Core glycosylation of collagen is initiated by two beta(1-O)galactosyltransferases

Schegg, B ; Hülsmeier, A J ; Rutschmann, C ; Maag, C ; Hennet, T

Abstract: Collagen is a trimer of three left-handed alpha-chains representing repeats of the motif Gly-X-Y, whereas (hydroxy)proline and (hydroxy)lysine residues are often found at positions X and Y. Selected hydroxylysine are further modified by the addition of galactose and glucose-galactose units. Collagen glycosylation takes place in the endoplasmic reticulum before triple helix formation and is mediated by beta(1-O)galactosyl- and alpha(1-2)glucosyltransferase enzymes. We have identified two collagen galactosyltransferases using affinity chromatography and tandem-MS protein sequencing. The two collagen beta(1-O)galactosyltransferases corresponded to the GLT25D1 and GLT25D2 proteins. Recombinant GLT25D1 and GLT25D2 enzymes showed a strong galactosyltransferase activity towards various types of collagen and towards the serum mannose-binding lectin MBL, which contains a collagen domain. Amino acid analysis of the products of GLT25D1 and GLT25D2 reactions confirmed the transfer of galactose to hydroxylysine residues. The GLT25D1 gene is constitutively expressed in human tissues, whereas the GLT25D2 gene is only expressed at low levels in the nervous system. The GLT25D1 and GLT25D2 enzymes are similar to CEECAM1, to which we could not attribute any collagen galactosyltransferase activity. The GLT25D1 and GLT25D2 genes now allow addressing the biological significance of collagen glycosylation and the importance of this post-translational modification in the etiology of connective tissue disorders.

DOI: https://doi.org/10.1128/MCB.02085-07

Core glycosylation of collagen is initiated by two β(1-O)galactosyltransferases

Running Title: Collagen galactosylation

Belinda Schegg, Andreas J. Hülsmeier, Christoph Rutschmann, Charlotte Maag, Thierry Hennet§

Institute of Physiology and Center for Integrative Human Physiology, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich

§Corresponding author: Thierry Hennet
Institute of Physiology, University of Zurich,
Winterthurerstrasse 190, CH-8057 Zurich, Switzerland
Tel: +41-1-635-5080, Fax: +41-1-635-6814
E-mail: thennet@access.uzh.ch

Word counts: 1514 words in the Materials and Methods
2516 words in Introduction, Results and Discussion

Abbreviations: CEECAM1, cerebral endothelial cell adhesion molecule-1; ColGalT, collagen galactosyltransferase; ER, endoplasmic reticulum; Hyl, hydroxylysine; GHyl, galactosylated hydroxylysine; GGHyl, glucosyl-galactosylated hydroxylysine; MBL, mannose-binding lectin

Keywords: extracellular matrix, glycosylation, hydroxylation, mass-spectrometry, endoplasmic reticulum

Data deposition: The nucleotide sequences reported in this paper correspond to the GenBank/EBI Data Bank entries with the accession numbers NM_024656, NM_015101 and NM_016174.

Abstract

Collagen is a trimer of three left-handed alpha-chains representing repeats of the motif Gly-X-Y, whereas (hydroxy)proline and (hydroxy)lysine residues are often found at positions X and Y. Selected hydroxylysine are further modified by the addition of galactose and glucose-galactose units. Collagen glycosylation takes place in the endoplasmic reticulum before triple helix formation and is mediated by β(1-O)galactosyl- and α(1-2)glucosyltransferase enzymes. We have identified two collagen galactosyltransferases using affinity chromatography and tandem-MS protein sequencing. The two collagen β(1-O)galactosyltransferases corresponded to the GLT25D1 and GLT25D2 proteins. Recombinant GLT25D1 and GLT25D2 enzymes showed a strong galactosyltransferase activity towards various types of collagen and towards the serum mannose-binding lectin MBL, which contains a collagen domain. Amino acid analysis of the products of GLT25D1 and GLT25D2 reactions confirmed the transfer of galactose to hydroxylysine residues. The GLT25D1 gene is constitutively expressed in human tissues, whereas the GLT25D2 gene is only expressed at low levels in the nervous system. The GLT25D1 and GLT25D2 enzymes are similar to CEECAM1, to which we could not attribute any collagen galactosyltransferase activity. The GLT25D1 and GLT25D2 genes now allow addressing the biological significance of collagen glycosylation and the importance of this post-translational modification in the etiology of connective tissue disorders.

Introduction

Collagens are the most abundant proteins in the human body. To date, 29 types of collagen have been described, which are encoded by at least 44 genes (21, 37, 45). Collagens are characterized by domains representing repeats of the triplet Gly-X-Y, where proline and lysine are often found at the positions X and Y. The Gly-X-Y repeats are not confined to collagens, but are also found in several proteins, such as the hormone adiponectin (29), the mannose-binding lectin (MBL) (11), the C1q complement protein (35), the COLQ subunit of the acetylcholine esterase complex (4) and the surfactant proteins SP-A and SP-D (11).

After synthesis in the endoplasmic reticulum (ER), three procollagen subunits associate to build a right-handed triple helix. However, before the formation of the triple helix structure, the nascent procollagen polypeptides undergo several post-translational modifications. These modifications comprise the hydroxylation of selected proline (20) and lysine (33) residues, which are catalyzed by three prolyl-4-hydroxylases (17), one prolyl-3-hydroxylase (46) and three lysyl hydroxylases (43). Hydroxylysine can be further modified by the addition of the monosaccharide Gal(β1-O) or the disaccharide Glc(α1-2)Gal(β1-O) (39).

Whereas the glycosylation of collagen has been first described by Grassmann and Schleich in 1935 (9) and the structure of the glycan determined by Spiro in 1967 as being Glc(α1-2)Gal(β1-O)Hyl (40), the molecular nature of the collagen glycosyltransferase enzymes has remained elusive up to now. Collagen galactosyltransferase (ColGalT) and glucosyltransferase activities have been characterized using partially purified proteins (24, 31, 32), which appeared to be

instable. Recently, the lysyl hydroxylase-3 LH3 enzyme has been shown to catalyze a modest galactosyl- and glucosyltransferase activity, suggesting that this enzyme is a combined hydroxylase and glycosyltransferase (12).

Prolyl and lysyl hydroxylation contribute to the stability of the collagen triple helix whereas hydroxylysine is essential for the cross-linking of collagen molecules, thus ensuring the strength of collagen fibrils (28). By contrast, the biological significance of collagen glycosylation is still unclear. The collagen domain of adiponectin and mannose-binding lectin also carry glycosylated hydroxylysine residues, which appear important for the oligomerization and proper secretion of these proteins (6, 29).

The importance of collagen post-translational modifications is reflected by the diseases caused by defective collagen modifying enzymes. Mutations of the *LH1* lysyl hydroxylase-1 gene lead to the connective tissue disorder Ehlers-Danlos syndrome type-VI (14) and mutations in the *LH2* lysyl hydroxylase-2 gene lead to the Bruck syndrome (44). The deficiency of the prolyl-3-hydroxylase-1 gene causes a severe form of osteogenesis imperfecta (5). The availability of the collagen glycosyltransferase genes will enable the comprehensive investigation of this post-translational modification in cellular and animal models and possibly in human diseases.

Materials and Methods

Affinity chromatography. Collagen glycosyltransferases were enriched by affinity chromatography as described by Myllyla *et al.* (1, 22, 24). Briefly, 10-day-old chicken embryos were homogenized in 225 mM mannitol, 75 mM sucrose, 50 µM DTT and 50 mM Tris-HCl, pH 7.4 at 4 °C and centrifuged at 15,000 x *g* for 40 min. Supernatants were filtered and proteins precipitated in 60 % $(NH_4)_2SO_4$. The pellets obtained after 20 min centrifugation at 15,000 x *g* were dissolved in 0.2 M NaCl, 50 µM DTT, 1 % glycerol, 20 mM Tris-HCl, pH 7.4 and dialyzed overnight against 2.5 l of enzyme buffer (0.15 M NaCl, 10 mM $MnCl_2$, 50 M DTT, 1 % glycerol, 50 mM Tris-HCl, pH 7.4). The chicken protein extracts were loaded on a column of agarose-bound bovine achilles collagen type-I fragments as described previously (32). The column was washed with 5 volumes of enzyme buffer containing 500 µM UDP followed by elution with 0.1 % acetic acid. Collected fractions were immediately neutralized with 1 M Tris pH 8.0.

