

## Structural and functional analysis of cystatin E reveals enzymologically relevant dimer and amyloid fibril states

Elfriede Dall<sup>1</sup>, Julia C. Hollerweger<sup>1</sup>, Sven O. Dahms<sup>1</sup>, Haissi Cui<sup>2</sup>, Katharina Häußermann<sup>2</sup>, and Hans Brandstetter<sup>1\*</sup>

From the <sup>1</sup>Department of Biosciences, University of Salzburg, A-5020 Salzburg, Austria; <sup>2</sup>Center for Integrated Protein Science Munich, Technical University of Munich, D-85748, Germany

Running title: *Cystatin E transforms to multimers with distinct functions*

\*To whom correspondence should be addressed: Hans Brandstetter: Department of Biosciences, University of Salzburg, A-5020 Salzburg, Austria; [hans.brandstetter@sbg.ac.at](mailto:hans.brandstetter@sbg.ac.at); Tel. +43 (662)8044-7278; Fax. +43 (662)8044-7209.

**Keywords:** enzyme inhibitor, conformational change, amyloid, cysteine protease, protein stability

---

### ABSTRACT

Protein activity is often regulated by altering the oligomerization state. One mechanism of multimerization involves domain swapping, wherein proteins exchange parts of their structures and thereby form long-lived dimers or multimers. Domain swapping has been specifically observed in amyloidogenic proteins, for example the cystatin superfamily of cysteine protease inhibitors. Cystatins are twin-headed inhibitors, simultaneously targeting the lysosomal cathepsins and legumain, with important roles in cancer progression and Alzheimer's disease. Although cystatin E is the most potent legumain inhibitor identified so far, nothing is known about its propensity to oligomerize. In this study we show that conformational destabilization of cystatin E leads to the formation of a domain-swapped dimer with increased conformational stability. This dimer was active as a legumain inhibitor by forming a trimeric complex. By contrast, the binding sites towards papain-like proteases were buried within the cystatin E dimer. We also showed that the dimers could further convert to

amyloid fibrils. Unexpectedly, cystatin E amyloid fibrils contained functional protein, which inhibited both legumain and papain-like enzymes. Fibril formation was further regulated by glycosylation. We speculate that cystatin amyloid fibrils might serve as a binding platform to stabilize the pH-sensitive legumain and cathepsins in the extracellular environment, contributing to their physiological and pathological functions.

---

Cysteine proteases are key regulators in many physiological processes. By consequence, dysregulation of protease activity can have severe effects resulting in a variety of pathologies, including cancer and Alzheimer's disease (1,2). Therefore, proteases must be regulated delicately and on different levels. Cystatins are inhibitors specifically controlling the activity of cysteine proteases. The cystatin superfamily shares a common fold that is characterized by a 5-stranded antiparallel  $\beta$ -sheet that is enwrapping a central  $\alpha$ -helix. It is further organized into three subfamilies:

(i) stefins, (ii) cystatins, and (iii) kininogens, fetuins and non-inhibitory cystatins (3,4). Family 1 cystatins (stefins) are mainly localized intracellularly and ubiquitously expressed in most cell types (3,5). They are potent inhibitors of papain (family C1) and the papain-like cathepsins with  $K_i$  values in the low nM range (6,7). The interaction of stefins with papain is mediated by a tripartite wedge-shaped structure formed by the N-terminus (Ser1-Val10, cystatin C numbering) and two hairpin loops (loop L1 and L2). Essentially, the N-terminus binds to the non-primed side, while the two adjacent hairpin loops occupy the primed-substrate binding sites.

Family 2 cystatins resemble the largest subfamily of the cystatin-fold with seven members identified so far. Contrasting the stefins, selected family 2 cystatins (C, E/M, F) harbor, in addition to their papain-binding site, a legumain binding site (8-10). Human legumain is a caspase-like cysteine protease (family C13) that mainly localizes to the endo-lysosomal system, where it plays an important function for the processing of antigens for presentation on the MHCII complex (11). On a pathophysiological level, legumain has been implicated in various disorders, including cancers and Alzheimer's disease (12-14). Under these conditions legumain was found translocated to the nucleus, cytoplasm and extracellularly. Because of its strict specificity for cleaving after asparagine residues, it is synonymously referred to as the asparaginyl-endopeptidase (AEP) (15,16). This strict preference is exploited by the legumain inhibitory cystatins C, E and F, which use a conserved Asn39 residue, localized on a reactive center loop different from the papain inhibitory site to specifically bind to the legumain active site (9,17). Furthermore, the interaction with legumain involves an additional legumain exosite loop (LEL) inserted between cystatin strands  $\beta_3$  and  $\beta_4$ . Complex formation leads to conformational stabilization of the pH-sensitive legumain at near neutral pH. Unlike family 1 cystatins, legumain inhibitory cystatins are secreted outside the cell and are in some cases glycosylated (10,18-20). While cystatin C is ubiquitously expressed in different human tissues, cystatin E/M is mainly localized to skin epithelia, emphasizing its role in cutaneous biology (5,10,21). Co-localization of human cystatin E (hCE) and legumain has been reported in hair follicles (22).

Cystatins not only encode a high intrinsic variability because of their function as dual protease inhibitors but also because of their ability to transform to distinct oligomerization states upon conformational destabilization. Factors triggering this oligomerization include N-terminal truncation by proteolytic enzymes, acidic pH, heating and point mutations. These cause dimer formation via a domain swapping mechanism (23-25). Essentially, the N-terminal segment, comprising  $\beta_1$ ,  $\alpha$ ,  $\beta_2$  up to the L1-loop, of one monomer exchanges with that of a second monomer (26). Consequently, the papain inhibitory site becomes inaccessible, whereas the legumain inhibitory site remains intact. Cystatin C oligomerization leads to the formation of amyloid deposits in the brain at advanced age (25). A naturally occurring L68Q variant was identified in the cerebral fluid of patients suffering from hereditary cystatin C angiopathy (HCC, Iceland disease), which accelerates this process significantly (6,27). Similarly, N-terminally truncated cystatin C, lacking the first 10 amino acids of the native sequence, was isolated from cystatin C amyloid deposits (28). This truncation was associated with proteolytic processing by proteases released to the cerebrospinal fluid and similarly results in accelerated formation of amyloid depositions (29). Stefin B was also reported to form amyloid fibrils and is an  $A\beta$ -binding protein and therefore supposed to play a role in Alzheimer's disease (30-32).

Both legumain and cystatins became attractive drug targets due to their relevance in different types of cancer and dementia. Among the cystatins, the family two cystatins became especially interesting, because of their function as dual protease inhibitors and because they are secreted to the extracellular space, where legumain and cathepsins are similarly observed under pathophysiologic conditions. Cystatin E is the most potent physiologic legumain inhibitor, binding 100fold more tightly as compared to cystatin C (7). Thereby it is associated with a tumor suppressor function in prostate cancer, melanoma and oral carcinoma cells (33-35). Furthermore, cystatin E has been observed co-localized with legumain in the extracellular environment under normal and under pathophysiologic conditions (22,36). Notably, not only co-localization but also co-trafficking of

legumain together with cystatin E from outside the cell to inside a cell has been reported (37).

However, cystatin E's physico-chemical properties including its propensity to form dimers and higher oligomers were hardly studied so far. Therefore, we set out to challenge the conformation of cystatin E by different triggering factors including N-terminal truncation, pH, and heating. Furthermore, we analyzed the effect thereof on its 3d structure and inhibitory function.

## Results

### *Cystatin E forms a dimer under destabilizing conditions*

To mimic the effect of N-terminal proteolytic processing, we recombinantly produced an N-terminally truncated human cystatin E ( $\Delta$ hCE) variant, lacking the first 10 amino acids following the mature N-terminus ( $\Delta$ (Arg4-Leu13)hCE;  $\Delta$ hCE; Fig. S1). Size exclusion experiments (SEC) revealed that while full-length, wild-type hCE migrated at the expected elution volume of the monomeric 15 kDa protein, the N-terminally truncated  $\Delta$ hCE variant showed an additional peak eluting at the size of a dimer (Fig. 1A). This observation was already a first indication for a similar tendency of hCE to oligomerize as described for other family members. As this oligomerization tendency correlated with conformational destabilizations, we next investigated the effect of heating on full-length hCE. Interestingly, we found that cystatin E monomer efficiently converted to a dimeric form upon incubation at  $\geq 70$  °C (Fig. 1A,B). Using SEC we could determine the transition temperature where 50% conversion was reached, to be 65 °C (Fig. 1B). This was further cross-confirmed by Thermofluor experiments which revealed a melting temperature of monomeric hCE of 65 °C (Fig. 1C, black curve). In this particular case, thermally induced increase in fluorescence at 65 °C corresponded to the transition to dimeric hCE, rather than to protein unfolding. To highlight the relevance of these experiments, we used cystatin C as a standard to compare the thermal energy barriers for dimer formation between the two family members. Significantly, cystatin C showed qualitatively and quantitatively the same behavior as hCE with 50% conversion to dimeric hCC at 65 °C and 100% conversion at 70 °C (Fig. 1B). Similarly, we tested the effect of pH on hCE dimerization. Indeed, incubation of monomeric

hCE at pH 3.0 could significantly reduce the transition temperature. SEC experiments revealed efficient conversion to dimeric hCE already at 60 °C (Fig. 1D). Additionally, using the Thermofluor method we observed a reduction of the transition temperature by 8 °C upon incubation at pH 3.5 (Fig. 1C, grey curve).

### *The cystatin E dimer is the thermodynamically preferred conformation*

Based on these experiments we concluded that conversion of monomeric to dimeric hCE is associated with an energy barrier that needs to be overcome. The energy barrier can be reduced by destabilizing conditions like N-terminal truncation, low pH and heating. Similarly, long term incubation at ambient temperature will have the same effect. To test whether the hCE dimer is a stable folding state by itself, we examined melting curves of monomeric and dimeric hCE (Fig. 1C). Dimeric hCE was generated by incubation of the monomer at 80 °C. Interestingly, the dimer showed an increased melting temperature, indicating that once the energy burden of dimerization has been overcome, the dimer is the structurally more stable conformation. Additionally, we found the monomer-dimer transition to be irreversible, i.e. dimeric hCE could not be converted back to monomeric hCE. Monomeric hCE thus represents a metastable state which is kinetically preferred in the folding process over the more stable dimeric hCE. Together, these observations were consistent with a cystatin C-like mode of dimerization, which is mediated by domain swapping (26).

### *The cystatin E dimer forms via domain swapping*

To unveil the molecular mechanism of hCE dimerization, we determined the crystal structure of dimeric hCE which was preincubated at 80 °C. Cystatin E crystals diffracted to 2.9 Å resolution and revealed dimer formation via domain swapping (Fig. 2A,B and Table S1). The N-terminal region  $\beta$ 1- $\alpha$ 1- $\beta$ 2-L1 of one cystatin E molecule (hCE) swapped out and integrated at the equivalent position of a second cystatin E molecule (hCE') and *vice versa*. To allow this sub-domain movement, loop L1, which connects  $\beta$ -strands 2 and 3 in the monomer, adopted an extended conformation. Two elongated  $\beta$ -strands  $\beta$ II and  $\beta$ III composed of the monomeric  $\beta$ 2-L1- $\beta$ 3 of monomer A and B, respectively, aligned in an

antiparallel fashion and thereby bridged the two monomers (Fig. 2B,C and Fig. S2A,B). The monomeric loops L1 and L1' of either hCE are integrated within the resulting extended  $\beta$ -sheet, with the local symmetry dyad passing through them (Fig. 2C). Consequently, dimeric hCE has integrated four additional L1, L1'-derived hydrogen bonds, contributing to its higher thermal stability and energetically more stable conformation as compared to the monomeric structures (Fig. 2C). Additionally, the monomer-dimer transition is accompanied by a conformational relaxation of the strictly conserved Val57, relaxing the unfavorable monomer conformation,  $(\phi|\psi) = (-121^\circ|-144^\circ)$ , to a favorable dimer conformation,  $(\phi|\psi) = (-140^\circ|137^\circ)$  (Fig. S1; pdb entry 4n6l) (38). Together, the L1-L1'  $\beta$ -sheet formed a hinge region connecting two hCE monomers.

