Crystal Structure of Human Cystatin D, a Cysteine Peptidase Inhibitor with Restricted Inhibition Profile

Marcia Alvarez-Fernandez[+], Yu-He Liang[§], Magnus Abrahamson[+,¶] and Xiao-Dong Su[‡, §,¶]

[+] Department of Clinical Chemistry, Institute of Laboratory Medicine, Lund University, SE-221 85 Lund, Sweden,

[§] The National Laboratory of Protein Engineering and Plant Genetic Engineering, Peking University, 100871 Beijing, China and

[‡] Department of Molecular Biophysics, Center for Chemistry and Chemical Engineering, Box 124, SE-221 00 Lund, Sweden

All manuscript correspondence should be addressed to:

Dr. Magnus Abrahamson
Dept. of Clinical Chemistry
Lund University Hospital
S-221 85 Lund
SWEDEN
Telephone: +46 46 17 34 45
FAX: +46 46 18 91 14
E-mail: Magnus.Abrahamson@klinkem.lu.se

RUNNING TITLE: Crystal structure of cystatin D

SUMMARY

Cystatins are natural inhibitors of papain-like (family C1) and legumain-related (family C13) cysteine peptidases. Cystatin D is a type 2 cystatin, a secreted inhibitor found in human saliva and tear fluid. Compared to its homologues, cystatin D presents an unusual inhibition profile with a preferential inhibition cathepsin S > cathepsin H > cathepsin L, and no inhibition of cathepsin B or pig legumain. To elucidate the structural reasons for this specificity, we have crystallized recombinant human Arg26-cystatin D and solved its structures at room temperature and at cryo conditions to 2.5 and 1.8 Å resolution, respectively. Human cystatin D presents the typical cystatin fold, with a five-stranded anti-parallel β-sheet wrapped around a five-turn α-helix. The structures reveal differences in the peptidase-interacting regions when compared to other cystatins, providing plausible explanations to the restricted inhibitory specificity of cystatin D for some papain-like peptidases, and its lack of reactivity towards legumain-related enzymes.

INTRODUCTION

Cystatins are natural inhibitors of family C1 (papain-like) cysteine peptidases. In mammals, cystatins inhibit peptidases such as cathepsins B, H, K, L, and S both intra- and extracellularly following a reversible, tight-binding mechanism (1). The family C1 enzymes are involved in the normal lysosomal turnover of proteins, but are also implicated in many disease processes, such as tumor invasion and connective tissue destruction at inflammation (2-4).

The cystatins constitute a superfamily of related proteins. The mammalian superfamily members are of three major types (1,5,6). Type 1 cystatins (also called stefins) are primarily cytoplasmatic, single-domain proteins composed of approximately 100 amino acid residues, with no disulfide bridges and no signal peptide. Type 2 cystatins are secreted inhibitors, also single-domain proteins but about 120 residues long, and present two well-conserved disulfide bridges and typical signal peptides. Type 3 cystatins, or kininogens, are multidomain proteins presenting three tandemly repeated type 2 cystatin-like domains.

Chicken egg-white (CEW[1]) cystatin, an avian type 2 cystatin, was the first cysteine peptidase inhibitor for which the three-dimensional structure was determined by X-ray crystallography (7). The structures of two type 1 cystatins have also been determined: human cystatin A (or stefin A) by both NMR spectroscopy (8) and, recently, by X-ray crystallography of a complex with cathepsin H (9), and human cystatin B (or stefin B) in complex with papain by X-ray crystallography (10). All three cystatins show the same overall structure, with a five-stranded antiparallel β-sheet wrapped around a five-turn α-helix. In these cystatin structures the papain-binding site is a tripartite, wedge-shaped edge, formed by the N-terminal segment and the first and second hairpin loops, called L1 and L2. There is also a NMR model for the plant inhibitor, oryzacystatin, which shows the same "cystatin fold" as the animal cystatins (11). In addition, the structure of a dimeric form of human cystatin C has been published (12). Although this dimeric form of cystatin C is inactive as a papain inhibitor due to shedding of the binding site, each of the two domains formed by 3D subdomain swapping adopt the monomeric cystatin fold.

Despite these quite extensive structural data, detailed knowledge of what determines the specificity profiles of different cystatins is lacking. Cystatin D is a type 2 cystatin so far only found in human saliva and tear fluid (13). It is produced as a preprotein of 142 amino acid residues, of

which the first 20 constitute a typical signal peptide (14). Cystatin D was originally found as the product of a gene segment displaying a high degree of homology to the human cystatin C gene (15). Its complete amino acid sequence displays 55% identical residues compared to the cystatin C sequence, with all sequence motifs known to be essential for cysteine peptidase inhibition well conserved (14). However, the inhibition profile of cystatin D for human family C1 peptidases is clearly different from that of, e.g., cystatin C (1,16). Unlike the latter, cystatin D is unable to inhibit cathepsin B. Besides it shows a preferential inhibition of cathepsin S over cathepsins H and L (16). By a site-directed mutagenesis approach to alter residues in the N-terminal segments of cystatin D and C, it has been shown that these residues can interact with the non-primed substrate pockets of the enzymes in a substrate-like manner (17). Moreover, evidence was presented that N-terminal sequence differences partly explain the specificity differences between cystatins D and C. However, by analysis of engineered hybrid cystatins it was apparent that structural differences also in the frame-work cystatin molecule must have a large effect on the inhibitory specificity of cystatin D (17).

It was recently reported that some type 2 cystatins can inhibit mammalian legumain, a cysteine peptidase of family C13, through a novel reactive site located on the opposite side to the papain-binding site (18), in a loop referred to as the back-side loop (BSL). This site results in tight reversible inhibition of pig legumain and is active on human cystatins C, E/M and F, but not on cystatins A, B and D. Thus, also with respect to inhibition of family C13 enzymes, cystatin D displays a more restricted and specific inhibition profile than other type 2 cystatins.

In the present study, we have crystallized and determined the three-dimensional structure of recombinant human cystatin D, with the aim to clarify the structural reasons for its selectivity at target enzyme inhibition.

EXPERIMENTAL PROCEDURES

Expression and purification of recombinant cystatin D

Human 'Arg26-cystatin D' (one of the two allelic variants present in approx. equal proportions in the population (19)) was overexpressed in an *E. coli* expression system as described before (13). After expression, the protein was purified by anion exchange chromatography on a Q-Sepharose column [30 x 300 mm^2] (Amersham Pharmacia Biotech, Uppsala, Sweden), followed by size exclusion chromatography (SEC) on a Superdex 75 10/30 column (Amersham Pharmacia Biotech) connected to a FPLC system. The anion exchange chromatography was performed using 20 mM ethanolamine, pH 9.0, containing 1 mM benzamidinium chloride as elution buffer, and the SEC using 50 mM Tris buffer, pH 7.5, with 150 mM NaCl. The fractions of highest purity were pooled and dialyzed against 100 mM Tris buffer, pH 7.5. The protein solution was then concentrated using a Vivaspin column with cut-off limit of 5000 Da (Vivascience, Lincoln, UK), to a final concentration of approximately 8 mg/ml.

Protein concentrations were determined by UV absorption spectroscopy at 280 nm using ε = 18,200 $M^{-1}cm^{-1}$ as extinction coefficient ($A_{280,\ 0.1\%}$ = 1.29) (17). Purity of the protein in SEC fractions was determined by size- and charge-separating electrophoreses, in 16.5% SDS-PAGE gels (20) and 1% agarose gels (21), respectively.

Crystallization

Crystallization plates were prepared using the hanging-drop vapor diffusion method in 24-well VDX-plates (Hampton Research, Laguna Nigel, CA). Initial screening of crystallization conditions, at 18°C, was done using the Crystal Screen (CS) kits 1 and 2 (22,23) (Hampton Research). Five-μL droplets were used in the initial screens (2.5 μL protein solution and 2.5 μL precipitant solution) and 6-10 μL droplets in optimization trials. Reservoirs contained 750 μL in initial screens and 1000 μL at optimization.

