Biochem. J. (1987) 246, 799–802 (Printed in Great Britain)

Amino acid sequence of a cystatin from venom of the African puff adder (*Bitis arietans*)

Anka RITONJA,*† Herbert J. EVANS,‡§ Werner MACHLEIDT* and Alan J. BARRETT‡‖
*Institut für Physiologische Chemie, Physikalische Biochemie und Zellbiologie der Universität München, Goethestrasse 33, D-8000 München 2, Federal Republic of Germany, and ‡Biochemistry Department, Strangeways Laboratory, Cambridge CB1 4RN, U.K.

The amino acid sequence of a cystatin from the venom of the African puff adder (*Bitis arietans*) is reported. It shows the protein to be more closely related to the Type 2 cystatins than the others, and yet it is only 31–37% identical to the known Type 2 cystatins, and differs strikingly in the insertion of a six-residue segment.

INTRODUCTION

A cystatin-like inhibitor of cysteine proteinases has been isolated from the venom of the puff adder (*Bitis arietans*) (Evans & Barrett, 1987). In the present paper we report the amino acid sequence of the inhibitor and discuss its structural and evolutionary relationship to other members of the cystatin superfamily.

EXPERIMENTAL

Materials

The cystatin-like cysteine proteinase inhibitor from the venom of the African puff adder (*Bitis arietans*) was purified as described (Evans & Barrett, 1987). Iodo-[^{3}H]acetic acid was from Amersham, carboxypeptidase B (di-isopropylphosphofluoridate-treated) from Sigma, and β-trypsin was prepared as described by Strop & Cechova (1981).

Methods

The native inhibitor was reduced with 2-mercapto-ethanol and carboxymethylated with iodo[^{3}H]acetic acid, cleaved with CNBr or oxidized with performic acid by standard procedures. Peptides were purified by gel filtration on Sephacryl S-200 in 80% formic acid and/or by reverse phase h.p.l.c. on a Gynkothek RP C_{18} column by use of gradients of acetonitrile in 0.1% trifluoroacetic acid. Carboxypeptidase B digestion was as described by Ambler (1972).

Amino acid analysis was done with a Kontron Liquimat II analyser using fluorescence detection after post-column reaction with *o*-phthalaldehyde. Sequences were determined by automated solid-phase Edman degradation using a non-commercial sequencer. The amino acid phenylthiohydantoin derivatives were identified and quantified by reverse-phase h.p.l.c. operating on-line to the sequencer (Machleidt, 1983). Prior to sequencing, the peptides were covalently coupled to porous glass supports as described by Machleidt *et al.* (1986). Cysteine was identified as the *S*-[^{3}H]-carboxymethylcysteine phenylthiohydantoin derivative by liquid-scintillation counting.

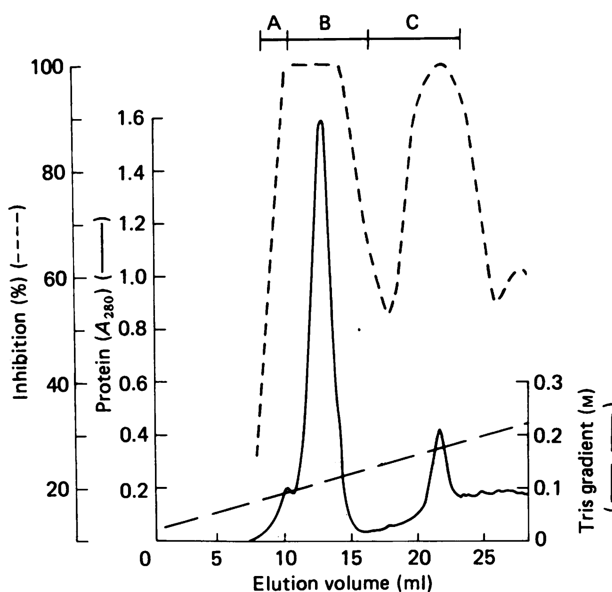

Fig. 1. F.p.l.c. chromatogram of isoinhibitors

The affinity-purified puff adder cystatin was run on the Mono-Q f.p.l.c. column, and fractions were assayed, exactly as described by Evans & Barrett (1987). Bars mark the pools A, B and C of material used for amino acid sequencing.

