Cloning and characterization of gp36, a human mucin-type glycoprotein preferentially expressed in vascular endothelium

Gert ZIMMER*, Frank OEFFNER†, Veronika VON MESSLING*, Thomas TSCHERNIG‡, Hermann-Josef GRÖNE§, Hans-Dieter KLENK† and Georg HERRLER*¹

*Institut für Virologie, Tierärztliche Hochschule Hannover, Bünteweg 17, D-30559 Hannover, Germany, †Institut für Virologie, Philipps-Universität Marburg, Robert-Koch-Straße 17, D-35037 Marburg, Germany, ‡Institut für Anatomie II, Medizinische Hochschule Hannover, Carl-Neuberg-Straße 1, D-30623 Hannover, Germany, and §Deutsches Krebsforschungszentrum, Abteilung Allgemeine Experimentelle Pathologie, Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany

A mucin-type glycoprotein has been described in murine, rat and canine tissues as a differentiation antigen and influenza-virus receptor. We have cloned a cDNA from human placenta RNA encoding the corresponding human protein, a type-I integral membrane protein of 162 amino acids. Madin–Darby canine kidney cells transfected with the cDNA clone directed the cell-surface expression of a 36-kDa O-glycosylated sialoglycoprotein, gp36, and two minor isoforms of 28 and 70 kDa. gp36 has a broad tissue distribution with strong expression in lung, placenta and skeletal muscle, as shown by PCR screening of different

cDNA libraries. Immunohistochemical detection of gp36 in cryo-sections of human placenta, kidney, lung and nasal polyps showed that the glycoprotein is expressed at the apical plasma membrane of vascular endothelial cells. Expression of gp36 was not restricted to endothelial cells, as alveolar epithelial cells were found to express gp36 as well.

Key words: alveolar epithelium, O-glycosylation, membrane protein, placenta.

INTRODUCTION

In two previous studies, we described a sialoglycoprotein with an apparent molecular mass of 40 kDa, gp40, a major apical-cell-surface protein of type-I Madin–Darby canine kidney (MDCK) cells [1,2]. This glycoprotein has drawn our attention because it functions as a main receptor for the influenza C virus in this cell line. The efficient binding of the influenza C virus to gp40 is explained in part by the presence of 9-O-acetylated sialic acids, the receptor determinant for this virus. In addition, the high number of O-glycans attached to gp40 provides favourable conditions for a multivalent interaction with the virus. Influenza C virus is a human pathogen that infects the upper respiratory tract. It is not known whether influenza C virus uses a gp40 homologue as a receptor in human tissues.

gp40 represents the canine homologue of a cell-surface antigen expressed by various epithelial and non-epithelial cell types in rat and mouse. The mouse homologue was first identified as an early-response protein (OTS-8) that was induced by phorbol ester in an osteoblastic cell line [3]. The protein has also been described as a cell-surface antigen (gp38) that is expressed by type-I thymus epithelial cells and by stromal cells in T-dependent areas of peripheral lymphoid tissues [4,5]. In the rat, the gp40 homologue has been reported to be a marker (E11 antigen) for the late steps in the differentiation pathway of the osteoblastic lineage [6]. Another group has identified the rat homologue as a differentiation antigen (T1 α) of lung alveolar cells. This protein is not expressed by type-II alveolar cells, the stem cells for the alveolar epithelium, but is induced upon their differentiation into type-I alveolar cells [7,8]. In addition, T1 α shows developmentally regulated expression in rat embryonic brain [8,9]. Finally, the rat

homologue has been shown to be present in glomerular epithelial cells. Interestingly, this protein was found to be down-regulated in puromycin aminonucleoside nephrosis, a rat model of human minimal-change nephropathy [10].

Because of the characteristics of the rat, murine and canine homologues of this mucin-type glycoprotein – broad tissue distribution, strongly regulated expression, potential virus receptor – it was of interest to identify the human homologue. We found several human cDNA sequences that showed significant homology with the three sequences from rat, mouse and dog. These cDNAs were part of human gene-transcript catalogues from different tissues, including human brain, fetal kidney, heart, uterus and testis tumour. Without exception, the sequences were incomplete and did not encode the total open reading frame. In the present study, we describe the molecular cloning of a full-length cDNA of the human gp40 homologue, gp36. The production of polyclonal and monoclonal antibodies directed against the ectodomain of gp36 allowed the biochemical characterization and the histological localization of this hitherto unknown human cell-surface antigen.

MATERIALS AND METHODS

Oligonucleotides

The following oligonucleotides were used: GSP (gene-specific primer)-1 (5'-AAGCACGTTGGCAGCAGGGCGTAACC), GSP-2 (5'-CACTTGTTGCCACCAGAGTTGTCAAGCC), 5'-anchor primer (5'-CUACUACUACUAGGCCACGCGTCGACTAGTACGGGIIGGGIIGGGIIG), where I is deoxyinosine and U is deoxyuridine, 5'-adaptor primer (5'-CUACUACUACUAGGCCACGCGTCGACTAGTAC), GSP-3 (5'-

Abbreviations used: MDCK, Madin–Darby canine kidney; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; 5'-RACE, rapid amplification of 5'-cDNA ends; MBP, maltose-binding protein; GSP, gene-specific primer.

¹ To whom correspondence should be addressed (e-mail herrler@viro.tiho-hannover.de).