X-ray data collection and processing

Room temperature (RT) data were collected on a Mar image plate system (Marresearch GmbH, Hamburg, Germany) mounted on a Rigaku RU-200 rotating anode generator operating at 50 kV, 90

mA. Crystals were mounted in a quartz capillary for data collection. A full data set was collected from a single crystal.

Cryo-conditions for data collection were worked out using sucrose as cryo protectant. The crystal was equilibrated with the mother liquor in the presence of 15% sucrose for a few minutes. The crystal was then mounted in a nylon CryoLoop (Hampton Research) and flash-cooled directly in a cold nitrogen stream at about 100K. Diffraction data were collected at the crystallographic beamline BL711 at the MAX-II synchrotron lab in Lund (Sweden) using a Mar345 image plate detector (X-ray Research GmbH, Norderstedt, Germany). A typical exposure time was 60 s per frame with 1° oscillation. All data sets were processed using the DENZO and SCALEPACK packages (24).

Structure determination

The structure of cystatin D was solved by molecular replacement methods. For the RT data, this was done by using CEW cystatin[2] as search model. Different modifications, such as poly Ala, poly Ser, were tried. The programs AMoRe (25) and Crystallography & NMR System (CNS) (26) were used for the replacement search of the data between 15.0 and 4.0 Å. The molecular replacement solution was refined using the program CNS on the complete RT data (30.0 to 2.5 Å). The refined RT structure was then used as model for the rigid-body refinement on the cryo data.

Structural alignment and graphical illustrations

Multiple sequence alignment of cystatins with known structures was initially done by the GCG (Genetics Computer Group) Wisconsin Package software. The alignment was modified using the multiple structure alignment obtained with the program MAPS (Multiple Alignment of Protein Structures)[3]. The structures used in the alignment were obtained from the Protein Data Bank (27): CEW cystatin[2], dimeric human cystatin C[4], cystatin A[5], cystatin B[6], and oryzacystatin[7]. If not otherwise indicated, the amino acid numbering used is that of human cystatin C[8], as previously used for cystatin D and other human type 2 cystatins (13,28,29). Graphical representations were prepared with the programs MOLMOL (30) and GRASP (31).

RESULTS AND DISCUSSION

Crystallization of cystatin D and crystal data collection

Recombinant human cystatin D crystals appeared during the first week in CS-kit 1 condition 39 (100 mM Na-HEPES buffer, pH 7.5, with 2% (w/v) PEG400 and 2 M $(NH_4)_2SO_4$), at 18°C. Finer grids based on this condition were settled at the same temperature by using either Tris or HEPES as buffer at a pH interval between 6.5 and 8.0 and by varying the ammonium sulfate (0.4-2.4 M) and PEG (1-2% (w/v)) concentrations. Crystals were obtained under several conditions. They were stable and presented typical shapes as long rods or plates. Two of the well-diffracting crystals were used for structure determination. These crystals were grown at 18°C in 100 mM Tris, pH 7.5, with 2.4 M $(NH_4)_2SO_4$ and 2.5% PEG 400.

A room temperature data set was collected from a plate-shaped crystal with dimensions about 0.5 x 0.3 x 0.1 mm^3. The crystal diffracted beyond 2.5 Å and a full data set could be collected from a single crystal (Table I). A second similar crystal soaked in 15% sucrose was used to collect a data set at about 100 K, giving diffraction beyond 1.8 Å (Table I).

<Table I around here>

Molecular replacement was used as method to solve the structure of cystatin D from the RT data set. This was accomplished using the crystal structure of CEW cystatin as search model. Using AMoRe (25) and CNS with data collected between 15.0 and 4.0 Å, we obtained the same rotational solutions, which were well above background regardless of which model was used. From the extinction list, two axes were clearly shown as screw axes. Thus, the space groups $P2_12_12_1$ and $P2_12_12$ were both tested for translational search. The space group $P2_12_12$ gave the correct solution. The rigid body refinement using CNS (32) with the RT data (30.0 to 2.5 Å) lowered the R_{cryst}/R_{free} from 0.438/0.438 to 0.349/0.342, respectively. The simulated annealing method was then applied for further refinement. The maps were calculated and inspected, and the residues of the search model were changed to the correct ones. Composite-omit-maps were then calculated to remove model bias. A total of 112 residues, from position Ala10 to Val120 (human cystatin C numbering, Figs. 1A, 2), are included in the final RT structure model. No electron density was detected for the residues in the N-terminal segment before Ala10 (Fig. 1A). This region must thus be disordered in

structure. The flexible region around residues 80-84 was difficult to build in before the composite-omit-maps were made. Thirty-two water molecules were added to the model where strong difference densities (>3σ) were shown and the hydrogen-bond geometry was good. The individual B-factor refinement was applied to the final RT model.

The structure solution was straightforward for the cryo data after the RT structure was refined. The cryo structure had slight but significant changes in the cell dimensions (Table I). The refined RT model was then used for the rigid-body refinement on the cryo data in the resolution range from 30.0 to 1.8 Å. It was trivial to perform the subsequent refinement steps by CNS and to add a total of 85 water molecules to the cryo model (Table I).

Despite the significant unit cell changes, particularly on the length of b-axis, the RT and cryo structures turned out to be very similar with root mean square deviation (RMSD) on the Cα trace of 0.36 Å and an overall RMSD with side-chains of 0.91 Å. At the C1 peptidase binding region, the RMSD values when comparing the cryo and RT structures are 0.23, 0.13 and 0.63 Å for the main-chain atoms of the N-terminal part (amino acid residues Gly11-Ala15), the L1 (Gln55-Gly59) and the L2 (Val104-Asp108) loops, respectively. The relatively large RMSD value for L2 is attributable to the contribution from Pro105. At the putative legumain (C13 peptidase) binding site, the RMSD value for the main-chain atoms of the BSL (Val37-Glu41) is 0.15 Å.

The strong similarity between the RT and cryo structures can also be seen from a B-factor plot (web figure).

Overall structure

Human cystatin D adopts the so-called 'cystatin fold' (Fig. 1B), as its five homologues with known structures (7,8,10,11). The core structure is built from a five-stranded antiparallel β-sheet (consisting of β1: Ile13-Thr16, β2: Ser44-Ile57, β3: Val60-Thr71, β4: Glu95-Val104, β5: Lys109-Lys119) that is wrapped around a five-turn α-helix (Lys21-Lys36) (Figs. 1B, 2).

Comparison with CEW cystatin (Figs. 1C, D) revealed some notable differences in the overall structure of cystatin D: 1) The loop at the C-terminal end of the α-helix is larger than that in CEW cystatin, which deforms the last turn of the helix. 2) Cystatin D does not present a bulge in the middle of the second strand of the β-sheet around position 49. 3) There is no helix in the appendix

loop of cystatin D. Instead, it presents a disordered conformation. This is also the case for the monomeric domains in the crystal structure of dimeric human cystatin C (12). 4) Marked differences in the putative peptidase-interacting regions of cystatin D are observed (see below).

Comparison of the electrostatic potential surfaces of the two proteins (Fig. 1E) revealed further differences. Cystatin D has a narrower and more elongated shape than CEW cystatin. This might be a result of the missing bulge in the β2 strand, straightening up the β-sheet in cystatin D. Also, the two cystatins differ quite significantly with respect to the charge distribution on their surfaces. In CEW cystatin, positive and negative charges are evenly distributed on the protein surface. In cystatin D, however, the surface presents some strongly (mainly negatively) charged areas whereas other areas are pronounced hydrophobic.

