Molecular plasticity controls enzymatic activity

DOI: 10.1002/anie.200((will be filled in by the editorial staff))

Multiple Stable Conformations Account for Reversible Concentration-Dependent Oligomerization and Autoinhibition of a Metamorphic Metallopeptidase**

Mar López-Pelegrín [#,‡], Núria Cerdà-Costa [#,‡], Anna Cintas-Pedrola [#], Fátima Herranz-Trillo [$], Pau Bernadó [$], Juan R. Peinado [#,&], Joan L. Arolas [#], and F. Xavier Gomis-Rüth [#,]*

In general, the native fold of a protein in a given environment is unique and at a global free-energy minimum [1]. However, some proteins spontaneously undergo substantial fold switching and reversibly transit between several conformers: "metamorphic" proteins [2]. Unfortunately, identifying and examining such proteins is a challenge because they are highly dynamic and impossible to identify *a priori* [3]. In contrast, minor rearrangement often occurs in single-domain enzymes upon binding of substrates, as shown for proteolytic enzymes of the metallopeptidase (MP) class [4]. As to enzymatic activity, an increase in enzyme concentration usually raises activity, as more substrate can be bound and turned over [5]. Here we describe a metamorphic minimal selective and specific caseinolytic metallopeptidase, *selecase*, which shows reversible and concentration-dependent equilibrium between different discrete states and an associated loss of enzymatic activity due to autoinhibition.

We recently discovered a family of soluble minimal MPs named "minigluzincins" and characterized two of them, proabylysin and projannalysin, but we only isolated them as inactive zymogens, each in a single conformation [6]. In the present study, we introduce selecase from *Methanocaldococcus jannaschii* as a novel family member. We recombinantly produced and purified selecase (see Experimental Procedures [EP] and Supplemental Results and Discussion [SRD] in the Supporting Information for details). In contrast to the other minigluzincins, the 110-residue full-length selecase corresponded to a mature, fully active MP with narrow and selective—hitherto unreported—substrate specificity that cleaved bovine milk casein at a single site on its α_{s1} chain (Suppl. Fig. 1 and Suppl. Tables 1-2).

Selecase was extremely soluble in aqueous buffer and did not precipitate at 130mg/ml. Thus, we studied the concentration-dependent enzymatic activity of selecase on a peptide that mimics the casein cleavage site (peptide CCS). Normally, peptide-bond hydrolysis by MPs is an ordered single-displacement reaction, which follows simple Michaelis-Menten kinetics [7]. This entails that higher enzyme concentrations enhance the initial rate of reaction in the pre-steady state following a hyperbolic curve until a plateau is reached upon saturation [5]. This is found *e.g.* with tobacco-etch virus proteinase, which is widely used in biotechnology (Fig. 1a).

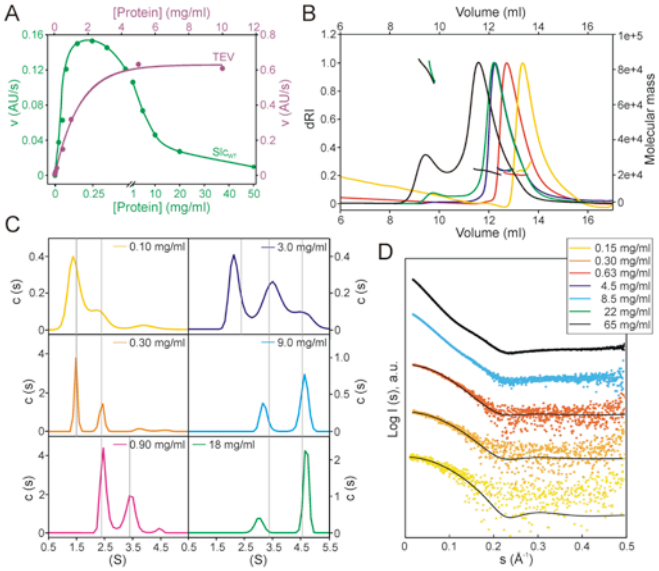


Figure 1. **A polyoligomeric metallocaseinase with abnormal activity. (A)** Proteolytic activity of wild-type selecase on peptide CCS (green curve). Tobacco-etch virus proteinase mutant S219V, which shows comparable catalytic efficiency to selecase but normal concentration-dependent activity, is shown for comparison (purple curve). **(B)** SEC-MALLS of selecase at selected initial concentrations (0.15-65mg/ml; see also Suppl. Fig. 2b). The peak pattern moves towards smaller elution volumes with increasing protein concentration, thus suggesting protein oligomerization. Curves are colored according to the inset in panel (D). **(C)** Analytical ultracentrifugation curves at six selected concentrations depicting the concentration-dependent oligomeric populations. Essentially, monomers are predominantly found at 0-0.3mg/ml; dimers at 0.3-2mg/ml; tetramers at 2-6mgl; and octamers at >6mg/ml. **(D)** SAXS intensity profiles, I(s), as a function of the momentum transfer, s, measured for wild-type selecase at selected concentrations (see Suppl. Fig. 3 for all curves). Profiles have been displaced along the I(s) axis for comparison. The experimental scattering curves at the three lowermost concentrations studied indicate a mixed population of monomers and dimers based on the crystallographic structures of slc_1 and slc_2 (black curves).

Surprisingly, although selecase activity did indeed increase with concentration between 0.025-0.25mg/ml, it fell sharply thereafter to become only residual at 50mg/ml. Most interestingly, this inactive concentrated selecase regained maximal activity following simple dilution with buffer. Accordingly, selecase showed reversible enzymatic autoinhibition due to changes in concentration—and not to inhibition by the substrate or any other reagent—, which to our knowledge is novel for peptidases.

Subsequently, we explored oligomerization of selecase in solution in the concentration range 0.15-65mg/ml using several biophysical techniques (see EP and SRD for full details). Briefly, calibrated size-exclusion chromatography (SEC) revealed monomers, dimers, tetramers, and octamers in variable amounts depending on the concentration (Suppl. Fig. 2a). SEC-MALLS, which combines SEC with multi-angle laser light scattering (MALLS), revealed two average populations of ~25KDa and ~80KDa, possibly corresponding, respectively, to dimeric and octameric selecase but also additional

[#] Proteolysis Lab; Molecular Biology Institute of Barcelona, CSIC; Barcelona Science Park; c/ Baldiri Reixac, 15-21; 08028 Barcelona (Spain).

[*] Corresponding e-mail: fxgr@ibmb.csic.es
Homepage: http://www.ibmb.csic.es/home/xgomis.

[$] Centre de Biochimie Structurale; INSERM U1054; CNRS UMR 5048; Université Montpellier 1 and 2; 29 rue de Navacelles; 34090 Montpellier (France).

[‡] These authors share first authorship. [&] Present address: Department of Medical Sciences; University of Castilla-La Mancha; E-13071 Ciudad Real (Spain).

[**] See Supporting Information for Experimental Procedures, Supplemental Results and Discussion, Acknowledgements, Supplemental References, 8 Supplemental Tables, 7 Supplemental Figures, and Legends of 3 Supplemental Movies.

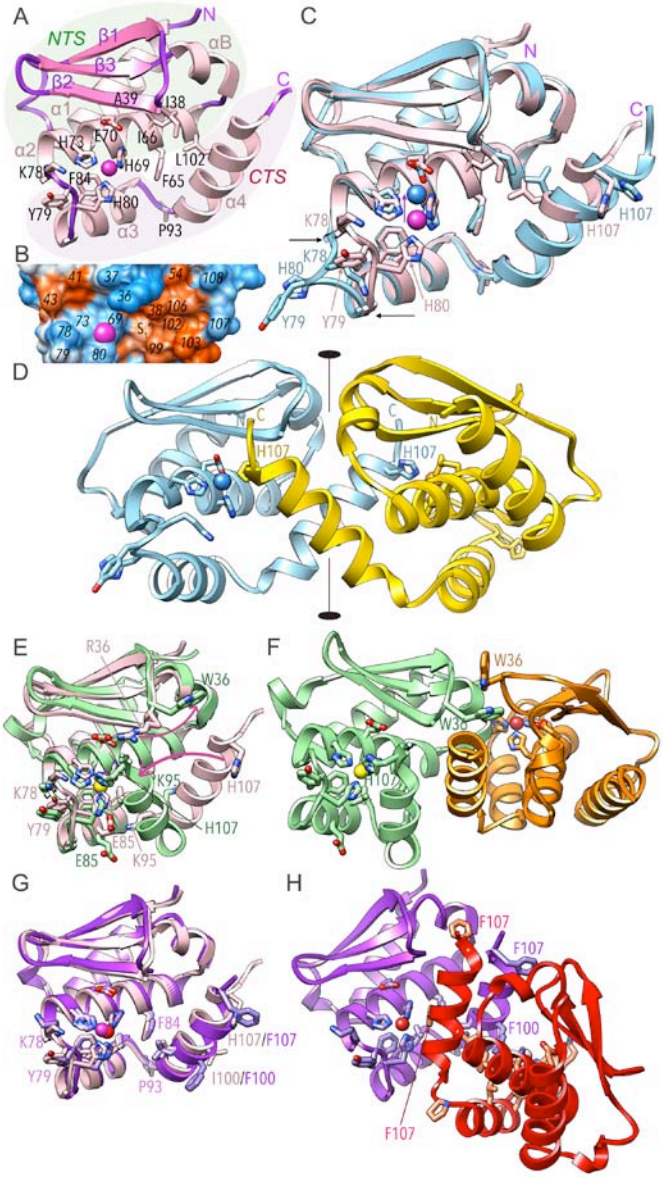


Figure 2. **Competent monomer vs. incompetent dimers. (A)** Ribbon plot of slc_1 in standard orientation [8]. Helices (α1, αB, and α2-α4) are shown in pink, β-strands (β1-β3) in magenta, and loops/coils in purple. For extent and nomenclature of regular secondary structure elements, see Suppl. Table 7. Selected residues are shown for their side chains, as is the catalytic metal ion (magenta sphere). The NTS and the CTS are shown over light green and light purple background, respectively. **(B)** Surface representation of slc_1 colored according to Kyte-Doolittle hydrophobicity (blue=hydrophilic over white to orange=hydrophobic) in the same orientation as in (A) showing the active-site cleft with the hydrophobic S_1' specificity pocket. **(C)** Superposition of slc_1 (in pink) and the slc_2 monomer (in cyan). Depicted are the respective metal ions, which are shifted relative to each other (purple arrow). Horizontal black arrows pinpoint the anchor points around which the conformational rearrangement occurs. **(D)** Overall structure of symmetric dimeric slc_2—chains in cyan and gold—depicted so that the crystallographic dyad (black horizontal ellipses joined by a line) is in the plane of the picture. **(E)** Superposition of slc_1 (in pink) and molecule B of the $R^{36}W$ selecase dimeric mutant ($slc_{2'}$; in light green). Magenta arrows pinpoint the side chain movement at position 36 owing to the mutation and the 50º-rotation of C-terminal helix α4. **(F)** Structure of the asymmetric dimer of $slc_{2'}$ consisting of helix-rotated molecule B (green) and close-to-native molecule A (orange). Both active-site clefts are blocked but following different mechanisms. Note the two W^{36} side chains at the interface. **(G)** Superposition of slc_1 (in pink) and one of the two equivalent close-to-native monomers of selecase $I^{100}F+ H^{107}F$ dimeric mutant ($slc_{2'}$; in purple). **(H)** Inactive dimer of $slc_{2'}$(in purple and red).

species such as monomers and tetramers (Fig. 1b and Suppl. Fig. 2b). Sedimentation velocity analytical ultracentrifugation revealed the concentration-dependent presence of four oligomeric species, which would be consistent with monomers, dimers, tetramers, and octamers. This was backed by equilibrium velocity experiments showing concentration-dependent average masses ranging between monomers+dimers and octamers (Fig. 1c and Suppl. Table 3). Chemical crosslinking experiments followed by SDS-PAGE, in turn, showed monomers, dimers, monomer-dimer complexes, and tetramers. Higher oligomerization species were not detected due to intrinsic experimental limitations (Suppl. Fig. 2c). Circular dichroism spectra of selecase, with either zinc or nickel in the catalytic site, displayed the typical shape of well-folded mostly α-helical proteins (Suppl. Fig. 2d). Finally, SAXS revealed that the protein did not aggregate at concentrations of up to 65mg/ml (Suppl. Table 4, Fig. 1d, Suppl. Figs. 3 and 4). These results further showed that the relative population of the oligomeric species in solution was concentration dependent. In addition, single value decomposition analysis of the SAXS dataset indicated that four species (monomers, dimers, tetramers, and octamers) were present.

Summarizing, biophysical analyses in solution indicated the presence of mixtures of monomers, dimers, tetramers and octamers, with higher concentrations leading to greater oligomerization but not indiscriminate aggregation or precipitation. The concentrations at which monomeric selecase was predominant coincided with those of maximal enzymatic activity (0.2-0.3mg/ml; Fig. 1a), thus indicating that the monomer is the active species and that oligomers correspond to self-inhibiting species in all cases (see below). This would explain why higher enzyme concentrations yielded lower activity (Fig. 1a) and is reminiscent of previous reports on oligomerization inhibiting activity of phospholipase A2 [9]. Notably, simple dilution with buffer reversed oligomerization to yield monomers and restore activity.

To identify the molecular determinants of this behavior, we crystallized and solved the structure of wild-type selecase (see EP, SRD and Suppl. Table 5). We obtained three crystal forms—orthorhombic, tetragonal, and hexagonal—, which serendipitously corresponded, respectively, to monomeric (slc_1), dimeric (slc_2) and tetrameric (slc_4) forms of selecase. This indicated that at least three of the oligomerization states found in solution had a counterpart in the form of a stable, isolatable species, each one favored by particular crystallization conditions. The crystal structure of monomeric slc_1 reveals—by comparison with several functional but otherwise unrelated MPs—that the overall architecture, the metal-binding site, and the active-site environment are consistent with a competent and functional mature enzyme (see [7b, 10]). This conclusion is supported by the enzymatic activity found for selecase in solution being associated with a monomeric species (see above). It is also reinforced by SAXS for which the experimental scattering curves at the three lowermost concentrations—covering the activity maximum of the enzyme—clearly indicated mixed populations of monomers and dimers based on the crystallographic coordinates of slc_1 and slc_2 (see below), with the monomeric fraction at the two lowermost concentrations reaching 70% (see Fig. 1d, SRD and Suppl. Figs. 3-4).

At 13.1 KDa, slc_1 is the smallest active peptidase structurally characterized to date and it has a compact globular shape of ~35-40Å in diameter (Fig. 2a). It consists of an upper N-terminal sub-domain (NTS; residues M^1-Y^{76}) and a lower C-terminal sub-domain (CTS; G^{77}–K^{109}), which are connected by a mostly hydrophobic interface (Suppl. Table 6) and separated by a horizontal central active-site cleft (Fig. 2b). The NTS is an α/β-sandwich, with a three-stranded mixed β-sheet (β1-β3; Suppl. Table 7) that forms the roof of the selecase moiety (Fig. 2a). Two roughly parallel α-helices ("backing helix" α1 and "active-site helix" α2) are attached to the convex surface of the sheet, which faces the central core of the protein. A short helical segment ("linking helix" αB) is inserted in the loop connecting strand β3 with helix α2 (Lβ3α2). Helix α2 roughly parallels the active-site cleft and ends with the last residue of the NTS at Y^{76}. It encompasses a metal binding motif, H^{69}-E^{70}-X-X-H^{73}, which is characteristic of MPs and includes two metal-binding histidines and a general base/acid glutamate essential for catalysis [11]. Residue H^{80}, imbedded within Lα2α3 of the CTS, is the third metal ligand. The CTS mainly consists of two helices ("glutamate helix" α3 and "C-terminal helix" α4), whose axes intersect at ~90º. Helix α3 contains F^{84} at the center of the "Ser/Gly-turn" [6, 11a], which creates a hydrophobic base for the metal-binding site and contributes to its stabilization. The active-site cleft of selecase is framed by helix α2; "upper-rim" strand β2 of the NTS sheet and the preceding "bulge-edge segment" (L^{34}-I^{38}); helices α3 and α4; and Lα2α3, in particular through the side chains of K^{78} and Y^{79}. The catalytic metal ion resides at the bottom left of the cleft (Fig. 2a,b). At its right, a deep hydrophobic S_1' pocket is shaped by I^{38}, A^{39}, F^{65}, I^{66},

L^{102} and the solvent-accessible ring surface of H^{69}. This pocket optimally accommodates a phenylalanine in the P_1' position of substrates as found at the casein cleavage site. The slc_1 moiety is held together by a central hydrophobic core, which traverses the entire molecule, and several of the contributing residues also shape the NTS-CTS interface (Fig. 2a and Suppl. Table 6).

The crystal structure of slc_2 shows a dimer (Suppl. Table 8), and superposition of slc_1 and slc_2 monomers reveals good overall fit, with only minor differences within the NTS (see Fig. 2c). However, major metamorphic rearrangement is observed around the metal-binding site (see Fig. 2c and Suppl. Movie 1). In slc_2, at the beginning of CTS, Lα2α3 folds outward between G^{77} and I^{81}, with a maximal displacement of ~7Å. This causes the third metal-binding protein residue in slc_1, H^{80}, to swing out and protrude from the molecular surface. This, in turn, leads to an upward shift of the catalytic metal towards the general base/acid E^{70} (Fig. 2c). Two selecase monomers associate through C_2 symmetry under occlusion of a large surface (2,130Å^2; see Suppl. Table 8 and Fig. 2d) and so the third metal-binding site is taken over by H^{107} from helix α4 of the symmetric molecule. Accordingly, this H^{80}/H^{107} ligand swap is an intermolecular event that yields a catalytically incompetent metal-binding site and a blocked active-site cleft in slc_2. This is consistent with oligomerization states higher than a monomer coinciding with inactive species in solution (see above).

