# Primary structure of a putative serine protease specific for IGF-binding proteins

Jürg Zumbrunn, Beat Trueb*

*M.E. Müller-Institut für Biomechanik, Universität Bern, Postfach 30, CH-3010 Bern, Switzerland*

Received 10 October 1996; revised version received 17 October 1996

**Abstract From a subtracted cDNA library we have isolated a cDNA clone coding for a novel transformation-sensitive protein which is expressed by human fibroblasts, but not by their matched SV40 transformed counterparts. This protein has a molecular mass of 51 kDa and is highly related to the HtrA family of serine proteases from bacteria. At the N-terminal end, it contains an IGF-binding domain which may modulate the activity of the associated serine protease. Our data are consistent with the assumption that the novel protein represents one of the proteases that regulate the availability of IGFs by cleaving IGF-binding proteins.**

*Key words:* Insulin-like growth factor; Insulin-like growth factor binding protein; Serine protease; Subtractive cDNA cloning; Transformation-sensitive protein; Human fibroblast

## 1. Introduction

Fibroblasts transformed by oncogenic viruses are often used by researchers as a model system to investigate the molecular events that lead to immortalization and phenotypic transformation of tumor cells. Usually, such fibroblasts exhibit dramatic alterations in their morphology, including a roundish shape, irregularities at the cell surface and disintegration of the cytoskeleton. It is well accepted that these changes are caused by subtle alterations in the synthesis and degradation of proteins (transformation-sensitive proteins) that are involved, directly or indirectly, in cell adhesion and attachment.

We have recently reported a novel approach to search for transformation-sensitive proteins [1]. Genes that were expressed by normal fibroblasts, but not by their SV40 transformed counterparts, were selected by subtractive cDNA cloning. This approach led to the isolation of a dozen differentially expressed cDNA clones, many of which coded for known transformation-sensitive proteins, including fibronectin, collagen VI and vinculin.

One of the clones obtained in this way was found to code for a novel serine protease. Here we present evidence that this protein might be involved in the cleavage of IGF-binding proteins.

*Corresponding author. Fax: (41) (31) 632 4999.

*Abbreviations:* DFP, diisopropyl fluorophosphate; IGF, insulin-like growth factor; IGFBP, insulin-like growth factor binding protein; RACE, rapid amplification of cDNA ends.

## 2. Materials and methods

### 2.1. Cell culture

Embryonic human lung fibroblasts (WI38) and their SV40 transformed counterparts (WI38-VA13) were purchased from the American Type Culture Collection (Rockville, MD). The cells were cultivated at 37°C under an atmosphere of 5% $CO_2$ in Dulbecco's modified Eagle's Medium supplemented with 9% fetal calf serum, 100 μg/ml streptomycin and 100 U/ml penicillin.

### 2.2. Northern blotting

RNA was extracted from confluent cell layers by the SDS/proteinase K method and purified on oligo(dT)-cellulose [1–3]. The samples (2 μg/lane) were resolved on a 1% agarose gel in the presence of 1 M formaldehyde and transferred to a Nylon membrane by vacuum blotting [3]. The membrane was hybridized overnight under standard conditions (42°C, 50% formamide) with a cDNA probe that had been labeled by the random primed oligolabeling method [4]. The blot was washed at regular stringency and exposed to X-ray film. The multiple tissue Northern blot used in this study contained poly(A) RNA (2 μg/lane) from eight different human tissues (Clontech, Palo Alto, CA) and was processed as described above.

### 2.3. Screening of cDNA libraries

A subtracted cDNA library was created with mRNA from human fibroblasts and their SV40 transformed counterparts by the biotin/streptavidin/phenol method [1,5]. Clone L56 was used to screen two commercial cDNA libraries, one prepared from human fibroblasts (HL1011), the other prepared from human placenta (HL1075b, Clontech) by the plaque hybridization technique [3,6]. Positive clones were picked and subcloned into the plasmid pUC19. To obtain the 5′-end of the L56 mRNA, the AmpliFINDER RACE Kit (Clontech) was utilized in combination with several synthetic oligonucleotide primers that had been designed according to the 5′ sequences of the cDNA clones.

