ADVANCES IN VIRUS RESEARCH, VOL. 40

STRUCTURE AND FUNCTION OF THE HEF GLYCOPROTEIN OF INFLUENZA C VIRUS

Georg Herrler and Hans-Dieter Klenk

Institut für Virologie
Philipps-Universität Marburg
D-3550 Marburg, Germany

I. INTRODUCTION

Soon after the first isolation of an influenza C virus from a patient (Taylor, 1949), it became obvious that this virus differs from other myxoviruses in several aspects. Pronounced differences have been observed in the interactions between the virus and cell surfaces, suggesting that influenza C virus attaches to receptors different from those recognized by other myxoviruses. While influenza A and B viruses agglutinate erythrocytes from many species, including humans, the spectrum of erythrocytes agglutinated by influenza C virus is much more restricted. Erythrocytes from rats, mice, and adult chickens are suitable for hemagglutination and hemadsorption tests; cells from other species, however, react not at all or only poorly with influenza C virus (Hirst, 1950; Minuse *et al.*, 1954; Chakraverty, 1974; Ohuchi *et al.*, 1978). Differences are also observed so far as hemagglutination inhibitors are concerned. A variety of glycoproteins have been shown to prevent influenza A and B viruses from agglutinating erythrocytes. In the case of influenza C virus, rat serum was for a long time the only known hemagglutination inhibitor (Styk, 1955; O'Callaghan *et al.*, 1980).

A difference in the receptors for influenza C virus and other myxoviruses was also suggested by studies on the receptor-destroying enzyme. The ability of influenza C virus to inactivate its own receptors was reported soon after the first isolation of this virus from a patient (Hirst, 1950). However, the influenza C enzyme did not affect the receptors of other myxoviruses and, conversely, the receptor-destroying enzyme of either of the latter viruses was unable to inactivate the receptors for influenza C virus on erythrocytes. While the enzyme of influenza A and B virus was characterized as a neuraminidase in the 1950s (Klenk *et al.*, 1955), even with refined methodology no such activity was detectable with influenza C virus (Kendal, 1975; Nerome, *et al.*, 1976).

It is now known that both the receptor-binding and receptor-destroying activities as well as the fusion activity of influenza C virus are mediated by the only glycoprotein present on the surface of the virus particle. The structure and functions of this protein, which is designated HEF, are reviewed in the following sections.

II. STRUCTURE

A. Primary Structure

For two strains of influenza C virus, the RNA segment containing the genetic information for HEF has been cloned and sequenced (Nakada *et al.*, 1984; Pfeifer and Compans, 1984). Sequence data for several strains have been obtained by direct sequencing of the viral RNA (Buonagurio *et al.*, 1985; Adachi *et al.*, 1989). The gene is 2070–2075 nucleotides in length and can code for a polypeptide of 654–655 amino acids (Fig. 1). The predicted polypeptide has a molecular weight of about 72,000. At the amino terminus there is a stretch of 12 hydrophobic amino acids, which may represent the signal sequence. Cleavage of this sequence results in a polypeptide with a molecular weight of about 70,500. Two additional hydrophobic sequences are located at positions 447–463 and 627–652. The former is probably involved in the fusion activity, as discussed in Section III,C. The hydrophobic amino acid sequence at the carboxy-terminal end is assumed to function as a membrane anchor, which is followed by a cytoplasmic tail of only three amino acids. While a homology of 30% has been observed between the hemagglutinins of influenza A and B viruses (Krystal *et al.*, 1982), no significant values of homology were found when these glycoproteins were compared with the HEF protein of influenza C virus. The similarity between the HEF sequence and the HA sequence is restricted to the presence of the three hydrophobic domains mentioned above. Using this criterion, sequence alignments have been reported, with six to nine cysteines being conserved in the glycoproteins of the three types of influenza virus (Nakada *et al.*, 1984; Pfeifer and Compans, 1984). Comparison of the other influenza C proteins with their influenza A and B counterparts also revealed only a very low degree of sequence similarity (Yamashita *et al.*, 1989). Together, the sequence data suggest that influenza A and B viruses are more closely related to one another than they are to influenza C virus.

B. Co- and Posttranslational Modifications

Among the modifications of the HEF polypeptide, gycosylation has been studied in greatest detail. In the presence of the inhibitor tunicamycin the unglycosylated form of the protein is obtained (Nagele, 1983; Hongo *et al.*, 1986a). This finding indicates that the native glycoprotein only contains N-linked oligosaccharides, while O-linked carbohydrate structures are absent. As indicated in Fig. 1, the amino acid sequence contains eight consensus sequences Asn–X–Ser/Thr suitable for the attachment of N-linked oligosaccharides (Nakada *et al.*, 1984; Pfeifer and Compans, 1984). Analysis of the synthesis of the influenza C glycoprotein in the presence of limiting concentrations of glycosylation inhibitors suggested the presence of seven oligosaccharides on the native protein (Nagele, 1983): six on the HEF_1 portion and only one on HEF_2.

