Volume 255, number 2, 211-214 FEB 07589 September 1989

Primary structure of a new cysteine proteinase inhibitor from pig leucocytes

Anka Ritonja, Majda Kopitar, Roman Jerala and Vito Turk

Department of Biochemistry, J. Stefan Institute, Jamova 39, 61000 Ljubljana, Yugoslavia

Received 27 July 1989

The primary structure of a pig leucocyte cysteine proteinase inhibitor, also called cathelin, was determined. The sequence was obtained from analyses of peptides isolated from the chymotryptic, endoproteinase Lys-C and protease V8 digests, and by analysis of the peptides derived from the hydrolysis of the aspartyl-prolyl bond of the carboxymethylated inhibitor. The inhibitor consists of 96 residues. The N-terminal residue of the inhibitor is pyrrolidonecarboxylic acid. The amino acid sequence of cathelin suggests the appearance of a new family of cysteine proteinase inhibitors.

Cathelin; Cysteine proteinase inhibitor; Primary structure; Cathepsin L inhibitor

1. INTRODUCTION

Protein inhibitors of cysteine proteinases are widely distributed in living organisms. Among them are cystatins, which form a superfamily of structurally homologous proteins. The cystatin superfamily can be grouped into three families: the stefins, the cystatins and the kininogens [1-4]. In addition to this, a sequence homology of human plasma histidine-rich glycoprotein (HRG) and human α-2 HS-glycoprotein (α-2 HSG) with members of cystatin superfamily was reported [5,6]. It was found that α-2 HSG and HRG are composed of two cystatin-like segments suggesting that both glycoproteins represent a new family within the cystatin superfamily. However, although structurally homologous to cystatins, both glycoproteins have no inhibitory activity against known cysteine proteinases. Statistical analysis of the similarities between the sequences of all homologous proteins, based on the presence of different copies of cystatin-like segments and the presence or absence of disulphide bonds, enables us to assign all known members of the cystatin superfamily into four types or families [7].

Recently, we isolated a new cysteine proteinase inhibitor from pig leucocytes [8]. The protein has a p*I* of 4.8 and molecular mass of about 11 kDa. It was found to be a potent, tight-binding inhibitor of cathepsin L and papain. No appreciable inhibition of other cysteine proteinases was observed.

The purpose of this paper is to present the primary structure of pig leucocyte cysteine proteinase inhibitor and to compare this sequence with those of cystatins. We discuss the possibility that this new inhibitor introduces a new family of cysteine proteinase inhibitors. The name cathelin for a new inhibitor of cathepsin L is proposed.

2. EXPERIMENTAL

The chemicals used for Edman degradation and amino acid analysis were of sequenal grade from Applied Biosystems. Iodo-[^{3}H]acetic acid was from Amersham, endoproteinase Lys-C from Boehringer Mannheim and protease V8 (*S. aureus*) from Miles Scientific (UK). TLCK treated α-chymotrypsin was prepared according to [9].

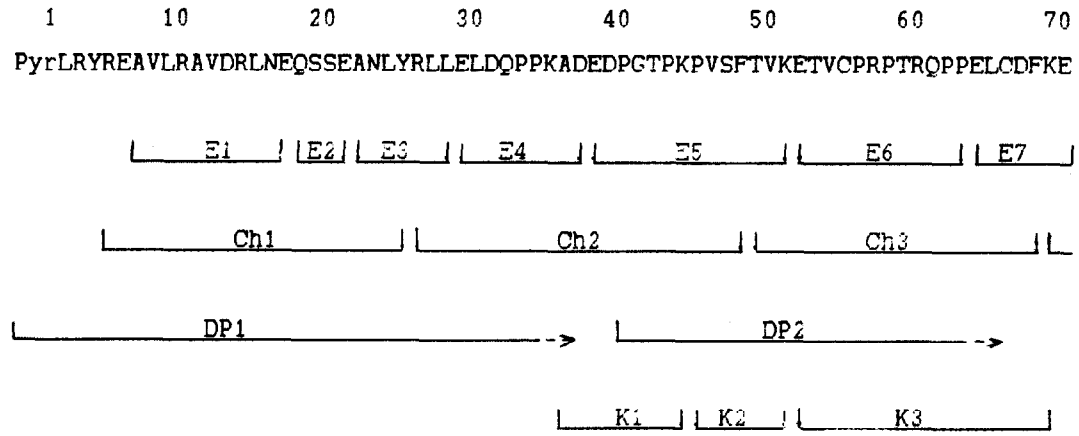
Correspondence address: A. Ritonja, Department of Biochemistry, J. Stefan Institute, Jamova 39, 61000 Ljubljana, Yugoslavia

