Enzyme from the medicinal leech (*Hirudo medicinalis*) that specifically splits endo-ε(-γ-Glu)-Lys isopeptide bonds: cDNA cloning and protein primary structure

Arkady Fradkov[a], Sergey Berezhnoy[a], Ekaterina Barsova[a], Lyudmila Zavalova[a], Sergey Lukyanov[a], Isolda Baskova[b], Eugene D. Sverdlov[c,*]

[a]*Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, 117871 Moscow, Russian Federation*
[b]*Biological Faculty, Moscow State University, 119899 Moscow, Russian Federation*
[c]*Institute of Molecular Genetics, Russian Academy of Sciences, 123182 Moscow, Russian Federation*

Received 24 April 1996; revised version received 27 May 1996

Abstract Earlier we detected a novel enzymatic activity in salivary gland secretion of the medicinal leech, splitting isopeptide bonds between the glutamine γ-carboxamide and lysine ε-amino group. This activity is due to destabilase. We described its partial amino acid sequence and sequences of two closely related cDNAs, but none of them perfectly matched the protein isolated. Here we report the isolation and sequence peculiarities of the third cDNA of the family as well as the complete sequence of the destabilase protein. The inferred mature protein product of this cDNA matches the independently determined destabilase protein sequence. It contains 115 amino acid residues including 14 highly conserved Cys residues and is formed from a precursor containing specific leader peptide.

Key words: Protein sequence; Gene family; Gene characterization; ε-(γ-Glu)-Lys; Isopeptidase; Medicinal leech

1. Introduction

Isopeptide bonds between side chains of Gln and Lys, ε-(γ-Glu)-Lys, are formed by specific transglutaminases and are important in vital processes such as the final steps of blood coagulation [1–4]. The presence of transglutaminases in diverse organisms, organs, tissues and body fluids has long suggested that natural isopeptidases that antagonize the action of transglutaminases might also exist. Recently we described such an enzyme, an endo-ε-(γ-Glu)-Lys isopeptidase also called destabilase. The enzyme specifically splits isopeptide bonds in diisopeptides [5], between fibrin chains, but leaves polypeptide chains intact [6,7]. It dissolves stabilized fibrin in vitro [8] and stimulates thrombolysis in experimental animals [9]. Due to its ability to dissolve stabilized fibrin in vitro [8] and thus to reverse the critical final step in blood coagulation [9], destabilase is a potential thrombolytic agent. In our previous work [10] the corresponding protein was purified, its partial amino acid sequence determined, and two cDNAs encoding closely related but still different proteins were cloned and sequenced, demonstrating the existence of the gene family for destabilase related proteins. The structural homology correlated with functional similarity of the destabilase and the product of one of the cDNAs cloned into a baculovirus expression vector and expressed in insect cells. However, the polypeptides encoded by these two cDNAs were only partially homologous to the independently isolated protein [10]. Here we describe the identification and sequence of a cDNA coding for the third member of the family with precisely the same sequence as the destabilase protein. We also report the sequences of the peptides forming together the complete sequence of the destabilase protein.

2. Materials and methods

2.1. General techniques

The oligonucleotides used as adapters and/or PCR amplification primers are listed in Table 1.

If not stated otherwise, standard procedures [11] were used. Total RNA was isolated from adult leeches using a guanidine isothiocyanate procedure [12]. The cDNA synthesis was carried out with a Marathon cDNA amplification kit (Clontech Laboratories, USA) according to the manufacturer's instructions.

The oligonucleotide duplex AP1/AP2 was ligated to the double stranded cDNA and the adapter ligated library was subsequently used for the PCR amplifications as described below.

2.2. PCR amplification of the destabilase cDNA fragments

PCR amplification of the adapter ligated cDNA library was carried out with DsA and DsB primers corresponding to amino acid residues 8–15 and 35–41 of the destabilase protein (Fig. 1B). The KlenTaq DNA polymerase (AB Peptides, USA) was used for amplification under the following conditions: 25 cycles, 94°C, 5 s; 50°C, 20 s; 72°C, 30 s. PCR amplifications were done with 100 ng of single stranded cDNA and 50 pmol of each of the primers in 50 ml of the PCR mixture (10 mM Tris-HCl, pH 8.0, 50 mM KCl, 2.5 mM $MgCl_2$, 0.01% gelatin, 200 M each dNTP, and 2.5 units of Taq DNA-polymerase (Perkin Elmer-Cetus). Total leech cDNA was used for PCR amplification.

