Biochem. J. (1984) **217**, 813–817
Printed in Great Britain

Cystatin

Amino acid sequence and possible secondary structure

Christian SCHWABE,* Angela ANASTASI,† Hal CROW,* J. Ken McDONALD* and Alan J. BARRETT†

**Department of Biochemistry, Medical University of South Carolina, 171 Ashley Avenue, Charleston, SC 29403, U.S.A., and †Department of Biochemistry, Strangeways Laboratory, Worts Causeway, Cambridge CB1 4RN, U.K.*

(*Received 7 July 1983/Accepted 14 October 1983*)

The amino acid sequence of cystatin, the protein from chicken egg-white that is a tight-binding inhibitor of many cysteine proteinases, is reported. Cystatin is composed of 116 amino acid residues, and the M_r is calculated to be 13143. No striking similarity to any other known sequence has been detected. The results of computer analysis of the sequence and c.d. spectrometry indicate that the secondary structure includes relatively little α-helix (about 20%) and that the remainder is mainly β-structure.

Relatively few protein inhibitors specific for cysteine proteinases are known, in contrast with the large number of inhibitors of serine proteinases (Laskowski & Kato, 1980), and for none of the cysteine proteinase inhibitors has the mechanism of inhibition been established. It is possible that some inhibitors of cysteine proteinases act by mechanisms closely analogous to those of serine proteinase inhibitors, as is suggested by the discovery that the bromelain inhibitor from *Ananas comosus* appears to be homologous with the serine proteinase inhibitors of the Bowman–Birk type from leguminous seeds (Ketcham *et al.*, 1978). On the other hand, quite different mechanisms, possibly involving disulphide exchange, for example, may exist.

As a step towards learning about these mechanisms, establishing evolutionary relationships among the inhibitors and exploring their biological functions, we have now determined the amino acid sequence of cystatin, the chicken egg-white inhibitor of cysteine proteinases that has recently been characterized (Anastasi *et al.*, 1983).

Materials and methods

Materials

Egg-white cystatin (form 1) was isolated as described by Anastasi *et al.* (1983). Reagents were of analytical or sequencing grade. All sequencer reagents and solvents were purchased from Beckman Instruments Co.

Reduction and carboxymethylation

Cystatin was reduced and carboxymethylated as described previously (Anastasi *et al.*, 1983), but sometimes with the use of iodo[2-^{3}H]acetic acid (code TRA.195; Amersham International).

Preparation of CNBr-cleavage peptides

CNBr cleavage of the reduced and carboxymethylated protein (15mg) was performed in 70% (v/v) formic acid at a protein concentration of 10mg/ml and a CNBr concentration of 20mg/ml (Findlay & Brew, 1972). After 24h at 20°C the cleaved protein was diluted 10-fold with distilled water and freeze-dried. This material was dissolved in 1ml of 5% (v/v) formic acid and run on a column (1.5cm × 80cm) of Sephadex G-50 in 5% formic acid at a flow rate of 10ml/h. The effluent was monitored by absorbance at 280nm and by the fluorescamine reaction (Udenfriend *et al.*, 1972). After a small early peak of uncleaved material, three further fluorescamine-reactive peaks were well resolved. The fractions containing the separated peptides, numbered CN-1, CN-2 and CN-3, were freeze-dried.

Preparation of tryptic peptides

Reduced carboxymethylated cystatin (6mg) and the CNBr-cleavage fragments (1–4 μmol) were dissolved in 2ml of 1% (w/v) NH_4HCO_3 and digested with trypsin as described previously (Anastasi *et al.*, 1983).

Tryptic peptides were separated by ion-

exchange chromatography on a column (1 cm × 55 cm) of Dowex 50 (X2) resin equilibrated with 0.2M-pyridine/acetic acid, pH 3.1, essentially by the method of Schroeder (1967). The column eluates were monitored by the fluorescamine assay. After being dried and freeze-dried from water, the peptides were checked for homogeneity by high-voltage paper electrophoresis at pH 3.5 in pyridine/acetic acid/water (1:10:90, by vol.), pH 3.5, and formic acid/acetic acid/water (52:29:919, by vol.), pH 1.9, and by *N*-terminal analysis.

Amino acid analysis

Samples were hydrolysed under N_2 for 24h in 6M-HCl for running on the Dionex 502 analyser.

