Molecular cloning and sequence analysis of the cDNA for ancrod, a thrombin-like enzyme from the venom of *Calloselasma rhodostoma*

Lo-Chun AU,*†§ Shwu-Bin LIN,‡ Jeng-Sheng CHOU,‡ Go-Wai TEH,* Kuo-Jui CHANG† and Chia-Min SHIH‡

*Department of Medical Research, Veterans General Hospital-Taipei, Taipei, †School of Medical Technology, National Yang Ming Medical College, Taipei, and ‡Graduate Institute of Medical Technology, Department of Clinical Pathology, College of Medicine, National Taiwan University, Taipei, Taiwan, Republic of China

The 1.54 kb cDNA for ancrod, a thrombin-like enzyme, was cloned from a λ ZAP cDNA library derived from the venom glands of *Calloselasma (Agkistrodon) rhodostoma*. The cDNA sequence reveals that ancrod is synthesized as a pre-zymogen of 258 amino acids, including a putative secretory peptide of 18 amino acids and a proposed zymogen peptide of 6 amino-acid residues. The amino-acid sequence of the predicted active form of the enzyme exhibits a high degree of sequence similarity to those of mammalian serine proteases (trypsin and pancreatic kallikrein) and other thrombin-like enzymes (batroxobin and

flavoxobin). Key amino-acid residues (His⁴³, Asp⁸⁸, Ser¹⁸² and Asp¹⁷⁶) that are thought to be involved in the substrate cleavage and in the substrate-binding reaction are conserved. Ancrod contains 13 cysteine residues. Based on alignment with the amino-acid sequences of trypsin and batroxobin, six disulphide bridges can be predicted to be present in the ancrod protein. The existence of a free cysteine, which changes the common sequence surrounding the Ser¹⁸² active site from Gly-Asp-Ser-Gly-Gly-Pro to Cys-Asp-Ser-Gly-Gly-Pro, is unusual for a serine protease.

INTRODUCTION

Snake venoms affect blood coagulation and platelet functions in a complex manner [1]. The venom of the snake *Calloselasma rhodostoma* (formerly known as *Agkistrodon rhodostoma*) contains both platelet-activating and platelet-inhibitory components, a haemorrhagic protein and a thrombin-like enzyme [1–7]. Thrombin-like enzymes [8–10] are serine proteases, which belong to the trypsin family [11]. In contrast with thrombin, which converts fibrinogen to fibrin by removing fibrinopeptides A and B, a thrombin-like enzyme releases only fibrinopeptide A from fibrinogen [12,13]. Parenteral administration of the enzyme causes the conversion of fibrin into a fibrin derivative that is rapidly degraded and eliminated through a fibrinolytic process in the bloodstream [12]. Ancrod is a highly active thrombin-like enzyme purified from the venom of *Calloselasma rhodostoma* [13]. Since ancrod has been commercially available for many years, numerous exploratory studies on its clinical application have been reported [14]. The results have indicated considerable potential value for this protein in the treatment of various thrombo-embolic states, and in the prevention of thrombosis. However, the molecular characteristics of ancrod have not been described. Previously, we have cloned and analysed the cDNA for a platelet-aggregation inhibitor and for a haemorrhagic protein that were derived from the venom-gland cDNA library of *Calloselasma rhodostoma* [15]. Here, we report the cloning and nucleotide-sequence analysis of the cDNA for ancrod, and the deduced amino-acid sequence.

MATERIALS AND METHODS

Isolation of cDNA clones for ancrod

A set of cDNAs derived from a poly(A)⁺ RNA preparation of the venom glands of *Calloselasma rhodostoma* were cloned into the Uni-ZAPTMXR vector [16] by inserting the cDNAs into

