Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Expression, purification, crystallization and preliminary X-ray diffraction studies of human cathepsin F complexed with an irreversible vinyl sulfone inhibitor

Joseph D. Ho, Yana Meltser, Joseph J. Buggy, James T. Palmer, Kyle C. Elrod, Hedy Chan, Kyle D. Mortara and John R. Somoza*

Celera Inc., 180 Kimball Way, South San Francisco, CA 94080, USA

Correspondence e-mail: john.somoza@celera.com

Received 12 July 2002
Accepted 24 September 2002

Cathepsin F is a cysteine protease believed to be involved in the antigen-presenting process of the class II major histocompatibility complex (MHC-II) in macrophages. It has been expressed, purified and crystallized. A complete data set to a resolution of 2.5 Å has been collected at room temperature. The Laue group was determined to be orthorhombic, space group $P2_12_12$, with unit-cell parameters a = 68.9, b = 104.8, c = 68.5 Å.

1. Introduction

Cathepsin F is a lysosomal papain-family cysteine protease that is expressed in many human tissues: medium to high expression levels are observed in heart, brain, skeletal muscle, kidney and pancreas, whereas lower levels are detected in placenta, lung and liver (Nägler *et al.*, 1999; Santamaría *et al.*, 1999; Wang *et al.*, 1998). Its significant level of expression in some cancer cell lines also opens the possibility that this enzyme may be involved in tumor progression (Santamaría *et al.*, 1999; Wang *et al.*, 1998).

Cathepsin F has the longest proregion of any member of the human papain-family enzymes. The proregion contains 251 amino-acid residues; multiple sequence alignments and homology profiles indicate that a portion of this region contains structural similarities with cysteine protease inhibitors of the cystatin superfamily (Nägler *et al.*, 1999; Wex, Levy *et al.*, 1999; Wex, Wex *et al.*, 1999). It has been suggested that this cystatin-like domain may function as an endogenous cysteine protease inhibitor and may exert physiological effects on other cysteine proteases. Experiments are currently in progress to elucidate the role of this domain (Nägler *et al.*, 1999; Wex, Levy *et al.*, 1999; Wex, Wex *et al.*, 1999). Aside from this intriguing feature, cathepsin F is similar to other cysteine proteases in that it is translated as a preproenzyme, processed to the corresponding proenzyme, targeted to the lysosome by the mannose-6-phosphate signal attached to it and activated by cleavage and dissociation from its proregion. Its proteolytic activity has an acid pH optimum (Wang *et al.*, 1998) and the mature region of the protein displays 34–42% identity when compared with mature human cathepsins H, K, L, V, O, S and W (Nägler *et al.*, 1999).

Recently, data obtained by Shi *et al.* (2000) suggest that cathepsin F may play a role in the processing of class II major histocompatibility complex (MHC-II) molecules in macrophages. Similar to cathepsin S, cathepsin F cleaves Iip10 (the 10 kDa remnant of the invariant chain) to generate CLIP (class II-associated invariant chain peptide), which remains bound to the class II binding groove until exchanged with an antigenic peptide. The importance of cathepsins S and F in MHC-II antigen presentation makes these two enzymes potential targets in treating diseases caused by hyperimmune responses, such as asthma or rheumatoid arthritis. A crystal structure of cathepsin S has been previously reported by McGrath *et al.* (1998). The structure of cathepsin F will provide additional insights important for the design and discovery of specific inhibitors for one or both of these enzymes.

In this report, we describe the expression, purification, activation, kinetic characterization, crystallization and preliminary X-ray diffraction studies for cathepsin F complexed with an irreversible vinyl sulfone inhibitor.

