INFLUENZA VIRUS HAEMAGGLUTININ SIGNAL SEQUENCE

J. McCAULEY+, J. BYE, K. ELDER, M. J. GETHING, J. J. SKEHEL+, A. SMITH and M. D. WATERFIELD
Imperial Cancer Research Fund, PO Box 123, Lincoln's Inn Fields, London, WC2A 3PX, and *+National Institute for Medical Research, Mill Hill, London, NW7 IAA, England*

Received 5 October 1979

1. Introduction

The haemagglutinin is the major glycoprotein of the influenza virus membrane. We have reported [1] that in cell-free protein synthesising systems prepared from wheat germ or from L cells, translation of haemagglutinin messenger RNA leads to the production of two haemagglutinin precursor polypeptides. In both systems a 63 000 dalton precursor was produced and in the L cell system a 75 000 dalton component was also detected. The synthesis of the latter precursor was enhanced by the addition of microsomal membranes to either cell free system and furthermore under these conditions of translation this polypeptide was shown to be glycosylated and to be included within the added membrane vesicles. Here, the results of analyses of the amino-terminal sequences of these two haemagglutinin precursors are reported. In addition the partial amino acid sequence of the 63 000 dalton component is confirmed and the complete amino-terminal sequence is deduced from the results of nucleotide sequence analyses of the virion RNA which is complementary in sequence to the haemagglutinin messenger RNA.

2. Methods

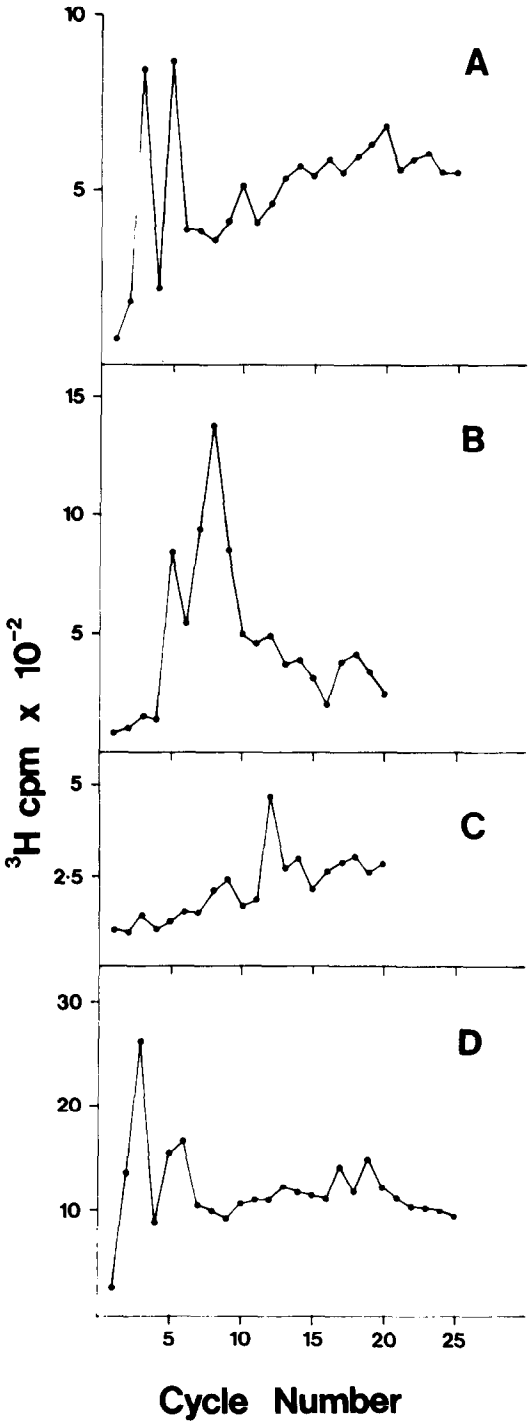
The conditions for cell-free protein synthesis were as in [1,2]. L-[4,5-^{3}H]Leucine (100 Ci/mmol), L-[4,5-^{3}H]isoleucine (80 Ci/mmol), L-[3,4 (*n*) -^{3}H]-valine (15 Ci/mmol) or L-[^{35}S]methionine (2000 Ci/mmol) from the Radiochemical Centre, Amersham, were added separately to wheat germ or nuclease-treated L-cell extracts containing RNA extracted from primary chick embryo fibroblasts infected with A/Japan/305/57 (H2Nl) influenza virus. A portion of the [^{35}S]methionine-labelled translation product was added to each [^{3}H]amino acid-labelled product and the polypeptides were separated by electrophoresis in 7.5% polyacrylamide gels. The gels were autoradiographed wet and the radioactive haemagglutinin precursors were eluted from the gel bands. Following addition of carrier myoglobin (85 nmol) the labelled polypeptides were precipitated and their amino acid sequence determined as in [2]. Virion RNAs were labelled at their 3′-termini using cytidine 3′,5′-[5′-^{32}P]bisphosphate from the Radiochemical Centre, Amersham and T_4-RNA ligase (P-L Biochemicals) as in [3] and subsequently purified by polyacrylamide gel electrophoresis [4]. Nucleotide sequence determinations were by the procedure in [5].