The nucleotide sequence presented here has been submitted to the EMBL/GenBank database under accession number Y07506

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/89/$3.50 © 1989 Federation of European Biochemical Societies

Cleavage at the aspartyl-prolyl bond in the carboxymethylated protein was performed by the procedure used by Fraser and co-workers [10], somewhat modified. Hydrolysis was performed for 70 h at 38°C in 50% acetic acid at a protein concentration of 3 mg/ml. Unblocking of the blocked α-amino group of the N-terminal amino acid residue in Asp-Pro peptide DP1 was performed by the modified procedure developed by Blomback and Doolittle [11]. Peptide DP1 (1 nmol) was dissolved in 50 μl of 1.0 M NaOH and the solution was maintained 12 h at room temperature and 5 h at 37°C. After this period, the reaction mixture was neutralized with trifluoroacetic acid and the peptide with the liberated α-amino group was separated from the unreacted peptide.

For the determination of free -SH groups, native and active



Ala	A	4	Cys	C	4	His	H	1	Met	M	0
Arg	R	7	Gln	Q	5	Ile	I	3	Phe	F	2
Asn	N	4	Glu	E	9	Leu	L	10	Pro	P	10
Asp	D	6	Gly	G	2	Lys	K	5	Pyr	-	1

Ser	S	7	Val	V	8
Thr	T	6			
Trp	W	0			
Tyr	Y	2			

Total: 96 residues, Mr: 10 834.

Fig.1. Amino acid sequence and composition of inhibitor, and strategy of the sequence determination. The amino acid sequence of *S*-[^{3}H]carboxymethylated protein was determined by automated Edman degradation of the four endoproteinase Lys-C-generated peptides (K1–K4), the nine protease V8-generated peptides (E1–E9), the four α-chymotrypsin generated peptides (Ch1–Ch4) and of two DP peptides (DP1 and DP2) obtained after acid hydrolysis of the Asp-Pro peptide bond. The peptide DP1 was sequenced after reactivation of the α-amino group of the N-terminal amino acid residue. Cysteine residues were identified by the radioactivity of the *S*-[^{3}H]carboxymethylated derivative.

inhibitor were alkylated with iodo-[^{3}H]acetic acid without prior reduction. The protein was reduced with 2-mercaptoethanol in 6 M guanidine hydrochloride, and carboxymethylated with iodo-[^{3}H]acetic acid.

Peptides were purified by gel chromatography and HPLC on an Applied Biosystems RP-18 column eluted with aqueous acetonitrile-containing trifluoroacetic acid. Samples were sequenced with an Applied Biosystems model 470 A gas-phase sequenator [12]. Phenylthiohydantoin derivatives were identified on line with the 120 A HPLC [13].

3. RESULTS AND DISCUSSION

Native and also reduced and carboxymethylated inhibitor failed to yield the N-terminal sequence indicating that its α-amino group was blocked. Reduced and carboxymethylated protein was hydrolyzed by endoproteinase Lys-C. The sequence of the K1 peptide showed the presence of an Asp-Pro bond (see fig.1). The acid-labile character of this bond was used for hydrolysis of the protein chain to obtain the 39-residue long N-terminally blocked peptide DP1 and 57-residue long C-terminal peptide DP2. The DP2 peptide contained 7 Pro residues amongst the first 24 residues and consequently the automated sequence stopped.

