Volume 324, number 2, 226–230 FEBS 12586 June 1993

Identification, sequence analysis, and characterization of cDNA clones encoding two granzyme-like serine proteinases from rat duodenum

Alexander Yu. Amerik, Serge V. Yarovoi, Vitali G. Grigorenko and Vladimir K. Antonov†

Shemyakin Institute of Bioorganic Chemistry, Russian Academy of Science, Moscow, 117871, Russian Fededration

Received 19 March 1993; revised version received 1 April 1993

Clones of cDNA encoding two serine proteinases were isolated from a cDNA library prepared from rat duodenum mRNA. The deduced amino acid sequences consisted of 248 residues and possessed a high level of homology to one another and to the sequences of granzymes, cathepsin G, and mast cell proteases I and II. Analysis of the enzymes' primary structures allowed the identification of the catalytic amino acid triad and the prediction of the substrate specificity. Northern blotting experiments showed that while one of these proteinases is expressed only in duodenum, the other enzyme is present in duodenum, lung, and spleen. It is supposed that these proteinases may play an important role in the function of an organism's defence systems.

cDNA cloning; Duodenum, Granzyme; Serine proteinase

1. INTRODUCTION

When the phenomenon of cell cytotoxicity was studied, it was shown that cytoplasmic granules of cytotoxic T-lymphocytes contain several serine proteinases called granzymes [1,2]. The actual physiological functions of granzymes are not known, but several lines of evidence suggest that these enzymes play an important role in target cell lysis [2]. Proteolytic enzymes, possessing structural homology with granzymes, are contained in neutrophil granules (cathepsin G) [3,4] and mast cell granules [5–7]. It was proposed that these enzymes may play an important role in defence against pathogens [3] and in hypersensitivity, anaphylactic, and inflammatory reactions [8]. Recently, the new serine proteinase (duodenase) was identified in the bovine duodenum [9]. This enzyme has a high level of homology of the N-terminal amino acid sequence with granzymes, cathepsin G, and mast cell proteases. In this communication we report about cloning and sequencing of the cDNA for the related proteinases from rat duodenum, analyze the tissue-specificity of their expression, and characterize the deduced amino acid sequences.

Correspondence address A.Yu. Amerik, Shemyakin Institute of Bioorganic Chemistry, Russian Academy of Science, Moscow, 117871, Russia. Fax. (7) (095) 310 70 07.

† Prof. V.K. Antonov passed away on the 26th of July 1992.

Abbreviations: CHYM A, chymotrypsinogen A, GLP I and II, granzyme-like proteins I and II; GRAN B, granzyme B; CAT G, cathepsin G; RMCP I and II, rat mast cell proteases I and II, DUODEN, duodenase; PCR, polymerase chain reaction.

The nucleotide sequences presented here have been submitted to the EMBL, GenBank, and DDBJ Nucleotide Sequence Databases under the accession numbers X66693 and X68657 (*R norvegicus* mRNAs for granzyme-like protein I and II, respectively)

2. MATERIALS AND METHODS

For RNA isolation from rat duodenum the guanidine thiocyanate method was used [10]. Poly(A)$^+$ RNA was selected and used as a template for cDNA synthesis [11]. Double-stranded cDNA was ligated with *Sma*I-digested, dephosphorylated plasmid pSP64 and used for the transformation of *E. coli* MH I cells. The library was blotted on nylon membrane and hybridized overnight with the ^{32}P-labeled multiprime probes or synthetic oligonucleotide probes.

The sequencing was performed on double- and single-stranded DNA templates according to the methods of Sanger et al [12].

For Nothern blot analysis poly(A)$^+$ RNA was electrophoresed on agarose–formaldehyde gels [11] and transferred to nylon membrane by vacuum blotting. Hybridization with ^{32}P-labeled cDNA probes was carried out in the presence of 50% formamide at 42°C overnight.

