Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Expression of a selenomethionyl derivative and preliminary crystallographic studies of human cystatin C

Maciej Kozak,[a]† Elzbieta Jankowska,[b] Robert Janowski,[a] Zbigniew Grzonka,[b] Anders Grubb,[c] Marcia Alvarez Fernandez,[c] Magnus Abrahamson[c] and Mariusz Jaskolski[a,d]*

[a]Department of Crystallography, Faculty of Chemistry, A. Mickiewicz University, Grunwaldzka 6, 60-780 Poznan, Poland, [b]Department of Chemistry, University of Gdansk, Poland, [c]Department of Clinical Chemistry, University of Lund, Sweden, and [d]Center for Biocrystallographic Research, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland

† Present address: Department of Macromolecular Physics, Faculty of Physics, A. Mickiewicz University, Poznan, Poland.

Correspondence e-mail: mariuszj@amu.edu.pl

Received 27 May 1999
Accepted 26 August 1999

Human cystatin C, a protein with amyloidogenic properties and a potent inhibitor of papain-like mammalian proteases, has been produced in its full-length form by recombinant techniques and crystallized in two polymorphic forms: cubic and tetragonal. A selenomethionyl derivative of the protein, obtained by *Escherichia coli* expression and with complete Met→Se-Met substitution confirmed by mass spectrometry, amino-acid analysis and X-ray absorption spectra, was crystallized in the cubic form. A truncated variant of the protein, lacking ten N-terminal residues, has also been crystallized. The crystals of this variant are tetragonal and, like the two polymorphs of the full-length protein, contain multiple copies of the molecule in the asymmetric unit, suggesting oligomerization of the protein.

1. Introduction

Cystatins are single-chain reversibly binding protein inhibitors of enzymes belonging to the papain family of cysteine proteases. They have been found in plants and animals (Turk & Bode, 1991) and in protozoa (Irvine *et al.*, 1992). Human cystatin C (HCC) is a low molecular weight protein (13 343 Da) composed of 120 amino-acid residues (Grubb & Löfberg, 1982; Fig. 1). It has high affinity for cathepsin B and other human cysteine proteases (reviewed in Abrahamson, 1993). HCC plays an important role in the development of cerebral haemorrhage in patients with a hereditary form of amyloid angiopathy, where a variant of cystatin C is deposited as amyloid fibrils in the cerebral arteries. This molecular pathology, caused by a point mutation and resulting in an amino-acid substitution in the HCC sequence (Leu68Gln; Fig. 1), leads to brain haemorrhage and death in young adults (Abrahamson *et al.*, 1992; Olafsson *et al.*, 1996). Crystallographic and NMR studies of three cysteine-protease inhibitors, chicken cystatin C (Bode *et al.*, 1988; Dieckmann *et al.*, 1993; Engh *et al.*, 1993), cystatin B (in complex with papain; Stubbs *et al.*, 1990) and cystatin A (Martin *et al.*, 1995), have shown similar overall structure, with three regions implicated in interactions with the target enzymes. These regions include the N-terminal segment and two hairpin loops. The 11 N-terminal amino-acid residues of HCC are important for its very high affinity binding to papain (K_i = 11 f*M*; Lindahl *et al.*, 1992) and to other cysteine proteases (Abrahamson *et al.*, 1987, 1991). It has been shown that leukocyte elastase can specifically cleave the single N-terminal Val10–Gly11 bond of HCC, resulting in seriously compromised affinities for such target enzymes as cathepsins B, H and L (Abrahamson *et al.*, 1991). Peptide inhibitors structurally based on the N-terminal segment of cystatin C show interesting biomedical activities, including protection against streptococcal infections in mice (Björck *et al.*, 1989) and inhibition of the replication of coronavirus and herpes simplex virus (Collins & Grubb, 1991; Björck *et al.*, 1990). Development of a second generation of more effective specific cysteine-protease peptide inhibitors would be greatly facilitated by a three-dimensional structure of HCC. Similarly, such a model would be very useful in the elucidation of the pathophysiological background of the cerebral haemorrhage produced by the Leu68Gln variant of HCC. Our initial attempts to solve the crystal structure of HCC by molecular replacement using as a probe the available crystallographic model of N-truncated chicken cystatin C (Bode *et al.*, 1988; 41% sequence identity, 62.5% homology) were unsuccessful. We have, therefore, produced an Se-Met derivative of HCC for crystal structure determination by the MAD method (Hendrickson, 1991).