2. Experimental procedures

2.1. Cloning and expression of human procathepsin F

Human procathepsin F (prohcatF) was cloned from a human brain Quick Clone cDNA library (Clontech) using the polymerase chain reaction (PCR) with the forward primer 5′-*Xho*I (5′-GGGGTATCT**CTCGAG**AAAA-GAGAGGCTGAAGCTGCCCCCGCCCAG-CCCCGAGCCGC-3′) and the reverse primer 3′-*Avr*II (5′-CTGTGA**CCTAGG**TCAGTCCA-CCACCGCCGA-3′) (restriction sites in bold) with 30 rounds of amplification using Platinum Pfx DNA polymerase (Life Technologies, Inc.). The ∼1.5 kbp PCR product was cloned using the Topo TA vector (Invitrogen) and the prohcatF sequence was verified by DNA-sequencing analysis. The prohcatF coding sequence was then subcloned from the Topo TA vector into the *Xho*I and *Avr*II sites of the

Pichia pastoris expression vector pPIC9 (Invitrogen) (prohcatFpPIC9), placing the prohcatF sequence after the strong secretory yeast α-factor pheromone signal sequence. This plasmid was linearized with *Sal*I, electroporated into *P. pastoris* strain GS115 (mut$^+$) using standard procedures (Invitrogen) and selected on media lacking histidine. 16 colonies were screened by standard pilot-scale expression (Invitrogen) in shaker flasks, but all were found to have no cathepsin F activity. However, Wang *et al.* (1998) had reported successful expression of a truncated version of prohcatF in *P. pastoris*. This truncated prohcatF starts at methionine 128 of the proregion. To create this version of prohcatF, prohcatFpPIC9 was PCR amplified with the forward primer 5′-M2*Xho*I (5′-GATC**CTCGAG**AAAAGAGAGGCTGAAGCTATTTCTTCTCTGTCCCAAAAC-3′) and the reverse primer 3′-M2*Not*I (5′-CTGTGA**GCGGCCGC**TCAGTCCACCACCGCCGA-3′) (restriction sites in bold). The resulting ∼1 kbp PCR product was digested and subcloned into the *Xho*I and *Not*I sites of pPIC9 (prohcatFm2pPIC9). The sequence of this clone was verified by DNA-sequence analysis. After pilot-scale protein expression, several clones were found with cathepsin F activity. ProhcatFm2 was then subcloned into the *Xho*I and *Not*I sites of a cloning vector, pBluescript SK (+) (Stratagene) (prohcatFm2pBS). This construct was subsequently used to generate the deglycosylated version of prohcatFm2 using the QuikChange site-directed mutagenesis procedures (Stratagene).

Human procathepsin F contains five potential *N*-linked glycosylation sites: two in the proregion at positions 141p (N_{141p}ET) and 176p (N_{176p}RT) and three in the mature region at positions 97 (N_{97}FS), 108 (N_{108}DS) and 170 (N_{170}RS). The two potential glycosylation sites in the proregion were left unmutated. The three sites in the mature region were rendered refractory to glycosylation by substitution of asparagine with glutamine. The N97Q, N108Q and N170Q mutations were each created by amplifying upstream and downstream of the mutagenic sites with antiparallel primers. N97Q was made with the forward primer 5′-CACATGCAGTCCTGC**CAG**TTCTCAGCAGAGAAG-3′ and its reverse complement; N108Q was made with the forward primer 5′-CCAAGGTCTACATC**CAG**GACTCCGTGGAGCTG-3′ and its reverse complement; N170Q was made with the forward primer 5′-GTGGGCTACGGC**CAG**CGCTCTGACGTTC-3′ and its reverse complement (mutations are in bold). By taking advantage of convenient restriction sites, fragments of the gene were then swapped to create a construct containing all three deglycosylation (DG) mutations (prohcatFm2DGpBS). The prohcatFm2DG sequence was verified by DNA-sequencing analysis and subcloned into pPIC9 (prohcatFm2DGpPIC9) for *P. pastoris* expression.

2.2. Fermentation

The fermentation conditions described by Invitrogen were used initially; however, owing to poor cell viability, a mix-feed fermentation protocol was developed to produce procathepsin F. 500 ml BMGY medium (Invitrogen) was placed in a 4 l baffled flask, inoculated with 1 ml of the glycerol stock of the selected clone and shaken overnight at 250 rev min^{-1} at 303 K. In the morning, another 500 ml of BMGY was added and the culture was shaken for a further 4 h. The 1 l inoculum was pumped into a sterilized reactor with 6 l of BMGY and 2 ml Antifoam 289 (Sigma). The resulting culture was grown overnight with the following control settings: agitation at 1000 rev min^{-1}, air at 10 l min^{-1} and pH 5.5 (controlled with 28% ammonium hydroxide). In the morning, glycerol feed [50%(*w*/*v*) glycerol in distilled deionized water with 12 ml l^{-1} Pichia Trace Metal Salts (Invitrogen)] was initiated at 1.4 ml min^{-1} for 4 h. At the end of the glycerol feed, the OD_{600} was about 200. Induction feed (37.5% glycerol, 25% methanol, in distilled deionized water with 12 ml l^{-1} Pichia Trace Metal Salts, filter sterilized) was then initiated at 0.7 ml min^{-1} and continued for 72 h. At the end of the induction, the OD_{600} was greater than 500. The supernatant with the expressed protein was harvested by centrifugation and filtered through a 0.2 µm cellulose acetate filter.

