr cleavage of S-carboxymethylated HA2 (800nmol) yields four CNBr peptides, CN1–CN4, which can be separated by gel filtration as shown in Fig. 4. The N-terminal peptide CN4 was not investigated further as its sequence has already been published by Waterfield *et al.* (1979) and shown to

Fig. 5. Chromatography of tryptic digests of peptides (0.6μmol, 18 h digestion) from variant-X-31 haemagglutinin chain HA2

For details, see the legend to Fig. 2. (a) Tryptic peptides from S-[2-¹⁴C]carboxymethylated peptide CN1; (b) tryptic peptides from peptide CN2.

Table 2. *Carbohydrate composition of the seven oligosaccharide units on X-31 haemagglutinin*

The values in parentheses are the numbers of residues per mol of peptide. Glucosamine was determined on an amino acid analyser and is assumed to be *N*-acetylated. Neutral sugars were determined by g.l.c. as their alditol acetates. The two different forms of peptide CN1.T4 from HA2 had the same amino acid composition, but differed in electrophoretic mobility at pH6.5. The location of the peptides in the sequence of HA1 and HA2 is shown in Figs. 3 and 6.

Chain	Peptide	Asparagine residue	Sugar composition (nmol/30nmol of peptide)				Carbohydrate type
			Glucosamine	Mannose	Galactose	Fucose	
HA1	CN1.T1b.Th1a	8	129 (4.3)	160 (5.3)	77 (2.6)	51 (1.7)	Complex
	CN1.T1b.Th4	22	129 (4.3)	98 (3.3)	63 (2.3)	57 (1.9)	Complex
	CN1.T1c	38	63 (2.1)	187 (6.2)	17 (0.6)	7.7 (0.3)	Simple and complex
	CN1.T3.Th5a	81	115 (3.8)	102 (3.4)	93 (3.1)	27 (0.9)	Complex
	CN1.T8	165	56 (1.9)	258 (8.6)			Simple
	CN3.T2.Th4	285	58 (1.9)	132 (4.4)			Simple
HA2	CN1.T4a	154	117 (3.9)	113 (3.8)	69 (2.3)	44 (1.5)	Complex
	CN1.T4b	154	96 (3.2)	86 (2.9)	45 (1.5)	26 (0.9)	Complex

glutinin the carbohydrate unit as asparagine-38 was also of the *N*-acetyl-lactosamine type, whereas in X-31 it closely resembles the oligomannoside type in composition, with less than molar ratios of galactose and fucose. The tryptic peptide containing the single *N*-acetyl-lactosamine type oligosaccharide unit on HA2 was isolated in two different forms, T4a and T4b. T4b contains significantly less (one residue each) of glucosamine, mannose, galactose and fucose than T4a (Table 2).

Discussion

There has been considerable speculation over the mechanisms of antigenic shift and drift in influenza virus [see Fazekas de St. Groth (1978) and Webster & Laver (1975) for reviews] and the number and nature of antigenic determinants on the influenza-virus haemagglutinin (Laver *et al.*, 1974; Virelizier *et al.*, 1974; Gerhard, 1978; Russell *et al.*, 1979). Consequently, several laboratories have been engaged in establishing the chemical basis of antigenic variation in influenza virus. The current availability of monoclonal hybridoma antibodies against the influenza haemagglutinin (Gerhard, 1976) and the recent advances in the elucidation of the primary (Ward & Dopheide, 1980; Dopheide & Ward, 1980; Waterfield *et al.*, 1979) and tertiary (Wilson *et al.*, 1980) structures of the different haemagglutinins and the genes that code for them (Porter *et al.*, 1979; Gething *et al.*, 1980; Min-Jou *et al.*, 1980; Sleight *et al.*, 1980; Both *et al.*, 1980; Air, 1980) are beginning to establish the chemical changes associated with antigenic variation.

Partial sequence analysis of several Hong Kong (H3) variants selected with monoclonal hybridoma

antibodies (Laver *et al.*, 1979; Webster *et al.*, 1980) and several natural field strains isolated between 1968 and 1977 (Laver *et al.*, 1980) showed that the haemagglutinins from these antigenically distinct viruses could differ by as few as one or two amino acid residues. They also showed that the peptides that changed in the variants selected with monoclonal antibodies also changed in the natural field strains, although in almost all cases the changes that occurred were different. However, in these studies not all of the sequence was accounted for, and not all changes observed could be established with certainty. The insoluble peptides that comprise residues 110–141 and 229–255 in HA1 were not recovered and the large soluble peptides that account for residues 1–50, 58–90, 156–201 and 270–299 in HA1 were not reliably analysed by single hydrolyses. Similar gaps also occurred in the analysis of HA2. To fully characterize these variants, complete amino acid sequences are required.