3. RESULTS AND DISCUSSION

3.1. *Isolation of cDNA clones encoding GLPs from rat duodenum*

For the cDNA library analysis, oligonucleotide CATGTAGGGGCGGGAGTGGGGCTTGGCCTC-ATGGCCCCC was synthesized. This probe corresponds to the N-terminal amino acid sequence of the duodenase [9] and to the nucleotide sequence of the human CCPX gene [13]. After hybridization analysis of 100,000 independent clones, two positive clones were identified. Determination of cDNA sequences of these clones allowed us to synthesize oligonucleotides CATGAAGCCGATCCCCACTCTCGACC and TACAGAGTCCAAACCCCACTCCCGGC and to use them as first primers in a polymerase chain reaction (PCR). Nu-

Published by Elsevier Science Publishers B V.

A. GRANZYME-LIKE PROTEIN I

```
TCTTCTAGAGCTGAAAAGAGAGCAAGGACAACACTCTCGACGGTGGGACCTAGGTGGCCTT   61

CCAGGGAAGATGAACCTCCTCTTGCTCCTGCTGACCGTCTCTCTGGCCCCCACGACAGAG  121
          M  N  L  L  L  L  L  L  T  V  S  L  A  P  T  T  E   -4

GCAGCGGAGATCATCGGGGGACATGAAGCCGATCCCCACTCTCGACCCTACATGGCCTAT  181
 A  A  E  I  I  G  G  H  E  A  D  P  H  S  R  P  Y  M  A  Y   17

CTTCAGTACAAGAATGAGGATTCTCGGGATACAATATGTGGTGGTTTCCTTATACGAGAG  241
 L  Q  Y  K  N  E  D  S  R  D  T  I  C  G  G  F  L  I  R  E   37

GACTTTGTGCTGACTGCTGCTCACTGTTCAGGGAGCAAAATAAATGTCACATTGGGGGCC  301
 D  F  V  L  T  A  A  H  C  S  G  S  K  I  N  V  T  L  G  A   57

CACAACATCAAAGAACAGGAGAAGACGCAGCAAGTCATCCCTGTGGTGAAAATCATTCCA  361
 H  N  I  K  E  Q  E  K  T  Q  Q  V  I  P  V  V  K  I  I  P   77

CACCCAGCGTATAATGCTAAGACAATCTCCAATGACATCATGCTATTAAAGCTGAAGAGT  421
 H  P  A  Y  N  A  K  T  I  S  N  D  I  M  L  L  K  L  K  S   97

AAGGCCAAGAGGACTAGAGCTGTGAAGACTCTCAGTCTGCCCAGAAGCAACTTCAAAGTG  481
 K  A  K  R  T  R  A  V  K  T  L  S  L  P  R  S  N  F  K  V  117

AAGCCAGGAGATGTGTGCTATGTGGCTGGTTGGGGAAAGCTGGGCCCAATGGGCAAATTC  541
 K  P  G  D  V  C  Y  V  A  G  W  G  K  L  G  P  M  G  K  F  137

CCAGACAAACTGCAAGAGGTTGAGCTAACAGTACAGGAGGATCAGGAGTGTGAGACTTAC  601
 P  D  K  L  Q  E  V  E  L  T  V  Q  E  D  Q  E  C  E  T  Y  157

TTAAAAAATGCTTACGACAAAGCCAATCAGATATGTGCGGGGGACCCAAAGATCAAATGT  661
 L  K  N  A  Y  D  K  A  N  Q  I  C  A  G  D  P  K  I  K  C  177

GCTTCCTTTCAGGGGGACTCTGGAGGGCCTCTTGTGTGTAAAAAAGTGGCCGCAGGCATC  721
 A  S  F  Q  G  D  S  G  G  P  L  V  C  K  K  V  A  A  G  I  197

GTCTCCTATGGACGTAAAGATGGTTCAACTCCACGGGCATTCACCAAAGTCTCGACTTTC  781
 V  S  Y  G  R  K  D  G  S  T  P  R  A  F  T  K  V  S  T  F  217

CTATCCTGGATAGAGGAAACTATGAAAAAGAGCTAACTACAAGATGCAACGTGGATCATT  841
 L  S  W  I  E  E  T  M  K  K  S                              228

TCCTGACTAACCACCTTCCCTATAGCTGAGTCCAGGATTGCTCTAGGATAGATGGCAGCA  901

ACTGAATAAAGCATTTTTTCTGAC                                      925
```