2. Materials and methods

2.1. Protein expression and purification

Production of wild-type human cystatin C was accomplished in a previously described *Escherichia coli* expression system (Abrahamson *et al.*, 1988). For production of Se-Met-enriched cystatin C, the plasmid pHD313 was transformed into competent *E. coli* BL834(DE3) cells (Novagen). Single colonies were picked and positive transformants were selected by their ability to express cystatin C in

small-scale cultures (3 h induction at 315 K in LB medium) as assessed by SDS–polyacrylamide gel electrophoresis. A subclone of the transformed *E. coli* BL834(DE3) cells was grown in minimal medium composed of 87 m*M* Na_2HPO_4, 46 m*M* KH_2PO_4, 18 m*M* NaCl, 7.5 m*M* $(NH_4)_2SO_4$, 0.2%(*w*/*v*) glucose, 1.7 m*M* $MgSO_4$, 0.117 m*M* $CaCl_2$, 0.015 m*M* $FeSO_4$, 0.075 m*M* thiamine and 0.3 m*M* methionine or selenomethionine (Sigma). Initial cultures were grown in minimal medium containing normal methionine. Prior to induction of expression, the cells were centrifuged for 15 min at 2700*g* and resuspended in Se-Met-containing minimal medium. Growth conditions and conditions for induction of expression were as described previously (Hall *et al.*, 1993), except that the culture time after induction of expression was prolonged to 6 h. Normal or Se-Met-enriched cystatin C was purified in a two-step procedure as described previously (Abrahamson *et al.*, 1988), resulting in a better than 95% pure protein preparation as assessed by SDS–polyacrylamide and agarose gel electrophoreses. The molecular mass of normal wild-type and Se-Met-enriched HCC was determined by electrospray mass spectrometry using a quadrupole mass spectrometer (Quattro; VG Masslab, Altrincham, Cheshire, England). The mass spectra of the selenomethionyl protein confirmed that the three Met residues (Fig. 1) have been fully substituted by Se-Met. A successful Met→Se-Met substitution was additionally confirmed by analysing the amino-acid composition of the Se-Met protein using the Biochrom 20 amino-acid analyser (Pharmacia Biotech) and its program for analysis of physiological amino acids. The Se-Met protein was hydrolyzed *in vacuo* for 24 h in 6 *M* HCl. For crystallized Se-Met HCC, X-ray absorption spectra confirmed the crystals to have a significant selenium content. Cystatin C devoid of ten N-terminal residues was obtained by incubation of recombinant wild-type human cystatin C with leukocyte elastase and was isolated as described previously (Abrahamson *et al.*, 1991).

```
1   SSPGKPPRLV GGPMDASVEE EGVRRALDFA VGEYNKASND
41  MYHSRALQVV RARKQIVAGV NYFLDVELGR TTCTKTQPNL
81  DNCPFHDQPH LKRKAFCSFQ IYAVPWQGTM TLSKSTCQDA
```

Figure 1
Amino-acid sequence of human cystatin C. The Met residues are shown in bold and the N-terminal residues, missing in the N-truncated variant, are shown in italics. Leu68, which in an amyloid-forming naturally occurring mutant is replaced by glutamine, is underlined.

2.2. Crystallization

Three variants of HCC were used in the present studies: (i) the native wild-type 120-residue protein, (ii) its Se-Met derivative and (iii) a truncated form of the native protein from which the first ten N-terminal amino acids have been removed by processing with leukocyte elastase (Fig. 1).

Two crystal forms (cubic and tetragonal) of the full-length protein and one crystal form (tetragonal) of the N-truncated protein were obtained. Both forms of the full-length protein were produced at 277 K using the vapour-diffusion method and the hanging-drop technique. Initial crystallization conditions (precipitant, pH) were determined using the sparse-matrix method (Jancarik & Kim, 1991) and Crystal Screen (Hampton Research). Crystals of the cubic form were grown using protein concentration of 10 mg ml^{-1}. Lyophilized protein was dissolved in 100 m*M* sodium acetate buffer pH 4.8 containing 20 m*M* $CaCl_2$. 4–6 µl droplets were equilibrated against 1 ml of reservoir with analogous buffer/$CaCl_2$ solution. After 2 d, the reservoir solution was supplemented with 100–200 µl MPD. Cube-shaped crystals (Fig. 2*a*) appeared within a week, reaching maximum dimensions of 0.5 × 0.5 × 0.5 mm after eight weeks. Crystals of the selenomethionyl derivative of HCC were grown under identical conditions. They reached maximum dimensions of 0.15 × 0.15 × 0.15 mm within about two weeks.