MS peptide analysis. The eluted fractions from the affinity chromatography were desalted and concentrated with Amicon Ultra 10 cartridges (Millipore). Portions of 2 µg protein were reduced in 0.6 M Tris pH 8.5, 50 mM DTT for 5 min at 80 °C and alkylated for 40 min at RT in the dark by the addition of iodoacetamide (Sigma-Aldrich, final concentration 200 mM) and desalted by adding 9 volumes of ice cold methanol for 18 h on ice. Alkylated proteins were digested for 18 h at 37 °C with 0.01 µg trypsin (Roche). ZipTip (Millipore) purified peptides were then analyzed by liquid chromatography-MS. The desalted peptide digest was adjusted to 0.2 % formic

acid, 3 % ACN and directly injected onto a custom packed 80 mm x 0.075 mm ProntoSil-Pur C18- AQ, 3 µm, 200 Å, column (Bischoff GmbH, Leonberg, Germany), connected to a LTQ-ICR-FT mass spectrometer (Thermo Scientific, Bremen, Germany). The peptides were eluted with a binary gradient of solvents A (3 % ACN, 0.2 % formic acid) and B (80 % ACN, 0.2 % formic acid) using an Eksigent-Nano-HPLC system (Eksigent technologies, Dublin, USA). The column was flushed for 16 min at a flow rate of 500 nl/min with 100 % buffer A. Buffer B was increased to 3 % over 5 min, to 60 % over 50 min, to 100 % over 3 min and held at 100 % for 7 min. During gradient elution, the flow rate was maintained at 200 nl/min. The mass spectral data were acquired in the mass range of 300-2000 m/z. Data dependent MS/MS spectra were recorded of up to four of the most intense ions with higher charge state than 1+ using collision-induced dissociation (CID). Target ions already selected for MS/MS were dynamically excluded for 60 s. Peptide signals exceeding 500 counts were subjected to CID with normalized collision energy of 32 %. MS and MS/MS data were searched using Mascot Server 2.1 (Matrix Science, London, UK) as the search engine. Modifications used include carbamidomethylation (Cys, fixed) and oxidation (Met, variable). The monoisotopic masses of +2, and +3 charged peptides were searched with a peptide tolerance of 2 ppm and MS/MS tolerance of 0.8 Da. MS/MS spectra were searched against the UniRef100 20051018 database (2764545 sequences; 1015909965 residues) downloaded from the European Bioinformatics Institute (http://www.ebi.ac.uk/uniprot/database/download.html) and the *Gallus gallus* predicted proteins database from the Ensemble Genome Browser

ftp://ftp.ensembl.org/pub/current_chicken/data/fasta/pep/Gallus_gallus.WASHUC1.jul.pep.abinitio.fa.gz,August 2005).

Cloning and protein expression. The *GLT25D2*, *LH3* and *MBL* cDNAs were purchased from the RZPD repository (Berlin, Germany). The *GLT25D1* and *cerebral endothelial cell adhesion molecule-1 (CEECAM1)* cDNAs were cloned by RT-PCR from human fibroblast total RNA using the primers 5'-ATCTGAATTCCCTTTAAGGCGCGGCCAGAGTC-3', 5'-ATGTCTAGATGGAGCCTGGGCCACCGATG-3' for *GLT25D1*, and 5'-CGTAGAATTCGAGAGCTCCGGGGGCCGCT3', 5'-GACTATCTAGAGTAGTGGCCTGCTCCTGGAC-3' (Microsynth, Switzerland) for *CEECAM1.* The RT-PCR products were subcloned as *EcoR*I-*Xba*I fragments into the pFastBacI baculovirus transfer vector (Invitrogen). The *MBL* cDNA was subcloned into the *EcoR*I site of the pFmel-protA vector (48) to yield a protein-A fusion protein. The corresponding 732 bp *MBL* fragment was amplified with the primers 5'-ATCGAATTCATGGTGGCAGCGTCTTACTC-3' and 5'-ATCGAATTCAGGAGGGCCTGAGTGATATG-3'. Recombinant baculoviruses were produced in *S. frugiperda* Sf9 cells as described previously (13). Protein-A tagged MBL was co-expressed together with LH3, purified from the supernatant of infected Sf9 cells by IgG Sepharose chromatography (48) and subsequently used as acceptor for the enzymatic activity assay. The expression of the recombinantly expressed enzymes was analyzed on a 10 % SDS-PAGE gel. Prior to electrophoresis, proteins were enriched by concanavalin-A Sepharose (GE Healthcare) chromatography.

Protein bands were excised from the SDS-PAGE gel, digested in gel with trypsin according to Shevchenko *et al* (34) and identified by MS peptide analysis.

Preparation of ColGalT acceptors. Bovine Achilles collagen type-I, bovine nasal septum collagen type-II, human placenta collagen type-III, -IV and -V (Sigma) were deglycosylated by trifluoromethane sulfonic acid (TFMS)-mediated cleavage (7, 38). Acceptor proteins (50 µg) were lyophilized followed by an incubation in a dry ice/ethanol bath for 20 min. Proteins were dissolved in 50 µl TFMS/toluene (16.6:1, v:v) (Sigma-Aldrich). Reactions were subsequently incubated at -20 °C for 24 h, then neutralized with 150 µl pyridine/H_2O (2:1, v:v) in the dry ice/ethanol bath followed by 15 min incubation on ice. The sample was mixed with 400 µl 50 mM ammonium acetate and dialyzed overnight against 2.5 l of 50 mM ammonium acetate.

Collagen glycosyltransferase assays. Baculovirus infected Sf9 cells were lyzed in 1 % Triton-X100/TBS pH 7.4 for 10 min on ice and postnuclear supernatant was used as enzyme source. Collagen was heat denatured for 10 min at 60 °C in sodium acetate pH 6.8 and rapidly cooled down to 0 °C before use. Assays were performed with 10 µl of Sf9 postnuclear supernatant in a final volume of 100 µl containing 0.5 mg/ml collagen acceptors, 60 µM UDP-Gal or UDP-Glc, 50,000 cpm UDP-[^{14}C]Gal or UDP-[^{14}C]Glc (GE Healthcare), 10 mM $MnCl_2$, 20 mM NaCl, 50 mM MOPS pH 7.4, 1 mM DTT. Reactions were incubated for 3 h at 37 °C and stopped by the addition of 500 µl of ice cold 5 % TCA / 5 % phosphotungstic acid. Enzymatic activity assays for Km

analysis were performed as described above but with varying amounts of either collagen type-I or UDP-Gal as substrates.

Amino acid analysis. The reaction products of the collagen galactosyltransferase assays were hydrolyzed in 4 M NaOH for 72 h at 105 °C and the resulting single amino acids were derivatized with Fmoc according to Bank *et al.* (2). RP-HPLC (LaChrom Hitachi, Merck) of single amino acids was performed on a ODS Hypersil column, 150 x 3 mm, 3 µm particle size (Thermo Electron Corporation) at 40 °C. The galactosylated Hyl (GHyl) and galactosyl-glucosylated Hyl (GGHyl) standards were kindly provided by Ruggero Tenni (University of Pavia) (42). Amino acids were separated at a flow rate of 0.2 ml/min using a gradient elution (Table S1) with the solvents 0.5 M citric acid, 5 mM $(CH_3)_4NCl$, pH 2.85 (A); 80 % of 20 mM sodium acetate trihydrate, 5 mM $(CH_3)_4NCl$, pH 4.5, 20 % of methanol (B); 100 % of ACN (C). Radiolabeled [^{3}H]Val and [^{14}C]Tyr (Moravek Biochemicals and Radiochemicals, USA) were used as internal standard. Radioactivity was counted in a β-counter (Tri-Carb 2900TR, Packard). For β-galactosidase digestion of GHyl, hydrolyzed amino acids were loaded on AG 50W-X8(H^+) (Bio-Rad), washed with 0.8 % acetic acid and eluted with 5 % ammonia. After removal of ammonia by lyophilization, the samples were digested with 10 mU of bovine testis β-galactosidase (QA-Bio, San Mateo, CA) in 100 mM sodium citrate, pH 4.3 for 16 h at 37 °C. Liberated Gal was separated from GHyl by passage through AG 50W-X8(H^+), whereas G-Hyl was released by elution with 5 % ammonia.

RNA interference. Lentivirus particles expressing the short hairpin RNA constructs TRCN0000034884, TRCN0000034885, TRCN0000034887 and TRCN0000034888 targeting human *GLT25D1* (MISSION shRNA NM_024656, Sigma) were produced in HEK293T cells as described previously (10). Aliquots of 500 µl of lentivirus-producing HEK293T cell supernatants were added to 60'000 HeLa cells for 24 h. Cells expressing the short hairpin RNA constructs were selected by 2.5 µg/ml of puromycin for 10 days. Silencing of the *GLT25D1* gene was monitored by quantitative RT-PCR (SYBR Green JumpStart Taq ReadyMix , Sigma) using the primers 5'-ATTGCGCGCCCACAGCAC-3' and 5'-GGTGGGAGCCGAGATGAAGC-3'. The expression of the GLT25D2 and of the GAPDH genes in HeLa cells was determined with the primers 5'-GATAACATTGACCAGGCTCAG-3', 5'-CCCAAAAGGATTGGCTCCAAC-3', 5'-ATGCTGGCGCTGAGTACGTCGTG-3' and 5'-GTGATGGCATGGACTGTGGTCAT-3', respectively.

Northern blotting. The *GLT25D1*, *GLT25D2* and *CEECAM1* cDNA probes were synthesized by PCR using the primers 5'-GATGAGGCCGAGAGCTTCATGC-3' and 5'-GCATGAAGCTCTCGGCCTCATC-3', 5'-AAGCAGGCATCCAGATGTACC-3' and 5'-TCCAGCTGAGCCTGGTCAATG-3', 5'-GTGGATGGCTGGATGCTCAAC-3' and 5'-GACTATCTAGAGTAGTGGCCTGCTCCTGGAC-3', respectively. The resulting 676 bp-long *GLT25D1*, 559 bp-long *GLT25D2* and 785 bp-long *CEECAM1* probes were labeled with α[^{32}P]dCTP (Hartman Analytic, Germany) by random priming (Stratagene). Multiple human tissue RNA blots (MTE array 3, BD Bioscience and First Choice Northern Human Blot 1, Ambion) were prehybridized with QuikHyb®

hybridization solution (Stratagene) containing 100 µg/ml ultra pure herring sperm DNA (Invitrogen) for 1 h at 65 °C, then hybridized with 5 x 10^5 cpm of each labeled probe overnight at 65 °C. The arrays were washed in 0.1 x SSC, 0.1 % SDS up to 60 °C and exposed on BioMax XAR film (Kodak) for 24 h at -80 °C.

