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*1, 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

**INTRODUCTION**

*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.

**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 pro-

**REFERENCES**

1. ANS, 8-anilinonaphthalene-1-sulfonic acid; BMRB, BioMagResBank; DTT, dithiothreitol; E-64, trans-epoxysuccinyl-L-leucylamido-(4-guanidino)butane; HSQC, heteronuclear single quantum correlation; IFTG, isopropyl β-D-thiogalactopyranoside; LTM, long-term memory; NOE, nuclear Overhauser effect; hetNOE, heteronuclear NOE; NOESY, nuclear Overhauser enhancement spectroscopy; SEC, size-exclusion chromatography.

2. Correspondence may be addressed to either of these authors (email tsgpc@life.nthu.edu.tw or sshu@gate.sinica.edu.tw).

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

Materials

ANS (8-anilinonaphthalene-1-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 D660 value of the culture reached 0.7. IPTG (final concentration, 1 mM) was added to the cell culture to initiate recombinant protein overexpression. 15N-labelled and 15N/13C-labelled crammer samples were obtained from E. coli cultures incubated in M9 medium containing (1 g/l) 15NH4Cl and/or (2 g/l) [13C]glucose/15NH4Cl (Cambridge Isotope Laboratories) containing 1 mM EDTA, 0.1 mM DTT and 0.1% Triton X-100. Purified crammer (∼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.

Characterization of the oligomeric states of crammer

Purified crammer (∼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.

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) was recorded as a function of pH using a flurometer (F-7000, Hitachi) between 290 and 400 nm (λex = 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 (λex = 365 nm).

NMR spectroscopy

Uniformly 13C/15N-labelled crammer (0.3 mM) was buffered in 10 mM citric acid-sodium phosphate containing 10% (v/v) 2H2O. The pH values of the samples were adjusted prior to NMR measurements at 25°C using AVANCE NMR spectrometers (Bruker Spectrospin) operating at 1H Larmor frequencies of 500 or 600 MHz, and a Varian INOVA NMR spectrometer equipped with a cryogenic probe head and operating at the 1H 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). 1H, 13C and 15N backbone and side-chain resonance assignments

Table 1 NMR constraints and structural statistics for C72S

<table>
<thead>
<tr>
<th>Distance constraints</th>
<th>RMSD, root mean square deviation.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total NOE</td>
<td>859</td>
</tr>
<tr>
<td>Intraresidue</td>
<td>341</td>
</tr>
<tr>
<td>Sequential ([I–J] = 1)</td>
<td>184</td>
</tr>
<tr>
<td>Medium range ([I–J] ≤ 4)</td>
<td>78</td>
</tr>
<tr>
<td>Long range ([I–J] &gt; 4)</td>
<td>61</td>
</tr>
<tr>
<td>Hydrogen bond constraints</td>
<td>61</td>
</tr>
<tr>
<td>Total d/hedral angle constraints</td>
<td>134</td>
</tr>
<tr>
<td>φ</td>
<td>67</td>
</tr>
<tr>
<td>ψ</td>
<td>67</td>
</tr>
<tr>
<td>RMSD (Å) with respect to the average structure</td>
<td></td>
</tr>
<tr>
<td>Well-defined regions (residues 8–49 and 52–74)</td>
<td>10.4 ± 0.1</td>
</tr>
<tr>
<td>Backbone</td>
<td>1.9 ± 0.4</td>
</tr>
<tr>
<td>Heavy atoms</td>
<td>2.2 ± 0.4</td>
</tr>
<tr>
<td>All residues (residues 1–80)</td>
<td>9.3 ± 0.6</td>
</tr>
<tr>
<td>Backbone</td>
<td>1.9 ± 0.4</td>
</tr>
<tr>
<td>Heavy atoms</td>
<td>2.2 ± 0.4</td>
</tr>
<tr>
<td>Total energy after water refinement (kcal/mol)</td>
<td>−2700 ± 100</td>
</tr>
<tr>
<td>RMSD from idealized covalent geometry</td>
<td></td>
</tr>
<tr>
<td>Bonds (Å)</td>
<td>0.0153 ± 0.0005</td>
</tr>
<tr>
<td>Angles (°)</td>
<td>2.10 ± 0.07</td>
</tr>
<tr>
<td>Impropers (°)</td>
<td>2.6 ± 0.1</td>
</tr>
<tr>
<td>Ramachandran statistics</td>
<td></td>
</tr>
<tr>
<td>Most favoured regions</td>
<td>85.2%</td>
</tr>
<tr>
<td>Additionally allowed regions</td>
<td>13.6%</td>
</tr>
<tr>
<td>Generously allowed regions</td>
<td>1.1%</td>
</tr>
<tr>
<td>Disallowed regions</td>
<td>0.0</td>
</tr>
<tr>
<td>BioMagResBank accession code</td>
<td>16719</td>
</tr>
<tr>
<td>PDB code</td>
<td>2195</td>
</tr>
</tbody>
</table>