Table 1
Oligonucleotides used for the destabilase cDNA isolation and cloning

Oligonucleotide designation	Oligonucleotide sequence (5′→3′)
AP1	GTAATACGACTCACTATAGGGCTC-GAGCGGCCGCCCGGGCAGGT
AP2	ACCTGCCCGGGC
AP3	GAGCGGCCGCCCGGGCAGGT
AP4	GTAATACGACTCACTATAGGGC
DsA	CGTTGTATTTGYCARGTNGARGG
DsB	GTATGGATCCTTWATYTGRTA
P1	ACTCAAGCTTCCAACATCCAT
P2	AGCGTCGACAGACTTCCTGCGTCCATG
P3	GCTCTAGAGATTGGAAGGTGCGGCA
P4	GCTCTAGAAGGTGGAAGGATGTAACAAT

Y = T+C, W = A+T+G, R = A+G, N = A+C+G+T.

*Corresponding author. Fax: (7) (95) 196-02-21.
E-mail: eds@glas.apc.org

PII S0014-5793(96)00644-8

The resulting product was cloned into a pBluescript SK vector (Stratagene, USA) linearized with *Bam*HI and *Sma*I restriction endonucleases.

2.3. Amplification of 5'- and 3'-termini of the destabilase cDNA

To obtain cDNA fragments covering the entire coding regions we used the Marathon RACE strategy originally reported earlier [13]. For all PCR reactions the hot start technique was applied with Taq-Start Antibodies (Clontech Laboratories), according to the instructions of the manufacturer.

The 5'-RACE procedure was implemented with an adapter ligated cDNA library with the nested PCR primer pairs AP1 and P1 for the first round and AP2 and P2 for the second round. The first round included 25 cycles under the following conditions: 95°C, 3 s; 60°C, 20 s; 72°C, 1 min.

The reaction mixture was 10-fold diluted and used for 20 cycles of amplification in the second round of PCR under the same conditions.

The 3'-RACE procedure was performed as described above with AP1 and P3 primers for the first round and AP2 and P4 pair for the second round.

The resulting products were cloned into a pBluescript SK vector linearized with *Xba*I/*Sal*I and *Xma*III/*Sal*I restriction endonucleases [14].

2.4. DNA sequencing

The plasmids were sequenced by the termination method using Sequenase Version 2.0 DNA polymerase (US Biochemicals) [15]. Both strands of the cDNA were sequenced. At least three independent clones for each PCR product were sequenced to determine the final sequence.

2.5. Protein primary structure determination

Isolation and purification of the destabilase protein, sequencing of its 36 N-terminal amino acids, and isolation and sequencing of four BrCN peptides (1–4, Fig. 1) have been described earlier [10]. To complete sequencing the longest BrCN peptide, 2 (Fig. 1B), started in [10], it was digested with *Staphylococcus aureus* glutamic acid specific protease V-8 ('Sigma') by the method described in [15]. Vertical arrows in Fig. 1B indicate positions of the splits. The shorter peptides were sequenced as described [10]. The sequence of 10 C-terminal amino acids of BrCN peptide 4 (Fig. 1B) was also determined as described in [10].

3. Results and discussion

3.1. The complete primary structure of the destabilase protein

Earlier we presented a partial amino acid sequence of the destabilase protein. The destabilase protein was split into four fragments with BrCN. Two of the fragments have been sequenced previously [10], but the longest one, 2, as well as C-terminal peptide 4 (Fig. 1B) remained unfinished. Now we have completed the sequencing of the peptides covering all 115 amino acids of the protein sequence. For this goal fragment 2 was digested with a glutamic acid specific protease, the resulting shorter fragments sequenced, and the whole structure of the fragment thus determined. The primary structures of the peptides sequenced are shown in Fig. 1B. Several questions still remain to be answered, all pertinent to posttranslational modifications of the protein. Our preliminary data based on mass spectrometry of the peptides implied that some covalently bound groups were associated with the protein. In particular, there is rather strong evidence that one of such groups is attached to the heptapeptide containing Thr^{64} (our unpublished data). The involvement of 14 cysteines of the protein polypeptide chain in disulfide bridge formation should also be investigated. The corresponding experiments are now in progress.

