the FEBS Journal

Unconventional translation initiation of human trypsinogen 4 at a CUG codon with an N-terminal leucine

A possible means to regulate gene expression

Attila L. Németh[1,*], Péter Medveczky[1,*], Júlia Tóth[1], Erika Siklódi[1], Katalin Schlett[2], András Patthy[3], Miklós Palkovits[4], Judit Ovádi[5], Natália Tőkési[5], Péter Németh[6], László Szilágyi[1] and László Gráf[1,3]

1 Department of Biochemistry, Eötvös Loránd University, Budapest, Hungary
2 Departments of Physiology and Neurobiology, Eötvös Loránd University, Budapest, Hungary
3 Biotechnology Research Group of the Hungarian Academy of Sciences, Budapest, Hungary
4 Laboratory of Neuromorphology, Department of Anatomy, Semmelweis University, Budapest, Hungary
5 Institute of Enzymology, Hungarian Academy of Sciences, Budapest, Hungary
6 Institute of Immunology and Biotechnology, University of Pécs, Hungary

Keywords
brain; protein synthesis; PRSS3; serine protease; translation initiation; trypsin 4

Correspondence
L. Gráf, Department of Biochemistry, Eötvös Loránd University, Pázmány Péter s. 1/C, Budapest H-1117, Hungary
Fax: +36 1 3812172
Tel: +36 1 3812171
E-mail: graf@ludens.elte.hu

*These authors contributed equally to this work

(Received 7 December 2006, accepted 18 January 2007)

doi:10.1111/j.1742-4658.2007.05708.x

Summary

Chromosomal rearrangements apparently account for the presence of a primate-specific gene (protease serine 3) in chromosome 9. This gene encodes, as the result of alternative splicing, both mesotrypsinogen and trypsinogen 4. Whereas mesotrypsinogen is known to be a pancreatic protease, neither the chemical nature nor biological function of trypsinogen 4 has been explored previously. The trypsinogen 4 sequence contains two predicted translation initiation sites: an AUG site that codes for a 72-residue leader peptide on *Isoform A*, and a CUG site that codes for a 28-residue leader peptide on *Isoform B*. We report studies that provide evidence for the N-terminal amino acid sequence of trypsinogen 4 and the possible mechanism of expression of this protein in human brain and transiently transfected cells. We raised mAbs against a 28-amino acid synthetic peptide representing the leader sequence of *Isoform B* and against recombinant trypsin 4. By using these antibodies, we isolated and chemically identified trypsinogen 4 from extracts of both post mortem human brain and transiently transfected HeLa cells. Our results show that *Isoform B*, with a leucine N terminus, is the predominant (if not exclusive) form of the enzyme in post mortem human brain, but that both isoforms are expressed in transiently transfected cells. On the basis of our studies on the expression of a series of trypsinogen 4 constructs in two different cell lines, we propose that unconventional translation initiation at a CUG with a leucine, rather than a methionine, N terminus may serve as a means to regulate protein expression.

Recent genome programmes have explored an increasing number of new genes with unknown function. The estimated 35 000 human genes encode more than 10^5 expressed proteins as the result of various mechanisms, such as alternative promotion of transcription, alternative splicing of the transcripts and alternative translational initiation. Chromosome rearrangement can also serve as a source for evolutionary heterogeneity.

Abbreviations
GFP, green fluorescent protein; PRSS3, protease serine 3.

An interesting example of such a mechanism is the occurrence of a trypsinogen gene in a chromosome different from the original locus for the major forms of pancreatic trypsinogens in chromosome 7. Uniquely in primates, a series of chromosome translocations between chromosome 7, 11 and 9 led to the formation of the protease serine 3 (*PRSS3*) gene that encodes, as a result of alternative splicing, both mesotrypsinogen and trypsinogen 4 [1,2]. Structurally, human mesotrypsinogen and human trypsinogen 4 differ only in their N-terminal sequences. Whereas mesotrypsinogen has a typical signal sequence (splice *Isoform C* in the Swiss-Prot database), the two isoforms of human trypsinogen 4 have highly charged N-terminal leader sequences. The predicted longer form (splice *Isoform A* in the Swiss-Prot database) contains a 72-residue N-terminal leader peptide, whereas the shorter form (splice *Isoform B* in the Swiss-Prot database) contains a 28-residue N-terminal leader peptide (Fig. 1A). The translation initiation site for splice *Isoform A* is an AUG codon, whereas the deduced initiation site for splice *Isoform B* is a CUG codon, as first proposed by Wiegand and coworkers [3].

During translational initiation in eukaryotes, the complex consisting of the small (40S) ribosomal subunit, Met-tRNAi and eIF2-GTP, usually enters at the 5′ end of the mRNA and scans the 5′ untranslated region until reaching the first AUG codon. Approximately 10% of eukaryotic mRNAs are not translated from the first AUG codon if it is in an unfavorable sequence context; instead, translation starts from the second or another downstream AUG codon. Furthermore, there are several well-demonstrated cases, first discovered among viral genes, of translation starting from a non-AUG codon [4]. Initiation from non-AUG codons in eukaryotes now includes eukaryotic