The tetragonal form of full-length human cystatin C was obtained using a modification of the conditions described above. Lyophilized protein was dissolved to a final concentration of 15 mg ml^{-1} in a solution containing 100 m*M* sodium acetate buffer pH 4.8, 20 m*M* $CaCl_2$ and 40–45% MPD. 10 µl droplets were equilibrated against reservoir solution (buffer, $CaCl_2$, MPD) for 2 d, after which time the concentration of MPD in the reservoir solution was increased to 50–60% (addition of 100–200 µl of MPD). Bipyramidal tetragonal crystals (Fig. 2*b*) appeared after 4–8 weeks, reaching maximum dimensions of 0.25 × 0.25 × 0.4 mm after a further 12 weeks.

Crystallization conditions for the N-truncated form of HCC were also determined using the sparse-matrix method and Crystal Screen. Protein samples (3 µl) at a concentration of 6 mg ml^{-1} were mixed with equal amounts of reservoir solution [0.4 *M* $(NH_4)H_2PO_4$]. Slab-like crystals (Fig. 2*c*) grew in hanging drops at room temperature within two weeks to maximum dimensions of 0.2 × 0.2 × 0.1 mm.

2.3. Data collection and processing

Diffraction data for the cubic form of the full-length protein were collected at 82 K using synchrotron radiation (Brookhaven National Laboratory, beamline X9B, λ = 0.98 Å) and a 345 mm MAR Research image-plate scanner. The crystal was mounted in a nylon-fibre loop and flash-frozen in a nitrogen-gas stream (Teng, 1990). A 3.1 Å data set was obtained by measuring the intensities of 68891 reflections. The final merged data set of 4071 unique reflections with $I/\sigma(I) > 0.0$ is 93% complete (95% in the last resolution shell) and is characterized by $R_{\rm int} = 0.086$ and $\langle I/\sigma(I)\rangle = 17.6$.

For the tetragonal form of full-length HCC, the data were collected at 120 K using synchrotron radiation (EMBL, c/o DESY, Hamburg, beamline BW7A, λ = 0.8885 Å) and a 180 mm MAR Research image-plate scanner. The crystal was mounted and flash-frozen as above. 192024 reflections with $I/\sigma(I) > 0.0$ were measured to 3.03 Å resolution. After merging, they were reduced to a unique data set consisting of 12273 reflections. This data set is 98% complete (98% in the last resolution shell) and is characterized by $R_{\rm int} = 0.123$ and $\langle I/\sigma(I)\rangle = 14.0$.

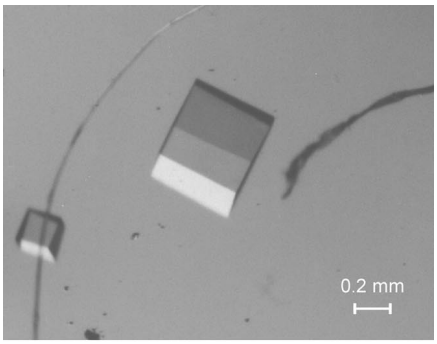
X-ray diffraction data for the crystals of N-truncated HCC were measured at 293 K using a 300 mm MAR Research detector and Cu $K\alpha$ radiation generated from an SRA2 rotating-anode generator (Siemens) operated at 45 kV and 112 mA. A total of 79910 reflections with $I/\sigma(I) > 0.0$ were measured and reduced to 22081 unique data extending to 2.7 Å. This data set is 84% complete (91% in the last resolution shell), with $R_{\rm int} = 0.108$ and $\langle I/\sigma(I)\rangle = 6.9$.

Indexing and integration of all images was performed in *DENZO* and scaling of the intensity data was performed in *SCALEPACK*; both are from the *HKL* program package (Otwinowski, 1993).

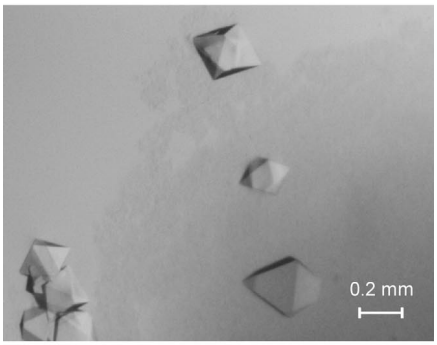
An X-ray absorption spectrum for selenomethionyl HCC was recorded using synchrotron radiation (EMBL, c/o DESY, Hamburg, beamline X31) and a crystalline sample of the protein (1 mm^3) packed in a thin-walled capillary. The spectrum had a steep absorption edge at $\lambda_e = 0.9803$ Å ($f'_{\rm min} = -8.87$e) and an absorption peak at $\lambda_p = 0.9800$ Å ($f''_{\rm max} = 9.75$e).