Three size classes of oligosaccharides—G_1, G_2, and G_3—have been resolved by gel chromatography (Nakamura *et al.*, 1979). Oligosaccharides corresponding to the two smaller size classes (i.e., G_2 and G_3) have also been observed in influenza A virus, while G_1 is restricted to influenza C virus. G_3 appears to represent the mannose-rich type of oligosaccharides. The oligosaccharides of size classes G_1 and G_2 have both been shown to contain *N*-acetylneuraminic acid, indicating that they are of the complex type. Because of the presence of sialic acid on the viral surface, influenza C virus is able to inhibit the hemagglutinating activity of influenza A viruses (Nerome *et al.*, 1976; Meier-Ewert *et al.*, 1978). The structure of the different oligosaccharides has not been determined. It has been suggested that some HEF polypeptides contain predominantly oligosaccharides of the larger size classes, while others are glycosylated with the smaller size classes (Nagele, 1983). This would provide an explanation as to why, after sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE), HEF is detected as a doublet band (Herrler *et al.*, 1979; Sugawara *et al.*, 1981).

```
                                                                             60
AGC AGA AGC AGG GGG TTA ATA ATG TTT TTC TCA TTA CTC TTG GTG TTG GGC CTC ACA GAG
                            Met Phe Phe Ser Leu Leu Leu Val Leu Gly Leu Thr Glu
                            ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                                                                            120
GCT GAA AAA ATA AAG ATA TGC CTT CAA AAG CAA GTG AAC AGT AGC TTC AGC CTA CAC AAT
Ala Glu Lys Ile Lys Ile Cys Leu Gln Lys Gln Val Asn Ser Ser Phe Ser Leu His Asn
 △                       ○                       ●
                                                                            180
GGC TTC GGA GGA AAT TTG TAT GCC ACA GAA GAA AAA AGA ATG TTT GAG CTT GTT AAG CCC
Gly Phe Gly Gly Asn Leu Tyr Ala Thr Glu Glu Lys Arg Met Phe Glu Leu Val Lys Pro

                                                                            240
AAA GCT GGA GCC TCT GTC TTG AAT CAA AGT ACA TGG ATT GGC TTT GGA GAT TCA AGG ACT
Lys Ala Gly Ala Ser Val Leu Asn Gln Ser Thr Trp Ile Gly Phe Gly Asp Ser Arg Thr
                             ●                                       □
                                                                            300
GAC AAA AGC AAT TCA GCT TTT CCT AGG TCT GCT GAT GTT TCA GCA AAA ACT GCT GAT AAG
Asp Lys Ser Asn Ser Ala Phe Pro Arg Ser Ala Asp Val Ser Ala Lys Thr Ala Asp Lys

                                                                            360
TTT CGT TTT TTG TCT GGT GGA TCC TTA ATG TTG AGT ATG TTT GGC CCA CCT GGG AAG GTA
Phe Arg Phe Leu Ser Gly Gly Ser Leu Met Leu Ser Met Phe Gly Pro Pro Gly Lys Val

                                                                            420
GAC TAC CTT TAC CAA GGA TGT GGA AAA CAT AAA GTT TTT TAT GAA GGA GTT AAC TGG AGT
Asp Tyr Leu Tyr Gln Gly Cys Gly Lys His Lys Val Phe Tyr Glu Gly Val Asn Trp Ser
                                                                     ●
                                                                            480
CCA CAT GCT GCT ATA AAT TGT TAC AGA AAA AAT TGG ACT GAT ATC AAA CTG AAT TTC CAG
Pro His Ala Ala Ile Asn Cys Tyr Arg Lys Asn Trp Thr Asp Ile Lys Leu Asn Phe Gln
                                         ●
                                                                            540
AAA AAC ATT TAT GAA TTG GCT TCA CAA TCA CAT TGC ATG AGC TTG GTG AAT GCC TTG GAC
Lys Asn Ile Tyr Glu Leu Ala Ser Gln Ser His Cys Met Ser Leu Val Asn Ala Leu Asp

                                                                            600
AAA ACT ATT CCT TTA CAA GTG ACT GCT GGG ACT GCA GGA AAT TGC AAC AAC AGC TTC TTA
Lys Thr Ile Pro Leu Gln Val Thr Ala Gly Thr Ala Gly Asn Cys Asn Asn Ser Phe Leu
                                                                 ●
                                                                            660
AAA AAT CCA GCA TTG TAC ACA CAA GAA GTC AAG CCT TCA GAA AAC AAA TGT GGG AAA GAA
Lys Asn Pro Ala Leu Tyr Thr Gln Glu Val Lys Pro Ser Glu Asn Lys Cys Gly Lys Glu

                                                                            720
AAT CTT GCT TTC TTC ACA CTT CCA ACC CAA TTT GGA ACC TAT GAG TGC AAA CTG CAT CTT
Asn Leu Ala Phe Phe Thr Leu Pro Thr Gln Phe Gly Thr Tyr Glu Cys Lys Leu His Leu

                                                                            780
GTG GCT TCT TGC TAT TTC ATC TAT GAT AGT AAA GAA GTG TAC AAT AAA AGA GGA TGT GAC
Val Ala Ser Cys Tyr Phe Ile Tyr Asp Ser Lys Glu Val Tyr Asn Lys Arg Gly Cys Asp

                                                                            840
AAC TAC TTT CAA GTG ATC TAT GAT TCA TTT GGA AAA GTC GTT GGA GGA CTA GAT AAC AGG
Asn Tyr Phe Gln Val Ile Tyr Asp Ser Phe Gly Lys Val Val Gly Gly Leu Asp Asn Arg

                                                                            900
GTA TCA CCT TAC ACA GGG AAT TCT GGA GAC ACC CCA ACA ATG CAA TGT GAC ATG CTC CAG
Val Ser Pro Tyr Thr Gly Asn Ser Gly Asp Thr Pro Thr Met Gln Cys Asp Met Leu Gln
                                         ■
                                                                            960
CTG AAA CCT GGA AGA TAT TCA GTA AGA AGC TCT CCA AGA TTC CTT TTA ATG CCT GAA AGA
Leu Lys Pro Gly Arg Tyr Ser Val Arg Ser Ser Pro Arg Phe Leu Leu Met Pro Glu Arg

                                                                           1020
AGT TAT TGC TTT GAC ATG AAA GAA AAA GGA CCA GTC ACT GCT GTC CAA TCC ATT TGG GGA
Ser Tyr Cys Phe Asp Met Lys Glu Lys Gly Pro Val Thr Ala Val Gln Ser Ile Trp Gly

                                                                           1080
AAA GGC AGA GAA TCT GAC TAT GCA GTG GAT CAA GCT TGC TTG AGC ACT CCA GGG TGC ATG
Lys Gly Arg Glu Ser Asp Tyr Ala Val Asp Gln Ala Cys Leu Ser Thr Pro Gly Cys Met

                                                                           1140
TTG ATC CAA AAG CAA AAG CCA TAC ATT GGA GAA GCT GAT GAT CAC CAT GGA GAT CAA GAA
Leu Ile Gln Lys Gln Lys Pro Tyr Ile Gly Glu Ala Asp Asp His His Gly Asp Gln Glu

                                                                           1200
ATG AGG GAG TTG CTG TCA GGA CTG GAC TAT GAA GCT AGA TGC ATA TCA CAA TCA GGG TGG
Met Arg Glu Leu Leu Ser Gly Leu Asp Tyr Glu Ala Arg Cys Ile Ser Gln Ser Gly Trp
                                                     ○
                                                                           1260
GTG AAT GAA ACC AGT CCT TTT ACG GAG AAA TAC CTC CTT CCT CCC AAA TTT GGA AGA TGC
Val Asn Glu Thr Ser Pro Phe Thr Glu Lys Tyr Leu Leu Pro Pro Lys Phe Gly Arg Cys
     ●                                                                       ○
                                                                           1320
CCT TTG GCT GCA AAG GAA GAA TCC ATT CCA AAA ATC CCA GAT GGC CTT CTA ATT CCC ACC
Pro Leu Ala Ala Lys Glu Glu Ser Ile Pro Lys Ile Pro Asp Gly Leu Leu Ile Pro Thr

                                                                           1380
AGT GGA ACC GAT ACC ACT GTA ACC AAA CCT AAG AGC AGA ATT TTT GGA ATC GAT GAC CTC
Ser Gly Thr Asp Thr Thr Val Thr Lys Pro Lys Ser Arg Ile Phe Gly Ile Asp Asp Leu
                                                    ~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                                   ↑
                                                                           1440
ATT ATT GGT TTG CTC TTT GTT GCA ATC GTT GAA ACA GGA ATT GGA GGC TAT CTG CTT GGA
Ile Ile Gly Leu Leu Phe Val Ala Ile Val Glu Thr Gly Ile Gly Gly Tyr Leu Leu Gly
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                                                                           1500
AGT AGA AAA GAA TCA GGA GGA GGT GTG ACA AAA GAA TCA GCT GAA AAA GGG TTT GAG AAA
Ser Arg Lys Glu Ser Gly Gly Gly Val Thr Lys Glu Ser Ala Glu Lys Gly Phe Glu Lys

                                                                           1560
ATT GGA AAT GAC ATA CAA ATT TTA AAA TCT TCT ATA AAT ATC GCA ATA GAA AAA CTA AAT
Ile Gly Asn Asp Ile Gln Ile Leu Lys Ser Ser Ile Asn Ile Ala Ile Glu Lys Leu Asn

                                                                           1620
GAC AGA ATT TCT CAT GAT GAG CAA GCC ATC AGA GAT CTA ACT TTA GAA ATT GAA AAT GCA
Asp Arg Ile Ser His Asp Glu Gln Ala Ile Arg Asp Leu Thr Leu Glu Ile Glu Asn Ala

                                                                           1680
AGA TCT GAA GCT TTA TTG GGA GAA TTG GGA ATA ATA AGA GCC TTA TTG GTA GGA AAT ATA
Arg Ser Glu Ala Leu Leu Gly Glu Leu Gly Ile Ile Arg Ala Leu Leu Val Gly Asn Ile
                                                                         ●
                                                                           1740
AGC ATA GGA TTA CAG GAA TCT TTA TGG GAA CTA GCT TCA GAA ATA ACA AAT AGA GCA GGA
Ser Ile Gly Leu Gln Glu Ser Leu Trp Glu Leu Ala Ser Glu Ile Thr Asn Arg Ala Gly

                                                                           1800
GAT CTA GCA GTT GAA GTC TCC CCA GGT TGC TGG ATA ATT GAC AAT AAC ATT TGT GAT CAA
Asp Leu Ala Val Glu Val Ser Pro Gly Cys Trp Ile Ile Asp Asn Asn Ile Cys Asp Gln
                                     ○                               ○
                                                                           1860
AGC TGT CAA AAT TTT ATT TTC AAG TTC AAC GAA ACT GCA CCT GTT CCA ACC ATT CCC CCT
Ser Cys Gln Asn Phe Ile Phe Lys Phe Asn Glu Thr Ala Pro Val Pro Thr Ile Pro Pro
     ○                               ●
                                                                           1920
CTT GAC ACA AAA ATT GAT CTG CAA TCA GAT CCT TTT TAC TGG GGA AGC AGC TTG GGC TTA
Leu Asp Thr Lys Ile Asp Leu Gln Ser Asp Pro Phe Tyr Trp Gly Ser Ser Leu Gly Leu
                                                    ~~~~~~~~~~~~~~~~~~~~~~~~~~~

                                                                           1980
GCA ATA ACT GCT ACT ATT TCA TTG GCA GCT TTG GTG ATC TCT GGG ATC GCC ATC TGC AGA
Ala Ile Thr Ala Thr Ile Ser Leu Ala Ala Leu Val Ile Ser Gly Ile Ala Ile Cys Arg
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                                                                           2040
ACT AAA TGA TTG AGA CAA TTT TGA AAA ATG GAT AAT GTG TTG GTC AAT ATT TTG TAC AGT
Thr Lys End

TTT ATA AAA AAC AAA AAT CCC CTT GCT ACT GCT
```

As discussed in Section II,D, glycosylation of HEF is important for the presentation of the antigenic epitopes. Furthermore, the carbohydrate side chains are crucial for the stability of the glycoprotein by protecting it from proteolytic degradation. In the presence of tunicamycin, virions are released from the infected cells; however, the virus particles are lacking surface proteins (Hongo *et al.*, 1986a).