Results

ColGalT identification. Most glycosylation pathways have been characterized at the molecular level over the past decades. However, the genes encoding the glycosyltransferases involved in the glycosylation of Hyl in collagen have remained unknown up to now. Here, we have applied a cloning strategy based on the enrichment of proteins by affinity chromatography, peptide sequencing and heterologous expression of isolated candidate proteins. The enrichment procedure followed the method of Myllyla *et al.* (32) using immobilized denatured collagen type-I to capture the ColGalT. Homogenates of 10-day-old chicken embryos were used as source of ColGalT enzyme as applied previously (1, 24). To identify potential ColGalT enzymes among the proteins enriched by the affinity chromatography, we selected proteins sharing sequence homology with known glycosyltransferases. We also narrowed down on proteins containing ER-localization motifs considering the cellular localization of collagen glycosylation and on proteins containing N-glycosylation sites since it was shown that the ColGalT activity could by enriched by concanavalin-A lectin chromatography (30).

One of the candidate proteins identified by tandem-MS peptide sequencing was the putative glycosyltransferase GLT25D2 (Fig. 1A, Fig. S1). GLT25D2 is a type-II transmembrane protein of 626 amino acids including four N-glycosylation sites and the ER-retention signal RDEL at the C-terminus (Fig. 1B). No enzymatic activity was attributed to GLT25D2 but database annotations pointed to sequence homology with bacterial enzymes involved in LPS biosynthesis. Similar proteins to the chicken GLT25D2 could be deduced from all metazoan genomes. In the human genome,

GLT25D2 was found to be strongly similar to two proteins, namely GLT25D1 and CEECAM1. The three proteins contained N-glycosylation sites and the ER retrieval signal RDEL at the C-terminus and shared above 50 % sequence identity (Fig. 2).

ColGalT activity. The putative ColGalT activity of GLT25D1, GLT25D2 and CEECAM1 was assayed by expressing the three proteins as recombinant baculovirus in Sf9 insect cells. Five types of collagen were tested as possible acceptor substrates. Because native collagen is readily glycosylated to varying extents, we also included deglycosylated collagen preparations in the assays. GLT25D1 and GLT25D2 showed a strong ColGalT activity on all deglycosylated collagen acceptors tested, whereas CEECAM1 did not show any activity (Table 1). As expected, the ColGalT activity of GLT25D1 and GLT25D2 was lower when using native collagen acceptors. Noteworthy, collagen type-IV and collagen type-V in the native form were hardly galactosylated by GLT25D1 and GLT25D2, suggesting that most Hyl residues were already glycosylated. In addition to true collagens, GLT25D1 and GLT25D2, but not CEECAM1, were able to transfer Gal to the serum protein MBL (Fig. 3), which contains four Hyl sites in its collagen domain, showing that ColGalT activity was not limited to large collagen acceptors. The nucleotide-sugars UDP-GlcNAc and UDP-GalNAc were also tested as possible donor substrates but no GlcNAc- or GalNAc-transfer to collagen could be detected (data not shown).

We also tested the ColGalT activity of the human LH3 enzyme, which had been reported to catalyze three reactions on collagen, namely the hydroxylation of Lys, plus the β(1-O)galactosylation and α(1-2)glucosylation of Hyl (12, 47). Surprisingly, we could not detect any significant ColGalT activity for LH3 under our assay

conditions using bovine Achilles collagen type-I as acceptor (Fig. 4A). However, as described previously (12, 47), we did measure a low collagen glucosyltransferase activity for LH3, whereas GLT25D1, GLT25D2 and CEECAM1 failed to show any significant collagen glucosyltransferase activity (Fig. 4B). Although no collagen glycosyltransferase activity could be attributed to CEECAM1, we did confirm that the recombinant protein was expressed in Sf9 cells like GLT25D1, GLT25D2 and LH3, as shown by SDS-PAGE (Fig. 4C). To confirm the identity of the GLT25D1, GLT25D2, CEECAM1 and LH3 proteins, the corresponding bands were excised from the gel, digested with trypsin and submitted to tandem-MS peptide sequencing (data not shown). However, even if shown to be expressed, it is still possible that the levels of CEECAM1 could be too low to detect activity.

Km values. The apparent Km of GLT25D1 and GLT25D2 was determined for the bovine collagen type-I acceptor and for UDP-Gal as donor substrate, since these values had been reported previously for the semi-purified ColGalT activity (23). The Km of GLT25D1 and GLT25D2 for the collagen type-I acceptor was 13.6 g/l and 9.8 g/l, respectively (Fig. 5A, C), whereas Myllyla *et al.* reported a Km of 150 g/l for the partially purified chicken ColGalT enzyme. The Km values for UDP-Gal were 18.77 µM for GLT25D1 and 33.53 µM for GLT25D2 (Fig. 5B, D). These values are comparable to those reported by Myllyla *et al*, who determined a Km for UDP-Gal of 30 µM for the partially purified chicken ColGalT (23).

ColGalT reaction products. The products of the GLT25D1 and GLT25D2 mediated ColGalT reactions were further analyzed to confirm the transfer of β-linked Gal to Hyl residues on collagen. The reaction products were hydrolyzed in 4 M NaOH to

yield single amino acids. After derivatization with Fmoc and separation by reverse phase HPLC, the amino acid profiles obtained from collagen type-I, collagen type-II and from the GLT25D1-, GLT25D2-reacted collagen acceptors were compared to a profile of authentic amino acid standards. The amount of GHyl and GGHyl was higher in collagen type-II than in collagen type-I as measured by the ratio GHyl/Hyl and GGHyl/Hyl, respectively (Fig. 6). This finding was in agreement with the values reported in the literature (16). The analysis of additional types of collagen, such as collagen type-IV and sponge collagen confirmed the variable extent of Hyl glycosylation across collagens (data not shown). The amino acid profiles obtained after GLT25D1 and GLT25D2 reactions in the presence of UDP-[^{14}C]Gal were further analyzed by β-counting (Fig. 6, lower panel). The [^{14}C]Gal signal co-migrated with the GHyl standard, indicating that GLT25D1 and GLT25D2 indeed represent true ColGalT enzymes. To demonstrate the β-linkage between Gal and Hyl, the GLT25D1 and GLT25D2 reaction products were digested with β-galactosidase. [^{14}C]Gal-Hyl was prepared by alkaline hydrolysis of GLT25D1 and GLT25D2 reaction products and samples of 25,000 cpm were incubated overnight with bovine testes β-galactosidase. More than 80% of the radioactivity was recovered as free [^{14}C]Gal after β-galactosidase digestion, which confirmed the β-configuration of the linkage catalyzed by the GLT25D1 and GLT25D2 enzymes (Table 2).

ColGalT gene silencing. To correlate the expression of the GLT25D1 enzyme with the endogenous ColGalT activity in human cells, we have silenced the *GLT25D1* gene in HeLa cells by RNA interference. HeLa cells only expressed the *GLT25D1* gene, whereas *GLT25D2* expression remained undetected by RT-PCR (Fig. 7A). The

GLT25D1 gene was silenced by introducing short hairpin RNA constructs into HeLa cells. The *GLT25D1* expression was reduced to 31%, 67% and 39% of normal mRNA levels when expressing combinations of short hairpin RNA probes (Fig. 7B). The endogenous ColGalT activity of HeLa cells was measured and found to reach 5.7 pmol·min^{-1}mg prot^{-1} in wildtype cells whereas it was decreased to 2.0, 3.7 and 2.1 pmol·min^{-1}mg prot^{-1} in *GLT25D1*-knockdowned cells. When normalizing the ColGalT activity of wildtype HeLa cells to 100%, this activity was decreased to 35%, 65% and 37% in *GLT25D1*-knockdowned cells (Fig. 7C). The direct comparison between the level of *GLT25D1* expression and ColGalT activity in HeLa cells showed a strong correlation (Fig. 7D), thereby establishing a link between the GLT25D1 enzyme and the ColGalT activity *in vivo*.

ColGalT gene expression. The expression of *GLT25D1*, *GLT25D2* and *CEECAM1* in human tissues was analyzed by Northern blotting on 10 human tissues. A single *GLT25D1* transcript of 3.7 kb was detected in all tissues, whereas mRNA levels were highest in placenta, heart, lung and spleen (Fig. 8A). By contrast, the *GLT25D2* gene yielded a 5.1 kb mRNA that was detected only in brain and skeletal muscle at lower levels (Fig. 8A). A 2.3 kb *CEECAM1* transcript was detected in most tissues investigated, while a second transcript of 5.9 kb was also detected in brain (Fig. 8A). The survey of *GLT25D1*, *GLT25D2* and *CEECAM1* gene expression was extended to 75 human tissues and cell types using a Northern dot blotting array. The broad expression of the *GLT25D1* gene was confirmed in fetal and adult human tissues (Fig. 8B, Fig. S2). Similarly, the narrow range of *GLT25D2* gene expression was also evident throughout the RNA array (Fig. 8B, Fig. S2). The *CEECAM1* gene, which

could not be related to a ColGalT activity, showed a widespread expression across tissues. Notably, *CEECAM1* was highly expressed in secretory tissues like salivary glands, pancreas and placenta and in the nervous system (Fig. 8B, Fig. S2). Additionally, the *GLT25D1* and *CEECAM1* genes, but not *GLT25D2*, were also expressed in various carcinoma cell lines (Fig. S2). This survey of ColGalT gene expression suggested that GLT25D1 represents the main source of ColGalT activity in human tissues, while GLT25D2 appears to be specialized to few cell types and possibly to few collagen acceptors.

