Identification of Endoglycan, a Member of the CD34/Podocalyxin Family of Sialomucins*

Christopher Sassetti, Annemieke Van Zante, and Steven D. Rosen‡

From the Department of Anatomy, Program in Immunology, and Cardiovascular Research Institute, University of California, San Francisco, California 94143-0452

CD34 and podocalyxin are structurally related sialomucins, which are expressed in multiple tissues including vascular endothelium and hematopoietic progenitors. These glycoproteins have been proposed to be involved in processes as diverse as glomerular filtration, inhibition of stem cell differentiation, and leukocyte-endothelial adhesion. Using homologies present in the cytoplasmic tails of these proteins, we have identified a novel member of this family, which we designate endoglycan. This protein shares a similar overall domain structure with the other family members including a sialomucin domain, but also possesses an extremely acidic amino-terminal region. In addition, endoglycan contains several potential glycosaminoglycan attachment sites and is modified with chondroitin sulfate. Endoglycan mRNA and protein were detected in both endothelial cells and CD34- bone marrow cells. Thus, CD34, podocalyxin, and endoglycan comprise a family of sialomucins sharing both structural similarity and sequence homology, which are expressed by both endothelium and multipotent hematopoietic progenitors. While the members of this family may perform overlapping functions at these sites, the unique structural features of endoglycan suggest distinct functions for this molecule.

CD34 and podocalyxin are evolutionarily related glycoproteins, which share a similar overall domain structure as well as significant sequence homology. Structurally, the extracellular region of each of these molecules is dominated by an amino-terminal mucin-like domain, which is densely substituted with sialylated O-linked carbohydrates. This extensive glycosylation causes mucin domains to adopt an extended, rodlike structure and presumably globular domain. This domain may fold into an immunoglobulin-like structure as the positions of 2 of the cysteines are conserved in the C2 set of the immunoglobulin superfAMILY (3). The cytoplasmic domains of these proteins are 73–76 amino acids in length and highly conserved between species orthologs. It is also in this region that the highest homology between CD34 and podocalyxin is found (4).

CD34 is expressed by multipotent hematopoietic progenitors, but is lost during differentiation and is not present on mature hematopoietic cells (5–7). This property makes CD34 a useful marker for the identification and purification of progenitor cells. CD34+ cells isolated from bone marrow or cord blood are clinically important, since a small number of such cells are able to reconstitute hematopoiesis after myeloablative therapy (8). Expression of CD34 in an immature hematopoietic cell line has been shown to inhibit its differentiation (9), suggesting that one possible function for CD34 is to maintain the undifferentiated phenotype of progenitor cells. The chicken ortholog of podocalyxin, known as thrombomucin, is also expressed on multipotent hematopoietic progenitors as well as thrombocytes (10), and rat podocalyxin is found on platelets (11).

Podocalyxin was originally described as the major sialoprotein on the podocytes of the kidney glomerulus. At this site, podocalyxin is concentrated on the interdigitating secondary foot processes of these cells, where it is thought to maintain the filtration slits between these processes via charge repulsion (12). The importance of the podocyte’s anionic character has been demonstrated by neutralization experiments using either polycations or desialylation. Either treatment disrupts the glomerular filter and induces proteinuria (13, 14).

In addition to these sites, CD34 and podocalyxin are both broadly expressed on the luminal surface of vascular endothelium (15–17). The function of these mucins on unactivated endothelium is not clear, but when properly glycosylated in high endothelial venules (HEV),1 -1 both can function as ligands for the leukocyte adhesion molecule, l-selectin (4, 18). HEV are specialized postcapillary venules present in lymph nodes and Peyer’s patches, which support the recruitment of blood-borne lymphocytes into the lymphoid tissue (19). The adhesion between lymphocytes and HEV is initiated by the lectin domain of l-selectin binding to carbohydrates presented by specific endothelial ligands (20). Since all of the biochemically defined HEV-expressed ligands for l-selectin are sialomucins (4, 21), it is likely that the ability of these structures to present multivalent carbohydrate ligands to the clustered lectin domains of l-selectin is critical for the formation of a high avidity interaction.

Thus, podocalyxin and CD34 are capable of promoting lymphocyte-endothelial adhesion when appropriate glycoforms are expressed by HEV, but these proteins are implicated in differ-

* This work was supported by Merit Award R37 GM23547 (to S. D. R.) from the National Institutes of Health and Predoctoral Fellowship 97-411 (to C. S.) from the American Heart Association, California Affiliate. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBankTM/EBI Data Bank with accession number(s) AF219137.

† To whom correspondence should be addressed: Dept. of Anatomy, University of California, San Francisco, CA 94143-0452. E-mail: sdr@tisa.ucsf.edu.

‡ To whom correspondence should be addressed: Dept. of Anatomy, University of California, San Francisco, CA 94143-0452. E-mail: sdr@tisa.ucsf.edu.

1 The abbreviations used are: HEV, high endothelial venules; AD/ IgG, acidic domain/IgG chimera; EG/IgG, endoglycan/IgG chimera; EST, expressed sequence tag; GAG, glycosaminoglycan; GST, glutathione S-transferase; HMEC, human microvascular endothelial cell(s); HPRT, hypoxanthine phosphoribosyltransferase; HuVEC, human umbilical vein endothelial cell(s); OSCE, O-sialoglycoprotein endoprotease; PBS, Dulbecco’s phosphate-buffered saline; RT, reverse transcriptase; PCR, polymerase chain reaction; PAGE, polyacrylamide gel electrophoresis; CHO, Chinese hamster ovary; bp, base pair(s); kb, kilobase pair(s).
ent specialized functions at other sites (i.e. anti-adhesion in the case of podocalyxin on podocytes, and inhibition of differentiation in the case of CD34 on hematopoietic cells). Therefore, these two related sialomucins appear to be multifunctional proteins whose function depends, at least in part, on tissue-specific glycosylation.

