

# TIN-ag-RP, a Novel Catalytically Inactive Cathepsin B-Related Protein with EGF Domains, Is Predominantly Expressed in Vascular Smooth Muscle Cells<sup>†,‡</sup>

Thomas Wex,<sup>§,||,⊥</sup> Alex Lipyansky,<sup>§,||</sup> Natascha C. Brömme,<sup>§</sup> Heike Wex,<sup>@</sup> Xiu Qin Guan,<sup>§</sup> and Dieter Brömme<sup>\*,§</sup>

Department of Human Genetics, Mount Sinai School of Medicine, New York, New York 10029, and Department of Pediatric Hematology and Oncology, Otto-von-Guericke University, Magdeburg, Germany

Received September 26, 2000; Revised Manuscript Received November 16, 2000

**ABSTRACT:** A human cDNA of 2166 bp encoding a novel cathepsin B-related protein was isolated and characterized. The amino acid sequence of the predicted protein of 467 aa was 46% identical with that of human tubulointerstitial nephritis antigen (TIN-ag), and therefore, the protein was tentatively designated as the TIN-ag-related protein (TIN-ag-RP). The amino acid sequence of TIN-ag-RP is composed of a 21 aa long signal sequence, a 181 aa long N-terminal domain containing two epidermal growth factor-like domains, a follistatin motif, and a 265 aa long cathepsin B-like domain. Interestingly, a serine residue has replaced the active site cysteine residue in the cathepsin B-like domain, resulting in a proteolytically inactive protein. Evolutionary analysis revealed that a distinct family of “TIN-ag-like” proteins had evolved in vertebrates. Northern blot analysis revealed a single TIN-ag-RP transcript of 2.4 kb in various tissues with the highest transcript levels detected in aorta, heart, placenta, skeletal muscle, kidney, and a colorectal adenocarcinoma cell line. Using a polyclonal anti-TIN-ag-RP antibody, TIN-ag-RP expression was predominantly seen in vascular smooth muscle (VSM) cells, but also in cardiac and skeletal muscle cells as well as in kidney. Interestingly, uterine smooth muscle cells completely lacked TIN-ag-RP expression, implying a regulated gene expression. Localization studies in HeLa cells stably transfected with TIN-ag-RP cDNA showed that TIN-ag-RP is glycosylated and actively secreted, a finding in line with its proposed function as a structural or regulatory protein similar to TIN-ag.

Inflammatory tubulointerstitial nephritis is a long-known kidney disease characterized by tubular destruction and intensive IgG and C3 deposits along the tubulo basement membrane (TBM)<sup>1</sup> (1, 2). Autoimmune tubulointerstitial nephritis (TIN) is generally characterized by the presence of immune complexes between autoantibodies and a kidney TBM component (1, 3). Biomedical studies and purification of potential target antigens from basement membranes of different animals led to the detection of multiple protein

forms ranging from 27 to 300 kDa (1, 4–7). This approach resulted in the identification of a human glycoprotein, termed the TIN antigen (TIN-ag), that was strongly recognized by anti-TBM antibodies (5). Recently, cDNAs encoding rabbit, mouse, and human TIN-ag orthologues have been cloned, and it was shown that these proteins are exclusively expressed in the TBM (8–10). Analyses of the TIN-ag amino acid sequences revealed a multidomain protein structure, including N-terminal EGF-like domains, a follistatin motif, and a C-terminal cathepsin B-like domain (10, 11). Multidomain or chimeric proteins are typical for extracellular matrix components (12–15).

Here, we report the cloning and characterization of a novel TIN-ag-related protein (TIN-ag-RP). Like TIN-ag, TIN-ag-RP is composed of several distinct protein domains, namely, two EGF-like domains and a proteolytically inactive cathepsin B-like domain. In contrast to TIN-ag, the gene and protein expression of TIN-ag-RP is not limited to kidney. TIN-ag-RP is expressed in a variety of human organs with its highest levels found in vascular and other smooth muscle cells. On the basis of its structural features and tissue distribution, TIN-ag-RP can be classified as a novel putative extracellular matrix protein that together with TIN-ag forms an evolutionally conserved group of TIN-ag-like proteins.

## EXPERIMENTAL PROCEDURES

*Isolation and Cloning of Human TIN-ag-RP cDNA from a Placenta cDNA Library.* Human TIN-ag-RP cDNA was originally identified by analyzing expressed sequence tags

<sup>†</sup> This work was supported by a fellowship (We/2170/2-1) from the Deutsche Forschungsgemeinschaft and in part by the personal funds of Doris Brömme-Weber and Hans-Jürgen Brömme.

<sup>‡</sup> The amino acid and cDNA sequence of TIN-ag-RP was deposited in GenBank (AF236150).

<sup>\*</sup> To whom correspondence should be addressed. Phone: (212) 659-6753. Fax: (212) 849-2508. E-mail: Dieter.Bromme@mssm.edu.

<sup>§</sup> Mount Sinai School of Medicine.

<sup>||</sup> These authors contributed equally to this work.

<sup>⊥</sup> Present address: Clinic for Gastroenterology, Hepatology and Infectious Diseases, University of Magdeburg, Leipziger Str. 44, D-39120 Magdeburg, Germany.

<sup>@</sup> Otto-von-Guericke University.

<sup>1</sup> Abbreviations: aa, amino acid; bp, base pair(s); CTLA, cytotoxic T cell-associated antigen; DMEM, Dulbecco's modified Eagle's medium; ECM, extracellular matrix; FCS, fetal calf serum; FITC, fluorescein isothiocyanate; HRP, horseradish peroxidase; mAb, monoclonal antibody; PAGE, polyacrylamide electrophoresis; PBS, phosphate-buffered saline; PBS/TT, phosphate-buffered saline containing Tween 20 and Triton X-100; SM, smooth muscle; TBM, tubulo basement membrane; TBS, Tris-buffered saline; TBS/TT, Tris-buffered saline containing Tween 20 and Triton X-100; TIN-ag-RP, tubulointerstitial nephritis antigen-related protein; UTR, untranslated region; VSM, vascular smooth muscle.

Table 1: Primers Used for Cloning and Expression of TIN-ag-RP

primer	primer sequence	position within the cDNA (residue nos.)
1	5'-cca-gaa-cct-gct-gtc-ttg-tga-ca-3'	822-844
2	5'-tgt-cac-aag-aca-gca-ggt-tct-gg-3'	844-822
3	5'-gcc-acc-atg-tgg-cga-tgt-cca-ctg-3'	58-72
4	5'-gca-gcc-tca-gtg-atg-acc-cat-gtc-3'	1464-1441
5	5'-cct-ttt-ccc-ccg-atc-caa-g-3'	345-364
6	5'-gcc-agc-ctc-gtc-tcg-ttc-3'	963-945
7	5'-aaa-aag-ctt-gcc-acc-atg-tgg-cga-tgt-cca-ctg-3' ( <i>Hind</i> III site)	58-72
8	5'-aaa-tct-aga-gcc-tat-ccc-atc-tgc-tgt-cct-cca-gtc-ata-ctg-gcc-atg-tga-tga-ccc-atg-tcc-tcc-atg-ccc-acg-cg-3' (T7 epitope and <i>Xba</i> I site)	1455-1430