2.3. Purification

Clarified *P. pastoris* supernatant was concentrated to 1 l and diafiltered first into 50 m*M* sodium acetate pH 4.8, 1 m*M* EDTA, 1 m*M* DTT, 150 m*M* NaCl and then into 50 m*M* sodium acetate pH 4.0, 1 m*M* EDTA, 1 m*M* DTT, 150 m*M* NaCl (buffer *A*) using a Millipore Pellicon 2 Cassette (10 kDa molecular-weight cutoff PLCGC-C membrane). The resulting supernatant was clarified by centrifugation at ∼15 000*g* and 0.2 µm filtration. The ultrafiltration and diafiltration procedures were performed at 277 K and all subsequent purification steps were performed at room temperature. The supernatant was loaded onto an HR 16/10 column (Pharmacia) with 20 ml of Source 15S resin (Pharmacia), washed with buffer *A* and eluted with a 0.15–1 *M* NaCl gradient in 40 ml with 50 m*M* sodium acetate pH 4.0, 1 m*M* EDTA, 1 m*M* DTT, 1 *M* NaCl (buffer *B*). All steps were performed at 1 ml min^{-1}. Procathepsin F eluted between 0.6 and 0.8 *M*

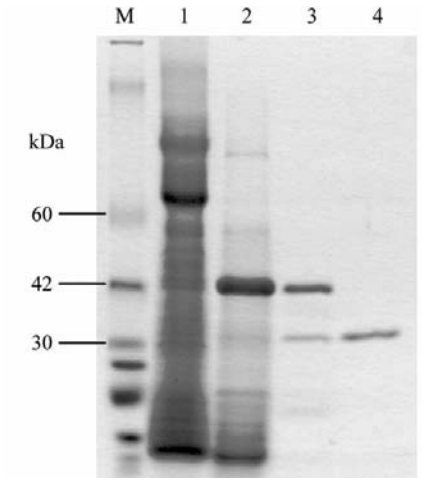

Figure 1
Coomassie-stained reducing SDS–PAGE showing the purity of cathepsin F at each purification step. Lane M, molecular-weight markers; lane 1, fermentation supernatant; lane 2, after first column (cation exchange; Source 15S); lane 3, after second column (gel filtration; Superdex 75); lane 4, after activation with immobilized pepsin and filtering the pepsin away.

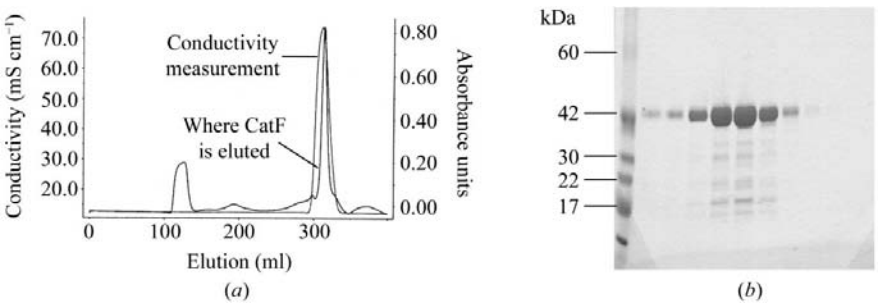

Figure 2
(*a*) Chromatogram of the gel-filtration step, showing that cathepsin F is eluted along with the salt. (*b*) Coomassie-stained reducing SDS–PAGE of the cathepsin F peak.

NaCl. At this point, procathepsin F was approximately 70% pure on SDS–PAGE (Fig. 1, lane 2).

