The amino acid sequence of a novel inhibitor of cathepsin D from potato

Anka Ritonja[1], Igor Križaj[1], Pika Meško[1], Majda Kopitar[1], Peter Lučovnik[1], Borut Štrukelj[1], Jože Pungerčar[1], David J. Buttle[2], Alan J. Barrett[2] and Vito Turk[1]

[1]*Department of Biochemistry, Jožef Stefan Institute, Jamova 39, 61000 Ljubljana, Yugoslavia* and [2]*Biochemistry Department, Strangeways Research Laboratory, Worts Causeway, Cambridge CB1 4RN, UK*

Received 23 April 1990

The amino acid sequence of a cathepsin D inhibitor isolated from potato is described It was determined by analysis of peptides generated by use of the glycine-specific proteinase PPIV. The order of the peptides was established by examination of tryptic peptides derived from the two cyanogen bromide peptides The inhibitor comprises 187 amino acid residues, and has a calculated M_r of 20 450

Aspartic proteinase inhibitor, Amino acid sequence; Cathepsin D

1. INTRODUCTION

In contrast to the widespread distribution of inhibitors of serine, cysteine and metallo-proteinases in nature, very few naturally occurring inhibitors of aspartic proteinases are known. Apart from the low M_r inhibitors related to pepstatin (reviewed in [1]), there is an inhibitor of pepsin and gastricsin from the roundworm, *Ascaris lumbricoides* [2]. A cathepsin D inhibitor from potato tubers was discovered and well characterized by Keilova and co-workers [3,4], and very recently, its complete amino acid sequence was determined [5]. We now report on the structure of a novel inhibitor of cathepsin D (NID), also from potato.

2. EXPERIMENTAL

The inhibitor of cathepsin D (NID), with p*I* 8.5 and apparent K_i in the order of 10^{-9} was prepared from potato tubers as described [6].

The papain proteinase IV (PPIV) was isolated as previously described [7]. β-Trypsin was prepared according to [8]. The chemicals used for Edman degradation and amino acid analysis were of Sequenal grade from Applied Biosystems. 4-Vinylpyridine monomer was from Fluka, and the Chromspher C8 HPLC column from Chrompack. All other chemicals were of analytical grade.

The protein was reduced with 2-mercaptoethanol in 6 M guanidine hydrochloride, and alkylated with 4-vinylpyridine monomer [9].

Hydrolysis of *S*-pyridylethylated protein with PPIV was performed in 0.1 M potassium phosphate buffer, pH 6.5, containing 10 mM dithiothreitol. After 16 h of digestion at room temperature (about 20°C), the reaction was stopped by the addition of conc trifluoroacetic acid to a final concentration of 20% (v/v). All other conditions for the digestion were as reported [10].

Reversible blocking of amino groups with maleic anhydride, and unblocking, were as described [11].

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

Peptides were purified by gel chromatography, and HPLC on the Chrompack column eluted with aqueous acetonitrile containing trifluoroacetic acid.

The reduced and alkylated protein and peptides were hydrolyzed with 6 M HCl at 110°C for 24 and 72 h. Amino acid analysis was done with the Beckman 118CL analyser, by post-column fluorescence detection after reaction with *o*-phthalaldehyde.

Samples were sequenced with an Applied Biosystems 475A liquid-phase sequenator. Phenylthiohydantoin derivatives were identified on-line with the 120A HPLC [12].

3. RESULTS AND DISCUSSION

The native, active inhibitor gave a single sequence up to 43 residues (Fig. 1). The first set of peptides was generated by digestion of the reduced, alkylated protein with the glycine-specific proteinase, PPIV. The alkylated inhibitor was insoluble, but nevertheless, PPIV completed the hydrolysis of the molecule so that all of the peptides except G5 + G6 and G6 were found in the soluble fraction of the digest, in an equimolar ratio (Fig. 1). Peptides G5 + G6 and G6 were insoluble, but were dissolved in 50% trifluoroacetic acid, and successfully sequenced to give the structure of the very hydrophobic core of the inhibitor molecule. The yield of peptides G5 and G6 was about 60% of that of the uncleaved G5 + G6. CNBr cleavage at the single Met residue yielded two peptides, CN1 and CN2. Peptide CN1 was maleylated for arginine-specific tryptic cleavage, and peptide CN2 was digested directly with trypsin. The sequences of the resulting peptides (R1–R4 from CN1, and T1–T4 from CN2) overlapped those of the G-series of peptides from PPIV digestion, and showed their order (Fig. 1). Peptides G9 and T4 were concluded to represent the C-terminus, since the C-terminal Val residue would not have been a site of cleavage by trypsin.

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

Fig. 1. Amino acid sequence and composition of inhibitor and strategy of the sequence determination. The amino acid sequence of *S*-pyridylethylated inhibitor was determined by automatic Edman degradation of the whole molecule (intact protein), the 9 glycine-specific proteinase PPIV peptides (G1–G9), 2 CNBr peptides (CN1 + CN2), 4 tryptic peptides from the maleylated CN1 peptide (R1–R4) and 4 tryptic peptides deriving from tryptic hydrolysis of CN2 peptide (T1–T4). The residues not directly identified in the individual peptides are marked (X). Asn-19 is the point of attachment of carbohydrate (indicated *).