The analyses were fitted to integer numbers of residues, and *F* values obtained for the goodness of fit, by the method of Alt *et al.* (1975).

Sequence analysis

Sequence analyses were performed on a Beckman 890C automatic protein sequencer by using 0.1M-Quadrol program 122974 in combination with Polybrene (2mg) (Klapper *et al.*, 1978), which aids the retention of the protein in the sequencer cup during the wash cycle. The anilinothiazolinone derivatives of amino acids were manually converted into their phenylthiohydantoins by treatment with 1M-HCl containing 0.1% (v/v) 2-mercaptoethanol for 10min at 80°C, and identified by high-pressure liquid chromatography on a Du Pont system in combination with a Waters automatic sample injector (WISP). Generally, 5 μl of amino acid phenylthiohydantoin sample (containing 0.05–3.0nmol) in acetonitrile was injected for each analysis. Separation was on a Zorbax C_{18} column (4.6mm × 15cm), with a Whatman guard column. An exponential gradient (16min) was generated from 10% acetonitrile in 10mM-sodium acetate (solvent A) and 100% acetonitrile (solvent B), running from 2% to 35% of solvent B. The amino acid phenylthiohydantoin derivatives were detected by their absorption at 254nm. At a flow rate of 1.5ml/min, the column re-equilibration time between runs was 10min. The identification of glutamic acid, glutamine, threonine and glycine was confirmed by use of a 5min isocratic run at 15% solvent B (H. Crowe & C. Schwabe, unpublished work). The identification of carboxymethylcysteine was confirmed by radioactivity of the residue obtained from the carboxy[^{3}H]methylated protein. The water-soluble residues arginine and histidine were identified with a 10min gradient of 2–55% solvent B.

Determination of c.d. spectrum

The c.d. spectrum of cystatin was determined in a Cary model 60 spectropolarimeter equipped with a model 6002 CD attachment. The instrument was calibrated with twice-sublimed (+)-camphor-10-sulphonic acid (Eastman). The protein concentration was 0.54mg/ml (in 50mM-ammonium bicarbonate buffer, pH 7.5), the light-path 0.02cm and the scale expansion 0.02° full scale. Data are expressed as mean residue ellipiticity $[\theta]_\lambda$, in units of degreees·cm^2·dmol^{-1} in accordance with the relation $[\theta]_\lambda = (\psi \cdot M_0)/(10 \cdot l \cdot c)$ where ψ is the observed ellipticity, l the pathlength in cm, c the protein concentration in g/ml, and M_0 the mean residue mass (113.347 for cystatin).

Results

Cystatin contains two methionine residues/molecule, and the expected three peptides formed by CNBr cleavage (CN-1, CN-2 and CN-3) were isolated by gel chromatography. Table 1 shows the results of amino acid analysis of the CNBr-cleavage peptides, in comparison with the composition of cystatin deduced from the final sequence.

Fig. 1 shows the complete sequence of cystatin form 1 deduced from automatic sequence analysis of the three CNBr-cleavage peptides and the tryptic peptides. The yield of amino acid phenylthiohydantoin derivatives at each step of sequence determination varied with the *N*-terminal residues, so that the average repetitive yield ranged from 70 to 95%. All of the CNBr-cleavage and tryptic peptides yielded single unambiguous sequences during the automatic sequencing runs, confirming the independent evidence of homogeneity from high-voltage electrophoresis. A single sequence run with peptide CN-1 (20nmol) gave residues Ser-1 to Gln-26, and residues 1–3 were recognized as the *N*-terminal tripeptide previously identified in the intact protein (Anastasi *et al.*, 1983). Similarly, runs with peptides CN-2 (40nmol) and CN-3 (20nmol) gave residues 30–85 and 90–116 respectively, although the identifications of His-84 and Gln-116 were not completely certain. Sequencing of the two methionine-containing tryptic peptides (T-4 and T-12) from intact reduced carboxymethylated cystatin provided residues 27–29 and 86–89, and confirmed the overlaps between CNBr-cleavage peptides. The tryptic peptides obtained by digestion of each of the individual CNBr-cleavage peptides were also sequenced, and in all cases confirmed the identification of residues in the runs with the intact peptides.

Cystatin contains no free thiol (Anastasi *et al.*, 1983), and the four cysteine residues towards the *C*-terminus undoubtedly form two disulphide bonds (J. K. McDonald, unpublished work). No carbohydrate has been detected in cystatin (Anastasi *et al.*, 1983).