CACTTCTCGAGAAAGGTCACCGTGGATTCTGAATG), GSP-4 (5'-CGATTGAATTCCAGCTCAGAATCTTGCTGCTCGGC), GSP-5 (5'-TGACATGAATTCGAAGGAGCCAGCACAGGCCAGCCAG), GSP-6 (5'-TGACATCTGCAGTCATGTTGACAAACCATCTTTCTCAACTG), GSP-7 (5'-TGAGGTACCAGCTCAGAATCTTGCTGCTCGG), GSP-8 (5'-TGAGGATCCAAGGTCACCGTGGATTCTGAATG), GSP-9 (5'-GGCTTGACAACTCTGGTGGCAACAAGTG-3'), glycine aldehyde-3-phosphate dehydrogenase (GAPDH)-1 (5'-CGACACCCACTCCTCCAC), GAPDH-2 (5'-AGGGGTCTACATGGCAACTG), T7 (5'-GTAATACGACTCACTATAGGGC) and T3 (5'-AATTAACCCTCACTAAAGGG). Underlined nucleotides are recognition sites for restriction endonucleases (see below).

Biochemical reagents

The following reagents were from Boehringer Mannheim (Mannheim, Germany): sialidase from Newcastle disease virus, poly(ethylene glycol) 1500, hygromycin B, blocking reagent, O-glycosidase, digoxigenin-labelled lectins from *Maackia amurensis*, *Sambuca nigra* and *Datura stramonium*, and peroxidase-conjugated antibodies directed to digoxigenin and to mouse immunoglobulin from sheep. The following reagents were from Sigma (Deisenhofen, Germany): Protein A-Sepharose, Freund's adjuvant and peroxidase-labelled lectins from *Dolichos biflorus*, *Vicia villosa*, *Maclura pomifera*, *Arachis hypogaea* and *Artocarpus integrifolia*. The following reagents were from Amersham Buchler (Braunschweig, Germany): enhanced chemiluminescence (ECL) detection reagent, FITC-linked anti-rabbit immunoglobulin from donkey, FITC-linked anti-mouse immunoglobulin from sheep, streptavidin-horseradish peroxidase and prestained protein molecular-mass markers. *Vibrio cholerae* sialidase was obtained from Behringwerke AG (Marburg, Germany). Reverse transcriptase from moloney murine leukaemia virus was obtained from Gibco-BRL (Karlsruhe, Germany). *Pfu* polymerase was purchased from Stratagene (Heidelberg, Germany). Sulpho-succinimido-biotin was obtained from Pierce (Rockford, IL, U.S.A.), amylose resin and factor Xa were from New England Biolabs (Schwalbach, Germany), and peroxidase-conjugated antibody from pig directed to rabbit immunoglobulin was a product of Dako (Glostrup, Denmark). The Antibody-Multiplier-System was obtained from Linaris (Wertheim, Germany). Monoclonal antibody directed to CD34 and a rabbit polyclonal antibody directed to von Willebrand factor were purchased from Novocastra Laboratories (Newcastle upon Tyne, U.K.).

Cloning of gp36

Several partial cDNA clones of the putative human gp40 homologue missing the 5' end of the open reading frame exist in the GenBank/EMBL data libraries (accession numbers R16862, R59547 and F06720). To obtain the missing end, a rapid amplification of 5'-cDNA ends (5'-RACE) procedure, according to Frohman et al. [11], was performed using a commercially available kit (Gibco-BRL). First-strand cDNA was synthesized from 1 µg of human placenta poly(A)⁺ RNA hybridized to 2.5 pmol of gene-specific oligonucleotide GSP-1 (nucleotides 342–368 of R59547) by using an RNaseH-negative derivative of moloney murine leukaemia virus reverse transcriptase. The mRNA template was then degraded with RNaseH, the unincorporated deoxynucleotides, primer and proteins were removed, and the cDNA was tailed at the 5'-end with deoxycytosine by terminal transferase. The product was used as a template for the amplification of the gp36 5' terminus by PCR, using the

nested gene-specific oligonucleotide GSP-2 (nucleotides 29–56 of R59547) and the PCR anchor primer. The latter binds to the oligo-deoxycytosine region at the 5' end (35 cycles, 95 °C for 15 s, 60 °C for 30 s and 72 °C for 2 min). The PCR product was diluted 1:50 and reamplified (25 cycles, 95 °C for 15 s, 64 °C for 30 s and 72 °C for 2 min). The 470-bp PCR product was subjected to cycle sequencing using the DyeDeoxy Terminator AmpliTaqFS kit (Perkin-Elmer). Both DNA strands were sequenced using either the GSP-2 primer (5 pmol) or the adaptor primer (5 pmol), which corresponds to the very 5' end of the anchor. A cDNA encoding the complete open reading frame of gp36 was generated by reverse transcriptase PCR. Reverse transcription was performed using the enzyme from murine leukaemia virus (2.5 units/µl), 1 µg of human placenta poly(A)⁺ RNA as a template and the antisense GSP-3 oligonucleotide (2.5 µM) for priming (nucleotides 450–472 of R59547). This antisense primer contained an *Xho*I site at the 5' terminus. The reaction was stopped by heat denaturation and the first-strand cDNA was supplemented with GSP-4 primer (1 µM). This primer was designed by taking advantage of the sequence information that we obtained by sequencing the 5'-RACE product (see above), and contained an *Eco*RI site at the 5' terminus. The PCR product was digested with *Eco*RI and *Xho*I and ligated into the corresponding sites of pBluescript KS(+/-) (Stratagene) to give pBS-gp36. The gp36 open reading frame was verified by cycle sequencing. Both DNA strands were sequenced using either the T7 or T3 primers, which bind to the flanking arms of the pBluescript vector.