As in slc_2, the protomer of tetrameric slc_4 shows good overall fit with slc_1 within the NTS, including the position and conformation of most side chains at the NTS-CTS interface. However, both major displacement and drastic conformational rearrangement are observed in the CTS (see Fig. 3a and Suppl. Movie 2). The segment of the active-site helix with the first two metal ligands undergoes slight displacement (Fig. 3a). Downstream loop Lα2α3 and glutamate helix α3—which is virtually unchanged in both slc_1 and slc_2—unfold and give rise to strands β4 and β5, which adopt a canonical β-ribbon structure (Fig. 3a and Suppl. Table 7). Such long stretches of a protein only rarely undergo such dramatic transitions [12]. The β-ribbon protrudes away from the molecular moiety (Fig. 3a,b), as a result of which metal-ligand H^{80} shifts ~16Å from its position in slc_1 and no longer binds the metal. In contrast with α3, the C-terminal helix α4 keeps its helical structure but is displaced ~30Å apart on average in slc_4 (Fig. 3a). Overall, this metamorphic structural transition of selecase is stabilized by the association of four monomers in the crystal (Fig. 3b-d, Suppl. Table 8, and Suppl. Movie 2), which would explain tetrameric oligomerization in solution (see above). The oligomer is a compact, almost spherical self-inhibitory particle of ~60-75Å in diameter (Fig. 3c,d). One monomer (chain A) interacts through D_2 symmetry—by hiding a total interface of 9,850Å^2—with two neighboring molecules (chains B and D) through mixed hydrophobic/hydrophilic contacts, and with one opposite monomer (chain C) through hydrophobic contacts (Suppl. Table 8). Two large elliptical openings (minor axis ~16Å, major axis ~21Å; Fig. 3d and Suppl. Movie 2) on opposite faces of the particle are framed by upper-rim strands β2 and Lβ5α4 of two vicinal monomers (AB and CD). Access to the particle lumen through these entrances is limited by the respective β-ribbons, which protrude away from the particle surface and do not contact each other. The central lumen of the particle features a channel of ~50Å in length and ~15Å in diameter and allocates two internal dimetallic zinc-binding sites. Each of them results from the fusion of two neighboring metal sites as originally found in slc_1 (chains AD and BC, respectively), with the two metal ions of each site ~3Å apart (Fig. 3b,d). Overall, this new conformation radically alters the structural segments that shape the S_1' pocket and the active-site cleft in competent slc_1 and, thus, indicates that, like slc_2, the tetrameric slc_4 structure corresponds to an inactive species. This, again, is consistent with tetramers coinciding with inactive species in solution.

Given the importance of the C-terminal helix α4 and loop Lβ1β2 in oligomerization, we selected residues R^{36}, I^{100}, I^{103}, and H^{107}, which had been observed to participate in dimerization in slc_2 and

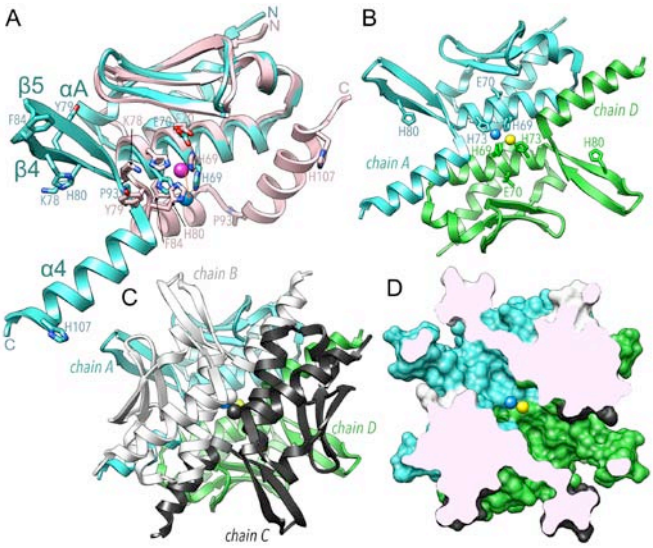


Figure 3. **A compact autoinhibitory tetrameric particle. (A)** Superposition of slc_1 and slc_4 monomers in pink (magenta metal ion) and turquoise (blue metal ion), respectively, in the view of Fig. 2a. Only the distinct secondary structure elements of slc_4 are labeled (see also Suppl. Table 7). Relevant residues undergoing major rearrangement are displayed for both structures and labeled. The metal is shifted downwards (red arrow). **(B)** Within the slc_4 tetramer, two neighbor monomers as in (A), in turquoise (chain A; metal in blue) and light green (chain D; metal in yellow), bind over a crystallographic dyad perpendicular to the plane of the picture. This gives rise to a non-functional dimetallic zinc site bound by H^{69} and H^{73} of either monomer. **(C)** Two dimers as in (B), in turquoise/light green (chains A and D) and white/dark gray (chains B and C), associate face to face under a relative 90º-rotation to yield the overall tetrameric particle, with two dimetallic zinc sites in the particle lumen. **(D)** Surface representation of (C) after clipping off the frontal part to delineate the central particle channel. Only the dimetallic site depicted in (B) is shown for clarity.

tetramerization in slc_4 (Suppl. Table 8), and generated a total of seven single, double and triple point mutants in an attempt to ablate the interactions responsible for oligomerization and thus obtain monomeric forms. In addition, we constructed two deletion mutants targeting α4, respectively lacking four (slcΔC4) and eight (slcΔC8) C-terminal residues. Moreover, we cloned two close orthologs from *Methanotorris igneus* and *Methanocaldococcus fervens*, which can be envisaged as natural fivefold and 19-fold point mutants of selecase (see EP and SRD for full details). All protein variants were produced, purified, and concentrated similarly to the wild type except for slcΔC4, which was obtained with lower yields and could only be maximally concentrated to 5.0mg/ml, and slcΔC8, which was insoluble and was discarded. This finding pointed to a stabilizing effect of helix α4 on the whole protein despite its overall flexibility in the various structures analyzed. Despite differences in the oligomer populations, all mutants revealed a concentration-dependent equilibrium between monomers, dimers, tetramers and octamers and a reduction in activity as concentration increased, similar to the wild type (Suppl. Fig. 5). These results indicate that selecase is highly plastic, which allows it to adapt to potentially deleterious point mutations and retain its capacity to oligomerize.

This plasticity is backed by further structural studies. Out of all the aforementioned mutants and orthologs, we managed to crystallize variants R^{36}W (hereafter $slc_{2'}$) and I^{100}F+H^{107}F (hereafter $slc_{2''}$) and solved their crystal structures (see Figs. 2e-h and Suppl. Movie 3). Most interestingly, $slc_{2'}$ showed a novel dimeric quaternary structure, distinct from slc_2, which displayed each protomer in a different conformation despite chemical identity of the molecules. One molecule (A) essentially displays the conformation of functional monomeric slc_1, including the metal site and the active-site cleft. It only differs significantly from the latter at Lβ1β2, which, owing to the side-chain replacement at position 36, causes the entire loop and thus the latter side chain to undergo major rearrangement towards the molecular moiety. The other molecule (B) also essentially coincides with slc_1 but just after the glutamate helix. Hereafter, a 90º-rotation around bond *N*-*Cα* of K^{95} results in C-terminal helix α4 being rotated as a rigid body by 50º so as to approach and thus sterically block its own active site cleft on its primed side. This further causes H^{107} to bind the catalytic metal, as observed in slc_2, except that here this is an intramolecular rather than an intermolecular event (compare Figs. 2d

and 2e,f). This novel conformation of a selecase variant in molecule B is stabilized by an asymmetric interaction between C-terminal helices with molecule A triggered by an edge-to-face interaction of the W^{36} side chains (Fig. 2f). This arrangement, in turn, causes the active-site cleft of molecule A to be blocked for substrate access by helix αB of molecule B, with Y^{57} of the latter interacting with the S_1' pocket of molecule A. The metal-binding site of the latter, in contrast, is unaffected. Accordingly, $slc_{2'}$ corresponds—like slc_2—to an inhibited conformation.

As to $slc_{2''}$, superposition of the two essentially identical monomers in the asymmetric unit onto slc_1 revealed a conformation that was close to the functional wild-type monomer, except that the end of the C-terminal helix was slightly unwound and more flexible owing to the two point mutations (Fig. 2g). However, the two phenylalanine residues at positions 100 and 107 make two $slc_{2''}$ monomers symmetrically bind mainly through their respective C-terminal helices, which run roughly parallel to each other. As a result the non-primed sides of the active-site clefts are occluded and the phenylalanine rings at position 100 penetrate, respectively, the S_1' specificity pocket of the symmetric partner, as this residue matches the specificity of the enzyme. Further symmetric contacts are observed between the F^{107} side chain of one molecule and loop Lβ1β2 of the other, which enhance the overall flexibility of these regions. Accordingly, the structure of $slc_{2''}$ provides yet another mechanism of inhibition of selecase: in this case merely by shielding of the cleft (Fig. 2h). Thus, the two crystal structures of $slc_{2'}$ and $slc_{2''}$ may represent genuine dimeric conformations of the mutants triggered by the respective side-chain replacements, as none of the respective structures was trapped in crystals of wild-type protein. This entails that replacement of just one and two residues, respectively, leads to two new structures of selecase (thus totaling five), supporting the metamorphic character of this protein.

Summarizing, we succeeded in identifying and probing for the first time the structural transitions of a natural metamorphic protein with a multi-funnel folding energy landscape. Although metamorphic proteins may be encoded by a relevant fraction of all genomes, the lack of bioinformatics or structural approaches to identify them from the sequence restricts their discovery to serendipity. Consistently, to our knowledge 3D structural evidence for their existence has only been published for two natural proteins [2, 13], which just flip between two folds: ubiquitin protein ligase inhibitor Mad2 [14] and the chemokine lymphotactin [15]. In selecase, the energy basins are occupied by distinct fully-structured and stable states and not by unfolded species or molten globules (Fig. 4 and Suppl. Movies 1-3). One conformer is catalytically competent and the others are incompetent but they coexist in equilibrium. These transitions between species are triggered by major rearrangement after residue G^{77} at the NTS-CTS interface, and they mainly affect the CTS. This is consistent with each sub-domain corresponding to a distinct folding unit or foldon [16] and the sub-domain interface acting as a reversible zipper. The high flexibility of CTS was further verified by computational analysis of local conformational frustration and inter-domain flexibility assessment based on the elastic network model (see SRD and Suppl. Fig. 7). In addition, the thermodynamic consistency of interconversion was further backed by calculated geometric and thermodynamic parameters of solvation free energy of folding and of dissociation, as well as compactness, for wild-type selecase structures (see SRD and Suppl. Table 8). The inherent flexibility of the CTS avoids kinetic trapping in an irreversible misfolded state during conversion between alternate conformers through the protein-protein interactions of oligomeric species as previously suggested for metamorphic proteins [13]. In our view, it is a striking observation that simple dilution/concentration of a sample at room temperature triggers fold switches that cause repacking of a hydrophobic core and exposure of new binding surfaces, which in turn generate spontaneous conversion between active monomers and inactive oligomers. This finding indicates that the energy barriers separating the minima are surmountable and that interconversion may proceed without passing through fully unfolded states [2], as suggested by the finding of largely conserved NTS foldons. Finally, our results also provide the first evidence for a peptidase with reversible, strictly concentration-dependent reduction of activity upon concentration increase, which is triggered by sequestering the competent conformation in incompetent but structured oligomers. This system affords a switch that provides a unique and reversible mechanism of control of catalytic activity in nature.

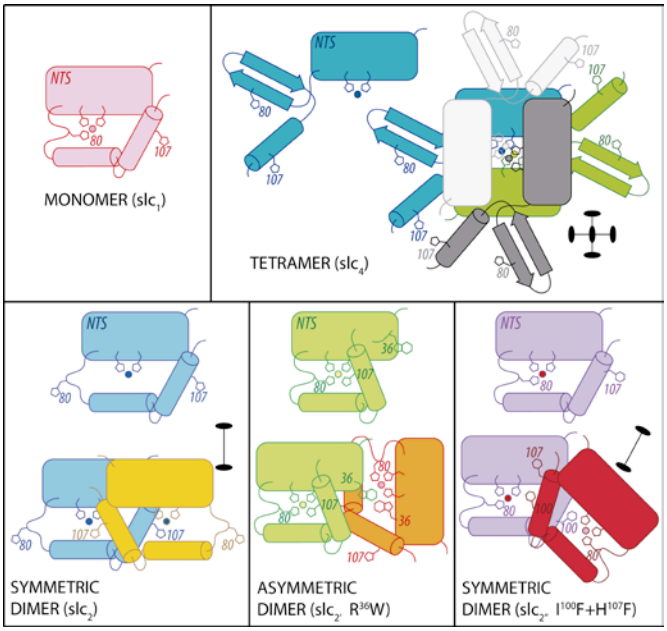


Figure 4. Scheme illustrating the topology of the distinct selecase structures reported. A black ellipse stands for a dyad vertical to the plane, two black ellipses connected by a line stand for a dyad in the plane. NTS, N-terminal sub-domain; histidines H^{69}, H^{73}, H^{80} and H^{107} are shown. In $slc_{2'}$, R^{36} is replaced by tryptophan; in $slc_{2''}$, I^{100} and H^{107} are replaced by phenylalanine.

Experimental Section

Details on the experimental procedures are provided in the Supporting Information.

Received: ((will be filled in by the editorial staff))
Published online on ((will be filled in by the editorial staff))

Keywords: metamorphic protein · metallopeptidase · protein folding · caseinolytic enzyme · proteolytic enzyme · crystal structure

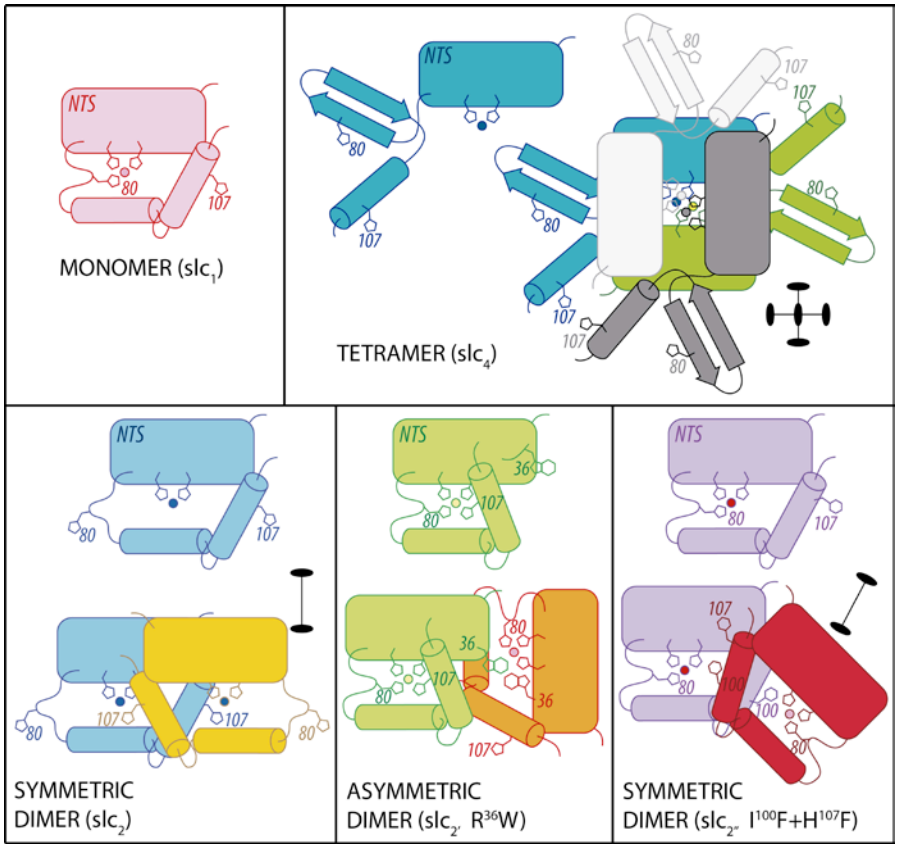
[1] C. B. Anfinsen, *Science* **1973**, *181*, 223-230.
[2] A. G. Murzin, *Science* **2008**, *320*, 1725-1726.
[3] S. C. Goodchild, P. M. G. Curmi, L. J. Brown, *Biophys. Rev.* **2011**, *3*, 143-153.
[4] P. Towler, B. Staker, S. G. Prasad, S. Menon, J. Tang, T. Parsons, D. Ryan, M. Fisher, D. Williams, N. A. Dales, M. A. Patane, M. W. Pantoliano, *J. Biol. Chem.* **2004**, *279*, 17996-18007.
[5] V. Henri, *C. R. Acad. Sci. Paris* **1902**, *135*, 916-919.
[6] M. López-Pelegrín, N. Cerdà-Costa, F. Martínez-Jiménez, A. Cintas-Pedrola, A. Canals, J. R. Peinado, M. A. Martí-Renom, C. López-Otín, J. L. Arolas, F. X. Gomis-Rüth, *J. Biol. Chem.* **2013**, *288*, 21279-21294.
[7] a) L. Polgár, in *Proteolytic enzymes - Tools and targets*. (Eds.: E. E. Sterchi, W. Stöcker), Springer Verlag, Berlin/Heidelberg, **1999**, pp. 148-166; b) B. W. Matthews, *Acc. Chem. Res.* **1988**, *21*, 333-340.
[8] F. X. Gomis-Rüth, T. O. Botelho, W. Bode, *Biochim. Biophys. Acta* **2012**, *1824*, 157-163.

[9] a) T. L. Hazlett, E. A. Dennis, *Biochemistry* **1985**, *24*, 6152-6158; b) D. H. Fremont, D. H. Anderson, I. A. Wilson, E. A. Dennis, N. H. Xuong, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 342-346.
[10] a) W. Bode, F. X. Gomis-Rüth, R. Huber, R. Zwilling, W. Stöcker, *Nature* **1992**, *358*, 164-167; b) F. X. Gomis-Rüth, *J. Biol. Chem.* **2009**, *284*, 15353-15357.
[11] a) F. X. Gomis-Rüth, *Crit. Rev. Biochem. Mol. Biol.* **2008**, *43*, 319-345; b) N. Cerdà-Costa, F. X. Gomis-Rüth, *Prot. Sci.* **2014**, *23*, 123-144.
[12] X. Zhou, F. Alber, G. Folkers, G. H. Gonnet, G. Chelvanayagam, *Proteins* **2000**, *41*, 248-256.
[13] P. N. Bryan, J. Orban, *Curr. Opin. Struct .Biol.* **2010**, *20*, 482-488.
[14] M. Mapelli, L. Massimiliano, S. Santaguida, A. Musacchio, *Cell* **2007**, *131*, 730-743.
[15] R. L. Tuinstra, F. C. Peterson, S. Kutlesa, E. S. Elgin, M. A. Kron, B. F. Volkman, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 5057-5562.
[16] A. R. Panchenko, Z. Luthey-Schulten, P. G. Wolynes, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 2008-2013.

Molecular plasticity controls enzymatic activity

Mar López-Pelegrín [#], Núria Cerdà-Costa [#], Anna Cintas-Pedrola, Fátima Herranz-Trillo, Pau Bernadó, Juan R. Peinado, Joan L. Arolas, and F. Xavier Gomis-Rüth * _____ **Page – Page**

Multiple Stable Conformations Account for Reversible Concentration-Dependent Oligomerization and Autoinhibition of a Metamorphic Metallopeptidase.



