The acid-stable proteinase inhibitor of human mucous secretions (HUSI-I, antileukoprotease)

Complete amino acid sequence as revealed by protein and cDNA sequencing and structural homology to whey proteins and Red Sea turtle proteinase inhibitor

Ursula Seemüller, Marianne Arnhold, Hans Fritz, Karin Wiedenmann*, Werner Machleidt*, Regina Heinzel+, Heribert Appelhans+, Hans-Günter Gassen+ and Friedrich Lottspeich°

*Abteilung für Klinische Chemie und Klinische Biochemie in der Chirurgischen Klinik Innenstadt, Universität München, Nussbaumstrasse 20, D-8000 München 2, *Institut für Physiologische Chemie, Physikalische Biochemie und Zellbiologie, Universität München, Goethestrasse 33, D-8000 München 2, +Institut für Organische Chemie und Biochemie der Technischen Hochschule Darmstadt, Petersenstrasse 22, D-6100 Darmstadt* and *°Max-Planck-Institut für Biochemie, Am Klopferspitz, D-8033 Martinsried, FRG*

Received 21 January 1986

The complete amino acid sequence of human antileukoprotease has been determined by direct sequencing of the inhibitory active protein isolated from seminal plasma (HUSI-I) and by sequence analysis of cDNA reverse-transcribed from mRNA prepared from cervical tissue. The inhibitor (M_r 11 726) consists of 107 amino acid residues including 16 cysteines presumably forming disulfide bonds. The molecule comprises two consecutive domains which are homologous to each other, to the second domain of the basic protease inhibitor from Red Sea turtle (chelonianin) and to both domains of the whey proteins of rat and mouse. Both domains contain a pattern of cysteines known as the 'four-disulfide-core' that has also been found in wheat germ agglutinin and neurophysin.

Antileukoprotease *HUSI-I* *Mucous secretion* *Proteinase inhibitor* *Sequence homology* *Whey protein*

1. INTRODUCTION

Human mucous fluids, such as seminal plasma, cervical mucus, bronchial and nasal secretions and tears, contain acid-stable proteinase inhibitors with strong affinity for trypsin and chymotrypsin as well as for neutrophil lysosomal elastase and cathepsin G [1–5]. Isolation of these inhibitors is complicated by the occurrence of multiple forms probably resulting from limited proteolysis at an early stage of sample collection and handling [3–7]. Most of the secretions contain inflammatory cells including neutrophils liberating lysosomal proteinases during their disintegration [3,8]. The inhibitors may be further modified by limited cleavage(s) when isolated via trypsin or chymotrypsin immobilized on a water-insoluble matrix [5–8]. Limited proteolytic degradation of the virgin inhibitor produced and secreted by epithelial mucus cells [3] could explain the discrepancies between several recent reports on biochemical and physicochemical properties of the mucus-derived so-called antileukoproteases [1–5,9].

The antileukoprotease of human seminal plasma, HUSI-I (human seminal proteinase inhibitor I), has been most extensively characterized [1,6–8]. HUSI-I proved to be similar to the antileukoproteases from other mucous secretions by the following criteria: acid stability, inhibition

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50

spectrum and molecular mass [3], immunological cross-reactivity [3,7,8], amino acid composition [3] and partial amino acid sequences [10]. We have now been able to establish the complete amino acid sequence of HUSI-I by comparison of protein sequences obtained from inhibitory active material with a cloned cDNA sequence reverse-transcribed from mRNA isolated from human cervical tissue. Surprisingly, HUSI-I shows a close structural homology to whey proteins from rat and mouse of so far unknown function, and to the basic protease inhibitor from Red Sea turtle.

2. EXPERIMENTAL

2.1. *Isolation of HUSI-I*

HUSI-I was isolated from human seminal plasma as described by Schiessler et al. [6,7] and further purified by reversed-phase high-performance liquid chromatography (HPLC) on a Supelco LC-18-DB column using a gradient from 0.05% trifluoroacetic acid in water to 0.05% trifluoroacetic acid in acetonitrile. Part of the material was repurified by affinity chromatography on a column of chymotrypsin-Sepharose essentially as in [5].

2.2. *Amino acid sequence analysis*

S-carboxymethylation of the reduced inhibitor, performic acid oxidation, cyanogen bromide cleavage, citraconylation and cleavage with trypsin were performed following standard protocols. Peptides were separated by gel chromatography on Bio-Gel P-30 in 50% acetic acid (cyanogen bromide fragments) or Sephadex G-50 in 0.1 M ammonium bicarbonate (tryptic fragments) followed by reversed-phase HPLC using a similar system as described above.