Bacterial expression of gp36 ectodomain

A cDNA fragment encoding the gp36 ectodomain (amino acids 21–128) was amplified by PCR from pBS-gp36 (40 ng) using *Pfu* polymerase. The sense primer used (GSP-5) bound to codons 21–29 and contained an *Eco*RI site at the 5' end. The antisense primer used (GSP-6) bound to codons 120–128 and contained a *Pst*I site at the 5' terminus. The PCR product was digested with *Eco*RI and *Pst*I and was subcloned into the corresponding sites of pMAL-c2 (New England Biolabs). The complete sequence of the insert was verified by cycle sequencing to exclude the possibility of amplification errors. The gp36 ectodomain was expressed in *Escherichia coli* (XL1-Blue MRF', Stratagene) as a maltose-binding protein (MBP)-gp36 fusion protein by induction with 1 mM isopropyl β-D-galactoside. The fusion protein in the bacterial lysates was adsorbed to amylose resin. The resin was washed with Tris-buffered saline and the fusion protein was eluted at 4 °C with Tris-buffered saline containing 20 mM maltose. About 120 mg of pure MBP-gp36 fusion protein were obtained from a 1-l culture. The fusion protein was analysed by factor-Xa cleavage, SDS/PAGE and Western blotting.

Production of antibodies

A polyclonal antiserum against gp36 ectodomain was raised in rabbits by injection of 400 µg of affinity-purified MBP-gp36 fusion protein in complete Freund's adjuvant followed by two booster injections in incomplete adjuvant at three-weekly intervals. For production of monoclonal antibodies, female Balb/c mice (8-week old) were immunized with MDCK cells stably expressing gp36 (2 × 10⁶ cells/injection). Three subcutaneous injections were given at 2-week intervals using the Antibody-Multiplier-System as adjuvant. Three days before the mice were killed a fourth injection without Antibody-Multiplier-System was given intraperitoneally. The spleen cells were fused with the P3X63Ag8.653 myeloma cells according to Köhler and Milstein [12] using poly(ethylene glycol) 1500 as fusing

agent. After hypoxanthine/aminopterin/thymidine selection, hybridoma supernatants were tested for antibody production by an ELISA using MBP-gp36 as antigen. Clones positive in the ELISA were further characterized by immunofluorescence and Western blotting using MDCK/gp36 cells (see below). The positive clone 18H5 used throughout this study was recloned twice by limiting dilution.

Stable expression of gp36 in MDCK cells

A cDNA fragment encoding the complete open reading frame of gp36 was amplified from pBS-gp36 by PCR using *Pfu* polymerase (25 cycles, 94 °C for 30 s, 60 °C for 20 s, 72 °C for 2 min). The sense primer (GSP-7) and the antisense primer (GSP-8) were designed to contain at their 5' termini a *KpnI* and a *BamHI* restriction site, respectively. The PCR product was digested with *BamHI* and *KpnI* and ligated into the corresponding sites of the pCEP4 expression vector (Invitrogen). The complete insert was verified by sequencing to exclude amplification errors. MDCK type-II cells (1×10^6 cells) were transfected with 40 µg of recombinant plasmid using the calcium phosphate method [13]. For stable expression, the transfected cells were selected with hygromycin B at 250 µg/ml. Resistant cells were subcloned and assayed for gp36 expression by immunofluorescence and Western blotting. A positive cell clone, designated MDCK/gp36, was used throughout this study.

Cell-surface labelling and immunodetection of gp36

Biotinylation of cell-surface proteins and immunoprecipitation were performed in principle as previously described [1,2]. The polyclonal gp36 antiserum was used at a dilution of 1:100. Immunofluorescence analysis of MDCK cells was carried out as reported in a previous study [2]. For this purpose, the polyclonal gp36 antiserum was used at a final concentration of 1:400, whereas the monoclonal antibodies were used at 1:20. For Western-blot analysis, 7×10^6 of either MDCK cells or MDCK/gp36 cells were lysed in 500 µl of Laemmli SDS sample buffer containing 200 mM dithiothreitol and boiled at 95 °C for 5 min. The solubilized proteins were run on a 15% polyacrylamide gel and blotted on to nitrocellulose according to Kyhse-Anderson [14]. The gp36 antiserum was used at a concentration of 1:5000. A peroxidase-conjugated antibody from pig directed to rabbit immunoglobulin was used as a secondary antibody at a dilution of 1:2000. 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 autoradiographic film. For immunohistological detection of gp36, frozen sections (6 µm) of human lung, placenta, kidney and nasal polyps were fixed with 3% paraformaldehyde for 20 min at room temperature, rinsed with PBS containing 0.1 M glycine and incubated with hybridoma supernatants (1:10). Bound primary antibodies were detected using either an FITC-conjugated antibody from donkey directed to mouse immunoglobulin (1:200) or a peroxidase-conjugated antibody from sheep directed to mouse immunoglobulin (1:300). In kidney sections, the alkaline phosphatase/anti-alkaline phosphatase technique with 3-amino-9-ethylcarbazole as a substrate was used for detection of the primary antibody. The monoclonal antibody directed to CD34 was diluted 1:100 for indirect immunofluorescence and the polyclonal antibody directed to von Willebrand factor was diluted 1:400. For detection of the latter antibody, an FITC-conjugated antibody directed to rabbit immunoglobulin was used (1:200).