Using homologies present in CD34 and podocalyxin, we have identified a third member of this gene family. This protein shares a similar overall structure with the other family members except for the presence of an extremely acidic amino-terminal domain. As is the case for CD34 and podocalyxin, this protein is detected in both endothelial cells and early hematopoietic progenitors. Due to its endothelial expression and extensive glycosylation, which includes a sialomucin domain and glycosaminoglycan chains, we propose the name endoglycan for this novel gene. We collectively refer to this gene family as the “CD34 family.”

**EXPERIMENTAL PROCEDURES**

Isolation of Endoglycan cDNA— Rapid Screen Human fetal brain cDNA library was purchased from Origene (Rockville, MD). 96 pools of approximately 500,000 cDNA clones each were screened by PCR using the following primers based on an EST corresponding to mouse endoglycan: sense, ATGAATTCTGGATATGACCAATTTGGC; antisense, ATGGATCCGAGCCTCTTCCCCGCTGC (all of the underlined sequences specified in this section were added for cloning purposes and are not present in the endoglycan cDNA). The same screen was then applied to 96 subpools of 100 clones, each derived from two of the positive pools, and then to individual clones from a positive subset. In this way a partial clone was obtained, which corresponded to bases 227–2269 in the sequence of Fig. 1. This sequence was used to identify a human EST, which corresponded to bases 104–386. In order to obtain a full-length clone, the same fetal brain library was rescreened along with a human placenta library (Origene) using the following primers, which amplify the extreme 5′ end of the human EST sequence: sense, GCTGGGTCTGATGAGCCTGG; antisense, TACTGTCCTTACTGCGGCTGC. Positive pools were then screened with the same antisense primer and a vector primer (vector primer 3, Origene) in order to identify the longest clone. Screening of subpools was performed, as above, resulting in isolation of a clone from the brain library with the sequence shown in Fig. 1.

**Northern Blots**— Multiple-tissue Northern blots were purchased from CLONTECH, hybridized with an endoglycan probe, stripped, and rehybridized with a β-actin probe. The endoglycan probe consisted of bases 1013–1525, which had been amplified from a cDNA clone by PCR using the following primers: sense, AATGGATCCGAGCCTCTTCCCGCTGC; antisense, ATGGATCCGAGCCTCTTCCCCGCTGC. Positive pools were then rescreened with the same antisense primer and a vector primer (vector primer 3, Origene) in order to identify the longest clone. Screening of subpools was performed, as above, resulting in isolation of a clone from the brain library with the sequence shown in Fig. 1.

**Reverse Transcriptase-PCR**—HUVEC for RT-PCR experiments were purchased from Clonetics and grown according to the manufacturer’s recommendations. Total RNA was prepared using RNAzol B (Tel-Test, Friendswood, TX). First-strand cDNA was prepared from 2 μg of RNA primed with random hexamers using AMV-RT (Life Technologies, Inc.). A 414-bp endoglycan cDNA fragment was amplified from HUVEC cDNA using the primers based on the murine endoglycan sequence given above. A 420-bp podocalyxin cDNA fragment was amplified using the following primers: sense, ATGAATTCTGGATATGACCAATTTGGC; antisense, ATGGATCCGAGCCTCTTCCCCGCTGC. 45 cycles of PCR were performed on cDNA corresponding to 20 ng of input RNA using Advantage cDNA polymerase (CLONTECH). Cryopreserved CD34+ bone marrow cells were purchased from Poietics (Gaithersburg, MD). cDNA was prepared from total RNA extracted from 60,000 cells, as above. A 519-bp endoglycan fragment and a 289-bp HPRT fragment were amplified using the following primers:

for endoglycan, sense (ATAGGATCCGAGCCTCTTCCCCGCTGC) and antisense (ATAGATTCCTGGAGCTTCGCTTGTCGTG); for HPRT, sense (CTCGTGTGATTACATCAAAAGACCTG) and antisense (TCAACACTTCGGTTGCTCTCTG). 35 cycles of PCR were performed on cDNA corresponding to 2000 input cells. The resulting DNA was electro- phoresed on 1% agarose and visualized with ethidium bromide.

**Metabolic Labeling with [32P]**—[32P]dATP was supplied with the Strip-EZ DNA kit (Ambion, Austin, TX). Hybridization and washing was performed as recom- mended by the blot manufacturer. The final wash was performed in 0.1× SSC, 0.1% SDS at 50 °C. Blots were stripped using the Strip-EZ DNA kit (Ambion).