the *Eco*RI and *Xho*I sites of the Uni-ZAPTMXR phage DNA. For convenience, the Uni-ZAPTMXR vector was excised *in vivo* to generate a pBluescript SK–phagemid cDNA library by co-transfecting the Uni-ZAPTMXR vector and an R408 helper phage [17] into *Escherichia coli* XL1-Blue host cells according to the manufacturer's instructions (Stratagene, La Jolla, CA, U.S.A.). *E. coli* XL1-Blue was then infected with the rescued phagemid. The infected bacteria were plated out on to Luria broth/ampicillin plates. Ampicillin-resistant colonies were transferred on to nitrocellulose membrane and prepared for colony hybridization [18]. A 20meric oligonucleotide probe (see Figure 1) was synthesized according to the nucleotide sequence of a batroxobin cDNA [8]. The probe was ³²P-end-labelled with T4 polynucleotide kinase and [γ -³²P]ATP [19]. The membrane replicas were pre-hybridized at 65 °C for 3 h in hybridization buffer [0.45 M sodium chloride, 0.045 M sodium citrate, 0.02 M potassium phosphate, pH 6.8, 0.1 % (w/v) SDS, 0.02 % (w/v) ficoll and 0.02 % (w/v) polyvinylpyrrolidone] plus 100 μ g/ml sheared and denatured herring-sperm DNA. The membrane replicas were then hybridized at 54 °C for 8 h in 30 ml of hybridization buffer containing 5 pmol of the labelled oligonucleotide probe. Subsequent washes of the hybridized membrane filters were carried out in 6 $\times$ SSC (1 $\times$ SSC is 0.15 M NaCl/0.015 M sodium citrate) and 0.1 % (w/v) SDS as follows: four times at 4 °C for 5 min each, once at room temperature (5 min) and once at 50 °C (2 min). The positive clones were detected by autoradiography, and were further purified by isolating single cell colonies.

Determination of the nucleotide sequence of the cDNA insert

The phagemids of positive clones were purified by alkaline lysis of the host cell, followed by equilibrium centrifugation in a CsCl/ethidium-bromide gradient [18]. The nucleotide sequence of the cDNA insert was determined by the dideoxy method using the 'walking primer' strategy (Figure 2).

§ To whom correspondence should be addressed.

The nucleotide sequence reported in Figure 3 will appear in the GenBank Database under accession no. L07308.

Figure 1 Oligonucleotide probe (20mer; boxed) synthesized according to the nucleotide sequence of the N-terminal amino-acid residues of batroxobin [8]

Figure 2 Partial restriction map and sequencing strategy used to obtain the nucleotide sequence of the cDNA clone J-5

Arrows indicate directions and sequenced regions. The longest open reading frame is shown as a filled box.

RESULTS AND DISCUSSION

Screening of a cDNA library

Based on the nucleotide sequence of the N-terminal amino-acid residues of batroxobin [8], a 20meric oligonucleotide was synthesized (see Figure 1) and was used as a hybridization probe in the screening of a phagemid-based cDNA library as described in the Materials and methods section. Three positive clones (designated J-2, J-3 and J-5) were obtained from the screening of about 1000 colonies. The size of the cDNA insert in each positive clone was determined by PCR using a pair of primers flanking the cloning sites of the pBluescript SK-phagemid DNA. The results indicated that J-2, J-3 and J-5 were independent and overlapping clones carrying cDNA inserts of 1.4, 1.3 and 1.5 kb respectively.

Nucleotide-sequence analysis

Clone J-5, which carries the longest cDNA insert, was completely sequenced (Figure 2). Most regions of the cDNA were sequenced in both directions. On analysis of the nucleotide sequence obtained, a long open reading frame was observed (Figure 2). In this open reading frame, the translation-initiation site is assigned to the first methionine codon at nucleotides 178–180, which is followed by an uninterrupted coding region, which putatively codes for 258 amino-acid residues (Figure 3). The 3' non-coding region contained a poly(A)⁺ sequence and a putative polyadenylation signal sequence AATAAA (nucleotides 1519–1524), which was located at a site 14 nucleotides upstream from the poly(A)⁺ sequence (Figure 3). Two putative glycosylation sites, Asn-Xaa-Thr, were located at amino-acid residues 148–150 and 229–231. Ancrod is known to be a glycoprotein [20].

