Mechanisms of proteolytic activity regulation exerted via a unique propeptide in matrix metalloproteinases and intra/intermolecular interactions in a novel family of minimal gluzincins

Maria del Mar López Pelegrín

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (**www.tdx.cat**) i a través del Dipòsit Digital de la UB (**diposit.ub.edu**) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (**www.tdx.cat**) y a través del Repositorio Digital de la UB (**diposit.ub.edu**) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (**www.tdx.cat**) service and by the UB Digital Repository (**diposit.ub.edu**) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

UNIVERSITAT DE BARCELONA

Departament de Bioquímica i Biologia Molecular

Facultat de Farmàcia

CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

INSTITUT DE BIOLOGIA MOLECULAR DE BARCELONA

Departament de Biologia Estructural

Proteolysis Lab

Mechanisms of proteolytic activity regulation exerted via a unique propeptide in matrix metalloproteinases and intra/intermolecular interactions in a novel family of minimal gluzincins

Maria del Mar López Pelegrín

Barcelona, Novembre 2015

UNIVERSITAT DE BARCELONA

Facultat de Farmàcia

Programa de Doctorat de Biotecnologia 2011-2012

Mechanisms of proteolytic activity regulation exerted via a unique propeptide in matrix metalloproteinases and intra/intermolecular interactions in a novel family of minimal gluzincins

Memòria presentada per Maria del Mar López Pelegrín per optar al títol de doctor per la Universitat de Barcelona.

Maria del Mar López Pelegrín

Supervisors:

Prof. F. Xavier Gomis-Rüth

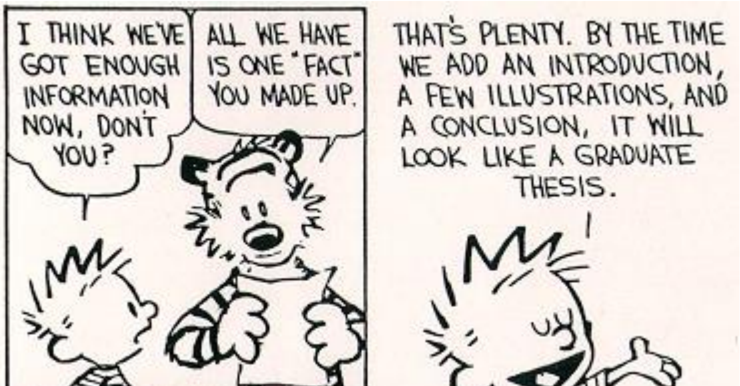
Dr. Joan López Arolas

Tutora:

Dr. Josefa Badia Palacín

Barcelona, Novembre 2015

Als meus pares,


I THINK WE'VE GOT ENOUGH INFORMATION NOW, DON'T YOU?
ALL WE HAVE IS ONE "FACT" YOU MADE UP.
THAT'S PLENTY. BY THE TIME WE ADD AN INTRODUCTION, A FEW ILLUSTRATIONS, AND A CONCLUSION, IT WILL LOOK LIKE A GRADUATE THESIS.

AGRAÏMENTS

En primer lloc vull agrair al Prof. F. Xavier Gomis-Rüth la oportunitat que em va brindar al acceptar-me al seu laboratori per a realitzar aquest doctorat. Sense les seves idees, motivació, temps i entusiasme aquest treball no hagués estat possible.

Al Dr. Joan López Arolas li agraeixo haver codirigit aquesta tesi i haver-me guiat de manera tant pacient durant aquests anys al laboratori, el seu consell, persistència i suport incondicional. Ha estat una peça clau en la feina presentada així com un pilar personal al llarg d'aquesta etapa.

A la Dra. Núria Cerdà Costa pel que m'ha ensenyat i per a fer tant agradable l'experiència de treballar colze a colze per a que els nostres "peipers" tiressin endavant.

A tots els companys de cri3 (passats i presents) els estic agraïda per les idees i consells científics, pel suport moral i les bones estones compartides entre les quatre parets del nostre lab. Pels "picos-de-las-cuatro", abraçadetes carrega-piles, estones de pesca a 20ºC (gràcies Tibi!), cafès d'emergència i alguna que altra canya pels voltants del Parc. També pel suport en moments no tant bons, que han estat part del procés necessari que m'ha portat avui aquí i han viscut en primera persona. Destaquen Pedro, Irens, Tibi, Lauris, Sora, Joan, Iñaki y Sergio...sempre disposats a llançar un salvavides als meus "mares de dudas" i a arrancar-me tants somriures durant les jornades de treball. Per tot el que m'han aportat a nivell personal més enllà del professional.

També a tots els membres dels altres Cri's, de l'IBMB, de la PAC i del PCB amb qui he tingut la oportunitat de treballar, que malauradament no puc mencionar un per un, per haver enriquit de manera significativa aquests anys de doctorat.

Un agradecimiento especial al Prof. Miguel Alcalde Galeote y a todo su grupo en el ICP (Madrid) por enseñarme tanto, por la motivación y la fantástica experiencia profesional y personal que fue realizar una estancia en su laboratorio…por hacerme sentir como en casa durante esos meses.

A tots els meus amics i amigues fora del laboratori, en especial les de tota la vida (Gaby, Adri Moya i Miquel, Laura, Silvia...) per animar-me, confiar en mi i comprendre la importància dels meus estudis quan no sempre els he pogut dedicar el temps que hagués volgut. A Mònica Pardo Muñoz per ser un pilar en totes les etapes que hem viscut juntes i les que ens queden per viure!

Finalment (last but not least!), agraeixo a la meva família tots els ànims, la paciència (...!) i la confiança dipositada. Per educar-me i estimar-me com ho han fet, per estar al meu costat en tot moment i per la força que sempre em donen per seguir avançant…a ells, els dedico aquesta tesi.

ABSTRACT

Metallopeptidases are major players in the physiology and pathology of all living organisms. Their exquisite regulation is therefore essential for proper function and to prevent misdirected temporal and/or spatial proteolytic activity, which may lead to disease. This regulation is achieved through a wide variety of mechanisms, including their biosynthesis as inactive precursors, also known as zymogens. In the present thesis, various mechanisms of metallopeptidase regulation were studied by using a combination of biochemical, biophysical, and structural techniques. In the first project, the crystal structure of a zymogen fragment of *Tannerella forsythia* karilysin revealed the shortest propeptide reported to date for a metallopeptidase. Additional biochemical and biophysical assays allowed ascertaining the importance of the propeptide in protein expression, folding and stability. In the second project, a novel family of minimal prokaryotic metallopeptidases termed "minigluzincins" was discovered, providing a minimal soluble scaffold for gluzincin metallopeptidases and integral-membrane metallopeptidases. Two members of this family, called proabylysin and projannalysin, showed two unique zymogenic forms of latency maintenance exerted, respectively, via intramolecular and intermolecular interactions through their C-terminal segments. In the third project, a further member of this novel family, called selecase, evinced selective and specific proteolytic activity against casein. A set of biophysical and structural studies showed that selecase reversibly commutes between several conformations of defined three-dimensional structure, which are associated with loss of enzymatic activity due to autoinhibition (i.e. active monomers vs. inactive dimers, tetramers and octamers). Overall, the present thesis contributes substantially to the field broadening previous knowledge at the molecular level on metallopeptidases and their regulatory mechanisms, which paves the way for the design of specific inhibitors to modulate their activity as part of therapeutic approaches.

ABSTRACT (in Catalan)

Les metal·lopeptidases participen de manera decisiva en la fisiologia i patologia de tots els organismes vius. La seva estricta regulació és, per tant, essencial pel correcte funcionament d'aquestes i per a prevenir una activitat proteolítica inadequada en el temps i/o espai que podria causar malalties. Aquesta regulació s'aconsegueix mitjançant un ampli ventall de mecanismes, incloent la seva biosíntesis com a precursors inactius, també coneguts com a zimògens. En la present tesi s'han estudiat diversos mecanismes de regulació de metal·lopeptidases, tot combinant tècniques bioquímiques, biofísiques i estructurals. En el primer projecte, l'estructura cristal·lina d'un fragment zimogènic de la proteïna "karilysin" de *Tannerella forsythia* ha revelat el propèptid més curt descrit fins al moment per una metal·lopeptidasa. Assajos bioquímics i biofísics addicionals han permès determinar la importància del propèptid en l'expressió, plegament i estabilitat de la proteïna. En el segon projecte, s'ha descobert una nova família de metal·lopeptidases mínimes d'origen procariota, que s'han anomenat "minigluzincins", les quals proporcionen un model estructural per a metal·lopeptidases pertanyents al clan de les gluzincines i per a d'altres integrals de membrana. Dues membres d'aquesta família, anomenades "proabylysin" i "projannalysin", han exhibit formes úniques de manteniment de latència exercides, respectivament, de forma intra- i intermolecular mitjançant llurs segments C-terminals. En el tercer projecte, una altra membre d'aquesta nova família, anomenada "selecase", ha presentat una activitat proteolítica selectiva i específica sobre caseïna. La conjunció d'estudis duts a terme, tant biofísics com estructurals, ha evidenciat que "selecase" transita de manera reversible entre diverses conformacions d'estructura tridimensional definida, associades amb una disminució d'activitat enzimàtica deguda a autoinhibició (és a dir, monòmers actius i dímers, tetràmers i octàmers inactius). En definitiva, aquesta tesi contribueix de manera substancial a ampliar el coneixement previ sobre metal·lopeptidases a nivell molecular i a entendre els mecanismes que en regulen l'activitat, facilitant així el disseny d'inhibidors específics com a part d'aproximacions terapèutiques.

TABLE OF CONTENTS

INTRODUCTION

RESULTS

INDEX OF FIGURES AND TABLES

ABBREVIATIONS

AAA proteases	ATPases associated with diverse cellular activities
ADAM	A disintegrin and metalloproteinase
AP	Aminopeptidase
AUC	Analytical ultracentrifugation
C-	Carboxy-
CD	Catalytic domain
CTS	C-terminal subdomain
EC	Enzyme Commission
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EGTA	Ethyleneglycoltetraacetic acid
FACE-1	Farnesylated-protein converting enzyme 1
FTP	Fungalysin/thermolysin prodomain
IMMP	Integral membrane metallopeptidase
k_{cat}	Catalytic constant
K_M	Michaelis-Menten constant
LA4H	Leukotriene A4 hydrolase
LAS	Lysostaphin, D-Ala-D-Ala carboxypeptidase and sonic hedgehog
MALLS	Multiangle laser light scattering
MES	2-(N-morpholino)ethanesulfonic acid
MMP	Matrix metalloproteinase
MP	Metallopeptidase
N-	Amino-
NTS	N-terminal subdomain
PAGE	Polyacrilamide gel electrophoresis
PD	Prodomain
PDB	Protein Data Bank
SAXS	Small-angle X-ray scattering
S2P	Site-2 peptidase
SDS	Sodium dodecylsulfate
SEC	Size-exclusion chromatography
SP	Signal peptide
TIMP	Tissue inhibitor of matrix metalloproteinase
TLP	Thermolysin-like peptidase
TMH	Transmembrane helix
ZBCS	Zinc-binding consensus sequence

ABBREVIATIONS

The twenty physiological and proteinogenic amino acids with their respective three-letter and one-letter codes

Alanine	**Ala**	**A**
Arginine	**Arg**	**R**
Asparagine	**Asn**	**N**
Aspartic acid	**Asp**	**D**
Cysteine	**Cys**	**C**
Glutamic acid	**Glu**	**E**
Glutamine	**Gln**	**Q**
Glycine	**Gly**	**G**
Histidine	**His**	**H**
Isoleucine	**Ile**	**I**
Leucine	**Leu**	**L**
Lysine	**Lys**	**K**
Methionine	**Met**	**M**
Phenylalanine	**Phe**	**F**
Proline	**Pro**	**P**
Serine	**Ser**	**S**
Threonine	**Thr**	**T**
Tryptophan	**Trp**	**W**
Tyrosine	**Tyr**	**Y**
Valine	**Val**	**V**

THESIS OUTLINE

The present Ph.D. thesis addresses the study of metallopeptidases and the mechanisms involved in their regulation using a combination of biochemical, biophysical, and structural techniques. It is organized in five main sections: Introduction, Objectives, Results, General Discussion, and Concluding Remarks.

The **Introduction** provides a context for peptidases in general and metallopeptidases in particular, with special emphasis on their mode of action, methods used for their classification, biological roles, and mechanisms responsible for their proteolytic activity regulation. A description of the relevant families of metallopeptidases is included, together with structural features of their respective clans and previous knowledge concerning their regulation by zymogenicity. The **Objectives** of the present research are subsequently summarized.

The **Results** obtained along the thesis belong to three separate but related projects, which are presented as a compendium of publications in subsections **Project 1**, **Project 2** and **Project 3.** The resulting publications are:

Project 1: López-Pelegrín M, Ksiaszek M, Karim AY, Guevara T, Arolas JL, Potempa J & Gomis-Rüth FX. A novel mechanism of latency in matrix metalloproteinases. J Biol Chem 2015; 290(8): 4728-40.

Project 2: López-Pelegrín M, Cerdà-Costa N, Martínez-Jiménez F, Cintas-Pedrola A, Canals A, Peinado JR, Marti-Renom MA, López-Otín C, Arolas JL & Gomis-Rüth FX. A novel family of soluble minimal scaffolds provides structural insight into the catalytic domains of integral membrane metallopeptidases. J Biol Chem 2013; 288(29): 21279-94.

Project 3: López-Pelegrín M, Cerdà-Costa N, Cintas-Pedrola A, Herranz-Trillo F, Bernadó P, Peinado JR, Arolas JL & Gomis-Rüth FX. Multiple stable conformations account for reversible concentration-dependent oligomerization and autoinhibition of a metamorphic metallopeptidase. Angew Chem Int Ed Engl 2014; 53(40): 10624-30.

The **General Discussion** integrates the above-mentioned results and opens new research avenues for the work presented. The thesis ends with the **Concluding Remarks** for each project.

INTRODUCTION

1. PEPTIDASES

Peptidases are enzymes that catalyze the hydrolysis of the amide (peptide) bond in proteins or peptide substrates. Although the terms "proteolytic enzyme" and "protease" are recommended and used as synonymous of peptidase, it has to be noted that peptide bonds can be cleaved by other catalytic mechanisms which do not involve hydrolysis (Rawlings et al, 2011). In addition, the term "proteinase" is further used for peptidases that show specificity for intact proteins rather than peptides (Barrett & McDonald, 1986).

Most peptidases are relatively nonspecific and may target indiscriminately multiple substrates, which results in massive degradation as observed during digestion and processing of misfolded proteins, and in tissue remodeling and development (extensive proteolysis). Alternatively, some peptidases catalyze highly-selective cleavage of one or a small number of peptide bonds (limited proteolysis) that results in their own activation or inactivation as well as of other enzymes, DNA repressors, and peptidic hormones (Cerdà-Costa & Gomis-Rüth, 2014; López-Otín & Bond, 2008).

Although the simplest peptidase in its mature form consists of a single protein domain corresponding to the catalytic unit, most of them are multimodular proteins with different domains or multiple copies of similar domains. They are connected to the catalytic domain (CD) and confer special features to the molecule, allowing all types of sophisticated functions to be carried out. Additionally, these non-catalytic specialized functional modules can regulate the activity of peptidases, affecting substrate specificity, guiding cellular localization, modifying kinetic properties, changing sensitivity to endogenous inhibitors, and preventing untimely activation (see section 1.4). For instance, gelatinases A and B (also known as matrix metalloproteinase (MMP) 2 and 9, respectively) comprise three repeats of a fibronectin type II-like domain inserted into the peptidase domain, which are essential for efficient gelatinolytic and elastolytic activities (Nagase & Woessner, 1999).

1.1. Active site

The three-dimensional structures of peptidases show that the active site is commonly located in a groove, usually called cleft, on the surface of the molecule between adjacent structural (sub)domains. The active site of an enzyme accomplishes a double function: (i) substrate binding and (ii) reaction catalysis. The efficiency of both actions determines the overall activity of the enzyme towards a particular substrate, i.e. it determines the specificity of the enzyme. Binding sites arranged along the groove on one or both sides of the cleavage site are responsible for this specificity, even though exosites and allosteric sites may have a great influence too (Drag & Salvesen, 2010).

Schechter and Berger formulated in 1967 a general nomenclature for subsite position within the active-site cleft, which numbers the residues upstream and downstream of the scissile bond. The cleavage site is located between substrate residues P_1 and P_1' (P for peptide) and numbering increases both in the amino- (N-) terminal direction of the cleaved peptide bond (P_2, P_3, P_4, etc.) and in the carboxy- (C-) terminal direction (P_1', P_2', P_3', etc.). The positions of the substrate match the cognate sites of the enzyme, which are termed S_1, S_2, S_3, etc., and S_1', S_2', S_3', etc. (S for subsites) (Fig. 1).

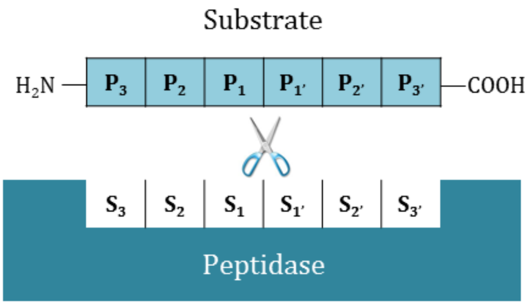


Fig. 1. Subsites of a peptidase-substrate complex. Schematic representation of the active site of a peptidase which is composed of subsites S_3-$S_{3'}$. The cleaved peptide bond (the scissile bond) is indicated by scissors. The positions on the peptide (substrate) are counted from the cleavage point and thus have the same numbering as the subsites on the enzyme. Figure adapted from (Schechter & Berger, 1967).

1.2. Classification

Peptidases can be classified according to different criteria:

1.2.1. Specificity

Peptidases were initially divided on the basis of the reaction they catalyze into two broad groups: (i) endopeptidases, i.e. those that target internal peptide bonds and (ii) exopeptidases, those that act on a free N- or C-terminus. Exopeptidases can be further assorted into aminopeptidases (APs) and carboxypeptidases depending on whether they cleave off N-terminal or C-terminal residues, respectively (Barrett et al, 2013).

This distinction is still maintained in the recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB), which can be consulted at http://www.chem.qmul.ac.uk/iubmb/enzyme/. This Committee is the successor to the Enzyme Commission (EC), which in 1961 devised a classification system for all kinds of enzymes based strictly on the type of reaction they catalyze, assigning them code numbers prefixed by EC. Accordingly, peptidases, already a subclass within hydrolases (EC 3 hydrolases; Subclass 3.4), can be sorted in 14 sub-subclasses (Table 1). Despite being a rather complex system for peptidase classification, the EC system plays an important role in nomenclature since it gives a unique EC number to each peptidase that can be used as an unambiguous reference to this enzyme when names may be ambiguous. Moreover, the system provides a recommended name for each peptidase as well as a list with other names that may be found in the literature referring to the same peptidase.

Table 1. Classification of peptidases according to the EC recommendations

Sub-subclass	Type of peptidase
EC 3.4.11	Aminopeptidases
EC 3.4.13	Dipeptidases
EC 3.4.14	Dipeptidyl-peptidases and tripeptidyl-peptidases
EC 3.4.15	Peptidyl-dipeptidases
EC 3.4.16	Serine-type carboxypeptidases
EC 3.4.17	Metallocarboxypeptidases
EC 3.4.18	Cysteine-type carboxypeptidases
EC 3.4.19	Omega peptidases
EC 3.4.21	Serine endopeptidases
EC 3.4.22	Cysteine endopeptidases
EC 3.4.23	Aspartic endopeptidases
EC 3.4.24	Metalloendopeptidases
EC 3.4.25	Threonine endopeptidases
EC 3.4.99	Endopeptidases of unknown catalytic mechanism

1.2.2. Catalytic type

Peptidases can also be classified according to the chemical groups responsible for catalysis into different catalytic types: serine, threonine, cysteine, aspartic, glutamate, and metallopeptidases. In the first three, the nucleophile group that initiates the attack on the peptide bond is the hydroxyl on the side chain of a serine or threonine, or the thiol on the side chain of a cysteine located in the active site of the peptidase ("protein nucleophiles") (Rawlings & Barrett, 2013) (Fig. 2). In contrast, in aspartic, glutamate and the nucleophile is a water molecule, bound and activated in the catalytic site by the side chains of aspartates and glutamates in aspartic and glutamate peptidases, respectively, or by one or two metal ions in metallopeptidases.

The first catalytic types identified in 1993 by Rawlings & Barrett were serine, cysteine, aspartic, and metallopeptidases (Rawlings & Barrett, 1993). These catalytic types were complemented in 1995 with the discovery of peptidases which had an N-terminal threonine acting as a nucleophile (Seemuller et al, 1995), and in 2004, when certain fungal endopeptidases currently known as eqolysins were discovered to be glutamate peptidases (Fujinaga et al, 2004). More recently, a type of proteolytic enzymes which employs an asparagine as nucleophile has been described (Rawlings et al, 2011). In the autotransporter haemoglobin protease (Hbp) from *Escherichia coli*, the side chain of an active-site asparagine attacks the main-chain carbonyl carbon and creates a stable succinimide ring that leads to self-cleavage and triggers the release of the N-terminal passenger domain from the barrel domain (Rawlings et al, 2011; Tajima et al, 2010). A similar mechanism of autoproteolysis has been observed in the maturation of some viral capsid proteins and intein-containing proteins. This type of peptide-bond cleavage does not involve hydrolysis, and thus these proteolytic enzymes should not be termed peptidases but "lyases" (EC subclass 4.3) according to the EC recommendations.

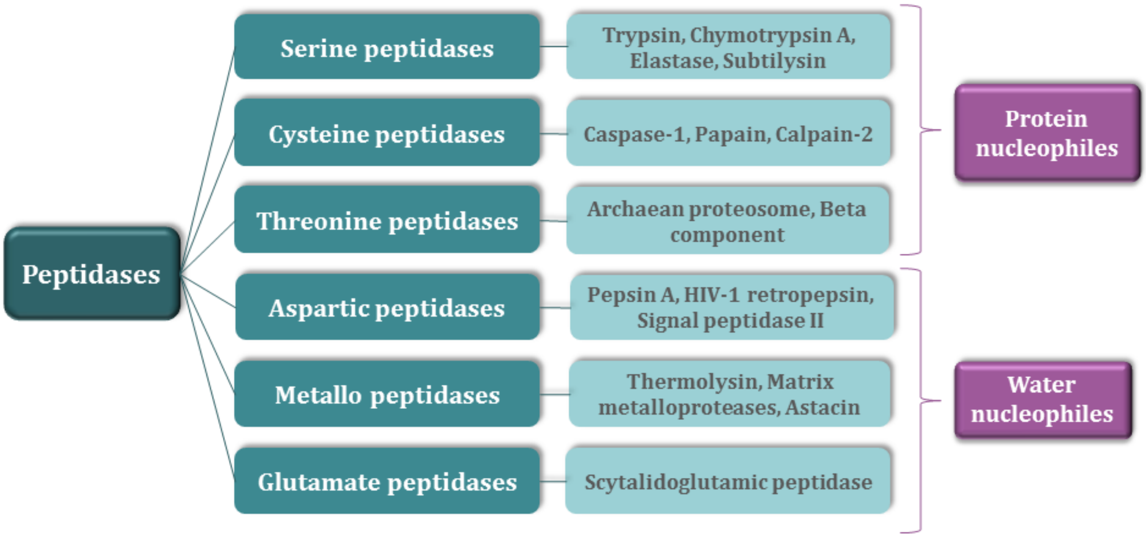


Fig. 2. Classification of peptidases according to the chemical groups responsible for the catalysis. Diagram depicting the six current established classes of peptidases and selected members of each class.

1.2.3. Sequence homology and structural similarity

The MEROPS database at http://merops.sanger.ac.uk provides a sequence-based classification of peptidases. In this hierarchical system, peptidases with homologous sequences are grouped into families, which are further assembled into clans (Rawlings & Barrett, 1993). Peptidases in the same family show homology in the part of the molecule bearing the active-site residues, the "peptidase unit". A family is built up around a peptidase that has been characterized in detail at the biochemical level, which is termed the "type example". The naming of the families follows the system introduced by Rawlings & Barrett, in which each family is named with a letter denoting the catalytic type (S, C, T, A, G, M, N or U, for serine, cysteine, threonine, aspartic, glutamic, metallo, asparagine or unknown, respectively), followed by a number assigned sequentially as the families are recognized in the MEROPS classification (Rawlings et al, 2014b). If a family is overseen the identifier is not re-used and for this reason the numerical identifiers are not correlative. As of August 2015, there are 253 families of peptidases in MEROPS database. In addition, some families are divided into subfamilies because there is evidence of an ancient divergence within the family; for instance, family S1 of serine peptidases has been subdivided into six (S1A to S1F).

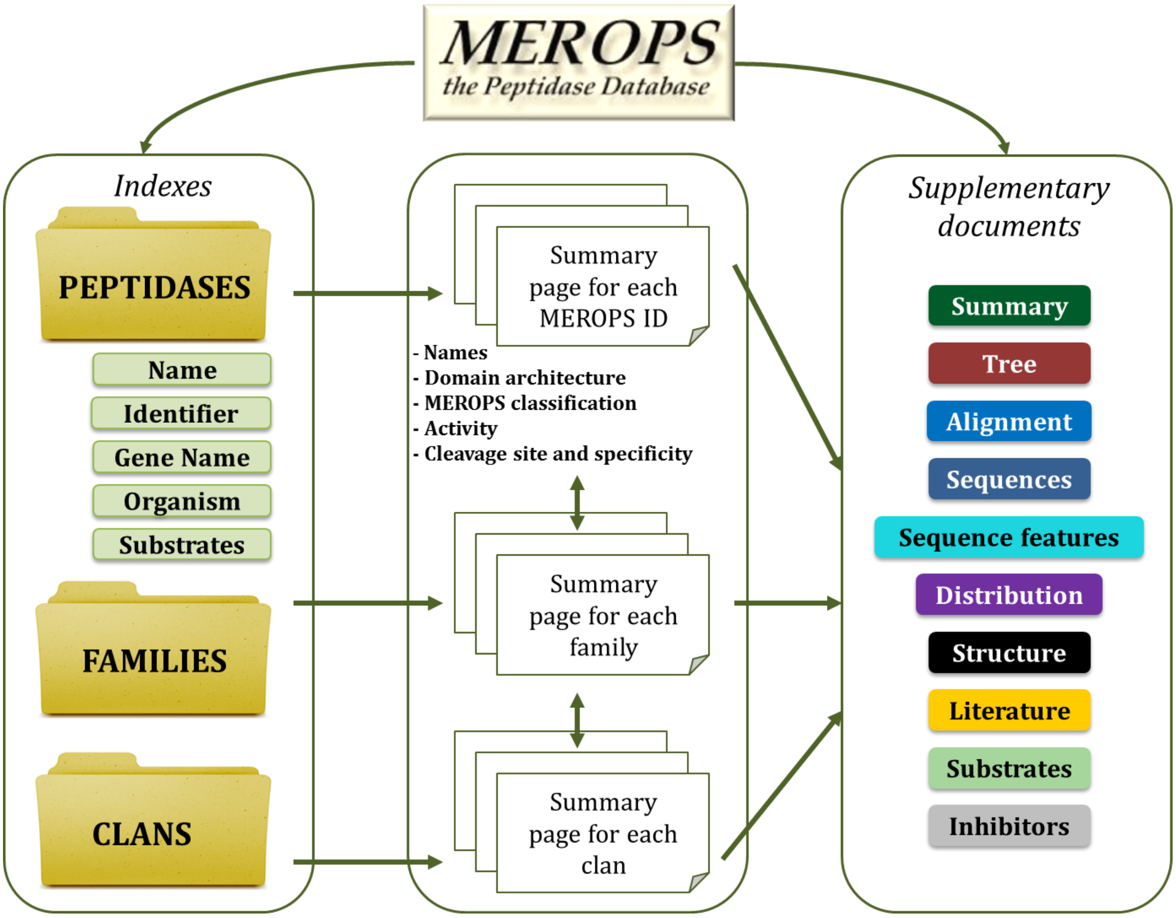


Fig. 3. Classification of peptidases according to the MEROPS database. Simplified diagram of the organization within the MEROPS database. An index of peptidases by name or MEROPS identifier gives access to a set of summary pages, each of which describes a single peptidase. Families and clans are equally linked to individual summary pages containing information on classification and nomenclature, and hypertext links to the relevant entries in online databases for genetics, protein and nucleic acid sequence data and tertiary structure. Figure adapted from (Rawlings & Barrett, 1999)

Nevertheless, it is well established that distant relationships cannot be inferred clearly from the primary structures and therefore fold similarities, which persist in evolution much longer than similarities in amino acid sequence, can be used to reflect evolutionary relationships between groups of proteins. The MEROPS database gathers distantly related groups into clans. The best evidence to support the formation of a clan is similarity in three-dimensional structures when the data are available, but the arrangement of catalytic residues in the polypeptide chains, as well as limited similarities in amino acid sequence around the catalytic amino acids are also taken into account. Each clan has a two-letter identifier, with the first letter indicating the catalytic type of the gathered families followed by an arbitrary second capital letter.

The MEROPS database not only provides a classification system but also acts as a catalogue, including summary pages with detailed information of each peptidase, family and clan. Summaries describe the classification and nomenclature of the peptidases and offer links to supplementary pages showing sequence identifiers, the structure if known, literature references and more. Thus, the database functions as an organizational framework around which a variety of related information is assembled (Rawlings et al, 2014a) (Fig. 3).

1.3. Biological significance

Peptidases are not mere protein or peptide-degrading enzymes but extremely important molecules distributed across all kingdoms of living organisms and involved in vital processes such as DNA replication and transcription, cell-cycle regulation, processing of peptide hormones, release of cytokines and growth factors, cell-cell fusion, cell adhesion and migration, viral polyprotein processing, bacterial cell wall biosynthesis, metabolism of antibiotics, ovulation and fertilization, embryonic development, morphogenesis, angiogenesis, alimentary uptake, bone formation, tissue remodeling, neuronal growth, immune and inflammatory responses, homeostasis, blood coagulation and fibrinolysis, and aging and apoptosis (López-Otín & Bond, 2008; Neurath & Walsh, 1976; Turk, 2006). Consistently, any alterations in proteolytic systems contribute to pathological conditions that range from arthritis, osteoporosis, fibrosis, and cancer to autoimmune, cardiovascular and neurological diseases (Gomis-Rüth, 2008; Ketelhuth & Back, 2011; Nagase, 2001; Nalivaeva et al, 2012).

Peptidase activity is not restricted to the interior of organisms because they have important exogenous functions. Some bacteria secrete peptidases as virulence factors during host infection that degrade essential components of connective host tissues in the dermis (e.g. collagen and elastin) and directly or indirectly contribute to inflammation in the host. They also inactivate key proteins in host defense, disarm the complement system, and prevent the infiltration of the infected sites with immune cells by inactivation of chemoattractants. These strategies are required for bacterial invasion, survival, proliferation and colonization of the host in a hostile environment (Jusko et al, 2012; Koziel & Potempa, 2013; Miyoshi et al, 1997; Potempa & Pike, 2005; Wu & Chen, 2011). Peptidases are also found in the venoms used by predators such as in poisonous snakes, scorpions and spiders, or, conversely, in poison-mediated defense strategies against predators such as in scorpion fish (Carrijo et al, 2005; Fox & Serrano, 2009; Rusmili et al, 2014).

1.4. Activity regulation

Since peptidases catalyze essentially irreversible reactions, their expression and activity must be strictly regulated spatially and temporally in order to prevent inappropriate proteolysis. Proteolytic activity can be regulated by different mechanisms:

1.4.1. Zymogenicity

The most extended mechanism of peptidase regulation is the secretion of peptidases as inactive precursors or zymogens. In many cases, zymogenicity is exerted by N-terminal prosegments or prodomains (PDs) that prevent access of substrates to the active-site cleft of the enzyme (Fig. 4). Full activity is achieved upon PD removal, usually by means of a proteolytic cleavage that can be either autolytic (i.e. self-cleavage) or heterolytic (i.e. catalyzed by another peptidase) in a process termed maturation. Such PDs often fold independently and guide the folding process of the cognate peptidase domains, thus acting as intramolecular chaperones (Bryan, 2002; Eder & Fersht, 1995; Khan & James, 1998; O'Donohue & Beaumont, 1996). They can also participate in the compartmentalization of the mature enzyme (Baliga et al, 2003; Demidyuk et al, 2010b).

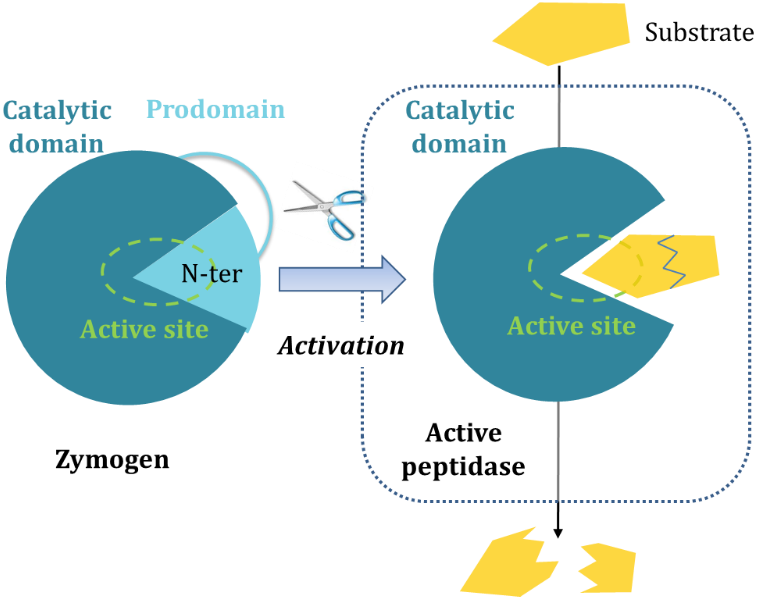


Fig. 4. Proteolytic activity regulation by zymogenicity. Schematic representation of prodomain removal from the active site of a zymogen by means of heterolytic or autolytic cleavage (represented by scissors), which next results in an active peptidase. Figure adapted from (Deu et al, 2012).

Peptidases implicated in proteolytic pathways such as blood coagulation and fibrinolysis, gastrulation, and apoptosis are usually expressed as zymogens. Subsequently, the action of one peptidase upon the latent form of another one amplifies the original signal (in a cascade-like-manner) and enables additional regulatory points in the pathway. The first event in such proteolytic pathways, the actual sensing of the signal, can involve additional factors or cytosolic platforms such as the apoptosome (Riedl & Salvesen, 2007). Zymogenicity also plays a central role in maintaining microbial virulence factors in an inactive form inside the pathogen until secretion at the site of infection (Veillard et al, 2013).

1.4.2. Inhibitors

Peptidase inhibitors are, together with zymogens, major regulators of peptidase activity. Peptidase inhibitors can be classified into endogenous or exogenous depending on whether their action takes place over peptidases from the same organism or from a different one, respectively. Most known endogenous peptidase inhibitors are proteins even though some microorganisms produce small non-protein inhibitors (e.g. pepstatin A, chymostatin, leupeptin, antipain, phosphoramidon and bestatin) (Bode & Huber, 2000; Rawlings, 2010; Sabotic & Kos, 2012). The number of peptidase inhibitors isolated and identified so far is extremely large, as it is the number of three-dimensional structures available. The MEROPS database collects them in families and clans in a similar way as it does with peptidases (http://merops.sanger.ac.uk/inhibitors/) (Rawlings et al, 2004). Again, as in peptidases, many criteria can be taken into account for their classification. One of them distinguishes between reversible and irreversible inhibitors or molecular traps (Fig. 5).

The vast majority of peptidase inhibitors present a critical segment that binds to the peptidase and blocks the active site, thus preventing access to the substrate in a typically reversible process. Reversible inhibitors can be classified by their mechanism of action into catalytically (canonical) and noncatalytically (noncanonical) competent inhibitors (Farady & Craik, 2010). Canonical inhibitors follow a standard mechanism and bind to the peptidase in a substrate-like manner, by inserting a “reactive loop” into the active site. While bound to the peptidase, the “scissile bond” is hydrolyzed very slowly. Since products from the reaction are not released, the amide bond can be eventually religated (Zakharova et al, 2009). Noteworthy, most reversible inhibitors do not directly block the catalytic residues but interact with subsites adjacent to the active site in a noncatalytically competent manner. An example of such mechanism is provided by cystatins, which inhibit papain-like cysteine peptidases (Bode & Huber, 2000). Similarly, tissue inhibitors of matrix metalloproteinases (TIMPs) function not only via hairpin loops and N-terminal residues over peptidase substrate-binding pockets but also interfering with the catalytic metal ion (Brew et al, 2000). Moreover, several reversible peptidase inhibitors target additional secondary-binding sites outside the active site (exo-sites) that are critical for peptidase activity. Irreversible or molecular trap inhibitors are typically large molecular-weight proteins, which act as a substrate and use catalytic machinery of the enzyme to trap and irreversibly inhibit it. There are only two large families of trapping inhibitors: the universal inhibitors called macroglobulins (MEROPS database family I39) and the serine peptidase inhibitors called serpins (MEROPS database family I4). In the macroglobulin family of inhibitors (α_2-macroglobulins and relatives), the cleavage of reactive (bait) loops of the inhibitor by the peptidase triggers conformational changes that sequester the peptidase within the rearranged inhibitor molecule. Although there are only a few examples of covalent macroglobulin-peptidase complexes, the peptidase generally remains intact (Laskowski & Kato, 1980; Luthy et al, 1973). Therefore, while large protein substrates are occluded from the active site, small-molecule substrates can still be hydrolyzed by peptidases complexed with the inhibitor (Garcia-Ferrer et al, 2015; Marrero et al, 2012). In the case of the serpin family, a covalently linked acyl-enzyme complex is formed with the target peptidase after the cleavage of the reactive loop. Next, a large conformational change takes place, which translocates the peptidase to the distal side of the inhibitor and the resulting steric collisions completely distort the active site of the peptidase (Whisstock & Bottomley, 2006; Ye & Goldsmith, 2001).

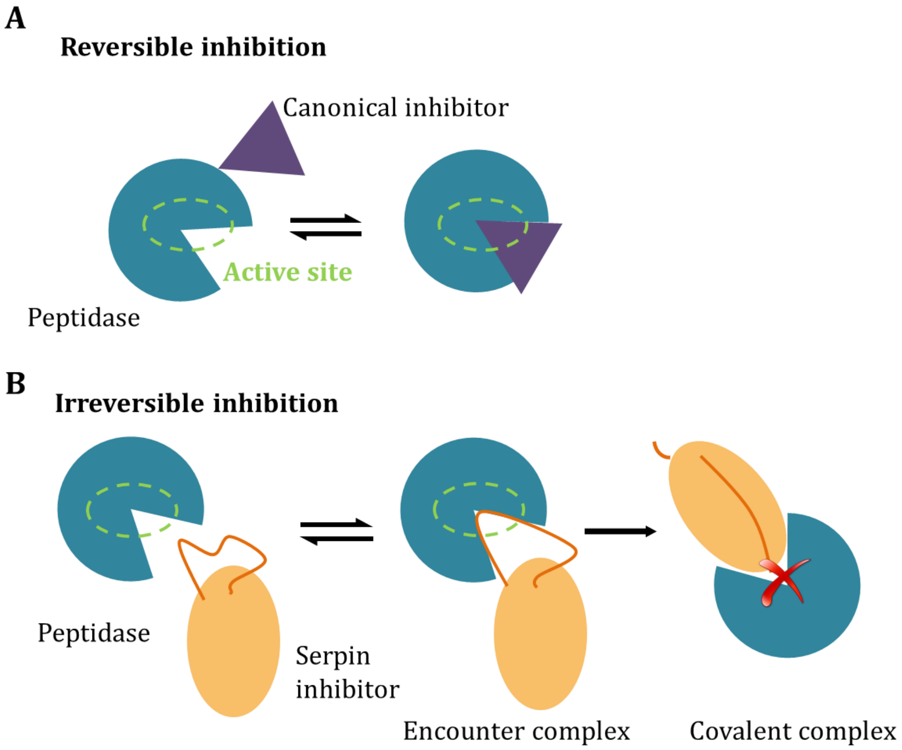


Fig. 5. Proteolytic activity regulation by reversible and irreversible inhibitors or molecular traps. **(A)** Schematic representation of a canonical inhibitor, which plugs the active site of the peptidase and prevents substrate access in a reversible manner. **(B)** Schematic representation of a serpin inhibitor, which binds the peptidase through a reactive loop in a substrate-like manner. It then completes the inhibition process by causing a conformational change and disrupting the active site of the peptidase irreversibly. Figure adapted from (Beinrohr et al, 2008).

1.4.3. Transcriptional and post-transcriptional regulation

The amount of transcripts coding for a peptidase depends on gene-copy number and on many factors that affect its transcription. Post-transcriptional events such as alternative splicing are thus determinant. A good example is that of MMPs, which are not expressed under normal quiescent conditions, but their transcription can be induced in tumor or stromal cells by various signals such as cytokines, growth factors, chemokines, and oncogene products (Overall & López-Otín, 2002).

1.4.4. Post-translational modifications

Peptidases endure many post-translational modifications such as glycosylation, metal binding, and disulfide bridging, which can modulate their activity. This type of modulation can be illustrated with the N-glycosylation of ADAM-8 (A Disintegrin And Metalloproteinase 8) recently shown to be required for its correct autocatalytic processing, localization, stability, and activity in a certain type of breast cancer cells, where it is abundantly expressed (Srinivasan et al, 2014). Another work reported the effect of site-directed mutagenesis on N-glycosylation and glycosylphosphatidylinositol sites over the expression of peptidase leishmanolysin (McGwire & Chang, 1996).

1.4.5. Cofactor requirement

Cofactors such as organic molecules, proteins, lipids, ions, and glycosaminoglycans are essential for the catalysis of certain enzymes and they have been described to modulate the activation of a number of peptidases. For instance, during blood coagulation, tissue factor serves as a cofactor for factor VIIa (“a” stands for activated) in the activation of the serine peptidase zymogen factor X via limited proteolysis. In turn, factor Xa uses activated coagulation factor Va as a cofactor during the formation of the prothrombin complex that eventually leads to the formation of thrombin and fibrin (Chen et al, 2006; Nesheim et al, 1981; Rosing et al, 1980; Schuijt et al, 2013).

1.4.6. Allosteric regulation

The interaction of a peptidase with specific modulators (proteins or small molecules), which bind to a site distinct from the active site and, thus, not involved in substrate recognition, induces conformational changes on the former and alters its function (either by enhancing or inhibiting it). Allosteric regulators do not need to be chemically similar to the substrate in order to compete against it for gaining access to the active site (Laskowski et al, 2009; Shen, 2010). Specific domain interactions of zymogens with other proteins or glycosaminoglycans can facilitate proteolytic removal or displacement of the PD from the active site. For example, glycosaminoglycans have been found to contribute to the autocatalytic activation of cysteine peptidases cathepsins, especially cathepsin K, by converting zymogens into better substrates. They favor an open conformation of the zymogen and thereby promote activation not only at acidic pH but also at values closer to neutral pH (Novinec et al, 2014).

1.4.7. Protein-protein interactions and oligomerization

Proteins bind to each other through a combination of hydrogen bonds, van der Waals contacts, and salt bridges. These interactions take part in zymogen activation, substrate accessibility, and binding to receptors and inhibitors, among other processes. Protein association can hence give rise to stable high-molecular-mass oligomeric structures as observed in the proteasome, tripeptidyl peptidase II, and tricorn peptidase complexes, as well as in secreted or membrane-bound meprin (Bertenshaw et al, 2003). Other oligomeric complexes have a transient nature and their association controls the activation of one of the constituting proteins in a certain location or time point. This is the case for the activation of MMP-2 by membrane-type 1 (MT1) MMP, which involves the formation of a complex between MT1-MMP, TIMP-2, and the MMP-2 zymogen. This ternary complex promotes the activation of MMP-2 by a TIMP-2-free MT1-MMP molecule *in trans* (Toth et al, 2000). A well-studied example of enhanced enzyme-substrate contact occurs during blood fibrinolysis, when the very inefficient serine peptidase tissue-type plasminogen activator (PA) is targeted to the fibrin clot surface, which serves as a template onto which both proteins bind. The ternary complex involving tissue-type PA, fibrin, and plasminogen enhances the activation efficiency of the latter (Hoylaerts et al, 1982).

Protein assembly functions to maintain peptidases in a catalytically competent form, as in the active heparin-stabilized tetramer of human mast-cell β-tryptase, whose tetramers rapidly convert into inactive monomers after dissociation of heparin (Schechter et al, 1995; Sommerhoff et al, 2000). In contrast, it has been described that oligomerization might inhibit activity, such as in phospholipase A2 (Fremont et al, 1993; Hazlett & Dennis, 1985).

1.4.8. Compartmentalization and trafficking

Proteolysis is finally regulated by the different distributions in space and time of the modulators described above (peptidases, activators, cofactors, inhibitors, substrates, etc.). Compartmentalization and peptidase trafficking between subcellular compartments or from intracellular to extracellular localization sequester enzymes into specific environments with proper biochemical conditions that promote optimal catalytic activity. These conditions include calcium concentrations, redox conditions ranging from oxidizing (in the compartments of the secretory pathway and extracellular space) to reducing (within endo-lysosomal compartments), and pH values ranging from basic to highly acidic. This is the case for the cysteine peptidases, cathepsins, whose premature activation is prevented by an N-terminal PD that dissociates from the CD once the zymogen is transported to the acidic lysosomal lumen (Mach, 2002; Pungercar et al, 2009; Wiederanders et al, 2003). In addition, these peptidases have such strict pH-requirements for both proteolytic activity and stability that efficient proteolysis mostly takes place within the borders of acidic lysosomes, at pH near 5 (Jordans et al, 2009).

2. METALLOPEPTIDASES

Metallopeptidases (MPs) are peptidases that require divalent metal ions to carry out their catalytic function. Although the catalytic metal ion found in the active site of MPs is usually zinc, it may likewise be cobalt, manganese or nickel. Most MPs require only one metal ion for catalysis (known as mononuclear or monometallic MPs) but others, like cobalt and manganese MPs, possess two metal ions that participate in the reaction. The catalytic metal ion is usually coordinated by three protein residues and a solvent molecule, which is essential for the hydrolysis of the peptide bond. In many peptidases two of these coordinating residues are histidines found within a HEXXH motif or short zinc-binding consensus sequence (ZBCS) (one-letter residue code; zinc-binding residues underlined; X for any residue) (Cerdà-Costa & Gomis-Rüth, 2014). A third coordinating residue is located downstream to this motif and can be a glutamic acid, a histidine, or an aspartic acid. Eventually, as it happens in some MMP zymogens, the zinc ion of the active site is coordinated by an unpaired cysteine of the PD, which keeps the enzyme inactive. Upon activation, the PD is removed and a water molecule replaces the cysteine to give rise to an active MMP (Nagase, 2001; Nagase & Woessner, 1999).

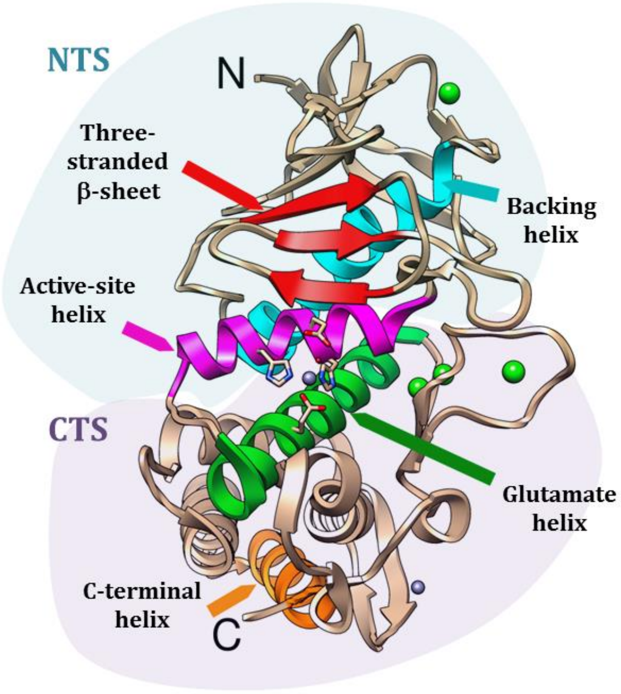


Fig. 6. Standard orientation of metallopeptidases. Richardson-type plot depicting the catalytic domain of *Bacillus thermoproteolyticus* thermolysin, which is divided into an N-terminal subdomain (NTS) and a C-terminal subdomain (CTS) by a central active-site cleft. Consensus secondary structure elements are colored and labeled. The metal ion is shown as a grey sphere and is bonded by relevant residues from the HEXXH motif of an “active-site helix” and the glutamate of a “glutamate helix”, which are shown as sticks. Figure adapted from (Gomis-Rüth et al, 2012).

A standard orientation was defined for structural representations of MPs, in which substrate binding as well as structural details of the catalytic metal-binding site are best visualized (Gomis-Rüth et al, 2012). This entails a frontal view of the horizontally-aligned active-site cleft, with the substrate bound from left (N-terminal non-primed side) to right (C-terminal primed side) and the catalytic metal ion residing at the cleft bottom at roughly half width. With this view most MPs are bifurcated by the cleft into an upper N-terminal subdomain (NTS), which includes the HEXXH motif embedded in a nearly horizontal “active-site helix”, and a lower C-terminal subdomain (CTS) that includes the third zinc ligand. The NTS additionally shows an oblique “backing helix” behind the “active-site helix”, and a mixed β-sheet of at least three strands paralleling the cleft (Fig. 6). The lowermost “upper-rim” strand of the sheet runs antiparallel to the substrate bound in the cleft and therefore contributes both to delimitating the cleft top and to binding of the substrate main-chain on its non-primed side through β-ribbon-like interactions (Gomis-Rüth et al, 2012; Schechter & Berger, 1967).

2.1. Reaction mechanism

The first three-dimensional structures of MPs determined were those of bovine pancreatic carboxypeptidase A and *Bacillus thermoproteolyticus* thermolysin (Lipscomb et al, 1968; Matthews et al, 1972). In the following decades, a large number of structures have become available, providing a consensus model mechanism of reaction for MPs. Thus, in mononuclear MPs, the catalytic mechanism is based on the activity of a water molecule, which coordinates the zinc and is further bound to and polarized by a general base/acid (the glutamic acid within the HEXXH motif) (Fig. 7A). Upon substrate binding to the active-site cleft, the water molecule is slightly displaced from the zinc ion and the scissile carbonyl oxygen of the substrate is further bound and polarized by the zinc (Fig. 7B). The hydrolysis of a peptide bond is mediated by the nucleophilic attack of this water molecule on the carbonyl carbon of the scissile bond, giving rise to a *gem*-diolate tetrahedral reaction intermediate (Fig. 7C), which is stabilized not only by the metal but also by neighboring protein residues (histidines, arginines, and tyrosines). The intermediate evolves under assistance of the general base/acid, which abstracts a proton from the water molecule and then transfers it to the nitrogen of the leaving group. This eventually leads to peptide-bond disruption (Fig. 7D).

2.2. Classification

Similarly to what happens with peptidases in general, there are different ways of classifying MPs. As of August 2015 (Release 9.13), the MEROPS database includes 71 families of MPs and 45 subfamilies. In addition, 15 clans cluster families according to evolutionary relationships.

Nevertheless, it is not always possible to assign a clan for every family as occurs with seven families, namely M73, M77, M79, M82, M87, M88, and M96.

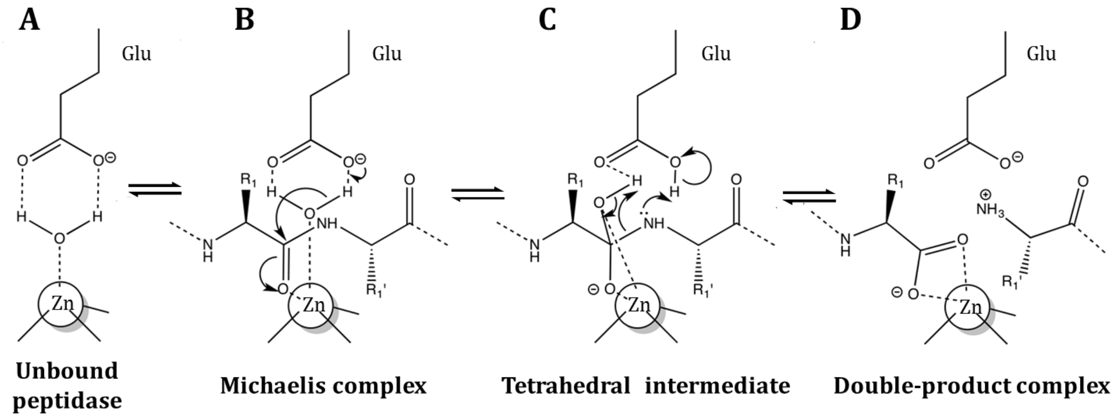


Fig. 7. Generally-accepted catalytic mechanism of mononuclear metallopeptidases. The catalytic solvent molecule is bound first to the catalytic metal ion (white sphere) and the general base/acid (glutamic acid; Glu) in the active site in the absence of a peptidic substrate **(A)**. Once the substrate is accommodated in the cleft and the Michaelis complex is formed **(B)**, the polarized solvent molecule attacks the scissile carbonyl group, which leads to a tetrahedral reaction intermediate **(C)**. The latter resolves in scissile bond breakage and double proton transfer to the newly formed α-amino group to render a double-product complex **(D)** (Matthews, 1988). Figure adapted from (Cerdà-Costa & Gomis-Rüth, 2014).