**HUVEC (Clonetics)** were grown to confluence in 200 cm² flasks as recommended. Cells were then washed with PBS and cultured for 16 h in sulfate-free Dulbecco’s modified Eagle’s medium containing 10% dialyzed fetal calf serum (Life Technologies, Inc.) and 10 ng/ml epidermal growth factor, 1 μg/ml streptomycin, and 100 units/ml penicillin. After transfection, cells were cultured in Opti-MEM (Life Technologies, Inc.) supplemented with 2 mM l-glutamine, 100 units/ml penicillin, 100 μg/ml streptomycin, and 200 μCi of Na2[35S]SO4 (ICN, Costa Mesa, CA) for 15 h. 1 ml of 4-methylumbelliferyl-β-D-xyloside (Sigma), 10 mM 4-methylumbelliferyl-β-D-mannoside (Sigma), and 1 mg/ml heparin (Sigma) was included in the above medium. For 48 h. 1 mM 4-methylumbelliferyl-β-D-glucoside (Sigma) was added to the medium for fluorescence microscopy. Fusion proteins were isolated by incubating aliquots of conditioned medium for 10 μl of protein A-Sepharose (Repligen, Cambridge, MA) for 2 h at 4 °C. Precipitates were washed with PBS, eluted with SDS-PAGE sample buffer containing 1% mercaptoethanol, boiled for 5 min, and analyzed by SDS-PAGE as described on 7.5% SDS-polyacrylamide gels. Gels were stained with Gel Code Blue (Pierce), destained, incubated with Fluorohance (Research Products International Corp, Mount Prospect, IL) for 30 min, and dried for fluorography.

**HUVEC** for Northern blot analysis were isolated from umbilical cords (San Francisco General Hospital) following published procedures (22) except that collagenase A (Roche Molecular Biochemicals) was used. HUVEC were grown in EGM medium (Clonetics, Walkersville, MD), and mRNA was isolated using Oligotex Direct kit (Qiagen, Valencia, CA). 5 μg of mRNA was electrophoresed on a 1% agarose/formaldehyde gel and transferred to Hybond N+ filters (Amersham Pharmacia Biotech). Blots were hybridized as above with an endoglycan probe consisting of bases 1–592 that had been excised from the cDNA clone by restriction digestion with EcoRI and BamHI.

**Antibody Preparation**—A fusion protein consisting of the predicted extracellular domain of endoglycan (bases 9–1511) fused to glutathione S-transferase (GST; bases 258–917 of pGEX-2T vector, Amersham Pharmacia Biotech) was constructed in the PEAK10 vector (Edge Bio-systems, Gaithersburg, MD). The endoglycan fragment was amplified by PCR using the following primers: sense (ATAGGATCCGAGCCTCTTCCCCGCTGC) and antisense (ATAGGATCCGAGCCTCTTCCCCGCTGC). 35 cycles of PCR were performed on cDNA corresponding to 2000 input cells. The resulting DNA was electro- phoresed on 1% agarose and visualized with ethidium bromide.
Identification of Endoglycan

**RESULTS**

Isolation of Endoglycan cDNA—The highest sequence homology between CD34 and podocalyxin occurs in the cytoplasmic domain containing with 0.6 mM EDTA in PBS (without Ca\(^{2+}\) and Mg\(^{2+}\)) for 20 min at room temperature. For staining with rabbit antibodies, cells were incubated with 10 μg/ml affinity-purified anti-endoglycan antibody or normal rabbit IgG (Sigma) in PBS containing 1% bovine serum albumin (Sigma), 2% normal goat serum, and 0.2% sodium azide (staining buffer). Bound antibodies were detected with Cy 3-conjugated goat anti-rabbit IgG (Zymed Laboratories Inc., South San Francisco, CA) in staining buffer. Staining with mouse IgG1 (Zymed Laboratories Inc.) was identical except the staining buffer contained 2% normal rabbit serum instead of goat serum.

Peptide Microsequencing—A cDNA encoding the AD/IgG fusion protein was transfected into COS-7 cells with LipofectAMINE, and cells were cultured in Opti-MEM (Life Technologies, Inc.) supplemented with glutamine, penicillin, and streptomycin. The conditioned medium was centrifuged at 20,000 g for 15 min at 4 °C, and Tris, pH 8.0, and sodium azide were added to final concentrations of 50 mM and 0.02% (dashed underline), predicted transmembrane domain (boxed underline), cysteines in the membrane proximal globular domain (circled underline), and potential sites for glycosaminoglycan addition (double underline), N-linked carbohydrate addition (triangles), and tyrosine sulfation (asterisk) are indicated.

**Fig.1.** Human endoglycan cDNA and deduced amino acid sequence. Amino acids identified by amino-terminal sequence analysis of the mature protein are boxed. The signal sequence (underlined), predicted transmembrane domain (boxed underline), cysteines in the membrane proximal globular domain (circled underline), and potential sites for glycosaminoglycan addition (double underline), N-linked carbohydrate addition (triangles), and tyrosine sulfation (asterisk) are indicated.
Identification of Endoglycan

Fig. 2. Detection of endoglycan mRNA in human tissues. A, Northern blot analysis of mRNA isolated from various tissues. Northern blots were hybridized with an endoglycan probe (top), followed by a β-actin probe (bottom). B, endoglycan mRNA was detected in HUVEC by reverse transcriptase-PCR. A 141-bp fragment of endoglycan or a 420-bp fragment of podocalyxin was amplified from HUVEC cDNA. +/−RT indicates the presence or absence of reverse transcriptase in the cDNA synthesis reactions. The absence of PCR products in −RT lanes demonstrates that RNA and not genomic DNA was the template.

mic domains of these proteins. In order to identify additional members of this gene family, we searched (tblastx) (27) the GenBank expressed sequence tag (EST) library using the peptide sequence of the cytoplasmic domain of human podocalyxin as a probe. Several overlapping mouse EST sequences were identified encoding a peptide sequence that was 44% identical to a 113-amino acid region encompassing the transmembrane and cytoplasmic domains of podocalyxin.