Discussion

The identification of the GLT25D1 and GLT25D2 ColGalT enzymes and of the similar, yet inactive, relative CEECAM1 protein raises the question of a possible restricted specificity towards collagen acceptor substrates. Previous work has shown that ColGalTs recognizes collagen peptides of at least 500-600 Da and that Hyl alone is not a suitable acceptor (23, 31). The different activity levels of recombinant GLT25D1 and GLT25D2 enzymes towards the various types of collagen tested support the idea of a differential substrate recognition. The nature of the substrate recognition may be complex and may include glycosylated residues as part of the motifs recognized. In fact, GLT25D2 was more active towards native collagen type-I and type-II than towards the deglycosylated forms of these proteins (Table 1), which indicates that glycan chains somehow affected the recognition of the acceptor substrate by GLT25D2. Along the same line, the apparent inactivity of CEECAM1 may be due to stringent structural requirements regarding the recognition of collagen peptides. A similarly complex mechanism of substrate recognition has been described for core glycosyltransferases acting on mucin proteins. Some members of the polypeptide N-acetylgalactosaminyltransferase family recognize peptide acceptors including serine or threonine residues near residues that were previously glycosylated by other N-acetylgalactosaminyltransferases (3). ColGalT assays with synthetic peptides including Hyl and GHyl at various positions will certainly answer the question of the acceptor substrate recognition.

Alternatively, it is possible that CEECAM1 represents a ColGalT acting on a limited set of substrates. The screening of additional proteins including collagen domains

like adiponectin, the acetylcholine esterase complex COLQ, and the complement protein C1q might confirm this eventuality. Finally, CEECAM1 may have lost any enzymatic activity over the course of evolution. In fact, CEECAM1 was first described as an adhesion protein (41), which might function as a carbohydrate-binding protein at the cell surface. However, the presence of the C-terminal RDEL motif would suggest that CEECAM1 is maintained in the ER.

The lysyl hydroxylase LH3 protein has been previously reported to possess three enzymatic activities, namely a lysyl hydroxylase, a ColGalT and a collagen glucosyltransferase activity (47). The glycosyltransferase activities attributed to LH3 were very low, casting doubt on their biological significance (27). The ColGalT activity of LH3 reported previously reached approximately twice the levels of endogenous ColGalT activity measured in Sf9 cells (47). It is possible that we could not distinguish the ColGalT activity of LH3 from the background activity levels in our assays. By comparison, the strong ColGalT activity described here for GLT25D1 and GLT25D2 implies that these proteins indeed represent true ColGalT enzymes. The dual glycosyltransferase activity of LH3 certainly requires closer attention as it is expected that the catalysis of both β(1-O) and α(1-2) linkages would require distinct domains responsible for the retaining α(1-2) and inverting β(1-O) glycosyltransferase activities (26).

The identification of the *GLT25D1* and *GLT25D2* genes as encoding two ColGalT enzymes opens new ways to investigate the biological significance of collagen glycosylation. Genes similar to the human *GLT25D1* and *GLT25D2* genes can be found in all metazoan genomes sequenced to date. For example, the *C. elegans* gene

D2045.9 represents the probable ortholog to the human *GLT25D1* and *GLT25D2* genes. Knockdown of the *D2045.9* gene by RNA interference yields multiple abnormalities like deformed mating organs, slow growth and uncoordinated locomotion (15, 36). By comparison, the loss of the lysyl hydroxylase gene *let-268* leads to a lethal phenotype associated to a defect in collagen type-IV secretion (21, 25). The conservation of collagen glycosylation throughout animals and the essential role of collagen glycosylation in worms emphasize the importance of this modification. In humans, the *GLT25D1* and *GLT25D2* genes are found on human chromosome 19p13 and chromosome 1q25, respectively. The involvement of ColGalTs in the pathogenesis of connective tissue disorders linked to chromosomes 19p13 and 1q25, such as psoriasis (19) and epidermolysis bullosa (8, 18) can now be straightforwardly documented.

Acknowledgements

We thank Prof. Dr. Beat Steinmann (University Children's Hospital Zurich, Switzerland) for helpful discussions, Dr. B. Roschitzki (Functional Genomic Center Zurich, FGCZ, Switzerland) and the FGCZ for their support and Prof. Dr. R. Tenni (University of Pavia, Italy) for the G-Hyl and GG-Hyl amino acid standards. This work was supported by Telethon Action Suisse, the Wolfermann-Nägeli Foundation and by grants from the Swiss National Science Foundation (PP00A-106756 and 3100A0-116039).

References

1. **Anttinen, H.** 1976. Stimulation of collagen galactosyltransferase and glucosyltransferase activities by lysophosphatidylcholine. Biochem J **160:**29-35.
2. **Bank, R. A., B. Beekman, R. Tenni, and J. M. TeKoppele.** 1997. Pre-column derivatisation method for the measurement of glycosylated hydroxylysines of collagenous proteins. J Chromatogr B Biomed Sci Appl **703:**267-272.
3. **Bennett, E. P., H. Hassan, U. Mandel, E. Mirgorodskaya, P. Roepstorff, J. Burchell, J. Taylor-Papadimitriou, M. A. Hollingsworth, G. Merkx, A. G. van Kessel, H. Eiberg, R. Steffensen, and H. Clausen.** 1998. Cloning of a human UDP-N-acetyl-alpha-D-Galactosamine:polypeptide N-acetylgalactosaminyltransferase that complements other GalNAc-transferases in complete O-glycosylation of the MUC1 tandem repeat. J Biol Chem **273:**30472-30481.
4. **Bon, S., A. Ayon, J. Leroy, and J. Massoulie.** 2003. Trimerization domain of the collagen tail of acetylcholinesterase. Neurochem Res **28:**523-535.
5. **Cabral, W. A., W. Chang, A. M. Barnes, M. Weis, M. A. Scott, S. Leikin, E. Makareeva, N. V. Kuznetsova, K. N. Rosenbaum, C. J. Tifft, D. I. Bulas, C. Kozma, P. A. Smith, D. R. Eyre, and J. C. Marini.** 2007. Prolyl 3-hydroxylase 1 deficiency causes a recessive metabolic bone disorder resembling lethal/severe osteogenesis imperfecta. Nat Genet **39:**359-365.
6. **Colley, K. J., and J. U. Baenziger.** 1987. Post-translational modifications of the core-specific lectin. Relationship to assembly, ligand binding, and secretion. J Biol Chem **262:**10296-10303.

7. **Edge, A. S., C. R. Faltynek, L. Hof, L. E. Reichert, Jr., and P. Weber.** 1981. Deglycosylation of glycoproteins by trifluoromethanesulfonic acid. Anal Biochem **118:**131-137.

8. **Gay, S., J. D. Fine, and J. S. Storer.** 1988. Autoantibodies to extracellular collagen matrix components in epidermolysis bullosa and other bullous diseases. Arch Dermatol Res **280:**333-337.

9. **Grassmann W, S. H.** 1953. Uber den Kohlenhydratgehalt des Kollagens II. Biochem Z **277:**320-328.

10. **Haeuptle, M. A., F. M. Pujol, C. Neupert, B. Winchester, A. J. Kastaniotis, M. Aebi, and T. Hennet.** 2008. Human RFT1 deficiency leads to a disorder of N-linked glycosylation. Am J Hum Genet **82:**600-6.

11. **Hakansson, K., and K. B. Reid.** 2000. Collectin structure: a review. Protein Sci **9:**1607-1617.

12. **Heikkinen, J., M. Risteli, C. Wang, J. Latvala, M. Rossi, M. Valtavaara, and R. Myllyla.** 2000. Lysyl hydroxylase 3 is a multifunctional protein possessing collagen glucosyltransferase activity. J Biol Chem **275:**36158-36163.

13. **Hennet, T., A. Dinter, P. Kuhnert, T. S. Mattu, P. M. Rudd, and E. G. Berger.** 1998. Genomic cloning and expression of three murine UDP-galactose: beta-N-acetylglucosamine beta1,3-galactosyltransferase genes. J Biol Chem **273:**58-65.

14. **Ihme, A., L. Risteli, T. Krieg, J. Risteli, U. Feldmann, K. Kruse, and P. K. Muller.** 1983. Biochemical characterization of variants of the Ehlers-Danlos syndrome type VI. Eur J Clin Invest **13:**357-362.

15. **Kamath, R. S., A. G. Fraser, Y. Dong, G. Poulin, R. Durbin, M. Gotta, A. Kanapin, N. Le Bot, S. Moreno, M. Sohrmann, D. P. Welchman, P. Zipperlen, and J. Ahringer.** 2003. Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. Nature **421:**231-237.

16. **Kivirikko, K. I., and R. Myllyla.** 1979. Collagen glycosyltransferases. Int Rev Connect Tissue Res **8:**23-72.

17. **Kukkola, L., R. Hieta, K. I. Kivirikko, and J. Myllyharju.** 2003. Identification and characterization of a third human, rat, and mouse collagen prolyl 4-hydroxylase isoenzyme. J Biol Chem **278:**47685-47693.

18. **Lane, E. B., and W. H. McLean.** 2004. Keratins and skin disorders. J Pathol **204:**355-366.

19. **Lee, Y. A., F. Ruschendorf, C. Windemuth, M. Schmitt-Egenolf, A. Stadelmann, G. Nurnberg, M. Stander, T. F. Wienker, A. Reis, and H. Traupe.** 2000. Genomewide scan in german families reveals evidence for a novel psoriasis-susceptibility locus on chromosome 19p13. Am J Hum Genet **67:**1020-1024.

20. **Mizuno, K., T. Hayashi, and H. P. Bachinger.** 2003. Hydroxylation-induced stabilization of the collagen triple helix. Further characterization of peptides with 4(R)-hydroxyproline in the Xaa position. J Biol Chem **278:**32373-32379.

21. **Myllyharju, J., and K. I. Kivirikko.** 2004. Collagens, modifying enzymes and their mutations in humans, flies and worms. Trends Genet **20:**33-43.

