Aust. J. Biol. Sci., 1980, **33**, 137–51

Amino Acid Sequence of Cyanogen Bromide Fragment CN2 from a Hong Kong Influenza Haemagglutinin Heavy Chain

Colin W. Ward, Theo. A. A. Dopheide and Adam S. Inglis

Division of Protein Chemistry, CSIRO, Parkville, Vic. 3052.

Abstract

The amino acid sequence of cyanogen bromide peptide CN2 from the heavy chain (HA_1) of the haemagglutinin of the Hong Kong variant A/Memphis/102/72 has been obtained by direct, automated sequence analysis on the whole fragment and by manual dansyl–Edman degradation of tryptic, peptic and chymotryptic peptides. It was found to contain 92 amino acid residues, including a large, insoluble, tryptic core peptide (residues 62–87). It did not contain any half-cystine residues or carbohydrate. The determination of its structure was complicated by the presence of an Asn-Ile bond at positions 48–49 which was readily cleaved by both trypsin and chymotrypsin.

Other keywords: comparative haemagglutinin sequences; glycosylation sites; sequence changes; antigenic variation.

Introduction

Antigenic change is a characteristic feature of influenza virus. Fresh outbreaks of influenza are usually associated with the appearance of variants that have haemagglutinin and sometimes neuraminidase proteins serologically different from those present in pre-existing strains. Two kinds of antigenic variation have been observed in influenza virus: major changes, termed antigenic shift, and minor changes, termed antigenic drift. Major changes in the haemagglutinin have occurred at irregular intervals (1889, 1918, 1957, 1968 and 1977; see Dowdle 1976), with minor antigenic changes every 1 or 2 years.

Peptide maps (Laver and Webster 1968, 1972) and recent studies on the primary structure of the haemagglutinins from Asian (Waterfield *et al.* 1979) or Hong Kong variants (Laver *et al.* 1979, 1980; Ward and Dopheide 1979*a*) have shown that these antigenic changes are associated with corresponding major (shift) or minor (drift) changes in amino acid sequence of the haemagglutinin.

To assist this characterization of the chemical changes associated with antigenic variation we are determining the amino acid sequence of the haemagglutinin from the Hong Kong variant A/Mem/102/72. We have shown that contrary to earlier reports the molecular weight of the haemagglutinin heavy chain (HA_1) is only 46 000–47 000 including approximately 24% by weight of carbohydrate (MW 11 500), giving an apoprotein molecular weight of 34 500–35 500 (see Ward and Dopheide 1976; Dopheide and Ward 1978*b*). Amino acid analysis based on this molecular weight indicated HA_1 contains approximately 321–331 amino acid residues including four

methionines. Cyanogen bromide cleavage of HA_1 at these methionine residues yields five peptides: CN1, CN2, CN3, CN4 and CN5. The amino acid sequence of the three peptides, CN4–CN3–CN5, which comprise the *C*-terminal 68 residues of HA_1 have been described previously (Dopheide and Ward 1978*a*, 1978*b*) as have the sequences of the peptides comprising the *N*-terminal 153 residues of the haemagglutinin light chain HA_2 (Dopheide and Ward 1979; Ward and Dopheide 1979*b*). This paper describes the amino acid sequence of the 92-residue cyanogen bromide fragment CN2 which extends from residue 169 to 260 in HA_1 from A/Memphis/72.

Materials and Methods

The virus used was a recombinant containing the haemagglutinin (H3) of A/Memphis/102/72 and the neuraminidase (N1) of A/Bel/42. The procedures employed in virus cultivation in 11-day chick embryos, virus purification following adsorption–elution from chick red blood cells, haemagglutinin isolation by electrophoresis on cellulose acetate blocks, and the separation of HA_1 and HA_2 by guanidine hydrochloride gradient centrifugation are described by Ward and Dopheide (1976).

Reduction, *S*-carboxymethylation, and the isolation of cyanogen bromide peptides were as described previously (Dopheide and Ward 1978*a*).

Tryptic and chymotryptic digestion of CN2 were carried out essentially as described by Ward and Dopheide (1979*b*), CN2 being first put into suspension by treatment with one-sixth final reaction volume of 3% (v/v) triethylamine. Peptic digestion of CN2 (300 nmol) was carried out in 5% formic acid at 37°C for 3 h using 100 μg pepsin. Thermolytic and *Staphylococcus aureus* protease digestion of tryptic peptide T10 was carried out essentially as described by Ward and Dopheide (1979*b*). Peptides were purified by gel filtration, high-voltage electrophoresis, or both (Ward and Dopheide 1979*b*). Peptides were assigned a prefix according to the digest from which they were obtained (T, tryptic; C, chymotryptic; Th, thermolytic; *S.a.*, *Staphylococcus aureus* protease; P, peptic) and a number (e.g. T1, T2, T3) according to the order in which these peptides occur in the final sequence. Peptides resulting from partial cleavages sharing common regions of sequence were denoted 1a, 1b, 1c etc.

Amino acid analysis, *C*-terminal amino acid analysis, and manual dansyl–Edman degradations were carried out as described previously (Ward and Dopheide 1979*b*). Automated sequencing of whole CN2 was done using a protein programme (Inglis *et al.* 1979) and that of the 25-residue chymotryptic peptide C7a was done using a peptide programme (Crewther and Inglis 1975). Derivatives of 3-phenyl-2-thiohydantoin (PTH) were identified by thin-layer chromatography (Inglis and Nicholls 1973), high-pressure liquid chromatography (Zimmerman *et al.* 1977) or amino acid analysis after back-hydrolysis with HI (Smithies *et al.* 1971).

