Biochem. J. (2012) **442**, 563–572 (Printed in Great Britain) doi:10.1042/BJ20111360

online data

A molten globule-to-ordered structure transition of *Drosophila melanogaster* crammer is required for its ability to inhibit cathepsin

Tien-Sheng TSENG*[1], Chao-Sheng CHENG*[1], Dian-Jiun CHEN*, Min-Fang SHIH*, Yu-Nan LIU*, Shang-Te Danny HSU*†[2] and Ping-Chiang LYU*‡[2]

*Institute of Bioinformatics and Structural Biology, National Tsing Hua University, Hsinchu, 30013, Taiwan, †Institute of Biological Chemistry, Academia Sinica, Taipei 115, Taiwan, and ‡Graduate Institute of Molecular Systems Biomedicine, China Medical University, Taichung, 40402, Taiwan

Drosophila melanogaster crammer is a novel cathepsin inhibitor that is involved in LTM (long-term memory) formation. The mechanism by which the inhibitory activity is regulated remains unclear. In the present paper we have shown that the oligomeric state of crammer is pH dependent. At neutral pH, crammer is predominantly dimeric *in vitro* as a result of disulfide bond formation, and is monomeric at acidic pH. Our inhibition assay shows that monomeric crammer, not disulfide-bonded dimer, is a strong competitive inhibitor of cathepsin L. Crammer is a monomeric molten globule in acidic solution, a condition that is similar to the environment in the lysosome where crammer is probably located. Upon binding to cathepsin L, however, crammer undergoes a molten globule-to-ordered structural transition. Using high-resolution NMR spectroscopy, we have shown that a cysteine-to-serine point mutation at position 72 (C72S) renders crammer monomeric at pH 6.0 and that the structure of the C72S variant highly resembles that of wild-type crammer in complex with cathepsin L at pH 4.0. We have determined the first solution structure of propeptide-like protease inhibitor in its active form and examined in detail using a variety of spectroscopic methods the folding properties of crammer in order to delineate its biomolecular recognition of cathepsin.

Key words: cathepsin, crammer, long-term memory (LTM), molten globule, propeptide-like protease inhibitor.

INTRODUCTION

Drosophila melanogaster has been developed as a model system for studying learning and memory because of its short lifespan and the relatively simple and facile nervous system [1,2]. To date, several genes have been identified to be involved in the formation of *Drosophila* olfactory memory [2–4], but little is known about the genetic basis and mechanisms that contribute to LTM (long-term memory) formation. One *Drosophila* mutant, crammer, exhibits a specific LTM defect [5,6]. The overexpression of crammer in glial cells impairs LTM, suggesting that the expression level of crammer is of functional importance with regard to LTM formation. The crammer gene encodes a cysteine protease inhibitor, and potential targets include cathepsins. The structural properties of crammer are hitherto uncharacterized. Crammer shares approximately 45 % primary sequence identity with the proregions of the *Bombyx mori* and *D. melanogaster* cysteine proteases, suggesting that crammer belongs to a class of cysteine protease inhibitors that have propeptide-like inhibitory activity [7]. Such inhibitors, originally identified in mouse-activated T-cells and mast cells, are also known as cytotoxic T-lymphocyte antigen (CTLA) 2α and 2β [8], and they exhibit inhibitory activities against papain and cathepsin L [9,10]. Similar inhibitors have also been identified in other organisms such as *B. mori* [11–13].

Using yeast two-hybrid assay, a number of crammer-interacting proteins have been identified. These include cathepsins B and L, and capping protein β [14,15]. Cathepsins are synthesized as zymogens, each of which contains an N-terminal proregion and a mature protein sequence. The proregion contains a signal peptide and a propeptide. The propeptide is required for intracellular targeting [16], protein folding [17] and enzyme inhibition [18]. Removal of the propeptide by other proteases [19,20] or by autocleavage at acidic pH [21] activates cathepsins. As the name suggests, inhibition by a propeptide-like protease inhibitor is linked to its sequence and/or structural similarity to the proregion of its target protease. Crammer contains a GNFD motif, GXNX(F/L)XD, which is highly conserved among propeptide-like inhibitors [13] and is found in the proregions of most cysteine proteases [22]. The GNFD motif is essential for auto-activation and protein folding [22], but its inhibitory role by propeptide-like inhibitors is unclear. Crammer also contains a consensus ERFNIN motif, $EX_3RX_3(F/Y)X_2(N/S)X_3IX_3N$, which is essential for protease inhibition [8,9]. Finally, crammer contains a conserved motif at the N-terminus, YKX_4KXY, which serves as a lysosomal-targeting sequence [23,24], suggesting that crammer is a lysosomal protein.

Despite the biological significance of propeptide-like protease inhibitors, there has hitherto been no structural information available. Although monomeric and dimeric crammer have been reported to inhibit cathepsin [15], the molecular mechanism of its inhibitory mechanism remains elusive. To clarify how the structure of crammer relates to its function, in the present paper we report detailed enzyme kinetic analysis and the solution structure of crammer using heteronuclear NMR spectroscopy. In light of the emerging roles of cathepsins B and L in neurodegenerative diseases [25–27], understanding the molecular basis of the protease inhibitory activity of crammer will provide valuable insight into the development of treatments for these diseases.

Abbreviations used: ANS, 8-anilinonaphthalene-l-sulfonic acid; BMRB, BioMagResBank; DTT, dithiothreitol; E-64, *trans*-epoxysuccinyl-L-leucylamido-(4-guanidino)butane; HSQC, heteronuclear single quantum correlation; IPTG, isopropyl β-D-thiogalactopyranoside; LTM, long-term memory; NOE, nuclear Overhauser effect; hetNOE, heteronuclear NOE; NOESY, nuclear Overhauser enhancement spectroscopy; SEC, size-exclusion chromatography.

[1] These authors contributed equally to this work.

[2] Correspondence may be addressed to either of these authors (email lslpc@life.nthu.edu.tw or sthsu@gate.sinica.edu.tw).

The structural co-ordinates reported will appear in the PDB under accesion code 2L95.

EXPERIMENTAL

Materials

ANS (8-anilinonaphthalene-l-sulfonic acid) and E-64 [*trans*-epoxysuccinyl-L-leucylamido-(4-guanidino)butane] were purchased from Sigma–Aldrich. 2-Mercaptoethanol, chloramphenicol and sodium azide were supplied by Merck Research Laboratories. Ampicillin, DTT (dithiothreitol) and IPTG (isopropyl β-D-thiogalactopyranoside) were obtained from MDBio (Taipei, Taiwan). EDTA and Blue Dextran 2000 were purchased from USB. Glacial acetic acid and Triton X-100 were supplied by ECHO Chemical and Amresco respectively. All chemicals were of analytical grade.

Protein expression and purification

A two-step PCR was used to synthesize the crammer gene [28]. The primers were obtained from Mission Biotechnology. The crammer gene was assembled and amplified by PCR [28]. The final product was cloned into a pAED4 vector (Dr Min Lu, Weill Medical College of Cornell University, New York, NY, U.S.A.) and expressed in *Escherichia coli* Rosetta (DE3) strain (Merck). Site-specific point mutations were prepared with the QuikChange site-directed mutagenesis kit (Stratagene) and expressed and purified following the same protocol as described for wild-type crammer. The sequences of the recombinant genes were verified by DNA sequencing (Mission Biotechnology).

E. coli cells harbouring pAED4 that contain the crammer or mutant genes were cultured in Luria-Bertani medium containing 100 mg/ml ampicillin and 30 mg/ml chloramphenicol at 37 °C until the D_{600} value of the culture reached 0.7. IPTG (final concentration, 1 mM) was added to the cell culture to initiate recombinant protein overexpression. ^{15}N-labelled and ^{15}N/^{13}C-labelled crammer samples were obtained from *E. coli* cultures incubated in M9 medium containing (1 g/l) $^{15}NH_4Cl$ and/or (2 g/l) [^{13}C]glucose/$^{15}NH_4Cl$ (Cambridge Isotope Laboratories) [29]. Whole cells were lysed with glacial acetic acid and centrifuged at 30700 **g** for 20 min at 4 °C to remove the cell debris. The supernatant was dialysed against MilliQ water, and the precipitate was removed by centrifugation (30700 **g** for 10 min at 4 °C). Recombinant crammer was purified by an 1100 Series RP-HPLC system (Agilent Technologies) using a C_{18} semi-preparative column (Nacalai). A linear water/acetonitrile gradient (29–37 %) was used for protein separation, and purified protein fractions were characterized by an Autoflex III MALDI–TOF (matrix-assisted laser-desorption ionization–time-of-flight) mass spectrometer (Bruker Daltonics). Protein concentrations were determined using the Protein Assay reagents (Bio-Rad Laboratories) and bovine serum albumin of known concentrations as the calibration standards.

Characterization of the oligomeric states of crammer

Purified crammer ($\sim$1 mg) was dissolved in buffers of varying pH values [50 mM citric acid-sodium phosphate (pH 3.0–5.0), or 50 mM Tris/HCl (pH 6.0–8.0)] for SEC (size-exclusion chromatography) analysis. The protein samples were loaded on to a 16/60 Superdex 75 pg column connected to an FPLC system (AKTAprime, Amersham Biosciences). To identify the oligomeric state of crammer *in vivo*, crammer was extracted from the heads of wild-type *D. melanogaster,* which were frozen in liquid nitrogen and detached from the fly bodies by vigorously shaking with pre-chilled 25- and 40-mesh sieves. The frozen heads were pulverized and suspended in 50 mM sodium acetate (pH 5.0), containing 1 mM EDTA, 0.1 % Triton X-100 and 0.5 mM sodium azide [15,30]. A polyclonal antibody against crammer (LTK Bio-Laboratories) was used to probe the oligomeric state of crammer.

Table 1 NMR constraints and structural statistics for C72S

RMSD, root mean square deviation.