Analysis with PDBePISA (39) and PDBsum (40) revealed that the hCE dimer interface is built up by 45 hydrogen bonds, 361 non-bonded contacts and 2 charge driven interactions. This results in a buried surface area (BSA) of the dimer interface of 6710  $\text{\AA}^2$  and an interface area of approx. 3354  $\text{\AA}^2$ . The high content in non-bonded contacts is also reflected in a very low solvation free energy gain  $\Delta^1G$  upon complex formation of -43.5 kcal/M. Furthermore the free energy of complex dissociation  $\Delta G^{\text{diss}}$  of 55.4 kcal/M indicates that the dimer is thermodynamically stable.

### ***The cystatin E dimer adopts a structurally distinct conformation***

From a structural perspective, the hCE dimer was assembled from two hCE monomer conformers that are connected via their L1 loops. When we compared the structural similarity of the two half domains hCEE' and hCE'E within the dimer, we observed a relatively low root-mean-square-deviation of the  $C^\alpha$  positions (rmsd = 0.33  $\text{\AA}$ ). The deviations between the two subdomains may relate to different packing environments within the crystal lattice. Interestingly, superposition of dimeric hCE with monomeric hCE revealed an average r.m.s deviation of 0.66  $\text{\AA}$ , indicating significant structural differences in hCE monomer as compared to the corresponding dimeric subdomain (Fig. 2A). Structural differences accumulated from a number of small rearrangements that accompanied the integration of the N-terminal  $\beta 1$ - $\alpha 1$ - $\beta 2$  region from one

molecule into the second. Therefore, the cystatin E monomer and dimer are two structurally distinct sub-species; the dimer is not the sole combination of two monomers.

When we superposed hCE and hCC monomer structures, we found an r.m.s. deviation of the  $C^\alpha$  positions of 1.3  $\text{\AA}$  (Fig. S2C). Interestingly, the difference between the hCE and hCC dimer structures is increased by two-fold (r.m.s.d. 2.8; Fig. S2D). Most of this high difference can be accounted for by the flexibility of the connecting hinge region (L1 – L1'). Small structural changes within this region translate into large motions in the peripheral regions, e.g. the legumain binding site. Given the flexibility of the connecting hinge, differences in crystal packing may also contribute to differences in the relative spatial orientation of the two subdomains, both in the hCE and hCC dimer structures (25). Therefore, we analyzed the crystal packing of the hCE dimer crystals. Remarkably, we discovered the assembly of two hCE dimers to a tetramer (Fig. S2E). A similar assembly was previously described for stefin B (pdb 2oct) and cystatin C (pdb 1g96). The stefin B tetramer forms via a hand shaking mechanism of the L2 loops that is triggered by a proline switch (Pro105 cystatin C numbering; Fig. S2F). Pro105 is conserved throughout the whole cystatin family. However, the L2 loop is shortened by 3 amino acids in the family 2 cystatins (Fig. S1). Hence, a direct handshake between two L2 loops is not possible. Nevertheless, the overall assembly (nearly parallel  $\beta 2$ - $\beta 3$  strands) is similar between stefin B, cystatin C and cystatin E and therefore suggests a conserved mechanism of oligomerization. The dimers interact via hydrophobic contacts involving residues of the former L1 loops. The hCE tetramer is visible in the crystal structure, however not in SEC experiments, indicating a low affinity of this assembly.

### ***The cystatin E dimer is a functional legumain inhibitor***

The hCE dimer is a unique folding state that encodes for two legumain binding sites on one molecule. In contrast to monomeric hCE, the legumain binding sites of the dimer are built up by contributions of both hCE chains, e.g. legumain binding site 1 is built up from the RCL' (reactive center loop) of one chain, hCE', and the legumain exosite loop (LEL) of the second chain, hCE, and

*vice versa* (Fig. 2B, 3A). Despite this large structural rearrangement, the local conformations of the legumain RCL and LEL were virtually identical in monomeric and dimeric hCE structures (Fig. 3B). Thus, we hypothesized that dimeric hCE is still functional as a legumain inhibitor. Indeed, when we incubated legumain with dimeric hCE, we observed a complete loss of its Ala-Ala-Asn-AMC substrate turnover (Fig. 3C). The crystal structure of dimeric hCE uncovered two symmetric and functional legumain binding sites, either of which is accessible for legumain binding. To test whether a simultaneous binding of two legumain molecules by one hCE dimer, i.e. a tetrameric assembly is sterically feasible we prepared a model of such a complex based on the crystal structures of the hCE dimer and the legumain-hCE complex (pdb entry 4n6o (17)). First, we superposed the hCE dimer onto the legumain-hCE complex (Fig. 3D). We found virtually identical intermolecular contacts for either of the two possible docking modes, i.e. docking legumain to the hCEE' or to the hCE'E sub-molecule was equally well accessible. However, from these docking models it became immediately clear that simultaneous binding to both sites is sterically impossible. When legumain binds to the hCEE' site, the free hCE'E site is in close proximity to the legumain insertion loop on legumain and *vice versa*. Hence, a trimer composed of one hCE dimer and one legumain (AEP) seemed to be the most likely assembly (Fig. 3D). To test this hypothesis, we performed size exclusion experiments of a preformed legumain-hCE dimer complex. Indeed, we observed a peak at the size of a trimer, but not at the expected size of a tetramer (Fig. 3E). Additionally, SDS-PAGE revealed a two-fold higher hCE concentration in the fractions of the legumain-hCE dimer complex (AEP-hCE<sub>d</sub>) as compared to legumain-hCE monomer complex (AEP-hCE<sub>m</sub>; Fig. 3F). The legumain concentration was identical in both experiments. The difference in hCE concentration within the fractions containing the hCE-legumain complex indicates a 2:1 stoichiometry rather than 2:2. Thus, cystatin E virtually loses 50% of its legumain binding sites via domain swapping. The observation of symmetric legumain binding sites within the cystatin E dimer suggested that binding to both sites is possible at the same affinity. Furthermore, structurally identical binding sites in the monomer and dimer implied

similar affinity constants towards legumain. Indeed, we could determine a  $K_I$  of  $10.7 \pm 5.6$  pM for monomeric hCE and  $13.5 \pm 6.7$  pM for dimeric hCE. Molar concentrations of monomeric and dimeric hCE were calculated assuming molecular weights of 15 and 30 kDa, respectively. Even though docking models of the hCE dimer – legumain complex showed some minor clashes between the unbound legumain binding site on the hCE dimer and legumain (Fig. 3D), these did not translate into weakened inhibition constants. (Mostly) unhindered complex formation is possible due to the flexible hinge region (L1 – L1').

Similar  $K_I$  constants of dimeric vs. monomeric cystatin E implied that the mode of binding of dimeric hCE to legumain is qualitatively similar to binding of monomeric hCE. To further test this conclusion, we investigated features that are characteristic for the hCE–legumain complex. Firstly, we could previously show that interaction of the LEL on cystatin E with the legumain primed substrate binding site has a positive effect on its thermal stability at near neutral pH, where the isolated enzyme by itself is unstable (17,41). The same is true for the legumain–hCE dimer complex (Fig. 3G). Secondly, we observed pH-dependent cleavage and religation of monomeric hCE at the conserved Asn39 residue located on the legumain reactive center loop of hCE by legumain (17). Cleavage happens at  $\text{pH} < 5.5$ , religation is observed upon incubation of cleaved hCE at  $\text{pH} > 6.5$ . We observed the same pH-dependent behavior when dimeric hCE was used (Fig. 3H). Together, these observations are in nice agreement with the crystal structure of dimeric hCE that uncovered structurally and functionally uncompromised legumain binding sites.

#### ***The cystatin E dimer is not functional as papain inhibitor***

Contrasting the legumain inhibitory site, the papain inhibitory site structurally differs in monomeric and dimeric hCE (Fig. 2A,B). Upon dimerization the L1 loop, which is essential for binding to the papain S1' site, got incorporated into the newly formed  $\beta$ II- $\beta$ III-sheet. Thus, loop L1 is not available for binding anymore and the papain-inhibitory site is destroyed by the dimer interface. In agreement with this observation, dimeric hCE did not show inhibition of papain and human cathepsin S in a peptide substrate

hydrolysis assay (Fig. 3C). Consequently, dimerization of cystatin E led to 100% loss of its cathepsin inhibitory function.

#### ***Glycosylation on the L2 loop is compatible with dimerization***

Native cystatin E harbors an N-glycosylation site on its L2 loop (Asn112). Since both glycosylated and non-glycosylated hCE were previously reported *in vivo*, we were interested in the relevance of glycosylation for hCE dimerization (20,42). The crystal structure of the hCE dimer suggested no negative effect of glycosylation on dimer formation (Fig. 4A). To test whether this is the case, we incubated glycosylated hCE produced in LEXSY (*Leishmania tarentolae* expression system) at 85 °C and subjected it to SEC. Indeed, we observed a shift corresponding to dimeric, glycosylated hCE (Fig. 4B). In agreement with the domain swapping mechanism, this dimer similarly inhibited legumain but not papain in a fluorescent peptide substrate assays (Fig. 4C). Furthermore, Thermofluor experiments of glycosylated hCE revealed a similar thermal unfolding behavior as observed for unglycosylated hCE (Fig. S3).

#### ***Cystatin E is resistant to hetero-dimerization***

Since the overall structures of stefin B, cystatin C and E dimers are superficially similar, we questioned whether dimerization may not be restricted to homo-dimer formation, but hetero-dimerization might be possible. To test this, we performed pull-down assays using a cystatin E construct carrying a C-terminal strep-tag and wild-type cystatin C. Both proteins were co-incubated at 70 °C to ensure efficient domain swapping. Subsequently, we immobilized hCE on a streptactin resin via its strep-tag. If hetero-dimerization had occurred we would have expected co-purification of the untagged hCC. However, using SDS-PAGE we could not detect co-migration of hCC, solely hCE homo-dimer had bound (Fig. S4). This observation can probably be understood by the low sequence identity of only 30% between cystatin E and C. Sequence variations on the  $\alpha$ 1-helix, which is a critically connecting element, may result in steric clashes, thereby preventing hetero-dimerization. Inspection of the hCE and hCC monomer structures revealed e.g. a <sup>hCE</sup>Ala29Phe<sup>hCC</sup> variation which may be in steric conflict with Ser<sup>114</sup>-Gln-Leu<sup>116</sup> ( $\beta$ 5) on hCE. Thus, the affinity for homo-dimer formation is

supposedly much higher than for hetero-dimerization. It might be possible to outcompete this affinity difference by having a large excess of hCC relative to hCE.

#### ***Cystatin E forms amyloid fibrils***

Cystatin E domain swapping results in the formation of an energetically more favorable conformation. In the simplest situation, two domain swapped monomers assemble to form a dimer. However, domain swapping is not restricted to dimerization, but in principle also allows the formation of multimers via concerted swapping reactions (chain reaction). The existence of such multimers was previously reported for stefin B and cystatin C and resulted in the assembly of highly ordered fibrils (26,27,30). Thermal denaturation curves of monomeric cystatin E revealed a first transition/partial unfolding between 60 and 70 °C which corresponded to dimer formation (Fig. 1C). Additionally, we observed a second transition between 80 and 90 °C both for monomeric and dimeric hCE, pointing towards the presence of another conformational state like fibrils, analogous to cystatin C. To test this hypothesis, we incubated cystatin E at a temperature above the second transition point (90 °C). Thereby, we observed the formation of an insoluble protein pellet. To test whether this pellet contains misfolded, aggregated protein or rather structured amyloid fibrils, we performed xray diffraction experiments. Interestingly, we observed two diffraction maxima at 10 Å and 4.7 Å resolution, which are indicative for cross beta structures, which typically build up amyloid fibrils (Fig. 5A). Furthermore, we could confirm the presence of fibrils using thioflavin T, a dye specifically binding to amyloid structures (Fig. 5B). Even more, negative staining TEM experiments also revealed the presence of fibril structures (Fig. S5). In agreement with thermal denaturation curves, fibril formation is more efficient at 90 °C as compared to 80 °C because the energy barrier is only partially overcome at 80 °C (Fig. 1C, 5B).