A different approach divides mononuclear MPs into tribes, clans and families by taking into account their particular metal-binding signature (residues coordinating the catalytic metal and general acid/base) and overall fold similarity (Fig. 8). Accordingly, MPs containing zinc as a catalytic metal and the short ZBCS, HEXXH, were grouped into a tribe termed “zincins” (Bode et al, 1993). MPs exhibiting an inverted zinc-binding signature (HXXEH) were grouped into another tribe termed “inverzincins” (Becker & Roth, 1992). Other tribes of MPs have been described over the years: αβα-exopeptidases, which comprise the funnelin clan of metallocarboxypeptidases (Gomis-Rüth, 2008; Gomis-Rüth et al, 2012); LAS MPs, named according to the founding group members lysostaphin, D-Ala-D-Ala carboxypeptidase and sonic hedgehog (Bochtler et al, 2004; Marcyjaniak et al, 2004); and relatives of hydrogenase maturating factor (HybD) (Fritsche et al, 1999).

Most mononuclear MPs belong to the zincin tribe of MPs. Depending on the nature and position of the third zinc ligand they can be further classified into clans (Fig. 8). Traditionally and most-extensively studied clans are the “gluzincins” and the “metzincins” (Bode et al, 1993; Cerdà-Costa & Gomis-Rüth, 2014; Hooper, 1994). In metzincins, the third zinc-ligand is a conserved histidine or aspartic acid located within an extended ZBCS (HEXXHXXGXXH/D+M), whereas in gluzincins this role is accomplished by a glutamic acid (HEXXH+E). The rest of clans within the zincin tribe share an aspartic acid as a third zinc ligand but differ in the nature of the residue below the metal and in their overall fold: in aspzincins it is a conserved alanine (Fushimi et al, 1999); in

S2P-zincins (see section 4.1.2. for nomenclature) it is a proline (Feng et al, 2007); and in FtsH-like AAA MPs it is apparently not a protein residue but rather a solvent molecule (Bieniossek et al, 2006; Langklotz et al, 2012).

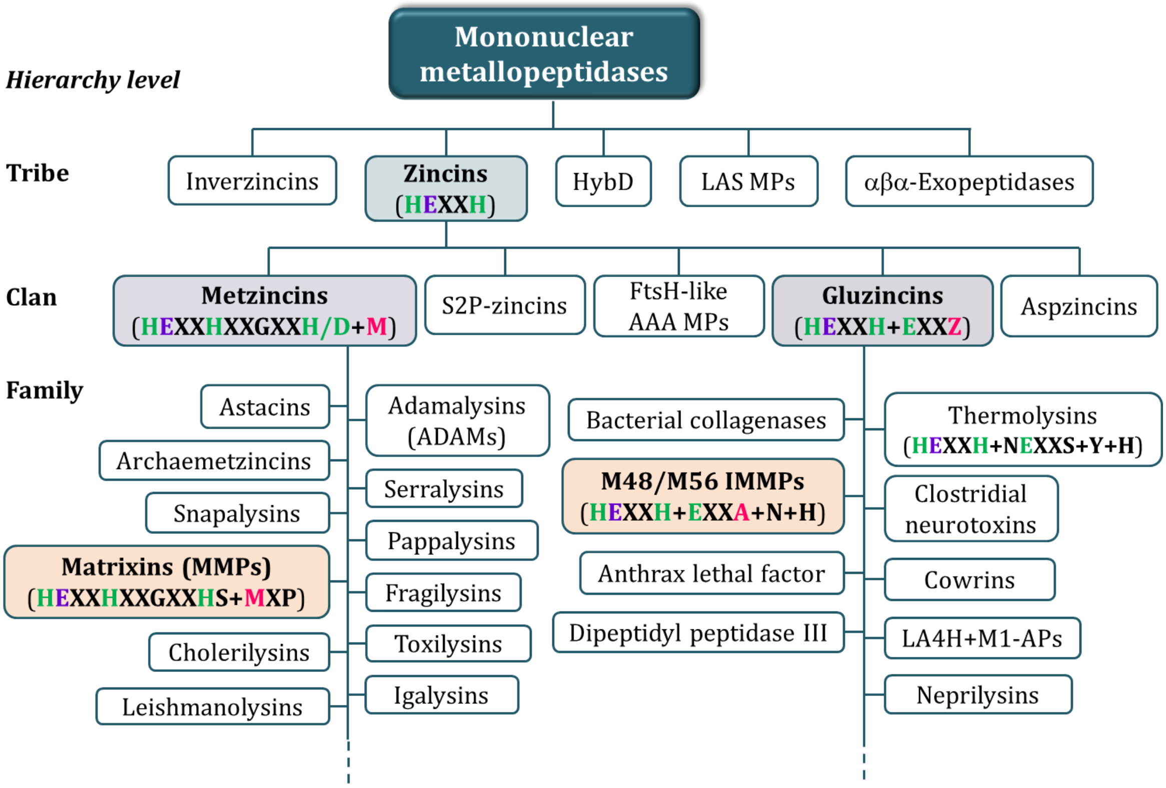


Fig. 8. Classification of mononuclear metallopeptidases. Scheme of mononuclear metallopeptidases (MPs), which are divided into different tribes characterized at the structural level: inverzincins, zincins, relatives of hydrogenase maturating factor (HybD), LAS MPs, and αβα-exopeptidases. These tribes are further subdivided into clans, which in turn give rise to families. The tribe, clans, and families including the peptidases studied in the present thesis are highlighted in blue, violet, and orange, respectively. Selected residues of these categories are detailed in bold. Metal-binding residues are shown in green, general base/acid residues in violet, and residues/molecules occupying the position of the "Met-turn" or "Ser/Gly-turn" beneath the metal site in pink. Other residues engaged in substrate binding, stabilization of the reaction intermediate, and/or catalysis are further shown in black, except for X, which stands for any residue and is only used as a spacer within motifs. In gluzincins Z stands for A/S/G/T. Figure adapted from (Cerdà-Costa & Gomis-Rüth, 2014).

2.3. Activity inhibition

MPs have essential roles in the physiology and pathology of living organisms, so that they require precise regulation (Grötzinger & Rose-John, 2013; Nagase & Murphy, 2013; Noel & Sounni, 2013; Stöcker & Gomis-Rüth, 2013). PDs and endogenous inhibitors are major regulators of MP activity, as for peptidases in general, but there are other specific mechanisms through which MPs can be inhibited. These strategies can involve peptidic or small-molecule inhibitors, chelating agents, and excess of metal.

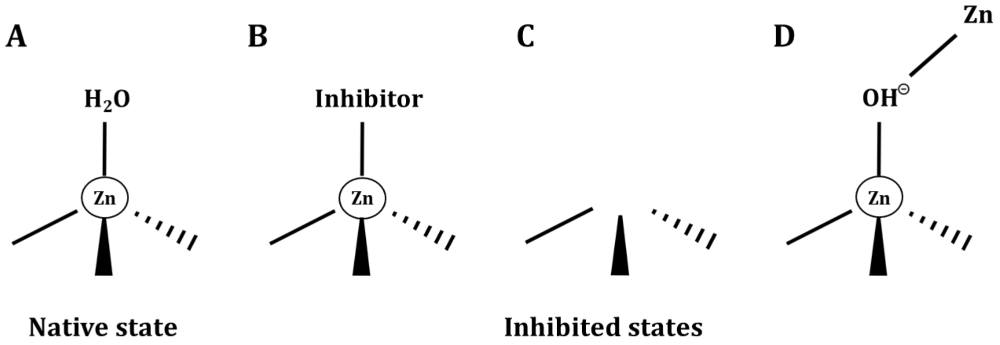


Fig. 9. Proteolytic activity inhibition in zinc-dependent peptidases. **(A)** Native state. **(B)** Water displacement by a specific inhibitor designed to bind the zinc. **(C)** Removal of the zinc ion by metal chelators. **(D)** Binding of a second zinc ion to the catalytic metal-bound water.

Synthetic small molecule inhibitors displace the catalytic water molecule from the coordination sphere of the zinc and bind tightly to it, mimicking the Michaelis complex and thus preventing substrate binding (Fig. 9B). They usually contain a metal-binding group such as carboxylate (R-COOH), thiol (R-SH), phosphate (R-PO_3) or hydroxamic acid (R-CO-NH-OH) groups, to which a series of other chemical groups are attached to fit in the subsites of the MP and provide specificity to the inhibitor (Georgiadis & Dive, 2015; Verma, 2012). Chelating agents such as ethylenediaminetetraacetic acid (EDTA), ethyleneglycoltetraacetic acid (EGTA), and 1,10-phenantroline, in turn, are able to bind with high affinity and sequester the catalytic metal in a bi-, tri- or tetradentate fashion to form stable ring structures that render inactive metal-depleted enzymes (Fig. 9C). EDTA and EGTA inactivate not only zinc-dependent MPs but also some calcium-stabilized peptidases from other classes. Accordingly, 1,10-phenanthroline is preferred as MP inhibitor since it has much higher affinity for zinc (2.5×10^6 M^{-1}) than for calcium (3.2 M^{-1}); i.e. 1 mM 1,10-phenanthroline generally inactivates an MP, whereas it does not remove calcium from a calcium-binding protein. This is commonly a reversible process in which the elimination of the chelating agent and the addition of the metal leads to the recovery of proteolytic activity (Beynon & Bond, 2001). Finally, high concentrations of zinc, in the millimolar range, often inhibit MPs. Inhibition results from the formation of zinc monohydroxide that bridges the catalytic zinc ion and a side-chain in the active site of the enzyme (Fig. 9D). This inhibition is competitive with substrates, as thoroughly studied with the funnelin metallocarboxypeptidase A from bovine (Mock & Wang, 1999). It was demonstrated that this exopeptidase binds a second zinc in the pH range 7-10 when using equilibrium dialysis. Structural studies further revealed that the second zinc binds to the enzyme active site, establishing a distorted tetrahedrally coordinated complex, which involves the catalytic glutamic acid, a water molecule, a chloride ion, and a hydroxide ion (Gómez-Ortiz et al, 1997). The inhibitory zinc holds the hydroxide at nearly the same location as a previously observed active-site water molecule and probably perturbs the substrate positioning and stereochemical rearrangements required for substrate cleavage during catalysis. Besides carboxypeptidase A, several zincin MPs have been shown to be inhibited by excess of zinc ion, namely thermolysin, human neutrophil collagenase, and angiotensin converting enzyme, among others (Bunning et al, 1983; Holmquist & Vallee, 1974; Mallya & Van Wart, 1989).

3. METZINCIN METALLOPEPTIDASES

The metzincin clan of MPs comprises a C-terminally extended ZBCS, HEXXHXXGXXH/D+M (one-letter residue code; zinc-binding residues underlined; X for any residue), with a third zinc-binding histidine or aspartic acid (see section 2.2 and Fig. 8). In addition, metzincins show a conserved glycine (only pappalysins have an asparagine here, see Fig. 10) between the second and third metal ligands and a conserved methionine in the third position of a 1,4-β-turn just underneath the catalytic ion, to which the clan owes part of its name, viz. methionine-containing zincins. The so-called "Met-turn" forms a hydrophobic pillow below the catalytic zinc (Bode et al, 1993; Gomis-Rüth, 2003; Stöcker et al, 1995).

3.1. Classification and main common structural features

Individual metzincin families are mainly distinguished by the residue following the third zinc-coordinating residue and the residues surrounding the methionine in the "Met-turn" (Fig. 8). To date, more than 300 deposited three-dimensional structures provide detailed knowledge at the molecular level for the 12 current metzincin families: astacins (Bode et al, 1992); serralysins (Hamada et al, 1996); leishmanolysins (Schlagenhauf et al, 1998); snapalysins (Kurisu et al, 1997); MMPs *alias* matrixins (Tallant et al, 2010b); adamalysins *alias* ADAMs (Gomis-Rüth et al, 1993); pappalysins (Tallant et al, 2006); archaemetzincins (Waltersperger et al, 2010); igalysins (unpublished results, see (Cerdà-Costa & Gomis-Rüth, 2014); toxilysins (Ng et al, 2013); fragilysins (Goulas et al, 2011); and cholerilysins (Yu et al, 2012). In addition to those structurally characterized, other families may enlarge the clan based on the presence of the characteristic extended ZBCS pattern with small variants in sequences reported from genome sequencing projects. These families have been tentatively referred to as gametolysins, thuringilysins, coelilysins, ascomycolysins, and helicolysins (Gomis-Rüth, 2003; Gomis-Rüth, 2009).

The available structural information reveals that metzincin CDs share a common scaffold and active-site environment, into which each family has grafted characteristic and distinguishing structural elements**.** Their globular CDs span 130-260 residues and are divisible into an upper NTS and a lower CTS with respect to a central active-site cleft, with the catalytic zinc found at its bottom (Cerdà-Costa & Gomis-Rüth, 2014). Some common structural elements of most metzincins include five β-strands and three long α-helices arranged in the following identical sequential order: βI, αA, βII, βIII, βIV, βV, αB and αC (Fig. 10).

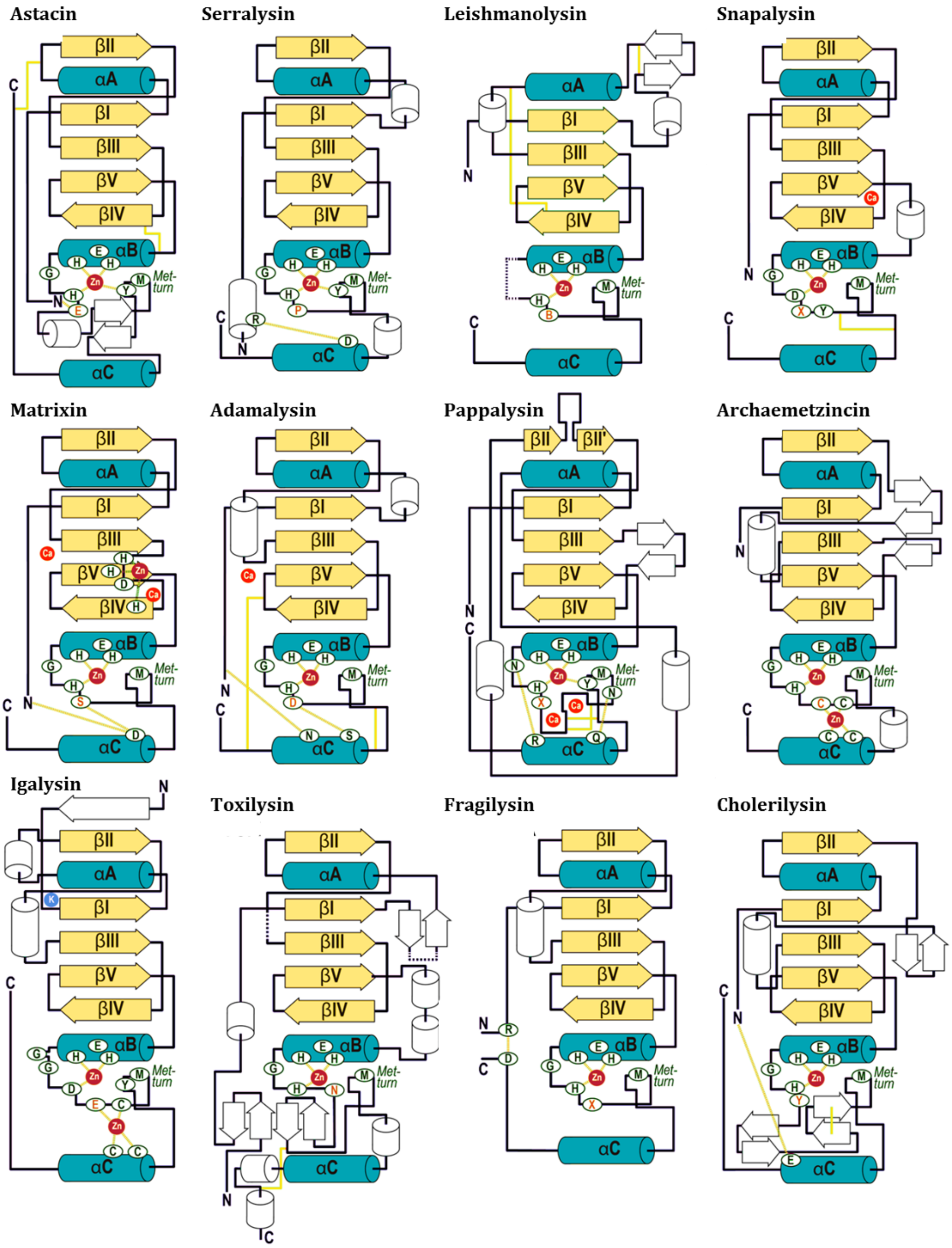


Fig. 10. Topology of metzincin catalytic domains. Scheme showing the regular secondary-structure elements (strands as arrows and helices as rods) of a representative member of each of the 12 structurally-characterized metzincin families (top left to right bottom): *Astacus astacus* astacin, *Pseudomonas aeruginosa* serralysin, *Leishmania major* leishmanolysin, *Streptomyces caespitosus* snapalysin, human matrixin MMP-8, *Crotalus adamanteus* adamalysin, the pappalysin *Methanosarcina acetivorans* ulilysin, *Methanopyrus kandleri* archaemetzincin, *Bacteroides ovatus* igalysin, *Escherichia coli* toxilysin, *Bacteroides fragilis* fragilysin, and *Escherichia coli* cholerilysin. The common strands and helices are shown in yellow and blue, respectively, and labeled βI-βV and αA-αC. Hydrogen bonds/metal-ligand bonds and disulfide bonds are shown as dashed and yellow solid lines, respectively. Disordered segments or points of insertion of extra domains are indicated by black dashes. Family-specific residues are in red. B stands for bulky, hydrophobic residues and X for any residue. Figure adapted from (Cerdà-Costa & Gomis-Rüth, 2014).

3.1.1. The N-terminal subdomain

The consensus NTS generally displays a five-stranded twisted β-sheet (βI to βV) on the top (except leishmanolysin, which has four strands) flanked by two long α-helices, αA and αB (also referred to as the "backing helix" and the "active-site helix" respectively), with the former helix inserted between strands βI and βII (Gomis-Rüth, 2003). All strands but the fourth one (βIV) are parallel to each other and to a putative substrate bound in the cleft. Antiparallel strand βIV forms the lower-edge of this subdomain and creates an upper-rim or northern-wall of the active site, running antiparallel to a substrate when bound to the cleft, and anchoring it through inter-main-chain interactions. Thereafter, the chain turns back to βV strand, which also binds the substrate mainly on its non-primed side. The loop segment connecting βIII and βIV (LβIIIβIV), also known as the "bulge-edge segment", generates protruding structural elements, which affect enzyme-substrate interactions on the primed side of the active-site clefts (mainly subsites S_1' and S_2') and therefore has a role in the substrate specificity of each family.

3.1.2. The active site

All metzincin families exhibit similar active sites. Helix αB, which encompasses the first half of the extended ZBCS, is topologically equivalent in the prototype structures of the distinct metzincin families (Gomis-Rüth, 2009) (Fig. 11). The two zinc-binding histidines within the motif are separated by a single helix turn, which allows coordination of the catalytic zinc by the two respective imidazole side chains. At the end of helix αB, the polypeptide chain takes a sharp turn downward mediated by the glycine of the consensus sequence and leaves the NTS (see Fig. 10).

3.1.3. The C-terminal subdomain

The consensus CTS is shorter than the NTS and contains few regular secondary-structure elements: mainly a C-terminal helix (αC) at the end of the polypeptide chain and a conserved 1,4-β-turn ("Met-turn") (Gomis-Rüth, 2003; Tallant et al, 2010a). The CTS starts after the above-mentioned glycine-mediated sharp turn, with the main chain leading to the third zinc ligand, which approaches the metal from below. The latter is immediately followed by a conserved "family-specific" residue and separated from the downstream-located "Met-turn" by "connecting segments" ranging from six to 77 amino acids. The "Met-turn" contains a strictly invariant methionine at position three, which forms a hydrophobic floor or pillow for the zinc-binding site, although there is no direct contact neither with the zinc nor with the substrate.

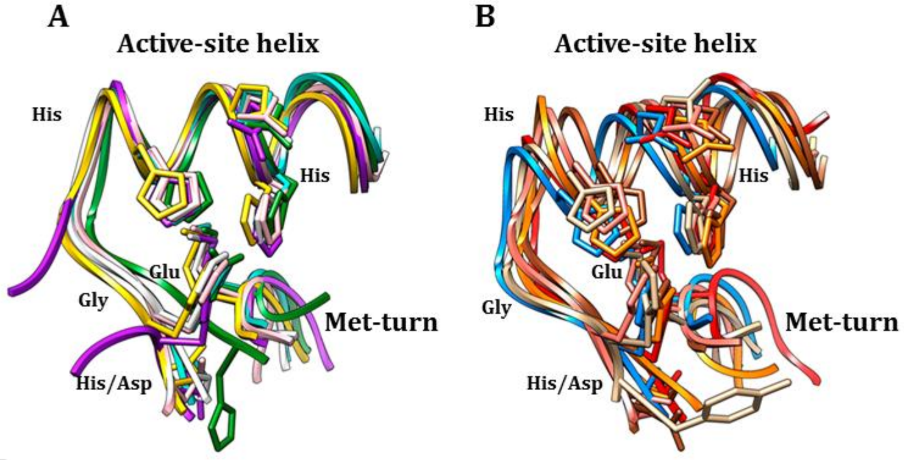


Fig. 11. Common structural elements in metzincin metallopeptidases. **(A)** Superposition of the active-site helix and the downstream chain segment until the family-specific residue *plus* the “Met-turn” of astacin (white), serralysin (pink), leishmanolysin (purple; an insertion is found between the second and third zinc-binding histidines), snapalysin (green), matrixin (turquoise), and adamalysin (gold) in two orthogonal views. The catalytic metal ions have been omitted for clarity. **(B)** Same as (A) but showing pappalysin (salmon), archaemetzincin (blue), igalysin (orange), toxilysin (sienna), fragilysin (red), and cholerilysin (tan). Figure adapted from (Cerdà-Costa & Gomis-Rüth, 2014).

3.1.4. Matrix metalloproteinases

MMPs, also called matrixins and vertebrate collagenases, constitute a separate family within the metzincin clan of MPs and are collected in family M10 of the MEROPS database (Bode et al, 1993; Gomis-Rüth, 2009; Murphy & Nagase, 2008; Stöcker & Bode, 1995; Tallant et al, 2010b; Visse & Nagase, 2003). They are secreted or membrane-bound peptidases and were initially identified as the active agent engaged in tail resorption during frog metamorphosis (Gross, 2004; Gross & Lapière, 1962; Lapière Ch, 2005). Although MMPs are mainly known for their role in broad-spectrum turnover of extracellular matrix (ECM) components, they also carry out selective limited proteolysis to activate or inactivate other proteins and enzymes, and are responsible for the shedding of membrane-anchored forms into circulation. Therefore, their activity must be subjected to exquisite spatial and temporal control; otherwise their degrading potential leads to uncontrolled proteolysis and tissue destruction, apoptosis and inflammation, and eventually to pathologies such as ulcer, rheumatoid and osteoarthritis, stroke, cardiovascular disease, fibrosis, Alzheimer, and cancer (Freije et al, 2003; Malemud, 2006; Nagase & Murphy, 2013; Nissinen & Kahari, 2014).

Active MMPs are naturally kept in check by α_2-macroglobulin (non-specifically) and by their endogenous, co-secreted or ECM-anchored 22-to-29-kDa TIMPs (Brew et al, 2000). Therapeutic strategies directed to combat dysregulation of MMPs include the design of drugs to target their CDs, for which detailed three-dimensional structural knowledge is essential.

Despite being widely distributed, these enzymes have been mostly studied in mammals, with structural information available for human, mouse, and pig CDs of MMPs see Table I in Cerdà-Costa & Gomis-Rüth, 2014). MMPs are distinguished by numbers (termed MMP-1 to MMP-28, with some numbers unassigned since they were later shown to be identical to other members). They might have likely evolved from a single-domain protein, which underwent successive rounds of duplication, gene fusion, and exon shuffling events to generate the multidomain architecture and functional diversity currently exhibited by MMPs (Fanjul-Fernández et al, 2010). These soluble or membrane-bound multidomain proteins usually include a 20-residue signal peptide (SP) for secretion, an ~80-residue PD, a ~160-residue zinc- and calcium-dependent CD, a linker region of variable length, and a ~200-residue hemopexin-like C-terminal domain for collagen binding, zymogen activation and dimerization (Nagase et al, 2006). In addition to mammals, MMPs have been studied (at least at the mRNA level) in several classes of animals and plants where their distribution is consistent with a Darwinian tree-based pathway. As to lower eukaryotes and prokaryotes, open reading frames with high sequence similarity to MMP CDs have been found in the genomes of fungi, viruses, archaea, and bacteria but have not been studied biochemically nor characterized at the structural level with the exception of the bacterial MMP karilysin (Cerdà-Costa et al, 2011; Karim et al, 2010).

3.1.5. Bacterial matrix metalloproteinases

Karilysin is an MP secreted by *Tannerella forsythia*, a pathogenic bacterium that inhabits the human gingival crevice together with *Porphyromonas gingivalis* and *Treponema denticola*. They are collectively referred to as “the red complex” and are the major responsible for periodontitis, the most infective chronic inflammatory disease known to mankind, with 10-15% of adults affected worldwide (Guevara et al, 2013; Petersen & Ogawa, 2005; Socransky et al, 1998). These pathogens secrete a vast armamentarium of proteolytic enzymes that has been thoroughly characterized for the latter two organisms (Eley & Cox, 2003). Nevertheless, very little is known about peptidases produced by *T. forsythia*, and as a matter of fact, of about two dozens of genes putatively encoding for secreted peptidases in its genome, only the cysteine peptidase PrtH, the trypsin-like serine peptidase BspA, and the MP karilysin have been hitherto characterized at the protein level (Guevara et al, 2013; Karim et al, 2010; Saito et al, 1997; Sharma, 2010).

Karilysin is a 472-residue protein comprising a 20-residue SP, a potential 14-residue PD (or propeptide owing to its short length), a 161-residue CD (18 kDa), and a 277-residue C-terminal domain (30 kDa) of unknown function (Fig. 12A). The recombinant full-length protein is expressed in *E. coli* without the SP and gives rise to the CD (hereafter Kly18) through sequential autolysis (Fig. 12B). The first cleavage at N^{34}-Y^{35} (numbering according to UniProt database code D0EM77) removes the N-terminal propeptide (Q^{21}-N^{34}) and generates a fully active 48-kDa variant (hereafter

Kly48). Subsequent cleavages downstream of the CD give rise to 38-kDa and 18-kDa active forms (Kly38 and Kly18, respectively).

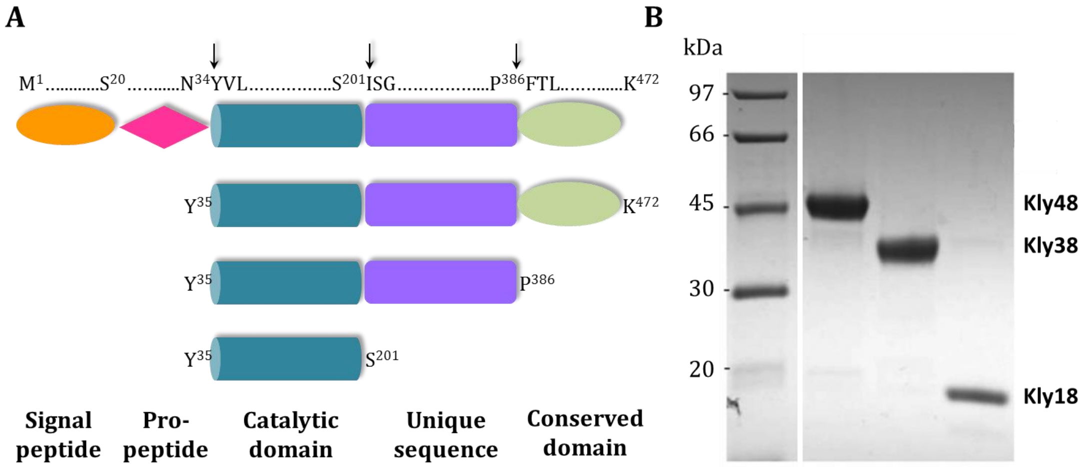


Fig. 12. Domain architecture and autolytic activation of karilysin. **(A)** Schematic representation of the domain structure of full-length karilysin (Kly) including autolytic cleavage sites indicated by arrows. **(B)** Purified forms of Kly48, Kly38, and Kly18 analyzed by sodium dodecylsulfate polyacrilamide gel electrophoresis (SDS-PAGE). Figure adapted from (Karim et al, 2010).

Analysis of the peptide bonds hydrolyzed during autolytic activation, together with extensive subsite mapping using a cellular library of peptide substrates, revealed that karilysin has a strong preference for medium-sized to bulky hydrophobic residues in S_1'; in particular for leucine, tyrosine and methionine (Karim et al, 2010). This is reminiscent of the specificity of mammalian MMPs, which show a similar preference for cleavage (Woessner & Nagase, 2000). Additional proteolytic assays evinced that karilysin efficiently degrades elastin, fibrinogen and fibronectin, supporting a role in periodontitis progression. Its role in the virulence of *T. forsythia* was further proved by its ability to inactivate a component of the immune system that targets bacteria and other microorganisms, the antimicrobial peptide LL-37 (Karim et al, 2010; Koziel et al, 2010). Moreover, the activity and thermal stability of the enzyme is considerably increased by calcium, and as expected for MPs, it is completely inhibited by EDTA and 1,10-phenantroline.

Karilysin shows the ZBCS characteristic of metzincins (Gomis-Rüth, 2003; Gomis-Rüth, 2009; Tallant et al, 2010a). Kly18 was crystallized under different conditions and shows a twisted five-stranded β-sheet (strands βI–βV), three α-helices, a magnesium ion, a structural zinc ion, and a catalytic zinc ion (Cerdà-Costa et al, 2011; Guevara et al, 2013). The catalytic zinc is tetrahedrically coordinated by H^{155}, H^{159}, H^{165} and a molecule of water. As described above for metzincins (see 3.1), the three-dimensional structure of Kly18 can be divided by the active-site cleft into a large NTS (Y^{35}-G^{162}) and a small CTS (I^{163}–S^{201}) (Fig. 13A,B). The NTS comprises a β-sheet whose strands are all parallel to a substrate bound to the cleft except for the outermost strand βIV, which forms the upper rim of the crevice and runs antiparallel. The concave site of the β-sheet accommodates the “backing helix” (αA) and runs across the back surface of the molecule from top right to bottom left. Additionally, the “active-site helix” (αB) runs horizontally across the center of the molecule, contributing to the back wall of the active-site cleft. After βIV, LβIVβV, and βV, the polypeptide enters a large loop LβVαB, which contributes to cleft delimitation and substrate binding. This loop opens out into the “active-site helix” αB, which includes the first half of the ZBCS, with E^{156} acting as the general-base/acid required for catalysis. After helix αB, the NTS ends at the glycine of the consensus sequence (G^{162}), and the polypeptide enters the CTS reaching for the third zinc binding residue. Downstream follows the “Met-turn”, which is stabilized by a double main-chain interaction with the side-chain of D^{188} within the “C-terminal helix” αC. Thereafter the chain enters the “specificity loop” (G^{179}–Q^{183}), which delimitates the back of the S$_1$′ subsite or specificity pocket of Kly18 (Fig. 13A,B). At N^{186}, the chain enters the “C-terminal helix”, which leads into the C-terminus on the molecular surface.

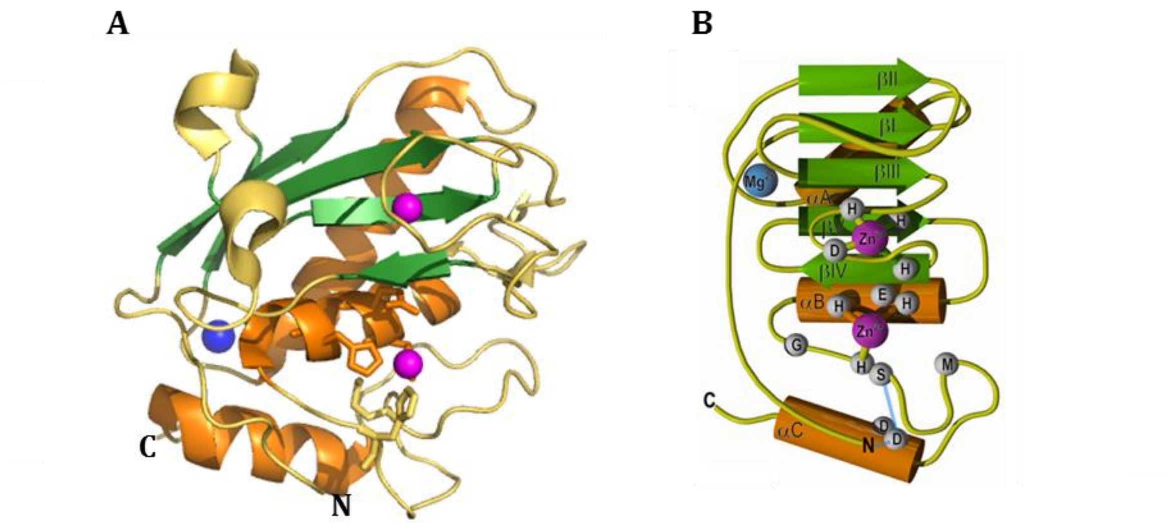


Fig. 13. Structure of the karilysin catalytic domain. **(A)** Richardson-type plot of the catalytic domain (CD) of karilysin (Kly18) (PDB code 2XS4). The regular secondary-structure elements are shown as green ribbons (β-strands βI to βV) and orange arrows (helices αA to αC). Residues participating in catalytic-zinc binding, as well as the general base/acid glutamate, and the “Met-turn” methionine are displayed as sticks. The catalytic and structural zinc ions as well as the magnesium ion are shown as magenta and blue spheres, respectively. **(B)** Topology diagram of Kly18 illustrating the regular secondary-structure elements and cations with the same color code as in (A). Hydrogen bonds and salt bridges are displayed as blue dashed lines. Figure panel adapted from (Cerdà-Costa et al, 2011).

Kly18 has comparable S$_1$’ subsite specificity with MMPs and structurally matches most of the characteristics of mammalian MMP CDs. Superposition of Kly18 with the main chain from several MMP CDs showed not only common regular secondary-structure elements but also common catalytic residues and “Met-turn” (Cerdà-Costa et al, 2011) (Fig. 14A). Moreover, similarity searches revealed that Kly18 is sequentially and evolutionarily closer to winged insects and vertebrate MMPs than to other bacterial counterparts, with sequence identities above 40%, despite belonging to a bacterium from the phylum *Bacteroidetes* (Fig. 14B).

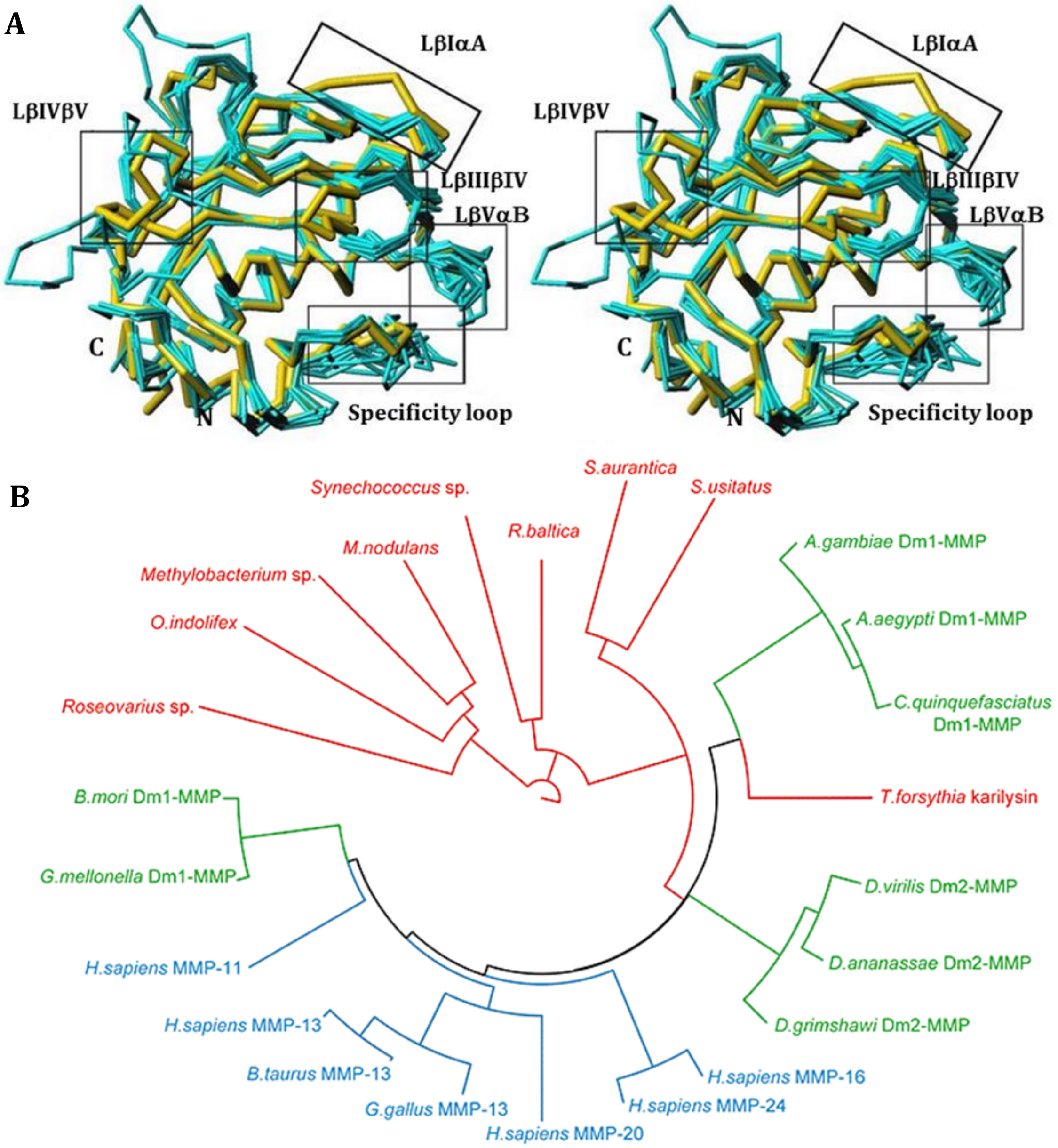


Fig. 14. Structural, evolutionary, and sequence relationships of the karilysin catalytic domain. **(A)** Superposition in stereo of the main chain of the catalytic domain (CD) of karilysin (Kly18; in yellow) and the CDs of MMP-1, -3, -7, -8, -9 (without fibronectin-type II insertions), -10, -11, -12, -13, and -16 (all in cyan). Kly18 segments showing significant deviations from the common chain trace are framed and labeled. **(B)** Phylogenetic tree of selected MMP CDs in relation with Kly18 sequence. For further details on the analysis, refer to Cerdà-Costa et al. 2011. Sequences and tree branches are colored red (bacteria), green (winged insects), and blue (vertebrates). Figure adapted from (Cerdà-Costa et al. 2011).

Accordingly, it was hypothesized that karilysin may be the result of horizontal gene transfer of an MMP gene co-opted from an external source (Cerdà-Costa et al, 2011). Mammals or mosquitoes feeding on human blood would be likely candidates due to the intimate coexistence between *T. forsythia* and the human blood-irrigated gingival crevice. Subsequently, Kly18 would have evolved in a bacterial environment, where it was furnished with unique flanking domains to give rise to full-length karilysin.

3.2. Activity regulation by zymogenicity

Molecular information about the mechanisms that control the activity of peptidases is essential to understand the role of a certain peptidase in a given biological or pathological process. However, structural information is still scarce for many of the regulatory systems. For instance, although most peptidases are biosynthesized as inactive precursors, among metzincins only zymogens from the MMP, astacin and fragilysin families have been structurally analyzed to date.

3.2.1. Zymogenicity in matrix metalloproteinases

With over 200 structures reported, MMPs constitute the most studied family of metzincins in structural terms. Nevertheless, only four structures have been determined with the presence of the PD: proMMP-1 (Protein Data Bank [PDB] code 1SU3; (Jozic et al, 2005), proMMP-2 (PDB code 1EAK); (Morgunova et al, 1999), proMMP-3 (PDB code 1SLM); (Becker et al, 1995), and proMMP-9 (PDB code 1L6J); (Elkins et al, 2002). Overall, these structures revealed that mammalian MMP zymogens have a pre-formed competent CD and an N-terminal globular PD which spans between 66 and 91 residues. In contrast to what has been observed in other zymogens, the MMP PD does not act as a chaperone for the *in vitro* refolding of the catalytic moiety (Milla et al, 2006). The typical MMP CD consists of a five-stranded β-sheet, three α-helices, and a connective loop (Fig. 15A). It contains two zinc ions (one catalytic and one structural) and up to three calcium ions which stabilize the structure. The PD is ellipsoidal and folded into a three-helix bundle with a left-handed twist that creates the scaffold to position a peptide in extended conformation to block the active-site cleft. This peptide starts in the outermost right edge of the cleft and inhibits the catalytic zinc through a cysteine imbedded in a conserved motif, PRCGXPD. This sequence is called the "cysteine-switch" or "velcro" sequence and runs through the active-site cleft in the opposite direction to a substrate bound to the cleft. The central cysteine in the "cysteine-switch" replaces the catalytic solvent molecule and coordinates the catalytic zinc ion (Fig. 15B). Upon activation of a proMMP, the cysteine-zinc interaction is disrupted, which enables the zinc ion to recruit a water molecule necessary for peptide hydrolysis (Springman et al, 1990; Van Wart & Birkedal-Hansen, 1990).

Activation of mammalian proMMPs is induced by cleavage in a so-called “bait region” by several peptidases such as trypsin, plasmin, and other MMPs. The “bait region” is located in a flexible region between helix 1 and 2 of the PD, and it has been seen at molecular level only in the three-dimensional structure of proMMP-2 where it is stabilized by a disulfide bond (Morgunova et al, 1999). Cleavage of the “bait region” removes a small part of the PD and destabilizes its arrangement, allowing further cleavages in a mechanism referred to as “stepwise activation”. Eventually, the final cleavage site X-F/Y becomes accessible for processing and next it takes place the dissociation of the cysteine-zinc interaction. Ultimately, the new competent N-terminus forms a salt bridge with a conserved aspartic acid of the final C-terminal helix (Nagase, 1997).

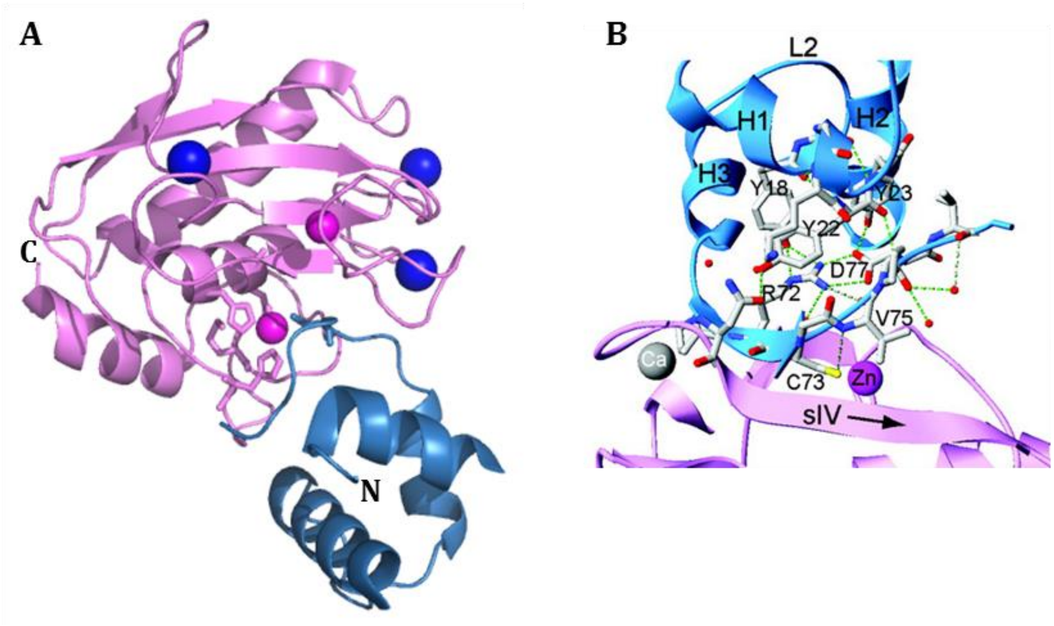


Fig. 15. Structure of the matrix metalloproteinase 1 zymogen. **(A)** Richardson-type plot of promatrix metalloproteinase-1 (proMMP-1) (PDB code 1SU3). The prodomain (PD) and the mature peptidase are depicted in blue and violet, respectively. The regular secondary-structure elements are shown as ribbons (helices α1-α6) and arrows (strands βI-βV). Residues participating in catalytic-zinc binding, as well as the general base/acid glutamate, and the “Met-turn” methionine are displayed as sticks. The catalytic and structural zinc ions as well as the calcium ions are shown as magenta and blue spheres, respectively. The cysteine from the PD engaged in the “cysteine-switch” is also shown as a stick. **(B)** Close-up view of (A) in a different orientation depicting the catalytic zinc ion and its interactions with residues from the PD. Figure panel adapted from (Jozic et al, 2005).

Disruption of the cysteine-zinc interaction may be also triggered by allosteric perturbation of the PD by chemicals such as chaotropic agents, sodium dodecyl sulfate, and SH-reactive agents, which can be used to activate proMMPs *in vitro* (Nagase, 1997; Ra & Parks, 2007). The “cysteine-switch” mechanism could be shared, with some variations, by other metzincin families for which conserved cysteines were observed upstream of the CD, such as leishmanolysins (motif HRCIHD; (Gomis-Rüth, 2003; Schlagenhauf et al, 1998)), pappalysins (motif CG; (Tallant et al, 2006)), and ADAMs (motif PKMCGV; (Milla et al, 2006)).

3.2.2. Zymogenicity in astacin and fragilysin

A distinct mechanism for latency maintenance in metzincin MPs was inferred from the three-dimensional structures of the zymogens of astacin and fragilysin, in which an aspartic acid residue coordinates the catalytic zinc, thus featuring an “aspartate-switch” (Goulas et al, 2011; Guevara et al, 2010). Proastacin from the European crayfish *Astacus astacus* shows the shortest structurally PD reported for an MP, which spans 34 residues and contains a conserved aspartic acid (D^{21}; numbering according to the mature enzyme, UniProt database code Q9U918). Astacin PD, in accordance to what has been described for proMMPs, does not have an intramolecular chaperone function, as revealed by refolding experiments with the recombinantly expressed mature protein (Reyda et al, 1999; Yiallouros et al, 2002). The PD runs across the front surface of the CD in the inverse direction of a substrate, blocking the active-site cleft (Guevara et al, 2010) (Fig. 16A). The interaction between the PD and the CD buries an interface of 1,580 $Å^2$ and is based on 75 close contacts (<4 Å), among them 20 hydrogen bonds, two organometallic interactions, and hydrophobic contacts between 27 pairs of residues.

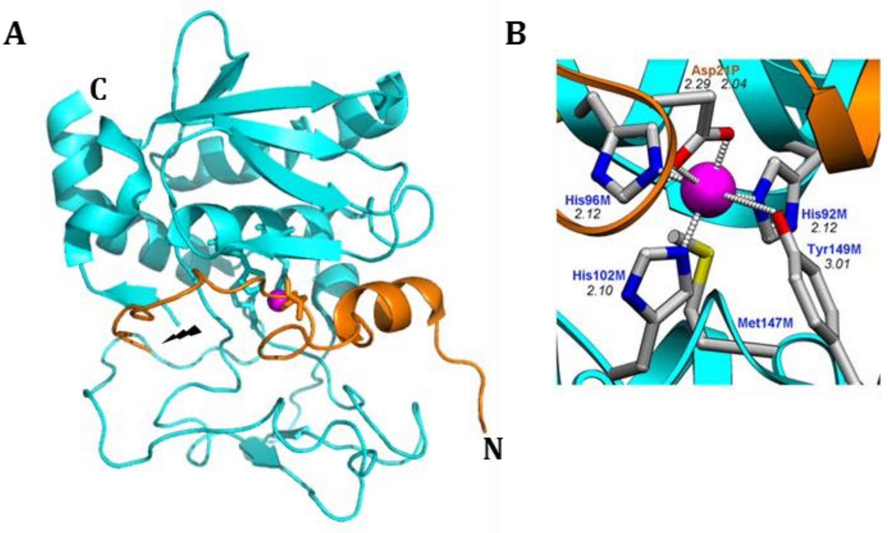


Fig. 16. Structure of the astacin zymogen. **(A)** Richardson-type plot of proastacin (PDB code 3LQ0). The prodomain (PD) and the mature peptidase are depicted in orange and cyan, respectively. The regular secondary-structure elements are shown as ribbons (helices α1-α4) and arrows (strands β1-β9). Residues participating in catalytic-zinc binding, as well as the general base/acid glutamate, and the “Met-turn” methionine are displayed as sticks model. The zinc ion is shown as a magenta sphere. The aspartic acid residue from the PD accounting for the “aspartate-switch” is also shown as a stick model. The activation cleavage site is also indicated. **(B)** Close-up view of (A) depicting the catalytic zinc ion with its protein ligands. The respective bonding distances (in Å) are shown, as it is the “Met-turn” methionine. Figure panel adapted from (Guevara et al, 2010).

The zinc-binding aspartic acid is provided by a wide loop immediately downstream of a PD helix that occupies the primed side of the cleft. The PD runs up along the molecular surface to the active-site cleft and from E^6 on, the chain runs through the active-site cleft approximately until R^{25}, adopting a helical conformation (helix α1) until Y^{12}. Since substrates usually bind to active-site clefts in extended conformation, the helical structure of the PD may contribute to the prevention of autolysis. At N^{14}, the chain protrudes outward from the protein moiety and enters a wide Z-shaped loop, directly in front of the zinc. This loop includes the conserved D^{21}, which coordinates the catalytic zinc ion from the top in a bidentate manner and replaces the zinc-bound solvent molecule in the mature enzyme, thus accounting for the "aspartate-switch" mechanism (Fig. 16B).

Proastacin activation *in vivo* is understood to be a two-step process involving successive cleavages by trypsin and active astacin until the mature N-terminus is generated (Yiallouros et al, 2002). The activation also entails conformational changes in mature astacin. In the first activation step, endogenous trypsin might perform cleavages on certain exposed regions of the PD, a process reminiscent of the "bait-region" cleavage in MMPs. In the second activation step, premature active astacin produces subsequent cleavages that eventually give rise to the new competent N-terminus at Ala^{1}.

In contrast to proastacin, fragilysin from the enteropathogenic bacterium *Bacteroides fragilis* shows the largest PD hitherto structurally reported for an MP (194 out of 397 residues in total; UniProt database code O86049). It inhibits the catalytic moiety through an "aspartate-switch" mechanism similar to that of the astacin family. In contrast to proMMPs and proastacin, the PD functions as a chaperone assisting in the folding and stabilization of the CD. The PD does not cap the CD but is attached to its right lateral surface and prevents the access to the active-site cleft through the C-terminal segment, which runs in extended conformation across the entire catalytic front in the opposite orientation to a substrate (Goulas et al, 2011; Tallant et al, 2010b) (Fig. 17A). The PD is formed by a large twisted antiparallel β-sheet, which vertically traverses the whole domain and has concave and convex sides. The interaction between the PD and the CD exhibits good shape complementarity and covers an area of 1,996 $Å^2$ at the protein interface, with 63 residues involved and 98 close contacts. The most important segment for latency is that encompassing α3, Lα3β11, and β11, which traverses the front of the CD from right to left. It establishes a parallel β-sheet interaction on the nonprimed side of the cleft with "upper-rim strand" β15 of the CD. It also approaches the beginning of helix α5 above the cleft, as well as Lβ14β15, Lβ15β16, Lβ16α6, "active-site helix" α6, and the segment connecting the "Met-turn" with "C-terminal helix" α7. A prominent bulge preceding β11 gives rise to a tight 1,4-turn spanning Y^{191}-D^{194}, whose side chains penetrate the catalytic moiety. The former residue occupies the primed side of the cleft and the latter coordinates the catalytic zinc ion in a bidentate manner in substitution of the solvent molecule usually found in mature CDs (Fig. 17B).