22. **Myllyla, R., H. Anttinen, L. Risteli, and K. I. Kivirikko.** 1977. Isolation of collagen glucosyltransferase as a homogeneous protein from chick embryos. Biochim Biophys Acta **480:**113-121.

23. **Myllyla, R., L. Risteli, and K. I. Kivirikko.** 1975. Assay of collagen-galactosyltransferase and collagen-glucosyltransferase activities and preliminary characterization of enzymic reactions with transferases from chick-embryo cartilage. Eur J Biochem **52:**401-410.

24. **Myllyla, R., L. Risteli, and K. I. Kivirikko.** 1976. Collagen glucosyltransferase. Partial purification and characterization of the enzyme from whole chick embryos and chick-embryo cartilage. Eur J Biochem **61:**59-67.

25. **Norman, K. R., and D. G. Moerman.** 2000. The let-268 locus of Caenorhabditis elegans encodes a procollagen lysyl hydroxylase that is essential for type IV collagen secretion. Dev Biol **227:**690-705.

26. **Qasba, P. K., B. Ramakrishnan, and E. Boeggeman.** 2005. Substrate-induced conformational changes in glycosyltransferases. Trends Biochem Sci **30:**53.

27. **Rautavuoma, K., K. Takaluoma, K. Passoja, A. Pirskanen, A. P. Kvist, K. I. Kivirikko, and J. Myllyharju.** 2002. Characterization of three fragments that constitute the monomers of the human lysyl hydroxylase isoenzymes 1-3. The 30-kDa N-terminal fragment is not required for lysyl hydroxylase activity. J Biol Chem **277:**23084-23091.

28. **Reiser, K., R. J. McCormick, and R. B. Rucker.** 1992. Enzymatic and nonenzymatic cross-linking of collagen and elastin. Faseb J **6:**2439-2449.

29. **Richards, A. A., T. Stephens, H. K. Charlton, A. Jones, G. A. Macdonald, J. B. Prins, and J. P. Whitehead.** 2006. Adiponectin multimerization is dependent on conserved lysines in the collagenous domain: evidence for regulation of

multimerization by alterations in posttranslational modifications. Mol Endocrinol **20:**1673-1687.

30. **Risteli, L.** 1978. Further characterization of collagen galactosyltransferase from chick embryos. Biochem J **169:**189-196.

31. **Risteli, L., R. Myllya, and K. I. Kivirikko.** 1976. Partial purification and characterization of collagen galactosyltransferase from chick embryos. Biochem J **155:**145-153.

32. **Risteli, L., R. Myllyla, and K. I. Kivirikko.** 1976. Affinity chromatography of collagen glycosyltransferases on collagen linked to agarose. Eur J Biochem **67:**197-202.

33. **Risteli, M., O. Niemitalo, H. Lankinen, A. H. Juffer, and R. Myllyla.** 2004. Characterization of collagenous peptides bound to lysyl hydroxylase isoforms. J Biol Chem **279:**37535-37543.

34. **Shevchenko, A., H. Tomas, J. Havlis, J. V. Olsen, and M. Mann.** 2006. In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat Protoc **1:**2856-2860.

35. **Shinkai, H., and K. Yonemasu.** 1979. Hydroxylysine-linked glycosides of human complement subcomponent C1q and various collagens. Biochem J **177:**847-852.

36. **Simmer, F., C. Moorman, A. M. van der Linden, E. Kuijk, P. V. van den Berghe, R. S. Kamath, A. G. Fraser, J. Ahringer, and R. H. Plasterk.** 2003. Genome-wide RNAi of C. elegans using the hypersensitive rrf-3 strain reveals novel gene functions. PLoS Biol **1:**77-84.

37. **Soderhall, C., I. Marenholz, T. Kerscher, F. Ruschendorf, J. Esparza-Gordillo, M. Worm, C. Gruber, G. Mayr, M. Albrecht, K. Rohde, H. Schulz, U. Wahn, N. Hubner, and Y. A. Lee.** 2007. Variants in a novel epidermal collagen gene (COL29A1) are associated with atopic dermatitis. PLoS Biol **5:**19521961.

38. **Sojar, H. T., and O. P. Bahl.** 1987. Chemical deglycosylation of glycoproteins. Methods Enzymol **138:**341-350.

39. **Spiro, R. G.** 1969. Characterization and quantitative determination of the hydroxylysine-linked carbohydrate units of several collagens. J Biol Chem **244:**602-612.

40. **Spiro, R. G.** 1967. The structure of the disaccharide unit of the renal glomerular basement membrane. J Biol Chem **242:**4813-4823.

41. **Starzyk, R. M., C. Rosenow, J. Frye, M. Leismann, E. Rodzinski, S. Putney, and E. I. Tuomanen.** 2000. Cerebral cell adhesion molecule: a novel leukocyte adhesion determinant on blood-brain barrier capillary endothelium. J Infect Dis **181:**181-187.

42. **Tenni, R., D. Rimoldi, G. Zanaboni, G. Cetta, and A. A. Castellani.** 1984. Hydroxylysine glycosides: preparation and analysis by reverse phase high performance liquid chromatography. Ital J Biochem **33:**117-127.

43. **Valtavaara, M., C. Szpirer, J. Szpirer, and R. Myllyla.** 1998. Primary structure, tissue distribution, and chromosomal localization of a novel isoform of lysyl hydroxylase (lysyl hydroxylase 3). J Biol Chem **273:**12881-12886.

44. **van der Slot, A. J., A. M. Zuurmond, A. F. Bardoel, C. Wijmenga, H. E. Pruijs, D. O. Sillence, J. Brinckmann, D. J. Abraham, C. M. Black, N. Verzijl,**

J. DeGroot, R. Hanemaaijer, J. M. TeKoppele, T. W. Huizinga, and R. A. Bank. 2003. Identification of PLOD2 as telopeptide lysyl hydroxylase, an important enzyme in fibrosis. J Biol Chem **278:**40967-40972.

45. **Veit, G., B. Kobbe, D. R. Keene, M. Paulsson, M. Koch, and R. Wagener.** 2006. Collagen XXVIII, a novel von Willebrand factor A domain-containing protein with many imperfections in the collagenous domain. J Biol Chem **281:**3494-3504.

46. **Vranka, J. A., L. Y. Sakai, and H. P. Bachinger.** 2004. Prolyl 3-hydroxylase 1, enzyme characterization and identification of a novel family of enzymes. J Biol Chem **279:**23615-23621.

47. **Wang, C., H. Luosujarvi, J. Heikkinen, M. Risteli, L. Uitto, and R. Myllyla.** 2002. The third activity for lysyl hydroxylase 3: galactosylation of hydroxylysyl residues in collagens in vitro. Matrix Biol **21:**559-566.

48. **Zhou, D., M. Malissard, E. G. Berger, and T. Hennet.** 2000. Secretion and purification of recombinant beta1-4 galactosyltransferase from insect cells using pFmel-protA, a novel transposition-based baculovirus transfer vector. Arch Biochem Biophys **374:**3-7.

Table 1. ColGalT activity measured in Sf9 cell lysates. Values represent the average ± S.D. of 4 assays.

Collagen	TFMS [a]	Sf9 [b]	GLT25D1	GLT25D2	CEECAM1
			(pmol·min^{-1}·mg prot^{-1})		
no acceptor		1.5 ± 0.6	1.9 ± 0.7	0.9 ± 0.4	0.5 ± 0.1
Type I	-	10.5 ± 1.8	43.7 ± 12.0	57.2 ± 8.5	8.7 ± 2.0
	+	5.0 ± 3.0	61.4 ± 18.2	18.0 ± 5.4	2.7 ± 1.9
Type II	-	6.5 ± 0.8	36.6 ± 3.7	35.1 ± 3.8	3.3 ± 1.3
	+	3.4 ± 0.8	67.0 ± 16.5	22.1 ± 8.8	2.8 ± 0.8
Type III	-	2.9 ± 0.9	31.7 ± 5.8	25.2 ± 6.8	1.8 ± 0.8
	+	11.0 ± 3.6	244.7 ± 25.7	102.1 ± 17.6	6.5 ± 1.7
Type IV	-	1.7 ± 1.4	3.6 ± 2.2	3.6 ± 0.9	1.9 ± 0.7
	+	12.9 ± 4.6	433.2 ± 49.6	164.2 ± 43.3	8.4 ± 2.4
Type V	-	2.3 ± 0.7	2.7 ± 0.3	2.3 ± 0.4	0.7 ± 0.1
	+	2.7 ± 1.6	18.6 ± 3.6	4.7 ± 1.5	2.0 ± 1.1

[a] TFMS-mediated deglycosylation

[b] Sf9 cells infected with an empty baculovirus

Table 2. β-Galactosidase digestion of GLT25D1 and GLT25D2 reaction products. Values represent the average of 2 independent experiments.

		Input (cpm)	Products (cpm)	
Enzyme	**β-Galactosidase**	**G-Hyl**	**G-Hyl**	**Gal**
GLT25D1	+	25,000	3,780	19,376
GLT25D1	-	25,000	20,996	1,073
GLT25D2	+	25,000	3,528	20,362
GLT25D2	-	25,000	20,313	1,029

Figure Legends

Figure 1: ColGalT identification by mass spectrometry. Proteins isolated by affinity chromatography were analyzed by LC-MS. (A) Peptide fragment spectra of two peptides identifying GLT25D2, (B) protein sequence of the *Gallus gallus* GLT25D2. The two identifying peptides are shaded in grey, the four potential N-glycosylation sites are underlined and the ER retrieval signal is shown in bold.