Amino acid sequences were determined by automated solid-phase Edman degradation in a non-commercial sequencer performing on-line detection of the released amino acid phenylthiohydantoin derivatives (PTH) by HPLC. Prior to sequencing the peptides were coupled to *p*-phenylene diisothiocynate-activated aminopropyl glass and/or to aminopropyl glass using hydroxybenzotriazole-catalysed carbodiimide activation in anhydrous dimethyl formamide following the procedures described [11].

2.3. *cDNA cloning and sequence analysis*

mRNA was prepared from human cervix uteri tissue samples, enriched by oligo(dT) chromatography and reverse-transcribed into cDNA [12]. The double-stranded cDNA was tailed with dCTP and annealed to the plasmid pBR322 dG-tailed at its *Pst*I site. After transformation of *E. coli* DH 1 strain [13] the tetracycline-resistant colonies were screened by colony hybridisation [14] with a chemically synthesized deoxyoligonucleotide (mixed probe) derived from the C-terminal portion of the known protein sequence. From the plasmid DNA of the positive colonies *Pst*I fragments were prepared, subcloned into pUC18 and sequenced by the Maxam and Gilbert technique. Details of this work will be published separately (Heinzel, R. et al., in preparation).

2.4. *Sequence comparisons*

Sequence comparisons with the contents of the Protein Sequence Database provided by the National Biomedical Research Foundation, Washington, DC (released on 25 February 1985) were made with a VAX computer using the SEARCH and RELATE programs [15].

3. RESULTS

3.1. *Amino acid sequence determination*

Sequence analysis (fig.1) revealed that the isolated active inhibitor was not an intact polypeptide chain but rather consisted of 2–3 preformed proteolytic fragments. A mixture of two parallel sequences starting with Ser (n-1) and Tyr (n-2) was obtained by direct Edman degradation of this material. N-terminal sequencing of the performic acid-oxidized inhibitor yielded a mixture of the same two sequences (p-1, p-2) plus an additional sequence obviously resulting from acidolytic cleavage of the labile Asp-49–Pro-50 peptide bond (p-3). Attempts to separate the preformed proteolytic fragments by gel filtration and reversed-phase HPLC were only partially successful. A pure fragment was obtained, however, comprising the N-terminus of the inhibitor molecule (p-4). A second pure fragment yielded a homogeneous sequence extending from Met-73 to close to the C-terminal end of the molecule (c-3).

The sequence of residues 36–107 was determined from tryptic fragments (t-1–t-5) of the per-