Molecular plasticity controls enzymatic activity: the native fold of a protein in a given environment is normally unique and in a global free-energy minimum. Some proteins, however, spontaneously undergo substantial fold switching to reversibly transit between defined conformers, the "metamorphic" proteins. Here, we present a minimal metamorphic, selective and specific caseinolytic metallopeptidase, selecase, which reversibly transits between several different states of defined three-dimensional structure, which are associated with loss of enzymatic activity due to autoinhibition. The latter is triggered by sequestering the competent conformation in incompetent but structured dimers, tetramers, and octamers. This system, which is compatible with a discrete multi-funnel energy landscape, affords a switch that provides a reversible mechanism of control of catalytic activity unique in nature.

SUPPORTING INFORMATION

Table of Contents

1. Experimental Procedures 2
1.1. Protein production and purification 2
1.2. Proteolytic and inhibitor assays 3
1.3. Concentration dependence of proteolytic activity 3
1.4. Size-exclusion chromatography and multi-angle laser light scattering 4
1.5. Analytical ultracentrifugation 4
1.6. Chemical crosslinking experiments 4
1.7. Circular dichroism spectroscopy 5
1.8. Small-angle X-ray scattering 5
1.9. Crystallization and data collection 5
1.10. Structure solution and refinement 6
1.11. Bioinformatics 7
2. Supplemental Results and Discussion 8
3. Acknowledgments 13
4. Supplemental References 13
5. Supplemental Tables 15
Supplemental Table 1 15
Supplemental Table 2 16
Supplemental Table 3 17
Supplemental Table 4 18
Supplemental Table 5 19
Supplemental Table 6 20
Supplemental Table 7 21
Supplemental Table 8 22
6. Supplemental Figures 23
Supplemental Figure 1 23
Supplemental Figure 2 24
Supplemental Figure 3 25
Supplemental Figure 4 26
Supplemental Figure 5 27
Supplemental Figure 6 28
Supplemental Figure 7 29
7. Legends to Supplemental Movies 30

1. EXPERIMENTAL PROCEDURES

1.1. Protein production and purification — The gene coding for selecase (110 residues, 13.1KDa; "uncharacterized protein MJ1213" according to UniProt (UP) sequence database access code Q58610) was amplified from *Methanocaldococcus jannaschii* genomic DNA. It was cloned at *Bam*HI and *Hind*III restriction sites into vector pPROEX-HTa (Invitrogen), which confers resistance towards ampicillin and attaches an N-terminal hexahistidine-tag (His_6-tag) followed by a tobacco etch virus (TEV) proteinase recognition site among other residues (M-S-Y-Y-H-H-H-H-H-H-D-Y-D-I-P-T-T-E-N-L-Y-F-Q-↓-G-A-M-D-P; the arrow shows the TEV proteinase cleavage site). This gave rise to tagged 16.5-KDa selecase. The gene was also cloned using *Nde*I and *Xho*I restriction sites into vector pCRI7 (modified from pET28a vector; Novagen), which confers resistance towards kanamycin and does not attach extra residues (13.1-KDa selecase). Single-residue point mutants were generated using the QuikChange Site-Directed Mutagenesis Kit (Stratagene) according to the manufacturer's instructions using either the pPROEX-HTa (point mutant $E^{70}A$) or the pCRI7 (point mutants $R^{36}W$, $I^{100}F$, $I^{103}E$, and $H^{107}F$) constructs as template. Multiple-residue point mutants ($I^{100}F+H^{107}F$, $I^{103}E+H^{107}F$, $I^{100}F+I^{103}E+H^{107}F$) were similarly generated starting from single-residue mutants. Deletion variants without the last four (slcΔC4; lacking H^{107}-Q-K-K^{110}; superscripted residue numbering of selecase according to UP 58610) and eight residues (slcΔC8; lacking I^{103}-N-K-L-H-Q-K-K^{110}) were prepared with the pCRI7 construct as template. Synthetic genes encoding two selecase orthologs from *Methanotorris igneus* (114 residues, 13.6KDa; UP F6BCT9) and *Methanocaldococcus fervens* (118 residues, 14.1KDa; UP C7P7R3) were purchased from GeneArt (Invitrogen) and cloned into vector pCRI7 as described above. All constructs were verified by DNA sequencing. Proteins were produced by heterologous overexpression in *Escherichia coli* BL21 (DE3) cells, which were grown at 37°C in Luria Bertani medium supplemented with either 100μg/ml ampicillin (pPROEX-HTa vector) or 30μg/ml kanamycin (pCRI7 vector). Cultures were induced at an A_{600} of 0.8 with 0.2-1mM isopropyl-β-D-thiogalactopyranoside and incubated either for 5h at 37°C or overnight at 18°C.

Purification and processing of 16.5-KDa selecase and its $E^{70}A$ mutant was performed as follows. After centrifugation at 7,000x*g* for 30min at 4°C, the pellet was washed twice with buffer A (50mM Tris-HCl, 500mM NaCl, pH8.0), resuspended in the same buffer plus 10mM imidazole, and supplemented with EDTA-free protease inhibitor cocktail tablets and DNase I (both Roche Diagnostics). Cells were lysed using a cell disrupter (Constant Systems) at a pressure of 1.35Kbar, and the cell debris was removed by centrifugation at 50,000x*g* for 1h at 4°C. The supernatant was filtered (0.22μm pore size; Millipore), loaded onto a HiTrap Chelating HP column (GE Healthcare) charged with nickel sulfate, and the fusion protein was eluted using a gradient of buffer A plus 10-500mM imidazole. Subsequently, the sample was dialyzed overnight at 20°C against buffer B (50mM Tris-HCl, 250mM NaCl, 1mM dithiothreitol, pH8.0) in the presence of His_6-tagged TEV proteinase at a proteinase:sample molar ratio of 1:200. This cleavage left peptide G-A-M-D-P at the N-terminus of the protein (13.6-KDa selecase). The digested sample was passed several times through nickel-nitrilotriacetic acid resin (Invitrogen) previously equilibrated with buffer A plus 10mM imidazole to remove His_6-containing molecules. The flow-through was collected, concentrated by ultrafiltration, and further purified by size-exclusion chromatography (SEC) on a HiLoad 26/60 Superdex 75 column (GE Healthcare) previously equilibrated with buffer C (20mM Tris-HCl, 500mM NaCl, pH7.5). Some preparations of 16.5-KDa selecase underwent spontaneous autolytic cleavage in the N-terminal fusion region during purification (Suppl. Fig. 1b), thus giving rise to a variant that contained seven extra N-terminal residues (F-Q-G-A-M-D-P; 13.8-KDa selecase).

Untagged wild-type 13.1-KDa selecase, its single- and multiple-residue point mutants, the C-terminal deletion variant slcΔC4, and the *M. igneus* and *M. fervens* orthologs were purified as follows. After centrifugation at 7,000x*g* for 30min at 4°C, the pellet was washed twice with buffer A and resuspended in the same buffer supplemented with EDTA-free protease inhibitor cocktail tablets and DNase I. Cells were lyzed at 4ºC using a cell disrupter as described above, and the cell debris was removed by centrifugation at 50,000x*g* for 1h at 4°C. The supernatant was incubated for 1h either at 65°C (wild type) or 50ºC (mutants, deletion variants and orthologs), and the precipitate was removed by centrifugation at 50,000x*g* for 1h at 4°C. The supernatant was filtered (0.22μm pore size) and dialyzed overnight at room temperature against buffer D (20mM Tris-HCl, 150mM NaCl, pH7.0). The protein was subsequently purified by cation exchange chromatography using a HiLoad 16/10 SP Sepharose HP column (GE Healthcare), and by SEC on a HiLoad 26/60 Superdex 75 column (GE Healthcare), previously equilibrated with buffer C.

Protein identity and purity were assessed by 15% Tricine-SDS-PAGE stained with Coomassie blue and mass spectrometry (MS). Ultrafiltration steps were performed with Vivaspin 15 and Vivaspin 500 filter devices of 5-KDa cut-off (Sartorius Stedim Biotech). Approximate protein concentration was determined by measuring A_{280} in a spectrophotometer (NanoDrop) and the calculated absorption coefficient, $E_{0.1\%} = 0.45$, except for mutant $R^{36}W$ ($E_{0.1\%} = 0.87$) and the ortholog from *M. fervens* ($E_{0.1\%} = 0.53$). N-terminal sequencing through Edman

degradation, peptide-mass fingerprinting of tryptic protein digests, and MS analyses were carried out at the Proteomics Facilities of Centro de Investigaciones Biológicas (CIB; Madrid, Spain) and Vall d'Hebron Institute of Oncology (Barcelona, Spain).

1.2. Proteolytic and inhibitor assays **–** Proteolytic activities were assayed at 37°C in buffer E (50mM MES, 150mM NaCl, pH5.5), buffer F (50mM HEPES, 150mM NaCl, pH7.5) or buffer G (50mM CHES, 150mM NaCl, pH9.5) at 20μM final protein concentration unless otherwise stated. To provide reproducible and comparable activity, selecase was subjected to four consecutive overnight dialysis steps at 4°C against buffer C plus, respectively, 10mM 1,10-phenantroline; 1mM 1,10-phenantroline; 1mM $ZnCl_2$; and 10μM $ZnCl_2$.

Proteolytic activity against the fluorescein conjugates BODIPY FL casein, DQ gelatin, and DQ bovine serum albumin (all from Invitrogen) was measured according to the manufacturer's instructions using a microplate fluorimeter (FLx800, BioTek or Infinite M200, Tecan). Assays with natural protein substrates (at 0.5mg/ml; all from Sigma) included bovine plasma fibronectin, bovine muscle actin, human plasma fibrinogen, cold-water fish-skin gelatin, bovine milk casein, and bovine milk α-casein. In general, reactions were carried out overnight in buffer F at 37°C, 65°C, and 80°C at an enzyme:substrate ratio of 1:5 (w/w) for the first two substrates and 1**:**10 (w/w) for the others. Cleavage was assessed by 15% Tricine-SDS-PAGE stained with Coomassie blue. For the particular case of cleavage of bovine milk α-casein depicted in Suppl. Fig. 1a, the substrate was incubated with 13.6-KDa selecase (at 25μg/ml; peptidase:substrate molar ratio 1:20; identical results for 13.1-KDa selecase, data not shown). PageRuler Plus Prestained Protein Ladder (Fermentas) was used as molecular mass standard. Activity from contaminating TEV peptidase, which is a cysteine peptidase, was ruled out as this protease did not react with α-casein in a control assay (data not shown) and was insensitive to general metallopeptidase (MP) inhibitors, which however completely abolished cleavage (see below and Suppl. Table 1). In addition, proteolytic activity was assayed in a zymogram of bovine milk α-casein (Suppl. Fig. 1b). Samples were prepared without reducing or boiling. Following electrophoresis on 12.5% Glycine-SDS-PAGE, SDS was removed from the gel with two washes in 2.5% (w/v) Triton X-100 for 20min. The zymogram was incubated for 48h at 37°C in buffer F and subsequently stained with Coomassie blue.

Peptidolytic activity was further tested on ten fluorogenic peptide substrates of sequence: Abz-K-D-E-S-Y-R-K(dnp) (Abz, aminobenzoyl; dnp, 2,4-dinitrophenylamino); Abz-T-V-L-E-R-S-K(dnp); Abz-D-Y-V-A-S-E-K(dnp); Abz-Y-G-K-R-V-F-K(dnp); Abz-V-K-F-Y-D-I-K(dnp); Dabcyl-L-A-R-V-E-Edans (Dabcyl, *p*-dimethyl(aminophenyl)azobenzoate; Edans, 2-aminoethylamino-1-naphthalene sulfonate); Abz-G-I-V-R-A-K(dnp); Mca-P-L-G-L-Dap(dnp)-A-R-NH_2 (Mca, 7-methoxycoumarin-4-acetyl; Dap, L-diaminopropionyl); Mca-R-P-K-R-V-E-Nva-W-R-K(dnp)-NH_2 (Nva, norvaline); and Dnp-P-L-G-L-W-A-(*D*)R-NH_2 (all from Bachem; see Ref. [1] for details on the first six substrates). Further activity assays were performed with the fluorogenic peptide substrate Abz-E-L-A-Y-F-Y-P-E-K(dnp) (peptide CCS; purchased from GL Biochem Ltd), which mimics the cleavage site of bovine milk α_{s1}-casein (Suppl. Fig. 1c; see also section 2). Reactions were monitored in a microplate fluorimeter (Infinite M200, Tecan) at enzyme:substrate molar ratios of 1:0.25, 1:0.5, 1:1.25, and 1:6.25. Kinetic parameters of peptide CCS cleavage were obtained (k_{cat} = 4±0.2 x $10^{-4}s^{-1}$; K_m = 3±0.3μM; k_{cat}/K_m = $133M^{-1}s^{-1}$) using the substrate in buffer F at 37°C according to the Michaelis-Menten equation implemented in the SIGMAPLOT v.10.0 program. Peptide CCS was also used to determine the optimal pH for activity (Suppl. Fig. 1d) in buffer 150mM NaCl with 50mM MES for pH5.5-6.5; 50mM HEPES for pH7.0-8.0; and 50mM CHES for pH8.5-9.5, respectively. Carboxypeptidase activity against the chromogenic substrates *N*-(3-[2-furyl]acryloyl)-F-F-OH (Bachem), *N*-(3-[2-furyl]acryloyl)-G-L-A-OH (Bachem), and *N*-(3-[2-furyl]acryloyl)-L-G-P-A-OH (Sigma) was tested using a microplate spectrophotometer (PowerWave XS, BioTek) at enzyme:substrate molar ratios of 1:5 and 1:25. Aminopeptidase activity was assayed with fluorogenic substrates F-Amc (Amc, 7-amino-4-methylcoumarin), (*H*)T-Amc, and Y-Amc (all from Bachem), and with the *p*-nitroanilide (pNA) derivatives of a representative set of natural L-amino acids and peptides (from Bachem): A-pNA, M-pNA, L-pNA, K-pNA, V-pNA, (*H*)I-pNA, (*H*)G-pNA, N-acetyl-F-pNA, A-A-pNA, *N*-benzyloxycarbonyl-V-G-R-pNA, *N*(α)-benzoyl-I-E-G-R-pNA, and *N*-succinyl-A-A-P-F-pNA. Reactions were monitored in a microplate fluorimeter (FLx800, BioTek) or a PowerWave XS spectrophotometer (BioTek) at enzyme:substrate molar ratios of 1:125 and 1:250.

For inhibition assays, selecase (at 2μM in buffer F) was incubated for 30min with inhibitors of different protease classes, and the remaining proteolytic activity on peptide CCS was measured at 37°C (Suppl. Table 1). Activity of metal-substituted selecase variants was assayed using the latter substrate and protein (at 20μM) subjected to, first, metal removal through sequential dialysis against 10mM and 1mM 1,10-phenanthroline and, next, sequential dialysis against 1mM and 10μM of zinc, cobalt, magnesium, manganese, calcium, cadmium, copper, and nickel chloride or sulfate salts (Suppl. Table 2).

1.3. Concentration dependence of proteolytic activity **—** Proteolytic activity against peptide CCS (at 10μM) based on the initial rate of reaction was determined as described in section 1.2 for 13.1-KDa selecase at

0.025, 0.050, 0.075, 0.15, 0.25, 0.35, 0.50, 1.0, 5.0, 10, 20, and 50mg/ml in buffers C and H (20mM Tris-HCl, 150mM NaCl, pH7.5; see Fig. 1a). Aprotinin (from Sigma), which was used as an internal molecular-mass standard in SEC (see next section), did not interfere with catalytic activity (data not shown). TEV proteinase mutant S219V, produced and purified according to [2], was used as a control for a normally behaving peptidase in a concentration-dependent activity assay at 0.025, 0.050, 0.075, 0.15, 0.50, 1.0, 5.0, and 10mg/ml in buffer I (20mM Tris-HCl, 150mM NaCl, 10% glycerol, 1mM dithiothreitol, pH8.0) at 30°C against the fluorogenic peptide substrate Abz-E-N-L-Y-F-Q-E-G-K(dnp) (GL Biochem Ltd) (also at 10μM; see Fig. 1a), which was chosen because it was reported to be cleaved with a similar efficiency to that of selecase with peptide CCS [3]. The activity of this TEV proteinase was adjusted to a hyperbolic curve with program SIGMAPLOT (Fig. 1a). The kinetic parameters were k_{cat} = 0.014±10^{-3}s^{-1}; K_m = 94±22μM; k_{cat}/K_m = 150M^{-1}s^{-1} (compare with values of section 1.2 for selecase; see also [3] for other TEV variants and substrates). In addition, a sample of selecase at 5.0mg/ml, which was only residually active, was diluted to 0.025mg/ml in buffer and analyzed for proteolytic activity against peptide CCS. This sample revealed activity equivalent to that of selecase at 0.025mg/ml that had not been concentrated before (data not shown), thus showing that concentration-dependent inactivation was reversible. Finally, selecase single- and multiple-residue mutants, deletion variant slcΔC4, and the orthologs from *M. igneus* and *M. fervens* were assayed for activity as described above for peptide CCS at 0.25, 0.50, 2.5 and 5.0mg/ml in buffer H.

1.4. Size-exclusion chromatography and multi-angle laser light scattering **—** 13.1-KDa selecase at 0.15, 0.50, 1.0, 5.0, 10, 25, and 50mg/ml in buffers C and H plus 30, 75, 165, 180, 180, 180, and 180μg of aprotinin, respectively, as internal calibrator were analyzed at room temperature by SEC on a Superdex 75 10/300 GL column (GE Healthcare) connected to an ÄKTA Purifier System (GE Healthcare) and previously equilibrated with the same buffer (Suppl. Fig. 2a; curve at 25mg/ml omitted for clarity). Passage through this chromatographic column led to dilution of the samples. Equivalent results were obtained with 13.6-KDa selecase (data not shown). The column was calibrated with the following protein standards: conalbumin (75KDa; 9.99ml), ovalbumin (44KDa; 10.76ml), carbonic anhydrase (29KDa; 11.85ml), ribonuclease A (13.7KDa; 13.33ml), and aprotinin (6.5KDa; 15.18ml). The theoretical migration volumes of monomeric, dimeric and tetrameric selecase were calculated according to the elution volumes of the protein standards and corrected according to the hydrodynamic parameters calculated with the HYDRAPRO program [4] using the coordinates of the respective structures (see section 1.10). The theoretical reference value for an octamer was derived from the calibration curve. In addition, a sample of selecase at 25mg/ml was loaded onto this column previously equilibrated with buffer H, collected at 0.3mg/ml, and subsequently re-loaded at this protein concentration onto the same column. This assay revealed a chromatogram comparable to a protein sample at 0.3mg/ml that had not been concentrated before (data not shown), thus revealing that concentration-dependent oligomerization is reversible. Furthermore, nickel-selecase (see section 1.2) was concentrated to 11mg/ml in the presence of peptide A-Y-F-Y-P (purchased from GL Biochem Ltd.) at 1mM in buffer H and was analyzed by SEC on a Superdex 75 10/300 GL column previously equilibrated with the same buffer and using 180μg of aprotinin as internal calibrator. The chromatogram mainly corresponded to a monomeric species (Suppl. Fig. 2a). Finally, selecase single- and multiple-residue mutants, deletion variant slcΔC4, and the orthologs from *M. igneus* and *M. fervens* were likewise analyzed by SEC at 0.50mg/ml and 5.0mg/ml in buffer H (Suppl. Fig. 5a,b).