Figure 2: Protein alignment. The three putative human ColGalT enzymes share a high degree of sequence identity (63 % between GLT25D1 and GLT25D2, 50% between GLT25D2 and CEECAM1, 55 % between GLT25D1 and CEECAM1). The proteins include the C-terminal RDEL ER retrieval motif. Black squares represent amino acids identical or similar in all 3 proteins; grey squares represent amino acids identical or similar in 2 of the proteins.

Figure 3: ColGalT activity towards MBL. MBL was produced in Sf9 cells co-infected with a baculovirus expressing LH3. ColGalT activity assay was performed as described in the *Experimental procedures* section. Bars indicate the means of four assays. Error bars indicate the SD.

Figure 4: Time course of baculovirus mediated protein expression in Sf9 cells. (A) ColGalT activity and (B) collagen glucosyltransferase activity measured in cells expressing GLT25D1 (▲), GLT25D2 (✶), CEECAM1 (□), LH3 (●). Bovine Achilles

collagen type-I was used as acceptor substrate. The activity measured in Sf9 infected with an empty baculovirus is shown in both panels with filled squares (■).Values indicate the means of four assays. Error bars indicate the SD. (C) SDS-PAGE of recombinantly expressed proteins. Arrows indicate the recombinant protein bands, as confirmed by LC-MS mediated protein sequencing.

Figure 5: Determination of the apparent Km of GLT25D1 and GLT25D2. (A) Lineweaver Burk blot for GLT25D1 on collagen, with the calculated Michaelis-Menten constant of 13.6 g/l. (C) Lineweaver Burk blot for GLT25D2 on collagen type-I with the calculated Michaelis-Menten constant of 9.8 g/l. (B) Lineweaver Burk blot for GLT25D1 on UDP-Gal with the calculated Michaelis-Menten constant of 18.77 µM . (D) Lineweaver Burk blot for GLT25D2 on UDP-Gal with the calculated Michaelis-Menten constant of 33.53 µM.

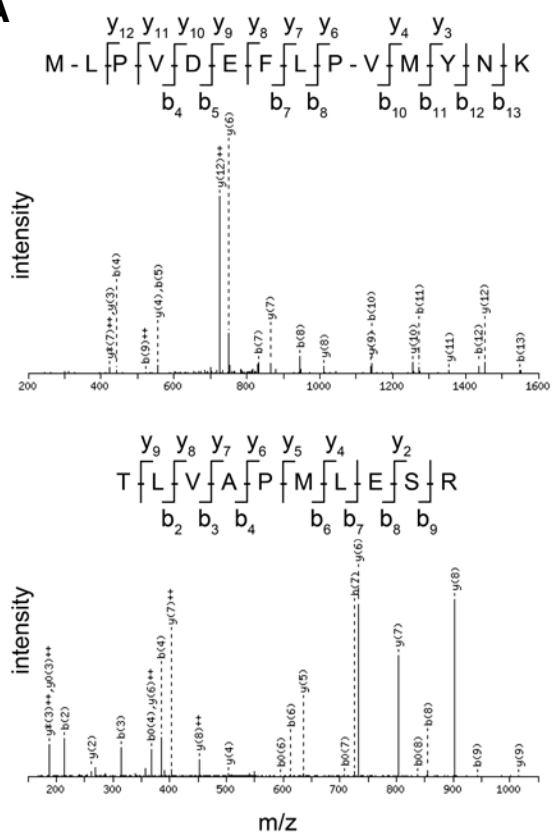
Figure 6: Product identification by RP-HPLC. The first panel represents an amino acid standard containing the standards for G-Hyl and GG-Hyl. The second and the third panels show the amino acid profiles of bovine collagen type-I and type-II hydrolysates, respectively. The lower two panels show the radioactive trace obtained after reaction of collagen type-I with GLT25D1 and GLT25D2. [^{3}H]Val and [^{14}C]Tyr were used as internal amino acid standards. Amino acids are marked in single-letter code. Hyp, hydroxyproline.

Figure 7: Silencing of the *GLT25D1* gene. (A) RT-PCR detection of *GLT25D1* and *GLT25D2* expression in HeLa cells. (B) mRNA *GLT25D1* levels in wildtype HeLa cells (black bar) and in *GLT25D1*-silenced HeLa cells (KD#1-3, white bars). (C) Relative ColGalT activity in wildtype HeLa cells (black bar) and in *GLT25D1*-silenced HeLa cells (white bars). (D) Comparison between *GLT25D1* mRNA levels and ColGalT activity in wildtype HeLa cells (set to 100%) and in *GLT25D1*-silenced HeLa cells.

Figure 8: Tissue Northern blotting. (A) The mRNA expression patterns of *GLT25D1*, *GLT25D2* and *CEECAM1* were analyzed in 10 human tissues and (B) in 36 human tissues and cell lines, whereas a representative collection of additional tissues and cell types is shown in Fig. S2. PBL, peripheral blood leukocytes.

Figure 1

A



B

```
XGSIADHWSEKKSMCISNGRKLILIASRTVSILTIVTHRA     1
SYPGASGWLPLEQMTWHQLPSAVQGCSAILQYTAVGEGPR    41
ILQPSEHGQEGKGSLLQLQQPPAFALFCTLETLIMKYLPL    81
FANGGGWVATDHNADNTTAILREWLKNVQNLYHDVEWRPM   121
EDPQSYPEEMGPKHWPSSRFTHVMKLRQAALRAAREKWSD   161
YVLFLDTDNLLTNPETLNLLIAENKTLVAPMLESRFLYSN   201
FWCGITPQAGGWGYYKRTLDYPLIREWKRTGCFAVPMIHS   241
TFLIDLRKEASTKLMFYPPHQDYTWSFDDIMVFAFSSRQA   281
GIQMFICNREHYGFLPMPLKSHQTLQEETENFVHTLIEAM   321
SKWLCPAVLEPPVSICRHVQLYQSSVYLQVMVGISALFQS   361
IVLPLSPLSMSLFLRSALTRWDLMKALNTSQLKALSIDML   401
PGYRDPYSSRPLTRGEIGCFLSHYYIWKEVVNRGLEKTLV   441
IEDDVRFEHQFKRKLMKLMDDIEQAQLDWELIYIGRKRMQ   481
VQQPEKAVPNVMNLVEADYSYWTLGYAISFQGAQKLIGAE   521
PFSKMLPVDEFLPVMYNKHPVAKYMEYYESRDLKAFSAEP   561
LLVYPTHYTGQPGYLSDTETSTIWDNETVSTDWDRTHSWK   601
SRQQGQIHSEAQNKDALPPQSSLNAPSSRDEL           641
```

Figure 2

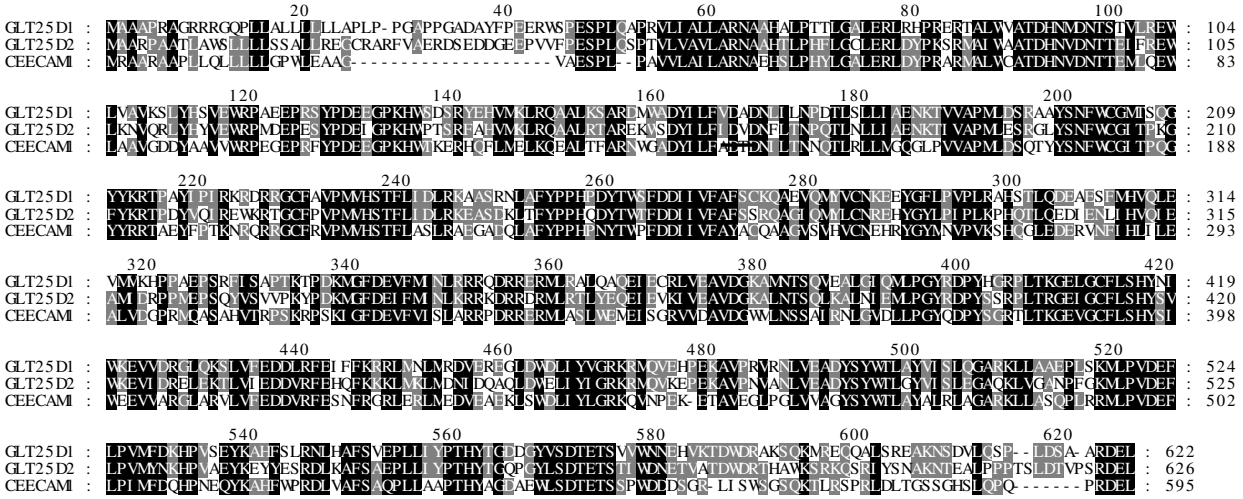


Figure 3

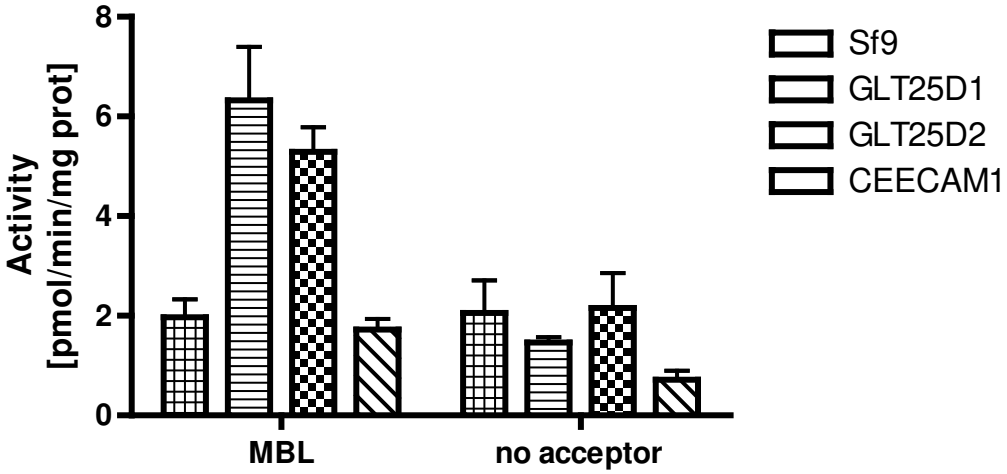

Activity
[pmol/min/mg prot]
8
6
4
2
0
MBL
no acceptor
Sf9
GLT25D1
GLT25D2
CEECAM1