3. Results and discussion

The amino-terminal sequence of the 75 000 dalton glycosylated precursor has isoleucine residues at positions 3 and 5 equivalent to those of the HA_1 polypeptide component of the haemagglutinin from virus particles [6]. This was determined using [^{3}H]isoleucine-labelled material produced in an L cell system containing dog pancreas microsomal membranes (fig.1A). The amino-terminal sequence of the [^{3}H]-isoleucine-labelled 63 000 dalton precursor produced in an L cell system without added membranes was, however, quite different since radioactivity was detected in residues 2, 3, 5, 6, 17 and 19 (fig.1D).

In subsequent experiments the amino-terminal

 Elsevier/North-Holland Biomedical Press



sequence of the 63 000 dalton polypeptide produced in wheat germ extracts was analyzed and the results shown in fig.1B,C were obtained. In [^{3}H]leucine- and [^{3}H]valine-labelled products leucine residues were detected at positons 5, 7 and 8 and valine at position 12. In contrast to the results obtained for the amino-terminal sequence of the 63 000 dalton product of the L cell system these results indicate that leucine rather than isoleucine is present at position 5. However, since it can also be seen from the results in fig.1D that the yield of [^{3}H]isoleucine from the 63 00 dalton L cell system product in cycle 3 was ~2-times that obtained in cycle 5 it is possible that a mixture of polypeptides with the amino-terminal sequences of both the 75 000 dalton and the 63 000 dalton precursors was present in this product. If this is the case then the partial amino-terminal sequence of the 63 000 dalton haemagglutinin precursor is:

NH_2–?–Ile–Ile–?–Leu–Ile–Leu–Leu–?–?–?–
Val–?–?–?–?–Ile–?–Ile–

Furthermore since the isoleucine residues at positions 17 and 19 of this precursor have the same spacing as those at positions 3 and 5 of the viral HA_1 polypeptide, the primary translation product of haemagglutinin messenger RNA appears to contain an amino-terminal extension 14 amino acids in length.

The results of the nucleotide sequence analysis presented in fig.2 support this proposal. Since the sequence determined is complementary to that of the haemagglutinin mRNA, the first initiation codon is at residues 44–46 (fig.3) and the nucleotide sequence data indicate the sequence of the amino-terminal 45 amino acid residues. The protein sequence thus deduced confirms the partial amino acid sequence

Fig.1. Amino-terminal sequence analysis of influenza virus haemagglutinin synthesised in vitro. The recovery of [^{3}H]-amino acid at each cycle of Edman degradation in the automatic sequencer for the following preparations was obtained by the methods in [11]. (a) [^{3}H]Isoleucine-labelled 75 000 dalton glycosylated precursor synthesised in L cell extracts in the presence of microsomal membranes. (b) [^{3}H]Leucine-labelled 63 000 dalton precursor synthesised in wheat germ extracts. (c) [^{3}H]Valine-labelled 63 000 dalton precursor synthesised in wheat germ extracts. (d) [^{3}H]Isoleucine-labelled 63 000 dalton precursor synthesised in L cell extracts.

