

Molecular characterization of gp40, a mucin-type glycoprotein from the apical plasma membrane of Madin–Darby canine kidney cells (type I)

Gert ZIMMER*, Friedrich LOTTSPEICH†, Andrea MAISNER*, Hans-Dieter KLENK*, and Georg HERRLER*¹

*Institut für Virologie der Philipps-Universität Marburg, Robert-Koch-Straße 17, D-35037 Marburg, Germany and †Max-Planck-Institut für Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, Germany

gp40 has been recently identified as a major apical cell-surface sialoglycoprotein of type-I Madin–Darby canine kidney cells, a cell line widely used for the study of polarized transport. The determination of two internal amino acid sequences of the purified glycoprotein by Edman degradation enabled us to isolate the cDNA encoding the 18.6 kDa protein backbone of gp40. Sequence analysis revealed that gp40 is a type-I membrane protein which has several characteristics in common with glycoporphin A and other mucin-type glycoproteins. At least 14 serine/threonine residues were found to be used for O-glycosylation. No potential sites for N-glycosylation were detected. gp40 turned out to represent the canine homologue of

a cell-surface antigen expressed by various epithelial and non-epithelial cells in rat and mouse. Potential O-glycosylation sites, transmembrane and cytoplasmic domains were found to be highly conserved in the three species. gp40 was detected in canine lung, intestine, kidney, brain and heart but not in liver and spleen. The subline II of Madin–Darby canine kidney cells was found not to express gp40. Stable expression of gp40 in transfected type-II cells revealed that gp40 is predominantly delivered to the apical plasma membrane. N-Glycans and a glycosylphosphatidylinositol anchor, both proposed apical targeting signals, are absent from gp40, indicating that other determinants are responsible for its polarized transport.

INTRODUCTION

Polarized epithelial cells are characterized by two distinct plasma-membrane domains differing in composition and function [1–3]. A typical feature of the apical membrane is the high proportion of glycosphingolipids in its outer leaflet. It is thought that these lipids protect the membrane against the harsh conditions of the external environment with which the apical domain is in contact. In addition, this domain harbours unique proteins with specific functions characteristic of the respective cell type. In contrast, the basolateral domain exhibits mostly constitutive functions such as nutrient uptake from the circulation, cell–matrix attachment and growth control. Therefore, this plasma-membrane domain contains proteins that are found in many epithelial and non-epithelial cells. The lateral membrane contains epithelium-specific proteins that are responsible for intercellular contacts and junctions. Most important for the maintenance of cell polarity are the tight junctions because they prevent the intermixing of apical and basolateral components.

Madin–Darby canine kidney (MDCK) cells present one of the best-studied model systems for the understanding of cell polarity and protein transport. The cell line was isolated from dog kidney by Madin and Darby in 1958 and is thought to originate from the distal tubule or the collecting duct of the nephron [4–6]. Two sublines of MDCK cells have been described [5,7,8]: MDCK I cells are derived from an early passage of the cell population and MDCK II cells predominate in later passages [5,8]. The two sublines are distinct and stable in culture. They can be distinguished by morphological characteristics [5,9], but the most striking difference between the two strains is their transepithelial electrical resistance. MDCK I cells resemble tight epithelia by developing a very high electrical resistance whereas much lower values are measured across MDCK II cell monolayers [8]. Other

differences between the two MDCK sublines have been described. For example, MDCK I cells display an adenylate cyclase sensitivity to adrenaline, vasopressin, and prostaglandin E₁, which is not found in MDCK II cells. On the other hand, alkaline phosphatase, γ -glutamyltransferase and Na⁺/K⁺-ATPase activities have been detected only in MDCK II cells [8]. The glycosphingolipid composition of the two sublines has been shown to differ as well [10]. The Forssman antigen has only been detected in MDCK II cells, indicating that different sets of glycosyltransferases are expressed in MDCK I and II cells. We have reported that MDCK I but not MDCK II cells are susceptible to infection by influenza C virus [11,12]. The resistance of MDCK II cells to infection was shown to be due to the lack of 9-O-acetylated sialic acid, the receptor determinant for influenza C virus. The protein composition of the apical and basolateral plasma membranes of polarized MDCK cells has hardly been characterized at all. A few antibodies directed to cell-surface antigens have been described [13–15], but none of these proteins has so far been cloned. Most studies on protein trafficking in MDCK cells have been carried out using a restricted set of heterologous proteins including influenza A virus haemagglutinin and vesicular stomatitis virus G-protein [16].

We recently identified a 40 kDa sialoglycoprotein that is the major cell-surface protein of MDCK I cells recognized by influenza C virus via 9-O-acetylated sialic acids [17]. This protein, designated gp40, was characterized as an O-glycosylated (mucin-type) integral membrane protein, which predominantly resides in the apical plasma membrane of filter-grown cells. gp40 was shown to be subjected to constitutive endocytosis with internalization kinetics comparable with those of influenza C virus. To date, endocytic receptors of the apical plasma membrane have not been described and the signals responsible for their concentration in coated pits are not known. Furthermore

Abbreviations used: MDCK cells, Madin–Darby canine kidney cells; OG, n-octyl glucoside; WGA, wheat germ agglutinin; DABCO, 1,4-bis[2-(2,2,2-octyl)octane]; ECL, enhanced chemiluminescence; RT, reverse transcriptase; TBS, Tris-buffered saline (50 mM Tris/HCl/150 mM NaCl, pH 7.4); GPI, glycosylphosphatidylinositol; PMA, phorbol 12-myristate 13-acetate.

¹ To whom correspondence should be addressed.

information on apical sorting signals is scarce. Although proteins attached to the membrane via a glycosylphosphatidylinositol (GPI) anchor are delivered to the apical surface [16,18,19], this does not explain the targeting of transmembrane proteins. N-Glycosylation has been shown to be a potential apical transport signal for secretory proteins [20,21]. However, the role of N-glycosylation in apical transport of membrane proteins has not been elucidated and the situation of proteins that are predominantly or even exclusively O-glycosylated remains unclear. gp40 has several interesting properties that make it a promising subject for the analysis of apical transport and endocytosis. In the present study, we report its purification, molecular cloning and tissue distribution. We show that gp40 is a type-I membrane protein which is heavily O-glycosylated but lacks N-glycans. We found it to be the canine homologue of a protein previously described as an antigen in a variety of murine and rat cell systems of epithelial and non-epithelial origin. MDCK II cells are shown not to express endogenous gp40, providing the possibility of analysing gp40 transport in this closely related polarized cell line. We demonstrate that MDCK II cells transformed to stably express gp40 deliver this glycoprotein to the apical plasma membrane.

MATERIALS AND METHODS

Materials

The following reagents were from Boehringer-Mannheim GmbH (Mannheim, Germany): n-octyl glucoside (OG), hygromycin B, Complete[™] protease inhibitor cocktail, positively charged nylon membrane. The following reagents were purchased from Sigma (Deisenhofen, Germany): concanavalin A-agarose, wheat germ agglutinin (WGA)-agarose, Jacalin agarose, sialidase from *Clostridium perfringens*, methyl α -D-galactopyranoside, N-acetyl-D-glucosamine, methyl α -D-mannopyranoside, Protein A-Sepharose, PMA, 1,4-diazabicyclo[2.2.2]octane (DABCO), n-lauroylsarcosine. The following reagents were from Amersham Buchler GmbH (Braunschweig, Germany): enhanced chemiluminescence (ECL) detection reagent, streptavidin-horseradish peroxidase, biotinylated anti-rabbit immunoglobulin from donkey, FITC-linked anti-rabbit immunoglobulin from donkey, prestained protein molecular-mass markers. Sulphosuccinimido-biotin was obtained from Pierce Chemical Co., and Mowiol 40-88 was from Aldrich Chemie (Steinheim, Germany). *Taq* DNA polymerase (5 units/ μ l) and murine leukaemia virus reverse transcriptase (RT; 50 units/ μ l) were purchased from Perkin-Elmer.

Cell culture

MDCK I and MDCK II cells were provided by Dr. K. Simons (EMBL, Heidelberg, Germany) and were grown in Eagle's minimal essential medium (Gibco-BRL) supplemented with 10% foetal calf serum (Gibco-BRL). Porcine kidney cells (LLC-PK1) and human malignant melanoma cells (SK-MEL-28) were obtained from ATCC and maintained in Dulbecco's minimal essential medium (Gibco-BRL) supplemented with 10% foetal calf serum. For studies of cell polarity, 0.4 μ m pore size polycarbonate filters (Costar, Cambridge, MA, U.S.A.) were used. MDCK II cells stably expressing gp40 (see below) were seeded 4 or 5 days before experiments (2×10^6 cells per 24 mm unit and 2×10^5 cells per 7.5 mm unit). The polarization of the monolayer was assessed by measuring the electrical resistance between the apical and basolateral compartments of the filter chamber using a Millicell-ERS instrument (Millipore). Only filter cultures with an electrical resistance of at least 130 Ω cm² were used for experiments.

