Sequence and topology of a model intracellular membrane protein, E1 glycoprotein, from a coronavirus

John Armstrong, Heiner Niemann*, Sjef Smeekens††, Peter Rottier† & Graham Warren

European Molecular Biology Laboratory, Postfach, Universitiit, Giessen, FRG
† Institute for Virology, Veterinary Faculty, State University of Utrecht, 3509 TD Utrecht, The Netherlands

In the eukaryotic cell, both secreted and plasma membrane proteins are synthesized at the endoplasmic reticulum, then transported, via the Golgi complex, to the cell surface1-4. Each of the compartments of this transport pathway carries out particular metabolic functions5-9, and therefore presumably contains a distinct complement of membrane proteins. Thus, mechanisms must exist for localizing such proteins to their respective destinations. However, a major obstacle to the study of such mechanisms is that the isolation and detailed analysis of such internal membrane proteins pose formidable technical problems. We have therefore used the E1 glycoprotein from coronavirus MHV-A59 as a viral model for this class of protein. Here we present the primary structure of the protein, determined by analysis of cDNA clones prepared from viral mRNA.

In combination with a previous study of its assembly into the endoplasmic reticulum membrane10, the sequence reveals several unusual features of the protein which may be related to its intracellular localization.

The coronaviruses are a diverse class of enveloped RNA viruses of considerable medical and agricultural significance; they also provide a model for the study of persistent viral infections (see ref. 10 for review). In contrast to many enveloped viruses, the coronavirus mouse hepatitis virus (MHV) A59 buds inside the cell, into the lumen of the endoplasmic reticulum11-14. The assembled virion then appears to travel, via the Golgi complex, to the cell surface. Of the two viral membrane proteins, the smaller one, E1, is necessary for formation of the envelope, while the larger one, E2, reaches the cell surface as part of the budded virion12.

The mRNAs of MHV-A59 form a 'nested set': the seven RNAs share the 3' region of the positive-stranded genome, but extend to different lengths towards the 5' end15-18. From each RNA, only the 5' gene is translated19,20. In addition, a non-coding 'leader' sequence of approximately 70 bases, from the 5' end of the genome, is common to the mRNAs18,21,22. The E1 gene is second from the 5' end and is therefore translated from the second smallest mRNA, RNA 6 (refs 19, 20). The sequence of the 3'-terminal gene, encoding the viral nucleocapsid protein, has been determined previously23,24.

Copy DNA clones spanning the E1 gene were prepared by two methods23-25 and sequenced in the vectors M13mp8 (ref. 26) or pEMBL8 (ref. 27) by the chain-termination method28. Two versions were found, in two different clones, for the sequence immediately upstream from the E1 initiator codon. The shorter one is shown in Fig. 1; in the second clone, an additional copy of the pentanucleotide ATCTA was found between nucleotides 65 and 66, making the sequence similar to that of the region adjacent to the nucleocapsid gene of another strain of MHV29. This difference could represent a mutation; alternatively, it may reflect heterogeneity in the normal mRNA population. Indirect support for the latter possibility comes from the observation that a RNase T1 oligonucleotide from this region hybridizes to viral RNA in infected cells30.

The sequence in Fig. 1 represents the point of fusion between the 5' leader sequence and the coding portion of the RNA. The fusion is thought to occur by 'jumping' of the viral RNA polymerase to particular sites on its genome-length, negative-stranded template; the resumption of transcription then produces each of the subgenomic mRNAs23,31,32. Thus, it seems possible that the polymerase may jump to more than one point on the template for each mRNA, generating variable numbers of the repeated pentanucleotide AUCUA in the resulting transcript.

The sequence of the E1 cDNA and protein extending to the initiator codon of the adjacent nucleocapsid gene32. Proposed membrane-spanning regions are overlined.

Two versions were found, in two different clones, for the sequence immediately upstream from the E1 initiator codon. The shorter one is shown in Fig. 1; in the second clone, an additional copy of the pentanucleotide ATCTA was found between nucleotides 65 and 66, making the sequence similar to that of the region adjacent to the nucleocapsid gene of another strain of MHV29. This difference could represent a mutation; alternatively, it may reflect heterogeneity in the normal mRNA population. Indirect support for the latter possibility comes from the observation that a RNase T1 oligonucleotide from this region hybridizes to viral RNA in infected cells30.