Another posttranslational modification of HEF is the proteolytic cleavage of the precursor polypeptide HEF_0 into the cleavage products HEF_1 and HEF_2. As discussed in Section III,C, this modification is required for viral fusion activity. Cleavage is caused by a cellular protease. Some cultured cells [e.g., chick embryo fibroblasts, LLC-MK2, or Madin–Darby canine kidney (MDCK) cells] are lacking an appropriate enzyme or have only low amounts of it. On the surface of virions released by such cells, the glycoprotein HEF is found predominantly in the uncleaved form, which can be cleaved *in vitro* by incubation with trypsin and elastase (Compans *et al.*, 1977; Herrler *et al.*, 1979; Sugawara *et al.*, 1981). Influenza C viruses grown in embryonated eggs or primary chick kidney cells contain most of their glycoprotein molecules in the cleaved form. The cleavage products are detected after SDS–PAGE only in the presence of reducing agents (Herrler *et al.*, 1979).

This observation indicates that the two polypeptides are held together by disulfide bonds, as observed with several viral surface glycoproteins which are proteolytically cleaved. The disulfide bonds contribute to a unique electrophoretic behavior of HEF which is not observed with the glycoproteins of other influenza viruses. Under nonreducing conditions the electrophoretic mobility of HEF_0 suggests a molecular weight of about 100,000, which is not in accord with the size deduced from the sequence data. In the presence of reducing agents, the electrophoretic migration of the uncleaved glycoprotein suggests a molecular weight of about 80,000, which is in the size range expected for the glycosylated HEF_0. A shift from the 100K form to the 80K form is also observed under nonreducing conditions after proteolytic cleavage of HEF_0 into the disulfide-bonded products HEF_1 and HEF_2

FIG. 1. DNA sequence of gene segment 4 of influenza C/JHB/1/66 and its translation in open reading frame 1 (Pfeifer and Compans, 1984). The sequence is written in message sense. Hydrophobic sequences are marked with wavy lines. The predicted HEF_1–HEF_2 cleavage site is indicated by an arrow. The predicted cleavage site of the leader peptide is indicated by an open triangle. Solid circles indicate potential glycosylation sites. Open circles indicate cysteine residues conserved among hemagglutinins of influenza A, B, and C viruses. The active-site serine (amino acid 71) is indicated by an open square. The mutation site of a mutant with increased receptor-binding efficiency is marked with a solid square (Thr 284).

(designated $HEF_{1,2}$). No evidence for the release of a polypeptide could be obtained, which would explain the shift in the molecular weight by about 20,000 (Meier-Ewert *et al.*, 1980, 1981a,b).

Therefore, it is assumed that the uncleaved glycoprotein has a peculiar conformation which is maintained by disulfide bonds. This conformation may allow only association with a reduced amount of SDS, thereby causing aberrant electrophoretic migration behavior. The conformational constraint is released either by abolishing the disulfide bonds or by proteolytic cleavage of HEF_0 into HEF_1 and HEF_2. It is not known whether the formation of disulfide bonds is a co- or posttranslational modification of the glycoprotein. The proteolytic cleavage was found to be a late modification. In pulse–chase experiments hardly any cleaved glycoprotein was detectable in infected chick kidney cells. Therefore, the proteolytic cleavage may occur only shortly before virus particles are released by budding.

A modification of the influenza C glycoprotein, which has been described only recently, is the acylation with fatty acids (Veit *et al.*, 1990). The acyl chains are attached presumably to cysteine residues, as indicated by their release after treatment with either hydroxylamine or mercaptoethanol. Such a labile thioester-type linkage has been found on many acylated glycoproteins of both viral and cellular origin. In all cases tested palmitic acid was the predominant fatty acid. The HEF glycoprotein was unique in this respect, because stearic acid was detected as the prevailing fatty acid. The reason for this difference in the acylation is unknown. Cysteine residues in the cytoplasmic tail have been identified as fatty acid attachment sites for several glycoproteins. The cytoplasmic domain of HEF is very short. It contains only a single cysteine, which, therefore, is the candidate for attachment of stearic acid. So far, no biological function can be attributed to the fatty acid of the influenza C glycoprotein.

C. *Supramolecular Structure of the HEF Spike*

A characteristic feature of influenza C virus was revealed by electron microscopy long before the proteins had been analyzed. The surface projections are usually arranged in a reticular structure consisting mainly of hexagons (Flewett and Apostolov, 1967), which can be seen on both filamentous and spherical particles (Fig. 2). A single spike protein is observed on each of the six vertices of the hexagons (Herrler *et al.*, 1981). Values for the length of individual spikes are in the range of 8–15 nm. The low-resolution structure of the spikes determined by electron microscopy indicated that the influenza C glycoprotein is a trimer (Hewat *et al.*, 1984). The trimeric structure was confirmed when the sedimentation of the glycoprotein in sucrose

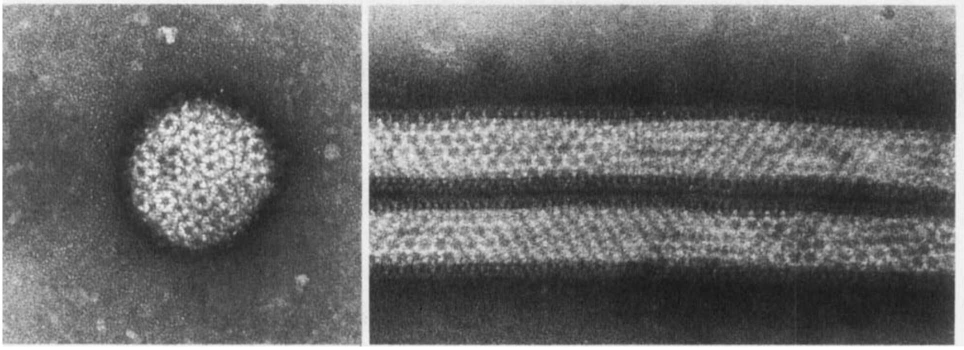

FIG. 2. Electron micrograph of influenza C virions. Both a spherical and a filamentous particle are shown. (Adapted from Herrler *et al.*, 1981.)

gradients was analyzed (Formanowski and Meier-Ewert, 1988; Formanowski *et al.*, 1989). Glycoprotein, which has been released from the viral surface by bromelain treatment, is still found as a trimer, indicating that the membrane anchor and the cytoplasmic tail are not essential for maintaining this structure. Calcium ions, however, appear to play an important role. On sucrose gradient centrifugation of bromelain-released HEF, trimers were detected only in the presence of calcium ions. When calcium-deficient buffers were used, the glycoprotein dissociated into monomers (Formanowski and Meier-Ewert, 1988; Formanowski *et al.*, 1989).

Lateral interactions between trimeric spike glycoproteins are probably involved in the formation of the hexagonal array on the viral surface. This is suggested by the finding that the reticular arrangement is sometimes maintained after removal of the spikes from the viral membrane by either protease treatment or spontaneous release (Herrler *et al.*, 1981). Lateral interactions are also suggested by electron micrographs of detergent-isolated spikes. On removal of the detergent, membrane glycoproteins (e.g., the influenza A hemagglutinin) form rosettes, where the proteins are connected at a central point via their hydrophobic membrane anchor. In contrast, the influenza C glycoproteins are arranged in an elongated beetlelike structure with individual spikes standing side by side (Formanowski *et al.*, 1989, 1990). The lateral interactions are not dependent on the proteolytic cleavage of the glycoprotein. The hexagonal array is observed with virus containing predominantly HEF_0 as well as with virus containing $HEF_{1,2}$. Glycoprotein in the uncleaved form also maintains the reticular pattern at pH 5.0. Cleaved glycoprotein, however, undergoes a major conformational change at low pH, resulting in the loss of the regular hexagonal arrangement of the spikes (Hewat *et al.*, 1984; Formanowski *et al.*, 1990).

It has been reported that crystals of bromelain-released HEF have been obtained, which are suitable for X-ray diffraction studies (Rosenthal *et al.*, 1990). Thus, there is hope that in the near future the three-dimensional structure of the influenza C glycoprotein will be known, which would represent major progress toward understanding the structure–function relationship.