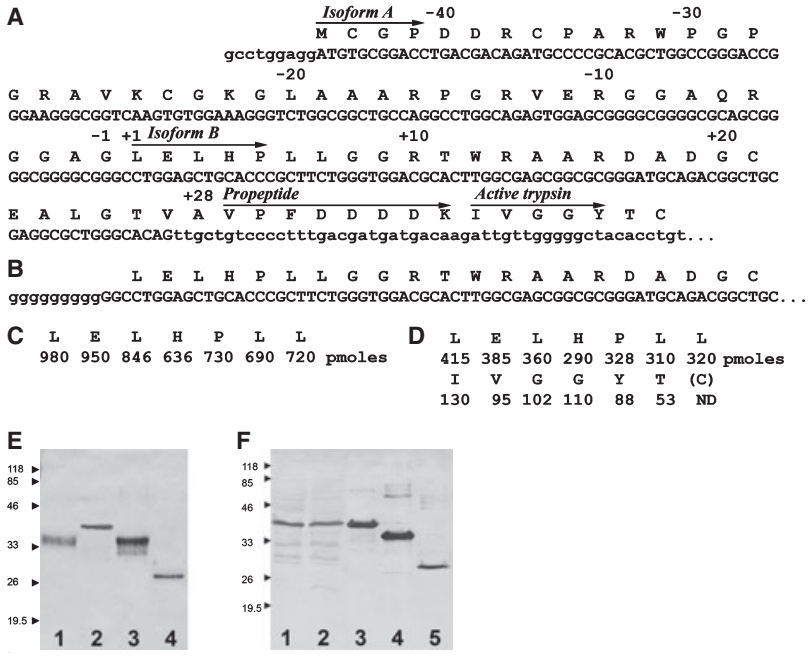

Fig. 1. (A) The partial nucleotide sequence of exon 1 (upper case), the 5′ end of exon 2 (lower case) and the deduced amino acid sequences of human trypsinogen 4. The numbering of the sequence is according to a previous publication [3]. (B) The 5′ end of the human trypsinogen 4 cDNA, as determined by 5′ rapid amplification of cDNA ends (5′-RACE) PCR. (C) N-terminal amino acid sequence of human trypsinogen 4 isolated from a 71.0 g occipital cortex sample by a mAb 1/B1 column. (D) N-terminal sequences of human trypsinogen 4 (isolated from a 71.5 g frontal cortex sample by a mAb 1/B1 column) after enterokinase digestion. The amount of N-terminal amino acids detected is indicated below the sequences. (E) Western blot (with mAb 1/B1) of human trypsinogen 4 isolated from an occipital cortex sample of human brain by a mAb p28 column. Protein molecular weight markers are indicated on the left side. Lane 1, trypsinogen 4, isolated from human brain, by mAb p28 immunoaffinity chromatography. Lane 2, recombinant tag-p72-trypsinogen 4. Lane 3, recombinant tag-p28-trypsinogen 4. Lane 4, recombinant trypsin 4. (F) Searching for proteolytic activity in human brain samples. Protein molecular weight markers are indicated on the left side. Lane 1, recombinant tag-p72 trypsinogen 4 incubated for 2 h at 37 °C with total brain extract. Lane 2, recombinant tag-p72 trypsinogen 4, incubated for 2 h at 37 °C with brain extract previously passed through a mAb 1/B1 immunoaffinity column. In both cases, trypsinogen 4 was recovered by mAb 1/B1 immunoaffinity chromatography. The blot was developed with mAb 1/B1. Lane 3, recombinant tag-p72-trypsinogen 4. Lane 4, recombinant tag-p28-trypsinogen 4. Lane 5, recombinant trypsin 4.

genes, such as proto-oncogenes, genes for transcription factor kinases and growth factors [5–10]. Originally, it was thought that regardless of the initiation codon used, methionine should be the initiating residue, and those few cases in which the reported initiator amino acid was not a methionine were limited to viral genes containing an internal ribosome entry site upstream of the nonconventional start codon [11,12]. Recently, however, Schwab and coworkers [13,14] have shown that leucine can also be translated as an initiator amino acid by using a CUG codon in short cryptic peptides in antigen-presenting cells. In this case, the leucine start does not depend on an internal ribosome entry site-like mRNA structure, and its translational efficiency is enhanced by a nucleotide context slightly different from the consensus Kozak sequence [15].

As a biochemical approach to determine the exact N-terminal sequence of trypsinogen 4, mAbs were raised against human trypsin 4 (obtained by enterokinase activation of recombinant human trypsinogen 4) and a synthetic fragment of the N-terminal 28-amino acid leader peptide of trypsinogen 4. We used these mAbs to purify and chemically identify trypsinogen 4 from human brain and from transiently transfected human cell lines. Our results show that *Isoform B* of trypsinogen 4, with a leucine N terminus, is the predominant (if not the exclusive) form of the enzyme in human brain, whereas both *Isoform B* and *Isoform A* can be extracted from transfected cells. Here we report, from amino acid sequencing, that although the N-terminal residue of the longer *Isoform A* isolated from the transiently transfected cells is methionine, as expected, the N-terminal amino acid of *Isoform B*, isolated from human brain and transiently transfected cells, is leucine.

Results

Determination of the 5′-terminal sequence of human trypsinogen 4 mRNA

In the original publication reporting the cloning of trypsinogen 4 cDNA, no ATG codon was found, even in the longest cDNA [3]. We repeated this experiment, with several brain samples, under slightly different conditions. We used C-tailing and an inosine containing abridged anchor primer, according to the 5′-rapid amplification of cDNA ends (5′-RACE System; Gibco-BRL), whereas in the original publication, G-tailing was used. Nevertheless, we obtained practically the same result, because the 5′ end of the cDNA was only four bases upstream from the putative CTG translation start codon (Fig. 1B) of *Isoform B*. Several attempts to isolate a cDNA containing the first upstream in-frame ATG codon were unsuccessful. It is interesting to note that the longest transcript deposited in the GenBank database (accession no. BI823946) also lacks the ATG (−44) start codon and starts from the third (G) nucleotide of the above-mentioned ATG codon.

Isolation and chemical identification of trypsinogen 4 from human brain

Different antihuman trypsinogen 4 mAbs were raised separately against recombinant human trypsin 4 (mAb 1/B1, mAb 6/B7) and the 28-amino acid leader peptide (mAb p28). Although all of these antibodies react with the leader peptide containing forms of trypsinogen, activated trypsin is only recognized by antibodies 1/B1 and 6/B7. Two antibodies – mAb 1/B1 and mAb p28 – were immobilized separately on cyanogen bromide activated Sepharose 4B. Pilot studies on the isolation of trypsinogen 4 from extracts of four different regions of human brain (see the Experimental procedures) showed that from all samples, and by both immunoaffinity columns, proteins of the same molecular size were isolated. The size and immunoreactivity of this protein corresponded to those of recombinant tag-p28 trypsinogen 4. This is illustrated in lane 1 of Fig. 1E, which shows a western blot (detected by mAb 1/B1) of trypsinogen 4, which was isolated via a mAb p28 column from a sample of human occipital cortex.

Affinity-purified proteins from three different brain regions were sequenced. In each case, we identified leucine as the only N-terminal amino acid of the isolated protein, irrespective of the specificity of the immobilized antibody (Fig. 1C). In order to prove the integrity of the isolated protein, human trypsinogen 4, isolated from a sample of the frontal cortex, was subjected to enterokinase digestion; N-terminal sequencing revealed the presence of both trypsin 4 with the N-terminal isoleucine and intact *Isoform B* starting with leucine (Fig. 1D).

Searching for a protease with potential processing activity in brain extract

To demonstrate the absence of protease activity capable of cleaving the trypsinogen leader sequence during the isolation process, we added recombinant *Isoform A* (tag-p72-trypsinogen 4) (Fig. 1F) to homogenized human brain samples and incubated them, without inhibitors, at 37 °C for 2 h. Then, the sam-

ples were centrifuged at 100 000 ***g*** for 20 min and the supernatants were subjected to immunoaffinity chromatography using immobilized mAb 1/B1. Western blotting of the eluted material clearly showed that the isolated protein was mostly intact (Fig. 1F, lane 1). Although faint bands indicated some proteolytic breakdown, no traces of a fragment, corresponding to *Isoform B* of trypsinogen 4, was found. Similar results were obtained when tag-p72-trypsinogen 4 was added to a brain homogenate that had been previously passed through an immunosorbent column (Fig. 1F, lane 2).