Human cystatin D is present in two natural forms due to a gene polymorphism (19). It has been shown that this variation neither significantly affects the enzyme-binding properties of the inhibitor nor has drastic effects on protein stability (16), but the structural consequences of the variation have not been elucidated. The form crystallized here, 'Arg26-cystatin D', in which the 26th residue of the 122-residue predicted mature cystatin D sequence (13) is Arg, has a population frequency of 0.45 (19). As the predicted amino acid sequence of mature cystatin D is one residue longer in the N-terminal than the reference sequence of cystatin C[8], the polymorphic residue is number 25 in an alignment of cystatin sequences (Fig. 2). The other natural form of human cystatin D (population frequency of 0.55) has Cys in this position unlike other type 2 cystatins, for which Arg25 is well conserved (Fig. 2). The present structure shows that the Arg25 residue in cystatin D is situated in the second turn of the α-helix. Although it seems to be exposed at the surface of the protein, its side-chain is undoubtly oriented towards the cavity formed by the bent L2 loop, as revealed by the density map. The side-chain is likely 'trapped' by a salt bridge with the side-chain of either Glu103 or Asp108 in the proximity of the L2 loop. The electron density for the amine groups further out in the side-chain of Arg25 is weak or almost absent in both the cryo and the RT structures. This indicates a large flexibility in the conformation of these amine groups, reflected by high B-factor values, suggesting that they can alternate between the two anchoring sites formed by Glu103 and Asp108. Similarly, the conserved Arg residue in CEW cystatin seems to form a hydrogen bond with the Ser residue in position 103 (Ser101 in CEW cystatin numbering). By

homology, the conserved Arg25 residue in the type 2 cystatins S, SA and SN could also form a salt bridge with Glu103 or Glu108 found in their sequences. The Arg25 residue thus appears to be an important factor for stabilization of the α-helix of some type 2 cystatins. In the cystatin D variant with Cys as residue 25, the bridge between the α-helix and the L2 loop cannot be formed and its stabilizing feature would be lost. This could be an explanation to the fact that the Cys variant is connected to both lower expression yields in *E. coli* and lower purification yields from saliva than the Arg variant (16). As pointed out by Balbin *et al.,* the unpaired Cys residue may be involved in disulfide exchange with other proteins in saliva (16). The present results demonstrate that the Cys side-chain with its thiol group indeed could be exposed to allow this.

<Fig. 1 around here>

<Fig. 2 around here>

The binding site for papain-like peptidases

The papain-binding site in cystatins is constituted by three well conserved segments: the flexible N-terminal part, a hairpin loop in the central region (L1) and another towards the C-terminal end (L2) of the sequence (33-42). These segments form a tripartite wedge-shaped edge (7,8,10,11,43) that enters the catalytic site in a substrate-like manner (10,44). In the cystatin D structure, the papain-binding segments are located in such a tripartite wedge (Fig. 1) in accordance with the other cystatin structures.

The density map for the N-terminal segment in cystatin D is poorly defined and, therefore, the structure model gives little information about the potential for interactions between the N-terminal segment of this cystatin and target family C1 peptidases. In the RT model, the first residue for which clear electron density is shown is Ala10. In the cryo model for cystatin D, this residue is Gly11. In the structures of CEW cystatin and human cystatin C, the first residues with clear density are Gly11 (Gly9 in CEW cystatin numbering) and Val10, respectively (7,12). This highlights a similar, very flexible N-terminal segment in all type 2 cystatins. For the other parts of the potential papain-binding site, there are some relevant differences in the cystatin D structure compared to that of CEW cystatin, however.

<Fig. 3 around here>

The L1 loop of the binding site contains the conserved 'cystatin motif', QXVXG, and is located between strands β2 and β3. L1 in cystatin D is larger and adopts a more 'squared' form than in CEW cystatin (Fig. 3A). This broader loop is a consequence of the missing β-bulge around position 49 in the β2 strand. This bulge is present in all cystatin structures solved so far, i.e., in CEW cystatin, human cystatins C, A and B, as well as in the plant cystatin, oryzacystatin. The missing bulge causes a displacement of one amino acid residue in the β2 strand of cystatin D, when comparing the sequence-based alignments of earlier publications (29) to the three-dimensional structure alignment based on the present results (Fig. 2). Both alignments come in tune again after Gly59, at the end of the papain-binding 'cystatin motif'. Thus, the extra residue is placed in the L1 loop, changing its morphology (Fig. 3A).

The L2 loop of the binding site is, as for CEW cystatin, a five-residue long hairpin loop between strands β4 and β5. The L2 appears to be more bent over the α-helix than in the crystal structure of CEW cystatin (Fig. 3B). This causes a deviation of approximately 30 degrees of the side chains of Pro105 and Trp106 from the wedge-shaped edge formed by the three segments in CEW cystatin.

From the electrostatic point of view, the charge distribution on the papain-binding site (N-term, L1 and L2) of CEW cystatin is completely hydrophobic. In cystatin D, however, the very negative charges of Glu107 and Asp108 in L2 (Fig. 1E) disrupt the otherwise hydrophobic wedge.

These quite drastic structural and electrostatic differences in the papain-binding site of cystatin D compared to other cystatins are likely the key to understand the inhibition profile of cystatin D (16). In analogy with other cystatins, there is a substantial amount of kinetic data indicating that cystatin D inhibits papain by a one-step reaction (40), suggesting that no significant structural modifications occur in the inhibitor or the peptidase upon interaction. Based on a model of the papain inhibition by cystatin D, simulated by substituting cystatin D in the crystal structure of the cystatin B (stefin B) complex with papain (10) and assuming that both cystatins dock into the peptidase active site cleft in a similar way, it seems that cystatin D should fit into the active cleft of papain-like enzymes in an analogous way as cystatin B does: 1) The N-terminal segment should enter the enzyme's narrow cleft as Gly11 fits in the S_1 subsite of the enzyme; 2) L1 should be in close contact with the residues forming the S_1' pocket in the enzyme and; 3) L2 should interact with

the wider part of the cleft where the conserved Trp106 in cystatin D may interact with the imidazole rings of Trp177 and Trp181 in papain, as most likely is the case for the corresponding Trp residue in other type 2 cystatins.

Cystatin D and papain should fit well together from the electrostatic point of view, as the very hydrophobic active site cleft of papain and the rather hydrophobic wedge of cystatin D should complement each other without any substantial hindrance. However, sterical hindrances caused by the local topology of cystatin D are likely the main factor weakening the binding. If cystatin D would fit into the active site groove in the same way as cystatin B does, in order to achieve the largest surface of contact with the enzyme, we would expect the side-chain of Val57a in the inhibitor to collide with the walls of the cleft, most likely with Trp177. This means that cystatin D might not be able to enter the enzyme cleft as deeply as other cystatins probably do. This results in lost contacts between the other cystatin parts involved in enzyme binding and the peptidase and, consequently, decreases the papain affinity of cystatin D compared to CEW cystatin and human cystatin C (1). Moreover, even if L1 would not be as protruding as it is, the more bent L2 might not be able to make as many contacts with papain as expected by analogy to other cystatins' mode of enzyme binding.

Previous studies have shown that a truncated form of human cystatin C, lacking the first ten residues of the N-terminal segment, has three orders of magnitude lower affinity for papain and other family C1 peptidases than full-length cystatin C (34,38). The same large decrease in target enzyme affinity is observed for its W106G variant (39). The affinity of cystatin D for papain (K_i 1.9 nM; reviewed in ref. (1)) is five orders of magnitude lower than that observed for wildtype cystatin C. These kinetic data suggest that sterical hindrances due to the larger L1 loop of cystatin D destabilize not only contacts in this region but also in the N-terminal region and/or in the L2 loop. In addition, the bent L2 loop most likely disfavors interactions between the imidazole rings in the inhibitor (Trp106) and the enzyme (Trp177 and Trp181).