Distance constraints	
Total NOE	859
Intraresidue	341
Sequential ($\|I–J\| = 1$)	184
Medium range ($\|I–J\| \leqslant 4$)	78
Long range ($\|I–J\| > 4$)	61
Hydrogen bond constraints	61
Total dihedral angle constraints	134
Φ	67
Ψ	67
RMSD (Å) with respect to the average structure	
Well-defined regions (residues 8–49 and 52–74)	
Backbone	0.4 ± 0.1
Heavy atoms	0.8 ± 0.2
All residues (residues 1–80)	
Backbone	1.9 ± 0.4
Heavy atoms	2.2 ± 0.4
Total energy after water refinement (kcal/mol)	-2700 ± 100
RMSD from idealized covalent geometry	
Bonds (Å)	0.0153 ± 0.0005
Angles (°)	2.10 ± 0.07
Impropers (°)	2.6 ± 0.1
Ramachandran statistics	
Most favoured regions	85.2%
Additionally allowed regions	13.6%
Generously allowed regions	1.1%
Disallowed regions	0.0
BioMagResBank accession code	16719
PDB code	2L95

Spectroscopic characterization of crammer

Far-UV CD spectra were acquired using an Aviv CD spectrometer (Model 202, Aviv Biomedical). Wild-type crammer and C72S were buffered in 10 mM citric acid-sodium phosphate (pH 2.0–12.0) and the protein concentrations were set to 30 μM. The far-UV CD spectra were recorded at 20 °C with a wavelength range between 260 and 190 nm, using a 1-mm path length cuvette. The helical contents of the samples were estimated using the CDNN software [31]. Thermal denaturation of crammer was carried out by recording a series of CD spectra from 4 °C to 96 °C, with an increment of 2 °C. Intrinsic fluorescence of C72S (0.02 mg/ml) were recorded as a function of pH using a fluorimeter (F-7000, Hitachi) between 290 and 400 nm ($\lambda_{ex} = 280$ nm). For ANS fluorescence measurements as a function of pH values, the protein concentrations were 30 μM and that of ANS was 20 μM. The wavelength range was set to 385–600 nm ($\lambda_{ex} = 365$ nm).

NMR spectroscopy

Uniformly ^{13}C/^{15}N-labelled crammer (0.3 mM) was buffered in 10 mM citric acid-sodium phosphate containing 10 % (v/v) ^2H_2O. The pH values of the samples were adjusted prior to NMR measurements at 25 °C using AVANCE NMR spectrometers (Bruker Spectrospin) operating at ^{1}H Larmor frequencies of 500 or 600 MHz, and a Varian INOVA NMR spectrometer equipped with a cryogenic probe head and operating at the ^{1}H Larmor frequency of 700 MHz. The spectra were processed using TopSpin (Bruker Spectrospin), NMRPipe [32] and Sparky (T. D. Goddard and D. G. Kneller, University of California, San Francisco). ^{1}H, ^{13}C and ^{15}N backbone and side-chain resonance assignments

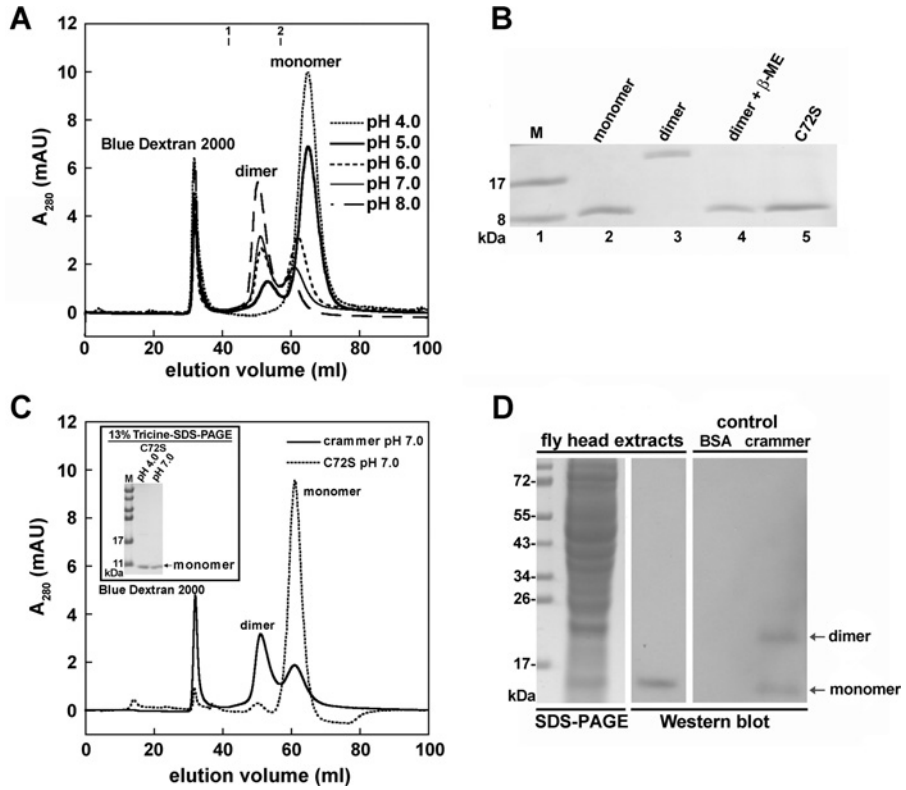

Figure 1 Crammer dimerization as a function of pH

(**A**) SEC is used to monitor the dimerization of crammer as a function of pH. Blue Dextran 2000 as the internal standard is eluted at the void volume. Molecular mass standards: 1, chymotrypsinogen A (25 kDa) and 2, RNase A (13.7 kDa). (**B**) 13 % (w/v) Tricine-SDS/PAGE. Lane 1 contains molecular size markers (M). Lanes labelled monomer and dimer are HPLC-purified proteins, and the proteins contained in last two lanes are as labelled in the Figure. β-ME, 2-mercaptoethanol. (**C**) The results from SEC and Tricine-SDS/PAGE suggest that C72S at neutral pH exists in a monomeric form. (**D**) Crammer extracted from *Drosophila* heads is monomeric in the absence of 2-mercaptoethanol. The control lane labelled 'crammer' contains the recombinant protein, which exists as both monomer and dimer as visualized by the antibody against crammer. In (**B**) and (**D**) the molecular mass is given in kDa on the left-hand side. AU, arbitrary unit.

for spectra of crammer at pH 3.0 and pH 6.0 were obtained from a set of heteronuclear multidimensional NMR spectra [33,34]. Secondary structure propensity was calculated using the SSP program [35] using the assigned $^{13}C\alpha$, $^{13}C\beta$ and $^{13}C'$ chemical shifts. For the hetNOE [heteronuclear NOE (nuclear Overhauser effect)] experiments, ^{1}H presaturation was achieved using a train of weak saturation pulse intervals during the 5-s recycle delay. The hetNOE value is defined as the ratio of the peak intensities recorded with and without ^{1}H saturation, i.e. $\text{hetNOE} = I^{sat}/I^{ref}$, where I^{sat} is the peak intensity with saturation, and I^{ref} is the peak intensity without saturation [36]. To study the binding of crammer to *Drosophila* cathepsin L, the ^{1}H-^{15}N-HSQC (heteronuclear single quantum correlation) spectrum of ^{15}N-labelled crammer and an excess of *Drosophila* cathepsin L in 10 mM citric acid/sodium phosphate buffer (pH 4.0) was recorded.

Solution structure determination

^{15}N- and ^{13}C-edited NOESY-HSQC spectra of crammer were recorded at pH 6.0 to obtain distance constraints through manual assignments of the NOE cross-peaks (Table 1). The chemical shifts of the $^{1}H^{N}$, ^{15}N, $^{1}H\alpha$, $^{13}C\alpha$, $^{13}C\beta$ and $^{13}C'$ resonances were used as the inputs for TALOS+ [37] to obtain backbone torsion angle constraints. CNS 1.1 was used for restrained molecular dynamics simulations to generate the solution structures [38]. The final ensemble consists of 11 lowest energy structures and was evaluated using the PROCHECK-NMR package [39] and Discovery Studio 2.0 (Accelrys). The solvent accessibility was calculated using Discovery Studio 2.0. PyMOL (http://www.pymol.org) was used for molecular visualization.

Inhibition assay

The cathepsin-inhibition assay was carried out according to the procedure described by Comas et al. [5]. SEC was used to purify cathepsin from *Drosophila* heads. Each SEC fraction was subject to protease activity assays using Z-Phe-Arg-AMC as substrate (Calbiochem) [40,41]. The amount of released AMC was measured by its fluorescence emission at 440 nm ($\lambda_{ex} = 380$ nm). The purified *Drosophila* cathepsin was identified using cathepsin B/L specific inhibitors [15,42,43] and the enzymatically active substance in the *Drosophila* head extract was mainly cathepsin L (Supplementary Figure S1 at http://www.BiochemJ.org/bj/442/bj4420563add.htm). Out of 6 g of *Drosophila* heads, we could obtain about 1 mg of active cathepsin L. We also constructed and expressed recombinant *Drosophila* cathepsin B using an *E. coli* expression system (described in detail in the Supplementary online data at http://www.BiochemJ.org/bj/442/bj4420563add.htm). The concentrations of active *Drosophila* cathepsins L and B were determined using E-64 [44]. For the inhibition assays, 70 mg of recombinant crammer was dissolved in 10 mM citric acid/sodium phosphate (pH 4.0) to yield a stock solution of 0.5 mM. Various concentrations of crammer taken from a stock solution were incubated with cathepsin (40 nM) in 100 mM sodium acetate (pH 5.0),

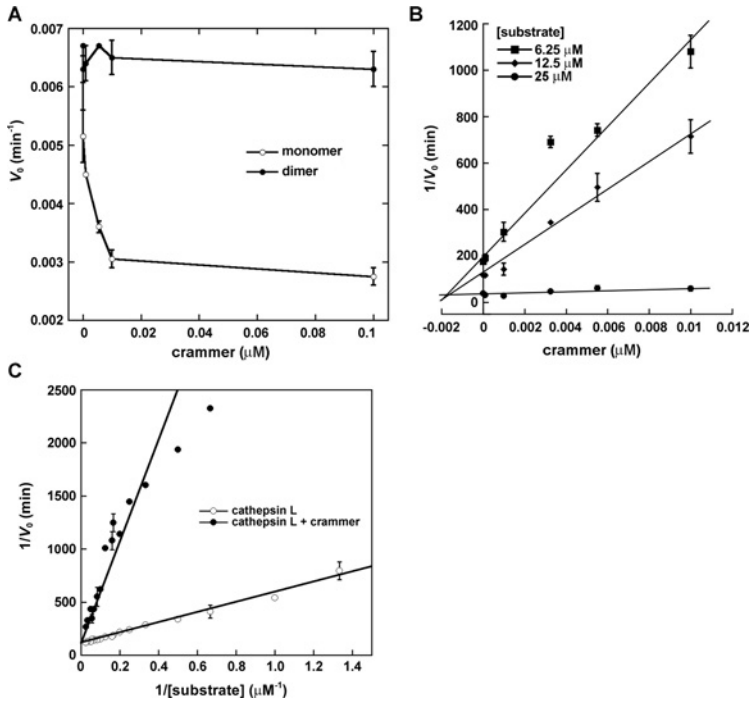

Figure 2 Inhibitory activity of crammer against *Drosophila* cathepsin L

(**A**) Plot of the initial rate (V_0) compared with crammer concentration. Monomeric crammer markedly inhibits *Drosophila* cathepsin L. (**B**) Dixon plot analysis of inhibition of *Drosophila* cathepsin L. This analysis is used to determine the inhibition constant (K_i) of the monomeric crammer. (**C**) Michaelis–Menten kinetics. Lineweaver–Burk analysis is utilized to determine kinetic parameters for *Drosophila* cathepsin L inhibition. The K_m value for cathepsin L increases in the presence of crammer, whereas the V_{max} value is unchanged.

containing 1 mM EDTA and 2 mM DTT. Z-Phe-Arg-AMC (18 μM) was then added to detect residual enzyme activity. Michaelis–Menten kinetics was used to determine the values of K_m and V_{max} of *Drosophila* cathepsin L in the presence or absence of crammer. The concentration of crammer was 10 nM. The non-linear curves were fit using KaleidaGraph (Synergy Software).