In order to obtain a full-length human cDNA corresponding to this gene, PCR primers were designed based on the available mouse sequence, which would be predicted to amplify a product corresponding to the human ortholog of this gene but not to the two known family members. This approach took advantage of the observation that the transmembrane domains of CD34 and podocalyxin differ from each other, but are highly conserved between species orthologs (16, 28–30). Thus, one primer was based on sequence in the predicted transmembrane domain and the other was based on a region in the cytoplasmic tail, which is conserved between this protein and podocalyxin (H-DNPTLDV, see Fig. 8). These primers amplified the same sized product from mouse and human cDNA libraries and were used to screen a human fetal brain cDNA library by the strategy described under “Experimental Procedures.” One full-length and one partial clone (bases 227–2269) were obtained and sequenced (Fig. 1). The full-length cDNA contains a single open reading frame, encoding a protein of 605 amino acids, followed by a 3′-untranslated region of 322 bp and a poly(A) tail. The start codon is identified in a strong context for translation initiation (31) and is followed by a hydrophobic region of 32 amino acids, which satisfies the criteria for a cleavable signal peptide (32).

Hydropathy analysis (33) of the derived amino acid sequence predicted a single transmembrane domain of 25 residues near the carboxyl terminus. Overall, the cDNA encodes a type I transmembrane protein with a similar domain structure to CD34 and podocalyxin. The 80-amino acid cytoplasmic domain is similar in length to the other family members (73 for CD34 and 76 for podocalyxin) and shares significant homology with both (58% identity with human podocalyxin and 33% with human CD34). The 500-amino acid extracellular region contains a membrane-proximal, cysteine-containing and presumably globular domain. This domain is similar to those found in the other family members, except that only three cysteines are present as compared with six and four in CD34 and podocalyxin, respectively. Amino-terminal to this structure is a domain of 156 amino acids, which contains 36% serine, threonine, and proline. This is twice the average content of these residues in human proteins (34) and is characteristic of mucin-like domains. At the amino terminus of the predicted protein is a highly acidic domain of 161 amino acids (after signal peptide cleavage), which contains 30% acidic residues and is not found in the other family members. This domain is particularly rich in glutamate, containing three polyglutamate tracts of 5–11 residues.

Inspection of the sequence revealed many potential sites for post-translational modification (Fig. 1). In addition to the dense O-linked carbohydrate expected to be present in the mucin-like domain, four potential sites of N-linked carbohydrate addition are present. The two tyrosines in the acidic amino-terminal domain are potential sites of sulfation (35). Six serine-glycine and two serine-alanine pairs, which are potential glycosaminoglycan attachment sites (36), are distributed throughout the extracellular region. Intracellularly, two potential sites for casein kinase II phosphorylation ((S/T)XX(D/E)) are found. The product of this cDNA will be referred to as endoglycan, based on the characterization provided below.

Tissue Distribution of Endoglycan mRNA—Northern blots containing poly(A)+ RNA from different human tissues were probed with a fragment of the endoglycan cDNA. A major 2.5-kb band was detected in several tissues. This mRNA was most prominent in brain but was also detected in pancreas, kidney, liver, and all hematopoietic and lymphoid tissues that were tested. (Fig. 2A). We suspected that this broad expression pattern could be indicative of endothelial expression, since both CD34 and podocalyxin are widely expressed by vascular endothelium (15–17). To address this possibility, a Northern blot
proteins were analyzed by SDS-PAGE and autoradiography. 

The presence of endoglycan mRNA in HUVEC was verified using reverse transcriptase-PCR. Endoglycan specific primers amplified a fragment of the expected size (Fig. 2B), demonstrating the presence of endoglycan mRNA in these cells. In the same experiment, a band of similar intensity was amplified with primers specific for podocalyxin. Similar results were obtained when RNA from the human microvascular endothelial cell line HMEC-1 (37) was probed for endoglycan mRNA by Northern blotting and RT-PCR (data not shown).

Biochemical Characterization of Recombinant Endoglycan—
IgG fusion proteins consisting of either the entire extracellular domain (amino acid 1–497) of endoglycan (EG/IgG) or only the amino-terminal acidic domain (amino acid 1–206, AD/IgG) were constructed and expressed in COS-7 cells. Secreted fusion proteins were then purified from the conditioned medium using protein A-Sepharose. The amino terminus of the mature protein was determined by subjecting the purified AD/IgG protein to amino-terminal sequence analysis. Two related sequences were found in similar amounts, GSDEP and SDEPG (Fig. 1). This result indicated that an approximately equimolar mixture of two differentially processed forms of the protein were present: one with an amino terminus of glycine 33 and the other beginning at serine 34, and thus confirmed that the preceding residues constitute a cleavable signal peptide.

We next investigated which of the many potential post-translational modifications were actually present on the endoglycan fusion proteins. Since both tyrosine sulfate and sulfated glycosaminoglycan modifications were suspected, the transfected cells were labeled with Na$_2^{35}$SO$_4$, and the purified fusion proteins were analyzed by SDS-PAGE and autoradiography. $^{35}$SO$_4$ was incorporated into both full-length EG/IgG and the AD/IgG but not into the IgG tail expressed alone (Fig. 3A), although similar amounts of all three proteins were observed by Coomassie Blue staining of the gels (data not shown).