(ESTs) of the GenBank database to identify novel thiol-dependent cathepsins. The human EST clone (accession number R48402) isolated from a breast cDNA library was significantly homologous to the region encoding the active site of cathepsin B. On the basis of the cDNA sequence of this clone, forward and reverse primers (primers 1 and 2 in Table 1) were designed that were separately used in combination with vector-based primers to screen several human  $\lambda$ gt-10 or  $\lambda$ gt-11 cDNA libraries. PCR amplifications were performed with *Pfu* polymerase (Stratagene, La Jolla, CA). Two PCR fragments were obtained using a human placenta cDNA library as a template (Clontech, Palo Alto, CA). The fragments contained either the 5'-region of the ORF, including 54 bp of the 5'-UTR, or the 3'-region of the ORF, including a 3'-UTR that was ~700 bp in length. Finally, primers flanking the ORF (primers 3 and 4 in Table 1) were designed and used to clone the cDNA containing the complete ORF from a placenta cDNA library. PCR products were cloned into Bluescript SK<sup>+</sup> and subsequently sequenced using a series of primers derived from either the vector or internal cDNA sequences. All nucleotide sequences were identified in both strands. Sequence analyses were performed using an Applied Biosystems model 377 automated sequencer.

**Amino Acid Sequence Alignments and Phylogenetic Analysis.** Multiple sequence alignments of amino acid sequences of papain-like cathepsins, cathepsin L-related proteins, and several TIN-ag-like proteins and EST clones were created using the ClustalW program (<http://dot.imgen.bcm.tmc.edu:9331/multi-align/Options/clustalw.html>). Protein sequences were retrieved from the GenBank Database {human (h) cathepsin K (S79895), h-cathepsin O (X77383), h-cathepsin C (U79415), h-cathepsin L (X12451), h-cathepsin H (X16832), h-cathepsin B (M14221), h-cathepsin S (M90696), h-cathepsin V (AF070448), h-cathepsin F (AF132894), h-cathepsin X (AF032906), h-cathepsin W (lymphopain) (AF055903), rat testin (U16858), mouse CTLA-alpha (NM007796), h-TIN antigen (AB022277), murine (mu) TIN antigen (AF153366), cysteine protease-like protein (*Caenorhabditis elegans*, termed ceTIN-ag) (Z81070), putative EST clones encoding TIN-ag-RP orthologues [mouse (AA288712), bovine (AW465388), and rat (AW141719)], and the gene product of *Drosophila melanogaster* (TIN-ag-like) (AAF46818)}. The phylogenetic dendrogram was created and verified using the Jones (JTT) algorithm of the Puzzle 4.0 program package (16).

**Northern Blot Analysis.** An ~620 bp PCR product (primers 5 and 6) of the TIN-ag-RP cDNA was labeled with [ $\alpha$ -<sup>32</sup>P]-dCTP using the READY TO GO kit (Amersham/Pharmacia, Piscataway, NJ). The probe was used for hybridization of

multiple-tissue Northern blots (Clontech). After prehybridization in ExpressHyb solution (Clontech), blots were hybridized with the probe at 68 °C for 1 h. After hybridization, blots were washed in 2× SSC containing 0.05% SDS at room temperature for 1 h and then at 50 °C for an additional 60 min in 0.1× SSC containing 0.1% SDS and finally subjected to autoradiography. RNA integrity and equal loading were confirmed by using GAPDH cDNA as a control (Clontech).

**Expression of Recombinant TIN-ag-RP in Escherichia coli and Generation of Polyclonal Anti-TIN-ag-RP Antibodies.** To obtain material for immunization, a fusion protein containing the entire ORF of TIN-ag-RP and an N-terminally located vector-encoded T7 epitope (pET17, Novagen, Madison, WI) was expressed in *E. coli* BL21 cells. The recombinant fusion product with an expected molecular mass of ~52 kDa was gel-purified and injected into New Zealand rabbits (Robert Sargeant, Ramona, CA). For lowering background staining, the polyclonal anti-TIN-ag-RP sera were further purified by affinity chromatography using nitrocellulose-blotted recombinant *E. coli*-expressed TIN-ag-RP as bait. Briefly, membrane strips with blotted TIN-ag-RP were blocked in TBS containing 7.5% nonfat milk for 1 h, rinsed twice with TBS containing 0.15% Tween 20 and 0.05% Triton X-100 (TBS/TT), and then incubated with 1 mL of diluted sera (1:60 in TBS/TT) at room temperature for 3 h. After four washing steps with TBS/TT (15 min each), antibodies were eluted from the membrane-bound TIN-ag-RP with 2 M glycine and 150 mM NaCl (pH 2.5) for 15 min, and then immediately neutralized with Tris base. As a negative control, prebleed sera were subjected to the same procedure.

**Expression of Recombinant TIN-ag-RP in HeLa Cells.** Cloned TIN-ag-RP cDNA was used as a template for PCR amplifications with *Pfu* polymerase (Stratagene) using primers 7 and 8 (Table 1) which contained restriction sites for cloning and the T7 epitope as a C-terminal extension. PCR amplification was performed as described previously (17). The PCR product that was obtained was digested with *Hind*III and *Xba*I, gel-purified using GeneClean (Bio 101, Vista, CA), and subcloned into the corresponding sites of the expression vector pcDNA.3 (Invitrogen, Carlsbad, CA). Human HeLa cells were maintained in Dulbecco's modified Eagle's medium supplemented with 20 mM glutamine, 100 units of penicillin, 100 units of streptomycin (all from Fisher Scientific), and 10% heat-inactivated fetal bovine serum (FBS, Gemini Bio-Products, Calabasas, CA) at 37 °C in a humidified 5% CO<sub>2</sub> atmosphere. HeLa cells were grown to 25% confluency in a six-well plate, washed twice with DMEM lacking all supplements (DMEM/LS), and then

subjected to the Lipofectamine-mediated transfection that was carried out according to the manufacturer's protocol. Briefly, 1  $\mu$ g of plasmid DNA from a midikit preparation (Qiagen, Hilden, Germany) was resuspended in 100  $\mu$ L of DMEM/LS, mixed with 6  $\mu$ L of Lipofectamine dissolved in 100  $\mu$ L of DMEM/LS, and subsequently incubated at room temperature for 40 min. After formation of the complex, 800  $\mu$ L of DMEM/LS was added, and cells were covered with 1 mL of transfection mixture per well. Cells were incubated in the CO<sub>2</sub> incubator for 5–6 h, and then 5 mL of complete DMEM containing 20% FCS per well was added. After overnight culture, the medium was replaced, and cells were cultured for an additional 48 h before the G418-mediated selection process was initiated. Transfected HeLa cells were grown in complete DMEM containing 100–500  $\mu$ g/mL G418 (Gemini Bio-Products) for 2–3 weeks, and single G418-resistant clones were established.