### 2.4. DNA sequencing

The sequences of the cDNA inserts were determined on both strands by the dideoxy chain termination method [7] using Sequenase 2.0 (USB, Cleveland, OH). The derived amino acid sequences were compared with all entries of the Swissprot databank (release 34.0) using the GCG Computer Program Package (University of Wisconsin, Madison, WI).

## 3. Results

### 3.1. Isolation of cDNA clones

Subtractive cDNA cloning starting from mRNA of normal and SV40 transformed fibroblasts yielded a dozen of differentially expressed cDNA clones. One of these clones, termed L56, was further investigated in this report. The insert of this clone hybridized to an mRNA of ~2300 nucleotides present in normal human fibroblasts, but missing in their matched SV40 transformed counterparts (Fig. 1, left). As demonstrated on a Northern blot with RNA from eight different human tissues, this mRNA was strongly expressed in placenta, moderately in brain, liver and kidney, and weakly in lung, skeletal muscle, heart and pancreas (Fig. 1, right).


*PII* S0014-5793(96)01229-X

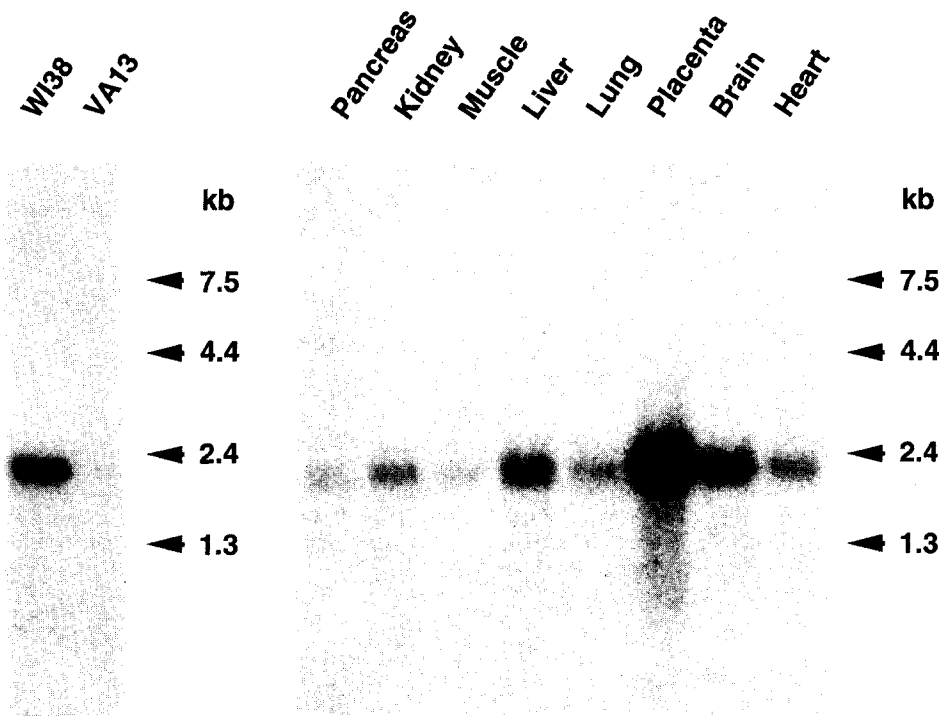


Fig. 1. Northern blot analysis. RNA from human fibroblasts (WI38), their SV40 transformed counterparts (VA13) and eight human tissues was resolved on agarose gels, transferred to Nylon membranes and hybridized with the radiolabeled insert of clone L56. The migration positions of standard RNAs are indicated in the margin.

Since the insert of the original clone was only 228 bp long and did not show much similarity with any of the entries of the EMBL databank, it was used as a probe to screen two commercial cDNA libraries, one prepared from human fibroblasts, the other from human placenta. 14 overlapping cDNA clones were found, but none contained the entire reading frame of the encoded protein. We therefore utilized the RACE technique to select clones that specifically contained the 5′-end of the mRNA. After several rounds of primer extension and amplification by the polymerase chain reaction, 16 additional cDNA clones were obtained. Altogether our clones covered ~2100 bp.