D. Antigenic Epitopes

Among monoclonal antibodies directed against the HE protein, two groups have been distinguished. Group A antibodies inhibited the hemagglutinating and hemolytic activities and neutralized the infectivity of influenza C virus, whereas group B antibodies lacked any of these reactivities (Sugawara *et al.*, 1986). Analysis of antigenic variants selected for resistance against either of these monoclonal antibodies suggested the presence of four antigenic epitopes: A-1, A-2, B-1, and B-2 (Sugawara *et al.*, 1988). Competitive binding assays indicated that sites A-1 and A-2 may be located close to one another (Sugawara *et al.*, 1988). Both A epitopes were shown to be sensitive to denaturing conditions (e.g., treatment with SDS). Therefore, on Western blots, HEF was detected only by group B antibodies. These results indicate that sites B-1 and B-2 are sequence-dependent epitopes, whereas sites A-1 and A-2 are conformation-dependent epitopes. The conformation of both A epitopes was found to be dependent on the glycosylation of HEF. The nonglycosylated form of the protein synthesized in the presence of tunicamycin was recognized by group B antibodies, while group A antibodies reacted only poorly or not at all (Hongo *et al.*, 1986b; Sugawara *et al.*, 1988). The antigenic sites are presumably different from the functional epitopes of the receptor-binding and receptor-destroying activities. The ability of several monoclonal antibodies to inhibit the hemagglutinating activity of influenza C virus (Sugawara *et al.*, 1986, 1988; Vlasak *et al.*, 1987; Herrler *et al.*, 1988a) may be due to steric hindrance. Some of the antibodies caused partial inhibition of the receptor-destroying enzyme, when the esterase activity was determined with large substrates, but no inhibitory effect was observed when small substrates were used (Vlasak *et al.*, 1987; Herrler *et al.*, 1988a; Hachinohe *et al.*, 1989).

The antigenic variation among different strains of influenza C virus is less pronounced than in the case of influenza A viruses. Several reports revealed a high degree of cross-reactivity between different strains, irrespective of the time and place of isolation (Czekalowski and Prasad, 1973; Chakraverty, 1974, 1978; Meier-Ewert *et al.*, 1981c;

Kawamura *et al.*, 1986). Using monoclonal antibodies, it was possible, however, to demonstrate antigenic variation (Sugawara *et al.*, 1986, 1988; Adachi *et al.*, 1989). The low extent of variation is not due to a low capacity to produce antigenic variants. Escape mutants resistant against monoclonal antibodies have been obtained with a frequency similar to values reported for influenza A virus (Sugawara *et al.*, 1988). Maybe the immune selection is less pronounced in the case of influenza C virus. This may also explain why no antigenic drift has been observed with this group of viruses. Among the antigenic variants arising within influenza A viruses, one usually becomes dominant and replaces the older ones. In contrast, analyses of different influenza C strains indicate that several antigenic variants cocirculate (Adachi *et al.*, 1989). This conclusion is supported by studies on the genetic variation in the HEF as well as the *NS* gene of influenza C virus (Buonagurio *et al.*, 1985, 1986; Kawamura *et al.*, 1986; Adachi *et al.*, 1989).

III. FUNCTIONS

A. Receptor-Destroying Enzyme

Although the receptor-destroying enzyme of influenza C virus was described by Hirst in 1950, more than 30 years passed before its specificity was elucidated. Identification of the enzyme activity was accomplished by analyzing its effect on hemagglutination inhibitors. Rat serum has long been known for its inhibitory activity (Styk, 1955; O'Callaghan *et al.*, 1980). Two components of rat serum have been shown to account for most of the hemagglutination inhibition activity: murinoglobulin and α_1-macroglobulin (Herrler *et al.*, 1985b; Kitame *et al.*, 1985). The carbohydrate portion of the latter compound was found to consist primarily of N-linked biantennary oligosaccharides (Herrler *et al.*, 1985b). The only effect of the influenza C enzyme on these oligosaccharides was a change in the terminal sialic acid residue. While the native macroglobulin contained 40% of its sialic acid as *N*-acetyl-9-*O*-acetylneuraminic acid, this amount was reduced to 10% after treatment with the receptor-destroying enzyme.

Concomitant with the decrease of the 9-*O*-acetylated sialic acid, an increase of *N*-acetylneuraminic acid was observed (Herrler *et al.*, 1985c). The same effect was obtained with bovine submandibulary mucin, another hemagglutination inhibitor. In both cases the change in the sialic acid was paralleled by loss of the inhibitory activity, indicating that the receptor-destroying enzyme of influenza C virus is a sialate 9-*O*-acetylesterase (Fig. 3). The enzyme has been shown to be

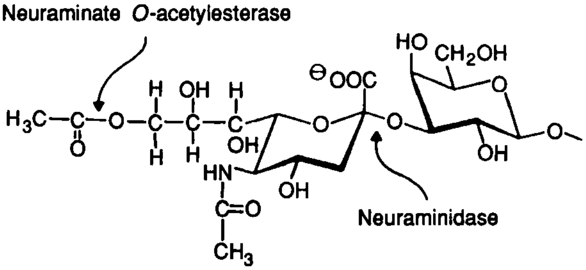

FIG. 3. Structure of *N*-acetyl-9-*O*-acetylneuraminic acid connected to galactose via an α2,3-linkage. The sites of action of the acetylesterase of influenza C virus and the neuraminidase of influenza A and B viruses are shown.

a function of HEF by several approaches: expression of the cloned HEF gene in vertebrate cells (Vlasak *et al.*, 1987) and analysis of the purified protein after isolation by detergent (Herrler *et al.*, 1988a) or protease treatment (Formanowski and Meier-Ewert, 1988).

The influenza C esterase belongs to the class of serine hydrolases which are inhibited by diisopropyl fluorophosphate (DFP). The inhibitor abolishes the enzyme activity without affecting the hemagglutinating activity (Muchmore and Varki, 1987). This finding suggests that the active site of the esterase and the receptor-binding site are different epitopes on the influenza C glycoprotein. There is some information on the amino acids which are crucial for the formation of the active site of the esterase. From the knowledge about other serine hydrolases (e.g., trypsin and chymotrypsin), it is expected that the enzyme mechanism involves a charge relay system, which is accomplished by a catalytic triad composed of the amino acids serine, histidine, and aspartic acid (Kraut, 1977). Taking advantage of the fact that DFP binds covalently to the serine in the active site of serine hydrolases, amino acid 71 of HEF has been identified as active-site serine (Herrler *et al.*, 1988b; Vlasak *et al.*, 1989). This amino acid is part of the sequence Phe–Gly–Asp–Ser–Arg (Fig. 1). While the motif Gly–Asp–Ser is found in the active site of many serine hydrolases, including trypsin and chymotrypsin, the following arginine residue has been detected so far only in the active site of the acetylesterases of influenza C virus and coronaviruses (see Section IV and Fig. 5). From inhibition studies with arginine-specific modifying reagents, it has been suggested that this arginine residue may be important for substrate recognition, possibly interacting with the carboxyl group of *N*-acetyl-9-*O*-acetylneuraminic acid (Neu5,9Ac$_2$) (Hayes and Varki, 1989).

Analysis of a series of compounds revealed that the esterase of influenza C virus has a high specificity for O-acetyl groups, Neu5,9Ac$_2$

TABLE I

SUBSTRATE SPECIFICITY OF THE ACETYLESTERASE OF INFLUENZA C/JHB/1/66[a]

Substrate	Relative cleavage rate (%)
N-Acetyl-9-*O*-acetylneuraminic acid	100
N-Acetyl-4-*O*-acetylneuraminic acid	3
N-Glycolyl-9-*O*-acetylneuraminic acid	33
N-Acetyl-7-*O*-acetylneuraminic acid	—
Bovine submandibular gland mucin	30
Rat serum glycoprotein	90
Rat erythrocytes	25
Equine submandibular gland mucin	—
4-Methylumbelliferyl acetate	220
4-Methylumbelliferyl butyrate	14
4-Nitrophenyl acetate	3500
α-Naphthyl acetate	2200

[a] From Schauer *et al.* (1988).

being hydrolyzed at the highest rate among all natural substrates tested (Schauer *et al.*, 1988) (see Table I). Some aromatic acetates (e.g., 4-nitrophenyl acetate or α-naphthyl acetate) are cleaved at higher rates than Neu5,9Ac$_2$ (Vlasak *et al.*, 1987; Schauer *et al.*, 1988; Wagaman *et al.*, 1989). These compounds are substrates for many serine hydrolases, including proteases, and therefore are not suited for determination of the enzyme specificity. They enable, however, fast and sensitive assays. α-Naphthyl acetate has been shown to be useful for cytochemical detection of influenza C-infected cells (Wagaman *et al.*, 1989). Treatment of erythrocytes with influenza C virus has been reported to change the reactivity of the cells with lectins specific for *N*-acetylgalactosamine (Luther *et al.*, 1988). From this finding it has been inferred that the receptor-destroying enzyme is able to release the acetyl residue of *N*-acetylgalactosamine. However, there is no direct chemical evidence supporting this conclusion.