B. GRANZYME-LIKE PROTEIN II

```
CCCTGAAGAGGATGTTCCTGTTCCTGTTCTTCCTGGTGGCCATCCTACCAGTCAACACT   59
          M  F  L  F  L  F  F  L  V  A  I  L  P  V  N  T   -5

GAAGGAGGAGAGATCATATGGGGTACAGAGTCCAAACCCCACTCCCGGCCCTACATGGCA  119
 E  G  G  E  I  I  W  G  T  E  S  K  P  H  S  R  P  Y  M  A   16

TTCATAAAGTTTTATGATAGTAATTCAGAACCCCATCACTGTGGCGGTTTCCTGGTGGCA  179
 F  I  K  F  Y  D  S  N  S  E  P  H  H  C  G  G  F  L  V  A   36

AAAGACATCGTAATGACAGCAGCTCACTGTAATGGAAGAAATATAAAAGTAACCTTAGGT  239
 K  D  I  V  M  T  A  A  H  C  N  G  R  N  I  K  V  T  L  G   56

GCTCACAATATCAAGAAACAAGAAAACACCCAGGTTATCTCTGTTGTAAAAGCCAAACCT  299
 A  H  N  I  K  K  Q  E  N  T  Q  V  I  S  V  V  K  A  K  P   76

CACGAGAACTATGACAGAGATTCACATTTTAATGACATCATGCTCCTGAAGTTGGAACGC  359
 H  E  N  Y  D  R  D  S  H  F  N  D  I  M  L  L  K  L  E  R   96

AAAGCTCAACTCAATGGTGTTGTGAAGACTATTGCCCTTCCTAGGAGCCAGGACTGGGTG  419
 K  A  Q  L  N  G  V  V  K  T  I  A  L  P  R  S  Q  D  W  V  116

AAACCTGGGCAGGTGTGCACAGTGGCAGGTTGGGGACGCTTGGCCAATTGTACTTCGTCT  479
 K  P  G  Q  V  C  T  V  A  G  W  G  R  L  A  N  C  T  S  S  136

AACACACTTCAAGAAGTGAATCTAGAAGTTCAGAAAGGCCAGAAGTGCCAAGACATGTCC  539
 N  T  L  Q  E  V  N  L  E  V  Q  K  G  Q  K  C  Q  D  M  S  156

GAAGACTACAACGACTCCATCCAGCTTTGTGTGGGAAACCCCAGCGAGGGGAAGGCTACT  599
 E  D  Y  N  D  S  I  Q  L  C  V  G  N  P  S  E  G  K  A  T  176

GGTAAGGGAGACTCAGGGGGTCCCTTTGTGTGCGATGGAGTGGCCCAGGGCATTGTCAGT  659
 G  K  G  D  S  G  G  P  F  V  C  D  G  V  A  Q  G  I  V  S  196

TATCGCTTGTGTACTGGGACACTTCCTCGAGTATTCACCAGAATCTCCAGCTTTATACCG  719
 Y  R  L  C  T  G  T  L  P  R  V  F  T  R  I  S  S  F  I  P  216

TGGATTCAGAAAACAATGAAAGTCCTTCAACAATCCTAGAACACAAAACCTGTGTCTGGG  779
 W  I  Q  K  T  M  K  V  L  Q  Q  S                           228

CCAATGTCCAGCATCCTGGGGTATGGCTATCTGAGTCTTAATAAAGAAATCTGTCTGCAG  839

GAAAAAAAAAAAAAAAAAA                                           858
```

Fig. 1. The sequences of GLP I and II cDNAs and their protein trans- lation. The numbering of the GLP amino acid sequences start with the first residue of the mature rat proteinases as determined by amino acid sequence analysis of the bovine duodenase [9].