To further purify the protein, it was passed at 1 ml min^{-1} through two XK 16/100 columns (Pharmacia) linked in series containing 360 ml of Superdex 75 gel-filtration resin (Pharmacia). The buffer used for the isocratic elution was 50 m*M* sodium acetate pH 4.0, 1 m*M* EDTA, 1 m*M* DTT, 50 m*M* NaCl. Surprisingly, procathepsin F did not elute at the theoretical molecular weight; instead, it was found at the end of the elution along with the NaCl (Fig. 2*a*). The resulting protein was at least 95% pure as judged by SDS–PAGE (Fig. 2*b*).

2.4. Activation

Procathepsin F was converted to the active form of the enzyme by digesting the proregion with pepsin. To activate cathepsin F while minimizing possible detrimental autocatalysis, the irreversible vinyl sulfone cathepsin F inhibitor, HCl.Morph-pip-Phe-NvaVSPh (Fig. 3), was added at a 10:1 molar ratio of inhibitor to procathepsin F. The inhibitor was not included while activating the sample used for kinetic analysis. 2 ml of immobilized pepsin gel (~4 mg of pepsin) (Pierce) per ~3 mg of procathepsin F was then mixed in a maximum volume of 25 ml in a 50 ml conical tube. The 50 ml conical tube with the mixture was placed horizontally on a circular-rotation incubator rotating at 75 rev min^{-1} at 310 K for 2 h. In order to completely activate the cathepsin F to a single species as judged by SDS–PAGE, greater than 50% loss of cathepsin F was unavoidable. After activation, the pepsin gel was separated from the protein solution by passing the mixture through a fritted 10 ml column (Biorad). The protein solution now contained the activated cathepsin F (hcatFDG) with the inhibitor bound and retained 95% purity as verified by SDS–PAGE (Fig. 1, lane 4). N-terminal sequencing revealed the sequence NH_2-KSVG-DLAPP, which is six residues upstream from the cleavage site predicted by Wang *et al.* (1998). The average protein yield was 1 mg of purified and activated cathepsin F per 6 l of *P. pastoris* fermentation supernatant.

2.5. Enzyme-kinetics analysis comparing deglycosylated and wild-type cathepsin F

Wild-type human cathepsin F was obtained from Dr Dieter Brömme (Mount Sinai School of Medicine, NY, USA). The deglycosylated and wild-type enzymes were active-site titered with E-64, a potent irreversible inhibitor of cathepsin F (Katunuma & Kominami, 1995), and characterized using the fluorogenic substrate Z-Phe-Arg-AMC (Bachem). All enzyme-activity measurements were performed at room temperature using 96-well kinetic plate readers. Reaction velocities were monitored at varying substrate concentrations (1.4, 1.6, 2.5, 3.3, 5.0, 10.0, 11.1, 12.5, 14.3, 16.6 and 20.0 µ*M*) by following the hydrolysis of the aminomethylcoumarin substrate ($ex_{355\lambda}$, $em_{460\lambda}$) over 5 min. Substrate was diluted in buffer (50 m*M* MES pH 6.5, 100 m*M* NaCl, 2.5 m*M* EDTA, 2.5 m*M* DTT, 0.01% BSA and 10% DMSO) and reactions were initiated by the addition of enzyme (1 n*M*) in a final reaction volume of 100 µl. The values of K_m and k_{cat} were derived from least-squares regression fit to the Michaelis–Menten equation using the *Enzyme Kinetics* v. 1.5 software package (Trinity Software Inc.).