The additional sequenal data were obtained from the peptides after protease V8 hydrolysis. The overlaps of the peptides were provided by the set of α-chymotryptic peptides. Val is considered to be the C-terminal residue of the inhibitor, based on the specificity of endoproteinase Lys-C and the proteinase V8. The protein contains four Cys residues, presumably forming two disulphide bonds because no free -SH group was found. To obtain the N-terminal sequence of the protein, the DP1 peptide was used for basic hydrolysis. Reactivation of α-amino group occurred about 50%, and its N-terminal sequence yielded a Glu residue. It was thus confirmed that pyrrolidonecarboxylic acid is the N-terminal residue of the inhibitor.

We compared the amino acid sequence of cathelin with all known proteins of the cystatin superfamily. The closest however, is the identity with bovine kininogen segment 1 and bovine colostrum cystatin (fig.2). Percentages (in parentheses) of identical residues when the sequences are aligned as in fig.2 are: bovine segments 1, 2 and 3 (22, 16 and 14), bovine colostrum cystatin (21) and chicken cystatin (16). Based on this low number, we consider this protein inhibitor as a member of a new family of cysteine proteinase inhibitors [18].

The most highly conserved residue Gly 9 (chicken cystatin numbering) found in all known sequences of inhibitory cystatins [1,4], recently identified as being important for the inhibitory activity of cystatins [19], is not present in cathelin. Its N-terminal part is 17 residues shorter than that of

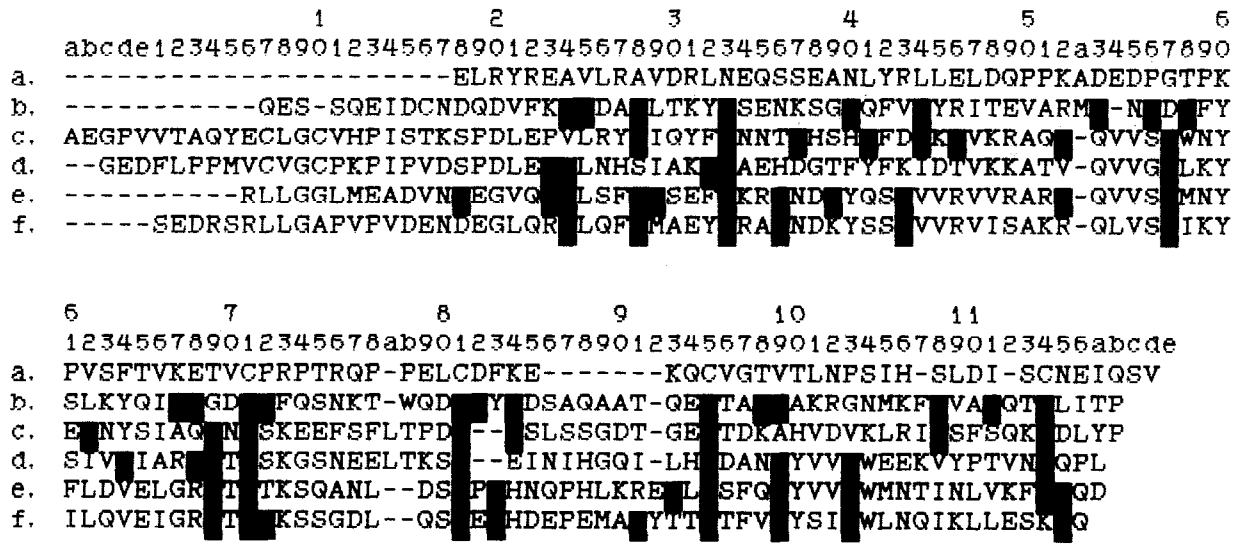


Fig.2. Amino acid sequence of the pig leucocyte inhibitor cathelin (a), aligned with those of bovine kininogen segments 1 (b), 2 (c) and 3 (d) [14], bovine colostrum cystatin (e) [15], and chicken cystatin (f) [16,17]. Residues are numbered according to chicken cystatin. Alignment was done as described in [18].

chicken cystatin (see fig.2) and is thus missing the whole part containing the Gly 9 residue. The shorter N-terminal part could not be the product of a proteinase digest because its N-terminal residue is pyrrolidonecarboxylic acid. Another important fact is that the highly conserved sequence QXVXG (residues 53–57) (see fig.2 and [1,4]) is not present in leucocyte inhibitor with the exception of the Gly 57 residue. Both highly conserved regions important for the proposed interaction of cystatin with cysteine proteinases of the papain family [19] are not present in pig leucocyte inhibitor cathelin, indicating that the interaction should be different. Recently, by X-ray diffraction studies, the model for the interaction of chicken cystatin and papain was proposed [20]. The exact interaction of pig leucocyte cathelin with cysteine proteinases remains to be established by the same approach from enzyme inhibitor crystals.