Both endoglycan fusion proteins migrated with apparent molecular weights much greater than those predicted for their respective peptides (EG/IgG = 78 kDa, AD/IgG = 44 kDa). The apparent molecular mass of 180 kDa for EG/IgG was more than twice this value, consistent with extensive post-translational modifications. To determine if the S-T-P-rich domain of endoglycan was mucin-like in structure, we employed O-sialoglycoprotein endopeptidase (OSGE), a protease that specifically degrades sialomucin domains (38, 39). As seen with the other family members (4), OSGE completely degraded EG/IgG (Fig. 3C), confirming the sialomucin-like character of this protein. As this enzyme requires its substrates to be sialylated (38), this experiment also provides evidence that endoglycan is modified with sialic acid containing glycans. OSGE treatment had only a minimal effect on the AD/IgG protein that lacks the mucin domain, verifying the absence of contaminating proteases in the enzyme preparation (data not shown).

Each of the $^{35}$SO$_4$-labeled fusion proteins migrated as two distinct species by SDS-PAGE. In order to investigate whether one or both of these sulfated species was modified with glycosaminoglycan chains, the labeled proteins were digested with heparinas or chondroitinase prior to electrophoresis. The high molecular weight species in both cases was sensitive to chondroitinase ABC but not to heparinase I or III (Fig. 3C), demonstrating that endoglycan was modified with chondroitin sulfate in these cells. Furthermore, treatment of the cells with $\alpha$-D-xyloside, a competitive inhibitor of both heparan sulfate and chondroitin sulfate GAG chain addition, prevented the formation of the high molecular weight form of AD/IgG (Fig. 3B) without affecting the low molecular weight species. Since only one potential glycosaminoglycan addition site (SG/A) is present in the AD fragment, we suspect that at least this site (serine 79) was modified. In contrast, the lower molecular weight component of each fusion protein exhibited a sulfate
modification that was resistant to chondroitinase and hepari-
nase digestions, as well as xyloside treatment. The nature of
this modification has not yet been characterized, but may rep-
resent sulfation of either another type of carbohydrate or
tyrosine.

In order to determine if GAG modifications were present on
the other members of the CD34 family, CD34/IgG and podoca-
lyxin/IgG fusion proteins were produced in COS-7 cells and
metabolically labeled with Na35SO4. When compared with EG/
IgG, these fusion proteins incorporated significantly less35SO4.
Furthermore, the bands representing both CD34/IgG and podo-
calyxin/IgG were resistant to treatment with a mixture of chon-
droitinase and heparinases (Fig. 3). Thus, CD34 and podoca-
lyxin are not modified with GAG chains in COS-7 cells.

Generation of Antibodies and Characterization of the Endo-
thelial Form of Endoglycan—A glutathione S-transferase fu-
sion protein containing the full extracellular domain of en-
doglycan was expressed in mammalian cells, purified, and used
to immunize rabbits. Antibodies were then affinity-purified on
immobilized antigen. The anti-endoglycan antibody stained
CHO cells that had been transfected with endoglycan but not
CD34 or podocalyxin transfectants (Fig. 4A), indicating that
the antibody did not cross-react with the other family mem-
bers. The antibody stained intact HUVEC (Fig. 4B) and HMEC
data not shown), thus establishing the cell surface expression
of endoglycan on these endothelial cells.

To characterize the structure of the native form of endogly-
can on endothelial cells, HUVEC were labeled with Na 2
35SO4
and detergent lysates were immunoprecipitated with the anti-
endoglycan antibody. Two sulfated species corresponding to
165 and 200 kDa were specifically precipitated (Fig. 5). The
apparent molecular mass of the more abundant component
(165 kDa) was greater than 2.5-fold larger than the predicted
molecular mass of the peptide (62 kDa), indicating significant
glycosylation. Both of these species were OSGE- and chon-
droitinase ABC-sensitive but heparinase-insensitive (Fig. 5).
These results established that native endoglycan produced by
endothelial cells was a chondroitin sulfate-modified sialomu-
cin. While treatment with both heparinase and chondroitinase
removed nearly all of the35SO4 from endoglycan, a labeled
species corresponding to 130 kDa remained. Therefore, as was
seen with recombinant endoglycan from COS-7 cells, the endo-
thelial form of the protein contained a sulfate modification that
was independent of GAG chains.

The distribution of endoglycan in situ was studied by immu-
nohistochemical staining of human foreskin with the anti-en-
doglycan antibody. In order to identify endothelium, an anti-
body to PECAM-1 was used (40). Endoglycan and PECAM-1
were colocalized on the endothelium of many blood vessels, as
shown in Fig. 6 (arrows). In addition, the anti-endoglycan an-
tibody stained smooth muscle bundles throughout the dermis
(Fig. 6, arrowheads) as well as smooth muscle surrounding
arterial vessels (data not shown).

Expression of Endoglycan on Hematopoietic Cells—Both
CD34 and podocalyxin are found on multipotent hematopoietic
progenitors (5–7, 10). In order to determine if endoglycan was also present on these cells, we stained purified human CD34

bone marrow cells with the anti-endoglycan antibody. As shown in Fig. 7 (A and B), almost all of these purified cells expressed both CD34 and endoglycan. As we do not know whether all CD34

1

cells are isolated by this purification method, it is possible that a population of bone marrow cells expressing lower CD34 levels may exist that do not express endoglycan.

To verify endoglycan expression on CD34

1

bone marrow cells, RT-PCR was performed on RNA extracted from these cells (98% CD34

1

). As shown in Fig. 7C, a 521-bp fragment of endoglycan or a 289-bp fragment of HPRT was amplified from RNA isolated from 98% pure CD34

1

bone marrow cells by RT-PCR. +/-RT indicates the presence or absence of reverse transcriptase in the cDNA synthesis reactions. D, peripheral blood cells were stained with anti-endoglycan (thick line) or rabbit IgG (dashed line). Monocytes were identified by staining with CD14, and then gated by forward and side scatter.