#### ***hCE fibrils contain functional protein***

Based on the assumption that fibril formation is a result of concerted domain swapping reactions, it seemed possible that hCE fibrils contained structured, functional protein. Since domain swapping did not compromise the functionality of

the legumain binding site, we tested whether hCE fibrils can, like dimeric hCE, inhibit legumain activity. Indeed, upon co-incubation of legumain with hCE fibrils, we observed complete inhibition of the Ala-Ala-Asn-AMC substrate turnover (Fig. 6A). This observation was a first strong indicator for the presence of folded protein within hCE fibrils. However, inhibition might have also been caused by unspecific precipitation mediated by insoluble hCE fibrils. To test whether the inhibition we observed is via specific interaction of hCE with the legumain active site, we performed a pull-down assay. Specifically, we co-incubated the insoluble hCE fibrils with active site-free and active site-blocked legumain variants. To block the active site we used a covalent Ac-Tyr-Val-Ala-Asp-chloromethylketone (YVAD-cmk) inhibitor. hCE fibrils are insoluble and for that reason served as stationary phase to extract potentially bound legumain. If the interaction of legumain with hCE fibrils was specific, via its active site, we would have expected to see binding of active site-free legumain but not of the YVAD-cmk-blocked variant. Indeed, we observed co-migration of wild-type legumain with hCE fibrils, while active site-blocked legumain remained in the soluble fraction (Fig. 6B). Furthermore, we observed specific processing of hCE fibrils at the Asn39 cleavage site on the legumain reactive center loop. Both results supported the presence of functional, folded protein within hCE fibrils. To test the structural integrity of the legumain exosite loop, we collected thermal denaturation curves of legumain pre-incubated with hCE fibrils. Analogous to monomeric and dimeric hCE, we observed a stabilizing effect of hCE fibrils on legumain at pH 6.5 (Fig. S6). The effect is not as pronounced as for the legumain – hCE (dimer) complex. However, this can be understood by the complexity of the fibrils.

#### ***hCE fibrils inhibit papain(-like) enzymes***

In principle, both monomeric and dimeric hCE are suitable building blocks for legumain-inhibitory hCE fibrils. However, the situation is different for papain. Only monomeric hCE is a functional papain inhibitor. Surprisingly, co-incubation of papain with hCE fibrils led to a complete inhibition of its enzymatic activity (Fig. 6A). Thus, hCE fibrils must contain, to some extent, monomeric hCE protein, which has not undergone domain swapping. Presumably, hCE fibrils are

built up by domain-swapped monomers, but are at the same time heterogeneously decorated with monomeric, wild-type protein (Fig. 10).

#### ***Dimeric hCE further converts to fibrils***

Thermal denaturation experiments uncovered that dimeric hCE has a higher thermal stability as compared to monomeric hCE. However, both species showed a transition at  $T > 80$  °C, indicating that dimeric hCE can similarly be converted to amyloid fibrils when the energy burden is significantly lowered, e.g. by further increase in temperature (Fig. 1C). Indeed, we found fibril formation of dimeric hCE upon incubation at 90 °C as evidenced in a ThT test (Fig. 5B). Therefore, the dimer is not a dead end, but most likely a folding intermediate on the route to multimers. Since both monomeric and dimeric hCE are capable of fibril formation, both might potentially serve as a building block for fibrils. However, based on Thermofluor experiments we would rather suggest that the dimer is the critical intermediate. To study this hypothesis, we set up a fibril nucleation assay where we tested the ability of monomeric and dimeric hCE to extend preformed fibrils. Specifically, we co-incubated monomeric and dimeric hCE with purified fibrils at 90 °C for 10 min. Interestingly, we observed an increase in ThT signal for dimeric hCE spiked with fibrils as compared to the control reaction without pre-addition of fibrils (Fig. 7). Thus, multimerization of dimeric hCE can be triggered by providing preformed fibrils as a “folding” template.

#### ***Fibril formation is pH-dependent and incompatible with glycosylation***

Domain swapping is triggered by conformational destabilization of monomeric hCE. Besides heating, we could identify low pH as a trigger factor for hCE dimerization. Similarly, we tested the effect of acidic pH on fibril formation and thereby found an about four-fold increase in ThT-signal following incubation at pH 3.0 (Fig. 8A). Furthermore, Thermofluor experiments revealed a reduction of the fibril-transition temperature by approx. 8 °C at pH 3.5 (Fig. 1C). Consequently, changing the pH environment will have an effect on the hCE oligomerization state. Along that line, we were also curious about the effect of glycosylation on hCE multimerization. Although dimerization is possible for the glycosylated

variant, multimerization might be negatively affected because of potential steric conflicts. Indeed, we did not observe fibril formation of glyco-hCE upon incubation at elevated temperatures (Fig. 8B,C). This suggested to us a regulatory function of glycosylation for hCE function.

### Discussion

Cystatin E is an intrinsically stable protein with a melting temperature  $> 60$  °C. However, destabilization does not lead to hCE denaturation as would be the case for most other proteins, but results in the transition to an energetically more favorable conformation (Fig. 9). This second folding state, the hCE dimer, represents a structurally distinct conformation with increased conformational stability. The energy barrier that needs to be overcome to allow for domain swapping can be reduced by triggering factors. Within this study, we could identify N-terminal truncation, acidification and heating as accelerators of dimerization (Fig. 9). Similarly, proteolytic processing has previously been reported for hCC, presumably catalyzed by leucocyte elastase, and led to faster dimerization and amyloid deposit formation (25). Along that line, a different but similar strategy of controlling the oligomerization state via proteolytic processing evolved in another member of the family 2 cystatins, cystatin F. Cystatin F harbors two additional cysteine residues which form two intermolecular disulfide bonds and thereby connect two cystatin F monomers. That way, the papain inhibitory site is blocked in an inhibitory latent dimeric cystatin F. However, the legumain inhibitory site is freely accessible. Activation of cystatin F is mediated by disulfide reduction or proteolysis in the N-terminal region (19,43). Although nothing is known so far about proteolytic processing of hCE *in vivo*, a similar event seems likely because of the flexible loop structure of the N-terminal region, which makes it prone to proteolytic processing. *In vitro*, we could previously demonstrate processing of hCC and hCE on their legumain RCL by legumain (17). Similarly, destabilization on this site may have a negative effect on the stability of monomeric and dimeric hCE. Interestingly, this cleavage is most efficient at acidic pH, where dimerization is also more easily achievable. Acidification is a common phenomenon in the brain of patients suffering

from dementia, emphasizing its relevance for domain swapping and consequent aggregation (44,45). Likewise, the L68Q hCC variant found in the cerebral fluid of patients suffering from Iceland disease has a higher tendency to form dimers and multimers. Replacement of the hydrophobic Leu68 by the hydrophilic glutamine causes local destabilization of the monomer, thereby also reducing the energy barrier for domain swapping (25). Such a destabilizing mutation has not yet been identified for hCE, however, comparable natural variations may exist. Domain swapping results in a complete loss of hCE's papain inhibitory activity and 50% loss of the legumain binding capacity. Consequently, domain swapping of cystatins in general will lead to an increase in protease activity due to loss of the inhibitors. Domain swapping thus provides another strategy to regulate proteolytic activity and can cause dysregulation of protease activity under pathological conditions. Legumain can stabilize cystatin E in its monomeric and dimeric state by binding to the actual conformation and thereby preventing further domain swapping. However, papain(-like enzymes) can only bind and retain the cystatin E monomer.

Although the conversion of monomeric to dimeric hCE is kinetically trapped, dimeric hCE can further convert to fibrils. Thus, domain swapping seems to be key for multimer formation. While the addition of preformed fibrils to dimeric hCE led to an increase in ThT fluorescence, monomeric hCE was unaffected. This is further confirming the relevance of the hCE dimer as a folding intermediate on the route to higher cystatin multimers. Additionally, it implies some prion-like behavior encoded in hCE. Since the hCE dimer has already been prone to domain swapping, it seems plausible that it is a better acceptor for a folding template than the monomer. However, it remains unclear whether monomeric hCE directly converts to fibrils if enough energy is supplied (1-step mechanism), or whether the dimer is formed first but immediately converted to fibrils (2-step mechanism; Fig. 10). A number of intermediate states have been described for other cystatins already, including a molten globule state or a tetramer (46,47). Therefore, the full picture presumably is more complex and not restricted to one intermediate.



Amyloid proteins are generally considered to lose their native conformation, while forming well-ordered cross-beta structures (48). This allows them to assemble into large structures containing many copies of the same molecule. A prominent example of such a protein is the A $\beta$ -peptide, which plays a central role in the pathology of Alzheimer's disease (49). Surprisingly, we could show that hCE amyloid fibrils contain biologically functional protein. While the ability of hCE fibrils to inhibit legumain is in agreement with both domain-swapped and monomeric hCE, papain inhibition is only possible with monomeric hCE. Hence we concluded that hCE amyloid fibrils contain domain swapped hCE as building block to fibril formation, however are heterogeneously decorated with monomeric hCE (Fig. 10). Domain swapping may not be 100% efficient explaining how some monomeric protein remains to be integrated into fibrils. Survival of monomeric protein is presumably time-dependent. Consequently, longer incubation may result in less monomeric protein and less papain inhibition. Based on these observations we assume that hCC fibrils also contain, to some extent, functional protease inhibitor.

Cystatin fibrils might serve as a binding platform to stabilize the pH-sensitive legumain and cathepsins in the extracellular environment, thereby increasing their lifetime. Since cystatins are reversible inhibitors, fibrils might provide a strategy of storing enzymes for later action. Albeit we did not observe hetero-dimerization between family 2 members, the formation of mixed fibrils seems possible. Protofibrils of hCE and hCC may assemble into mixed larger structures.

Unlike the family 1 cystatins, family 2 cystatins are frequently glycosylated. Human cystatin E harbours an N-glycosylation site on the L2-loop, which is compatible with dimerization but abolished fibril formation. From a structural perspective, modification on the L2-loop is in principle compatible with domain swapping, but in steric conflict with the formation of large fibrils, where proteins need to be packed together tightly. Interestingly, both glycosylated and unglycosylated hCE were observed *in vivo* e.g. in breast cancer cell lines or cystatin E overexpressing HEK293 cells, suggesting different (pathophysiological-) functions for both variants (10,20,37). While glycosylated hCE will only be present as monomer or dimer, unglycosylated

hCE can potentially also be converted to amyloid fibrils. Moreover, human cystatin C also harbours a predicted glycosylation site, but different to the one present in hCE. Cystatin C harbours an O-glycosylation site on its N-terminal region (Ser2). Glycosylation at this site will likely have an effect on dimerization, because this part of the cystatin molecule is directly involved in domain swapping. Indeed, there is indirect evidence from literature that N-terminal glycosylation is incompatible or at least restricting hCC dimerization (50). While cystatin C may lose the O-glycosylation site by N-terminal proteolytic cleavage, the hCE N-glycosylation is not affected because of its location on the L2 loop (28,51). However, (de-)glycosidases might play a role as regulatory enzymes, switching glycosylated hCE to unglycosylated hCE. N-terminally truncated cystatin C was isolated from cystatin C amyloid deposits and results in accelerated formation of amyloid depositions (29).

Recently, cystatin E was also identified in the cerebrum (52). Thus, it is attractive to speculate about a potential role of domain-swapped cystatin E in neuronal function.

## Experimental Procedures

### Preparation of proteins

Human wild-type cystatin E (hCE) and cystatin C (hCC) constructs lacking the N-terminal signal sequence were cloned into the pET22b(+) vector (Novagen) as described earlier (17). A truncated version of hCE ( $\Delta$ hCE) was prepared using a forward primer carrying an NcoI restriction site (5'-ATGCCCATGGAAGTGTGCGCCGACGACCCGCAGGTGC-3') and a reverse primer carrying an XhoI restriction site (5'-ACGTCTCGAGCATCTGCACACAGTTGTGC-3'). The construct has residues Arg4-Leu13 deleted and due to the employed cloning strategy starts with the double mutation R14M-D15E. Full-length wild-type hCE, N-terminally truncated  $\Delta$ hCE and hCC had the native signal peptide replaced by the pelB leader sequence present on the pet22b(+) vector to allow expression to the *E.coli* periplasmic space. During secretion the signal peptide was removed thereby liberating the new Arg4 N-terminus of full-length wild type hCE (hCC numbering), Met14 of the truncated  $\Delta$ hCE variant and Ser1 of hCC. Additionally, the

expression constructs carried a C-terminal His<sub>6</sub>-tag for purification. Furthermore, another wild-type cystatin E construct carrying a C-terminal His<sub>6</sub>-tag followed by a strep-tag was prepared. For that purpose, a wild-type cystatin E construct was used as a template and the strep-tag was introduced following a protocol based on the inverse-PCR method. As primers we used TCAGTTCGAAAAGTGAGATCCGGCTGCTA ACAAAGCCCGAAAGG (forward) and GGGTGTGACCAGTGGTGGTGGTGGTGGTG CTCGAGCATCTGC (reverse). Correctness of expression constructs was confirmed via DNA-sequencing by Eurofins MWG Operon (Martinsried, Germany).