Analysis by glycosidases and lectins

Biotinylated cell-surface gp36 was immunoprecipitated from lysates of 7×10^6 MDCK/gp36 cells. The immunoprecipitates still bound to the Protein A-Sepharose beads were washed and suspended in 200 µl of 50 mM sodium acetate/150 mM NaCl/9 mM CaCl_2 (pH 5.5) containing 0.5 mM PMSF and 1 unit/ml sialidase from *V. cholerae* or from Newcastle disease virus. Controls were incubated with buffer alone. The beads were agitated in a thermomixer 5436 (Eppendorf) for 60 min at 37 °C. The beads were washed and suspended in 50 mM phosphate buffer, pH 6.8, containing 0.5 mM PMSF and 25 m-units/ml of O-glycosidase. The beads were incubated for 2 h at 37 °C with agitation and then collected by centrifugation. gp36 was eluted from the beads by boiling at 95 °C for 10 min in 2× concentrated Laemmli SDS sample buffer containing 200 mM dithiothreitol. The effect of glycosidase treatment on the mobility of the glycoprotein in SDS/PAGE was analysed by a modified Western-blot procedure using streptavidin-horseradish peroxidase (1:2000) for detection of the biotinylated gp36 [2]. Similar results were obtained when gp36 was eluted from the Protein A-Sepharose beads before enzyme treatment.

For lectin analysis, gp36 was immunoprecipitated from unlabelled MDCK/gp36-cell lysates. The immunoprecipitates were divided into two halves; one was treated with *V. cholerae* sialidase (see above), whereas the other was incubated with buffer alone. After a 60-min incubation, gp36 was eluted from the beads, run on a SDS/polyacrylamide (12%) gel and transferred on to nitrocellulose. The blot was blocked with 0.5% blocking reagent overnight at 4 °C and probed with the lectins from *Arac. hypogea* (peanut agglutinin, PNA, 1:2000), *Arto. integrifolia* (Jacalin, 1:2000), *Macl. pomifera* (*Macl. pomifera* agglutinin, MPA, 1:2000), *S. nigra* (*S. nigra* agglutinin, SNA, 1:1000), *Maa. amurensis* (*Maa. amurensis* agglutinin, MAA, 1:200), *Da. stramonium* (*Da. stramonium* lectin, DSL, 1:1000), *Vic. villosa* (*Vic. villosa* agglutinin, VVA, 1:1000) or *Do. biflorus* (*Do. biflorus* agglutinin, DBA, 1:1000). All lectins were diluted in PBS containing Mg^{2+} and Ca^{2+} (1 mM each) and incubated with the blots for 60 min at 4 °C. Thereafter the blots were washed four times with PBS containing 0.1% Tween 20. The peroxidase-labelled lectins (VVA, DBA, PNA, Jacalin, MPA) were detected directly using a chemiluminescent peroxidase substrate (see above). Digoxigenin-labelled lectins (MAA, SNA, DSL) were detected in a similar way after incubation with peroxidase-conjugated antibody directed to digoxigenin (1:5000, 60 min, 4 °C).

PCR screening of cDNA libraries

A panel of 16 individual cDNA libraries from eight different human tissues (Clontech) was screened by PCR using the gp36-specific primers GSP-9 and GSP-1 flanking a 340-bp fragment (30 cycles, 94 °C for 1 min, 68 °C for 1 min). As a control, a 276-bp fragment specific for GAPDH was amplified from each cDNA library using the primer pair GAPDH-1 and GAPDH-2 (25 cycles, 94 °C for 45 s, 60 °C for 45 s, 72 °C for 1 min).

RESULTS

Cloning of gp36

In the GenBank/EMBL data libraries the sequences of several cDNA fragments from different human tissues have been deposited that align with the 3' ends of gp40 and its murine and rat homologues. Using the RACE technique and placenta mRNA as

Figure 1 Comparison of gp36 with its putative homologues from rat, mouse and dog

The alignment was generated by the Clustal program in HUSAR [15] using the translated nucleotide sequences [EMBL/GenBank accession numbers were U32115 and U07797 (rat homologue), M73748 and M96645 (mouse homologue) and Z81018 (dog homologue)]. The predicted signal sequence of gp36 is underlined, the double underline indicates the putative membrane-spanning segment. In the homologues, only amino acids differing from the gp36 sequence are shown. Identical amino acids are represented by bars whereas dots indicate gaps in the sequences to allow optimal alignment.

Table 1 Sequence identity between gp36 and the related sequences from dog, rat and mouse

The translated nucleotide sequences shown in Figure 1 and the corresponding nucleotide sequences were compared using the GAP program in HUSAR [15]. The values are given as the percentages of identical amino acids and nucleotides.

Pairing	Identical amino acids (%)	Identical nucleotides (%)
Human-dog	67	78
Human-mouse	48	64
Human-rat	52	66
Dog-mouse	54	69
Dog-rat	54	60
Rat-mouse	78	86

a template we obtained an additional 290 bp at the 5' end of the putative human homologue, including the missing start methionine. Based on this additional sequence data, an oligonucleotide primer was designed that allowed the amplification of the full-length open reading frame by reverse transcriptase PCR. This cDNA has been cloned and its sequence has been deposited in the GenBank/EMBL data libraries under accession number AJ225022. Aligning the four sequences by the Clustal program revealed that the number of amino acids of the human protein (162 amino acids) was very similar to that of its counterparts from dog (169 amino acids), mouse (172 amino acids) and rat (166 amino acids) (Figure 1). Short gaps in the sequences, 1-9 amino acids long, account for these differences. The relationship between the cloned cDNA and each of the three other proteins is shown in Table 1. The highest degree of homology exists between the human and canine proteins, with 78% identity on the nucleic acid level. Like its homologues, the human protein has 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 amino