*3.2. Nucleotide and derived amino acid sequence*

The combined cDNA sequence contained a G/C-rich 5′ untranslated region of 38 bp (Fig. 2). The first ATG codon (positions 39–41) fulfilled the criteria of a typical start site for translation [8]. This start codon was followed by an open reading frame of 1440 bp which terminated in the stop codon TAG (positions 1479–1481). The 3′ untranslated region was 550 bp long and contained the sequence ATTAAA (positions 2007–2012) that resembled the consensus signal AATAAA for polyadenylation. This signal was followed after 17 bp by a poly(A) tail of ~80 nucleotides.

The open reading frame could be translated into an amino acid sequence of 480 residues (Fig. 2). The predicted protein had a molecular mass of 51 kDa with a calculated isolectric point of 7.7. The sequence started with a typical signal peptide which represented the most hydrophobic portion of the entire protein on a hydrophilicity blot according to Kyte and Doolittle [9]. The rules of Von Heijne [10] predicted that this signal peptide was cleaved after alanine at position 22 with a score of 13.5. The mature product would therefore start with a glutamine residue which might spontaneously convert into pyroglutamic acid and block the N-terminus of the polypeptide as observed with many extracellular matrix proteins.

*3.3. Similarity to other proteins*

A detailed comparison of the protein sequence with all entries of the Swissprot databank revealed a substantial similarity between the first 120 amino acid residues and the family of the IGF-binding proteins [11,12]. With IGFBP-3 the L56 protein shared 44% sequence identity and this increased to 58% when conservative amino acid substitutions were included (Fig. 3, top). Comparable similarities were noted between the L56 protein and the other members of the IGFBP family. The aligned segment corresponded to a cluster of 12 cysteine residues that are conserved in all IGF-binding proteins. 11 of these cysteines were also found in the L56 sequence and the only cysteine that did not line up occurred in the L56 sequence shifted by 7 residues. It is therefore likely that the N-terminus of the L56 protein represents an IGF-binding domain.

The C-terminal part of the L56 protein (amino acids 150–480) was found to be highly related to the family of the HtrA/Do proteases from bacteria [13–15]. A comparison with the HtrA protease from *E. coli* (Fig. 3, bottom) showed 35% sequence identity or 58% similarity if conservative amino acid replacements were included. Similar alignments were obtained with the corresponding proteases from *Haemophilus influenzae* and *Salmonella typhimurium*. All these proteases are required for the survival of bacteria at temperatures above 42°C. We therefore investigated the possibility whether the