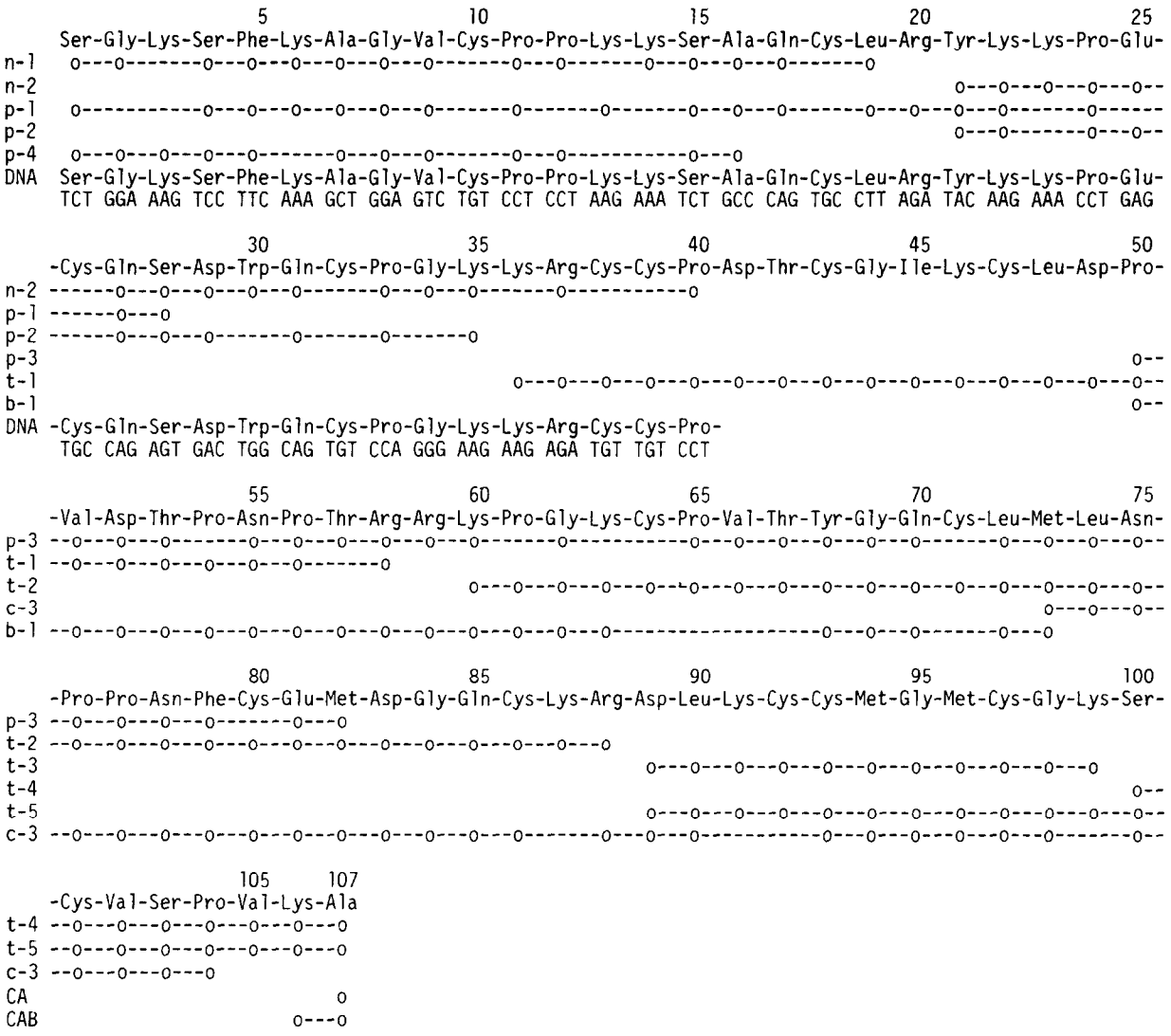

Fig.1. Amino acid sequence of HUSI-I as obtained by protein and cDNA sequencing. Protein sequences (O---O, wherein each 'O' denotes a residue that was unequivocally identified) were determined by automated Edman degradation of the whole inhibitor and its preformed proteolytic fragments (n-1, n-2, p-1, p-2, p-4, c-3), an acidolytic fragment (p-3), an acidolytically cleaved cyanogen bromide fragment (b-1), and tryptic fragments of the citraconylated inhibitor (t-1–t-5). The C-terminal end was identified by degradation with carboxypeptidases A (CA) and A plus B (CAB). Parallel sequences in the N-terminal region (residues 1–40) obtained from mixtures of unresolved fragments (n-1 plus n-2 and p-1 plus p-2 plus p-3) were aligned with the aid of the corresponding part of the cDNA-derived sequence (DNA).

formic acid-oxidized and citraconylated inhibitor. They were overlapped by an acidolytically cleaved cyanogen bromide fragment of the *S*-carboxymethylated protein (b-1) and the preformed chymotryptic fragment mentioned above (c-3). The C-terminal end of the inhibitor was confirmed by the results of carboxypeptidase digestions (cf. fig.1).

Unambiguous alignment of the various parallel sequences obtained from the N-terminal portion of the molecule (residues 1–40) was not possible without the information derived from the cor-

responding cDNA sequence (DNA in fig.1). The cDNA-coded amino acid sequence aligned the parallel sequences determined from mixtures of incompletely resolved fragments and is entirely consistent with the results of direct protein sequencing. Furthermore, comparison of the cDNA-derived and the direct protein sequences revealed that the cloned piece of cDNA codes for the total HUSI-I protein including a 25-residue leader sequence preceding the sequence of the mature protein. The codon for the C-terminal Ala residue of HUSI-I is followed by a stop codon (unpublished).

4. DISCUSSION

4.1. *Homogeneity of isolated active inhibitor preparations*

The HUSI-I preparation isolated in several batches could be reproducibly separated into 4 major and some minor inhibitory active forms. The major forms appeared homogeneous by the criteria of gel and ion-exchange chromatography, end-group analysis and polyacrylamide gel electrophoresis [6,7]. However, after reduction and separation in SDS gel electrophoresis a similar pattern of bands was always obtained corresponding to molecular masses of approx. 10.8 kDa (major band mostly) as well as 5.7, 4.3 and 3.4 kDa [7]. Obviously the secreted virgin inhibitor molecule is already proteolytically modified in the mucous fluids in vivo and during the course of sample collection and the first purification steps (cf. section 1). This view is supported by recent experiments showing that isolation of HUSI-I via immobilized anhydrotrypsin or HUSI-I-directed antibodies did not significantly change the protein band pattern in SDS electrophoresis (unpublished). It seems that limited degradation of the inhibitor in vivo cannot be avoided, causing extreme difficulties in the isolation of protein-chemically homogeneous material in amounts sufficient for detailed biochemical and structural studies.