RESULTS

Oligomeric state of crammer

Crammer has been reported to be either monomeric or dimeric [15], but the underlying mechanism for the monomer–dimer switch is hitherto unclear. We first examine the oligomeric state of crammer as a function of pH. SEC analysis reveals that crammer is predominately monomeric under acidic conditions, and it is dimeric at neutral and basic pH values (Figure 1A). After incubation with excess 2-mercaptoethanol, the dimeric form of wild-type crammer is dissociated into monomers as shown by Tricine-SDS/PAGE (Figure 1B). To confirm that the dimerization of crammer is a result of disulfide bond formation, we replace the only cysteine residue at position 72 with a serine, C72S, and indeed the dimeric form of crammer is no longer present in the SDS/PAGE gel (Figures 1B and 1C). An Ellman assay [45,46] shows that monomeric crammer has high free thiol content, but dimeric crammer does not (results not shown). Taken together, we conclude that dimeric crammer resulted from the formation of an intermolecular disulfide bond between crammer monomers. To investigate the oligomeric state of crammer *in vivo*, endogenous crammer was extracted from wild-type *Drosophila* heads and probed by Western blotting in the absence of 2-mercaptoethanol, and the results indicate that endogeneous crammer is monomeric, not dimeric, *in vivo* (Figure 1D).

Inhibitory assay

Deshapriya et al. [15] reported that both monomeric and dimeric crammer exhibit protease-inhibitory activity. To allow comparison of their results with those of the present study, recombinant crammer was incubated with *Drosophila* cathepsin L purified from fly head extracts to examine the enzyme activity (Figure 2A) following the protocol reported by Comas et al. [5]. The ZFR-AMC enzymatic assay indicates that only monomeric, not dimeric, crammer exhibits inhibitory activity against *Drosophila* cathepsin L. The inhibition of cathepsin L, monitored by fluorescence, gives a linear plot as a function of time (Supplementary Figure S2 at http://www.BiochemJ.org/bj/442/bj4420563add.htm) and a Dixon plot is used to determine the enzyme-inhibitor inhibition constant (K_i) [5,9,15]. The K_i value for monomeric crammer is 1.36 ± 0.67 nM (Figure 2B), which is comparable with those that have been reported for other strong cysteine protease inhibitors [9,13,15]. Additionally, recombinant wild-type crammer [5] and C72S in their monomeric forms also exhibit inhibitory activities against human cathepsin L (results not shown). Michaelis–Menten analysis reveals that the addition of crammer increases the Michaelis constant (K_m) of *Drosophila* cathepsin L from 3.85 μM to 20.06 μM without affecting the V_{max} value (0.006 min^{-1}) (Figure 2C), which is a hallmark of competitive inhibition [47]. In summary, monomeric crammer is a strong competitive inhibitor that blocks the cathepsin substrate-binding site.

Structural characterization

We first employed far-UV CD spectroscopy to examine the structure of crammer. The far-UV CD spectrum of crammer at pH 7.0 exhibits a strong signal at 222 nm, corresponding to an α-helical content of 63 ± 4%; crammer is less helical

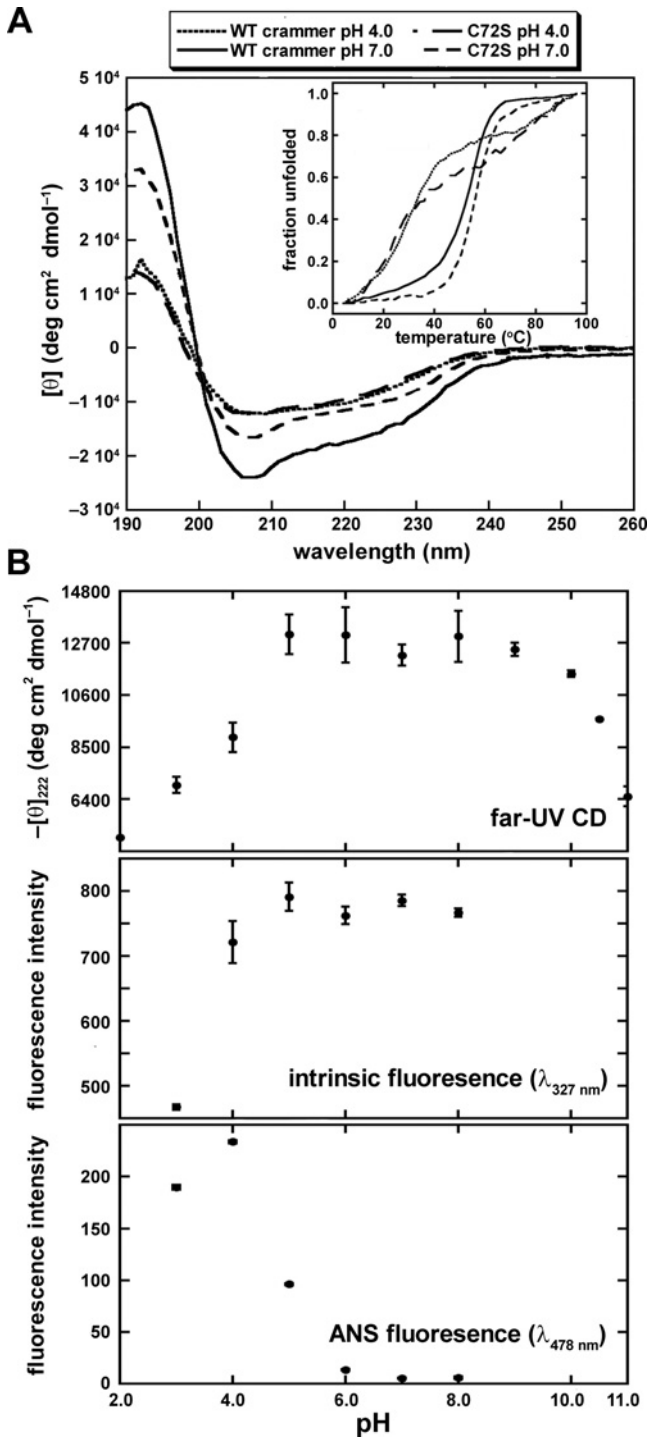

Figure 3 Spectral properties of crammer

(**A**) Far-UV CD of crammer at various solution pH values and temperature. WT, wild-type. Inset: thermal denaturation of crammer and C72S monitored by CD at 208 nm. (**B**) The spectral properties of C72S as a function of pH monitored by far-UV CD and intrinsic fluorescence spectroscopy and ANS binding.

at pH 4.0 (34 ± 4 %; Figure 3A). C72S appears to be slightly less helical than wild-type crammer at pH 7.0 (54 ± 3 %), but its α-helical content is similar to that of crammer at pH 4.0 (33 ± 1 %). When subject to thermal unfolding, crammer unfolds more cooperatively at pH 7.0 than it does at pH 4.0 (Figure 3A, inset) [48]. Interestingly, the secondary structure of C72S is less sensitive to pH (Figure 3A). For both wild-type and C72S, the ANS fluorescence negatively correlates with the CD signal at 222 nm as a function of pH, (Supplementary Figure S3 at http://www.BiochemJ.org/bj/442/bj4420563add.htm), indicating that both proteins partially unfold under acidic conditions.

We have employed far-UV CD, intrinsic tryptophan fluorescence and ANS-binding fluorescence spectroscopy to systematically monitor the folding of C72S as a function of pH (Figure 3B). The far-UV CD signal at 222 nm follows a bell-shaped distribution with a maximum plateau between pH 5 and 8 flanked by significant decreases at lower and higher pH values, suggesting that α-helix formation is associated with salt-bridge formation [49]. At neutral pH, most ionizable residues should be ionized, rendering salt-bridge formation more favourable. At high acidic or basic pH values, salt bridges are less likely to form due to overall protonation and deprotonation respectively [49]. It is noteworthy that that we observe blue shift of the intrinsic fluorescence of C72S on increasing pH values (results not shown), indicating that the local environments of the aromatic groups become more hydrophobic as a result of well-defined hydrophobic core formation at neutral pH values [7,50]. This finding is corroborated by the ANS data, which probe the extent to which hydrophobic regions are exposed (Figure 3B) [7,50]. At acidic pH values, ANS binds strongly to C72S, with fluorescent intensity peaking at pH 4.0. In the same assay, the fluorescence signal is lost at neutral pH, indicating the formation of a well-defined tertiary structure. Collectively, we conclude that crammer retains significant amount of helical structures under acidic conditions, but lacks a well-defined hydrophobic core as shown by the loss of cooperativity under thermal unfolding. This is good evidence to suggest that crammer is in a molten globular state under acidic conditions [51,52].

Protein folding

In order to obtain structural insight into the structural transition of crammer at a residue-specific level, high-resolution NMR spectroscopy was employed to investigate the solution structure of crammer. We have carried out a pH titration for C72S by recording a series of ^{1}H-^{15}N HSQC spectra at different pH values (Figure 4A and Supplementary Figure S4 at http://www.BiochemJ.org/bj/442/bj4420563add.htm). C72S exhibits poorly dispersed cross-peaks at pH 3.0 and 4.0, under which conditions C72S is predominantly molten globular. The observed resonances substantially broadened at pH 5.0. At pH values of 6.0 or higher, a new set of well-dispersed cross-peaks emerge, indicating the formation of a well-defined tertiary structure upon increase in pH, which is consistent with the global analysis by far-UV CD, intrinsic tryptophan fluorescence and ANS binding (Figure 3).