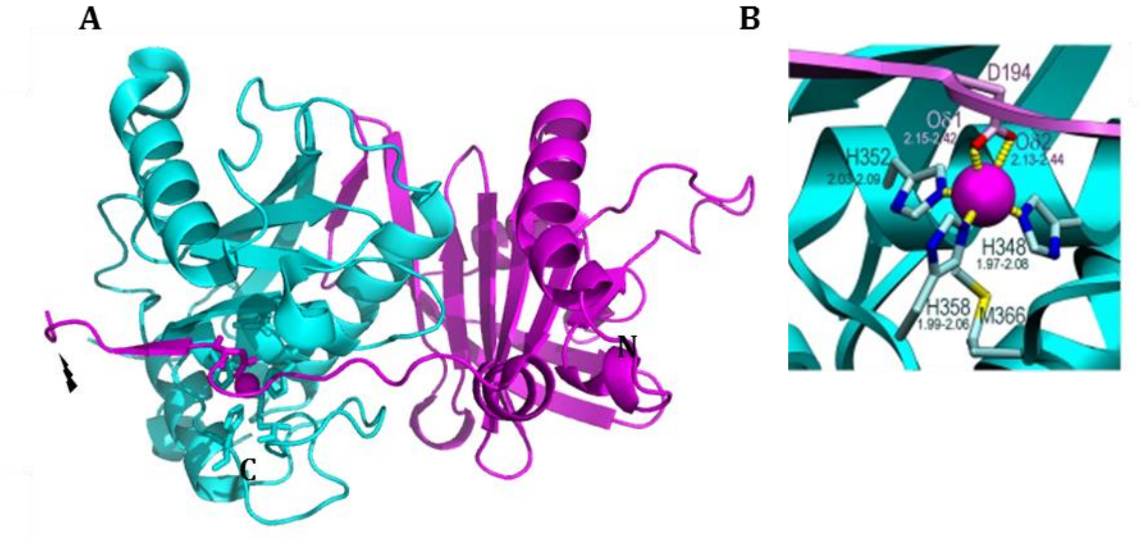


Fig. 17. Structure of the fragilysin zymogen. **(A)** Richardson-type plot of profragilysin (PDB code 3P24). The prodomain (PD) and the mature peptidase are depicted in magenta and cyan, respectively. The regular secondary-structure elements are shown as ribbons (helices α1–α7 and η1) and arrows (strands β1–β15). Residues participating in catalytic-zinc binding, as well as the general base/acid glutamate, and the "Met-turn" methionine are displayed as sticks. The zinc ion is shown as a magenta sphere. The aspartic acid residue from the PD accounting for the "aspartate-switch" is also shown as a stick. The activation cleavage site is indicated. **(B)** Close-up view of (A) depicting the catalytic zinc ion with its protein ligands. The respective bonding distances (in Å) are shown, as it is the "Met-turn" methionine. Figure panel adapted from (Goulas et al, 2011).

The activation of profragilysin is thought to be a heterolytic process most likely involving trypsin, which is widely expressed in stomach and small intestine. This hypothesis is further supported by the fact that the final activation cleavage site (R^{211}-A^{212}) matches trypsin's substrate specificity as evinced by limited proteolysis experiments that rendered the mature protein (Goulas et al, 2011).

4. GLUZINCIN METALLOPEPTIDASES

The gluzincin clan of MPs was named on the basis of a glutamic acid found as third metal-ligand residue in the ZBCS HEXXH+EXXZ(A,T,S,G) (one-letter residue code; zinc-binding residues underlined; X for any residue; Z for alanine, threonine, serine, or glycine) (Hooper, 1994) (see section 2.2). This glutamic acid is usually provided by a helix (or a short loop immediately preceding it) accordingly named "glutamate helix". In addition, the ZBCS of gluzincins shows a "Ser/Gly-turn" analogous to the "Met-turn" of metzincins, which provides the base for the metal-binding site (Gomis-Rüth, 2009).

4.1. Classification and main structural features

A large number of reported three-dimensional structures provides detailed molecular knowledge for nine of the current gluzincin families: bacterial collagenases (Eckhard et al, 2011); thermolysins (Matthews et al, 1972); cowrins (Gomis-Rüth, 2008); integral-membrane metallopeptidases (IMMPs) from MEROPS database families M48 and M56 (Pryor et al, 2013; Quigley et al, 2013); clostridial neurotoxins (Montecucco & Schiavo, 1993; Swaminathan, 2011); anthrax lethal factor and related proteins (Pannifer et al, 2001; Xu et al, 2012); leukotriene A4 hydrolase (LA4H), tricorn interacting factor F3, and MEROPS database M1-family APs (LA4H+M1-APs) (Kyrieleis et al, 2005; Tholander et al, 2008); dipeptidyl peptidase III and its structural relatives (Baral et al, 2008); and neprilysins (Oefner et al, 2000). In contrast to what happens in metzincins with the "Met-turn", the residue of the analogous "Ser/Gly-turn" varies among the different gluzincin families: while alanine is the most common residue, a threonine is found in clostridial neurotoxins and in some LA4H+M1-APs, a serine in thermolysins, and a serine or a glycine in cowrins.

4.1.1. Thermolysin, the prototype gluzincin metallopeptidase

Thermolysin from *B. thermoproteolyticus* is the founding member of gluzincins and the prototype of MEROPS database family M4 of peptidases, which also includes other MPs of similar sequence and domain structure, commonly known as thermolysin-like peptidases (TLPs). Since the structure of thermolysin was determined in 1972 (Matthews et al, 1972), its CD has been a reference for the analysis of subsequent structures from TLPs and other members of the gluzincin clan. The structure of the CD of thermolysin is bilobal, with α/β topology in the NTS and mainly helices in the CTS. The subdomains are separated by the active-site cleft. In thermolysin, the catalytic zinc ion is coordinated in an approximate tetrahedral geometry by two histidines within motif HELTH (residues 142-146; numbering according to the mature enzyme, UniProt database code P00800) that is common to the family members, E^{166} residue downstream in the chain, and a water molecule. Thermolysin also binds four calcium ions, Ca(1) and Ca(2) at a double binding-site near the active-site cleft, and Ca(3) and Ca(4) at exposed loops in the N-terminal and C-terminal lobes, respectively. Although the calcium ions do not have a catalytic role, they contribute to the enzyme's stability by protecting it from autolysis. Considering their high sequence similarity, most TLPs are expected to have similar three-dimensional structures; one major difference, however, seems to be the number of bound calcium ions. Accordingly, sequence comparisons indicate that the less stable TLPs lack binding sites for Ca(3) and Ca(4) (Van den Burg & Eijsink, 2013).

4.1.2. Integral-membrane gluzincin metallopeptidases

The gluzincin clan of MPs includes many members that are imbedded in membranes, anchored via multiple transmembrane regions, thus showing their active site either close to the membrane surface or within the membrane. They are known as integral membrane metallopeptidases (IMMPs) and their structural characterization is highly relevant; among many other specific aspects, to understand how certain folds favor the creation of suitable environments for the catalysis within lipidic membranes, their particular mode of action, and how certain mutations can affect their activity. Gluzincin IMMPs are mainly members of MEROPS database families M48 and M56, and they can be grouped together based on the common ZBCS HEXXH+EXXA+N+H (Fig. 8).

Obtaining crystals of membrane proteins with multiple transmembrane regions, in general, and IMMPs, in particular, is extremely challenging owing to low endogenous expression levels and technical difficulties related to their recombinant expression and purification. Therefore, limited structural information of this gluzincin family is currently available, with only two representative three-dimensional structures recently published: Ste24p from *Saccharomyces mikatae* and its human ortholog zinc MP STE24 (ZMSTE24), also known as farnesylated-protein converting enzyme 1 (FACE-1) (Pryor et al, 2013; Quigley et al, 2013). Both IMMPs belong to MEROPS database subfamily M48A and participate in the processing of prenylated substrates at a C-terminal motif known as CAAX, in which A is an aliphatic residue and the lipid is attached to the cysteine residue. Ste24p is essential for the complete maturation of the yeast mating pheromone a-factor (Fujimura-Kamada et al, 1997). It resides in the membrane of the endoplasmic reticulum with its N-terminus in the lumen and the active site and the C-terminus in the cytoplasm (Tam et al, 2001). The structure of Ste24p reveals a ring of seven transmembrane helices (TMHs) enclosing a voluminous cavity (~14,000 Å^3) containing the active site and substrate-binding groove (Pryor et al, 2013) (Fig. 18A). Other secondary-structure elements are found capping the structure on the luminal and cytosolic sides. On its luminal surface the structure is capped by two short α-helices, found in the loops between TMH II and III and between TMH VI and VII, together with other loops and the canted TMHs. The cytosolic part of the protein contains two additional domains, which form the substrate-binding groove: a mixed α helix–β strand-loop 5 domain and an α-helical

C-terminal domain blocking the large gaps between TMH V and VI and between TMH VII and I and the cytoplasmic surface of the structure (Fig. 18B).

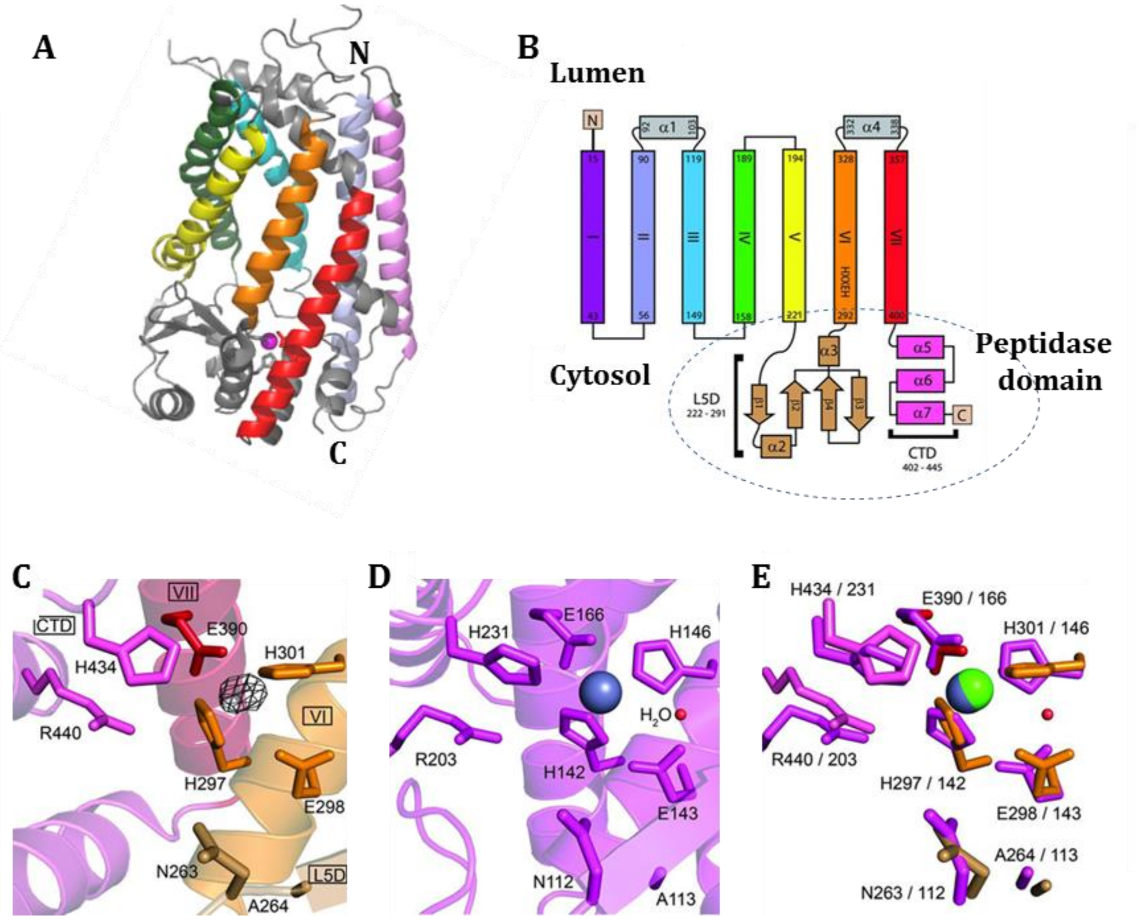


Fig. 18. Structure of Ste24p. **(A)** Richardson-type plot of Ste24p (PDB code 4IL3). The regular secondary-structure elements are shown as ribbons and arrows; the six transmembrane helices (TMHs) are colored. Residues participating in catalytic-zinc binding as well as the general base/acid glutamate are displayed as sticks. The zinc ion is shown as a magenta sphere. **(B)** Topology diagram of Ste24p illustrating the regular secondary-structure elements, with helices and strands represented by rectangles and arrows, respectively (same color code for TMHs as in [A]). The peptidase domain, which is located in the cytosol, is pinpointed within a blue circle. **(C)** Active site of Ste24p containing the HEXXH motif in helix VI (H^{297}-H^{301}), the catalytic residue (E^{390}), and substrate-coordinating residues N^{263}, A^{264}, H^{434}, and R^{440}. **(D)** Active site of thermolysin (PDB code 1LND) showing the HEXXH motif (H^{142}-H^{146}), the catalytic residue (E^{166}), and the substrate-coordinating residues N^{112}, A^{113}, R^{203}, and H^{231}. **(E)** Overlay of the Ste24p and thermolysin active sites evincing high similarity (root-mean-square deviation [rmsd] of overlaid residues: 0.5 Å). Figure panels (B,C,D,E) adapted from (Pryor et al, 2013).

The typical ZBCS HEXXH motif with the two histidine zinc ligands and the third-metal binding glutamic acid (E^{390}; residue numbering according to UniProt database code M4GGS2) are located in TMH VI and VII, respectively. Overall, the active site of Ste24p reveals striking structural resemblance of nearly all of the residues responsible for zinc coordination, catalysis, and substrate recognition with those of thermolysin (Pryor et al, 2013) (Fig. 18C,D,E).

FACE-1 is responsible for the maturation of prelamin A, the precursor of the nuclear intermediate filament protein, lamin A, which is important for the maintenance of the structure of the nucleus in multicellular organisms (Dechat et al, 2010; Dittmer & Misteli, 2011). FACE-1 dysfunction results in the accumulation of farnesylated unprocessed prelamin A in the nuclear envelope, thereby contributing to the appearance of lamin-processing diseases (laminopathies), such as the premature-aging disease progeria and metabolic disorders (Barrowman & Michaelis, 2009). Similarly to its yeast counterpart, FACE-1 is a transmembrane helical barrel enclosing a large water-filled chamber (12,000 $Å^3$), which crosses the membrane of the endoplasmatic reticulum (Quigley et al, 2013) (Fig. 19A). The chamber is sealed on the endoplasmic reticulum lumenal face of the membrane by three lumenal (L) α helices, LH1, LH2 and TMH7A (Fig. 19B). The nucleoplasmic/cytosolic side of the chamber is capped by the zinc MP domain, inserted between TMH5 and TMH6, with the catalytic site facing into the chamber. The zinc ion is coordinated by H^{335} and H^{339} from the HEXXH motif (residue numbering according to UniProt database code O75844) and E^{415} from TMH7 (Fig. 19C).

The MEROPS database family M48 encompasses other relevant gluzincin IMMPs, which have not been structurally characterized yet; i.e. *E. coli* HtpX and human mitochondrial Oma1, which represent prototypes of subfamilies M48B and M48C, respectively. HtpX participates in the proteolytic quality control of bacterial membrane proteins in conjunction with the ATP-dependent protein, FtsH (Akiyama, 2009; Sakoh et al, 2005). It is predicted to comprise four TMHs and a cytosolic catalytic moiety encompassing the HEXXH motif, but to date only a soluble non-catalytically competent fragment of an HtpX ortholog from *Vibrio parahemeolyticus* has been deposited in the PDB by a Structural Genomics Consortium (PDB code 3CQB; unpublished results) (Arolas et al, 2014). Oma1 is an ATP-independent IMMP involved in the maintenance of mitochondrial fusion processes. It participates in the constitutive proteolytic processing of the dynamin-related GTPase, optic atrophy 1 (Opa1). Under stress conditions, it also functions as a quality control peptidase converting Opa1 into short forms and inhibiting mitochondrial fusion. Disturbances in the dynamic balance between mithocondrial fusion and fission events cause various neurodegenerative diseases (Anand et al, 2014; Quirós et al, 2013). Oma1 is predicted to have two TMHs according to bioinformatics approaches but this topology and the localization of the active site still remain controversial (Arolas et al, 2014; Kaser et al, 2003).

Two other members within the gluzincin clan of MPs have been described as IMMPs: BlaR1 and MecR1 proteins from *Staphylococcus aureus*, which represent the archetypal models for MEROPS database family M56. These bacterial MPs are part of signal transduction systems that respond to the presence of β-lactam antibiotics by triggering the production of β-lactamases or penicillin-binding proteins (Sobhanifar et al, 2013; Zhang et al, 2001). BlaR1 and MecR1 are composed of an N-terminal IMMP domain facing the cytosol and a C-terminal extracellular sensor domain that binds environmental β-lactams (Mallorquí-Fernández et al, 2004). The binding of the antibiotic induces the activation of the MP moiety, which undergoes an autolytic cleavage that is followed by the cleavage of the repressor of the system, ultimately triggering production of lactamase. They are predicted to have four TMHs, with the HEXXH motif located between the third and fourth TMHs, and the third metal ligand predicted to be an aspartic acid located five residues C-terminal to the second zinc-binding histidine. The only structures reported to date correspond to the extracellular sensor domains of BlaR1 and MecR1 from *S. aureus*, while the MP domains still remain to be characterized (Marrero et al, 2006; Wilke et al, 2004).

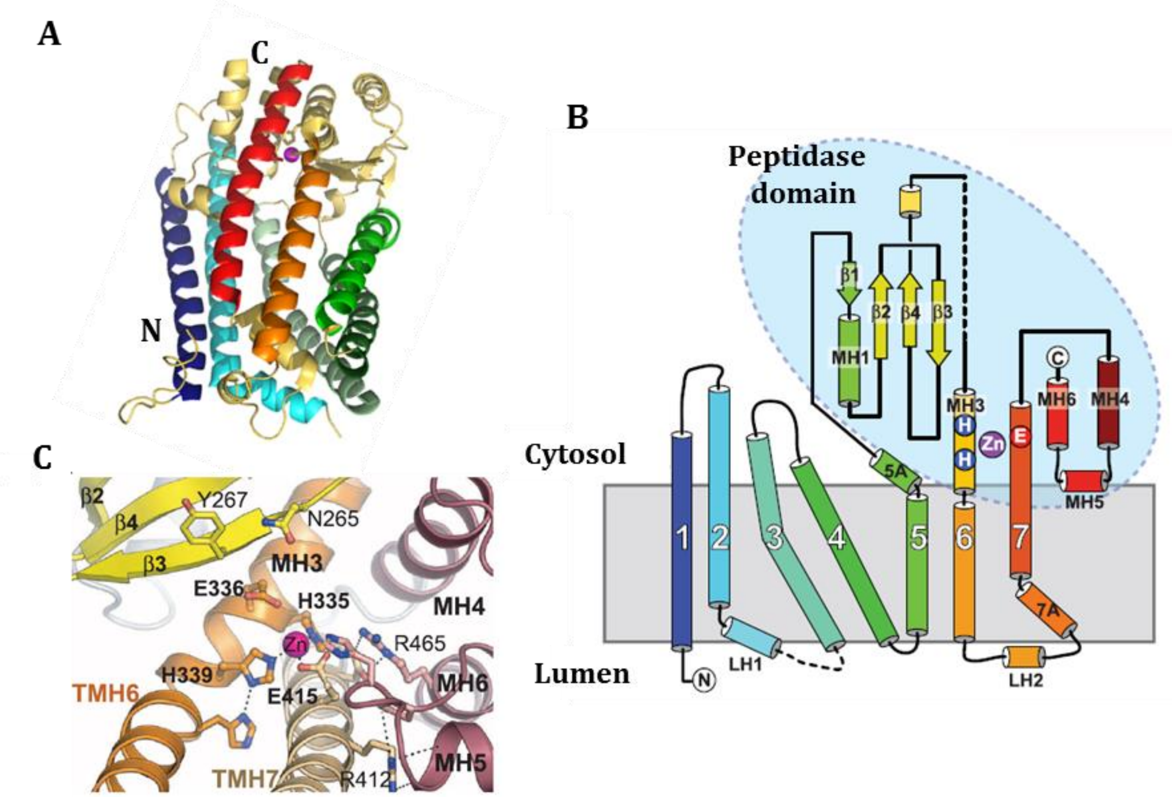


Fig. 19. Structure of FACE-1. **(A)** Richardson-type plot of farnesylated-protein converting enzyme 1 (FACE-1) (PDB code 2YPT). The regular secondary-structure elements are shown as ribbons and arrows; the six transmembrane helices (TMHs) are colored. Residues participating in catalytic-zinc binding as well as the general base/acid glutamate are displayed as sticks. The zinc ion is shown as a magenta sphere. **(B)** Topology diagram of FACE-1 illustrating the regular secondary-structure elements (same color code for TMHs as in [A]). The metallopeptidase domain, which is located in the cytosol, is highlighted in pale blue. **(C)** Detailed view of the active site showing as sticks the residues involved in zinc coordination, substrate binding, and catalysis. Figure panels (B,C) adapted from (Quigley et al, 2013).

The first three-dimensional structure of an IMMP to be determined was that of site-2 peptidase (S2P) from *Methanocaldococcus jannaschii* (MEROPS database family M50) (Feng et al, 2007). Despite not belonging to the gluzincin clan of MPs, S2P revealed for the first time common principles that govern peptide-bond catalysis in IMMPs, i.e. how hydrophilic water molecules enter the peptidase active site and how transmembrane substrates gain access to the active site. S2P cleaves the membrane-anchored sterol regulatory element-binding protein and releases a transcription factor which activates genes involved in synthesis and uptake of cholesterol and fatty acids (Zelenski et al, 1999). Human patients harboring reduction-of-function mutations in S2P exhibit an array of pathologies ranging from skin defects to neurological abnormalities (Rawson, 2013). The structure of S2P shows six TMHs (TMH1-6) and a zinc ion coordinated by three conserved residues in all SP2 proteins (H^{54} and H^{58} of TMH2, and D^{148} of TMH3; residue numbering according to UniProt database code Q57837) (Feng et al, 2007). The zinc ion of S2P is accessible to the cytoplasm by means of a narrow water-permeable channel found in the closed (to substrate) conformation of the enzyme. Three-dimensional structures of intramembrane-cleaving peptidases from other catalytic types have likewise been published. These include the rhomboid serine peptidase GlpG from *E. coli* (MEROPS database family S54) (Wang & Ha, 2007); the human aspartyl peptidase presenilin, which is the catalytic component of the γ-secretase complex and plays a central role in the pathogenesis of Alzheimer's disease (MEROPS database family A22) (Bai et al, 2015); and the related Signal Peptide Peptidase A from *E. coli*, responsible for cleaving the remnant SPs left behind in the membrane following Sec-dependent protein secretion (Kim et al, 2008).

4.2. Activity regulation by zymogenicity

In contrast to the numerous structures reported for CDs of gluzincin MPs, only a few structures have been determined for their zymogenic forms. Zymogenicity constitutes a major regulatory mechanism in soluble peptidases, where physical constraints as those provided by the membrane in IMMPs are absent. Within gluzincins, structural studies have been mainly centered on the characterization of TLP zymogens and their maturation mechanisms (Demidyuk et al, 2010a; Gao et al, 2010). TLPs are directly involved in several bacterial infections and a better understanding of their mechanism of action and latency will advance the development of therapeutics to treat the associated diseases.

While TLPs display CDs with largely similar sequences and structures, substantial differences exist in their N-terminal PDs. Accordingly, TLPs can be divided into two groups (Demidyuk et al, 2008). Representatives of the first group, whose prototype is thermolysin, show long PDs (~200 amino acids) and represent more than 70% of all TLPs. The rest of TLPs, whose representative is protealysin, belong to a second group that shows short PDs (~50 residues). Interestingly, the same species can produce enzymes with both types of PD, which may be an indication that PDs rather than CDs determine the biological function of M4 peptidases, by adapting the CDs of the peptidase to cell requirements. Thermolysin-type (or long) PDs contain two distinct regions of conservation: a fungalysin/thermolysin prodomain (FTP) and a PepSY domain (Demidyuk et al, 2008; Demidyuk et al, 2010a). While certain exceptions have been described (e.g. the neutral protease from *Bacillus* *stearothermophilus*), PDs can act as intramolecular chaperones, inhibit the cognate mature proteins *in trans*, and even influence their secretion (Chen & Inouye, 2008; Gao et al, 2010; Mansfeld et al, 2005; O'Donohue & Beaumont, 1996; Zhu et al, 1989). In particular, the contribution of the PD to the folding of the peptidase and its secretion has been related to a characteristic RY motif in the FTP domain (Braun et al, 2000; Gao et al, 2010). In addition, biochemical studies have suggested that long PDs might also mediate maturation and autoprocessing of TLPs (Braun et al, 2000; Marie-Claire et al, 1998; Nickerson et al, 2008).

The zymogen of vibriolysin TLP from *Pseudoalteromonas* sp. SM9913 (called MCP-02) was shown to undergo rapid self-cleavage between H^{204} and A^{205} (numbering according to UniProt database code A1DRD5), which results in an autoprocessed complex in which the PD becomes an inhibitor of the CD through the coordination of the central zinc via its C-terminal segment (Chang et al, 2007). The structure of an intermediate step of such autoprocessed complex provided for the first time insights into the maturation mechanism in TLPs with long PDs (Gao et al, 2010). In the MCP-02 autoprocessed complex, the 180-residue PD consists of two globular domains, FTP and PepSY, connected by a rigid hinge surrounding the CTS of the CD, which is already fully folded. The concave surfaces of both the FTP and PepSY domains are lined by four antiparallel β-strands, whereas their convex centers consist of two α-helices packed against the above β-sheets to form the hydrophobic core (Fig. 20A). The interface between the PD and the CD is rather large, with a buried surface area of 1,542 $Å^2$. Although most of the forces connecting them result from a complicated network of hydrogen bonds and salt bridges between the β-sheets of the FTP domain and the CTS of the CD, a second smaller interface is located between an extended loop of the PepSY domain and the NTS of the CD, with five hydrogen bonds formed between these domains. The C-terminus of the PD inserts into the catalytic cleft, and the C-terminal carboxylate group of H^{204} replaces the activated water molecule in the mature enzyme (Fig. 20B). The main interaction that holds the C-terminus of the PD within the catalytic cleft results from the short antiparallel β-sheets between Q^{203} and H^{204} in the PD, and W^{320} and F^{319} in the CD.

Biochemical studies suggest that thermolysin-like precursors with long propeptides are removed autocatalytically. In the case of thermolysin and metalloprotease from *Listeria monocytogenes*, the propeptide can be cleaved off only intramolecularly (Bitar et al, 2008; Marie-Claire et al, 1998) although intermolecular processing was found in other cases studied (Kawamoto et al, 1993; McIver et al, 1991; Toma et al, 1989). Autoprocessing in MCP-02 seems to be a combination of the activation of the active site and changes in the interaction between the PD and the CD. Following a selfcleavage of peptide bond H^{204}-A^{205}, a large conformational change from the zymogen to the autoprocessed complex occurs to release the N-terminus of the mature protein from the active site. A stepwise degradation of the PD follows and affects a part of the hydrophobic center in the FTP domain. As the FTP contributes to the majority of the interactions between the

PD and the CD, these cleavages result in the structural disassembly of the PD and ultimately to the release of the PD from the CD.

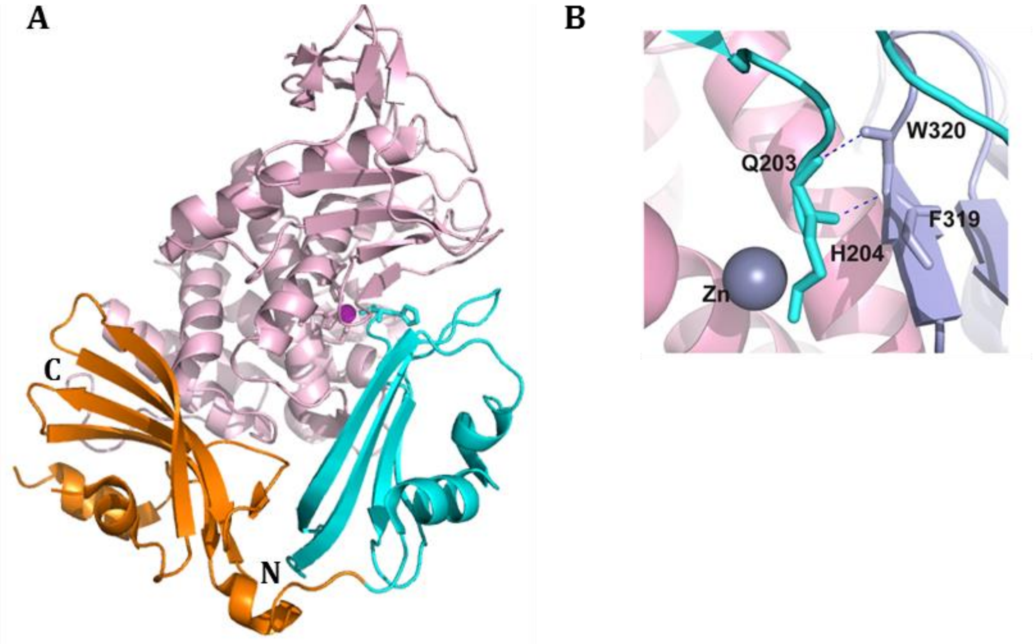


Fig. 20. Structure of the MCP-02 autoprocessed complex. **(A)** Richardson-type plot of proMCP-02 (PDB code 3NQY). The prodomain (fungalysin/thermolysin prodomain [FTP]+PepSY) and the mature peptidase are depicted in orange+cyan and light pink, respectively. The regular secondary-structure elements are shown as ribbons (helices) and arrows (strands). Residues participating in catalytic-zinc binding, as well as the general base/acid glutamate, and the "Ser/Gly-turn" serine are displayed as sticks. The zinc ion is shown as a magenta sphere. The histidine residue from the PD accounting for the inhibition is also shown as a stick. **(B)** Close-up view of (A) in a different orientation depicting the interaction of the catalytic zinc ion with its main ligands from the PD. Figure panel adapted from (Gao et al, 2010).

The first and only three-dimensional structure reported for a TLP zymogen is that of protealysin from *Serratia proteamaculans* (Demidyuk et al, 2010a). Here, a short PD forms an individual moiety whose N-terminal region occupies the active-site cleft and inhibits the enzyme. The C-terminal part of the PD, from residue 38 to residue 50 (numbering according to UniProt database code Q5MJ80), forms a flexible loop that connects with the CD (Fig. 21A). In the segment of the PD seen in the crystal structure (residues 6-37), two structural elements can be recognized: the N-terminal segment (residues 6-10) and a hairpin including two α-helices, α1 and α2 (residues 11-22 and 23-36, respectively). The key elements of the PD-CD interface in pro-protealysin include a highly conserved "PPL motif" in the PD and an "AYDD-hairpin" in the CD. The "PPL motif" (so-named because of two prolines and a leucine found in all protealysin-like enzymes) is a hydrophobic cluster of seven residues that directly interacts with the substrate-binding subsites of the active site and blocks access of the substrate (Fig. 21B). This element corresponds to residues 9 and 10 of the N-terminal segment and the N-terminal portion of α1 helix (residues 11-15). The chain continues with the C-terminal portion of helix α1, which hydrophobically interacts with the “AYDD hairpin” in the CTS of the mature moiety and with the NTS. The “AYDD hairpin” dams the active-site cleft on its subsites. Moreover, helix α2 of the PD contacts the surface of the NTS through three critical hydrogen bonds. With all these interactions, the PD fixes the N- and C-terminal domains of the pro-protealysin mature moiety in an open conformation of the active-site cleft that is unfavorable for catalysis. Therefore, the inhibitory effect in TLPs resulting from the presence of either long or short PDs causes both the loss of the water molecule and the blockage of the active-site cleft.

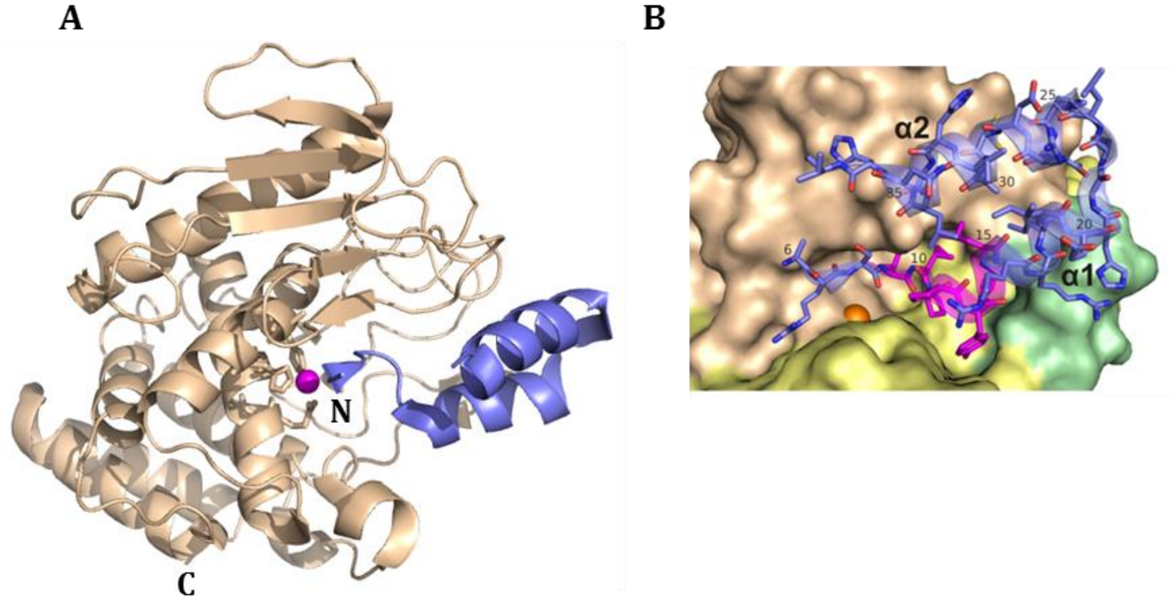


Fig. 21. Structure of the protealysin zymogen. **(A)** Richardson-type plot of pro-protealysin (PDB code 2VQX). The prodomain (PD) and the mature peptidase are depicted in slate blue and wheat, respectively. The regular secondary-structure elements are shown as ribbons (helices) and arrows (strands). Residues participating in catalytic-zinc binding, as well as the general base/acid glutamate, and the “Ser/Gly-turn” serine are displayed as sticks. The zinc ion is shown as a magenta sphere. Residues 38–51 of proPLN correspond to a flexible region that could not be identified in the electron density maps. **(B)** Surface representation depicting the interaction of the catalytic zinc ion (orange) with its main ligands from the PD (blue and magenta). Figure panel adapted from (Demidyuk et al, 2010b).

In contrast to thermolysin-like precursors, the maturation of protealysin-like peptidases has been scarcely studied. The *in vitro* inhibition of the maturation of protealysin by the Glu^{113}-to-Ala mutation points to autolytic processing (Gromova et al, 2009). The conversion of this protealysin mutant into the mature enzyme by incubation with the wild-type MP, in contrast, suggests an intermolecular process, which is structurally backed by the presence of cleavage sites for autolytic PD removal in the flexible loop at position 38-51 (Demidyuk et al, 2010a). Nevertheless, heterolytic PD removal is likely to occur in this loop as well, since it is exposed and therefore accessible to a wide range of endopeptidases.

OBJECTIVES

The study of the molecular mechanisms by which peptidases are kept inactive opens the way to the design of inhibitors in order to modulate their activity as part of novel therapeutic approaches. Accordingly, MMPs are major players in physiology and pathology and therefore highly valuable pharmaceutical targets. While the PD of a typical MMP is approximately 80 residues in length and contains the consensus "cysteine-switch" sequence PRCGXPD, the putative PD of karilysin spans only 14 residues (the shortest described for an MP) and lacks a cysteine. The **first project** of the present thesis (**Project 1**) aimed at understanding the (presumably different) zymogenic mechanism that controls the activation of this bacterial MMP using a combination of biochemical, biophysical, and structural techniques.

Membrane proteins are prime pharmaceutical targets but their study constitutes a major frontier in biochemistry due to the enormous technical hurdles associated. Accordingly, membrane proteins are dramatically under-represented in the structural database of the Protein Data Bank. In addition, recombinant expression trials of putative CDs of several membrane-located zincin MPs had failed in the past. The **second project** of the present thesis (**Project 2**) focused on the search of uncharacterized standalone hypothetic proteins with significant sequence similarity to CDs of members of MEROPS database families M48 and M56. These soluble scaffolds, termed "minigluzincins", were characterized using a combination of biochemical, biophysical, and structural techniques, in order to validate their suitability as structural models for further drug-design approaches.

The design of a minimal peptidase is an interesting topic both from a theoretical and biotechnological point of view, as it would simplify the production of peptidases for industrial applications. In addition, it is well known that in many MPs some domains or segments are not necessary for proteolytic activity. Furthermore, the results obtained from the second project defined a novel family of small-sized MPs, thus opening a window for their further study. Accordingly, the **third project** of the present thesis (**Project 3**) addressed the biochemical, biophysical, and structural characterization of the, to our knowledge, smallest active MP reported to date, which would represent a minimal unit for catalysis, "selecase".

RESULTS: PROJECT 1

"A Novel Mechanism of Latency in Matrix Metalloproteinases"

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 290, NO. 8, pp. 4728–4740, February 20, 2015

A Novel Mechanism of Latency in Matrix Metalloproteinases*

Received for publication, August 19, 2014, and in revised form, December 30, 2014 Published, JBC Papers in Press, January 2, 2015, DOI 10.1074/jbc.M114.605956

Mar López-Pelegrín[‡], **Miroslaw Ksiazek**[§], **Abdulkarim Y. Karim**[§1], **Tibisay Guevara**[‡], **Joan L. Arolas**[‡2], **Jan Potempa**[§¶3], **and F. Xavier Gomis-Rüth**[‡4]

From the [‡]Proteolysis Lab, Department of Structural Biology, Molecular Biology Institute of Barcelona, CSIC, Barcelona Science Park, c/Baldiri Reixac, 15-21, 08028 Barcelona, Catalonia, Spain, the [§]Department of Microbiology, Faculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian University, Ul. Gronostajowa 7, 30-387 Kraków, Poland, and the [¶]Oral Immunology and Infectious Disease, University of Louisville School of Dentistry, Louisville, Kentucky 40202

Background: Animal and plant matrix metalloproteinases (MMPs) are kept zymogenic through large prodomains and a cysteine-switch mechanism.

Results: Bacterial MMP karilysin has only a short N-terminal peptide upstream of the catalytic domain, which lacks cysteines.

Conclusion: This peptide inhibits through an aspartate-switch mechanism and also exerts other functions of authentic prodomains.

Significance: Karilysin is kept latent by a novel mechanism for MMPs.

The matrix metalloproteinases (MMPs) are a family of secreted soluble or membrane-anchored multimodular peptidases regularly found in several paralogous copies in animals and plants, where they have multiple functions. The minimal consensus domain architecture comprises a signal peptide, a 60–90-residue globular prodomain with a conserved sequence motif including a cysteine engaged in "cysteine-switch" or "Velcro" mediated latency, and a catalytic domain. Karilysin, from the human periodontopathogen *Tannerella forsythia*, is the only bacterial MMP to have been characterized biochemically to date. It shares with eukaryotic forms the catalytic domain but none of the flanking domains. Instead of the consensus MMP prodomain, it features a 14-residue propeptide, the shortest reported for a metallopeptidase, which lacks cysteines. Here we determined the structure of a prokarilysin fragment encompassing the propeptide and the catalytic domain, and found that the former runs across the cleft in the opposite direction to a bound substrate and inhibits the latter through an "aspartate-switch" mechanism. This finding is reminiscent of latency maintenance in the otherwise unrelated astacin and fragilysin metallopeptidase families. In addition, *in vivo* and biochemical assays showed that the propeptide contributes to protein folding and stability. Our analysis of prokarilysin reveals a novel mechanism of latency and activation in MMPs. Finally, our findings support the view that the karilysin catalytic domain was co-opted by competent bacteria through horizontal gene transfer from a eukaryotic source, and later evolved in a specific bacterial environment.

The matrix metalloproteinases (MMPs)[5] are a family of zinc- and calcium-dependent peptidases, which are grouped into the metzincin clan of metallopeptidases (MPs) together with other separate families such as the ADAMs/adamalysins, astacins, fragilysins, and serralysins (1–8). MMPs are found throughout animals and plants (9–12), where their distribution is consistent with a Darwinian tree-based pathway. In addition, polyplication has led to several paralogous MMP genes being present in the same organism: 24 in humans, 26 in sea urchin, 26 in zebrafish, seven in sea squirt, and two in fruit fly (11). In contrast, only a patchy phylogenetic distribution of genes encoding hypothetical orthologs has been found in viruses, *Bacteria*, *Archaea*, and fungi. Earlier studies of the relationship between mammalian MMPs and supposed prokaryotic orthologs included, as we now know, bacterial members of other metzincin families such as serralysins, fragilysins, and astacins (13–16). Accordingly, it was suggested that a primordial MMP may have arisen from an ancestor that is common to vertebrates, invertebrates, and plants but is not shared by earlier stages in evolution (11, 16–18). This entails that the hypothetical prokaryotic, viral, and fungal MMPs are incongruent with the tree of life or, more accurately, xenologs, *i.e.* the result of direct or indirect horizontal gene transfer from eukaryotic donors (9, 19, 20). This is reminiscent of the evolutionary origin postulated for fragilysin, which is the only molecular virulence factor described for enterotoxigenic *Bacteroides fragilis* and for which no similar proteins have been reported, not even from other *B. fragilis*

---

* This work was supported in part by European, United States American, Polish, Spanish, and Catalan Grants UMO-2012/04/A/NZ1/00051, UMO-2013/08/T/NZ1/00315, 2137/7.PR-EU/2011/2, 2975/7.PR/13/2014/2, DE09761, DE022597, FP7-HEALTH-2010-261460 "Gums&Joints," FP7-PEOPLE-2011-ITN-290246 "RAPID," FP7-HEALTH-2012-306029-2 "TRIGGER," BFU2012-32862, BIO2013-49320-EXP, CSD2006-00015, and 2014SGR9).

The atomic coordinates and structure factors (code 4R3V) have been deposited in the Protein Data Bank (http://wwpdb.org/).

[1] Present address: Dept. of Biology, College of Science, University of Salahaddin, Erbil, Kurdistan, Iraq.

[2] To whom correspondence may be addressed. Tel.: 34-934-020-187; Fax: 34-934-034-979; E-mail: jlacri@ibmb.csic.es.

[3] To whom correspondence may be addressed. Tel.: 502-852-5572; Fax: 502-852-5572; E-mail: jan.potempa@louisville.edu.

[4] To whom correspondence may be addressed: Tel.: 34-934-020-186; Fax: 34-934-034-979; E-mail: fxgr@ibmb.csic.es.

[5] The abbreviations used are: MMP, matrix metalloproteinase; MP, metallopeptidase; ADAM, a disintegrin and a metallopeptidase; HR3CP, human rhinovirus 3C proteinase; GST, glutathione *S*-transferase; Bistris propane, 1,3-bis-[tris(hydroxymethylmethylamino]propane; T_m, temperature of midtransition; Tricine, *N*-[2-hydroxy-1,1-bis(hydroxymethyl)ethyl]glycine; CSBZ, consensus sequence for zinc binding; NTS, N-terminal upper subdomain moiety; CD, catalytic domain; CTS, C-terminal lower subdomain; PDB, Protein Data Bank.

strains (21). Structural studies supported the view that the catalytic domain of this MP is the result of horizontal gene transfer of a member of the ADAM/adamalysin family, which has 38 orthologs in humans (8, 22–25), from a mammalian host to this bacterium, which thrives in the intestinal tract (26, 27).

Returning to MMPs, karilysin from the human periodontopathogen *Tannerella forsythia* is the only bacterial family member to have been analyzed biochemically to date (9, 28–33). In addition to karilysin, only MmpZ from *Bacillus anthracis* has been functionally assessed at the genetic level through knockout studies in *B. anthracis* cells, but it has not been isolated or characterized (34). Similarly to vertebrate MMPs, karilysin showed preference for medium-sized to bulky hydrophobic residues (leucine, tyrosine and methionine) in the specificity pocket, S_1' (Ref. 30; for active-site cleft subsite nomenclature, see Ref. 35). It inactivates antimicrobial peptide LL-37 and integrants of the complement system, including ficolin-2, ficolin-3, C4, and C5, by proteolysis and may thus contribute to evasion of the innate host immune response (29, 31). Karilysin is sequentially and evolutionarily closer to MMPs from winged insects that are transmission vectors of human diseases (47% sequence identity with Dm1 from *Aedes aegypti* and *Anopheles gambiae*; (9)) and mammals (44% identity with human MMP-11, -13, and -20 (9)) than to the few other bacterial sequences found in genomic sequences. Accordingly it was likewise suggested that it may be the result of horizontal gene transfer of an MMP gene from an animal to an intimate bacterial pathogen, which inhabits a biofilm on the tooth surface in humans (9).

The metzincins are characterized by a c̲onsensus s̲equence responsible for b̲inding of the catalytic z̲inc ion (CSBZ), H-E-*X*-*X*-H-*X*-*X*-(G/N)-*X*-*X*-(H/D) (amino acid one-letter code; *X* stands for any residue), and a conserved methionine-containing turn, the "Met-turn" (1–5, 36). In MMPs, the CSBZ encompasses three histidine zinc ligands, the general base/acid glutamate for catalysis, and a structurally relevant glycine (3). In addition, the distinct MMP paralogs are multidomain proteins that display a disparate domain organization that is the result of successive polyplication, gene fusion, and exon shuffling (11). The only domains common to all animal and plant MMPs are a signal peptide, which is removed after secretion, a prodomain and a catalytic domain, as found, *e.g.* in human MMP-7 and MMP-26, and in plant MMPs (12, 16, 18).

Most peptidases are biosynthesized as zymogens containing prosegments, which are required for latency maintenance to prevent unbridled activity but also sometimes to assist in proper folding of the usually downstream catalytic moieties (37–40). Metzincin exceptions lacking prosegments include the archaemetzincins, for which no hydrolytic activity has so far been reported, *i.e.* they might not need to be kept latent (41, 42); the toxilysin EcxA from *Escherichia coli*, whose soluble expression requires co-expression with its cognate EcxB subunit, thus pointing to a chaperone-like function for this ancillary subunit (43–45); the cholerilysin StcE from *E. coli*, for which an N-terminal immunoglobulin-like domain may assist the downstream catalytic moiety in proper folding (46); and igalysins, where an all-β-domain of similar topology to immunoglobulin-like domains is likewise found at the N terminus of the catalytic moiety (see Protein Data Bank (PDB) access codes 4DF9 and 3P1V and Ref. 5).

MMP prodomains (see Table 1 in Ref. 47) span 60–90 residues and include a conserved sequence motif, P-R-C-G-(V/N)-P-D, engaged in a "cysteine-switch" or "Velcro" mechanism of latency (10, 16, 48–51). It has been suggested that this mechanism may be shared by variants within other metzincin families, for which conserved cysteines were described upstream of the catalytic domain, such as the ADAMs/adamalysins (motif P-K-M-C-G-V (8, 52–54)), leishmanolysins (motif H-R-C-I-H-D (2)), and pappalysins (motif C-G (55)). In contrast, the 472 residues encoded by the karilysin gene (see UniProt sequence database access code D0EM77) only comprise a short 14-residue potential propeptide, which lacks cysteines, between the 20-residue signal peptide and the 161-residue mature catalytic moiety (Fig. 1*A*). A C-terminal domain of 277 residues of unknown function and sequence unrelated to any domain found in eukaryotic MMPs completes the protein. This strongly suggests a potentially different mechanism of latency maintenance, hitherto unseen not only in MMPs but also in metzincins in general, as the shortest prosegments described to date are those of members of the astacin family, which span >34 residues (7, 56–58).

We had previously determined the structure of the catalytic domain of karilysin (termed Kly18 (9)). To shed light on the molecular determinants of the first mechanism of latency maintenance of a bacterial MMP, in this work we assayed the possible function of the propeptide in folding, stability, and activity inhibition of Kly18. We further solved the x-ray crystal structure of an active-site mutant of a construct spanning the propeptide and Kly18 affecting the catalytic glutamate, pKly18-E156A, to circumvent autolysis. The mechanism derived was supported by site-directed mutagenesis and it is discussed in the context of general MMP latency maintenance.

EXPERIMENTAL PROCEDURES

Protein Production and Purification—The gene coding for full-length wild-type *T. forsythia* prokarilysin without the 20-residue signal peptide (hereafter pKly; 52 kDa; residues Gln^{21}-Lys^{472} according to UP D0EM77, see also Fig. 1*A*) was cloned at BamHI and XhoI restriction sites into vector pGEX-6P-1 (GE Healthcare) as described elsewhere (30). The resulting vector, pKAR1 (see Table 1 for an overview of vectors and constructs used), confers resistance toward ampicillin and attaches an N-terminal glutathione *S*-transferase (GST) moiety followed by a human rhinovirus 3C proteinase (HR3CP) recognition site (L-E-V-L-F-Q- ↓ -G̲-P̲; HR3CP cleavage leaves two extra residues, underlined, at the N terminus of the recombinant protein after digestion; three extra residues, L-G-S, are further present due to the cloning strategy). Single-residue point mutants pKly-Y35A and pKly-E156A (pKAR2 and pKAR3, respectively) were generated using the QuikChange Site-directed Mutagenesis Kit (Stratagene) according to the manufacturer's instructions as described (30). Double mutant pKly-D25A/Y35A (pKAR4) was similarly generated using pKAR2 as a template. Genes coding for the E156A-mutated catalytic domain of karilysin, with and without the propeptide

TABLE 1
Vectors and constructs

Name	Original vector	Antibiotic resistance[a]	Restriction sites	Insert	Fusion construct (N-terminal)	Additional N-terminal residues
pKAR1	pGEX-GP-1 (GE Healthcare)	*amp*	BamHI/XhoI	wt Gln^{21}-Lys^{472} (pKly)	GST + HR3CPr	G-P-L-G-S[b]
pKAR2	pGEX-GP-1 (GE Healthcare)	*amp*	BamHI/XhoI	Y35A mutant Gln^{21}-Lys^{472} (pKly-Y35A)	GST + HR3CPr	G-P-L-G-S[b]
pKAR3	pGEX-GP-1 (GE Healthcare)	*amp*	BamHI/XhoI	E^{156}A mutant Gln^{21}-Lys^{472} (pKly-E156A)	GST + HR3CPr	G-P-L-G-S[b]
pKAR4	pGEX-GP-1 (GE Healthcare)	*amp*	BamHI/XhoI	D25A/Y35A mutant Gln^{21}-Lys^{472} (pKly-D25A/Y35A)	GST + HR3CPr	G-P-L-G-S[b]
pKAR5	pGEX-GP-1 (GE Healthcare)	*amp*	BamHI/XhoI	E156A mutant Gln^{21}-Ser^{201} (pKly18-E156A)	GST + HR3CPr	G-P-[b]
pKAR6	pGEX-GP-1 (GE Healthcare)	*amp*	BamHI/XhoI	E156A mutant Tyr^{35}-Ser^{201} (Kly18-E156A)	GST + HR3CPr	G-P-[b]
pKAR7	pGEX-GP-1 (GE Healthcare)	*amp*	BamHI/XhoI	wt Gln^{21}-Ser^{201} (pKly18)	GST + HR3CPr	G-P-[b]
pKAR8	pGEX-GP-1 (GE Healthcare)	*amp*	BamHI/XhoI	Y35A mutant Gln^{21}-Ser^{201} (pKly18-Y35A)	GST + HR3CPr	G-P-[b]
pKAR9	pGEX-GP-1 (GE Healthcare)	*amp*	BamHI/XhoI	D25A/Y35A mutant Gln^{21}-Ser^{201} (pKly18-D25A/Y35A)	GST + HR3CPr	G-P-[b]
pKAR10	pCRI-7a (59)	*kan*	NcoI/XhoI	wt Gln^{21}-Ser^{201} (pKly18)	None	G-P-[b]
pKAR11	pCRI-7a (59)	*kan*	NcoI/XhoI	E156A mutant Gln^{21}-Ser^{201} (pKly18-E156A)	None	M-G-
pKAR12	pCRI-7a (59)	*kan*	NcoI/*XhoI*	wt Tyr^{35}-Ser^{201} (Kly18)	None	M-G-
pKAR13	pCRI-7a (59)	*kan*	NcoI/XhoI	E156A mutant Tyr^{35}-Ser^{201} (Kly18-E156A)	None	M-G-

[a] *amp*, ampicillin; GST, glutathione *S*-transferase; HR3CPr, recognition sequence for human rhinovirus 3C peptidase; *kan*, kanamycin; mut., mutant; wt, wild-type.
[b] After cleavage with HR3CP.

(hereafter pKly18-E156A and Kly18-E156A; 20 and 18 kDa; residues Gln^{21}-Ser^{201} and residues Tyr^{35}-Ser^{201}, respectively), were also cloned into vector pGEX-6P-1 (pKAR5 and pKAR6, respectively). Genes coding for pKly18 and its mutant proteins pKly18-Y35A and pKly18-D25A/Y35A were cloned into the same vector (pKAR7, pKAR8, and pKAR9, respectively) following a strategy previously described (59). Genes coding for pKly18, pKly18-E156A, Kly18, and Kly18-E156A were, furthermore, cloned at NcoI and XhoI restriction sites into vector pCRI-7a (59), which confers resistance toward kanamycin and does not attach fusion proteins (pKAR10-pKAR13, respectively). In these cases, the cloning strategy entailed that residues M-G were attached at the N terminus. All constructs were verified by DNA sequencing.