In view of these difficulties our approach was the combination of protein sequence data obtained from a mixture of multiple inhibitory active forms with the amino acid sequence derived from a cloned fragment of cDNA transcribed from mRNA isolated from cervical tissue. This strategy enabled us to elucidate the complete amino acid sequence of mature HUSI-I. Furthermore, the correspondence of the cDNA-derived and directly determined amino acid sequences indicates that the inhibitor synthesized in the cervical glands is structurally the same as that isolated from seminal plasma. From the close similarity of these inhibitors to the antileukoprotease from bronchial secretion [2–5,10] it can be expected that the latter inhibitor has basically the same structure, and very probably this also holds true for the antileukoproteases found in nasal secretion, tears

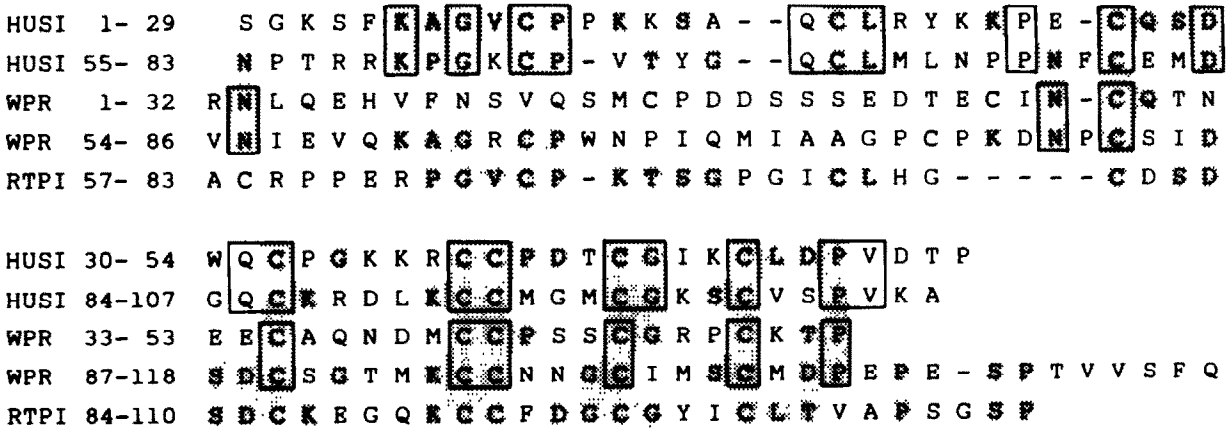

Fig.2. Sequence homologies between the two domains of HUSI-I, the two internal repeats of rat whey phosphoprotein, WPR [16], and the second domain of the basic protease inhibitor of Red Sea turtle, RTPI [18]. Residues identical among the internal repeats are boxed, sequence identities among the 3 proteins are shaded. Gaps (indicated by dashes) have been allowed to improve the alignment.

and the gut [3,8]. It may be assumed that the HUSI-I-type antileukoproteases found in human mucous secretions are encoded by a single gene expressed in all mucus epithelial cells.

4.2. *Two-domain structure of HUSI-I*

The HUSI-I sequence contains two consecutive domains of equal length which are homologous to each other (fig.2) suggesting that the present inhibitor has evolved by duplication of a single ancestral gene. Both domains contain the same characteristic pattern of 8 cysteine residues (see below), all of which are presumably involved in disulfide bonds. No direct experimental evidence is yet available for the location of the reactive site(s) on the two domains. Most probable candidates for P_1 and P_1' residues are Arg-20–Tyr-21 (first domain) and Leu-72–Met-73 or Met-73–Leu-74 (second domain) where limited cleavage of the inhibitor has been observed. One may speculate that these residues are part of an antitryptic site in the first domain and an anti-elastolytic and/or antichymotryptic site in the second domain.