Likewise, we 'docked' the structures of cystatin D and cathepsin B[9], again using the complex between cystatin B and papain as starting point for the model. Contrary to the inhibition of papain by cystatins, cathepsin B is inhibited in a two-step kinetic reaction (45). As proposed by Nycander *et al.*, the first step is regulated by the anchoring of the N-terminal part of the cystatin in the non-

primed S pockets. This is followed by displacement of the occluding loop of cathepsin B as the anchored cystatin pushes it away in order to introduce the L1 and L2 loops in the S' subsites. Structural studies have shown that the N-terminal segment of cystatin D is less favorable for this initial interaction than the N-terminal segment of cystatin C (39). In effect, when the N-terminal segment of cystatin D (the region until Gly11) was added on to a cystatin C frame work in a hybrid cystatin molecule, this variant showed 30 times lower affinity for cathepsin B than wildtype cystatin C. Still, it inhibited the enzyme (39). On the other hand, the introduction of the N-terminal segment of cystatin C into cystatin D in another hybrid molecule did not alter the inability of wildtype cystatin D to inhibit cathepsin B (39). The latter result is most likely due to that cystatin D, even if equipped with the more effective N-terminal segment of cystatin C, fails to push away the occluding loop and bind to the peptidase with the other two segments of the inhibitory wedge (the L1 and L2 loops). This is likely due to the side-chain of Val57a in the inhibitor being located too close to Trp221 in the S_1' pocket of cathepsin B. Furthermore, the L2 loop of cystatin D is marked by the presence of two negatively charged groups, i.e., in Glu107 and Asp108. These negatively charged side-chains would be situated in an unfavorable electrostatic environment established by the also negatively charged Asp224 in the active site cleft of cathepsin B, if cystatin D was forced to interact with the enzyme in the same way as cystatin B (stefin B) does with papain (10). In the same positions, CEW cystatin and human cystatin C present non-charged residues. In the case of cystatin B, although it presents a Glu residue in the loop, its positively charged His106 (His104 with CEW cystatin numbering) should fit well into the pocket.

The putative binding site for legumain-like peptidases

As recently reported, some type 2 cystatins are able to inhibit mammalian legumain (18), a lysosomal cysteine endopeptidase of family C13 (46), which shows preference for hydrolysis after an asparaginyl bond (47). The 'back-side loop' (BSL) at the end of the main α-helix (Fig 1A, B), containing residue Asn39, was identified as most likely being directly involved in legumain inhibition by these cystatins. In effect, Asn39 could be responsible for an inhibitory mechanism where cystatins inhibit mammalian legumain in a substrate-like manner (18). Like the inhibitorily active cystatins identified so far, human cystatins C, E/M and F, CEW cystatin and Bm-CPI-2, a

type 2 cystatin homologue from the filarial nematode parasite *Brugia malayi*[10], cystatin D presents an asparagine residue in this loop (Fig. 1, 2, 4). Still so, cystatin D is the only human type 2 cystatin investigated that cannot inhibit mammalian legumain (18). Assuming that Asn39 in other type 2 cystatins indeed is directly involved in legumain inhibition, we examined and compared the structures in the BSLs of cystatin D and CEW cystatin (Fig. 4), the latter being a tight-binding inhibitor of pig legumain (48). This was done in the attempt to provide a plausible explanation to why cystatin D is inactive as a legumain inhibitor.

For cystatin D, the structural differences observed were as follows: 1) The Asn residue in cystatin D is neither structurally conserved nor accessible at the surface of the protein. Instead, it is located at a position corresponding to residue 38 in the other type 2 cystatin structures (residue 36 in CEW cystatin) (Fig. 4A). Its side-chain points towards the core of the protein and is strongly hydrogen-bonded to Lys75 at the end of the third β-strand. 2) Structurally, there is an insertion of one amino acid residue in the loop. The isoleucine residue located between Val37 and Asn38 deviates the most from the Ala37 and Ser38 residues in the corresponding loop segment of CEW cystatin (Ala35 and Ser36 in CEW cystatin numbering). This Ile residue is surrounded by the hydrophobic environment provided by the side-chains in the C-terminal end of the α-helix and those in the end of the fifth β-strand. As a consequence of this stabilization, the Ile residue 'deforms' the loop and buries Asn38 deeper into the core. 3) Cystatin D presents a lysine residue at the position corresponding to residue 39 (37 in CEW cystatin), instead of the conserved asparaginyl residue believed to be involved in legumain inhibition in other type 2 cystatins (18). This lysine residue is oriented towards the solvent and, hence, accessible on the protein surface (Fig. 4A and B). Legumain activity is specific for the hydrolysis of substrates with an asparaginyl residue in the P_1 position (47), showing preference for Asn residues located in hydrophilic surface loops. Although accessible for enzyme binding, Lys39 is far from being adequate bait for the S_1 pocket in the enzyme. Thus, there is little reason to believe that legumain would show any affinity for the BSL in cystatin D.

One interesting possibility indicated by the finding that Lys39 is exposed in the BSL of cystatin D and situated as Asn39 in cystatin C, is that cystatin D may have evolved as an inhibitor of enzymes with preference for Lys binding in their S_1 pockets. The inhibition of legumain by some

cystatins proves that clan CD enzymes with overall similarity to legumain of family C13 could generally interact well with the BSL binding region. Good candidates for cystatin D target enzymes could e.g. be the Lys-gingipains of family C25 from the periodontal bacterium, *Porphyromonas gingivalis*, which have strict preference for Lys- bonds in substrate polypeptides (49,50). It is tempting to speculate that cystatin D acts as a physiological inhibitor of these or similar enzymes in saliva and hence could have a biological function to inhibit the growth and action of pathogenic oral bacteria.

<Fig. 4 around here>

Conclusions

In the present study, we have determined the structure of recombinant human cystatin D by X-ray crystallography under both RT and cryo conditions (2.5 and 1.8 Å resolution, respectively). This type 2 cystatin is not as widely distributed in the body as its homologue, cystatin C, but is rather restricted to saliva and tear fluid (13). This may point to a more restricted biological function of cystatin D than that of a general protector against papain-like lysosomal peptidases being released from, e.g., tumor cells or leaking from dying cells, as has been suggested for cystatin C. Besides, while cystatin C is considered as an 'universal' inhibitor, displaying inhibitory activity against all family C1 peptidases studied without relevant specificity, cystatin D shows a much more restricted inhibition profile with affinity for cathepsin S > cathepsin H > cathepsin L, and no inhibition of cathepsin B or pig legumain in family C13 (16,18). This restricted inhibition profile makes cystatin D a good target for structure-function studies aiming at an understanding of factors determining the inhibitory specificity of cystatins.

The crystal structures of cystatin D reveal no exceptional overall differences between this cystatin and its homologues. The 'cystatin fold' is rather well conserved, leaving the major structural differences to the most flexible parts of the protein, i.e., the peptidase-binding sites. Radical differences in the topology of the L1 loop in cystatin D, containing the conserved 'cystatin motif' involved in C1 peptidase-inhibition, is likely the major reason for the restricted inhibition profile of cystatin D with respect to interaction with family C1 peptidases. The larger L1 loop in

cystatin D might indicate an ability for a deeper and more selective interaction with a specific target enzyme, than we would expect for the more general inhibitor cystatin C.

Structural differences in the putative binding site for family C13 peptidases are clearly present in cystatin D, and most likely the reason why cystatin D is not an inhibitor of mammalian legumain. An equivalent to the Asn39 residue present in type 2 cystatins with ability to inhibit legumain is not present in the cystatin D structure. The Asn residue positioned in a nearby location in the cystatin D 'back-side loop' is not accessible at the surface but rather buried in the interior of the structure. This indirectly supports a model for legumain inhibition by cystatins that is relying on a substrate-like interaction between the Asn39 residue and the S_1 pocket of the active site cleft of the enzyme.

ACKNOWLEDGEMENTS

We wish to thank Drs. Matthias Bochtler and Maria Håkansson for helpful comments and fruitful discussions.