Proteins encoded by vectors pKAR1-pKAR9 were produced by heterologous overexpression in *E. coli* BL21(DE3) cells, which were grown at 37 °C in Luria-Bertani medium supplemented with 100 μg/ml of ampicillin. Cultures were induced at an A_{600} of 0.8 with 0.2 mM isopropyl β-D-thiogalactopyranoside and incubated overnight at 18 °C. Purification of wild-type and mutant pKly, and subsequent autolysis of the former to obtain Kly18, was achieved as described elsewhere (30). In turn, pKly18-E156A, Kly18-E156A, pKly18-Y35A, and pKly18-D25A/Y35A were purified as follows. After centrifugation at 7,000 × *g* for 30 min at 4 °C, the pellet was washed twice in 1× PBS, and resuspended in the same buffer supplemented with EDTA-free protease inhibitor mixture tablets and DNase I (both Roche Diagnostics). Cells were lysed using a cell disrupter (Constant Systems, Ltd.) at 1.35 Kbar, and the cell debris was removed by centrifugation at 40,000 × *g* for 1 h at 4 °C. The supernatant was filtered (0.22 μm pore size; Millipore), and incubated with glutathione-Sepharose 4B resin (GE Healthcare). The sample was washed first in 1× PBS and then in buffer A (50 mM Tris-HCl, 150 mM NaCl, pH 7.5), and eluted by incubation and cleavage with HR3CP at a 1:20 enzyme:substrate (w/w) ratio for 48 h at 4 °C. The protein was concentrated by ultrafiltration, and finally purified by size-exclusion chromatography on 16/600 or 10/300 Superdex 75 columns (GE Healthcare) previously equilibrated with buffer B (20 mM Tris-HCl, pH 8.0) or buffer C (20 mM Tris-HCl, 150 mM NaCl, pH 7.5).

Proteins encoded by vectors pKAR10–pKAR13 were produced in *E. coli* BL21(DE3) cells, which were grown at 37 °C in Luria-Bertani medium supplemented with 30 μg/ml of kanamycin. Cultures were induced at an A_{600} of 0.8 with 0.2–1 mM isopropyl β-D-thiogalactopyranoside and incubated either for 5 h at 37 °C or overnight at 18 °C. Cells were harvested by centrifugation at 7,000 × *g* for 30 min at 4 °C, washed in buffer A, resuspended in the same buffer, and further lysed in an ice-bath using a digital sonifier (Branson). After centrifugation at 15,000 × *g* for 30 min at 4 °C, both cell debris and supernatant were analyzed by 15% Tricine-SDS-PAGE stained with Coomassie Blue.

Protein identity and purity were assessed by mass spectrometry using an Autoflex Bruker apparatus and N-terminal sequencing through Edman degradation at the Proteomics Facility of Centro de Investigaciones Biológicas (Madrid, Spain). Ultrafiltration steps were performed with Vivaspin 15 and Vivaspin 4 filter devices of 5-kDa cut-off (Sartorius Stedim Biotech). Approximate protein concentration was determined by measuring A_{280} in a spectrophotometer (NanoDrop) using the calculated absorption coefficients $E_{0.1\%} = 2.32$ and 2.42 for pKly18-E156A and Kly18-E156A, respectively.

Autolytic Activation and Propeptide Inhibitory Activity Assays—Mutants pKly-Y35A (from pKAR2), pKly-D25A/Y35A (pKAR4), and pKly18-Y35A (pKAR8) were incubated in buffer B at 37 °C and at 0.4 mg/ml final protein concentration for up to 120 h to assay autolysis. Reactions were stopped at specific time points by boiling aliquots in reducing/denaturing buffer, and samples were further analyzed by 10% or 15% Tricine-SDS-PAGE stained with Coomassie Blue. Kly18, obtained by autolysis from pKAR1-encoded protein, was incubated at 0.025 μg/ml of final protein concentration for 30 min with 0.1–10 mM peptide Q-R-L-Y-D-N-G-P-L-T (purchased from GL Biochem Ltd.), which mimics the propeptide sequence. Proteolytic activity was subsequently measured at 37 °C in buffer C on substrate Mca-R-P-K-P-V-E-Nva-W-R-K(dnp)-NH_2 (Bachem; at 10 μM) in a microplate fluorimeter (Infinite M200, Tecan).

Thermal Shift Assays—Aliquots were prepared by mixing 7.5 μl of ×300 Sypro Orange dye (Molecular Probes) and 42.5 μl of either pKly18-E156A (from pKAR5) or Kly18-E156A (pKAR6) at 1–2 mg/ml in buffer C in the absence and presence of 1–5 mM

$CaCl_2$. Four replicates of each aliquot were analyzed in an iQ5 Multi-Color Real Time PCR Detection System (Bio-Rad) in 96-well PCR plates sealed with optical tape. Samples were heated from 30 to 95 °C at 0.5 °C/min, and the change in absorbance ($\lambda_{ex} = 490$ nm; $\lambda_{em} = 575$ nm) was monitored over time. The temperature of midtransition (T_m (60)) was determined for both proteins from the inflection point of each curve using iQ5 software.

Crystallization and Data Collection—Crystallization assays of pKAR5-encoded pKly18-E156A protein were carried out at the IBMB/IRB Crystallography Platform by the sitting-drop vapor diffusion method using 96 × 2-well MRC plates (Innovadyne). A TECAN Freedom EVO robot was used to prepare reservoir solutions, and a Phoenix/RE (Art Robbins) robot or a Cartesian Microsys 4000 XL (Genomic Solutions) was used for nanodrop dispensing. Crystallization plates were stored in Bruker steady-temperature crystal farms at 4 and 20 °C. Successful conditions were scaled up to the microliter range in 24-well Cryschem crystallization dishes (Hampton Research). The best crystals were obtained at 20 °C from drops containing protein solution (3.75 mg/ml in buffer B) and 100 mM Bistris propane, 200 mM potassium thiocyanate, 20% (w/v) polyethylene glycol 3350 (pH 7.5) as reservoir solution from 2:1-μl drops. Crystals were cryo-protected with 20% (v/v) glycerol. Diffraction datasets were collected at 100 K from liquid-N_2 flash cryo-cooled crystals (Oxford Cryosystems 700 series cryostream) on a Pilatus 6M pixel detector (from Dectris) at beam lines ID23-1 and ID29 of the European Synchrotron Radiation Facility (ESRF, Grenoble, France) within the Block Allocation Group "BAG Barcelona." Crystals contained two molecules per asymmetric unit. Diffraction data were integrated, scaled, merged, and reduced with programs XDS (61) and XSCALE (62) (see Table 2 for data processing statistics of the best dataset).

Structure Solution and Refinement—The structure of pKly18-E156A was solved by likelihood-scoring molecular replacement with the program PHASER (63) using the coordinates of the protein part only of mature wild-type Kly18 (PDB access code 4IN9 (64, 65)) as searching model. Two solutions were found at final Eulerian angles (α, β, γ, in °) of 285.8, 56.7, 97.2, and 76.9, 91.3, 284.2; and fractional cell coordinates (*x*, *y*, *z*) of 0.120, −0.017, 0.100, and 0.997, 0.332, and 0.608, respectively. These solutions gave initial *Z*-scores of 8.5 and 8.3 for the rotation functions, and 6.6 and 6.2 for the translation functions, respectively, as well as a final log-likelihood gain of 1,120. A subsequent density improvement step with ARP/wARP (66) rendered an electron density map that enabled straightforward chain tracing. Thereafter, manual model building with COOT (67, 68) alternated with crystallographic refinement with programs PHENIX (69) and BUSTER/TNT (70, 71), which included TLS refinement and automatically determined non-crystallographic restraints, until completion of the model. Both final model chains A and B contained residues Gln^{21}-Pro^{199}, as well as two zincs and one calcium each. Segment Val^{36}-Gly^{39} of chain A was continuous in the final Fourier map but ambiguous as to the position of the side chains. In addition, segments Gln^{38}-Ser^{40} and Ser^{54}-His^{57} of chain B were traced based on weak electron density to preserve chain continuity. Pro^{122} and Pro^{123} were linked by a *cis*-peptide bond. Three glycerol molecules and 226 solvent molecules completed the structure (see Table 2).

TABLE 2
Crystallographic data

Data	
Space group/cell constants (*a*, *b*, *c* in Å and β in °)	$P2_1$/36.18, 121.69, 41.88; 105.26
Wavelength (Å)	0.97242
No. of measurements/unique reflections	104,406/22,975
Resolution range (Å) (outermost shell)	60.8–2.01 (2.06–2.01)
Completeness (%)	99.8 (96.2)
R_{merge}[a]	0.105 (0.534)
$R_{r.i.m.}$ (=R_{meas})[a] /CC(1/2)[b]	0.116 (0.616)/99.6 (84.3)
Average intensity	10.3 (2.5)
B-factor (Wilson) ($Å^2$)/average multiplicity	29.0/4.5 (4.0)
Resolution range used for refinement (Å)	∞–2.01
No. of reflections used (test set)	22,047 (741)
Crystallographic R_{factor} (free R_{factor})[c]	0.165 (0.194)
No. of protein atoms[d]/solvent molecules/ligands/ions	2,843/226/3 glycerols/4 zinc, 2 calcium
Root mean square deviation target values bonds (Å)/angles (°)	0.010/0.98
Overall average *B*-factor ($Å^2$)	30.8
Model validation[e]	
Residues in favored regions/outliers/all residues	345 (97.5%)/0/355
Residues with bad bonds/bad angles/poor rotamers/Cβ deviation >0.25 Å	0/0/7(2.4%)/0

Values in parentheses refer to the outermost resolution shell.

[a] For details, see Refs. 110 and 111.

[b] According to Karplus and Diederichs (78).

[c] Crystallographic $R_{factor} = \Sigma_{hkl} \| F_{obs}| - k |F_{calc}\| / \Sigma_{hkl} |F_{obs}|$, where *k* is a scaling factor, and F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively. This factor is calculated for the working-set reflections; free R_{factor}, same for a test-set of reflections (>500) not used during refinement.

[d] Including atoms with double occupancy.

[e] According to MOLPROBITY (76).

Miscellaneous—Figures were prepared with CHIMERA (72). Structural superpositions were performed with SSM (73) within COOT, and with LSQKAB (74) and ROTMAT within the CCP4 suite of programs (75). Model validation was performed with MOLPROBITY (76). The interaction surface buried at the interface between the propeptide and the mature enzyme moiety was calculated with CNS version 1.3 (77). The final coordinates of pKly18-E156A are deposited with the PDB with code 4R3V.

RESULTS AND DISCUSSION

Roles of the Propeptide in Vitro and in Cellula—Wild-type karilysin is secreted as a zymogen with a 14-residue N-terminal propeptide (^{21}Q-R-L-Y-D-N-G-P-L-T-G-D-N-N^{34}), which is cleaved off at position Asn^{34}-Tyr^{35} during maturation (Fig. 1*A*). This is the primary activation cleavage and it releases an active 48-kDa form (Kly48 (30)). In recombinant protein production, subsequent cleavages within the C-terminal domain give rise to Kly38 and, finally, to a stable form of 18 kDa (Kly18), which corresponds to the isolated mature catalytic domain (CD) (5, 28, 30, 33). These cleavages were shown to be autolytic as activation was repressed by general chelating MP inhibitors and in the inactive active-site variant, E156A, which ablated the catalytic glutamate of the CSBZ (1, 5, 30, 79, 80). In addition, cleavage-site mutant Y35A, which does not match the specificity of the enzyme, was activated only slowly when compared with the wild-type (30, 33).

To assess whether the propeptide had a chaperone-like function on the downstream catalytic moiety, we cloned the genes encoding pKly18-E156A and Kly18-E156A in a vector that does not attach a fusion protein at the N terminus that would assist

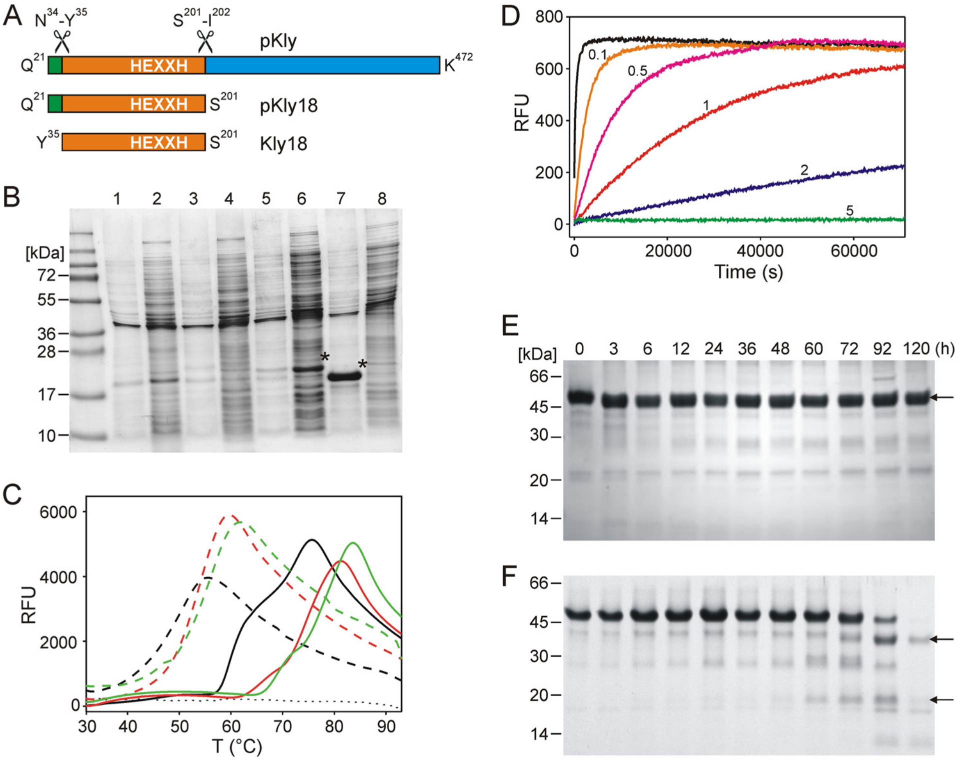


FIGURE 1. **Effect of the Kly18 propeptide *in vitro* and *in vivo*.** *A*, scheme depicting the domain structure of *T. forsythia* karilysin. Numbering according to UniProt D0EM77. *B*, SDS-PAGE of cultures of wild-type and E156A variants of pKly18 and Kly18. *Lanes 1* and *2*, insoluble and soluble fractions of wild-type pKly18 (from pKAR10), respectively. *Lanes 3* and *4*, insoluble and soluble fractions of wild-type Kly18 (pKAR12), respectively. *Lanes 5* and *6*, insoluble and soluble fractions of pKly18-E156A (pKAR11), respectively. *Lanes 7* and *8*, insoluble and soluble fractions of Kly18-E156A (pKAR13), respectively. Overexpressed proteins are labeled with an *asterisk*. *C*, unfolding transition curves showing temperature-dependent change in fluorescence of pKly18-E156A (pKAR5; *solid line*) and Kly18-E156A (pKAR6; *dashed line*) in the absence (*black curve*) and presence of $CaCl_2$ (1 mM, *red curve*; 5 mM, *green curve*). The *blank curve* is indicated with a *dotted line*. *D*, proteolytic activity of Kly18 (pKAR1) at 37 °C using substrate Mca-R-P-K-P-V-E-Nva-W-R-K(dnp)-NH_2 at 10 μM in the absence (0) and presence of 0.1, 0.5, 1, 2, and 5 mM propeptide mimic. *E* and *F*, stability of mutant pKly-Y35A (*E*) and mutant pKly-Y35A/D25A (*F*) over time. Kly48, Kly38, and Kly18 are indicated by *arrows*.

in proper folding (pKAR11 and pKAR13, respectively; see "Experimental Procedures"). We found that the active-site mutant pKly18-E156A was successfully overexpressed in soluble form (Fig. 1*B*). In contrast to the zymogen, Kly18-E156A was produced only in insoluble form (Fig. 1*B*). Moreover, when expressed from the pKAR6 vector, which attaches an N-terminal glutathione *S*-transferase fusion protein (see Table 1), Kly18-E156A was obtained with ~10 times lower yield than the proprotein (vector pKAR5). We conclude that the propeptide plays a major role in proper folding of Kly18 as previously described for other MPs such as fragilysin (26, 27), funnelin metallocarboxypeptidases (79, 81, 82), and ADAMs/adamalysins (54) but not for mammalian MMPs (83).

We further examined the effect of the propeptide in response to denaturation by a thermal shift assay following the thermofluor approach (60). Purified pKly18-E156A (pKAR5) showed two unfolding transitions compatible with unfolding of propeptide and CD, with T_m values of 60 ± 0.5 and 74 ± 0.5 °C (Fig. 1*C*). In contrast, the unfolding of purified Kly18-E156A (pKAR6) showed a single transition, with a T_m of 49 ± 2.2 °C. The addition of a physiological concentration of calcium resulted in a substantial increase in stability of both pKly18-E156A and Kly18-E156A. Accordingly, the former showed T_m values of 67.5 ± 1.7 and 76.5± 1.2 °C, and 69.5 ± 1.7 and 79 ± 2.2 °C, in the presence of 1 and 5 mM $CaCl_2$, respectively, whereas those of Kly18-E156A were 52.5 ± 1.2 and 54.5 ± 1 °C. This result is in agreement with the important role of calcium in Kly18 activity, as addition of 2–5 mM $CaCl_2$ is reported to enhance activity about three times (30). Thus, regardless of calcium, the 14-residue propeptide redounded to a dramatic increase in T_m, underpinning that it plays a major role in the thermal stability of the zymogen. Finally, we assayed the effect of a decapeptide spanning propeptide sequence Gln^{21}-Thr^{30} on the activity of purified mature Kly18 (from pKAR1) in the presence of a fluorogenic peptide substrate (Fig. 1*D*). We observed a weak but consistently concentration-dependent inhibitory effect as previously shown for other MPs when their propeptides or prodomains were added in *trans*, among others funnelins (79, 81), ADAMs/adamalysins (84), and mammalian MMPs (85–87). Summarizing, the propeptide of karilysin is the shortest currently described to date for an MP, and it exerts all roles, which collectively or selectively had been previously described for peptidase propeptides or prodomains: latency maintenance, folding assistance during biosynthesis, stability to thermal denaturation, and inhibition of peptidolytic activity (38, 39, 81).

Overall Structure of pKly18—Due to rapid autolytic processing of recombinant wild-type prokarilysin (30), crystals of pKly18 could only be obtained for an inactive variant affecting the catalytic glutamate (pKly18-E156A), as already reported for other MP zymogens (88–92). This protein crystallized as monoclinic crystals diffracting to 2-Å resolution with two molecules per asymmetric unit. These were essentially identical (Cα-atom root mean square deviation = 0.53 Å) except for segments Asn^{53}-His^{57} and Asn^{34}-Gly^{39}. The latter flanks the primary activation cleavage point and is flexible. It is stabilized through an interaction with segment Asn^{87}-Asn^{89} of the second molecule present in the asymmetric unit of the crystal although in different conformations in molecules A and B, so the discussion hereafter is centered on molecule A if not otherwise stated. When two values are indicated, they refer to both molecules.

The protein reveals a compact, almost spherical shape of ~40 Å in diameter and is subdivided into three moieties (Fig. 2*A*): the N-terminal propeptide (Gln^{21}-Asn^{34}), and a CD split into a larger N-terminal upper subdomain moiety (Tyr^{35}-Gly^{162}; NTS) and a smaller C-terminal lower subdomain moiety (Ile^{163}-Pro^{199}; CTS, see also Refs. 9 and 28)) if viewed in the standard orientation for MPs (35). NTS and CTS conform to the overall fold of vertebrate MMPs (47, 93) and are separated by a shallow active-site cleft. The NTS is an α/β-sandwich consisting of a twisted five-stranded pleated β-sheet (strands βI-βV; see Fig. 2*A*), which is parallel for its first four strands and antiparallel for its lowermost one, βIV. The sheet accommodates on its concave side two α-helices (the "backing helix" αA and the "active-site helix" αB; for numbering and extension of repeating secondary structure elements, see Fig. 2*C* of Ref. 9). The right-handed twist of the helices coincides with the right-handed twist of the sheet and both helices' axes intersect the strands of the sheet at an angle $\Omega \approx -35°$ (94). The two helices pack against each other interacting through Ala^{66}-Ala^{70} of αA and Leu^{149}-Ala^{154} of αB at a crossing angle $\Omega \approx -50°$, which corresponds to a class II helix interaction (94). The loop connecting strands βIII and βIV (LβIIIβIV) contains the "S-loop" (Gly^{100}-Leu^{115}), which encompasses first a binding site for a structural zinc (Zn^{998}) and, downstream, a binding site for a structural calcium (Ca^{997}; see Fig. 2*B*). The zinc is tetrahedrally coordinated by His^{102} Nε2, Asp^{104} Oδ2, His^{117} Nε2, and His^{133} Nδ1, whereas the calcium is octahedrally coordinated by six oxygens: Asp^{109} Oδ1, Gly^{110} O, Thr^{112} O, Ile^{114} O, Asp^{135} Oδ2, and Glu^{138} Oε2 (see Fig. 2*B* and its legend for details). The presence of calcium is consistent with its crucial role in catalysis (30) and in protein stability (see Fig. 1*C*). Such calcium is found in mammalian MMP structures (47, 93), but it was not found in previous mature Kly18 structures (see below and Refs. 9 and 28). At Gly^{162} of the CSBZ, the polypeptide chains take a sharp turn and enters the CTS (Fig. 2*A*), which mainly contains the "C-terminal helix" αC and the Met-turn, centered on Met^{173}, which forms a hydrophobic base for the catalytic metal-binding site and is required for its integrity in MMPs and metzincins in general (47).

The active-site cleft contains the catalytic zinc ion (Zn^{999}) at half-width coordinated by the three histidines of the CSBZ (His^{155}, His^{159}, and His^{165}) through their Nε2 atoms at distances 2.00–2.05Å (Fig. 2, *A* and *C*). The cleft is top-framed on its non-primed side (see Refs. 35 and 95)) by the "upper-rim strand" βIV of the NTS β-sheet, which in MMPs binds substrates in extended conformation from left to right through antiparallel β-ribbon-like interactions. On its primed side, the cleft is top-framed by the final stretch of the S-loop, termed the "bulge-edge segment" (Thr^{112}-Ley^{115}), and bottom-framed by the segment bridging the Met-turn and helix αC. This segment includes the "S$_1$′-wall forming segment" (Pro^{175}-Tyr^{177}) at the front and the "specificity loop" (Gly^{179}-Gln^{183}) at the back. Together with the first turn of the active-site helix αB, the latter structural elements contribute to the size and chemical nature of the S$_1$′ pocket, which confers specificity to Kly18 and also MMPs in general (47, 93), here for medium-sized to bulky hydrophobic residues (30). Side chains participating in pocket shaping include Leu^{115}, Ala^{116}, Thr^{151}, Val^{152}, His^{155}, Leu^{172}, Tyr^{177}, and Lys^{181}.

Inhibition by the Propeptide—The 14-residue propeptide starts at the front right and runs in extended conformation across the active-site cleft, thus blocking access to the cleft, though in the opposite direction to a substrate, *i.e.* right to left (Fig. 2, *A* and *C*). This reverse orientation of the propeptide in the cleft may contribute to attenuate autolysis, as previously suggested for zymogens of cysteine peptidases and mammalian MMPs (39). The interaction with the CD buries a surface of 2,100 ± 35 Å^2, which is much larger than the average of monomeric protein-protein domain intra-chain interfaces (1,193 Å^2 (96)) but is slightly lower than the range of typical MMP-protein inhibitor interaction surfaces (2,400–2,700 Å^2; see Ref. 97). The interaction includes 13 hydrogen bonds, a double salt bridge, one metallorganic bond, and hydrophobic carbon-carbon contacts between eight residues from the propeptide and 11 from the CD (see Table 3). Segments involved include almost the entire propeptide (Arg^{22}-Gly^{31}) and, from the CD, mainly Asn^{111}-Tyr^{120} from the bulge-edge segment and the upper-rim strand, and Pro^{175}-Tyr^{177} from the S$_1$′-wall forming segment. Further involved are Tyr^{106}, Ala^{124}, and Glu^{138} and the zinc-liganding histidine side chains. Four inter-main chain hydrogen bonds form on the primed side of the cleft (two with the S$_1$′-wall forming segment and two with the bulge-edge segment and strand βIV) and three more on the upstream non-primed side (with βIV and LβIVβV; Fig. 2*C*). In particular, Arg^{22} contacts the base of the S-loop: it doubly salt bridges Glu^{138}, which is also one of the calcium ligands (see above, Table 3 and Fig. 2*B*), and hydrogen bonds three carbonyl oxygens of the S-loop, Asn^{111}, Gly^{113}, and Thr^{112}, which, again, is also a calcium ligand. In addition, the Arg^{22} carbonyl oxygen binds the S$_1$′-wall forming segment and its side chain performs a hydrophobic interaction with Leu^{115}. Accordingly, this residue plays a major role in the stabilization of the Ca^{997} site and, thus, the zymogen in general, which explains its enhanced stability in response to thermal denaturation (see above). In addition, superposition of pKly18-E156A onto mature Kly18 in complex with a tetrapeptidic cleavage product in the primed side (see below) and human MMP-8 with a modeled substrate traversing its cleft based on inhibitor structures (98) indicates that Arg^{22} occupies the S$_3$′ position of the cleft.

However, the most important interaction of the propeptide with the CD is exerted by Asp^{25}, which approaches the catalytic zinc from the top and monodentately occupies through its Oδ1 atom the fourth position of the tetrahedral coordination sphere of the metal (2.00/2.04 Å apart; Fig. 2*C*) further to His^{155}, His^{159}, and His^{165} Nε2 atoms. The preceding carbonyl group of Tyr^{24} binds strand βIV, and its aromatic side chain penetrates the deep hydrophobic S$_1$′ pocket, mainly interacting with the

His^{155} ring face-to-face. The π-rings are ~3.5 Å apart and parallel but slightly displaced along the ring planes to form a half-overlapping sandwich, which gives rise to an optimal π-stacked structure (99). Downstream in the chain, Pro^{28} is in a pocket, probably S_2, framed by His^{159}, Glu^{164}, and Tyr^{120}, the latter two interact through a tight hydrogen bond (Tyr^{120} Oη-Glu^{164} Oε2,

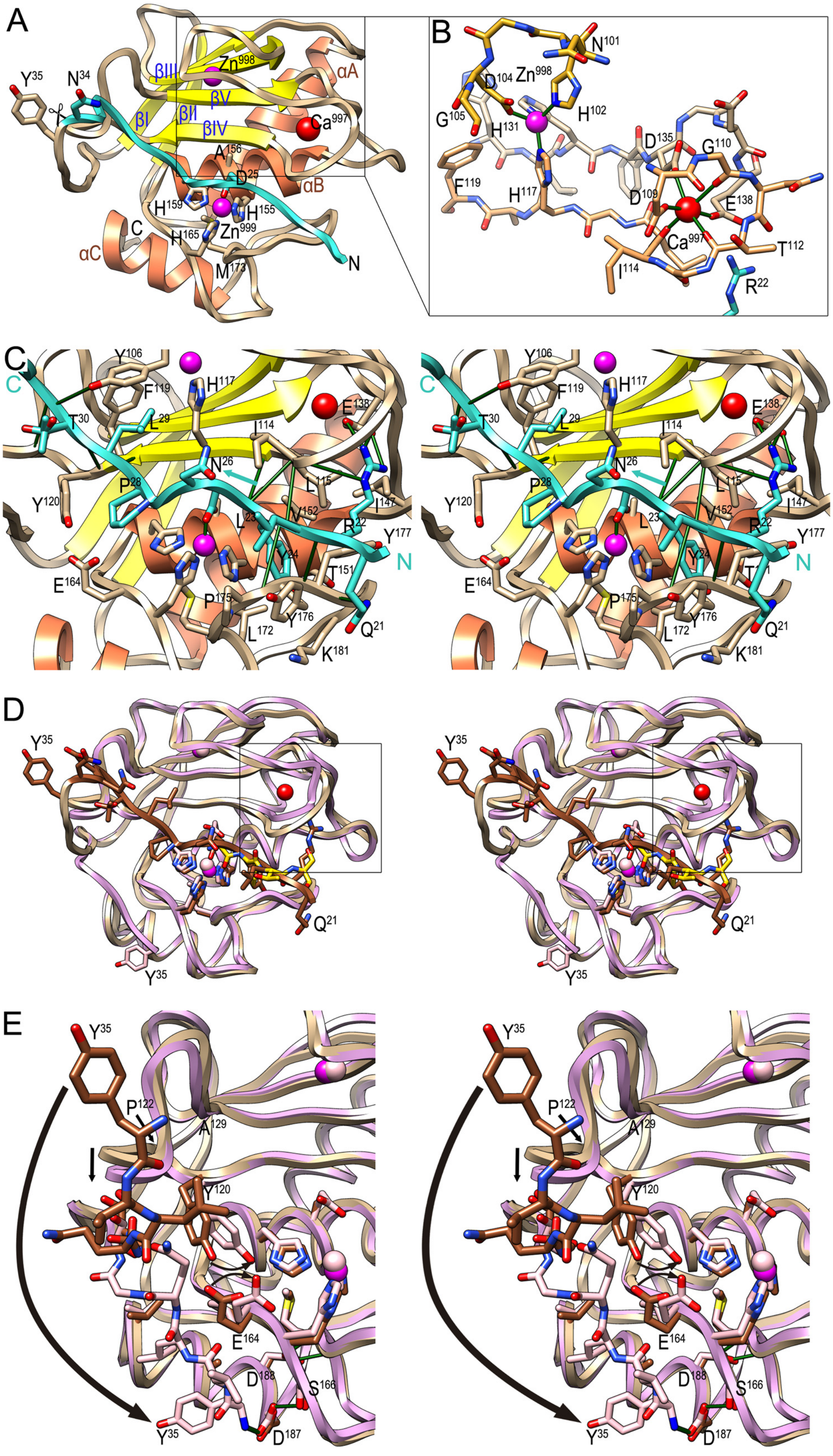


TABLE 3
Direct interactions between the propeptide (PP) and the catalytic domain (CD)

PP	CD	Distance (Å) molecule A / B	PP	CD
Hydrogen bonds			Hydrophobic carbon-carbon interactions	
R^{22} O	Y^{177} N	2.73 / 2.73	R^{22}	L^{115}
R^{22} Nη2	N^{111} O	3.19 / 3.19	L^{23}	Y^{176}
R^{22} Nη2	T^{112} O	3.20 / 3.21	Y^{24}	H^{155}
R^{22} Nη2	G^{113} O	3.24 / 3.22	Y^{24}	Y^{177}
R^{22} Nε	G^{113} O	3.18 / 3.19	D^{25}	I^{114}
Y^{24} N	P^{175} O	3.41 / 3.37	G^{27}	H^{159}
Y^{24} O	L^{115} N	2.82 / 2.79	G^{27}	H^{165}
Y^{24} O	A^{116} N	3.15 / 3.11	P^{28}	Y^{120}
L^{29} N	A^{118} O	2.81 / 2.81	P^{28}	H^{159}
L^{29} O	Y^{120} N	2.89 / 2.90	L^{29}	Y^{106}
T^{30} O	Y^{106} Oη	3.16 / 3.15	L^{29}	H^{117}
T^{30} Oγ1	Y^{106} Oη	3.75 / 4.18	L^{29}	F^{119}
G^{31} N	A^{124} O	2.93 / 2.98	T^{30}	Y^{120}

Salt bridges			Metallorganic bonds		
PP	CD	Distance (Å) molecule A / B	PP	CD	Distance (Å) molecule A / B
R^{22} Nη2	E^{138} Oε2	2.85 / 2.85	D^{25} Oδ1	Zn^{999}	2.00 / 2.04
R^{22} Nη1	E^{138} Oε1	3.08 / 3.01			

2.61 Å). Residue Leu^{29} is surrounded by the side chains of Tyr^{106}, His^{117}, and Phe^{119}, which may feature S$_3$ (Fig. 2*C*). After Gly^{31}, the polypeptide abandons the active-site cleft moving outward to reach the primary activation cleavage point, Asn^{34}-Tyr^{35} (Fig. 2*A*), after which the chain folds back toward the molecular moiety and enters strand βI of the NTS β-sheet.

A Novel Activation Mechanism in MMPs—Previous work had yielded three structures of mature wild-type Kly18 in complexes with tri- and tetrapeptidic cleavage products, as well as an inhibitory tetrapeptide in the non-primed side of the cleft (PDB 2XS3, 2XS4, and 4IN9 (9, 28)). These were obtained both in the presence and absence of magnesium and showed deviating chain traces for segment Asn^{53}-His^{57} (LβIαA) in the two molecules found in the asymmetric unit of the magnesium unbound structure (PDB 2XS3 (9)) and in the single molecules found in magnesium-bound (PDB 2XS4 (9)) and inhibitor-

FIGURE 2. **Overall structure of pKly18-E156A.** *A,* ribbon-type plot of pKly18-E156A in standard orientation (35). Depicted are the propeptide (ribbon and carbon atoms in *turquoise*) and the mature enzyme moiety (β-strands in *yellow* labeled βI-βV; α-helices in *salmon* labeled αA-αC; and coils and carbon atoms in *tan*). Further shown are the catalytic zinc ion (Zn^{999}; *bottom magenta sphere*), the structural zinc ion (Zn^{998}; *top magenta sphere*), and the structural calcium ion (Ca^{997}; *red sphere*), as well as the side chains of the three catalytic zinc ligands (His^{155}, His^{159}, and His^{165}), the Met turn methionine (Met^{173}), the alanine replacing the catalytic glutamate (Ala^{156}), and residues flanking the primary activation cleavage point, Asn^{34}-Tyr^{35}. *B,* close-up of the window of *A* as stick model highlighting the structural zinc and calcium sites. Protein segments depicted are Asn^{101}-Gly^{105} from the first part of the S-loop (carbons in *gold*), Asp^{109}-Phe^{119} from the second part of the S-loop (carbons in *sandy brown*), His^{131}-Glu^{138} from βV-LβVαB (carbons in *tan*), and the side chain of Arg^{22} from the propeptide (carbons in *turquoise*). The zinc is bound by His^{102} Nϵ2, Asp^{104} Oδ2, His^{117} Nϵ2, and His^{133} Nδ1 at distances 1.99–2.06 Å, and the calcium is bound by Asp^{109} Oδ1, Gly^{110} O, Thr^{112} O, Ile^{114} O, Asp^{135} Oδ2, and Glu^{138} Oϵ2 at distances 2.34–2.39 Å. These distances agree with standard zinc- (1.99–2.09 Å; (109)) and calcium-binding (2.36–2.39 Å; (109)) distance values for oxygens and nitrogens. *C,* close-up of *A* in wall-eye stereo centered on the catalytic zinc after a horizontal ~30° rotation upwards. Selected hydrogen and ionic bonds (see also Table 3) are depicted as *green lines*. Residues and ions labeled in *A* are not labeled here for clarity. The propeptide is shown in *cyan* to distinguish it from the mature catalytic moiety (in *tan*/*yellow*/*orange*) and its chain direction is pinpointed by a *cyan arrow* and labels of the N- and C-terminal parts depicted. *D,* superposition in wall-eye stereo of pKly18-E156A (*ribbon in tan* for the mature enzyme moiety and in *brown* for the propeptide, zinc ions in *magenta*, and calcium ion in *red*; stick model for the side chains of Ser^{20}-Tyr^{35} with carbons in *brown*) and Kly18 (ribbon and zinc ions in *pink*, see PDB 2XS3, molecule A (9)), which was obtained in a product complex with peptide A-F-T-S bound to the primed side of the cleft (stick model with carbons in *gold*). Tyr^{35} is shown for both structures. *E,* detail of *D* in wall-eye stereo depicting the large rearrangement of the N terminus at Tyr^{35} after maturation cleavage at Asn^{34}-Tyr^{35}. The α-amino group of Tyr^{35} makes a salt bridge with the side chain of Asp^{187} in the mature enzyme. Aside from Tyr^{120} and Glu^{164} (significantly) and Pro^{122}-Ala^{129} (slightly; see *black arrows*), maturation does not entail major conformational rearrangement of the rest of the structure.

bound crystals (PDB 4IN9 (28)). In addition, significant differences were also found in the second half of the S-loop including the bulge-edge segment, which was metal-free in all structures, as the aforementioned magnesium, which coincides with a potassium site in the inhibitor-bound form, was found on the opposite surface of the CD (see Fig. 1, *A* and *C*, in Ref. 9, and Fig. 1*A* in Ref. 28), in a place that suggests little if any functional or structural relevance. In these structures, either an outward- or an inward-folded flap was found for the S-loop (Fig. 1*E* in Ref. 9 and Fig. 1*D* in Ref. 28), which suggests intrinsic flexibility of this protein segment to adapt to different substrates. Among the distinct mature Kly18 coordinates, molecule A of the magnesium-unbound structure (PDB 2XS3) was chosen here for comparison with pKly18-E156A as it showed the lowest divergence in the overall chain trace (Fig. 2, *D* and *E*).

Superposition revealed that the mature CD is preformed in the zymogen and, with some notable local exceptions (see below), is simply uncovered during maturation by removal of the propeptide, as found in mammalian MMPs (47) and other MPs such as funnelins (79, 82). Removal occurs through cleavage at Asn^{34}-Tyr^{35}, which is solvent exposed on the molecular surface and thus readily accessible for processing (Fig. 2*A*). This explains why the wild-type zymogen undergoes rapid autolysis, so it cannot be isolated intact (see Ref. 30 and first section of "Results and Discussion"). This was the first cleavage observed *in vitro*, thus termed primary activation cleavage site, and no further cleavage was detected either within the propeptide or in the CD. The site is consistent with most vertebrate MMPs being activated at *X*-F/Y bonds, which are found at similar regions in all structures (10). Propeptide removal occurs under loss of a number of protein-protein interactions (see Table 3 and the preceding section), which explains why the mature enzyme is less stable to thermal denaturation (see first section of "Results and Discussion"). In particular, Arg^{22} plays a key role in stabilizing the Ca^{997} site (see above), and its removal may contribute to cation-site and S-loop flexibility, leading to metal loss. This site is easily created from the unbound form by two glycine-mediated main chain rotations (peptide flip of bond Thr^{112}-Gly^{113} and ~70° rotation of peptide bond Gly^{110}-Asn^{111}), so as to orient the carbonyl oxygens toward the interior, and cation binding should largely compensate for the energetic cost of such minor rearrangement. However, the finding that none of the mature Kly18 structures, which were partially obtained in the presence of calcium (9), contained an intact calcium site supports the requirement of Arg^{22} as an additional stabilizing factor for site integrity.

Activation further entails that the position occupied by Asp^{25} Oδ1 in the ligand sphere of the catalytic zinc (see the preceding section) is taken over by a catalytic solvent molecule, which renders a competent active site following an "aspartate-switch" mechanism. Such a competent zinc environment has also been reported for several mature MPs (see *e.g.* Refs. 64, 80, and 100). To date, aspartate-switch zymogenic mechanisms have been described only for astacins (7, 88) and fragilysins (26), which are only distantly related MPs grouped with MMPs within the metzincins. To verify the function of Asp^{25} in latency in pKly18, we used mutant pKly18-Y35A (from pKAR8), as the wild-type form (pKAR7) was insoluble. Although this mutant was produced with a yield similar to that of pKly18-E156A and was stable for several days, mutant pKly18-D25A/Y35A (pKAR9) was insoluble. We further assessed the function of Asp^{25} in full-length karilysin using the slowly autolytic mutant pKly-Y35A (pKAR2), as the reaction in the wild-type is too rapid (30). While pKly-Y35A was essentially intact after 5 days at 37 °C, pKly-D25A/Y35A (pKAR4) had been entirely transformed into the 38- and 18-kDa forms after this time (Fig. 1, *E* and *F*). Taken together, these results support the essential role of Asp^{25} in latency maintenance.

As to further changes upon maturation, segment Pro^{122}-Ala^{129} from LβIVβV is slightly shifted downwards by ~2 Å and the side chains of Tyr^{120} and Glu^{164} rotate toward the zinc site (Fig. 2*E*). Activation only entails major rearrangement of the new N-terminal segment Tyr^{35}-Ser^{40}, on the left surface (Fig. 2, *D* and *E*), which is rotated downward around bonds C-Cα and Cα-N of Ser^{40}. In this way, this segment nestles in a surface cavity framed by helix αC and the first segment of the CTS between Gly^{162}, and the "family specific residue," which is a serine in MMPs (1, 101) (here Ser^{166}). This entails that the new α-amino group of Tyr^{35}, which is translated 25 Å, establishes an intra-molecular salt bridge with Asp^{187} of αC, which is vaguely reminiscent of the activation of trypsin-like serine peptidases (102). Asp^{187}, in turn, is itself further bound to Ser^{166} and is adjacent to a second aspartate, Asp^{188}, which binds two main chain amides of the Met turn. This electrostatic network is characteristic of physiologically relevant mature MMPs, also referred to as "superactive forms" (47, 103). With the exception of the mature N-terminal fragment, the rest of this electrostatic network is already present in the zymogen (Fig. 2*E*).

Intensive studies of the activation of mammalian MMPs have produced the structures of pro-MMP-1 (PDB 1SU3 (90)), pro-MMP-3 (PDB 1SLM (104)), pro-MMP-9 (PDB 1L6J(105)), and pro-MMP-2 (PDB 1EAK; (89)). These studies revealed that the mammalian MMP zymogens contain a pre-formed competent protease moiety and true prodomains, which span between 66 and 91 residues, as shown for pro-MMP-2 (Fig. 3*A*) (47). The prodomains include elongated N-terminal extensions that may interact with ancillary domains, such as the fibronectin type II insertions found in MMP-2 and MMP-9, followed by globular cores of ~55 residues. These are made up of three α-helices that are arranged around a 3-fold axis with a left-handed twist.

The prodomain globular core serves as a scaffold to place a downstream peptide, which runs in extended conformation in the opposite direction to a bound substrate and thus blocks the active-site cleft (Fig. 3, *A* and *B*). This peptide encompasses the conserved motif involved in cysteine-switch or Velcro latency characteristic of animal and plant MMPs (48–50), ^{100}*P-R-C-G-N-P-D*^{106} (MMP-2 residues in italics; see PDB 1EAK and UP P08253), which is equivalent to pKly18 segment ^{23}L-Y-D-N-G-P-L^{29} (Fig. 3, *C* and *D*). Both the cysteine- and aspartate-switch motif show an intricate electrostatic network producing a unique scaffold to interact with the mature catalytic domain moiety. In contrast to pKly18, where the first cleavage occurs in the primary activation cleavage site, however, classical mammalian pro-MMPs are activated by conformational changes in the prodomain induced by cleavage in a so-called "bait region"

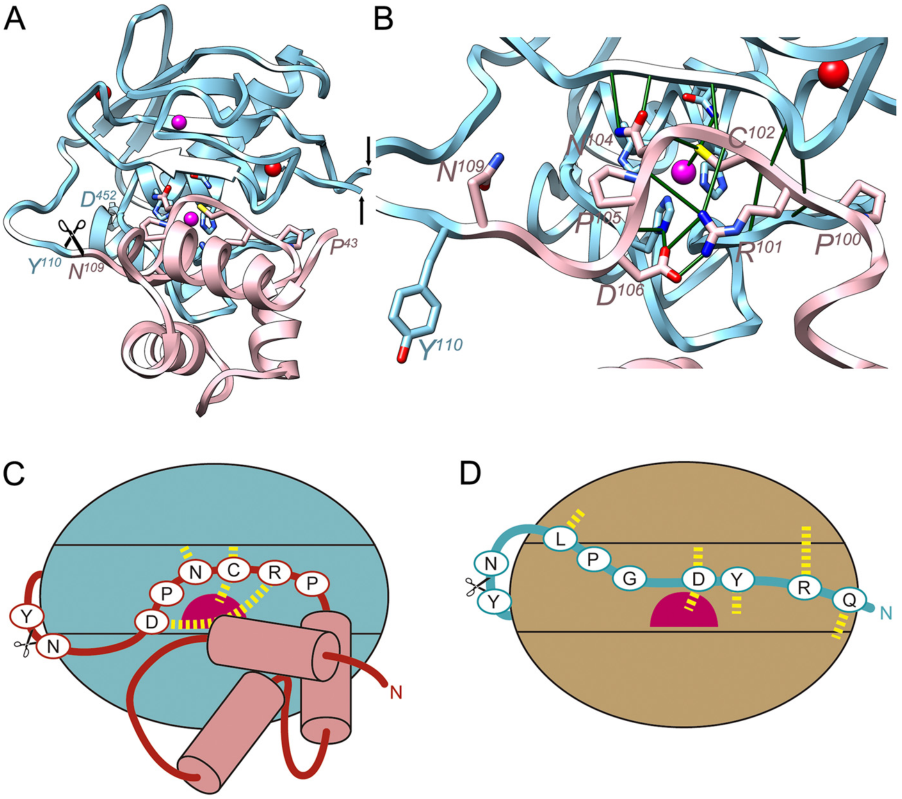


FIGURE 3. **Structural comparison with mammalian pro-MMPs.** *A,* schematic depicting of the structure of pro-MMP-2 (PDB 1EAK (89); MMP-2 residues in *italics* with superscripted numbering), shown only for its CD (*Tyr*^{110}-*Asp*^{52} in *cyan*; the fibronectin type-II domains spanning *Gln*^{219}-*Asp*^{92} have been omitted, the *black arrows* pinpoint the insertion points) and prodomain (*Pro*^{43}-*Asn*^{109} in *pink*, without the first 11 residues in extended conformation). The orientation displayed corresponds to that of Fig. 2*A* after applying a horizontal rotation of 15°. Residues of the conserved motif (*Pro*^{100}-*Asp*^{106}) key for structural integrity of the inhibitory segment are depicted for their side chains. *B*, close-up of *A* after removal of prodomain segment *Pro*^{43}-*Asn*^{66} to provide insight into the interactions of the conserved motif. Key electrostatic interactions are shown as *green lines*. The catalytic glutamate, *Glu*^{404}, is replaced by a glutamine, the histidines from the CSBZ are *His*^{403}, *His*^{407}, and *His*^{413}. *C* and *D*, scheme depicting the interaction modi of the propeptides of pro-MMP-2 through a cysteine-switch mechanism (*C*) and pKly18 through an aspartate-switch mechanism (*D*). The catalytic zinc ions are shown as *magenta spheres* and relevant interactions are shown as *yellow dashed lines*.

by several peptidases such as trypsin, plasmin, and other MMPs. Activation follows a "stepwise activation" process to eventually yield the final cleavage site *X*-F/Y accessible for processing and dissociation of cysteine and zinc to generate a functional active site (48, 49, 51, 106–108). As in Kly18, after cleavage at *Asn*^{109}-*Tyr*^{110}, the new N terminus is rearranged and participates in the electrostatic network centered on the conserved aspartate of helix αC, *Asp*^{346} in MMP-2.

Conclusions—This examination of the structure and function of the zymogen of the first bacterial MMP to be studied biochemically has uncovered several features of the activation mechanism of pKly18, which are shared with animal and plant MMPs: (i) the relevant cleavage site is *X*-F/Y; (ii) the scissile bond is located in similar regions of the structure; (iii) activation entails rearrangement of the segment equivalent to Tyr^{35}-Ser^{40} to yield a salt bridge between the new α-amino group and the first of two conserved aspartates in helix αC; (iv) this aspartate is bound to the family-specific serine; (v) the aspartate immediately downstream binds two main chain amides of the Met turn; (vi) the inhibitory segments run across the cleft in the opposite direction to a genuine substrate and metal blockage occurs through the side chain of an intervening residue, not through a chain terminus; and (vii) the catalytic moiety is largely preformed in the zymogen. All these features are related to the highly conserved CD itself. In contrast, all features of the mechanism related to the segment preceding this conserved CD diverge: (i) in pKly the propeptide spans just 14 residues and does not contain repetitive secondary structure elements, whereas eukaryotic MMPs feature a true protein prodomain that folds into a pseudosymmetric three-helix bundle followed by a segment in extended conformation; (ii) no relevant sequence similarity is found between the proregions; (iii) in eukaryotic MMPs activation occurs through a cysteine-switch mechanism exerted by residues from a conserved sequence motif, whereas in pKly18 this motif is absent and activation follows an aspartate-switch mechanism; (iv) multiple cleavages are apparently required in eukaryotic MMPs to liberate the CD, whereas a single cleavage suffices in pKly; and (v) the prodomain is not required for (re)folding of the catalytic moieties in eukaryotic MMPs, whereas it is in karilysin. In addition, pKly shares parts of its mechanism of latency with otherwise unrelated MPs from the astacin and fragilysin families. Accordingly, this overall novel mechanism unveiled for MMPs supports previous hypotheses, according to which Kly18 originated from an animal MMP CD co-opted through horizontal gene transfer by *T. forsythia*. This transfer was fostered by the intimate coexistence of the latter with the human blood-irrigated gingival crevice. Subsequently, Kly18 would have evolved in a bacterial environment, where it was furnished with unique flanking domains that contribute to a mechanism of zymogenicity similar to distantly related MPs only (9).

Acknowledgments—We are grateful to the joint IBMB/IRB Automated Crystallography Platform for assistance during crystallization experiments and Robin Rycroft for very helpful contributions to the manuscript. We acknowledge the help provided by local contacts at the ESRF synchrotron. The clone for production of human rhinovirus 3C proteinase was a generous gift from Arie Gerlof (EMBL, Hamburg). Funding for data collection was provided in part by ESRF. The Faculty of Biochemistry, Biophysics and Biotechnology of the Jagiellonian University is a beneficiary of structural funds from the European Union (POIG.02.01.00-12-064/08).