Cystatin E and C constructs were expressed in *E.coli* B121(DE3) cells. Briefly, the expression plasmid was transformed into B121(DE3) cells. For large scale expression, cells were grown in 2 l flasks filled with 600 ml LB medium (Carl Roth, Karlsruhe, Germany) supplemented with 100 µg/ml ampicillin at 37 °C with agitation at 220 rev/min until an OD<sub>600</sub> of 0.8 – 1.0 was reached. Expression was induced at 25 °C by the addition of 1 mM IPTG (isopropyl β-d-1-thiogalactopyranoside). After overnight expression, cells were harvested by centrifugation (10 min, 4000 rpm, 4 °C) and frozen. The periplasmic fraction containing recombinant proteins was extracted by cold osmotic shock. Briefly, frozen cell pellets of 1.8 l expression culture were gently resuspended in 150 ml lysis buffer composed of 30 mM Tris pH 7.5 and 20% sucrose and stirred for 20 min at ambient temperature. The lysate was centrifuged at 17500 g for 20 min at 4 °C. The supernatant containing soluble periplasmic proteins was batch-incubated with Ni-NTA Superflow resin (Qiagen, Hilden, Germany) for 30 min at 4 °C. Following a washing step using a buffer composed of 20 mM Tris pH 7.5, 300 mM NaCl and 10 mM imidazole, bound protein was eluted with washing buffer containing 250 mM imidazole. Elutions were concentrated using Amicon Ultra centrifugal filter units (MWCO: 3 kDa; Millipore) and subjected to further purification via size exclusion chromatography (SEC) utilizing the Äkta FPLC system equipped with a Superdex 75 10/300 GL column (GE Healthcare) equilibrated in a buffer composed of 20 mM citric acid pH 5.5 and 100 mM NaCl.

Glycosylated cystatin E (glycol-hCE) was produced using the *Leishmania tarentolae* expression system (LEXSY) following a protocol described previously (17).

Human wild-type legumain was cloned, expressed, purified and activated as described earlier (53).

#### ***Determination of oligomerization state using SEC***

To analyze the oligomerization state of different hCE variants after different treatments, 250 µl of sample were loaded on a S75 10/300 GL column equilibrated in a buffer composed of 20 mM citric acid pH 5.5 and 100 mM NaCl. To test the effect of heating, wild-type cystatin E was incubated at 37 °C, 60 °C, 70 °C, and 85 °C for 10 min at a concentration of 0.5 mg/ml in a buffer composed of 20 mM citric acid pH 5.5 and 100 mM NaCl. After 10 min incubation, the samples were chilled on ice for further 20 min. N-terminally truncated cystatin E was incubated at 4 °C and 37 °C prior to injection and wild-type cystatin C at 60 °C and 70 °C. Additionally, wild-type cystatin E was incubated at 60 °C in citric acid buffer at pH 3.0 to test the effect of pH on dimerization. To investigate the effect of glycosylation on dimerization, hCE produced in LEXSY was incubated at 80 °C for 10 min and subsequently analyzed by SEC.

To test the stoichiometry of the legumain – dimeric hCE complex, we incubated legumain (0.5 mg/ml) with dimeric hCE at a molar ratio of 1:1 (1 AEP : 1 hCE dimer) at pH 5.5 for 10 min on ice before injection of the sample on a S200 10/300 GL column. The hCE dimer was prepared by incubation of monomeric hCE at 80 °C for 10 min. Similarly, legumain was incubated with monomeric hCE at a molar ratio of 1:2 (1 AEP : 2 hCE monomers) for 10 min at pH 5.5. To calculate molar concentrations of monomeric and dimeric hCE, a molecular weight of 15 kDa and 30 kDa, respectively, was assumed. Both samples contained equal amounts of cystatin E either in monomeric or dimeric state. For all samples investigated, fractions were collected and analyzed on SDS-PAGE.

#### ***Thermofluor assays***

Thermal denaturation curves of different protein variants after different treatments were determined using the thermofluor method (54). Briefly, 1 mg/ml protein sample containing 50x SYPRO

Orange (Invitrogen) was added in a 1:10 ratio to 22.5  $\mu\text{l}$  of assay buffer composed of 50 mM citric acid pH 5.5/3.0 and 100 mM NaCl. Thermal denaturation curves were collected in a 7500 Real Time PCR System (Applied Biosystems) from 20 to 95  $^{\circ}\text{C}$ . The fluorescence data was analyzed as described previously (55). Protein samples investigated were monomeric hCE and dimeric hCE, which was prepared by incubation of monomeric hCE at 80  $^{\circ}\text{C}$  for 10 min. To test the stability of AEP in complex with monomeric hCE and dimeric hCE, a complex was prepared at pH 5.5 by mixing AEP and inhibitor in a 1:1 molar ratio, assuming 30 kDa as the molecular weight of dimeric hCE. Stability of AEP only and in complex with hCE was assayed in a buffer composed of 50 mM MES pH 6.5 and 100 mM NaCl. To test the effect of hCE fibrils on AEP stability, 1  $\mu\text{l}$  washed fibrils were added to the assay buffer (50 mM MES pH 6.5, 100 mM NaCl) before addition of AEP premixed with SYPRO Orange. Fibril preparation is described in section “Co-precipitation assay”.

#### ***Inhibition assays***

Inhibition of wild-type legumain was tested in legumain assay buffer (50 mM citric acid pH 5.5, 100 mM NaCl) containing 100  $\mu\text{M}$  Z-Ala-Ala-Asn-7-amino-4-methylcoumarin substrate (Z-AAN-AMC; Bachem). Assays were carried out in an Infinite M200 Plate Reader (Tecan). Briefly, the assay buffer was preincubated with 8 nM cystatin followed by the addition of 4 nM legumain. Increase in fluorescence was measured at 460 nm upon excitation at 380 nm at 37  $^{\circ}\text{C}$ . Inhibition of papain (EC 3.4.22.2; Merck, Darmstadt) and recombinant human cathepsin S was assayed in the same assay buffer containing 100  $\mu\text{M}$  Z-Phe-Arg-7-amino-4-methylcoumarin substrate (Z-FR-AMC; Bachem). The assay buffer was preincubated with 100 nM cystatin, followed by the addition of 50 nM papain or cathepsin S and fluorescence was similarly recorded at 460 nm. Dimeric cystatin E was prepared by incubation of monomeric cystatin E at 80  $^{\circ}\text{C}$  for 10 min. The sample was filtered to remove higher oligomers. Similarly, glycosylated cystatin E was also incubated at 80  $^{\circ}\text{C}$  for 10 min to generate the dimeric form. All experiments were carried out in triplicate.

#### ***Determination of $K_i$ values***

Inhibition constants of monomeric and dimeric hCE towards legumain were determined in assay buffer composed of 50 mM citric acid pH 5.5, 100 mM NaCl and 0.05% Tween-20 using the Morrison equation for tight binding inhibitors (56). Firstly, the  $K_M$  value of legumain towards the Z-AAN-AMC substrate was determined in assay buffer containing 3  $\mu\text{M}$  – 250  $\mu\text{M}$  substrate. The reactions were started by the addition of 2 nM enzyme. The  $K_M$  was calculated to be 52  $\mu\text{M}$  under these assay conditions. In a next step, the assay buffer containing 100  $\mu\text{M}$  Z-AAN-AMC substrate was preincubated with increasing concentrations of hCE ranging from 0.01 to 5 nM and the reaction was started by the addition of 2 nM legumain. Fluorescence was monitored at 460 nm and 37  $^{\circ}\text{C}$  for 10 min. Velocity of substrate turnover was calculated as fluorescence units per second and the data points were fit to the Morrison equation using GraphPad Prism version 7.0 (GraphPad Software, La Jolla California USA, www.graphpad.com). To calculate molar concentrations of dimeric hCE, a molecular weight of 30 kDa was assumed. All measurements were performed in triplicates.

#### ***Crystallization and structure solution of dimeric hCE***

Dimeric hCE was prepared by incubation of 1 mg/ml wild-type hCE at 80  $^{\circ}\text{C}$  for 10 min. Subsequently, the protein sample was filtered to remove higher oligomers and subjected to SEC using a S75 10/300 GL column equilibrated in a buffer composed of 20 mM citric acid pH 5.5 and 50 mM NaCl. Fractions containing dimeric protein were concentrated to approx. 30 mg/ml final concentration using Amicon Ultra centrifugal filter units (MWCO: 10 kDa; Millipore). Initial crystallization screening was performed in a sitting-drop vapour diffusion setup. 0.4  $\mu\text{l}$  of concentrated dimeric hCE were mixed with 0.4  $\mu\text{l}$  screen solution (Hampton Index HT or JBScreen Classic) and equilibrated with 60  $\mu\text{l}$  reservoir solution in 96-well Intelli Plates (Art Robbins Instruments) at 20  $^{\circ}\text{C}$ . After 1 – 2 weeks, crystals were observed in a condition composed of 20% PEG 8000 and 0.1 M Ches pH 9.5. Crystals were harvested after stepwise addition of a cryo-protectant solution containing 22% PEG 8000, 0.1 M Ches pH 9.5 and 20% glycerol and flash-frozen in liquid nitrogen. A native dataset was collected at 100 K on beamline ID23-2 (ESRF, Grenoble)

equipped with a Pilatus3\_2M detector to a resolution of 2.8 Å. 720 images were collected at a wavelength of 0.8726 Å at 0.5 ° oscillation range and 0.1 s exposure time.

Data processing was performed utilizing iMOSFLM (57) and SCALA from the CCP4 program suite (58). An initial search model was prepared using the crystal structure of monomeric hCE (pdb entry code 4n6l) by removing the N-terminal region up to loop L1. This model was used as an initial search model for molecular replacement using PHASER (59). Repeated cycles of manual rebuilding in COOT (60) and refinement using phenix.refine (61) were carried out. The atomic coordinates and experimental structure factors have been deposited with the Protein Data Bank ([www.pdb.org](http://www.pdb.org)) under the entry code 6FK0. Pymol (62) was used to create figures illustrating structures.

### ***Molecular modelling***

A model of dimeric hCE in complex with legumain was created using Topmatch (63). Specifically, the crystal structure of the legumain-hCE complex (pdb entry code 4n6o) served as a template to align the structure of dimeric hCE.

### ***Proteolysis and ligation assay***

To test proteolysis at the P1-Asn39 residue on cystatin E, monomeric and dimeric cystatin E were incubated with legumain in a 1:2 molar ratio (1 legumain : 2 hCE; assuming a molecular weight of 15 kDa for both monomeric and dimeric hCE) at pH 4.0 at 37 °C until 80% turnover was observed as judged by SDS-PAGE. Subsequently, the pH was shifted to 7.5 and the samples were incubated for 1 hour at 37 °C. As control samples we used legumain only and monomeric/dimeric hCE only. Progress of proteolysis and ligation was investigated after different time points via SDS-PAGE. Dimeric hCE was prepared from N-terminally truncated hCE.

### ***X-ray diffraction to test amyloid fibril formation***

Monomeric hCE was incubated at 90 °C for 10 min at 20 mg/ml protein concentration. Subsequently, the insoluble fraction potentially containing amyloid fibrils was harvested by centrifugation (16000 g, 10 min, 4 °C). To wash off residual monomeric protein, the pellet was resuspended in ddH<sub>2</sub>O and again harvested by centrifugation. The supernatant was discarded and

washed another 2 times. The pellet was then mounted at the edge of a quartz glass capillary and diffraction was assayed in-house using a Bruker Microstar rotating anode generator mounted with a Mar345dtb detector.

### ***Testing pH dependence of fibril formation***

To test for formation of amyloid fibrils, cystatin E was incubated at 80 °C for 10 min at a protein concentration of 10 mg/ml in a buffer composed of 50 mM citric acid and 100 mM NaCl (pH range 3 – 6) or 50 mM Tris and 100 mM NaCl (pH 7.0). Afterwards, samples were put on ice for at least 10 min before testing fibril formation using a Thioflavin T (ThT) test. Briefly, a ThT stock solution was prepared by dissolving 8 mg Thioflavin T (Sigma Aldrich) in 10 ml PBS buffer. A working solution was prepared freshly each day by diluting the ThT stock solution 1:50 in PBS buffer. 24 µl of working solution were mixed with 1 µl of sample in a 386 well black polystyrene plate (CORNING) and fluorescence was measured for 2 min at 25 °C with excitation at 440 nm and emission at 482 nm. Mean values and standard deviations from three measurements were calculated.