The results show that inhibitor exists as a single polypeptide chain with 187 residues, M_r 20450 (assuming 3 disulphide bonds) (Fig. 1). Residue 19 was the only position in which no normal phenylthiohydantoin derivative was detected. Hydrolysis of this derivative yielded aspartic acid, suggesting that it represents a glycosylated asparaginyl residue. This is consistent with the fact that the sequence Asn-Ser-Ser (residues 19–21) corresponds to a potential glycosylation site, the only one in the molecule.

The alignment of the sequence of the novel inhibitor of cathepsin D (NID) with those of related proteins (Fig. 2) shows many residues identical with those of potato cathepsin D inhibitor (PDI) (93%), miraculin (MRC) (30%), and soybean trypsin inhibitor (STI) (20%). The statistical test of relationship of the sequences as described in [16] shows them to be homologous, so it can be concluded that NID, PDI and miraculin are all members of the soybean trypsin inhibitor superfamily of proteins.

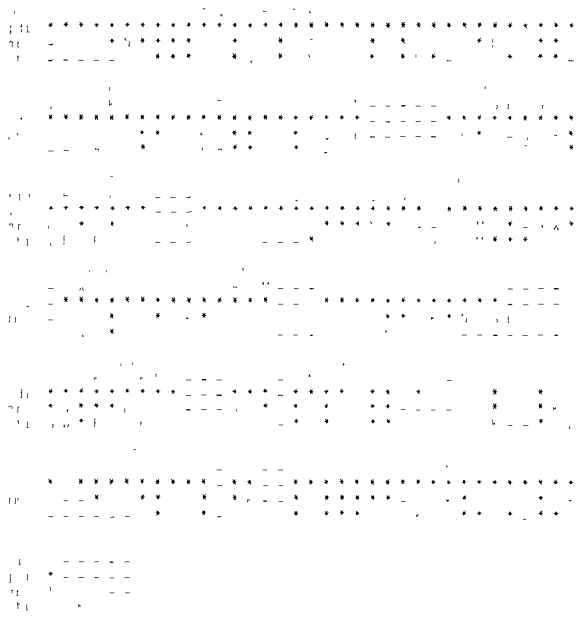

Fig. 2 Alignment of the sequence of the novel inhibitor of cathepsin D (NID) sequence with those of other plant proteins. Residues in the order sequences that are identical with those in NID are marked *, and the numbering is according to the sequence of NID. Key: nid, NID; pdi, cathepsin D inhibitor from potatoes [5]; mrc, miraculin [13]; and sti, soybean trypsin inhibitor [14]. Alignment was done as described in [15,16].

NID inhibits both trypsin and cathepsin D. The P_1 reactive site residue of STI is Arg-67 (in the numbering of Fig. 2), and the same is likely to be true of NID and PDI. STI does not inhibit cathepsin D (unpublished data). Therefore one could speculate that the reactive site for inhibition of cathepsin D is in the part of the sequence (residues 140–160) in which NID and PDI differ markedly from STI, and which apparently contains two disulfide loops.

Acknowledgement: 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. and Salvesen, G. (1986) Proteinase Inhibitors, Elsevier, Amsterdam
[2] Abu-Erreish, G.M. and Peanasky, R.J. (1974) J. Biol. Chem. 249, 1558–1569.
[3] Keilova, H. and Tomašek, V. (1976) Collect. Czech. Chem. Commun. 41, 489–497.
[4] Rupova, L., Keilova, H. and Tomašek, V. (1977) Collect. Czech. Chem. Commun. 42, 2279–2286
[5] Mareš, M., Meloun, B., Pavlik, M., Kostka, V. and Baudyš, M. (1989) FEBS Lett. 251, 94–98

[6] Meško, P., Kopitar, M., Ritonja, A., Križaj, I., Grebenc, M., Asmelash, B. and Turk, V. (1989) in: Proteinases and their Inhibitors: Recent Developments (Auerswald, E., Fritz, H. and Turk, V. eds) pp. 99–101, Kernforschungsanstalt Jülich GmbH, Julich.
[7] Buttle, D.J., Kembhavi, A.A., Sharp, S.L., Schute, R.E., Rich, D.H. and Barrett, A.J. (1989) Biochem. J. 261, 469–476.
[8] Strop, P. and Cechova, D. (1981) J. Chromatogr. 207, 55–62.
[9] Henschen, A. (1986) in: Advanced Methods in Protein Microsequence Analysis, pp. 244–255, Springer, Berlin.
[10] Buttle, D.J., Ritonja, A., Pearl, L.H., Turk, V. and Barrett, A.J. (1990) FEBS Lett. 260, 195–197.
[11] Butler, P.J.G. and Hartley, B.S. (1972) Methods Enzymol. 25, 191–199.
[12] Hunkapiller, M.W. and Hood, L.E. (1983) Methods Enzymol. 91, 486–493.
[13] Theerasilp, S., Hitotsuya, H., Nakajo, S., Nakaya, K., Nakamura, Y. and Kurihara, Y (1989) J. Biol. Chem. 264, 6655–6659.
[14] Kim, S.H., Hara, S., Hase, S., Ikemaka, T., Toda, H., Kitamura, K. and Kaizuma, N. (1985) J. Biochem. 98, 435–448.
[15] Dayhoff, M.O., Barker, W.C. and Hunt, L.T. (1983) Methods Enzymol. 91, 524–549.
[16] Rawlings, N.D. and Barrett, A.J. (1990) Biochem. J. 266, 622–624.