Amino-acid sequence analysis

The amino-acid sequence deduced from the coding region of the

1	31
TGA CTC TGC CAG CGT TTG AGA TTG TGG GCC ACC CAG CTG CTT AAT TTG ATC AAA TAA AGT	
61	91
GCT GCT TGA TCA AGA AGT CTC CTC TTG AGT TAT CTG ATT AGA TTG ATA CGG TAT CTC AAG	
121	151
GTT AAG TTT GGG ACT GGG ATC TTA CAG GCA AAG AGC TTT CTG TGC AGA GTT GAA GCT ATG	
181	211
GTG CTG ATC AGA GTG CTA GCA AAC CTT GTG ATA CTA CAG CTT TCT TAC GCA CAA AAG TCT	
241	271
val leu ile arg val leu ala asn leu val ile leu gln leu ser tyr ala gln lys ser	
301	331
TCT GAA CTG GTC ATT GGA GGT GAT GAA TGT AAC ATA AAT GAA CAT CGT TTC CTT GTA GCC	
361	391
ser glu leu val ile gly gly asp glu cys asn ile asn glu his arg phe leu val ala	
421	451
CAA AGT GTA AAA TTT GAC GAT GAG CAG GAA AGA TTC CCA AAG GAG AAG CAC TTT ATT CGC	
481	511
TGT AAC AAA CCC CGT ACC AGA TGG GGC GAG GAC ATC ATG TTG ATC AGG CTG AAC AAA CCT	
541	571
GTT AAC AAC AGT GAA CAC ATC GCT CCT CTC AGC AGC TGC CCT AAC CCT CCC ATT GTG GGC	
601	631
TCA GTT TGC CGT GTT ATG GGA TGG GGC TCA ATC AAT AAA TAT ATA GAC GTT TTG CCC GAT	
661	691
GAA CCT CGT TGT GCT AAT ATT AAC CTG TAC AAT TAC ACG GTG TGT CGT GGA GTT TTT CCA	
721	751
AGG ATA CCA AAG AAA AGC AAA ATA TTG TGT GCA GGT GAC CTG CAA GGA CGC CTA GAT TCA	
781	811
TGT CAC TGT GAC TCT GGG GGA CCT CTC ATT TGT AGT GAA GAA TTC CAT GGC ATT GTA TAT	
841	871
CGG GGA CCC AAT CCT TGT GCC CAA CCA GAT AAG CCT GCC CTC TAC ACC AAC ATC TTC GAT	
901	931
CAT CTT CAC TGG ATC CTT AGC ATT ATG GCA GGA AAT GCA ACT TGC TAT CCA TAA AAC CTT	
961	991
TTG AAA TAG TTA AGT GGA GAA AAT GTA ACA TAT TAG TAA ATC TCT TCT ATA TCC TTG CAT	
1021	1051
TGG AAC ATA TTC CCA GCT GTA AGC CTT TTT AGA CTC AAA TAG GAC TAC CTT TGG AGT AAG	
1081	1111
AAG TGC TCA AAA TAG TGC TGC AGG GAT CAT TTC CCA TTT AAT TTC AGT TTA AAA CAG TCT	
1141	1171
CCA TAG ATT GGA GGA CTG TTT TAG GGT TAG GTG CAA AAT TTT CTG ACT CTA AAA TGG ACC	
1201	1231
ATT CCA ATC TTT TAA CCT CTG AAT ATT TTC AAT TCT TGT CAG CTT CTG GGG CAG ATG CTC	
1261	1291
TCT GAG CTT GTC TTC TGG CAG ACA TTT CAT TAC CCA GCT AGG TAA CAT CAT CAG TGC TAG	
1321	1351
AAT ATT TTT TTC TAT TGG TAT TTC TGT GGC ATT TAC AAT ACG CTC TTA TGG AGT TAT CCA	
1381	1411
GTC ACC CCA CAA ACA TAT ACA TAT CCC CGG GTC CCA CTG TTG CCT AAA AAG GAT CCC AGA	
1441	1471
TTA ACC CCC CAT TTC CCA ATC ACT AAA TAG AAT CTT TTG AGA ATC ATG TTT TCA TGT AAA	
1501	1531
TTC TTA GGT ATC CAC AAC AAT AAA TTG TAT AAA TCA TCA AAA AAA AAA AAA AAA AAA	

Figure 3 Nucleotide sequence of the ancrod cDNA J-5 and the predicted amino-acid sequence of the ancrod precursor protein

The nucleotide residues are numbered on the left and the amino-acid residues are numbered on the right. The amino-acid numbering starts at the N-terminal amino acid, valine of ancrod. The putative pre-peptide and pro-peptide consist of amino acid residues –24 to –7 and –6 to –1 respectively. Arrows indicate predicted cleavage sites. Asterisks indicate putative amino-acid residues that are involved in catalytic action. The putative glycosylation sites are shown in boxes and the polyadenylation-signal sequence, AATAAA, is underlined.