Results

Tryptic Peptides of CN2

CN2 was digested with trypsin, the reaction mixture lyophilized and resuspended in 1·5 ml 0·01 M NH_4HCO_3–10% (v/v) isopropanol, and the insoluble material removed by centrifugation (50 000 *g* for 30 min). The soluble tryptic peptides were first fractionated by gel filtration (Fig. 1), the major peaks pooled as shown, and further separated by high-voltage paper electrophoresis at pH 6·5 and, when necessary, at pH 1·9. The major peptides present in the pooled fractions are shown in Fig. 1.

The insoluble tryptic material was fractionated by gel filtration on a single column of Sephadex G50 fine (150 by 0·9 cm) in 50% formic acid. It contained one major peptide, T10, as well as small amounts of partial-degradation products.

The amino acid composition, electrophoretic mobilities, yields and positions in the final sequence of the 13 tryptic peptides which comprise CN2 are shown in Table 1 and the amino acid sequences of these peptides in Table 2. T1 was the only peptide with *N*-terminal proline and was presumably the *N*-terminal peptide of CN2

(Dopheide and Ward 1978*b*). The major complication was the occurrence of substantial chymotryptic activity in the trypsin preparation used, which resulted in the production of several peptides lacking *C*-terminal basic residues, such as T2, T3, T6 and T7. The amino acid sequences of all tryptic peptides except T3 and T10 were determined without difficulty. For peptide T3, dansyl–Edman degradation gave the tentative sequence shown in Table 2 and selective tritiation established that Tyr was the *C*-terminal residue. No acid or amide assignments were obtained for positions 10, 11, 12 and 13. For the insoluble peptide T10, dansyl–Edman degradation gave the sequence of the first 10 residues, while thermolytic digestion gave four peptides whose amino acid compositions and sequences are shown in Table 1 and Table 2 respectively. Th1, Th2 and Th3 comprised the *N*-terminal 22 residues of T10, and Th4 accounted for the *C*-terminal four residues. Dansyl–Edman degradation gave the first 12 residues of peptide Th3 and selective tritiation established that Leu was the *C*-terminal residue. The complete sequence of T10 was established after the analysis of peptides produced by peptic digestion of whole CN2.

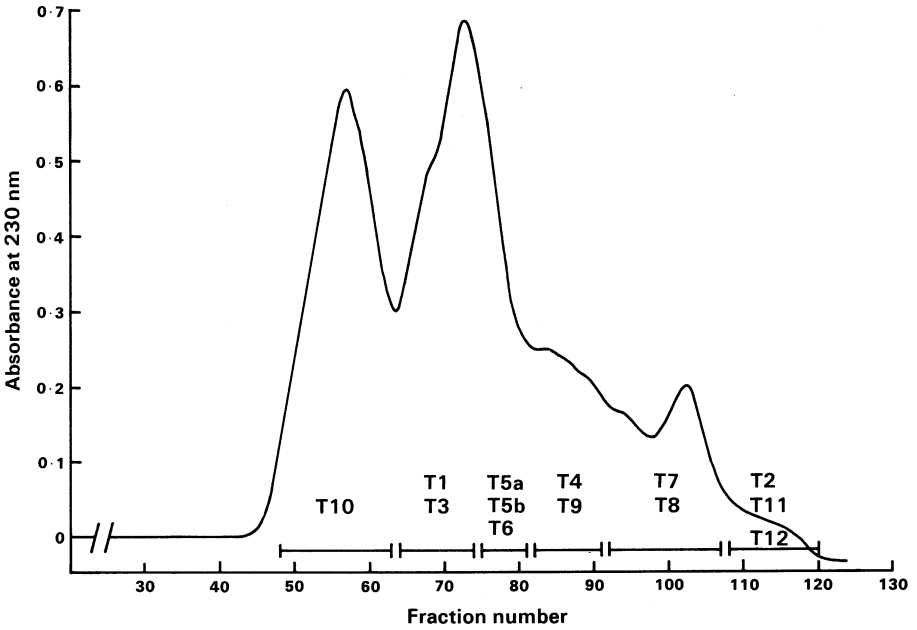

Fig. 1. Chromatography of tryptic digest of CN2 (0·7 μmol) on two Sephadex G50 fine and two Sephadex G15 superfine columns (each 150 by 0·9 cm) connected in series. The columns were eluted with 0·01 M NH_4HCO_3–10% (v/v) isopropanol at a flow rate of 4·4 ml/h. Fractions (4·4 ml) were pooled as shown and the major peptides present in each fraction are indicated.

Peptic Peptides of CN2

In view of the poor solubility of CN2 and the occurrence of several chymotryptic splits during tryptic digestion, CN2 was digested with pepsin in an attempt to obtain the overlaps required for the alignment of the final sequence. The peptic peptides were purified by gel filtration (Fig. 2) and paper electrophoresis at pH 6·5 and 1·9. Their amino acid compositions are shown in Table 3 and amino acid sequences in Table 4.