We have assigned the backbone resonances of C72S (H^N, Cα, Cβ, C' and N) at pH 3.0 to examine the solution structure of crammer in detail [BMRB (BioMagResBank) accession number 17367]. Although the far-UV CD and intrinsic fluorescence spectra as well as the chemical shift dispersions in the ^{1}H-^{15}N HSQC spectrum all indicate that C72S is largely unfolded at pH 4.0 and below, the secondary chemical shifts of the Cα, Cβ, and C' nuclei at pH 3.0 reflect a significant proportion of the molecule, particularly those in the N-terminal region, which exhibit substantial residual helical content (Figure 4B). Interestingly, α2 shifts significantly from residues 22–48 (pH 6.0) to 18–42 (pH 3.0), thus resulting in the connection of α2 with α1 to form a long stretch of α-helix at pH 3.0. We have also measured the ^{1}H-^{15}N hetNOEs of C72S at pH 6.0 and pH 3.0 to probe fast backbone dynamics on the femtosecond to picosecond timescale (Figure 4B) [36]. The overall hetNOE

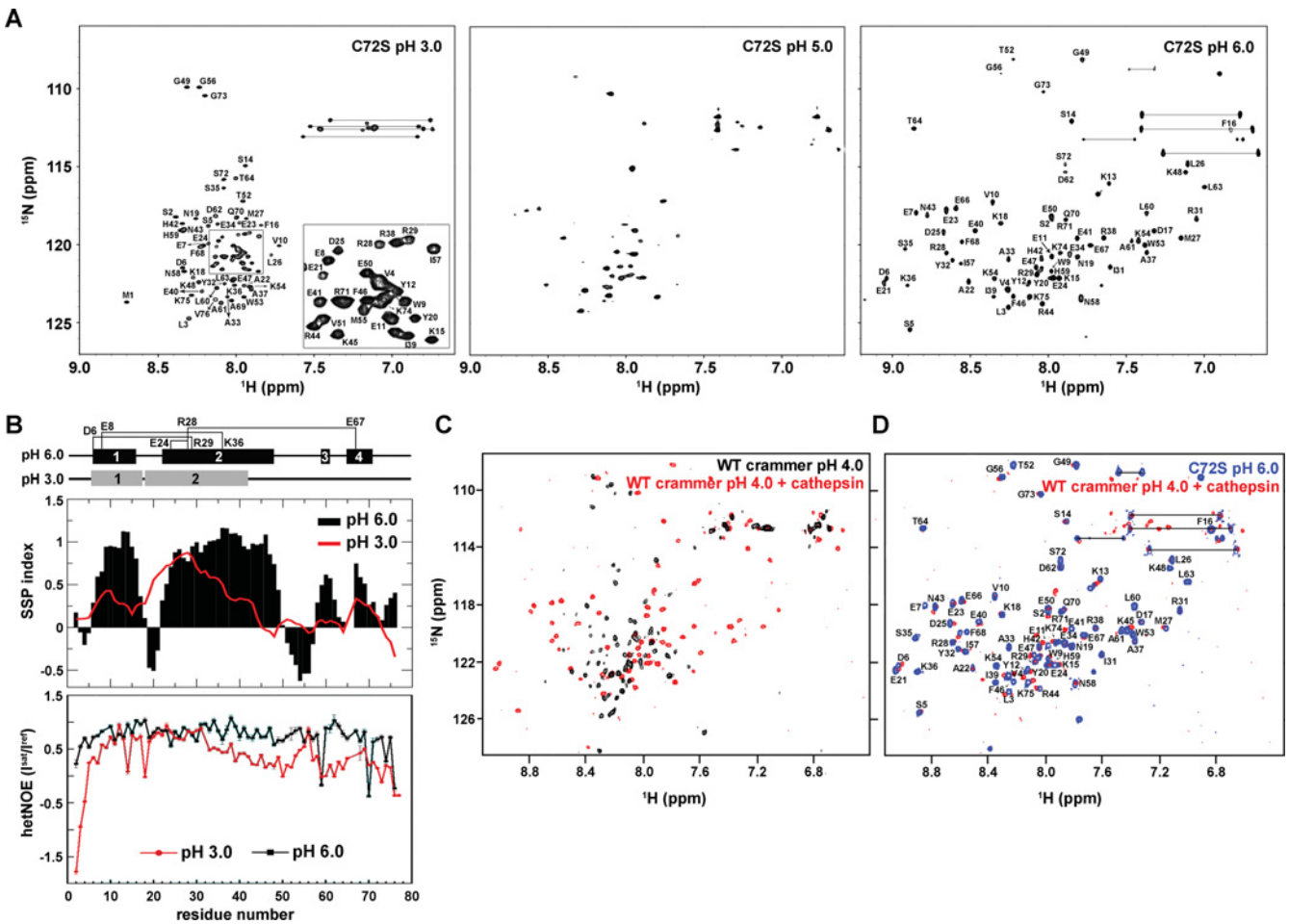

Figure 4 ^{1}H-^{15}N-HSQC Spectra

(**A**) ^{1}H-^{15}N-HSQC spectra of C72S as a function of pH. The assigned cross-peaks for C72S at pH 3.0 and pH 6.0 are labelled in the Figure. The inset shows an expanded view of part of the spectrum. (**B**) Secondary structure propensity and hetNOEs of C72S residues. Upper panel: secondary structure propensity (SSP) compared with C72S residue number. Red line, pH 3.0; black bars, pH 6.0. Positive values indicate α-helical propensity, and negative values indicate β-strand propensity. Lower panel: hetNOE values used to investigate the femtosecond-to-picosecond dynamics of C72S. Residues in folded regions have larger hetNOE values. The α-helical regions and the salt bridges are depicted schematically at the top of the Figure. (**C**) ^{1}H-^{15}N-HSQC spectra of wild-type (WT) crammer (pH 4.0) in the absence (black) or presence of *Drosophila* cathepsin L (red). (**D**) The spectra of the *Drosophila* cathepsin L/crammer complex (red) and C72S (blue). The C72S cross-peaks are labelled.

values are higher at pH 6.0 than those at pH 3.0, with average values of 0.72±0.05 and 0.34±0.02 respectively, indicating that the backbone conformation C72S is much more flexible, i.e. disordered, at pH 3.0, whereas at pH 6.0 the overall structure of C72S is compact and well defined. Despite the loss of compact hydrophobic core, the N-terminal region of C72S, especially α1 and the first half of α2, exhibits relatively higher hetNOE values (>0.6) that are comparable with those observed at pH 6.0. Taken together, the NMR data indicate that C72S retains a substantial amount of helical structure at acidic pH despite an ill-defined hydrophobic core (Figure 3B), further supporting our hypothesis that C72S is in a molten globular state under acidic conditions [53].

Intriguingly, although crammer is largely disordered at pH 4.0, the addition of cathepsin L purified from *Drosophila* head extract induces a molten globule-to-ordered structure transition as evidenced by the emergence of well-dispersed ^{1}H-^{15}N HSQC resonances (Figures 4C and 4D). The chemical shifts of these cross-peaks are very similar to those of C72S at pH 6.0 (Figure 4D), indicating that the solution structure of monomeric crammer C72S at pH 6.0 resembles that of cathepsin L-bound (activated) crammer at pH 4.0. We therefore use C72S at pH 6.0 as a mimic of cathepsin L-bounded crammer for detailed structural elucidation in the following section (Table 1).

Solution structure of C72S

We have obtained near complete backbone assignments of C72S except for those of Glu^{8}, Pro^{65} and Val^{76}–Asn^{79} (>91 % of all the expected H^{N}, Cα, Cβ, C' and N resonances). These unassigned residues all lay in the loop regions that are inherently flexible. The corresponding resonances are broadened beyond detection, which we attribute to the result of solvent exchange. The NMR chemical shift assignments of C72S have been deposited in the BMRB under accession number 16719. Using ^{13}C/^{15}N-edited NOESY spectra and backbone torsion angle restraints, we have determined the solution structure of C72S (Figure 5A and Table 1) which has been deposited in the PDB (PDB code 2L95).

The solution structure of C72S contains four α-helices [Asp^{6}–Phe^{16} (α1), Ala^{22}–Lys^{48} (α2), His^{59}–Ala^{61} (α3) and Pro^{65}–Arg^{71} (α4)] with α2 serving as the core structure that appears to maintain the overall tertiary structure (Figure 5A). The three-dimensional