**Immunofluorescence and Immunohistochemical Staining of TIN-ag-RP.** HeLa cells and transfectants were grown on positively charged slides for 24–48 h (BioGenex, San Ramon, CA). After being washed twice with PBS (pH 7.4), cells were immediately fixed with PBS (pH 7.4) containing 4% (w/v) paraformaldehyde (Fisher Scientific) at room temperature for 10 min. After three washing steps, cells were blocked using goat serum (BioGenex) at room temperature for 45 min, then rinsed twice with PBS (pH 7.4), and incubated with either the anti-TIN-ag-RP antibody (1:100) or the anti-T7 mab (1:10000; Novagen) at 4 °C overnight or at room temperature for 3 h. After the cells had been washed with PBS containing 0.1% Tween 20 (PBS/T) three times for 20 min, cells were treated with anti-mouse or anti-rabbit IgG-tetramethyl rhodamine isothiocyanate (TRITC, Sigma) at room temperature for 90 min, and subsequently washed another three times. Finally, cells were mounted using Fluoromount-G (Southern Biotechnology Association Inc., Birmingham, AL) and viewed with a fluorescence microscope (Eclipse E-800, Nikon).

For immunohistochemical staining of endogenous TIN-ag-RP, fresh tissue samples obtained from surgery were immediately fixed in 10% formaldehyde (Fisher Scientific) for 48 h. Then, samples were dehydrated in increasing concentrations of ethanol (75–100%), incubated in chloroform for 45 min, and embedded in Paraplast (Fisher Scientific) at 64 °C for 1 h using a Leica EG1160 cryostat (Leica, Wetzlar, Germany). Finally, 4–5  $\mu$ m sections were prepared using a Microtome 2030 apparatus (Leica), and subsequently subjected to immunohistochemical analysis that was carried out using the SUPER SENSITIVE READY-TO-USE kit (BioGenex) according to the manufacturer's protocol. Briefly, tissue sections were treated twice with xylene for 3 min, rehydrated in decreasing concentrations of ethanol, and finally incubated in PBS (pH 7.4) for 5 min. Then, specimens were treated with 3% peroxide block for 15 min, followed by incubation for 30 min in protein block solution. After the specimens had been rinsed, tissue sections were incubated with either the purified anti-TIN-ag-RP antibody or corresponding prebleed controls at 4 °C in a humidified chamber overnight. After three washing steps in PBS/T, the samples were incubated with biotinylated anti-IgG (LINK) for 20 min, washed three times in PBS/T, treated with horseradish peroxidase-conjugated streptavidin (LAB-EL) for 20 min, and then washed again three times. Finally,

immune complexes were detected using a DAB substrate solution for 3–10 min, followed by counterstaining with hematoxylin for 2 min. Tissue specimens were subjected to dehydration using an increasing concentration of ethanol (25–100%), mounted using aqueous mounting medium (Fisher Scientific), and finally viewed with a microscope (Eclipse E-800, Nikon).

**PAGE and Western Blot Analysis.** Tissue and cell samples were homogenized in 2 mL of ice-cold lysis buffer containing 2% SDS, 5% (v/w) mercaptoethanol, 10% (v/w) glycerol, and 62.5 mM Tris (pH 6.8) using a tissue grinder and then subjected to SDS–PAGE following the method of Laemmli (18). Briefly, samples were heated in reducing sample buffer at 95 °C for 5 min and centrifuged at 13000g for 3 min, and the resulting supernatant was subjected to 10% SDS–PAGE (all reagents from Fisher Scientific) or 4 to 20% gradient SDS–PAGE (Novex). Proteins were electroblotted onto a nitrocellulose membrane (Fisher Scientific). Membranes were blocked in 7.5% nonfat milk in TBS (pH 7.4) containing 0.15% Tween 20 and 0.05% Triton X-100 (TBS/TT) and incubated with the primary anti-TIN-ag-RP antibody at 4 °C overnight and subsequently with the HRP-labeled anti-rabbit conjugate at room temperature for 90 min. Immune complexes were detected using the lumilight chemiluminescence detection system (Roche, Indianapolis, IN) as recommended by the manufacturer. The directly HRP-labeled anti-T7 conjugate (Novagen) was used for the detection of the T7 epitope.

## RESULTS AND DISCUSSION

**Characterization of TIN-ag-RP cDNA.** For the identification of novel cathepsin-related proteins, the EST database at GenBank was searched using conserved peptide sequences surrounding the active site cysteine residue of papain-like proteases as Blast sequences. Human EST clone R48402 (position marked in Figure 1) was found to be similar to the corresponding region of cathepsin B-like proteases and, therefore, was chosen for further characterization. To obtain a full-length cDNA,  $\lambda$ -gt10 and  $\lambda$ -gt11 placenta cDNA libraries were screened by PCR using R48402- and vector-derived primers. PCR products were subcloned and used to identify the complete ORF within the cDNA. The cDNA contained a 54 bp 5'-UTR, an ORF of 1401 bp, and a 3'-UTR that was 711 bp in length, exhibiting a complete length of 2166 bp (Figure 1). The ORF depicted in Figure 1 starts at nucleotide 55 with a putative ATG translational start codon and ends at position 1456 with a stop codon. The stop codon is followed by a 3'-UTR region 711 bp in length, which contains the polyadenylation signal at positions 2145–2150. The ORF encodes a putative polypeptide of 467 amino acids with a calculated molecular mass of 52.4 kDa that was found to be homologous to human TIN-ag. On the basis of this information, the protein was tentatively named the tubulointerstitial nephritis antigen (TIN)-related protein (TIN-ag-RP). The translational start site conforms to the Kozak consensus sequence (GCCACCAtgG) in one of the two key positions at position –3 or 4 (19). TIN-ag-RP contains at position 4 a thymidine instead of the guanine nucleotide, which makes the putative translational start codon at nucleotide 55 a medium-strong initiation site (20). The utilization of this start codon is further supported by the presence of a classical signal peptide 21 aa in length (Figure 1).