### ***Testing the effect of glycosylation on fibril formation***

To test the effect of L2-loop-glycosylation of cystatin E on the formation of higher oligomers, we incubated cystatin E produced in LEXSY at 90 °C for 10 min at a concentration of 20 mg/ml in a buffer composed of 50 mM citric acid pH 5.5 and 100 mM NaCl to induce oligomerization. Subsequently, the sample was incubated on ice for at least 10 min before analysis in a ThT assay.

### ***Testing inhibition of legumain and papain by hCE fibrils***

Cystatin E fibrils were prepared by incubation of a 10 mg/ml sample at pH 3.0 and 80 °C for 10 min. The insoluble fraction was harvested by centrifugation (16000 g, 10 min), resuspended in washing buffer composed of 50 mM citric acid pH 5.5 and 100 mM NaCl and centrifuged again. This washing step was repeated 3 times. Afterwards the fibrils were resuspended in 30 µl wash buffer. To test inhibition of legumain and papain, the substrate solution described before was preincubated with 1 µl fibrils and the reaction started by the addition of the enzyme. In each

case, the final reaction volume was 50  $\mu$ l. Control reactions contained only the enzyme in substrate solution. All reactions were measured in triplicate.

#### ***Co-precipitation assay***

Cystatin E fibrils were prepared by incubation of monomeric hCE at 20 mg/ml concentration at 90 °C for 10 min. The insoluble fraction was harvested by centrifugation (16000 g, 10 min). To remove any residual monomeric or dimeric protein contaminants, the pellet was washed 3 times with a buffer composed of 50 mM citric acid pH 5.5, 100 mM NaCl. The final pellet was resuspended in 20  $\mu$ l wash buffer. 5  $\mu$ l fibrils were mixed with 5  $\mu$ l legumain (0.1 mg/ml final concentration) and incubated for 10 min at 20 °C. Afterwards the insoluble fraction was harvested by centrifugation and washed 2 times with washing buffer. Control reactions contained fibrils only, fibrils + AEP pre-complexed with the covalent Ac-Tyr-Val-Ala-Asp-chloromethylketone inhibitor (YVAD-cmk), and legumain only.

#### ***Testing fibril formation by addition of pre-formed fibrils***

Cystatin E fibrils were prepared and washed as described above. Dimeric hCE was prepared by incubation of monomeric protein at 80 °C for 10 min, followed by filtering. Monomeric and dimeric cystatin E (10  $\mu$ l, 10 mg/ml) were supplemented with 1  $\mu$ l cystatin E fibrils and incubated at 90 °C for 10 min. Control samples contained fibrils only or cystatin E monomer/dimer only. Afterwards, samples were incubated on ice for at least 10 min before setting up a ThT assay. The reaction was set up in a buffer composed of 50 mM citric acid pH 5.5 and 100 mM NaCl.

#### ***Cystatin E/C heterodimerization assay***

To test heterodimerization of cystatins E and C, cystatin E carrying a C-terminal strep-tag was co-incubated with cystatin C at 70 °C for 10 min.

Following further 10 min incubation on ice, the sample was loaded on a Strep-Tactin® Sepharose resin (iba, Germany) pre-equilibrated in wash buffer (100 mM Tris pH 8.0, 150 mM NaCl). The flow-through was collected and the resin washed 5 times with wash buffer. Finally, bound protein was eluted by applying 5 times 100  $\mu$ l elution buffer containing 100 mM Tris pH 8.0, 150 mM NaCl and 2.5 mM desthiobiotin. Control samples were cystatin E incubated at 37 °C (to confirm binding of monomeric protein; positive control), cystatin E incubated at 70 °C (to confirm binding of dimeric protein) and cystatin C incubated at 70 °C (to test unspecific binding of hCC; negative control). Fractions from different stages of purification were analyzed on SDS-PAGE.

#### ***Transmission electron microscopy***

Cystatin E amyloid fibrils were prepared from monomeric protein at a concentration of 20 mg/ml (in a buffer composed of 50 mM citric acid pH 5.5 and 100 mM NaCl) by incubation at 90 °C for 10 min. The insoluble fraction was harvested and washed as described above. Subsequently, the pellet containing hCE amyloid fibrils was resuspended in buffer composed of 20 mM MES pH 5.5 and 20 mM MgCl<sub>2</sub>. A 1:10 dilution was prepared and vortexed for 1 min. Subsequently, the solution was centrifuged for 10 min at room temperature and the supernatant was removed carefully and fixed by the addition of 0.1% glutaraldehyde (final concentration) for 10 min. 10  $\mu$ l sample were incubated on a Formvar carbon film coated 400 mesh copper grids (Electron Microscopy Sciences, München, Germany) grid for 2 min and stained with 2 % uranyl formiate solution containing 25 mM NaOH for 40 seconds. TEM imaging was performed using a Philips CM 100 transmission microscope operating at 100 kV. Images were acquired using an AMT 4 x 4 Megapixel CCD Camera. Imaging was performed at 28500 x magnification.

**Acknowledgements:** We would like to thank Hendrik Dietz (CIPSM), for providing access to the TEM-facility, Michael Groll (CIPSM) for support, Wait Tuck Soh for providing recombinant cathepsin S and Drs. Turk for valuable discussions. This work was supported by the Austrian Science Fund (FWF, project number W\_01213).

**Conflict of interest:** The authors declare that they have no conflicts of interest with the contents of this article.

**Author contributions:** E.D., J.C.H. and H.B. designed the experiments. K.H. and H.C. performed the TEM experiments. S.O.D. designed the x-ray experiment to test for fibril formation and contributed with valuable discussions. E.D. and J.C.H. performed all other experiments. E.D. and H.B. wrote the paper, and all authors reviewed the manuscript.

## References

1. Mason, S. D., and Joyce, J. A. (2011) Proteolytic networks in cancer. *Trends in cell biology* **21**, 228-237
2. Hasanbasic, S., Jahic, A., Karahmet, E., Sejranic, A., and Prnjavorac, B. (2016) The Role of Cysteine Protease in Alzheimer Disease. *Materia socio-medica* **28**, 235-238
3. Turk, V., and Bode, W. (1991) The cystatins: protein inhibitors of cysteine proteinases. *FEBS Lett* **285**, 213-219
4. Hedrich, J., Lottaz, D., Meyer, K., Yiallourous, I., Jahnen-Dechent, W., Stocker, W., and Becker-Pauly, C. (2010) Fetuin-A and cystatin C are endogenous inhibitors of human meprin metalloproteases. *Biochemistry* **49**, 8599-8607
5. Abrahamson, M., Barrett, A. J., Salvesen, G., and Grubb, A. (1986) Isolation of six cysteine proteinase inhibitors from human urine. Their physicochemical and enzyme kinetic properties and concentrations in biological fluids. *J Biol Chem* **261**, 11282-11289
6. Turk, V., Stoka, V., Vasiljeva, O., Renko, M., Sun, T., Turk, B., and Turk, D. Cysteine cathepsins: from structure, function and regulation to new frontiers. *Biochim Biophys Acta* **1824**, 68-88
7. Rawlings, N. D., Barrett, A. J., and Bateman, A. (2012) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* **40**, D343-350
8. Cheng, T., Hitomi, K., van Vlijmen-Willems, I. M., de Jongh, G. J., Yamamoto, K., Nishi, K., Watts, C., Reinheckel, T., Schalkwijk, J., and Zeeuwen, P. L. (2006) Cystatin M/E is a high affinity inhibitor of cathepsin V and cathepsin L by a reactive site that is distinct from the legumain-binding site. A novel clue for the role of cystatin M/E in epidermal cornification. *J Biol Chem* **281**, 15893-15899
9. Alvarez-Fernandez, M., Barrett, A. J., Gerhartz, B., Dando, P. M., Ni, J., and Abrahamson, M. (1999) Inhibition of mammalian legumain by some cystatins is due to a novel second reactive site. *J Biol Chem* **274**, 19195-19203
10. Ni, J., Abrahamson, M., Zhang, M., 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) Cystatin E is a novel human cysteine proteinase inhibitor with structural resemblance to family 2 cystatins. *J Biol Chem* **272**, 10853-10858
11. Manoury, B., Hewitt, E. W., Morrice, N., Dando, P. M., Barrett, A. J., and Watts, C. (1998) An asparaginyl endopeptidase processes a microbial antigen for class II MHC presentation. *Nature* **396**, 695-699
12. Haugen, M. H., Johansen, H. T., Pettersen, S. J., Solberg, R., Brix, K., Flatmark, K., and Maelandsmo, G. M. (2013) Nuclear legumain activity in colorectal cancer. *PloS one* **8**, e52980
13. Liu, C., Sun, C. Z., Huang, H. N., Janda, K., and Edgington, T. (2003) Overexpression of legumain in tumors is significant for invasion/metastasis and a candidate enzymatic target for prodrug therapy. *Cancer Res* **63**, 2957-2964
14. Zhang, Z., Obiany, O., Dall, E., Du, Y., Fu, H., Liu, X., Kang, S. S., Song, M., Yu, S. P., Cabrele, C., Schubert, M., Li, X., Wang, J. Z., Brandstetter, H., and Ye, K. (2017) Inhibition of delta-secretase improves cognitive functions in mouse models of Alzheimer's disease. *Nat Commun* **8**, 14740
15. Li, D. N., Matthews, S. P., Antoniou, A. N., Mazzeo, D., and Watts, C. (2003) Multistep autoactivation of asparaginyl endopeptidase in vitro and in vivo. *J Biol Chem* **278**, 38980-38990

16. Vidmar, R., Vizovisek, M., Turk, D., Turk, B., and Fonovic, M. (2017) Protease cleavage site fingerprinting by label-free in-gel degradomics reveals pH-dependent specificity switch of legumain. *The EMBO journal* **36**, 2455-2465
17. Dall, E., Fegg, J. C., Briza, P., and Brandstetter, H. (2015) Structure and mechanism of an aspartimide-dependent peptide ligase in human legumain. *Angew Chem Int Ed Engl* **54**, 2917-2921
18. Taupin, P., Ray, J., Fischer, W. H., Suhr, S. T., Hakansson, K., Grubb, A., and Gage, F. H. (2000) FGF-2-responsive neural stem cell proliferation requires CCg, a novel autocrine/paracrine cofactor. *Neuron* **28**, 385-397
19. Schuttelkopf, A. W., Hamilton, G., Watts, C., and van Aalten, D. M. (2006) Structural basis of reduction-dependent activation of human cystatin F. *J Biol Chem* **281**, 16570-16575
20. Sotiropoulou, G., Anisowicz, A., and Sager, R. (1997) Identification, cloning, and characterization of cystatin M, a novel cysteine proteinase inhibitor, down-regulated in breast cancer. *J Biol Chem* **272**, 903-910
21. Grubb, A. O. (2000) Cystatin C--properties and use as diagnostic marker. *Adv Clin Chem* **35**, 63-99
22. Cheng, T., van Vlijmen-Willems, I. M., Hitomi, K., Pasch, M. C., van Erp, P. E., Schalkwijk, J., and Zeeuwen, P. L. (2009) Colocalization of cystatin M/E and its target proteases suggests a role in terminal differentiation of human hair follicle and nail. *J Invest Dermatol* **129**, 1232-1242
23. Abrahamson, M., and Grubb, A. (1994) Increased body temperature accelerates aggregation of the Leu-68-->Gln mutant cystatin C, the amyloid-forming protein in hereditary cystatin C amyloid angiopathy. *Proc Natl Acad Sci U S A* **91**, 1416-1420
24. Ekiel, I., and Abrahamson, M. (1996) Folding-related dimerization of human cystatin C. *J Biol Chem* **271**, 1314-1321
25. Janowski, R., Abrahamson, M., Grubb, A., and Jaskolski, M. (2004) Domain swapping in N-truncated human cystatin C. *J Mol Biol* **341**, 151-160
26. Janowski, R., Kozak, M., Jankowska, E., Grzonka, Z., Grubb, A., Abrahamson, M., and Jaskolski, M. (2001) Human cystatin C, an amyloidogenic protein, dimerizes through three-dimensional domain swapping. *Nat Struct Biol* **8**, 316-320
27. Orlikowska, M., Jankowska, E., Kolodziejczyk, R., Jaskolski, M., and Szymanska, A. (2011) Hinge-loop mutation can be used to control 3D domain swapping and amyloidogenesis of human cystatin C. *Journal of structural biology* **173**, 406-413
28. Ghiso, J., Jenson, O., and Frangione, B. (1986) Amyloid fibrils in hereditary cerebral hemorrhage with amyloidosis of Icelandic type is a variant of gamma-trace basic protein (cystatin C). *Proc Natl Acad Sci U S A* **83**, 2974-2978
29. Yamada, M. (2000) Cerebral amyloid angiopathy: an overview. *Neuropathology* **20**, 8-22
30. Zerovnik, E., Pompe-Novak, M., Skarabot, M., Ravnihar, M., Musevic, I., and Turk, V. (2002) Human stefin B readily forms amyloid fibrils in vitro. *Biochim Biophys Acta* **1594**, 1-5
31. Zerovnik, E., Turk, V., and Waltho, J. P. (2002) Amyloid fibril formation by human stefin B: influence of the initial pH-induced intermediate state. *Biochemical Society transactions* **30**, 543-547
32. Skerget, K., Taler-Vercic, A., Bavdek, A., Hodnik, V., Ceru, S., Tusek-Znidaric, M., Kumm, T., Pitsi, D., Pompe-Novak, M., Palumaa, P., Soriano, S., Kopitar-Jerala, N., Turk, V., Anderluh, G., and Zerovnik, E. (2010) Interaction between oligomers of stefin B and amyloid-beta in vitro and in cells. *J Biol Chem* **285**, 3201-3210