Acknowledgements: We wish to express our gratitude to Dr A.J. Barrett for providing us with his manuscript prior to publication. We are grateful to B. Turk for his help in technical assistance in preparing this manuscript. This work was supported by the Research Council of Slovenia (Grant C1-0515-106 to V.T.) and the Federal Grant BB P-22 for Technological Development.

REFERENCES

[1] Barrett, A.J., Rawlings, N.D., Davies, M.E., Machleidt, W., Salvesen, G. and Turk, V. (1986) in: Proteinase Inhibitors (Barrett, A.J. and Salvesen, G. eds) pp. 515–569, North Holland, Amsterdam.

[2] Barrett, A.J., Fritz, H., Grubb, A., Isemura, S., Jarvinen, M., Katunuma, N., Machleidt, W., Muller-Esterl, W., Sasaki, M. and Turk, V. (1986) Biochem. J. 236, 312.

[3] Muller-Esterl, W., Fritz, H., Kellermann, J., Lottspeich, F., Machleidt, W. and Turk, V. (1985) FEBS Lett. 191, 221–226.

[4] Turk, V., Brzin, J., Sali, A. and Machleidt, W. (1986) in: Cysteine Proteinases and their Inhibitors (Turk, V. ed.) pp. 429–441, Walter de Gruyter, Berlin.

[5] Koide, T. and Odani, S. (1987) FEBS Lett. 216, 17–21.

[6] Elzanowsky, A., Barker, W.C., Hunt, L.T. and Seibel-Ross, E. (1988) FEBS Lett. 227, 167–170.

[7] Rowlings, N.D. and Barrett, A.J. (1988) J. Mol. Evol., in press.

[8] Kopitar, M., Ritonja, A., Popovic, T., Gabrijelcic, D., Krizaj, I. and Turk, V., submitted.

[9] Shaw, E. (1967) Methods Enzymol. 11, 677–686.

[10] Fraser, K.J., Poulsen, K. and Haber, E. (1972) Biochemistry 11, 4974–4977.

[11] Blomback, B. and Doolittle, R.F. (1963) Acta Chem. Scand. 17, 1816–1819.

[12] Hewick, R.M., Hunkapiller, M.W., Hood, L.E. and Dreyer, W.J. (1981) J. Biol. Chem. 256, 7990–7997.

[13] Hunkapiller, M.W. and Hood, L.E. (1983) Methods Enzymol. 91, 486–493.

[14] Kitamura, N., Takagaki, Y., Furuto, S., Tanaka, T., Nawa, H. and Nakanishi, S. (1983) Nature 305, 545–549.

[15] Hirado, M., Tsunasawa, S., Sakiyama, F., Niinobe, M. and Fujii, S. (1985) FEBS Lett. 186, 41–45.

[16] Turk, V., Brzin, J., Longer, M., Ritonja, A., Eropkin, M., Borchart, U. and Machleidt, W. (1983) Hoppe-Seyler's Z. Physiol. Chem. 364, 1487–1496.

[17] Schwabe, C., Anastasi, A., Crow, H., McDonald, J.K. and Barrett, A.J. (1984) Biochem. J. 217, 813–817.

[18] Dayhoff, M.O., Barker, W.C. and Hunt, L.T. (1983) Methods Enzymol. 91, 524–545.

[19] Abrahamson, M., Ritonja, A., Brown, M.A., Grubb, A., Machleidt, W. and Barrett, A.J. (1987) J. Biol. Chem. 262, 9688–9694.

[20] Bode, W., Engh, R., Musil, D., Thiele, U., Huber, R., Karshikov, A., Brzin, J., Kos, J. and Turk, V. (1988) EMBO J. 7, 2593–2599.