J-5 cDNA (Figure 3) was compared with that of the batroxobin precursor (see Figure 4) [8]. Batroxobin is a thrombin-like enzyme that is derived from snake venom of *Bothrops atrox moojeni*. On alignment using the method of Pearson and Lipman [21], 60% sequence identity (70% similarity considering conservative substitutions) was found between the J-5-derived amino-acid sequence and that of batroxobin [8]. This implies that the

Figure 4 Aligned amino-acid sequences of the ancrod precursor, batroxobin precursor, flavoxobin, bovine thrombin, bovine trypsin and human kallikrein

The numbering starts from the N-terminal residue of ancrod. Gaps have been introduced to optimize the sequence similarity. Asterisks indicate the amino-acid residues that are involved in the catalytic action.

cDNA clone J-5 does encode for ancrod, which is also a thrombin-like serine protease. Moreover, a 93% sequence identity between the first 30 amino-acid residues of the N-termini has allowed us to predict that ancrod is also synthesized as a pre-zymogen before maturation, as is batroxobin and other secretory serine proteases [8,22–24]. The precursor peptide includes a signal peptide (pre-peptide) and a zymogen peptide (pro-peptide). It is known that signal peptides generally contain a region that is rich in hydrophobic amino-acid residues with large side-chains, and terminate in a residue having a small side-chain (Ala, Ser, or Gly) [25]. Therefore, we predict that a possible site for the cleavage of the signal peptide in the precursor is present after the Ala residue at position -7 (that is, the signal peptide covers the sequence from amino acids -24 to -7; Figure 3). The zymogen peptide is known to contain a region that is rich in hydrophilic amino-acid residues [8,22,23] and the first N-terminal amino acid of

ancrod is valine [20]. The hydrophilic peptide (amino acids -6 to -1) is presumably a zymogen peptide.

Figure 4 also shows a high sequence similarity between ancrod and other mature serine proteases [9,22,23]. Based on sequence similarity, we could deduce the catalytic amino-acid residues to be His⁴³, Asp⁸⁸ and Ser¹⁸² (Figure 4). The location of Asp¹⁷⁶ at a position equivalent to the bottom of the substrate-binding pocket is consistent with the fact that ancrod cleaves a fibrinopeptide A fragment from fibrinogen after the positively charged arginine residue [13].

Ancrod contains 13 cysteine residues. Based on comparison of the aligned sequences of batroxobin, flavoxobin and trypsin, six possible disulphide bridges appear to be Cys⁷–Cys¹⁴¹, Cys²⁸–Cys⁴⁴, Cys⁷⁸–Cys²³², Cys¹²⁰–Cys¹⁸⁸, Cys¹⁵²–Cys¹⁶⁷ and Cys¹⁷⁸–Cys²⁰³.

The existence of a free cysteine (Cys¹⁸⁰), which changes the common sequence surrounding the active site Ser¹⁸² from Gly-Asp-Ser-Gly-Gly-Pro to Cys-Asp-Ser-Gly-Gly-Pro, is unusual for a serine protease. To ensure that amino-acid residues unique to ancrod (such as Cys¹⁸⁰, Lys⁴², Arg¹⁷⁴) found in the sequence alignment (Figure 4) were not artefacts derived from the cDNA cloning process, an independent clone J-2 was also sequenced. The sequence of J-2 was consistent with that of the J-5 clone. The replacement of Gly¹⁸⁰ by a cysteine residue in the common sequence around the serine residue of the active site does not seem to decrease the thrombin-like proteolytic activity of ancrod compared with that of batroxobin [8].