## A

gcgcggccagcctgggccccagccacaccttaccaggccgagcagccacctgtgg 60  
 cgatgtccactggggctactgtgtgtgtgcgctggctggccactgtgtgtgtgtg 120  
 R C P L G L L L L L P L A G H L A L G **W**  
 cagcagggctgtggcggcggagcagcagcgggtgtgtgtgtgtgtgtgtgtgtgt 180  
 Q Q G R G R R E L A P G L H L R G I R D  
 gcgggagggcggctgt 240  
 A G G R Y C Q E Q D L C C R G R A D D C  
 gccctgccctacctggggcgcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 300  
 A L P Y L G A I C Y C D L F C **M R T V S**  
 gactgtgccctgt 360  
 D C C P D F W D F C L G V P P P P P P I  
 caaggatgtatgcagtgaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 420  
 Q G C M H G G R I Y P V L G T Y W D N C  
 aacgtgtgcactgtccaggagcagcagcagcagcagcagcagcagcagcagcagc 480  
 N R C T C Q E N R Q W Q C D Q E P C L V  
 gatccagacatgatcaagcagcagcagcagcagcagcagcagcagcagcagcagc 540  
 D P D M I K A I N Q G N Y G W Q A G **M E**  
 agcgcccttggggcgcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 600  
 A F W G M T L D E G I R Y R L G T I R  
 ccattctctgt 660  
 P S S V M N M H E I Y T V L N P P P P V  
 ctctccagcagcctgt 720  
 L P T A F E A S E K W P N L I H E P L D  
 caagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 780  
 Q G N C A G S W A P S T A A V A S D R V  
 tcaatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 840  
 S I H S L G H M T P V L S P Q N L L S C  
 gac 900  
 D T H Q Q Q G C R G R L D G A W F L C  
 catccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 960  
 R R R G V V S D H C Y P P S G R E R D E  
 gctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1020  
 A G P A P P C M M H S R A M G R G K R Q  
 gccactgccactgccactgccactgccactgccactgccactgccactgccactgcc 1080  
 A T A H C P N S Y V N N N D I Y Q V T P  
 gtctaccgctcgctccac 1140  
 V Y R L G S N D K E I M K E L M E N G P  
 gtccacagcctcatggaggtgcagcagcagcagcagcagcagcagcagcagcagcagc 1200  
 V Q A L M E V H E D F F L Y K G G I Y S  
 cacacgccagtgagccttggggagcagagagacacacacacacacacacacacacac 1260  
 H T P V S L G R P E R Y R R H G T S S V  
 aagatcacagtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1320  
 K I T G W G E E T L P D G R T L K Y W T  
 gcggcacaactcctggggccagcctggggcgagagggggccactccgcatcgctgcgggc 1380  
 A A S W G P A W G E R G H F R I V R G  
 gtcatagtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1440  
 V N E C D I E S F V L G V W G R V G M E  
 gacatgggtcatcactgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1500  
 D M G H H  
 agggcggggcagagggcccaatggggcggtgacccagcctgcggccagagagccggc 1560  
 ggccagcggggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggc 1620  
 ctgggagcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggc 1680  
 caggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggc 1740  
 gattaccaaaagcaggacacacctcaagtttccagcccaatacccccaatactctgtat 1800  
 tcttt 1860  
 ccatcagggctcactgtaacctccgactcctgggttcaagtgaacctccacactcagcct 1920  
 ctcaagtagctgggactacaggtgacacacacacacacacacacacacacacacacacac 1980  
 aagaggggggtctcactgt 2040  
 acctgcctccgctcccaagt 2100  
 tatt 2160  
 tgatgg

## B



FIGURE 1: Nucleotide sequence, predicted amino acid sequence, and schematic presentation of human TIN-ag-RP. (A) The amino acid sequence is shown in single-letter code below the nucleotide sequence. The cleavage site of the signal peptide is indicated (▼). The translation start codon, the stop codon, two putative N-linked glycosylation sites, and the residues generally forming the active site in papain-like proteases [His400, Asn425, and the cysteine replacing the serine residue (Ser229)] are underlined and in boldface. The polyadenylation site is double-underlined. The sequence of the original EST clone (R48402) is double-underlined. (B) Schematic representation of TIN-ag-RP domain organization. The three major domains, the signal peptide (SP), and the cysteine-rich and cathepsin B-like domain are denoted. The localization of both N-linked glycosylation sites is denoted by diamonds within the column.

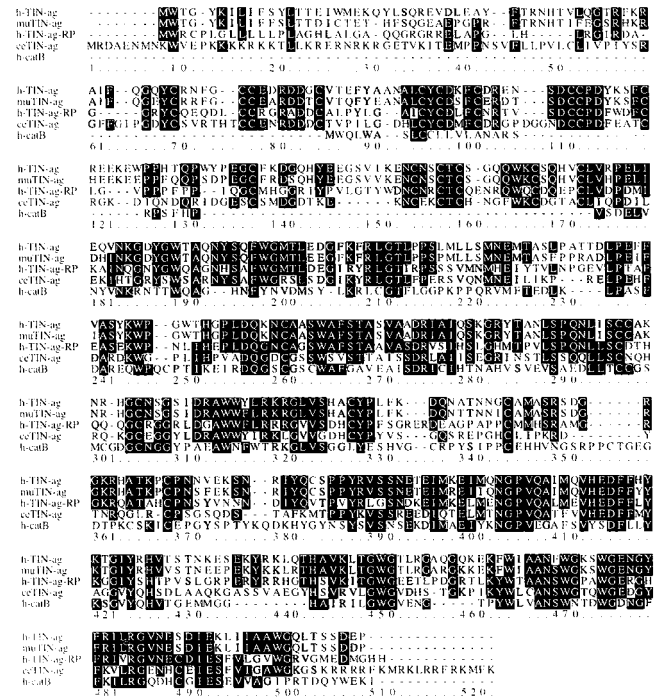


FIGURE 2: Multiple-amino acid sequence alignment of human TIN-ag-RP with human TIN-ag, TIN-ag orthologues of *C. elegans* and *D. melanogaster*, and human mature cathepsin B. Sequence alignments were determined using the CLUSTAL W program package (see Experimental Procedures). Identical amino acid residues are darkly shaded; similar amino acids and unrelated residues are shown with a white background.

Furthermore, TIN-ag-RP was found to contain a non-protease domain of 181 aa (Figure 1, amino acids 22–202) and a cathepsin B-related domain 265 aa in length (Figure 1A, amino acids 203–467) with predicated molecular masses of 20.4 and 29.8 kDa, respectively. The TIN-ag-RP cDNA-derived protein sequence contains two putative N-linked glycosylation sites in the non-protease domain that together with the presence of the signal peptide suggests that the protein might be targeted to the secretory pathway.

**Sequence and Structural Similarities to Other Proteins.** To investigate the relatedness of TIN-ag-RP to other proteins, homology and pattern analyses as well as multiple protein sequence alignments were performed using the BLAST, CLUSTAL-W, and PUZZLE programs (Figure 2). The highest degree of homology was found between the TIN-ag-RP and several TIN-ag mammalian orthologues whose amino acid sequences are 36–57% identical throughout (Table 2). In addition, related proteins were identified in the genomes of *C. elegans* (accession number Z81070) and *D. melanogaster* (AAF46818). An evolutionary analysis revealed that these two proteins together with TIN-ag and TIN-ag-RP orthologues form a separate branch within the dendrogram (Figure 3). On the basis of these data, we suggest that these proteins form a family of “TIN-ag-like” proteins. At present, TIN-ag and TIN-ag-RP represent the only known human members of this novel protein family. The proposed relatedness among the TIN-ag-like proteins is further supported by their conserved genomic organization as shown recently (21). The fact that both *C. elegans* and *D. melanogaster* have just one TIN-ag-like gene but mammals (*Bos taurus*, *Mus musculus*, *Rattus* sp., and *Homo sapiens*) express

Table 2: Sequence Comparison of Human TIN-ag-RP with TIN-ag Orthologues and Cathepsin B-like Proteases<sup>a</sup>

	level of identity (similarity) of amino acid sequences to human TIN-ag-RP (%)				
	mu-TIN-ag	h-TIN-ag	ce-TIN-ag	h-catB	bo-catB
prepro-protein (467 aa)	46.0 (55.7)	45.8 (55.3)	41.5 (52.2)	24.0 (32.1)	23.8 (31.9)
prepro-peptide (202 aa)	36.1 (45.0)	36.6 (47)	37.6 (44.6)	10.9 (15.8)	9.9 (15.3)
mature cathepsin-like domain (265 aa)	53.6 (64.5)	52.8 (61.5)	44.6 (58.1)	34.0 (44.5)	34.3 (44.5)

<sup>a</sup> Relative values of amino acid sequence identity and similarity (in parentheses) were calculated after pairwise alignment of each protein sequence (mu-TIN-ag, murine TIN-ag; h-TIN-ag, human TIN-ag; ce-TIN-ag, *C. elegans* TIN-like protein; h-catB, human cathepsin B; bo-catB, bovine cathepsin B) with the human TIN-ag-RP sequence using the ClustalW program (<http://dot.imgen.bcm.tmc.edu:9331/multi-align/Options/clustalw.html>).