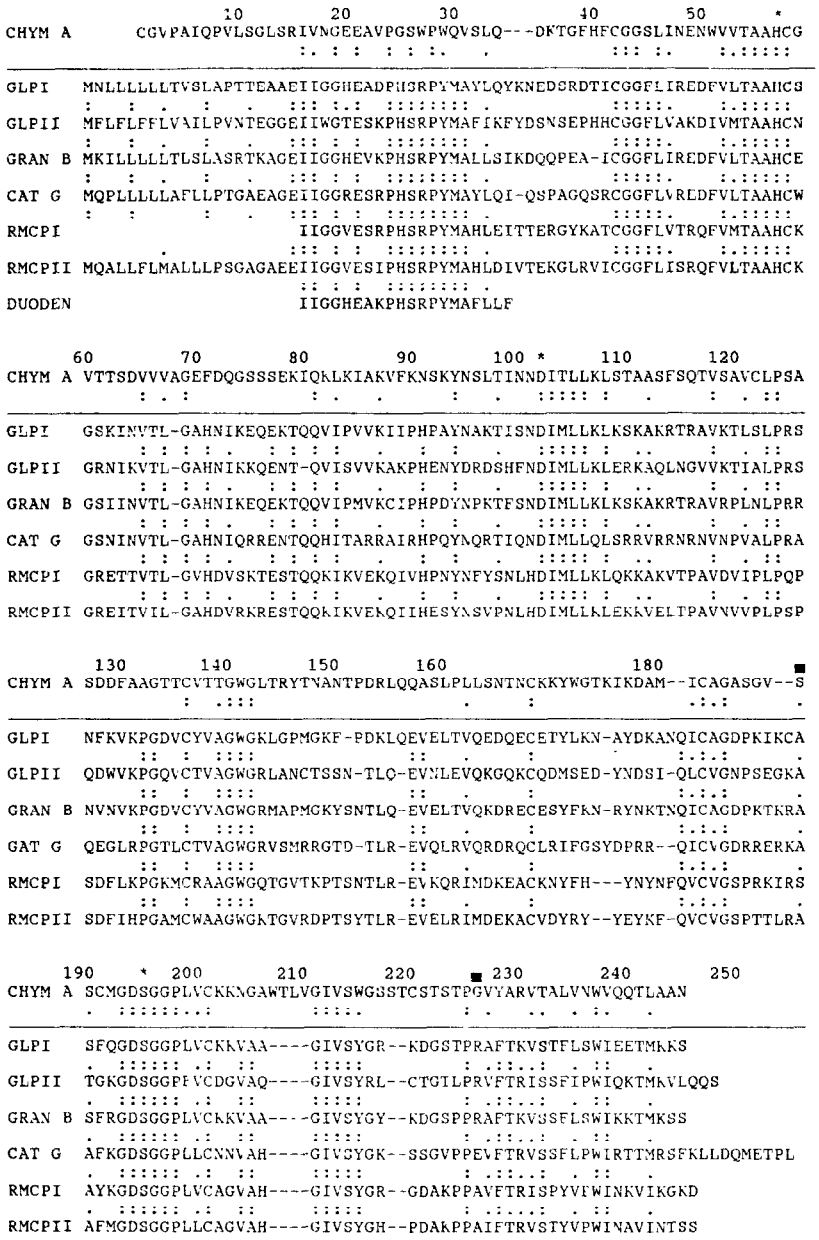

Fig. 2. Alignment of the predicted GLP amino acid sequences with the sequences of related serine proteinases. The alignment makes use of the following sequences: bovine CHYM A [16]; rat GLP I and II (this study); murine GRAN B [17]; human CAT G [4]; RMCP I and II [5–7]; bovine DUODEN [9]. The residues histidine (H), aspartic acid (D), and serine (S), which form the catalytic triad of serine proteinases, are indicated by *. The residues determining the primary substrate specificity [18] are marked by ■

cleotide ACACACGAGGGG$^{T}_{A}$CC$^{T}_{A}$CCGGAGTCCCC corresponding to the conservative amino acid sequence -Gly-Asp-Ser-Gly-Gly-Pro-Leu-Val-Cys- that surrounds the catalytically active serine residue [14] was used as a second primer. In both cases, after 25 PCR cycles, approximately 540 bp DNA fragments were amplified. The PCR fragments obtained were used as nucleotide probes for further analysis of cDNA library. Several clones containing cDNA insertions for each enzyme were identified and sequenced. Based on these data the full cDNA coding region sequences for two serine proteinases were reconstructed. Corresponding amino acid sequences deduced showed a high level of homology to granzymes. We chose to call these enzymes 'granzyme- like proteins I and II' (Fig. 1).