Apart from DFP the esterase activity of influenza C virus is also inhibited by diethyl-4-nitrophenyl phosphate and some isocoumarins (Schauer *et al.*, 1988; Vlasak *et al.*, 1989). Inhibition of the esterase by DFP or isocoumarins has been reported to reduce the infectivity of the virus (Muchmore and Varki, 1987; Vlasak *et al.*, 1989). This finding may suggest that the receptor-destroying enzyme is required for virus entry into cells. However, both the hemagglutination (i.e., receptor-binding) and hemolytic (i.e., fusion) activities are not affected by the inactivation of the esterase (Muchmore and Varki, 1987; Vlasak *et al.*, 1989). Thus, more experiments are necessary to show whether the

TABLE II

RECEPTOR SPECIFICITY OF INFLUENZA A, B, AND C VIRUSES[a]

Sialic acid on human erythrocytes	HA titer (HA units/ml)[b]		
	C/JHB/1/66	B/HK/8/73	A/PR/8/34
Native	0	64	256
Asialo	0	0	0
Neu5Ac			
α2,3Galβ1,3GalNAc	0	2	256
α2,3Galβ1,4GlcNAc	0	128	128
α2,6Galβ1,4GlcNAc	0	64	128
Neu5Gc			
α2,6Galβ1,4GlcNAc	0	2	0
Neu5,9Ac$_2$			
α2,3Galβ1,3GalNAc	128	0	0
α2,3Galβ1,4GlcNAc	128	0	0
α2,6Galβ1,4GlcNAc	128	0	0

[a] Adapted from Rogers *et al.* (1986).

[b] 0, HA titer <2.

reduction of the infectivity is correlated with the inhibition of the enzyme activity or whether it is due to an indirect effect of the inhibitor.

B. Receptor-Binding Activity

The identification of the receptor-destroying enzyme as a sialate 9-*O*-acetylesterase implied that Neu5,9Ac$_2$ (see Fig. 3) is a crucial component of the cellular receptors for influenza C virus (Herrler *et al.*, 1985c). Direct evidence for the role of Neu5,9Ac$_2$ as a receptor determinant was provided by studies with erythrocytes which had been modified to contain only a single type of sialic acid. Influenza C virus was able to agglutinate erythrocytes which had been sialylated with Neu5,9Ac$_2$, but not cells containing *N*-acetyl- or *N*-glycolylneuraminic acid (Rogers *et al.*, 1986) (see Table II). On the basis of these results, it was possible to explain previous observations which seemed to argue against an involvement of sialic acid in the attachment of influenza C virus to cells. The resistance of the erythrocyte receptors to periodate treatment (Ohuchi *et al.*, 1978) is due to a greatly reduced oxidation of Neu5,9Ac$_2$ by periodate compared to Neu5Ac (Haverkamp *et al.*, 1975). The difficulty in inactivating the influenza C receptors with viral and bacterial neuraminidases (Hirst, 1950; Kendal, 1975; Herrler *et al.*, 1985a) is explained by the relative resistance of Neu5,9Ac$_2$ to the action of these enzymes (Corfield *et al.*, 1981). The importance of

$Neu5,9Ac_2$ as a receptor determinant is not restricted to erythrocytes. 9-*O*-Acetylated sialic acid is also required for influenza C virus to initiate the infection of cultured cells (Herrler and Klenk, 1987a). In fact, lack of this type of sialic acid is a major reason for the resistance of many cell lines to influenza C virus.

Insertion of artificial receptors into the plasma membrane of cultured cells rendered several resistant cells sensitive to an influenza C infection. Moreover, an increase in the yield of virus was observed with cells which usually produce only low amounts of virus (Herrler and Klenk, 1987a). The presence of 9-*O*-acetylated sialic acid appears to be the major factor in determining whether a glycoconjugate can serve as a receptor for influenza C virus. Erythrocytes which have been resialylated to contain $Neu5,9Ac_2$ were agglutinated by influenza C virus regardless of whether the sialic acid molecule was attached to galactose via an α-2,3 or α-2,6 linkage (Rogers *et al.*, 1986) (see Table II). The sialyltransferase specific for the latter linkage type only acts on glycoproteins. Therefore, receptors generated by this enzyme are glycoproteins. On the other hand, it has been shown that bovine brain gangliosides can also serve as receptors for influenza C virus, although the active species among the mixture of gangliosides has not been determined (Herrler and Klenk, 1987a,b). Thus, both glycoproteins and glycolipids can be used as receptors by influenza C virus, provided they contain $Neu5,9Ac_2$. A larger number of glycoconjugates must be analyzed, however, in order to know whether factors other than the presence of $Neu5,9Ac_2$ are important for the receptor function of a glycoconjugate.

It has been suggested that, in addition to $Neu5,9Ac_2$, influenza C virus may also recognize *N*-acetylgalactosamine (Luther *et al.*, 1988). The conclusion is based on the finding (mentioned in Section III,A) that erythrocytes treated with influenza C virus differ from control cells in their reactivity with lectins specific for *N*-acetylgalactosamine. However, direct evidence for such a receptor specificity is lacking.

The amino acids involved in the receptor-binding site of HEF have not been determined. Valuable information should be obtained by the analysis of mutants with a change in the receptor-binding activity. A mutant has been described which has an expanded cell tropism due to a more efficient recognition of $Neu5,9Ac_2$-containing receptors compared to the parent virus (Szepanski *et al.*, 1989). Sequence analysis of this mutant indicated that a single point mutation (Thr 284 to isoleucine; see Fig. 1) is responsible for the change in the receptor-binding activity (Szepanski *et al.*, 1991). Interestingly, the mutation site is located next to a sequence (Gly–Asn–Ser–Gly) which, in similar form

(Gly–Gln–Ser–Gly), is also found in several subtypes of influenza A hemagglutinins (Fig. 1). The homologous sequence in the H3 subtype composing amino acids 225–228 has been shown to be part of the receptor-binding pocket (Weis *et al.*, 1988). These data suggest that the amino acids Gly 279 to Thr 284 may be constituents of the receptor-binding site of HEF and that the mutant is altered at this site. The observation that these amino acids are located on the unfolded polypeptide at a distance far from Ser 71 at the catalytic center of the esterase, together with the DFP effects (Section III,A), supports the notion that receptor binding and receptor inactivation are exerted by different structural domains of HEF.

Another example of an influenza C virus with a change in the receptor-binding activity has been reported (Camilleri and Maassab, 1988). Virus isolated from persistently infected MDCK cells was found to be more sensitive to the action of hemagglutination inhibitors than was wild-type virus. Sequence analysis of more mutants or variants of this type should help further define the receptor-destroying and receptor-binding sites of HEF. Obviously, however, final answers to these problems can be given only when the three-dimensional structure of the glycoprotein is available. The importance of individual amino acids involved in the formation of the functional epitopes can then be evaluated by site-directed mutagenesis.

The ability of the influenza C glycoprotein to attach to Neu5,9Ac$_2$-containing receptors can be used as powerful tool to detect 9-*O*-acetylated sialic acid (Muchmore and Varki, 1987). The ability of influenza C virus to agglutinate erythrocytes from an adult chicken, but not those from a 1-day-old chicken, was the basis for the discovery that Neu5,9Ac$_2$ is a differentiation marker on chicken erythrocytes, which has been confirmed by chemical analysis of the sialic acids on these cells (Herrler *et al.*, 1987). The sensitivity of the receptor recognition by influenza C virus is evident from studies with human erythrocytes. By chemical analysis only Neu5Ac has been detected, not Neu5,9Ac$_2$ (Shukla and Schauer, 1982). Agglutination and binding studies indicate, however, that erythrocytes from some individuals contain low levels of 9-*O*-acetylated sialic acid on their surface (Ohuchi *et al.*, 1978; Nishimura *et al.*, 1988).

C. Membrane Fusion

The fusion activity of influenza C virus was first demonstrated with erythrocytes. Microscopic observation of virus-induced cell fusion and photometric detection of hemolysis indicated that the virus is able to fuse with mouse and chicken erythrocytes (Ohuchi *et al.*, 1982; Kitame *et al.*, 1982). Recently, the fusion between virus membranes and artificial membranes has been monitored using a resonance energy assay (Formanowski *et al.*, 1990). In contrast to the hemagglutinating (Herrler *et al.*, 1979; Sugawara *et al.*, 1981) and esterase activities of HEF (Herrler *et al.*, 1988a), the fusion activity requires the proteolytic cleavage of HEF_0 into polypeptides HEF_1 and HEF_2 (Ohuchi *et al.*, 1982; Kitame *et al.*, 1982), described in Section II,B. The dependence of the influenza C virus-induced fusion on the cleavage of HEF indicated that this activity is a function of the surface glycoprotein. Virus with uncleaved glycoprotein can be rendered fusiogenic by *in vitro* cleavage of HEF. Both trypsin and elastase have been shown to be effective in this respect, whereas other proteases (e.g., chymotrypsin and thermolysin) were unable to activate the glycoprotein (Kitame *et al.*, 1982; Ohuchi *et al.*, 1982; Formanowski *et al.*, 1990).