4.3. *Sequence homology of HUSI-I with whey proteins and the basic protease inhibitor from Red Sea turtle*

A computer search revealed significant sequence homologies of HUSI-I with whey proteins of rat [16] and mouse [17] and with the second domain of the basic protease inhibitor from the Red Sea turtle, also known as chelonianin [18]. A striking identity is found in the pattern of cysteine residues (fig.2). Besides in chelonianin and whey proteins, a similar arrangement of cysteines has been found in several proteins of different (or unknown) biological function such as wheat germ agglutinin, snake neurotoxins, ragweed pollen allergen Ra5, and neurophysin. The so-called 'four-disulfide core' seems to be a common pattern of cysteine residues forming a typical 3-dimensional fold of disulfide loops which is supposed to stabilize a domain too small to have a sufficient hydrophobic core [19]. The possibility that these structures may be involved in binding to a membrane-bound receptor molecule should also be considered for the antileukoproteases.

Chelonianin is a two-domain proteinase inhibitor, the first domain being of the Kunitz type whereas the second, subtilisin-inhibiting domain could not be assigned to any of the existing inhibitor families [18]. It is this second domain that is homologous to the two domains of HUSI-I and, as was known before [16], to both domains of the whey proteins. Attempts to detect an inhibitory activity of rat whey protein have failed [16]. This would be consistent with the fact that practically no homology with HUSI-I is found in the putative active-site regions, but the target enzymes of the whey proteins may be different from the proteinases so far tested.

In conclusion, HUSI-I is a two-domain inhibitor constituting a new family of protein inhibitors of proteinases also comprising the second domain of chelonianin. An outstanding structural feature of this new inhibitor family is the pattern of cysteine residues underlying the four-disulfide core and shared with a group of small proteins of very different biological function.

ACKNOWLEDGEMENTS

This work was supported by grants (C-2, D-2) of the Sonderforschungsbereich 207 of the University of Munich. We are grateful to Professor H. Graeff, Technical University of Munich (SFB 207, grant A-2), for supplying us with cervical tissue samples.

REFERENCES

[1] Schiessler, H., Arnhold, M., Ohlsson, K. and Fritz, H. (1976) Hoppe-Seyler's Z. Physiol. Chem. 357, 1251–1260.
[2] Ohlsson, K., Tegner, H. and Akesson, U. (1977) Hoppe-Seyler's Z. Physiol. Chem. 358, 583–589.
[3] Schiessler, H., Hochstrasser, K. and Ohlsson, K. (1978) in: Neutral Proteases of Human Polymorphonuclear Leucocytes (Havemann, K. and Janoff, A. eds) pp.195–207, Urban & Schwazenberg, Baltimore.
[4] Kueppers, F. and Bromke, B.J. (1983) J. Lab. Clin. Med. 101, 747–757.
[5] Smith, C.E. and Johnson, D.A. (1985) Biochem. J. 225, 463–472.
[6] Schiessler, H., Fink, E. and Fritz, H. (1976) Methods Enzymol. 45, 847–859.
[7] Schiessler, H. (1976) Adv. Androl. 5, 45–102.
[8] Fritz, H., Schiessler, H., Schill, W.-B., Tschesche, H., Heimburger, N. and Wallner, O. (1975) in: Proteases and Biological Control (Reich, E. et al. eds) pp.737–766, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

[9] Hochstrasser, K., Albrecht, G.J., Schoenberger, Oe.L., Rasche, B. and Lempart, K. (1981) Hoppe-Seyler's Z. Physiol. Chem. 362, 1369–1375.
[10] Klasen, E.C. and Kramps, J.A. (1985) Biochem. Biophys. Res. Commun. 128, 285–289.
[11] Machleidt, W. (1983) in: Modern Methods in Protein Chemistry (Tschesche, H. ed.) pp.267–302, De Gruyter, Berlin.
[12] Gubler, U. and Hoffman, E.J. (1983) Gene 25, 263–269.
[13] Hanahan, D. (1983) J. Mol. Biol. 166, 557–580.
[14] Wallace, R.B., Johnson, M.J., Hirose, T., Miyake, T., Kawashima, E.H. and Itakura, K. (1981) Nucleic Acids Res. 9, 879–894.
[15] Dayhoff, M.O., Barker, W.C. and Hunt, L.T. (1983) Methods Enzymol. 91, 524–545.
[16] Dandekar, A.M., Robinson, E.A., Appella, E. and Qasba, P.-K. (1982) Proc. Natl. Acad. Sci. USA 79, 3987–3991.
[17] Hennighausen, L.G. and Sippel, A.E. (1982) Nucleic Acids Res. 10, 2677–2684.
[18] Kato, I. and Tominaga, N. (1979) Fed. Proc. 38, 832 (Abstr.).
[19] Drenth, J., Low, B.W., Richardson, J.S. and Wright, C.S. (1980) J. Biol. Chem. 255, 2652–2655.