3.2. *Characterization of the GLP I and II amino acid sequences*

Analysis of the amino acid sequences showed that the rat duodenal proteinases consist of 248 amino acid res-

Table I

Homologies between the amino acid sequences for the several related serine proteinases

	CHYM A (%)	GLP I (%)	GLP II (%)	GRAN B (%)	CAT G (%)	RMCP I (%)	RMCP II (%)
CHYM A	100.0	34.8	29.0	32.3	31 4	31.5	31.0
GLP I		100.0	51.0	77.8	54.5	46.5	47 2
GLP II			100.0	53.3	50.4	48.4	43 4
GRAN B				100.0	56.1	48.9	48.4
CAT G					100.0	47 3	50.8
RMCP I						100.0	72.8
RMCP II							100.0

idues and have calculated molecular masses of 27,239 Da and 27,465 Da for GLP I and GLP II, respectively. Predicted sequences start with a hydrophobic region (Fig. 1) which represents a typical signal peptide [15], indicating that the GLPs are translocated across the lipid membrane. The high level of homology between GLPs and granzymes, cathepsin G, and mast cell proteases (Fig. 2 and Table I) suggests that GLPs are contained in cytoplasmic granules of intestinal cells which take part in defence systems function.

Comparison of the GLP amino acid sequences with primary structures of the related serine proteinases allowed the identification of the catalytic amino acid triad found at homologous positions flanked by well-conserved peptide segments (Fig. 2).

The structural similarity between the GLPs and granzymes permits some assumptions about their substrate specificity. A molecular model for granzyme B was constructed by Murphy et al. [18]. According to this model the critical role of the amino acid residues 189 and 226 (using numeration based on bovine chymotrypsinogen) is postulated. These residues take part in forming the S pockets of serine proteinases. Granzyme B has Ala and Arg residues at these positions (Fig. 2), and it has a preference for aspartic acid at the P_1 residue of the substrate [19]. GLPs have the same amino acid residues in the corresponding positions (Fig. 2) and it can be expected that the primary substrate specificity of these enzymes is close to that of granzyme B.

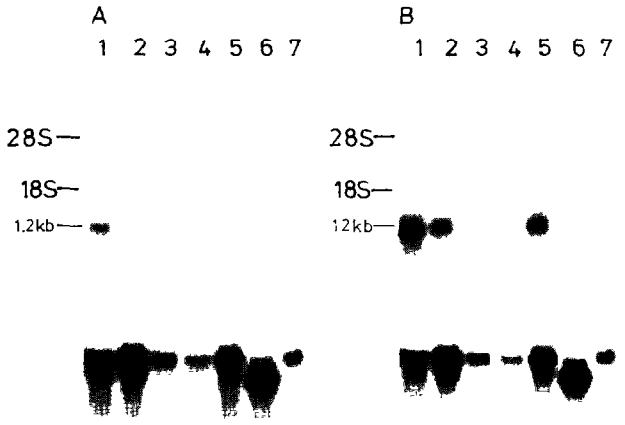

Fig. 3. Northern blot analysis of poly(A)$^+$ RNA from various rat organs. The cDNA fragments of GLP I (A) and GLP II (B) were used as probes; the same blots were reprobed with an actin probe (shown below each panel). In each blot the poly(A)$^+$ RNA from the various organs are presented in the same order: duodenum (1), lung (2), brain (3), liver (4), spleen (5), heart (6), and thymus (7). Size of the GLP transcripts was determined by comparison to 28S and 18S ribosomal RNA bands

3.3. *Tissue specificity of GLP gene expression*

The occurence of GLP I and II RNA transcripts in different organs were tested using Northern blotting (Fig. 3). The analysis indicated that the GLP mRNAs are approximately 1.2 Kb long.