Figure 4

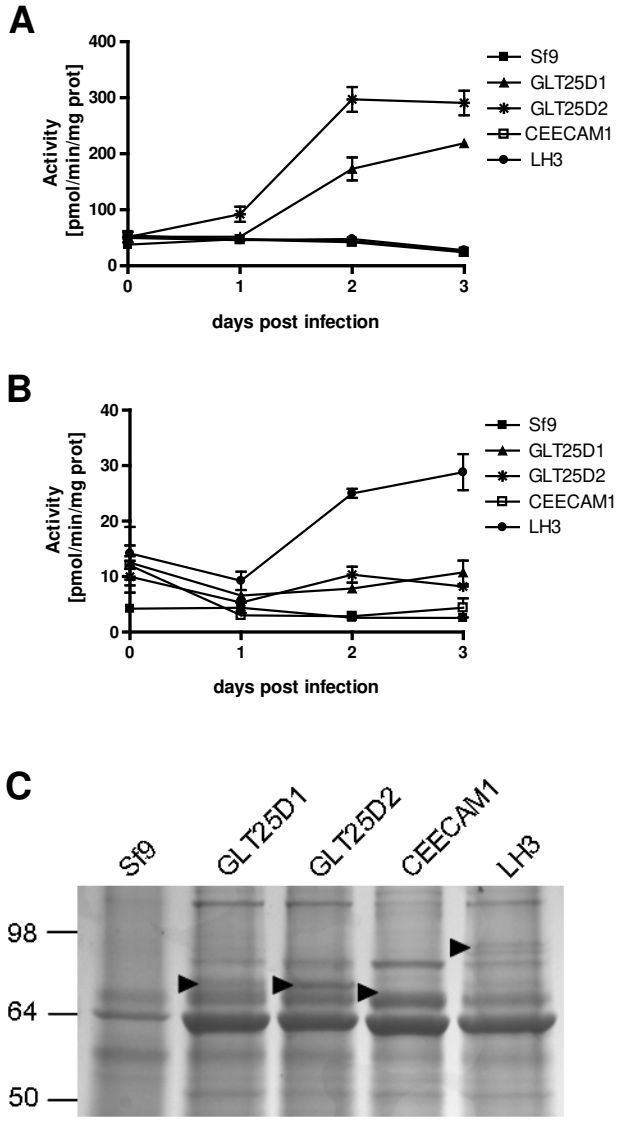

A
Activity
[pmol/min/mg prot]
400
300
200
100
0
0
1
2
3
days post infection
Sf9
GLT25D1
GLT25D2
CEECAM1
LH3
B
Activity
[pmol/min/mg prot]
40
30
20
10
0
0
1
2
3
days post infection
Sf9
GLT25D1
GLT25D2
CEECAM1
LH3
C
Sf9
GLT25D1
GLT25D2
CEECAM1
LH3
98
64
50

Figure 5

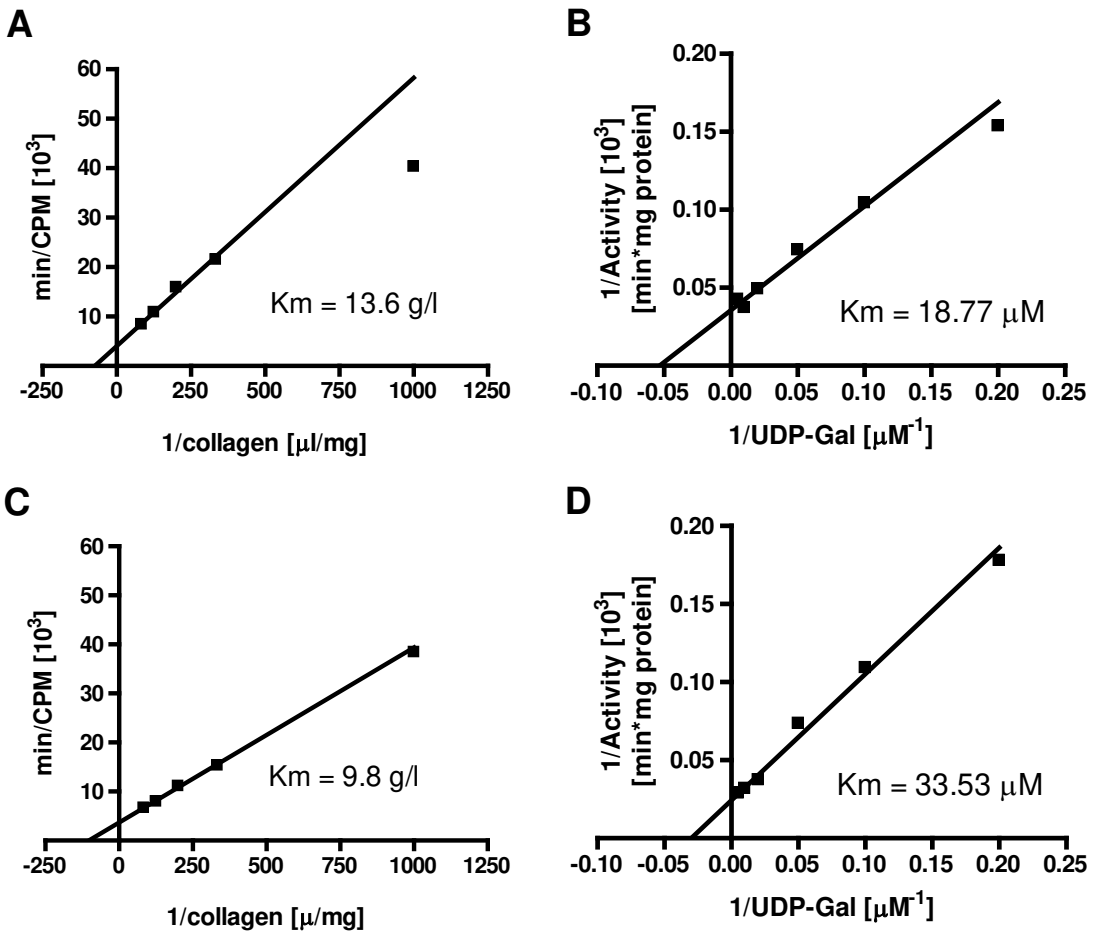

A
min/CPM [10^3]
60
50
40
30
20
10
Km = 13.6 g/l
-250
0
250
500
750
1000
1250
1/collagen [µl/mg]
B
1/Activity [10^3] [min*mg protein]
0.20
0.15
0.10
0.05
Km = 18.77 µM
-0.10
-0.05
0.00
0.05
0.10
0.15
0.20
0.25
1/UDP-Gal [µM^-1]
C
min/CPM [10^3]
60
50
40
30
20
10
Km = 9.8 g/l
-250
0
250
500
750
1000
1250
1/collagen [µ/mg]
D
1/Activity [10^3] [min*mg protein]
0.20
0.15
0.10
0.05
Km = 33.53 µM
-0.10
-0.05
0.00
0.05
0.10
0.15
0.20
0.25
1/UDP-Gal [µM^-1]