© The Authors. Journal compilation © 2012 Biochemical Society
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 $^{15}$N 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. 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-15N-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$_\alpha$, $^{13}$N, $^1$H$_\text{NZ}$, $^{13}$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),...
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 $1.36 \pm 0.67$ 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 $3.85 \mu$M to $20.06 \mu$M without affecting the $V_{max}$ value ($0.006 \text{ min}^{-1}$) (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 $\alpha$-helical content of $63\pm4\%$; crammer is less helical...
Cathepsin inhibition by crammer

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 1H-15N 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 (HN, 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 1H-15N 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 1H-15N 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
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 1H,15N 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-bound 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 Glu8, Pro65 and Val86–Asn89 (>91% of all the expected H2, 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 1C/15N-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 [Asp6–Phe16 (α1), Ala22–Lys48 (α2), His59–Ala61 (α3) and Pro65–Arg71 (α4)] with α2 serving as the core structure that appears to maintain the overall tertiary structure (Figure 5A). The three-dimensional
The 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 (Asp6–Arg29 and Glu8–Lys36), and two that connect α2 and α4 (Glu24–Arg28 and Arg28–Glu67) (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 (Trp9, Tyr12, Phe16, Tyr20, Tyr32 and Phe68) 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, $P = V_0 t + (V_s - V_0) \times \frac{1 - \exp(-k_{obs} t)}{k_{obs}}$, 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, 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.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Crammer</th>
<th>W9A</th>
<th>Y12A</th>
<th>F16A</th>
<th>Y20A</th>
<th>C72S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crammer K_i value</td>
<td>0.79 ± 0.35 μM</td>
<td>N.D.</td>
<td>18.94 ± 2.24 μM</td>
<td>2.26 ± 0.23 μM</td>
<td>5.40 ± 1.25 μM</td>
<td>1.87 ± 0.63 μM</td>
</tr>
</tbody>
</table>

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, Ser72, located in the flexible C-terminal region, is more than 95% solvent accessible. Cys72 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 (YXXKXY) 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
acids, 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 proregen (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 proregen 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 proregen at acidic pH [59]. When cleaved, the proregen exists as a molten globule [60], as we have shown here for crammer under acidic conditions. Nonetheless, the cleaved proregen 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 proregen 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 proregen 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 analyzed the data; and Tien-Sheng Tseng, Chao-Sheng Cheng, Shang-Te Danny Hsu and Ping-Chiang Lyu wrote the paper.

**REFERENCES**


**FUNDING**

This work was supported by the National Science Council, Taiwan [grant numbers NSC 90-2811-M-007-073, 99-3112-B-007-003, 99-2627-B-007-010 and 100-2113-M-001-031-MY2].
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′-CGGA TCC GACCCGA-TGAATCTATTGCTCCTG-3′ (BamHI restriction site underlined) and 5′-GA TCTCGAG TTACAGCTTGGGCAGACCCGCC-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 NaH2PO4 (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 accession code 2L95.
To identify the active substance in our Drosophila head extract, we used cathepsin B/L-specific inhibitors. A cathepsin L-specific inhibitor (S3576, \{tert-buty1-(2S)-1-\{2-[2-(2-ethylanilino)-2-oxoethyl]sulfanylcarboxybenzyl\}hydrazinyl\}-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.

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.

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