REFERENCES

1. Bode, W., Gomis-Rüth, F. X., and Stöcker, W. (1993) Astacins, serralysins, snake venom and matrix metalloproteinases exhibit identical zinc-binding environments (HEXXHXXGXXH and Met-turn) and topologies and should be grouped into a common family, the "metzincins". *FEBS Lett.* **331,** 134–140
2. Gomis-Rüth, F. X. (2003) Structural aspects of the *metzincin* clan of metalloendopeptidases. *Mol. Biotechnol.* **24,** 157–202
3. Gomis-Rüth, F. X. (2009) Catalytic domain architecture of metzincin metalloproteases. *J. Biol. Chem.* **284,** 15353–15357
4. Stöcker, W., Grams, F., Baumann, U., Reinemer, P., Gomis-Rüth, F. X., McKay, D. B., and Bode, W. (1995) The metzincins: topological and sequential relations between the astacins, adamalysins, serralysins, and matrixins (collagenases) define a superfamily of zinc-peptidases. *Protein Sci.* **4,** 823–840
5. Cerdà-Costa, N., and Gomis-Rüth, F. X. (2014) Architecture and function of metallopeptidase catalytic domains. *Protein Sci.* **23,** 123–144
6. Balaban, N. P., Rudakova, N. L., and Sharipova, M. R. (2012) Structural and functional characteristics and properties of metzincins. *Biochemistry* **77,** 119–127
7. Stöcker, W., and Gomis-Rüth, F. X. (2013) Astacins:proteases in development and tissue differentiation. in *Proteases: Structure and Function* (Brix, K., and Stöcker, W., eds) pp. 235–263, Springer Verlag, Vienna
8. Gomis-Rüth, F. X. (2013) Zinc adamalysins. in *Encyclopedia of Metalloproteins* (Uversky, V. N., Kretsinger, R. H., and Permyakov, E. A., eds) pp. 2345–2349, Springer Verlag, Heidelberg
9. Cerdà-Costa, N., Guevara, T., Karim, A. Y., Ksiazek, M., Nguyen, K. A., Arolas, J. L., Potempa, J., and Gomis-Rüth, F. X. (2011) The structure of the catalytic domain of *Tannerella forsythia* karilysin reveals it is a bacterial xenologue of animal matrix metalloproteinases. *Mol. Microbiol.* **79,** 119–132
10. Sang, Q. A., and Douglas, D. A. (1996) Computational sequence analysis of matrix metalloproteinases. *J. Protein Chem.* **15,** 137–160
11. Fanjul-Fernández, M., Folgueras, A. R., Cabrera, S., and López-Otín, C. (2010) Matrix metalloproteinases: evolution, gene regulation and functional analysis in mouse models. *Biochim. Biophys. Acta* **1803,** 3–19
12. Marino, G., and Funk, C. (2012) Matrix metalloproteinases in plants: a brief overview. *Physiol. Plant* **145,** 196–202
13. de Souza, S. J., and Brentani, R. (1993) Sequence homology between a bacterial metalloproteinase and eukaryotic matrix metalloproteinases. *J. Mol. Evol.* **36,** 596–598
14. Murphy, G. J., Murphy, G., and Reynolds, J. J. (1991) The origin of matrix metalloproteinases and their familial relationships. *FEBS Lett.* **289,** 4–7
15. Jackson, B. C., Carpenter, C., Nebert, D. W., and Vasiliou, V. (2010) Update of human and mouse forkhead box (FOX) gene families. *Hum. Genomics* **4,** 345–352
16. Massova, I., Kotra, L. P., Fridman, R., and Mobashery, S. (1998) Matrix metalloproeinases: structures, evolution, and diversification. *FASEB J.* **12,** 1075–1095
17. Angerer, L., Hussain, S., Wei, Z., and Livingston, B. T. (2006) Sea urchin metalloproteases: a genomic survey of the BMP-1/tolloid-like, MMP and ADAM families. *Dev. Biol.* **300,** 267–281
18. Das, S., Mandal, M., Chakraborti, T., Mandal, A., and Chakraborti, S. (2003) Structure and evolutionary aspects of matrix metalloproteinases: a brief overview. *Mol. Cell. Biochem.* **253,** 31–40
19. Keeling, P. J. (2009) Functional and ecological impacts of horizontal gene transfer in eukaryotes. *Curr. Opin. Genet. Dev.* **19,** 613–619
20. Anderson, M. T., and Seifert, H. S. (2011) Opportunity and means: horizontal gene transfer from the human host to a bacterial pathogen. *mBio* **2,** e00005–00011
21. Moncrief, J. S., Duncan, A. J., Wright, R. L., Barroso, L. A., and Wilkins, T. D. (1998) Molecular characterization of the fragilysin pathogenicity islet of enterotoxigenic *Bacteroides fragilis*. *Infect. Immun.* **66,** 1735–1739
22. Klein, T., and Bischoff, R. (2011) Active metalloproteases of the A Disintegrin and Metalloprotease (ADAM) family: biological function and structure. *J. Proteome Res.* **10,** 17–33
23. Takeda, S., Takeya, H., and Iwanaga, S. (2012) Snake venom metalloproteinases: structure, function and relevance to the mammalian ADAM/ADAMTS family proteins. *Biochim. Biophys. Acta* **1824,** 164–176
24. Edwards, D. R., Handsley, M. M., and Pennington, C. J. (2008) The ADAM metalloproteinases. *Mol. Aspects Med.* **29,** 258–289
25. Gomis-Rüth, F. X., Meyer, E. F., Kress, L. F., and Politi, V. (1998) Structures of adamalysin II with peptidic inhibitors: implications for the design of tumor necrosis factor α convertase inhibitors. *Protein Sci.* **7,** 283–292
26. Goulas, T., Arolas, J. L., and Gomis-Rüth, F. X. (2011) Structure, function and latency regulation of a bacterial enterotoxin potentially derived from a mammalian adamalysin/ADAM xenolog. *Proc. Natl. Acad. Sci. U.S.A.* **108,** 1856–1861
27. Goulas, T., and Gomis-Rüth, F. X. (2013) 186 Fragilysin. in *Handbook of Proteolytic Enzymes* (Rawlings, N. D., and Salvesen, G. S., eds) 3rd Ed., pp. 887–891, Academic Press, Oxford
28. Guevara, T., Ksiazek, M., Skottrup, P. D., Cerdà-Costa, N., Trillo-Muyo, S., de Diego, I., Riise, E., Potempa, J., and Gomis-Rüth, F. X. (2013) Structure of the catalytic domain of the *Tannerella forsythia* matrix metallopeptidase karilysin in complex with a tetrapeptidic inhibitor. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **69,** 472–476
29. Jusko, M., Potempa, J., Karim, A. Y., Ksiazek, M., Riesbeck, K., Garred, P., Eick, S., and Blom, A. M. (2012) A metalloproteinase karilysin present in the majority of *Tannerella forsythia* isolates inhibits all pathways of the complement system. *J. Immunol.* **188,** 2338–2349
30. Karim, A. Y., Kulczycka, M., Kantyka, T., Dubin, G., Jabaiah, A., Daugherty, P. S., Thogersen, I. B., Enghild, J. J., Nguyen, K. A., and Potempa, J. (2010) A novel matrix metalloprotease-like enzyme (karilysin) of the periodontal pathogen *Tannerella forsythia* ATCC 43037. *Biol. Chem.* **391,** 105–117
31. Koziel, J., Karim, A. Y., Przybyszewska, K., Ksiazek, M., Rapala-Kozik, M., Nguyen, K. A., and Potempa, J. (2010) Proteolytic inactivation of LL-37 by karilysin, a novel virulence mechanism of *Tannerella forsythia*. *J. Innate Immun.* **2,** 288–293
32. Potempa, J., Gomis-Rüth, F. X., and Karim, A. Y. (2013) Karilysin in *Handbook of Proteolytic Enzymes* (Rawlings, N. D., and Salvesen, G. S., eds) 3rd Ed., pp. 883–886, Academic Press, Oxford
33. Skottrup, P. D., Sørensen, G., Ksiazek, M., Potempa, J., and Riise, E. (2012) A phage display selected 7-mer peptide inhibitor of the *Tannerella forsythia* metalloprotease-like enzyme karilysin can be truncated to Ser-Trp-Phe-Pro. *PloS One* **7,** e48537
34. Pomerantsev, A. P., Pomerantseva, O. M., Moayeri, M., Fattah, R., Tallant, C., and Leppla, S. H. (2011) A *Bacillus anthracis* strain deleted for six proteases serves as an effective host for production of recombinant proteins. *Protein Expr. Purif.* **80,** 80–90
35. Gomis-Rüth, F. X., Botelho, T. O., and Bode, W. (2012) A standard orientation for metallopeptidases. *Biochim. Biophys. Acta* **1824,** 157–163
36. Stöcker, W., and Bode, W. (1995) Structural features of a superfamily of zinc-endopeptidases: the metzincins. *Curr. Opin. Struct. Biol.* **5,** 383–390
37. Stroud, R. M., Kossiakoff, A. A., and Chambers, J. L. (1977) Mechanisms of zymogen activation. *Annu. Rev. Biophys. Bioeng.* **6,** 177–193

38. Bryan, P. N. (2002) Prodomains and protein folding catalysis. *Chem. Rev.* **102,** 4805–4816
39. Khan, A. R., and James, M. N. (1998) Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. *Protein Sci.* **7,** 815–836
40. Baker, D., Shiau, A. K., and Agard, D. A. (1993) The role of proregions in protein folding. *Curr. Opin. Cell Biol.* **5,** 966–970
41. Graef, C., Schacherl, M., Waltersperger, S., and Baumann, U. (2012) Crystal structures of archaemetzincin reveal a moldable substrate-binding site. *PloS One* **7,** e43863
42. Waltersperger, S., Widmer, C., Wang, M., and Baumann, U. (2010) Crystal structure of archaemetzincin AmzA from *Methanopyrus kandleri* at 1.5-Å resolution. *Proteins* **78,** 2720–2723
43. Ng, N., Littler, D., Le Nours, J., Paton, A. W., Paton, J. C., Rossjohn, J., and Beddoe, T. (2013) Cloning, expression, purification and preliminary x-ray diffraction studies of a novel AB(5) toxin. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **69,** 912–915
44. Ng, N. M., Littler, D. R., Paton, A. W., Le Nours, J., Rossjohn, J., Paton, J. C., and Beddoe, T. (2013) EcxAB Is a founding member of a new family of metalloprotease AB toxins with a hybrid cholera-like B subunit. *Structure* **21,** 2003–2013
45. Gomis-Rüth, F. X. (2013) A different look for AB5 toxins. *Structure* **21,** 1909–1910
46. Yu, A. C., Worrall, L. J., and Strynadka, N. C. (2012) Structural insight into the bacterial mucinase StcE essential to adhesion and immune evasion during enterohemorrhagic *E. coli* infection. *Structure* **20,** 707–717
47. Tallant, C., Marrero, A., and Gomis-Rüth, F. X. (2010) Matrix metalloproteinases: fold and function of their catalytic domains. *Biochim. Biophys. Acta* **1803,** 20–28
48. Springman, E. B., Angleton, E. L., Birkedal-Hansen, H., and Van Wart, H. E. (1990) Multiple modes of activation of latent human fibroblast collagenase: evidence for the role of a Cys73 active-site zinc complex in latency and a "cysteine switch" mechanism for activation. *Proc. Natl. Acad. Sci. U.S.A.* **87,** 364–368
49. Van Wart, H. E., and Birkedal-Hansen, H. (1990) The cysteine switch: a principle of regulation of metalloproteinase activity with potential applicability to the entire matrix metalloproteinase gene family. *Proc. Natl. Acad. Sci. U.S.A.* **87,** 5578–5582
50. Vallee, B. L., and Auld, D. S. (1990) Zinc coordination, function, and structure of zinc enzymes and other proteins. *Biochemistry* **29,** 5647–5659
51. Rosenblum, G., Meroueh, S., Toth, M., Fisher, J. F., Fridman, R., Mobashery, S., and Sagi, I. (2007) Molecular structures and dynamics of the stepwise activation mechanism of a matrix metalloproteinase zymogen: challenging the cysteine switch dogma. *J. Am. Chem. Soc.* **129,** 13566–13574
52. Loechel, F., Overgaard, M. T., Oxvig, C., Albrechtsen, R., and Wewer, U. M. (1999) Regulation of human ADAM 12 protease by the prodomain: evidence for a functional cysteine switch. *J. Biol. Chem.* **274,** 13427–13433
53. Grams, F., Huber, R., Kress, L. F., Moroder, L., and Bode, W. (1993) Activation of snake venom metalloproteinases by a cysteine switch-like mechanism. *FEBS Lett.* **335,** 76–80
54. Leonard, J. D., Lin, F., and Milla, M. E. (2005) Chaperone-like properties of the prodomain of TNFα-converting enzyme (TACE) and the functional role of its cysteine switch. *Biochem. J.* **387,** 797–805
55. Tallant, C., García-Castellanos, R., Seco, J., Baumann, U., and Gomis-Rüth, F. X. (2006) Molecular analysis of ulilysin, the structural prototype of a new family of metzincin metalloproteases. *J. Biol. Chem.* **281,** 17920–17928
56. Gomis-Rüth, F. X., Trillo-Muyo, S., and Stöcker, W. (2012) Functional and structural insights into astacin metallopeptidases. *Biol. Chem.* **393,** 1027–1041
57. Sterchi, E. E., Stöcker, W., and Bond, J. S. (2008) Meprins, membrane-bound and secreted astacin metalloproteinases. *Mol. Aspects Med.* **29,** 309–328
58. Bond, J. S., and Beynon, R. J. (1995) The astacin family of metalloendopeptidases. *Protein Sci.* **4,** 1247–1261
59. Goulas, T., Cuppari, A., García-Castellanos, R., Snipas, S., Glockshuber, R., Arolas, J. L., and Gomis-Rüth, F. X. (2014) The pCri system: a vector collection for recombinant protein expression and purification. *PloS One* **9,** e112643
60. Ericsson, U. B., Hallberg, B. M., Detitta, G. T., Dekker, N., and Nordlund, P. (2006) Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal. Biochem.* **357,** 289–298
61. Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 125–132
62. Kabsch, W. (2010) Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 133–144
63. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674
64. Bode, W., Gomis-Rüth, F. X., Huber, R., Zwilling, R., and Stöcker, W. (1992) Structure of astacin and implications for activation of astacins and zinc-ligation of collagenases. *Nature* **358,** 164–167
65. Gomis-Rüth, F. X., Stöcker, W., Huber, R., Zwilling, R., and Bode, W. (1993) Refined 1.8-Å X-ray crystal structure of astacin, a zinc-endopeptidase from the crayfish *Astacus astacus* L. Structure determination, refinement, molecular structure and comparison with thermolysin. *J. Mol. Biol.* **229,** 945–968
66. Cohen, S. X., Ben Jelloul, M., Long, F., Vagin, A., Knipscheer, P., Lebbink, J., Sixma, T. K., Lamzin, V. S., Murshudov, G. N., and Perrakis, A. (2008) ARP/wARP and molecular replacement: the next generation. *Acta Crystallogr. D Biol. Crystallogr.* **64,** 49–60
67. Debreczeni, J. É., and Emsley, P. (2012) Handling ligands with Coot. *Acta Crystallogr. D Biol. Crystallogr.* **68,** 425–430
68. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 486–501
69. Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68,** 352–367
70. Smart, O. S., Womack, T. O., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vonrhein, C., and Bricogne, G. (2012) Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr. D Biol. Crystallogr.* **68,** 368–380
71. Blanc, E., Roversi, P., Vonrhein, C., Flensburg, C., Lea, S. M., and Bricogne, G. (2004) Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2210–2221
72. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera: a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612
73. Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2256–2268
74. Kabsch, W. (1976) A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. A* **32,** 922–923
75. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67,** 235–242
76. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 12–21
77. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR System: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54,** 905–921

78. Karplus, P. A., and Diederichs, K. (2012) Linking crystallographic model and data quality. *Science* **336,** 1030–1033
79. Gomis-Rüth, F. X. (2008) Structure and mechanism of metallocarboxypeptidases. *Crit. Rev. Biochem. Mol. Biol.* **43,** 319–345
80. Matthews, B. W. (1988) Structural basis of the action of thermolysin and related zinc peptidases. *Acc. Chem. Res.* **21,** 333–340
81. Vendrell, J., Querol, E., and Avilés, F. X. (2000) Metallocarboxypeptidases and their protein inhibitors. Structure, function and biomedical properties. *Biochim. Biophys. Acta* **1477,** 284–298
82. Arolas, J. L., and Gomis-Rüth, F. X. (2013) Zinc metallocarboxypeptidases. in *Encyclopedia of Metalloproteins* (Uversky, V. N., Kretsinger, R. H., and Permyakov, E. A., eds) pp. 2473–2479, Springer Verlag, Heidelberg, Germany
83. Wetmore, D. R., and Hardman, K. D. (1996) Roles of the propeptide and metal ions in the folding and stability of the catalytic domain of stromelysin (matrix metalloproteinase 3). *Biochemistry* **35,** 6549–6558
84. Gonzales, P. E., Solomon, A., Miller, A. B., Leesnitzer, M. A., Sagi, I., and Milla, M. E. (2004) Inhibition of the tumor necrosis factor-α-converting enzyme by its prodomain. *J. Biol. Chem.* **279,** 31638–31645
85. Fotouhi, N., Lugo, A., Visnick, M., Lusch, L., Walsky, R., Coffey, J. W., and Hanglow, A. C. (1994) Potent peptide inhibitors of stromelysin based on the prodomain region of matrix metalloproteinases. *J. Biol. Chem.* **269,** 30227–30231
86. Stetler-Stevenson, W. G., Talano, J. A., Gallagher, M. E., Krutzsch, H. C., and Liotta, L. A. (1991) Inhibition of human type IV collagenase by a highly conserved peptide sequence derived from its prosegment. *Am. J. Med. Sci.* **302,** 163–170
87. Woessner, J. F., Jr., and Nagase, H. (2000) *Matrix metalloproteinases and TIMPs (Protein Profile Series)* (Sheterline, P., ed) Oxford University Press, New York
88. Guevara, T., Yiallouros, I., Kappelhoff, R., Bissdorf, S., Stöcker, W., and Gomis-Rüth, F. X. (2010) Proenzyme structure and activation of astacin metallopeptidase. *J. Biol. Chem.* **285,** 13958–13965
89. Morgunova, E., Tuuttila, A., Bergmann, U., Isupov, M., Lindqvist, Y., Schneider, G., and Tryggvason, K. (1999) Structure of human pro-matrix metalloproteinase-2: activation mechanism revealed. *Science* **284,** 1667–1670
90. Jozic, D., Bourenkov, G., Lim, N. H., Visse, R., Nagase, H., Bode, W., and Maskos, K. (2005) X-ray structure of human pro-MMP-1: new insights into procollagenase activation and collagen binding. *J. Biol. Chem.* **280,** 9578–9585
91. Demidyuk, I. V., Gromova, T. Y., Polyakov, K. M., Melik-Adamyan, W. R., Kuranova, I. P., and Kostrov, S. V. (2010) Crystal structure of protealysin precursor: insights into propeptide function. *J. Biol. Chem.* **285,** 2003–2013
92. Gao, X., Wang, J., Yu, D. Q., Bian, F., Xie, B. B., Chen, X. L., Zhou, B. C., Lai, L. H., Wang, Z. X., Wu, J. W., and Zhang, Y. Z. (2010) Structural basis for the autoprocessing of zinc metalloproteases in the thermolysin family. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 17569–17574
93. Bode, W., and Maskos, K. (2003) Structural basis of the matrix metalloproteinases and their physiological inhibitors, the tissue inhibitors of metalloproteinases. *Biol. Chem.* **384,** 863–872
94. Chothia, C., Levitt, M., and Richardson, D. (1977) Structure of proteins: packing of α-helices and pleated sheets. *Proc. Natl. Acad. Sci. U.S.A.* **74,** 4130–4134
95. Schechter, I., and Berger, A. (1967) On the size of active site in proteases:. I. papain. *Biochem. Biophys. Res. Commun.* **27,** 157–162
96. Jones, S., Marin, A., and Thornton, J. M. (2000) Protein domain interfaces: characterization and comparison with oligomeric protein interfaces. *Protein Eng.* **13,** 77–82
97. Batra, J., Soares, A. S., Mehner, C., and Radisky, E. S. (2013) Matrix metalloproteinase-10/TIMP-2 structure and analyses define conserved core interactions and diverse exosite interactions in MMP/TIMP complexes. *PloS One* **8,** e75836
98. Grams, F., Reinemer, P., Powers, J. C., Kleine, T., Pieper, M., Tschesche, H., Huber, R., and Bode, W. (1995) X-ray structures of human neutrophil collagenase complexed with peptide thiol inhibitors: implications for substrate binding and rational drug design. *Eur. J. Biochem.* **228,** 830–841
99. Hunter, C. A., and Sanders, J. K. M. (1990) The natutre of π-π interactions. *J. Am. Chem. Soc.* **112,** 5525–5534
100. Gomis-Rüth, F. X., Kress, L. F., and Bode, W. (1993) First structure of a snake venom metalloproteinase: a prototype for matrix metalloproteinases/collagenases. *EMBO J.* **12,** 4151–4157
101. Jiang, W., and Bond, J. S. (1992) Families of metalloendopeptidases and their relationships. *FEBS Lett.* **312,** 110–114
102. Bode, W. (1979) Aktivierung, Aktivität und Inhibierung des Rindertrypsins. *Die Naturwissenschaften* **66,** 251–258
103. Reinemer, P., Grams, F., Huber, R., Kleine, T., Schnierer, S., Piper, M., Tschesche, H., and Bode, W. (1994) Structural implications for the role of the N terminus in the "superactivation" of collagenases: a crystallographic study. *FEBS Lett.* **338,** 227–233
104. Becker, J. W., Marcy, A. I., Rokosz, L. L., Axel, M. G., Burbaum, J. J., Fitzgerald, P. M., Cameron, P. M., Esser, C. K., Hagmann, W. K., Hermes, J. D., and Springer, J. P. (1995) Stromelysin-1: three-dimensional structure of the inhibited catalytic domain and of the C-truncated proenzyme. *Protein Sci.* **4,** 1966–1976
105. Elkins, P. A., Ho, Y. S., Smith, W. W., Janson, C. A., D'Alessio, K. J., McQueney, M. S., Cummings, M. D., and Romanic, A. M. (2002) Structure of the C-terminally truncated human ProMMP9, a gelatin-binding matrix metalloproteinase. *Acta Crystallogr. D Biol. Crystallogr.* **58,** 1182–1192
106. Cha, J., Pedersen, M. V., and Auld, D. S. (1996) Metal and pH dependence of heptapeptide catalysis by human matrilysin. *Biochemistry* **35,** 15831–15838
107. Nagase, H. (1997) Activation mechanisms of matrix metalloproteinases. *Biol. Chem.* **378,** 151–160
108. Nagase, H., Enghild, J. J., Suzuki, K., and Salvesen, G. (1990) Stepwise activation mechanisms of the precursor of matrix metalloproteinase 3 (stromelysin) by proteinases and (4-aminophenyl)mercuric acetate. *Biochemistry* **29,** 5783–5789
109. Harding, M. M. (2006) Small revisions to predicted distances around metal sites in proteins. *Acta Crystallogr. D Biol. Crystallogr.* **62,** 678–682
110. Weiss, M. S. (2001) Global indicators of X-ray quality. *J. Appl. Crystallogr.* **34,** 130–135
111. Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62,** 72–82

RESULTS: PROJECT 2

"A Novel Family of Soluble Minimal Scaffolds Provides Structural Insight into the Catalytic Domains of Integral Membrane Metallopeptidases"

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 288, NO. 29, pp. 21279–21294, July 19, 2013
© 2013 by The American Society for Biochemistry and Molecular Biology, Inc. Published in the U.S.A.

A Novel Family of Soluble Minimal Scaffolds Provides Structural Insight into the Catalytic Domains of Integral Membrane Metallopeptidases*

Received for publication, April 11, 2013, and in revised form, May 13, 2013 Published, JBC Papers in Press, June 3, 2013, DOI 10.1074/jbc.M113.476580

Mar López-Pelegrín[‡1], Núria Cerdà-Costa[‡1], Francisco Martínez-Jiménez[§¶], Anna Cintas-Pedrola[‡], Albert Canals[‖], Juan R. Peinado[‡2], Marc A. Marti-Renom[§¶], Carlos López-Otín[**], Joan L. Arolas[‡3], and F. Xavier Gomis-Rüth[‡4]

From the [‡]Proteolysis Laboratory, Department of Structural Biology, Molecular Biology Institute of Barcelona, Consejo Superior de Investigaciones Científicas, c/Baldiri Reixac, 15-21, 08028 Barcelona, the [§]Genome Biology Group, Centre Nacional d'Anàlisi Genòmic, c/Baldiri Reixac, 4, 08028 Barcelona, the [¶]Gene Regulation, Stem Cells and Cancer Program, Center for Genomic Regulation, c/Dr. Aiguader, 88, 08003 Barcelona, the [‖]Molecular Biology Institute of Barcelona, Consejo Superior de Investigaciones Científicas and Institute for Research in Biomedicine, c/Baldiri Reixac, 10-12, 08028 Barcelona, and the [**]Departamento de Bioquímica y Biología Molecular and Instituto Universitario de Oncología, Universidad de Oviedo, 33006 Oviedo, Spain

Background: Structural characterization of integral-membrane (IM) metallopeptidases (MPs) faces enormous technical hurdles.
Results: We have discovered a novel family of minimal MPs, minigluzincins, and determined the crystal structures of the zymogens of two family members.
Conclusion: Minigluzincins are valid models for catalytic domains of M48 and M56 family IMMPs.
Significance: They provide a high resolution scaffold for the design of small molecule inhibitors of IMMPs.

In the search for structural models of integral-membrane metallopeptidases (MPs), we discovered three related proteins from thermophilic prokaryotes, which we grouped into a novel family called "minigluzincins." We determined the crystal structures of the zymogens of two of these (*Pyrococcus abyssi* proabylysin and *Methanocaldococcus jannaschii* projannalysin), which are soluble and, with ~100 residues, constitute the shortest structurally characterized MPs to date. Despite relevant sequence and structural similarity, the structures revealed two unique mechanisms of latency maintenance through the C-terminal segments previously unseen in MPs as follows: intramolecular, through an extended tail, in proabylysin, and crosswise intermolecular, through a helix swap, in projannalysin. In addition, structural and sequence comparisons revealed large similarity with MPs of the gluzincin tribe such as thermolysin, leukotriene A4 hydrolase relatives, and cowrins. Noteworthy, gluzincins mostly contain a glutamate as third characteristic zinc ligand, whereas minigluzincins have a histidine. Sequence and structural similarity further allowed us to ascertain that minigluzincins are very similar to the catalytic domains of integral membrane MPs of the MEROPS database families M48 and M56, such as FACE1, HtpX, Oma1, and BlaR1/MecR1, which are provided with trans-membrane helices flanking or inserted into a minigluzincin-like catalytic domain. In a time where structural biochemistry of integral-membrane proteins in general still faces formidable challenges, the minigluzincin soluble minimal scaffold may contribute to our understanding of the working mechanisms of these membrane MPs and to the design of novel inhibitors through structure-aided rational drug design approaches.

The plasma membrane is the key information interface between the cytosol and cellular compartments or the extracellular environment, and whole genome analyses have revealed that 20–30% of proteins in humans, bacteria, and fungi are polytopic integral membrane proteins (1–4). They are of vital importance to living cells through their key roles in communication and transport, which explains why more than 60% of current pharmaceutical drug targets are membrane proteins (5, 6). Among integral membrane proteins are members of the peptidase class of enzymes, including integral-membrane metallopeptidases (IMMPs)[5] from families M48 and M56 (according to the MEROPS database (7)), which are α-helical for their transmembrane parts, as found in the otherwise unrelated rhodopsins, transporters, and channels (8).

---

* This work was supported in part by European, Spanish, and Catalan Agency Grants FP7-HEALTH-F3-2009-223101 "AntiPathoGN," FP7-HEALTH-2010-261460 "Gums&Joints," FP7-PEOPLE-2011-ITN-290246 "RAPID," FP7-HEALTH-2012-306029-2 "TRIGGER," BFU2010-19310, BFU2012-32862, CSD2006-00015, Fundació "La Marató de TV3" Grants 2009-100732 and 2009SGR1036, postdoctoral JAE contract from Consejo Superior de Investigaciones Científicas (co-funded by FSE), and two FPI Ph.D. fellowships from the Spanish Ministry for Science and Technology, currently part of the Ministry of Economy and Competitiveness.

The atomic coordinates and structure factors (codes 4JIU and 4JIX) have been deposited in the Protein Data Bank (http://wwpdb.org/).

[1] Both authors contributed equally to this work.

[2] Present address: Dept. of Medical Sciences, University of Castilla-La Mancha, 13071 Ciudad Real, Spain.

[3] To whom correspondence may be addressed. Tel.: 34-93-4020187; Fax: 34-93-4034979; E-mail: jlacri@ibmb.csic.es.

[4] To whom correspondence may be addressed. Tel.: 34-93-4020186; Fax: 34-93-4034979; E-mail: xgrcri@ibmb.csic.es.

[5] The abbreviations used are: IMMP, integral-membrane metallopeptidase; Abz, aminobenzoyl; Amc, 7-amino-4-methylcoumarin; CD, catalytic domain; dnp, 2,4-dinitrophenylamino; FACE1, farnesylated protein-converting enzyme 1; Mca, 7-methoxycoumarin-4-acetyl; MP, metallopeptidase; NTS, N-terminal subdomain; ONp, *p*-nitrophenyl ester; PDB, Protein Data Bank; pNA, *p*-nitroanilide; TEV, tobacco etch virus; TMH, transmembrane helix; UP, UniProt sequence database access code; CHES, 2-(cyclohexylamino)ethanesulfonic acid; Tricine, *N*-[2-hydroxy-1,1-bis(hydroxymethyl)ethyl]glycine; r.m.s.d., root mean square deviation.

M48 metallopeptidases (MPs) are subdivided into M48A, M48B, and M48C, whose prototypes are, respectively, *Saccharomyces cerevisiae* Ste24p, *Escherichia coli* HtpX, and human mitochondrial Oma1 peptidase. Ste24p was discovered as an enzyme required for maturation of the mating pheromone a-factor in yeast (9). Its mammalian ortholog, farnesylated protein-converting enzyme 1 (FACE1 also known as Zmpste24), performs a critical cleavage that removes the hydrophobic farnesyl-modified tail of prelamin A (10). Knock-out mice show growth retardation, muscular dystrophy, and premature death; similarly, disruptive mutations in humans have drastic consequences for health and life span (11–13). HtpX, together with the ATP-dependent protein FtsH, participates in the quality control system of bacterial membrane proteins, which is essential for growth and survival of the cell (14–17). Failure of this system under environmental stress conditions leads to the accumulation of misfolded membrane proteins. Finally, Oma1 helps in maintaining the integrity of the mitochondrial inner membrane (18, 19). It plays a crucial role in the proteolytic inactivation of the dynamin-related GTPase Opa1, and its loss causes obesity and defective thermogenesis in mice (20). Family M56, in turn, includes BlaR1 from *Staphylococcus aureus* and *Bacillus licheniformis*, and MecR1 from *S. aureus*. They function as parts of the signal transduction systems that trigger bacterial resistance to β-lactam antibiotics, a phenomenon that poses a serious threat to animal and human health and is exerted through the synthesis of a β-lactamase or a penicillin-binding protein (21–28). BlaR1 and MecR1 are composed of an N-terminal IMMP domain facing the cytosol and a C-terminal extracellular sensor domain that binds environmental β-lactams (29–32).

The strategic and widespread biological relevance of IMMPs underpins the need to acquire detailed knowledge about them at the molecular level, and atomic structural information, as provided by x-ray diffraction or nuclear magnetic resonance, makes a major contribution to this type of knowledge (33, 34). However, structural biochemistry of integral membrane proteins in general faces formidable conceptual and technical challenges due to the low concentrations at which they are naturally present in organisms, the difficulty in finding adequate recombinant overexpression systems that yield sufficient amounts of native-like protein, and their insolubility in strictly aqueous media, which requires screening for detergents and lipids that mimic the native membrane environment (35). Consequently, the only IMMP structures published to date are that of the intramembrane site-2 protease (36) from the hyperthermophilic archaeon *Methanocaldococcus jannaschii* (MEROPS family M50), an ortholog of a human enzyme that releases the N-terminal transcription factor domain from membrane-bound sterol regulatory element-binding proteins (36–38) and is unrelated to M48 and M56 IMMPs, and of human and yeast FACE1/Ste24p (Protein Data Bank (PDB) codes 4AW6 and 4IL3 (39, 40)), the only functional, truly integral-membrane representatives of families M48 and M56. In such a scenario, strategies aimed at obtaining high resolution structural information on soluble, correctly folded fragments of IMMP target proteins, such as globular catalytic domains (CDs) inserted into the overall transmembrane scaffold, may prove helpful in both the study of catalytic mechanisms and the design of drug-like inhibitors. The presence in the amino acid sequence of M48 and M56 IMMPs of large segments encompassing a zinc-binding motif (HE*XX*H) characteristic of the gluzincin and metzincin clans of the zincin tribe of MPs (see Refs. 41–43 and Fig. 2 in Ref. 44) between transmembrane helices (TMHs) provides a useful starting point for the identification of such CDs, which are mostly cytoplasmic (Fig. 1). In Ste24p and FACE1, these lie between the fifth and sixth of their seven TMHs; in HtpX, these lie between the second and third of its four putative TMHs; and in BlaR1/MecR1, these lie between the third and fourth of their four possible IMMP TMHs (26). By contrast, Oma1 is predicted to have only two transmembrane segments according to bioinformatics approaches (Fig. 1), but this topology and the localization of the active site still remain controversial (18).

In this work, we aimed to study the fold of CDs of M48 and M56 IMMPs to provide insights into the minimal scaffold required for their activity. As extensive expression trials of the putative CDs of several members of both families failed,[6] we set about searching for uncharacterized stand alone hypothetic proteins with significant sequence similarity to these CDs using various bioinformatics approaches. We identified, cloned, and overexpressed three sequences from the archaeal thermophiles *Pyrococcus abyssi* and *M. jannaschii*, and the thermophilic bacterium *Aquifex aeolicus*. We further purified and crystallized the first two, determined and analyzed their crystal structures, and functionally verified their zymogenic forms. Finally, we validated their suitability as valuable structural models for the CDs of M48 and M56 IMMPs.

EXPERIMENTAL PROCEDURES

Bioinformatics Sequence Analyses—A search with a partial sequence of the M56 family protein BlaR1 from *S. aureus* (residues K133–L250; UniProt (UP) sequence database access code P18357) and MecR1 (L139–L253; UP P0A0A9), putatively encompassing their TMH-flanked CDs, was carried out in 2004 using ProDom protein domain family database (release 2004.1), and identified ProDom family PD020519 as containing this sequence stretch. This family consisted of several BlaR1/MecR1-type sequences (plus others annotated as related to *E. coli* HtpX) that had been found using the PSI-BLAST program (45) with UP Q57587 (annotated as archaeal uncharacterized protein MJ0123 from the thermoautotrophican archaeon *M. jannaschii*, hereafter referred to as projannalysin) as the query. In contrast to multidomain BlaR1/MecR1 relatives, UP Q57587 included only domain PD020519 and spanned 110 residues. Further single-domain hit sequences were the uncharacterized proteins UP Q9V1Y2 from the thermoautotrophican archaeon *P. abyssi* (105 residues, hereafter referred to as proabylysin) and UP O66407 from the bacterial thermophile *A. aeolicus* (110 residues). In the current release of ProDom (2010.1), the former family PD020519 has expanded dramatically, which has led to the three archaeal sequences being currently grouped into a separate family, PDB0S8B5, which also includes three additional sequences.

---

[6] A. Marrero, N. Cerdà-Costa, J. L. Arolas, and F. X. Gomis-Rüth, unpublished results.

The MEROPS database (release 9.7, as of November, 2012) was downloaded and used for a sequence-based search using PSI-BLAST. Starting with the three aforementioned sequences, the top 20 hits were selected (with PSI-BLAST *E*-values ≤0.005) for each search, resulting in a total of 69 unique homologous sequences. A multiple sequence alignment of all the sequences selected was obtained using the PHYLEMON2 server (46). A final consensus multiple sequence alignment was trimmed using the TRIMAl method (47) and was subsequently used for searching the most likely model following the PROTEST methodology (48). BLOSUM62 was the matrix used by PHYML (49) to build a phylogenetic tree, which was then represented using the FIGTREE program. Sequence-based predictions of regular secondary structure elements were performed with JPRED3 (50).

Protein Production and Purification—Synthetic genes coding for the *A. aeolicus* protein (which contained a lysine-to-alanine mutation at position 2 due to the cloning strategy), proabylysin (lysine-to-valine mutation at position 2), and projannalysin (lysine-to-alanine mutation at position 2) were purchased from GeneArt (Invitrogen) and cloned into vectors pCRI7 and pCRI8 (both modified from pET-28a (Novagen), which confers resistance to kanamycin) using NcoI and XhoI restriction sites. While pCRI7 does not append any extra residue to the protein of interest, vector pCRI8 attaches an N-terminal hexahistidine tag followed by a tobacco etch virus (TEV) protease recognition site. C-terminal deletion mutants lacking one to five residues of proabylysin were cloned into vector pCRI7, and a C-terminal deletion mutant lacking five residues of projannalysin was cloned into vector pCRI8. In addition, the genes coding for proabylysin and its aforementioned C-terminal deletion mutants were cloned into vector pCRI6 (derived from plasmid pETM30, which confers resistance to kanamycin), which attaches an N-terminal hexahistidine tag followed by glutathione *S*-transferase (GST) and a TEV protease recognition site, and into a variant of vector pCRI7, into which an extra N-terminal streptavidin tag was introduced by PCR (hereafter referred to as vector pCRI7s). All constructs were verified by DNA sequencing. Proteins were produced by heterologous overexpression in *E. coli* BL21 (DE3) cells, which were grown at 37 °C in Luria Bertani medium supplemented with 30 μg/ml kanamycin. Cultures were induced at an A_{600} of 0.8 with 0.2–1 mM isopropyl-β-D-thiogalactopyranoside and kept growing either for 5 h at 37 °C or overnight at 18 °C.

The purification of proabylysin (cloned into the pCRI7 vector) was performed as follows. After centrifugation at 7,000 × *g* for 30 min at 4 °C, the cell pellet was washed twice with buffer A (50 mM Tris-HCl, 500 mM NaCl, pH 8.0) and resuspended in the same buffer supplemented with EDTA-free protease inhibitor mixture tablets (Roche Diagnostics) and DNase I (Roche Diagnostics). Cells were lysed at 4 °C using a cell disrupter (Constant Systems) at a pressure of 1.35 kbar, and the cell debris was removed by centrifugation at 50,000 × *g* for 1 h at 4 °C. The supernatant was incubated for 1 h at 85 °C, and the precipitated material was removed by further centrifugation at 50,000 × *g* for 1 h at 4 °C. The supernatant was filtered (0.22-μm pore size, Millipore) and dialyzed overnight at room temperature against buffer B (20 mM Tris-HCl, pH 8.0). The protein was subsequently purified by anion exchange chromatography using a HiLoad 16/10 Q-Sepharose HP column followed by a Mono Q 4.6/100PE column (both from GE Healthcare) and finally polished with a HiLoad 16/60 Superdex 75 size exclusion chromatography column (GE Healthcare) previously equilibrated with buffer C (20 mM Tris-HCl, 150 mM NaCl, pH 7.4). The purification of the proabylysin C-terminal deletion mutants was carried out in the same way except that the supernatant was incubated for 1 h at 50–65 °C instead of 85 °C. The purification of proabylysin and its mutants with an N-terminal GST fusion protein or a streptavidin tag (encoded by vectors pCRI6 or pCRI7s vector, respectively, see above) was also performed in the same way except that the heat shock step was replaced by affinity chromatography purification on a GSTrap HP (GE Healthcare) or Strep-Tactin-Sepharose (IBA) column, respectively, according to the manufacturers' instructions.

The purification of projannalysin and its C-terminal deletion mutant (both cloned into the pCRI8 vector) was performed as follows. After centrifugation at 7,000 × *g* for 30 min at 4 °C, the pellet was washed twice with buffer A, resuspended in the same buffer plus 10 mM imidazole, and supplemented with EDTA-free protease inhibitor mixture tablets and DNase I (both Roche Diagnostics). Cells were lysed using a cell disrupter as described above, and the cell debris was removed by centrifugation at 50,000 × *g* for 1 h at 4 °C. The supernatant was filtered, incubated with nickel-nitrilotriacetic acid resin (Invitrogen) previously equilibrated with buffer A plus 10 mM imidazole, and the fusion protein eluted using buffer A plus 350 mM imidazole. The sample was then dialyzed overnight at 20 °C against buffer D (50 mM Tris-HCl, 250 mM NaCl, 1 mM dithiothreitol, pH 8.0) in the presence of hexahistidine-tagged TEV protease at an enzyme/substrate ratio of 1:50 (w/w). Cleavage left the dipeptide glycine-proline at the N terminus of both proteins. Digested samples were passed several times through nickel-nitrilotriacetic acid resin previously equilibrated with buffer A plus 10 mM imidazole to remove noncleaved hexahistidine-containing molecules. Flow-throughs were collected, concentrated by ultrafiltration, and further purified by size exclusion chromatography on a HiLoad 16/60 Superdex 75 column previously equilibrated with buffer C.

Protein identity and purity were assessed by mass spectrometry and 15% Tricine/SDS-PAGE stained with Coomassie Blue. Ultrafiltration steps were performed with Vivaspin 15 and Vivaspin 500 filter devices of 5-kDa cutoff (Sartorius Stedim Biotech). Protein concentration was determined by measuring the absorbance at 280 nm using a spectrophotometer (NanoDrop) and calculated absorption coefficients $E_{0.1\%}$ of 0.98 and 0.34 for proabylysin and projannalysin, respectively. N-terminal sequencing through Edman degradation, peptide mass fingerprinting of tryptic protein digests, and mass spectrometry analyses were carried out at the proteomics facilities of the Centro de Investigaciones Biológicas and Vall d'Hebron Institute of Oncology. Melting temperatures (T_m) through differential scanning fluorimetry (thermofluor) were determined by using Sypro Orange dye (Invitrogen) and an iCycler iQ real time PCR detection system (Bio-Rad) as published previously (51, 52). Limited proteolysis trials to remove C-terminal peptides from full-length proabylysin and projannalysin were undertaken in buffer C by overnight incubation at 37 or 80 °C with trypsin, subtilisin, thermolysin (all from Sigma), or ulilysin (produced in-house according to Ref. 53) at different peptidase/substrate ratios (1:1, 1:5, 1:20, 1:50, 1:100, 1:200, and 1:500 (w/w)) and monitored by mass spectrometry.

Proteolytic and Inhibitory Activity Assays—Proteolytic activities were assayed at 37 °C in buffer E (50 mM MES, 150 mM NaCl, pH 5.5), buffer F (50 mM HEPES, 150 mM NaCl, pH 7.5), or buffer G (50 mM CHES, 150 mM NaCl, pH 9.5) at a final protein concentration of 50 μg/ml for proabylysin and projannalysin, unless otherwise stated.

Proteolytic activity against the fluorescein conjugates BODIPY FL casein, DQ gelatin, and DQ BSA (all from Invitrogen) was tested according to the manufacturer's instructions using a microplate fluorimeter (FLx800, BioTek or Infinite M200, Tecan). Assays with natural protein substrates (at 0.25 or 0.50 mg/ml) included bovine plasma fibronectin, bovine muscle actin, human plasma fibrinogen, cold water fish skin gelatin, bovine milk casein, and bovine milk α-casein (all from Sigma). Reactions were carried out in buffer F at 37, 65, and 80 °C overnight and 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. Proteolytic activity was further tested on 10 fluorogenic peptide substrates of sequence: Abz-KDESYRK-(dnp) (where Abz is aminobenzoyl, and dnp is 2,4-dinitrophenylamino); Abz-TVLERSK-(dnp); Abz-DYVASEK-(dnp); Abz-YGKRVFK-(dnp); Abz-VKFYDIK-(dnp); Dabcyl-LARVE-Edans (where Dabcyl is *p*-dimethyl(aminophenyl)azobenzoate, and Edans is 2-aminoethylamino-1-naphthalene sulfonate); Abz-GIVRAK-(dnp) (Bachem); Mca-PLGL-Dap(dnp)-AR-NH_2 (where Mca is 7-methoxycoumarin-4-acetyl, and Dap is L-diaminopropionyl) (Bachem); Mca-RPKRVE-Nva-WRK-(dnp)-NH_2 (where Nva is norvaline) (Bachem); and Dnp-PLGLWA-DR-NH_2 (Bachem) (see Ref. 54 for details on the first six substrates). Reactions were performed at enzyme/substrate molar ratios of 1:0.25, 1:0.5, 1:1.25, and 1:6.25 and monitored in a microplate fluorimeter (Infinite M200, Tecan). In addition, proteolytic assays with DQ BSA and the fluorogenic peptides were performed at enzyme/substrate molar ratios of 1:0.1 and 1:0.5, respectively, as described above at high temperatures (up to 80 °C) or in the presence of 0 – 4 M urea or 0 – 4 M guanidinium hydrochloride.

Esterase activity against the chromogenic substrates A-ONp (where ONp is *p*-nitrophenyl ester), H-ONp, and L-ONp (all from Bachem), and carboxypeptidase activity against the chromogenic substrates *N*-(3-[2-furyl]acryloyl)-FF-OH (Bachem), *N*-(3-[2-furyl]acryloyl)-GLA-OH (Bachem), and *N*-(3-[2-furyl]-acryloyl)-LGPA-OH (Sigma) were tested using a microplate spectrophotometer (PowerWave XS, BioTek) at enzyme/substrate molar ratios of 1:25 and 1:125 for ester substrates and 1:25 for carboxypeptidase substrates. Finally, aminopeptidase activity was assayed with fluorogenic substrates F-Amc (where Amc is 7-amino-4-methylcoumarin), (H)T-Amc, and Y-Amc (all from Bachem), and with the chromogenic *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-VGR-pNA, *N*(α)-benzoyl-IEGR-pNA, and *N*-succinyl-AAPF-pNA. Reactions were monitored in a microplate fluorimeter (FLx800, BioTek) or PowerWave XS spectrophotometer (BioTek) at enzyme/substrate molar ratios of 1:125 and 1:250.

Crystallization and X-ray Diffraction Data Collection—Crystallization assays were carried out by the sitting-drop vapor diffusion method using 96× 2-well MRC plates (Innovadyne) and a Cartesian (Genomic Solutions) or a Phoenix (Art Robbins) nanodrop robot at the IBMB/IRB high throughput automated crystallography platform. Crystallization plates were stored in Bruker steady-temperature crystal farms at 4 and 20 °C. Successful hits were scaled up to the microliter range with 24-well Cryschem crystallization dishes (Hampton Research). The best crystals of proabylysin appeared at 20 °C from equivolumetric drops containing protein solution (2 mg/ml in buffer C) and 100 mM MES, 200 mM sodium/potassium phosphate, 2.6 M sodium chloride, pH 6.5, as reservoir solution. The best crystals of projannalysin were obtained at 4 °C from drops containing protein solution (20 mg/ml in buffer C) and 100 mM Tris-HCl, 200 mM lithium sulfate monohydrate, 25% (w/v) polyethylene glycol 3,350, pH 8.5, as reservoir solution at a 2:1 volume ratio. Crystals were cryo-protected with successive passages through reservoir solution containing increasing concentrations of glycerol, up to 10 and 20% (v/v) for projannalysin and proabylysin, respectively. A heavy ion derivative of projannalysin was prepared by soaking native crystals for 5 min in reservoir solution containing 5 mM mercury(II) acetate.

Complete diffraction datasets were collected at 100 K from liquid N_2 flash-cryo-cooled crystals (Oxford Cryosystems 700 series cryostream) on a Pilatus 6-m pixel and an ADSC Q315R CCD detector at beam lines ID29 (proabylysin) and ID23-1 (native and derivative projannalysin), respectively, at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) within the Block Allocation Group "BAG Barcelona." Crystals of proabylysin were primitive orthorhombic, with one molecule per asymmetric unit, whereas crystals of projannalysin were trigonal with two molecules per asymmetric unit. Diffraction data were integrated, scaled, merged, and reduced with programs XDS (55) and XSCALE (56) or SCALA (57), the latter within the CCP4 suite of programs (see Table 1) (58).

Structure Solution and Refinement—The structure of proabylysin was solved by single wavelength anomalous diffraction by using the diffraction data to 1.6 Å resolution of a crystal collected at the zinc K-edge peak wavelength, as inferred from a previous XANES fluorescence scan, which enabled program SHELXD (59, 60) to identify the zinc ion present in the asymmetric unit. Subsequent phasing with this site using program SHELXE (60, 61) and higher resolved diffraction data to 1.15 Å resolution collected at a remote wavelength as a pseudo-native dataset yielded a suitable electron density map for chain tracing on a Silicon Graphics Octane2 workstation using the TURBO-FRODO program (62) and on a Linux-operated PC using the COOT program (63, 64). Model building alternated with crystallographic refinement with program BUSTER (65), which

TABLE 1
Crystallographic data

Dataset	*P. abyssi* proabylysin (UP Q9V1Y2)		*M. jannaschii* projannalysin (UP Q57587)	
	Native	Zinc[a] (absorption peak)	Native	Mercury[a] (absorption peak)
Space group	$P2_12_12_1$	$P2_12_12_1$	$P3_121$	$P3_121$
Cell constants (*a, b,* and *c*)	34.62, 44.66, 72.00 Å	34.68, 44.90, 72.23 Å	76.59, 76.59, 124.32 Å	76.6, 76.6, 124.8 Å
Wavelength	0.9540 Å	1.2821 Å	0.9790 Å	1.0044 Å
No. of measurements/unique reflections	495,316/40,510	99,212/15,317	305,164/29,001	179,641/31,910
Resolution range (outermost shell)[b]	44.7 to 1.15 Å (1.21 to 1.15 Å)	44.9 to 1.60 Å (1.69 to 1.60 Å)	45.4 to 2.00 Å (2.05 to 2.00 Å)	89.1 to 2.40 Å (2.46 to 2.40 Å)
Completeness	100.0% (100.0%)	98.9% (92.9%)	99.1% (98.6%)	99.6% (96.6%)
R_{merge}[c]	0.072 (0.892)	0.059 (0.374)	0.042 (0.851)	0.048 (0.541)
$R_{\text{r.i.m.}}$ (= R_{meas})[c] (/$R_{\text{p.i.m.}}$[c])	0.075 (0.934)/0.021 (0.272)	0.069 (0.453)/0.035 (0.250)	0.044 (0.910)	0.054 (0.615)
Average intensity $(\langle\langle\langle I\rangle/\sigma(\langle I\rangle)\rangle\rangle)$	18.5 (3.4)	15.5 (4.5)	30.3 (2.8)	24.2 (3.0)
B-Factor (Wilson) (Å^2)/average multiplicity	11.0/12.2 (11.3)	18.2/6.5 (5.2)	49.4/10.5 (7.9)	54.4/5.6 (4.4)
Resolution range used for refinement	∞ −1.15 Å		∞ −2.00 Å	
No. of reflections used (test set)	39,668 (778)		28,924 (751)	
Crystallographic R_{factor} (free R_{factor})[d]	0.147 (0.192)		0.183 (0.213)	
No. of protein atoms/solvent molecules/	900/47/		1,815/194/	
Neutral ligands/ Ionic ligands	1 $(CH_2OH)_2CHOH$/ 1 Zn^{2+}		7 $(CH_2OH)_2CHOH$/ 1$C(CH_2OH)_3NH_3^+$, 2 Zn^{2+}, 3 Cl^-, 4 SO_4^{2-}	
r.m.s.d. from target values				
Bonds/angles	0.014 Å/2.24°		0.009 Å/0.96°	
Average *B*-factors for protein atoms	17.7 Å^2		54.9 Å^2	
Main-chain conformational angle analysis[e]				
Residues in favored regions/outliers/all residues	103/0/103		211/0/212	

[a] Friedel mates were treated as separate reflections.

[b] Values in parentheses refer to the outermost resolution shell.

[c] $R_{\text{merge}} = \Sigma_{hkl}\Sigma_i|I_i(hkl) - \langle I(hkl)\rangle|/\Sigma_{hkl}\Sigma_i I_i(hkl)$; $R_{\text{r.i.m.}} = \Sigma_{hkl}(n_{hkl}/(n_{hkl}-1))^{1/2}\,\Sigma_i|I_i(hkl) - \langle I(hkl)\rangle|/\Sigma_{hkl}\Sigma_i I_i(hkl)$; $R_{\text{p.i.m.}} = \Sigma_{hkl}(1/(n_{hkl}-1))^{1/2}\,\Sigma_i\,|I_i(hkl) - \langle I(hkl)\rangle|/\Sigma_{hkl}\Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the *i*th intensity measurement, and n_{hkl} is the redundancy of reflection *hkl,* including symmetry-related reflections, and $\langle I(hkl)\rangle$ is its average intensity. $R_{\text{r.i.m.}}$ (also known as R_{meas}) and $R_{\text{p.i.m.}}$ are improved multiplicity-weighted indicators of the quality of the data, the redundancy-independent merging *R* factor and the precision-indicating merging *R* factor. The latter is computed after averaging over multiple measurements (for details, see Refs. 97, 98).

[d] Crystallographic $R_{\text{factor}} = \Sigma_{hkl}\|F_{\text{obs}}| - k|F_{\text{calc}}\|/\Sigma_{hkl}|F_{\text{obs}}|$, where *k* is a scaling factor, and F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively. This factor is calculated for the working set reflections; free R_{factor} is the same for a test set of reflections (>500) not used during refinement.

[e] Data were according to MOLPROBITY (71).

included TLS refinement, until completion of the model. The final refinement was performed with SHELXL (61) and included anisotropic refinement of atomic displacement parameters. It included residues M^1–V^{105} (superscripted residue numbering for proabylysin), one zinc ion, one glycerol molecule, and 147 solvent molecules (see Table 1).

The structure of projannalysin was solved by single wavelength anomalous diffraction by using the diffraction data to 2.4 Å resolution of a crystal collected at the mercury L_{III}-edge peak wavelength, as determined from a previous XANES fluorescence scan, which enabled program SHELXD to find six mercury sites. Visual inspection of these sites on a graphic display with TURBO-FRODO enabled the identification of two clusters of three sites, which allowed the associated dimeric noncrystallographic symmetry operator to be derived. Subsequent phasing with these sites using program SHELXE and higher resolved native diffraction data to 2.0 Å resolution, followed by density modification with 2-fold averaging with the DM program (66), yielded an electron density map suitable for model building, which proceeded as for proabylysin. Crystallographic refinement was performed with BUSTER. The final model of projannalysin included residues G^{-2}–I^{110} (superscripted residue numbering for projannalysin in italics) of molecule A and thus included the N-terminal residues G^{-2} and P^{-1}, which precede M^1 as a result of the cloning strategy and TEV cleavage (see above). In addition, residues K^7–I^{110} of molecule B plus seven glycerols, one Tris and two zinc cations, four sulfate and three chloride anions, and 194 solvent molecules completed the model of projannalysin (see Table 1).

Figures were prepared with the CHIMERA program (67). Interaction surfaces (taken as the total surface area buried in a complex) were calculated with CNS taking a probe radius of 1.4 Å (68). Structural superpositions were performed with the SSM routine (69) within COOT. Surface complementarity was computed with SC (70) within CCP4 with default parameters. Model validation was performed with MOLPROBITY (71) and the WHATCHECK routine of WHATIF (72). The final coordinates of proabylysin and projannalysin have been deposited with the PDB (access codes 4JIU and 4JIX, respectively).

Structural Similarity Analyses—The MAMMOTH program (73) was used with default parameters to search for proabylysin and projannalysin homologous structures in the PDB (74) as of November, 2012. Structural similarity searches were also carried out with the DALI program (75).