Here we present the complete amino acid sequence (328 residues) of the haemagglutinin heavy chain (HA1) from the early Hong Kong variant X-31 and the sequence of 188 of the 221 residues from the light chain (HA2). The highly aggregated hydrophobic region (residues 180–212) near the C-terminal end of HA2 was not resolved by peptide sequencing. The sequence was found to be almost identical with that of A/Mem/72, including the location of the seven oligosaccharide units at asparagine residues 8, 22, 38, 81, 165 and 285 in HA1 and 154 in HA2. The results show that there are only ten sequence differences (at residues 2, 3, 122, 144, 155, 158, 188, 207, 242 and 275) in the HA1 polypeptides of X-31 and A/Mem-72 and two differences (residues 150 and 216) in HA2 (see Figs.

HA1		20	40	60	80	100	120
HK/1/68 (X-31)	QDLPCNDNSTATLCGLHHAV	PNGTLVKTIITDQIEVNTAT	ELVQSSSTGKICNNPHRILD	GIDCTLIDALLGPHCDVFC	NETWDLFVERSKAFSNCYPY	DVPDYASLSLVAASSCTLEF	
NT/60/68	QDLPCNDNSTATLCGLHHAV	PNGTLVKTIITDQIEVNTAT	ELVQSSSTGKICNNPHRILD	GIDCTLIDALLGPHCDVFC	NETWDLFVERSKAFSNCYPY	DVPDYASLSLVAASSCTLEF	
Aichi/2/68 (X-31)	QDLPCNDNSTATLCGLHHAV	PNGTLVKTIITDQIEVNTAT	ELVQSSSTGKICNNPHRILD	GIDCTLIDALLGPHCDVFC	NETWDLFVERSKAFSNCYPY	DVPDYASLSLVAASSCTLEF	
Mem/1/71	QDLPCNDNSTATLCGLHHAV	PNGTLVKTIITDQIEVNTAT	ELVQSSSTGKICNNPHRILD	GIDCT	SKAFSNCYPY	DVPDYASLR	
Mem/102/72	QDLPCNDNSTATLCGLHHAV	PNGTLVKTIITDQIEVNTAT	ELVQSSSTGKICNNPHRILD	GIDCTLIDALLGPHCDVFC	NETWDLFVERSKAFSNCYPY	DVPDYASLSLVAASSCTLEF	
Vic/3/75	QDLPCNDNSTATLCGLHHAV	PNGTLVKTIITDQIEVNTAT	ELVQSSSTGKICNNPHRILD	GIDCTLIDALLGPHCDVFC	NETWDLFVERSKAFSNCYPY	DVPDYASLSLVAASSCTLEF	
		140	160	180	200	220	240
HK/1/68 (X-31)	ITEGFTWTGVTQNGCSNACK	RGPCSGFFSRNLWLTGSGST	YPLVNVTPMNNDFDKLYIW	GIHHPSTNOEQTSLYVQASG	RVTVSTRSQQTIIIPNIGSR	PWVRGLSSRSISYWTIVKPG	
NT/60/68	ITEGFTWTGVTQNGCSNACK	RGPCSGFFSRNLWLTGSGST	YPLVNVTPMNNDFDKLYIW	GIHHPSTNOEQTSLYVQASG	RVTVSTRSQQTIIIPNIGSR	PWVRGLSSRSISYWTIVKPG	
Aichi/2/68 (X-31)	ITEGFTWTGVTQNGCSNACK	RGPCSGFFSRNLWLTGSGST	YPLVNVTPMNNDFDKLYIW	GIHHPSTNOEQTSLYVQASG	RVTVSTRSQQTIIIPNIGSR	PWVRGLSSRSISYWTIVKPG	
Mem/1/71	ITEGFTWTGVTQNGCSNACK	RGPCSGFFSRNLWLTGSGST	YPLVNVTPMNNDFDKLYIW	GIHHPSTNOEQTSLYVQASG	RVTVSTRSQQTIIIPNIGSR	PWVRGLSSRSISYWTIVKPG	
Mem/102/72	ITEGFTWTGVTQNGCSNACK	RGPCSGFFSRNLWLTGSGST	YPLVNVTPMNNDFDKLYIW	GIHHPSTNOEQTSLYVQASG	RVTVSTRSQQTIIIPNIGSR	PWVRGLSSRSISYWTIVKPG	