Purification of gp40

About 10^{10} MDCK I cells were grown to confluence in 145 mm Petri dishes, washed twice with ice-cold PBS and scraped with a rubber 'policeman' into 10 ml of ice-cold PBS. The cells were pelleted by centrifugation (600 g; 10 min; 4 °C), suspended in 50 ml of Tris-buffered saline (TBS: 50 mM Tris/HCl, 150 mM NaCl, pH 7.4), slowly added to 1 litre of 20 mM Tris/HCl, and incubated on ice for 20 min. The cells were disrupted by 30 strokes in a Dounce homogenizer and the nuclei were removed by low-speed centrifugation (270 g; 10 min; 4 °C). Membranes were pelleted from the supernatant by ultracentrifugation (105 000 g; 90 min; 4 °C) and suspended in 250 ml of ice-cold TBS supplemented with Complete[™] protease inhibitor cocktail by passing them three times through a 22-gauge needle. An equal volume of 1.5% OG in TBS was added and solubilization of membrane proteins was allowed to proceed for 2 h at 4 °C. The membrane lysate was clarified by centrifugation (105 000 g; 60 min; 4 °C). The lysate was pumped at 10 ml/h over a column (10 mm \times 60 mm) packaged with 5 ml of concanavalin A-agarose. The column was washed with 100 ml of TBS containing 1% OG at a flow rate of 25 ml/h, and glycoproteins bound to concanavalin A were eluted with 25 ml of TBS containing 1% OG and 0.4 M methyl mannopyranoside at 0.1 ml/min. The flow-through was combined with the wash fraction and passed over a WGA-agarose column of similar size. The column was washed as above and glycoproteins bound by WGA were eluted with 25 ml of TBS containing 1% OG and 100 mM GlcNAc. The eluate was dialysed against water, lyophilized and dissolved in 10 ml of Mes buffer (50 mM Mes, 100 mM NaCl, 20 mM CaCl₂, pH 6.5) containing 0.5% OG. Insoluble material was removed by centrifugation (105 000 g; 60 min; 4 °C), and the supernatant was treated with 2 units of *Clostridium perfringens* sialidase for 90 min at 37 °C in the presence of 1 mM PMSF, 1 μ g/ml leupeptin and 1 μ g/ml pepstatin. To this fraction was added 4 ml of a 50% slurry of Jacalin agarose and the mixture was incubated overnight at 4 °C while rotating head over tail. The Jacalin agarose was collected in a column and washed with 100 ml of TBS containing 1% OG. Glycoproteins bound to Jacalin were eluted with 25 ml of TBS containing 1% OG and 0.4 M methyl galactopyranoside. The eluate was diluted to 70 ml with distilled water and subjected to ultrafiltration by centrifugation over membranes with a 10 kDa cut-off (Centriprep-10, Microcon-10; Amicon, Beverly, MA, U.S.A.). The retentate (250 μ l) was separated by SDS/PAGE (10% gel; 0.75 mm) under reducing conditions, and gp40 was excised after Coomassie Blue staining. Purification of gp40 was monitored by analysing each fraction by SDS/PAGE and silver staining [22,23] and by an influenza C virus overlay assay [17].

Amino acid sequence analysis

For internal amino acid sequence analysis the Coomassie-stained gp40 was excised from the SDS/polyacrylamide gel and cleaved in the gel with Endoproteinase LysC (Boehringer, Tutzing, Germany) as described by Eckerskorn and Lottspeich [24]; the resulting peptides were eluted and separated by reversed-phase HPLC on a Superspher 60 RP select B column (125 mm \times 2 mm) (Merck, Darmstadt, Germany) using 0.1% trifluoroacetic acid in water as solvent A and 0.1% trifluoroacetic acid in acetonitrile as solvent B. A gradient was performed from 0 to 60% solvent B in 60 min at a flow rate of 300 μ l/min. Detection was at 206 nm and collection of the peptides was performed manually. The main peak in the chromatogram was further cleaved with Endoproteinase GluC (Boehringer), and the resulting peptides

were separated under the same conditions as described above. Amino acid sequence analyses were performed on a model 471A pulsed liquid-phase sequencer (Applied Biosystems, Weiterstadt, Germany) according to the instructions of the manufacturer.

RT-PCR and molecular cloning of gp40

Based on the sequence information obtained by Edman degradation and taking into account the preferred codon usage in dogs [25], two degenerate oligonucleotides were designed and used for amplification of a partial cDNA fragment by RT-PCR. The sense primer 5'-TGAGAATTCCCIGA(C/T)GA(C/T)AT-(C/T)AT(C/T)CCIGGIGT(T/G/C)GA-(A/G)GA-3' encodes the sequence PDDIIPGVED (Figure 3, residues 28–37) and the antisense primer 5'-TGAGAATTCTGIGA(T/C)TC(T/C)TC-(T/C)TTIGC(G/A)TGIAC-3' encodes the sequence VHAK-EEXQ (residues 97–105), where X was assumed to be a serine residue. Reverse transcription was performed at 42 °C for 15 min using the enzyme from murine leukaemia virus (2.5 units/ μ l), 5 μ g of MDCK I total RNA as template, and the antisense oligonucleotide (2.5 μ M) for priming. The reaction was stopped by heat denaturation at 99 °C for 5 min. The first-strand cDNA was supplemented with the sense primer (1 μ M), and a 248 bp cDNA fragment was amplified by PCR (35 cycles: 94 °C for 1 min, 40 °C for 1 min, and 72 °C for 2 min). The PCR fragment was subjected to cycle sequencing using the DyeDeoxy Terminator AmpliTaqFS Kit (Perkin-Elmer) and either the sense or the antisense oligonucleotide (5 pmol) as sequencing primer. Gel electrophoresis and data analysis were performed on a 377 DNA sequencer (Perkin-Elmer). Using the 248 bp fragment as template, a hybridization probe was generated by incorporation of digoxigenin-dUTP by PCR (PCR DIG Probe Synthesis Kit; Boehringer-Mannheim). Analysis of gp40 expression in MDCK II cells by RT-PCR was performed as described above. As a positive control, a 450 bp cDNA fragment of VIP36 [26] was amplified using the oligonucleotide 5'-TCCATGCCCTCTGGACTTTCAAGG-CAGC-3' (nucleotides 241–270) as sense primer and the oligonucleotide 5'-AGAGTAGCGCACAGC-TAGGAAGGTATCATG-3' (nucleotides 661–690) as antisense primer (35 cycles: 94 °C for 1 min, 60 °C for 1 min, 72 °C for 2 min).

An oriented lambda MDCK I cDNA library was constructed in the UNI-ZAP XR vector (Stratagene, La Jolla, CA, U.S.A.) according to the instructions of the manufacturer. For first-strand cDNA synthesis, 5 μ g of polyadenylated RNA was used. The RNA was prepared from MDCK I cells that had been treated overnight with 0.1 μ g/ml PMA to stimulate gp40 transcription. The primary cDNA library contained about 600000 recombinants with an average insert size of 1.2 kb. The library was amplified once to have a titre of 1.8×10^9 plaque-forming units/ml. A total of 75000 phage plaques were analysed. The phage were transferred to positively charged nylon membranes, subjected to alkaline denaturation, neutralization and UV cross-linking of the DNA to the membranes. The filters were prehybridized for 2 h at 68 °C in solution containing $5 \times$ SSC (where $1 \times$ SSC is 0.15 M NaCl/0.015 M sodium citrate), 0.1 % (w/v) *N*-lauroylsarcosine, 0.02 % (w/v) SDS and 1 % (w/v) blocking reagent (Boehringer-Mannheim). Hybridization was carried out overnight at 68 °C in the same solution containing 20 ng/ml digoxigenin-labelled probe (denatured at 100 °C for 8 min). The filters were washed twice for 15 min at room temperature with 50 ml of $2 \times$ SSC/0.1 % SDS and twice for 15 min at 40 °C with 50 ml of $0.5 \times$ SSC/0.1 % SDS per 100 cm² filter. An enzyme immunoassay was used for luminescent detection of the hybridized probe (DIG Luminescent Detection

Kit; Boehringer-Mannheim). Of 20 positive plaques, ten were further analysed. The phage were plated at low density (100 plaque-forming units/84 mm dish) and rehybridized as above. Four clones that reacted again with the probe were isolated, amplified and subjected to *in vivo* excision according to the instructions of the manufacturer (Stratagene). The clones were analysed by restriction digestion and cycle sequencing.