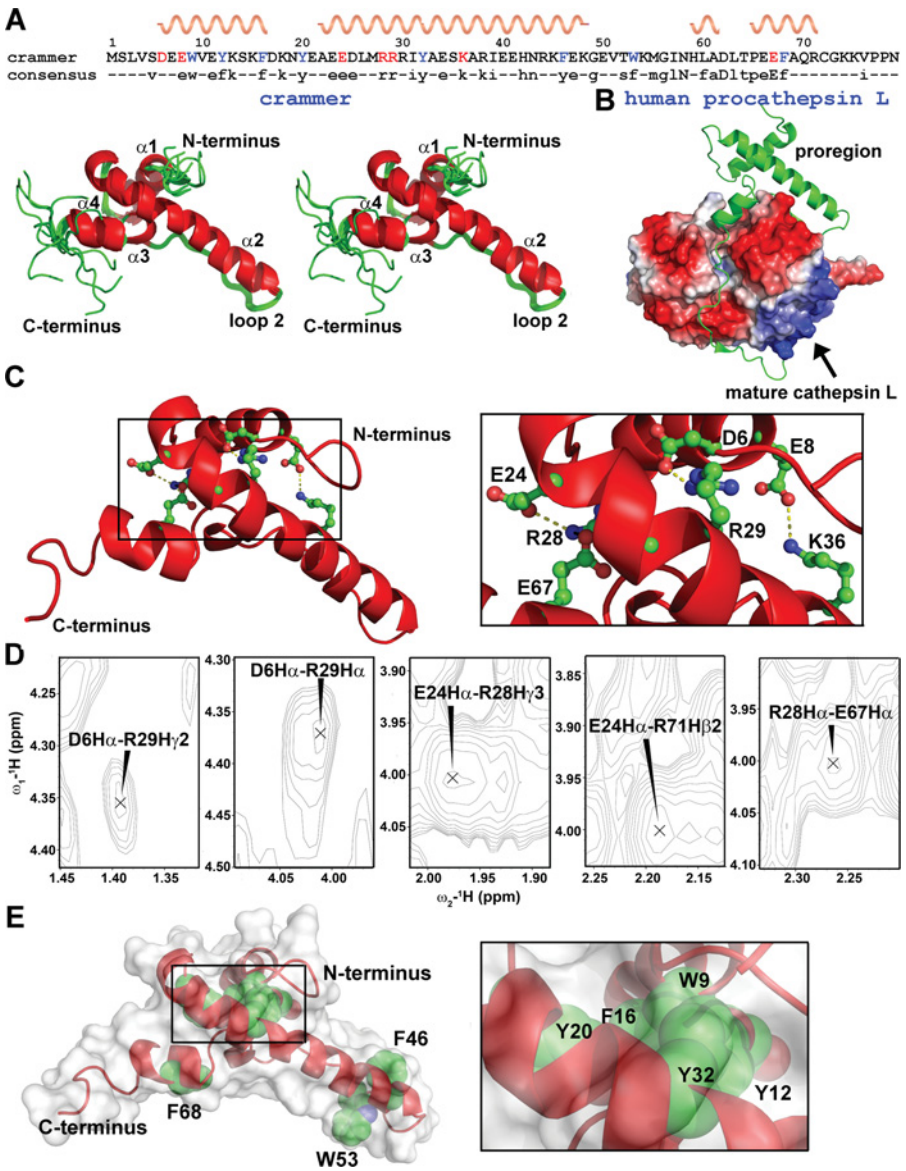

Figure 5 Solution structure of C72S

(**A**) Stereo view of eleven lowest-energy structures that are superpositioned on their backbone atoms (N, Cα and C'). The helices α1, α2, α3 and α4 are labelled. The consensus sequence is shown at the top of the Figure, with the helical regions indicated. (**B**) Structure of human procathepsin L (PDB code 1CS8). The proregion is shown in ribbon representation. (**C**) The crammer residues involved in salt bridges are shown as ball-and-stick representation in the inset. (**D**) NOE cross-peaks for the charged side-chains that are involved in salt-bridge formation. (**E**) The aromatic residues in C72S are shown in spheres to indicate the hydrophobic core packing. The protein surface is shown in white.

fold of crammer is very similar to that of the proregion of procathepsin L (Figure 5B) [54]. Structural alignment of the two gives a moderate pairwise positional root mean squared deviation of 4.09 Å (1 Å = 0.1 nm) for the Cα atoms within the secondary structure elements. The deviation is largely due to the result of different relative orientations of the individual α-helices (Figure 5A).

There are two salt bridges that connect α1 and α2 (Asp^{6}–Arg^{29} and Glu^{8}–Lys^{36}), and two that connect α2 and α4 (Glu^{24}–Arg^{28} and Arg^{28}–Glu^{67} (Figure 5C). Their presence is confirmed by long-range NOEs between their respective side-chains (Figure 5D). The ERFNIN motif is located in α2 and is surrounded by α1, α3, α4 and loop 2, which is a relatively long loop. As ERFNIN is located in the central region of the core structure, it may be an important folding element. Also important for folding are six well-conserved aromatic residues (Trp^{9}, Tyr^{12}, Phe^{16}, Tyr^{20}, Tyr^{32} and Phe^{68}) that form a small hydrophobic core for structure stabilization (Figure 5E).

To examine the importance of these aromatic residues in enzyme inhibition, we first constructed recombinant *Drosophila* cathepsins B and L (see the Supplementary online data). Recombinant cathepsin L could not be expressed despite the use of several expression vectors and *E. coli* strains (results not shown). We therefore focused on the enzymatic kinetics of cathepsin B inhibition by different concentrations of crammer (Figure 6A). On the basis of non-linear regression analysis, the k_{obs} values exhibit a linear relationship with respect to crammer concentrations (Figure 6B), implying a single-step, reversible and slow-binding inhibition mechanism [15,55]. Individual alanine replacements of these aromatic residues decrease the inhibitory

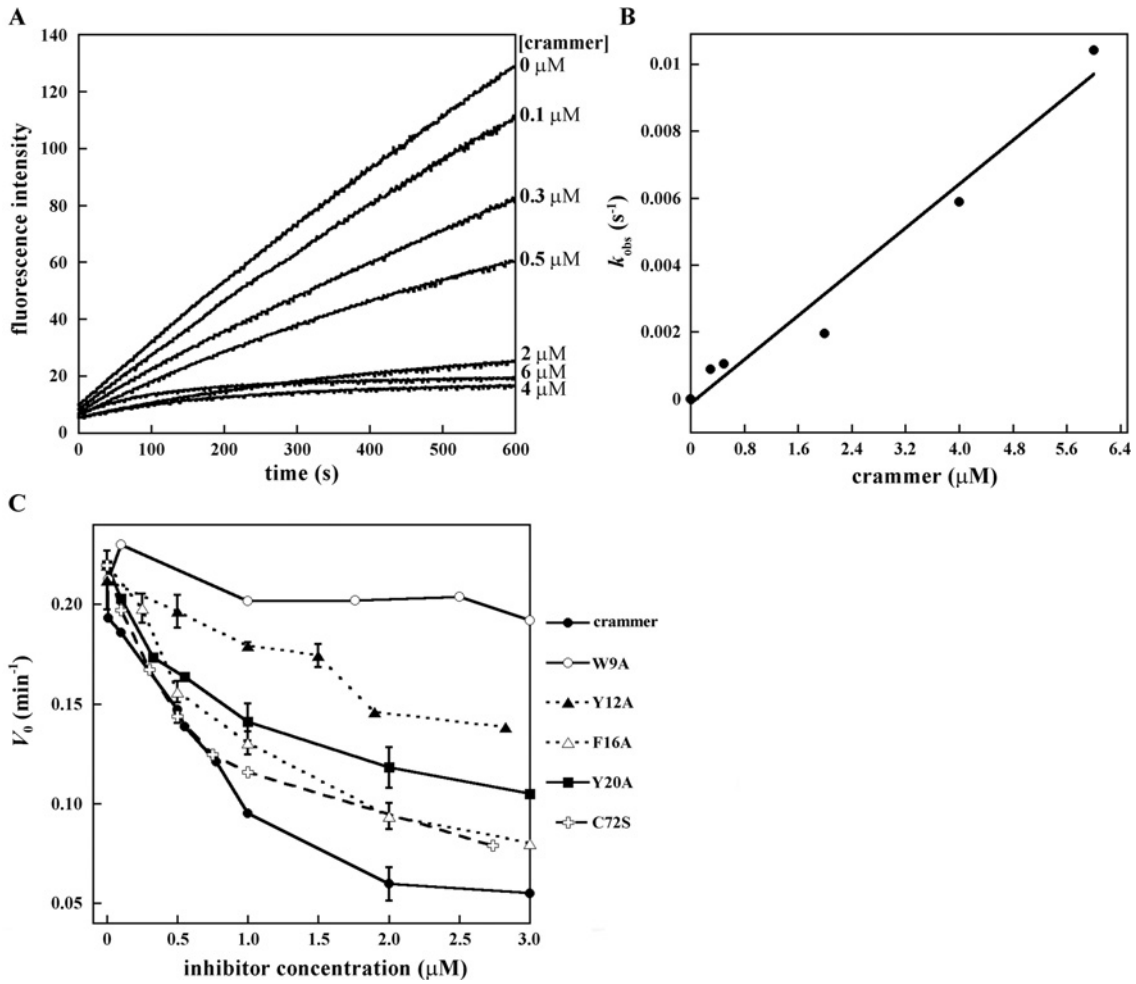

Figure 6 Inhibitory activities of crammer and its mutants against *Drosophila* cathepsin B

(**A**) Progress curves for the cathepsin B inhibition by different concentrations of crammer are non-linear as a function of time, implying a typical slow-binding inhibition. The observed fluorescence signals are fitted using the equation, $[P] = V_s \times t + (V_0 - V_s) \times [1 - \exp(-k_{obs} \times t)]/k_{obs}$ [15,55], where P is the amount of product formed at time t, V_0 and V_s are the initial and steady-state velocities respectively, and k_{obs} is the rate constant for inhibition. (**B**) Plot of k_{obs} compared with crammer concentrations for the inhibition of cathepsin B. On the basis of a regression analysis [15,55], the K_i values can be obtained. (**C**) Plots of the initial rates of fluorescence increase as a function of inhibitor (crammer variants) concentrations. The conserved aromatic residues are replaced with alanine to clarify their potential roles in cathepsin B inhibition.

Table 2 Cathepsin B inhibitor inhibition constants

The K_i values are determined by the V_0 and V_s values. N.D., not determined.

Enzyme	Crammer	W9A	Y12A	F16A	Y20A	C72S
Cathepsin B K_i value	0.79 ± 0.35 μM	N.D.	18.94 ± 2.24 μM	2.26 ± 0.23 μM	5.40 ± 1.25 μM	1.87 ± 0.63 μM

ability (Figure 6C and Table 2). Wild-type crammer displays a maximal *Drosophila* cathepsin B inhibition with a K_i value of 0.79 ± 0.35 μM. In contrast, W9A, Y12A, and Y20A exhibit weak or medium inhibition activities. The removal of these aromatic side chains are expected to perturb the hydrophobic core, and the loss of structural integrity in turn diminishes the inhibitory activities against cathepsin B. This is particularly pronounced for W9A (Figures 5 and 6), lending support to the importance of these aromatic residues in protein folding. Finally, Ser^{72}, located in the flexible C-terminal region, is more than 95 % solvent accessible. Cys^{72} in crammer is also probably highly exposed, which would allow for intermolecular disulfide bond formation.