Table 1. Amino acid compositions of tryptic peptides from CN2 and thermolytic peptides from tryptic peptide T10

Values are expressed as residues per mole. Values in parentheses are those found in sequence. Hydrolyses were performed in 5·7 M HCl–0.004 M thioglycollic acid at 108°C for 24 h. No corrections for destruction or incomplete hydrolysis have been made. Electrophoretic mobilities (M) are expressed relative to aspartic acid (Offord 1966) for which $M = -1$. Yield represents nanomoles of purified peptide recovered from tryptic digestion of 700 nmole CN2 or 100 nmole T10

Amino acid	T1	T2	T3	T4	T5a	T5b	T6	T7	T8	T9	T10	T10 Th1	T10 Th2	T10 Th3	T10 Th4	T11	T12	CN2
Asp	4·8(5)		1·1(1)				1·0(1)				3·7(4)			3·8(4)				11·0(11)
Thr	0·1		1·7(2)		1·9(2)	1·8(2)	1·0(1)				1·0(1)		0·8(1)					5·5(6)
Ser	0·1		1·7(2)	0·9(1)	1·2(1)	1·0(1)	0·9(1)	0·8(1)		1·8(2)	1·8(2)	1·1(1)		1·0(1)				8·1(10)
Glu	0·2		2·9(3)	1·0(1)			2·2(2)				0·2							6·3(6)
Pro	1·2(1)		1·1(1)				1·2(1)	1·1(1)			2·1(2)			1·2(1)	1·0(1)			5·5(6)
Gly	0·1		1·0(1)	1·0(1)				1·0(1)		1·0(1)	2·8(2)		0·5	1·8(2)		1·0(1)		8·2(7)
Ala	0·1			1·0(1)							1·0(1)		0·3		0·9(1)			3·1(2)
Val			0·9(1)	1·0(1)	2·0(2)	1·9(2)			0·9(1)		1·5(2)			1·8(2)				5·6(7)
Ile			0·9(1)				1·6(2)	1·1(1)			5·2(6)	0·9(1)	0·9(1)	2·1(3)	1·1(1)			9·2(10)
Leu	0·2	1·0(1)	1·2(1)							1·1(1)	2·1(2)			1·8(2)				4·9(5)
Tyr	0·2	1·0(1)	1·0(2)								0·8(1)		1·0(1)			1·0(1)		3·6(4)
Phe	1·0(1)															1·0(1)		2·3(2)
His			1·6(2)															2·0(2)
Lys	0·9(1)				1·0(1)	1·0(1)					1·0(1)			1·1(1)		1·0(1)		4·3(4)
Arg				1·0(1)		0·9(1)		1·1(1)	1·1(1)	1·1(1)	1·0(1)				1·0(1)			5·3(5)
Trp			(1)					(1)			(1)		(1)					(3)
Hse[A]																	1·0(1)	0·9(1)
M	−0·57	0	−0·06	+0·46	+0·31	+0·47	0	+0·46	+0·78	+0·53	—	0	0	−0·21	+0·5	+0·42	0	
Yield	167	220	108	414	89	270	276	149	198	341	317	20	18	22	20	111	100	
S. P.[B]	1–8	9–10	11–27	28–33	34–39	34–40	41–48	49–54	55–56	57–61	62–87	62–63	64–67	68–83	84–87	88–91	92	

[A] Homoserine. [B] Sequence position.

The peptides recovered did not account for the complete sequence of CN2 but did allow some partial ordering of the tryptic peptides. Peptides P1, P2, P3 and P4 permitted the tryptic peptides T1, T2, T3 and T4, which make up the *N*-terminal

Table 2. Amino acid sequences of tryptic peptides from CN2 and thermolytic peptides from T10
Residue determined as the dansyl derivative, ⟶; residue determined as the PTH, ⟶; residue determined by amino acid analysis after back-hydrolysis with HI, ⇢; residue determined by selective tritiation, ⟵

Peptide	Amino acid sequence
T1	Pro Asn Asn Asp Asn Phe Asp Lys
T2	Leu Tyr
T3	Ile Trp Gly Val His His Pro Ser Thr Asx Glx Glx Glx Thr Ser Leu Tyr
T4	Val Gln Ala Ser Gly Arg
T5a	Val Thr Val Ser Thr Lys
T5b	Val Thr Val Ser Thr Lys Arg
T6	Ser Gln Gln Thr Ile Ile Pro Asn
T7	Ile Gly Ser Arg Pro Trp
T8	Val Arg
T9	Gly Leu Ser Ser Arg
T10	Ile Ser Ile Tyr Trp Thr Ile Val Lys Pro (Asx$_4$, Ser, Pro, Gly$_3$, Ala, Val, Ile$_2$, Leu$_2$)Arg
T10 Th1	Ile Ser
T10 Th2	Ile Tyr Trp Thr
T10 Th3	Ile Val Lys Pro Gly Asx Ile Leu Val Ile Asx Ser (Asx$_2$, Gly) Leu
T10 Th4	Ile Ala Pro Arg
T11	Gly Tyr Phe Lys
T12	HSer

of T10. Definite acid/amide assignments were obtained for residue 3 in P8b and 33 residues, to be aligned (see Fig. 4) and resolved the uncertainties in the acid/amide assignments of T3. Peptide P10a provided the overlap between T10 and T11, and in conjunction with peptides P7, P8a, P8b and P9 established the overall sequence

residue 3 in P9 and tentative evidence was obtained for residues 5 and 7. The drop in yield (observed qualitatively, dansyl procedure) after the fifth Edman cycle of peptide P9 suggested the cyclic imide rearrangement of an asparaginyl–glycine sequence at residues 5 and 6. If residue 5 is asparagine then, from the electrophoretic mobility of peptide P9, residue 7 must be aspartic acid. These assignments were verified, as mentioned below, by sequenator degradation of chymotryptic peptide C7a. The peptides which accounted for the other regions of CN2 were either not recovered or contained no new information (P5, P6a, P6b, P10b and P11).