Vic/3/75	ITEGFTWTGVTQNGCSNACK	RGPCSGFFSRNLWLTGSGST	YPLVNVTPMNNDFDKLYIW	GIHHPSTNOEQTSLYVQASG	RVTVSTRSQQTIIIPNIGSR	PWVRGLSSRSISYWTIVKPG	
		260	280	300	320		
HK/1/68 (X-31)	DVLVINSNGNLIAPRGYFKM	RTGKSSIMRSDAPIDTCISE	CITPNGSIPNDKPFQNVNKI	TYGACPKYVKQNTLKLATCM	RNVPEKQT		
NT/60/68	DVLVINSNGNLIAPRGYFKM	RTGKSSIMRSDAPIDTCISE	CITPNGSIPNDKPFQNVNKI	TYGACPKYVKQNTLKLATCM	RNVPEKQT		
Aichi/2/68 (X-31)	DVLVINSNGNLIAPRGYFKM	RTGKSSIMRSDAPIDTCISE	CITPNGSIPNDKPFQNVNKI	TYGACPKYVKQNTLKLATCM	RNVPEKQT		
Mem/1/71	DVLVINSNGNLIAPRGYFKM	RTGKSSIMR	I	TYGACPKYVKQNTLKLATCM	RNVPEKQT		
Mem/102/72	DVLVINSNGNLIAPRGYFKM	RTGKSSIMRSDAPIDTCISE	CITPNGSIPNDKPFQNVNKI	TYGACPKYVKQNTLKLATCM	RNVPEKQT		
Vic/3/75	DVLVINSNGNLIAPRGYFKM	RTGKSSIMRSDAPIDTCISE	CITPNGSIPNDKPFQNVNKI	TYGACPKYVKQNTLKLATCM	RNVPEKQT		
HA2		20	40	60	80	100	120
HK/1/68 (X-31)	GLFGAIAAGFIENGWEGMIDG	WYGRFRHONSEGTQQAADLKS	TQAAIDQINGKLN RVIEKTN	EKPHQIEKEFSEVEGR IQDL	EKYVEDTKIDLWSYNAELLV	ALENQHTIDLT DSENNKLF	
NT/60/68	GLFGAIAAGFIENGWEGMIDG	WYGRFRHONSEGTQQAADLKS	TQAAIDQINGKLN RVIEKTN	EKPHQIEKEFSEVEGR IQDL	EKYVEDTKIDLWSYNAELLV	ALENQHTIDLT DSENNKLF	
Aichi/2/68 (X-31)	GLFGAIAAGFIENGWEGMIDG	WYGRFRHONSEGTQQAADLKS	TQAAIDQINGKLN RVIEKTN	EKPHQIEKEFSEVEGR IQDL	EKYVEDTKIDLWSYNAELLV	ALENQHTIDLT DSENNKLF	
Mem/1/71	GLFGAIAAGFIENGWEGMIDG	HONSEGTQQAADLKS	TQAAIDQINGKLN RVIEKTN	EKPHQIEKEFSEVEGR IQDL	EKYVEDTK	LFE	
Mem/102/72	GLFGAIAAGFIENGWEGMIDG	WYGRFRHONSEGTQQAADLKS	TQAAIDQINGKLN RVIEKTN	EKPHQIEKEFSEVEGR IQDL	EKYVEDTKIDLWSYNAELLV	ALENQHTIDLT DSENNKLF	
Vic/3/75	GLFGAIAAGFIENGWEGMIDG	WYGRFRHONSEGTQQAADLKS	TQAAIDQINGKLN RVIEKTN	EKPHQIEKEFSEVEGR IQDL	EKYVEDTKIDLWSYNAELLV	ALENQHTIDLT DSENNKLF	
		140	160	180	200	220	
HK/1/68 (X-31)	KTRRQLRENAEEMNGCFKI	YHKCDNACIESIRNGTYDHD	VYRDEALNNRFQIKGVELKS	GKDWILWISFAISCFLLCV	VLLGFIMMACQGNIRCNIC I		
NT/60/68	KTRRQLRENAEEMNGCFKI	YHKCDNACIESIRNGTYDHD	VYRDEALNNRFQIKGVELKS	GKDWILWISFAISCFLLCV	VLLGFIMMACQGNIRCNIC I		
Aichi/2/68 (X-31)	KTRRQLRENAEEMNGCFKI	YHKCDNACIESIRNGTYDHD	VYRDEALNNRFQIKGVELKS	GKDWILWISFAISCFLLCV	VLLGFIMMACQGNIRCNIC I		
Mem/1/71	KTRRQLR	I	YHKCDNACIESIR	FQIKGVELK	GNIR		
Mem/102/72	KTRRQLRENAEEMNGCFKI	YHKCDNACIESIRNGTYDHD	VYRDEALNNRFQIKGVELKS	GKDWILWISFAISCFLLCV	VLLGFIMMACQGNIRCNIC I		
Vic/3/75	KTRRQLRENAEEMNGCFKI	YHKCDNACIESIRNGTYDHD	VYRDEALNNRFQIKGVELKS	GKDWILWISFAISCFLLCV	VLLGFIMMACQGNIRCNIC I		