Figure 2 Western-blot analysis of gp36 from transfected MDCK cells

Cell lysates were prepared from MDCK/gp36 cells (lane a) and MDCK type-II cells (lane b). The cell proteins were separated by SDS/PAGE under reducing conditions and transferred on to nitrocellulose. gp36 was detected by incubating the blots with an antiserum directed to MBP-gp36 followed by an incubation with a peroxidase-conjugated antibody directed to rabbit immunoglobulin. The relative positions of standard proteins of the indicated molecular masses (in kDa) are shown on the left.

acid positions 20 and 21 [16]. A hydrophobic stretch of 25 amino acids near the C-terminus is able to form a membrane-spanning segment, which is followed by a short cytoplasmic tail of nine amino acids. The highest degree of sequence conservation was found in the C-terminal parts of the proteins, especially in the transmembrane and cytoplasmic domains. In the ectodomain, sequence conservation was restricted to only short amino acid stretches. The ectodomains are characterized by a strikingly high proportion of serine and threonine residues. In the human protein, 31 of these potential O-glycosylation sites are present, and 15 of them are conserved in the four species. Like the canine and rat homologues, the human protein does not contain any potential N-glycosylation sites.

Biochemical characterization of gp36

For characterization of the encoded protein, a polyclonal antibody was produced by immunization of rabbits with recombinant antigen. The putative ectodomain of the gp40 homologue was cloned in-frame with the MBP. This fusion protein was expressed in *E. coli*, purified and used for immunization. We obtained an antiserum that reacted specifically with non-permeabilized MDCK cells stably expressing the human gp40 homologue (results not shown). No staining was observed when the parental cells were analysed or when preimmune serum was used. The accessibility of the protein for the antibody without permeabilization of the cells shows that the protein is expressed at the cell surface with its N-terminal portion facing the outside of the cell, as would be expected from a type-I membrane protein. The expression of gp36 in recombinant MDCK cells was further analysed by subjecting cell lysates to SDS/PAGE and Western-blot analysis. As shown in Figure 2, the gp36-specific antiserum recognized a major band with an apparent molecular mass of 36 kDa and two minor proteins with apparent molecular masses of 28 and 70 kDa, respectively. In order to determine whether all three bands were expressed on the cell surface, the surface proteins were labelled using a biotinylating reagent. Upon SDS/PAGE and transfer on to nitrocellulose, detection of

Figure 3 Detection of cell-surface biotinylated gp36 in transfected MDCK cells

MDCK type-I cells (lane a), MDCK type-II cells (lane b) and MDCK type-II cells stably expressing either the canine homologue gp40 (lane c) or gp36 (lane d) were surface-labelled with biotin at 4 °C. Cell lysates were run on SDS/PAGE and transferred on to nitrocellulose. Biotin residues of the labelled cell-surface proteins were detected using streptavidin-horseradish peroxidase and a chemiluminescent peroxidase substrate. Only the very dominant cell-surface proteins were visualized because of the very short time of exposure to an autoradiographic film. The positions of gp36, gp40 and their putative dimer forms (*) are indicated on the right. The relative positions of standard proteins of the indicated molecular masses (in kDa) are shown on the left.

Figure 4 Detection of O-linked oligosaccharides in gp36 by glycosidase treatment

MDCK/gp36 cells were surface-labelled with biotin at 4 °C. gp36 was immunoprecipitated from the cell lysates using an antiserum directed to MBP-gp36. The immunoprecipitates were treated with buffer (lane a), *V. cholerae* sialidase (lane b), Newcastle disease virus sialidase (lane c), and *V. cholerae* sialidase plus O-glycosidase for 3 h (lane d) or 6 h (lane e) as described in the Materials and methods section. The digested samples were separated by SDS/PAGE, transferred on to nitrocellulose, and the biotinylated gp36 was visualized using streptavidin-peroxidase. The relative positions of standard proteins of the indicated molecular masses (in kDa) are shown on the left.

the biotinylated proteins was mediated by streptavidin-peroxidase using a chemiluminescent peroxidase substrate. Strong labelling of the 36- and 70-kDa bands indicated that these two forms of gp36 were expressed on the cell surface (Figure 3). On the other hand, the 28-kDa protein appeared to be an intracellular form, because it was not accessible to surface biotinylation. Figure 3 also shows that both the 36- and 70-kDa bands had an increased electrophoretic mobility compared with their counterparts of the canine homologue. This difference in the apparent molecular mass explains the designation gp36 for the human homologue and gp40 for the canine homologue. The antiserum raised against the gp36 fusion protein did not react with gp40, and an antibody directed to gp40 did not precipitate the 36-kDa protein (results not shown). This finding may be explained by the

Figure 5 Binding of lectins to immunoprecipitated gp36

gp36 was immunoprecipitated from lysates of MDCK/gp36 cells and treated with either *V. cholerae* sialidase (+) or with enzyme buffer alone (—). The digested samples were separated by SDS/PAGE under reducing conditions, transferred on to nitrocellulose and incubated with the indicated lectins, which were conjugated with either peroxidase or digoxigenin. Detection of the bound lectins was performed as described in the Materials and methods section.

relatively low sequence conservation in the ectodomains of these two proteins (see above).