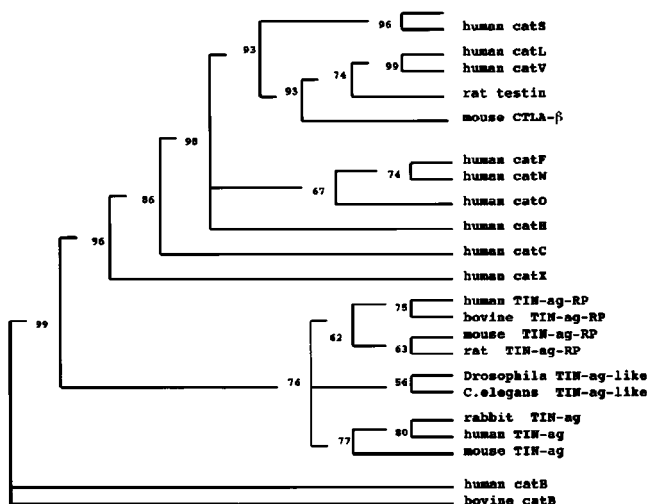


FIGURE 3: Phylogenetic tree based on deduced amino acid sequences of human papain-like proteins, including 11 human cathepsins, TIN-ag, and TIN-ag-RP orthologues, including related EST clones. The dendrogram was created and verified using the PUZZLE program (see Experimental Procedures). Support for the internal branches of the unrooted quartet puzzling tree topology is shown in percent. The dendrogram was obtained using the complete amino acid sequences of all proteins except the EST clones encoding putative TIN-ag-RP orthologues from *M. musculus*, *Rattus* sp., *B. taurus*, and *D. melanogaster*. This quartet puzzling tree is completely resolved (1000 puzzling steps, 7315 quartets analyzed, 5.1% unresolved quartets).

TIN-ag and TIN-ag-RP implies that the common ancestral gene between TIN-ag and TIN-ag-RP split late in evolution.

**Protein Domains in TIN-ag-RP.** Homology searches using either the complete amino acid sequence or portions of TIN-ag-RP revealed that TIN-ag-RP represents a “mosaic protein” composed of two major domains, namely, a cysteine-rich region 94 aa in length (amino acids 47–140) in the N-terminal non-protease domain and a 265 aa long cathepsin B-like domain (Figure 1B). The cysteine-rich domain was homologous to the “von Willebrand factor” (accession numbers U66246 and AF052036), several SCO-spondin-mucin-like proteins (AF01521, AF043909, and AC004877), and notch-like proteins (A49128 and M33874). Notably, all these proteins are homologous to the cysteine-rich N-terminus of TIN-ag-RP (amino acids 47–140) that contains 16 cysteine residues. Due to overlapping motifs, TIN-ag-RP contains two putative C-6 motifs and a C-8 EGF-like motif (22) that partially overlap with a follistatin-like domain (Figure 1B). Follistatin modules are present in all members of the EGF-like growth factor family, including EGF, transforming growth factor, and “heparin-binding EGF-like growth factor” (23). Notably, both TIN-ag (10) and TIN-ag-RP exhibit a similar follistatin module, but the spacing of their cysteine residues differs slightly from those of other

follistatin domains (data not shown). In addition, TIN-ag-RP contains a complete conserved lipocalin motif (residues 67–80) which is homologous to the lipocalin protein family. Lipocalins represent small extracellular proteins which are involved in the binding of small hydrophobic molecules and complex formation of extracellular proteins (24).

The largest protein domain of TIN-ag-RP is the 265 aa long cathepsin B-like domain (Figure 1B). Overall, the amino acid sequence of this cathepsin B-like domain is ~34% identical and ~45% similar to that of cathepsin B. Interestingly, the peptide sequences of TIN-ag-RP homologous to the active site histidine and asparagine residues in human papain-like cathepsins are conserved, whereas the region equivalent to the active site cysteine residue motif was found to be changed. Since site-directed mutagenesis of the active site cysteine residue of papain-like proteases resulted in a complete loss of enzymatic activity (25, 26), it can be assumed that TIN-ag-RP lacks significant proteolytic activity, a fact that was confirmed by biochemical analysis of recombinant TIN-ag-RP. Substrate assays at acidic and neutral pH values using synthetic and protein substrates revealed no protease activity for TIN-ag-RP (not shown).

Notably, TIN-ag-like proteins are not the only proteins that contain catalytically inactive domains of proteases. Mammalian testin is homologous to cathepsin L, but does not exhibit enzymatic activity (27). Like the TIN-ag-like protein family, testin contains a serine instead of a cysteine residue at the site homologous to the catalytic center of cysteine proteases. Recently, an aortic, catalytically inactive carboxypeptidase-like ECM protein was shown to be predominantly expressed in VSM, where it was found to be upregulated during VSM differentiation (28).

**Tissue Distribution of TIN-ag-RP.** Northern blot analysis of total RNA from multiple human tissues revealed moderate and high transcript levels in a variety of human tissues and cell lines. The colorectal adenocarcinoma W480 exhibited the strongest hybridization signal, followed by aorta (not shown), heart, placenta, and kidney (Figure 4). The majority of samples such as skeletal muscle, pancreas, lung, lymph nodes, bone marrow, and thyroid exhibited moderate levels of the TIN-ag-RP transcript. Human colon, small intestine, ovary, spleen, testis, and prostate revealed low transcript levels for TIN-ag-RP. No TIN-ag-RP transcripts were detected in hematopoietic cells and leukaemic cell lines. Hybridization signals revealed mostly a single transcript with an approximate size of about 2.4 kb that roughly matches the length of the cloned TIN-ag-RP cDNA [2166 bp excluding the poly(A) tail]. However, the carcinoma cell line W480 and the heart tissue revealed an additional minor band at 3.5 kb that might represent a second transcript or partially unspliced RNA (Figure 4).

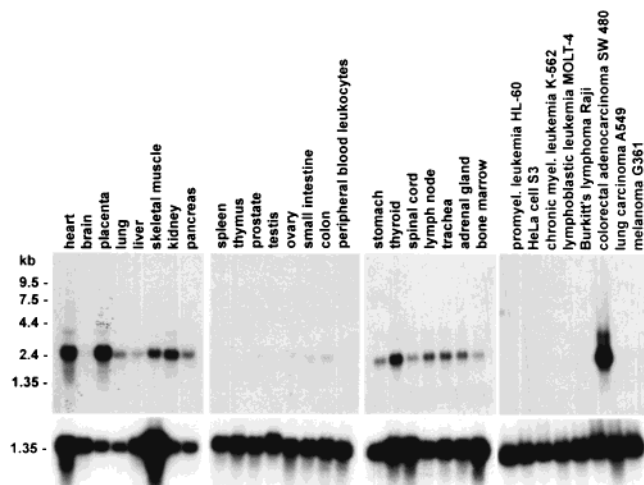


FIGURE 4: Northern blot analysis of human TIN-ag-RP in human tissues, cells, and tumor cell lines. Nitrocellulose blots containing total RNA from multiple human tissues, cells, and tumor cell lines (Clontech) were hybridized with an [ $\alpha$ - $^{32}$ P]dCTP-labeled TIN-ag-RP and GAPDH probe.