Fig. 7. Comparative amino acid sequences (single letter code) of haemagglutinin HA1 and HA2 chains from six different Hong Kong influenza A variants

The location of the amino acid substitutions are indicated (shaded). The sources for the sequence data are as follows: Hong Kong/1/68 (X-31), Verhoeyen *et al.*, 1980; NT/60/68, Both *et al.*, 1980; Aichi/2/68 (X-31), the present paper; Mem/1/71, Air, 1980; Mem/102/72, Ward & Dopheide, 1980; Dopheide & Ward, 1980; Sleight *et al.*, 1980; Vic/3/75, Min-Jou *et al.*, 1980. Additional partial sequence data is also available for Eng/87/69, Eng/187/70, Eng/42/72, Port Chalmers/1/73 and Texas/1/77 (Laver *et al.*, 1980).

3 and 6). Of the two differences found in HA2, that at position 216 does not occur in any other Hong Kong strain examined (Both *et al.*, 1980; Min-Jou *et al.*, 1980). These strains, in common with X-31, contain arginine at this position. With regard to the ten heavy-chain differences, three (at positions 144, 155 and 207) were detected in the comparative peptide-'map' analyses of Laver *et al.* (1980) and one (at residue 275) in the partial sequence analysis of Waterfield *et al.* (1979). The threonine/serine substitutions at positions 270 and 279 found by Waterfield *et al.* (1979) were not present in our isolate of X-31.

The results also show that although the distribution of oligosaccharide units on X-31 HA1 and HA2 and the amino acid sequences around these glycosylation sites are identical with that found in the haemagglutinin from A/Memphis/102/72, the sugar compositions for the homologous oligosaccharide units are not (see Table 2 and Ward *et al.*, 1980b). Similar heterogeneity of carbohydrate groups has been reported for many glycoproteins (see Gottschalk, 1969) and is illustrated in the present paper by the two different forms of the *N*-acetyl-lactosamine-type oligosaccharide on X-31 HA2. The nature of the attached carbohydrate groups at each position on X-31 HA1 and HA2 are the same as those on A/Memphis/102/72 except for that at asparagine-38. In A/Mem/72 this oligosaccharide was of the *N*-acetyl-lactosamine type with the composition $\text{Fuc}_{0.6}\text{Gal}_{2.5}\text{Man}_{5.4}\text{GlcNAc}_4$, whereas in X-31 it has the composition $\text{Fuc}_{0.3}\text{Gal}_{0.6}\text{Man}_6\text{GlcNAc}_7$. Since both *N*-acetyl-lactosamine- and oligomannoside-type carbohydrates are synthesized via a common mannose-rich intermediate (see Schacter *et al.*, 1979), this suggests that only partial processing of the precursor may have occurred. Alternative possibilities are that both *N*-acetyl-lactosamine and oligomannoside units are attached to the same locus on different haemagglutinin molecules, which seems unlikely (see Rosner *et al.*, 1980), or that the peptide containing this carbohydrate unit was contaminated with another *N*-acetyl-lactosamine-type glycopeptide (although this was not apparent from the end-group or amino acid analysis).