13.1-KDa selecase at 0.15, 0.30, 0.63, 1.2, 4.5, 8.5, 22, 46, and 65mg/ml in buffer H was analyzed at room temperature by SEC (same column as above but connected to a Waters Alliance apparatus) coupled to a multi-angle laser light scattering (MALLS) device connected to DAWN HELEOS II and Optilab T-rEX (refractometer with EXtended range) detectors (from Wyatt Technology). SEC-MALLS was similarly performed for nickel-selecase at 0.25, 1.0, 10, and 50mg/ml. Data were processed with ASTRA 6 software for direct determination of the absolute molar mass of samples and the differential refractive index (dRI) (Fig. 1b and Suppl. Fig. 2b). The experiments were performed at UVHCI-EMBL (Grenoble, France).

1.5. Analytical ultracentrifugation **—** Analytical ultracentrifugation (AUC) of 13.1-KDa selecase at 0.03, 0.10, 0.30, 0.90, 3.0, 9.0, and 18mg/ml in buffer H was carried out at 20°C using a Beckman Coulter Optima XL-I analytical ultracentrifuge with an An50 Ti 8-hole rotor at the Analytical Ultracentrifugation and Light Scattering Facility of CIB (Madrid, Spain). AUC was similarly performed for nickel-selecase at 0.03, 0.10, 0.30, 0.90, and 3.0mg/ml. Sedimentation velocity experiments were performed at 48,000rpm, with buffer density ρ=1.00499g/ml and viscosity η=0.010214Pose (V_{bar} = 0.7655ml/g), and the data were analyzed using program SEDFIT v14.1 ([5]; Suppl. Table 3). Sedimentation equilibrium data were acquired at 11,000, 16,000, and 20,000rpm and analyzed using HeteroAnalysis v1.1.44 software ([6]; Suppl. Table 3)

1.6. Chemical crosslinking experiments **—** 13.1-KDa selecase at 0.15, 0.50, 2.0, 5.0, 10, 20, and 40mg/ml in buffer F was incubated with a collection of crosslinking reagents, namely glutaraldehyde, formaldehyde, 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC; exceptionally in buffer E), dimethyl

adipimidate (DMA), dimethyl pimelimidate (DMP), and bis(sulfosuccinimidyl)suberate (BS^3) (the first two from Sigma, the others from Pierce) at different protein:crosslinker ratios (1-to-20-fold molar excess of crosslinker), at different temperatures (4°C, room temperature, 37°C, and 50°C), and for different time periods (10min-12h). The crosslinking solutions were always prepared immediately before use. The reactions were quenched with denaturing loading buffer and analyzed by 15% Tricine-SDS-PAGE stained with Coomassie blue.

1.7. Circular dichroism spectroscopy — 13.1-KDa selecase containing either zinc or nickel in the active site (see section 1.2) were prepared at 0.10mg/ml and 0.50mg/ml in buffer H for far-UV (180-260nm) and near-UV (250-330nm) circular dichroism spectroscopy. Measurements were carried out in a Jasco J-815 spectrometer at 20°C using 1-mm and 1-cm path length cells for far-UV and near-UV, respectively. The obtained spectra were analyzed with program Spectra Analysis v2.09.03.

1.8. Small-angle X-ray scattering — 13.1-KDa selecase at 0.15, 0.30, 0.63, 1.2, 1.9, 4.5, 8.5, 15, 22, 46, and 65mg/ml in buffer H was analyzed by small-angle X-ray scattering (SAXS). SAXS data were collected at 20°C on a Pilatus 1M pixel detector (from Dectris) at beam line BM29 [7] of the European Synchrotron Radiation Facility (ESRF, Grenoble, France) within the Block Allocation Group "BAG Barcelona." SAXS was similarly performed with nickel-selecase at 0.15, 0.25, 0.50, 1.0, 2.0, 5.0, 10, 25, and 50mg/ml. Data were recorded using a robot sample changer at a sample-to-detector distance of 2.85m, thus covering the range of momentum transfer $0.0036<s<0.501Å^{-1}$ ($s = 4\pi\sin\theta/\lambda$, where 2θ is the scattering angle and $\lambda = 0.991Å$ is the X-ray wavelength). To assess radiation damage, ten successive one-second exposures were compared and no significant changes were observed (data not shown). Buffer subtraction was performed using standard protocols with program PRIMUS [8]. The forward scattering $I(0)$, as well as the radius of gyration (R_g), were calculated using the Guinier approximation assuming that, at very small angles ($s<1.3/R_g$), the intensity, $I(s)$, can be represented as $I(s)=I(0)\cdot\exp(-(sR_g)^2/3)$. The pair-distance distribution function, $P(r)$, from which the maximum particle dimensions (D_{max}) were estimated, was computed with program GNOM [9]. See Suppl. Table 4.

Single Value Decomposition (SVD) analysis as implemented in program MATLAB (MathWorks, Natick, Mass.) was used to assess the number of species required to describe the SAXS dataset of selecase at the distinct concentrations. The eigenvalues, the shape of the eigenvectors, and the capacity to describe the complete dataset with an increasing amount of species were used as criteria to assess the number of species present. Low-concentration curves (0.15, 0.30, and 0.63mg/ml) were univocally identified as corresponding to a monomer-dimer equilibrium with program OLIGOMER [10] using the crystal structures reported in the present work (slc_1 and slc_2; see sections 1.9 and 1.10). Computation of the theoretical scattering profiles from the crystallographic models was performed with program CRYSOL [11].

1.9. Crystallization and data collection — Crystallization screenings were performed at the IBMB/IRB Crystallography Platform (PAC, Barcelona) by the sitting-drop vapor diffusion method using 96x2-well MRC plates (Innovadyne). A TECAN Freedom EVO robot was used to prepare reservoir solutions, and a Cartesian Microsys 4000 XL (Genomic Solutions) robot or a Phoenix/RE (Art Robbins) robot was used for nanodrop dispensing. Crystallization plates were stored in Bruker steady-temperature crystal farms at 4ºC or 20°C. Successful hits were scaled up to the microliter range with 24-well Cryschem crystallization dishes (Hampton Research) whenever possible. Both wild-type selecase variants assayed (13.1-KDa- and 13.6-KDa-selecase) behaved equivalently in crystallization studies. Dimeric wild-type selecase (slc_2) and tetrameric wild-type selecase (slc_4) crystallized from protein solutions at 54mg/ml and 9.0-100mg/ml in buffer C, respectively. Best crystals of slc_2 were tetragonal and appeared at 20°C in equivolumetric drops containing protein solution and 100mM sodium cacodylate, 30% (v/v) 2-methyl-2,4-pentanediol, 3% (w/v) polyethylene glycol 8,000, pH6.5 as reservoir solution. Crystals of slc_4 were hexagonal and appeared at 16-20°C in equivolumetric drops containing protein solution and several reservoir solutions, among them 100mM Bis-Tris propane, 200mM sodium citrate, 18-24% (w/v) polyethylene glycol 3,350, pH7.0-9.0 as the best. Monomeric wild-type slc_1 yielded orthorhombic crystals at 20°C in equivolumetric drops containing nickel-selecase protein solution (obtained as mentioned in section 1.2; see also section 2) at 11mg/ml in buffer C with pentapeptide A-Y-F-Y-P (GL Biochem Ltd.) at 5mM, and 100mM Tris-HCl, 200mM sodium acetate, 30% (w/v) polyethylene glycol 4,000, pH8.5; or 100mM Tris-HCl, 30% (w/v) polyethylene glycol 3,500, pH8.5 as reservoir solution. Monoclinic crystals of dimeric selecase mutant $R^{36}W$ ($slc_{2'}$) were serendipitously obtained directly from protein solution at 5.0mg/ml in buffer H stored at 4ºC for one month. Crystals of orthorhombic dimeric selecase mutant $I^{100}F+H^{107}F$ ($slc_{2''}$) were obtained at 5.0-11mg/ml in equivolumetric drops with several different reservoir solutions, both at 4ºC and 20ºC (all crystals tested corresponded to the same crystal form). The best conditions contained 100mM sodium cacodylate, 1M sodium acetate trihydrate, pH6.5 as reservoir solution. All crystals were cryo-protected by rapid passage through drops containing increasing concentrations of glycerol (up to 20% (v/v)).

Complete diffraction datasets were collected at 100K from liquid-N_2 flash cryo-cooled crystals (Oxford Cryosystems 700 series cryostream) on a MarCCD detector (from Marresearch) at beam line ID23-2 of ESRF

(Grenoble, France) within the Block Allocation Group "BAG Barcelona" (slc_1, $slc_{2'}$ and slc_4) or on a Pilatus 6M pixel detector (from Dectris) at beam line XALOC of synchrotron ALBA (Barcelona, Spain; slc_2 and $slc_{2''}$). All crystal types contained one molecule per asymmetric unit except the mutant ones, which harbored four ($slc_{2'}$; two dimers) and two molecules ($slc_{2''}$; one dimer). Diffraction data were integrated, scaled, merged, and reduced with programs XDS [12] and XSCALE [13] or SCALA [14] within the CCP4 suite of programs [15] (see Suppl. Table 5 for data processing statistics).

1.10. Structure solution and refinement — The structure of tetragonal slc_2 was the first to be solved, by likelihood-scoring molecular replacement using program PHASER [16] and a searching model obtained by trimming the side chains of proabylysin (PDB 4JIX; [17]) with CHAINSAW [18] according to a sequence alignment performed with MULTALIN [19]. The final refined solution corresponded to Euler angles α=165.1º, β=83.7º, γ=25.8º and fractional cell coordinates x=0.52, y=0.18, z=-0.14 and had initial Z-scores for the rotation and translation functions of 5.7 and 12.5, respectively. Subsequent model building with the COOT program [20] alternated with crystallographic refinement with PHENIX [21] and BUSTER/TNT [22], which included TLS refinement, until the final refined slc_2 model was obtained. The latter consisted of residues D^3-K^{110}, one zinc ion, three glycerol molecules, and 30 solvent molecules.

The structure of hexagonal slc_4 was solved thereafter with PHASER and the coordinates of the protein part only of tetragonal slc_2 fragment D^3-K^{109} excepting loop Y^{76}-I^{81}. The final refined solution corresponded to Euler angles α=10.2º, β=27.5º, γ=154.3º and fractional cell coordinates x=0.46, y=0.59, z=0.17, and had initial Z-scores for the rotation and translation functions of 3.5 and 8.3, respectively. Visual inspection of the rotated and translated molecule with COOT revealed clashes and regions lying outside of the Fourier map for the C-terminal part of the model, so the first refinement step was performed with fragment D^3-K^{75} only. This was followed by a density modification and model extension step with the AUTOBUILD protocol of PHENIX [23], which produced an improved Fourier map. Thereafter, model completion and refinement proceeded as for slc_2. The final model of slc_4 comprised residues K^2-K^{109}, one zinc cation, one tentative sodium cation, four glycerol molecules, and one solvent molecule. The sodium ion does not play an essential structural or functional role and its—tentative—presence may be attributed to the crystallization conditions.

The structure of orthorhombic slc_1 was also solved with PHASER using protein fragment D^3-K^{75} of slc_2 as a searching model. The final refined solution corresponded to Euler angles α=266.3º, β=125.9º, γ=134.1º and fractional cell coordinates x=-0.32, y=0.35, z=0.31, and had initial Z-scores for the rotation and translation functions of 6.6 and 7.9, respectively. A Fourier map calculated with the appropriately rotated and translated model was then subjected to density modification and model extension with both ARP/wARP [24] and the AUTOBUILD routine of PHENIX. Models were then built based on both resulting Fourier maps. Subsequently, model completion and refinement proceeded as for slc_2 and slc_4. The pentapeptide used to maintain selecase in its monomeric quaternary structure (see sections 1.9 and 2) was apparently bound with very low occupancy, so in the final Fourier map the corresponding region was conservatively interpreted as housing solvent molecules. Therefore, the final model of slc_1 comprised protein residues M^1-K^{109}, one nickel cation mimicking the active-site zinc, one glycerol molecule, and 58 solvent molecules.

The monoclinic structure of dimeric protein variant $slc_{2'}$ (mutant $R^{36}W$) was solved by likelihood-scoring molecular replacement with slc_2 fragment D^3-Y^{76} using program MOLREP [25] taking advantage of a local two-fold axis revealed by a previously calculated self rotation function at angles θ=170º, φ=0º, χ=180º, respectively (peak height 78% of origin peak). A Fourier map calculated with the four appropriately rotated and translated polypeptide chains in the asymmetric unit was then subjected to density modification and model extension with the AUTOBUILD routine of PHENIX. Subsequently, model completion and refinement proceeded as mentioned above. The final model of $slc_{2'}$ contained residues K^2-K^{109} of molecule A and M^1-K^{109} of molecule B, corresponding to one dimer, and R^4-K^{109} of molecule C and D^3-K^{109} of molecule D, the latter two corresponding to the second dimer. In addition, four catalytic zinc ions, four glycerol molecules, one tentative chloride anion, and 149 solvent molecules completed the model. The chloride ion does not play an essential structural or functional role and its—tentative—presence may be attributed to the crystallization conditions. Dimer AB was better defined by the final Fourier map than dimer CD, as revealed by the respective average thermal displacement parameters (A/B: 47.0/50.0Å^2; C/D: 81.4/71.5Å^2). Superposition of both dimers with program COOT revealed they are essentially identical, with a core *rmsd* of 0.43Å, so presentation of results and discussion will center on dimer AB if not otherwise stated.

Finally, the orthorhombic dimeric structure of $slc_{2''}$ (mutant $I^{100}F$+$H^{107}F$) was likewise solved with PHASER using protein fragment D^3-K^{75} of slc_2 as a searching model. The two final refined solutions corresponded to Euler angles α=83.8º, β=63.9º, γ=135.9º and α=191.4º, β=128.1º, γ=319.7º, and to fractional cell coordinates x=0.08, y=0.01, z=0.98 and x=0.72, y=0.75, z=0.29, respectively, and had initial Z-scores for the rotation/translation functions of 5.3/4.7 and 8.2/7.6, respectively. A Fourier map calculated with the appropriately

rotated and translated models was then subjected to density modification and model extension with ARP/wARP. Thereafter, model completion and refinement proceeded as mentioned above. The final model of $slc_{2''}$ comprised residues M^{1}-Q^{108} of both molecules A and B, two zinc cations, five acetate anions, one glycerol, and 251 solvent molecules. Superposition of both monomers with program COOT revealed they are essentially identical, with a core *rmsd* of 0.37Å.

Suppl. Table 5 provides a summary of the final model refinement processes.

1.11. Bioinformatics — Figures were prepared with the CHIMERA program [26]. Close contacts (<4Å) were determined with CNS [27]. Interaction surfaces and theoretic values of the standard Gibbs free energy (ΔG) were computed with the PISA program [28] at http://www.ebi.ac.uk/pdbe/pisa with default parameters. Quaternary analyses were likewise performed with PISA. Structural superpositions were performed with the SSM routine [29] within COOT. Surface complementarity was computed with program SC [30] within CCP4 with default parameters and protein residues only. Local configurational frustration was analyzed with the FRUSTRATROMETER [31] at http://lfp.qb.fcen.uba.ar/embnet. Inter-domain flexibility was determined with HINGEPROT [32] with parameters GNM=20 and ANM=36 at http://www.prc.boun.edu.tr/appserv/prc/hingeprot/index.html. Model validation was performed with MOLPROBITY [33], COOT, the WHATCHECK routine of WHATIF [34], and the wwPDB Validation Server at http://wwpdb-validation.wwpdb.org/validservice. Secondary structure prediction was performed with JPRED3 [35]. The final coordinates of slc_1, slc_2, $slc_{2'}$, $slc_{2''}$, and slc_4 have been deposited with the PDB (access codes 4QHF, 4QHG, 4QHH, 4QHI, and 4QHJ, respectively).

2. SUPPLEMENTAL RESULTS AND DISCUSSION

Overall, four N-terminally differing variants of wild-type selecase were obtained (see section 1.1): 16.5-KDa selecase, which contained 28 extra residues including a His_6-tag and a TEV proteinase recognition sequence preceding the first methionine of the natural protein; 13.8-KDa selecase, which resulted from occasional autolytic cleavage of 16.5-KDa selecase and had the extra N-terminal residues F-Q-G-A-M-D-P (see below); 13.6-KDa selecase, which was produced by TEV proteinase processing of 16.5-KDa selecase and had extra residues G-A-M-D-P; and the natural 13.1-KDa selecase, which did not have any additional residues. All variants showed equivalent proteolytic activity and susceptibility towards inhibitors, as well as crystallizability (here, only 13.1-KDa and 13.6-KDa selecase variants were tested). Cleavage activity for selecase was only detected with bovine milk α_{s1}-casein, at bond Y_{159}-F_{160} (see UP P02662) within segment E-L-A-Y-↓-F-Y-P (Suppl. Fig. 1a,b), as determined by N-terminal sequencing. This segment is close to an epitope responsible for allergenicity (N_{154}-F_{168}) of this casein chain, and proteolysis herein decreased immuno-reactivity of this protein [36]. This points to a potential biotechnological application of selecase in the treatment of alimentary casein and in other casein-related applications. Moreover, 16.5-KDa selecase underwent spontaneous autolysis within the TEV proteinase recognition sequence in some preparations at a site (E-N-L-Y-↓-F-Q-G) that is similar to the casein cleavage site. However, while a fluorogenic peptide mimicking the casein cleavage site (peptide CCS) was likewise efficiently cleaved ($k_{cat}/K_m = 133 M^{-1}s^{-1}$), preferentially at pH values close to neutrality (see section 1.2 and Suppl. Fig. 1c,d), fluorogenic peptides encompassing two sequences recognized by TEV proteinase (E-N-L-Y-F-Q-S and E-N-L-Y-F-Q-E) were not cleaved by selecase, which indicates that autolytic cleavage occurs only slowly and at high concentrations. We conclude that selecase is a selective and specific caseinolytic endopeptidase.