Finally, we speculate that ancrod may also possess a cysteine protease-like activity. Cysteine proteases share some similarities of reaction mechanism with serine proteases [26], but they are generally active in a higher pH range. The -SH group of a free cysteine residue plays a role that is similar to that of the -OH group of the active-site serine. Cys¹⁸⁰, which is located close to the catalytic cave, may be coupled with His⁴³, as generally occurs in cysteine proteases to cleave peptide-bonds.

We thank Dr. Ming-Yi Liao for providing us with venom glands. We also thank Dr. Tur-Fu Huang for helpful advice and Dr. Kong-Bung Choo for critical reading of the manuscript. This work was supported by grant NSC 82-0418-B075-016-BC from the National Science Council of the Republic of China.

REFERENCES

- Ouyang, C., Teng, C. M. and Huang, T. F. (1987) *Asia Pac. J. Pharmacol.* **2**, 169–179
- Esnouf, M. P. and Tunnah, G. W. (1967) *Br. J. Haematol.* **13**, 581–590
- Pizzo, S. V., Schwartz, M. L., Hill, R. L. and McKee, P. (1972) *J. Biol. Chem.* **247**, 637–645
- Ouyang, C., Hwang, L. J. and Huang, T. F. (1983) *Toxicon* **21**, 25–33
- Ouyang, C., Hwang, L. J. and Huang, T. F. (1985) *T'ai-Wan I Hsueh Hui Tsa Chih* **84**, 1196–1206
- Ouyang, C., Yeh, H. i. and Huang, T. F. (1986) *Toxicon* **24**, 633–643
- Huang, T. F., Wu, Y. I. and Ouyang, C. (1987) *Biochim. Biophys. Acta* **925**, 248–257
- Itoh, N., Tanaka, N., Mihashi, S. and Yamashina, I. (1987) *J. Biol. Chem.* **262**, 3132–3135
- Shieh, T. C., Kawabata, S. I., Kihara, K., Ohno, M. and Iwanaga, S. (1988) *J. Biochem. (Tokyo)* **103**, 596–605
- Pirkle, H., Theodor, I. and Lopez, R. (1989) *Thromb. Res.* **56**, 159–168
- Itoh, N., Tanaka, N., Funakoshi, I., Kawasaki, T., Mihashi, S. and Yamashina, I. (1988) *J. Biol. Chem.* **263**, 7628–7631
- Stocker, K. (1978) *Handb. Exp. Pharmacol.* **46**, 451–484
- Ewart, M. R., Hatton, M. W. C., Basford, J. M. and Dodgson, K. S. (1970) *Biochem. J.* **118**, 603–609
- Latallo, Z. S. (1983) *Thromb. Haemostasis* **50**, 604–609
- Au, L. C., Huang, Y. B., Huang, T. F., Teh, G. W., Lin, H. H. and Choo, K. B. (1991) *Biochem. Biophys. Res. Commun.* **181**, 585–593
- Greener, A. (1990) *Strategies in Molecular Biology* **3**, 5–6
- Russel, M., Kidd, S. and Kelly, M. R. (1986) *Gene* **45**, 333–338

- 18 Sambrook, J., Fritsch, E. F., Maniatis, T. (1989) in *Molecular Cloning, A Laboratory Manual*, 2nd edn., chapter 1, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
- 19 Maxam, A. M. and Gilbert, W. (1980) *Methods Enzymol.* **65**, 507–521
- 20 Hatton, M. W. C. (1973) *Biochem. J.* **131**, 799–807
- 21 Pearson, W. R. and Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2444–2448
- 22 MacDonald, R. J., Stary, S. J. and Swift, G. H. (1982) *J. Biol. Chem.* **257**, 9724–9732
- 23 Swift, G. H., Dargon, J. C., Ashley, P. L., Scott, W., Cummings, S. W. and MacDonald, R. J. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 7263–7267
- 24 Degen, S. J. F., MacGillivray, R. T. A. and Davie, E. W. (1983) *Biochemistry* **22**, 2087–2097
- 25 Steiner, D. F., Quinn, P. S., Chan, S. J., Marsh, J. and Tager, H. (1980) *Ann. N.Y. Acad. Sci.* **343**, 1–16
- 26 Takio, K., Towatari, T., Katunuma, N., Teller, D. C. and Titani, K. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3666–3670

Received 21 January 1993/15 March 1993; accepted 22 March 1993