Antigenic analysis of Hong Kong field strains with a panel of monoclonal hybridoma antibodies (Webster & Laver, 1980) indicates that there were two antigenic variants of H3N2 viruses, Hong Kong/1/68 and Aichi/2/68, co-circulating in 1968, and both are referred to as supplying the haemagglutinin gene for the recombinant virus X-31 (Kilbourne, 1969; Min-Jou *et al.*, 1980). The antigenic relationships suggested that Hong Kong/1/68 gave rise to the early England strains (Eng/878/69, England/187/70), whereas A/Aichi/2/68 gave rise to the American isolate A/Mem/1/71. The comparative

sequences summarized in Fig. 7, in conjunction with the partial sequence data of Laver *et al.* (1980), confirm this suggestion. The X-31 haemagglutinin sequence described in the present paper is identical with the partial sequence data for A/Mem/1/71 (Air, 1980), except for residue-150 in HA2. It differs from the haemagglutinin-sequence data for both the X-31 strain examined by Verhoeven *et al.* (1980), and the early Australian isolate A/NT/60/68 (Both *et al.*, 1980) at positions 2, 31, 78 and 158 in HA1, and from the X-31 strain of Verhoeven *et al.* (1980), at position 132 in HA2.

The sequence homologies in conjunction with the antigenic data of Webster & Laver (1980) suggest that the X-31 haemagglutinin described in the present paper is that of A/Aichi/2/68, whereas that described by Verhoeven *et al.* (1980) may be A/Hong Kong/1/68.

The data in Table 7 also show that antigenic drift does not involve the sequential replacement, at a single locus, of hydrophobic amino acid residues of increasing side-chain surface area (Fazekas de St. Groth, 1978). Instead, as first reported by Laver *et al.* (1980), once an amino acid residue changed during antigenic drift, it did not change again in subsequent mutant generations. Rather, amino acid substitutions accumulated at other loci as antigenic drift proceeded.

We thank Mr. R. W. O'Brien for technical assistance, Dr. A. S. Inglis for the automated sequenator run and Dr. C. M. Roxburgh for the carbohydrate analyses and help with some amino acid analyses.

References

- Air, G. M. (1980) in *Structure and Variation in Influenza Virus* (Laver, G. & Air, G., eds.), pp. 135–146, Elsevier, New York
- Albersheim, P., Nevins, D. J., English, P. D. & Karr, A. (1967) *Carbohydr. Res.* **5**, 340–345
- Allen, A. K. & Neuberger, A. (1975) *FEBS Lett.* **60**, 76–80
- Both, G. W., Sleight, M. J., Bender, V. J. & Moss, B. A. (1980) in *Structure and Variation in Influenza Virus* (Laver, G. & Air, G., eds.), pp. 81–89, Elsevier, New York
- Dopheide, T. A. & Ward, C. W. (1978) *Eur. J. Biochem.* **85**, 393–398
- Dopheide, T. A. & Ward, C. W. (1979) *Virology* **92**, 230–235
- Dopheide, T. A. & Ward, C. W. (1980) in *Structure and Variation in Influenza Virus* (Laver, G. & Air, G., eds.), pp. 21–26, Elsevier, New York
- Drzeniek, R., Seto, J. T. & Rott, R. (1966) *Biochim. Biophys. Acta* **128**, 547–558
- Easley, C. W. (1965) *Biochim. Biophys. Acta* **107**, 386–388
- Fazekas de St. Groth, S. (1978) in *The Influenza Virus Hemagglutinin* (Laver, W. G., Bachmayer, H. & Weil, R., eds.), pp. 25–48, Springer-Verlag, Vienna