Selecase active-site mutant $E^{70}A$, which ablated the catalytically essential general base/acid glutamate of the zinc-binding signature (H-E-X-X-H) that is characteristic of most MPs [37], was inactive against casein and peptide CCS (Suppl. Fig. 1a-c). General MP inhibitors (EDTA, 1,10-phenanthroline and excess zinc; see Suppl. Fig. 1a,b and Suppl. Table 1) and some matrix metalloproteinase/ADAM inhibitors (batimastat and MMP inhibitor III were the best) abolished selecase activity (Suppl. Table 1). Metals other than zinc did not restore activity of metal-depleted apo-selecase with the exception of cobalt, which enhanced its activity (Suppl. Table 2) as reported for other MPs such as carboxypeptidase A, thermolysin, and astacin [38]. In particular, nickel-selecase was inactive. These studies show that selecase is a MP.

Selecase was characterized for concentration-dependent oligomerization behavior in solution using a battery of complementary biophysical techniques: SEC (see section 1.4), SEC-MALLS (section 1.4), AUC (section 1.5), chemical crosslinking followed by SDS-PAGE (section 1.6), circular dichroism spectroscopy (section 1.7), and SAXS (section 1.8):

The peak pattern of selecase analyzed by both calibrated SEC and SEC-MALLS moved towards smaller elution volumes with increasing protein concentration, indicating ordered protein oligomerization but not indiscriminate aggregation or precipitation (Suppl. Fig. 2a,b). In SEC, only concentrations at or below ~0.2mg/ml contained essentially monomeric selecase (Suppl. Fig. 2a). In SEC-MALLS, the calculated absolute mass of the different peaks revealed two main average values of ~25KDa and ~80KDa, which would be consistent with dimeric and octameric oligomers, respectively. However, the substantial difference in the elution volumes of protein at 0.15mg/ml and 1.2mg/ml indicated the presence of additional species such as monomeric selecase at lower concentrations. In addition, the amplitude of the peaks at higher concentrations suggested the presence of tetramers. SEC-MALLS experiments performed in parallel with nickel-selecase showed essentially the same results (data not shown).

AUC sedimentation velocity and sedimentation equilibrium analyses determined the population of species and average mass at the different protein concentrations, with predominance of monomers at 0-0.3mg/ml, dimers at 0.3-2mg/ml, tetramers at 2-6mg/ml, and octamers at >6mg/ml. Sedimentation velocity experiments performed using zinc-selecase at 0.03-0.1mg/ml showed a majority species with sedimentation coefficient of 1.5S. This value was initially assigned by the software to dimeric selecase. However, in zinc-containing proteins this value would correspond to monomeric selecase as previously reported [5, 39]. This was further confirmed by the use of nickel-selecase at this same concentration range, which resulted in a coefficient of 1.3S that unequivocally corresponds to a monomer. As in the sedimentation velocity experiments, the use of nickel-selecase was more conclusive in the equilibrium velocity experiments than zinc-selecase in showing the prevalence of monomeric protein at low concentrations, with average masses of 14.6KDa and 15.4KDa at 0.1mg/ml and 0.3mg/ml, respectively (see also Fig. 1c and Suppl. Table 3).

Overall, crosslinking experiments followed by SDS-PAGE revealed the presence of monomers, dimers, monomers linked to dimers, and tetramers. Formaldehyde, EDC, DMA, DMP, and BS^3 showed the presence of monomers and dimers in reactions with protein concentration ≥0.5mg/ml. Formaldehyde and BS^3 further showed

the formation of a monomer linked with a dimer, and of a tetramer (see Suppl. Fig. 2c). Unfortunately, selecase tended to precipitate at high protein and crosslinker concentrations (20mg/ml and 5-fold molar excess, respectively), which prevented detection of higher oligomeric species than tetramers and the overall quantitative assessment of concentration-dependent oligomerization by this technique. We attribute this precipitation to the high number of lysines in selecase (17 out of 110 residues), which are targeted by crosslinking reagents.

Circular dichroism spectra displayed the typical shape (minima at 208nm and 222nm) of a well-folded, mostly α-helical protein. In addition, zinc- and nickel-selecase showed very similar spectra, thus confirming that metal replacement had no significant influence in the overall conformation of the protein (Suppl. Fig. 2d). Oligomerization could not be followed by this technique as protein concentrations higher than 0.50mg/ml gave an increase in noise (HT[V]) that precluded measurements (CD[mdeg]) at the far-UV region. The near-UV region, which is sensitive to changes in tertiary structure [40], was also explored but did not show any significant signal (data not shown). This is consistent with the protein not adopting a single structure, which would be compatible with the different oligomeric populations of selecase at 0.50mg/ml.

SAXS experiments (Fig. 1d, Suppl. Table 4 and Suppl. Figs. 3-4), in turn, revealed that even at high concentrations (up to 65mg/ml) the protein did not aggregate. The experimental curves showed systematic variations in data when increasing protein concentration, i.e. the values for the radius of gyration (R_g), the forward scattering ($I(0)$/concentration), and the maximum dimension of the particle (D_{max}) displayed a continuous growth, which confirmed that in solution the relative population of the oligomeric species was dependent on concentration. After Single Value Decomposition (SVD) analysis of the complete SAXS dataset, inspection of the eigenvector shapes and eigenvalues, as well as successive improvement of data description with an increasing number of eigenvectors indicated the presence of four species, which were assigned to monomers, dimers, tetramers, and octamers. In particular, the curves at the three lowermost concentrations assayed (0.15, 0.30, and 0.63mg/ml) indicated a mixed population of monomers and dimers based on the crystallographic structures slc_1 and slc_2 (see section 1.10 and below). A satisfactory description of the three curves, with χ^2 values of 0.78, 0.83, and 0.83, respectively, was obtained with monomer fractions of 0.70±0.10, 0.69±0.04, and 0.35±0.02, respectively (Fig. 1c). This result largely coincides with those of AUC at low concentrations (see Suppl. Table 3). Experimental SAXS curves at higher concentrations (1.2–65mg/ml) could not be interpreted in terms of the crystallographic structures due to the non-crystallized octamers present in solution. Analysis of SAXS curves measured for nickel-selecase rendered very similar results to those of zinc-selecase (data not shown). No signs of aggregation were observed here either at concentrations up to 50mg/ml. The SAXS profiles showed a concentration-dependent increase in the derived values of R_g, D_{max}, and $I(0)$/concentration. The SVD analysis indicated a mixture of monomers, dimers, tetramers, and octamers. Finally, the OLIGOMER analysis of the first five curves (from 0.15mg/ml to 2.0mg/ml) could be properly described as a combination of slc_1 and slc_2 with relative populations for the monomer ranging from 0.94±0.06 to 0.23 ±0.01 at the minimum and maximum concentrations, respectively.

Taken together, biophysical approaches in solution converged at the presence of discrete monomers, dimers, tetramers, and octamers, with higher concentrations leading to higher degrees of oligomerization but not indiscriminate aggregation or precipitation. This was consistent with the protein being concentrated in ultracentrifugal devices beyond 130mg/ml without precipitation. Concentrations at which monomeric selecase was predominant coincided with those of maximal enzymatic activity (0.2-0.3mg/ml; Fig. 1a), thus indicating that the monomer is the active species and that oligomers correspond in all cases to self-inhibiting species. This explained why higher enzyme concentrations yielded lower activity (Fig. 1a). Finally, simple dilution with buffer reversed oligomerization to yield monomers and restore activity.

To identify the molecular determinants of this oligomeric behavior, we crystallized and solved the structure of wild-type selecase (Suppl. Table 5). It has been reported that crystalline forms of the MP carboxypeptidase displayed enzymatic activity [41], so crystal structures of enzymes adequately represent the active forms in solution. We obtained three crystal forms—orthorhombic, tetragonal, and hexagonal—, which serendipitously corresponded, respectively, to monomeric (slc_1), dimeric (slc_2) and tetrameric (slc_4) forms of selecase. This indicated that at least three of the oligomerization states found in solution had a counterpart in the form of a stable, isolatable species, each one favored by particular crystallization conditions, which would mimic the effect of "growing" concentration and drive the system into the different oligomeric states [42]. The concentrations at which monomeric selecase was found in solution were insufficient for crystallization studies (generally >1-2mg/ml). To overcome this and obtain crystals of the monomeric form we replaced the catalytic zinc ion with divalent nickel, which rendered the enzyme inactive (see above and Suppl. Table 2). In general, this metal replacement does not affect the overall active-site geometry in MPs due to the shared divalent charge and similar ionic radii (nickel, 0.69Å; zinc, 0.74Å; see e.g. [38b]), so monomeric nickel-selecase can be considered as a valid structural model for the monomeric active zinc enzyme. This was supported by circular dichroism

experiments, analytical ultracentrifugation, SEC-MALLS, and SAXS, which indicated equivalence of zinc- and nickel-selecase (see above). We then incubated diluted monomeric nickel-selecase with an excess of a pentapeptide spanning the casein cleavage sequence (A-Y-F-Y-P) to saturate the active-site cleft and maintain its overall competent conformation, and subsequently concentrated the sample. This strategy yielded monomeric selecase at 11mg/ml, a concentration at which the protein is normally oligomeric (Suppl. Fig. 2a).

The crystal structures of active monomeric wild-type selecase (slc_1), and inactive dimeric (slc_2) and tetrameric wild-type selecase (slc_4; see Suppl. Table 5) show that both the N- and the C-terminus are flexible and surface located, thus explaining why N-terminally differing selecase variants give rise to equivalent enzymatic activity and crystals. In addition, interactions important for molecular cohesion in slc_1 are provided by a charge-relay system involving salt bridges K^2-D^{60}, E^{17}-R^{64}, and E^{61}-R^{64} (all within a so-called N-terminal sub-domain (NTS), see main text), which is stabilized by the aliphatic side-chain part of K^2 residing on a hydrophobic pillow made by L^7, L^{55}, and I^{63}. Other cohering contacts include the central hydrophobic core of the protein (Suppl. Table 6), salt bridges E^{62}-K^{105} and K^{75}-E^{87} (both at the interface between NTS and the so-called C-terminal sub-domain (CTS), see main text), and hydrogen bonds at the NTS-CTS interface between the side chains of R^{64} and K^{105} and the main-chain carbonyl oxygens of L^{91} and Y^{57}, respectively. In slc_2, the N-terminal stretch is somewhat more flexible than in slc_1, as a result of which the charge-relay system is absent. Instead, D^{60} interacts with R^{64}, which becomes reoriented for its side chain and no longer bridges E^{61} with E^{17}. This, in turn, causes E^{61} to hydrogen-bond N^{94}. In addition, major differences are found at N^{22}-N^{26}, which is part of Lα1β1 in both slc_1 and slc_2. In slc_4 the same differences as in slc_2 are observed with respect to the NTS charge relay chain of slc_1. There is just one additional interaction between R^{64} and N^{13} when compared with slc_2. Given that both slc_2 and slc_4 are catalytically inert as inferable from the conformation of their active sites (see main text), the two constellations found for this charge relay in slc_1 and slc_2/slc_4 could make the difference between competence and incompetence, respectively, in wild-type selecase. In addition, major differences are found in slc_4 at L^{20}-N^{26}, which in the tetrameric protein adopts a helical structure (helix αA, see Suppl. Table 7). However, given that this loop also differs from slc_1 in slc_2, we attribute these differences to intrinsic variability of this surface-located region.

The major conformational rearrangement observed in the three wild-type selecase structures mainly affects the CTS and the central trigger point is the sub-domain interface at G^{77}. In all cases, the inactive species are characterized either by non-functional metal-binding sites and/or blocked active-site clefts—mainly on their non-primed sides—by repositioning of the C-terminal helices, which would hamper proper binding of peptide substrates. Interestingly, the position and side chain conformation of upstream residue Y^{76} are still very similar in all three structures. This indicates that selecase is composed of two basic folding units (foldons), respectively spanning the NTS and the CTS, so the mainly hydrophobic interface between sub-domains would act as a zipper to trigger reversible conformational rearrangements. The high mobility of the CTS was further confirmed by computational analysis of local conformational frustration, which identifies sites of mobility of a protein around its native basin [31]. These calculations identified segments spanning almost the entire CTS as highly frustrated in the three different wild-type selecase monomers (Suppl. Fig. 7). Consistently, analysis of inter-domain flexibility based on the elastic network model revealed potential hinge motions at the sub-domain junctions of each of the three structures at K^{75}-Y^{76} (data not shown).

The structural bases of the transitions between competent monomeric and incompetent dimeric/tetrameric autoinhibitory assemblies of wild-type selecase are consistent with calculated geometric and thermodynamic parameters (see Suppl. Table 8). In accordance with enzymatic activity corresponding to the most stable structure, slc_1 has the lowest calculated solvation free energy of folding (ΔG=–109kcal/mol; [43]), which lies in the range described for well-folded structures (~–1kcal/mol per residue; [43]). Compactness, which further contributes to stability, is reflected by the slc_1 monomer having the smallest accessible surface of all three conformers (6,999Å^2). The slc_2 protomer shows a similar value of ΔG (–105kcal/mol) but a larger overall surface (7,274Å^2; 4% larger than in slc_1). Here, the incompetent quaternary structure is stabilized through dimerization, which yields a value of ΔG of –217kcal/mol for the dimer. This is more than twice the value of the slc_2 monomer and coherent with a value of ΔG_{diss}, the free energy of assembly dissociation [28], of 24kcal/mol, which, in turn, is consistent with typical free energies of dimerization in protein-protein complexes ranging between –11 and –22kcal/mol [44]. Finally, the slc_4 protomer has the lowest value of monomeric ΔG (–92kcal/mol) and largest accessible surface (8,696Å^2; 24% larger than in slc_1). Its overall stability is provided by the tetrameric arrangement and reflected by ΔG and ΔG_{diss} values of –424kcal/mol (4.6 times higher than that of the slc_4 monomer) and 14kcal/mol, respectively.

Given the importance of the C-terminal helix α4 and loop Lβ1β2 in the dimerization of slc_2 and tetramerization of slc_4 (see main text), we selected residues R^{36}, I^{100}, I^{103}, and H^{107}, which had been observed to participate in dimerization in slc_2 and tetramerization in slc_4, and generated a total of seven single, double and triple point mutants (R^{36}W, I^{100}F, I^{103}E, H^{107}F, I^{100}F+H^{107}F, I^{103}E+H^{107}F, I^{100}F+I^{103}E+H^{107}F) to try to ablate the

interactions responsible for oligomerization and thus obtain monomeric forms. In addition, we constructed two deletion variants affecting α4, respectively lacking four (slcΔC4) and eight (slcΔC8) C-terminal residues. All protein variants were produced, purified, and concentrated similarly to the wild type except for slcΔC4, which was obtained with lower yields and could only be maximally concentrated to 5.0mg/ml, and slcΔC8, which was insoluble under the expression conditions tested and was thus discarded for further studies (see section 1.1). This finding pointed to a stabilizing effect of helix α4 on the whole protein despite its overall flexibility in the distinct structures analyzed (see main text). Inspection of the elution profiles of mutant and deletion variants in calibrated SEC at low concentration (0.50mg/ml) revealed no significant difference with the monomer-dimer equilibrium of the wild-type (Suppl. Fig. 5a). At high concentrations (5.0mg/ml; Suppl. Fig. 5b), all species displayed variable populations of dimers, tetramers, and octamers, with the notable exception of selecase mutants $H^{107}F$ and $I^{100}F+H^{107}F$, which were mainly tetramers. Cleavage assays at 0.25mg/ml and 5.0mg/ml, in turn, revealed activity similar to the wild type of all forms except for mutants $H^{107}F$ and $I^{100}F+H^{107}F$, which were less active because their oligomerization equilibrium is shifted towards tetramers at lower concentrations than the other forms (Suppl. Fig. 5c,d). In all cases, however, the activity at high concentration was lower than that at low concentration, as observed for the wild type. Accordingly, although mutation of the residues engaged in oligomerization in wild-type selecase may produce a certain rearrangement in the distribution of the distinct oligomers, the four species are still observed and this is reflected by diminished activity upon concentration increase as in the wild type. Taken together, these results indicate that selecase is highly plastic, so it can adapt to potentially deleterious mutations affecting segments engaged in oligomerization and retain its capacity to oligomerize. This flexibility even tolerates ablation of four residues of the C-terminal helix, but ablation of eight residues destabilizes the protein and renders it insoluble. This is reminiscent of the related minigluzincin, proabylysin [17]. Out of all these variants, we managed to crystallize mutants $R^{36}W$ (termed $slc_{2'}$; section 1.9) and $I^{100}F+H^{107}F$ (termed $slc_{2''}$; section 1.9) and solve their structures (section 1.10 and Suppl. Table 5). Interface analysis with program PISA (see 1.11) clearly suggested that the oligomerization state of $slc_{2'}$ was a dimer, with an interface area (dimer AB) of 1,152Å^2, an Δ^iG value (solvation free energy gain upon dimerization) of –15.4kcal/mol, a Δ^iG P-value (probability of getting a lower Δ^iG value) of 0.085, and a CSS score (complexation significance score; indicates for the contribution of an interface to assembly formation) of 0.818. The same analysis for $slc_{2''}$ likewise clearly suggested a dimer, with an interface area of 950Å^2, an Δ^iG value of –17.9kcal/mol, a Δ^iG P-value of 0.028, and a CSS score of 0.636. Structural details are discussed in the main text.

We further identified two close selecase orthologs from *M. igneus* and *M. fervens*, which had not been found in previous searches for minigluzincins [17]. These were much closer to selecase (81% and 95% sequence identity, respectively) than any other of the minigluzincin orthologs previously studied and could actually be considered as natural fivefold and 19-fold point mutants, respectively (Suppl. Fig. 6). We produced and purified them similarly to wild-type selecase (section 1.1), and analyzed their concentration-dependent oligomerization behavior and activity, as with the aforementioned selecase point mutants. They likewise evinced concentration-dependent oligomerization, which was similar to selecase at low concentration (0.50mg/ml) for the *M. fervens* ortholog but rather displaced towards dimers in the *M. igneus* form (Suppl. Fig. 5a). At high concentrations (5.0mg/ml), both were mainly dimeric-tetrameric (Suppl. Fig. 5b). Interestingly, while both showed less activity against peptide CCS at high concentration than at low concentration as in selecase and its mutants (Suppl. Fig. 5c,d), the *M. igneus* ortholog was much more active than the *M. fervens* form. This is consistent with the lower degree of identity with selecase of the latter than the former, thus pointing to potential differences in substrate specificity and activity.