3.2. Cloning and analysis of the third type of destabilase related cDNAs (Ds3)

In our recent work [10] we were able to clone two closely related cDNAs, Ds1 and Ds2, encoding two homologous but not identical proteins which were in turn partially homologous to the destabilase protein. Therefore we came to the conclusion that there exists a destabilase gene family of at least three members corresponding to three different proteins. However, we failed to isolate the cDNA precisely encoding the identified protein. In search for the missing cDNA we prepared a new cDNA library ligated to adapter AP1/AP2 and used it for PCR amplification with two degenerate primers corresponding to positions 8–15 (RCICQVEG, primer DsA) and 35–41 (YQIKEPY, primer DsB) of the destabilase mature protein (Fig. 1B). Twenty-four individual clones containing amplified 115 bp long fragments were isolated and sequenced. Eighteen of them contained a characteristic sequence encoding the NNE tripeptide distinguishing the destabilase protein from the amino acid sequences inferred from Ds1 and Ds2 cDNAs [10]. The sequence obtained was used for specific PCR amplifications of the corresponding 5'- and 3'-termini of the mRNA according to the Marathon cDNA amplification procedure [13]. Finally, the whole coding sequence matching the sequence of the destabilase protein was obtained (Fig. 1A). The open reading frame was 135 amino acids long followed by UGA stop codons and a polyadenylation signal AAUAAA.

Table 2 shows the identity between the primary structures of the three destabilase mRNAs and proteins.

Ds3 mRNA in its coding part was most similar to Ds1 mRNA and markedly different from Ds2. This suggests that Ds1 and Ds3 genes had a common ancestor that had diverged from the Ds2 gene earlier in the evolution. The remarkable identity in the 5'-noncoding regions of Ds1 and Ds3 mRNAs as opposed to the totally different 5'-noncoding sequence of the Ds2 mRNA is in line with the hypothesis. The 3'-noncoding regions of all three mRNAs were completely different. All the mRNA contexts surrounding initiation AUG codons did not completely match the consensus GCCA/GCCAUGG [20] but had A residues at an important position +4. The G+C content of Ds3 mRNA in coding region is 51.0, as compared to 50.0 and 44.5 for Ds1 and Ds2 mRNAs, respectively. In all cases the G+C levels of the noncoding parts were lower as compared to the coding regions, in accord with the general rules of the G+C content distribution [16].

Table 2
Identity between mRNAs and protein primary structures for the three destabilase forms

Forms under comparison	Percentage of identity between:		
	mRNAs (coding parts)	Protein precursors	Mature proteins
Ds1-Ds2	62.5	60.0	66.1
Ds1-Ds3	88.2	79.3	75.7
Ds2-Ds3	64.0	62.2	68.7