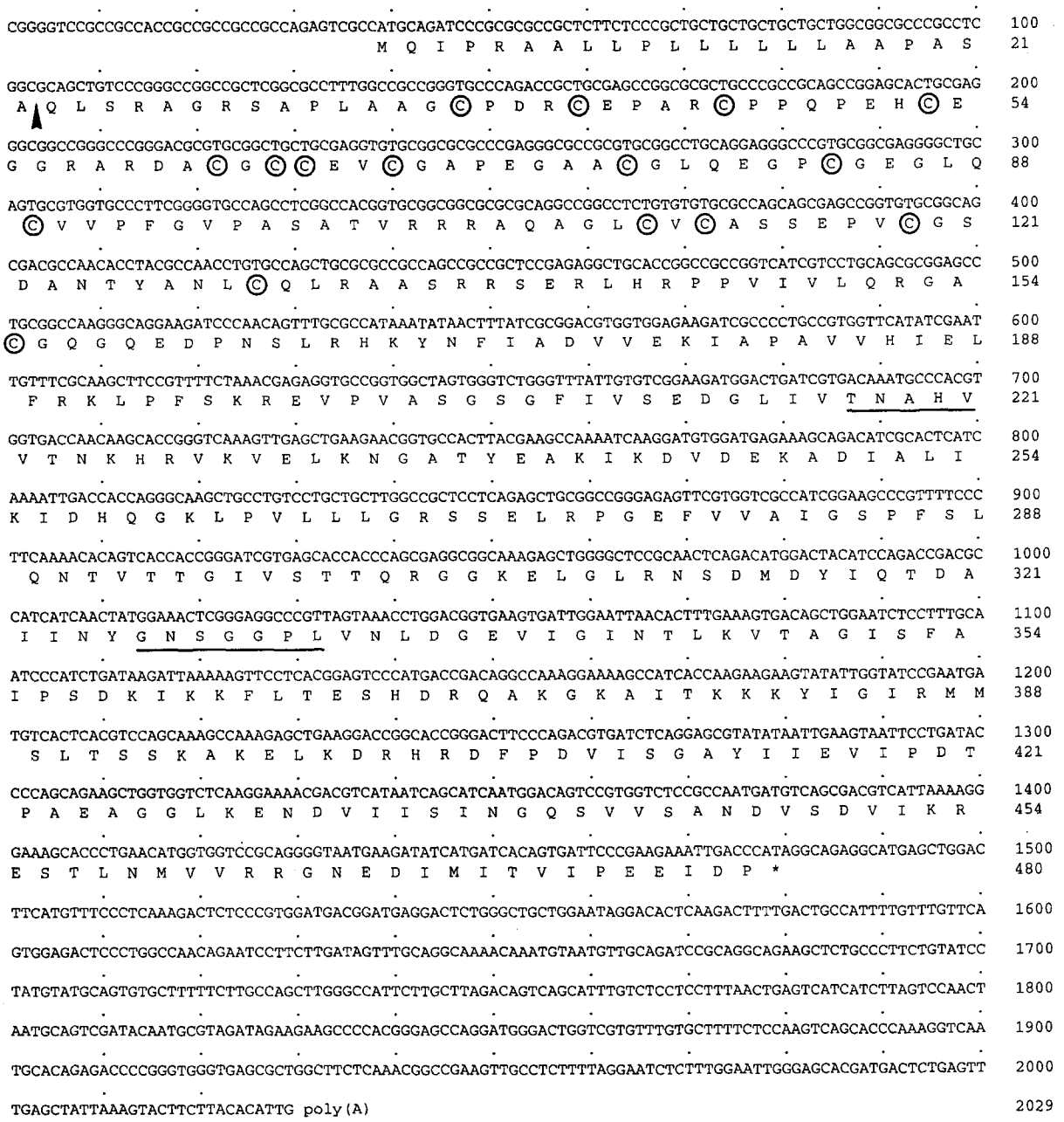

Fig. 2. Complete nucleotide and derived amino acid sequence of the novel transformation-sensitive protein (databank accession no. Y07921). The signal peptidase cleavage site is indicated by an arrow, sequences surrounding the active serine and histidine residue are underlined; cysteine residues are encircled.

L56 protein might represent a heat shock protease. However, no change in the level of the L56 mRNA was noted when human fibroblasts were shifted from 37 to 41°C for 4 h (not shown). The HtrA/Do proteases belong to the family of the serine proteases since they possess the amino acid sequence GNSGGAL in their active site. The corresponding sequence GNSGGPL was found in the L56 protein at position 326–332. For catalytic activity, the HtrA proteases require, in addition to the active serine, a conserved histidine which occurs in the *E. coli* sequence at position 109 (TNNHV). This histidine was also found in the human sequence at position 220 (TNAHV). Thus, there remains little doubt that the human L56 protein represents a functional serine protease.

The last 40 amino acid residues of the IGF-binding domain and the first 20 residues of the protease domain were found to share a striking similarity with the Kazal type of serine protease inhibitors (Fig. 3, middle) [16]. An alignment of the human pancreatic trypsin inhibitor [17] with this region of the L56 sequence revealed 32% sequence identity or 53% sequence similarity. With human follistatin which harbors several Kazal domains [18], the identity was even 42% (60% similarity). Protease inhibitors of the Kazal type possess a conserved tyrosine residue and six cysteine residues which are arranged in three disulfide bonds. The tyrosine residue and four of the cysteine residues were conserved in the L56 protein. It is therefore possible that the related segment of the L56 protein will fold into a tertiary structure resembling that of trypsin inhibitors.