DISCUSSION

Using Western blotting, we have established that endogenous crammer (isolated from fruit fly head) is monomeric *in vivo* (Figure 1D). Crammer contains a lysosomal-localization motif (YKX_4KXY) that probably targets crammer into the lysosome, a highly acidic (<pH 5.0) cellular compartment where crammer is expected to be predominantly monomeric (Figure 1A). Although Deshapriya et al. [15] reported that both monomeric and dimeric forms of crammer can inactivate cathepsin, our inhibitory assay has shown that only monomeric, not dimeric, crammer is a strong natural inhibitor (nanomolar range) of cathepsin L. Under

acidic conditions, monomeric crammer contains a substantial amount of helical conformation, but lacks a compact hydrophobic core (Figures 3 and 4A), indicating that crammer is a molten globule under acidic conditions. When bound to cathepsin L, however, crammer adopts an ordered structure that gives rise to well-dispersed NMR resonances which highly resemble those of C72S at pH 6.0 (Figures 4C and 4D). Supported by the spectral similarity, we therefore use C72S at pH 6.0 to mimic the cathepsin L-bound form of wild-type crammer at physiological, acidic conditions, for detailed structural elucidations.

Despite the reports of several cysteine protease structures [56,57], there is hitherto no structural information of protease proregions in isolation and that of propeptide-like protease inhibitors. The solution structure of C72S is the first high-resolution structure for a propeptide-like protease inhibitor. Indeed, the structure of C72S resembles that of the human cathepsin L proregion (PDB code 1CS8; Figure 5B) [54]. The proregion C-terminal residues of procathepsin L blocks the cathepsin active site [54], as do those of crammer. The two structures also share the same topology; however, the sequence composition and chain length between the two proteins are very different, as reflected by structural differences for α3 and the N- and C-termini. C72S has an additional helical segment (α3), encompassing residues 59–61, that is absent in the proregion of human cathepsin L. In particular, the C-terminal tail of the proregion is longer and is free of cysteine.

Human cathepsins B and L have been associated with certain neurodegenerative diseases, e.g. Parkinson's disease [58]. In the fruit fly, the overexpression of crammer impairs LTM [5,6], providing evidence of causal relationship between cysteine protease activities and neurophysiology in humans and flies. In general, activation of human cathepsin requires cleavage of its proregion at acidic pH [59]. When cleaved, the proregion exists as a molten globule [60], as we have shown here for crammer under acidic conditions. Nonetheless, the cleaved proregion does not inactivate cathepsin [61]. Although much experimental work is required to further establish the activation mechanism of *Drosophila* cathepsin, our data and that of others have suggested that human and *Drosophila* cathepsins may be activated by the same mechanism.

Our bioinformatic analysis suggests that crammer is localized in the lysosome. Although both crammer and the cleaved proregion of human cathepsin share molten globule-like structural features under acidic conditions *in vitro*, crammer is a strong inhibitor of cathepsin under acidic conditions, whereas the proregion is not [61]. Importantly, the tight binding of crammer to cathepsin is associated with a molten globule-to-ordered structure transition. It remains to be seen if such a structural transition occurs when other propeptide-like protease inhibitors bind to their targets. This structural transition is also regulated by pH. Once crammer is in other organelles with higher pH values, crammer probably presents as an inactive dimer against cathepsin. Hence, the switch from monomeric to dimeric crammer is probably an alternative mechanism for the cathepsin regulation. This pH-mediated regulation is also observed in other proteins [62]. However, the detailed molecular mechanism of this regulation in crammer requires further investigation.

AUTHOR CONTRIBUTION

Tien-Sheng Tseng, Chao-Sheng Cheng, Shang-Te Danny Hsu and Ping-Chiang Lyu designed research; Tien-Sheng Tseng, Chao-Sheng Cheng. Dian-Jiun Chen, Min-Fang Shih, Yu-Nan Liu and Shang-Te Danny Hsu performed research; Tien-Sheng Tseng, Chao-Sheng Cheng, Shang-Te Danny Hsu and Ping-Chiang Lyu analysed the data; and Tien-Sheng Tseng, Chao-Sheng Cheng, Shang-Te Danny Hsu and Ping-Chiang Lyu wrote the paper.

FUNDING

This work was supported by the National Science Council, Taiwan [grant numbers NSC 99-2811-M-007-073, 99-3112-B-007-003, 99-2627-B-007-010 and 100-2113-M-001-031-MY2].

REFERENCES

1 Tully, T. (1987) *Drosophila* learning and memory revisited. Trends Neurosci. **10**, 330–334
2 Waddell, S. and Quinn, W. G. (2001) Flies, genes, and learning. Annu. Rev. Neurosci. **24**, 1283–1309
3 Davis, R. L. (2005) Olfactory memory formation in *Drosophila*: from molecular to systems neuroscience. Annu. Rev. Neurosci. **28**, 275–302
4 Keene, A. C. and Waddell, S. (2007) *Drosophila* olfactory memory: single genes to complex neural circuits. Nat. Rev. Neurosci. **8**, 341–354
5 Comas, D., Petit, F. and Preat, T. (2004) *Drosophila* long-term memory formation involves regulation of cathepsin activity. Nature **430**, 460–463
6 Krashes, M. J. and Waddell, S. (2008) Rapid consolidation to a radish and protein synthesis-dependent long-term memory after single-session appetitive olfactory conditioning in *Drosophila*. J. Neurosci. **28**, 3103–3113
7 Jerala, R., Zerovnik, E., Kidric, J. and Turk, V. (1998) pH-induced conformational transitions of the propeptide of human cathepsin L. A role for a molten globule state in zymogen activation. J. Biol. Chem. **273**, 11498–11504
8 Denizot, F., Brunet, J. F., Roustan, P., Harper, K., Suzan, M., Luciani, M. F., Mattei, M. G. and Golstein, P. (1989) Novel structures CTLA-2α and CTLA-2β expressed in mouse activated T cells and mast cells and homologous to cysteine proteinase proregions. Eur. J. Immunol. **19**, 631–635
9 Delaria, K., Fiorentino, L., Wallace, L., Tamburini, P., Brownell, E. and Muller, D. (1994) Inhibition of cathepsin L-like cysteine proteases by cytotoxic T-lymphocyte antigen-2β. J. Biol. Chem. **269**, 25172–25177
10 Kurata, M., Hirata, M., Watabe, S., Miyake, M., Takahashi, S. Y. and Yamamoto, Y. (2003) Expression, purification, and inhibitory activities of mouse cytotoxic T-lymphocyte antigen-2α. Protein Expr. Purif. **32**, 119–125
11 Yamamoto, Y., Watabe, S., Kageyama, T. and Takahashi, S. Y. (1999) A novel inhibitor protein for *Bombyx* cysteine proteinase is homologous to propeptide regions of cysteine proteinases. FEBS Lett. **448**, 257–260
12 Yamamoto, Y., Watabe, S., Kageyama, T. and Takahashi, S. Y. (1999) Purification and characterization of *Bombyx* cysteine proteinase specific inhibitors from the hemolymph of *Bombyx mori*. Arch. Insect Biochem. Physiol. **42**, 119–129
13 Kurata, M., Yamamoto, Y., Watabe, S., Makino, Y., Ogawa, K. and Takahashi, S. Y. (2001) *Bombyx* cysteine proteinase inhibitor (BCPI) homologous to propeptide regions of cysteine proteinases is a strong, selective inhibitor of cathepsin L-like cysteine proteinases. J. Biochem. **130**, 857–863
14 Giot, L., Bader, J. S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y. L., Ooi, C. E., Godwin, B., Vitols, E. et al. (2003) A protein interaction map of *Drosophila melanogaster*. Science **302**, 1727–1736
15 Deshapriya, R. M., Takeuchi, A., Shirao, K., Isa, K., Watabe, S., Murakami, R., Tsujimura, H. and Yamamoto, Y. (2007) *Drosophila* CTLA-2-like protein (D/CTLA-2) inhibits cysteine proteinase 1 (CP1), a cathepsin L-like enzyme. Zoolog. Sci. **24**, 21–30
16 Hanewinkel, H., Glossl, J. and Kresse, H. (1987) Biosynthesis of cathepsin B in cultured normal and I-cell fibroblasts. J. Biol. Chem. **262**, 12351–12355
17 Tao, K., Stearns, N. A., Dong, J., Wu, Q. L. and Sahagian, G. G. (1994) The proregion of cathepsin L is required for proper folding, stability, and ER exit. Arch. Biochem. Biophys. **311**, 19–27
18 Groves, M. R., Coulombe, R., Jenkins, J. and Cygler, M. (1998) Structural basis for specificity of papain-like cysteine protease proregions toward their cognate enzymes. Proteins **32**, 504–514
19 Nishimura, Y., Kawabata, T. and Kato, K. (1988) Identification of latent procathepsins B and L in microsomal lumen: characterization of enzymatic activation and proteolytic processing *in vitro*. Arch. Biochem. Biophys. **261**, 64–71
20 Rowan, A. D., Mason, P., Mach, L. and Mort, J. S. (1992) Rat procathepsin B. Proteolytic processing to the mature form *in vitro*. J. Biol. Chem. **267**, 15993–15999
21 Khan, A. R. and James, M. N. (1998) Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. Protein Sci. **7**, 815–836
22 Vernet, T., Berti, P. J., de Montigny, C., Musil, R., Tessier, D. C., Menard, R., Magny, M. C., Storer, A. C. and Thomas, D. Y. (1995) Processing of the papain precursor. The ionization state of a conserved amino acid motif within the Pro region participates in the regulation of intramolecular processing. J. Biol. Chem. **270**, 10838–10846
23 McIntyre, G. F., Godbold, G. D. and Erickson, A. H. (1994) The pH-dependent membrane association of procathepsin L is mediated by a 9-residue sequence within the propeptide. J. Biol. Chem. **269**, 567–572
24 Wiederanders, B., Kaulmann, G. and Schilling, K. (2003) Functions of propeptide parts in cysteine proteases. Curr. Protein Pept. Sci. **4**, 309–326

25 Qiao, L., Hamamichi, S., Caldwell, K. A., Caldwell, G. A., Yacoubian, T. A., Wilson, S., Xie, Z. L., Speake, L. D., Parks, R., Crabtree, D. et al. (2008) Lysosomal enzyme cathepsin D protects against α-synuclein aggregation and toxicity. Mol. Brain **1**, 17

26 Hook, V. Y. (2006) Unique neuronal functions of cathepsin L and cathepsin B in secretory vesicles: biosynthesis of peptides in neurotransmission and neurodegenerative disease. Biol. Chem. **387**, 1429–1439

27 Sevenich, L., Pennacchio, L. A., Peters, C. and Reinheckel, T. (2006) Human cathepsin L rescues the neurodegeneration and lethality in cathepsin B/L double-deficient mice. Biol. Chem. **387**, 885–891