33. Briggs, J. J., Haugen, M. H., Johansen, H. T., Riker, A. I., Abrahamson, M., Fodstad, O., Maelandsmo, G. M., and Solberg, R. (2010) Cystatin E/M suppresses legumain activity and invasion of human melanoma. *BMC Cancer* **10**, 17
34. Vigneswaran, N., Wu, J., Nagaraj, N., James, R., Zeeuwen, P., and Zacharias, W. (2006) Silencing of cystatin M in metastatic oral cancer cell line MDA-686Ln by siRNA increases cysteine proteinases and legumain activities, cell proliferation and in vitro invasion. *Life Sci* **78**, 898-907
35. Pulukuri, S. M., Gorantla, B., Knost, J. A., and Rao, J. S. (2009) Frequent loss of cystatin E/M expression implicated in the progression of prostate cancer. *Oncogene* **28**, 2829-2838
36. Wallin, H., Apelqvist, J., Andersson, F., Ekstrom, U., and Abrahamson, M. (2017) Low-level internalization of cystatin E/M affects legumain activity and migration of melanoma cells. *J Biol Chem* **292**, 14413-14424
37. Smith, R., Johansen, H. T., Nilsen, H., Haugen, M. H., Pettersen, S. J., Maelandsmo, G. M., Abrahamson, M., and Solberg, R. (2012) Intra- and extracellular regulation of activity and processing of legumain by cystatin E/M. *Biochimie* **94**, 2590-2599
38. Rodziewicz-Motowidlo, S., Iwaszkiewicz, J., Sosnowska, R., Czaplewska, P., Sobolewski, E., Szymanska, A., Stachowiak, K., and Liwo, A. (2009) The role of the Val57 amino-acid residue in the hinge loop of the human cystatin C. Conformational studies of the beta2-L1-beta3 segments of wild-type human cystatin C and its mutants. *Biopolymers* **91**, 373-383
39. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-797
40. Laskowski, R. A., Jablonska, J., Pravda, L., Varekova, R. S., and Thornton, J. M. (2018) PDBsum: Structural summaries of PDB entries. *Protein Sci* **27**, 129-134
41. Dall, E., and Brandstetter, H. (2013) Mechanistic and structural studies on legumain explain its zymogenicity, distinct activation pathways, and regulation. *Proc Natl Acad Sci U S A* **110**, 10940-10945
42. Lunde, N. N., Haugen, M. H., Bodin Larsen, K. B., Damgaard, I., Pettersen, S. J., Kasem, R., Rut, W., Drag, M., Poreba, M., Johansen, H. T., and Solberg, R. (2017) Glycosylation is important for legumain localization and processing to active forms but not for cystatin E/M inhibitory functions. *Biochimie* **139**, 27-37
43. Hamilton, G., Colbert, J. D., Schuettelkopf, A. W., and Watts, C. (2008) Cystatin F is a cathepsin C-directed protease inhibitor regulated by proteolysis. *The EMBO journal* **27**, 499-508
44. Yates, C. M., Butterworth, J., Tennant, M. C., and Gordon, A. (1990) Enzyme activities in relation to pH and lactate in postmortem brain in Alzheimer-type and other dementias. *J Neurochem* **55**, 1624-1630
45. Pirchl, M., and Humpel, C. (2009) [Does acidosis in brain play a role in Alzheimer's disease?]. *Neuropsychiatr* **23**, 187-192
46. Jenko Kokalj, S., Guncar, G., Stern, I., Morgan, G., Rabzelj, S., Kenig, M., Staniforth, R. A., Waltho, J. P., Zerovnik, E., and Turk, D. (2007) Essential role of proline isomerization in stefin B tetramer formation. *J Mol Biol* **366**, 1569-1579
47. Staniforth, R. A., Giannini, S., Higgins, L. D., Conroy, M. J., Hounslow, A. M., Jerala, R., Craven, C. J., and Waltho, J. P. (2001) Three-dimensional domain swapping in the folded and molten-globule states of cystatins, an amyloid-forming structural superfamily. *The EMBO journal* **20**, 4774-4781

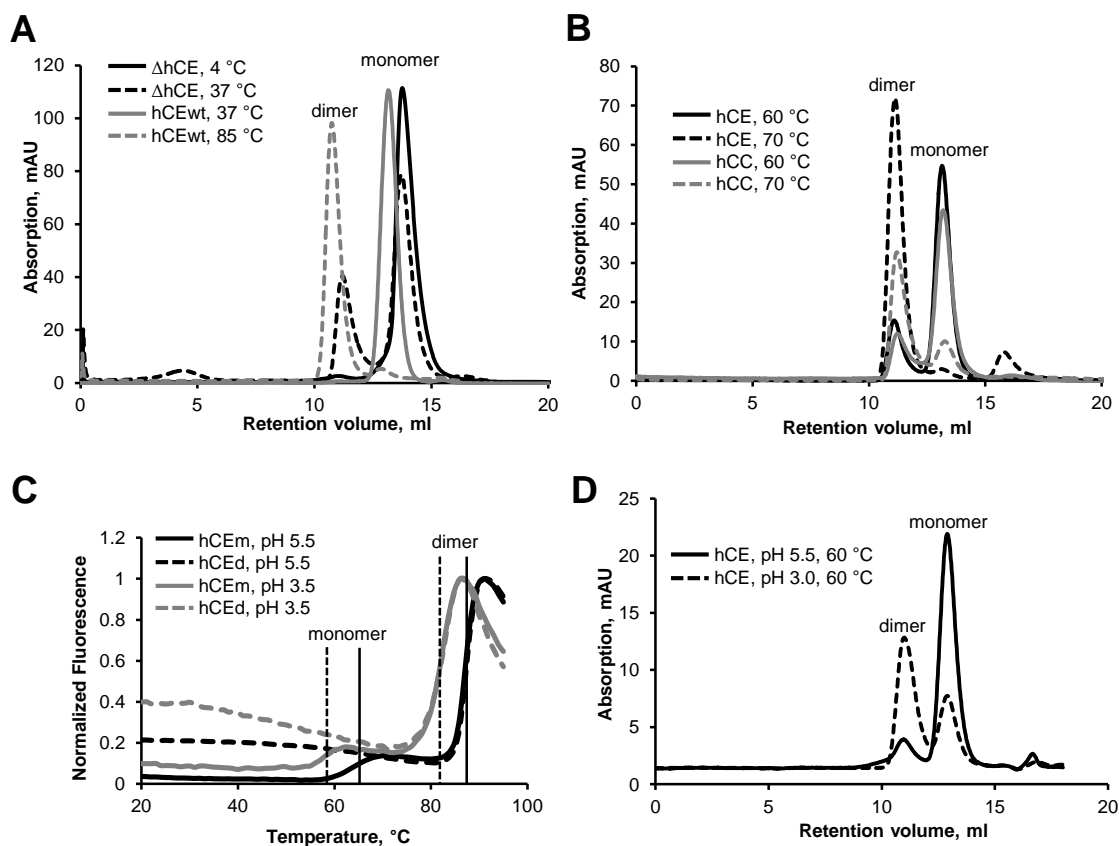
48. Knowles, T. P., Vendruscolo, M., and Dobson, C. M. (2014) The amyloid state and its association with protein misfolding diseases. *Nat Rev Mol Cell Biol* **15**, 384-396
49. O'Brien, R. J., and Wong, P. C. (2011) Amyloid precursor protein processing and Alzheimer's disease. *Annu Rev Neurosci* **34**, 185-204
50. Mi, W., Pawlik, M., Sastre, M., Jung, S. S., Radvinsky, D. S., Klein, A. M., Sommer, J., Schmidt, S. D., Nixon, R. A., Mathews, P. M., and Levy, E. (2007) Cystatin C inhibits amyloid-beta deposition in Alzheimer's disease mouse models. *Nat Genet* **39**, 1440-1442
51. Nguyen, A., and Hulleman, J. D. (2016) Evidence of Alternative Cystatin C Signal Sequence Cleavage Which Is Influenced by the A25T Polymorphism. *PloS one* **11**, e0147684
52. Yamane, T., Takeuchi, K., Yamamoto, Y., Li, Y. H., Fujiwara, M., Nishi, K., Takahashi, S., and Ohkubo, I. (2002) Legumain from bovine kidney: its purification, molecular cloning, immunohistochemical localization and degradation of annexin II and vitamin D-binding protein. *Biochim Biophys Acta* **1596**, 108-120
53. Dall, E., and Brandstetter, H. (2012) Activation of legumain involves proteolytic and conformational events, resulting in a context- and substrate-dependent activity profile. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **68**, 24-31
54. Ericsson, U. B., Hallberg, B. M., Detitta, G. T., Dekker, N., and Nordlund, P. (2006) Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal Biochem* **357**, 289-298
55. Niesen, F. (2010) Excel script for the analysis of protein unfolding data acquired by Differential Scanning Fluorimetry (DSF). (Niesen, F. ed., 3.0 Ed., Structural Genomics Consortium, Oxford
56. Williams, J. W., and Morrison, J. F. (1979) The kinetics of reversible tight-binding inhibition. *Methods in enzymology* **63**, 437-467
57. Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R., and Leslie, A. G. (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr* **67**, 271-281
58. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-242
59. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674
60. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132
61. Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **68**, 352-367
62. DeLano, W. L. (2002) PyMol. *DeLano Scientific, San Carlos, Ca, USA*
63. Sippl, M. J. (2008) On distance and similarity in fold space. *Bioinformatics* **24**, 872-873

**FOOTNOTES**

Funding was provided by the Austrian Science Fund (FWF project W\_01213).

