Volume 229, number 1, 63–67 FEB 05610 February 1988

Complementary DNA and derived amino acid sequence of the precursor of one of the three protein components of the inter-α-trypsin inhibitor complex

Wolfgang Gebhard, Thomas Schreitmüller, Karl Hochstrasser* and Elmar Wachter+

*Abteilung für Klinische Chemie und Klinische Biochemie in der Chirurgischen Klinik Innenstadt der Universität München, Nußbaumstr. 20, D-8000 München 2, *Klinik und Poliklinik für Hals-, Nasen- und Ohrenkranke der Universität München, Klinikum Großhadern, Marchioninistr. 15, D-8000 München 70* and *+Institut für Physiologische Chemie, Physikalische Biochemie und Zellbiologie der Universität München, Goethestr. 33, D-8000 München 2, FRG*

Received 4 January 1988

Inter-α-trypsin inhibitor is composed of three distinct protein components. These protein components stem from independently encoded and proteolytically processed precursor proteins. Only the structure of the protein component responsible for the inhibitory activity has been established so far. We now present the complete amino acid sequence of the precursor of the second protein component derived from cloned cDNA. The precursor molecule includes both a signal peptide and a propeptide sequence and seems to be further processed prior to the assembly of the inter-α-trypsin inhibitor complex.

Proteinase inhibitor; Inter-α-trypsin inhibitor; Propeptide; Posttranslational processing; (Human serum)

1. INTRODUCTION

For years it has been recognized that inter-α-trypsin inhibitor (ITI) is a single-chain protein (review [1]). However, recently we could show that peptides obtained from purified ITI derive from proteins encoded on three different mRNAs [2]. Therefore, ITI is rather a complex of three different protein components than a single-chain protein. The inhibitor part of ITI originally perceived as its N-terminal portion and called H130 was found to be encoded on a 1.2 kb mRNA [2] identical to that described for the urinary trypsin inhibitor (UTI) [3]. The amino acid sequences of H130 as well as of UTI, and probably also of serum trypsin inhibitor (STI) [4], are identical. All these molecules are formed by proteolytic processing of one and the same precursor protein with N-terminal α_1-microglobulin and C-terminal Kunitz-type inhibitor domains. Only the inhibitor domains become part of the mature ITI complex.