2.6. Crystallization and preliminary data collection

The purified complex (~1 mg) was placed in an Amicon Ultrafree-15 centrifugal filter unit with a Biomax-10 membrane, washed into the crystallization buffer (50 m*M* sodium acetate pH 4.0, 1 m*M* EDTA and 150 m*M* NaCl) to remove excess irreversible inhibitors and concentrated to approximately 3 mg ml^{-1}. The complex was crystallized using the hanging-drop vapor-diffusion method. Factorial screening (Jancarik & Kim, 1991) using kits obtained from Hampton Research (Laguna Niguel, CA, USA) was used to identify lead conditions. The best results were obtained using a reservoir that consisted of 0.2 *M* magnesium acetate, 0.1 *M* sodium cacodylate pH 6.0–6.6 and 25%(*w*/*v*) PEG 8000 at 291 K (crystals were also obtained under similar conditions using PEG 4000 and PEG 3350). The drop consisted of 0.5 µl of the well solution and 1 µl of a solution consisting of approximately 3 mg ml^{-1} complex. A 2.5 Å data set was obtained at room temperature. These data were recorded on an R-AXIS IV++ (MSC Inc.) image-plate system using 1.54 Å X-rays produced by a Rigaku RU-H3R (MSC Inc.) rotating-anode generator operating at 50 kV and 100 mA and equipped with Osmic mirrors (Osmic Inc.). The data were reduced using the *DENZO* and *SCALEPACK* packages (Otwinowski & Minor, 1993).

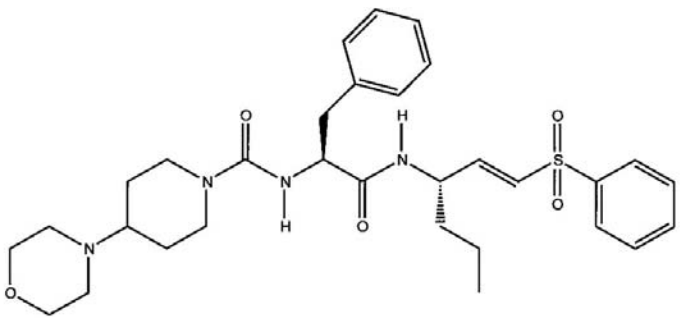

Figure 3
4-Morpholin-4-yl-piperidine-1-carboxylic acid [1-(3-benzenesulfonyl-1-propyl-allylcarbamoyl)-2-phenylethyl]-amide hydrochloride (HCl.Morph-pip-Phe-NvaVSPh).

3. Results and discussion

The human procathepsin F sequence includes five potential glycosylation sites. The three potential glycosylation sites in the mature domain were mutated from Asn to Gln (a methylene addition to the side chain) to avoid potential crystallization problems arising from heterogeneous glycosylation. After the protein was purified and activated, the kinetics of the deglycosylated cathepsin F and wild-type enzymes were measured and compared. The measured Michaelis constant for the deglycosylated form was twofold higher than that of the wild-type enzyme (K_m of 43 *versus* 17 µ*M*) and the k_{cat} was twofold lower (1.5 *versus* 2.9 s^{-1}). Thus, the deglycosylated enzyme is catalytically competent, although it displays a modest fourfold reduction in catalytic efficiency for hydrolysis of this substrate. Although the loss of carbohydrate might attribute to the slightly lower catalytic efficiency of the deglycosylated enzyme, alignment of the sequence of mature cathepsin F with the sequences and structures of mature cathepsin S (McGrath *et al.*, 1998) and K (McGrath *et al.*, 1997) showed that the glycosylation sites were all located at the surface of the enzyme, away from the active site.

The expression yield of procathepsin F (prohcatFm2DG) was relatively low (~0.7 mg l^{-1}) compared with that of procathepsin K in the same expression system (~38 mg l^{-1}; Linnevers *et al.*, 1997); therefore, finding an efficient first purification

step was crucial. A construct of procathepsin F with an N-terminal His tag was made and expressed; however, the N-terminus of the protein was found to be degraded during expression. Binding to both cation- and anion-exchange resins was explored at different pH values, but contaminating proteins of greater quantity competed with procathepsin F. Eventually, a condition for binding to cation-exchange resin was found. In the cation-exchange step, inclusion of 150 m*M* NaCl in the buffer-exchanged fermentation supernatant prevented binding of a major contaminating protein (of ∼60 kDa) (Fig. 1), allowing procathepsin F to be retained on the resin. During the elution, a rather short 0.15–1 *M* NaCl gradient (40 ml on a 20 ml column) was used to generate a small volume of pooled fractions.