RESULTS

The snake cystatin was isolated as described in the preceding paper (Evans & Barrett, 1987); the final f.p.l.c. Mono Q anion exchange column resolved three protein peaks (A, B, C in Fig. 1) which were analysed by *N*-terminal amino acid sequencing through 60–61 steps (Fig. 2). The results indicate that the peaks contain

† Permanent address: Department of Biochemistry, J. Stefan Institute, Jamova 39, YU-61000 Ljubljana, Yugoslavia.
§ Permanent address: Department of Biochemistry, Medical College of Virginia, Virginia Commonwealth University, PO Box 614, Richmond, VA 23298-0001, U.S.A.
‖ To whom correspondence and reprint requests should be addressed.

```
                                           ↓
            1        10        20        30
Peak A      IPGGLSPRDVTDPDVQEAAAFAVEKYNARSKNDYY

              40        50     ↓  60
            FKERRVVEAQSQVVSGVKYYLTMEL

                                           ↓
            1        10        20        30
Peak B      IPGGLSPRDVTDPDVQEAAAFAVEKYNAGSKNDYY

              40        50     ↓  60
            FKERRVVEAQSQVVSGVKYYLMMELL
                                 T
                                           ↓
            1        10        20        30
Peak C      IPGGLSPRDVTDPDVQEAAAFAVEKYNAGSKNDYY

              40        50     ↓  60
            FKERRVVEAQSQVVSGVKYYLTMELL
```

Fig. 2. *N*-Terminal amino acid sequences of the isoinhibitors obtained from peaks A, B and C in Fig. 1

It can be seen that the forms differ in residues at two positions, 29 and 57, in the first 60 residues.

slightly different isoforms differing in one or two amino acids within the first 60 residues. Residue 29 was arginine or glycine, and residue 57 methionine or threonine. There was no heterogeneity at the *N*-terminus, indicative of aminopeptidase-like degradation such as has been found previously for several other cystatins (Barrett *et al.*, 1986*b*). Material from peak B was used for determination of the complete amino acid sequence (Fig. 3).

The protein consists of a single polypeptide chain of 111 residues, and contains no histidine. Single residues were found at all positions except 57, where methionine and threonine were present in approximately equal amounts, indicating the presence of two isoforms differing in this single position. A molecular mass of 12650 was calculated from the sequence of the Thr-57 form.

N-Terminal solid-phase Edman degradation of the native inhibitor (NI in Fig. 3) resulted in a continuous sequence of 61 residues showing a single Asp–Pro bond (positions 12–13), equimolar amounts of methionine and threonine in position 57, and a single methionine in position 58. The reduced and *S*-carboxymethylated inhibitor was cleaved with CNBr, and the fragments containing *C*-terminal homoserine (CN-1) were selectively coupled to aminopropyl glass without prior separation. CNBr cleavage was almost complete, but about 40% of the acid-labile Asp_{12}–Pro_{13} bond was hydrolysed by the formic acid used in the cleavage procedure, so that two sequences were obtained corresponding to residues 1–57 (1–58 from the Thr-57/Met-58 form) and 13–57 (13–58 for the Thr-57/Met-58 form). The second CNBr fragment (CN-2) was isolated by reverse phase h.p.l.c., and sequenced in two separate experiments after immobilization via carboxyl side chains and via lysyl side chains. Within the usual limits of error, the amino acid composition of CN-2 obtained by hydrolysis (Table 1) matched the composition calculated from the sequence. The *C*-terminal lysyl residue was confirmed by carboxypeptidase B hydrolysis of CN-2 and also of the performic-acid-oxidized, uncleaved protein. A few gaps in the automated degradation sequences, notably at the two tryptophan residues, were filled by use of sequence data from peptides generated by hydrolysis of the reduced and *S*-carboxymethylated inhibitor with *β*-trypsin. Eleven tryptic peptides were isolated by reverse phase h.p.l.c. and sequenced (T-1 to T-11 in Fig. 3).