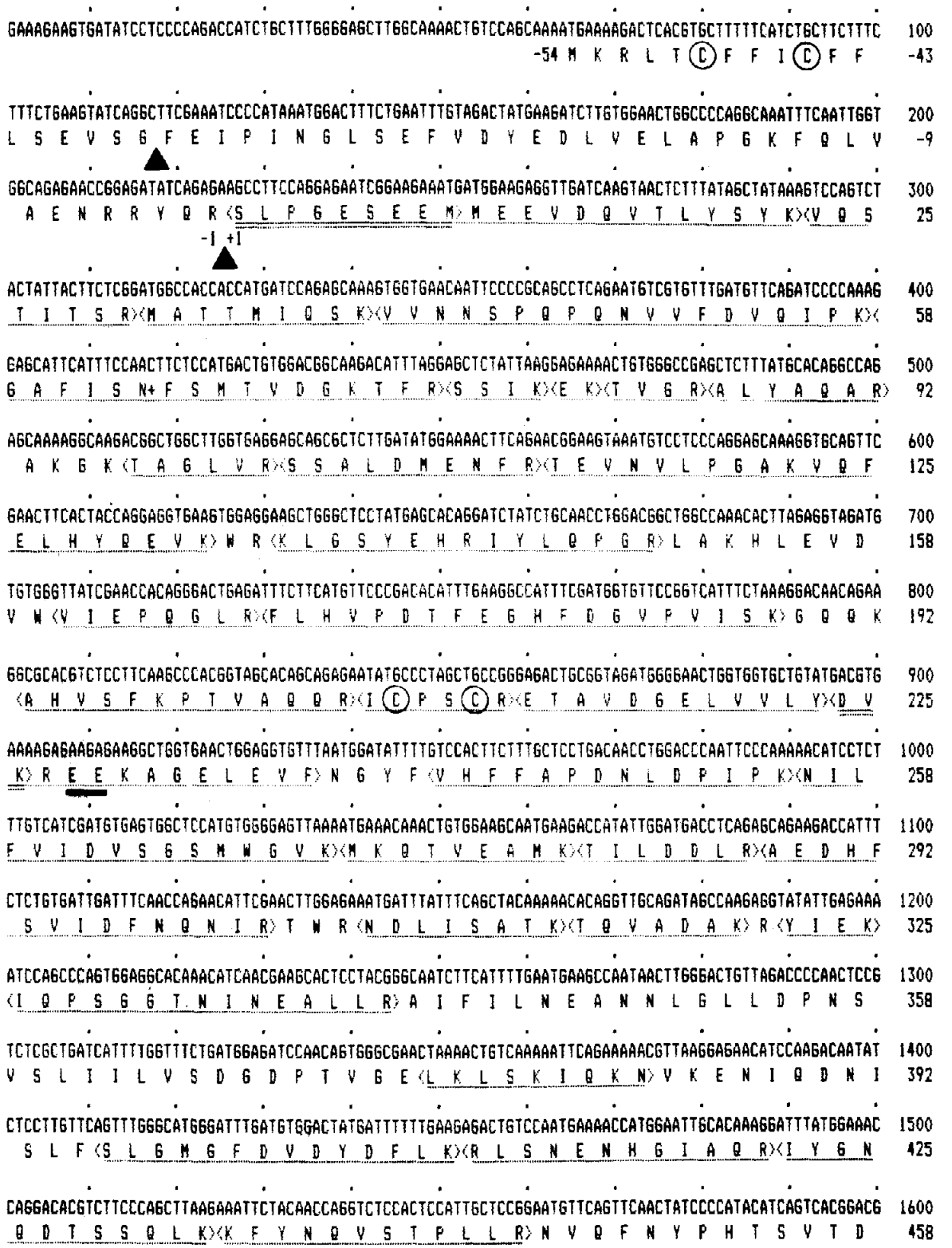
The complete amino acid sequence of the precursor of the second protein component of the ITI complex now could be derived from its cDNA sequence and confirmed by partial amino acid sequencing of purified human ITI.

Correspondence address: W. Gebhard, Abteilung für Klinische Chemie und Klinische Biochemie in der Chirurgischen Klinik Innenstadt der Universität München, Nußbaumstr. 20, D-8000 München 2, FRG

2. MATERIALS AND METHODS

The same human liver cDNA library from which we isolated a partial clone [2] was further screened with a 153 bp *Eco*RI-*Cla*I DNA fragment (fig.1) obtained by digestion of the corresponding DNA, separation on a 4% polyacrylamide gel, electroelution with a BIOTRAP (Schleicher & Schuell) according to the manufacturers protocol, and labeling using the oligolabeling kit of Boehringer Mannheim. After a prehybridisation step with 3 × SSC, 3 × Denhardt, 0.1% SDS and salmon