The GLP I RNA transcript was found only in duodenum (Fig. 3) which contains a large number of mucosal mast cells [20]. Brunet *et al.* [21] showed that some mast cell populations contain an RNA transcript close to murine granzyme B mRNA. By virtue of this it seems to be possible that GLP I is expressed in mucosal mast cells. At the same time, cytotoxic intestinal intraepithelial lymphocytes [22] also can be considered as a potential source of GLP I. The level of GLP II RNA transcript is highest in duodenum, slightly lower in lung and spleen, and insignificant in liver. All of these organs contain a number of different cells involved in immune response. It is difficult at this point to indicate the specific cells in which the enzyme is expressed; however, macrophages appear to be a good candidate for this role.

Finally, it is necessary to note that the GLP transcripts were detected in those organs where the organism encounters a lot of alien antigenes. It seems that the GLPs are involved in the organism's protection, but details of their action are unclear.

Acknowledgements: We thank L. M. Ginodman D Sci. for critical reading of the manuscript, G.V. Petukhova for help with PCR technique, and N.S. Bystrov for providing the nucleotide probes We are greatly indebted to G.D. Graff for his assistance in the preparation of the manuscript. This work was supported by the Human Genome Project of Russian Academy of Sciences.

REFERENCES

[1] Masson, D. and Tschopp, J. (1987) Cell 49, 679–685.
[2] Jenne, D.E. and Tschopp, J. (1988) Curr. Topics Microbiol. Immunol. 140, 33–47.
[3] Starkey, P.M. (1977) in: Proteinase in Mammalian Cells and Tissues (Barrett, A.J. Ed.) pp. 57–89, North-Holland, Amsterdam.
[4] Salvesen, G., Farley, D., Shuman, J., Przybyla, A., Reilly, C and Travis, J. (1987) Biochemistry 26, 2289–2293.
[5] Woodbury, R.G., Katunuma, N., Kobayashi, K., Titani, K. and Neurath, H. (1978) Biochemistry 17, 811–819.
[6] Trong, H.L., Parmelee, D.C., Walsh, K.A., Neurath, H. and Woodbury, R.G. (1987) Biochemistry 26, 6988–6994.
[7] Benfey, Ph.N., Yin, F.H. and Leder, Ph. (1987) J. Biol. Chem. 262, 5377–5384.
[8] Woodbury, R G , Trong, H.L. and Neurath, H. (1987) Acta Histochem. Cytochem. 20, 261–269.
[9] Antonov, V.K., Vorotyntseva, T.I. and Zamolodchikova, T.S. (1992) Dokl. Acad. Nauk. 324, 1318–1322.
[10] Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W (1979) Biochemistry 18, 5294–5299.
[11] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: a Laboratory Manual (2nd edition), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[12] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[13] Meier, M., Kwong, P.C., Fregeau, C.J., Atkinson, E.A., Burrington, M., Ehrman, N., Sorensen, O., Lin, C.C., Wilkins, J and Bleackley, R.C. (1990) Biochemistry 29, 4042–4049.
[14] Brenner, S. (1988) Nature 334, 528–530.
[15] von Heijne, G. (1986) Nucleic Acids Res. 14, 4683–4690.
[16] Hartley, B.S. (1964) Nature 201, 1284–1288.
[17] Lobe, C.G , Finlay, B.B., Paranchych, W., Paetkau, V.H. and Bleackley, R.C. (1986) Science 232, 858–861.
[18] Murphy, M.E.P., Moult, J., Bleackley, R.C., Gershenfeld, H., Weissman, I.L. and James, M.N.G (1988) Proteins: Struct. Funct. Genet. 4, 190–204
[19] Odake, S., Kam, C.-M., Narasimhan, L., Poe, M., Blake, J.T., Krahenbuhl, O., Tschopp, J. and Powers, J.C. (1991) Biochemistry 30, 2217–2227.
[20] Enerback, L. (1986) in: Mast Cells Differentiation and Heterogeneity (Befus, A.D., Bienenstock, J. and Denburg, J.A. Eds.) pp. 1–26, Raven Press, NY.
[21] Brunet, J.-F., Denizot, F., Suzan, M., Haas, W., Mencia-Huerta, J.-M., Berke,G., Luciani, M -F. and Golstein, P. (1987) J Immunol. 138, 4102–4105.
[22] Goodman, T. and Lefrancois, L. (1988) Nature 333, 855–858.