The cysteine residues were identified by the radioactivity of the *S*-[^{3}H]carboxymethylcysteine phenylthiohydantoin derivatives. After *S*-carboxymethylation of the reduced inhibitor, three cysteine residues were found, in positions 65, 81, and 94. Treatment of the unreduced

```
                  10                  20                  30
I P G G L S P R D V T D P D V Q E A A A F A V E K Y N A G S K N D Y Y F K
o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o
                                    NI
o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o
                                    CN-1
o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-X o-o-o-o-o-X o-o-o-o-o-X
    T-1                       T-2                     T-3         T-4

    40                  50                  60                  70
E R R V V E A Q S Q V V S G V K Y Y L M M E L L K T T C K K T V G R P K G
o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o
                        NI
o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o
                CN-1                                      CN-2
o-o   o-o-o-o-o-o-o-o-o-o-o-o-X o-o-o-o-o-o-o-o-X o-o-o-X   o-o-o-o-o-X o
T-5             T-6                 T-7 A/B        T-8          T-9

          80                  90                  100                 110
Y Q E I Q N C N L P P E N Q Q E E I T C R F E V W S R P W L P S T S L T K
o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-X-o-o-o-X-o-o-o-o-o-o-o-X
                                    CN-2
o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o
                T-10                                  T-11

                                                                        o
                                                                     CP-B
```

Fig. 3. Amino acid sequence of puff adder cystatin showing the strategy of sequence determination

Bars indicate the sequences obtained from the native inhibitor (NI), from CNBr fragments (CN-1, CN-2), from *β*-tryptic peptides (T-1 to T-11) and by carboxypeptidase B digestion (CPB). Residues that were directly identified in Edman degradations are marked (○), and residues that were not identified are marked (×). The sequences of peptides T-7A and T-7B were YYLMMELLK and YYLTMELLK, respectively.

inhibitor with iodo[^{14}C]acetate, in 6 M-guanidine hydrochloride, resulted in no significant incorporation of radioactivity, indicating that no free thiol group existed in the inhibitor sample used for sequencing. After reduction and *S*-carboxymethylation, approximately the same specific radioactivity was found in the tryptic peptides T-8 (containing Cys-65) and T-10 (containing Cys-81 and Cys-94). This was consistent with the possibility that Cys-65 and Cys-81 had been linked in a disulphide bond, and labelled following reduction, whereas Cys-94 had been oxidized, or carboxymethylated by the 1 mM-iodoacetic acid used during the isolation of the inhibitor (Evans & Barrett, 1987), and was not labelled.

Table 1. Amino acid composition of snake cystatin and its CNBr fragment CN-2

The composition calculated from the amino acid sequence is compared with the results of amino acid analysis after 24 h or 72 h (for Val, Ile) hydrolysis. No corrections were made for losses of labile amino acids (Ser, Thr). Cys was determined as *S*-carboxymethylcysteine. Abbreviation: n.d., not determined.

Amino acid	Residues: Snake cystatin, Calculated	Snake cystatin, Hydrolysis	CN-2, Calculated	CN-2, Hydrolysis
Asx	9	8.85	3	2.93
Thr	7	5.93	6	4.89
Ser	7	5.20	3	2.64
Glx	17	16.78	10	9.81
Pro	8	n.d.	5	n.d.
Gly	6	6.32	2	2.17
Ala	6	5.87	–	0.10
Cys	3	1.41	3	1.54
Val	10	9.72	2	1.90
Met	2	1.20	–	0.06
Ile	3	2.73	2	1.80
Leu	7	6.87	5	4.75
Tyr	6	5.63	1	0.55
Phe	3	2.97	1	1.00
Lys	9	8.77	5	4.90
His	–	0.11	–	0.17
Arg	6	5.82	3	3.15
Trp	2	n.d.	2	n.d.
Total	111		53	

DISCUSSION

We compared the amino acid sequence of the puff adder cysteine proteinase inhibitor with the available sequences of cystatins (Fig. 4). It was clear that the venom inhibitor is a member of the cystatin superfamily, there being many identical residues. It is notable that the residue Gly-9 (chicken cystatin numbering), recently identified as being important for the inhibitory activity of cystatins, again appears in the snake cystatins, as in all other known sequences of inhibitory cystatins (Abrahamson *et al.*, 1987).

The proteins of the cystatin superfamily have been divided into three families (Barrett *et al.*, 1986*a*; Barrett, 1987), and it is clear that the sequence of the snake cystatin is most like that of the Type 2 cystatins, with 31–37% of identical residues, the greatest similarity being to beef colostrum cystatin. Nevertheless, the snake protein differs in two important respects from the Type 2 cystatins described previously.

One of the differences between the sequence of the puff adder cystatin and those of other Type 2 cystatins is in the segment 78–85 (chicken cystatin numbering), in which there is an insertion of six residues into the presumptive disulphide loop 71–81, followed by deletion of two residues beyond the loop. Presumably, this part of the sequence is of little functional importance in the cystatin superfamily, since it is largely deleted from the Type 1 cystatins.