Preparation of antibodies

A multiple antigenic peptide was synthesized by anchoring the residues 26–40 of gp40 (VRPDDIIPGVEDSVV) on to an immunogenically inert core molecule of radially branching lysine dendrites [27,28]. After synthesis, the whole multiple antigenic peptide macromolecule was cleaved off with trifluoroacetic acid, desalted and used for immunization of rabbits. The antigen was given at doses of 2 mg at 3-weekly intervals. After four injections, the rabbits were bled. For immunofluorescence analysis, the serum was affinity-purified on immobilized gp40 using a published method [29].

Immunofluorescence analysis

MDCK cells grown on glass coverslips were fixed with 4 % paraformaldehyde in PBS for 20 min at room temperature. Aldehyde groups were quenched with 0.1 M glycine in PBS for 5 min at room temperature and the cells were rinsed in PBS with three changes. For intracellular staining, the cells were permeabilized with 0.2 % saponin in PBS for 5 min at room temperature. For cell-surface staining, fixed cells were directly incubated with affinity-purified anti-gp40 serum in PBS containing 3 % BSA (PBS-B) for 60 min at 37 °C and washed with PBS. The primary antibody was detected using a FITC-labelled antibody from donkey directed to rabbit immunoglobulin (1:100 in PBS-B) for 30 min at 37 °C. The samples were washed and then mounted in Mowiol and 2.5 % DABCO. The coverslips were viewed and photographed with a Zeiss Axiophot microscope equipped with UV optics. For confocal immunofluorescence analysis, MDCK II cells were grown on 7.5 mm Transwell units. Fixing and staining were performed in principle as described above except that longer incubation times (120 min) were chosen to allow primary and secondary antibody to penetrate the filter. The confocal laser-scanning microscope (Carl Zeiss, Oberkochen, Germany) worked with the blue line (408 nm wavelength) of an argon laser.

Cell-surface labelling and immunoprecipitation

Cell-surface proteins were labelled with biotin as previously described [17]. Labelled cells were lysed in 1 ml of RIPA buffer (1 % Triton X-100, 1 % deoxycholate, 0.1 % SDS, 200 mM NaCl, 20 mM Tris/HCl, pH 7.5, 10 mM EDTA, 10 mM iodoacetamide, 1 mM PMSF, 50 units/ml aprotinin). After incubation on ice for 30 min, insoluble material was removed by centrifugation (105000 g; 30 min; 4 °C). The cell lysate received 100 μ l of a 50 % slurry of prewashed Protein A-Sepharose and 20 μ l of anti-gp40 serum. In the control, 20 μ l of preimmune serum was added. After agitation overnight at 4 °C, the Sepharose beads were washed four times with RIPA buffer. The precipitated gp40 was eluted by boiling the beads for 10 min in $2 \times$ concentrated reducing SDS sample buffer (100 mM Tris/HCl, pH 6.8, 4 % SDS, 10 % glycerol, 0.02 % Bromphenol Blue, 200 mM dithiothreitol). The immunoprecipitates were separated by SDS/PAGE (10 % gel) and transferred to nitrocellulose by the semi-dry blot technique [30]. Detection of biotinylated gp40

by streptavidin–horseradish peroxidase and a chemiluminescent peroxidase substrate (ECL detection reagent) was performed as previously described [17].

Stable expression of gp40 in MDCK II cells

A cDNA fragment containing the complete open reading frame of gp40 was amplified by PCR (30 cycles: 94 °C for 40 s, 60 °C for 20 s, and 72 °C for 1 min) using Pfu polymerase. The sense primer used (5'-TGCAAGCTTCAGCGCCGAGACAACCGCTGG) binds to nucleotides 145–165 and contains at the 5' end a *Hind*III site with three additional nucleotides. The antisense primer used (5'-TGACTCGAGGTCATCTTCTCCACATGCTTG) binds to nucleotides 768–788 and contains at the 5' end a *Xho*I site with three additional nucleotides. The PCR product was digested with *Xho*I and *Hind*III and ligated into the corresponding sites of the pCEP4 expression vector (Invitrogen, Leek, The Netherlands). The complete sequence of the insert was verified to exclude the possibility of amplification errors. MDCK II cells (1×10^6 cells) were transfected with 40 μ g of recombinant plasmid using the calcium phosphate method [31]. For stable expression, the transfected cells were selected with hygromycin B at 250 μ g/ml. Resistant cells were subcloned and assayed for gp40 expression by immunofluorescence and immunoprecipitation.

Western-blot analysis of canine tissues

Biopsy samples were prepared from a dog immediately after death and kept frozen at -80 °C. The samples were ground in liquid nitrogen using mortar and pestle. The dry powder (0.1 g) was homogenized in 2 ml of TBS supplemented with protease inhibitor cocktail using a T8 Ultra-Turrax (IKA Labortechnik, Staufen, Germany). Proteins were solubilized by adding 100 μ l of 10% Triton X-114 to 900 μ l of homogenate. After 90 min at 4 °C, insoluble material was removed by centrifugation (14000 g; 10 min; 4 °C). The supernatant was subjected to temperature-induced phase separation [32,33], and gp40 was precipitated from the resulting detergent phase by immobilized WGA as previously described [17]. gp40 was detected by a standard Western-blot technique using anti-gp40 serum (1:200), biotinylated anti-rabbit immunoglobulin antibody (1:500) and streptavidin–horseradish peroxidase (1:1000). All incubations were carried out at 4 °C for 60 min each. A chemiluminescent substrate (ECL detection reagent) was used to detect horseradish peroxidase by short exposure to an autoradiography film.

RESULTS

Purification of gp40

In a previous study, we showed by a virus-overlay assay that gp40 is the major MDCK I cell-surface protein recognized by influenza C virus [17]. We also demonstrated that gp40 is a mucin-type sialoglycoprotein that is heavily O-glycosylated but with no evidence of N-glycans. In this study, we took advantage of this knowledge to purify the glycoprotein from an OG extract of MDCK I cell membranes by lectin affinity chromatography. The purification of gp40 was monitored using the virus-overlay assay (Figure 1A) and silver staining of polyacrylamide gels (Figure 1B). To separate gp40 from N-glycosylated proteins we used concanavalin A, a lectin that interacts with N-linked oligosaccharides of the high-mannose and hybrid type [34,35]. In the next step, the flow-through was chromatographed on immobilized WGA, which has affinity for GlcNAc and sialic acid residues [36]. Lane f shows that apart from gp40 only a few other glycoproteins were bound by WGA and subsequently eluted with

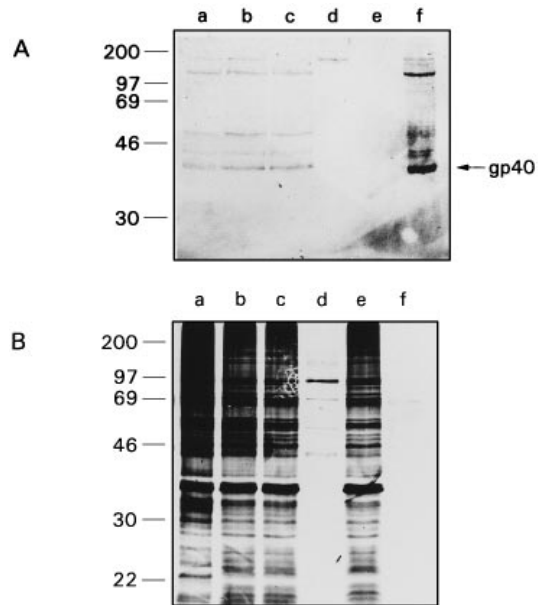


Figure 1 Purification of gp40

Aliquots of the following fractions were separated by SDS/PAGE: lane a, microsomes; lane b, OG lysate; lane c, flow-through of concanavalin A–agarose chromatography; lane d, proteins bound by concanavalin A; lane e, flow-through of WGA–agarose chromatography; lane f, proteins bound by WGA. Molecular masses are indicated on the left in kDa. (A) The proteins were transferred to nitrocellulose and probed with influenza C virus. Bound virus was detected using a chromogenic esterase substrate which was cleaved by the viral O-acetyltransferase. The position of gp40 is indicated. (B) The polyacrylamide gel was silver stained as described by Merrill et al. [22]. Because of its mucin-type character, gp40 is not stained by this procedure (compare with Figure 2).