An additional characteristic of the fusion activity is pH dependence. Similar to influenza A and B viruses and several other viruses, influenza C virus causes fusion only at a low pH. Optimal pH values for hemolysis of erythrocytes vary between 5.0 and 5.7, depending on the virus strain. Optimal fusion between virus and unilamellar liposomes was detected in the range of 5.6–6.1. Several changes have been observed when the glycoprotein is shifted from neutral to acidic pH values: (1) The glycoprotein becomes susceptible to trypsin digestion; (2) the endogenous tryptophan fluorescence decreases; and (3) the hexagonal arrangement of the surface projections disappears (Formanowski *et al.*, 1990). These changes, which were only observed with virus containing the cleaved HEF (i.e., $HEF_{1,2}$), suggest that exposure to a low pH results in a conformational change of the glycoprotein.

The characteristics of the influenza C virus-induced fusion described so far (i.e., a dependence on both proteolytic cleavage and low pH and a conformational change at low pH) are very similar to those reported for the fusion activity of influenza A and B viruses. It is therefore likely that fusion occurs by a similar mechanism for all influenza viruses. With influenza A virus it is widely accepted that the conformational change observed at acidic pH results in the exposure of the amino-terminal portion of the membrane-bound cleavage product (HA_2). This part of the protein is made up of a stretch of hydrophobic amino acids, which probably interact with the cellular membrane, thereby inducing fusion between the viral envelope and the membrane of the target cell. This model is also applicable to influenza C virus.

Differences between influenza A and C viruses have been observed so far as the kinetics of the fusion process are concerned. In the case of influenza A virus, the conformational change is fast and a later step is rate limiting. With influenza C virus the conformational change has been found to be a rate-limiting step (Formanowski *et al.*, 1990). The reason for the delayed conformational change may be related to the

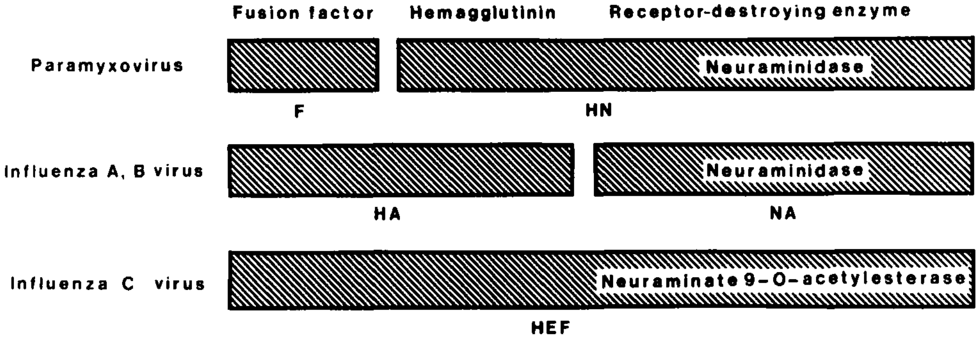

FIG. 4. The glycoproteins of paramyxoviruses (HN and F), influenza A and B viruses (HA and NA), and influenza C virus (HEF), illustrating differences in the distribution of biological activities (i.e., fusion, hemagglutination, and receptor inactivation). The sizes of the boxes representing the individual glycoproteins are not proportional to the molecular weights.

hexagonal arrangement of the spikes. The close packing of the glycoproteins might be a hindrance in adopting the conformation required for fusion.

In the course of virus infection, the viral fusion activity is crucial for the penetration of enveloped viruses. Viruses with a pH-dependent fusion activity are generally assumed to enter a cell via endosomes. The acidic pH within such vesicles triggers the fusion reaction, resulting in the release of the nucleocapsid into the cytoplasm. This may also apply to influenza C virus, although no evidence has been presented to support this assumption. In any case the fusion activity is crucial for the infectivity of the virus. Virus with uncleaved glycoprotein is lacking not only fusion activity, but also infectivity (Herrler *et al.*, 1979; Sugawara *et al.*, 1981). Restoration of the fusion activity *in vitro* by proteolytic cleavage of the glycoprotein is accompanied by restoration of the infectivity.

Due to the characteristics of the glycoprotein HEF, influenza C virus is unique among myxoviruses. Influenza A and B viruses as well as paramyxoviruses differ from influenza C virus in the specificity of the receptor-binding activity (Neu5Ac versus Neu5,9Ac$_2$) and the receptor-destroying enzyme (neuraminidase versus acetylesterase). In addition, HEF is responsible for three activities (receptor binding, receptor inactivation, and fusion), while both paramyxoviruses and influenza A and B viruses have two surface glycoproteins for these activities (Fig. 4). The unique characteristics of the influenza C glycoprotein are reflected in the designation "HEF," which has been proposed to indicate that this protein can function as a hemagglutinin, as an esterase, and as fusion factor (Herrler *et al.*, 1988a). Others have chosen the designation "HE" (Vlasak *et al.*, 1987), which ignores the fusion activity. In addition, there is an HE protein present on some coronaviruses (Cavanagh *et al.*, 1990). This protein, described in Section IV, has hemagglutinating and esterase activities, but no fusion activity. Thus, "HEF" is an appropriate designation for the influenza C glycoprotein, to distinguish it from the coronavirus glycoprotein.

IV. RELATIONSHIP BETWEEN HEF AND THE CORONAVIRUS GLYCOPROTEIN HE

For many years ortho- and paramyxoviruses have been thought to be the only animal viruses containing receptor-destroying enzymes. Prompted by a sequence similarity between an open reading frame on the genome of mouse hepatitis virus and the HEF gene of influenza C virus (Luytjes *et al.*, 1988) it was found that bovine coronavirus (BCV) is able to inactivate its own receptors on erythrocytes (Vlasak *et al.*, 1988a). The enzyme turned out to be a sialate 9-*O*-acetylesterase similar to the receptor-destroying enzyme of influenza C virus. In fact, the coronavirus enzyme was able to inactivate the receptors for influenza C virus on erythrocytes, and the esterase of influenza C virus inactivated the receptors for BCV (Vlasak *et al.*, 1988a), indicating that both viruses use the same receptor determinant for attachment to cells (i.e., Neu5,9Ac$_2$). This conclusion was confirmed by resialylation studies with erythrocytes and has been extended to a porcine coronavirus, hemagglutinating encephalomyelitis virus (HEV) (Schultze *et al.*, 1990).

An acetylesterase activity has been reported not only for BCV but also for HEV and some strains of mouse hepatitis virus (Yokomori *et al.*, 1989; Schultze *et al.*, 1991a; Pfleiderer *et al.*, 1991). The acetylesterase activity of BCV has been shown to be a function of a surface glycoprotein which is detected as a disulfide-linked dimer with a molecular weight of about 140,000. (Vlasak *et al.*, 1988b; Schultze *et al.*, 1991a). The same protein has been identified previously as a hemagglutinin (King *et al.*, 1985), and therefore the designation "HE" has been chosen to indicate its dual function as hemagglutinin and esterase (Cavanagh *et al.*, 1990). Similar to its influenza C counterpart, the esterase of coronaviruses is a serine esterase which can be inhibited by DFP (Vlasak *et al.*, 1988b; Schultze *et al.*, 1990, 1991a). Inhibition of the enzyme activity results in a dramatic reduction of infectivity, suggesting an important role for the esterase in an early stage of the infection (Vlasak *et al.*, 1988b). The importance of the HE protein has also been demonstrated with monoclonal antibodies which

```
             .         .         .         .         .         .
1   MFSSLLLVLGLTEAEKIKICLQKQVNSSFSLHNGFGGNLYATEEKRMFELVKPKAGASVL  60
    ~~~~~~~~~~~~              || ||  |    |   |   ||||  |    | |
2                            MGSTCIAMAPRTLLLLIGCQLVFGFNEPLNIVSHL  35
                                 ~~~~~~~~~~~~~

             .▽        .         .         .         .         .
1   NQSTWIGFGDSRTD......KSNSAFPRSADVSAK.TADKFRFLSGGSLMLSMFGPPGKV 113
    ||  || ||||| |       ||  | |          |||   || ||| ||
2   ND.DWFLFGDSRSDCTYVENNGHPKLDWLDLDPKLCNSGKISAKSGNSLFRSFHFTDF..  92
           -----

             .         .         .         .         .         .
1   DYLYQGCGKHKVFYEGVNWSPHAAINCY....RKNWTDIKLNFQKNIYELASQSHCMSLV 169
     | | | | | ||||||||||| || |      | | | |  |   |||   | ||||||
2   .YNYTGEGDQIVFYEGVNFSPNHGFKCLAYGDNKRWMGNKARFYARVYEKMAQYRSLSFV 151
               -------

             .         .         .         .         .         .
1   NALDKTIPLQVTAGTAGNCNNSFLKNPALYTQEVKPSENKCGKENLAFFTLPTQFGTYEC 229
    |            || ||     | || |   |   ||  || |          |   ||
2   NVPYAYGGKAKPTSICKH.KTLTLNNPTFISKESNYVDYYYESE........ANFTLAGC 202

             .         .         .         .         .         .
1   KLHLVASCYFIYDSKEVYNKRGCDNYF..QVIYDSFGKVVGGLDNRVSPYTGNSGDTPTM 287
       ||| | |   |||     |   ||  |  ||    || |||  |       |    |
2   DEFIVPLCVFNGHSKGSSSDPANKYYMDSQSYYNMDTGVLYGFNSTLDVGNTAKDPGLDL 262

             .         .         .         .         .         .
1   QCDMLQLKPGRYSVRSSPRFLLMPERSYCFDMKEK.GPVTAVQSIWGKGRESDYAVDQAC 346
     |  | | || |   |   || || |  ||    |  ||  ||| |   | ||     ||
2   TCRYLALTPGNYKAVSLEYLLSLPSKAICLRKPKRFMPVQVVDSRWNSTRQSDNMTAVAC 322

             .         .         .         .         .         .
1   LSTPGCMLIQKQKPYIGEADDHHGDQEMRELLSGLDYEARCISQSG..WVNETSPFTEKY 404
          |         |  | ||||  ||||||||  |  || | |  | || |     |
2   QLPYCFFRNTSADYSGGTHDVHHGDFHFRQLLSGLLLNVSCIAQQGAFLYNNVSSSWPAY 382
                         ----     -----

             .         .         .         .         .         .
1   LLPPKFGRCPLAAKEESIPKIPDGLLIPTSGTDTTVTKPKSRIPGIDDLIIGLLFVAIVE 464
         || || ||  | || |     ||    |  |      ~~~~~~~~~~~~~~~~~
2   ....GYGQCPTAANIGYMAPVCIYDPLPVVLLGVLLGIAVLIIVFLILYFMTDSGVRLHE 438
                             ~~~~~~~~~~~~~~~~~~~~~~~~~~

             .         .         .         .         .         .
1   TGIGGYLLGSRKESGGGVTKESAEKGFEKIGNDIQILKSSINIAIEKLNDRISHDENAIR 524

2   A 439

             .         .         .         .         .         .
1   DLTLEIENARSEALLGELGIIRALLVGNISIGLQESLWELASEITNRAGDLAVEVSPGCW 584

             .         .         .         .         .         .
1   IIDNNICDQSCQNFIFKFNETAPVPTIPPLDTKIDLQSDPFYWGSSLGLAITATISLAAL 644
                                              ~~~~~~~~~~~~~~~~~~

             .
1   VISGIAICRTK 655
    ~~~~~~~~
```