Isolation and chemical identification of trypsinogen 4 from transiently transfected HeLa cells

Transfection experiments, using several constructs (Fig. 2A,3A), were used in different cell lines. We transiently transfected HeLa cells with $p72^{M}T4$, $p72^{M}p28^{L(TTG)}T4$ and $p28^{L}T4$ constructs, and the cells were stained with antihuman trypsinogen 4 p28 (mAb p28; data not shown). Cell lysates from transfected cells were examined by western blotting (Fig. 2B) and subjected to immunoaffinity chromatography on a mAb 1/B1 column. Immunoreactive proteins, eluted in single fractions from the immunoaffinity columns, were analyzed by N-terminal amino acid sequencing (Fig. 2C,D). Western blots, together with the N-terminal amino acid sequences of trypsinogens isolated from 6×10^6 cells transfected with the $p72^{M}T4$ construct, indicated that both *Isoforms A* and *B* of human trypsinogen 4 were expressed. Cells transfected with the $p72^{M}p28^{L(TTG)}$ construct, however, expressed only the longer isoform (*Isoform A*) and no traces of the shorter isoform (*Isoform B*). By contrast, the expression of only *Isoform B* was detected in the cells transfected with the $p28^{L}T4$ construct (Fig. 2B) and leucine was identified as the sole N-terminal amino acid of this protein (Fig. 2D).

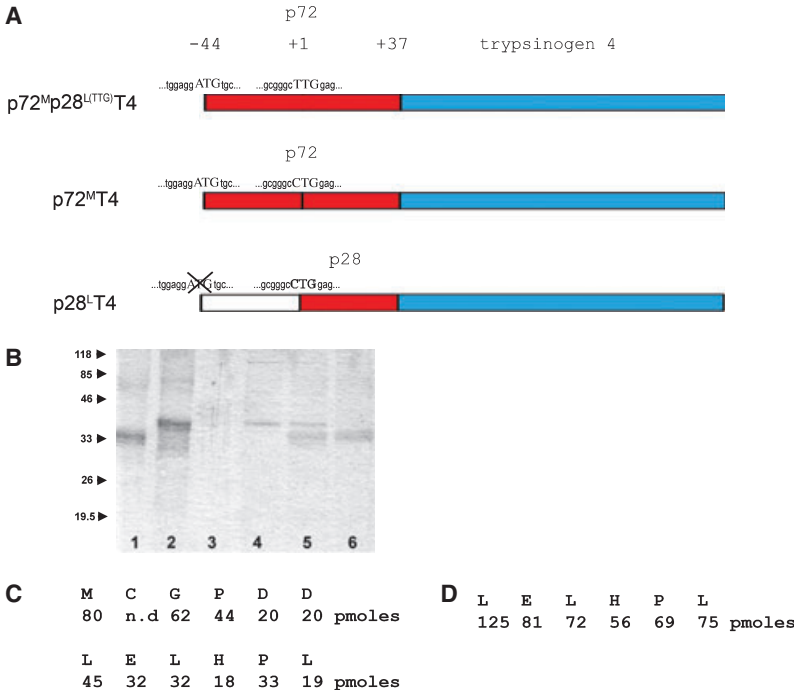

Fig. 2. (A) Schematic representation of the gene constructs used for expression of different isoforms of human trypsinogen 4 ($p72^{M}p28^{L(TTG)}T4$, $p72^{M}T4$, $p28^{L}T4$) in HeLa cells. The white box indicates nontranslated regions caused by the deletion of the AUG initiation codon, active trypsin 4 is represented by the blue box. (B) Western blot of human trypsinogen 4, detected by using mAb p28. Protein molecular weight markers are indicated on the left side. Lane 1, recombinant tag-p28-trypsinogen 4. Lane 2, recombinant tag-p72-trypsinogen 4. Lane 3, nontransfected, control HeLa cells 4. Lane 4, trypsinogen 4 detected from $p72^{M}p28^{L(TTG)}T4$ transfected HeLa cells. Lane 5, trypsinogen 4 detected from $p72^{M}T4$ transfected HeLa cells. Lane 6, trypsinogen 4 detected from $p28^{L}T4$ transfected HeLa cells. (C) N-terminal amino acid sequence of human trypsinogen 4 isolated from HeLa cells transiently transfected with $p72^{M}T4$ plasmid. (D) N-terminal amino acid sequence of human trypsinogen 4 isolated from HeLa cells transiently transfected with $p28^{L}T4$ plasmid. The amount of N-terminal amino acids detected is indicated below the sequences.