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50



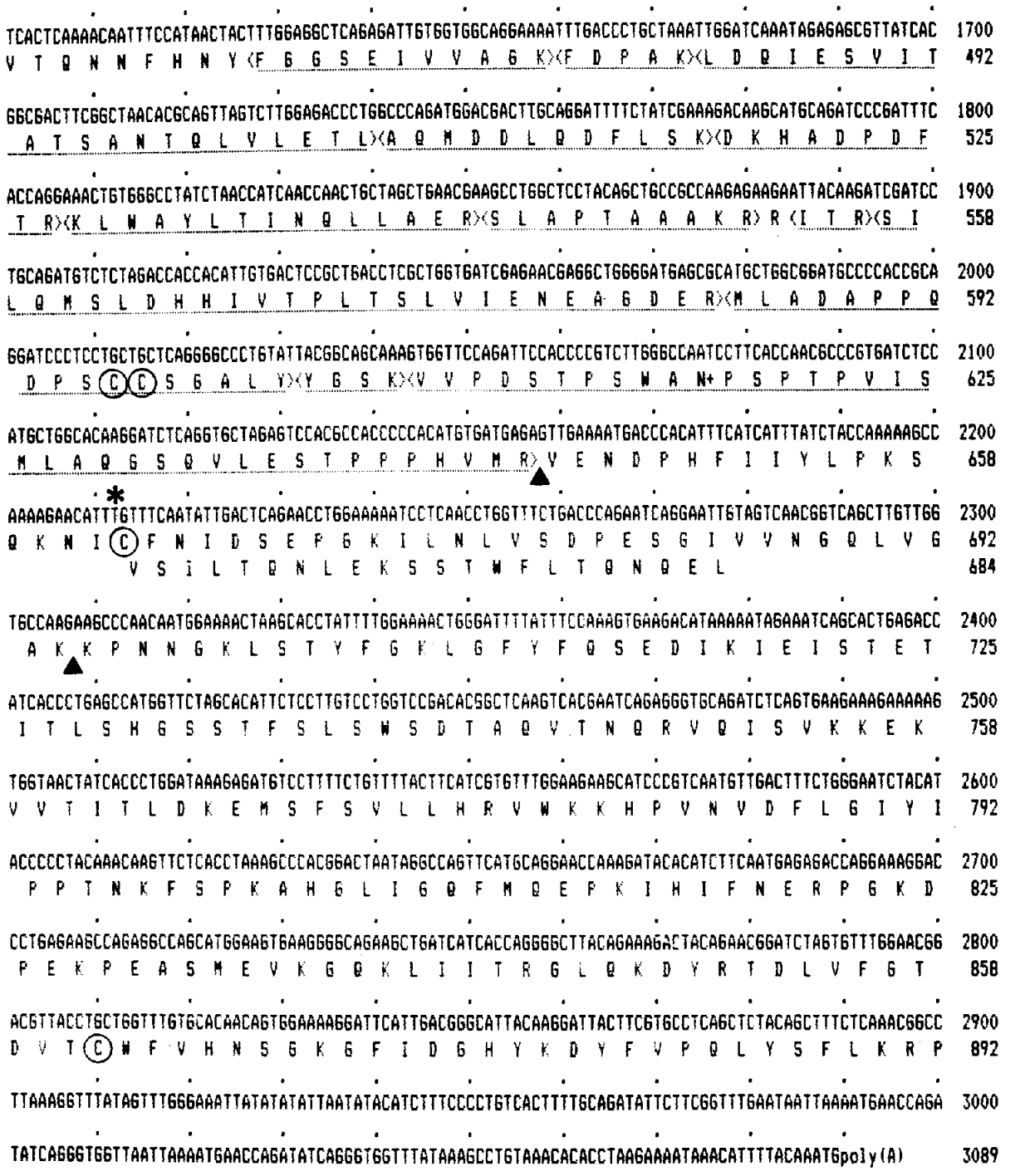


Fig.1. cDNA-derived complete amino acid sequence of the precursor of the second protein component of ITI. Partially or completely amino acid sequenced peptides obtained by cleavage of purified ITI with either trypsin, chymotrypsin, *Staphylococcus aureus* V8 proteinase or CNBr are underlined and shown in brackets < >. * marks a deleted T in one of the clones. The resultant alternative reading frame is shown in the line below. N+: Asn-*N*-acetylgalactosamine. 4-Carboxyglutamic acid residues are underlined by a thick bar. Signal peptide and propeptide cleavage sites as well as potential C-terminal processing sites are indicated by triangles.

sperm DNA at a concentration of 200 μg/ml for 2 h at 68°C plaque hybridisation was performed in the same buffer overnight at a temperature gradient from 68 to 37°C. The applied radioactivity was 1×10^6 cpm/ml. The nitrocellulose filters (BA85, Schleicher & Schuell) were washed at room temperature twice with 3 × SSC, 0.1% SDS, twice with 1 × SSC, 0.1% SDS and once with 0.1 × SSC, 0.1% SDS [1 × SSC: 0.15 M NaCl/0.015 M sodium citrate (pH 7); 1 × Denhardt: bovine serum albumin, ficoll, and polyvinylpyrrolidone, 0.02% each]. Autoradiography was performed with Kodak XAR film at −70°C using DuPont amplifier screens overnight. Positive plaques were rescreened once.

Sequence data were collected and analysed using the MicroGenie® Sequence Analysis Program [5], version 4.0, and GenBank® and NBRF sequence databanks, releases 10/87.