To confirm Northern blot analysis data at the protein level, a polyclonal anti-TIN-ag-RP antibody was generated. The anti-TIN-ag-RP antibody was characterized using cell extracts and culture supernatants of HeLa-D4 cells which were stably transfected with the pcDNA3.1 vector containing T7-tagged TIN-ag-RP cDNA. As shown in Figure 5, the immunohistochemical staining of anti-TIN-ag-RP antibodies revealed a similar vesicular staining pattern in HeLa-D4 cells that has been previously shown using the anti-T7 mab. In Western analysis, both antibodies detected a major band at 55–57 kDa (Figure 5). In addition to this major band, the polyclonal anti-TIN-ag-RP antibody recognized two minor bands at 40 and 20 kDa which were not recognized by the anti-T7 mab. Since the T7 epitope was C-terminally fused with TIN-ag-RP, these two proteins might represent proteolytically modified TIN-ag-RP-derived fragments lacking the C-terminus. Neither wild-type HeLa nor mock-transfected HeLa cells revealed any signals using anti-TIN-ag-RP antibodies (data partially shown in Figure 5).

To identify endogenous TIN-ag-RP, several human tissue specimens were subjected to immunohistochemistry and Western blot analysis. As shown in Figure 6, the largest

amount of TIN-ag-RP was detected in total cell lysates from placenta and aorta, whereas human liver and small intestine revealed small but detectable amounts of TIN-ag-RP. Interestingly, two samples of colon adenomas revealed moderate to high levels of expression of TIN-ag-RP, indicating that in neoplastic tissue an upregulation of TIN-ag-RP occurs. This is in accordance with the finding that the colorectal adenocarcinoma W480 cell line exhibited a high level of expression of the TIN-ag-RP transcript (Figure 4). Notably, HeLa-D4 cells and placenta exhibited minor bands at lower molecular masses, indicating a proteolytic degradation or processing.

Immunohistochemical studies on human tissue specimens revealed high expression levels of TIN-ag-RP in vascular smooth muscle (VSM) cells, but other muscle cells, including visceral SM, skeletal, and cardiac muscle cells, also contained considerable amounts of TIN-ag-RP. As shown in panels 1, 3, and 5 of Figure 7, the tunica media of blood vessels expressed TIN-ag-RP at high levels (brownish staining), whereas the inner layer of endothelial cells ("E") was found to be negative. The corresponding prebleed of the anti-TIN-ag-RP antibody, used as a negative control, did not exhibit any signals (panel 2 of Figure 7). This strong expression of TIN-ag-RP in the tunica media of the vasculature system may be responsible for the ubiquitous distribution of the mRNA in the multiple-tissue Northern blots. Analyses of adult heart tissue revealed that cardiac muscle (CM) cells also express TIN-ag-RP, whereas surrounding fibrocytes (F) were negative for TIN-ag-RP (panels 3 and 4 of Figure 7). Analogous analyses of embryonic tissues confirmed the expression of TIN-ag-RP in the tunica media of blood vessels, heart tissue (data not shown), and skeletal muscle (panel 6 of Figure 7). In contrast, the visceral SM cells of the uterus did not exhibit any TIN-ag-RP expression, although VSM cells of the same section were strongly positive (panel 5 of Figure 7). In addition to this muscle-specific TIN-ag-RP expression, significant staining was observed in cells forming the collecting tubules in the kidney medulla and tubules and glomerular capillaries in the renal corpuscles of the cortex (panels 7 and 8 of Figure 7). TIN-ag, the other family member of TIN-ag-related proteins, is specifically expressed in the basal membrane of the epithelium of Bowman's capsule of early stage glomeruli and

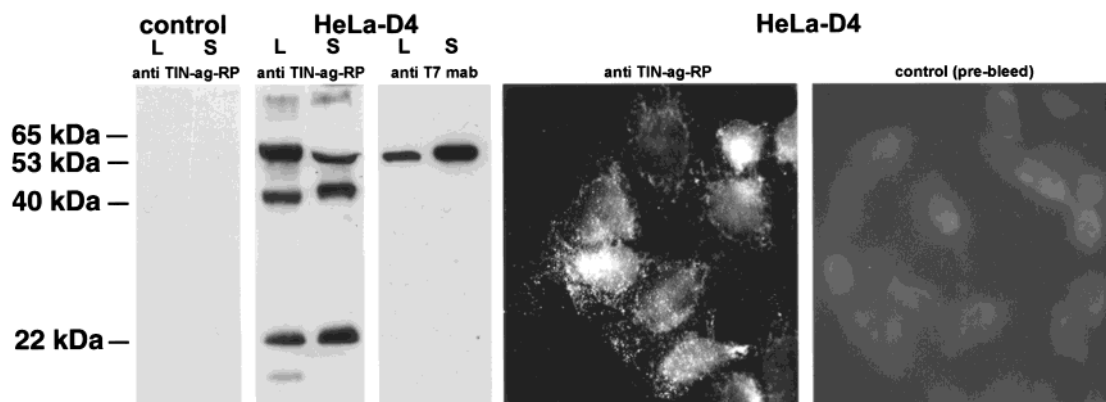


FIGURE 5: Characterization of the polyclonal anti-TIN-ag-RP antibody. (Left) The anti-TIN-ag-RP antibody used in Western blot analysis recognized three molecular forms of recombinant TIN-ag-RP (55–57, 40, and 20 kDa) in stably transfected HeLa-D4 cells, whereas the vector-transfected control did not exhibit any signals. The anti-T7 mab detected only the 55–57 kDa protein (L, lysate; S, supernatant). (Right) Immunofluorescence staining of TIN-ag-RP revealed a vesicular staining in HeLa-D4 cells. Purified preimmune serum was used as a control.



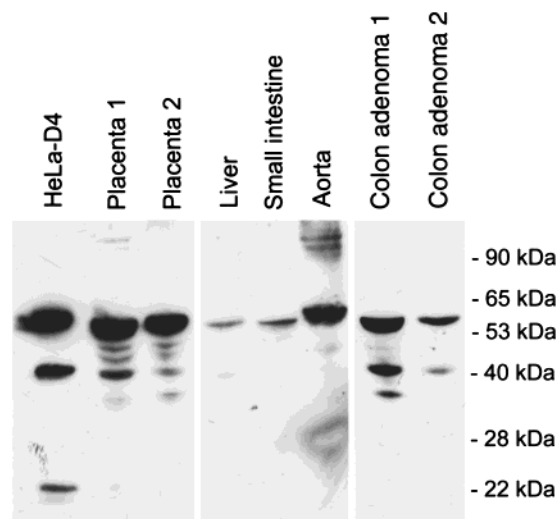


FIGURE 6: Western blot analysis of TIN-ag-RP-expressing HeLa cells, human tissue samples, and colon adenomas. Equal amounts of total protein extract from the specimen identified on the top of each lane were separated in a 4 to 20% gradient SDS-PAGE and transferred onto nitrocellulose, and finally, the TIN-ag-RP was identified by immunohistochemistry using the polyclonal anti-TIN-ag-RP antibody. As a positive control, the 50-fold diluted total cell lysate from stably TIN-ag-RP-transfected HeLa-D4 cells was used.

primitive proximal tubuli (9). The possibility that the polyclonal antibody raised against TI-ag-RP shows some cross reactivity with the kidney-specific TIN-ag cannot be unambiguously excluded. However, it has been described that TIN-ag is primarily expressed in fetal tubulogenesis (9, 11), whereas our kidney sample was derived from an adult individual. A weaker staining was also found in enterocytes of the small intestine as well as in cytotrophoblastic and syncytiotrophoblastic cells of the placenta (data not shown).