3. Results and discussion

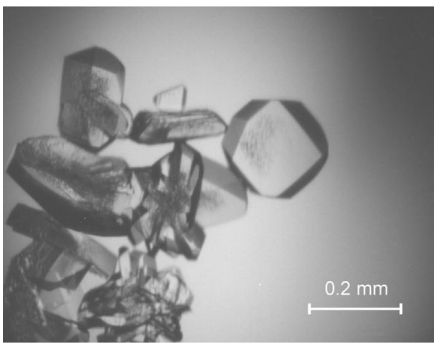
The masses, confirmed by mass spectrometry, for the native and Se-Met full-length HCC (13343 and 13484 Da, respectively) indicate that in variant (ii) all three S-methionine residues in the sequence (Fig. 1) were substituted by Se-methionine. Additionally, the amino-acid compositions of normal wild-type and Se-Met-enriched cystatin C were analyzed by amino-acid analysis after hydrolysis. Se-Met (Sigma) showed a retention time in a coupled HPLC chromatogram that coincided with the retention time for leucine. Upon analysis, the hydrolyzed Se-Met-enriched recombinant cystatin C showed a complete loss of the methionine peak in the chromatogram, with a concomitant increase of the leucine peak compared with normal recombinant cystatin C. The X-ray absorption spectrum, in which the Se *K* absorption edge is clearly visible, also confirms the presence of selenium.



(*a*)



(*b*)



(*c*)

Figure 2
Single crystals of HCC: (*a*) cubic and (*b*) tetragonal for the full-length protein and (*c*) tetragonal for the N-terminally truncated protein.

The cubic crystals of full-length HCC were very unstable in the X-ray beam at room temperature. An attempt to collect X-ray diffraction data at 297 K was unsuccessful, as the crystals rapidly lost their diffraction power and stopped diffracting within 1 h. Lowering the temperature of the environment to 282 K significantly increased the stability of the crystals and allowed an incomplete data set to be collected to 4.5 Å from one crystal. At 82 K, the crystals were stable enough to allow the measurement of a complete data set to 3.1 Å resolution. These cubic crystals belong to the *I*432 space group and have a unit-cell parameter $a = 140.4$ Å.

Attempts to collect useful MAD data for the cubic crystals of the selenomethionyl derivative of HCC have failed because of the very limited resolution (worse than 4.5 Å). This might be attributable to the relatively small size of the Se-Met HCC crystals and to the weak intensity of the synchrotron beamline available for these experiments.

The crystals of the tetragonal form of full-length HCC were more stable in the beam. A room-temperature diffraction experiment resulted in an incomplete data set to 4.3 Å resolution, while at 120 K a complete data set to 3.03 Å resolution was obtained. The crystals belong to the $P4_12_12$ space group (or its enantiomorph $P4_32_12$) and have unit-cell dimensions $a = 91.5$, $c = 144.5$ Å.

Crystals of the N-terminally truncated HCC were very stable in the X-ray beam and diffracted to 2.7 Å resolution. They belong to the tetragonal system, with unit-cell parameters $a = 68.8$, $c = 206.5$ Å. The symmetry and systematic absences are consistent with the $P4_2$ space group.

Analysis of the Matthews volume (Matthews, 1968) for the cubic form of full-length HCC indicates rather unambiguously that two monomers of the protein are present in the asymmetric unit (2.16 $Å^3\,Da^{-1}$). In the large asymmetric unit of the tetragonal form of the full-length protein, between three (3.78 $Å^3\,Da^{-1}$) and seven (1.62 $Å^3\,Da^{-1}$) monomers can be accommodated. In the asymmetric unit of the tetragonal form of N-truncated HCC, between five (3.66 $Å^3\,Da^{-1}$) and 11 (1.66 $Å^3\,Da^{-1}$) independent monomers may be present. The propensity of HCC to crystallize with multiple copies of the molecule in the asymmetric unit, in combination with the additional possibilities offered by the point-symmetry elements of the unit cells, may be indicative of the tendency of the protein to oligomerize. Such oligomerization might reflect the amyloid-forming property of Leu68Gln cystatin C, as earlier observations demonstrate that both wild-type and Leu68Gln-substituted cystatin C are capable of forming dimers (Abrahamson & Grubb, 1994; Ekiel & Abrahamson, 1996; Ekiel *et al.*, 1997).