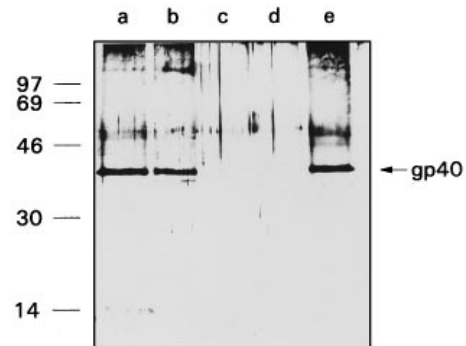


Figure 2 Chromatography of gp40 on immobilized Jacalin

Proteins eluted from WGA–agarose (lane a) were either left untreated (lanes b and c) or were treated with sialidase (lanes d and e) before chromatography on Jacalin agarose (lanes b and d, flow-through; lanes c and e, proteins bound by Jacalin). The fractions were analysed by SDS/PAGE and silver staining as described by Merrill et al. [23]. This silver stain also reacts with highly O-glycosylated proteins like gp40 (position indicated). Molecular masses are indicated on the left in kDa.

100 mM GlcNAc. Most other proteins of the extract did not bind to the column resulting in considerable enrichment of gp40 by this purification step. For further purification of gp40, we used Jacalin, a lectin recognizing O-glycans of the sequence Gal β 1-3GalNAc-Ser/Thr. Jacalin has been reported to tolerate this sequence with an additional sialic acid α 2,3-linked to Gal [37]. However, when the fraction eluted from WGA was applied

```

1      GTCAGCTCGGCGGGCGGGCTGCGGGAGGAGAACTTGGGGACCCCTGGGGC
53     CCGGGCTCTGCTCCCGGGCTTCCAGCCCTTCCCTGGCTCCCGGGCTGGGGC
113    CTGGCCGAGAGCTGGACCCCCCAGCCGCGGAGACAACCGCTGGCCCCAGGGACCGG
173    GCAGCCGACGAGATGTGGAGGTTGCCAGTCTCTCTTGGTTTGGGGCGGGCGGGCTC
1      M W R V P V L L L V L G G A G L
233    CGGGTCCCGCAGCAGGACCCAGCAGTCCCGCCAGATGACATAATACCAGGTGTAGAA
17     R V P A A G A S T V R P D D I I P G V E
      #
393    GACAGCGTGTGACCCAGGACAGACAGCAGTGTGACGCGCAGGACAGATAAC
37     D S V V T P G T E D S V V T P G A E D N
      #
353    GTGGTGACTGACGGTGCCACTGAAGACGCTTATGAGTCTGGCTTAACCCCTCTGGTGACA
57     V V T D G A T E E P Y E S G L T P L V T
      #
413    AAAAATACAGAGAGTGTAAACAGACTTACACTTAGAGGATGGGCCGACTCAAGAAAGCACA
77     K N T E S V T D L H L E D G P T O E S T
      #
473    GTCATGCCAAAGAGAGAGCCAGACACCAACCTGAATAGTAGTACTGACTACTCTCC
97     V H A K E E S Q S T T T L N V V T S H S
      #
533    AGAGAGAAAGTAGGTGAAGACACCGGAGACACCGGTTGAGAAAGATGGCTGGCAACAGTG
117    R E K V G E D T E T T V E K D G L A T V
      #
593    ACCCTGGTGGGATCATAATGGAGCTTACTAGCCATGGATTCATTGGTGGGATCATC
137    T L V G I I V G V L L A I G F I G G I I
653    ATGTGGTGTGCTCGAAAAATGTCCAGGAGTACTCGCCCTAAAGAGCTGAAGAATTGTGC
157    L V V A R K M S G R Y S P *
713    CCTTACCACACATTTAAAAGAGAGTTCCGACTTCCCGTCCCAAGCATGTGGGAG
773    AAGATGACCCATGGAAATGCTTGGCGATCCCACTACCAAAATCCATGGTATCCCTCCACT
833    TACCAAAAGTAACCAAGGGAAAGACAGTTCACAAGACTTGTCTCTAGGCAGGCATCT
893    GTCTTCTGAAAATGTTTTAAATATAAATTTGCAAAAGAGGATTCAAAAGACACAATGC
953    CTAGAAAATGGAGAAAACCACTAACTGATTTGTAACAAAGACAATATAATTAATTTG
1013   ATGTAGAAATATGCTCCCATCTCCGTTCTGACCAATCTCTCGTATTGTTAAAATGCAA
1073   TCAAGCTGGAAATATTTATACAGTGGAGCCCTTGAGTGTGAGTGGCAGCTTCATAAACA
1133   GCAGAAGCATTTTGAATTTCCGATCTTGTGAATGTACACATCCCAATCTGGCCCTGCTT
1193   TTTTAAAGCCTAACTTCATGTGAATAATCAATAGAAAAATCTAACTTGTATAATATGT
1253   AATGACCAACTGCCTTCCCTGGCCAGATAAATAATTTCAAAGGGCCCGGTGTAGGTC
1313   CATAAAGGGGAAAGTCTATGGTTTGGGAGAAANGGATTTGTTTATCCGCAATAAAGTCT
1373   CATCGGGGGCANGCTCCTAAGGCTTGTCTTATTGCTTCTCCAGCCGGAATGGGCTTT
1433   GGTTGAACTCAAANGCTTTGCTCCCGGCTTGGCCGCTTCCCTCCNTTCCCC
1493   CTCAAAGAAACGAAATGGGGGAAANCCCCAAAACCTC

```

Figure 3 Nucleotide sequence of gp40 cDNA and amino acid sequence of the encoded protein

The two peptides originally identified by Edman degradation are underlined. Those residues that appeared as gaps during microsequencing are marked #. The shaded regions indicate the proposed signal sequence and the putative membrane-spanning segment. The asterisk denotes the stop codon. These sequence data are available from EMBL/GeneBank under accession number Z81018.

to Jacalin agarose, most gp40 was not bound by the lectin (Figure 2, compare lane b with lane c). Only when gp40 was treated with sialidase and rechromatographed was it totally retained by Jacalin and could be eluted from the column with 400 mM galactopyranoside (Figure 2, compare lane d with lane e). This indicates that Jacalin has a higher affinity for the desialylated carbohydrate structure. The glycoprotein fraction was concentrated and subjected to SDS/PAGE. gp40 was excised from the Coomassie-stained polyacrylamide gel and enzymically digested in the gel slice using the proteases Lys-C and Glu-C. Two major peptides were recovered, separated by reverse-phase HPLC and their amino acid sequences determined by gas-phase Edman degradation (Figure 3). At 14 positions of the sequences we obtained gaps, probably because of post-translational modifications of the corresponding amino acids that rendered them insoluble in the organic transfer solvents used.

Molecular cloning and sequence analysis of gp40

A partial cDNA of gp40 was generated by RT-PCR using two degenerate oligonucleotide primers that were designed on the basis of the sequences PDDIIPGVED (residues 28–37) and VHAKEESQ (residues 97–105; Figure 3). The amplified 248 bp

cDNA fragment encoded a stretch of 33 amino acids of peptide 1 and 27 amino acids of peptide 2 linked by 17 amino acids not determined by Edman degradation. The 248 bp cDNA was labelled with digoxigenin and used to screen a MDCK I cDNA library that we constructed in the UNI-ZAP XR vector. Hybridization of about 75000 plaques under low stringency revealed 20 positive clones of which ten were further analysed. Four clones were still positive when rehybridized and all of them contained inserts of approx. 2500 bp. Determination of their nucleotide sequence revealed that, in one case, a full-length cDNA coding for gp40 was obtained. This cDNA contained an open reading frame encoding 169 amino acids (Figure 3), including the two peptide sequences determined by Edman degradation (underlined). gp40 shows the typical structure of a type-I membrane protein. It contains a putative N-terminal signal sequence with a highly scoring predicted signal sequence cleavage site between position 16 and 17 [38]. A stretch of 23 hydrophobic amino acids near the C-terminus is likely to form a membrane-spanning segment which is followed by a short cytoplasmic tail of nine amino acids. A serine within this tail at position 164 is part of a potential protein kinase C phosphorylation site (SGR) as well as part of a potential cAMP-dependent phosphorylation site (RKMS). The cleaved form of gp40 has a calculated molecular mass of only 16800 Da, indicating that carbohydrates and other post-translational modifications account for about 58% of the molecular mass of the mature glycoprotein. Whereas potential N-glycosylation sites and cysteine residues are not found in the sequence, potential O-glycosylation sites are very abundant. In the extracellular domain, about 25% of the amino acids represent either serine or threonine residues. At least 14 of the 31 potential O-glycosylation sites are believed to be actually used because the 14 gaps we observed during Edman degradation turned out to be either serine or threonine. Tandem repeats, a characteristic feature of many although not all mucin-type glycoproteins, are not found in gp40. However, the sequence PGVEDSVVT (residues 33–41) located near the N-terminus is tandemly repeated three times in a modified form. The threonine residue of this motif is most likely O-glycosylated in each repeat, since it appeared as a gap during Edman microsequencing.