The sequence in Fig. 1 represents the point of fusion between the 5' leader sequence and the coding portion of the RNA. The fusion is thought to occur by 'jumping' of the viral RNA polymerase to particular sites on its genome-length, negative-stranded template; the resumption of transcription then produces each of the subgenomic mRNAs23,31,32. Thus, it seems possible that the polymerase may jump to more than one point on the template for each mRNA, generating variable numbers of the repeated pentanucleotide AUCUA in the resulting transcript.

Figure 1 shows the amino acid sequence encoded by the E1 gene. The predicted molecular weight of the protein is 26,000, slightly higher than that observed by gel electrophoresis33, but consistent with the unusual electrophoretic behaviour of this34, and other, hydrophobic proteins. Several features of the protein, when assembled into membranes in the virus35, or in vitro36, are reflected in membrane proteins. E1 is known to lack a cleaved 'signal peptide'37, the N-terminal region of the sequence contains no good candidate for a cleavage site38. Second, the N-terminal region bears O-linked sugars33, which, uniquely among viral proteins so far studied, are the only known post-translational modification to E1. Assuming that the terminal Met is removed37, the N-terminal sequence is Ser-Ser-Thr-Thr, which is identical to the O-glycosylated amino terminus of M-type glycophorin A (ref. 38). The O-linked sugars of E1 are the
selves identical to those found in glycoporphin. Third, most of the protein is resistant to proteolysis when assembled in the membrane. Only 2.5 kilodaltons of polypeptide from the N-terminus are cleavable on the luminal side of the membrane (or outside the virion) and 1.5 kilodaltons from the C-terminus from the cytoplasmic (or intra-virion) side, suggesting that the protein is largely buried in the membrane. In the sequence, a run of 22 uncharged residues from positions 26 to 47 represents a potential membrane-spanning region; residues 1–25 correspond to the portion removable by protease. A further sequence of uncharged residues, positions 57–106, is sufficiently long to cross the membrane twice more. If this region is divided in two, and each half plotted as an α-helical ‘wheel’, all the polar side chains of both sections cluster within 140°. Thus, a plausible conformation for this region is two hairpinned helices in the membrane, with adjacent polar faces (Fig 2a). There are no other long hydrophobic sequences, implying that the region from residues 107 to 190 is either folded in the membrane to neutralize charges, or, more likely, is adjacent to the membrane but resistant to proteolysis. These features are summarized in Fig. 2b.

Which, if any, of these various features might be responsible for the protein’s intracellular localization? We do not know, for example, whether the protein has an active ‘signal’ causing its arrest on the transport pathway, or, alternatively, if it lacks a signal for onward transport; nor do we know whether a sorting process might operate on one or the other side of the membrane. The availability of a cDNA clone for the protein presents the opportunity to investigate these questions by allowing expression of the cloned DNA and in vitro mutagenesis.

This approach has already been applied to two other viral glycoproteins, to investigate the importance of their cytoplasmic domains for transport to the cell surface, yielding opposite conclusions. An intrinsic problem with the method, however, is the difficulty of distinguishing specific effects due to alterations at the site of mutagenesis, from a general structural disruption of the molecule. In this respect the E1 protein may be advantageous in that it provides the possibility of creating a more ‘active’ phenotype in the mutated molecule: specifically, particular alterations to the protein may result in its transport to the cell surface.

We thank Willy Spann for communicating results before publication, G. Heisterberg-Moutsis (G. B. F. Braunschweig) for help with oligonucleotide synthesis, Ben van der Zeijst for discussion and Annie Steiner for preparing the manuscript. J.A. was supported by fellowships from the Royal Society and the European Molecular Biology Organization, H. N. by the Deutsche Forschungsgemeinschaft, SFB47 (Virologie), Teilprojekt B3, and P. R. by a short-term EMBO fellowship. Some of these results have been presented in preliminary form elsewhere.

Received 19 December 1983; accepted 21 March 1984.