The sequence was analysed by use of a DEC-

Table 1. *Amino acid analysis of the three CNBr-cleavage peptides, in comparison with the composition deduced from the final sequence*

The three CNBr-cleavage peptides corresponded to segments of the sequence as follows: CN-1, residues 1–29; CN-2, residues 30–89; CN-3, residues 90–116 (see Fig. 1). The analyses of the individual peptides are normalized with respect to glutamic acid, and expressed as residues/molecule. The recoveries of the labile residues serine and threonine are uncorrected for losses, and naturally tryptophan did not survive acid hydrolysis. The low recovery of valine was attributable to incomplete hydrolysis of bonds involving this residue during 24h. The M_r value given for the intact protein is for the native disulphide-bonded molecule, taking account of the finding that, of the 26 dicarboxylic amino acid residues, 11 are amidated.

	Composition found			Composition from sequence
	Peptide CN-1	Peptide CN-2	Peptide CN-3	Cystatin
Asx	4.85	4.96	1.01	10
Thr	1.44	0.02	2.34	5
Ser	5.21	1.75	2.59	12
Glx	8.00	7.00	3.00	16
Pro	1.89	2.59	1.11	5
Gly	3.02	2.37	0.1	5
Ala	3.13	3.61	0.94	7
Cys	1.85	0.1	1.67	4
Val	2.49	2.04	1.42	9
Met	0.08	0	0	2
Ile	2.01	0.20	1.71	6
Leu	2.44	4.81	2.78	10
Tyr	2.92	0.3	1.82	5
Phe	1.0	1.39	1.01	3
His	0.88	0.04	–	1
Lys	4.06	0.69	2.99	7
Trp	0	0	0	1
Arg	4.59	3.59	0	8
Residues found	50 + Met 1	36 + Met 1	25 + Trp 1	116
Total residues	51	37	26	–
M_r	5700	4000	2800	13143
F factor	0.276	0.219	0.284	–

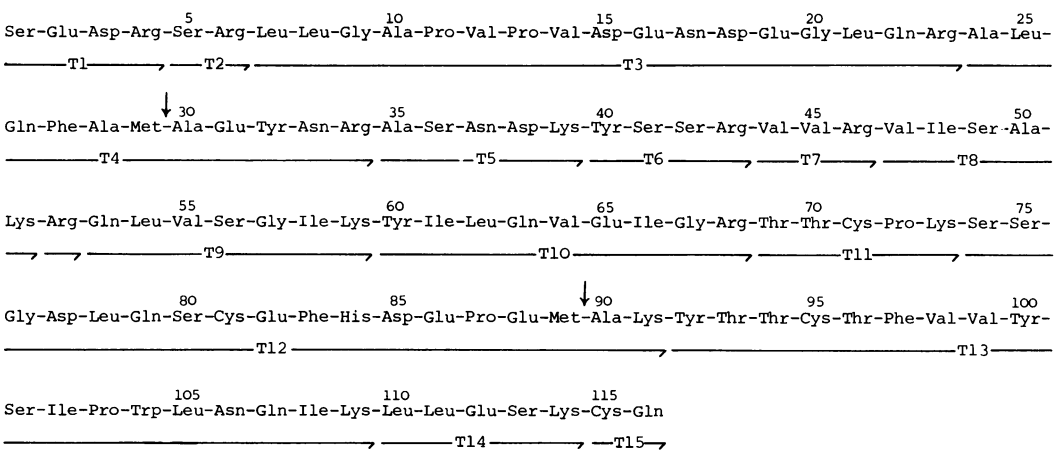

Fig. 1. *Amino acid sequence of chicken egg-white cystatin (form 1)*

The points of CNBr cleavage at the two methionine residues are marked with vertical arrows, and the 15 tryptic peptides (T1–T15) plus arginine are also indicated and numbered. All of the larger peptides were sequenced, and the compositions of the small tryptic peptides T2, T6 and T7, as well as free arginine, were verified by amino acid analysis.