A search in the EMBL data library revealed that gp40 is homologous to a protein found in mouse and rat. Furthermore we found homology (72% identity, 82% similarity) to a human cDNA sequence that is part of a human gene transcript catalogue. The alignment of the translated sequences of all four proteins is shown in Figure 4. It is evident that the human sequence is incomplete. The N-terminal part of the protein including the start methionine is missing. In contrast, the alignment of gp40 with the mouse and rat proteins extended over the whole reading frame. While the rat and mouse sequences with 78% identity (92% similarity) are closely related to each other, the sequence identity between gp40 and either of the two proteins is only 54% (70% similarity). The highest degree of sequence conservation is found in the C-terminal part of the proteins especially in their transmembrane and cytoplasmic domains (83% and 67% identity respectively). In the extracellular domain, sequence conservation is restricted to short amino acid stretches which in many cases show an accumulation of serine and threonine residues (e.g. residues 96–97, 106–109, 114–117, 133–137).

gp38, the mouse homologue of gp40, was identified as a cell-surface marker of thymus epithelial type-I cells [40] and was also described as being expressed by stromal cells in T-cell-dependent areas of peripheral lymphoid tissues [41]. In mouse osteoblastic cells, the glycoprotein, designated OTS-8, was induced by phorbol ester treatment [42]. In the rat, the protein was reported to be a

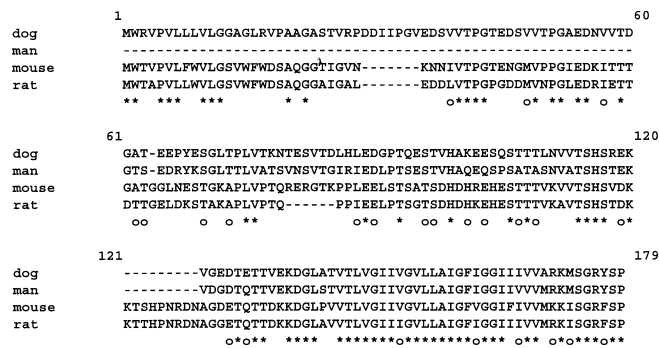


Figure 4 Comparison of gp40 with its putative homologues from rat, mouse and man

The alignment was generated by the Clustal program in HUSAR [39] using the translated nucleotide sequences [EMBL/GenBank accession numbers: U32115 and U07797 (rat homologue), M73748 and M96645 (mouse homologue), and R16862 and R59547 (human homologue)]. The asterisks denote positions with identical amino acids, and the circles indicate positions occupied by similar amino acids.

marker (E11) for the transition from the osteoblast to the osteocyte [43]. It was also described as a differentiation antigen (T1 α) on rat lung alveolar cells and was found to be developmentally expressed in embryonic brain [44,45].

Absence of gp40 expression in the type-II subline of MDCK cells

To confirm that the cloned cDNA encodes gp40, we produced a monospecific antibody against a synthetic peptide corresponding

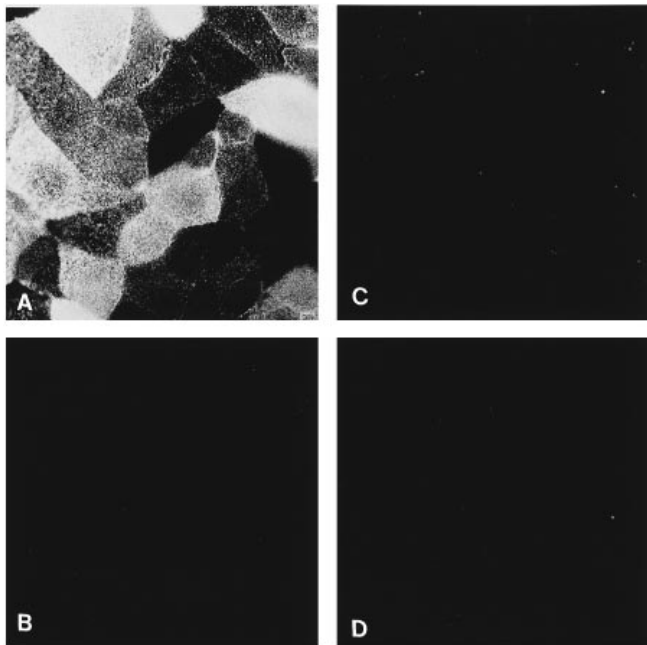


Figure 5 Cell-surface staining of MDCK cells using a monospecific antibody directed against gp40

MDCK I (A, B, C) and MDCK II cells (D) were fixed and incubated with affinity-purified anti-gp40 serum (A, C, D) or preimmune serum (B). Antibody binding to gp40 can be competed for by the peptide used for immunization (C). The first antibody was detected using a FITC-conjugated anti-rabbit immunoglobulin serum.

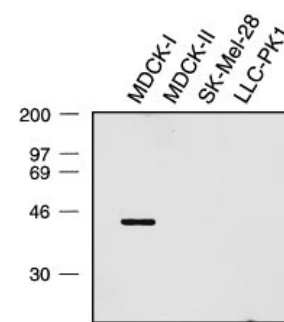


Figure 6 Immunoprecipitation of cell-surface gp40

The cell lines indicated on top were grown in Petri dishes (about 7×10^8 cells per dish) and surface labelled with biotin at 4 °C. gp40 was immunoprecipitated from the cell lysates using a monospecific anti-gp40 serum. The immunoprecipitates were run on SDS/PAGE and transferred to nitrocellulose. Biotin residues were detected using streptavidin-horseradish peroxidase. Molecular masses are indicated on the left in kDa.

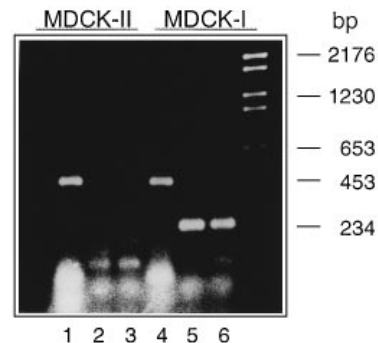


Figure 7 RT-PCR analysis of gp40 expression in the two MDCK sublines

MDCK I (lanes 4–6) and MDCK II cells (lanes 1–3) were left untreated (lanes 1,3,4,6) or were treated overnight with 0.1 μ g/ml PMA (lanes 2 and 5). Total RNA was prepared, reverse-transcribed, and used for PCR amplification of a 248 bp cDNA fragment of gp40 (lanes 2,3,5,6) and a 450 bp cDNA fragment of VIP36 as positive control (lanes 1 and 4) as described in the Materials and methods section.

to a region near the N-terminus of gp40. This region contains no serine and threonine residues and hence would be less likely to be masked by carbohydrate chains. The serum we obtained was affinity-purified on immobilized gp40 and used for immunofluorescence studies. Figure 5(A) shows that the antigen was recognized on the plasma membrane of MDCK I cells. Staining of the cells was specific as neither the preimmune serum (Figure 5B) nor the antibody preincubated with the peptide used for immunization (Figure 5C) showed any reaction with the cell surface. In contrast with MDCK I cells, the type-II subline of MDCK cells was negative for the antigen and the plasma membrane (Figure 5D) as well as intracellular compartments (not shown). Immunoprecipitation analysis revealed that the antibody recognizes a cell-surface-labelled protein of 40 kDa in MDCK I cells but not in MDCK II or other cell lines tested (Figure 6). In addition, the immunoprecipitated glycoprotein was recognized by influenza C virus using the virus-overlay assay (not shown). These results established that the isolated cDNA encodes the protein core of gp40. Our observation that the antibody did not recognize gp40 in MDCK II cells suggested that the glycoprotein may not be expressed in this subline. As shown in Figure 7, gp40 mRNA could not be detected in MDCK

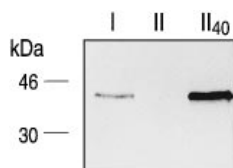


Figure 8 Recombinant expression of gp40 in MDCK II cells

MDCK I cells (I), MDCK II cells (II) and MDCK II cells transfected with pCEP4-gp40 (II₄₀) were surface-labelled with biotin and subjected to immunoprecipitation as described in Figure 6.