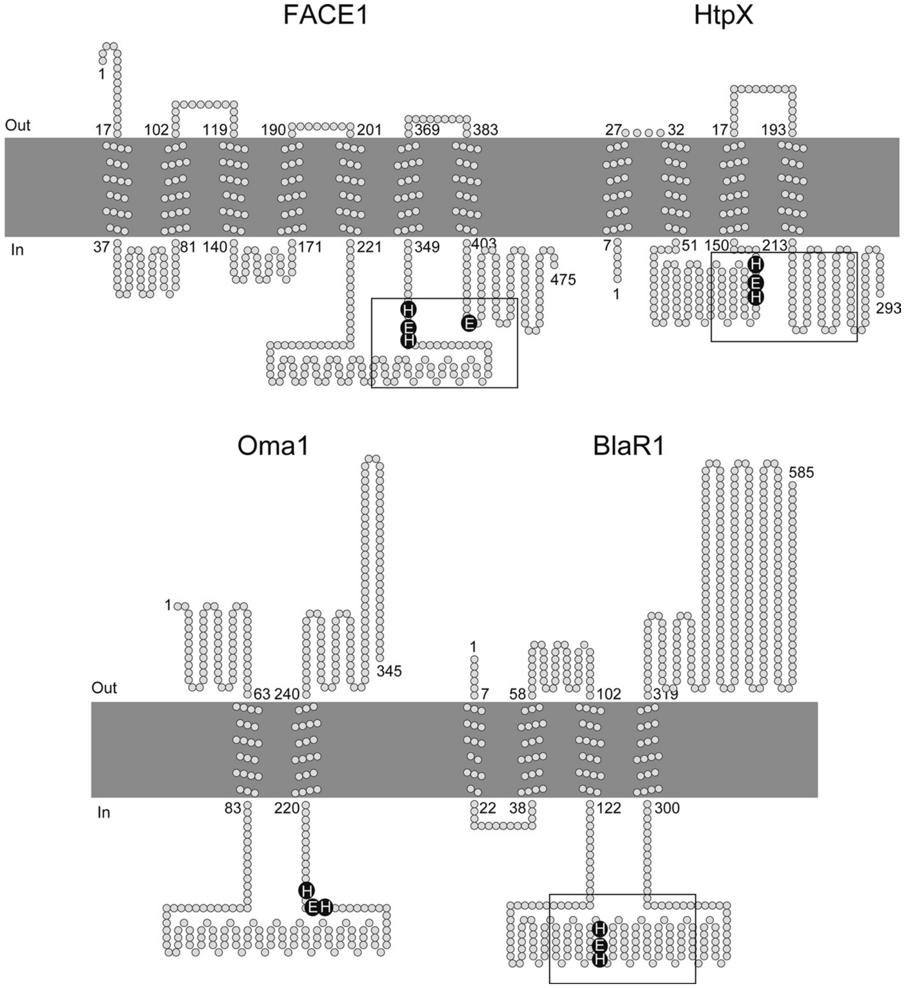


FIGURE 1. **Transmembrane topologies of selected M48 and M56 integral membrane metallopeptidases.** The predicted transmembrane topologies of *E. coli* HtpX (UP P23894), *S. cerevisiae* Oma1 (UP P36163), and *S. aureus* BlaR1 (UP P18357) were obtained using the TOPCONS program (99) and represented using the TOPO2 server. The topology of human FACE1 (UP O75844) was derived from the experimental coordinates (PDB 4AW6). The N- and C-terminal residues, as well as the flanking transmembrane residues, are labeled. FACE1 is located in the human endoplasmic reticulum and Golgi apparatus membrane, HtpX in the bacterial cell inner membrane, Oma1 in the mitochondrial inner membrane, and BlaR1 in the bacterial cell membrane. The zinc-binding signature HE*XX*H residues are shown in *black circles* as well as the third protein ligand of FACE1, a glutamate. The segments approximately corresponding to the catalytic domains are framed where known or predictable.

RESULTS AND DISCUSSION

Bioinformatics Searches Reveal Minimal Putative Metallopeptidases—We performed bioinformatics searches to identify orthologs of M48 and M56 IMMPs containing only the CD by using the sequence stretch of BlaR1/MecR1 from *S. aureus* (family M56) contained between two predicted TMHs and encompassing the HE*XX*H zinc-binding motif (see "Experimental Procedures" and Fig. 1). Three top hits were found, which were uncharacterized proteins from the thermophiles *A. aeolicus*, *P. abyssi*, and *M. jannaschii*. They contained the zinc-binding motif and were closely related, aligning with 28–31% pairwise sequence identity (Fig. 2*A*). With a size of 105–110 residues and the absence of predicted signal peptides and transmembrane segments, they would represent a "minimal" scaffold for a soluble MP. A profile-based sequence alignment against the MEROPS database revealed that these sequences were evolutionarily related to M56-IMMPs and also close to M48-IMMPs (Fig. 2*B*).

Proteolytic Assays Suggest Zymogenic Forms—The three thermophilic proteins were recombinantly overexpressed in *E. coli*. The proteins from *P. abyssi* (proabylysin) and *M. jannaschii* (projannalysin) proved to be soluble and were readily purified to homogeneity. Conversely, the ortholog from *A. aeolicus* was insoluble under all expression conditions tested, and no suitable refolding strategies could be found, and so it was discarded. Full-length proabylysin and projannalysin were correctly folded as assessed by size exclusion chromatography, which unambiguously revealed that they eluted as a monomer and dimer, respectively, at protein concentrations ranging from 0.1 to 2 mg/ml for the former and 0.1 to 40 mg/ml for the latter. Notably, proabylysin was stable for long periods only at a maximum concentration of ~0.6 mg/ml at 20 °C. Both proteins were inactive against a large battery of protein and peptide substrates (see "Experimental Procedures"). As *P. abyssi* and *M. jannaschii* are hyperthermophilic organisms whose growing temperature and pressure can reach 95 °C and 200 atmospheres, respectively, we also performed activity assays with these two proteins at high temperature (up to 80 °C) and/or in the presence of chaotropic agents (urea and guanidinium hydrochloride) to mimic the physiological conditions. How-

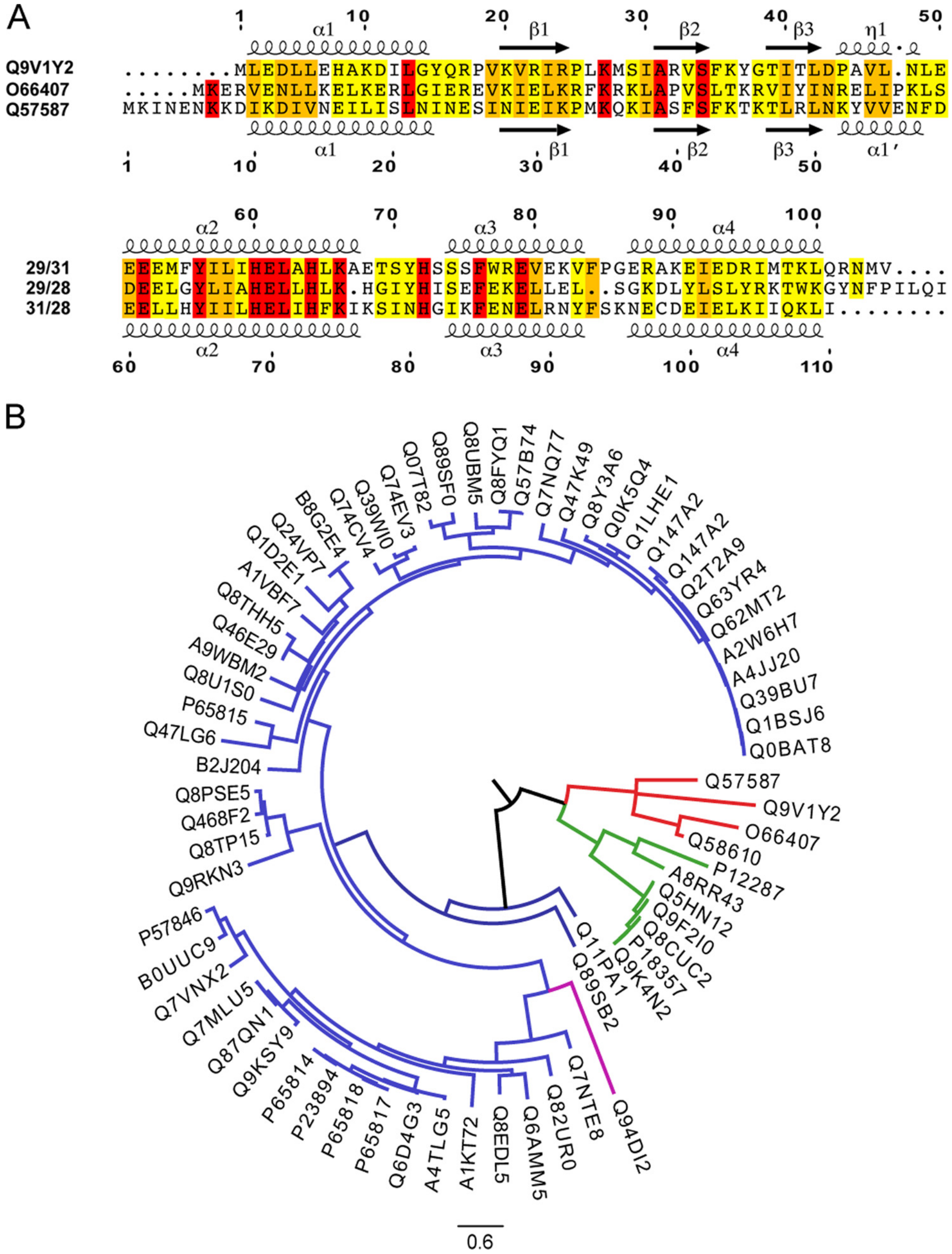
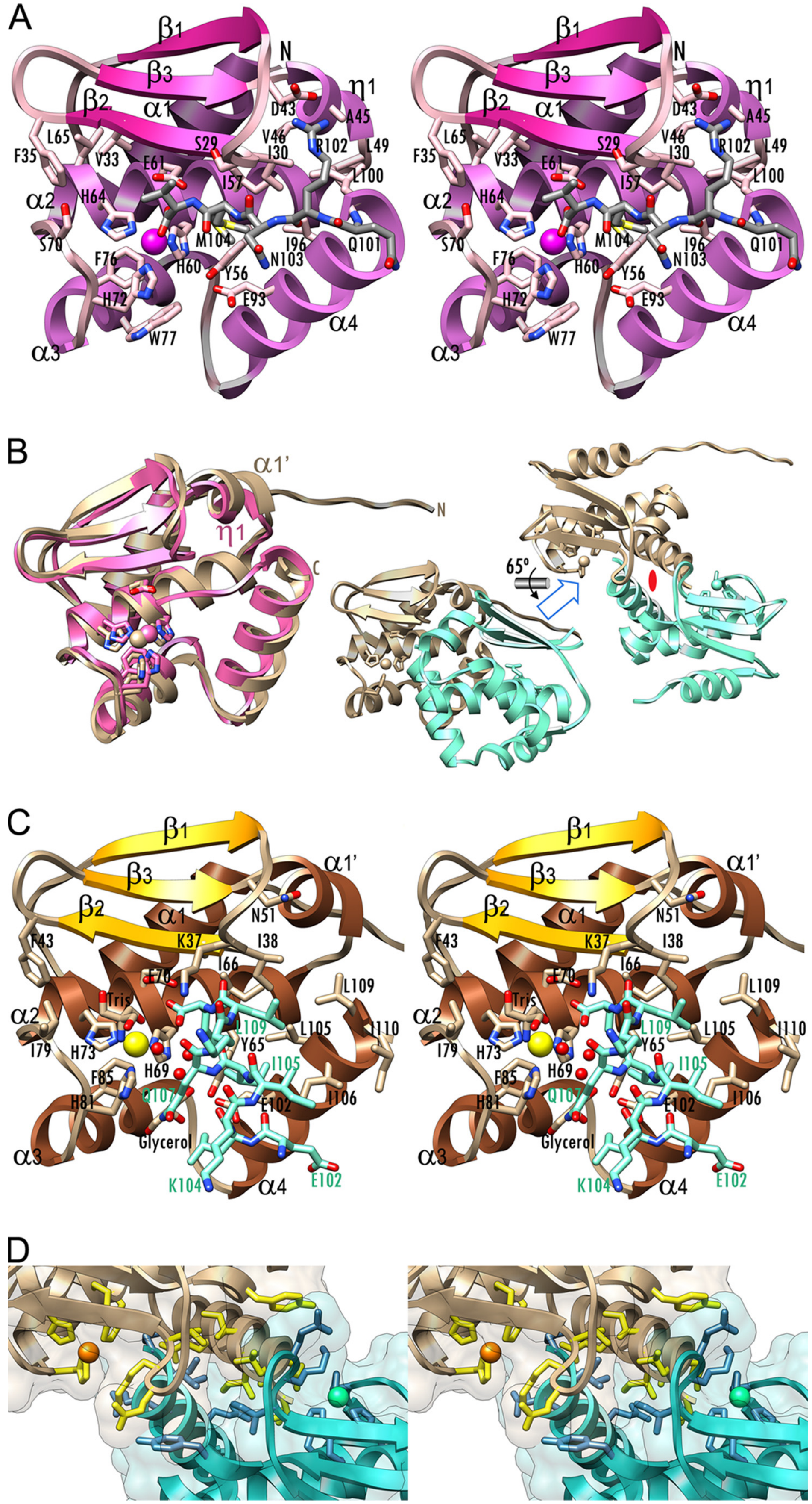


FIGURE 2. **Bioinformatics studies of the target proteins.** *A,* structure-based sequence alignment (from *top to bottom*) of *P. abyssi* UP Q9V1Y2, *A. aeolicus* UP O66407, and *M. jannaschii* UP Q57587 performed using the MULTALIN program (100) and represented using the ESPRIPT server. Residues with 100, 75, and 50% identity are highlighted through *red, orange,* and *yellow background*, respectively. The sequence identities of each protein (in %) with the other two proteins (*top to bottom*) are shown at the beginning of the second alignment block. Numbering and regular secondary structural elements correspond to *P. abyssi* UP Q9V1Y2 (*top*) and *M. jannaschii* UP Q57587 (*bottom*). *B,* sequence-based phylogenetic tree of the three sequences from *A* and 69 homologous sequences from MEROPS database represented by their UniProt codes. The tree shows three major groups, including HtpX-/HtpX-2-like proteins of family M48 (in *blue*), BlaR1/MecR1-, or PenR1-like proteins of family M56 (in *green*), an unassigned member of family M48 (in *magenta*), and the three sequences of *A* in *red*. The tree revealed another orthologous sequence from *M. jannaschii* (UP 58610) not analyzed in the present work. The *scale bar* represents PHYML branch length.

ever, we did not observe any activity for these proteins under these conditions either (data not shown), strongly suggesting that they were zymogens, as indicated by structural studies (see below).

Structure of Proabylysin Indicates a Novel Intramolecular Mechanism of Latency—The crystal structure of proabylysin was determined by single wavelength anomalous diffraction of the intrinsic zinc and refined with diffraction data to 1.15 Å resolution (see "Experimental Procedures"). The structure shows a compact globular and almost spherical shape ~35 Å in diameter (Fig. 3*A*). A shallow active-site cleft is carved into the front surface of the molecule when viewed in standard orientation (44) and divides it into an upper N-terminal subdomain (NTS; M^1–E^{68}) and a lower C-terminal subdomain (T^{69}–V^{105}). The NTS starts on the right top rear of the molecule in the form of an α-helix, hereafter referred to as the "backing helix" α1 (for nomenclature and extension of regular secondary structure elements, see Fig. 2*A*), which runs diagonally downward and ends in a loop connecting α1 with the first strand β1 (Lα1β1) of a front-twisted three-stranded β-sheet (β1–β3). The top strands


A
β1
β3
N
η1
β2
α1
D43
A45
L65
S29
V46
R102
L49
F35
V33
E61
I30
L100
α2
H64
I57
S70
M104
I96
Q101
F76
H60
Y56
N103
H72
E93
α4
α3
W77
B
α1'
N
η1
C
65°
C
β1
β3
α1'
β2
α1
N51
F43
K37
I38
I66
E70
L109
Tris
α2
L109
L105
I110
Y65
I79
H73
I105
H69
F85
I106
Q107
H81
E102
α3
Glycerol
K104
α4
E102
D

(β1 and β3) are parallel, and the lowermost (β2), which creates an "upper rim" of the active-site cleft, is antiparallel to the cleft and thus to any substrate bound to it. Preceding the latter strand is the "bulge-edge segment" (L^{26}–I^{30}), which contributes to shaping the top of the cleft on its primed side (for cleft and substrate subsite nomenclature, see Refs. 44, 76). After β3, a short 3_{10}-helix (η1; the "linking helix") at the right top of the molecule links the sheet with the "active-site helix" α2, which contains the short zinc-binding consensus sequence ^{60}HE*XX*H^{64} (see below). The helices and the sheet of the NTS contribute through their inner surfaces to an internal hydrophobic core, which traverses the entire protein moiety and glues the structure together. After α2, the polypeptide chain enters the C-terminal subdomain through Lα2α3, which leads to the "glutamate helix" α3. The latter is termed thus due to its topological equivalence with a similar helix in gluzincins (see below), and it lies in a horizontal plane that is roughly parallel to that of helix α2 but with the helix axis vertically rotated backwards by ~50°, so that α3 ends with V^{83} at the back molecular surface. At F^{84}, the chain turns frontward and, after a loop that forms the bottom of the molecule (P^{85}–R^{86}), opens out into the "C-terminal helix" α4, which runs obliquely and creates the lower right part of the molecule (Fig. 3*A*). The last residue of α4, L^{100}, resides in a hydrophobic pocket created by the side chains of A^{45}, I^{30}, I^{57}, and I^{96}, which feature the rightmost end of the internal hydrophobic core of the NTS (Fig. 3*A*). From Q^{101} onward, the polypeptide chain enters the primed side of the active-site cleft in the extended conformation of, but reverse orientation to, a true peptidic substrate and reaches the zinc site with the C-terminal carboxylate of V^{105}. This C-terminal tail establishes the inter-main chain hydrogen bonds with residues from the bulge edge segment and the start of upper rim strand β2 (N^{103}O-N^{30}N, 2.87 Å; V^{105}N-A^{31}O, 2.92 Å).

The catalytic zinc ion resides at the bottom of the active-site crevice, somewhat displaced from halfway toward the nonprimed side of the cleft (Fig. 3*A*). The metal is coordinated in a tetrahedral fashion by the Nϵ2 atoms of the two histidines of the consensus sequence, H^{60} (2.02 Å apart) and H^{64} (2.00 Å), one of the two C-terminal carboxylate oxygens of V^{105} (2.02 Å), and the Nϵ2 atom of H^{72} (2.02 Å apart), which is provided by Lα3α4 two positions ahead of the start of the glutamate helix α4. The other carboxylate oxygen of V^{105} is further way from the metal (2.69 Å) and in a geometrically unfavorable conformation for binding (Fig. 3*A*). By contrast, this oxygen contacts in a distorted bidentate manner the two carboxylate oxygens of the general base/acid glutamate E^{61} (2.68 and 3.28 Å apart), thus indicating that either of the carboxylate groups must be in protonated state.

The insertion of the C-terminal tail also provides information on the distinct subsites on the primed side of the cleft. M^{104} nestles into what is most likely the $\mathbf{S_1{}'}$ subsite of the cleft (cleft and substrate subsites in boldface type), which is usually the major determinant for specificity in endolytic MPs (44). The methionine side chain is present in double conformation, and the surrounding residues give rise to a site that is wide and hydrophobic, thus suggesting that mature abylysin could easily accommodate bulky hydrophobic residues in $\mathbf{P_1{}'}$, as observed in the otherwise unrelated matrix metalloproteinases, snapalysins, and ADAMs/adamalysins, which are metzincins (77–79). The $\mathbf{S_1{}'}$ pocket is framed at its back by residues from the start of the active-site helix (Y^{56}, I^{57}, and H^{60}), and at its right and bottom by residues from the C-terminal helix (E^{93} and I^{96}). The antepenultimate residue of the tail, N^{103}, possibly at $\mathbf{P_2{}'}$, points toward the bulk solvent, suggesting that this site is not relevant for specificity as long as the extended conformation of the substrate is maintained. R^{102}, in turn, probably occupies $\mathbf{S_3{}'}$, which would be shaped by the side chains of I^{30}, A^{45}, and L^{100}, and the main chain of the bulge-edge segment at P^{25}–I^{30}. The guanidinium group of R^{102} establishes two hydrogen bonds with the main chain carbonyl oxygens of L^{26} and S^{29} (2.50 and 2.92 Å apart, respectively), and a double salt bridge with D^{43} at the end of Lβ3η1. The R^{102} Nη1 and Nη2 atoms are, respectively, 3.01 and 3.02 Å apart from the carboxylate oxygens of the aspartate. As found for M^{104} in $\mathbf{S_1{}'}$, R^{102} fits neatly into the potential $\mathbf{S_3{}'}$ site, suggesting that mature abylysin could have a preference for arginine in $\mathbf{P_3{}'}$. A look at the putative nonprimed side of the cleft, in turn, suggests that V^{105} is in $\mathbf{S_1}$. It has its side chain pointing toward the bulk solvent, thus suggesting that any residue of a potential substrate could be found in $\mathbf{P_1}$. Upstream of $\mathbf{P_1}$, a substrate could be accommodated in a rather shallow $\mathbf{S_2}$ or $\mathbf{S_3}$ site, putatively framed by the side chains of V^{33}, F^{35}, H^{64}, L^{65}, and S^{70} (Fig. 3*A*).

Overall, the structure of the globular part of proabylysin spanning M^{1}–L^{100} indicates a competent MP moiety, active-site cleft, and metal-binding site, including the pivotal glutamate general base/acid required for catalysis, as found in most active enzymes of this class (see also below). By contrast, the insertion of the C-terminal tail into the active-site cleft in the reverse orientation of a substrate, thus blocking structural elements essential for catalysis and access of true substrates, provides a structural explanation for the lack of hydrolytic activity of the protein as it would actually correspond to a zymogen.

FIGURE 3. **Structures of proabylysin and projannalysin.** *A,* cross-eye stereo ribbon plot of proabylysin in standard orientation (44), with coils, strands (labeled β1–β3, see also Fig. 2*A*), and helices (α1–α4 and η1) in *pale pink, magenta*, and *pale purple*, respectively. Selected residues are shown for their side chains as *sticks* with *pink* carbons and labeled, as is the N terminus of the protein. The C-terminal segment running across the cleft (Q^{101}–V^{105}) is depicted as a *stick model* with *gray carbons*. *B,* superposition of the proabylysin (*pink*) and projannalysin (*tan*) monomers, whose termini are labeled. For both structures, the metal-binding residues, the general base/acid glutamate, and the Ser/Gly turn phenylalanine are depicted as *sticks* with carbons colored as the respective ribbons (*left panel*). Structure of the projannalysin dimer is in standard orientation (*i.e.* as in *A*), with one monomer in *tan* and the other in *aquamarine*, and after a horizontal 65° rotation, which depicts the view along the noncrystallographic 2-fold axis shown as a *red ellipse* (*right panel*). *C,* cross-eye stereo ribbon plot of projannalysin in standard orientation, with coils, strands (labeled β1–β3, see also Fig. 2*A*), and helices (α1, α1′, and α2–α4) in *tan, orange,* and *brown,* respectively. Selected residues, as well as a Tris molecule coordinating the metal (*yellow sphere*) and a glycerol at the dimer interface, are shown as *sticks* with *tan* carbons and labeled in *black*. Three waters are shown as *red spheres*. The C-terminal segment (E^{102}–I^{110}) of the symmetry-related molecule in the dimer is shown as a *stick model* with carbons and labels in *turquoise*. *D,* close-up view in cross-eye stereo of the two projannalysin monomers displayed in Fig. 3*B* (*right panel*) superposed with the respective semi-transparent Connolly surfaces. The side chains of the residues involved in dimerization are shown as *yellow* and *blue sticks*, respectively, for molecules *A* and *B*.

The latency mode has been previously unseen for structurally characterized MPs, which are normally kept latent by N-terminal pro-domains or propeptides, as in astacin (80), meprin β (81), fragilysin (52), funnelin carboxypeptidases (82), matrix metalloproteinases (77), ADAMs/adamalysins (78), and thermolysin (83). Accordingly, the present structure represents a novel mechanism of intramolecular latency maintenance in MPs.

Structure of Projannalysin Suggests a Novel Intermolecular Mechanism of Latency—The crystal structure of projannalysin was determined by single wavelength anomalous diffraction of a mercury derivative and refined with diffraction data to 2.0 Å resolution (see "Experimental Procedures"). In one of the two molecules found in the asymmetric unit of the crystal, molecule A, the N-terminal stretch of projannalysin protrudes away from the molecular moiety and is defined from the first residue of the construct, G^{-2}, to N^4, due to an interaction with a crystallographic neighbor (Fig. 3*B*, *left*, *tan*). The final electron density is somewhat weaker at E^5 and N^6, but it becomes fully defined again from K^7 onward, which is the first residue of the globular moiety. By contrast, in molecule B the polypeptide chain is defined from K^7 onward only (Fig. 3*B*, *right*, *turquoise*).

In accordance with a 31% sequence identity (see Fig. 2*A*), projannalysin and proabylysin exhibit a very similar overall shape and practically overlapping chains from the beginning of helix $\alpha1$ to the end of helix $\alpha4$ (M^1–L^{100} of proabylysin and D^9–L^{109} of projannalysin; *Z* value according to the DALI program of 16.0; r.m.s.d. of 1.5 Å). All the regular secondary structural elements are conserved with the sole exception of proabylysin 3_{10}-helix $\eta1$, which has an extra residue in projannalysin and becomes here an α-helix (termed $\alpha1'$, K^{52}–N^{57}; see Figs. 2*A* and 3, *B* and *C*). After L^{109}, the last residue of projannalysin projects away from the molecular body, pointing in the opposite direction to the proabylysin C-terminal tail and becoming engaged in binding of a neighboring molecule (see below and Fig. 3, *B* and *C*).

Consistent with the dimeric behavior observed in size exclusion chromatography, projannalysin is also a dimer in the crystal structure (Fig. 3*B*, *right*). Two molecules interact through an interaction surface of ~1,980 Å², which is above the range generally described for protein-protein complexes (1,250–1,750 Å² (84)), and indicates that such a dimer is likely to occur *in vivo*. The interaction surface shows a surface complementarity (= 0.67) that lies within the range reported for antibody/antigen interfaces (0.64–0.68 (70)). The interaction is nearly symmetric and results from 32 intermolecular contacts (<4 Å), among them nine hydrogen bonds and hydrophobic interactions between seven molecule A and six molecule B residues (Fig. 3*D*). Participating segments include L$\beta1\beta2$, L$\beta3\alpha1'$, $\alpha2$, L$\alpha2\alpha3$, and $\alpha4$ of either molecule (Fig. 3*D*), and the catalytic metal ions are 24 Å apart. Most noteworthy is the crosswise interaction between the active site of one molecule and the C-terminal helix $\alpha4$ of the other (Fig. 3, *B* and *C*). The latter blocks access to the cleft and thus provides the structural explanation for the lack of hydrolytic activity of projannalysin. Like proabylysin, it is a zymogen, but by contrast latency is exerted crosswise intermolecularly by the swap of a C-terminal α-helix. This also entails significant differences in the zinc-binding site. In projannalysin, the metal is surrounded by a distorted octahedral hexa-coordination sphere, which is rather unusual for zinc according to MESPEUS database (see Table 2 in Ref. 85). Protein ligands equivalent to proabylysin are the Nϵ2 atoms of H^{69} (2.1/2.0 Å apart from the metal in molecules A/B, respectively) and H^{73} (2.1/2.1 Å) of the zinc-binding sequence, and H^{81} of L$\alpha2\alpha3$ (2.1/2.1 Å). In addition, the metal is distinctly bound by a solvent molecule (2.2/2.1 Å apart) and, in molecule A, a Tris molecule through its amino group (2.2 Å) and one of its hydroxyl groups (1.9 Å) in a bidentate fashion. In molecule B, this ligand is replaced by a glycerol, two of whose hydroxyl oxygens are 2.1 and 2.2 Å apart from the metal, respectively. The solvent molecule and H^{73} Nϵ2 are in apical positions and the other four ligands in a plane with the metal (Fig. 3*C*). Accordingly, and in contrast to proabylysin, the two carboxylate oxygens of the C terminus (here, I^{110} of the neighboring molecule) are far from the zinc ion (4.1–4.2 Å). Instead, they contact the carboxylate of the general base/acid glutamate, E^{70} (2.6/2.5 Å in molecule A/B, respectively) *plus* the zinc-binding hydroxyl of the Tris (glycerol) molecule (both 2.6 Å apart), and the zinc-bound water (2.7/2.6 Å), respectively. This architecture indicates that either of the carboxylate groups of E^{70} or L^{109} must be in protonated state, as found in proabylysin (see above). The structure of the dimeric zymogenic complex further reveals that the $\mathbf{S_1}'$ pocket is partially occupied by the side chain of the last residue, I^{110}, instead of the penultimate residue as seen in proabylysin. The pocket is very similar in depth and hydrophobicity in both zymogens and framed in projannalysin by A^{39} (A^{31} in proabylysin), I^{38} (I^{30}), I^{66} (I^{58}), Y^{65} (Y^{56}), I^{105} (I^{96}), L^{109} (L^{110}), and E^{102} (E^{93}).

As for proabylysin, the globular part of projannalysin spanning D^9–L^{109} indicates a competent fold of catalytic moiety, active-site cleft, metal-binding site, and general base/acid glutamate, as usual for active MPs (see also below). By contrast, the insertion of the C-terminal part of helix $\alpha4$ into the active-site of a neighbor contributes to another novel C-terminal latency mechanism for this type of enzymes, which is homodimeric and occurs crosswise.

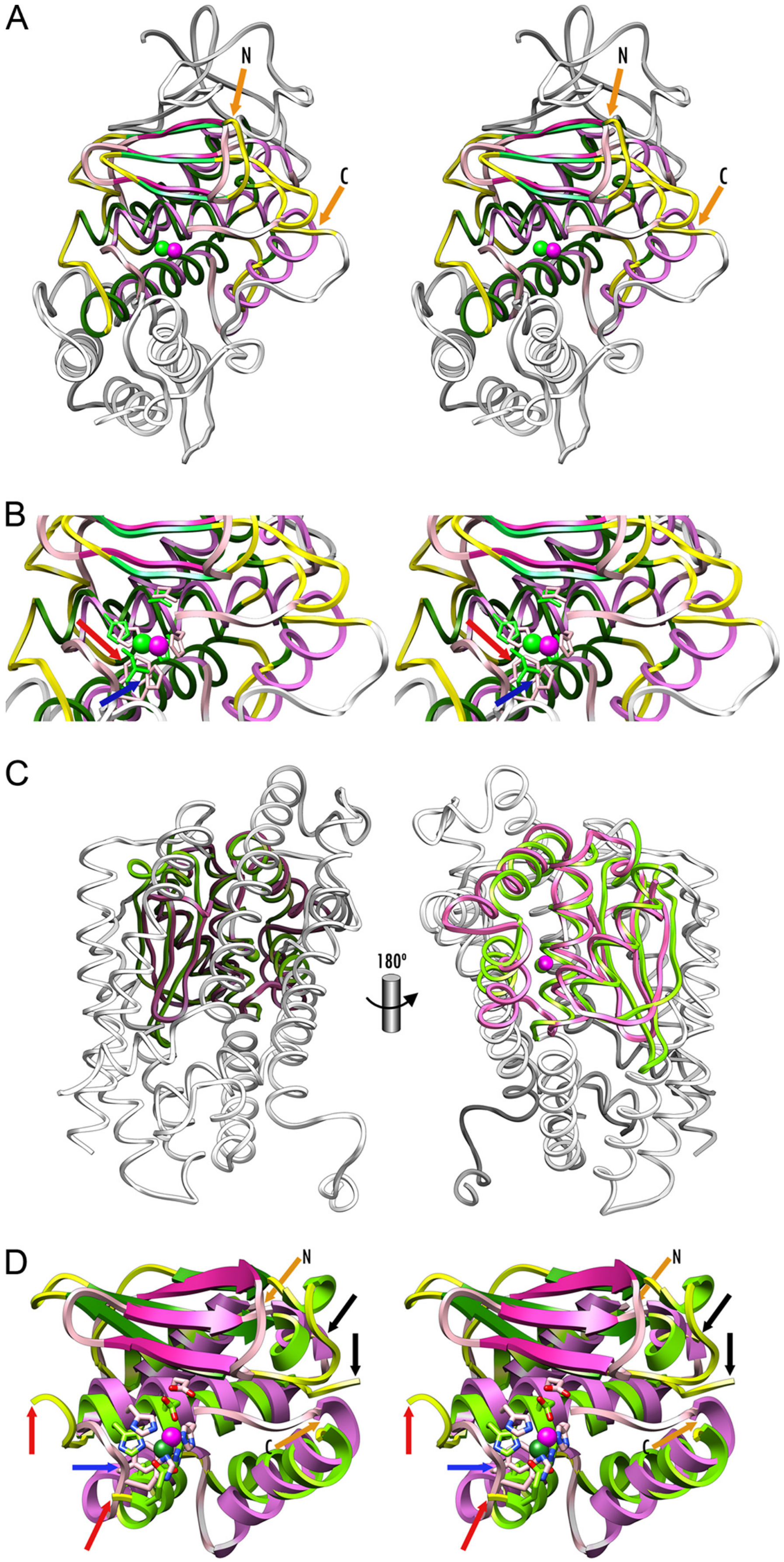
Activation Trials Point to an Intricate Mechanism of Peptidase Activation—To provide insight into the mechanism of activation of these MPs, we attempted to subject them to limited proteolysis to remove the zymogenic segments by assaying them with a collection of standard serine proteinases and metalloproteinases, which only yielded intact proteins or resulted in complete degradation (data not shown). We thus designed a range of constructs lacking C-terminal residues of proabylysin and projannalysin. Proabylysin constructs in which the last five residues were sequentially deleted showed dramatically decreasing expression and solubility levels when compared with the wild-type protein, thus precluding purification and/or functional assays for most of the constructs. Only those lacking the last one and two residues, respectively, could be purified. However, they showed a significantly decreased melting temperature (T_m) in a thermofluor assay (85 and 65 °C, respectively; see "Experimental Procedures") when compared with that of the wild type (95 °C), which is indicative of structurally labile and inactive forms. Assays of fusion constructs of C-terminal truncations with N-terminal GST or streptavidin tag, which rendered only soluble aggregates as assessed by size exclusion chromatography (data not shown), were also unsuccessful. Similar negative results were obtained with a projannalysin mutant lacking the last five residues (data not shown), which theoretically should avoid inhibition but not affect the overall protein moiety as inferred from the crystal structure (see above). In conclusion, we failed to obtain active abylysin and jannalysin. This strongly suggests that both proteins likely need the assistance of a specific protease for activation and/or a conformational change that is brought about in the intracellular living environment but that we were unable to reproduce *in vitro*. Nevertheless, the possibility that these proteins are inactive ancestors of metallopeptidases cannot be ruled out.

Proabylysin and Projannalysin Are Structurally Related to Gluzincins—Thermolysin from *Bacillus thermoproteolyticus* was the first metalloendopeptidase structurally characterized, back in 1972 (86), and is the founding member of the gluzincins (see Refs. 41, 42 and also see Table 2 in Ref. 44). Superposition of both proabylysin and projannalysin onto thermolysin reveals that the polypeptide chains are very similar along almost the entire structure of the former (M^1–K^{99}) and V71–P195 of the latter (Fig. 4, *A* and *B*), including the position of the N and C termini of the superposed parts and all regular secondary structure elements, with the exception of the linking and C-terminal helices, which are replaced by loops in the *Bacillus* enzyme. This similarity is reflected by a DALI *Z* value of 3.5, an r.m.s.d. of 2.9 Å, 83 topologically aligned residues but only 8% sequence identity. In particular, the active-site and glutamate helices overlap despite being, respectively, one and two turns shorter in the present structures (Fig. 4*A*; see also preceding sections). Overall, these two helices are the major characteristic features of gluzincins and provide the three protein zinc ligands, usually two histidines and a glutamate (H142, H146, and E166 in thermolysin). This shortening of the present structures entails that the third protein zinc ligand, a histidine, is found in the loop preceding the glutamate helix. Nevertheless, the position of this third ligand is practically the same, and the Cα atoms of proabylysin H^{72} and thermolysin E166 are only 1.4 Å apart (Fig. 4*B*). In addition, the three-stranded β-sheets and the backing helices fit well on top of each other in both structures, both in sequence, position, and orientation, although they are linked by longer loops in thermolysin (Fig. 4*A*). A further feature of gluzincins is the “Ser/Gly turn” within the glutamate helix, downstream of the metal-binding glutamate, which contains a residue, usually a serine or a glycine, whose side chain is found immediately below the metal-binding site (82). This structural element is reminiscent of the “Met turn,” which has a strictly conserved methionine and is characteristic of metzincins (51). Such a Ser/Gly turn is also present in proabylysin and projannalysin, centered on a phenylalanine (F^{76} and F^{85}, respectively), which is also found in the *A. aeolicus* protein (Fig. 2*A*). In general, structurally characterized gluzincins are much larger than the minimal core found in proabylysin and projannalysin; thermolysin spans 316 residues and cowrin metallocarboxypeptidases, which also show the aforementioned structural similarity, span ~500–700 (82, 86). In particular, both thermolysins and cowrins possess a downstream “tyrosine helix” below the glutamate helix (82).

Another structurally characterized gluzincin family consists of leukotriene A4 hydrolase (PDB 3B7S (87)) and related metalloaminopeptidases such as cold-active aminopeptidase (PDB 3CIA (88)), endoplasmic reticulum aminopeptidase 1 (PDB 3QNF and 3MDJ (89, 90)), tricorn interacting factor F3 (PDB 1Z5H (91)), M1 alanyl aminopeptidase (PDB 3EBI (92)), and aminopeptidase N (PDB 2HPO (93)). Like cowrins (82), these are large multidomain enzymes spanning ~600–950 residues that contain an inserted CD bearing a greater resemblance to proabylysin and projannalysin than that of thermolysin (DALI *Z* values between 7.2 and 6.4). They even include the linking and C-terminal helices, although the latter is rotated frontward by ~45–90°. This family has a glutamate as the third zinc ligand and a threonine or alanine in the Ser/Gly turn.

Overall, the close general similarity of proabylysin and projannalysin to distinct active gluzincin families strongly suggests that the globular parts of these two zymogens are in an overall competent conformation (see also preceding sections) and that latency is exerted through C-terminal segments by sterically blocking access to preformed CDs. Accordingly, the thermophilic proteins studied here would represent a minimal structural and functional core of the gluzincin clan, and as such, we propose the family name “minigluzincins” for them. The fact that minigluzincins have a histidine instead of a glutamate as third zinc ligands and a phenylalanine in the Ser/Gly turn reflects variability among gluzincins that is also observed in metzincins, which generally have a histidine as the third zinc ligand, although snapalysins and, possibly, thuringilysins have an aspartate (94).

Minigluzincins Are Models for Catalytic Domains of M48 and M56 Metallopeptidases—Despite the obvious structural similarity with the aforementioned gluzincins, automated structure-similarity searches of proabylysin and projannalysin against the PDB identified three other more closely related structures: a putative peptidase from *Geobacter sulfurreducens* (PDB 3C37; DALI *Z* score 7.4; r.m.s.d. 3.1 Å; 98 aligned residues; 13% sequence identity); a fragment of HtpX from *Vibrio parahemeolyticus* (PDB 3CQB; DALI *Z* score 6.8; r.m.s.d. 2.1 Å; 71 aligned residues; 15% sequence identity); and human FACE1 (PDB 4AW6; (40); DALI *Z* score 4.6; r.m.s.d. 2.8 Å; 96 aligned residues; 18% sequence identity; see Fig. 4, *C* and *D*). The latter is very similar to its yeast ortholog Ste24p (PDB 4IL3 (39)). PDB entry 3C37 was deposited with the PDB by a structural genomics consortium but is not yet published. It corresponds to a soluble 253-residue protein tagged by the depositors as an M48 family peptidase, although it is not an IMMP. It is a nonfunctional protein whose putative catalytic metal ion is bound in a canonical manner by three residues provided by an active-site and a glutamate helix and also by a fourth histidine ligand coming from a second protein chain, so that the cleft is blocked. In the absence of further data and a publication, crystallization artifacts cannot be ruled out in this case. PDB entry 3CQB, in turn, is an artificial fragment between two TMHs of a *Vibrio* relative of M48 family member HtpX from *E. coli*, which was likewise deposited by a structural genomics consortium and is not yet published. It lacks a functional active-site cleft and metal-binding site, and it could potentially correspond to the NTS


A
N
C
N
C
B
C
180°
D
N
C
N
C

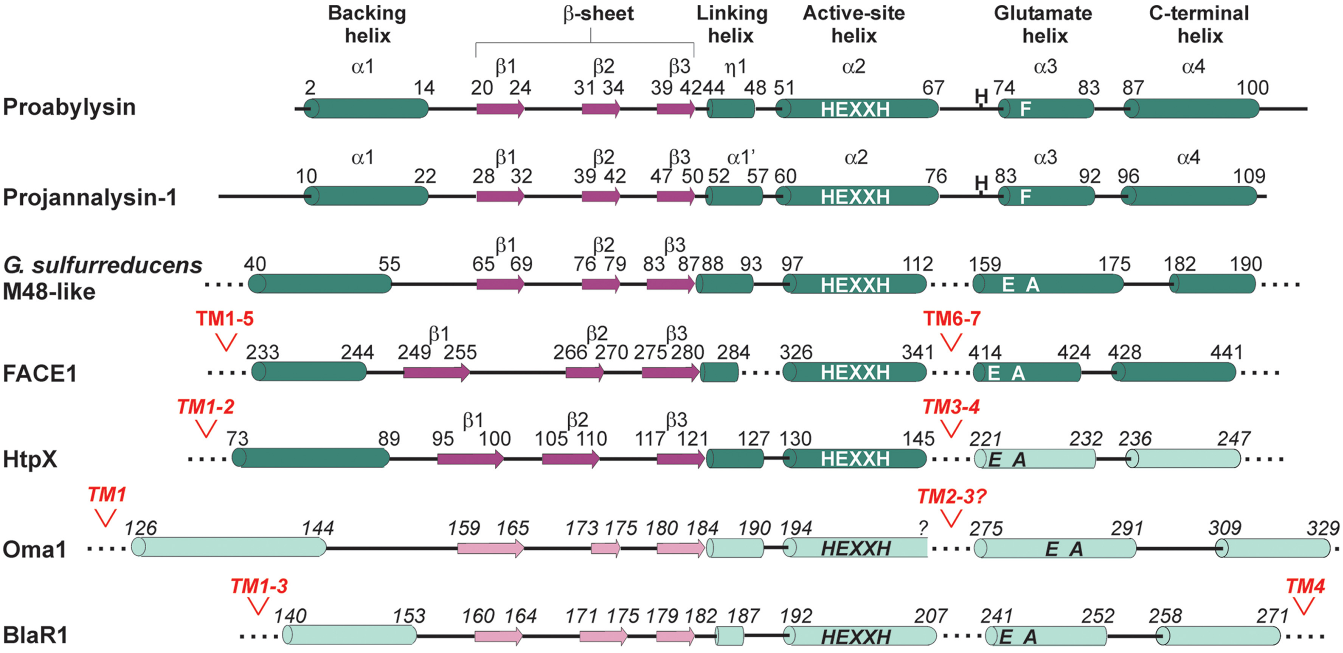


FIGURE 5. **Common topology of minigluzincins and selected M48 and M56 metallopeptidases.** The actual regular secondary structural elements of the experimental structures of proabylysin (UP Q9V1Y2; this work), projannalysin (UP Q57587; this work), *G. sulfurreducens* M48-like peptidase (UP Q74D82; PDB 3C37), human FACE1 (UP O75844; PDB 4AW6), and a truncated form of *V. parahemeolyticus* HtpX containing only its N-terminal subdomain (UP Q87QN1; PDB 3CQB) are represented with *dark green rods* (helices) and *dark purple arrows* (strands) with the residues flanking each secondary structure element. In addition, the predicted regular secondary structure elements of *E. coli* HtpX C-terminal subdomain (UP P23894), yeast Oma1 (UP P36163), and *S. aureus* BlaR1 (UP P18357) are represented with *light green rods* (helices) and *light purple arrows* (strands). The protein numbering, the HEXXH motif, the third zinc-binding residue, the Ser/Gly turn residue, and the transmembrane segments (*TM*) taken either from the experimental structures or the predictions are indicated by either *regular* or *italic letters*, respectively.

of the CD of this M48 family member. Finally, PDB entries 4AW6 and 4IL3 correspond to human and yeast FACE1/Ste34p, the only two true IMMPs currently available for their three-dimensional structures in addition to site-2 protease MP from *M. jannaschii* (36). The latter belongs to family M50, is not a gluzincin (see Fig. 2 in Ref. 44), and is only very distantly related to proabylysin (PDB 3B4R; DALI *Z* score 1.2; r.m.s.d. 3.5; 51 aligned residues; 10% sequence identity) and its structural relatives. Human and yeast FACE1/Ste24p represent the only M48-IMMPs structurally characterized to date. Further entries that are similar to minigluzincins are two closely related hypothetical proteins reported by structural genomics consortia: ybeY from *E. coli* (1XM5 (95)) and AQ-1354 from *A. aeolicus* (PDB 1OZ9 (96)), for which no function has been described. These are single domain proteins of ~150 residues that contain a third metal-binding histidine and a methionine and phenylalanine, respectively, in the Ser/Gly turn. They possess a linking helix but, despite rather large DALI *Z* scores when compared with proabylysin (6.9 and 7.6, respectively), they lack the C-terminal helix of minigluzincins and have additional regular secondary structure elements inserted into the basic scaffold.

Superposition of proabylysin and projannalysin onto the putative peptidase of *G. sulfurreducens*, human FACE1, and truncated HtpX reveals that all the common regular secondary structural elements shared by minigluzincins with thermolysin and other gluzincins are also found in the aforementioned M48-related structures. Taken together, this ascribes M48-IMMPs to the gluzincin clan of MPs. In addition, the closer structural similarity of minigluzincins to the aforementioned three M48-like structures than to other gluzincins (see *Z* scores

FIGURE 4. **Structural similarities of minigluzincins.** *A*, cross-eye stereo plot depicting the superposition of proabylysin and *B. thermoproteolyticus* thermolysin (PDB 4TLN (101)) as ribbons colored in *pale pink, magenta,* and *pale lilac* (proabylysin) and *yellow, light green*, and *dark green* (thermolysin) for the coils, strands, and helices of the common stretches, respectively, whose ends are highlighted by *orange arrows* (M^1–K^{99} of proabylysin and V71–P195 of thermolysin). The orientation corresponds to that of the standard orientation of MPs (44) as in Fig. 4*A*. The corresponding zinc ions are shown as *magenta* and *green spheres*, respectively. The remaining 70-residue N-terminal and 121-residue C-terminal stretches of thermolysin are shown in *white*. The characteristic calcium ions of thermolysin have been omitted for clarity. *B*, close-up view of *A* in cross-eye stereo depicting the metal-binding sites of proabylysin and thermolysin. The residues of the HEXXH sequence are shown for their side chains (in *pink* for proabylysin (H^{60}, E^{61}, and H^{64}) and in *green* for thermolysin (H142, E143, and H146)). In addition, the respective downstream third metal-binding residues (H^{72} of proabylysin and E166 of thermolysin; pinpointed by a *blue arrow*) and the residue of the Ser/Gly turn (F^{76} of proabylysin and S169 of thermolysin; pinpointed by a *red arrow*) are depicted. *C*, superposition of proabylysin and M48 family human FACE1 protein (PDB 4AW6) in two views resulting from a 90° counterclockwise in-plane rotation of the orientation in *A* (*left panel*) and a subsequent vertical 180° rotation (*right panel*). The structurally equivalent parts and the catalytic metal ions are displayed in *hot pink* (proabylysin) and *green* (FACE1), and the remaining parts of FACE1 are shown as a *white coil*. *D*, close-up view in cross-eye stereo of (*C, left*) after a 90° clockwise in-plane rotation, *i.e.* in standard orientation, showing only the common parts (M^1–Q^{101} of proabylysin and K233-G343 + E415-N442 of FACE1). Coils, strands, and helices are shown, respectively, in *pink, magenta,* and *hot pink* (proabylysin) and in *yellow, green,* and *light green* (FACE1). The catalytic metal, the residues engaged in metal binding, and the general base/acid glutamate of each structure (H^{60}, E^{61}, H^{64}, and H^{72} of proabylysin and H335, E336, H339, and E415 of FACE1) are shown for their side chains with *pink* and *green carbons*, respectively. *Orange arrows* pinpoint the N and C termini of the common structure parts. *Red arrows* pinpoint the positions between the second and third zinc ligands, at which FACE1 has two of its TMHs inserted (segment H344-F414). These are actually a continuation of the active-site and glutamate helices. *Black arrows* indicate the residues flanking the disordered segment in FACE1 (E286-K321), which coincides with proabylysin linking helix η1. A *blue arrow* highlights the position of the Ser/Gly turn residue (F^{76} in proabylysin and A418 in FACE1).

above) is reflected by the fact that the linking helix (in PDB 3C37, 3CQB, and 4AW6) and the C-terminal helix (in PDB 3C37 and 4AW6), absent in thermolysin and cowrins, are found in both groups (see Figs. 4, *C* and *D,* and 5). Furthermore, in the absence of structural information on the IMMP part of M56 proteins, independent secondary-structure prediction of full-length BlaR1/MecR1, together with predictions for M48 HtpX and Oma1 and the experimental structures of the two minigluzincin zymogens and their three structural relatives, reveals that the CDs are most likely very similar in M48- and M56-IMMPs. This is further supported by the fact that minigluzincins were actually found in a search with BlaR1/MecR1 sequences and by evolutionary studies (see above), which indicate that minigluzincins are even closer to M56- than to M48-IMMPs (Fig. 2*B*).

Conclusions—This work describes a novel family of minimal ~100-residue soluble MPs, the minigluzincins, which evince unique zymogenic structures that are maintained in their inactive state by intra- or intermolecular C-terminal segments. The minigluzincins share very similar chain traces with thermolysin and other gluzincins such as leukotriene A4 relatives and cowrins, but they bear even higher similarity with M48 and M56 IMMPs, thus representing a minimal scaffold for the CDs of a large cohort of integral-membrane proteins. The characterizing feature of the M48 and M56 family CDs is a compact domain containing a backing helix, a three-stranded β-sheet, a linking helix, an active-site helix providing two metal ligands, a glutamate helix providing a third metal ligand, and a Ser/Gly turn, as well as a C-terminal helix (Fig. 5). Into this minimal common scaffold, each specific family member has inserted large upstream, internal, and downstream structural elements, including TMHs, to yield the specific functionality as membrane-embedded MPs. Altogether, structural knowledge of minigluzincins may be helpful in the construction of shorter soluble variants of IMMPs and thus facilitate the design of drugs to modulate their function.

Acknowledgments—We are indebted to Tibisay Guevara for excellent assistance during crystallization experiments. We also thank the Automated Crystallography Platform at IBMB/IRB. We acknowledge the help provided by local contacts at the ESRF Synchrotron (Grenoble, France). Funding for traveling and data collection was provided in part by ESRF.