REFERENCES

1. Abrahamson, M., Alvarez-Fernandez, M., and Nathanson, C. M. (2003) *Biochem. Soc. Symp.* **70**, 179-199
2. Sloane, B. F., Moin, K., Krepela, E., and Rozhin, J. (1990) *Cancer Metastasis Rev.* **9**, 333-352
3. Mort, J. S., Recklies, A. D., and Poole, A. R. (1984) *Arthritis Rheum.* **27**, 509-515
4. Buttle, D. J., Burnett, D., and Abrahamson, M. (1990) *Scand. J. Clin. Lab. Invest.* **50**, 509-516
5. Barrett, A. J., Rawlings, N. D., Davies, M. E., Machleidt, W., Salvesen, G., and Turk, V. (1986) in *Proteinase inhibitors* (Barrett, A. J., and Salvesen, G., eds) Vol. 12, pp. 515-569, Elsevier Science Publishers BV, New York
6. Rawlings, N. D., and Barrett, A. J. (1990) *J. Mol. Evol.* **30**, 60-71
7. Bode, W., Engh, R., Musil, D., Thiele, U., Huber, R., Karshikov, A., Brzin, J., Kos, J., and Turk, V. (1988) *EMBO J.* **7**, 2593-2599
8. Martin, J. R., Craven, C. J., Jerala, R., Kroon-Zitko, L., Zerovnik, E., Turk, V., and Waltho, J. P. (1995) *J. Mol. Biol.* **246**, 331-343
9. Jenko, S., Dolenc, I., Guncar, G., Dobersek, A., Podobnik, M., and Turk, D. (2003) *J. Mol. Biol.* **326**, 875-885
10. Stubbs, M. T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarcic, B., and Turk, V. (1990) *EMBO J.* **9**, 1939-1947
11. Nagata, K., Kudo, N., Abe, K., Arai, S., and Tanokura, M. (2000) *Biochemistry* **39**, 14753-14760
12. Janowski, R., Kozak, M., Jankowska, E., Grzonka, Z., Grubb, A., Abrahamson, M., and Jaskolski, M. (2001) *Nat. Struct. Biol.* **8**, 316-320
13. Freije, J. P., Balbin, M., Abrahamson, M., Velasco, G., Dalboge, H., Grubb, A., and Lopez-Otin, C. (1993) *J. Biol. Chem.* **268**, 15737-15744
14. Freije, J. P., Abrahamson, M., Olafsson, I., Velasco, G., Grubb, A., and Lopez-Otin, C. (1991) *J. Biol. Chem.* **266**, 20538-20543
15. Abrahamson, M., Olafsson, I., Palsdottir, A., Ulvsbäck, M., Lundwall, Å., Jensson, O., and Grubb, A. (1990) *Biochem. J.* **268**, 287-294
16. Balbin, M., Hall, A., Grubb, A., Mason, R. W., Lopez-Otin, C., and Abrahamson, M. (1994) *J. Biol. Chem.* **269**, 23156-23162
17. Hall, A., Ekiel, I., Mason, R. W., Kasprzykowski, F., Grubb, A., and Abrahamson, M. (1998) *Biochemistry* **37**, 4071-4079
18. Alvarez-Fernandez, M., Barrett, A. J., Gerhartz, B., Dando, P. M., Ni, J., and Abrahamson, M. (1999) *J. Biol. Chem.* **274**, 19195-19203

19. Balbin, M., Freije, J. P., Abrahamson, M., Velasco, G., Grubb, A., and Lopez-Otin, C. (1993) *Hum. Genet.* **90**, 668-669
20. Laemmli, U. K. (1970) *Nature* **227**, 680-685.
21. Jeppson, J. O., Laurell, C. B., and Franzen, B. (1979) *Clin. Chem.* **25**, 629-638
22. Jancarik, J., and Kim, S.-H. (1991) *J. Appl. Cryst.* **24**, 409-411
23. Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y., and McPherson, A. (1994) *Acta Cryst.* **D50**, 414-423
24. Otwinowski, Z., and Minor, W. (eds) (1996) *Processing of X-ray diffraction data collected in oscillation mode* Vol. 276. Meth. Enzymol. Edited by Carter, C. W., and Sweet, R. M., Academic Press
25. Navaza, J. (1994) *Acta Cryst.* **50**, 157-163
26. Brünger, 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., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Cryst.* **D54**, 905-921
27. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) *Nucleic Acids Res.* **28**, 235-242
28. Ni, J., Abrahamson, M., Zhang, M., Alvarez-Fernandez, M. A., Grubb, A., Su, J., Yu, G. L., Li, Y., Parmelee, D., Xing, L., Coleman, T. A., Gentz, S., Thotakura, R., Nguyen, N., Hesselberg, M., and Gentz, R. (1997) *J. Biol. Chem.* **272**, 10853-10858
29. Ni, J., Fernandez, M. A., Danielsson, L., Chillakuru, R. A., Zhang, J., Grubb, A., Su, J., Gentz, R., and Abrahamson, M. (1998) *J. Biol. Chem.* **273**, 24797-24804
30. Koradi, R., Billeter, M., and Wuthrich, K. (1996) *J. Mol. Graph.* **14**, 51-55, 29-32
31. Nicholls, A., Sharp, K. A., and Honig, B. (1991) *Proteins* **11**, 281-296
32. Brünger, A. T. (1992) *XPLOR Version 3.1 A system for X-ray crystallography and NMR*, Yale University Press, New Haven
33. Abrahamson, M., Ritonja, A., Brown, M. A., Grubb, A., Machleidt, W., and Barrett, A. J. (1987) *J. Biol. Chem.* **262**, 9688-9694
34. Abrahamson, M., Mason, R. W., Hansson, H., Buttle, D. J., Grubb, A., and Ohlsson, K. (1991) *Biochem. J.* **273**, 621-626
35. Auerswald, E. A., Genenger, G., Assfalg-Machleidt, I., Machleidt, W., Engh, R. A., and Fritz, H. (1992) *Eur. J. Biochem.* **209**, 837-845
36. Björk, I., Brieditis, I., and Abrahamson, M. (1995) *Biochem. J.* **306**, 513-518
37. Björk, I., Brieditis, I., Raub-Segall, E., Pol, E., Håkansson, K., and Abrahamson, M. (1996) *Biochemistry* **35**, 10720-10726
38. Hall, A., Dalboge, H., Grubb, A., and Abrahamson, M. (1993) *Biochem. J.* **291**, 123-129

39. Hall, A., Håkansson, K., Mason, R. W., Grubb, A., and Abrahamson, M. (1995) *J. Biol. Chem.* **270**, 5115-5121
40. Lindahl, P., Nycander, M., Ylinenjarvi, K., Pol, E., and Björk, I. (1992) *Biochem. J.* **286**, 165-171
41. Lindahl, P., Ripoll, D., Abrahamson, M., Mort, J. S., and Storer, A. C. (1994) *Biochemistry* **33**, 4384-4392
42. Mason, R. W., Sol-Church, K., and Abrahamson, M. (1998) *Biochem. J.* **330**, 833-838
43. Dieckmann, T., Mitschang, L., Hofmann, M., Kos, J., Turk, V., Auerswald, E. A., Jaenicke, R., and Oschkinat, H. (1993) *J. Mol. Biol.* **234**, 1048-1059
44. Bode, W., Engh, R., Musil, D., Laber, B., Stubbs, M., Huber, R., and Turk, V. (1990) *Biol. Chem. Hoppe-Seyler* **371**, 111-118
45. Nycander, M., Estrada, S., Mort, J. S., Abrahamson, M., and Björk, I. (1998) *FEBS Lett.* **422**, 61-64
46. Chen, J. M., Dando, P. M., Stevens, R. A., Fortunato, M., and Barrett, A. J. (1998) *Biochem. J.* **335**, 111-117
47. Dando, P. M., Fortunato, M., Smith, L., Knight, C. G., McKendrick, J. E., and Barrett, A. J. (1999) *Biochem. J.* **339**, 743-749
48. Chen, J. M., Dando, P. M., Rawlings, N. D., Brown, M. A., Young, N. E., Stevens, R. A., Hewitt, E., Watts, C., and Barrett, A. J. (1997) *J. Biol. Chem.* **272**, 8090-8098
49. Pike, R., McGraw, W., Potempa, J., and Travis, J. (1994) *J. Biol. Chem.* **269**, 406-411
50. Pavloff, N., Pemberton, P. A., Potempa, J., Chen, W. C., Pike, R. N., Prochazka, V., Kiefer, M. C., Travis, J., and Barr, P. J. (1997) *J. Biol. Chem.* **272**, 1595-1600
51. Grubb, A., and Löfberg, H. (1982) *Proc. Natl. Acad. Sci. U. S. A.* **79**, 3024-3027
52. Musil, D., Zucic, D., Turk, D., Engh, R. A., Mayr, I., Huber, R., Popovic, T., Turk, V., Towatari, T., Katunuma, N., and Bode, W. (1991) *EMBO J.* **10**, 2321-2330
53. Manoury, B., Gregory, W. F., Maizels, R. M., and Watts, C. (2001) *Curr. Biol.* **11**, 447-451