Figure 6

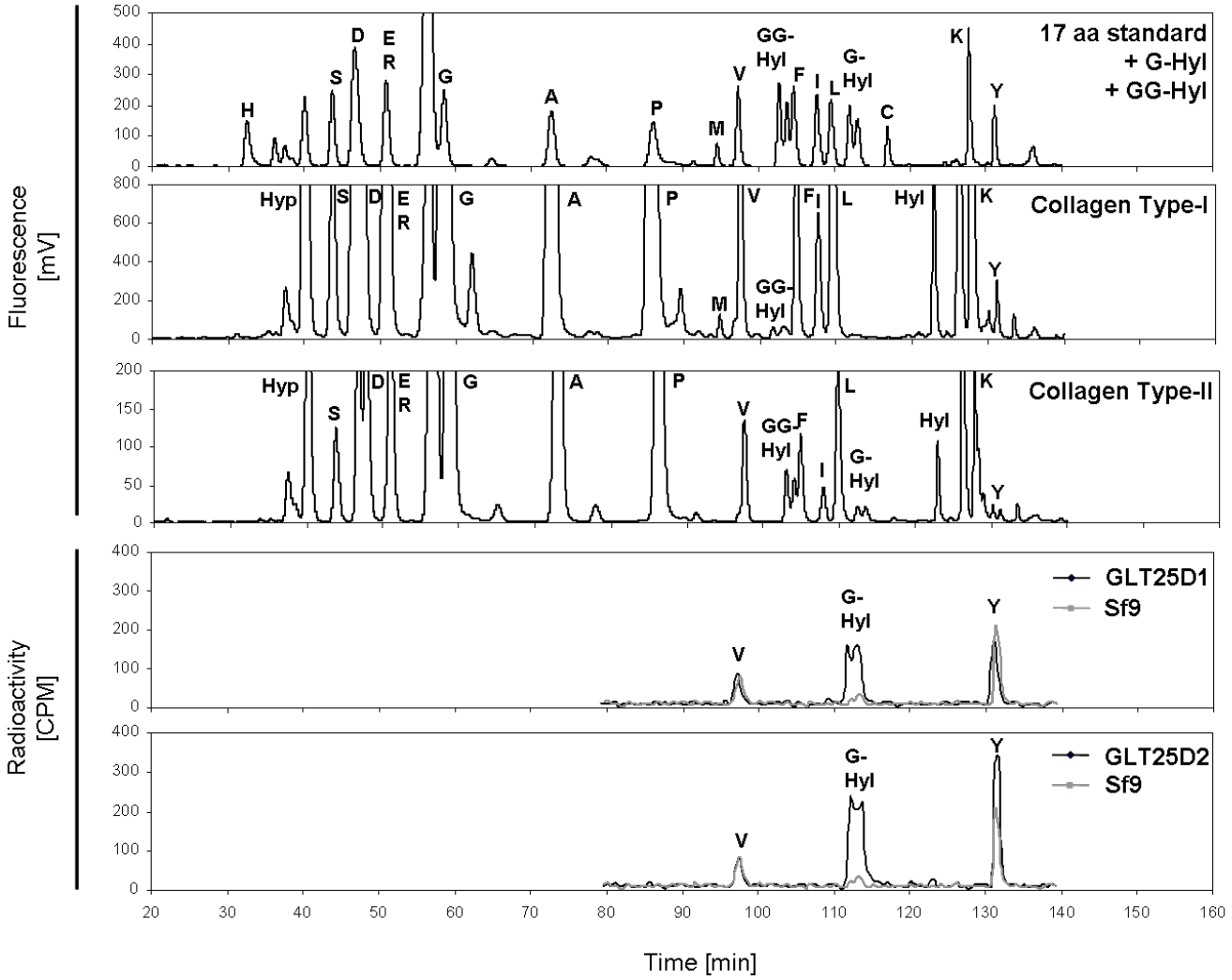

Fluorescence [mV]
17 aa standard + G-Hyl + GG-Hyl
Collagen Type-I
Collagen Type-II
Radioactivity [CPM]
GLT25D1
Sf9
GLT25D2
Sf9
Time [min]

Figure 7

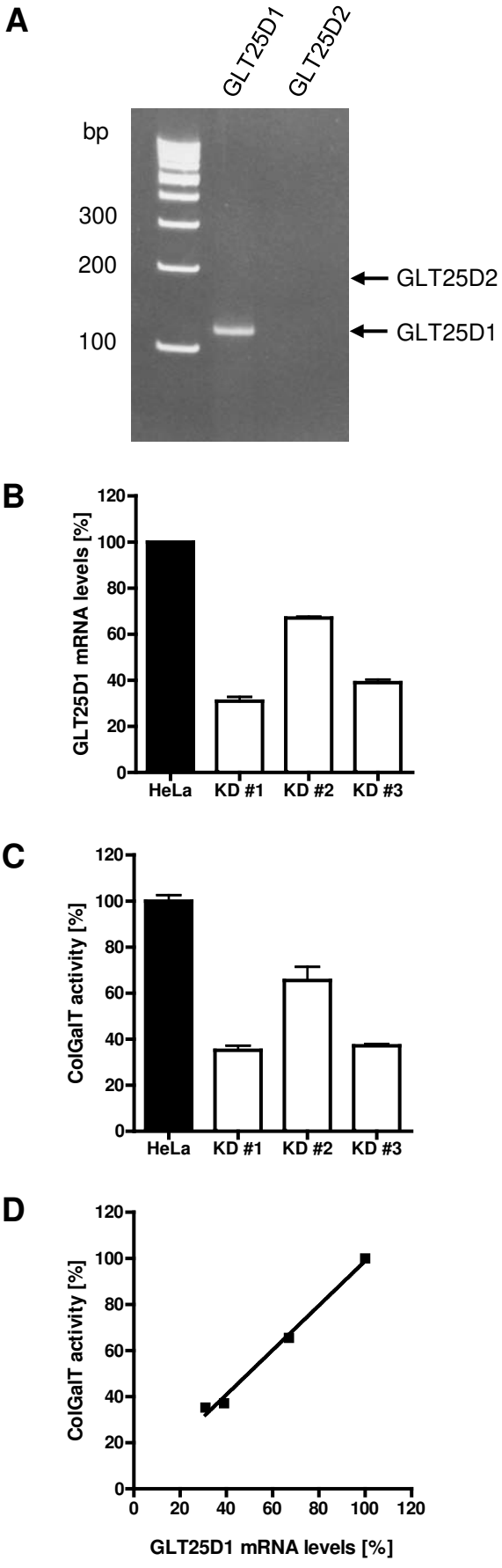
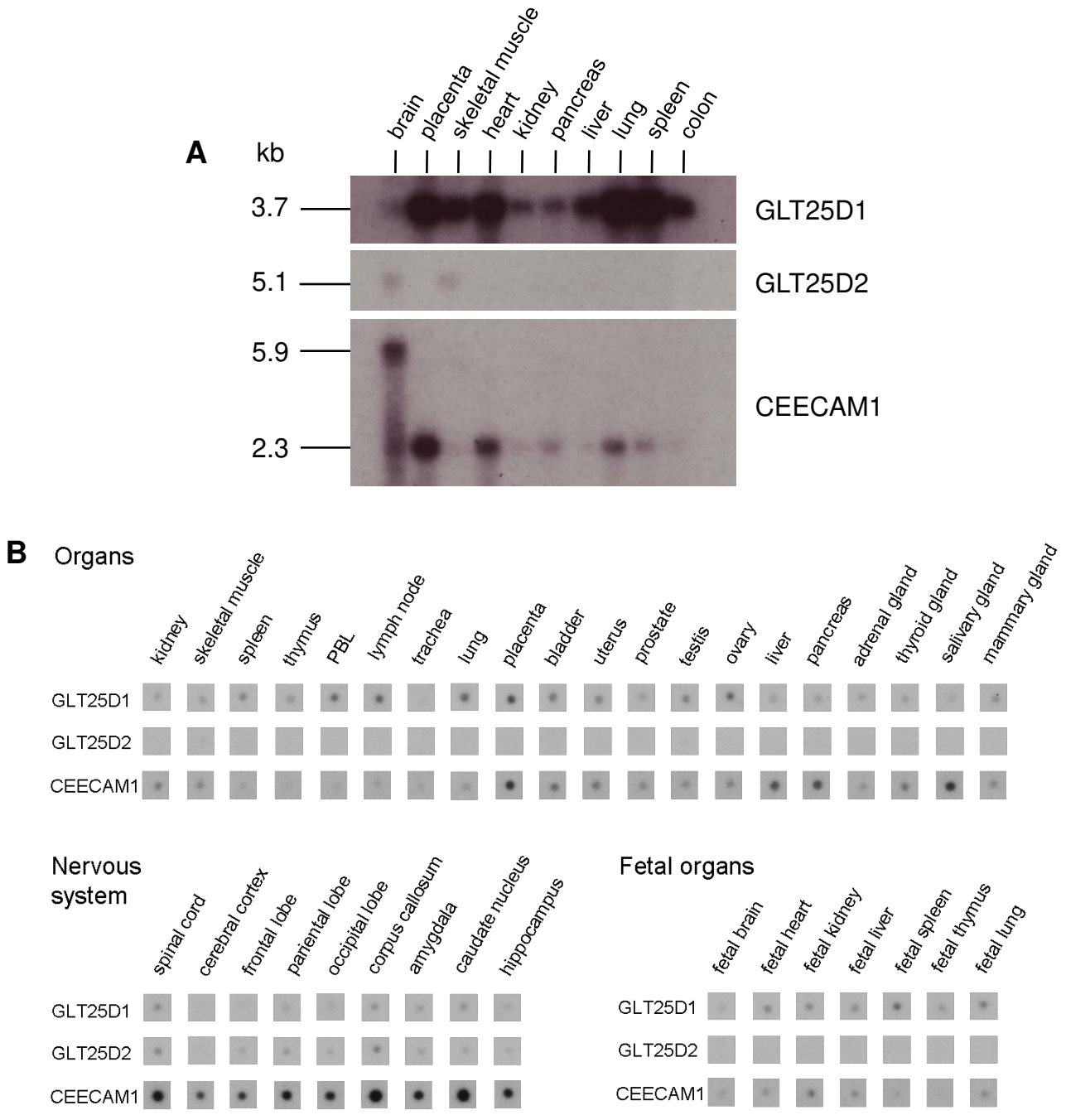

A
GLT25D1
GLT25D2
bp
300
200
100
GLT25D2
GLT25D1
B
GLT25D1 mRNA levels [%]
120
100
80
60
40
20
0
HeLa
KD #1
KD #2
KD #3
C
ColGalT activity [%]
120
100
80
60
40
20
0
HeLa
KD #1
KD #2
KD #3
D
ColGalT activity [%]
120
100
80
60
40
20
0
0
20
40
60
80
100
120
GLT25D1 mRNA levels [%]

Figure 8


A
kb
brain
placenta
skeletal muscle
heart
kidney
pancreas
liver
lung
spleen
colon
3.7
5.1
5.9
2.3
GLT25D1
GLT25D2
CEECAM1
B
Organs
kidney
skeletal muscle
spleen
thymus
PBL
lymph node
trachea
lung
placenta
bladder
uterus
prostate
testis
ovary
liver
pancreas
adrenal gland
thyroid gland
salivary gland
mammary gland
GLT25D1
GLT25D2
CEECAM1
Nervous system
spinal cord
cerebral cortex
frontal lobe
parietal lobe
occipital lobe
corpus callosum
amygdala
caudate nucleus
hippocampus
GLT25D1
GLT25D2
CEECAM1
Fetal organs
fetal brain
fetal heart
fetal kidney
fetal liver
fetal spleen
fetal thymus
fetal lung
GLT25D1
GLT25D2
CEECAM1