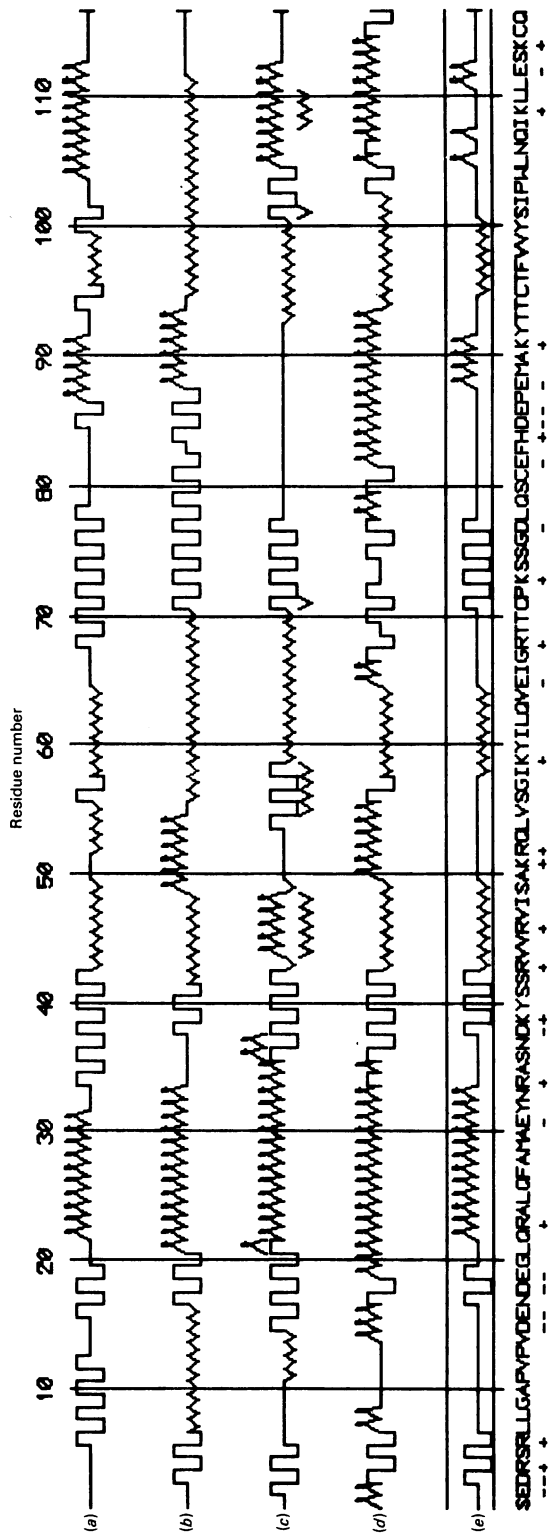

Fig. 2. *Computer prediction of secondary structure from the sequence of cystatin*

The algorithms used to make the four predictions were those of (*a*) Burgess *et al.* (1974), (*b*) Chou & Fasman (1979*a*,*b*), (*c*) Nagano (1977) and (*d*) Robson & Suzuki (1976). Line (*e*) shows those parts of the structure for which three of the four predictions agree. The sequence, indicated here in one-letter notation, corresponds to that in Fig. 1.

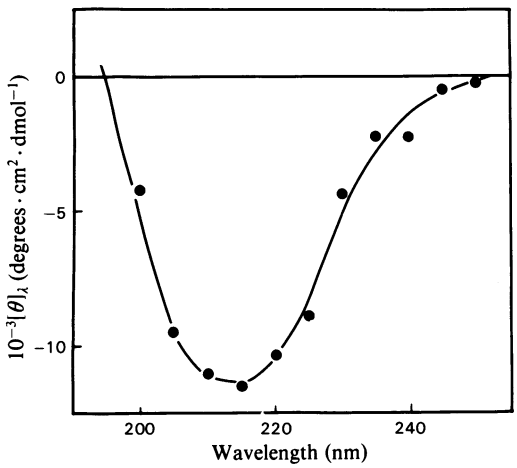

Fig. 3. *C.d. spectrum of native cystatin at pH7.5*
See the text for details.

System 10 computer running a FORTRAN program (Dzionara *et al.*, 1977; Rawlings *et al.*, 1983) to apply the methods of Burgess *et al.* (1974), Robson & Suzuki (1976), Nagano (1977) and Chou & Fasman (1979*a*,*b*) to predict the secondary structure of the protein. The results are presented in Fig. 2. The Chou & Fasman (1979*a*,*b*) method predicted 21% α-helix, 42% β-sheet, 24% β-turn and 12% random coil.