REFERENCES

1. Wallin, E., and von Heijne, G. (1998) Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.* **7,** 1029–1038
2. Sonoda, Y., Newstead, S., Hu, N. J., Alguel, Y., Nji, E., Beis, K., Yashiro, S., Lee, C., Leung, J., Cameron, A. D., Byrne, B., Iwata, S., and Drew, D. (2011) Benchmarking membrane protein detergent stability for improving throughput of high resolution x-ray structures. *Structure* **19,** 17–25
3. Almén, M. S., Nordström, K. J., Fredriksson, R., and Schiöth, H. B. (2009) Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC Biol.* **7,** 50
4. Kim, H., Melén, K., Osterberg, M., and von Heijne, G. (2006) A global topology map of the *Saccharomyces cerevisiae* membrane proteome. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 11142–11147
5. Terstappen, G. C., and Reggiani, A. (2001) *In silico* research in drug discovery. *Trends Pharmacol. Sci.* **22,** 23–26
6. Davey, J. (2004) G-protein-coupled receptors: new approaches to maximise the impact of GPCRS in drug discovery. *Expert Opin. Ther. Targets* **8,** 165–170
7. Rawlings, N. D., Barrett, A. J., and Bateman, A. (2012) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* **40,** D343–D350
8. Tusnády, G. E., Dosztányi, Z., and Simon, I. (2004) Transmembrane proteins in the Protein Data Bank: identification and classification. *Bioinformatics* **20,** 2964–2972
9. Fujimura-Kamada, K., Nouvet, F. J., and Michaelis, S. (1997) A novel membrane-associated metalloprotease, Ste24p, is required for the first step of NH_2-terminal processing of the yeast α-factor precursor. *J. Cell Biol.* **136,** 271–285
10. Freije, J. M., Blay, P., Pendás, A. M., Cadiñanos, J., Crespo, P., and López-Otín, C. (1999) Identification and chromosomal location of two human genes encoding enzymes potentially involved in proteolytic maturation of farnesylated proteins. *Genomics* **58,** 270–280
11. Barrowman, J., and Michaelis, S. (2009) ZMPSTE24, an integral membrane zinc metalloprotease with a connection to progeroid disorders. *Biol. Chem.* **390,** 761–773
12. Varela, I., Pereira, S., Ugalde, A. P., Navarro, C. L., Suárez, M. F., Cau, P., Cadiñanos, J., Osorio, F. G., Foray, N., Cobo, J., de Carlos, F., Lévy, N., Freije, J. M., and López-Otín, C. (2008) Combined treatment with statins and aminobisphosphonates extends longevity in a mouse model of human premature aging. *Nat. Med.* **14,** 767–772
13. Freije, J. M., and López-Otín, C. (2012) Reprogramming aging and progeria. *Curr. Opin. Cell Biol.* **24,** 757–764
14. Akiyama, Y. (2009) Quality control of cytoplasmic membrane proteins in *Escherichia coli*. *J. Biochem.* **146,** 449–454
15. Bieniossek, C., Schalch, T., Bumann, M., Meister, M., Meier, R., and Baumann, U. (2006) The molecular architecture of the metalloprotease FtsH. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 3066–3071
16. Sakoh, M., Ito, K., and Akiyama, Y. (2005) Proteolytic activity of HtpX, a membrane-bound and stress-controlled protease from *Escherichia coli*. *J. Biol. Chem.* **280,** 33305–33310
17. Suno, R., Niwa, H., Tsuchiya, D., Zhang, X., Yoshida, M., and Morikawa, K. (2006) Structure of the whole cytosolic region of ATP-dependent protease FtsH. *Mol. Cell* **22,** 575–585
18. Kaser, M., Kambacheld, M., Kisters-Woike, B., and Langer, T. (2003) Oma1, a novel membrane-bound metallopeptidase in mitochondria with activities overlapping with the m-AAA protease. *J. Biol. Chem.* **278,** 46414–46423
19. McBride, H., and Soubannier, V. (2010) Mitochondrial function: OMA1 and OPA1, the grandmasters of mitochondrial health. *Curr. Biol.* **20,** R274–R276
20. Quirós, P. M., Ramsay, A. J., Sala, D., Fernández-Vizarra, E., Rodríguez, F., Peinado, J. R., Fernández-García, M. S., Vega, J. A., Enríquez, J. A., Zorzano, A., and López-Otín, C. (2012) Loss of mitochondrial protease OMA1 alters processing of the GTPase OPA1 and causes obesity and defective thermogenesis in mice. *EMBO J.* **31,** 2117–2133
21. Mallorquí-Fernández, G., Marrero, A., García-Piquè, S., García-Castellanos, R., and Gomis-Rüth, F. X. (2004) Staphylococcal methicillin resistance: fine focus on folds and functions. *FEMS Microbiol. Lett.* **235,** 1–8
22. Fuda, C. C., Fisher, J. F., and Mobashery, S. (2005) β-Lactam resistance in *Staphylococcus aureus*: the adaptive resistance of a plastic genome. *Cell. Mol. Life Sci.* **62,** 2617–2633
23. Llarrull, L. I., Testero, S. A., Fisher, J. F., and Mobashery, S. (2010) The future of the β-lactams. *Curr. Opin. Microbiol.* **13,** 551–557
24. Llarrull, L. I., and Mobashery, S. (2012) Dissection of events in the resistance to β-lactam antibiotics mediated by the protein BlaR1 from. *Staphylococcus aureus*. *Biochemistry* **51,** 4642–4649
25. Zhu, Y., Englebert, S., Joris, B., Ghuysen, J. M., Kobayashi, T., and Lampen, J. O. (1992) Structure, function, and fate of the BlaR signal transducer involved in induction of β-lactamase in *Bacillus licheniformis*. *J. Bacteriol.* **174,** 6171–6178
26. Hardt, K., Joris, B., Lepage, S., Brasseur, R., Lampen, J. O., Frère, J. M., Fink, A. L., and Ghuysen, J. M. (1997) The penicillin sensory transducer,

BlaR, involved in the inducibility of β-lactamase synthesis in *Bacillus licheniformis* is embedded in the plasma membrane via a four-α-helix bundle. *Mol. Microbiol.* **23,** 935–944
27. Berzigotti, S., Benlafya, K., Sépulchre, J., Amoroso, A., and Joris, B. (2012) *Bacillus licheniformis* BlaR1 L3 loop is a zinc metalloprotease activated by self-proteolysis. *PLoS ONE* **7,** e36400
28. Hao, H., Dai, M., Wang, Y., Huang, L., and Yuan, Z. (2012) Key genetic elements and regulation systems in methicillin-resistant *Staphylococcus aureus*. *Future Microbiol.* **7,** 1315–1329
29. Marrero, A., Mallorquí-Fernández, G., Guevara, T., García-Castellanos, R., and Gomis-Rüth, F. X. (2006) Unbound and acylated structures of the MecR1 extracellular antibiotic-sensor domain provide insights into the signal-transduction system that triggers methicillin resistance. *J. Mol. Biol.* **361,** 506–521
30. Birck, C., Cha, J. Y., Cross, J., Schulze-Briese, C., Meroueh, S. O., Schlegel, H. B., Mobashery, S., and Samama, J.-P. (2004) X-ray crystal structure of the acylated β-lactam sensor domain of BlaR1 from *Staphylococcus aureus* and the mechanism of receptor activation for signal transduction. *J. Am. Chem. Soc.* **126,** 13945–13947
31. Wilke, M. S., Hills, T. L., Zhang, H. Z., Chambers, H. F., and Strynadka, N. C. (2004) Crystal structures of the apo and penicillin-acylated forms of the BlaR1 β-lactam sensor of *Staphylococcus aureus*. *J. Biol. Chem.* **279,** 47278–47287
32. Kerff, F., Charlier, P., Colombo, M.-L., Sauvage, E., Brans, A., Frère, J.-M., Joris, B., and Fonzé, E. (2003) Crystal structure of the sensor domain of the BlaR penicillin receptor from *Bacillus licheniformis*. *Biochemistry* **42,** 12835–12843
33. Kuntz, I. D. (1992) Structure-based strategies for drug design and discovery. *Science* **257,** 1078–1082
34. Greer, J., Erickson, J. W., Baldwin, J. J., and Varney, M. D. (1994) Application of the three-dimensional structures of protein target molecules in structure-based drug design. *J. Med. Chem.* **37,** 1035–1054
35. Arinaminpathy, Y., Khurana, E., Engelman, D. M., and Gerstein, M. B. (2009) Computational analysis of membrane proteins: the largest class of drug targets. *Drug Discov. Today* **14,** 1130–1135
36. Feng, L., Yan, H., Wu, Z., Yan, N., Wang, Z., Jeffrey, P. D., and Shi, Y. (2007) Structure of a site-2 protease family intramembrane metalloprotease. *Science* **318,** 1608–1612
37. Wolfe, M. S. (2009) Intramembrane proteolysis. *Chem. Rev.* **109,** 1599–1612
38. Wolfe, M. S. (2009) Intramembrane-cleaving proteases. *J. Biol. Chem.* **284,** 13969–13973
39. Pryor, E. E., Jr., Horanyi, P. S., Clark, K. M., Fedoriw, N., Connelly, S. M., Koszelak-Rosenblum, M., Zhu, G., Malkowski, M. G., Wiener, M. C., and Dumont, M. E. (2013) Structure of the integral membrane protein C*AAX* protease Ste24p. *Science* **339,** 1600–1604
40. Quigley, A., Dong, Y. Y., Pike, A. C., Dong, L., Shrestha, L., Berridge, G., Stansfeld, P. J., Sansom, M. S., Edwards, A. M., Bountra, C., von Delft, F., Bullock, A. N., Burgess-Brown, N. A., and Carpenter, E. P. (2013) The structural basis of ZMPSTE24-dependent laminopathies. *Science* **339,** 1604–1607
41. Bode, W., Gomis-Rüth, F. X., and Stöckler, W. (1993) Astacins, serralysins, snake venom, and matrix metalloproteinases exhibit identical zinc-binding environments (HE*XX*H*XX*G*XX*H and Met-turn) and topologies and should be grouped into a common family, the 'metzincins'. *FEBS Lett.* **331,** 134–140
42. Hooper, N. M. (1994) Families of zinc metalloproteases. *FEBS Lett.* **354,** 1–6
43. McKerrow, J. H. (1987) Human fibroblast collagenase contains an amino acid sequence homologous to the zinc-binding site of *Serratia* protease. *J. Biol. Chem.* **262,** 5943–5943
44. Gomis-Rüth, F. X., Botelho, T. O., and Bode, W. (2012) A standard orientation for metallopeptidases. *Biochim. Biophys. Acta* **1824,** 157–163
45. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402
46. Sánchez, R., Serra, F., Tárraga, J., Medina, I., Carbonell, J., Pulido, L., de María, A., Capella-Gutíerrez, S., Huerta-Cepas, J., Gabaldón, T., Dopazo, J., and Dopazo, H. (2011) Phylemon 2.0: a suite of web-tools for molecular evolution, phylogenetics, phylogenomics, and hypotheses testing. *Nucleic Acids Res.* **39,** W470–W474
47. Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009) TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25,** 1972–1973
48. Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2011) ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27,** 1164–1165
49. Guindon, S., and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52,** 696–704
50. Cole, C., Barber, J. D., and Barton, G. J. (2008) The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.* **36,** W197–W201
51. Tallant, C., García-Castellanos, R., Baumann, U., and Gomis-Rüth, F. X. (2010) On the relevance of the Met-turn methionine in metzincins. *J. Biol. Chem.* **285,** 13951–13957
52. Goulas, T., Arolas, J. L., and Gomis-Rüth, F. X. (2011) Structure, function, and latency regulation of a bacterial enterotoxin potentially derived from a mammalian adamalysin/ADAM xenolog. *Proc. Natl. Acad. Sci. U.S.A.* **108,** 1856–1861
53. Tallant, C., García-Castellanos, R., Seco, J., Baumann, U., and Gomis-Rüth, F. X. (2006) Molecular analysis of ulilysin, the structural prototype of a new family of metzincin metalloproteases. *J. Biol. Chem.* **281,** 17920–17928
54. Botelho, T. O., Guevara, T., Marrero, A., Arêde, P., Fluxà, V. S., Reymond, J. L., Oliveira, D. C., and Gomis-Rüth, F. X. (2011) Structural and functional analyses reveal that *Staphylococcus aureus* antibiotic resistance factor HmrA is a zinc-dependent endopeptidase. *J. Biol. Chem.* **286,** 25697–25709
55. Kabsch, W. (2010) *XDS*. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 125–132
56. Kabsch, W. (2010) Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 133–144
57. Evans, P. R. (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. D Biol. Crystallogr.* **67,** 282–292
58. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67,** 235–242
59. Sheldrick, G. M. (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 479–485
60. Sheldrick, G. M. (2011) The *SHELX* approach to experimental phasing of macromolecules. *Acta Crystallogr. Sect. A* **67,** C13
61. Hübschle, C. B., Dittrich, B., and Sheldrick, G. M. (2012) *SHELXE*–a GUI for *SHELXL*. *Acta Crystallogr. Sect. A* **68,** s121
62. Carranza, C., Inisan, A.-G., Mouthuy-Knoops, E., Cambillau, C., and Roussel, A. (1999) *AFMB Activity Report 1996–1999*, pp. 89–90, CNRS-UPR 9039, Marseille
63. Debreczeni, J. É., and Emsley, P. (2012) Handling ligands with Coot. *Acta Crystallogr. D Biol. Crystallogr.* **68,** 425–430
64. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 486–501
65. Smart, O. S., Womack, T. O., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vonrhein, C., and Bricogne, G. (2012) Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr. D Biol. Crystallogr.* **68,** 368–380
66. Cowtan, K. (2010) Recent developments in classical density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 470–478
67. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,**

1605–1612

68. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54,** 905–921
69. Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2256–2268
70. Lawrence, M. C., and Colman, P. M. (1993) Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234,** 946–950
71. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 12–21
72. Vriend, G. (1990) What if: a molecular modelling and drug design program. *J. Mol. Graph.* **8,** 52–56
73. Ortiz, A. R., Strauss, C. E., and Olmea, O. (2002) MAMMOTH (matching molecular models obtained from theory): an automated method for model comparison. *Protein Sci.* **11,** 2606–2621
74. Berman, H. M., Battistuz, T., Bhat, T. N., Bluhm, W. F., Bourne, P. E., Burkhardt, K., Feng, Z., Gilliland, G. L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, H., Westbrook, J. D., and Zardecki, C. (2002) The Protein Data Bank. *Acta Crystallogr. D Biol. Crystallogr.* **58,** 899–907
75. Holm, L., and Rosenström, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38,** W545–W549
76. Schechter, I., and Berger, A. (1967) On the size of active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27,** 157–162
77. Tallant, C., Marrero, A., and Gomis-Rüth, F. X. (2010) Matrix metalloproteinases: fold and function of their catalytic domains. *Biochim. Biophys. Acta* **1803,** 20–28
78. Takeda, S. (2009) Three-dimensional domain architecture of the ADAM family proteinases. *Semin. Cell Dev. Biol.* **20,** 146–152
79. Takeda, S., Takeya, H., and Iwanaga, S. (2012) Snake venom metalloproteinases: Structure, function, and relevance to the mammalian ADAM/ADAMTS family proteins. *Biochim. Biophys. Acta* **1824,** 164–176
80. Guevara, T., Yiallouros, I., Kappelhoff, R., Bissdorf, S., Stöcker, W., and Gomis-Rüth, F. X. (2010) Proenzyme structure and activation of astacin metallopeptidase. *J. Biol. Chem.* **285,** 13958–13965
81. Arolas, J. L., Broder, C., Jefferson, T., Guevara, T., Sterchi, E. E., Bode, W., Stöcker, W., Becker-Pauly, C., and Gomis-Rüth, F. X. (2012) Structural basis for the sheddase function of human meprin β-metalloproteinase at the plasma membrane. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 16131–16136
82. Gomis-Rüth, F. X. (2008) Structure and mechanism of metallocarboxypeptidases. *Crit. Rev. Biochem. Mol. Biol.* **43,** 319–345
83. Gao, X., Wang, J., Yu, D. Q., Bian, F., Xie, B. B., Chen, X. L., Zhou, B. C., Lai, L. H., Wang, Z. X., Wu, J. W., and Zhang, Y. Z. (2010) Structural basis for the autoprocessing of zinc metalloproteases in the thermolysin family. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 17569–17574
84. Janin, J., and Chothia, C. (1990) The structure of protein-protein recognition sites. *J. Biol. Chem.* **265,** 16027–16030
85. Harding, M. M. (2004) The architecture of metal coordination groups in proteins. *Acta Crystallogr. D Biol. Crystallogr.* **60,** 849–859
86. Matthews, B. W., Jansonius, J. N., Colman, P. M., Schoenborn, B. P., and Dupourque, D. (1972) Three-dimensional structure of thermolysin. *Nat. New Biol.* **238,** 37–41
87. Tholander, F., Muroya, A., Roques, B. P., Fournié-Zaluski, M. C., Thunnissen, M. M., and Haeggström, J. Z. (2008) Structure-based dissection of the active site chemistry of leukotriene A4 hydrolase: implications for M1 aminopeptidases and inhibitor design. *Chem. Biol.* **15,** 920–929
88. Bauvois, C., Jacquamet, L., Huston, A. L., Borel, F., Feller, G., and Ferrer, J. L. (2008) Crystal structure of the cold-active aminopeptidase from *Colwellia psychrerythraea*, a close structural homologue of the human bifunctional leukotriene A4 hydrolase. *J. Biol. Chem.* **283,** 23315–23325
89. Nguyen, T. T., Chang, S. C., Evnouchidou, I., York, I. A., Zikos, C., Rock, K. L., Goldberg, A. L., Stratikos, E., and Stern, L. J. (2011) Structural basis for antigenic peptide precursor processing by the endoplasmic reticulum aminopeptidase ERAP1. *Nat. Struct. Mol. Biol.* **18,** 604–613
90. Kochan, G., Krojer, T., Harvey, D., Fischer, R., Chen, L., Vollmar, M., von Delft, F., Kavanagh, K. L., Brown, M. A., Bowness, P., Wordsworth, P., Kessler, B. M., and Oppermann, U. (2011) Crystal structures of the endoplasmic reticulum aminopeptidase-1 (ERAP1) reveal the molecular basis for N-terminal peptide trimming. *Proc. Natl. Acad. Sci. U.S.A.* **108,** 7745–7750
91. Kyrieleis, O. J., Goettig, P., Kiefersauer, R., Huber, R., and Brandstetter, H. (2005) Crystal structures of the tricorn interacting factor F3 from *Thermoplasma acidophilum*, a zinc aminopeptidase in three different conformations. *J. Mol. Biol.* **349,** 787–800
92. McGowan, S., Porter, C. J., Lowther, J., Stack, C. M., Golding, S. J., Skinner-Adams, T. S., Trenholme, K. R., Teuscher, F., Donnelly, S. M., Grembecka, J., Mucha, A., Kafarski, P., Degori, R., Buckle, A. M., Gardiner, D. L., Whisstock, J. C., and Dalton, J. P. (2009) Structural basis for the inhibition of the essential *Plasmodium falciparum* M1 neutral aminopeptidase. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 2537–2542
93. Addlagatta, A., Gay, L., and Matthews, B. W. (2006) Structure of aminopeptidase N from *Escherichia coli* suggests a compartmentalized, gated active site. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 13339–13344
94. Gomis-Rüth, F. X. (2003) Structural aspects of the *metzincin* clan of metalloendopeptidases. *Mol. Biotechnol.* **24,** 157–202
95. Zhan, C., Fedorov, E. V., Shi, W., Ramagopal, U. A., Thirumuruhan, R., Manjasetty, B. A., Almo, S. C., Fiser, A., Chance, M. R., and Fedorov, A. A. (2005) The ybeY protein from *Escherichia coli* is a metalloprotein. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **61,** 959–963
96. Oganesyan, V., Busso, D., Brandsen, J., Chen, S., Jancarik, J., Kim, R., and Kim, S. H. (2003) Structure of the hypothetical protein AQ_1354 from *Aquifex aeolicus*. *Acta Crystallogr. D Biol. Crystallogr.* **59,** 1219–1223
97. Weiss, M. S. (2001) Global indicators of x-ray quality. *J. Appl. Cryst.* **34,** 130–135
98. Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62,** 72–82
99. Bernsel, A., Viklund, H., Hennerdal, A., and Elofsson, A. (2009) TOPCONS: consensus prediction of membrane protein topology. *Nucleic Acids Res.* **37,** W465–W468
100. Corpet, F. (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* **16,** 10881–10890
101. Holmes, M. A., and Matthews, B. W. (1981) Binding of hydroxamic acid inhibitors to crystalline thermolysin suggests a pentacoordinate zinc intermediate in catalysis. *Biochemistry* **20,** 6912–6920

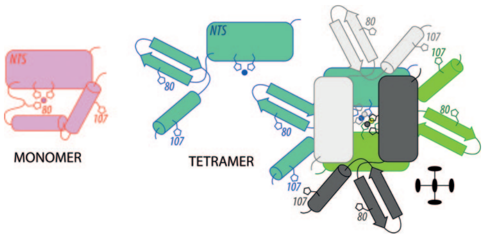
RESULTS: PROJECT 3

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

A Journal of the Gesellschaft Deutscher Chemiker
Angewandte Chemie
GDCh
International Edition

Enzymatic Activity

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



Shape shifting: A minimal metamorphic, selective, and specific caseinolytic metallopeptidase, selecase, reversibly transits between several different states of defined three-dimensional structure (monomer and tetramer represented in picture). The competent conformation is sequestered in incompetent but structured dimers, tetramers, and octamers, which are associated with loss of enzymatic activity due to autoinhibition.

M. López-Pelegrín, N. Cerdà-Costa, A. Cintas-Pedrola, F. Herranz-Trillo, P. Bernadó, J. R. Peinado, J. L. Arolas, F. X. Gomis-Rüth* ______ **10624 – 10630**

Keywords: metallopeptidases · metamorphic proteins · protein folding · protein structures

2014 – 53/40

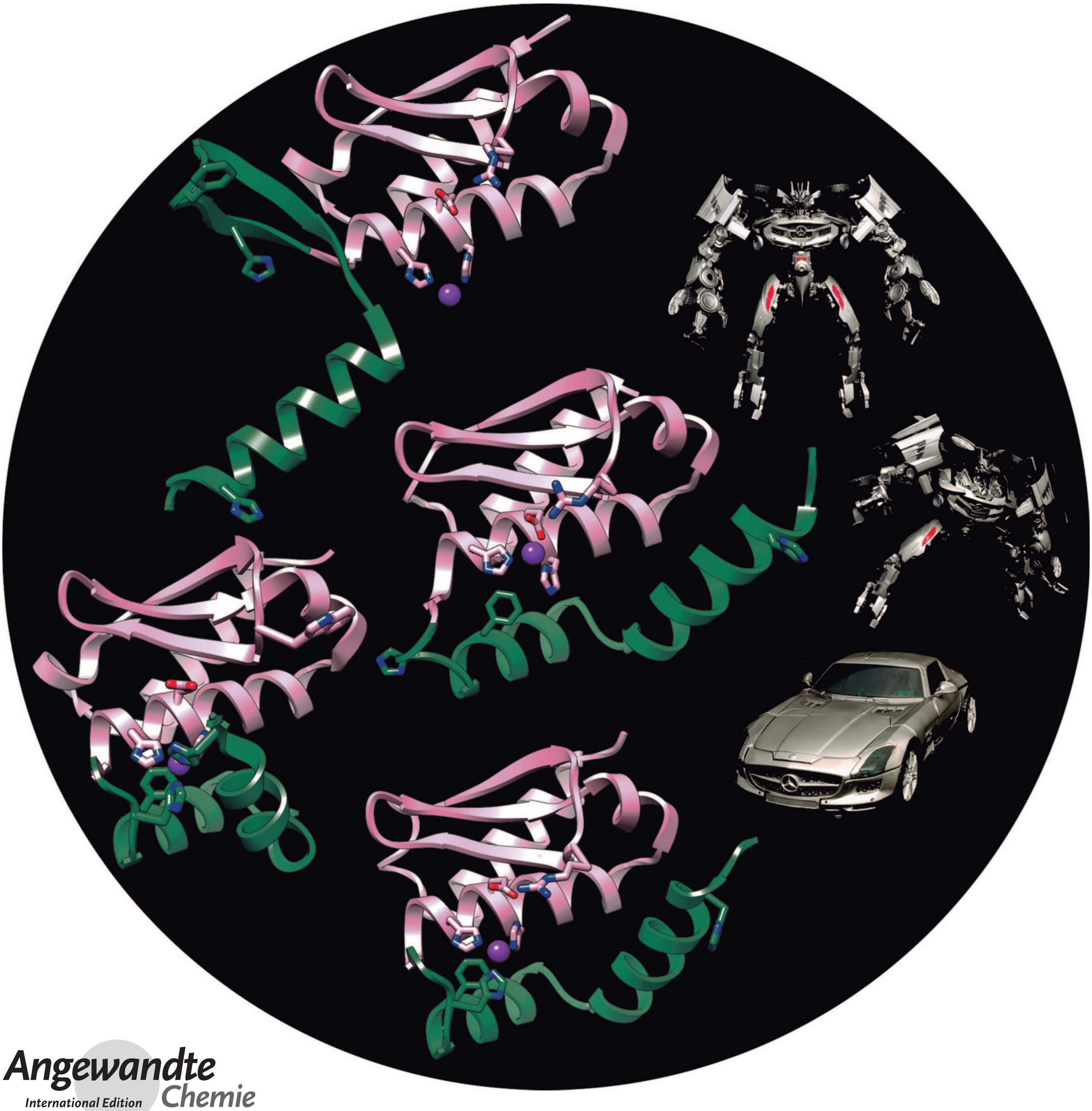
WILEY-VCH

Enzymatic Activity

DOI: 10.1002/anie.201405727

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**



Abstract: *Molecular plasticity controls enzymatic activity: the native fold of a protein in a given environment is normally unique and at 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 multifunnel energy landscape, affords a switch that provides a reversible mechanism of control of catalytic activity unique in nature.*

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] 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 increases activity, as more substrate can be bound and turned over.[5] Here we describe a metamorphic minimal *sel*ective and specific *case*inolytic metallopeptid*ase*, *selecase*, which shows a 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 the 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. Figure 1 and Suppl. Tables 1 and 2).

Selecase was extremely soluble in aqueous buffer and did not precipitate at 130 mg mL^{-1}. 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 that 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 for example, with tobacco-etch virus proteinase, which is widely used in biotechnology (Figure 1 a).

Surprisingly, although selecase activity did indeed increase with concentration between 0.025–0.25 mg mL^{-1}, it fell sharply thereafter to become only residual at 50 mg mL^{-1}. 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. This, to our knowledge, is novel for peptidases.

Subsequently, we explored the oligomerization of selecase in solution in the concentration range 0.15–65 mg mL^{-1} 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. Figure 2 a). SEC-MALLS, which combines SEC with multi-angle laser light scattering (MALLS), revealed two average populations with molecular weights of 25 KDa and 80 KDa, possibly corresponding to dimeric and octameric selecase, respectively, along with additional species such as monomers and tetramers (Figure 1 b and Suppl. Figure 2 b). 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 (Figure 1 c 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. Figure 2 c). The 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. Figure 2 d). Finally, SAXS revealed that the protein did not aggregate at concentrations of up to 65 mg mL^{-1} (Suppl. Table 4, Figure 1 d, Suppl. Figures 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

---

[*] M. López-Pelegrín,[+] Dr. N. Cerdà-Costa,[+] Dr. A. Cintas-Pedrola,
Dr. J. R. Peinado,[$] Dr. J. L. Arolas, Prof. Dr. F. X. Gomis-Rüth
Proteolysis Lab, Molecular Biology Institute of Barcelona
CSIC, Barcelona Science Park
c/Baldiri Reixac, 15–21, 08028 Barcelona (Spain)
E-mail: fxgr@ibmb.csic.es
Homepage: http://www.ibmb.csic.es/home/xgomis
F. Herranz-Trillo, Dr. P. Bernadó
Centre de Biochimie Structurale, INSERM U1054
CNRS UMR 5048, Université Montpellier 1 and 2
29 rue de Navacelles, 34090 Montpellier (France)
[$] Present address: Department of Medical Sciences
University of Castilla-La Mancha, 13071 Ciudad Real (Spain)
[+] These authors contributed equally to this work.

Supporting information for this article (including experimental procedures, supplemental results and discussion, acknowledgments, supplemental references, tables, figures, and movies) is available on the WWW under http://dx.doi.org/10.1002/anie.201405727.

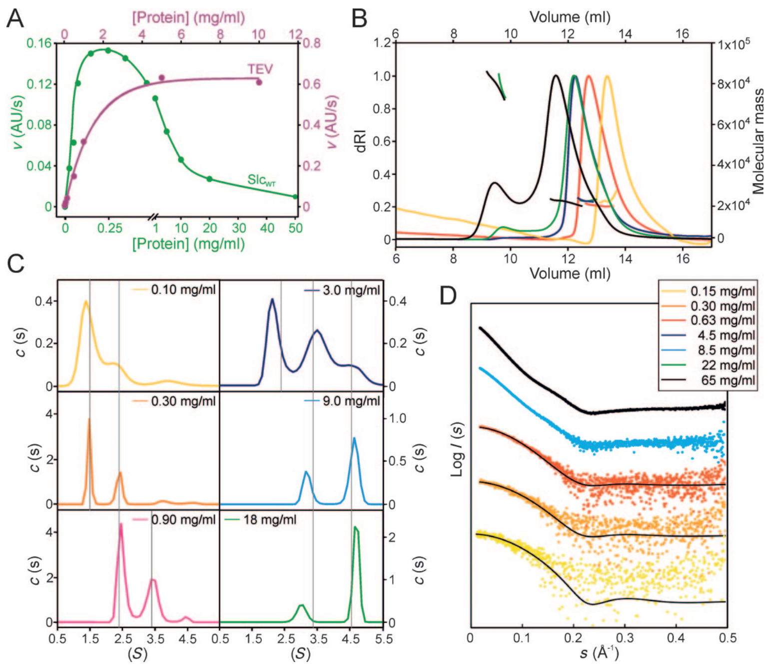


Figure 1. A polyoligomeric metallocaseinase with abnormal activity. A) Proteolytic activity of wild-type selecase on peptide CCS (green curve). Tobacco-etch virus proteinase mutant S219V, which shows comparable catalytic efficiency to selecase but normal concentration-dependent activity, is shown for comparison (purple curve). B) SEC-MALLS of selecase at selected initial concentrations (0.15–65 $\mathrm{mg\,mL^{-1}}$; see also Suppl. Figure 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). dRI = differential refracting index. C) Analytical ultracentrifugation curves at six selected concentrations depicting the concentration-dependent oligomeric populations. Essentially, monomers are predominantly found at 0–0.3 $\mathrm{mg\,mL^{-1}}$; dimers at 0.3–2 $\mathrm{mg\,mL^{-1}}$; tetramers at 2–6 $\mathrm{mg\,mL^{-1}}$; and octamers at >6 $\mathrm{mg\,mL^{-1}}$. S = sedimentation coefficient, $c(S)$ = continuous sedimentation coefficient discribution. D) SAXS intensity profiles, $I(s)$, as a function of the momentum transfer, s, measured for wild-type selecase at selected concentrations (see Suppl. Figure 3 for all curves). Profiles have been displaced along the $I(s)$ axis for comparison. The experimental scattering curves at the three lowest concentrations studied indicate a mixed population of monomers and dimers based on the crystallographic structures of slc$_1$ and slc$_2$ (black curves).

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 (0.2–0.3 $\mathrm{mg\,mL^{-1}}$; Figure 1 a), 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 (Figure 1 a) and is reminiscent of previous reports on oligomerization inhibiting the 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 to monomeric (slc$_1$), dimeric (slc$_2$), and tetrameric (slc$_4$) forms of selecase, respectively. 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 Refs. [7b,10]). This conclusion is supported by the enzymatic activity found for selecase in solution associated with a monomeric species (see above). It is also reinforced by SAXS for which the experimental scattering curves at the three lowest 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 lowest concentrations reaching 70 % (see Figure 1 d, SRD, and Suppl. Figures 3 and 4).

At 13.1 KDa, slc$_1$ is the smallest active peptidase structurally characterized to date and it has a compact globular shape 35–40 Å in diameter (Figure 2 a). It consists of an upper N-terminal subdomain (NTS; residues M^1–Y^{76}) and a lower C-terminal subdomain (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 (Figure 2 b). 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 (Figure 2 a). 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 roughly 90°. 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; the "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

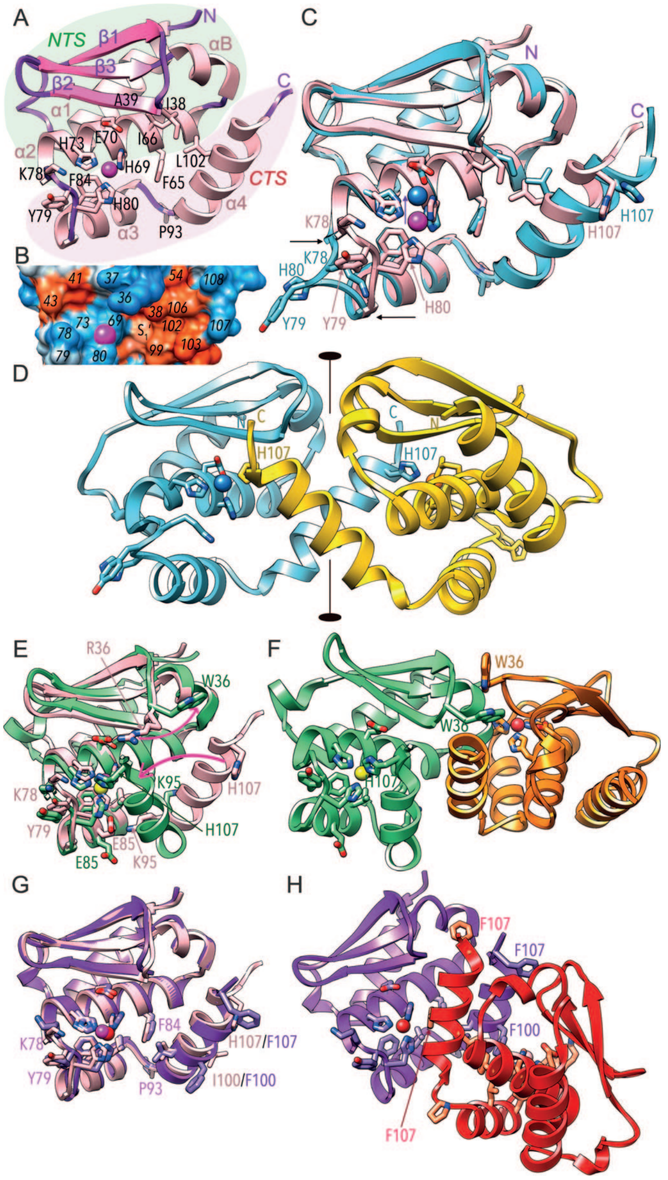


Figure 2. Competent monomer versus incompetent dimers. A) Ribbon representation of slc$_1$ in standard orientation.[8] Helices (α1, αB, and α2–α4) are shown in pink, β-strands (β1–β3) in magenta, and loops/coils in purple. For extent and nomenclature of regular secondary structure elements, see Suppl. 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 50° 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).

the cleft (Figure 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 (Figure 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 Figure 2c). However, major metamorphic rearrangement is observed around the metal-binding site (see Figure 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 7 Å. 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} (Figure 2c). Two selecase monomers associate through C_2 symmetry under occlusion of a large surface (2130 Å^2; see Suppl. Table 8 and Figure 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 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 Figure 3a and Suppl. Movie 2). The segment of the active-site helix with the first two metal ligands undergoes slight displacement (Figure 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 (Figure 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 (Figure 3a,b), as a result of which metal-ligand H^{80} shifts roughly 16 Å 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 about 30 Å apart on average in slc$_4$ (Figure 3a). Overall, this metamorphic structural transition of selecase is stabilized by the association of four monomers in the crystal (Figure 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 60–75 Å in diameter

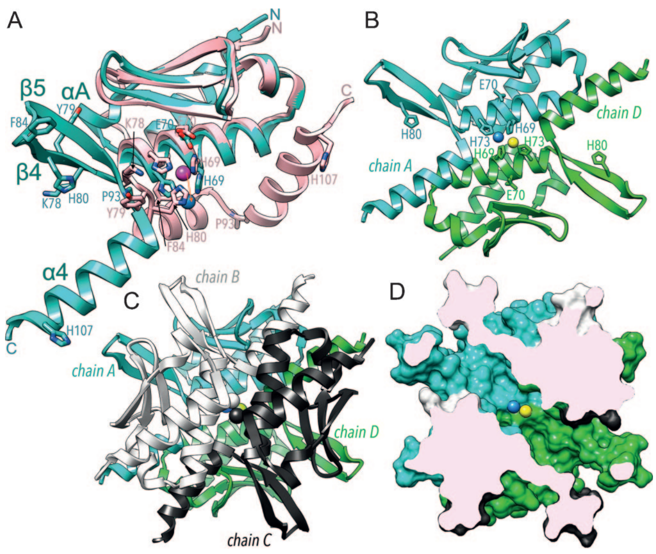


Figure 3. A compact autoinhibitory tetrameric particle. A) Superposition of slc$_1$ and slc$_4$ monomers in pink (magenta metal ion) and turquoise (blue metal ion), respectively, in the view of Figure 2a. Only the distinct secondary structure elements of slc$_4$ are labeled (see also Suppl. Table 7). 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 nonfunctional 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 90° 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.

(Figure 3c,d). One monomer (chain A) interacts through D_2 symmetry—by hiding a total interface of 9850 Å^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 ≈16 Å, major axis ≈21 Å; Figure 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 50 Å in length and 15 Å 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 roughly 3 Å apart (Figure 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 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, lacking four (slcΔC4) and eight (slcΔC8) C-terminal residues. Moreover, we cloned two close orthologues 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 5.0 mg mL^{-1}, 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 displayed a concentration-dependent equilibrium between monomers, dimers, tetramers, and octamers and a reduction in activity as concentration increased, similar to the wild-type (Suppl. Figure 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 orthologues, 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 Figures 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 the 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 only until the glutamate helix. Thereafter, a 90° rotation around bond N–Cα of K^{95} results in C-terminal helix α4 being rotated as a rigid body by 50° 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 Figure 2d–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 (Figure 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.

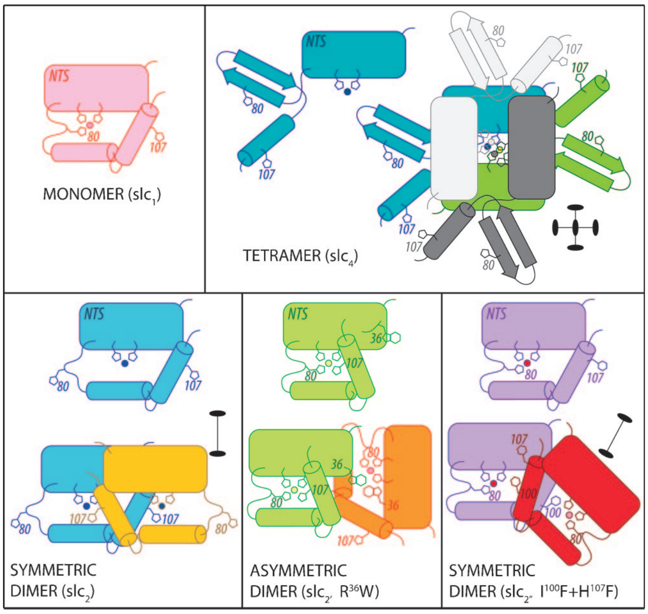


Figure 4. Scheme illustrating the topology of the distinct selecase structures reported. A black ellipse stands for a dyad vertical to the plane, two black ellipses connected by a line stand for a dyad in the plane. NTS, N-terminal subdomain; histidines H^{69}, H^{73}, H^{80}, and H^{107} are shown.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

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 that of 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 (Figure 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 nonprimed sides of the active-site clefts are occluded and the phenylalanine rings at position 100 penetrate 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 the shielding of the cleft (Figure 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 corresponding structures was trapped in crystals of the wild-type protein. This implies that replacement of just one and two residues leads to two new structures of selecase (thus totaling five), supporting the metamorphic character of this protein.

Summarizing, we have succeeded in identifying and probing for the first time the structural transitions of a natural metamorphic protein with a multifunnel folding energy landscape. Although metamorphic proteins may be encoded by a relevant fraction of all genomes, the lack of bioinformatics and 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 (Figure 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 subdomain corresponding to a distinct folding unit or foldon[16] and the subdomain interface acting as a reversible zipper. The high flexibility of CTS was further verified by computational analysis of local conformational frustration and assessment of interdomain flexibility based on the elastic network model (see SRD and Suppl. Figure 7). In addition, the thermodynamic consistency of interconversion was further backed by the calculated geometric and thermodynamic parameters of the solvation free energy of folding and of dissociation, as well as compactness, for wild-type selecase structures (see SRD and Suppl. Table 8). Owing to inherent flexibility of the CTS, it 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 the repacking of a hydrophobic core and exposure of new binding surfaces, which in turn generate the 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 a reversible, strictly concentration-dependent reduction of activity at higher concentrations, which is triggered by the sequestering of 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.

Received: May 28, 2014

Keywords: metallopeptidases · metamorphic proteins · protein folding · protein structures

---

[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. Lett.* **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. Hebd. Seances Acad. Sci.* **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, Berlin, **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 Proteins Proteomics* **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, *Protein Sci.* **2014**, *23*, 123–144.

[12] X. Zhou, F. Alber, G. Folkers, G. H. Gonnet, G. Chelvanayagam, *Proteins Struct. Funct. Genet.* **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.

SUPPORTING INFORMATION

Table of Contents

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}$$\mathrm{s}^{-1}$; K_{m}=94±22µM; $k_{\mathrm{cat}}/K_{\mathrm{m}}$=150$\mathrm{M}^{-1}\mathrm{s}^{-1}$ (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 section 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 $\Delta^{i}G$ value (solvation free energy gain upon dimerization) of –15.4kcal/mol, a $\Delta^{i}G$ P-value (probability of getting a lower $\Delta^{i}G$ 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 $\Delta^{i}G$ value of –17.9kcal/mol, a $\Delta^{i}G$ 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
0.03mg/ml 0.10mg/ml 0.30mg/ml 0.90mg/ml 3.0mg/ml 9.0mg/ml 18mg/ml	slc_1 (76.0%), slc_2 (24%) slc_1 (70.6%), slc_2 (29.4%) slc_1 (55.9%), slc_2 (33.7%), slc_4 (10.4%) slc_2 (53.0%), slc_4 (34.5%), slc_8 (12.5%) slc_2 (24.8%), slc_4 (47.2%), slc_8 (28%) slc_4 (32.6%), slc_8 (67.4%) slc_4 (20.9%), slc_8 (79.1%)
Equilibrium velocity	Average mass (Da) ($Ln(Mw/Mw_{app})=7.86 \times \varphi$)
0.03mg/ml 0.10mg/ml 0.30mg/ml 0.90mg/ml 3.0mg/ml 9.0mg/ml 18mg/ml	16,866±192 21,364±96 32,313±204 41,135±164 55,833±240 75,875±356 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
$\alpha1$ (I^{6}, L^{7}, I^{10}, L^{11}, T^{14}, I^{15}, L^{18})	$L\alpha2\alpha3$ (L^{81})
$L\alpha1\beta1$ (L^{20}, I^{27})	$\alpha3$ ($\underline{F}^{84}$, $\underline{L}^{88}$, $\underline{L}^{91}$)
$\beta1$ (I^{29}, I^{31})	$L\alpha3\alpha4$ ($\underline{F}^{92}$)
$\beta2$ ($\underline{I}^{38}$, A^{39}, I^{41})	$\alpha4$ ($\underline{K}^{95}$, $\underline{I}^{98}$, $\underline{L}^{99}$, $\underline{L}^{102}$, $\underline{L}^{106}$)
$L\beta2\beta3$ (L^{43})	
$\beta3$ (I^{48}, I^{50})	
αB ($\underline{I}^{54}$, L^{55}, $\underline{Y}^{57}$, $\underline{L}^{58}$)	
$\alpha2$ (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)

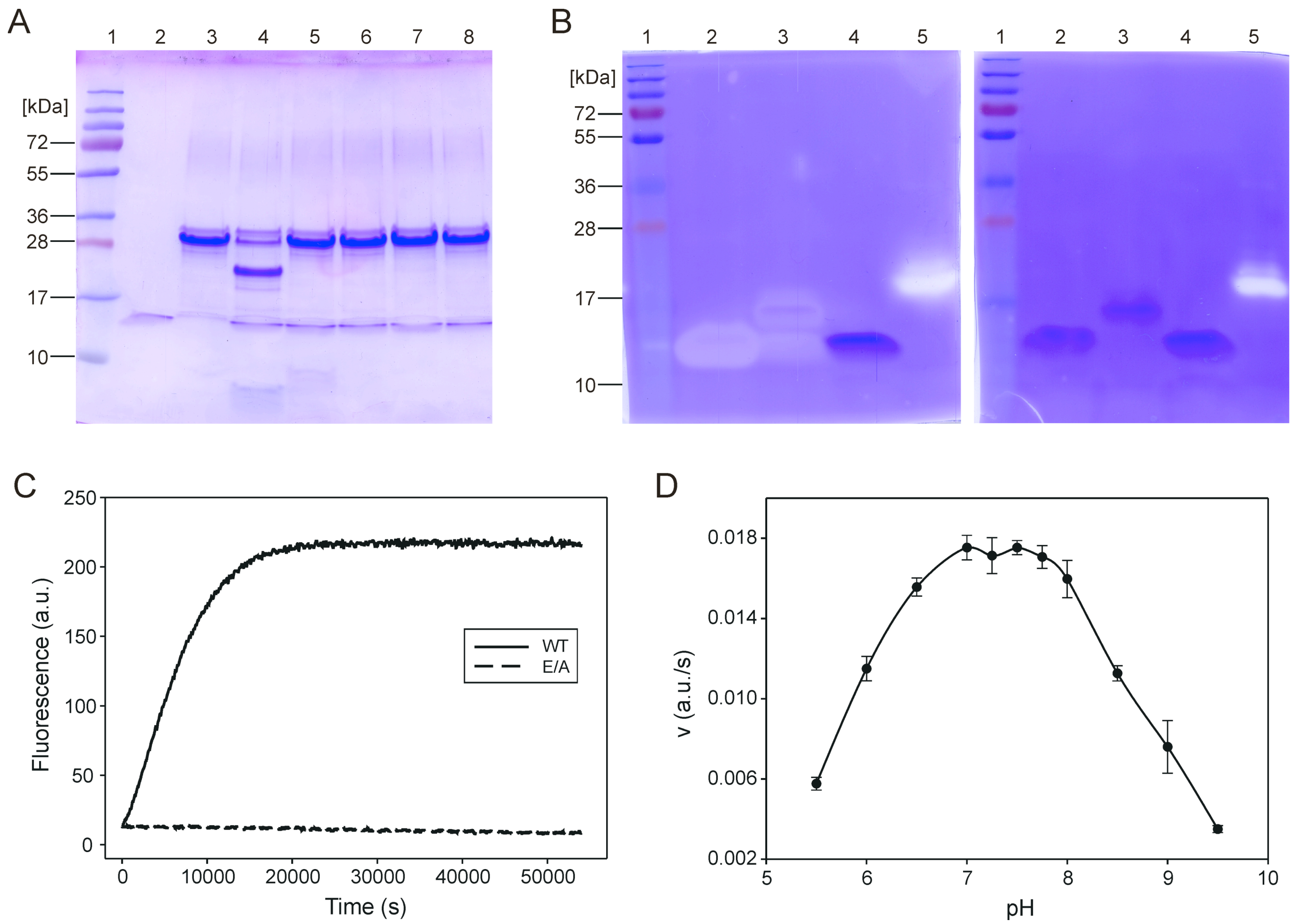
40 symmetric close contacts (incl. 2 hydrogen bonds, 1 metallorganic contact, 1 metal-metal contact, and 25 hydrophobic carbon-carbon contacts by 17 residues of each molecule). Segments involved in oligomerization: 17-19, 38, 54-76, 93-106, and 999 (metal ion). Sc=0.65

Molecular and oligomerization analysis [d]

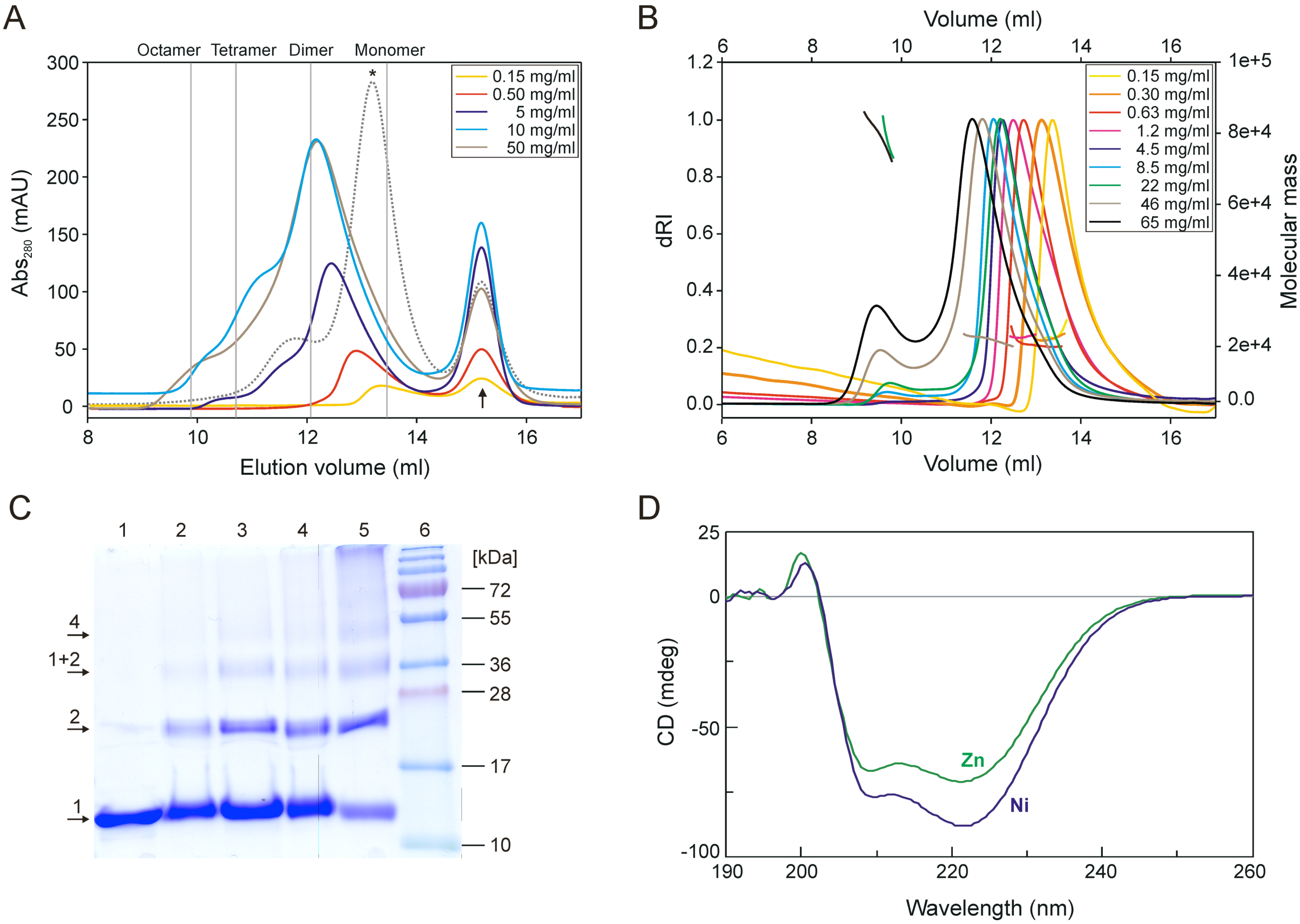
	No. of atoms	No. of residues	Total surface (Å^2)	ΔG [d,e] (kcal/mol)	Residues (%) in regular secondary structure (α / β / total)	Calculated oligomeric state [d] / ΔG_{diss}(kcal/mol) [f]	Interface buried (Å^2)
Slc_1 (monomer)	914	109	6,999	–109	60 / 14 / 73	1 / -	
Slc_2 (monomer)	907	108	7,274	–105	66 / 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±350Å^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$_4$.

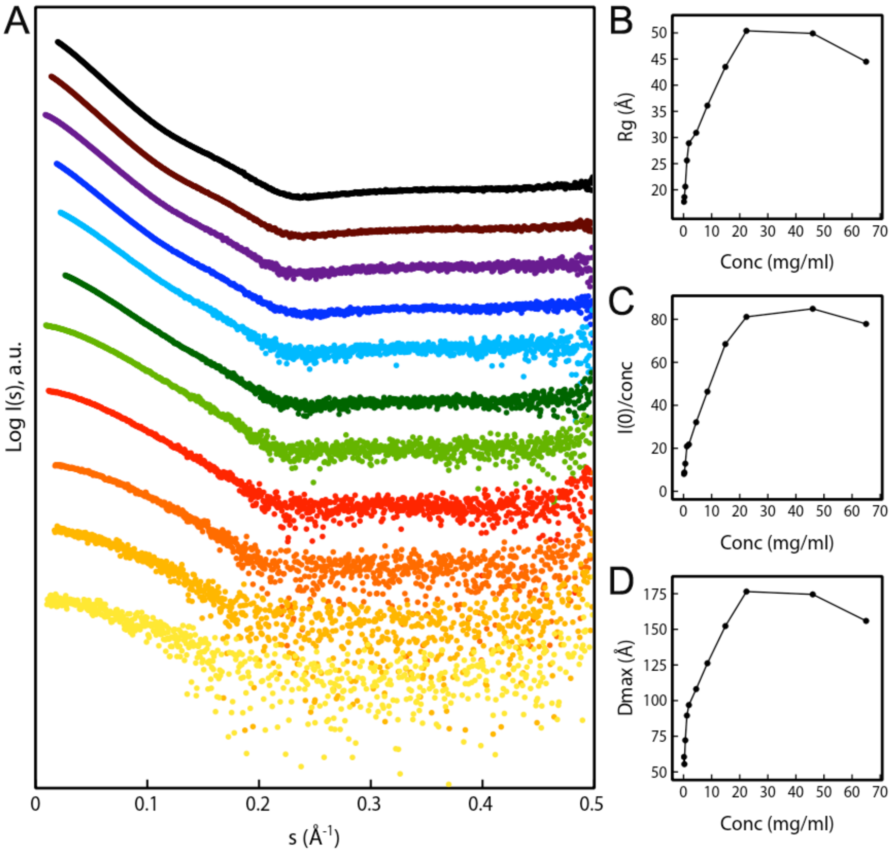
6. SUPPLEMENTAL FIGURES



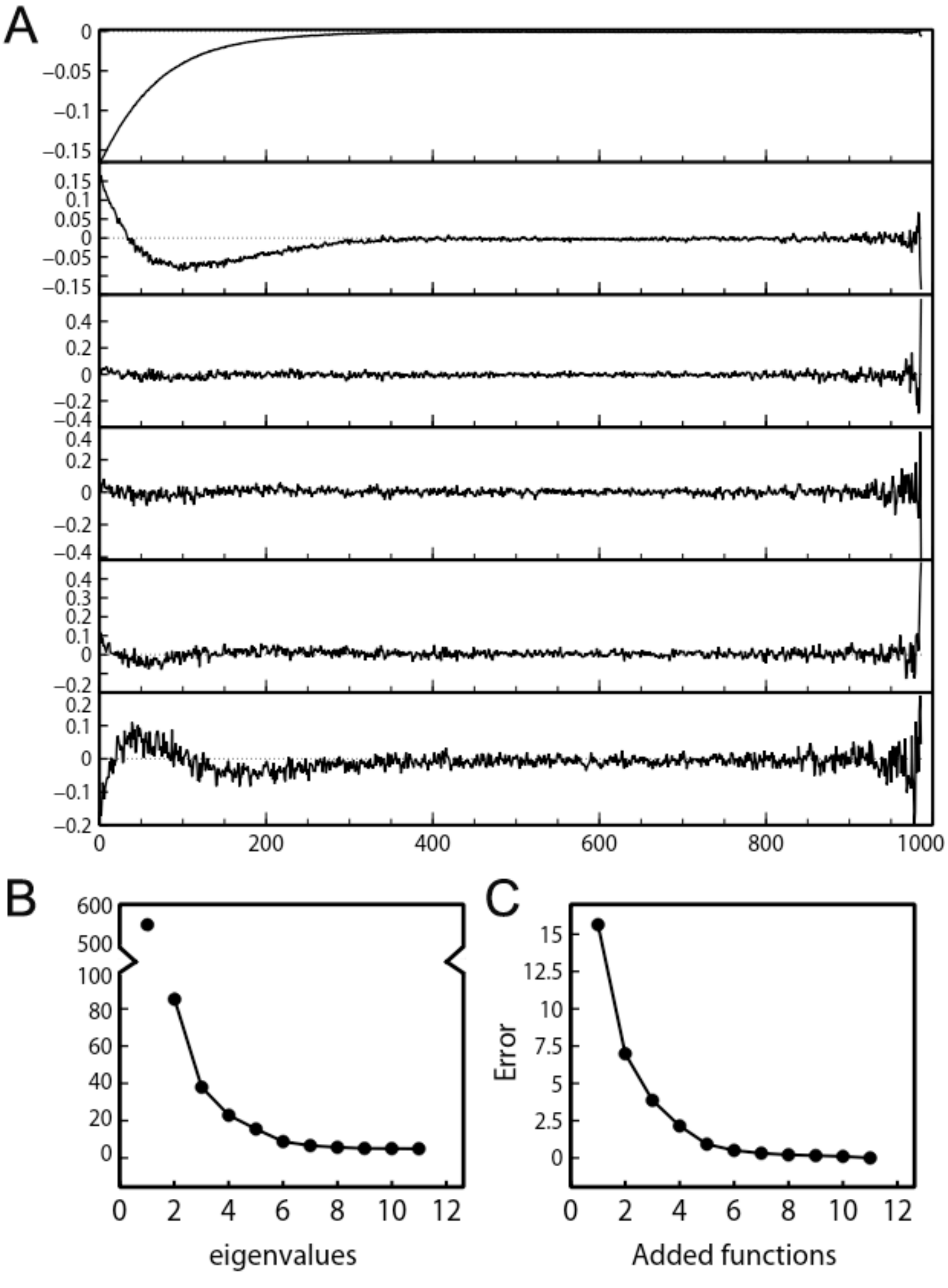
Suppl. Figure 1 – (A) SDS-PAGE of α-casein incubated with selecase (see section 1.2). *Lane 1*, molecular mass standard. *Lane 2*, selecase. *Lane 3*, α-casein consisting of α_{s2}-casein (top light band; 25KDa) and α_{s1}-casein (bottom strong band; 22-23.7KDa). *Lane 4*, α-casein incubated with wild-type 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.