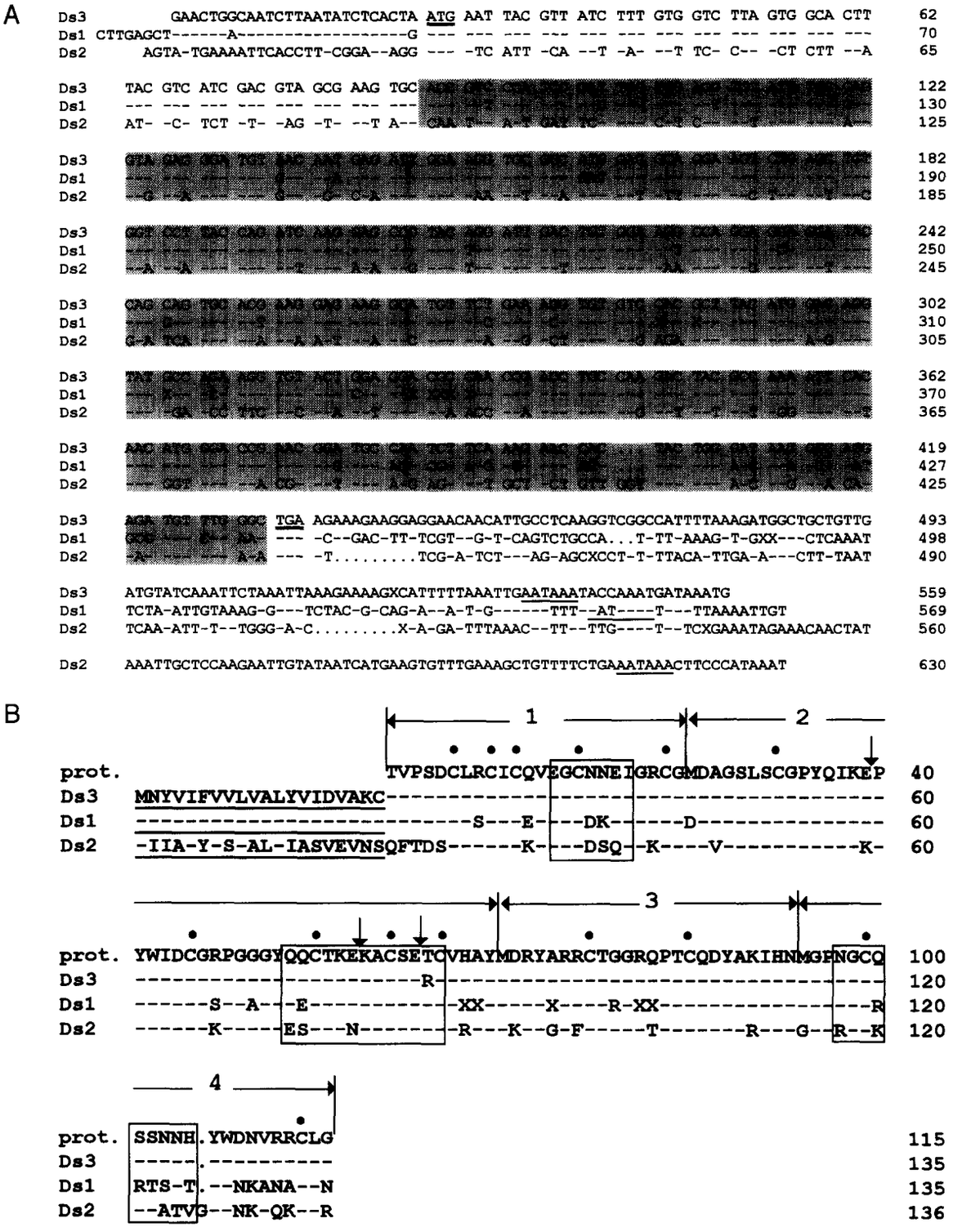

Fig. 1. Alignments of cDNA nucleotide and inferred amino acid sequences of Ds1, Ds2 and Ds3, and destabilase protein. A: Alignment of Ds1, Ds2 and Ds3 cDNA nucleotide sequences (EMBL accession numbers U24121 and U24122 for Ds1 and Ds2. Ds3 is being prepared for submission). Hyphens indicate identical nucleotides. Initiation and stop codons as well as putative polyadenylation signals are underlined. Coding sequences of the mature proteins are shown in the shaded box. B: Alignment of the amino acid sequences deduced from Ds1, Ds2, and Ds3 cDNAs and the amino acid sequence of destabilase protein (prot.). Hyphens indicate identity with the corresponding residues in the Ds3 sequence in the signal part (underlined) and with the destabilase protein in mature protein regions. Regions enriched in hydrophilic residues are boxed. The positions of 14 conserved cysteine residues are marked with black circles. The peptides resulting from BrCN splitting of the destabilase protein are numbered above the protein amino acid sequence. Vertical arrows indicate the positions of the longest BrCN peptide digestion with *Staphylococcus* protease V-8.

3.3. *The destabilase protein, its precursor and comparison with other members of the protein family*

The comparison of the deduced amino acid sequence with the N-terminal structure of the protein allows one to conclude that the destabilase is translated as a precursor containing a 20 amino acids long potential signal peptide (Fig. 1B). Signal peptides of this kind are cleaved from the precursor by signal peptidases [17,18]. This finding correlates with destabilase secretion and is a common feature of all three members of the family. Although the sequences of the three signal peptides were not identical, all of them retained major characteristics of the leader part of the protein precursors transported to the endoplasmic reticulum: they were enriched in apolar amino acids and separated from the mature protein by polar amino acid residues.

The mature destabilase Ds3 protein consists of 115 amino acid residues. It shares 75.7 and 66.1% identity with Ds1 and Ds2 polypeptides, respectively. Like the two other proteins it is characterized by a high cysteine content (14 of the 115 amino acid residues, in contrast to an average cysteine content of about 3%). The positions of Cys residues were conserved in all three amino acid sequences, which suggests their functional and structural importance. Three very hydrophilic regions were present in the protein at positions 14–20, 53–65, and 97–105 of the mature polypeptide. The regions are very conservative among the three proteins and might be directly involved in the isopeptidase mechanism.

Only one difference was detected between the inferred protein sequence and that obtained by direct peptide sequencing: at position 64 of the mature protein (Fig. 1B) the deduced protein chain contains an Arg residue in contrast to a Thr residue in the isolated protein. At present we do not know the reason for this discrepancy. Several cDNA clones sequenced independently gave the same Arg residue at this point. Possibly, the sequence of the peptide has to be reconfirmed at this position because according to the mass spectrometry data Thr^{64} can be modified by a group attached to it. Such a modification could lead to erroneous residue determination. The substitution Arg→Thr can also be caused by just one point mutation AGG (Arg)→ACG (Thr). It should be mentioned that two other cDNAs encode Thr at this position.

As well as mRNA data, differences in the three protein structures (the distribution of hydrophobic and hydrophilic regions, overall content of basic and acidic amino acids, etc.) are also in favor of early evolutionary divergence of members of the destabilase gene family, and possibly suggest some functional differences between the enzymes as to substrate specificity or kinetic behavior.