28 Xiong, A. S., Yao, Q. H., Peng, R. H., Li, X., Fan, H. Q., Cheng, Z. M. and Li, Y. (2004) A simple, rapid, high-fidelity and cost-effective PCR-based two-step DNA synthesis method for long gene sequences. Nucleic Acids Res. **32**, e98

29 Liu, Y. N., Lai, Y. T., Chou, W. I., Chang, M. D. and Lyu, P. C. (2007) Solution structure of family 21 carbohydrate-binding module from *Rhizopus oryzae* glucoamylase. Biochem. J. **403**, 21–30

30 Qian, F., Bajkowski, A. S., Steiner, D. F., Chan, S. J. and Frankfater, A. (1989) Expression of five cathepsins in murine melanomas of varying metastatic potential and normal tissues. Cancer Res. **49**, 4870–4875

31 Bohm, G., Muhr, R. and Jaenicke, R. (1992) Quantitative analysis of protein far UV circular dichroism spectra by neural networks. Protein Eng. **5**, 191–195

32 Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR **6**, 277–293

33 Sattler, M., Schleucher, J. and Griesinger, C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Prog. Nucl. Magn. Reson. Spectrosc **34**, 93–158

34 Hsu, S. T., Cabrita, L. D., Christodoulou, J. and Dobson, C. M. (2009) ^{1}H, ^{15}N and ^{13}C assignments of domain 5 of *Dictyostelium discoideum* gelation factor (ABP-120) in its native and 8M urea-denatured states. Biomol. NMR Assign. **3**, 29–31

35 Marsh, J. A., Singh, V. K., Jia, Z. and Forman-Kay, J. D. (2006) Sensitivity of secondary structure propensities to sequence differences between α- and γ-synuclein: implications for fibrillation. Protein Sci. **15**, 2795–2804

36 Kay, L. E., Torchia, D. A. and Bax, A. (1989) Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. Biochemistry **28**, 8972–8979

37 Shen, Y., Delaglio, F., Cornilescu, G. and Bax, A. (2009) TALOS + : a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J. Biomol. NMR **44**, 213–223

38 Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S. et al. (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. Sect. D Biol. Crystallogr. **54**, 905–921

39 Laskowski, R. A., Macarthur, M. W., Moss, D. S. and Thornton, J. M. (1993) Procheck: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. **26**, 283–291

40 Barrett, A. J. (1980) Fluorimetric assays for cathepsin B and cathepsin H with methylcoumarylamide substrates. Biochem. J. **187**, 909–912

41 Barrett, A. J. and Kirschke, H. (1981) Cathepsin B, Cathepsin H, and cathepsin L. Methods Enzymol. **80**, 535–561

42 Shah, P. P., Myers, M. C., Beavers, M. P., Purvis, J. E., Jing, H., Grieser, H. J., Sharlow, E. R., Napper, A. D., Huryn, D. M., Cooperman, B. S. et al. (2008) Kinetic characterization and molecular docking of a novel, potent, and selective slow-binding inhibitor of human cathepsin L. Mol. Pharmacol. **74**, 34–41

43 Myers, M. C., Shah, P. P., Diamond, S. L., Huryn, D. M. and Smith, III, A. B. (2008) Identification and synthesis of a unique thiocarbazate cathepsin L inhibitor. Bioorg. Med. Chem. Lett. **18**, 210–214

44 Barrett, A. J., Kembhavi, A. A., Brown, M. A., Kirschke, H., Knight, C. G., Tamai, M. and Hanada, K. (1982) L-*trans*-epoxysuccinyl-leucylamido(4-guanidino)butane (E-64) and its analogues as inhibitors of cysteine proteinases including cathepsins B, H and L. Biochem. J. **201**, 189–198

45 Ellman, G. L. (1959) Tissue sulfhydryl groups. Arch. Biochem. Biophys. **82**, 70–77

46 Riddles, P. W., Blakeley, R. L. and Zerner, B. (1983) Reassessment of Ellman's reagent. Methods Enzymol. **91**, 49–60

47 Dixon, M. (1953) The determination of enzyme inhibitor constants. Biochem. J. **55**, 170–171

48 Schulman, B. A., Kim, P. S., Dobson, C. M. and Redfield, C. (1997) A residue-specific NMR view of the non-cooperative unfolding of a molten globule. Nat. Struct. Biol. **4**, 630–634

49 Lyu, P. C., Gans, P. J. and Kallenbach, N. R. (1992) Energetic contribution of solvent-exposed ion pairs to α-helix structure. J. Mol. Biol. **223**, 343–350

50 Chevigne, A., Dumez, M. E., Dumoulin, M., Matagne, A., Jacquet, A. and Galleni, M. (2010) Comparative study of mature and zymogen mite cysteine protease stability and pH unfolding. Biochim. Biophys. Acta **1800**, 937–945

51 Christensen, H. and Pain, R. H. (1991) Molten globule intermediates and protein folding. Eur. Biophys. J. **19**, 221–229

52 Dhulesia, A., Cremades, N., Kumita, J. R., Hsu, S. T. D., Mossuto, M. F., Dumoulin, M., Nietlispach, D., Akke, M., Salvatella, X. and Dobson, C. M. (2010) Local cooperativity in an amyloidogenic state of human lysozyme observed at atomic resolution. J. Am. Chem. Soc. **132**, 15580–15588

53 Baker, D., Sohl, J. L. and Agard, D. A. (1992) A protein-folding reaction under kinetic control. Nature **356**, 263–265

54 Coulombe, R., Grochulski, P., Sivaraman, J., Menard, R., Mort, J. S. and Cygler, M. (1996) Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment. EMBO J. **15**, 5492–5503

55 Fox, T., de Miguel, E., Mort, J. S. and Storer, A. C. (1992) Potent slow-binding inhibition of cathepsin B by its propeptide. Biochemistry **31**, 12571–12576

56 Podobnik, M., Kuhelj, R., Turk, V. and Turk, D. (1997) Crystal structure of the wild-type human procathepsin B at 2.5 Å resolution reveals the native active site of a papain-like cysteine protease zymogen. J. Mol. Biol. **271**, 774–788

57 Kaulmann, G., Palm, G. J., Schilling, K., Hilgenfeld, R. and Wiederanders, B. (2006) The crystal structure of a Cys^{25}→Ala mutant of human procathepsin S elucidates enzyme-prosequence interactions. Protein Sci. **15**, 2619–2629

58 Felbor, U., Kessler, B., Mothes, W., Goebel, H. H., Ploegh, H. L., Bronson, R. T. and Olsen, B. R. (2002) Neuronal loss and brain atrophy in mice lacking cathepsins B and L. Proc. Natl. Acad. Sci. U.S.A. **99**, 7883–7888

59 Mason, R. W., Gal, S. and Gottesman, M. M. (1987) The identification of the major excreted protein (MEP) from a transformed mouse fibroblast cell line as a catalytically active precursor form of cathepsin L. Biochem. J. **248**, 449–454

60 Gutierrez-Gonzalez, L. H., Rojo-Dominguez, A., Cabrera-Gonzalez, N. E., Perez-Montfort, R. and Padilla-Zuniga, A. J. (2006) Loosely packed papain prosegment displays inhibitory activity. Arch. Biochem. Biophys. **446**, 151–160

61 Roche, L., Tort, J. and Dalton, J. P. (1999) The propeptide of *Fasciola hepatica* cathepsin L is a potent and selective inhibitor of the mature enzyme. Mol. Biochem. Parasitol. **98**, 271–277

62 Bullough, P. A., Hughson, F. M., Skehel, J. J. and Wiley, D. C. (1994) Structure of influenza haemagglutinin at the pH of membrane fusion. Nature **371**, 37–43

Received 22 August 2011/29 November 2011; accepted 9 December 2011
Published as BJ Immediate Publication 9 December 2011, doi:10.1042/BJ20111360

Biochem. J. (2012) **442**, 563–572 (Printed in Great Britain) doi:10.1042/BJ20111360

online data

SUPPLEMENTARY ONLINE DATA

A molten globule-to-ordered structure transition of *Drosophila melanogaster* crammer is required for its ability to inhibit cathepsin

Tien-Sheng TSENG*[1], Chao-Sheng CHENG*[1], Dian-Jiun CHEN*, Min-Fang SHIH*, Yu-Nan LIU*, Shang-Te Danny HSU*†[2] and Ping-Chiang LYU*‡[2]

*Institute of Bioinformatics and Structural Biology, National Tsing Hua University, Hsinchu, 30013, Taiwan, †Institute of Biological Chemistry, Academia Sinica, Taipei 115, Taiwan, and ‡Graduate Institute of Molecular Systems Biomedicine, China Medical University, Taichung, 40402, Taiwan

EXPERIMENTAL

Expression and purification of *Drosophila* cathepsin B

The *Drosophila* cathepsin B gene was amplified from cDNA libraries using the primers 5′-CGGATCCGACCCGATGAATCTATTGCTCCTG-3′ (BamHI restriction site underlined) and 5′-GATCTCGAGTTACAGCTTGGGCAGACCCGCC-3′ (XhoI restriction site underlined). The PCR amplification program was performed by 29 cycles of 30 s at 95 °C, 30 s at 54 °C and 1 min at 72 °C. The PCR product was cloned into a pET-32a(+) vector (Novagen) to obtain a *Drosophila* cathepsin B construct containing a proregion and a mature protein. The sequence of the recombinant genes was verified by DNA sequencing (Mission Biotechnology).

E. coli BL21-Gold® (DE3) cells (Stratagene) harbouring a plasmid with the *Drosophila* cathepsin B gene were cultured in Luria–Bertani medium containing 50 mg/ml ampicillin at 37 °C until the D_{600} value of the culture was 0.6. The protein was then expressed by adding IPTG (final concentration = 1 mM). Whole cells were then harvested by centrifugation at 4000 ***g*** for 20 min, and lysed by sonication. The lysates were centrifuged at 16000 ***g*** for 20 min at 4 °C. The supernatant was removed and the precipitate was resuspended in 20 ml of 6 M guanidine buffer [50 mM Tris/HCl (pH 8.0), 150 mM NaCl, 5 mM EDTA, 10 mM DTT and 6 M GdnHCl (guanidinium chloride)]. The solution was then diluted into 1 litre of 50 mM Tris/HCl (pH 8.5), 150 mM NaCl, 5 mM EDTA, 10 mM reduced glutathione, 1 mM oxidized glutathione and 0.5 M arginine overnight. After refolding, the protein solution was concentrated and dialysed against 25 mM NaH_2PO_4 (pH 7.0) and 0.5 M NaCl at 4 °C. To autoprocess cathepsin B, the protein solution was adjusted to pH 4.5 with acetic acid, and 5 mM EDTA and 5 mM DTT were added into a solution to incubate at 37 °C for 1 h.