Finally, despite overall structure and sequence similarity of selecase with projannalysin and proabylysin (35% and 40% respective identity), the latter two minigluzincins could only be produced, isolated, and characterized as single conformers. These corresponded to inactive zymogens in which access of substrates to the cleft was blocked [17]. These zymogens were self-inhibited in two distinct and novel manners and any attempts to produce mutants and variants that would show activity or conformational changes failed despite extensive trials, so a function distinct to proteolysis for these two proteins cannot be ruled out. Interestingly, proabylysin is maintained latent inside the molecule by its own C-terminal stretch, which penetrates the—overall competent—active-site cleft in extended conformation and contacts the catalytic metal through its C-terminal carboxylate group. In contrast, projannalysin—also in an overall competent conformation for its metal site and cleft—forms a dimer that is similar to that of slc_2. In the former, however, the C-terminal carboxylate approaches the metal—as in proabylysin—and there is no ligand swap as in slc_2. Overall, this points to a unique mechanism for latency maintenance via oligomerization through a fold-switch capacity of the specific metallocaseinase selecase that is shared with very close orthologs from *M. igneus* and *M. fervens* (~80-95% sequence identity) but not with other more distant homologs (~35-40% sequence identity), not even from the same organism (projannalysin). This system of autoinhibition, together with the highly selective specificity of selecase, has enabled, for the first time

to our knowledge, intracellular overproduction in *E. coli* of a mature active peptidase in soluble functional state without affecting host cell growth.

3. ACKNOWLEDGMENTS

We are indebted to Tibisay Guevara for her outstanding dedication during crystallization experiments and assistance throughout the life of the project, and to Robin Rycroft and two anonymous reviewers for substantial contributions to the manuscript. We are also grateful to Marc Jamin and Francine Gerard for their generous help with SEC-MALLS experiments and further thank the IBMB/IRB Crystallography Platform (PAC) in Barcelona and the Analytical Ultracentrifugation and Light Scattering Facility of CIB in Madrid. We used platforms of the Grenoble Instruct Center (ISBG; UMS3518 CNRS-CEA-UJF-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB). This study was supported in part by grants from European, Spanish, and Catalan agencies (FP7-HEALTH-2010-261460 "Gums&Joints"; FP7-PEOPLE-2011-ITN-290246 "RAPID"; FP7-HEALTH-2012-306029-2 "TRIGGER"; BFU2012-32862; CSD2006-00015; Fundació "La Marató de TV3" grant 2009-100732; 2009SGR1036; and "Pot d'Idees" FGB301793) and FPI Ph.D. fellowships from the former Spanish Ministry for Science and Technology, currently of Economy and Competitiveness, to M.L.-P. and A.C.-P. P.B. acknowledges funds from ANR-CHEX (project SPIN-HD) and ATIP-Avenir. We acknowledge the help provided by ESRF and ALBA synchrotron local contacts. The very rapid SAXS beam-time allocation by ESRF is specially appreciated. Funding for travelling and synchrotron data collection was provided in part by ESRF.

4. SUPPLEMENTAL REFERENCES

[1] T. O. Botelho, T. Guevara, A. Marrero, P. Arêde, V. S. Fluxa, J. L. Reymond, D. C. Oliveira, F. X. Gomis-Rüth, *J. Biol. Chem.* **2011**, *286*, 25697-25709.

[2] J. Phan, A. Zdanov, A. G. Evdokimov, J. E. Tropea, H. K. Peters, 3rd, R. B. Kapust, M. Li, A. Wlodawer, D. S. Waugh, *J Biol Chem* **2002**, *277*, 50564-50572.

[3] R. B. Kapust, J. Tozser, T. D. Copeland, D. S. Waugh, *Biochem. Biophys. Res. Commun.* **2002**, *294*, 949-955.

[4] A. Ortega, D. Amoros, J. Garcia de la Torre, *Biophys. J.* **2011**, *101*, 892-898.

[5] P. Schuck, *Biophysical journal* **2000**, *78*, 1606-1619.

[6] H. Zhao, C. A. Brautigam, R. Ghirlando, P. Schuck, *Current protocols in protein science / editorial board, John E. Coligan ... [et al.]* **2013**, *Chapter 20*, Unit20 12.

[7] P. Pernot, P. Theveneau, T. Giraud, R. Nogueira Fernandes, D. Nurizzo, D. Spruce, J. Surr, S. McSweeney, A. Round, F. Felisaz, L. Foedinger, A. Gobbo, J. Huet, C. Villard, F. P. Cipriani, *J. Phys. Conf. Ser.* **2010**, *247*, 012009.

[8] P. V. Konarev, V. V. Volkov, A. V. Sokolova, M. H. J. Koch, D. I. Svergun, *J. Appl. Cryst.* **2003**, *36*, 1277-1282.

[9] D. I. Svergun, *J. Appl. Cryst.* **1992**, *25*.

[10] M. V. Petoukhov, D. Franke, A. V. Shkumatov, G. Tria, A. G. Kikhney, M. Gajda, C. Gorba, H. D. T. Mertens, P. V. Konarev, D. I. Svergun, *J. Appl. Cryst.* **2012**, *45*, 342-350.

[11] D. I. Svergun, C. Barberato, M. H. J. Koch, *J. Appl. Cryst.* **1995**, *28*, 768-773.

[12] W. Kabsch, *Acta Crystallogr. sect. D* **2010**, *66*, 125-132.

[13] W. Kabsch, *Acta Crystallogr. sect. D* **2010**, *66*, 133-144.

[14] P. R. Evans, *Acta Crystallogr. sect. D* **2011**, *67*, 282-292.

[15] M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, K. S. Wilson, *Acta Crystallogr. sect. D* **2011**, *67*, 235-242.

[16] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, *J. Appl. Crystallogr.* **2007**, *40*, 658-674.

[17] M. López-Pelegrín, N. Cerdà-Costa, F. Martínez-Jiménez, A. Cintas-Pedrola, A. Canals, J. R. Peinado, M. A. Martí-Renom, C. López-Otín, J. L. Arolas, F. X. Gomis-Rüth, *J. Biol. Chem.* **2013**, *288*, 21279-21294.

[18] N. Stein, *Acta Crystallogr. sect. D* **2008**, *41*, 641-643.

[19] F. Corpet, *Nucl. Acids Res.* **1988**, *16*, 10881-10890.

[20] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta Crystallogr. sect. D* **2010**, *66*, 486-501.

[21] P. V. Afonine, R. W. Grosse-Kunstleve, N. Echols, J. J. Headd, N. W. Moriarty, M. Mustyakimov, T. C. Terwilliger, A. Urzhumtsev, P. H. Zwart, P. D. Adams, *Acta Crystallogr. sect. D* **2012**, *68*, 352-367.

[22] O. S. Smart, T. O. Womack, C. Flensburg, P. Keller, W. Paciorek, A. Sharff, C. Vonrhein, G. Bricogne, *Acta Crystallogr. sect. D* **2012**, *68*, 368-380.

[23] T. C. Terwilliger, R. W. Grosse-Kunstleve, P. V. Afonine, N. W. Moriarty, P. H. Zwart, L. W. Hung, R. J. Read, P. D. Adams, *Acta Crystallogr. sect. D* **2008**, *64*, 61-69.

[24] G. Langer, S. X. Cohen, V. S. Lamzin, A. Perrakis, *Nat. Protoc.* **2008**, *3*, 1171-1179.
[25] A. Vagin, A. Tepliakov, *Acta Crystallogr. sect. D* **1997**, *66*, 22-25.
[26] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605-1612.
[27] A. T. Brünger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J.-S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, G. L. Warren, *Acta Crystallogr. sect. D* **1998**, *54*, 905-921.
[28] E. Krissinel, K. Henrick, *J. Mol. Biol.* **2007**, *372*, 774-797.
[29] E. Krissinel, K. Henrick, *Acta Crystallogr. sect. D* **2004**, *60*, 2256-2268.
[30] M. C. Lawrence, P. M. Colman, *J. Mol. Biol.* **1993**, *234*, 946-950.
[31] M. Jenik, R. G. Parra, L. G. Radusky, A. Turjanski, P. G. Wolynes, D. U. Ferreiro, *Nucl. Acids Res.* **2012**, *40*, W348-W351.
[32] U. Emekli, D. Schneidman-Duhovny, H. J. Wolfson, R. Nussinov, T. Haliloglu, *Proteins* **2008**, *70*, 1219-1227.
[33] V. B. Chen, W. B. Arendall, 3rd, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, *Acta Crystallogr. sect. D* **2010**, *66*, 12-21.
[34] G. Vriend, *J. Mol. Graph.* **1990**, *8*, 52-56.
[35] C. Cole, J. D. Barber, G. J. Barton, *Nucl. Acids Res.* **2008**, *36*, W197-W201.
[36] S. El-Ghaish, H. Rabesona, Y. Choiset, M. Sitohy, T. Haertle, J. M. Chobert, *J. Dairy Res.* **2011**, 1-8.
[37] aW. Bode, F. X. Gomis-Rüth, W. Stöcker, *FEBS Lett.* **1993**, *331*, 134-140; bF. X. Gomis-Rüth, *Crit. Rev. Biochem. Mol. Biol.* **2008**, *43*, 319-345; cB. W. Matthews, *Acc. Chem. Res.* **1988**, *21*, 333-340; dN. Cerdà-Costa, F. X. Gomis-Rüth, *Prot. Sci.* **2014**, *23*, 123-144.
[38] aD. S. Auld, B. Holmquist, *Biochemistry* **1974**, *13*, 4355-4361; bF. X. Gomis-Rüth, H. Nar, F. Grams, I. Yallouros, U. Küsthardt, R. Zwilling, W. Bode, W. Stöcker, *J. Biol. Chem.* **1994**, *269*, 17111-17117; cB. Holmquist, B. L. Vallee, *J. Biol. Chem.* **1974**, *249*, 4601-4607.
[39] T. M. Laue, B. D. Shah, T. M. Ridgeway, S. L. Pelletier, in *Analytical ultracentrifugation in biochemistry and polymer science*. (Eds.: S. E. Harding, A. J. Rowe, J. C. Horton), The Royal Society of Chemistry, Cambridge, UK, **1992**, pp. 90-125.
[40] S. M. Kelly, T. J. Jess, N. C. Price, *Biochim. Biophys. Acta* **2005**, *1751*, 119-139.
[41] W. N. Lipscomb, *Proc. Natl. Acad. Sci. USA* **1973**, *70*, 3797-3801.
[42] A. Fersht, *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*., W.H. Freeman & Co., New York, **1999**.
[43] D. Eisenberg, A. D. McLachlan, *Nature* **1986**, *319*, 199-203.
[44] C. R. Cantor, P. R. Schimmel, *Biophysical chemistry. Part I: the conformation of biological macromolecules*., W.H. Freeman & Co., New York, **1980**.
[45] I. de Diego, F. T. Veillard, T. Guevara, B. Potempa, M. Szutkowska, J. Potempa, F. X. Gomis-Rüth, *J. Biol. Chem.* **2013**, *288*, 14287-14296.
[46] K. Diederichs, P. A. Karplus, *Nat. Struct. Biol.* **1997**, *4*, 269-275.
[47] P. A. Karplus, K. Diederichs, *Science* **2012**, *336*, 1030-1033.
[48] I. W. Davis, A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. Bryan Arendall, 3rd, J. Snoeyink, J. S. Richardson, D. C. Richardson, *Nucl. Acids Res.* **2007**, *35*, W375-W383.
[49] J. Janin, C. Chothia, *J. Biol. Chem.* **1990**, *265*, 16027-16030.
[50] P. Gouet, X. Robert, E. Courcelle, *Nucl. Acids Res.* **2003**, *31*, 3320-3323.

5. SUPPLEMENTAL TABLES

Suppl. Table 1. Inhibition of wild-type selecase activity.

Inhibitor	Concentration (mM)	Specificity	Relative activity (%)
None	—	—	100
EDTA	1	Metallopeptidases	0
$ZnCl_2$	1	Metallopeptidases	3
Phosphoramidon	0.25	Metallopeptidases	39
Actinonin	0.25	Metallopeptidases	6
GM1489	0.05	Metallopeptidases	24
MMP inhibitor III	0.05	Metallopeptidases	5
MMP inhibitor V	0.25	Metallopeptidases	11
Batimastat	0.005	Metallopeptidases	2
Marimastat	0.25	Metallopeptidases	45
PMSF	1	Serine peptidases	82
Pefabloc	1	Serine peptidases	78
Benzamidine	1	Serine peptidases	66
Iodoacetamide	1	Cysteine peptidases	79
E-64	0.1	Cysteine peptidases	94
Pepstatin A	0.1	Aspartic peptidases	75

Values as mean of three independent measurements, SD within ±5. Selecase concentration: 2µM.

Suppl. Table 2. Effect of divalent ions on wild-type apo-selecase reactivation.

Cation	Final concentration (mM)	Relative activity (%)
Zn^{2+}	0.01	100
Co^{2+}	0.01	162
Mg^{2+}	0.01	4
Mn^{2+}	0.01	5
Ca^{2+}	0.01	4
Cd^{2+}	0.01	5
Cu^{2+}	0.01	16
Ni^{2+}	0.01	3

Values as mean of three independent measurements, SD within ±5. Selecase concentration: 20μM

Suppl. Table 3. Analytical ultracentrifugation data of wild-type selecase.	
Sedimentation velocity	Population of species
0.03 mg/ml	slc_1 (76.0%), slc_2 (24%)
0.10 mg/ml	slc_1 (70.6%), slc_2 (29.4%)
0.30 mg/ml	slc_1 (55.9%), slc_2 (33.7%), slc_4 (10.4%)
0.90 mg/ml	slc_2 (53.0%), slc_4 (34.5%), slc_8 (12.5%)
3.0 mg/ml	slc_2 (24.8%), slc_4 (47.2%), slc_8 (28%)
9.0 mg/ml	slc_4 (32.6%), slc_8 (67.4%)
18 mg/ml	slc_4 (20.9%), slc_8 (79.1%)
Equilibrium velocity	Average mass (Da) ($Ln(Mw/Mw_{app}) = 7.86 \times \varphi$)
0.03 mg/ml	16,866 ± 192
0.10 mg/ml	21,364 ± 96
0.30 mg/ml	32,313 ± 204
0.90 mg/ml	41,135 ± 164
3.0 mg/ml	55,833 ± 240
9.0 mg/ml	75,875 ± 356
18 mg/ml	86,902 ± 460

Suppl. Table 4. Analysis of SAXS diffraction data of wild-type selecase.

Conc. (mg/ml)	R_g (Å)	$I(0)$/conc	D_{max} (Å)
0.15	17.7	8.07	60.5
0.30	18.6	8.75	65.4
0.63	20.6	12.86	72.2
1.2	25.6	20.92	89.5
1.9	28.9	21.79	96.9
4.5	30.9	32.11	108.1
8.5	36.1	46.35	126.2
15	43.5	68.47	152.3
22	50.4	81.14	176.5
46	50.0	83.80	179.5
65	47.3	79.44	181.0

Suppl. Table 5. Crystallographic data.

Dataset	Tetragonal slc_2	Hexagonal slc_4	Orthorhombic slc_1	Monoclinic $slc_{2'}$	Orthorhombic $slc_{2''}$
Space group / cell constants (a, b, c, in Å)	$P4_32_12$/ 43.53, 43.53, 128.13	$P6_422$ / 102.2, 102.2, 59.4	$P2_12_12$ / 51.26, 77.35, 31.11	$P2_1$ / 37.69, 101.09, 76.57, β=99.8º	$P2_12_12_1$ / 49.39, 49.84, 104.84
Wavelength (Å)	0.9795	0.8726	0.8726	0.8726	0.9795
No. of measurements / unique reflections	67,065 / 6,813	84,097 / 3,962	54,472 / 7,658	169,251 / 25,107	303,440 / 26,633
Resolution range (Å) (outermost shell) [a]	43.5 – 2.20 (2.35 – 2.20)	49.3 – 3.00 (3.16 – 3.00)	42.7 – 2.10 (2.20 – 2.10)	42.0 – 2.30 (2.44 – 2.30)	52.4 – 1.75 (1.85 – 1.75)
Completeness (%)	99.5 (97.6)	99.7 (99.7)	99.9 (99.3)	99.8 (99.7)	99.2 (95.1)
R_{merge} [b]	0.061 (0.649)	0.058 (0.881)	0.070 (0.868)	0.101 (0.998)	0.054 (0.851)
R_{meas} [c] / CC(1/2) [d]	0.064 (0.728) / 1.000 (0.829)	0.060 (0.914) / 1.000 (0.886)	0.076 (0.935) / 0.999 (0.852)	0.110 (1.083) / 0.999 (0.810)	0.056 (0.928) / 1.000 (0.701)
Average intensity [e]	23.4 (2.3)	32.4 (2.5)	17.7 (2.6)	15.0 (2.1)	24.4 (2.1)
B-Factor (Wilson) (Å^2) / Average multiplicity	49.0 / 9.8 (4.5)	114.2 / 21.2 (13.4)	45.5 / 7.1 (7.2)	51.0 / 6.7 (6.6)	35.8 / 11.4 (6.2)
Resolution range used for refinement (Å)	41.2 – 2.20	49.3 – 3.00	42.7 – 2.10	42.0 – 2.30	52.4 – 1.75
No. of reflections used (among them, test set)	6,765 (483)	3,952 (489)	7,658 (494)	25,107 (739)	26,632 (749)
Crystallographic R_{factor} (free R_{factor}) [b]	0.219 (0.242)	0.200 (0.257)	0.208 (0.281)	0.222 (0.247)	0.185 (0.196)
No. of protein atoms / ions	914 / 1 Zn^{2+}	906 / 1 Zn^{2+}, 1 Na^+	914 / 1 Ni^{2+}	3,618 / 4 Zn^{2+}, 1 Cl^-	1,864 / 2 Zn^{2+}
solvent molecules / ligands	33 / 2 GOL	1 / 4 GOL	58 / 1 GOL	149 / 4 GOL	251 / 5 ACT, 1 GOL
Rmsd from target values					
bonds (Å) / angles (°)	0.010 / 0.99	0.010 / 1.13	0.010 / 1.01	0.009 / 1.11	0.010 / 1.03
Aver. B-factors protein atoms (Å^2)	50.9	125.2	49.7	62.3	36.6
Residue main-chain conformational angle and side-chain rotamer analysis [f]					
favored regions / outliers / all residues /	101 / 1 / 106	102 / 0 / 106	105 / 0 / 107 /	414 / 0 / 422	211 / 0 / 212
bad rotamers (%) / MolProbity score	12.6 / 2.27 (69th percent.)	5.83 / 2.14 (99th percent.)	1.92 / 1.26 (100th percent.)	6.59 / 2.02 (90th percent.)	3.40 / 1.76 (79st percent.)