FOOTNOTES

* This work was supported by grants from the Crafoord Foundation, the A. Österlund Foundation, the Swedish Science Council (project no. 09915) and the Faculty of Medicine at the University of Lund. In part it was sponsored by the Commission of the European Communities, specific RTD program "Quality of Life and Management of Living Resources", QLRT-2001-01250, "Novel non-antibiotic treatment of staphylococcal diseases". X.-D.S. was supported by the Swedish Cancer Society, SBNet (Swedish structural biology network), Peking University, and by grants from the National High Technology and Development Program of China (863 program 2002BA711A13) and the National Science Fund of China (NSFC) for Distinguished Young Scholars (30325012). The costs of publication of this article...

The atomic coordinates and structure factors of the cryo and room temperature structures of cystatin D have been deposited to the Research Collaboratory for Structural Bioinformatics Protein Databank = PDB #1ROA andPDB#1RN7, respectively.

¶ To whom correspondence should be addressed. E-mail: Magnus.Abrahamson@klinkem.lu.se (M.A.) or su-xd@pku.edu.cn (X.-D.S.).

[1] The abbreviations used are: CEW cystatin, chicken egg-white cystatin; SDS-PAGE, SDS-polyacrylamide gel electrophoresis; L1, first hairpin loop; L2, second hairpin loop; BSL, back-side loop; RMSD, root mean square deviation; RT, room temperature; PDB_id, Protein Database identification number; SEC; size exclusion chromatography.

[2] Research Collaboratory for Structural Bioinformatics Protein Databank = PDB #1CEW (7)

[3] Lu, G. (1998) in manuscript. Web server: http://bioinfo1.mbfys.lu.se/TOP/maps.html

[4] Research Collaboratory for Structural Bioinformatics Protein Databank = PDB #1G96 (12)

[5] Research Collaboratory for Structural Bioinformatics Protein Databank = PDB #1DVD (8)

[6] Research Collaboratory for Structural Bioinformatics Protein Databank = PDB #1STF (10)

[7] Research Collaboratory for Structural Bioinformatics Protein Databank = PDB # 1EQK (11)

[8] Human cystatin C numbering (51) is used for cystatin D and other cystatins in this paper.

[9] Research Collaboratory for Structural Bioinformatics Protein Databank = PDB #1HUC (52)

[10] GenBank = GenBank Accession Number AF015263 (53)

FIGURE LEGENDS

Fig. 1. The cystatin D structure and comparison with CEW cystatin. *A*. The amino acid sequence of the cystatin D form crystallized, recombinant human 'Arg26-cystatin D'. The recombinant protein has two extra residues (Ala-Pro) in the N-terminal, but is otherwise identical to one of the two natural forms of cystatin D, with Arg in position 26 of the 122-residue mature protein sequence. The 12 first residues in the N-terminal segment of the recombinant protein are shown in *grey*, to indicate the poor electron density for this segment. The secondary structure elements are indicated in *yellow* for α-helix and *blue* for β-sheet. The motifs known to be important for the inhibition of papain-like enzymes by other cystatins are marked by *red* boxes and some well conserved residues in these are indicated according to human cystatin C numbering. The putative legumain-binding site (BSL) is also indicated and the presence of an Asn residue in this loop is pointed out (*underlined*). The position of the residue varying due to a gene polymorphism (Cys/Arg) is marked by an *arrowhead*. ***B*.** Ribbon representation of the cryo structure of human cystatin D viewed from the front. The α-helix is marked in *yellow* and the β-sheet in *blue*. The three segments involved in papain binding, formed by the N-terminal segment (*N-term),* the first and second hairpin loops (*L1* and *L2*), are indicated. The 'back-side loop' (*BSL*) involved in legumain inhibition by other type 2 cystatins is also indicated. ***C*.** and ***D*.** Representation of the aligned structures of human cystatin D (in *magenta*) and CEW cystatin (in *cyan*) are viewed from the front and from the C-terminal end of the α-helix, respectively. ***E*.** The surface rendering overlapped with structures for charge distribution on CEW cystatin (*left*) and the cystatin D cryo structure model (*right*). Color scale on the top shows charge intensity as indicated by the values. The protease binding loops are labeled as in ***B***. The illustrations were made by the programs MOLMOL and GRASP.

Fig. 2. Alignment of cystatins with determined structures. The sequence alignment shown is based on a structural alignment performed by MAPS of the human cystatin D structure and those known for type 1 and 2 cystatins (human cystatin A, human cystatin B, CEW cystatin, human cystatin C from the dimer structure, and the plant cystatin, oryzacystatin). α-helices and β-sheets are

indicated in *yellow* and *blue*, respectively. The conserved papain-binding site is marked by boxes in *magenta*. The *red* asterisk indicates the position of the Asn residue, necessary for legumain inhibition. *Arrows* indicate the two conserved disulfide bridges in type 2 cystatins. Human cystatin C amino acid numbering was used with the letter "a" indicating residues inserted in the cystatin D structure compared to the other cystatins structures. The residues closest in space at structural alignment are aligned in this figure, but it should be noted that structural positions of cystatin D residues 37-37a-38 all deviate quite much from those for residues 37-38 in CEW cystatin and human cystatin C. Which of the three cystatin D residues that should be seen as the inserted one could therefore not be predicted with certainty (see Fig. 4). Similarly, in the larger L1 loop of cystatin D, residues 57a-58 correspond to residue 58 of the other two type 2 cystatins, with the latter located in a position intermediate to those of cystatin D residues 57a and 58 (see Fig. 3).

Fig. 3. The papain-binding sites of cystatin D and CEW cystatin. Stereo views of the aligned segments of human cystatin D (in *magenta*) and CEW cystatin (in *cyan*) involved in inhibition of C1 peptidases. ***A.*** Front view of the first hairpin loop, L1. ***B.*** The second hairpin loop, L2, view from the top (N-terminal end) of the aligned α-helices. Labels are in the corresponding color.

Fig. 4. The putative legumain-binding site of cystatin D and CEW cystatin. Stereo views of the 'back-side loops' of cystatin D (in *magenta*) and CEW cystatin (in *cyan*). ***A.*** Side view of the α-helix end. ***B.*** View from the bottom of the α-helix. Residue labels are in the same color as for the corresponding protein.

Web figure. The B-factor plot for the RT (*in red*) and cryo (*in green*) structures of cystatin D. The conserved segments involved in the papain-binding ability are indicated (*red arrows*). Numbering according to the respective structures deposited to the Research Collaboratory for Structural Bioinformatics Protein Databank= PDB #1RN7 and PDB#1ROA.