II cells by RT-PCR, confirming our assumption. The gp40 homologue in mouse, OTS-8, has been reported to be inducible in osteoblastic cells by treatment with phorbol esters [42]. In MDCK-II cells, however, we could not find any effect of this agent on gp40 expression (Figure 7). In contrast, phorbol ester treatment stimulated gp40 expression in MDCK I cells, which was analysed by Northern blot (not shown).

Apical localization of gp40 stably expressed in MDCK II cells

As MDCK II cells do not express endogenous gp40, this well-established polarized cell line should provide an excellent cell system to study the polarized transport of gp40 by recombinant expression. Therefore we established a MDCK II line that stably expressed gp40. These cells retained the polarized phenotype of the parental cells as indicated by measurement of their transepithelial electrical resistance. There was apparently no difference in molecular mass between endogenous gp40 of MDCK I cells and gp40 expressed in MDCK II cells (Figure 8), indicating that the glycoprotein was correctly processed in MDCK II cells. Cell-surface biotinylation indicated that gp40 is transported to the MDCK II cell surface. To determine the cell-surface distribution of gp40 between the apical and basolateral plasma membrane, filter-grown cells were selectively biotinylated at either domain. Aliquots of cell lysate were separated by SDS/PAGE, transferred to nitrocellulose, and the labelled apical and basolateral proteins were detected using streptavidin–peroxidase (Figure 9A). An immunoprecipitation step had been omitted from this procedure since in our experience this technique provides only semiquantitative data. However, because of its strong expression, gp40 could easily be detected among the labelled proteins. Using this approach, gp40 showed a predominantly apical (80%) localization. For comparison, membrane cofactor protein (CD46) was analysed as a basolateral marker protein [46] and the haemagglutinin of influenza A virus as an apical marker protein. Under the same experimental setting, 95% of CD46 was detected on the basolateral surface and 90% of the haemagglutinin on the apical surface of MDCK cells (not shown). In another approach, the cell-surface distribution of gp40 was analysed by indirect immunofluorescence using a confocal laser-scanning microscope. Filter-grown MDCK II cells stably expressing gp40 were fixed without permeabilization of the plasma membrane and were incubated from both the apical and basolateral side with affinity-purified anti-gp40 and an FITC-conjugated secondary antibody. In Figure 9(B), horizontal sections (parallel to the filter) are shown that represent (from top) the apical, central and basal portions of the cells. The brightest fluorescence was observed in the apical section. The section through the centre sometimes showed a weak honeycomb-like pattern, indicating that gp40 is present on the lateral plasma membrane to some extent. Almost no fluorescence was detected

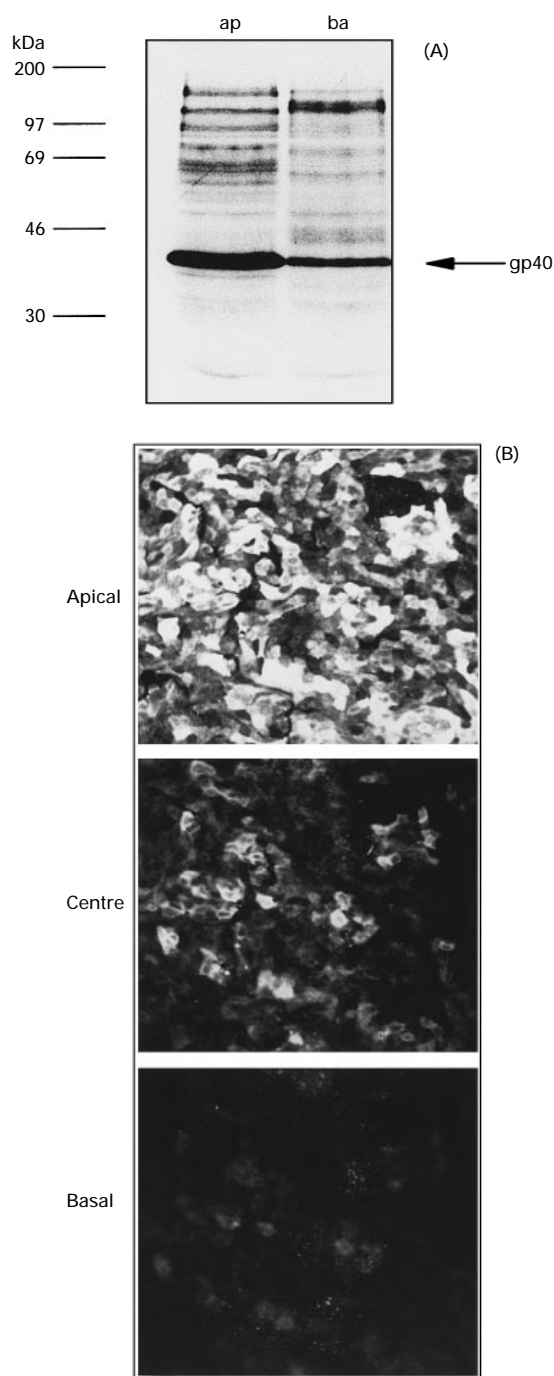


Figure 9 Polarized cell-surface distribution of gp40

A MDCK II cell clone stably expressing gp40 was grown on porous filters. (A) The cells were labelled either at the apical (ap) or the basolateral plasma membrane (ba) with biotin. Aliquots of cell lysate were run on SDS/PAGE, transferred to nitrocellulose and biotinylated proteins were detected using streptavidin–horseradish peroxidase. The position of gp40 is indicated. (B) The cells were fixed and both the apical and basolateral surface were incubated with affinity-purified anti-gp40 serum without permeabilization. FITC-conjugated anti-rabbit immunoglobulin serum was used for detection of the first antibody. The samples were viewed with a laser scanning microscope. Confocal micrographs of horizontal focal planes through the apical surface, the centre and the basal surface are shown.

in the basal section. Taken together, our findings show that gp40 is a predominantly apical membrane protein not only in MDCK I cells [17] but also in a heterologous cell system.

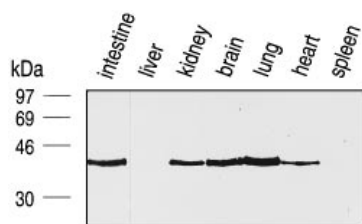


Figure 10 Tissue distribution of gp40

Biopsy samples from different canine organs (indicated on top) were extracted with Triton X-114. The extract was subjected to temperature-induced phase separation, and the detergent phase was used for precipitation with immobilized WGA. The precipitates were analysed for gp40 by Western blot using a monospecific anti-gp40 serum.

Detection of gp40 in different canine tissues

As the pattern of cell-surface proteins of a continuous cell line may differ from that of the parental cells, we asked whether gp40 can be detected not only in MDCK I cells but also in canine kidney biopsy specimens. To prove this, biopsy samples of various canine organs were extracted with Triton X-114, and membrane proteins were enriched in the detergent phase after temperature-induced phase separation. After precipitation with immobilized WGA, gp40 was immunologically detected using the Western-blot technique (Figure 10). In this way, we found that gp40 has a relatively broad tissue distribution, with a high level of expression in lung and brain. Somewhat lower levels were observed in kidney, intestine and heart whereas no expression was detected in liver and spleen.

DISCUSSION

The designation 'mucin' is generally used to refer to very large secreted O-glycosylated proteins with a carbohydrate content of typically more than 50% [47]. Cell-surface proteins have been described that are glycosylated in a similar way but which are membrane-associated and which are much smaller than the secreted mucins. Well-known examples of these cell-surface mucin-like glycoproteins are glycophorin A and leukosialin (sialophorin, CD43), the most dominant sialoglycoproteins of erythrocytes and leucocytes respectively. We have recently described a 40 kDa cell-surface glycoprotein of MDCK I cells, gp40, that also has characteristics of a mucin-type glycoprotein [17]. The primary structure of gp40, which has been determined in the present study by Edman degradation and molecular cloning, established its mucin-like character. Typical features are the high content of serine and threonine residues in the extracellular domain of gp40 and the clustered arrangement of several of these potential O-glycosylation sites. Edman degradation of two peptides from the extracellular domain suggested that at least 14 threonine/serine residues are O-glycosylated. There are six additional potential O-glycosylation sites located in a section of the extracellular domain that had not been subjected to Edman degradation. It is not possible to predict exactly from the primary sequence whether these sites are actually used for O-glycosylation since consensus sequences specifying O-glycosylation have not been established [48]. In accordance with the resistance of gp40 to N-glycosidase F [17], its primary sequence does not contain any potential N-glycosylation site, indicating that the difference of 58% in molecular mass between the mature glycoprotein and its protein backbone can be attributed mostly if not exclusively to O-linked glycans. For

comparison, glycophorin A contains one N-linked and 16 O-linked oligosaccharides contributing to 60% of its molecular mass [49]. Other similarities between gp40 and glycophorin A are their small size and the lack of tandem repeats, a feature found in many mucin-like glycoproteins. Like glycophorin A [50], leukosialin [51] and MUC1 [52], the primary structure of gp40 reveals a typical type-I membrane protein [53]. This confirms our previous observation that gp40 is enriched in the detergent phase after temperature-induced Triton X-114 phase separation [17].