No clear-cut homology to destabilase was found in public databases [19]. Some short polypeptide stretches homologous to segments of different proteins were discussed previously. In view of PMSF inhibition of destabilase activity (our unpublished data), a serine residue might be a part of the active center. Short motifs loosely homologous to the known consensus GDSGGP [20] of some serine proteinases were found at positions 4–8 (DSCLR), 15–21 (GCDSQIG) and 26–34 (DVGSLSCGP). All three proteins contain the only histidine residue (His^{92}) that might be involved in the active center function. The destabilase could be reasonably suggested to share homology with transglutaminases. But no math was found, even with the consensus sequence of transglutaminase active centers (YGQCWVFA [4]).

Homologies were seen between residues 31–62 of Ds2 and residues 243–274 of the thrombospondin related anonymous protein (TRAP) [21] and between residues 2–15 of Ds1 and positions 4–17 of a novel lysozyme from *Asterias rubens* [22], although the meaning of these homologies is unclear.

The discovery of the gene family and proteins with apparently similar functional ability to digest isopeptide bonds brings up many questions on their role in the life of the leech. In this connection it would be necessary to study kinetic and regulatory properties of the different forms, tissue specificity of their expression, their balance in the course of development, and, most importantly, whether digestion of isopeptide bonds is the only function of these enzymes. If similar enzymes and genes exist also in other organisms remains to be answered as well.

Acknowledgements: This work was supported by the International Science Foundation, the State National Programme 'New Techniques in Bioengineering' and the Russian Foundation for Basic Research. We thank Dr. B. Glotov (Institute of Molecular Genetics, Moscow, Russia) for his comments and advice. E.D. Sverdlov is also a Visiting Research Professor in the Department of Biomedical Engineering of Boston University, Boston, MA, USA.

References

[1] Lorand, L. and Conrad, S.M. (1984) Mol. Cell. Biochem. 58, 9–35.
[2] Shainoff, J.R., Urbanic, D.A. and DiBello, P.M. (1991) J. Biol. Chem. 266, 6429–6437.
[3] Tamaki, T. and Aoki, N. (1982) J. Biol. Chem. 257, 14767–14772.
[4] Aeschlimann, D. and Paulsson, M. (1994) Thromb. Haemostas. 71, 402–415.
[5] Baskova, I.P., Timokhina, E.A., Nikonov, G.I. and Stepanov, V.M. (1990) Biokhimia (USSR) 55, 771–775.
[6] Zavalova, L.L., Nikonov, G.I., Kuzina, E.V., Popova, G.Yu. and Baskova, I.P. (1991) Biokhimiya (USSR) 56, 115–124.
[7] Zavalova, L.L., Kuzina, E.V., Levina, N.B. and Baskova,I.P. (1993) Thromb. Res. 71, 241–244.
[8] Baskova, I.P. and Nikonov, G.I. (1985) Biokhimiya (USSR) 50, 363–375.
[9] Baskova, I.P. and Nikonov, G.I. (1991) Blood Coagulat. Fibrinolysis 2, 167–172.
[10] Zavalova, L., Lukyanov, S., Baskova, I., Snezhkov, E., Akopov, S., Berezhnoy, S., Bogdanova, E., Barsova, E. and Sverdlov, E.D. (1996) Mol. Gen. Genet. (submitted).
[11] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[12] Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. 162,156–159.
[13] Chenchik, A., Diachenko, L., Tarabykin, V., Lukyanov, S. and Siebert, D. (1996) BioTechniques (in press).
[14] Tabor, S. and Richardson, C.C. (1987) Proc. Natl. Acad. Sci. USA 84, 4767–4771.
[15] Houmard, J. and Drapeau,G.R. (1972) Proc. Natl. Acad. Sci. USA 69, 3506–3509.
[16] Bernardi, G. (1989) Annu. Rev. Genet. 23, 637–661.
[17] High, S. (1992) BioEssays 14, 535–540.
[18] Dalbey, R.E. and von Heijne, G. (1992) Trends Biochem. Sci. 17, 474–478.
[19] Protein Identification Research, Release 41 (1994).
[20] Dixon, M. and Webb, E.C. (1979) Enzymes, 3d edn., Longmans, London.
[21] Robson, K.J.H., Hall, J.R.S., Jennings, M.W., Harris, T.J.R., Marsh, K., Newbold, C.I., Tate, V.E. and Weatherall, D.J. (1988) Nature 335, 79–82.
[22] Jolles, J. and Jolles, P. (1975) Eur. J. Biochem. 54, 19–23.