The open reading frame of gp36 encodes a polypeptide of only 17.8 kDa, which is predicted to be processed further by cleavage of its signal peptide to give a 15.6-kDa protein. The discrepancy between the deduced molecular mass and the apparent size of the mature protein suggests that post-translational modifications have occurred. Whereas N-glycosylation could be excluded (see above), O-glycosylation was suggested by the presence of multiple serine and threonine residues in the ectodomain. To get evidence for the presence of O-linked oligosaccharides on gp36, we analysed the protein by digestion with sialidase and O-glycosidase. Following surface biotinylation, cellular proteins were immunoprecipitated and incubated with the respective enzyme. The effect of the enzyme treatment on the electrophoretic mobility of gp36 was determined as described for Figure 3 (Figure 4). Treatment of gp36 with sialidase from *V. cholerae* or Newcastle disease virus caused a slightly decreased mobility of the protein. This unusual behaviour, which is often observed with heavily O-glycosylated sialoglycoproteins [17], is due to the fact that negative charge is lost when sialic acids are removed. When the sialidase-treated sample was further incubated with O-glycosidase, the gp36 protein was detected as a broad band in the 26–30-kDa range. Since O-glycosidase cleaves exclusively O-linked oligosaccharides containing the Gal β 1,3GalNAc sequence [18], it is expected that O-glycans that are structurally unrelated still remained attached to gp36. It should be noted that the cell-surface-biotinylated 70-kDa form was not immunoprecipitated. It appears that this band represented a dimer of gp36 that was very unstable under these conditions. The presence of O-linked oligosaccharides on gp36 was further confirmed when gp36 was analysed for reactivity with various lectins (Figure 5). Sialidase-treated gp36 was recognized by the lectins Jacalin, PNA and MPA. Whereas MPA recognizes terminal galactose and galactosamine residues [19], Jacalin and PNA are specific for the Gal β 1,3GalNAc sequence [20,21], also known as the T-antigen or core-I structure of O-glycans. Sialidase pretreatment of gp36 was an absolute requirement for recognition. From this we conclude that the O-linked glycans of gp36 are mostly if not completely sialylated. The two lectins with specificity for terminal sialic residues, MAA and SNA, showed a different reactivity towards gp36. SNA, which has affinity for sialic acids in α -2-

Figure 6 Screening of different human cDNA libraries for the presence of gp36 cDNA

A panel of 16 different cDNA libraries from eight different human tissues was screened by PCR: brain (lane 1), placenta (lane 2), skeletal muscle (lane 3), kidney (lane 4), heart (lane 5), lung (lane 6), liver (lane 7), pancreas (lane 8), negative control (lane 9) and pool (10). Either a 340-bp fragment of gp36 (upper panel) or a 276-bp fragment of GAPDH was amplified (lower panel). Only the PCR products from eight representative libraries are shown. The DNA molecular-size markers (M) on the left and right represent DNA fragments of 2176, 1766, 1230, 1033, 653, 517, 453, 394, 298, 234/220 and 54 bp.

6 glycosidic linkage [22], did not recognize the glycoprotein, whereas MAA, which binds sialic acids in α 2-3 glycosidic linkage, did [23,24]. The lectins VVA, DBA and DSL, with affinity for terminal GalNAc or GlcNAc residues [25–27], did not bind to gp36 (results not shown), indicating that O-glycans of the core II, III and IV types are not present. With no lectin was the 28-kDa intracellular form of gp36 detected.

Tissue distribution and cell-type-specific expression of gp36

In order to get information about the tissue distribution of gp36, cDNAs derived from eight different human tissues were analysed by PCR for gp36-specific sequences. The results are shown in Figure 6. Clear evidence was obtained for the presence of gp36 mRNA in placenta, lung, skeletal muscle and heart. Weak but distinct bands were detected in the brain, kidney and liver samples. In pancreas, gp36 cDNA was below the detection limit. This result indicates that gp36 has a broad tissue distribution. For detection of gp36 on the cellular level, histological sections of placenta, lung and kidney were stained with a monoclonal antibody (mAb18H5) directed towards the gp36 ectodomain. In cryo-sections of human placenta, mAb18H5 reacted specifically with the endothelial cells of vascular capillaries, whereas trophoblasts and other cells were not recognized (Figures 7A and 7B). Only the apical (luminal) plasma membranes of the endothelial cells were stained, indicating that gp36 is expressed by these cells in a polarized fashion. The cells that were reactive with mAb18H5 were also found to be positive for von Willebrand factor and CD34, two markers of endothelial cells (Figure 7C and 7D). However, the CD34 antigen, a mucin-type membrane protein like gp36, was not totally restricted to the apical plasma membrane but was also detected at the basolateral plasma membrane, whereas the antibody directed towards von Willebrand factor, a secretory protein, stained the endothelium

Figure 7 Localization of gp36, CD34 and von Willebrand factor in human placental sections by indirect immunofluorescence

Cryo-sections of human placenta were stained with monoclonal antibodies directed to gp36 (A,B), CD34 (C) or with a polyclonal antibody directed to von Willebrand factor (D). The primary antibodies were detected using FITC-labelled antibodies directed to mouse or rabbit immunoglobulins, respectively. (A) Several horizontal sections of chorionic zots bordered by an epithelium of trophoblasts showing weak autofluorescence. The bright fluorescence indicates binding of the monoclonal antibody to the gp36 antigen, expressed by vascular endothelial cells.

in a spotted manner. In kidney sections, the endothelia of the preglomerular artery, arteriole and peritubular capillaries were found to be positive for gp36, whereas tubular epithelial cells were negative (results not shown). Several cells in the glomeruli were also stained, but resolution of the cryo-sections did not allow us to identify the nature of these cells definitely. Endothelial cells in lung and nasal polyps also showed strong expression of gp36 (results not shown). In addition, gp36 was also found on the apical plasma membrane of the alveolar epithelium.