Table I

Data collection and refinement statistics

Parameter	Crystal I	Crystal II
Detection type	Room temperature	Cryo (about 100K)
	$\lambda = 1.5418$	$\lambda = 0.9979$
Space group	$P2_12_12$	$P2_12_12$
Unit cell parameters	$a = 34.90$	$a = 34.05$
	$b = 84.37$	$b = 81.72$
	$c = 47.65$	$c = 46.74$
	$\alpha = \beta = \gamma = 90°$	$\alpha = \beta = \gamma = 90°$
No. waters	32	85
No. non-H protein atoms	912	907
Diffraction limit (Å)	30-2.5	30-1.8
Mosaicity (from Denzo)	0.433	0.366
Solvent content (%)	55.4	51.9
R-merge (%)	7.8 (25.2)	5.8 (34.2)
$I/\sigma(I)$	25.1 (8.0)	17.5 (3.0)
No. unique reflections	4955	12036
Completeness (%)	94.5 (96.8)	94.9 (98.1)
R_{free}	0.229	0.278
R_{conv}	0.191	0.250
Averaged B-factor (Å²)		
All atoms	44.27	33.47
Main chain	38.92	27.77
Side chains	49.06	37.33
Solvent	48.44	42.34
Ramachandran plot statistics (%)		
Most favored region	87	90
Additional allowed region	11	9
Generously allowed region	1	0
Disallowed region	1	1

Figure 1

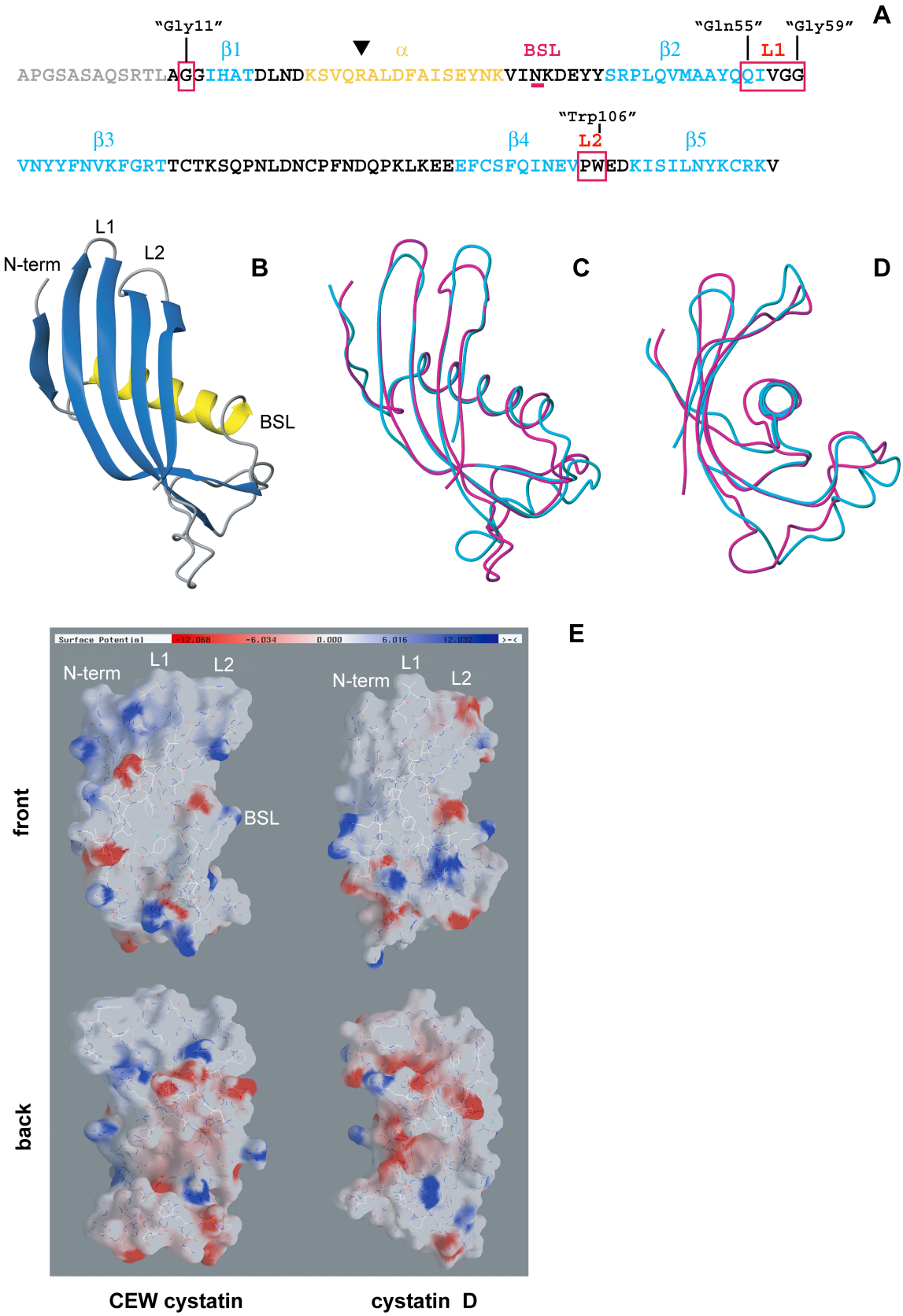

A
"Gly11"
"Gln55"
"Gly59"
β1
α
BSL
β2
L1
APGSASAQSRTLAGGIHATDLNDKSVQRALDFAISEYNKVINKDEYYSRPLQVMAAYQQIVGG
"Trp106"
β3
β4
L2
β5
VNYYFNVKFGRTTCTKSQPNLDNCPFNDQPKLKEEEFCSFQINEVPWEDKISILNYKCRKV
L1
N-term
L2
B
BSL
C
D
E
Surface Potential
-12.068
-6.034
0.000
6.016
12.032
N-term
L1
L2
BSL
front
back
CEW cystatin
cystatin D