Other materials and methods were as described [2].

3. RESULTS

cDNA sequencing of the full-length cDNA clone coding for α_1-microglobulin and the Kunitz inhibitor domains revealed that the 5′-untranslated region of the cDNA is 15 bp longer than that reported in the literature [3]. The cDNA starts with the sequence GGCTCTTCTC GTTGC followed by the published sequence and therefore has a length of 1228 bp preceding the poly(A) tail.

The cDNA coding for the second protein component of the inter-α-trypsin inhibitor complex could be derived from two overlapping clones and comprises 3089 bp and a poly(A) tract. The cDNA sequence (fig.1) deviates in three positions (857, 864, 917) from the partial sequence previously reported (fig.1 in [2], positions 1, 8 and 62). These deviations were recognized as part of the cloning linker sequence, as a correctly determined but artificially surplus nucleotide, and an error in typing, respectively.

After a 5′-non-coding region of 63 bp a large open reading frame of 2838 bp encodes a protein of 946 amino acid residues with a calculated molecular mass of 106 447 Da. This is in relatively good agreement with the observed value of about 95 kDa obtained after in vitro translation of baboon liver mRNA and immunodetection of ITI precursor proteins [6].

There is no striking homology to any other protein in the database we used. The significance of some short and distant relationships of parts of the molecule to other proteins or parts thereof is difficult to judge without further information.

4. DISCUSSION

Eukaryotic signal peptides vary considerably with respect to length and sequence. The main common denominator appears to be a more basic N-terminal region and a central clustering of hydrophobic amino acids followed by a more polar C-terminal region. Analysis of this common pattern allowed the deduction of some rules for the prediction of the cleavage site between a signal peptide and the exported protein [7]. According to these rules and the prediction of a characteristic β-turn at this position in the precursor of the second ITI component cleavage after the glycine residue −37 is most likely (fig.2). The signal peptide contains an unusually high content of phenylalanine rather than leucine residues.

After CNBr cleavage of purified ITI, isolation and amino acid sequencing of the peptides, the N-terminus of the mature second ITI component is clearly identified by the absence of a methionine residue in the precursor molecule upstream of the peptide S-L-P-G-E-S-E-E-M (fig.2). Therefore, the amino acid sequence from positions −36 to −1 must represent a propeptide. Processing just behind the arginine residue in position −1 would be plausible. Actually, β-turns are predicted at this site (fig.2).

Obviously, further proteolytic processing occurs. Peptides, enzymatically derived from isolated ITI, are densely distributed over only part of the coding region but do not cover the large C-terminal stretch of the precursor protein. Although the C-terminal amino acid of the mature protein could not yet be determined with certainty, of the potential mono- or dibasic cleavage sites [8] the arginyl bond at position 644 seems to be a conceivable candidate for processing by the same trypsin-like enzyme which cleaves off the propeptide. The mature protein would thus have 4 cysteine residues arranged in pairs.

One of four clones covering the DNA sequence region around nucleotide position 2212 (asterisk in fig.1) clearly has a deletion of a T in this position as verified by sequencing in both strands. This leads to an interruption of the altered open reading frame shortly after this event (second reading frame in fig.1, positions 663–684). The existence of such a precursor protein remains to be proven. C-terminal processing would be possible at the same