**Expression of Recombinant TIN-ag-RP.** HeLa cells were stably transfected with a TIN-ag-RP T7 epitope cDNA construct, and a G418-resistant clone (HeLa-D4) was established (Figure 6). As shown in Figure 8, recombinant TIN-ag-RP was mainly found in the culture medium. The finding that the level of cellular expression (lysate) did not show a significant increase over a 4 day culture experiment suggests that TIN-ag-RP is constitutively secreted. The secretion of TIN-ag-RP was completely blocked by the addition of tunicamycin that inhibits N-linked glycosylation (29). This finding is in line with the fact that glycosylation is required for entering the secretory pathway (30). In addition, the *in vivo* inhibition of N-linked glycosylation in HeLa-D4 cells revealed a 3–4 kDa shift in the molecular mass of TIN-ag-RP detected in the cellular lysate, showing that both putative N-linked glycosylation sites (marked in Figure 1) are used *in vivo*. In addition, the secretion provided indirect proof for the cleavage of the signal peptide of TIN-ag-RP. To identify its exact processing site, secreted TIN-ag-RP was purified using the C-terminal T7 epitope as a target and subjected to N-terminal sequencing. Edman degradation revealed an N-terminal sequence of X-Q-Q-G-R-G (performed by the DNA/protein technology center at Rockefeller University, New York, NY). On the basis of these results, the cleavage site between the signal peptide and the cysteine-rich domain of TIN-ag-RP was found between Gly21 and Ala22, confirming the putative cleavage site predicted by the Heijne's algorithm.

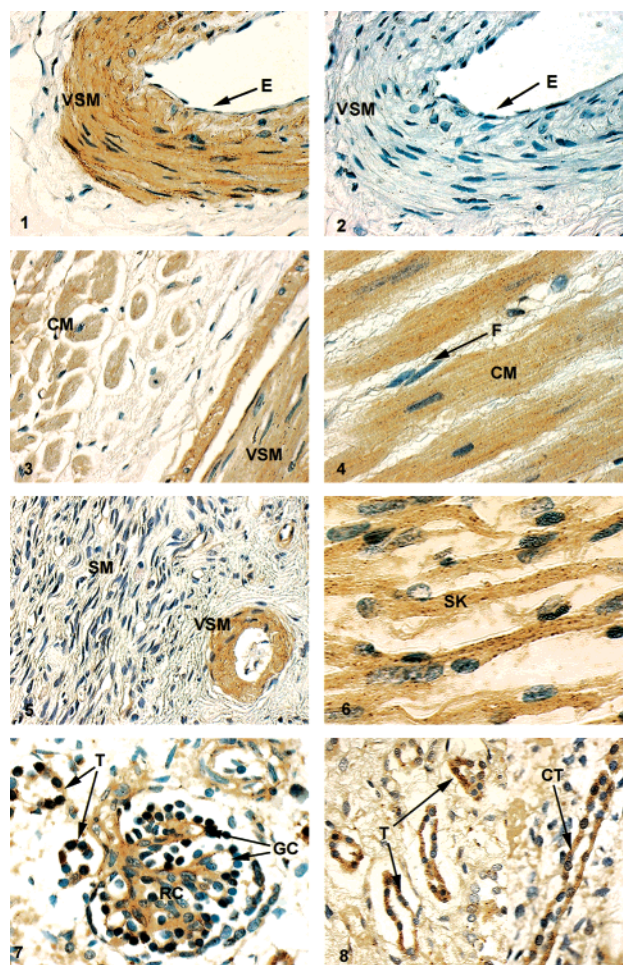


FIGURE 7: Immunohistochemical localization of endogenous TIN-ag-RP in human tissue specimens. TIN-ag-RP was detected using the affinity-purified polyclonal anti-TIN-ag-RP antibody (1:200 to 1:400 dilution). The antigen-antibody complex was identified by a peroxidase reaction resulting in a brownish staining pattern. Preimmune serum which was identically treated using nitrocellulose blots containing recombinant TIN-ag-RP was used as a negative control (1:50 dilution). The following specimens were analyzed: (1) detail of artery, vascular smooth muscle (VSM) cells and endothelial cells (E); (2) negative control for panel 1; (3) heart, cross section of cardiac muscle fibers (HM) and longitudinal section of blood vessel with VSM; (4) adult heart, longitudinal section of cardiac muscle fibers (CM) and surrounding connective tissue with fibrocytes (F); (5) uterus, interlacing bundles of smooth muscle (SM) cells which are separated from the VSM of the blood vessel by connective tissue; (6) embryonic skeletal muscle (10th week of gestation), skeletal muscle cells (SK) with multiple nuclei; (7) kidney cortex with renal corpuscle (RC) and glomerular capillaries (GC); and (8) kidney medulla with longitudinal cross section of tubules (T, left side) and longitudinal cross section of collecting tubules (CT, right side). Magnifications were 600-fold for panels 1, 2, 4, and 6–8 or 200-fold for panels 3 and 5.

**Functional Implications and Conclusion.** The presence of the two major structural domains, namely, the cysteine-rich region, including EGF-like and follistatin-like motifs, as well as the cathepsin B-like domain, suggests pleiotropic functions for TIN-ag-RP. EGF-like domains have been identified in a variety of other proteins, including cell surface receptors, matrix proteins, and enzymes (12, 31). Interestingly, there seems to be no common function of EGF-like domains in these proteins. Numerous reports have shown that EGF-like domains are involved in the regulation of differentiation and proliferation as well as blood coagulation by mediating

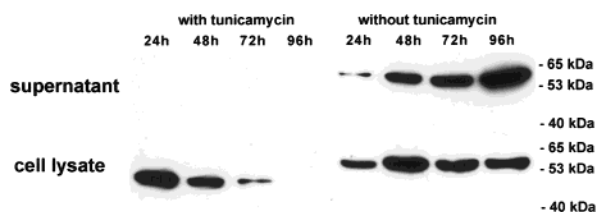


FIGURE 8: Expression and deglycosylation of recombinant TIN-ag-RP in HeLa-D4 cells. HeLa-D4 cells were grown in either the absence or presence of 10  $\mu$ g/mL tunicamycin. Total cell lysate (L) and culture supernatant (S) were harvested as indicated, and then subjected to Western blot analysis using the anti-T7 mab.

protein–protein interactions (15, 32). Since the EGF- and follistatin-like domains of TIN-ag-RP are ~75% similar to the consensus sequence of both modules and have some additional or missing spacing between their cysteine residues, it is unlikely that they mediate a direct interaction with a growth hormone receptor of the EGF family. However, it has been shown that EGF-like motifs, as present in the TIN-ag-RP, can efficiently mediate protein–matrix interactions between extracellular matrix fibers and growth factor binding proteins such as latent TGF- $\beta$  binding proteins (LTBPs) and IGF-binding proteins (IGFBPs) (33, 34). Since TIN-ag was shown to interact with type IV collagen and laminin (11, 35), it is likely that TIN-ag-RP also will bind related proteins present in the ECM. If, in analogy to the LTBPs and IGFBPs, the EGF- and follistatin-like domains of TIN-ag-RP also represent an ECM binding region, TIN-ag-RP would be targeted to the ECM compartment by its EGF-like domains. The function of the catalytically inactive cathepsin B-like domain remains unknown. It might be speculated that it retains substrate binding properties which might be utilized for specific ECM–TIN-ag-RP interactions. For example, ECM components such as vitellin, collagen type X, and aggrecan have been shown to be substrates of cathepsin B-like proteases (36–38).