Although gp40 has several structural characteristics in common with glycophorin A and other cell-surface mucin-like glycoproteins, it did not show any significant homology with these molecules. gp40 was rather found to be the canine homologue of the murine and rat cell-surface antigen designated OTS-8, E11, gp38 or T1 α [40–45]. In accordance with our results in the canine system, the homologues were found to be strongly expressed in type-I lung epithelial cells, but were also detected in a variety of other epithelial and non-epithelial cells. The murine homologue has also been found in glomeruli and tubules of kidney [40], providing further evidence that MDCK I cells originate from tubulus epithelial cells. Despite its broad tissue distribution, gp40 is differentially expressed by even closely related cell types. This is evident not only from the two sublines of MDCK cells, but was also noted for rat lung epithelial cells of type I and II [45]. Nose et al. [42] reported that treatment of osteoblastic cells with phorbol esters resulted in an induction of the gp40 homologue OTS-8. However, phorbol ester did not induce gp40 expression in MDCK II cells and likewise it had no effect on T1 α expression in freshly isolated type-II lung epithelial cells [45]. In MDCK I cells, however, phorbol ester treatment stimulated gp40 expression. We also observed that gp40 transcription is significantly elevated in MDCK I cells cultured on plastic dishes compared with MDCK I cells grown on porous polycarbonate filters (G. Zimmer and G. Herrler unpublished work). Interestingly, Rishi et al. [45] found that T1 α is rapidly induced, when type-II lung epithelial cells are cultured under conditions where they flatten and lose expression of many of their known cell markers. These findings indicate that regulation of gp40 or its homologues may vary in a cell-specific manner.

The amino acid homology between gp40 and its homologues is not uniformly distributed over the whole molecule. In the ectodomain, the overall homology is rather low (24% identity, 38% similarity between the homologues of dog, mouse and rat). Sequence conservation in this domain is restricted only to short amino acid stretches, most of which contain several potential O-glycosylation sites. Of the 14 O-glycosylation sites that we have identified by Edman degradation, eight were found to be conserved between the three species, suggesting that not all O-glycosylation sites are of equal importance for the glycoprotein. The single potential N-glycosylation site in the mouse homologue is not conserved in the rat and dog homologues. In striking contrast with the ectodomains, the transmembrane domains of all four homologues reveal a very high degree of sequence conservation (78% identity, 91% similarity). In contrast, the transmembrane regions of many other proteins, for example glycophorin A and glycophorin C [49,54], do not show a similar sequence conservation. This suggests that the transmembrane domain of gp40 may have an important function in addition to anchoring the molecule in the plasma membrane. Functional importance is also expected from the short cytoplasmic domain, as this domain also shows a high degree of sequence conservation (67% identity, 100% similarity). Interestingly, potential O-glycosylation sites, transmembrane and cytoplasmic domains have also been found to be conserved between the human and murine homologues of MUC1 [55].

In a previous study, we found gp40 to be predominantly located on the apical plasma membrane of filter-grown MDCK I cells [17]. In transfected MDCK II cells stably expressing gp40, we observed a similar polarized distribution. Immunocytochemical analysis of rat lung tissue sections showed that the rat homologue T1 α is also expressed on the apical surface of alveolar type-I cells [44]. In MDCK cells, which have been intensively studied with regard to polarized protein transport, sorting of apical and basolateral proteins occurs in the *trans*-Golgi network, where they are segregated into distinct transport vesicles for direct delivery to the respective plasma membrane [3,56,57]. The signals responsible for basolateral sorting have been shown to reside in the cytoplasmic domains and often contain a critical tyrosine residue. In many cases the signals overlap with those for clathrin-coated pit localization [3]. Little information on apical sorting signals is available. Although proteins linked to the membrane via a GPI anchor are delivered to the apical plasma membrane, it is not established that GPI anchors themselves serve as apical signals. Recent studies showed that GPIs that are not linked to protein are distributed in a non-polarized fashion in MDCK cells [58]. Furthermore tail-minus forms of many basolateral transmembrane proteins are also transported to the apical surface [3]. These observations suggest that the unknown apical sorting signals reside in the ectodomains of these proteins and that they may be dominated by stronger basolateral signals in the cytoplasmic domains. Studies on the polarized transport of the mucin-like glycoprotein MUC1 using chimaeric proteins also suggest that the ectodomain of this glycoprotein contains the information for apical targeting [59]. Many secreted glycoproteins are released from MDCK cells into the apical medium [16]. Recent data suggest that the core of *N*-glycans is a potential determinant for apical transport of secreted proteins [20,21]. Interestingly, a protein with a putative lectin-like domain (VIP36) has been isolated from apical transport vesicles of MDCK cells [26]. Although its carbohydrate-binding activity has not yet been elucidated, it has been proposed that this protein is involved in the apical targeting of glycoproteins [60]. It is obvious from our data that *N*-glycans are not involved in the apical transport of gp40. If the apical targeting signal is located in the glycans, *O*-glycans have to contain this information. However, *O*-glycans are structurally very different from *N*-glycans. For example, they do not contain the typical core unit of *N*-glycans made up of two GlcNAc and three mannose residues. Therefore, if *O*-glycans represent apical sorting signals, more than one lectin-like molecule is expected to be involved in the sorting process.

The physiological role of gp40 and its homologues is not yet known. Mucins in general have a filamentous brush-like conformation resulting from their extensive *O*-glycosylation allowing them to extend high above the cell surface [61]. This highly negatively charged structure is relatively resistant to proteases and acts as a physical barrier protecting epithelial cells from environmental agents. On the other hand, mucin-type glycoproteins have been shown to be involved in cell-cell interactions. For example, they represent high-affinity ligands for the selectins, lectin-like proteins on leucocytes and epithelial cells that participate in leucocyte trafficking, thrombosis and inflammation [62]. A general characteristic of many cancer cells is an increased production of *O*-glycosylated proteins which have been implicated in their protection from the immune system [63]. Moreover, mucin-type glycoproteins have been proposed to influence the metastatic potential of tumour cells by either acting as anti-adhesion molecules [64] or mediating adhesion to endothelial cells via interaction with selectins [39,65]. Thus gp40 may have a protective function in some cell types and may be involved in cellular interactions in others. Although the physiological

function of gp40 has not yet been established, we have shown that it can serve as a receptor for influenza C virus on MDCK I cells [17]. The efficient binding of influenza C virus to gp40 is explained in part by the presence of 9-*O*-acetylated sialic acids, the receptor determinant for this virus. Another important point is the mucin-type character of gp40. The high number of *O*-glycans provides favourable conditions for a multivalent interaction and thus for strong binding of the virion to the cell surface. It has been known for a long time that the agglutination of erythrocytes by influenza viruses is due to the presence of another mucin-like molecule, glycophorin A. However, glycophorin A does not represent a physiological receptor, as influenza viruses do not infect erythrocytes but do infect the epithelia of the respiratory tract. Our observation that gp40 is also strongly expressed in lung tissue provides further evidence for its importance as an influenza virus receptor.

We thank Behringwerke AG, Marburg, for kindly providing biopsy samples from dog. This work was supported by a grant from Deutsche Forschungsgemeinschaft (SFB 286) to G.H.