DISCUSSION

The data presented in this study provide strong evidence that the cloned cDNA encodes for the human homologue of a protein that has been described previously in rat, mouse and dog as a differentiation marker of alveolar cells [8], as a down-regulated component in puromycin aminonucleoside nephrosis [10] and as a potential receptor for the influenza C virus [1]. A striking feature of all four homologues is their extraordinary high content of serine and threonine residues, suggesting that O-glycosylation may be responsible for the differences between the calculated and apparent molecular masses (more than 50%). In fact, using different lectins and glycosidases we were able to show that gp36 in transfected cells contains O-glycans. Accordingly, 14 serine/threonine residues of the canine homologue gp40 were identified by Edman degradation to be modified by O-glycosylation [2]. Therefore, it is reasonable to assume that all four homologues share this mucin-type character.

After stable transfection of MDCK cells with the gp36 cDNA, a 70-kDa protein resistant to reducing agents was detected at the cell surface in addition to the major 36-kDa protein. This high-molecular-mass form may represent a dimer of gp36 that is partly stable in SDS. Dimerization may be driven by aggregation of the hydrophobic membrane anchors, as has been shown for glycophorin A, another small mucin-like membrane protein [28–30]. In addition to the 36- and 70-kDa variants at the cell surface, an intracellular protein of 28 kDa was detected in transfected cells that did not react with any of the lectins used. This protein may represent the unglycosylated dimeric form of gp36. In general, precursors of mucin-type glycoproteins are only N-glycosylated or unglycosylated because the transfer of individual monosaccharides during the synthesis of O-glycans in the Golgi is a very fast process with no discrete intermediates detectable [31]. Only in some cases have partially O-glycosylated intermediates been described [32–34].

Immunohistological analysis revealed that gp36 is expressed in alveolar epithelial cells and by vascular endothelium of different tissues. Interestingly, the rat homologue has also been found to be expressed by alveolar epithelial cells; however, its expression in vascular endothelial cells has not been noticed [6,8–10]. The expression of gp36 in epithelial cells as functionally diverse as endothelial cells and alveolar epithelial cells makes it very difficult to assign a precise physiological function to gp36. Likewise, the physiological importance of the gp36 homologues has not been elucidated, although it has been suggested for the homologue gp40 that the mucin-type character of the protein might be related to its function [2]. In general, mucin-type glycoproteins have an extended brush-like conformation due to their extensive O-glycosylation [35]. This highly negatively charged structure is

relatively resistant to proteases and provides a physical barrier protecting epithelial cells from environmental agents. Interestingly, gp36, like its homologues, is selectively transported to the apical plasma membrane of most cell types [2,6,8,10]. In this position it faces either the external environment (lung alveolar epithelium) or a luminal compartment (endothelium), which is compatible with a protective function. An important feature of mucin-type glycoproteins is their high content of sialic acids, which contribute mainly to the very negative charge of the molecule. A high concentration of sialic acids on the cell surface is known to protect the cell from attack by the alternative complement pathway [36–40]. In addition, the repulsive forces provided by sialic acids are also known to block cell adhesion and cell aggregation [41,42]. These masking effects of sialic acids may be especially important for endothelial cells for two reasons. First, the luminal (apical) plasma membrane of endothelial cells is confronted directly with the complement system. Second, the adhesion of blood cells to the endothelium is usually not a desirable effect, as this may lead to thrombosis and arteriosclerosis. On the other hand, mucin-type glycoproteins have been shown to participate in cell–cell interactions. For example, CD34 on endothelial cells contains oligosaccharide motifs that are recognized by L-selectin [43], a lectin-like protein on leucocytes that is involved in leucocyte trafficking, thrombosis and inflammation [44]. It will be interesting to see if gp36 expressed by endothelial cells will also be recognized by selectins. We have shown recently that the canine homologue gp40 is the major cell-surface protein recognized by the influenza C virus in a canine renal epithelial cell line [2]. The clustered array of sialic acids presented by gp40 is thought to be the major determinant for the preferred binding of influenza C virus to this cell-surface molecule. For the same reason, the presence of gp36 in epithelial cells of the respiratory tract makes it a potential receptor molecule for influenza viruses and other viruses that bind to sialic acid residues. In summary, we have identified a membrane protein with diverse distribution in human tissues the physiological function of which might be related to its mucin-type character.

We thank Ina Trotz and Karin Westermann for excellent technical assistance. This work was supported by a grant from Deutsche Forschungsgemeinschaft (SFB 286) to G.H.