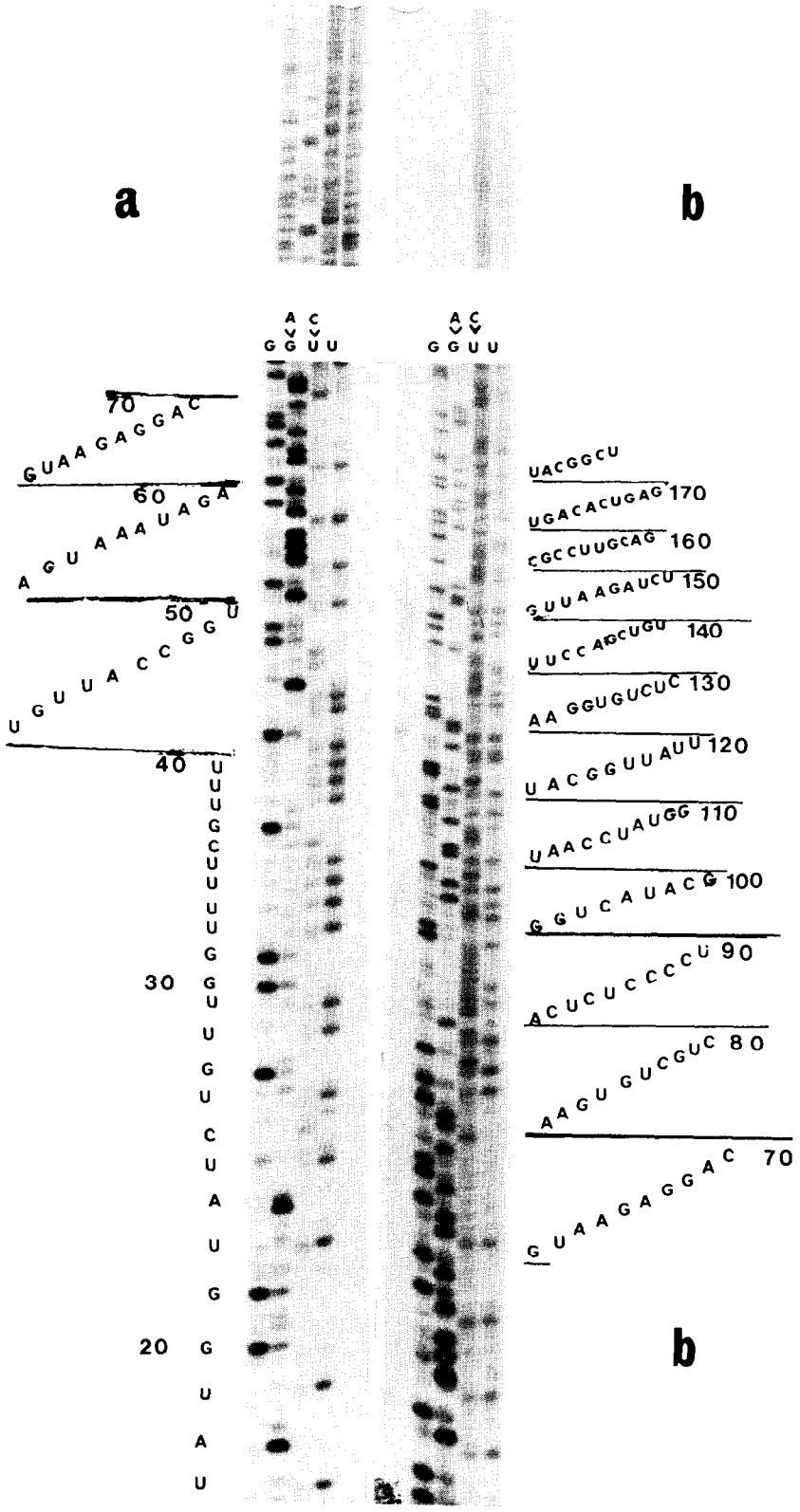


Fig.2.