The research of MJ was supported in part by an International Research Scholar's award from the Howard Hughes Medical Institute. The research was sponsored in part by grant 279/P04/97/13 from the State Committee for Scientific Research (KBN), Poland and by the Swedish Medical Research Council (projects 5196 and 9915). We thank Veronica Lindström, Anne-Cathrine Löfström and Dr Bernd Gerhartz for valuable assistance with protein expression and purification and Dr Gottfried Palm for assistance at the Brookhaven synchrotron.

References

Abrahamson, M. (1993). *J. Braz. Assoc. Adv. Sci.* **45**, 229–304.
Abrahamson, M., Dalbøge, H., Olafsson, I., Carlsen, S. & Grubb, A. (1988). *FEBS Lett.* **236**, 14–18.
Abrahamson, M. & Grubb, A. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 1416–1420.
Abrahamson, M., Jonsdottir, S., Olafsson, I., Jensson, O. & Grubb, A. (1992). *Hum. Genet.* **89**, 377–380.
Abrahamson, M., Mason, R. W., Hansson, H., Buttle, D. J., Grubb, A. & Ohlsson, K. (1991). *Biochem. J.* **273**, 621–626.
Abrahamson, M., Ritonja, A., Brown, M. A., Grubb, A., Machleidt, W. & Barrett, A. J. (1987). *J. Biol. Chem.* **262**, 9688–94.
Björck, L., Åkesson, P., Bohus, M., Trojnar, J., Abrahamson, M., Olafsson, I. & Grubb, A. (1989). *Nature (London)*, **337**, 385–386.
Björck, L., Grubb, A. & Kjellen, L. (1990). *J. Virol.* **64**, 941–943.
Bode, W., Engh, R., Musil, D., Thiele, U., Huber, R., Karshikov, A., Brzin, J., Kos, J. & Turku, V. (1988). *EMBO J.* **7**, 2593–2599.
Collins, A. R. & Grubb, A. (1991). *Antimicrob. Agents Chemother.* **35**, 2444–2446.
Dieckmann, T., Mitschang, L., Hofmann, M., Kos, J., Turk, V., Auerswald, E. A., Jaenicke, R. & Oschkinat, H. (1993). *J. Mol. Biol.* **234**, 1048–1059.
Ekiel, I. & Abrahamson, M. (1996). *J. Biol. Chem.* **271**, 1314–1321.
Ekiel, I., Abrahamson, M., Fulton, D. B., Lindahl, P., Storer, A. C., Levadoux, W., Lafrance, M., Labelle, S., Pomerleau, Y., Groleau, D., LeSauteur, L. & Gehring, K. (1997). *J. Mol. Biol.* **271**, 266–277.
Engh, R. A., Dieckmann, T., Bode, W., Auerswald, E. A., Turk, V., Huber, R. & Oschkinat, H. (1993). *J. Mol. Biol.* **234**, 1060–1069.
Grubb, A. & Löfberg, H. (1982). *Proc. Natl Acad. Sci. USA*, **79**, 3024–3027.

Hall, A., Dalbøge, H., Grubb, A. & Abrahamson, M. (1993). *Biochem. J.* **291**, 123–129.
Hendrickson, W. A. (1991). *Science*, **254**, 51–58.
Irvine, J. W., Coombs, G. H. & North, M. J. (1992). *FEMS Microbiol. Lett.* **96**, 67–72.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **21**, 916–924.
Lindahl, P., Abrahamson, M. & Björk, I. (1992). *Biochem. J.* **281**, 49–55.
Martin, J. R., Craven, C. J., Jerala, R., Kroon-Zitko, L., Zerovnik, E., Turk, V. & Waltho, J. P. (1995). *J. Mol. Biol.* **246**, 331–343.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Olafsson, I., Thorsteinsson, L. & Jensson, O. (1996). *Brain Pathol.* **6**, 121–126.
Otwinowski, Z. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.
Stubbs, M. T., Laber, B., Bode, W., Huber, R., Jerela, R., Lenarèiè, B. & Turku, V. (1990). *EMBO J.* **9**, 1939–1990.
Teng, T.-Y. (1990). *J. Appl. Cryst.* **23**, 387–391.
Turk, V. & Bode, W. (1991). *FEBS Lett.* **285**, 213–219.