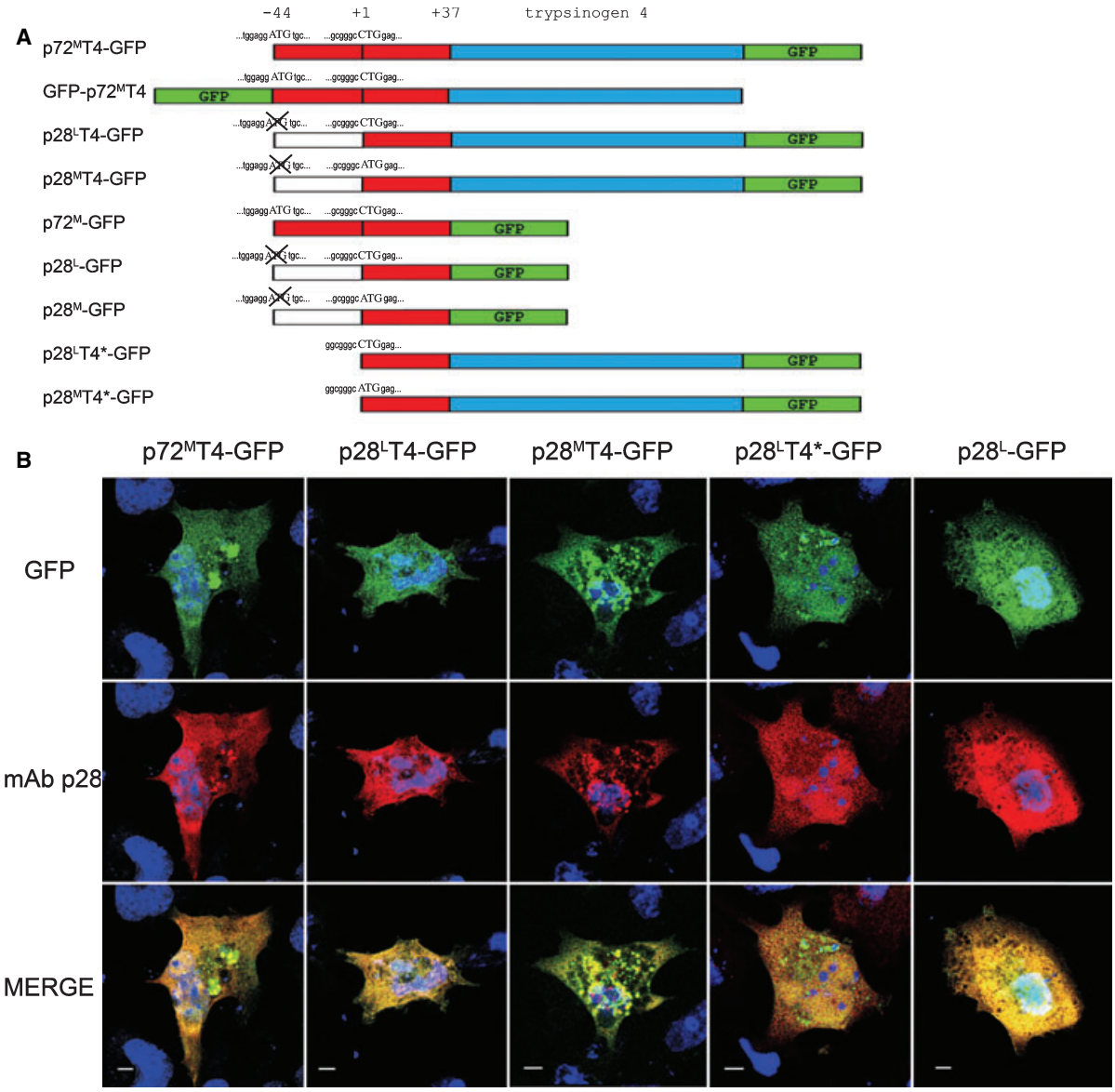

Fig. 3. (A) Green fluorescent protein (GFP)-fused plasmid constructs. Amino acid numbering and the indicated sequences are as described in Fig. 1A. The white box indicates nontranslated regions caused by the deletion of the initiation AUG codon; active trypsin 4 is represented by a blue box. p72^{M}T4-GFP, p72 form of trypsinogen 4 with a C-terminal GFP fusion protein; GFP-p72^{M}T4, p72 form of trypsinogen 4 with an N-terminal GFP fusion protein; p28^{L}T4-GFP, p28 form of trypsinogen 4 with a deleted ATG(–44) codon, CTG(+1) coding for a leucine initiator amino acid and a C-terminal GFP fusion protein; p28^{M}T4-GFP, p28 form of trypsinogen 4 with a deleted ATG(–44) codon, a mutated ATG(+1) coding for methionine as the initiator amino acid and a C-terminal GFP fusion protein; p72^{M}-GFP, p72 leader peptide with a C-terminal GFP fusion protein without the trypsinogen 4 catalytic domain; p28^{L}-GFP, p28 leader peptide with a deleted ATG(–44) codon, CTG(+1) coding for a leucine initiator amino acid and a C-terminal GFP fusion protein without the trypsinogen 4 catalytic domain; p28^{M}-GFP, p28 leader peptide with a deleted ATG(–44) codon, a mutated ATG(+1) coding for methionine as the initiator amino acid and a C-terminal GFP fusion protein without the trypsinogen 4 catalytic domain; p28^{L}T4*-GFP, p28 form of trypsinogen 4 with a deleted 5′-UTR sequence between ATG(–44) and GGG(–3), leaving only a 7 bp upstream sequence before CTG(+1) coding for a leucine initiator amino acid and the C-terminal GFP fusion protein; p28^{M}T4*-GFP, p28 form of trypsinogen 4 with a deleted 5′-UTR sequence between ATG(–44) and GGG(–3), leaving only a 7 bp upstream sequence before a mutated ATG(+1) coding for a methionine initiator amino acid and C-terminal GFP fusion protein. (B) Representative pictures from U87 human astroglioma cells, transiently transfected with different constructs, as indicated above the pictures. In each case, single optical sections taken by confocal microscopy are presented. GFP labelling (green) always colocalized with mAb p28 immunostaining (red). Cell nuclei were stained with Draq5 (blue). Depending on the constructs and the relative trypsinogen 4 expression levels, aggregation of GFP-labelled proteins were occasionally observed. Bars indicate 5 μm.

Transient transfection of human U87 glioblastoma cells with trypsinogen 4 GFP-fused constructs

We expressed human trypsinogen 4, fused with GFP reporter protein, in the U87 human glioblastoma cell line. Cells were transiently transfected with the constructs depicted in Fig. 3A and were immunostained, 24h post-transfection, with mAbs raised against the activated protease, mAb 1/B1 (data not shown), or against the 28-residue leader peptide, mAb p28 (Fig. 3B). The GFP reporter protein always colocalized with the immunostaining, with antibodies recognizing either the p28 leader sequence (mAb p28) or the protease domain (mAb 1/B1), indicating that trypsinogen 4 was localized mainly in an inactive form in the transfected cells. The observed localization of trypsinogen 4 was the same when GFP was fused to the N- or the C-terminal end of the molecule (data not shown).

We determined the number of cells showing GFP fluorescence by visual inspection of pictures taken from several microscopic fields. We consider the percentage of the GFP-positive cells as a measure of relative expression, because all experimental parameters, number of cells transfected, amount of plasmid, incubation time, etc., were essentially identical at each transfection (see the Experimental procedures).

In the case of constructs using an AUG initiation codon at site −44, the relative expression levels of *Isoform A* (expressed together with *Isoform B*; Fig. 2B, lane 5) were elevated compared with *Isoform B* with an AUG start codon (p72^{M}T4-GFP versus p28^{M} T4-GFP or p72^{M}-GFP versus p28^{M}-GFP; Table 1). The expression level of *Isoform B* with the wild-type CUG initiation codon was lower than that of *Isoform A* (expressed together with *Isoform B*) (p72^{M}T4-GFP versus p28^{L}T4-GFP; Table 1) and was dependent on the length of the wild-type upstream sequence preceding the CUG codon (p28^{L}T4-GFP versus p28^{L}T4*-GFP; Table 1). Nevertheless, protein expression was detected in all cases in which the CUG initiation codon was employed. Analogous constructs with the AUG initiation codon (p28^{M}T4-GFP versus p28^{M}T4*-GFP) did not show dependence on the length of the 5′-UTR region.

Table 1. Relative expression levels in U87 cells transiently transfected with green fluorescent protein (GFP)-fusion constructs. Plasmids used for transfections are as depicted in Fig. 3A, with the exception of pAcGFP-N1, indicating the cloning vector without any trypsinogen constructs. The percentage of GFP-positive cells was determined by comparing the number of cells showing GFP fluorescence with the total number of 4′-6-diamidino-2-phenylindole-positive cell nuclei in each microscopic field. Averages were calculated from three randomly chosen fields, and then the values of at least three independent transfection experiments were averaged (percentage ± standard deviation). GFP-positive cells were identified by visual inspection of pictures taken at identical exposure settings and were verified by inspecting the number of cells immunostained with mAb p28.