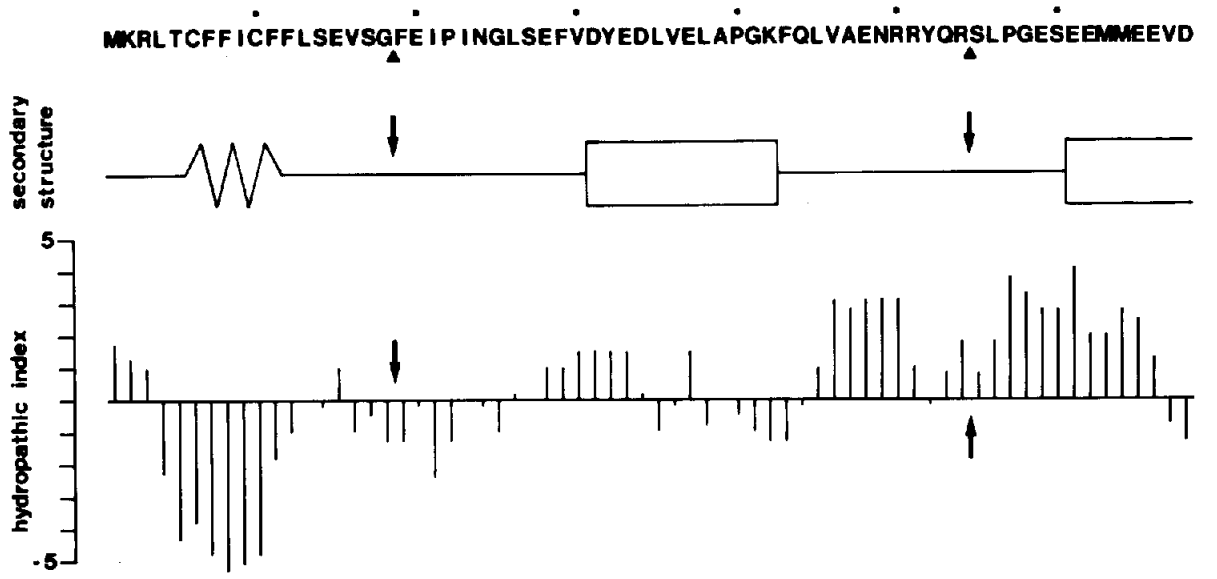


Fig.2. Signal peptide, propeptide and N-terminal protein sequence of the mature second ITI protein, including its secondary structure plot [9] and hydropathy plot [10]. The boxes, zig-zags, flat lines and arrows represent regions of α-helices, β-strands, random coils and β-turns, respectively.

amino acid residue thus producing the same mature protein as discussed above.

The proteolytically trimmed nude second protein of the ITI complex would have a calculated molecular mass of 66 323 Da. Like the first protein component of ITI which represents the Kunitz-type inhibitor domains the second protein component is also glycosylated. There are three putative glycosylation sites (N-X-T/S) at amino acid residues 42, 64 and 617. Only two (64, 617) were found to be derivatized (fig.1). In experiments to separate ITI into its components one can observe values of 78 and 85 kDa for those parts of the ITI complex that do not represent the Kunitz-type inhibitor domains [11]. This is in accordance with the present data.

Acknowledgements: This research was supported by the Bundesminister für Forschung und Technologie, 03 8578 3, and Bayer AG, Wuppertal. We thank Professor H. Fritz for his valuable advice and support, and G. Leysath, G. Zuckriegl, G. Schweiger and I. Rossmanith for expert technical assistance.

REFERENCES

[1] Gebhard, W. and Hochstraßer, K. (1986) in: Proteinase Inhibitors (Barrett, A.J. and Salvesen, G. eds) pp. 389–401, Elsevier, Amsterdam, New York.
[2] Schreitmüller, T., Hochstraßer, K., Reisinger, P.W.M., Wachter, E. and Gebhard, W. (1987) Biol. Chem. Hoppe-Seyler 368, 963–970.
[3] Kaumeyer, J.F., Polazzi, J.O. and Kotick, M.P. (1986) Nucleic Acids Res. 14, 7839–7849.
[4] Maruyama, M., Yamamoto, T., Sumi, H., Tsushima, H., Mihara, H. and Minamino, N. (1986) Enzyme 35, 225–231.
[5] Queen, C. and Korn, L.J. (1984) Nucleic Acids Res. 12, 581–599.
[6] Bourguignon, J., Vercaigne, D., Sesboüé, R., Martin, J.P. and Salier, J.-P. (1983) FEBS Lett. 162, 379–383.
[7] Von Heijne, G. (1986) Nucleic Acids Res. 14, 4683–4690.
[8] Schwartz, T.W. (1986) FEBS Lett. 200, 1–10.
[9] Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) J. Mol. Biol. 120, 97–120.
[10] Hopp, T.P. and Woods, K.R. (1981) Proc. Natl. Acad. Sci. USA 78, 3824–3828.
[11] Gebhard, W. and Schreitmüller, T. (1988) Biol. Chem. Hoppe-Seyler, submitted.