[a] Values in parentheses refer to the outermost resolution shell. [b] For definitions, see Table 1 in [45]. [c] According to Diederichs & Karplus [46]. [d] According to Karplus & Diederichs [47]. [e] According to program XDS/XSCALE. [f] According to MOLPROBITY [48]. GOL, glycerol; ACT, acetate.

Suppl. Table 6. Residues contributing to the hydrophobic core and the sub-domain interface in slc_1.

NTS	CTS
α1 (I^{6}, L^{7}, I^{10}, L^{11}, T^{14}, I^{15}, L^{18})	Lα2α3 (L^{81})
Lα1β1 (L^{20}, I^{27})	α3 ($\underline{F^{84}}$, $\underline{L^{88}}$, $\underline{L^{91}}$)
β1 (I^{29}, I^{31})	Lα3α4 ($\underline{F^{92}}$)
β2 ($\underline{I^{38}}$, A^{39}, I^{41})	α4 ($\underline{K^{95}}$, $\underline{I^{98}}$, $\underline{L^{99}}$, $\underline{L^{102}}$, $\underline{L^{106}}$)
Lβ2β3 (L^{43})	
β3 (I^{48}, I^{50})	
αB ($\underline{I^{54}}$, L^{55}, $\underline{Y^{57}}$, $\underline{L^{58}}$)	
α2 (I^{63}, $\underline{F^{65}}$, $\underline{I^{66}}$, L^{67}, $\underline{A^{68}}$, L^{71}, $\underline{L^{72}}$, L^{74}, $\underline{Y^{76}}$)	

The respective regular secondary-structure elements are shown (see also Suppl. Table 7). Residues contributing to the NTS-CTS interface are underlined.

Suppl. Table 7. Regular secondary structure elements of selecase.

Residue range	Name	Comments
R^{4}-N^{19}	"backing helix" α1	
N^{21}-N^{26}	αA	In slc_4, $slc_{2'}$ and $slc_{2''}$
N^{28}-K^{32}	β1	
I^{38}-S^{42}	β2	
K^{46}-I^{50}	β3	
K^{51}-L^{58}	"linking helix" αB	
D^{60}-Y^{76}	"active-site helix" α2	Till K^{75} in slc_4
K^{78}-N^{82}	β4	Only in slc_4
N^{82}-L^{91}	"glutamate helix" α3	Not in slc_4
E^{87}-L^{91}	β5	Only in slc_4
N^{94}-Q^{108}	"C-terminal helix" α4	Till H^{107} in slc_1, $slc_{2'}$ (B) and $slc_{2''}$ From K^{95} onwards in $slc_{2'}$ (B)

The nomenclature has been chosen to be compatible with that of projannalysin and proabylysin (see PDB 4JIX and 4JIU; [17]). Helix αB was termed α1' in projannalysin and topologically corresponded to 3_{10}-helix η1 in proabylysin.

Suppl. Table 8. Oligomerization and thermodynamic parameters of wild-type selecase.

Pairwise close contacts [a] and surface complementarity (Sc) [b]

Slc_2 (molecule A – molecule B; sym. op. –Y, –X, –Z+1/2 [c])

34 symmetric close contacts (incl. 8 hydrogen bonds, 1 metallorganic contact, and 34 hydrophobic carbon-carbon contacts by 5 residues of each molecule). Segments involved in dimerization: 35-39, 53-57, 69-70, and 100-110. Sc=0.57

Slc_4 (molecule A – molecule B; sym. op. –Y+1, –X, –Z+4/3)

42 symmetric close contacts (incl. 10 hydrogen bonds and 6 hydrophobic carbon-carbon contacts by 5 residues of each molecule). Segments involved in oligomerization: 34-38, 51-57, 69-70, and 91-107. Sc=0.76

Slc_4 (molecule A – molecule C; sym. op. Y, X+1, –Z+4/3)

9 symmetric close contacts (all are hydrophobic carbon-carbon contacts made by 6 residues of each molecule). Segments involved in oligomerization: 95-106. Sc=0.80

Slc_4 (molecule A – molecule D; sym. op. –X, –Y+1, Z)

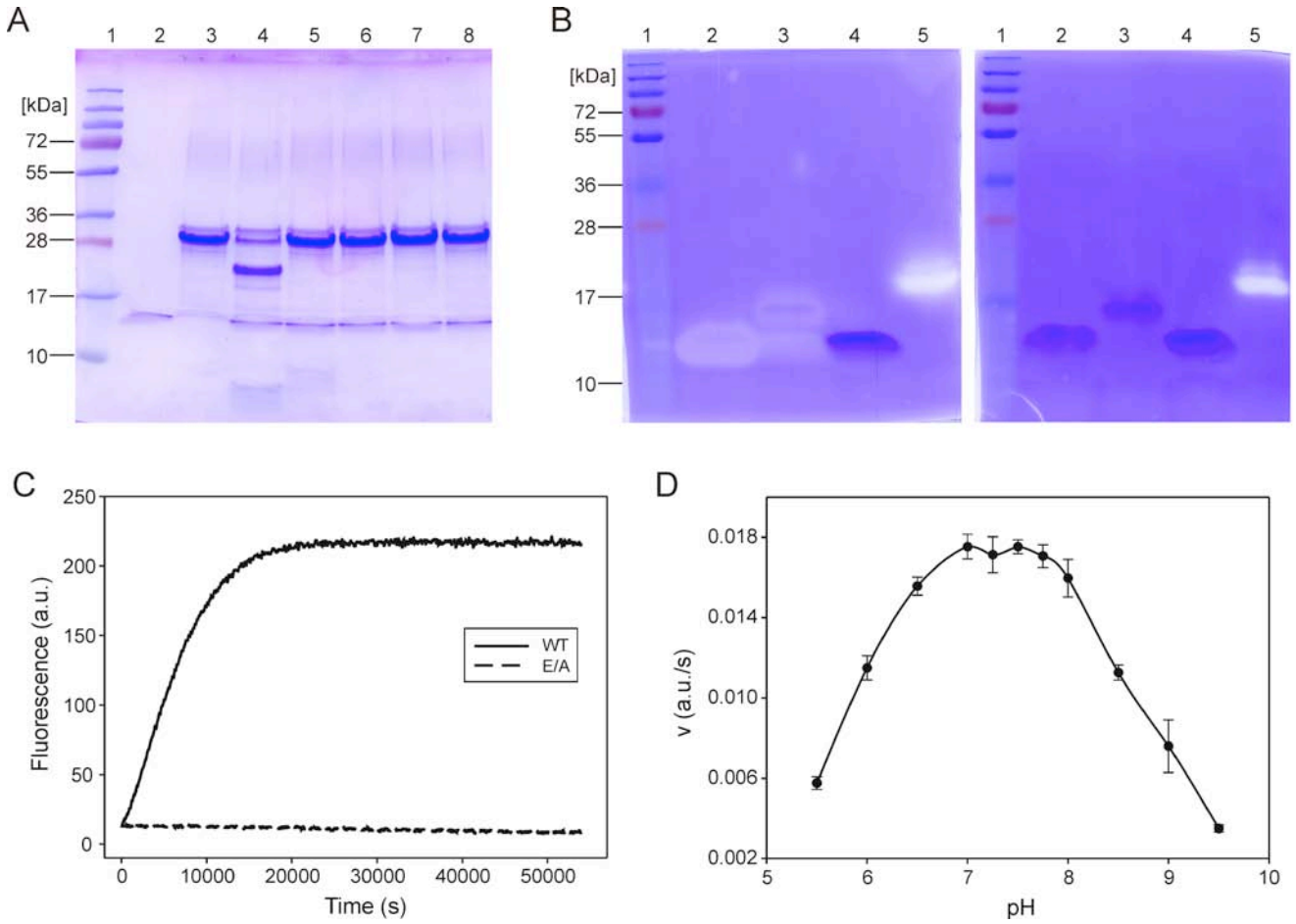
40 symmetric close contacts (incl. 2 hydrogen bonds, 1 metallorganic contact, 1 metal-metal contact, and 25 hydrophobic carbon-carbon contacts by 17 residues of each molecule). Segments involved in oligomerization: 17-19, 38, 54-76, 93-106, and 999 (metal ion). Sc=0.65

Molecular and oligomerization analysis [d]

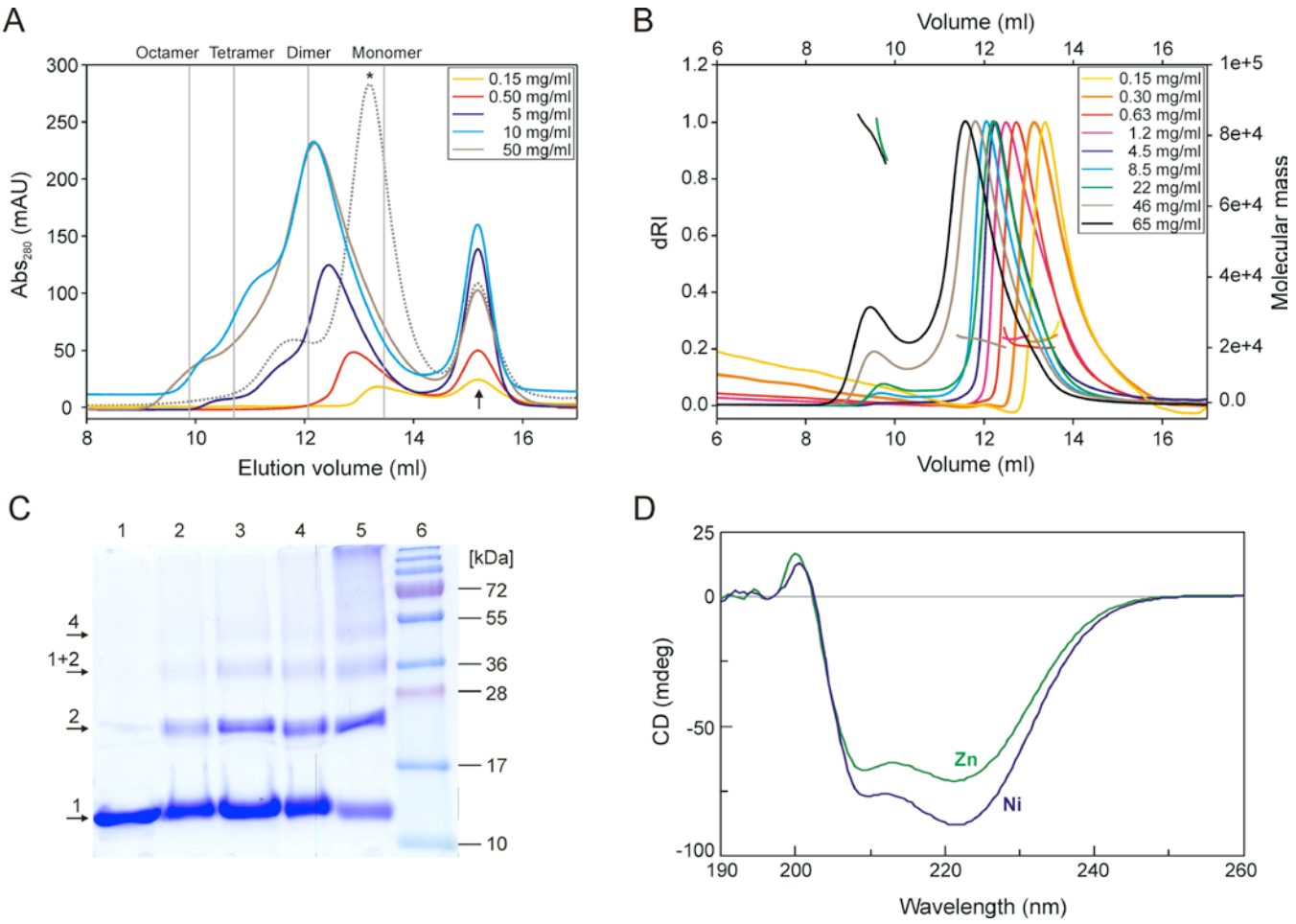
	No. of atoms	No. of residues	Total surface ($Å^2$)	ΔG [d,e] (kcal/mol)	Residues (%) in regular secondary structure (α / β / total)	Calculated oligomeric state [d] / ΔG_{diss}(kcal/mol) [f]	Interface buried ($Å^2$)
Slc_1 (monomer)	914	109	6,999	-109	60 / 14 / 73	1 / -	
Slc_2 (monomer)	907	108	7,274	-105	66 / 14/ 75	2 / 24	
Slc_4 (monomer)	906	108	8,696	-92	61 / 23 / 80	4 / 14	
Slc_2 (dimer)	1814	216	12,683	-217			2,130 [g]
Slc_4 (tetramer)	3,624	432	25,704	-424			9,850

[a] According to program CNS [27]. [b] According to program SC [30]. [c] Sym. op. is the crystallographic operator that generates the symmetric molecule interacting with molecule A in the respective oligomer. Molecules B and D are neighbors of molecule A and molecule C is in opposite position it. [d] According to computational crystal-packing analysis with the PISA program, which has a success rate of 80-90% in the recovery of functional biological assemblies [28]. [e] ΔG is the solvation free energy of folding. [f] ΔG_{diss} is the free energy of assembly dissociation. $\Delta G_{diss} > 0$ is indicative of a thermodynamically stable assembly (see [28]). [g] This value is above the range generally described for protein-protein complexes (1600 ±350$Å^2$; [49]). Data on $slc_{2'}$ and $slc_{2''}$ have not been included as these variants are not the same chemical species as slc_1, slc_2, and slc_4.

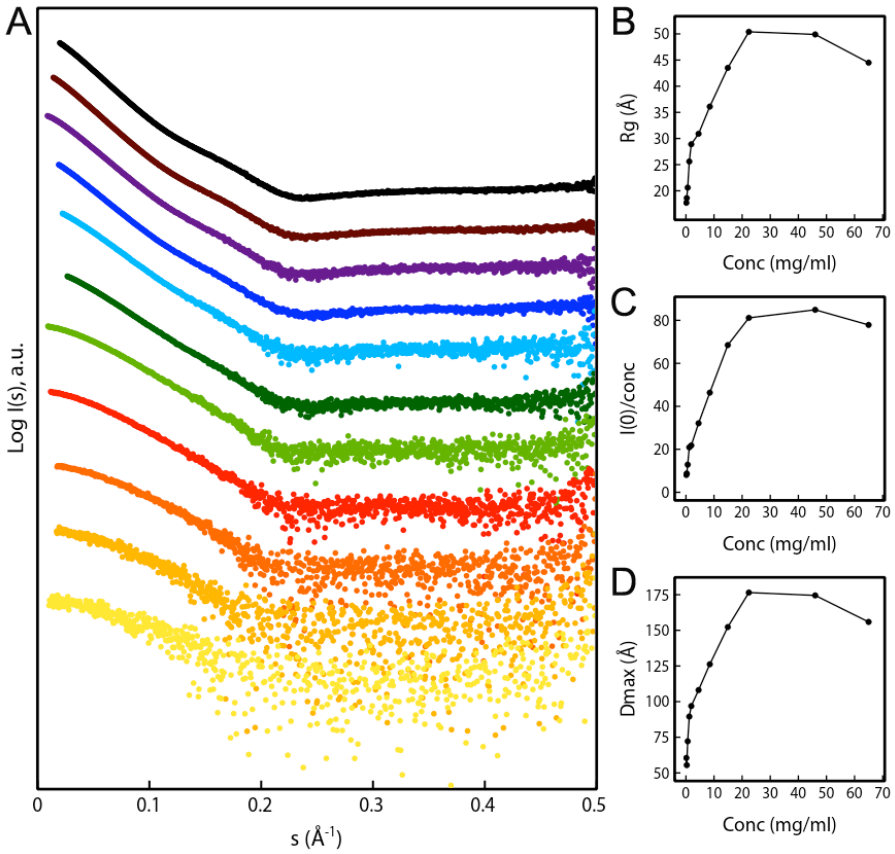
6. SUPPLEMENTAL FIGURES



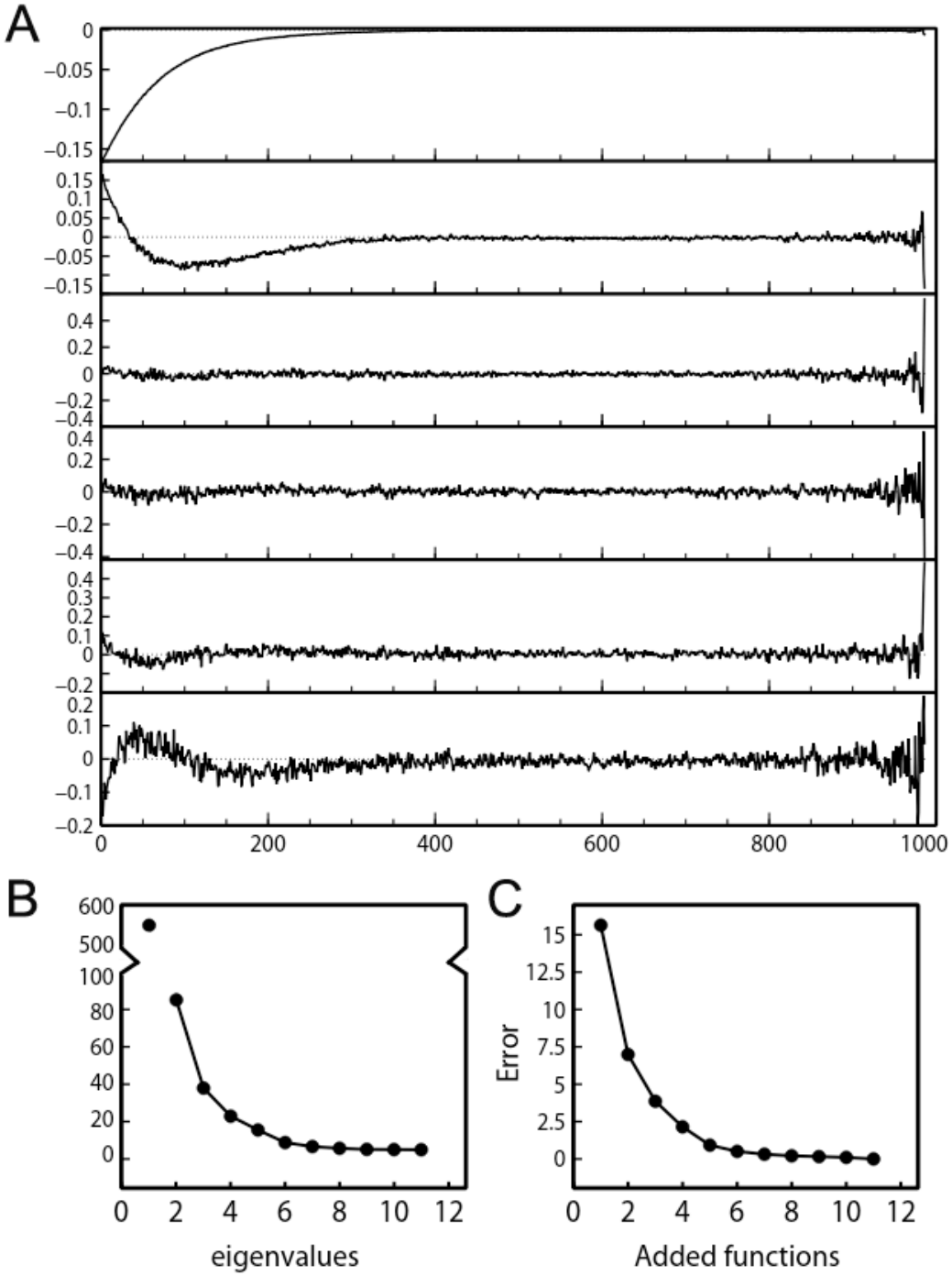
Suppl. Figure 1 – (A) SDS-PAGE of α-casein incubated with selecase (see section 1.2). *Lane 1*, molecular mass standard. *Lane 2*, selecase. *Lane 3*, α-casein consisting of α_{s2}-casein (top light band; 25KDa) and α_{s1}-casein (bottom strong band; 22-23.7KDa). *Lane 4*, α-casein incubated with wild-type selecase leads to cleavage of α_{s1}-casein at a single site. *Lanes 5, 6, 7*, and *8*, α-casein incubated, respectively, with inactive active-site mutant $E^{70}A$ and wild-type selecase in the presence of EDTA, $ZnCl_2$, and 1,10-phenanthroline (all at 1mM). **(B)** Coomassie-stained zymogram of α-casein incubated for 48h at 37°C with 13.6-KDa and 16.5-KDa selecase (2.5µg; same for 13.1-KDa selecase, data not shown) in the absence (left panel) and presence (right panel) of 5mM 1,10-phenanthroline. Caseinolytic activity is detected as light lysis zones. *Lane 1*, molecular mass standard. *Lane 2*, active 13.6-KDa selecase. *Lane 3*, active 16.5-KDa selecase and its likewise active partial autolytic cleavage product, 13.8-KDa selecase (left panel only). *Lane 4*, inactive $E^{70}A$ 13.6-KDa selecase mutant. *Lane 5*, bovine trypsin (0.025µg) as positive control. **(C)** Proteolysis of peptide CCS at 10µM over time after incubation at 37°C in 50mM HEPES, 150mM NaCl, pH7.5 with either wild-type (WT) or $E^{70}A$-mutant (E/A) 13.6-KDa selecase at 2µM. Equivalent results were obtained with 13.1-KDa selecase (data not shown). **(D)** pH-Dependence of selecase cleavage activity on peptide CCS.