Construct	Percentage of GFP-positive cells
pAcGFP-N1	27.8 ± 5.0
p72^{M}T4-GFP	21.0 ± 2.5
p28^{L}T4-GFP	7.6 ± 1.8
p28^{M}T4-GFP	14.6 ± 2.9
p72^{M}-GFP	23.0 ± 3.9
p28^{L}-GFP	13.8 ± 2.4
p28^{M}-GFP	15.1 ± 0.9
p28^{L}T4*-GFP	2.0 ± 1.0
p28^{M}T4*-GFP	17.2 ± 3.8

Discussion

The occurrence of human trypsinogen 4 was first revealed in human brain [3], but later it was also found in human epithelial cells from prostate, colon and airway, and in several different tumors [16]. Trypsinogen 4 has two distinctive features: it contains an unusual mutation (Gly193 to Arg), responsible for its unique enzymatic properties [17–19]; and has an unconventional leader sequence (Fig. 1A).

By sequencing trypsinogen 4 samples isolated from human brain following only a short (2–5 h) post mortem delay, we found no traces of *Isoform A* beginning with Met(−44) (Fig. 1A). This result contrasts with the predictions of two isoforms based on the analysis of the *PRSS3* gene [3]. Instead, in each case we identified only the sequence corresponding to *Isoform B* beginning with leucine (Fig. 1C). We were unable to isolate *Isoform A* from any parts of the human brain. However, we cannot exclude the presence of the longer isoform in certain tissues; in addition, we found that *Isoform A* was expressed in cells transfected with the construct containing the full-length *Isoform A* gene (p72^{M}T4) (Fig. 2B). In accordance with previously published data [3,16], we were unable to detect any mRNA containing the upstream AUG codon for Met(−44), (Fig. 1B). As we were working with human brain samples, degradation of RNA owing to the post mortem delay is a possibility.

In theory, our finding that the zymogen form of trypsinogen 4 possesses a leucine N terminus has two explanations: either a hitherto-unknown proteolytic processing mechanism is responsible for cleaving the leader sequence, or leucine is, in fact, the initiator amino acid. The first possibility appeared to be

unlikely in the light of our *in vitro* experiments, according to which extracts from post mortem human brain samples did not convert the recombinant *Isoform A* (tag-p72-trypsinogen 4) to an *Isoform B*-like protein (Fig. 1F). Furthermore, expression of *Isoform B* with a leucine N terminus was also detected in HeLa cells transfected with a construct containing the full-length *Isoform A* gene ($p72^{M}T4$) (Fig. 2B,C). Importantly, we also detected N-terminal leucine in trypsinogen 4 isolated from cells transfected with a gene construct that lacks the upstream AUG codon for Met(−44) ($p28^{L}T4$) (Fig. 2B,D). All results, listed above, support our proposal that CUG is the initiation codon directing the incorporation of leucine, rather than methionine, into *Isoform B* of trypsinogen 4. The most convincing experimental evidence in favor of this hypothesis, however, came from a comparison of the expressed proteins from HeLa cells transfected with constructs $p72^{M}p28^{L(TTG)}T4$ and $p72^{M}T4$, respectively (Fig. 2B, lane 4 versus lane 5). In cells transfected with the former construct, in which CTG encoding Leu28 was replaced with TTG, another codon for Leu, only *Isoform A* was formed, whereas in cells transfected with the original construct containing the wild-type DNA sequence for trypsinogen 4, both *Isoforms A* and *B* were detected. The exclusive interpretation of this experiment is that the generation of *Isoform B* of trypsinogen 4 occurs at the level of translation and not post-translationally.

Recently, Schwab and coworkers [13,14] presented a similar case in eukaryotes: during the antigen presentation by Class I major histocompatibility complex molecules, the synthesis of a cryptic peptide was initiated with leucine by using CUG as the initiation codon without an upstream internal ribosomal entry site. The contextual sequence requirements for non-AUG initiation are not fully understood, and the critical nature of nucleotides that surround the non-AUG triplet is controversial [6]. The Kozak context of the *PRSS3 Isoform B* CUG initiation codon (GCGGGCcugG) resembles the optimal context of an AUG codon (GCCRCCaugG) [15]. In the experiment of Schwab and coworkers, however, the optimal context of CUG initiation was TCCACCcugG, different from that of *PRSS3*. In the present study, shortening of the wild-type 5′-UTR region upstream of the CUG initiation codon in the $p28^{L}T4$*-GFP construct to only seven nucleotides led to significantly decreased, but not abolished, expression of the GFP-fused enzyme (Table 1). A decrease in the expression level was not observed when the 5′-UTR region was removed preceding the AUG initiation codon. This finding indicates the important, but not exclusive, role of the 5′-UTR region beyond the Kozak region in translation initiation from the CUG initiation codon. There is an ≈30-nucleotide-long GC-rich region preceding the CUG start codon that might have a role in recognition of the suboptimal translational initiation site. Irrespective of the length of the 5′-UTR region, we found that the relative expression levels were lower in both U87 and HeLa cells when leucine was used as the initiator amino acid (Table 1); this suggests that CUG translational initiation may control the expression level. Thus, one is tempted to speculate that under physiological conditions, the translational initiation of human trypsinogen 4 with a Leu N terminus may function to keep the expression of the protein at a relatively low level.

The first exon of trypsinogen 4 is derived from the noncoding first exon of *LOC120224*, a chromosome-11 gene [2]. *LOC120224* codes for a widely conserved transmembrane protein of unknown function. The missing upstream AUG initiation codon in the *LOC120224* transmembrane protein does not necessarily mean that translation starts from a downstream AUG, as predicted by genome and mRNA analysis, but raises the possibility that the translated form may have used a CUG start codon with an N-terminal leucine amino acid. Our present study indicates that non-AUG translation initiation may be operable more often than anticipated. This may have a great impact on the analysis of genes on the basis of genome sequencing.

It has been suggested that human trypsinogen 4 plays functional roles in human cancer and metastasis [20–22], amyloid fragment production in aged astrocytes [23], or in epithelial tissues modulating protease-activated receptor-2 and -4 activity [16,24]. More recently, in an *in vitro* study we have shown that recombinant human trypsin 4 selectively clips residues 80–97 from human myelin basic protein [25], indicating a possible link to the development of multiple sclerosis [26,27]. It is a possibility to consider, that a significant release and activation of trypsinogen 4 would occur only under pathological conditions when the trypsinogen 4-expressing cells undergo damage in the human brain. Until clinical experiments support or deny this hypothesis, the biological function of human trypsinogen 4 remains in doubt.

Experimental procedures

Human brain samples

Tissue samples were obtained from the Human Brain Tissue Bank, Budapest. Brains were removed from the

skull, rapidly frozen on dry ice and stored at −70 °C until dissection. Samples from four different human brains were used for human trypsinogen 4 isolation in separate experiments. These were as follows: 77.3 and 76.9 g samples of the occipital and temporal cortex, respectively, from an 81-year-old woman, a 71.5 g sample of frontal cortex from an 83-year-old man and a 71.0 g sample of occipital cortex from a 85-year-old man with a short (2–5 h) post mortem delay.