REFERENCES

- 1 Simons, K. and Fuller, S. D. (1985) *Annu. Rev. Cell Biol.* **1**, 243–288
- 2 Simons, K. and Wandinger-Ness, A. (1990) *Cell* **62**, 207–210
- 3 Matter, K. and Mellman, I. (1994) *Curr. Opin. Cell Biol.* **6**, 545–554
- 4 Rindler, M. J., Chuman, L. M., Shaffer, L. and Saier, M. H. J. (1979) *J. Cell Biol.* **81**, 635–648
- 5 Valentich, J. D. (1981) *Ann. N.Y. Acad. Sci.* **372**, 384–405
- 6 Herzlinger, D. A., Easton, T. G. and Ojakian, G. K. (1982) *J. Cell Biol.* **93**, 269–277
- 7 Simmons, N. L. (1981) *J. Membr. Biol.* **59**, 105–114
- 8 Richardson, J. C. W., Scalera, V. and Simmons, N. L. (1981) *Biochim. Biophys. Acta* **673**, 26–36
- 9 von Bonsdorff, C.-H., Fuller, S. D. and Simons, K. (1985) *EMBO J.* **4**, 2781–2792
- 10 Hansson, G. C., Simons, K. and van Meer, G. (1986) *EMBO J.* **5**, 483–489
- 11 Herrler, G. and Klenk, H. D. (1987) *Virology* **159**, 102–108
- 12 Schultze, B., Zimmer, G. and Herrler, G. (1996) *J. Gen. Virol.* **77**, 2507–2514
- 13 Vega-Salas, D. E., Salas, P. J. I., Gundersen, D. and Rodriguez-Boulan, E. (1987) *J. Cell Biol.* **104**, 905–916
- 14 Ojakian, G. K. and Schwimmer, R. (1988) *J. Cell Biol.* **107**, 2377–2387
- 15 LeBivic, A., Sambuy, Y., Mostov, K. and Rodriguez-Boulan, E. (1990) *J. Cell Biol.* **110**, 1533–1539
- 16 Rodriguez-Boulan, E. and Powell, S. K. (1992) *Annu. Rev. Cell Biol.* **8**, 395–427
- 17 Zimmer, G., Klenk, H.-D. and Herrler, G. (1995) *J. Biol. Chem.* **270**, 17815–17822
- 18 Brown, D. A., Crise, I. W. and Rose, J. K. (1989) *Science* **245**, 1499–1501
- 19 Lisanti, M. P., Caras, I. W., Davitz, M. A. and Rodriguez-Boulan, E. (1989) *J. Cell Biol.* **109**, 2145–2156
- 20 Scheiffele, P., Peränen, J. and Simons, K. (1995) *Nature (London)* **378**, 96–98
- 21 Fiedler, K. and Simons, K. (1995) *Cell* **81**, 309–312
- 22 Merrill, C. R., Switzer, R. C. and van Keuren, M. L. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4335–4339
- 23 Merrill, C. R., Goldman, D., Sedman, S. A. and Ebert, M. H. (1981) *Science* **211**, 1437–1438
- 24 Eckerskorn, C. and Lottspeich, F. (1989) *Chromatographia* **28**, 92–94
- 25 Wada, K., Wada, Y., Ishibashi, F., Gojobori, T. and Ikemura, T. (1992) *Nucleic Acids Res.* **20**, 2111–2118
- 26 Fiedler, K., Parton, R. G., Kellner, R., Etzold, T. and Simons, K. (1994) *EMBO J.* **13**, 1729–1740
- 27 Posnett, D. N., McGrath, H. and Tam, J. P. (1988) *J. Biol. Chem.* **263**, 1719–1725
- 28 Tam, J. P. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 5409–5413
- 29 Jöns, T. and Drenckhahn, D. (1992) *EMBO J.* **11**, 2863–2867
- 30 Kyhse-Anderson, J. (1984) *J. Biochem. Biophys. Methods* **10**, 203–209
- 31 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- 32 Bordier, C. (1981) *J. Biol. Chem.* **256**, 1604–1607
- 33 Lisanti, M. P., Sargiacomo, M., Graeve, L., Saltiel, A. R. and Rodriguez-Boulan, E. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 9557–9561
- 34 Baenziger, J. U. and Fiete, D. (1979) *J. Biol. Chem.* **256**, 2400–2407
- 35 Ogata, S., Muramatsu, T. and Kobata, A. (1975) *J. Biochem (Tokyo)* **78**, 687–696

- 36 Gallagher, J. T., Morris, A. and Dexter, T. M. (1985) *Biochem. J.* **231**, 115–122
- 37 Sastry, M. V., Banerjee, P., Patanjali, S. R., Swamy, M. J., Swarnalatha, G. V. and Surolia, A. (1986) *J. Biol. Chem.* **261**, 11726–11733
- 38 von Heijne, G. (1986) *Nucleic Acids Res.* **14**, 4683–4690
- 39 Senger, M., Glatting, K. H., Ritter, O. and Suhai, S. (1995) *Comput. Methods Programs Biomed.* **46**, 131–141
- 40 Farr, A. G., Nelson, A. and Hosier, S. (1992) *J. Histochem. Cytochem.* **40**, 651–664
- 41 Farr, A. G., Berry, M. L., Kim, A., Nelson, A. J., Welch, M. P. and Aruffo, A. (1992) *J. Exp. Med.* **176**, 1477–1482
- 42 Nose, K., Saito, H. and Kuroki, T. (1990) *Cell Growth Differ.* **1**, 511–518
- 43 Wetterwald, A., Hofstetter, H., Cecchini, M. G., Lanske, B., Wagner, C., Fleisch, H. and Atkinson, M. (1996) *Bone* **18**, 125–132
- 44 Dobbs, L. G., Williams, M. C. and Gonzalez, R. (1988) *Biochim. Biophys. Acta* **970**, 146–156
- 45 Rishi, A. K., Joyce-Brady, M., Fisher, J., Dobbs, L. G., Floros, J., VanderSpek, J., Brody, J. S. and Williams, M. C. (1995) *Dev. Biol.* **167**, 294–306
- 46 Maisner, A., Liszewski, M. K., Atkinson, J. P., Schwartz-Albiez, R. and Herrler, G. (1996) *J. Biol. Chem.* **271**, 18853–18858
- 47 Carraway, K. L. and Hull, S. R. (1991) *Glycobiology* **1**, 131–138
- 48 Gooley, A. A. and Williams, K. L. (1994) *Glycobiology* **4**, 413–417
- 49 Pisano, A., Redmond, J. W., Williams, K. L. and Gooley, A. A. (1993) *Glycobiology* **3**, 429–435
- 50 Tomita, M. and Marchesi, V. T. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 2964–2968
- 51 Shelley, C. S., Remold-O'Donell, E., Davis, A. E. I., Bruns, G. A. P., Rosen, F. S., Carroll, M. C. and Whitehead, A. S. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2819–2823
- 52 Gendler, S. J., Lancaster, C. A., Taylor-Papadimitriou, J., Duhig, T., Peat, N., Burchell, J., Pemberton, L., Laline, E. and Wilson, D. (1990) *J. Biol. Chem.* **265**, 15286–15293
- 53 Wickner, W. and Lodish, H. F. (1985) *Science* **230**, 400–407
- 54 Dahr, W. and Beyreuther, K. (1985) *Biol. Chem. Hoppe-Seyler* **366**, 1067–1070
- 55 Spicer, A. P., Parry, G., Patton, S. and Gendler, S. J. (1991) *J. Biol. Chem.* **266**, 15099–15109
- 56 Wandering-Ness, A., Bennett, M. K., Antony, C. and Simons, K. (1990) *J. Cell Biol.* **111**, 987–1000
- 57 Mostov, K., Apodaca, G., Aroeti, B. and Okamoto, C. (1992) *J. Cell Biol.* **116**, 577–583
- 58 van 't'Hof, W., Rodriguez-Boulan, E. and Menon, A. K. (1995) *J. Biol. Chem.* **270**, 24150–24155
- 59 Pemberton, L. F., Rughetti, A., Taylor-Papadimitriou, J. and Gendler, S. J. (1996) *J. Biol. Chem.* **271**, 2332–2340
- 60 Fiedler, K. and Simons, K. (1994) *Cell* **77**, 625–626
- 61 Jentoft, N. (1990) *Trends Biochem. Sci.* **15**, 291–294
- 62 Varki, A. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 7390–7397
- 63 Bhavanandan, V. P. and Furukawa, K. (1991) in *Biology of the Sialic Acids*, (Rosenberg, A., ed.), pp. 145–196, Plenum Press, New York
- 64 Ligtenberg, M. J. L., Buijs, F., Vos, H. L. and Hilkens, J. (1990) *Cancer Res.* **52**, 2318–2324
- 65 Irimura, T., Nakamori, S., Matsushita, S., Taniuchi, Y., Todoroki, N., Tsuji, T., Izumi, Y., Kawamura, Y., Hoff, S. D., Cleary, K. R. and Ota, D. M. (1993) *Cancer Biol.* **4**, 319–324
- 66 Sawada, R., Tsuboi, S. and Fukuda, M. (1994) *J. Biol. Chem.* **269**, 1425–1431