The abbreviations used are: hCE, human cystatin E; hCC, human cystatin C; hCF, human cystatin F; AEP, asparaginyl endopeptidase; RCL, reactive center loop; LEL, legumain exosite loop; SEC, size exclusion chromatography; YVAD-cmk, Ac-Tyr-Val-Ala-Asp-chloromethylketone; AAN-AMC, Z-Ala-Ala-Asn-7-amino-4-methylcoumarin; Z-FR-AMC, Z-Phe-Arg-7-amino-4-methylcoumarin; glyco-hCE, glycosylated cystatin E

## Figures

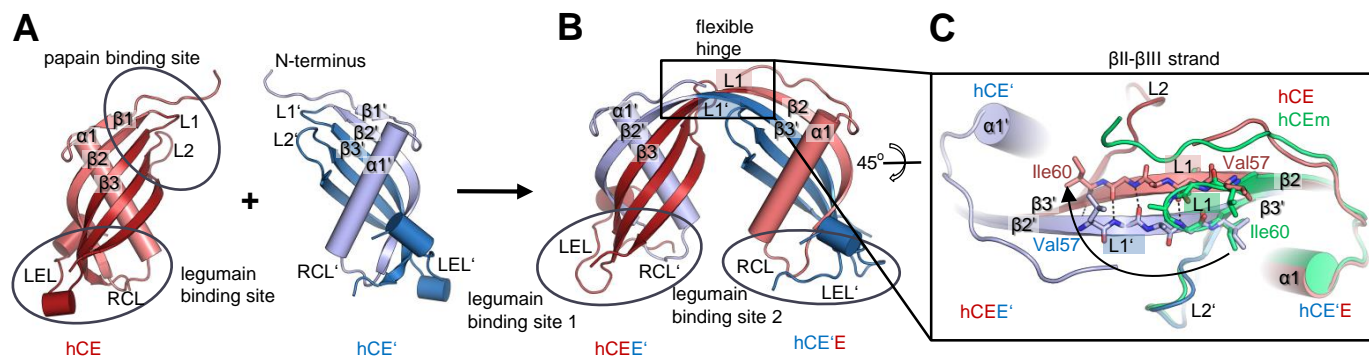
**Figure 1. Dimerization of cystatin E (hCE) is triggered by destabilization.**

**A)** Dimer formation of N-terminally truncated ( $\Delta$ hCE) and wild-type (hCEwt) hCE was assayed following incubation at indicated temperatures using size exclusion chromatography (SEC).

**B)** Cystatin E and cystatin C (hCC) were incubated at indicated temperatures before injection onto the SEC column. At 70 °C > 90% conversion to the dimeric form could be observed.

**C)** Thermal denaturation curves were collected for monomeric (hCEm) and dimeric (hCEd) hCE at indicated pH values. Melting Temperatures ( $T_M$ ) could be determined to be 65 °C (transition 1) and 87 °C (transition 2) at pH 5.5 for hCEm and hCEd, respectively. Acidic pH led to a reduction of  $T_M$  (dashed, vertical lines). Transition 1 was only observed for monomeric hCE and corresponds to the conversion of monomeric to dimeric hCE.

**D)** Incubation of hCE at pH 3.0 led to more efficient conversion to the dimeric form.

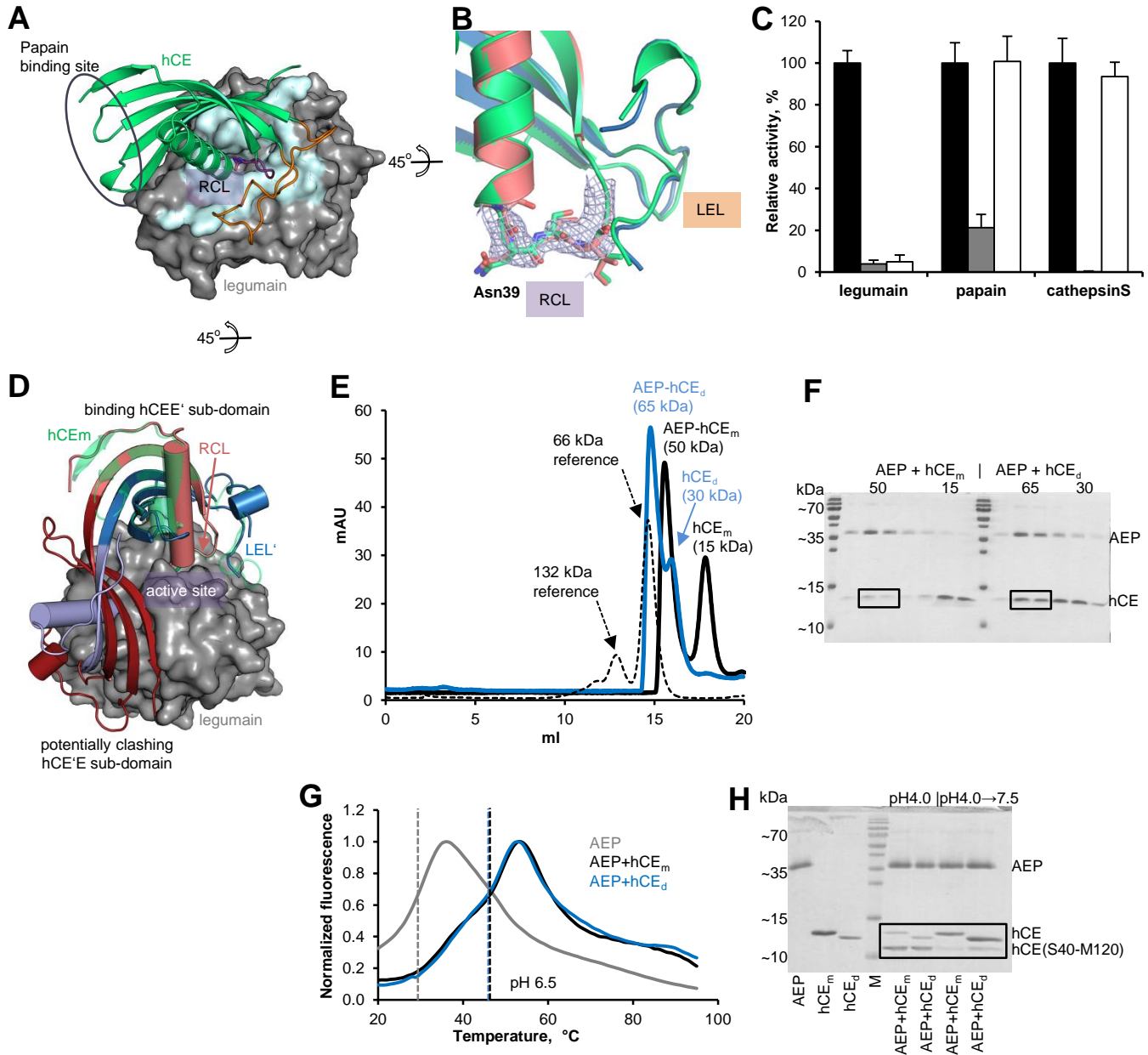


**Figure 2. The cystatin E dimer forms via domain swapping.**

A) Crystal structure of monomeric hCE (pdb 4n6l) with exposed papain- and legumain-binding sites. The regions of two monomers (red and blue) that undergo domain swapping are shown in light colors.

B) Crystal structure of the hCE dimer illustrated in cartoon representation. The dimer is composed of two hCE monomers where the N-terminal region (light blue) of molecule hCE' swapped out and integrated into the equivalent position on molecule hCE and *vice versa*. Thereby, two symmetric subunits hCEE' and hCE'E are formed.

C) Top view on the flexible hinge region formed by the former L1 loops. Upon domain swapping, L1 (light red) and L1' (light blue) rotated out by 90 ° and thereby formed the  $\beta$ II- $\beta$ II-strand connecting  $\beta 2$  to  $\beta 3$  and  $\beta 2'$  to  $\beta 3'$  respectively. The structure of monomeric hCE (green cartoon) was superposed onto the hCE'E subdomain.



**Figure 3. The legumain-inhibitory site on dimeric hCE is accessible and fully functional.**

**A)** Crystal structure of the hCE – legumain complex (pdb 4n6o). hCE is shown in green cartoon, the reactive center loop (RCL) harboring the P1-Asn39 residue in purple, the legumain exosite loop (LEL) in orange and legumain in grey surface. The contact area on legumain is shown in light blue.

**B)** Superposition of hCE monomer (green) and dimer (red and blue) in cartoon representation. Intactness of the RCL was confirmed by the continuous electron density map ( $F_o - F_c$  omit map, contoured at  $2.0 \sigma$ ).

**C)** Inhibition of papain, cathepsin S and legumain by monomeric (grey) and dimeric (white) cystatin E was assayed using fluorogenic FR-AMC (papain, cathepsin S) and AAN-AMC (legumain) substrates. Activities were normalized to control reactions harboring enzyme only (black).

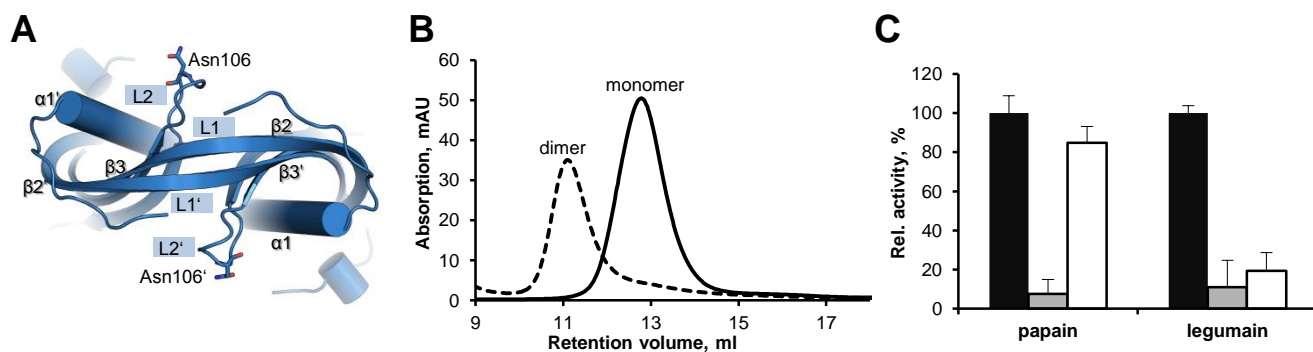
**D)** Model of a legumain–hCE dimer complex. Legumain is illustrated as grey surface, dimeric hCE as blue-red cartoon and monomeric hCE as green cartoon. Since hCEE' and hCE'E are symmetric subdomains they are in principle both capable of binding to the legumain active site.

**E)** Legumain (AEP) was incubated with monomeric (hCE<sub>m</sub>) and dimeric (hCE<sub>d</sub>) hCE and complex formation was investigated via SEC. BSA (bovine serum albumin) monomer (66 kDa) and dimer (132 kDa) peaks served as a reference (black, dashed line)

**F)** SDS-PAGE of the peak fractions from the experiment described in E).

**G)** Thermal denaturation curves of legumain alone (grey curve) and pre-complexed with monomeric hCE (black curve) and dimeric hCE (blue curve) were recorded at pH 6.5 following the Thermofluor method. Dashed lines indicate melting temperatures ( $T_M$ ).

**H)** SDS-PAGE showing legumain alone (AEP), monomeric hCE (hCE<sub>m</sub>), dimeric hCE (hCE<sub>d</sub>), molecular weight marker (M), legumain incubated with monomeric hCE (AEP+hCE<sub>m</sub>) or dimeric hCE (AEP+hCE<sub>d</sub>) at pH 4.0 and after a subsequent shift to pH 7.5. Incubation of hCE with legumain at pH 4.0 leads to cleavage after P1-Asn39, which is visible in the (S40-M120)-hCE band. Subsequent shift to neutral pH (7.5) led to conversion of cleaved hCE to intact hCE upon religation of the Ser38-Asn39 peptide bond on hCE. Dimeric hCE was prepared from N-terminally truncated hCE.

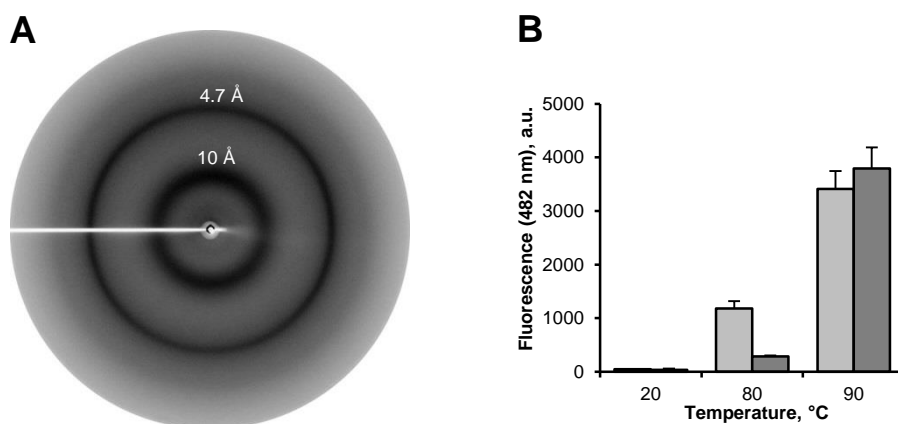


**Figure 4. Glycosylation is compatible with hCE dimerization.**

A) Top-view on the hCE dimer structure in cartoon representation. The Asn112 residues located on the L2 loops that are prone to glycosylation are shown in sticks.