REFERENCES

- Zimmer, G., Klenk, H.-D. and Herrler, G. (1995) *J. Biol. Chem.* **270**, 17815–17822
- Zimmer, G., Lottspeich, F., Maisner, A., Klenk, H.-D. and Herrler, G. (1997) *Biochem. J.* **326**, 99–108
- Nose, K., Saito, H. and Kuroki, T. (1990) *Cell Growth Differ.* **1**, 511–518
- Farr, A. G., Berry, M. L., Kim, A., Nelson, A. J., Welch, M. P. and Aruffo, A. (1992) *J. Exp. Med.* **176**, 1477–1482
- Farr, A. G., Nelson, A. and Hosier, S. (1992) *J. Histochem. Cytochem.* **40**, 651–664
- Wetterwald, A., Hofstetter, H., Cecchini, M. G., Lanske, B., Wagner, C., Fleisch, H. and Atkinson, M. (1996) *Bone* **18**, 125–132
- Williams, M. C. and Dobbs, L. G. (1990) *Am. J. Respir. Cell Mol. Biol.* **2**, 533–542
- 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
- Williams, M. C., YuXia, C., Hinds, A., Rishi, A. K. and Wetterwald, A. (1996) *Am. J. Respir. Cell Mol. Biol.* **14**, 577–585
- Breiteneder-Geleff, S., Matsui, K., Soleiman, A., Meraner, P., Poczewski, H., Kalz, R., Schaffner, G. and Kerjaschki, D. (1997) *Am. J. Pathol.* **151**, 1141–1152
- Frohman, M. A., Dush, M. K. and Martin, G. R. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8998–9002
- Köhler, G. and Milstein, C. (1975) *Nature (London)* **256**, 495–497
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: a Laboratory Manual*, 2nd edn., Cold Spring Harbor Press, Cold Spring Harbor
- Kyhse-Anderson, J. (1984) *J. Biochem. Biophys. Methods* **10**, 203–209
- Senger, M., Glatting, K. H., Ritter, O. and Suhai, S. (1995) *Comput. Methods Programs Biomed.* **46**, 131–141

(B) A blood vessel at a higher magnification. Note that gp36 is located predominantly at the apical plasma membrane of the endothelium. For comparison, endothelial cells were stained with antibodies to CD34 (C) and von Willebrand factor (D), two markers for this cell type. The bars represent 40 μ m (A) and 25 μ m (B, C and D), respectively.

- 16 von Heijne, G. (1986) *Nucleic Acids Res.* **14**, 4683–4690
- 17 Segrest, J. P. and Jackson, R. L. (1972) *Methods Enzymol.* **28**, 54
- 18 Kobata, A. (1979) *Anal. Biochem.* **100**, 1–14
- 19 Bausch, N. J. and Poretz, R. D. (1977) *Biochemistry* **16**, 5790–5794
- 20 Sastry, M. V., Barnajee, P., Patanjali, S. R., Swamy, M. J., Swanalatha, G. V. and Suroolia, A. (1986) *J. Biol. Chem.* **261**, 11726–11733
- 21 Sueyoshi, S., Tsuji, T. and Osawa, T. (1988) *Carbohydr. Res.* **178**, 213–224
- 22 Shibuya, N., Goldstein, I. J., Broeknaert, W. F., Nsimba-Lubaki, M., Peeters, B. and Peumans, W. J. (1987) *J. Biol. Chem.* **262**, 1596–1601
- 23 Wang, W.-C. and Cummings, R. D. (1988) *J. Biol. Chem.* **263**, 4576–4585
- 24 Knibbs, R. N., Goldstein, I. J., Ratcliffe, R. M. and Shibuya, N. (1991) *J. Biol. Chem.* **266**, 83–88
- 25 Tollefsen, C. E. and Kornfeld, R. (1983) *J. Biol. Chem.* **258**, 5172–5176
- 26 Etzler, M. E. and Kabat, E. A. (1970) *Biochemistry* **9**, 869–877
- 27 Yamashita, K., Totani, K., Ohkura, T., Takasaki, S., Goldstein, I. J. and Kobata, A. (1987) *J. Biol. Chem.* **262**, 1602–1607
- 28 Treutlein, H. R., Lemmon, M. A., Engelman, D. M. and Brünger, A. T. (1992) *Biochemistry* **31**, 12726–12733
- 29 Lemmon, M. A., Flanagan, J. M., Hunt, J. F., Adair, B. D., Bormann, B.-J., Dempsey, C. E. and Engelman, D. M. (1992) *J. Biol. Chem.* **267**, 7683–7689
- 30 Lemmon, M. A., Flanagan, J. M., Treutlein, H. R., Zhang, J. and Engelman, D. M. (1992) *Biochemistry* **31**, 12719–12725
- 31 Carraway, K. L. and Hull, S. R. (1991) *Glycobiology* **1**, 131–138
- 32 Tooze, S. A., Tooze, J. and Warren, G. (1988) *J. Cell Biol.* **106**, 1475–1487
- 33 Spielman, J., Rockley, N. L. and Carraway, K. L. (1987) *J. Biol. Chem.* **262**, 269–275
- 34 Spielman, J., Hull, S. R., Sheng, Z., Kanterman, R., Bright, A. and Carraway, K. L. (1988) *J. Biol. Chem.* **263**, 9621–9629
- 35 Jentoft, N. (1990) *Trends Biochem. Sci.* **15**, 291–294
- 36 Fearon, D. T. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 1971–1975
- 37 Pangburn, M. K. and Müller-Eberhard, H. F. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 2416–2420
- 38 Kazatchkine, M. D., Fearon, D. T. and Austin, K. F. (1979) *J. Immunol.* **122**, 75–81
- 39 Michalek, M. T., Mold, C. and Bremer, E. G. (1988) *J. Immunol.* **140**, 1588–1594
- 40 Meri, S. and Pangburn, M. K. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 3982–3986
- 41 Schauer, R. (1982) *Adv. Carbohydr. Chem. Biochem.* **40**, 131–234
- 42 Schauer, R. (1985) *Trends Biochem. Sci.* **10**, 357–360
- 43 Baumhueter, S., Singer, M. S., Henzel, W., Hemmerich, S., Renz, M., Rosen, S. D. and Lasky, L. A. (1993) *Science* **262**, 436–438
- 44 Varki, A. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 7390–7397

Received 4 January 1999/23 March 1999; accepted 29 April 1999