BACKGROUND PAPER

CORONAVIRUS M AND HE: TWO PECULIAR GLYCOPROTEINS

Peter J.M.Rottier

Institute of Infectious Diseases and Immunology
Department of Virology
Veterinary Faculty
State University of Utrecht
The Netherlands

Enveloped viruses carry in their membranes one or more surface glycoproteins. In coronaviruses three such proteins have been identified: a small integral membrane protein (M), a large glycoprotein constituting the characteristic viral spikes (S), and an intermediate-sized glycoprotein (HE). This paper will summarize published knowledge on the properties and function of the small and the intermediate-sized membrane proteins. They will be referred to as M and HE, respectively.

THE HE GLYCOPROTEIN

With only a few exceptions (e.g. MHV-A59 and FIPV), coronaviruses possess haemagglutinating activity. For IBV and TGEV, whose virions lack a HE protein, this activity resides in the S protein. All other coronaviruses have a HE glycoprotein that has been shown or is supposed to represent the haemagglutinin. Interestingly, a second type of surface projection has been observed in coronaviruses possessing HE: a fringe of small granular projections situated at the base of the characteristic bulbous spikes. These short projections do indeed account for the haemagglutinating activity. Treatment of purified BCV with bromelain did not affect it. The enzyme degraded both the S and M protein while leaving HE intact; treated virions appeared to have lost the peplomers, whereas the fringe of short projections was still in place. Similar results were obtained with the human coronavirus OC43. Additional confirmation of the identity of HE as the haemagglutinin was recently obtained from experiments showing that monoclonal antibodies specific to the HE protein of BCV inhibited haemagglutination by this virus.

HE of coronaviruses is a glycoprotein of M_r approximately 60-65K with an apoprotein of about 45K. It is a typical class I membrane protein having a N-terminal signal sequence and a membrane anchor domain close to its C terminus. The protein is glycosylated only through N-linked oligosaccharides. The BCV HE sequence, which is the only HE sequence available so far, contains 9 potential N-glycosylation sites,the majority of which seems to be actually used. It is not known whether the protein is acylated or carries other modifications.

In infected cells HE is synthesized as a precursor glycoprotein that dimerizes rapidly by disulphide bond formation. The resulting homodimeric units are assembled into virions and transported out of the cell. En route to the plasma membrane part of the N-linked high-mannose oligosaccharides on the HE protein are processed to the complex type. Exactly how many homodimeric units make up one small projection has still to be determined.

Coronaviruses and Their Diseases
Edited by D. Cavanagh and T.D.K. Brown
Plenum Press, New York, 1990

The gene encoding the HE protein is located immediately upstream of the S gene and has probably been acquired by heterologous recombination. Interestingly, the indications for these features originally came from work with one of the few coronaviruses devoid of haemagglutinating activity, MHVA59. When the unique region of mRNA 2 of this virus was sequenced a second open reading frame was found downstream of the sequence encoding the 30K non-structural protein. This reading frame appeared to lack a translation initiation codon and is probably a pseudogene in MHV-A59. Most surprisingly, the sequence showed considerable homology to the HE haemagglutinin sequence of influenza C virus; at the amino acid level some 30% identity exists. It is believed that these influenza-like sequences were acquired during evolution by non-homologous recombination of some coronavirus with an influenza virus.

These observations prompted experiments to compare the biological properties of coronavirus HE to those of the influenza C HE protein. The latter is known to have both receptor binding and receptor-destroying activity, the receptor being characterized by the presence of a 9-O-acetylneuraminic acid modification. Indeed, the haemagglutinating coronaviruses HCV OC43 and BCV were found to recognize and bind to similarly sialylated receptors. In addition, receptor inactivation was demonstrated using BCV and shown to be of a type similar to that of influenza C, but distinct from that of influenza A. It is interesting to note that a similar observation was reported years ago for a mouse enteric coronavirus namely diarrhoea virus of infant mice (DVIM).

Further characterization of the receptor destroying activity showed that:

1. the activity is associated with the HE glycoprotein,
2. it is an acetylesterase activity that removes acetyl groups from O-acetylated sialic acids,
3. it is a serine esterase type of enzyme,
4. inactivation of the esterase activity strongly reduces the infectivity of the virus suggesting a role in one of the earliest stages of infection.

IBV and MHV-A59 being prototypes in the study of coronaviruses, interest in the HE glycoprotein has long lagged behind that for the other structural proteins. One realizes now that the presence of this protein in virions is the rule rather than the exception. Given the important features of the protein as they now emerge, focus on HE is expected to increase considerably. The central issues that will need to be addressed include:

1. the receptor binding activity: its function and, in relation to this, the function of the S glycoprotein in binding to the target cell; the further characterization of the receptor(s),
2. the receptor destroying activity: its mechanism and the precise role in coronavirus infection,
3. the biogenesis and structure of the small surface projections.

THE M GLYCOPROTEIN

Together with the nucleocapsid protein the small integral membrane protein M is the most abundant polypeptide constituent of coronavirions. M is a glycoprotein of M_r 25-30K that may either be N-glycosylated or carry only O-linked oligosaccharides as is the case with BCV, MHV and DVIM. The biological relevance of these differential glycosylations is obscure.