Taken together, TIN-ag-RP represents a modular protein structure typical for extracellular protein families that is composed of two major domains, the cysteine-rich N-terminus and a catalytically inactive cathepsin B-like domain. Its predominant expression in smooth muscle cells may suggest a structural and/or regulatory function of this protein in the development of SM cells, particularly in VSM cells.

## REFERENCES

- Wilson, C. B., Holdsworth, S. R., and Meale, T. J. (1981) *Aust. N. Z. J. Med.* 11, 94–100.
- Alpers, C. E. (1998) *Renal Failure* 20, 763–771.
- Wuthrich, R. P., and Sibalic, V. (1998) *Exp. Nephrol.* 6, 288–293.
- Yoshida, H., Wakashin, Y., Ueda, S., Azemoto, R., Iesato, K., Yamamoto, S., Mori, T., Ogawa, M., Mori, Y., and Wakashin, M. (1990) *Kidney Int.* 37, 1286–1294.
- Katz, A., Fish, A. J., Santamaria, P., Nevins, T. E., Kim, Y., and Butkowski, R. J. (1992) *Am. J. Med.* 93, 691–698.
- Butkowski, R. J., Kleppel, M. M., Katz, A., Michael, A. F., and Fish, A. J. (1991) *Kidney Int.* 40, 838–846.
- Joh, K., Takada, K., Ohkawa, K., Kanetsuna, Y., and Aizawa, S. (1998) *Int. J. Mol. Med.* 1, 223–226.
- Nelson, T. R., Charonis, A. S., McIvor, R. S., and Butkowski, R. J. (1995) *J. Biol. Chem.* 270, 16265–16270.
- Nelson, T. R., Butkowski, R. J., Michael, A. F., and Charonis, A. S. (1997) *Connect. Tissue Res.* 36, 223–229.
- Ikeda, M., Takemura, T., Hino, S., and Yoshioka, K. (2000) *Biochem. Biophys. Res. Commun.* 268, 225–230.
- Kanwar, Y. S., Kumar, A., Yang, Q., Tian, Y., Wada, J., Kashiwara, N., and Wallner, E. I. (1999) *Proc. Natl. Acad. Sci. U.S.A.* 96, 11323–11328.
- Aumailley, M., and Gayraud, B. (1998) *J. Mol. Biol.* 76, 253–265.
- Shaub, A. (1999) *Nat. Cell Biol.* 1, E173–E175.
- Wuthrich, R. P., and Sibalic, V. (1998) *Exp. Nephrol.* 6, 288–293.
- Engel, J. (1996) *Matrix Biol.* 15, 295–299.
- Strimmer, K., and von Haeseler, A. (1997) *Proc. Natl. Acad. Sci. U.S.A.* 94, 6815–6819.
- Wex, T., Levy, B., Smeeckens, S. P., Ansorge, S., Desnick, R. J., and Brömme, D. (1998) *Biochem. Biophys. Res. Commun.* 248, 255–261.
- Laemmli, U. K. (1970) *Nature* 227, 680–685.
- Kozak, M. (1987) *Nucleic Acids Res.* 15, 8125–8148.
- Kozak, M. (1996) *Mamm. Genome* 7, 563–574.
- Brömme, N. C., Wex, T., Wex, H., Levy, B., Lipynsky, A., and Brömme, D. (2000) *Biochem. Biophys. Res. Commun.* 271, 474–480.
- Engel, J. (1989) *FEBS Lett.* 251, 1–7.
- Uchida, T., Wada, K., Akamatsu, T., Yonezawa, M., Noguchi, H., Mizoguchi, A., Kasuga, M., and Sakamoto, C. (1999) *Biochem. Biophys. Res. Commun.* 266, 593–602.
- Flower, D. R. (1996) *Biochem. J.* 318, 1–14.
- Vernet, T., Khouri, H. E., Laflamme, P., Tessier, D. C., Musil, R., Gour-Salin, B. J., Storer, A. C., and Thomas, D. Y. (1991) *J. Biol. Chem.* 266, 21451–21457.
- Baker, S. C., Yokomori, K., Dong, S., Carlisle, R., Gorbelenya, A. E., Koonin, E. V., and Lai, M. M. (1993) *J. Virol.* 67, 6056–6063.
- Cheng, C. Y., Morris, I., and Bardin, C. W. (1993) *Biochem. Biophys. Res. Commun.* 191, 224–231.
- Layne, M. D., Endege, W. O., Jain, M. K., Yet, S. F., Hsieh, C. M., Chin, M. T., Perrella, M. A., Blonar, M. A. H. E., and Lee, R. T. (1998) *J. Biol. Chem.* 273, 15654–15660.
- Elbein, A. D. (1984) *CRC Crit. Rev. Biochem.* 16, 21–49.
- Pellerin, S., Lefeuvre, B., Wade, R. H., Savona, C., Chambaz, E. M., and Feige, J. J. (1993) *J. Biol. Chem.* 268, 18810–18817.
- Appella, E., Weber, I. T., and Blasi, F. (1988) *FEBS Lett.* 231, 1–4.
- Kuroda, S., and Tanizawa, K. (1999) *Biochem. Biophys. Res. Commun.* 265, 752–757.
- Bayes-Genis, A., Conover, C. A., and Schwartz, R. S. (2000) *Circ. Res.* 86, 125–130.
- Saharinen, J., Hyytiäinen, M., Taipale, J., and Keski-Oja, J. (1999) *Cytokine Growth Factor Rev.* 10, 99–117.
- Kalfa, T. A., Thull, J. D., Butkowski, R. J., and Charonis, A. S. (1994) *J. Biol. Chem.* 269, 1654–1659.
- Cho, W. L., Tsao, S. M., Hays, A. R., Walter, R., Chen, J. S., Snigirevskaya, E. S., and Raikhel, A. S. (1999) *J. Biol. Chem.* 274, 13311–13321.
- Sires, U. I., Schmid, T. M., Fliszar, C. J., Wang, Z. Q., Gluck, S. L., and Welgus, H. G. (1995) *J. Clin. Invest.* 95, 2089–2095.
- Fosang, A. J., Last, K., Gardiner, P., Jackson, D. C., and Brown, L. (1995) *Biochem. J.* 310, 337–343.

BI0022660