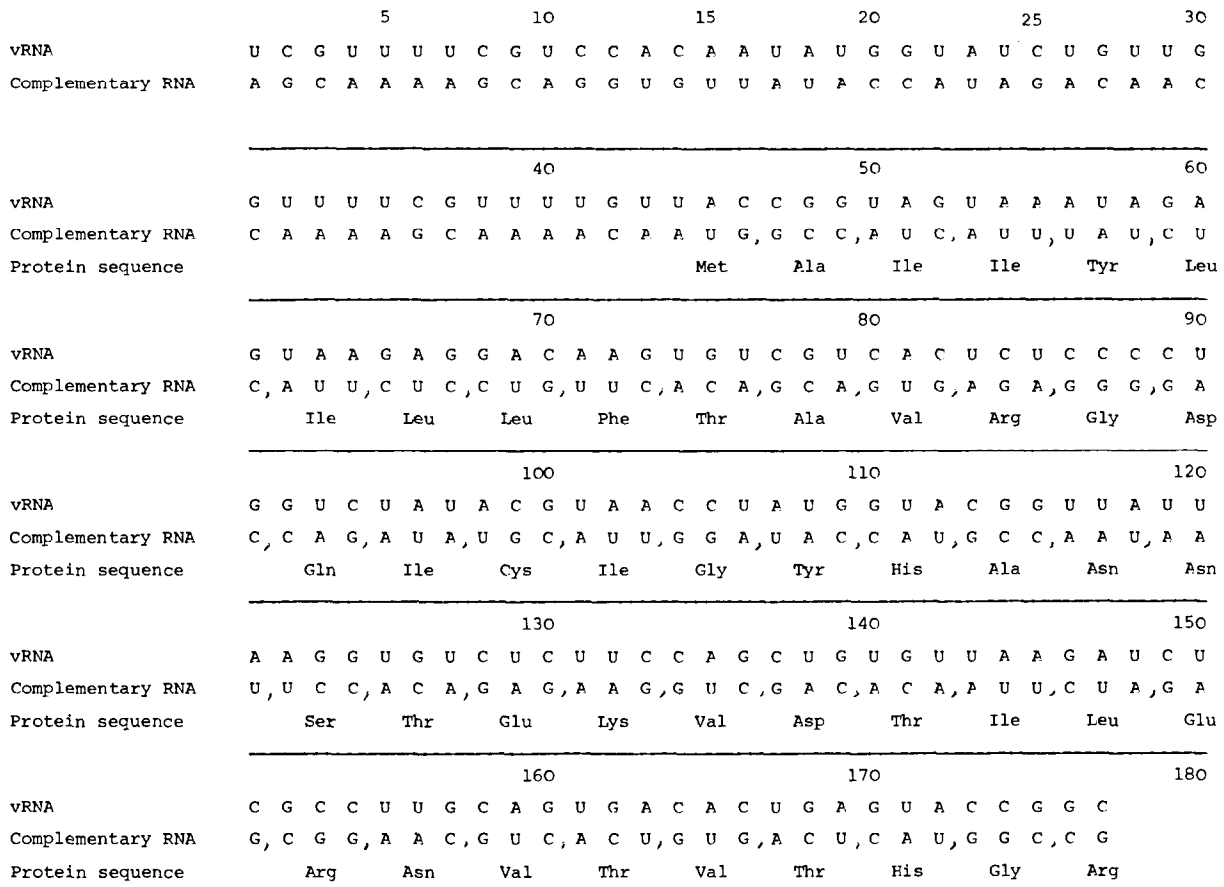


Fig.3. The 3′-terminal nucleotide sequence of the haemagglutinin gene, its complement and the amino-terminal sequence of the haemagglutinin.

determined for the 63 000 dalton component presented above with the provision that the initiator methionine residue is removed in vitro. In addition the nucleotide sequence establishes that the isoleucine residues detected at positions 17 and 19 of this component are equivalent to residues 3 and 5 of the HA_1 polypeptide isolated from virus particles. The amino acid sequence following the 14 residue amino-terminal extension is in agreement with that obtained by amino acid sequence analysis of the HA_1 polypeptide ([6], unpublished results).

The results presented here, therefore, indicate that the haemagglutinin of the influenza virus membrane is synthesized as a precursor which contains a hydrophobic amino terminal sequence not present in the virion glycoprotein. They support the concept of common initial events in the biosynthesis of secretory [7] and membrane proteins [8,9] according to which the hydrophobic amino-terminal extension may function as a signal peptide in facilitating association of the haemagglutinin precursor with the endoplasmic reticulum and its subsequent trans-mem-

Fig.2. Autoradiograph of the cleavage products of 3′-terminal labelled gene 4 of influenza virus. Reaction conditions were those in [5]. The cleavage products of the 4 reactions were fractionated by electrophoresis at 2000 V in gels which contained either (a) 12% acrylamide, 0.6% bisacrylamide, or (b) 8% acrylamide, 0.4% bisacrylamide. The gel dimensions were 960 × 200 × 0.5 mm.

brane movement. Taken together the nucleotide and the amino acid sequence data also indicate that the first AUG triplet of the haemagglutinin RNA transcript is indeed the initiating codon and they, therefore, imply that the 5'-terminal extensions present on those influenza virus RNA transcripts which function as messenger RNAs [10] are not translated.

Similar results to the nucleotide sequence presented here have been obtained by analysis of complementary DNA preparation by Dr G. Air (personal communication).

Acknowledgements

We thank D. Peattie for communicating her sequencing procedure before publication and D. Stevens for excellent assistance.

References

[1] Elder, K. T., Bye, J. M., Skehel, J. J., Waterfield, M. D. and Smith, A. E. (1979) Virology 95, 343–350.
[2] Paucha, E., Harvey, R. and Smith, A. E. (1978) J. Virol. 28, 154–170.
[3] England, T. E. and Uhlenbeck, O. C. (1978) Nature 275, 560–561.
[4] Skehel, J. J. and Hay, A. J. (1978) Nucleic Acid Res. 5, 1207–1219.
[5] Peattie, D. A. (1979) Proc. Natl. Acad. Sci. USA 76, 1760–1764.
[6] Waterfield, M. D., Espelie, K., Elder, K. and Skehel, J. J. (1979) Brit. Med. Bull. 35, 57–63.
[7] Blobel, G. and Dobberstein, B. (1975) J. Cell. Biol. 67, 835–851.
[8] Rothman, J. E. and Lodish, H. F. (1977) Nature 269, 775–780.
[9] Lingappa, V. R., Katz, F. N., Lodish, H. F. and Blobel, G. (1978) J. Biol. Chem. 253, 8667–8670.
[10] Bouloy, M. Plotch, S. J. and Krug, R. M. (1978) Proc. Natl. Acad. Sci. USA 75, 4886–4890.
[11] Paucha, E., Mellor, A., Harvey, R., Smith, A. E., Hewick, R. M. and Waterfield, M. D. (1978) Proc. Natl. Acad. Sci. USA 75, 2165–2169.