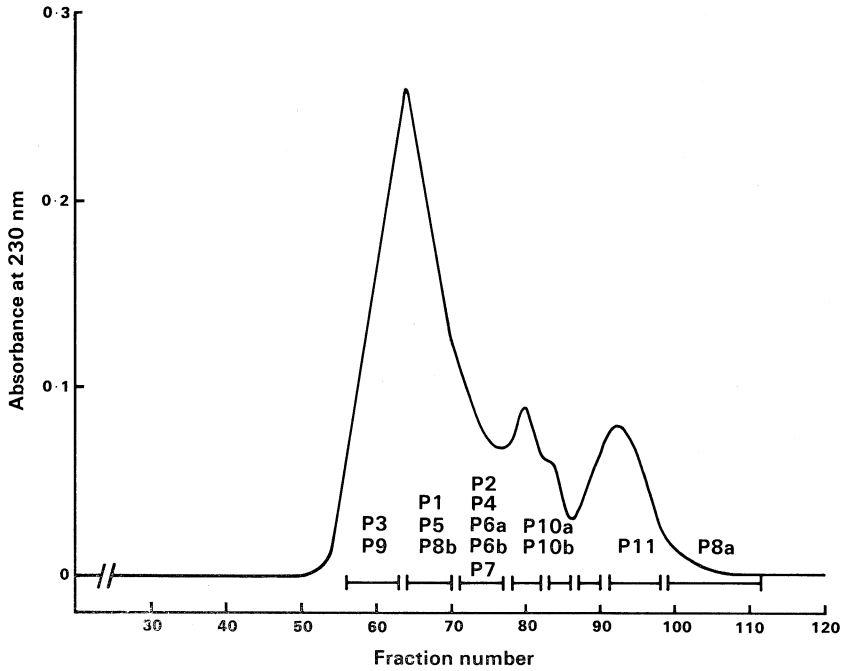

Fig. 2. Chromatography of peptic digest of CN2 (0·3 μmol) on four coupled Sephadex columns. Conditions were as described in the legend to Fig. 1.

Chymotryptic Peptides of CN2

The soluble peptides obtained after chymotryptic digestion of CN2 were separated by gel filtration (Fig. 3) and paper electrophoresis at pH 6·5 and 1·9. The insoluble fraction remaining after chymotryptic digestion was fractionated by gel filtration on Sephadex G50 (150 by 0·9 cm) in 50% formic acid to yield two major peptides. One was a large partial-digestion product of CN2 which accounted for residues 11–92 and the other was the 24-residue peptide C7a. The amino acid compositions of the 10 chymotryptic peptides are shown in Table 5 and their amino acid sequences in Table 6. Peptide C3 provided the overlap between T4 and T5a/b and, by amino acid composition, the overlap between T5a/b and T6. Peptide C5 provided the overlap between T8 and T9 while C6 provided the overlap between T9 and T10. C7a and C7b confirmed the alignment of T10 and T11 and automated sequenator analysis of C7a verified the acid/amide assignments for this region. C8 provided the final alignment of T11 with the *C*-terminal homoserine, T12. These peptides permitted

Table 3. Amino acid compositions of peptic peptides from CN2

Values are expressed as residues per mole. Yield represents nanomoles of purified peptide recovered after peptic digestion of 300 nmoles of CN2. For other details see legend to Table 1

Amino acid	P1	P2	P3	P4	P5	P6a	P6b	P7	P8a	P8b	P9	P10a	P10b	P11
Asp	4·7(5)	1·0(1)	1·2(1)						1·1(1)	1·0(1)	3·1(3)			
Thr			2·0(2)		1·8(2)				1·2(1)		0·2			
Ser			2·2(2)	1·1(1)	0·9(1)	2·0(2)	1·9(2)	1·0(1)			1·0(1)			
Glu			3·1(3)	1·1(1)							0·3			
Pro	1·1(1)		1·0(1)						1·0(1)	0·9(1)		1·0(1)		
Gly			1·3(1)	1·0(1)		0·9(1)	1·0(1)		1·1(1)	1·1(1)	1·2(1)	1·1(1)	1·0(1)	
Ala				1·0(1)							0·2	0·9(1)		
Val			1·1(1)	1·0(1)	2·1(2)				0·8(1)		1·0(1)			
Ile			0·9(1)					1·0(1)	1·6(2)	0·9(1)	0·9(1)	1·0(1)		
Leu		1·0(1)	1·0(1)			1·0(1)	1·0(1)		1·0(1)	1·1(1)	1·2(1)			
Tyr			1·1(1)	0·9(1)					0·9(1)			0·9(1)	1·0(1)	
Phe	1·0(1)												1·0(1)	0·9(1)
His			1·7(2)											
Lys		0·9(1)			1·0(1)				0·9(1)		0·2			1·1(1)
Arg				1·0(1)	1·0(1)		1·0(1)					1·2(1)		
Trp			0·7(1)						(1)					
M	−0·67	−0·47	−0·05	−0·34	+0·31	0	+0·53	0	0	−0·40	−0·31	+0·38	0	+0·56
Yield	30	75	82	80	44	110	21	42	34	30	57	28	14	20
S. P.[A]	1–6	7–9	10–26	27–33	34–40	57–60	57–61	62–63	65–75	71–75	76–83	84–89	88–90	90–91

[A] Sequence position.

the *N*-terminal 48 residues and the *C*-terminal 38 residues to be aligned with certainty (see Fig. 4) but did not provide the required overlaps into peptide T7.