RNA isolation, reverse transcription and 5′-RACE

Total RNA was isolated from 30 to 100 mg of human frontal cortex tissue samples, using TRI Reagent (Sigma, Budapest, Hungary) according to the manufacturer's instructions. First-strand cDNA was synthesized by priming with gene-specific primer 1 (5′- GGCTTTACACTCAGCCTGGG-3′). Reverse transcription was performed by the RevertAid H Minus First Strand cDNA Synthesis Kit (Fermentas, Vilnius, Lithuania). The synthesized cDNA was subjected to homopolymeric tailing to create a binding site for the abridged anchor primer on the 3′ end of the cDNA. PCR amplification of the C-tailed cDNA was performed with the abridged anchor primer (5′-GGCCACGCGTCGACTAGTACGGGII GGGIIGGGIIG-3′, 5′-RACE System; Gibco-BRL, Grand Island, NY, USA) and a nested gene-specific primer 2 (5′-GGAGAGTTTGATCAGCATGATGTC-3′) using *Taq* polymerase. The PCR products were cloned into pBluescript vector, via TA ligation, and then sequenced.

Cloning and expression of the *PRSS3* gene

The gene sequence coding for *Isoform B* of human trypsinogen 4 was cloned from human brain cDNA with the primers FP1 (5′-CGCATATGGAGCTGCACCCGCTTC TG-3′) and RP1 (5′-GACTGCAGGGATCCCGGGGG CTTTAGC-3′). The PCR product was subcloned into vector pET-15b (Novagen, Madison, WI, USA). This construct resulted in a fusion protein with a histidine tag at its N terminus (tag-p28-trypsinogen 4). As the mRNA corresponding to the *Isoform A* of human trypsinogen 4 could not be found with the 5′-RACE technique, the DNA sequence encoding the first exon was PCR amplified from genomic DNA using the forward and reverse primers FP2 (5′-CTGCATATGTGCGGACCTGACGACAGATGC-3′) and RP2 (5′-CTGCAGCAACTGTGCCCAGCGCCTCGC-3′), and then fused with the cloned *Isoform B* coding sequence using the naturally occurring *Alw*NI site. The gene was subcloned into the expression vector pET-15b (tag-p72-trypsinogen 4).

To express the splice *Isoforms A* and *B* of human trypsinogen 4 in *Escherichia coli*, 500 mL cultures of Rosetta™ (DE3)pLysS cells (Novagen), transformed with the constructs, were grown at 37 °C in Luria–Bertani medium containing ampicillin. Cells were harvested, and the isolation and refolding of the inclusion bodies were carried out, as described previously [17,28], with minor modification.

The full-length *Isoform A* gene was used as template in a PCR reaction, with Hu4-F1 (5′-GCGCAAGCTTCCTGGA GGATGTGCGGACCTGACGAC-3′) and Hu4-R1 (5′-GC CTGGATCCGAGCTGTTGGCAGCGATGG-3′) primers, for subcloning the *PRSS3* gene into pcDNA3 (Invitrogen, Carlsbad, CA, USA), pAcGFP1-N1 and pAcGFP1-C1 (BD Biosciences, Clontech, Mountain View, CA, USA) vectors at *Hin*dIII and *Bam*HI sites, resulting in $p72^{M}T4$, $p72^{M}T4$-GFP and GFP-$p72^{M}T4$ constructs, respectively. Hu4-F2 (5′-GCG CAAGCTTGCGGACCTGACGACAGATGC-3′) and Hu4R1 primers were used to amplify the *PRSS3* gene sequence lacking the initial ATG codon, which was subcloned into pcDNA3 and pAcGFP1-N1 vectors, resulting in $p28^{L}T4$ and $p28^{L}T4$-GFP constructs, respectively. The mutation Leu1 to Met was introduced by the megaprimer PCR reaction in $p28^{L}T4$-GFP, by using the mutagenic primer p28ATG (5′-GAGCTCCATGCCCGCCC-3′). The resulting construct ($p28^{M}T4$-GFP) lacked the initial ATG codon of *Isoform A*, and the initial CTG codon of *Isoform B* was mutated to ATG. The corresponding constructs were made lacking the trypsin catalytic domain using p72GFP primer (5′-GTCGGATCCTTGTCATCATCGTCAAAGG-3′), resulting in $p72^{M}$-GFP, $p28^{L}$-GFP and $p28^{M}$-GFP constructs in the pAcGFP-N1 expression vector. The silent mutation of the CTG initiation codon to TTG was introduced by the mutagenic primer, p28TTG (5′-GTGCAG CTCCAAGCCCGCCCC-3′), by using the megaprimer PCR method, and the gene harboring the mutation was cloned into the pcDNA3 vector. This construct is designated as $p72^{M}p28^{L(TTG)}T4$. The $p28^{L}T4$*-GFP and $p28^{M}T4$*-GFP constructs were made by removing the 5′-UTR region of *Isoform B* (that is part of the coding region of *Isoform A*) from the $p28^{M}T4$-GFP and $p28^{L}T4$-GFP constructs, using the p28Hind primer (5′-CGCAGCGAAGCTTGGCGGGC-3′). Only a seven-nucleotide-long sequence was left before the putative CUG initiator codon (underlined) to ensure the wild-type Kozak sequence was maintained.

Antibodies

Recombinant human trypsin 4, and the synthetic 28-amino acid leader peptide, were used to immunize female BALB/c mice (Charles River Laboratories, Raleigh, NC, USA). Antigen-specific B lymphocytes, prepared from the spleens of high-responder animals, were preselected using a method developed in our laboratory and published previously [29]. The fusion partner was the Sp-2/0 Ag14 (ATCC, Manassas, VA, USA) mouse myeloma cell line, and the hybridoma cells were prepared and cloned as described previously [30]. The selected clones of hybridomas were cultured in DMEM (Sigma) containing 10% fetal bovine serum (Gibco-BRL). The mass production of antibodies was performed by hybridoma fermentation (Harvest

Mouse; Serotec, Oxford, UK). mAbs were purified by Protein-G based Sepharose 4B affinity chromatography (Pharmacia, Upsalla, Sweden) and then concentrated by Amicon ultrafiltration (Millipore, Billerica, MA, USA). Different antigens were used to characterize immunoserologically the mAbs raised against the protease domain (mAb 1/B1 and mAb 6/B7) and the 28-amino acid leader peptide (mAb p28) of human trypsinogen 4 (data not shown).

Immunoaffinity media preparation

mAbs 1/B1 and p28 were immobilized separately on cyanogen bromide-activated Sepharose 4B (Pharmacia). Antibodies were dialyzed against the coupling buffer (0.1 M $NaHCO_3$, pH 8.3, containing 0.5 M NaCl) and mixed with the resin. Coupling efficiency proved to be > 90%. The stability of coupling was tested by washing the resin with the elution buffer of the chromatography (50 mM HCl). Under these conditions, the coupled antibody was not eluted from the column.