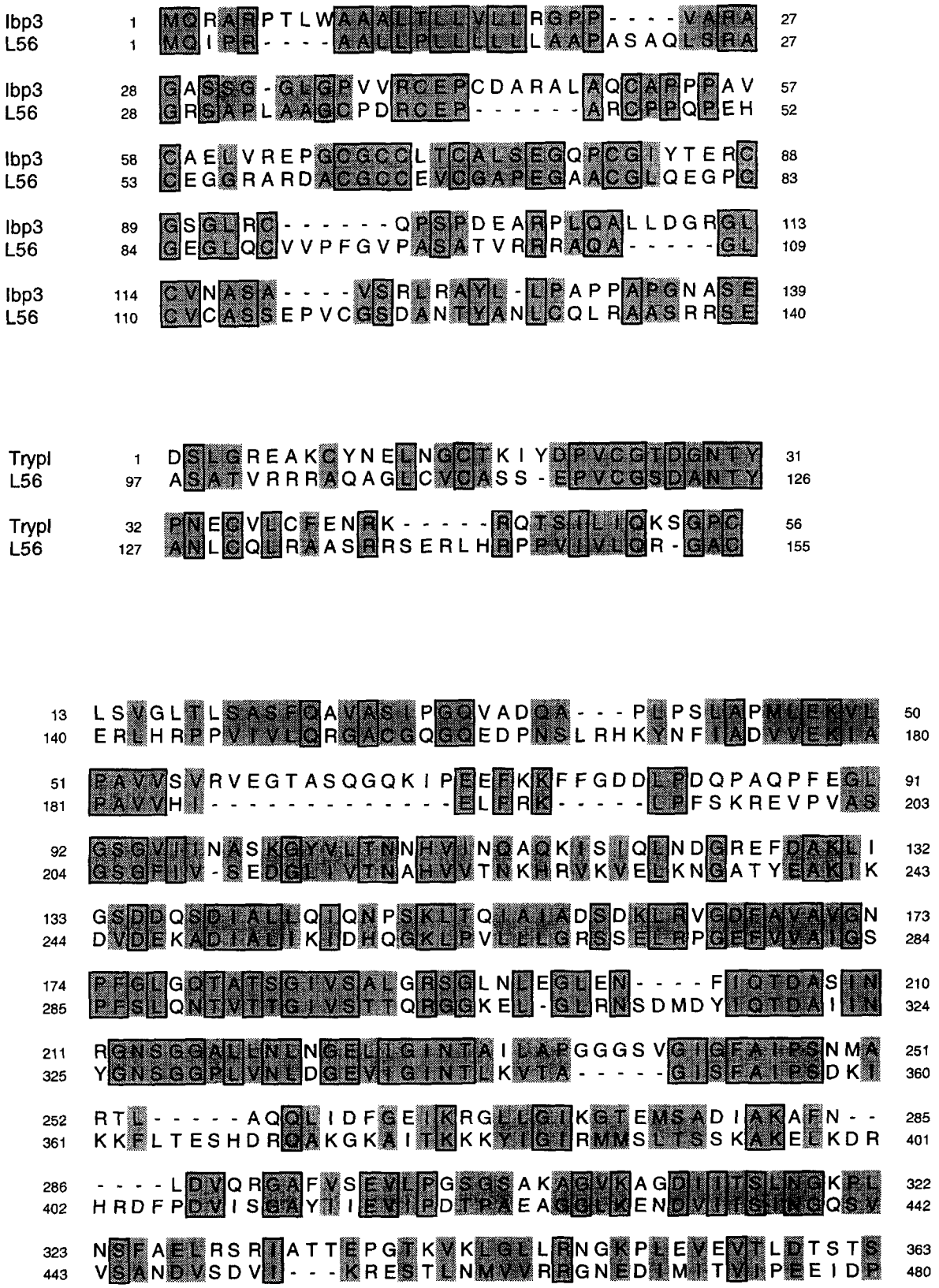


Fig. 3. Alignment of the amino acid sequences from the human L56 protein and the human IGFBP-3 (top), the human pancreatic trypsin inhibitor (middle) and the HtrA protease from *E. coli*. Identical residues are boxed; related residues are shaded.