Table 4. Amino acid sequence of some peptides from CN2

Residue determined as dansyl derivative, ⟶; residue determined as PTH, ⟶; residue determined by amino acid analysis after back-hydrolysis with HI, - - -; residue determined by selective tritiation, ⟵

Peptide	Amino acid sequence
P1	Pro (Asn Asn Asp Asn) Phe
P2	Asp Lys Leu
P3	Tyr Ile Trp Gly Val His His Pro Ser Thr Asp Gln Glu Gln Thr Ser Leu
P4	Tyr Val Gln Ala Ser Gly Arg
P5	Val (Thr Val Ser Thr Lys Arg)
P6a	Gly Leu Ser Ser
P6b	Gly (Leu, Ser, Ser, Arg)
P7	Ile Ser
P8a	Tyr (Trp, Thr, Ile, Val, Lys, Pro, Gly, Asx, Ile Leu)
P8b	Pro Gly Asp Ile Leu
P9	Val Ile Asn Ser Asx Gly Asx Leu
P10a	Ile Ala Pro Arg Gly Tyr
P10b	Gly Tyr Phe
P11	Phe Lys

Complete Sequence of CN2

The final structure of CN2 was established by automated sequenator analysis (67 cycles) on 600 nmoles of CN2. Positive identifications of PTH derivatives by thin-layer chromatography or amino acid analysis after back-hydrolysis with HI were obtained for 63 of the 67 cycles, confirming the arrangement of peptides T1–T4

and providing definite overlaps between peptides T5b, T6, T7 and T8. The complete sequence of CN2 and the peptides used to deduce the sequence are shown in Fig. 4.

Discussion

The primary structure of cyanogen bromide peptide CN2 described in this paper (Fig. 4) has been obtained by direct automated sequence analysis on the whole fragment and manual dansyl–Edman degradation of tryptic, peptic and chymotryptic peptides. All acid and amide assignments have been based on the direct identification of PTH derivatives and the final assignments in each peptide are in agreement with the expected net charges based on electrophoretic mobilities at pH 6·5 (Offord 1966).

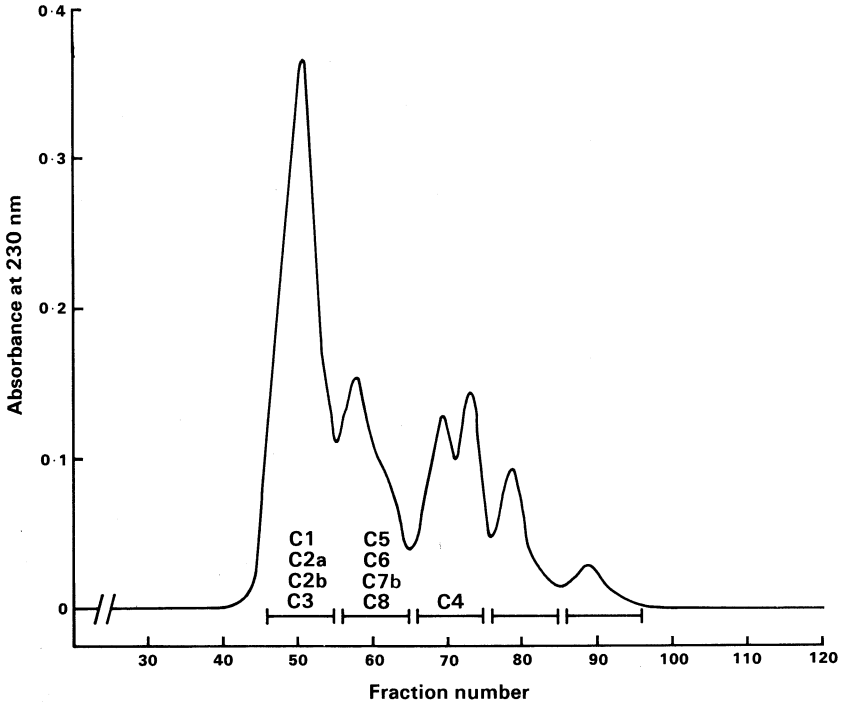

Fig. 3. Chromatography of chymotryptic digest of CN2 (1·0 μmol) on four coupled Sephadex columns. Conditions were as described in the legend to Fig. 1.

The major obstacle to the completion of the final peptide alignments was the presence of a labile Asn-Ile bond at residues 48–49 which was readily cleaved by both trypsin and chymotrypsin. The cleavage of this bond by TPCK-treated trypsin was also found in the comparative peptide-mapping analysis of HA_1 from several other Hong Kong field strains (Laver *et al.* 1980) and variants selected with monoclonal hybridoma antibodies (Laver *et al.* 1979) and presumably represents an intrinsic tryptic cleavage rather than that of contaminating chymotrypsin. The final overlap across this bond was obtained by automated sequence analysis of whole CN2 as shown in Fig. 4. Cyanogen bromide peptide CN2 was found to contain 92 amino acid residues including two histidine residues at positions 15 and 16 in CN2 and three tryptophan residues at positions 12, 54 and 66. It does not contain any half-cystine residues or carbohydrate.

Table 5. Amino acid compositions of chymotryptic peptides from CN2

Values are expressed as residues per mole. Yields represent nanomoles of purified peptide recovered after chymotryptic digestion of 1·0 μmole of CN2. For other details see legend to Table 1