Isolation of trypsinogen 4 from human brain

Samples were homogenized in five volumes of NaCl/P_i, pH 7.4 and the homogenate was centrifuged at 100 000 ***g*** for 20 min. The pellet was then rehomogenized in five volumes of NaCl/P_i, pH 7.4, containing 1% (v/v) Tween-20, 1 mM phenylmethanesulfonyl fluoride, 1 mM cystatin, 1 mM leupeptin and 1 mM EDTA as protease inhibitors. After centrifugation of the homogenate at 100 000 ***g*** for 20 min, the supernatant was immediately used for immunoaffinity chromatography. The total protein concentration was determined by the bicinchoninic acid method (Sigma).

Cell lines and transfections

HeLa and U87 were used for transfection assays. HeLa cells were grown in DMEM/F-12 medium supplemented with 10% fetal bovine serum, 1 mM sodium pyruvate, 100 $U \cdot mL^{-1}$ of streptomycin and 100 $\mu g \cdot mL^{-1}$ of penicillin, whereas U87 human glioblastoma cells were cultured in DMEM containing 10% fetal bovine serum, 4500 $mg \cdot mL^{-1}$ of glucose and 40 $\mu g \cdot mL^{-1}$ of gentamycin (all Sigma), in a humidified 37 °C incubator with 5% CO_2. For transfection assays, 10^5 HeLa or U87 cells were seeded onto poly L-lysine (Sigma)-coated 13 mm diameter glass coverslips in 24-well plates, transfected either with Fugene 6 (Roche, Mannheim, Germany; HeLa cells) or Lipofectamine 2000 (Gibco; U87 cells) transfection reagents, according to the manufacturers' instructions, and were processed 24 h after transfection. For immunofluorescence cell studies in HeLa or U87 cells, 250 ng or 1 μg of DNA was used, respectively.

Isolation of trypsinogen 4 from HeLa cells

A total of 6×10^6 HeLa cells, seeded into 60-mm Petri dishes and transfected with $p72^{M}T4$ or $p28^{L}T4$ plasmid constructs, were used for trypsinogen isolation in separate parallel experiments. Cells were homogenized in 5 mL of lysis buffer (1% Tween-20, 50 mM Tris/HCl, 150 mM NaCl, pH 8, containing 1 mM phenylmethanesulfonyl fluoride, 1 mM benzamidine, 1 mM cystatine, 1 mM leupeptin and 1 mM EDTA). Homogenized samples were incubated for 1 h on ice and then, after centrifugation (14 000 ***g***, 20 min, 4 °C) the supernatants were used for immunoaffinity chromatography.

Immunoaffinity chromatography

Supernatant fractions of human brain and transfected HeLa cell homogenates were passed through the 1.5–0.5 mL immunoaffinity column. The columns were washed three times with 10 mL of NaCl/P_i, pH 7.4, containing 1% Tween-20 and 150 mM NaCl. Elution was carried out with 50 mM HCl, and 1–0.5 mL fractions were collected. Fractions were screened for trypsin immunoreactivity by gel electrophoresis and western blotting.

Western blot analysis

Proteins were separated by SDS–PAGE (15% gel) and were transferred to nitrocellulose membranes (Pharmacia). Blots were blocked in NaCl/Tris-Tween buffer (20 mM Tris, pH 8.0, 150 mM NaCl, 0.05% Tween-20) at room temperature and then incubated with mAb 1/B1 or p28 (1 : 3000) overnight at 4 °C. After being washed for 3×5 min with NaCl/Tris-Tween, blots were incubated with biotin-conjugated anti-mouse secondary serum (B-7151; Sigma), at a 1 : 5000 dilution, in NaCl/Tris-Tween, for 1 h at room temperature. After washing, the blots were incubated with ExtrAvidin peroxidase conjugate (E-2886; Sigma), at a 1 : 3000 dilution, for 1 h at room temperature followed by a 5 min wash in NaCl/Tris. The color development reaction was carried out using diaminobenzidine (Sigma), in NaCl/Tris, in the presence of 0.4 mM $NiCl_2$ and 1.25% H_2O_2.

Amino acid sequence determination

Fractions containing human trypsinogen 4 immunoreactivity were freeze-dried, dissolved in 10 mM NH_4HCO_3 and subjected to N-terminal amino acid analysis in a Procise sequencer (ABI 494; Applied Biosystems, Foster City, CA, USA) employing an EDMAN DEGRADATION SEQUENATOR program.

Immunostaining

Transfected HeLa cells were fixed with cold methanol for 15 min, or with 4% paraformaldehyde for 20 min, at

room temperature. The staining patterns were similar with the different fixatives used. The cells were washed in NaCl/P_i containing 0.1% Triton-X-100, then blocked for 30 min in NaCl/P_i containing 0.1% Triton-X-100 and 5% goat serum (Sigma). Subsequently, the cells were stained with mAb p28 (1 : 1000), followed by fluorescein isothiocyanate or Texas-Red conjugated anti-mouse sera (Jackson Laboratories, Bar Harbor, ME, USA), all diluted in 0.1% Triton-X-100 containing 5% goat serum. After washing in NaCl/P_i, nuclei were counterstained with 4′,6-diamidino-2-phenylindole, and the coverslips were mounted on Crystal Mount medium (Biomeda Corp., Foster City, CA, USA). Transfected U87 cells were fixed by 4% paraformaldehyde in NaCl/P_i (pH 7.4) for 20 min at room temperature, permeabilized with 0.1% Triton-X-100 for 5 min and blocked by 2% BSA-NaCl/P_i-0.1% Na azide (blocking solution) for 1 h at room temperature. Cells were incubated with mAb p28 (1 : 1000), at 4 °C overnight, followed by anti-mouse biotin (1 : 1000; goat IgG; Jackson Laboratories), for 1.5 h at room temperature, and Extravidin-TRITC (1 : 1000; Sigma) for 1 h at room temperature. All antibodies were diluted in blocking solution. Nuclei were labeled by incubation with 4′,6-diamidino-2-phenylindole or DRAQ5 (fluorescent dies) for 10 min at room temperature (1 : 2000; BioStatus Ltd, Shepshed, UK), then the coverslips were washed and mounted using Mowiol 4.88 (Polysciences Gmbh, Eppelheim, Germany).

Fluorescence microscopy

Confocal microscopy was carried out by a 488 nm Argon laser, and by 546 nm and 633 nm Helium-Neon lasers, using the ×60 oil-immersion objective of an Olympus IX71 microscope equipped with FLUOVIEW500 software (Olympus, Tokyo, Japan). The sequential scanning mode was used during recordings to exclude potential cross-talk completely between different channels. For wide-field observations in HeLa cells, a Leica DMLS microscope (Leica Microsystems, Wetzlar, Germany), equipped with appropriate filter sets and a cooled CCD camera (SPOT; Digital Instruments, Buffalo, NY, USA), and a C-PLAN ×100 immersion objective, was used, and digital images were recorded with SPOT 4.0.2. To estimate the relative expression levels for different GFP-tagged constructs in U87 cells, the number of GFP-positive cells was compared with the total number of 4′-6-diamidino-2-phenylindole-positive cell nuclei. Digital images from at least three randomly chosen microscopic fields from transfected U87 cells were recorded with a ×20 objective on an Olympus BX-51 microscope fitted with a FLUOVIEW2 camera, and the numbers obtained were averaged. These values were determined in three independent transfection experiments, and the averages are shown in Table 1.