Figure 2

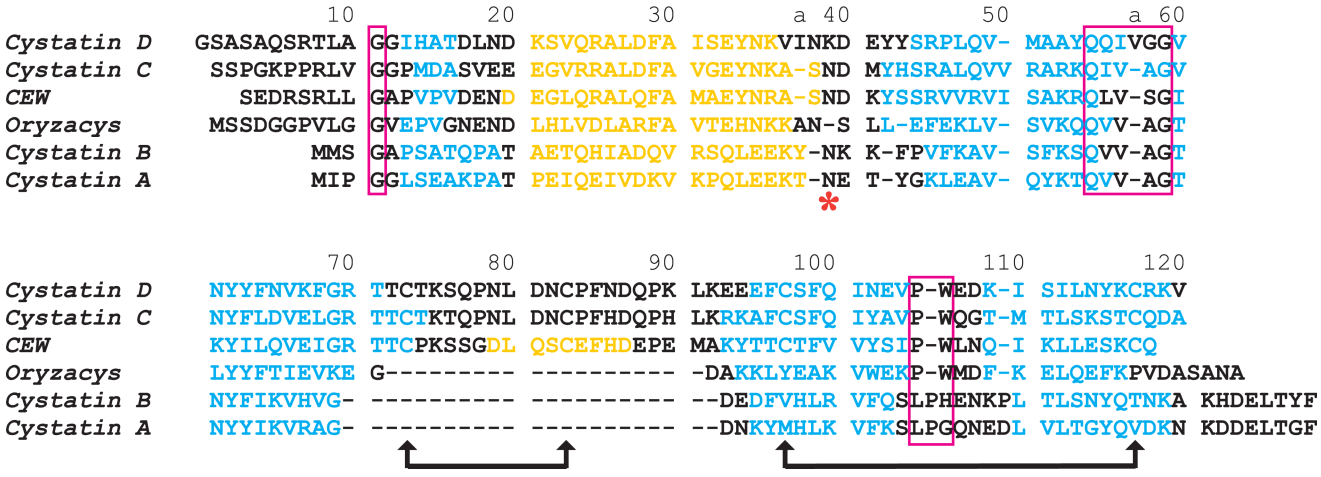

10 20 30 a 40 50 a 60
Cystatin D GSASAQSRTLA GGIHATDLND KSVQRALDFA ISEYNKVINKD EYYSRPLQV- MAAYQQIVGGV
Cystatin C SSPGKPPRLV GGPMDASVEE EGVRRALDFA VGEYNKA-SND MYHSRALQVV RARKQIV-AGV
CEW SEDRSRLL GAPVPVDEND EGLQRALQFA MAEYNRA-SND KYSSRVVRVI SAKRQLV-SGI
Oryzacys MSSDGGPVLG GVEPVGNEND LHLVDLARFA VTEHNKKAN-S LL-EFEKLV- SVKQQVV-AGT
Cystatin B MMS GAPSATQPAT AETQHIADQV RSQLEEKY-NK K-FPVFKAV- SFKSQVV-AGT
Cystatin A MIP GGLSEAKPAT PEIQEIVDKV KPQLEEKT-NE T-YGKLEAV- QYKTQVV-AGT
70 80 90 100 110 120
Cystatin D NYYFNVKFGR TTCTKSQPNL DNCPFNDQPK LKEEEFCSFQ INEVP-WEDK-I SILNYKCRKV
Cystatin C NYFLDVELGR TTCTKTQPNL DNCPFHDQPH LKRKAFCSFQ IYAVP-WQGT-M TLSKSTCQDA
CEW KYILQVEIGR TTCPKSSGDL QSCEFHDEPE MAKYTTCTFV VYSIP-WLNQ-I KLLESKCQ
Oryzacys LYYFTIEVKE G---------- ---------- -DAKKLYEAK VWEKP-WMDF-K ELQEFKPVDASANA
Cystatin B NYFIKVHVG- ---------- ---------- --DEDFVHLR VFQSLPHENKPL TLSNYQTNKA KHDELTYF
Cystatin A NYYIKVRAG- ---------- ---------- --DNKYMHLK VFKSLPGQNEDL VLTGYQVDKN KDDELTGF

Figure 3

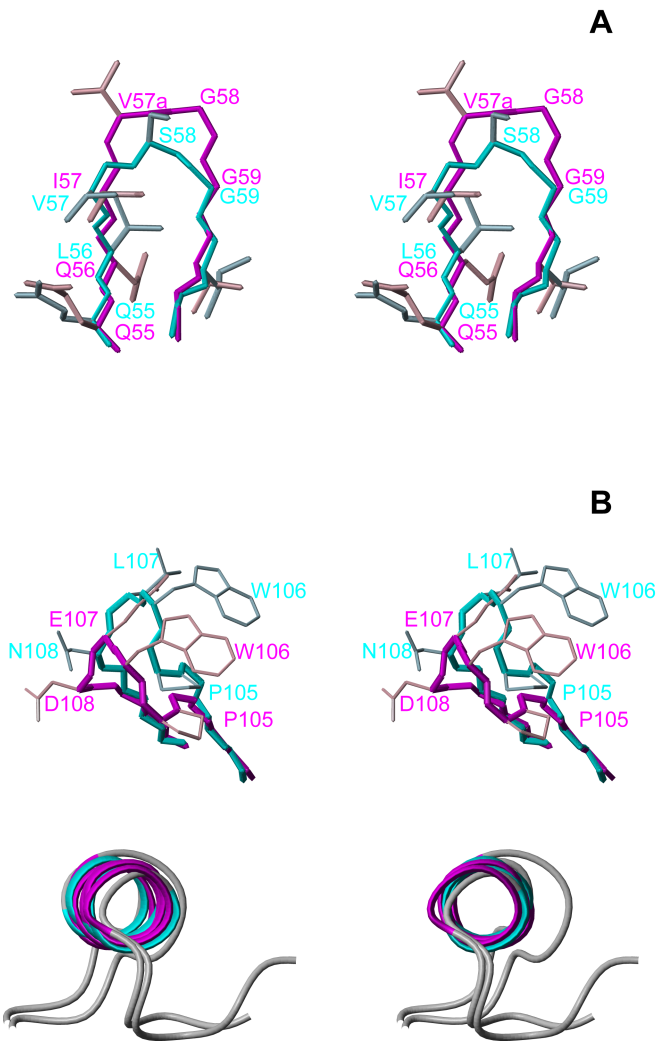
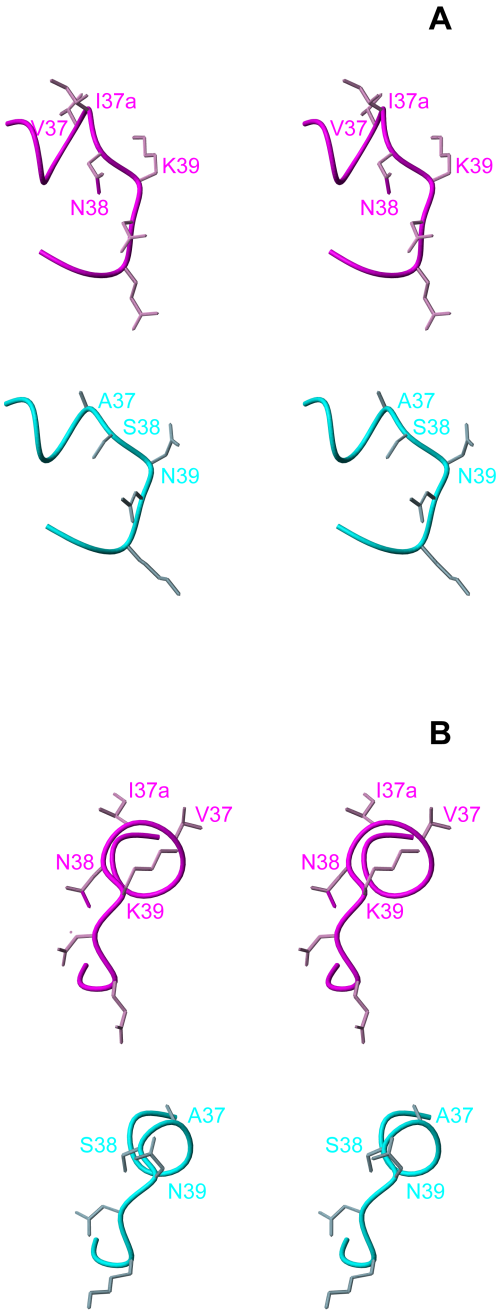
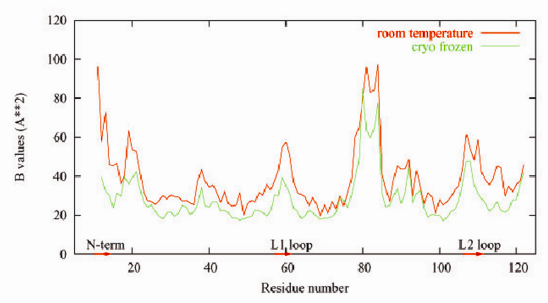

A
V57a
G58
S58
I57
V57
G59
G59
L56
Q56
Q55
Q55
V57a
G58
S58
I57
V57
G59
G59
L56
Q56
Q55
Q55
B
L107
W106
E107
N108
W106
P105
D108
P105
L107
W106
E107
N108
W106
P105
D108
P105

Figure 4


A
I37a
V37
K39
N38
I37a
V37
K39
N38
A37
S38
N39
A37
S38
N39
B
I37a
V37
N38
K39
I37a
V37
N38
K39
A37
S38
N39
A37
S38
N39

Web figure


room temperature
cryo frozen
B values (A**2)
Residue number
N-term
L1 loop
L2 loop
0
20
40
60
80
100
120