DISCUSSION

CD34, podocalyxin, and endoglycan represent a novel family of sialomucins, all of which are expressed on endothelial cells and hematopoietic precursors. Members of this CD34 family are defined by their overall domain structure as well as sequence homology (Fig. 8). The mucin-like domains of these proteins are 130–290 amino acids in length and show no obvious sequence conservation except for the high content of serine, threonine and proline residues. Unlike many other mucins, these domains do not exhibit any obvious sequence repeats (this study; Refs. 16 and 29). Dense O-glycosylation is inferred,
since the mature glycoproteins migrate with apparent molecular weights 2–3-fold larger than the predicted core proteins, and by their sensitivity to OSGE (Fig. 5; Ref. 4).

A membrane-proximal globular domain is predicted in all family members, although the sequence conservation is low (26–30% identity) and the number of cysteines varies (six in CD34, four in podocalyxin, and three in endoglycan). Since the relative positions of the amino-terminal two cysteines are conserved in all three family members, these residues are likely to be involved in an intrachain disulfide bond. The third cysteine (Cys-487) in endoglycan is not conserved and may be available for interchain disulfide bond formation involved in homodimerization or hetero-oligomerization.

Interestingly, the most significant sequence homology among CD34, podocalyxin, and endoglycan is found in their cytoplasmic domains with limited discernible conservation in the transmembrane or extracellular regions (Fig. 8, B and C). The cytoplasmic domains of these proteins are 33–58% identical, with three regions of 6–12 residues exhibiting greater than 50% similarity. In addition, endoglycan and podocalyxin share an HDNPTL/E/D/V motif that is not present in CD34 (Fig. 8C). The similarity in this region suggests that the intracellular domains serve a similar function in these proteins. Multiple lines of evidence support a function for the cytoplasmic domain of CD34 in cell signaling. Potential sites for serine/threonine phosphorylation are present in the cytoplasmic domains of all three family members (this study; Refs. 8 and 16) and CD34 is phosphorylated after protein kinase C activation (41). Additionally, in contrast to the full-length form, a natural splice variant of CD34 lacking most of the cytoplasmic domain does not inhibit the differentiation of an immature cell line when overexpressed (9). A splice variant of podocalyxin with a truncated cytoplasmic tail has also been described in chicken (10). These data suggest that the cytoplasmic domains of proteins in this family participate in cell signaling functions, and that alternative splicing in these regions may have important consequences.

Despite the relatedness of CD34 and podocalyxin, they are not genetically linked. The CD34 gene is located on 1q32 (42, 43) and podocalyxin is on 7q32–33 (44). The chromosomal location of endoglycan has not been determined.

Endoglycan exhibits several structural features that are not found in the other family members. Most striking is the presence of an extremely acidic amino-terminal domain characterized by several polyglutamate tracts. While polyglutamate occurs in many intracellular proteins involved in transcriptional regulation, this structure is rare in extracellular domains (45, 46). Endoglycan is also modified with chondroitin sulfate modifications. Glycosaminoglycan modifications have not been reported for CD34 or podocalyxin and were not found when these proteins were produced in COS-7 cells. A previous analysis of the carbohydrates carried by glomerular podocalyxin also failed to reveal significant amounts of glycosaminoglycans (47). Thus, GAG chain addition appears to be unique to endoglycan among the CD34 family members. Another sulfate modification is also present on endoglycan that is unrelated to glycosaminoglycans.

---

**Fig. 8. Overview of the CD34 family of sialomucins.**

*Fig. 8, A.* Comparison of the domain structure of the family members. Acidic refers to the unique 161-amino acid acidic domain of endoglycan. The mucin domains contain greater than 36% serine, threonine, and proline. This domain varies in length between family members (130 amino acids for CD34, 156 for endoglycan, and 290 for podocalyxin).

*Fig. 8, B.* Similarity plot of human podocalyxin (vertical) and human endoglycan (horizontal) showing the lack of homology except for the cytoplasmic domains. Sequences were analyzed with Compare and DotPlot programs (Genetics Computer Group, Madison, WI) using a window size of 40 and a stringency of 15.

*Fig. 8, C.* Alignment of the conserved cytoplasmic tails of the CD34 family members. Sequences shown begin at amino acid 281 of human CD34, 309 of mouse CD34, 525 of human endoglycan, 453 of human podocalyxin, and 477 of rabbit podocalyxin and continue to the carboxyl terminus of each protein. Note the high interspecies conservation of both CD34 (90% identity between human and mouse) and podocalyxin (96% identity between human and rabbit). Regions of greater than 50% similarity between family members are boxed.
glycan chains. While this modification remains uncharacterized, we suspect that it may represent sulfation of the two tyrosine residues in the acidic domain, which are in an acidic context that promotes sulfation (35).

Although the functions of CD34 and podocalyxin on unactivated endothelium remain unclear, several functions have been proposed for endothelial proteoglycans, for which endoglycan is a new example. In particular, proteoglycans have been proposed to act as presentation molecules or coreceptors for a number of growth factors and chemokines (48, 49). At sites of inflammation, chemokines derived from a tissue source or produced by endothelial cells are presented on the luminal surface of the vascular endothelium by binding to “presentation molecules.” Proteoglycans with unspecified protein cores have been implicated in this function (49, 50). These interactions are thought to allow chemokines to interact with blood-borne leukocytes without being diluted by the flowing blood (51).