Acknowledgements

This study was supported by Hungarian Research Grants OTKA to L. Gráf (T047154, TS 049812), L. Szilágyi (T037568) and J. Gergely (TS 0044711).

References

1 Roach JC, Wang K, Gan L & Hood L (1997) The molecular evolution of the vertebrate trypsinogens. *J Mol Evol* **45**, 640–652.
2 Rowen L, Williams E, Glusman G, Linardopoulou E, Friedman C, Ahearn ME, Seto J, Boysen C, Qin S, Wang K *et al.* (2005) Interchromosomal segmental duplications explain the unusual structure of PRSS3, the gene for an inhibitor-resistant trypsinogen. *Mol Biol Evol* **22**, 1712–1720.
3 Wiegand U, Corbach S, Minn A, Kang J & Müller-Hill B (1993) Cloning of the cDNA encoding human brain trypsinogen and characterization of its product. *Gene* **136**, 167–175.
4 Becerra SP, Rose JA, Hardy M, Baroudy BM & Anderson CW (1985) Direct mapping of adeno-associated virus capsid proteins B and C: a possible ACG initiation codon. *Proc Natl Acad Sci USA* **82**, 7919–7923.
5 Acland P, Dixon M, Peters G & Dickson C (1990) Subcellular fate of the int-2 oncoprotein is determined by choice of initiation codon. *Nature* **343**, 662–665.
6 Kozak M (1997) Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not generally affected by the nucleotides in positions +5 and +6. *EMBO J* **16**, 2482–2492.
7 Peabody DS (1989) Translation initiation at non-AUG triplets in mammalian cells. *J Biol Chem* **264**, 5031–5035.
8 Hann SR, King MW, Bentley DL, Anderson CW & Eisenman RN (1988) A non-AUG translational initiation in c-myc exon 1 generates an N-terminally distinct protein whose synthesis is disrupted in Burkitt's lymphomas. *Cell* **52**, 185–195.
9 Saris CJ, Domen J & Berns A (1991) The pim-1 oncogene encodes two related protein-serine/threonine kinases by alternative initiation at AUG and CUG. *EMBO J* **10**, 655–664.
10 Touriol C, Bornes S, Bonnal S, Audigier S, Prats H, Prats AC & Vagner S (2003) Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons. *Biol Cell* **95**, 169–178.
11 Sasaki J & Nakashima N (2000) Methionine-independent initiation of translation in the capsid protein of an insect RNA virus. *Proc Natl Acad Sci USA* **97**, 1512–1515.
12 Wilson JE, Pestova TV, Hellen CU & Sarnow P (2000) Initiation of protein synthesis from the A site of the ribosome. *Cell* **102**, 511–520.

13 Schwab SR, Li KC, Kang C & Shastri N (2003) Constitutive display of cryptic translation products by MHC class I molecules. *Science* **301**, 1367–1371.
14 Schwab SR, Shugart JA, Horng T, Malarkannan S & Shastri N (2004) Unanticipated antigens: translation initiation at CUG with leucine. *PLoS Biol* **2**, e366.
15 Kozak M (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283–292.
16 Cottrell GS, Amadesi S, Grady EF & Bunnett NW (2004) Trypsin IV, a novel agonist of protease-activated receptors 2 and 4. *J Biol Chem* **279**, 13532–13539.
17 Katona G, Berglund GI, Hajdú J, Gráf L & Szilágyi L (2002) Crystal structure reveals basis for the inhibitor resistance of human brain trypsin. *J Mol Biol* **315**, 1209–1218.
18 Szmola R, Kukor Z & Sahin-Tóth M (2003) Human mesotrypsin is a unique digestive protease specialized for the degradation of trypsin inhibitors. *J Biol Chem* **278**, 48580–48589.
19 Tóth J, Gombos L, Simon Z, Medveczky P, Szilágyi L, Gráf L & Málnási-Csizmadia A (2006) Thermodynamic analysis reveals structural rearrangement during the acylation step in human trypsin 4 on 4-methylumbelliferyl 4-guanidinobenzoate substrate analogue. *J Biol Chem* **281**, 12596–12602.
20 Diederichs S, Bulk E, Steffen B, Ji P, Tickenbrock L, Lang K, Zanker KS, Metzger R, Schneider PM, Gerke V *et al*. (2004) S100 family members and trypsinogens are predictors of distant metastasis and survival in early-stage non-small cell lung cancer. *Cancer Res* **64**, 5564–5569.
21 Marsit CJ, Karagas MR, Danaee H, Liu M, Andrew A, Schned A, Nelson HH & Kelsey KT (2006) Carcinogen exposure and gene promoter hypermethylation in bladder cancer. *Carcinogenesis* **27**, 112–116.
22 Yamashita K, Mimori K, Inoue H, Mori M & Sidransky D (2003) A tumor-suppressive role for trypsin in human cancer progression. *Cancer Res* **63**, 6575–6578.
23 Minn A, Schubert M, Neiss WF & Müller-Hill B (1998) Enhanced GFAP expression in astrocytes of transgenic mice expressing the human brain-specific trypsinogen IV. *Glia* **22**, 338–347.
24 Grishina Z, Ostrowska E, Halangk W, Sahin-Tóth M & Reiser G (2005) Activity of recombinant trypsin isoforms on human proteinase-activated receptors (PAR): mesotrypsin cannot activate epithelial PAR-1-2, but weakly activates brain PAR-1. *Br J Pharmacol* **146**, 990–999.
25 Medveczky P, Antal J, Patthy A, Kékesi K, Juhász G, Szilágyi L & Gráf L (2006) Myelin basic protein, an autoantigen in multiple sclerosis, is selectively processed by human trypsin 4. *FEBS Lett* **580**, 545–552.
26 O'Connor KC, Bar-Or A & Hafler DA (2001) The neuroimmunology of multiple sclerosis: possible roles of T and B lymphocytes in immunopathogenesis. *J Clin Immunol* **21**, 81–92.
27 Steinman L (1996) Multiple sclerosis: a coordinated immunological attack against myelin in the central nervous system. *Cell* **85**, 299–302.
28 Szilágyi L, Kénesi E, Katona G, Kaslik G, Juhász G & Gráf L (2001) Comparative in vitro studies on native and recombinant human cationic trypsins. Cathepsin B is a possible pathological activator of trypsinogen in pancreatitis. *J Biol Chem* **276**, 24574–24580.
29 Najbauer J, Tigyi GJ & Németh P (1986) Antigen-specific cell adherence assay: a new method for separation of antigen-specific hybridoma cells. *Hybridoma* **5**, 361–370.
30 Kohler G & Milstein C (1975) Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256**, 495–497.