- Gerhard, W. (1976) *J. Exp. Med.* **144**, 985
- Gerhard, W. (1978) in *The Influenza Virus Haemagglutinin* (Laver, W. G., Bachmayer, H. & Weil, R. eds.), pp. 15–24, Springer-Verlag, Vienna
- Gething, M.-J., Bye, J., Skehel, J. J., Waterfield, M. D. (1980) in *Structure and Variation in Influenza Virus* (Laver, G. & Air, G., eds.), pp. 1–10, Elsevier, New York
- Gottschalk, A. (1969) *Nature (London)* **222**, 452–454
- Holcomb, G. H., James, S. A. & Ward, D. M. (1968) *Biochemistry* **7**, 1291–1296
- Inglis, A. S., Strike, P. M., Osborne, W. C. & Burley, R. W. (1979) *FEBS Lett.* **97**, 179–182
- Kilbourne, E. (1969) *Bull. W.H.O.* **41**, 643–645
- Klenk, H.-D., Rott, R., Orlich, M. & Blödorn, J. (1975) *Virology* **68**, 426–439
- Laver, W. G. (1969) in *Fundamental Techniques in Virology* (Habel, K. & Salzman, N. P., eds.), pp. 82–86, Academic Press, New York
- Laver, W. G. & Kilbourne, E. D. (1966) *Virology* **30**, 493–501
- Laver, W. G. & Webster, R. G. (1968) *Virology* **34**, 193–202
- Laver, W. G. & Webster, R. G. (1972) *Virology* **48**, 445–455
- Laver, W. G., Downie, J. C. & Webster, R. G. (1974) *Virology* **59**, 230–244
- Laver, W. G., Air, G. M., Webster, R. G., Gerhard, W., Ward, C. W. & Dopheide, T. A. (1979) *Virology* **98**, 226–237
- Laver, W. G., Air, G. M., Dopheide, T. A. & Ward, C. W. (1980) *Nature (London)* **283**, 454–457
- Lazarowitz, S. G. & Choppin, P. W. (1975) *Virology* **68**, 440–454
- Min-Jou, W., Verhoeyen, M., Devos, R., Saman, E., Huylebroeck, D., van Rompuy, L., Fang, R. X. & Fiers, W. (1980) in *Structure and Variation in Influenza Virus* (Laver, G. & Air, G., eds.) pp. 63–68, Elsevier, New York
- Pereira, M. S. (1976) in *Influenza Virus, Vaccines and Strategy* (Selby, P., ed), pp. 25–31, Academic Press, New York
- Porter, A. G., Barber, C., Carey, N. H., Hallewell, R. A., Threlfall, G. & Emtage, J. S. (1979) *Nature (London)* **282**, 471–477
- Rosner, M. R., Grinna, L. S. & Robbins, P. W. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 67–71
- Russell, R. J., Burns, W. H., White, D. O., Anders, E. M., Ward, C. W. & Jackson, D. C. (1979) *J. Immunol.* **123**, 825–831
- Schacter, H., Narasimhan, S. & Wilson, J. R. (1979) in *Glycoconjugate Research* (Gregory, J. D. & Jeanloz, R. W., eds.), pp. 575–596, Academic Press, New York
- Schild, G. C., Oxford, J. S., Dowdle, W. R., Coleman, M., Pereira, M. S. & Chakraverty, P. (1974) *Bull. W.H.O.* **51**, 1–11
- Simpson, R. J., Neuberger, M. R. & Liu, T. Y. (1976) *J. Biol. Chem.* **251**, 1936–1940
- Sleigh, M. J., Both, G. W., Brownlee, G. G., Bender, V. J. & Moss, B. A. (1980) in *Structure and Variation in Influenza Virus* (Laver, G. & Air, G., eds.), pp. 69–78, Elsevier, New York
- Smith, W., Andrews, C. H. & Laidlaw, P. P. (1933) *Lancet* **ii**, 66–68
- Verhoeyen, M., Fang, R., Min-Jou, W., Devos, R., Huylebroeck, D., Saman, E. & Fiers, W. (1980) *Nature (London)* **286**, 771–776
- Virelizier, J. L., Postlethwaite, R., Schild, G. C. & Allison, A. C. (1974) *J. Exp. Med.* **140**, 1559–1570
- Ward, C. W. & Dopheide, T. A. (1979) *Virology* **95**, 107–118
- Ward, C. W. & Dopheide, T. A. (1980) *Virology* **103**, 37–53
- Ward, C. W., Dopheide, T. A. & Inglis, A. S. (1980a) *Aust. J. Biol. Sci.* **33**, 137–151
- Ward, C. W., Gleeson, P. A. & Dopheide, T. A. (1980b) *Biochem. J.* **189**, 649–652
- Waterfield, M. D., Espelie, K., Elder, K. & Skehel, J. J. (1979) *Br. Med. Bull.* **35**, 57–63
- Webster, R. G. & Laver, W. G. (1975) in *The Influenza Viruses and Influenza* (Kilbourne, E. D., ed.), pp. 269–314, Academic Press, New York
- Webster, R. G. & Laver, W. G. (1980) *Virology* **104**, 139–148
- Wiley, D. C. & Skehel, J. J. (1977) *J. Mol. Biol.* **112**, 343–347
- Wilson, I. A., Skehel, J. J. & Wiley, D. C. (1980) in *Structure and Variation in Influenza Virus* (Laver, G. & Air, G., eds.), pp. 339–348, Elsevier, New York