FIG. 5. Alignment of the amino acid sequences of the HEF protein of influenza C/JHB/1/66 (1) and the HE protein of mouse hepatitis virus, strain JHM (2). Every tenth position is indicated by a dot. Wavy lines indicate hydrophobic sequences. With HEF these compose the amino-terminal signal sequence (amino acids 1–12), the presumptive fusion peptide (amino acids 447–463), and the membrane anchor (amino acids 624–652). HE has only two hydrophobic regions, the signal peptide (amino acids 1–17) and the membrane anchor (amino acids 404–429), which do not align with the corresponding domains of HEF. The active-site serine of the acetylesterase, which is conserved in both proteins, is marked with an open triangle. Stretches of amino acids identical in both sequences are indicated by underscoring. Vertical lines indicate identical or related amino acids. The figure is based on an alignment of HEF_1 and HE. (Courtesy of S. G. Siddell.)

were shown to neutralize BCV both *in vivo* and *in vitro* (Deregt *et al.*, 1989).

When the amino acid sequence of the HE protein is aligned with the sequence of the influenza C glycoprotein HEF, homology is observed only with the HEF_1 cleavage product. There is no sequence on the HE protein which is related to the HEF_2 polypeptide (Fig. 5). This observation is not surprising, because HEF_2 is responsible for the fusion activity of influenza C virus, whereas, in the case of coronaviruses, fusion is a function not of HE, but of the S protein (reviewed by Spaan *et al.*, 1988). The homology between the amino acid sequences of HE and HEF_1 has been reported to be 30% (Luytjes *et al.*, 1988). The alignment indicates that there are many conservative substitutions. A few regions are completely identical in both sequences (Fig. 5). Among these is the sequence Phe–Gly–Asp–Ser–Arg, which, in the case of influenza C virus, has been shown to contain the active-site serine of the esterase (Herrler *et al.*, 1988b; Vlasak *et al.*, 1989).

It is interesting to note that, on the other hand, the putative constituent sequence of the HEF receptor-binding site (Gly 279–Thr 284) does not have a homologous counterpart in the HE sequence. This observation may be related to the recent finding that HE is not very efficient in agglutinating erythrocytes (Schultze *et al.*, 1991a) and that the major hemagglutinin of BCV is the peplomer glycoprotein S (Schultze *et al.*, 1991b). It has been argued that the extent of identity between HE and HEF_1 is high enough to rule out convergent evolution, and, therefore, it has been speculated that coronaviruses acquired the HE gene from influenza C virus by nonhomologous recombination between ancestors of both viruses (Luytjes *et al.*, 1988). However, acetylesterases are also found in cells. If coronaviruses actually acquired the esterase gene by a recombination event, the gene might as well be derived from a cellular gene. More information about the viral and cellular esterases is required to distinguish between these possibilities.

REFERENCES

Adachi, K., Kitame, F., Sugawara, K., Nishimura, H., and Nakamura, K. (1989). *Virology* **172,** 125–133.

Buonagurio, D. A., Nakada, S., Desselberger, U., Krystal, M., and Palese, P. (1985). *Virology* **146,** 221–232.

Buonagurio, D. A., Nakada, S., Fitch, W. M., and Palese, P. (1986). *Virology* **153,** 12–21.

Camilleri, J. J., and Maassab, H. F. (1988). *Intervirology* **29,** 178–184.

Cavanagh, D., Brian, D. A., Enjuanes, L., Holmes, K. V., Lai, M. M. C., Laude, H., Siddell, S. G., Spaan, W., Taguchi, F., and Talbot, P. J. (1990). *Virology* **176,** 306–307.

Chakraverty, P. (1974). *J. Gen. Virol.* **25,** 421–425.

Chakraverty, P. (1978). *Arch. Virol.* **58,** 341–348.

Compans, R. W., Bishop, D. H. L., and Meier-Ewert, H. (1977). *J. Virol.* **21,** 658–665.

Corfield, A. P., Michalski, J. C., and Schauer, R. (1981). *In* "Sialidases and Sialidoses. Perspectives in Inherited Metabolic Diseases" (G. Tettamanti, P. Durand, and S. Di Donato, eds.), Vol. 4, pp. 3–70. Edi Fermes, Milan, Italy.

Czekalowski, J. W., and Prasad, A. K. (1973). *Arch. Gesamte Virusforsch.* **42,** 215–227.

Deregt, D., Gifford, G. A., Ijaz, M. K., Watts, T. C., Gilchrist, J. E., Haines, D. M., and Babiuk, L. A. (1989). *J. Gen. Virol.* **70,** 993–998.

Flewett, T. H., and Apostolov, K. (1967). *J. Gen. Virol.* **1,** 297–304.

Formanowski, F., and Meier-Ewert, H. (1988). *Virus Res.* **10,** 177–192.

Formanowski, F., Wrigley, N. G., and Meier-Ewert, H. (1989). *In* "Genetics and Pathogenicity of Negative Strand Viruses" (B. W. J. Mahy and D. Kolakofsky, eds.), pp. 16–23. Elsevier, Amsterdam.

Formanowski, F., Wharton, S. A., Calder, L. J., Hofbauer, C., and Meier-Ewert, H. (1990). *J. Gen. Virol.* **71,** 1181–1188.

Hachinohe, S., Sugawara, K., Nishimura, H., Kitame, F., and Nakamura, K. (1989). *J. Gen. Virol.* **70,** 1287–1292.

Haverkamp, J., Schauer, R., Wember, M., Kamerling, J. P., and Vliegenthart, J. F. G. (1975). *Hoppe-Seyler's Z. Physiol. Chem.* **365,** 1575–1583.

Hayes, B. K., and Varki, A. (1989). *J. Biol. Chem.* **264,** 19443–19448.

Herrler, G., and Klenk, H.-D. (1987a). *Virology* **159,** 102–108.

Herrler, G., and Klenk, H.-D. (1987b). *In* "The Biology of Negative Strand Viruses" (B. W. J. Mahy and D. Kolakofsky, eds.), pp. 63–67. Elsevier, Amsterdam.

Herrler, G., Compans, R. W., and Meier-Ewert, H. (1979). *Virology* **99,** 49–56.

Herrler, G., Nagele, A., Meier-Ewert, H., Bhown, A. S., and Compans, R. W. (1981). *Virology* **113,** 439–451.

Herrler, G., Rott, R., and Klenk, H.-D. (1985a). *Virology* **159,** 102–108.

Herrler, G., Geyer, R., Müller, H.-P., Stirm, S., and Klenk, H.-D. (1985b). *Virus Res.* **2,** 183–192.