Proteins were purified by an AKTAprime system with a HiPrep® Sephacryl S-100 high-resolution gel-filtration column (Amersham Biosciences). The running buffer was 100 mM sodium acetate buffer (pH 5.0), containing 1 mM EDTA and 2 mM DTT. Each fraction was identified using Z-Phe-Arg-AMC substrate (Calbiochem) to confirm the cathepsin B activity.

[1] These authors contributed equally to this work

[2] Correspondence may be addressed to either of these authors (email lslpc@life.nthu.edu.tw or sthsu@gate.sinica.edu.tw).

The structural co-ordinates reported will appear in the PDB under accesion code 2L95.

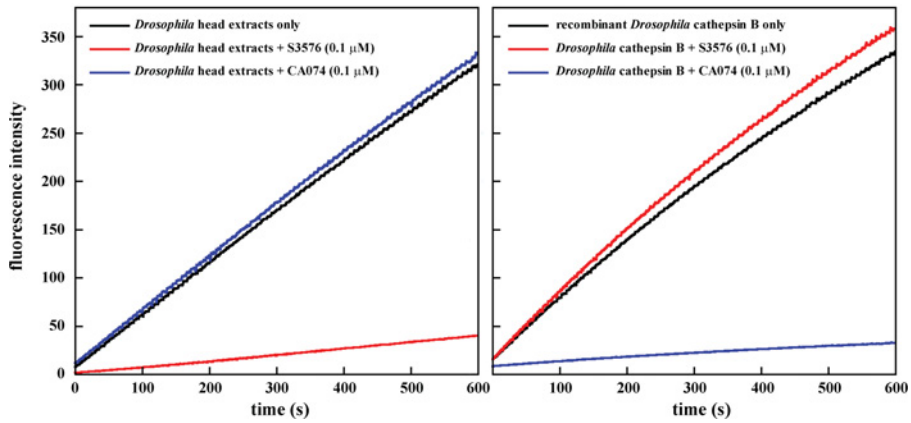

Figure S1 Progress curves for the inhibition of cathepsins

To identify the active substance in our *Drosophila* head extract, we used cathepsin B/L-specific inhibitors. A cathepsin L-specific inhibitor (S3576, {tert-butyl-[(2S)-1-{2-[2-(2-ethylanilino)-2-oxoethyl]sulfanylcarbonylhydrazinyl}-3-(1H-indol-3-yl)-1-oxopropan-2-yl]carbamate} shows strong inhibition against the *Drosophila* head extract (left-hand panel). The same use of S3576 does not inhibit recombinant *Drosophila* cathepsin B (right-hand panel). In contrast, the cathepsin B-specific inhibitor {CA074, [L-3-*trans*-(propylcarbamyl)oxirane-2-carbonyl]- L-isoleucyl-L-proline} exhibits opposite results in the same assay. This indicates that the enzymatically active substance in the *Drosophila* head extract is mainly cathepsin L.

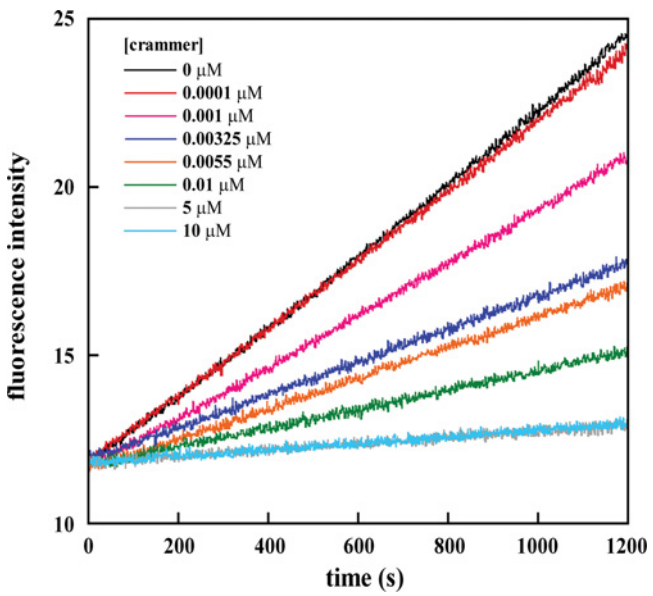

Figure S2 Progress curves for the inhibition of cathepsin L by crammer

Progress curves for the inhibition of *Drosophila* cathepsin L in the presence of various concentrations of crammer gave linear plots, suggesting a concentration-dependent inhibition of initial velocity of product formation. The crammer concentrations are labelled in the Figure.

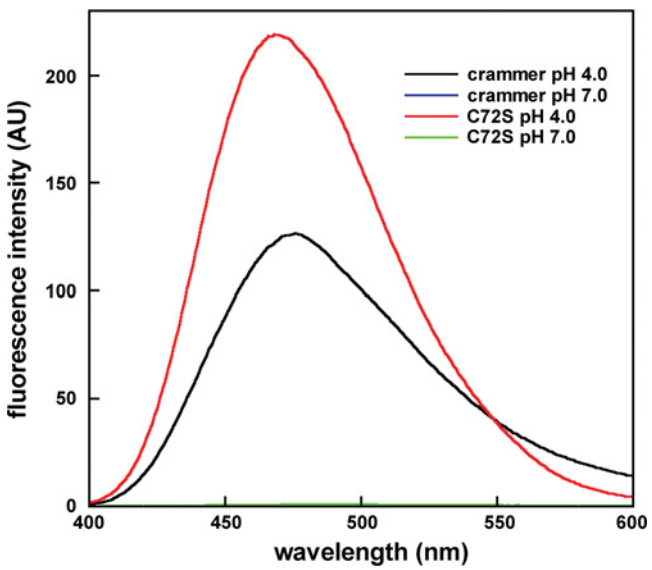

Figure S3 ANS binding

The exposure of hydrophobic core was investigated by ANS fluorescence spectra. The ANS intensity was recorded from 400 to 600 nm (λ_{ex} = 365 nm).

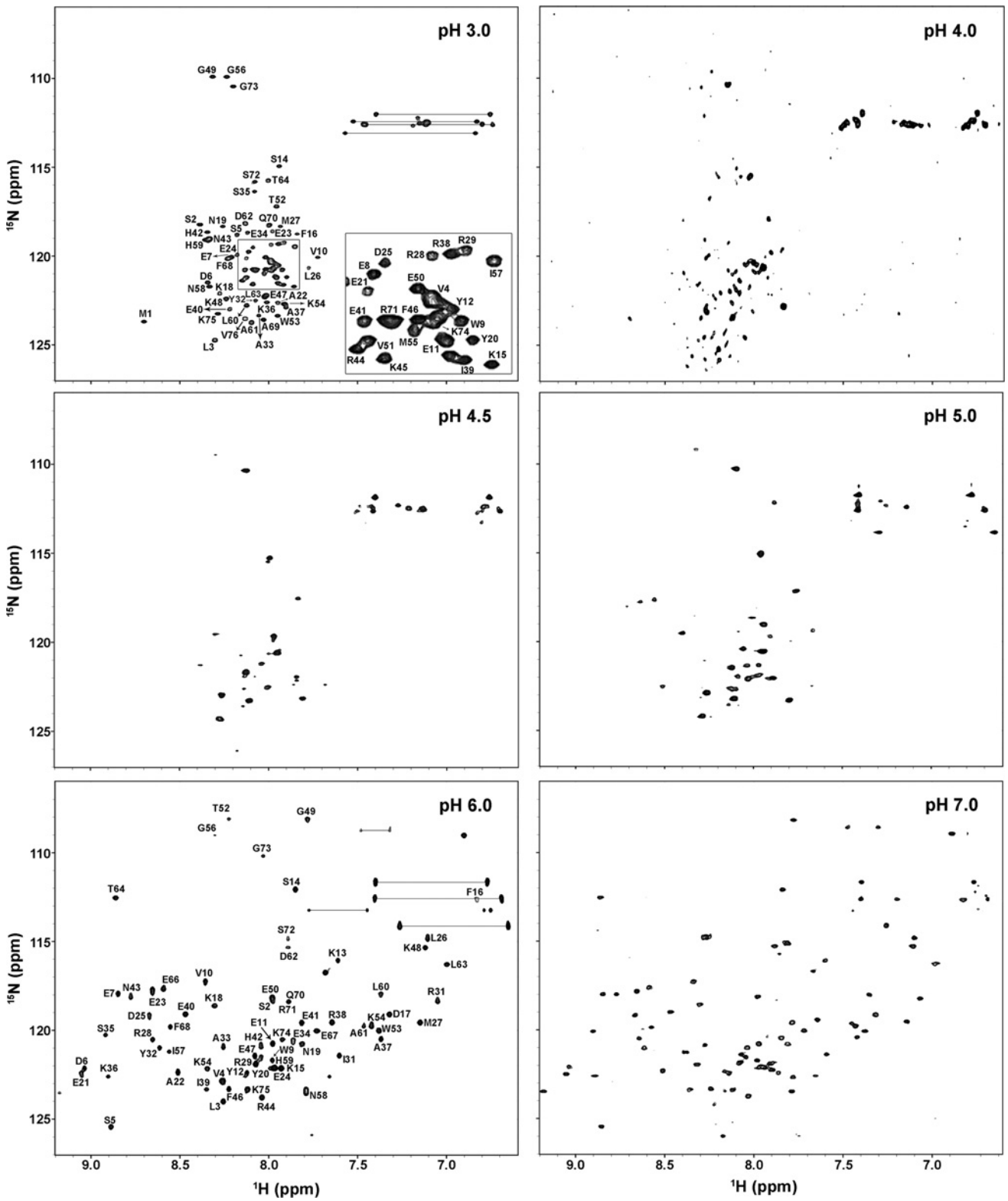

Figure S4 ^{1}H-^{15}N-HSQC spectra of C72S as a function of pH

All spectra were recorded at 25 °C, and the assigned cross-peaks for crammer at pH 3.0 and pH 6.0 are labelled in the Figure. The insert in the upper-left-hand panel shows an expanded view of part of the spectrum.

Received 22 August 2011/29 November 2011; accepted 9 December 2011
Published as BJ Immediate Publication 9 December 2011, doi:10.1042/BJ20111360