Amino acid	C1	C2a	C2b	C3	C4	C5	C6	C7a	C7b	C8
Asp	4·9 (5)	1·1 (1)	1·1 (1)	1·1 (1)				3·7 (4)		
Thr		1·8 (2)	1·7 (2)	2·5 (3)				1·0 (1)		
Ser		1·8 (2)	1·7 (2)	3·0 (3)	1·0 (1)		3·0 (3)	1·2 (1)		
Glu		2·9 (3)	3·0 (3)	2·8 (3)						
Pro	1·1 (1)	1·2 (1)	1·0 (1)	1·0 (1)	1·0 (1)			1·7 (2)	1·0 (1)	
Gly		1·4 (1)	1·2 (1)	1·1 (1)	1·1 (1)	1·1 (1)		2·8 (3)	1·1 (1)	
Ala				1·1 (1)				0·9 (1)	1·1 (1)	
Val		0·9 (1)	1·2 (1)	2·3 (3)		0·8 (1)		1·6 (2)		
Ile		0·9 (1)		2·0 (2)	1·0 (1)		1·6 (2)	3·3 (4)	0·8 (1)	
Leu	1·1 (1)	1·1 (1)	1·1 (1)			1·0 (1)		2·1(2)		
Tyr	0·9 (1)	0·9 (1)	0·9 (1)				0·9 (1)	1·0 (1)	1·0 (1)	
Phe	1·0 (1)							0·5 (1)	0·8 (1)	
His		1·9 (2)	2·0 (2)							
Lys	1·0 (1)			1·1 (1)				0·9 (1)		1·0 (1)
Arg				1·8 (2)	1·1 (1)	0·9 (1)	1·2 (1)	1·2 (1)	0·9 (1)	
Trp		(1)			0·8 (1)			(1)		
Hse[A]										0·8 (1)
M	−0·43	0	0	+0·30	+0·42	+0·49	+0·38	—	+0·38	+0·76
Yield	254	89	22	62	71	110	50	417	55	130
S. P.[B]	1–10	11–27	13–27	28–48	49–54	55–58	59–65	66–90	84–90	91–92

[A] Homoserine. [B] Sequence position.

Table 6. Amino acid sequences of chymotryptic peptides from CN2

Peptide C7a was analysed by automated sequence analysis. The sequences of the other peptides were determined manually. Identifications are as follows: ⟶ dansyl derivatives; ⟶ PTH derivatives; - - ⟶ amino acid analysis after HI hydrolysis of PTH derivatives; ⟵ selective tritiation

Peptide	Amino acid sequence
C1	Pro (Asn Asn Asp Asn Phe Asp Lys Leu Tyr)
C2a	Ile (Trp Gly Val His His Pro Ser Thr Asp Gln Glu Gln Thr Ser Leu Tyr)
C2b	Gly (Val His His Pro Ser Thr Asp Gln Glu Gln Thr Ser Leu Tyr)
C3	Val Gln Ala Ser Gly Arg Val Thr Val Ser Thr (Lys Arg Ser Gln Gln Thr Ile Ile Pro Asn)
C4	Ile Gly Ser Arg Pro Trp
C5	Val Arg Gly Leu
C6	Ser Ser Arg Ile Ser Ile Tyr
C7a	Thr Ile Val Lys Pro Gly Asp ILe Leu Val Ile Asn Ser Asn Gly Asp Leu Ile Ala (Pro Arg Gly Tyr Phe)
C7b	Ile Ala Pro Arg Gly Tyr Phe
C8	Lys HSer

The data presented in this paper and an earlier report (Dopheide and Ward 1978*a*) collectively describe the amino acid sequence of the four cyanogen bromide peptides which comprise the 160 residues at the *C*-terminal of A/Mem/72 HA_1. The relative

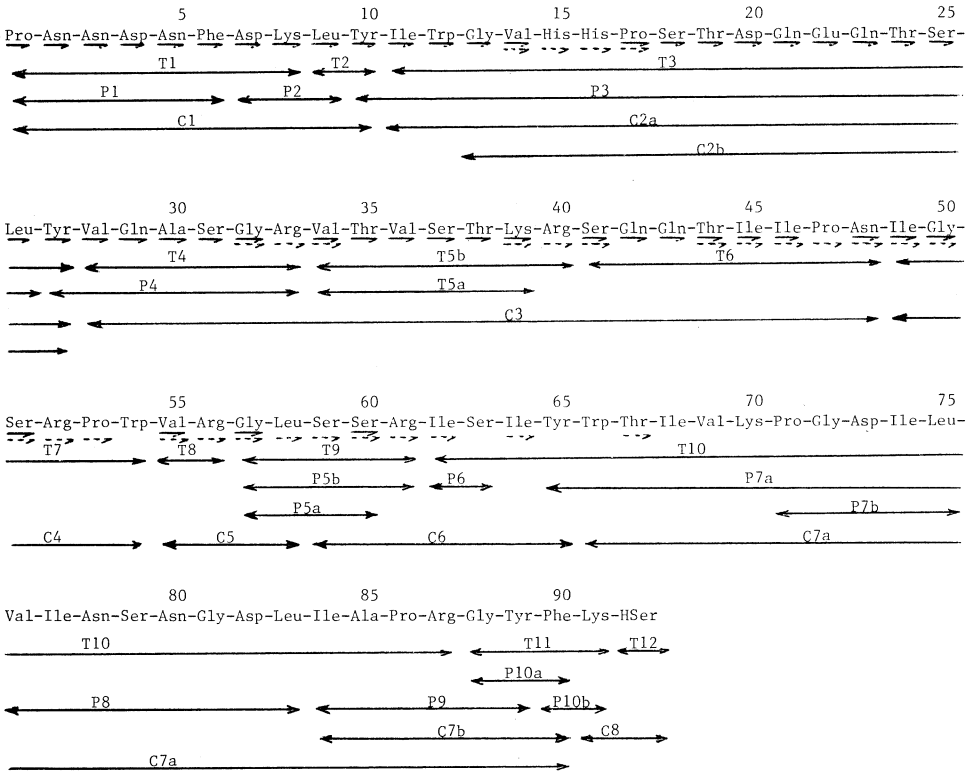

Fig. 4. Amino acid sequence of HA_1 CN2 showing the alignment of peptides used to determine the complete sequence. Residues determined by automated sequenator analysis of whole CN2 and identified by thin-layer chromatography of PTH (——➤) and by amino acid analysis after hydrolysis with HI (- - - - ➤) are shown.

sizes and positions of these cyanogen bromide peptides are shown in Fig. 5. CN2 extends from residues 169 to 260 in the complete sequence HA_1. Predictions of possible secondary structure using the algorithms of Chou and Fasman (1974) and

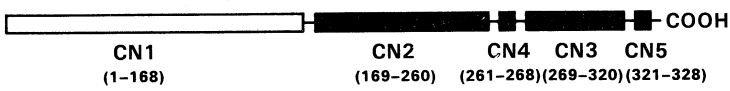

Fig. 5. Schematic representation of cyanogen bromide peptides of A/Mem./102/72 HA_1. The data in this paper and Dopheide and Ward (1978*a*) describe the amino acid sequence of the *C*-terminal 160 residues (solid boxes).