Although heparan sulfate proteoglycans have primarily been implicated as chemokine presentation molecules, several chemokines can bind to chondroitin sulfate (52, 53), suggesting that endoglycan could play a role in this process.

Cell surface proteoglycans are also essential for optimal responses to several growth factors. In many cases a co-receptor role has been proposed, in which growth factor binding to proteoglycans on a target cell promotes more efficient interactions between the growth factor and its signaling receptor. In this way, the proteoglycan co-receptor can contribute to the overall avidity of growth factor/receptor binding as well as to the specificity of the interaction (48, 54). The protein-tyrosine phosphatase ζ (PTPζ) is a transmembrane proteoglycan, which itself may be a signaling receptor for the growth factors prolactin and midkine in the central nervous system. High affinity binding of these ligands to PTPζ requires the presence of chondroitin sulfate chains (55, 56). These observations suggest that a role for endoglycan in growth factor signaling, either direct or indirect, merits investigation.

The ability of both CD34 and podocalyxin to serve as ligands for L-selectin when they are properly glycosylated (e.g. in HEV) raises the possibility that specific glycoforms of endoglycan may be able to perform this function as well. The posttranslational requirements involved in the binding of HEV ligands to L-selectin have been investigated extensively. Specifically, categorizing groups including galactose-6-O-SO3 and N-acetylgalcosamine-6-O-SO3 in the context of sialyl-Lewis x have been implicated in L-selectin ligand activity (57–59). Interestingly, a chondroitin sulfate-modified siamolin with L-selectin binding activity has been isolated from cultures of rat high endothelial cells (60). This material was resolved into 150- and >200-kDa species by SDS-PAGE, which is reminiscent of the 165- and >200-kDa species observed for HUVEC endoglycan. Preliminary immunohistochemistry indicates the presence of endoglycan on a subset of HEV in human tonsil. Consistent with this immunolocalization, endoglycan mRNA is detected in purified high endothelial cells by RT-PCR. Whether endoglycan is an HEV ligand for L-selectin is a subject for further study. There is also evidence for extralymphoid L-selectin ligands, which are induced on endothelium by proinflammatory stimuli (61, 62). Since biochemical identification of these ligands is lacking, endoglycan should be considered in this context as well.

Endoglycan is present on CD34+ bone marrow cells and is absent on most mature blood cells, except monocytes. This expression pattern not only provides another potential marker for the identification and purification of hematopoietic progenitors, but as observed with CD34, may have functional consequences as well. The determination of the relative expression levels of the CD34 family members during the differentiation of different lineages promises to provide interesting clues as to the function of this gene family in hematopoiesis.

Acknowledgments—We thank Valerie Tiger-Hughes at San Francisco General Hospital for invaluable assistance in obtaining umbilical cords. We thank Dr. Robert Atkins for providing the PHM5 antibody. Dr. David Simmons for providing the pGl1 plasmid, and Dr. Kelly McNagny for helpful comments on this manuscript.