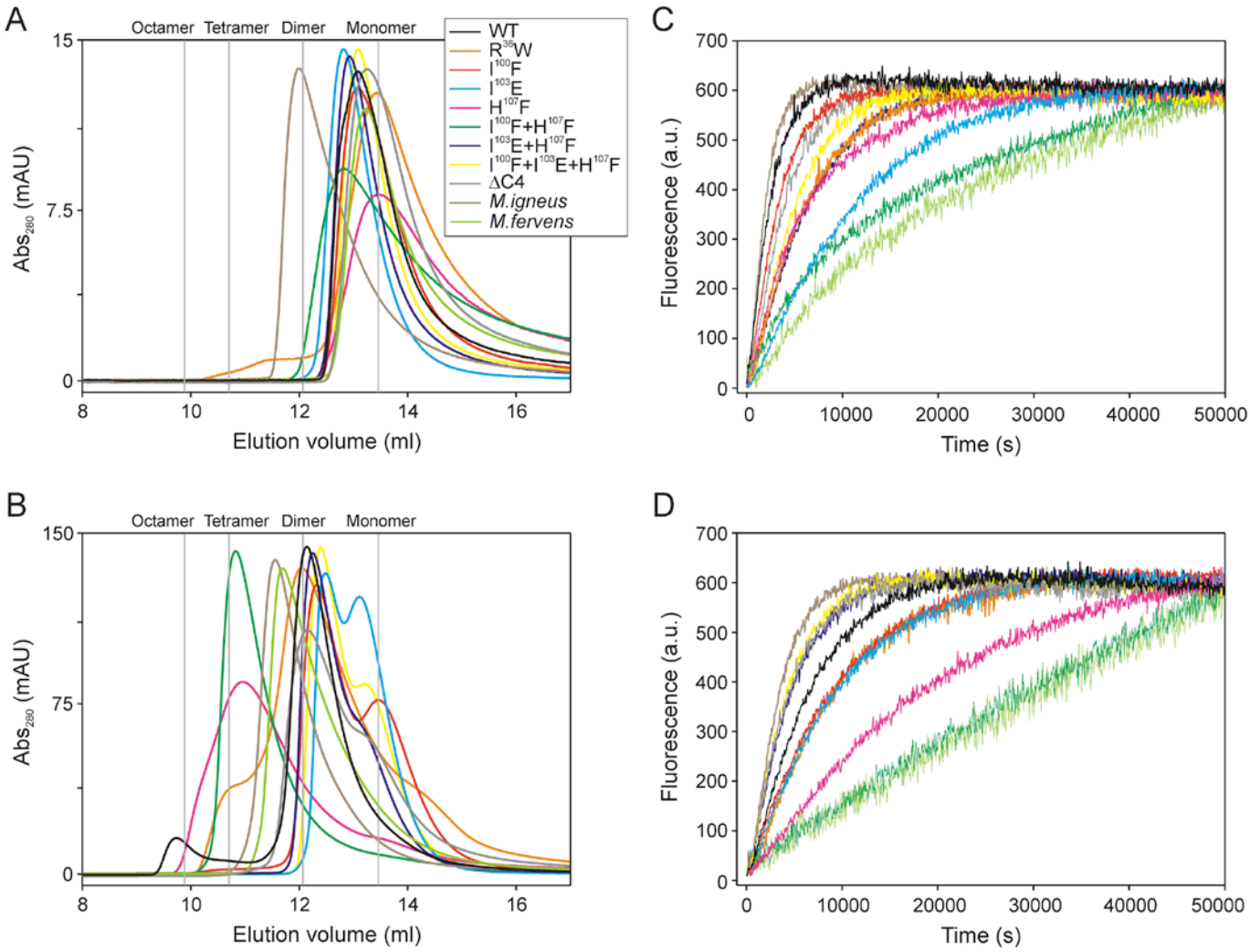
Suppl. Figure 2 – (A) Calibrated size-exclusion chromatography (SEC) of selecase at selected initial concentrations (0.15-50mg/ml; samples become diluted during this chromatography) plus aprotinin (arrow) as internal molecular-mass reference (6.5KDa; see section 1.4 for details). The peak pattern observed at each concentration corresponds to distinct populations of oligomeric selecase, for which the theoretical migration volumes are shown as vertical lines. Only concentrations at or below ~0.20mg/ml contain essentially monomeric selecase. The peak corresponding to the complex between nickel-selecase and a pentapeptide (see section 1.4) injected at 11mg/ml (asterisk; dotted line) elutes mainly as a monomer while unbound zinc-selecase at 10mg/ml (light blue trace) is mainly a dimer. **(B)** SEC coupled to multi-angle laser light scattering (SEC-MALLS) of selecase at selected initial concentrations (0.15-65mg/ml; samples become diluted during this chromatography). The peak pattern moves towards smaller elution volumes when increasing protein concentration, indicating protein oligomerization. Analysis of the absolute mass of the different peaks reveals two main species of ~25KDa and ~80KDa corresponding, respectively, to dimeric and octameric selecase. **(C)** Qualitative crosslinking experiment followed by SDS-PAGE depicting the population of selecase monomers (1), dimers (2), monomers linked with dimers (1+2), and tetramers (4) at increasing protein concentrations and 5-fold molar excess of crosslinker (BS^3) at room temperature (*Lanes 1-5*, 0.25, 0.50, 2.0, 10 [all 10'], and 10mg/ml [12h], *lane 6*, molecular mass standard). **(D)** Far-UV circular dichroism spectra of zinc- and nickel-selecase.



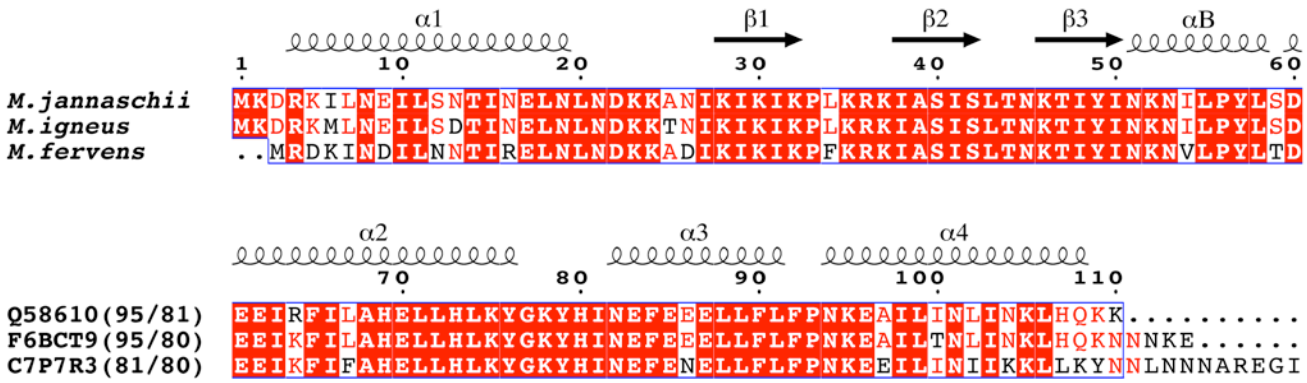
Suppl. Figure 3 – (A) SAXS intensity profiles, $I(s)$, as a function of the momentum transfer, s, measured for wild-type selecase at 11 concentrations (*bottom* to *top*): 0.15 (yellow), 0.30, 0.63, 1.2, 1.9, 4.5, 8.5, 15, 22, 46 and 65mg/ml (black). Profiles have been displaced along the $I(s)$ axis for comparison. Variation of the primary SAXS data parameters with concentration: **(B)** R_g, **(C)** $I(0)$/concentration, and **(D)** D_{max}.



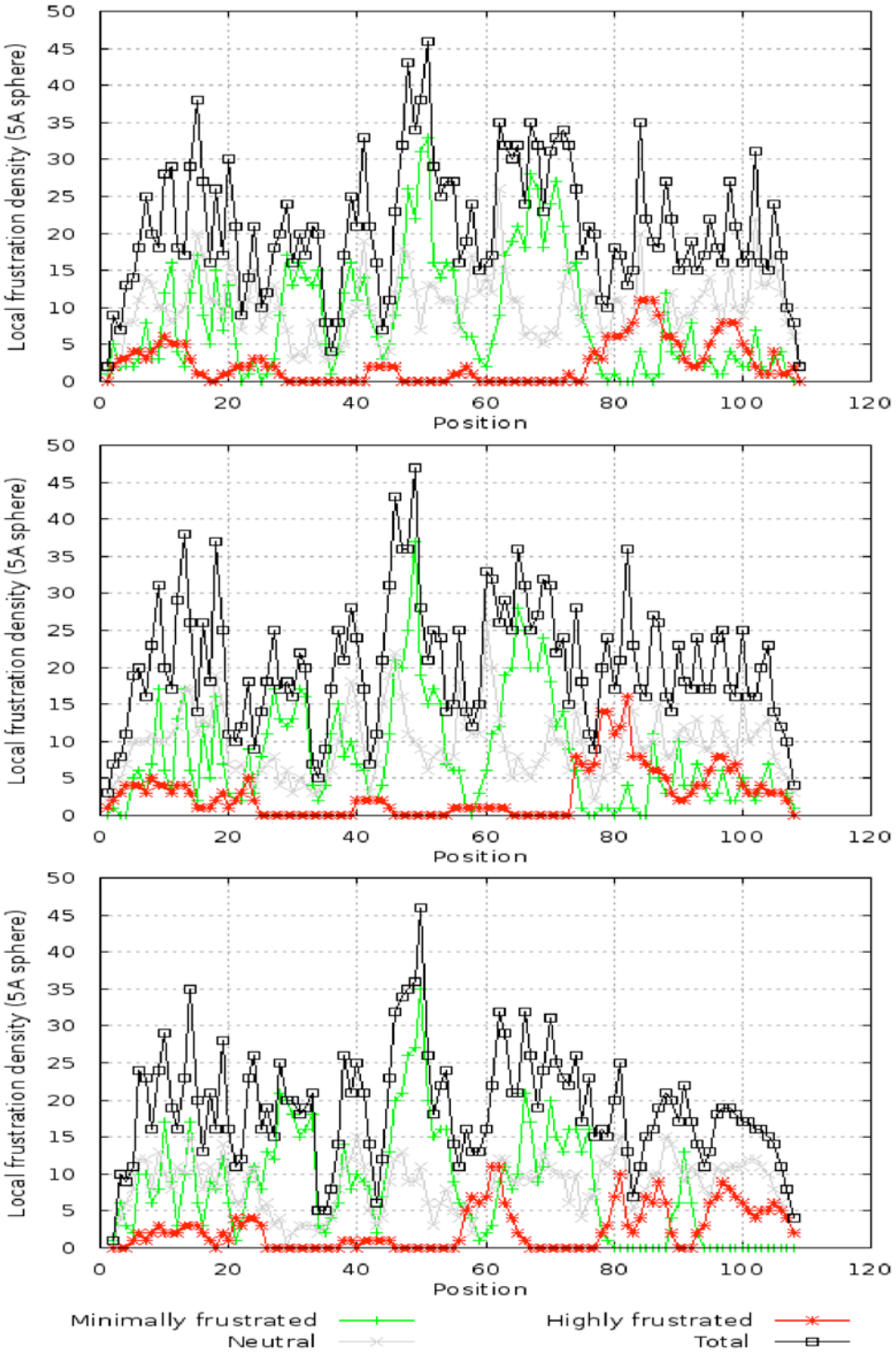
Suppl. Figure 4 – Single Value Decomposition (SVD) analysis of the SAXS data measured for selecase. Eigenvectors **(A)** derived from the analysis sorted by their corresponding eigenvalues **(B)**. **(C)** Agreement with the experimental data upon successive increase in the number of eigenvectors.



Suppl. Figure 5 – (A) Calibrated size-exclusion chromatography of seven selecase point mutants and a deletion variant at 0.50mg/ml initial concentration at injection (for color coding, see inset) and the orthologs from *M. igneus* (brown line) and *M. fervens* (light green line) as compared with the wild type (black line). See also Suppl. Fig. 2a. **(B)** Same as (A) but at 5.0mg/ml. **(C)** Cleavage activity of point mutants, deletion variant, and orthologs at 0.25mg/ml in front of peptide CCS compared with wild-type selecase. The curves have been normalized to the wild-type curve (in black), for which the scale on the ordinate is valid. Color coding as in (A). **(D)** Same as (C) but at 5.0mg/ml.



Suppl. Figure 6 – Sequence alignment of, *top* to *bottom,* selecase (UP Q58610) and two close orthologs from *M. igneus* (UP F6BCT9) and *M. fervens* (UP C7P7R3) performed using the MULTALIN program [19] at http://multalin.toulouse.inra.fr/multalin/multalin.html and represented using the ESPRIPT server [50] at http://espript.ibcp.fr/ESPript/ESPript. Numbering and regular secondary structure elements correspond to *M. jannaschii* slc_1 (see also Suppl. Table 7). The sequence identity of each protein with the other two is indicated, respectively, in the lower alignment block after the UP code.



Suppl. Figure 7 - Analysis of local configurational frustration of the final slc_1, slc_2, and slc_4 wild-type selecase monomers (*top* to *bottom*) with the FRUSTRATROMETER [31] at http://lfp.qb.fcen.uba.ar/embnet. Highly frustrated regions (red traces) include, respectively, segments 11-13, 79-90, and 95-100; segments 74-88 and 95-99; and segments 57-63, 80-81, 85-88, 95-101, and 103-106. The common consensus region of frustration spans almost the entire CTS. Data on $slc_{2'}$ and $slc_{2''}$ have not been included as these variants are not the same chemical species as slc_1, slc_2, and slc_4.

7. LEGENDS TO SUPPLEMENTAL MOVIES

Suppl. Movie 1 – Movie to illustrate the transition between slc_1 and slc_2 monomers. The ribbon plot of slc_1 is first shown in pink with its functional active-site environment. The cleft with the hydrophobic S_1' pocket is then depicted as a surface model colored according to Kyte-Doolittle hydrophobicity (blue=hydrophilic over white to orange=hydrophobic). The transition between slc_1 (in pink) and slc_2 (in blue) was generated with the morph option of program CHIMERA, i.e. it does not correspond to any empirically demonstrated trajectory. The transformation is correlated with a dimeric quaternary arrangement (chains in blue and yellow), which leads H^{107} from the C-terminal helix of the symmetric molecule to replace H^{80} in the metal coordination sphere.

Suppl. Movie 2 – Movie to illustrate the transition between slc_1 and slc_4 protomers. The ribbon plot of competent slc_1 is shown in pink and its transformation to slc_4 (in turquoise) was generated with the morph option of program CHIMERA, i.e. it does not correspond to any empirically demonstrated trajectory. The structure of slc_4 is only stable due to the tetrameric quaternary structure (chains in turquoise, green, white and black), which gives rise to an autoinhibitory particle. The latter is traversed by a long cleft, which harbors four metal ions arranged as two dimetallic sites (shown as red spheres). A cut-through view of the particle surface provides insight into the channel, shown for the wall created by the turquoise and green monomers. The associated dimetallic site is pinpointed by a blue and a yellow sphere. Finally, the ribbon plot of the latter two monomers and cations is shown, followed by a close up view of the dimetallic site and the metal-binding residues.

Suppl. Movie 3 – Movie to illustrate a possible transition between slc_1 and $slc_{2'}$ (molecule B) protomers. The ribbon plot of competent slc_1 is shown in pink and its transformation to $slc_{2'}$ molecule B (in green) was generated with the morph option of program CHIMERA, i.e. it does not correspond to any empirically demonstrated trajectory. The oligomeric structure of $slc_{2'}$ is a chemically identical but conformationally heterogenic dimer comprising a modified molecule B (green) and one with close-to-functional overall conformation similar to slc_1 (molecule A; orange ribbon). Dimerization is triggered by flipping in the tryptophan side chains of position 36 and results in H^{107} from the C-terminal helix of the modified molecule binding the catalytic zinc ion as fourth intramolecular protein ligand. Inhibition of the second molecule is achieved by blocking access to the active-site cleft within the dimer.