## 4. Discussion

We have isolated the cDNA for a novel transformation-sensitive protein from human fibroblasts by subtractive cDNA cloning. This protein comprises four distinct domains, namely a typical signal peptide, a domain related to IGFBPs,


Fig. 4. Domain structure of the novel serine protease. S, signal peptide; IGFB, domain related to IGFBPs; KI, domain related to the Kazal type of protease inhibitors.

a short linker segment related to the Kazal type of proteinase inhibitors and a large domain related to serine proteases (Fig. 4).

There is good evidence that the L56 protein represents a functional serine protease. Its amino acid sequence shares 58% similarity with the HtrA/Do family of proteases from bacteria. The HtrA proteases are typical serine proteases which can be inhibited by DFP or other trypsin inhibitors [13–15]. In addition to the conserved serine residue, they depend on the presence of a histidine residue in their catalytic domain. The similarity observed between the human and the bacterial protein extends over the entire length of the HtrA sequence and is particularly well pronounced in the regions surrounding the active serine and histidine residues.

The high sequence conservation between two proteins from bacteria and humans is quite extraordinary. Moderate similarities have previously been noted with proteins involved in fundamental biological processes such as translation of proteins at ribosomes or translocation of nascent polypeptides across biological membranes [19]. In most cases, however, the similarities were confined to the tertiary rather than the primary structure of the polypeptides. In spite of this astonishing similarity, we are convinced that we have not simply isolated a bacterial contamination from our cDNA library because of three reasons. (1) We have isolated our cDNA clones from three independent human libraries. (2) Our clones hybridized to a distinct mRNA present in RNA preparations from eight different human tissues. (3) Our cDNA sequences overlap with several sequence entries stored in the EST database of human expressed sequence tags.

The HtrA protein enables bacteria to survive at elevated temperature [13–15]. It seems unlikely that the human protein may serve a similar function since its expression is not induced by heat shock. Some clues to its function can be deduced from its unique primary structure. In contrast to the bacterial protein, the human protein contains an additional domain related to IGF-binding proteins at its N-terminal end [11,12]. It is therefore conceivable that the L56 protease will interact with IGFI or IGFII. Consequently this domain could play a regulatory role in the activation or inhibition of the serine protease. This hypothesis is supported by the fact that the end of the IGF-binding domain and the beginning of the protease domain resemble the Kazal type of protease inhibitors. Kazal domains occur in serine protease inhibitors (e.g. pancreatic trypsin inhibitor) as well as in follistatin-like molecules. Recently, it has been proposed that follistatin-like domains of extracellular matrix proteins may bind and store growth factors such as TGF-β and PDGF [20]. Thus, the Kazal domain of the L56 protein might either associate with the active site of the serine protease or cooperate with the IGF-binding domain in the interaction with growth factors.

In the literature there is one class of proteases (of unknown sequence) that appears to be functionally very similar to the L56 protein, namely the proteases that cleave IGF-binding proteins [11,12]. It is well documented that the activity of IGFI and IGFII is modulated by IGF-binding proteins which control the availability of these growth factors for their receptors. Six IGFBPs occurring in most tissues and body fluids have been characterized so far. They bind IGFI and IGFII with similar affinity and appear to function as a reservoir for growth factors. The IGFs are liberated from the complexes by specific proteases. The molecular nature, in particular the primary structures of these proteases, has remained obscure so far. Limited information is available on the biological activity of proteases that specifically cleave IGFBP-2 to IGFBP-5 [21–26]. All these proteases belong to the family of the serine proteases since they are inhibited by DFP or other serine protease inhibitors. Among the four proteases, the one specific for IGFBP-4 is most similar to our L56 protein. It is secreted by human fibroblasts and its activity is also markedly decreased in SV40 transformed cells. IGFI and IGFII have been demonstrated to activate this protease and two mechanisms have been put forward to explain this activation process [21,22]. (1) The growth factors bind directly to the protease, thereby inducing proteolytic activity. This mechanism is consistent with our structural data indicating that the L56 protease contains an IGF-binding domain at its N-terminus. (2) Binding of IGFs may enhance the susceptibility of the IGFBP to proteolytic cleavage. This mechanism is not supported by our results, but cannot be ruled out either.