Fasman *et al.* (1976) show that this *C*-terminal 160-residue section of HA_1 may be very rich in β-structure (residues 7–16, 23–30, 33–39, 41–46, 54–58, 62–70, 74–78, 105–111, 126–137 and 140–144) with two short sections of α-helix at the *C*-terminus (residues 145–150, 155–160).

Fig. 6. Sequence of the *C*-terminal 160 amino acid residues from A/Mem/102/72 HA_1 (Dopheide and Ward 1978*a*; this paper) and A/Jap/305/57 HA_1 (Waterfield *et al.* 1978, 1979). Regions of identity are shown in the boxes and gaps have been inserted to maximize the homology.

Some comparisons can now be drawn between the amino acid sequence of this *C*-terminal 160-residue region in A/Mem/72 and that of the homologous region in the haemagglutinin from the Asian variant A/Jap/57 (Waterfield *et al.* 1978, 1979). As shown in Fig. 6, for the 120 residues for which comparisons can be made the sequences differ markedly (approx. 50%) including the location of the glycosylated asparagine residues. The homologous Asn-Asn sequence shown at positions 2–3 in Fig. 6 is glycosylated in A/Jap/57 but not in A/Mem/72 owing to the absence of the required serine or threonine at position 5 (see Neuberger *et al.* 1972). Furthermore carbohydrate groups near the *C*-terminus of both haemagglutinins are attached to non-homologous asparagine residues and are of different types. The carbohydrate group present at Asn_{121} in A/Jap/57 is complex containing all four sugars—*N*-acetylglucosamine, galactose, mannose and fucose (Waterfield *et al.* 1979)—while that present at Asn_{117} in A/Mem/102/72 is simple containing two residues of *N*-acetylglucosamine and five residues of mannose (Dopheide and Ward 1978*a*). These comparisons indicate that the major differences in the type and distribution of oligosaccharide side chains on different influenza haemagglutinins depend primarily on differences in amino acid sequence. Mutational events which affect the occurrence of Asn-X-Ser/Thr sequences (Neuberger *et al.* 1972) will subsequently affect both the type and location of attached carbohydrate. As has been mentioned previously (Ward and Dopheide 1979*a*; Waterfield *et al.* 1979), the distribution of half-cystine residues in the Asian and Hong Kong HA_1 is identical and highly conserved areas of sequence occur, particularly towards their *C*-terminal regions.

The changes in amino acid sequence associated with antigenic drift in Hong Kong variants selected with monoclonal hybridoma antibodies (Laver *et al.* 1979) has implicated the direct or indirect involvement of the serine residue at position 37 in CN2, and some portion of peptide Ile_{49}–Arg_{56}, in one of the antigenic regions of the Hong Kong haemagglutinin. Similar analysis of natural field strains (Laver *et al.* 1980) showed that mutations had also occurred at Lys_{39}, Arg_{40}, Ile_{49} and peptide Gly_{57}–Arg_{61}, including in some strains Leu_{58} and Ser_{59} (numbers refer to position in CN2, see Figs 4 and 6). Jackson *et al.* (1979*a*, 1979*b*) could detect neither antigenic nor significant immunogenic activity in peptide CN2 which carries these residues, implying that these residues are not directly involved in an antigenic region. Alternatively it may simply reflect a loss of conformational integrity of this fragment following its isolation from the rest of the haemagglutinin. As found with other proteins, the absence of disulfide bridges in fragment CN2 would be a significant factor in its failure to re-attain its native conformation (Atassi *et al.* 1976; Johnson *et al.* 1978; Jackson *et al.* 1979*a*, 1979*b*).

Acknowledgments

We wish to thank Dr C. M. Roxburgh for the high-pressure liquid chromatographic identifications and for his help with some of the amino acid analyses, and Mr I. D. Heritage and Mr P. McK. Strike for technical assistance.

References

Atassi, M. Z., Lee, C. L., and Habeeb, A. F. S. A. (1976). Enzymic and immunochemical properties of lysozyme-XII. Delineation of the reactive site around the two central di-sulfides by immunochemical and conformational studies of derivatives of the two-disulfide peptides. *Immunochemistry* **13**, 7–14.

Chou, P. Y., and Fasman, G. D. (1974). Prediction of protein conformation. *Biochemistry* **13**, 222–44.

Crewther, W. G., and Inglis, A. S. (1975). Automatic procedures for determining amino acid sequences of peptides. *Anal. Biochem.* **68**, 572–85.

Dopheide, T. A. A., and Ward, C. W. (1978*a*). The carboxyl-terminal sequence of the heavy chain of a Hong Kong influenza haemagglutinin. *Eur. J. Biochem.* **85**, 393–8.