The gene encoding the M protein is located in the 3'-terminal region of the genome upstream from the N gene and is consequently expressed through one of the smaller subgenomic mRNAs. The M gene of several coronaviruses has been cloned and sequenced. The deduced amino acid sequences reveal common features but also some differences. The general structure of the mature protein is characterized by the presence of three successive stretches of hydrophobic amino acid residues in the N-terminal half of the molecule. This region is preceded by a hydrophilic N-terminus containing potential N- or O-glycosylation sites and followed by a more amphiphilic C-terminal half. Sequence homology among the various proteins is significant in some parts of the hydrophobic domains of the polypeptide, but is very high in the centre of the molecule where a stretch of 8 residues is identical among all viruses studied so far. A conspicuous feature in the sequences of the TGEV and FIPV M proteins is the presence of an N-terminal hydrophobic

extension. As will be discussed below, this extension apparently functions as a membrane insertion signal.

Investigations of the organization of M in microsomal membranes and in the virion membrane using proteases have led to a model in which the protein is anchored in the lipid bilayer by its three hydrophobic domains. Protruding from one side of the membrane into the lumen of the microsomes is the N-terminus of the molecule. In virions this side corresponds to the external domain which carries the oligosaccharides. On the opposite side of the membrane most of the C-terminal half appears to be protected from protease attack and supposedly is tightly associated with the polar surface of the bilayer. Only the extreme C-terminal tail is accessible.

Several studies have focussed on the way this complex protein is assembled into membranes. As noted above, coronavirus M proteins generally lack an amino-terminal signal sequence. Yet they are targeted to the membane in a signal recognition particle-dependent manner and inserted cotranslationally. Both the first and the third transmembrane domain may function as targeting signals as either of them is able to direct insertion and anchoring of mutant M proteins from which the other two hydrophobic segments are deleted. Membrane integration of TGEV and FIPV M is mediated by a N-terminal signal sequence. Apparently this signal is cleaved off. It was suggested that cleavage may occur posttranslationally. Although *in vitro* the amino-terminal signal sequence appeared not to be an absolute requirement, its presence greatly enhanced the correct membrane assembly of the polypeptide.

An interesting biological feature of the M protein is its restriction to internal cell membranes. In contrast to most other viral membrane glycoproteins M accumulates in coronavirus infected cells in the perinuclear region. It is not transported to the plasma membrane except when in virions which are subsequently shed from the cell. As was shown by electron microscopy with MHV-A59 the protein is largely retained in the so-called budding compartment located between the endoplasmic reticulum and Golgi system. In addition, some M also reaches the Golgi membranes as an integral membrane protein.

Intracellular restriction is an intrinsic property of the M protein and is not dependent on other coronaviral factors. Expression of M proteins from different coronaviruses in various cells has revealed that M is transported efficiently to the Golgi apparatus. In the case of MHV-A59 M protein the majority of the O-linked sugars on the expression product acquire the sialic acid modification indicating that the protein has reached the trans cisternae of the organelle.

Its accumulation in the Golgi apparatus makes the M protein attractive in studies of the targeting of membrane proteins. Deletion mutagenesis has shown that removal of transmembrane domains can have diverse effects. A mutant M protein having only the first hydrophobic region appeared to accumulate intracellularly in a manner indistinguishable from the wild type protein. In contrast, a mutant with only the third transmembrane domain was not retained in the Golgi region but transported to the plasma membrane. These observations suggest that information for retention in the Golgi apparatus resides in the first transmembrane domain.

The domain structure and the biological features of the M protein presumably reflect its functions in the coronavirus life cycle:

1. The hydrophilic N-terminal virion ectodomain elicits antibodies. These antibodies can neutralize the virus but only in a complement-dependent manner. The role of the O-linked sugars carried by the ectodomain of some coronavirus M proteins is still unknown.
2. The cluster of transmembrane segments has been related to the intracellular budding of coronaviruses. The perinuclear region where M accumulates in infected cells coincides with the location of virion formation. In addition, this hydrophobic domain of M may be involved in the interaction with the S protein. Since no complexes between M and S have been observed in detergent-disrupted cells or virus, interaction probably occurs at the level of the membrane.
3. The C-terminal half of the protein most likely interacts with the nucleocapsid and is thus of key importance for the budding process. Association of the nucleocapsid of detergent-solubilized MHV-A59 with M but not with S has been demonstrated. Both the extreme C-

terminal tail and the highly conserved domain in the center of the M protein are candidate domains for binding to genomic RNA and/or nucleocapsid protein.

As more information is becoming available, the various functions of M will soon be understood in greater molecular detail. As a consequence, some functions may need to be redefined such as the exclusive role of M in determining the intracellular budding site. The recent finding that M, when expressed from cloned cDNA, is transported beyond the budding compartment, indicates that in the coronavirus-infected cell some factor is required to retain M in this compartment. It may well turn out that the S protein serves such a function and that the budding site is determined by the concerted action of both viral membrane proteins. Support for this assumption is provided by the finding that S accumulates in this very region of the cell and is transported to the cell surface only very slowly unless it is incorporated in virions.

The M protein has a number of interesting features both as a virion protein and as a model membrane glycoprotein. Future research will certainly concentrate on both of these aspects. The expected main issues will be:

1. its role in viral budding, as part of the attempt to understand the process of virus assembly in molecular terms
2. the significance of the N-terminal signal peptide in the TGEV and FIPV M sequence
3. the function of O-linked vs. N-linked oligosaccharides
4. the mechanism of membrane integration
5. the identification of targeting information within the protein's structure.

Intracellular budding among positive strand RNA viruses is not restricted to coronaviruses; it has also been observed with toroviruses and arteriviruses. These viruses also share some replication characteristics and there are indications that they specify a membrane protein with features similar to M. It is an exciting idea, therefore, that these features may have more general relevance which makes the M protein all the more interesting.