The c.d. spectrum of native egg-white cystatin (Fig. 3) is compatible with the computer prediction of conformation. A broad featureless minimum at 215nm, cross-over at 195nm and rather low rotational power (as compared with known α-helix structures) strongly suggest a large contribution of β-structure to this spectrum. The absolute minimum shows a mean molar residue rotation of -11700 degrees·cm^2·dmol^{-1}.

Discussion

The M_r of cystatin form 1 can be calculated from the sequence to be 13143 (before reduction of disulphide bonds), a value somewhat lower than that of 15000 indicated by sodium dodecyl sulphate/polyacrylamide-gel electrophoresis (Anastasi *et al.*, 1983).

Unlike the bromelain inhibitor from *Ananas*, another small-M_r inhibitor of cysteine proteinases, cystatin shows no evidence of homology with the Bowman–Birk soya-bean family of serine proteinase inhibitors, nor have we detected any significant similarities to the sequences of other proteinase inhibitors. Indeed, comparison of the sequence with those of the 1081 other proteins in the Protein Sequence Database of 1978 by use of the RELATE program of Dayhoff *et al.* (1983) detected matches above the 3.0 S.D. unit level only with brown-trout histone H2B (3.21), human haptoglobin β-chain (3.20) and badger fibrinopeptide A (3.08). In a sample of this size, approximately 1.5 matches above 3.0 S.D. units would be expected by chance, and further examination of the sequences of the three proteins suggested no genuine relationship to cystatin. It therefore seems likely that cystatin represents a new superfamily of proteins. Clearly it will be of interest to discover whether or not the mammalian cytoplasmic inhibitors of cysteine proteinases, which resemble cystatin in M_r and other properties (Järvinen & Rinne, 1982), are evolutionarily related.

We are most grateful to Mr. Neil D. Rawlings and Mr. Keith Ashman for provision of the computer prediction of secondary structure of cystatin, and for the computer comparison of the sequence with other known sequences.

References

Alt, J., Heyman, E. & Krisch, K. (1975) *Eur. J. Biochem.* **53**, 357–369
Anastasi, A., Brown, M. A., Kembhavi, A. A., Nicklin, M. J. H., Sayers, C. A., Sunter, D. C. & Barrett, A. J. (1983) *Biochem. J.* **211**, 129–138
Burgess, A. W., Ponnuswamy, P. K. & Scheraga, H. A. (1974) *Isr. J. Chem.* **12**, 239–286
Chou, P. Y. & Fasman, G. D. (1979*a*) *Biophys. J.* **26**, 367–384
Chou, P. Y. & Fasman, G. D. (1979*b*) *Biophys. J.* **26**, 385–400
Dayhoff, M. O., Barker, W. C. & Hunt, L. T. (1983) *Methods Enzymol.* **91**, 524–545
Dzionara, M., Robinson, S. M. L. & Wittmann-Liebold, B. (1977) *Hoppe-Seyler's Z. Physiol. Chem.* **358**, 1003–1019
Findlay, J. B. C. & Brew, K. (1972) *Eur. J. Biochem.* **27**, 65–68
Järvinen, M. & Rinne, A. (1982) *Biochim. Biophys. Acta* **708**, 210–217
Ketcham, L. K., Barker, W. C. & Dayhoff, M. O. (1978) in *Atlas of Protein Sequence and Structure 1978* (Dayhoff, M. O. & Eck, R. V., eds.), pp. 131–143 (see p. 136), National Biomedical Research Foundation, Silver Spring
Klapper, D. G., Wilde, C. E. & Capra, J. D. (1978) *Anal.Biochem.* **85**, 126–131
Laskowsi, M., Jr. & Kato, I. (1980) *Annu. Rev. Biochem.* **49**, 593–626
Nagano, K. (1977) *J. Mol. Biol.* **109**, 251–274
Rawlings, N. D., Ashman, K. & Wittmann-Liebold, B. (1983) *Int. J. Pept. Protein Res.* in the press
Robson, B. & Suzuki, E. (1976) *J. Mol. Biol.* **107**, 327–356
Schroeder, W. (1967) *Methods Enzymol.* **11**, 351–361
Udenfriend, S., Stein, S., Böhlen, P., Dairman, W., Leinmgruber, W. & Weigele, M. (1972) *Science* **178**, 871–872