The second difference concerns the disulphide loops.

Fig. 4. Amino acid sequence of the puff adder cystatin, aligned with those of other members of the cystatin superfamily

The puff adder venom inhibitor sequence is aligned with those of human cystatin A (ha), rat cystatin A (ra), human cystatin B (hb), rat cystatin B (rb), human cystatin C (hc), beef colostrum cystatin (bc), human cystatin S (hs), human cystatin SN (hsn), chicken cystatin (cc), human kininogen segment 1 (residues 19–127) (hk1), kininogen segment 2 (residues 128–249) (hk2) and kininogen segment 3 (residues 250–373) (hk3), and the residues are numbered according to chicken cystatin. Gaps have been introduced where they increase the score of the alignment calculated as described by Dayhoff *et al.* (1983), with a gap penalty of −6. Residues identical to those in the venom cystatin are printed as blocks in the other sequences. Brackets indicate the disulphide bonds identified in several Type 2 cystatins, and inferred by homology and other indirect evidence for the kininogens. Original references to the sequences are given by Barrett *et al.* (1986*b*).

The Type 2 cystatins characteristically contain two disulphide loops (Fig. 4). The puff adder cystatin contains three of the cysteine residues expected to form these, but the fourth would be a few residues beyond the *C*-terminus found in the present study. One may speculate that proteolytic cleavage after Lys-111 (occurring *in vivo* or during commercial handling of the snake venom) gave rise to a short disulphide-linked fragment containing the fourth cysteine residue which subsequently was lost by reduction or oxidation of the exposed disulphide bond. No peptide fragment was detected after performic acid oxidation of the previously active inhibitor, so we can assume that snake cystatin containing only the one extended disulphide loop, Cys-65 to Cys-81, is still a tight-binding inhibitor of papain and related cysteine proteinases. We do not know, however, whether a *C*-terminal extension containing a fourth cysteine residue is present in the virgin inhibitor *in vivo*.

Puff adder cystatin is the first member of the cystatin superfamily to be definitely identified in reptiles. The biological reason for the presence of a cysteine proteinase inhibitor in snake venom is not obvious, but there is no evidence that it is connected to the toxicity of the venom; cystatins are found in saliva, milk and other body fluids of mammals (Barrett *et al.*, 1986*b*).

We thank Mr. N. D. Rawlings for help with the comparison of amino acid sequences. The work was partially supported by the Sonderforschungsbereich 0207 der Universität München (project C-2) and the Bundesministerium für Forschung und Technologie, GFR (project 8651).

REFERENCES

Abrahamson, M., Ritonja, A., Brown, M. A., Grubb, A., Machleidt, W. & Barrett, A. J. (1987) J. Biol. Chem., in the press

Ambler, R. P. (1972) Methods Enzymol. **25**, 143–154

Barrett, A. J. (1987) Trends Biochem. Sci., in the press

Barrett, A. J., Fritz, H., Grubb, A., Isemura, S., Järvinen, M., Katunuma, N., Machleidt, W., Müller-Esterl, W., Sasaki, M. & Turk, V. (1986*a*) Biochem. J. **236**, 312

Barrett, A. J., Rawlings, N. D., Davies, M. E., Machleidt, W., Salvesen, G. & Turk, V. (1986*b*) in Proteinase Inhibitors (Barrett, A. J. & Salvesen, G., eds.), pp. 515–569, Elsevier Science Publishers

Dayhoff, M. O., Barker, W. C. & Hunt, L. T. (1983) Methods Enzymol. **91**, 524–545

Evans, H. J. & Barrett, A. J. (1987) Biochem. J. **246**, 795–797

Machleidt, W. (1983) in Modern Methods in Protein Chemistry (Tschesche, H., ed.), pp. 267–302, Walter de Gruyter, Berlin

Machleidt, W., Borchart, U. & Ritonja, A. (1986) in Advanced Methods in Protein Microsequence Analysis (Wittmann-Liebold, B., Salnikow, J. & Erdmann, V. A., eds.), pp. 91–107, Springer, Heidelberg

Strop, P. & Cechova, D. (1981) J. Chromatogr. **207**, 55–62

Received 21 April 1987/15 July 1987; accepted 22 July 1987