In summary, all our data suggest that we have cloned a novel serine protease related to the proteases for IGFBPs. To support our hypothesis it is inevitable now to extend our studies to the protein level. Unfortunately, all efforts to express our cDNA clones in a bacterial expression system have failed so far. The expression of the protein, however, is absolutely necessary to demonstrate a proteolytic activity of the L56 protein and to verify the modulation of this activity by IGFs. The expression of the protein should also enable us to investigate the role of the novel protease in transformed cells.

*Acknowledgements:* We thank Dr. K.H. Winterhalter for his interest and for continuous support. This study was funded by grants from the ETH Zurich (0-20-854-94) and the Swiss National Science Foundation (31-40337.94).

## References

[1] Schenker, T., Lach, C., Kessler, B., Calderara, S. and Trueb, B. (1994) J. Biol. Chem. 269, 25447–25453.
[2] Rowe, D.W., Moen, R.C., Davidson, J.M., Byers, P.H., Bornstein, P. and Palmiter, R.D. (1978) Biochemistry 17, 1581–1590.
[3] Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1987) Current Protocols in Molecular Biology, Greene, New York.
[4] Feinberg, A. and Vogelstein, B. (1983) Anal. Biochem. 132, 6–13.
[5] Duguid, J.R. and Dinauer, M.C. (1990) Nucleic Acids Res. 18, 2789–2792.
[6] Koller, E., Winterhalter, K.H. and Trueb, B. (1989) EMBO J. 8, 1073–1077.
[7] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA, 74, 5463–5467.
[8] Kozak, M. (1987) Nucleic Acids Res. 15, 8125–8148.
[9] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.
[10] Von Heijne, G. (1986) Nucleic Acids Res. 14, 4683–4690.
[11] Zapf, J. (1995) Eur. J. Endocrinol. 132, 645–654.
[12] Bach, L.A. and Rechler, M.M. (1995) Diabetes Rev. 3, 38–61.
[13] Lipinska, B., Sharma, S. and Georgopoulos, C. (1988) Nucleic Acids Res. 16, 10053–10067.
[14] Seol, J.H., Woo, S.K., Jung, E.M., Yoo, S.J., Lee, C.S., Kim, K., Tanaka, K., Ichihara, A., Ha, D.B. and Chung, C.H. (1991) Biochem. Biophys. Res. Commun. 176, 730–736.

[15] Skorko-Glonek, J., Wawrzynow, A., Krzewski, K., Kurpierz, K. and Lipinska, B. (1995) Gene 163, 47–52.
[16] Laskowski, M. and Kato, I. (1980) Annu. Rev. Biochem. 49, 593–626.
[17] Horii, A., Kobayashi, T., Tomita, N., Yamamoto, T., Fukushige, S., Murotsu, T., Ogawa, M., Mori, T. and Matsubara, K. (1987) Biochem. Biophys. Res. Commun. 149, 635–641.
[18] Shimasaki, S., Koga, M., Esch, F., Cooksey, K., Mercado, M., Koba, A., Ueno, N., Ying, S.-Y., Ling, N. and Guillemin, R. (1988) Proc. Natl. Acad. Sci. USA 85, 4218–4222.
[19] Dobberstein, B. (1994) Nature 367, 599–600.
[20] Patthy, L. and Nikolics, K. (1993) Trends Neurosci. 16, 76–81.
[21] Fowlkes, J. and Freemark, M. (1992) Endocrinology 131, 2071–2076.
[22] Conover, C.A., Kiefer, M.C. and Zapf, J. (1993) J. Clin. Invest. 91, 1129–1137.
[23] Nam, T.J., Busby, W.H. Jr. and Clemmons, D.R. (1994) Endocrinology 135, 1385–1391.
[24] Claussen, M., Zapf, J. and Braulke, T. (1994) Endocrinology 134, 1964–1966.
[25] Gockerman, A. and Clemmons, D.R. (1995) Circ. Res. 76, 514–521.
[26] Angelloz-Nicoud, P. and Binoux, M. (1995) Endocrinology 136, 5485–5492.