REFERENCES

1. Li, F., Erickson, H. P., James, J. A., Moore, K. R., Cummings, R. D., and McEver, R. P. (1996) J. Biol. Chem. 271, 6342-6347
2. Cyster, J. G., Shotton, D. M., and Williams, A. F. (1991) Immuno. Rev. 111, 191-198
3. Civin, C. I., Strouse, L. C., Brownell, C., Fackler, M. J., Schwartz, J. F., and Champion, J. H. (1984) J. Immunol. 133, 157-165
4. Krause, D. S., Fackler, M. J., Civin, C. I., and May, W. S. (1996) Blood 87, 1-13
5. Fackler, M. J., Krause, D. S., Smith, O. M., Civin, C. I., and May, W. S. (1995) Blood 85, 3040-3047
6. McNagny, K. R., Pettersson, I., Rossi, F., Flamme, I., Shevchenko, A., Mann, M., and Graf, T. (1997) J. Cell Biol. 138, 1395-1407
7. Kostinin, A., Solin, M. L., Reivanen, J., Javonon, E., Väisänen, R., and Holhöfer, H. (1999) Am. J. Pathol. 154, 813-822
8. Kierjaschki, D., Sharkey, D. J., and Farquhar, M. G. (1994) J. Cell Biol. 98, 1591-1596
9. Andrews, P. M. (1979) Kidney Int. 15, 376-385
10. Seiler, M. W., Venkatraman, M. A., and Cotran, R. S. (1975) Science 189, 390-393
11. Fina, L., Melgada, H. V., Robertson, D., Bradley, N. J., Monaghan, P., Delia, D., Sutherland, D. R., Baker, M. A., and Greaves, M. F. (1990) Blood 75, 2417-2426
12. Kershaw, D. B., Beck, S. G., Wharram, B. L., Wiggins, J. E., Goyal, M., Thomas, P. E., and Wiggins, R. C. (1997) J. Biol. Chem. 272, 15708-15714
13. Kierjaschki, D., Poczewski, H., Dekan, G., Horvat, R., Balzar, E., Krafi, N., and Atkins, R. C. (1986) J. Clin. Invest. 78, 1142-1149
14. Baumbheer, S., Singer, M. S., Henzel, W., Hemmerich, S., Renz, M., Rosen, S. D., and Lasky, L. A. (1993) Science 262, 436-438
15. Girard, J. P., and Springer, T. A. (1995) Immunol. Today 16, 449-457
16. Hynes, R. O. (1992) J. Clin. Invest. 89, 1025-1030
17. Kierjaschki, D., Sharkey, D. J., and Farquhar, M. G. (1994) J. Cell Biol. 98, 1591-1596
18. Puri, K. D., Finger, E. B., Gaudernack, G., and Springer, T. A. (1995) J. Cell Biol. 131, 261-270
19. Jaffe, E. A., Nachman, R. L., Becker, C. G., and Minick, C. R. (1973) J. Clin. Invest. 52, 2745-2758
20. Mizushima, S., and Nagata, S. (1990) Nucleic Acids Res. 18, 5322
21. Simmons, D. L. (1993) in Cellular Interactions in Development: A Practical Approach (Hartley, D. A., ed) pp. 93-127, IRL Press, Oxford
22. Harlow, E., Lane, D. (1999) Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
23. Eaton, D. L., Wood, W. I., Eaton, D., Hass, E. F., Hollingshead, P., Wion, K., Muller, J., Lawes, R. M., Vehar, G. A., and Gorman, C. (1986) Biochemistry 25, 8343-8347
24. Alltchel, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) J. Mol. Biol. 215, 403-430
25. Brown, J., Greaves, M. F., and Melgada, H. V. (1991) Int. Immunol. 3, 175-184
26. Simoons, D. L., Satterthwaite, A. B., Tenen, D. G., and Seed, B. (1992) J. Immunol. 148, 267-271
27. Kershaw, D. B., Thomas, P. E., Wharram, B. L., Goyal, M., Wiggins, J. E., Whitehead, C. I., and Wiggins, R. C. (1995) J. Biol. Chem. 276, 29439-29446
28. Rehfeld, J. F. (1996) Mamm. Genomics 1, 563-574
29. Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997) Protein Eng. 10, 1-4
30. Kyte, J. and Doolittle, R. F. (1982) J. Mol. Biol. 157, 105-132
31. Doolittle, R. F. (1986) Of Urfs and Orfs, University Science Books, Mill Valley, CA
43. Howell, S. M., Molgaard, H. V., Greaves, M. F., and Spurr, N. K. (1991) *Hum. Genet.* **87**, 625–627
44. Kershaw, D. B., Wiggins, J. E., Wharram, B. L., and Wiggins, R. C. (1997) *Genomics* **45**, 239–240
45. Bannwarth, S., Giordanengo, V., Lesimple, J., and Lefebvre, J. C. (1998) *J. Biol. Chem.* **273**, 1911–1916
46. Fisher, L. W., McBride, O. W., Termine, J. D., and Young, M. F. (1990) *J. Biol. Chem.* **265**, 2347–2351
47. Dekan, G., Gabel, C., and Farquhar, M. G. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88**, 5398–5402
48. Carey, D. J. (1997) *Biochem. J.* **327**, 1–16
49. Middleton, J., Neil, S., Wintle, J., Clark-Lewis, I., Moore, H., Lam, C., Auer, M., Huh, E., and Rot, A. (1997) *Cell* **91**, 385–396
50. Weber, K. S., von Hundelshausen, P., Clark-Lewis, I., Weber, P. C., and Weber, C. (1999) *Eur. J. Immunol.* **29**, 700–712
51. Rot, A. (1992) *Immunol. Today* **13**, 291–294
52. Peterson, F., Brandt, E., Lindahl, U., and Spillmann, D. (1998) *J. Biol. Chem.* **274**, 12376–12382
53. Wolff, E. A., Greenfield, B., Taub, D. D., Murphy, W. J., Bennett, K. L., and Aruffo, A. (1999) *J. Biol. Chem.* **274**, 2518–2524
54. Kan, M., Wu, X., Wang, F., and McKeehan, W. L. (1999) *J. Biol. Chem.* **274**, 15947–15952
55. Maeda, N., Nishiwaki, T., Shintani, T., Hamanaka, H., and Noda, M. (1996) *J. Biol. Chem.* **271**, 21446–21452
56. Maeda, N., Ichihara-Tanaka, K., Kimura, T., Kudomatsu, K., Muramatsu, T., and Noda, M. (1999) *J. Biol. Chem.* **274**, 12474–12479
57. Hemmerich, S., Leffler, H., and Rosen, S. D. (1995) *J. Biol. Chem.* **270**, 12035–12047
58. Hemmerich, S., Bertozzi, C. R., Leffler, H., and Rosen, S. D. (1994) *Biochemistry* **33**, 4820–4829
59. Hemmerich, S., and Rosen, S. D. (1994) *Biochemistry* **33**, 4830–4835
60. Derry, C. J., Favezou, C., Mordasy, K. R., and Ager, A. (1999) *Eur. J. Immunol.* **29**, 419–430
61. Zakrzewicz, A., Grafe, M., Terbeck, D., Bongrazio, M., Auch-Schwelk, W., Walzog, B., Graf, K., Fleck, E., Ley, K., and Gaehtgens, P. (1997) *Blood* **89**, 3228–3235
62. Ley, K., Tedder, T. F., and Kansas, G. S. (1993) *Blood* **82**, 1632–1638
Identification of Endoglycan, a Member of the CD34/Podocalyxin Family of Sialomucins
Christopher Sassetti, Annemieke Van Zante and Steven D. Rosen

J. Biol. Chem. 2000, 275:9001-9010.
doi: 10.1074/jbc.275.12.9001

Access the most updated version of this article at http://www.jbc.org/content/275/12/9001

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts

This article cites 58 references, 32 of which can be accessed free at http://www.jbc.org/content/275/12/9001.full.html#ref-list-1