In the gel-filtration step, two XK 16/100 columns were poured with Superdex 75 gel-filtration resin (Pharmacia) and linked in series. The columns showed a normal separation profile when calibrated using the LMW Calibration Kit (Pharmacia) at neutral pH. Interestingly, instead of eluting at its theoretical molecular weight, procathepsin F eluted along with the salt fractions (Fig. 2*a*). This unusual interaction of the protein with the resin provided an excellent purification step that yielded a single band on SDS–PAGE (Fig. 2*b*). We suspect that the nature of the interaction is ion exchange because Sephacryl S-200 (a dextran-based gel-filtration media) was previously reported to have adsorptive properties at pH 3.5 and can be desorbed by elution with 0.5 *M* NaCl (Belew *et al.*, 1978). The interaction involves the proregion because after activating procathepsin F the mature protein elutes as expected for its molecular weight and does not exhibit any interaction with the resin (data not shown).

Table 1
Data-collection and processing statistics.

Values in parentheses are for the highest resolution shell (2.59–2.50 Å).

Space group	$P2_12_12$
Unit-cell parameters (Å)	a = 68.9, b = 104.8, c = 68.5
Resolution range (Å)	100.0–2.50
No. of observations	68358
No. of unique reflections	17490
Completeness (%)	98.9 (95.2)
$R_{sym}(I)$ (%)	14.3 (60.2)
$\langle I\rangle/\langle$error$\rangle$	10.5 (2.0)

We activated the purified procathepsin F by mixing the protein with immobilized pepsin gel to eliminate a purification step to remove pepsin. We also included an irreversible inhibitor to complex mature cathepsin F as soon as it was produced. During the activation process we discovered intermediate species of various sizes, but upon complete activation there was a single band which corresponded to mature cathepsin F (Fig. 1, lane 4). We stopped the activation process by filtering the immobilized pepsin gel away.

The complex was crystallized readily under the conditions described above. However, large diffraction-quality crystals were only obtained when the tray was left in the incubator undisturbed for at least two weeks.

A complete data set at 2.5 Å was obtained at room temperature. The autoindexing performed by *DENZO* showed that crystals were primitive orthorhombic and the systematic absences suggested that the space group was $P2_12_12$ (this space group was later confirmed by successfully identifying a molecular-replacement solution). Although the R_{merge} rose with redundancy, the signal-to-noise ratio of the high-resolution shell was still above 2. Other relevant diffraction data statistics are shown in Table 1.

The crystal structure of human cathepsin F complexed with the vinyl sulfone inhibitor will be published elsewhere. This structure will provide a high-resolution description of the active site and thereby contribute to the design of potent and selective inhibitors.

References

Belew, M., Porath, J., Fohlman, J. & Janson, J. C. (1978). *J. Chromatogr.* **147**, 205–212.

Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.

Katunuma, N. & Kominami, E. (1995). *Methods Enzymol.* **251**, 382–397.

Linnevers, C. J., McGrath, M. E., Armstrong, R., Mistry, F. R., Barnes, M. G., Klaus, J. L., Palmer, J. T., Katz, B. A. & Brömme, D. (1997). *Protein Sci.* **6**, 919–921.

McGrath, M. E., Klaus, J. L., Barnes, M. G. & Brömme, D. (1997). *Nature Struct. Biol.* **4**, 105–109.

McGrath, M. E., Palmer, J. T., Brömme, D. & Somoza, J. R. (1998). *Protein Sci.* **7**, 1294–1302.

Nägler, D. K., Sulea, T. & Ménard, R. (1999). *Biochem. Biophys. Res. Commun.* **257**, 313–318.

Otwinowski, Z. & Minor, W. (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.

Santamaría, I., Velasco, G., Pendás, A. M., Paz, A. & López-Otín, C. (1999). *J. Biol. Chem.* **274**, 13800–13809.

Shi, G. P., Bryant, R. A., Riese, R., Verhelst, S., Driessen, C., Li, Z., Brömme, D., Ploegh, H. L. & Chapman, H. A. (2000). *J. Exp. Med.* **191**, 1177–1186.

Wang, B., Shi, G. P., Yao, P. M., Li, Z., Chapman, H. A. & Brömme, D. (1998). *J. Biol. Chem.* **273**, 32000–32008.

Wex, T., Levy, B., Wex, H. & Brömme, D. (1999). *Biochem. Biophys. Res. Commun.* **259**, 401–407.

Wex, T., Wex, H. & Brömme, D. (1999). *Biol. Chem.* **380**, 1439–1442.