Herrler, G., Rott, R., Klenk, H.-D., Müller, H.-P., Shukla, A. K., and Schauer, R. (1985c). *EMBO J.* **4,** 1503–1506.

Herrler, G., Reuter, G., Rott, R., Klenk, H.-D., and Schauer, R. (1987). *Biol. Chem. Hoppe-Seyler* **368,** 451–454.

Herrler, G., Dürkop, I., Becht, H., and Klenk, H.-D. (1988a). *J. Gen. Virol.* **69,** 839–846.

Herrler, G., Multhaup, G., Beyreuther, K., and Klenk, H.-D. (1988b). *Arch. Virol.* **102,** 269–274.

Hewat, E. A., Cusack, S., and Verwey, C. (1984). *J. Mol. Biol.* **175,** 175–193.

Hirst, G. K. (1950). *J. Exp. Med.* **91,** 177–185.

Hongo, S., Sugawara, K., Homma, M., and Nakamura, K. (1986a). *Arch. Virol.* **89,** 171–187.

Hongo, S., Sugawara, K., Homma, M., and Nakamura, K. (1986b). *Arch. Virol.* **89,** 189–201.

Kawamura, H., Tashiro, M., Kitame, F., Homma, M., and Nakamura, K. (1986). *Virus Res.* **4,** 275–288.

Kendal, A. P. (1975). *Virology* **65,** 87–99.

King, B., Potts, B. J., and Brian, D. A. (1985). *Virus Res.* **2,** 53–59.

Kitame, F., Sugawara, K., Ohwada, K., and Homma, M. (1982). *Arch. Virol.* **73,** 357–361.

Kitame, F., Nakamura, K., Saito, A., Sinohara, H., and Homma, M. (1985). *Virus Res.* **3,** 231–244.

Klenk, E., Faillard, H., and Lempfrid, H. (1955). *Hoppe-Seyler's Z. Physiol. Chem.* **301,** 235–246.

Kraut, J. (1977). *Annu. Rev. Biochem.* **46,** 331–358.

Krystal, M., Elliot, R. M., Benz, E. W., Young, J. F., and Palese, P. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79,** 4800–4804.

Luther, P., Cushley, W., Hölzer, C., Desselberger, U., and Oxford, J. S. (1988). *Arch. Virol.* **101,** 247–254.

Luytjes, W., Bredenbeek, P. J., Noten, A. F., Horzinek, M. C., and Spaan, W. J. (1988). *Virology* **166,** 415–422.

Meier-Ewert, H., Compans, R. W., Bishop, D. H. L., and Herrler, G. (1978). *In* "Negative Strand Viruses and the Host Cell" (B. W. J. Mahy and R. D. Barry, eds.), pp. 127–133. Academic Press, New York.

Meier-Ewert, H., Herrler, G., Nagele, A., and Compans, R. W. (1980). *In* "Structure and Variation in Influenza Virus" (W. G. Laver and G. M. Air, eds.), pp. 357–366. Elsevier/North-Holland, Amsterdam.

Meier-Ewert, H., Nagele, A., Herrler, G., Basak, S., and Compans, R. W. (1981a). *In* "Replication of Negative Strand Viruses" (D. H. L. Bishop and R. W. Compans, eds.), pp. 173–180. Elsevier/North-Holland, Amsterdam.

Meier-Ewert, H., Nagele, A., Herrler, G., Basak, S., and Compans, R. W. (1981b). *In* "Genetic Variation among Influenza Virus" (D. P. Nayak and C. F. Fox, eds.), pp. 263–272. Academic Press, New York.

Meier-Ewert, H., Petri, G., and Bishop, D. H. L. (1981c). *Arch. Virol.* **67,** 141–147.

Minuse, E., Quilligan, J. J., and Francis, T., Jr. (1954). *J. Lab. Clin. Med.* **43,** 31–43.

Muchmore, E. A., and Varki, A. (1987). *Science* **236,** 1293–1295.

Nagele, A. (1983). Ph.D. thesis, Technical University, Munich.

Nakada, S., Creager, R. S., Krystal, M., Aaronson, R. P., and Palese, P. (1984). *J. Virol.* **50,** 118–124.

Nakamura, K., Herrler, G., Petri, T., Meier-Ewert, H., and Compans, R. W. 91979). *J. Virol.* **29,** 997–1005.

Nerome, K., Ishida, M., and Nakayama, M. (1976). *Arch. Virol.* **50,** 241–244.

Nishimura, H., Sugawara, K., Kitame, F., and Nakamura, K. (1988). *J. Gen. Virol.* **69,** 2545–2553.

O'Callaghan, R. J., Gohd, R. S., and Labat, D. D. (1980). *Infect. Immun.* **30,** 500–505.

Ohuchi, M., Homma, M., Muramatsu, M., and Ohyama, S. (1978). *Microbiol. Immunol.* **22,** 197–203.

Ohuchi, M., Ohuchi, R., and Mifune, K. (1982). *J. Virol.* **42,** 1076–1079.

Pfeifer, J. B., and Compans, R. W. (1984). *Virus Res.* **1,** 281–296.

Pfleiderer, M., Routledge, E., Herrler, G., and Siddell, S. (1991). *J. Gen. Virol.* **72,** 1309–1315.

Rogers, G. N., Herrler, G., Paulson, J. C., and Klenk, H.-D. (1986). *J. Biol. Chem.* **261,** 5947–5951.

Rosenthal, P. B., Formanowski, F., Skehel, J. J., Meier-Ewert, H., and Wiley, D. C. (1990). *Int. Congr. Virol., 8th* **P17–025,** 218.

Schauer, R., Reuter, G., Stoll, S., Posadas del Rio, F., Herrler, G., and Klenk, H.-D. (1988). *Biol. Chem. Hoppe-Seyler* **369,** 1121–1130.

Schultze, B., Gross, H.-J., Brossmer, R., Klenk, H.-D., and Herrler, G. (1990). *Virus Res.* **16,** 185–194.

Schultze, B., Wahn, K., Klenk, H.-D., and Herrler, G. (1991a). *Virology* **180,** 221–228.

Schultze, B., Gross, H. J., Brossmer, R., and Herrler, G. (1991b). Submitted for publication.

Shukla, A. K., and Schauer, R. (1982). *Hoppe-Seyler's Z. Physiol. Chem.* **363,** 255–262.

Spaan, W., Cavanagh, D., and Horzinek, M. C. (1988). *J. Gen. Virol.* **69,** 2939–2952.

Styk, B. (1955). *Folia Biol. (Prague)* **1,** 207–212.

Sugawara, K., Ohuchi, M., Nakamura, K., and Homma, M. (1981). *Arch. Virol.* **68,** 147–151.

Sugawara, K., Nishimura, H., Kitame, F., and Nakamura, K. (1986).*Virus Res.* **6,** 27–32.

Sugawara, K., Kitame, F., Nishimura, H., and Nakamura, K. (1988). *J. Gen. Virol.* **69,** 537–547.

Szepanski, S., Klenk, H.-D., and Herrler, H. (1989). *In* "Cell Biology of Virus Entry, Replication, and Pathogenesis" (R. W. Compans, A. Helenius, and M. B. A. Oldstone, eds.), pp. 125–134. Liss, New York.

Szepanski, S., Gross, H. J., Brossmer, R., Klenk, H.-D., and Herrler, G. (1991). Submitted for publication.

Taylor, R. M. (1949). *Am. J. Public Health* **39,** 171–178.

Veit, M., Herrler, G., Schmidt, M. F. G., Rott, R., and Klenk, H.-D. (1990). *Virology* **177,** 807–811.

Vlasak, R., Krystal, M., Nacht, M., and Palese, P. (1987). *Virology* **160,** 419–425.

Vlasak, R., Luytjes, W., Spaan, W., and Palese, P. (1988a). *Proc. Natl. Acad. Sci. U.S.A.* **85,** 4526–4529.

Vlasak, R., Luytjes, W., Leider, J., Spaan, W., and Palese, P. (1988b). *J. Virol.* **62,** 4686–4690.

Vlasak, R., Muster, T., Lauro, A. M., Powers, J. C., and Palese, P. (1989). *J. Virol.* **63,** 2056–2062.

Wagaman, P. C., Spence, H. A., and O'Callaghan, R. J., (1989). *J. Clin. Microbiol.* **27,** 832–836.

Weis, W., Brown, J. H., Cusack, S., Paulson, J. C., Skehel, J. J., and Wiley, D. C. (1988). *Nature (London)* **333,** 426–431.

Yamashita, M., Krystal, M., and Palese, P. (1989). *Virology* **171,** 458–466.

Yokomori, K., La Monica, M., Makino, S., Shieh, C.-K., and Lai, M. M. C. (1989). *Virology* **173,** 683–691.