Dopheide, T. A. A., and Ward, C. W. (1978*b*). Studies on the primary structure of the haemagglutinin from an H3N2 variant A/Memphis/102/72. In 'The Influenza Virus Haemagglutinin'. (Eds W. G. Laver, H. Bachmayer and R. Weil.) Topics in Inf. Dis. Vol. 3, pp. 191–201. (Springer-Verlag: Vienna.)

Dopheide, T. A. A., and Ward, C. W. (1979). A Hong Kong influenza haemagglutinin light chain: the amino acid sequence of cyanogen bromide fragments CN3 and CN4 and the *N*-terminal 45 residues. *Virology* **92**, 230–5.

Dowdle, W. R. (1976). Influenza: epidemic patterns and antigenic variation. In 'Influenza Virus, Vaccines and Strategy'. (Ed. P. Selby), pp. 17–21. (Academic Press: London.)

Fasman, G. D., Chou, P. Y., and Adler, A. J. (1976). Prediction of the conformation of the histones. *Biophys. J.* **16**, 1201–38.

Inglis, A. S., and Nicholls, P. W. (1973). Identification of phenylthiohydantoins of amino acids by thin layer chromatography. *J. Chromatogr.* **79**, 344–6.

Inglis, A. S., Strike, P. M., Osborne, W. C., and Burley, R. W. (1979). Sequenator determination of the amino acid sequence of Apovitellinin I from turkey's egg yolk. *FEBS Lett.* **97**, 179–82.

Jackson, D. C., Dopheide, T. A. A., Russell, R. J., White, D. O., and Ward, C. W. (1979*a*). Antigenic determinants of influenza virus haemagglutinin. II. Antigenic reactivity of the isolated *N*-terminal cyanogen bromide peptide of A/Memphis/72 haemagglutinin heavy chain. *Virology* **93**, 458–65.

Jackson, D. C., Brown, L. E., White, D. O., Dopheide, T. A. A., and Ward, C. W. (1979*b*). Antigenic determinants of influenza virus haemagglutinin. IV. Immunogenicity of cyanogen bromide fragments isolated from the haemagglutinin of A/Mem./72. *J. Immunol.* **123**, 2610–17.

Johnson, E. R., Anderson, W. L., Wetlaufer, D. B., Lee, C.-L., and Atassi, M. Z. (1978). Formation of three-dimensional structure in protein fragments. *J. Biol. Chem.* **253**, 3408–14.

Laver, W. G., and Webster, R. G. (1968). Selection of antigenic mutants of influenza viruses. Isolation and peptide mapping of their haemagglutinating proteins. *Virology* **34**, 193–202.

Laver, W. G., and Webster, R. G. (1972). Studies on the origin of pandemic influenza. II. Peptide maps of the light and heavy polypeptide chains from the haemagglutinin subunits of Hong Kong influenza. *Virology* **48**, 445–55.

Laver, W. G., Air, G. M., Webster, R. G., Gerhard, W., Ward, C. W., and Dopheide, T. A. A. (1979). Antigenic drift in type A influenza virus. Sequence changes in the haemagglutinin of Hong Kong (H3N2) variants selected with monoclonal hydridoma antibodies. *Virology* **98**, 226–37.

Laver, W. G., Air, G. M., Dopheide, T. A. A., and Ward, C. W. (1980). Amino acid sequence changes in the haemagglutinin of A/Hong Kong (N3N2) influenza virus during the period 1968–1977. *Nature (London)* **283**, 454–7.

Neuberger, A., Gottschalk, A., Marshall, R. D., and Spiro, R. G. (1972). In 'Glycoproteins'. (Ed. A. Gottschalk), 2nd edn, Part A, p. 78. (Elsevier: Amsterdam.)

Offord, R. E. (1966). Electrophoretic mobilities of peptides on paper and their use in the determination of amide groups. *Nature (London)* **211**, 591–3.

Smithies, O., Gibson, D., Fanning, E. M., Goodfliesh, R. M., Gilman, J. G., and Ballantyne, D. L. (1971). Quantitative procedures for use with the Edman–Begg sequenator. Partial sequences of two unusual immunoglobulin light chains, Rzf and Sac. *Biochemistry* **10**, 4912–21.

Ward, C. W., and Dopheide, T. A. A. (1976). Size and chemical composition of influenza virus haemagglutinin chains. *FEBS Lett.* **65**, 365–8.

Ward, C. W., and Dopheide, T. A. A. (1979*a*). Primary structure of the Hong Kong (H3) haemagglutinin. *Br. Med. Bull.* **35**, 51–6.

Ward, C. W., and Dopheide, T. A. A. (1979*b*). A Hong Kong influenza haemagglutinin light chain: amino acid sequence of cyanogen bromide fragment CN2. *Virology* **95**, 107–18.

Waterfield, M. D., Skehel, J. J., Nakashima, Y., Gurnett, A., and Bilham, T. (1978). Structural studies on the haemagglutinin of the Asian influenza virus Japan/305/57–Bellamy/42 (H2N1). Cyanogen bromide cleavage of the larger polypeptide chain HA_1. In 'The Influenza Virus Haemagglutinin'. (Eds W. G. Laver, H. Bachmayer and R. Weil). Topics in Inf. Dis. Vol. 3, pp. 167–79. (Springer-Verlag: Vienna.)

Waterfield, M. D., Espelie, K., Elder, K., and Skehel, J. J. (1979). Structure of the haemagglutinin of influenza virus. *Br. Med. Bull.* **35**, 57–63.

Zimmerman, C. L., Appella, E., and Pisano, J. J. (1977). Rapid analysis of amino acid phenylthiohydantoins by high-performance liquid chromatography. *Anal. Biochem.* **77**, 567–73.

Manuscript received 21 November 1979, accepted 30 January 1980