B) Glycosylated hCE produced in LEXSY was incubated at 20 °C (black curve) and 85 °C (dashed, black curve) for 10 min. Subsequently, both samples were injected onto a S75 10/300 GL column. Incubation at 85 °C led to a shift in the retention volume that corresponds to dimeric hCE.

C) Inhibition of papain and legumain by monomeric (grey) and dimeric (white) glycosylated cystatin E was assayed using fluorogenic FR-AMC (papain) and AAN-AMC (legumain) substrates. Activities were normalized to control reactions harboring enzyme only (black).

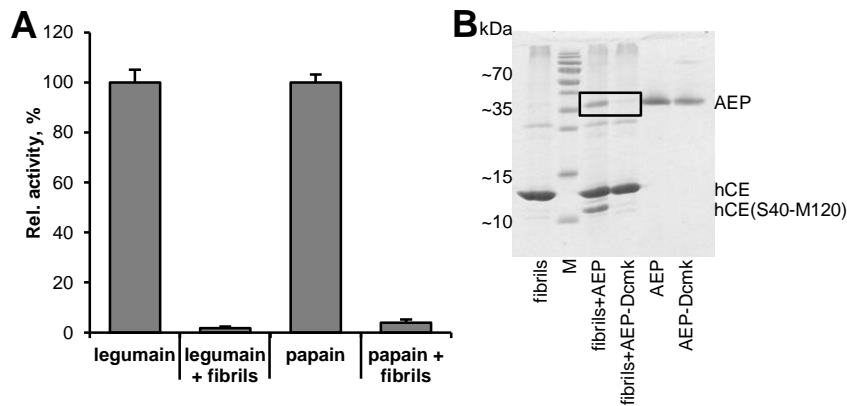


**Figure 5. Cystatin E forms cross-beta structures.**

A) X-ray diffraction experiments of insoluble hCE pellet revealed two rings at 10 Å and 4.7 Å resolution, which are characteristic for cross-beta structures.

B) Monomeric (light grey) and dimeric (dark grey) hCE were incubated at 20, 80 and 90 °C for 10 min. Subsequently, binding of thioflavin T (ThT) was measured as an increase in fluorescence detected at 482 nm.

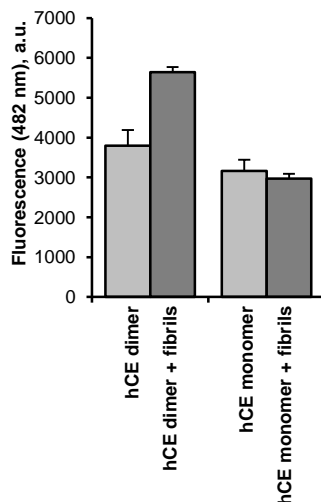




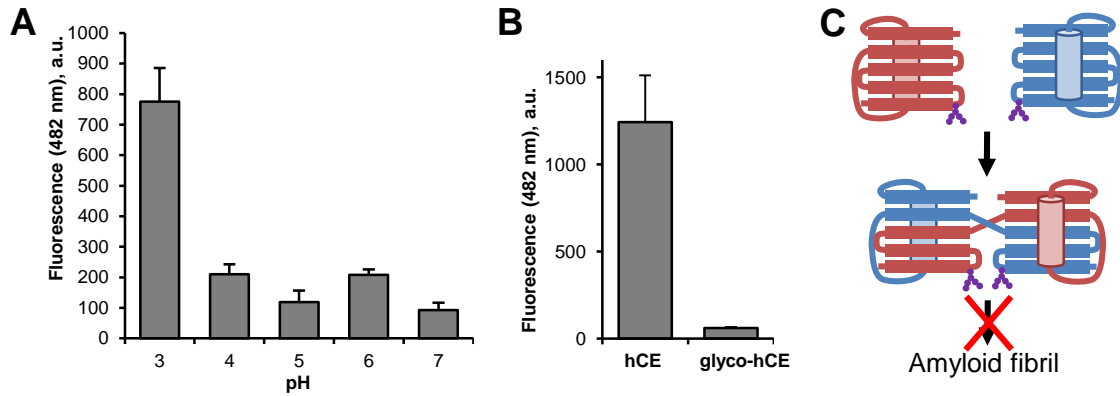
**Figure 6. hCE fibrils contain functional protein.**

**A)** Activity of legumain and papain was measured upon addition of hCE amyloid fibrils in a fluorescent substrate assay using FR-AMC (papain) and AAN-AMC (legumain) substrates. Control reactions contained only the respective enzyme.

**B)** SDS-PAGE showing a co-precipitation assay of hCE fibrils and legumain. Insoluble hCE fibrils were incubated with active-site free legumain (AEP) and active site-blocked legumain (AEP-Dcmk). Subsequently, the insoluble fraction was harvested by centrifugation and loaded on SDS-PAGE. Control reactions contained fibrils only, AEP only and AEP-Dcmk only. While active site-free legumain bound to hCE-fibrils, active site-blocked AEP-Dcmk did not bind. Additionally, a band corresponding to P1-Asn39 processed hCE was observed (hCE(S40-M120); C-terminal cleavage product).



**Figure 7. hCE fibrils serve as a template for dimeric hCE to bind.** hCE monomer and dimer were incubated at 90 °C for 10 min in the presence (dark grey) and absence (light grey) of hCE fibrils. Subsequently, ThT binding was measured as an increase in fluorescence at 482 nm. Control reactions contained ThT only and fibrils only.

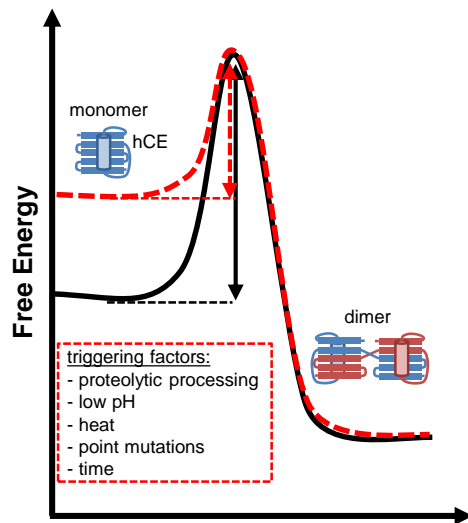


**Figure 8. hCE fibril formation is pH dependent and incompatible with glycosylation.**

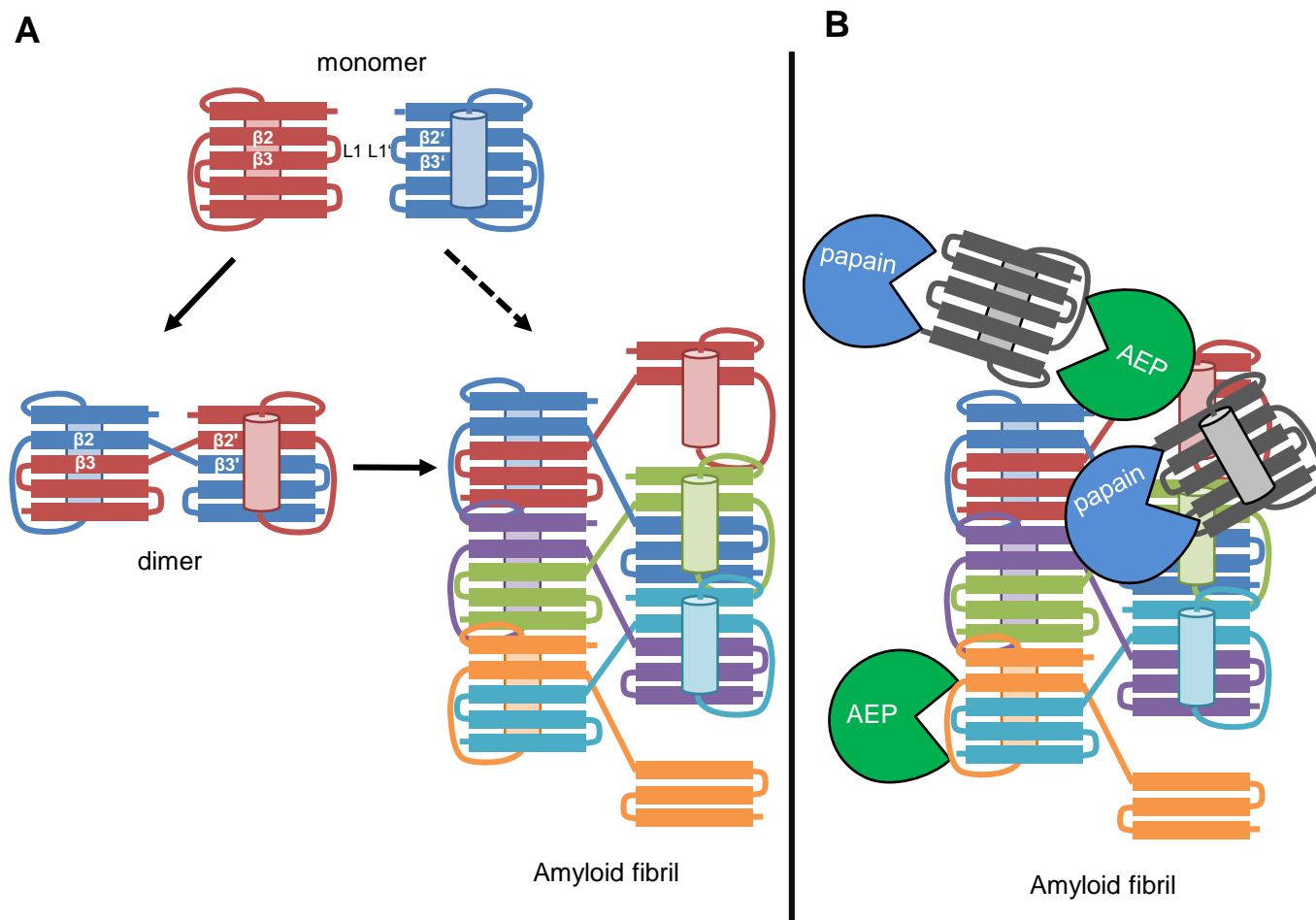
**A)** Monomeric hCE was incubated at indicated pH values for 10 min. Subsequently, binding of ThT was measured as an increase in fluorescence at 482 nm.

**B)** Glycosylated cystatin E (glyco-hCE) produced in LEXSY (*Leishmania tarentolae* expression system, Jena Bioscience) and unglycosylated hCE produced in *E.coli* were incubated at 90 °C for 10 min. Subsequently, binding of ThT was measured by monitoring an increase in fluorescence at 482 nm.

**C)** Model of glycosylated hCE forming a dimer but not higher oligomers.



**Figure 9. Conversion of monomeric hCE to the dimer needs a source of energy.** The cystatin E monomer is a stable folding state that can convert to a dimeric state if a certain energy barrier is overcome (black arrow). The dimer has a higher thermal and fold stability as compared to the monomer. Factors reducing the energy barrier (red arrow) are time, pH, proteolytic processing, mutations, and temperature.



**Figure 10. Model of cystatin E oligomerization.**

**A)** Monomeric hCE can convert to a dimer upon mild conformational destabilization. Dimeric hCE can further convert to ordered oligomers potentially via concerted domain swapping reactions. The conversion from monomer to dimer and dimer to oligomer needs a certain energy barrier to be overcome. The energy barrier can be overcome by mild destabilization by low pH, N-terminal truncation, heat and point mutations, amongst other factors. Presumably, dimeric hCE is a stable intermediate on the route to amyloid fibrils. Consequently, the conversion of monomeric hCE to fibrils very likely proceeds via the dimer.

**B)** hCE fibrils are functional as legumain and papain inhibitors. For that reason we suppose that they are heterogeneously composed of domain swapped and monomeric subunits. The presence of hCE monomers (dark grey units) allows for the inhibition of papain-like enzymes. Inhibition of legumain is possible both with monomeric and domain-swapped hCE.

**Structural and functional analysis of cystatin E reveals enzymologically relevant dimer and amyloid fibril states**

Elfriede Dall, Julia C. Hollerweger, Sven O. Dahms, Haissi Cui, Katharina Häußermann and Hans Brandstetter

*J. Biol. Chem.* published online July 2, 2018

---

Access the most updated version of this article at doi: [10.1074/jbc.RA118.002154](https://doi.org/10.1074/jbc.RA118.002154)

Alerts:

- [When this article is cited](#)
- [When a correction for this article is posted](#)

[Click here](#) to choose from all of JBC's e-mail alerts