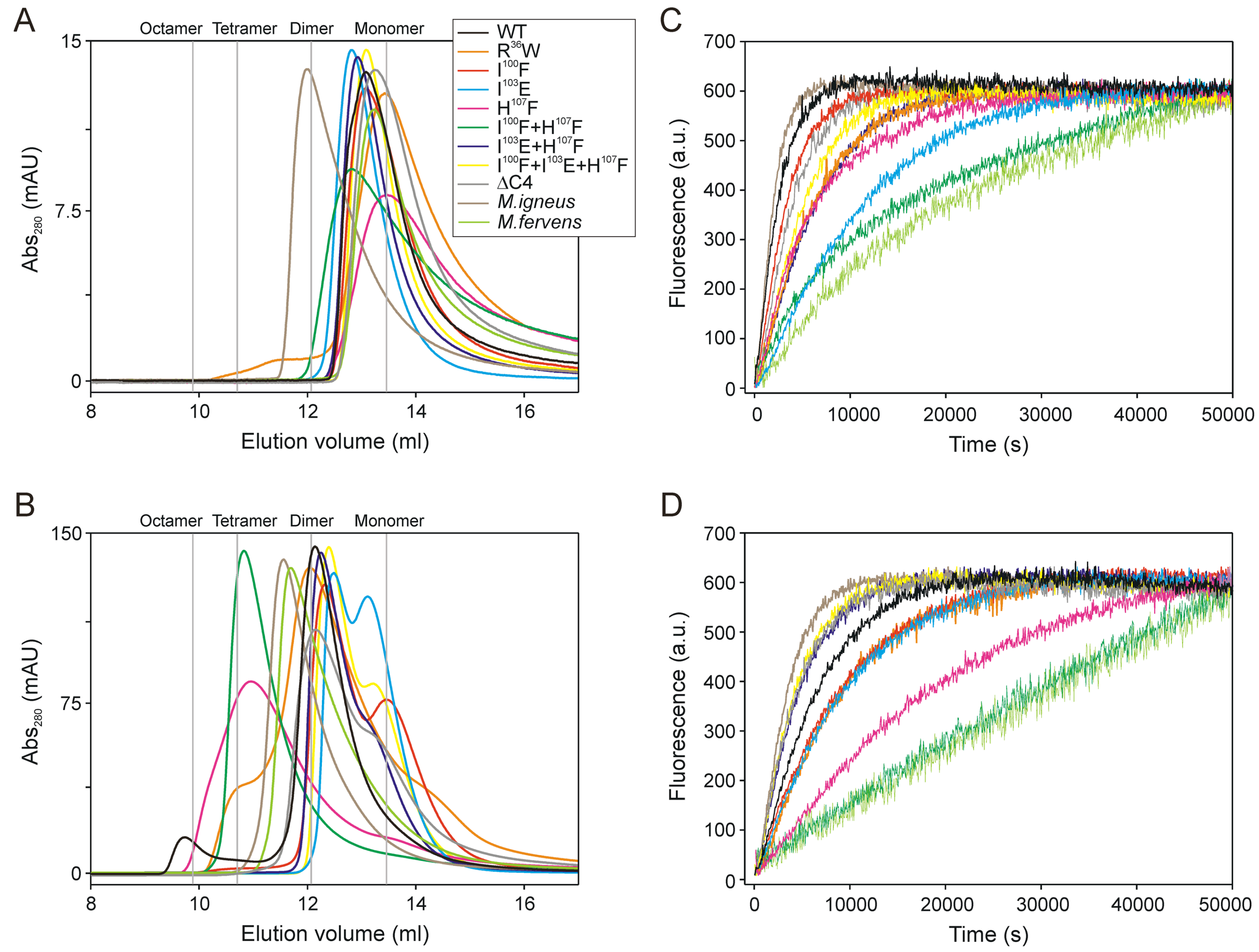
Suppl. Figure 2 – (A) Calibrated size-exclusion chromatography (SEC) of selecase at selected initial concentrations (0.15-50mg/ml; samples become diluted during this chromatography) plus aprotinin (arrow) as internal molecular-mass reference (6.5KDa; see section 1.4 for details). The peak pattern observed 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.



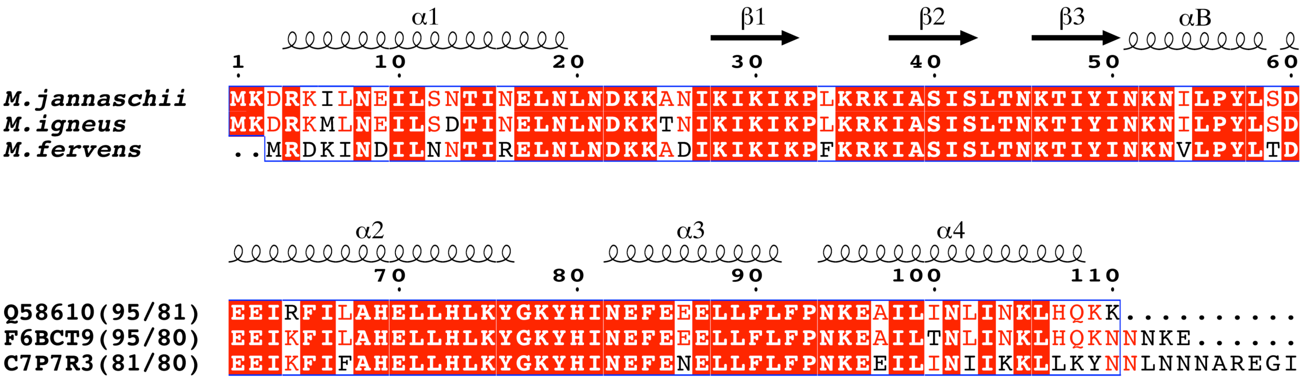
Suppl. Figure 3 – (A) SAXS intensity profiles, *I*(s), as a function of the momentum transfer, s, measured for wild-type selecase at 11 concentrations (*bottom* to *top*): 0.15 (yellow), 0.30, 0.63, 1.2, 1.9, 4.5, 8.5, 15, 22, 46, and 65mg/ml (black). Profiles have been displaced along the *I*(s) axis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



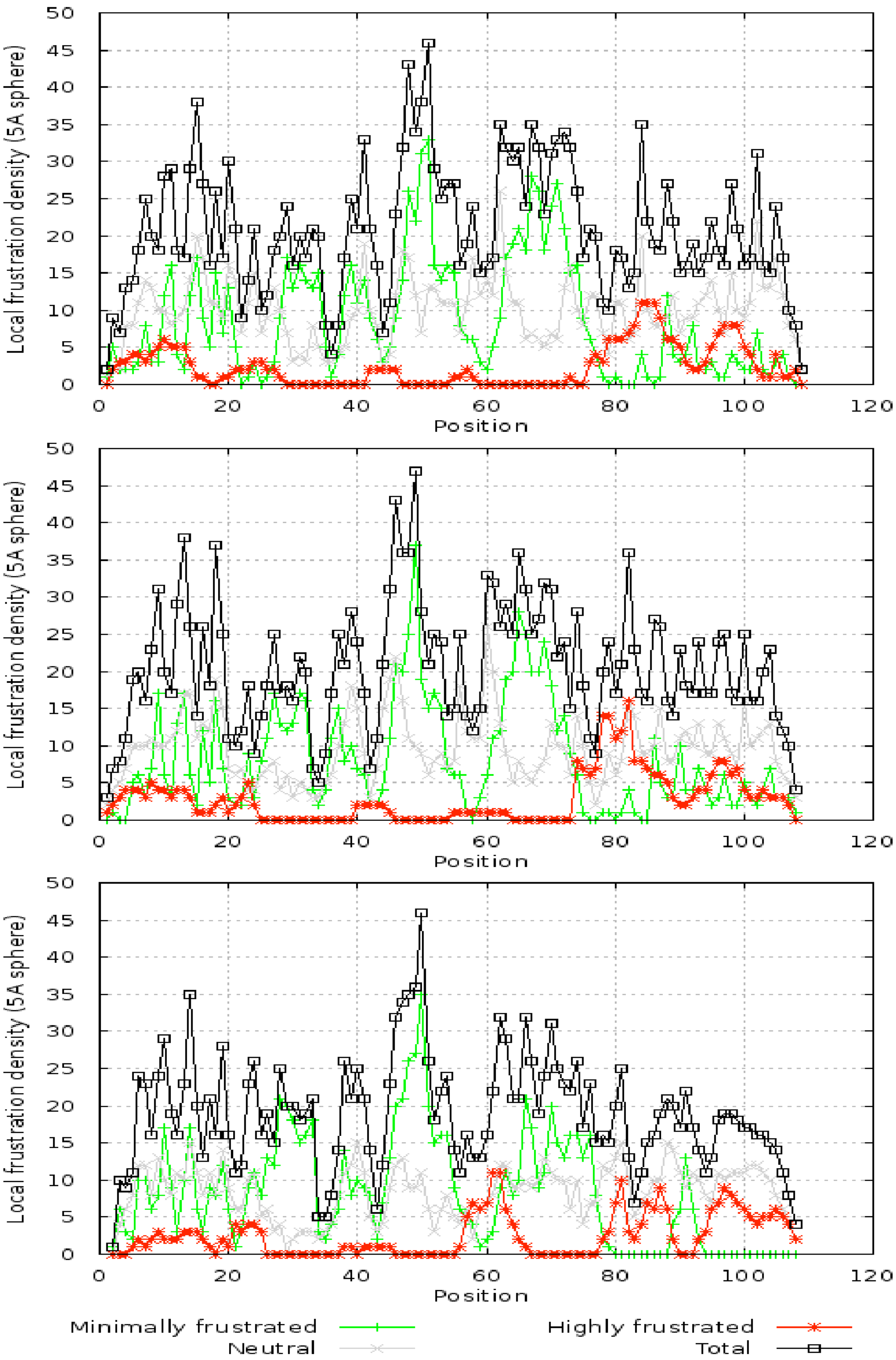
Suppl. Figure 4 – Single Value Decomposition (SVD) analysis of the SAXS data measured for selecase. Eigenvectors **(A)** derived from the analysis sorted by their corresponding eigenvalues **(B)**. **(C)** Agreement with the experimental data upon successive increase in the number of eigenvectors.



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$_4$.

GENERAL DISCUSSION

The present thesis focused on the study of four prokaryotic MPs, both from a biochemical and a biophysical/structural perspective, providing novel and peculiar information regarding the mechanisms that these MPs employ to maintain latency. Interestingly, these mechanisms include three new forms of zymogenicity exerted through short N- and C-terminal segments, which were previously unseen in MPs, and a concentration-dependent mechanism of regulation through oligomerization. In addition, the last protein to be studied, selecase, is, to our knowledge, the smallest active MP reported so far, thus advancing important insights into what a minimal scaffold with proteolytic activity might be.

In the first project, we structurally characterized the zymogen of the first bacterial MMP to be studied biochemically, karilysin from *T. forsythia* (Karim et al, 2010). Whereas the sequence and structure of its CD resemble that of eukaryotic counterparts, striking differences exist in the flanking domains, and most importantly, in the segment responsible for latency. Remarkably, karilysin shows the shortest PD described for an MP, which spans only 14 residues (accordingly referred to as propeptide) instead of the ~80-residues PD typical of mammalian MMPs (see “Introduction”, sections 3.1.4 and 3.1.5).

The crystal structure of prokarilysin revealed a CD almost identical to the mature portion (Kly18) previously determined (Cerdà-Costa et al, 2011), suggesting that it is preformed in the zymogen and just uncovered during maturation, as previously described for other MMPs (Tallant et al, 2010b). The short N-terminal propeptide lacks relevant secondary-structure elements and runs across the active-site cleft in an extended conformation but in the opposite direction of a genuine substrate. As in mammalian MMPs, the side chain of a conserved residue within the propeptide is found substituting the solvent molecule in the zinc coordination sphere, rendering a non-competent active site. Nevertheless, the conserved residue in prokarilysin is an aspartatic acid (D^{25}), in contrast to the typical cysteine residue involved in “cysteine-switch” or “velcro” latency of animal and plant MMPs (Rosenblum et al, 2007; Tallant et al, 2010b; Van Wart & Birkedal-Hansen, 1990). Thus, the inhibitory mechanism in karilysin rather follows an “aspartate-switch” mechanism similar to that described for the otherwise unrelated astacin and fragilysin zymogens (Goulas et al, 2011; Guevara et al, 2010). Based on *in vitro* experiments, maturation of prokarilysin occurs through a single autolytic cleavage at peptide bond N^{34}-Y^{35}, which is found in a solvent-exposed loop, without any other cleavages in the propeptide or CD. Upon propeptide removal, a number of protein-protein interactions are lost, which contribute to the stability of the zymogen. Propeptide residue D^{25} of the zinc ligand sphere is replaced by a solvent molecule, and a major rearrangement of the new N-terminus takes place, which yields a salt bridge between the α-amino group of Y^{35} and D^{187} in “C-terminal helix” αC. Although animal MMPs undergo a relevant cleavage in a similar site, X-F/Y, the cut only takes place after certain trimming clips and conformational changes are induced by a primary cleavage in a “bait region” (Jozic et al, 2005; Nagase, 1997). The combination of structural and functional data (both *in vitro* and *in cellula*) further allowed us to shed light on different roles that the propeptide of karilysin exerts, not only in latency maintenance, but also in folding assistance during biosynthesis, thermal stability, and inhibition of peptidolytic activity *in trans*.

Overall, this work proposes a novel mechanism of latency maintenance for MMPs, which shares some features with inhibition mechanisms described for the unrelated MPs astacin and fragilysin. Moreover, it supports a previous hypothesis that karilysin would have originated from a CD of an animal MMP co-opted through horizontal gene transfer by *T. forsythia*.

In the second project, we studied a set of short-sized (105-110 residues) single-domain, soluble potential MPs from thermophilic organisms which were found by bioinformatics searches. Although our primary aim was to discover uncharacterized standalone proteins with significant sequence similarity to CDs of IMMPs from MEROPS database families M48 and M56, the structural analysis of these proteins finally led to the proposal of a novel family of minimal MPs termed “minigluzincins” (Fig. 22). This newly proposed family of MPs will appear in upcoming release of the database as we also managed to confirm activity experimentally for one of its members (see below). More recently, a publication by Trame, Rawlings et al., reported the structure of a putative mini-zincin that, in contrast to minigluzincins, had neither a cation in the active site nor proven proteolytic activity (Trame et al, 2014).

We solved the crystal structures of the zymogens of *Pyrococcus abyssi* proabylysin and *Methanocaldococcus jannaschii* projannalysin. These structures confirmed their respective monomeric and dimeric behavior in size-exclusion chromatography and provided an explanation for the lack of proteolytic activity of both proteins in the tested conditions. In accordance with 31% sequence similarity, they exhibit a very similar overall shape and conserved secondary-structure elements with practically overlapping chains from the beginning of helix α1 to the end of helix α4 (M^{1}-L^{100} of proabylysin and D^{9}-L^{109} of projannalysin). While the conformation of these globular parts indicates a competent MP moiety, the inactive state of the zymogens is maintained by short C-terminal segments (after Q^{101} in proabylysin and L^{109} in projannalysin). This finding contrasts markedly with the N-terminal PDs or propeptides normally described for MPs. Noteworthy, the mechanism of C-terminal latency maintenance differs in both MPs, being intramolecular through an extended tail of five residues in proabylysin and a crosswise intermolecular interaction between neighboring molecules through a helix swap in projannalysin. Unfortunately, all our attempts to provide insights into the activation mechanisms of these MPs based on the removal of the zymogenic segments by limited proteolysis experiments or by design of C-terminal deletion mutants failed. Due to our inability to activate these MPs *in vitro*, we hypothesized that their activation might require either the participation of a specific proteolytic cleavage or a conformational change favored by the special environmental conditions in which the thermophilic organisms that produce them are found.

The globular parts of proabylysin and projannalysin not only encompass the ZBCS HEXXH but also display sequential and structural similarity to distinct active gluzincins that strongly support these short-sized MPs as structural and functional core scaffolds for the clan. Hence, we proposed the name minigluzincins for this newly discovered family whose members show a histidine as a third metal-binding residue (instead of the typical glutamic acid) and a phenylalanine on the "Ser/Gly turn", reflecting variability within the clan comparable to that observed among metzincins (Cerdà-Costa & Gomis-Rüth, 2014; Gomis-Rüth, 2003; Gomis-Rüth, 2009). Notably, minigluzincins bear an even higher structural resemblance to the CDs of some IMMPs from families M48/M56 than to active gluzincins in general, including secondary-structure elements (such as the linking helix and C-terminal helix) that are absent in gluzincins such as thermolysin and cowrins. This closer similarity is consistent with the fact that they were initially identified in the search for structural models of CDs of such IMMPs by bioinformatics.

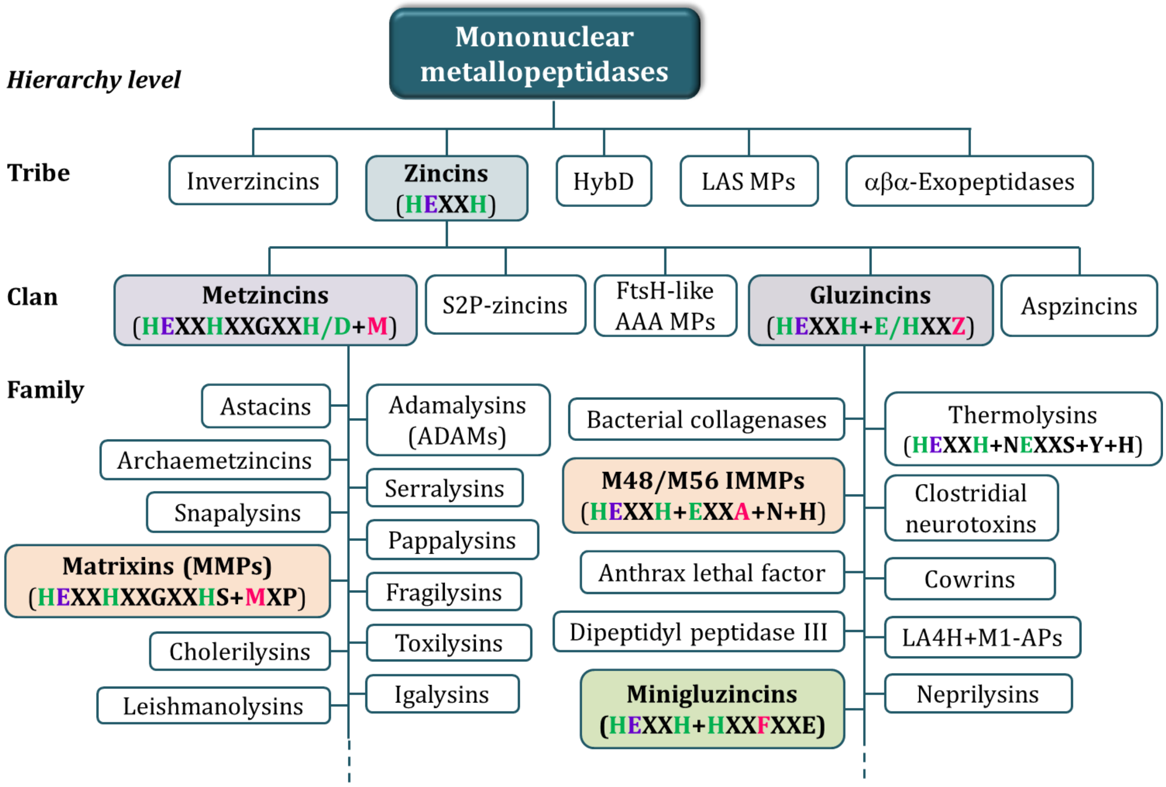


Fig. 22. Description of a novel family of metallopeptidases. Update of the previous classification of mononuclear metallopeptidases shown in Fig. 8 including the recently described family of minigluzincins (highlighted in green). Metal-binding residues are shown in green, general base/acid residues in violet, and residues/molecules occupying the position of the "Met-turn" or "Ser/Gly-turn" beneath the metal site in pink. Other residues engaged in substrate binding, stabilization of the reaction intermediate, and/or catalysis are further shown in black, except for X, which stands for any residue and is only used as a spacer within motifs. Accordingly, the consensus sequence for the gluzincin clan has been modified and Z stands now for A/F/S/G/T. Figure adapted from (Cerdà-Costa & Gomis-Rüth, 2014).

Taken together, the globular parts of the structures described above (leaving aside their C-terminal inhibitory segments) provide a minimal scaffold for the CDs of M48/M56-family IMMPs, into which the specific members have inserted other structural elements, such as transmembrane domains, that confer them particular functionality. Accordingly, this scaffold might contribute to the understanding of the working mechanisms of the above-mentioned IMMPs, and sets the path for the design of inhibitors aimed at modulating their activity. In addition, minigluzincins serve as a core scaffold for the whole gluzincin clan, and more generally, constitute the shortest peptidases structurally characterized so far, displacing the 132-residue-CD of *Streptomyces caespitosus* snapalysin from the metzincin clan (Harada et al, 1995; Kurisu et al, 2000; Trillo-Muyo et al, 2013). Therefore, they hold interest as the minimal set of residues and secondary structure arrangements required for a peptidase to be an independent competent unit able to carry out its catalytic activity.

In close relation with the idea of a minimal MP, the third project characterized the first active member of the minigluzincin family, a selective caseinolytic metallopeptidase from *M. jannaschii*, which we termed selecase. Selecase selectively and specifically cleaves bovine milk casein at a single site of its α_{S1} chain (Y^{159}-F^{160}), which served to design a derived fluorogenic peptidic substrate (Abz-E-L-A-Y-↓-F-Y-P-E-K(dnp)) to carry out activity assays. Unexpectedly, selecase did not follow a typical Michaelis-Menten kinetic reaction but showed a concentration-dependent enzymatic activity that increased upon protein concentration (up to 0.25 mg/ml) and progressively decreased thereafter. Most interestingly, the reaction was reversible given that maximal activity was regained by simple dilution with buffer. Accordingly, we described a reversible enzymatic autoinhibition mechanism triggered solely by changes in concentration and not related to the substrate or any other reagent, which is novel for peptidases.

We further explored the oligomerization of selecase in solution in a wide concentration range using several biophysical techniques, including size exclusion chromatography (SEC), SEC combined with multiangle laser light scattering (SEC-MALLS), crosslinking, analytical ultracentrifugation (AUC), and small-angle X-ray scattering (SAXS), which jointly allowed us to identify the existence of a concentration-dependent-equilibrium between at least four oligomeric species: monomers, dimers, tetramers, and octamers. Oligomerization studies, together with activity assays at different concentrations, pointed to the monomer as the only active species and the rest of oligomers as self-inhibiting species, providing an explanation for the low activity observed at high concentrations. This finding is reminiscent of the decreased activity of phospholipase A2 due to aggregation (Hazlett & Dennis, 1985). We next solved the crystal structures of monomeric, dimeric and tetrameric forms of selecase. Monomeric selecase revealed a compact globular shape divisible into upper NTS and lower CTS via the horizontal central active-site cleft. While most of its secondary-structure elements were common with the structures of its minigluzincin relatives (proabylysin and projannalysin), monomeric selecase displayed a competent overall architecture, metal-binding site and active-site cleft without any C-terminal segment obstructing the substrate access, so it stands out as the smallest active peptidase structurally characterized to date. The protomers of both dimeric and tetrameric selecase showed good overall fit and only minor differences within the NTS compared to monomeric selecase. However, the protomer of dimeric selecase suffered a major rearrangement around the metal-binding site and the CTS, which was even more drastic in the protomer of the tetramer. The association of protomers stabilized the altered structural segments into the respective self-inhibitory particles, which correspond to/coincide with inactive species in solution.

All our attempts to disrupt the interactions responsible for oligomerization through point mutagenesis of selected residues resulted in mutants exhibiting a similar oligomerization capacity to that of the wild-type protein in solution. Moreover, we obtained novel self-inhibiting dimeric conformations in the structures of two of the mutants, which we managed to crystallize, thus proving selecase to be a highly plastic protein. Even two orthlogs of selecase from *Methanotorris igneus* and *Methanocaldococcus fervens*, which could be considered as natural 5-fold and 19-fold point mutants, likewise evinced an equivalent concentration-dependent oligomerization behavior. Altogether, our findings revealed that catalytically active monomers of selecase coexist in equilibrium with incompetent oligomers and that the transition between these discrete states is triggered by rearrangements at the NTS-CTS interface. Both NTS and CTS can thus be considered as individual folding units or foldons, with oligomerization mainly affecting the latter. Importantly, this work succeeded in identifying and characterizing a metamorphic protein (i.e. a protein that spontaneously undergoes substantial fold switching and reversibly transits between several conformers), which is a challenge due to the highly dynamic nature of these molecules (Bryan & Orban, 2010; Goodchild et al, 2011; Murzin, 2008).

Leaving aside its interest as a metamorphic protein with an enzymatic activity reversibly controlled by structural transitions, selecase proved to be an extremely specific peptidase. Additionally, it is small, very stable, and able to function at elevated temperatures and within a fairly wide range of pH values. All these properties make this enzyme suitable for industrial/biotechnological applications involving the specific digestion of a polypeptide chain containing the recognition sequence of this enzyme, i.e. either directly against the α_{S1}-chain of bovine casein or against a polypeptide in which an engineered recognition sequence has been inserted. Accordingly, selecase might be used in several processes that employ casein as a primary source. For instance, in food industry selecase could help improve the organoleptic properties of milk-derived products or reduce the allergenicity of the epitope of the α_{S1}-chain of casein where this enzyme cleaves. Other industrial processes that use casein include the production of casein-derived glues, biofilms/biopolymers, or bioactive peptides. Casein is further used as an additive in several applications (such as in paints, cosmetics, etc.) and therefore the activity of selecase could modify their properties. Furthermore, this enzyme could be used in other biotechnological applications, which require the use of a highly specific cleavage such as the removal of a fusion tag/protein during recombinant protein production. In light of the above, we presented a European Patent application entitled “Novel proteolytic enzyme for specific cleavage and use thereof” to protect and exploit the use of selecase (E.P application No. 13328350.0., 2013).

In order to improve the applicability of selecase in the above-mentioned industrial processes or in laboratory-scale research, and considering the kinetic parameters that we obtained for the fluorogenic peptide cleavage ($K_{cat} = 4 \pm 0.2 \times 10^{-4}$ s^{-1} ; $K_m = 3 \pm 0.3$ μM; $K_{cat}/K_m = 133$ M^{-1} s^{-1}), it could be interesting to subject this enzyme to directed evolution. This kind of experiments would be oriented to obtain mutagenic variants of selecase with qualities that emphasize its properties. Therefore, hypothetic mutants with improved features such as increased activity against the target sequence, increased thermostability, and/or a reduced capacity to oligomerize would be selected. Moreover, any information derived from the characterization of these improved variants would deepen our knowledge concerning the structure-function relationship in selecase and would complement the biochemical and biophysical/structural studies that have been carried out in this last project of the present thesis.

CONCLUDING REMARKS

Project 1: *"A Novel Mechanism of Latency in Matrix Metalloproteinases"*

We have characterized a unique propeptide in MMPs. The bacterial MMP karilysin comprises the shortest PD reported for an MP (14-residue propeptide), which inhibits the CD through an "aspartate-switch" following a latency mechanism reminiscent of that of unrelated astacin and fragilysin MP families (Fig. 23). This propeptide is essential for bacterial protein expression and significantly contributes to protein stability.

Project 2: *"A Novel Family of Soluble Minimal Scaffolds Provides Structural Insight into the Catalytic Domains of Integral Membrane Metallopeptidases"*

We have described a novel family of minimal MPs. Newly discovered soluble, small-sized MPs (called minigluzincins) show two latency mechanisms through the C-terminal segments previously unseen in MPs: intramolecular in proabylysin, and crosswise intermolecular through dimerization in projannalysin (Fig. 23). Minigluzincins share large similarity with IMMPs, which are provided with TMHs flanking or inserted into a minigluzincin-like CD.

Project 3: *"Multiple Stable Conformations Account for Reversible Concentration-Dependent Oligomerization and Autoinhibition of a Metamorphic Metallopeptidase"*

Finally, we have discovered an MP showing molecular plasticity. The metamorphic, selective, and specific caseinolytic minigluzincin, selecase, reversibly transits between four states of defined three-dimensional structure (monomer, dimer, and tetramer represented in Fig. 23, *plus* octamer). The competent catalytic conformation is sequestered in incompetent but structured dimers, tetramers, and octamers, which are associated with loss of enzymatic activity due to autoinhibition.

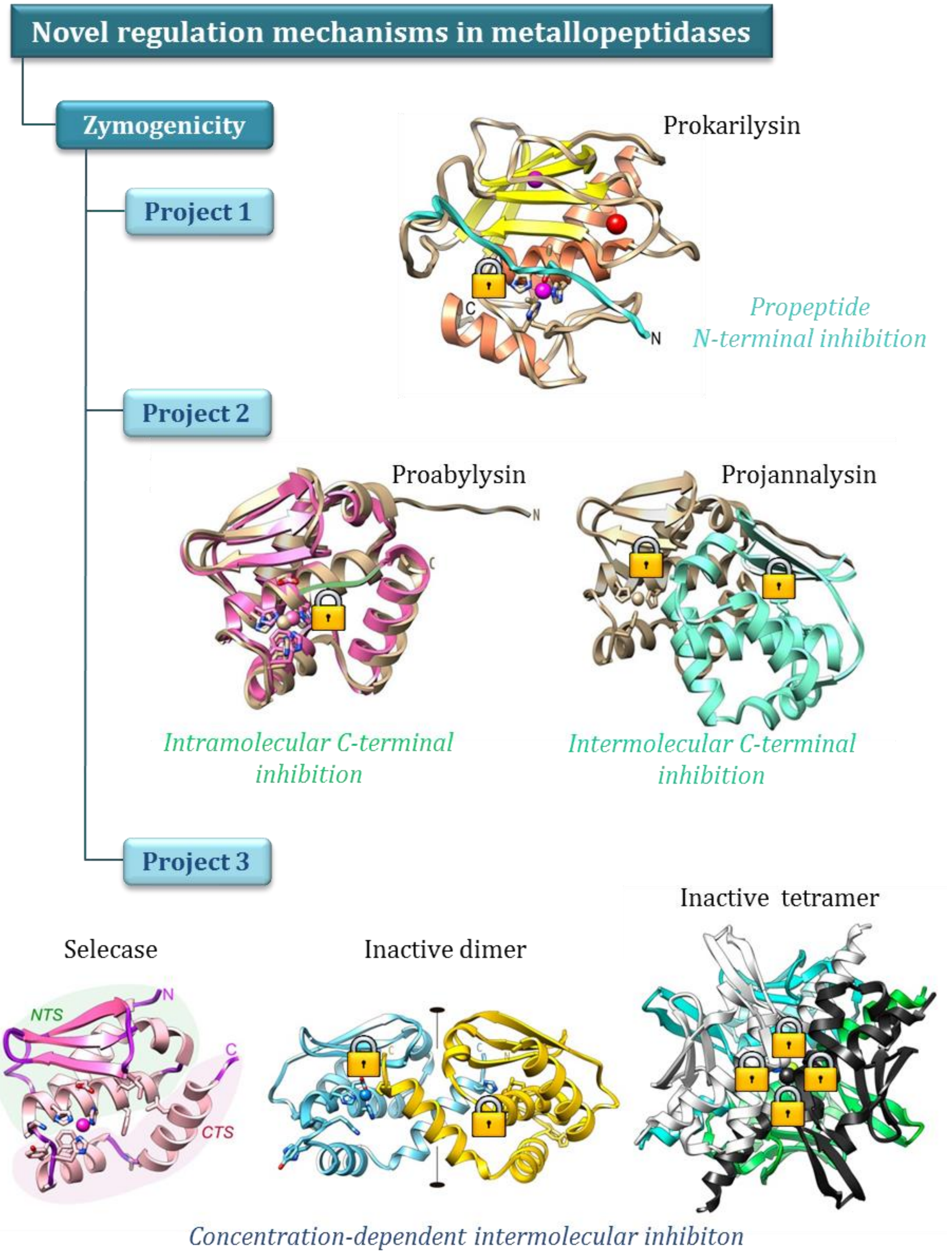


Fig. 23. Mechanisms of proteolytic activity regulation described along the present thesis. Schematic overview of a unique propeptide in matrix metalloproteinases (Project 1) and intra/intermolecular interactions in a novel family of minimal gluzincins (Project 2 and Project 3).

GENERAL REFERENCES

A

Akiyama Y (2009) Quality control of cytoplasmic membrane proteins in Escherichia coli. *J Biochem* **146:** 449-454

Anand R, Wai T, Baker MJ, Kladt N, Schauss AC, Rugarli E, Langer T (2014) The i-AAA protease YME1L and OMA1 cleave OPA1 to balance mitochondrial fusion and fission. *J Cell Biol* **204:** 919-929

Arolas JL, García-Castellanos R, Goulas T, Akiyama Y, Gomis-Rüth FX (2014) Expression and purification of integral membrane metallopeptidase HtpX. *Protein Expr Purif* **99:** 113-118

B

Bai XC, Yan C, Yang G, Lu P, Ma D, Sun L, Zhou R, Scheres SH, Shi Y (2015) An atomic structure of human gamma-secretase. *Nature* **525:** 212-217

Baliga BC, Colussi PA, Read SH, Dias MM, Jans DA, Kumar S (2003) Role of prodomain in importin-mediated nuclear localization and activation of caspase-2. *J Biol Chem* **278:** 4899-4905

Baral PK, Jajcanin-Jozic N, Deller S, Macheroux P, Abramic M, Gruber K (2008) The first structure of dipeptidyl-peptidase III provides insight into the catalytic mechanism and mode of substrate binding. *J Biol Chem* **283:** 22316-22324

Barrett AJ, McDonald JK (1986) Nomenclature: protease, proteinase and peptidase. *Biochem J* **237:** 935

Barrett AJ, Rawlings ND, Salvesen G, Fred Woessner J (2013) Introduction. In *Handbook of Proteolytic Enzymes*, Salvesen NDR (ed). Academic Press

Barrowman J, Michaelis S (2009) ZMPSTE24, an integral membrane zinc metalloprotease with a connection to progeroid disorders. *Biol Chem* **390:** 761-773

Becker AB, Roth RA (1992) An unusual active site identified in a family of zinc metalloendopeptidases. *Proc Natl Acad Sci U S A* **89:** 3835-3839

Becker JW, Marcy AI, Rokosz LL, Axel MG, Burbaum JJ, Fitzgerald PM, Cameron PM, Esser CK, Hagmann WK, Hermes JD, et al. (1995) Stromelysin-1: three-dimensional structure of the inhibited catalytic domain and of the C-truncated proenzyme. *Protein Sci* **4:** 1966-1976

Beinrohr L, Dobo J, Zavodszky P, Gal P (2008) C1, MBL-MASPs and C1-inhibitor: novel approaches for targeting complement-mediated inflammation. *Trends Mol Med* **14:** 511-521

Bertenshaw GP, Norcum MT, Bond JS (2003) Structure of homo- and hetero-oligomeric meprin metalloproteases. Dimers, tetramers, and high molecular mass multimers. *J Biol Chem* **278:** 2522-2532

Beynon RJ, Bond JS (2001) *Proteolytic Enzymes : a Practical Approach*, 2nd edn. Oxford ; New York: Oxford University Press.

Bieniossek C, Schalch T, Bumann M, Meister M, Meier R, Baumann U (2006) The molecular architecture of the metalloprotease FtsH. *Proc Natl Acad Sci U S A* **103:** 3066-3071

Bitar AP, Cao M, Marquis H (2008) The metalloprotease of Listeria monocytogenes is activated by intramolecular autocatalysis. *J Bacteriol* **190:** 107-111

Bochtler M, Odintsov SG, Marcyjaniak M, Sabala I (2004) Similar active sites in lysostaphins and D-Ala-D-Ala metallopeptidases. *Protein Sci* **13:** 854-861

Bode W, Gomis-Rüth FX, Huber R, Zwilling R, Stöcker W (1992) Structure of astacin and implications for activation of astacins and zinc-ligation of collagenases. *Nature* **358:** 164-167

Bode W, Gomis-Rüth FX, Stockler W (1993) Astacins, serralysins, snake venom and matrix metalloproteinases exhibit identical zinc-binding environments (HEXXHXXGXXH and Met-turn) and topologies and should be grouped into a common family, the 'metzincins'. *FEBS Lett* **331:** 134-140

Bode W, Huber R (2000) Structural basis of the endoproteinase-protein inhibitor interaction. *Biochim Biophys Acta* **1477:** 241-252

Braun P, Bitter W, Tommassen J (2000) Activation of Pseudomonas aeruginosa elastase in Pseudomonas putida by triggering dissociation of the propeptide-enzyme complex. *Microbiology* **146 (Pt 10):** 2565-2572

Brew K, Dinakarpandian D, Nagase H (2000) Tissue inhibitors of metalloproteinases: evolution, structure and function. *Biochim Biophys Acta* **1477:** 267-283

Bryan PN (2002) Prodomains and protein folding catalysis. *Chem Rev* **102:** 4805-4816

Bryan PN, Orban J (2010) Proteins that switch folds. *Curr Opin Struct Biol* **20:** 482-488

Bunning P, Holmquist B, Riordan JF (1983) Substrate specificity and kinetic characteristics of angiotensin converting enzyme. *Biochemistry* **22:** 103-110

C

Carrijo LC, Andrich F, de Lima ME, Cordeiro MN, Richardson M, Figueiredo SG (2005) Biological properties of the venom from the scorpionfish (Scorpaena plumieri) and purification of a gelatinolytic protease. *Toxicon* **45:** 843-850

Cerdà-Costa N, Gomis-Rüth FX (2014) Architecture and function of metallopeptidase catalytic domains. *Protein Sci* **23:** 123-144

Cerdà-Costa N, Guevara T, Karim AY, Ksiazek M, Nguyen KA, Arolas JL, Potempa J, Gomis-Rüth FX (2011) The structure of the catalytic domain of Tannerella forsythia karilysin reveals it is a bacterial xenologue of animal matrix metalloproteinases. *Mol Microbiol* **79:** 119-132

Chang AK, Park JW, Lee EH, Lee JS (2007) The N-terminal propeptide of Vibrio vulnificus extracellular metalloprotease is both an inhibitor of and a substrate for the enzyme. *J Bacteriol* **189:** 6832-6838

Chen VM, Ahamed J, Versteeg HH, Berndt MC, Ruf W, Hogg PJ (2006) Evidence for activation of tissue factor by an allosteric disulfide bond. *Biochemistry* **45:** 12020-12028

Chen YJ, Inouye M (2008) The intramolecular chaperone-mediated protein folding. *Curr Opin Struct Biol* **18:** 765-770

D

Dechat T, Adam SA, Taimen P, Shimi T, Goldman RD (2010) Nuclear lamins. *Cold Spring Harb Perspect Biol* **2:** a000547

Demidyuk IV, Gasanov EV, Safina DR, Kostrov SV (2008) Structural organization of precursors of thermolysin-like proteinases. *Protein J* **27:** 343-354

Demidyuk IV, Gromova TY, Polyakov KM, Melik-Adamyan WR, Kuranova IP, Kostrov SV (2010a) Crystal structure of the protealysin precursor: insights into propeptide function. *J Biol Chem* **285:** 2003-2013

Demidyuk IV, Shubin AV, Gasanov EV, Kostrov SV (2010b) Propeptides as modulators of functional activity of proteases. *Biomol Concepts* **1:** 305-322

Deu E, Verdoes M, Bogyo M (2012) New approaches for dissecting protease functions to improve probe development and drug discovery. *Nat Struct Mol Biol* **19:** 9-16

Dittmer TA, Misteli T (2011) The lamin protein family. *Genome Biol* **12:** 222

Drag M, Salvesen GS (2010) Emerging principles in protease-based drug discovery. *Nat Rev Drug Discov* **9:** 690-701

E

Eckhard U, Schonauer E, Nuss D, Brandstetter H (2011) Structure of collagenase G reveals a chew-and-digest mechanism of bacterial collagenolysis. *Nat Struct Mol Biol* **18:** 1109-1114

Eder J, Fersht AR (1995) Pro-sequence-assisted protein folding. *Mol Microbiol* **16:** 609-614

Eley BM, Cox SW (2003) Proteolytic and hydrolytic enzymes from putative periodontal pathogens: characterization, molecular genetics, effects on host defenses and tissues and detection in gingival crevice fluid. *Periodontol 2000* **31:** 105-124

Elkins PA, Ho YS, Smith WW, Janson CA, D'Alessio KJ, McQueney MS, Cummings MD, Romanic AM (2002) Structure of the C-terminally truncated human ProMMP9, a gelatin-binding matrix metalloproteinase. *Acta Crystallogr D Biol Crystallogr* **58:** 1182-1192

F

Fanjul-Fernández M, Folgueras AR, Cabrera S, López-Otín C (2010) Matrix metalloproteinases: evolution, gene regulation and functional analysis in mouse models. *Biochim Biophys Acta* **1803:** 3-19

Farady CJ, Craik CS (2010) Mechanisms of macromolecular protease inhibitors. *ChemBioChem* **11:** 2341-2346

Feng L, Yan H, Wu Z, Yan N, Wang Z, Jeffrey PD, Shi Y (2007) Structure of a site-2 protease family intramembrane metalloprotease. *Science* **318:** 1608-1612

Fox JW, Serrano SM (2009) Timeline of key events in snake venom metalloproteinase research. *J Proteomics* **72:** 200-209

Freije JM, Balbin M, Pendas AM, Sanchez LM, Puente XS, López-Otín C (2003) Matrix metalloproteinases and tumor progression. *Adv Exp Med Biol* **532:** 91-107

Fremont DH, Anderson DH, Wilson IA, Dennis EA, Xuong NH (1993) Crystal structure of phospholipase A2 from Indian cobra reveals a trimeric association. *Proc Natl Acad Sci USA* **90:** 342-346

Fritsche E, Paschos A, Beisel HG, Bock A, Huber R (1999) Crystal structure of the hydrogenase maturating endopeptidase HYBD from Escherichia coli. *J Mol Biol* **288:** 989-998

Fujimura-Kamada K, Nouvet FJ, Michaelis S (1997) A novel membrane-associated metalloprotease, Ste24p, is required for the first step of NH2-terminal processing of the yeast a-factor precursor. *J Cell Biol* **136:** 271-285

Fujinaga M, Cherney MM, Oyama H, Oda K, James MN (2004) The molecular structure and catalytic mechanism of a novel carboxyl peptidase from Scytalidium lignicolum. *Proc Natl Acad Sci U S A* **101:** 3364-3369

Fushimi N, Ee CE, Nakajima T, Ichishima E (1999) Aspzincin, a family of metalloendopeptidases with a new zinc-binding motif. Identification of new zinc-binding sites (His(128), His(132), and Asp(164)) and three catalytically crucial residues (Glu(129), Asp(143), and Tyr(106)) of deuterolysin from Aspergillus oryzae by site-directed mutagenesis. *J Biol Chem* **274:** 24195-24201

G

Gao X, Wang J, Yu DQ, Bian F, Xie BB, Chen XL, Zhou BC, Lai LH, Wang ZX, Wu JW, Zhang YZ (2010) Structural basis for the autoprocessing of zinc metalloproteases in the thermolysin family. *Proc Natl Acad Sci U S A* **107:** 17569-17574

Garcia-Ferrer I, Arede P, Gómez-Blanco J, Luque D, Duquerroy S, Castón JR, Goulas T, Gomis-Rüth FX (2015) Structural and functional insights into Escherichia coli alpha2-macroglobulin endopeptidase snap-trap inhibition. *Proc Natl Acad Sci U S A* **112:** 8290-8295

Georgiadis D, Dive V (2015) Phosphinic peptides as potent inhibitors of zinc-metalloproteases. *Top Curr Chem* **360:** 1-38

Gómez-Ortiz M, Gomis-Rüth FX, Huber R, Avilés FX (1997) Inhibition of carboxypeptidase A by excess zinc: analysis of the structural determinants by X-ray crystallography. *FEBS Lett* **400:** 336-340

Gomis-Rüth FX (2003) Structural aspects of the metzincin clan of metalloendopeptidases. *Mol Biotechnol* **24:** 157-202

Gomis-Rüth FX (2008) Structure and mechanism of metallocarboxypeptidases. *Crit Rev Biochem Mol Biol* **43:** 319-345

Gomis-Rüth FX (2009) Catalytic domain architecture of metzincin metalloproteases. *J Biol Chem* **284:** 15353-15357

Gomis-Rüth FX, Botelho TO, Bode W (2012) A standard orientation for metallopeptidases. *Biochim Biophys Acta* **1824:** 157-163

Gomis-Rüth FX, Kress LF, Bode W (1993) First structure of a snake venom metalloproteinase: a prototype for matrix metalloproteinases/collagenases. *EMBO J* **12:** 4151-4157

Goodchild S, Curmi PG, Brown L (2011) Structural gymnastics of multifunctional metamorphic proteins. *Biophys Rev* **3:** 143-153

Goulas T, Arolas JL, Gomis-Rüth FX (2011) Structure, function and latency regulation of a bacterial enterotoxin potentially derived from a mammalian adamalysin/ADAM xenolog. *Proc Natl Acad Sci U S A* **108:** 1856-1861

Gromova TY, Demidyuk IV, Kozlovskiy VI, Kuranova IP, Kostrov SV (2009) Processing of protealysin precursor. *Biochimie* **91:** 639-645

Gross J (2004) How tadpoles lose their tails: path to discovery of the first matrix metalloproteinase. *Matrix Biol* **23:** 3-13

Gross J, Lapière CM (1962) Collagenolytic activity in amphibian tissues: a tissue culture assay. *Proc Natl Acad Sci U S A* **48:** 1014-1022

Grötzinger J, Rose-John S (2013) ADAM proteases in physiology and pathophysiology: cleave to function in health or to cause disease. In *Proteases: Structure and Function*, Brix K, Stöcker W (eds), pp 303-318. Springer Vienna

Guevara T, Ksiazek M, Skottrup PD, Cerdà-Costa N, Trillo-Muyo S, de Diego I, Riise E, Potempa J, Gomis-Rüth FX (2013) Structure of the catalytic domain of the Tannerella forsythia matrix metallopeptidase karilysin in complex with a tetrapeptidic inhibitor. *Acta Crystallogr Sect F Struct Biology Cryst Commun* **69:** 472-476

Guevara T, Yiallouros I, Kappelhoff R, Bissdorf S, Stöcker W, Gomis-Rüth FX (2010) Proenzyme structure and activation of astacin metallopeptidase. *J Biol Chem* **285:** 13958-13965

H

Hamada K, Hata Y, Katsuya Y, Hiramatsu H, Fujiwara T, Katsube Y (1996) Crystal structure of Serratia protease, a zinc-dependent proteinase from Serratia sp. E-15, containing a beta-sheet coil motif at 2.0 A resolution. *J Biochem* **119:** 844-851

Harada S, Kinoshita T, Kasai N, Tsunasawa S, Sakiyama F (1995) Complete amino acid sequence of a zinc metalloendoprotease from Streptomyces caespitosus. *Eur J Biochem* **233:** 683-686

Hazlett TL, Dennis EA (1985) Aggregation studies on fluorescein-coupled cobra venom phospholipase A2. *Biochemistry* **24:** 6152-6158

Holmquist B, Vallee BL (1974) Metal substitutions and inhibition of thermolysin: spectra of the cobalt enzyme. *J Biol Chem* **249:** 4601-4607

Hooper NM (1994) Families of zinc metalloproteases. *FEBS Lett* **354:** 1-6

Hoylaerts M, Rijken DC, Lijnen HR, Collen D (1982) Kinetics of the activation of plasminogen by human tissue plasminogen activator. Role of fibrin. *J Biol Chem* **257:** 2912-2919

J

Jordans S, Jenko-Kokalj S, Kuhl NM, Tedelind S, Sendt W, Bromme D, Turk D, Brix K (2009) Monitoring compartment-specific substrate cleavage by cathepsins B, K, L, and S at physiological pH and redox conditions. *BMC Biochem* **10:** 23

Jozic D, Bourenkov G, Lim NH, Visse R, Nagase H, Bode W, Maskos K (2005) X-ray structure of human proMMP-1: new insights into procollagenase activation and collagen binding. *J Biol Chem* **280:** 9578-9585

Jusko M, Potempa J, Karim AY, Ksiazek M, Riesbeck K, Garred P, Eick S, Blom AM (2012) A metalloproteinase karilysin present in the majority of Tannerella forsythia isolates inhibits all pathways of the complement system. *J Immunol* **188:** 2338-2349

K

Karim AY, Kulczycka M, Kantyka T, Dubin G, Jabaiah A, Daugherty PS, Thogersen IB, Enghild JJ, Nguyen KA, Potempa J (2010) A novel matrix metalloprotease-like enzyme (karilysin) of the periodontal pathogen Tannerella forsythia ATCC 43037. *Biol Chem* **391:** 105-117

Kaser M, Kambacheld M, Kisters-Woike B, Langer T (2003) Oma1, a novel membrane-bound metallopeptidase in mitochondria with activities overlapping with the m-AAA protease. *J Biol Chem* **278:** 46414-46423

Kawamoto S, Shibano Y, Fukushima J, Ishii N, Morihara K, Okuda K (1993) Site-directed mutagenesis of Glu-141 and His-223 in Pseudomonas aeruginosa elastase: catalytic activity, processing, and protective activity of the elastase against Pseudomonas infection. *Infect Immun* **61:** 1400-1405

Ketelhuth DF, Back M (2011) The role of matrix metalloproteinases in atherothrombosis. *Curr Atheroscler Rep* **13:** 162-169

Khan AR, James MN (1998) Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. *Protein Sci* **7:** 815-836

Kim AC, Oliver DC, Paetzel M (2008) Crystal structure of a bacterial signal Peptide peptidase. *J Mol Biol* **376:** 352-366

Koziel J, Karim AY, Przybyszewska K, Ksiazek M, Rapala-Kozik M, Nguyen KA, Potempa J (2010) Proteolytic inactivation of LL-37 by karilysin, a novel virulence mechanism of Tannerella forsythia. *J Innate Immun* **2:** 288-293

Koziel J, Potempa J (2013) Protease-armed bacteria in the skin. *Cell Tissue Res* **351:** 325-337

Kurisu G, Kai Y, Harada S (2000) Structure of the zinc-binding site in the crystal structure of a zinc endoprotease from Streptomyces caespitosus at 1 A resolution. *J Inorg Biochem* **82:** 225-228

Kurisu G, Kinoshita T, Sugimoto A, Nagara A, Kai Y, Kasai N, Harada S (1997) Structure of the zinc endoprotease from Streptomyces caespitosus. *J Biochem* **121:** 304-308

Kyrieleis OJ, Goettig P, Kiefersauer R, Huber R, Brandstetter H (2005) Crystal structures of the tricorn interacting factor F3 from Thermoplasma acidophilum, a zinc aminopeptidase in three different conformations. *J Mol Biol* **349:** 787-800

L

Langklotz S, Baumann U, Narberhaus F (2012) Structure and function of the bacterial AAA protease FtsH. *Biochim Biophys Acta* **1823:** 40-48

Lapière Ch M (2005) Tadpole collagenase, the single parent of such a large family. *Biochimie* **87:** 243-247

Laskowski M, Jr., Kato I (1980) Protein inhibitors of proteinases. *Annu Rev Biochem* **49:** 593-626

Laskowski RA, Gerick F, Thornton JM (2009) The structural basis of allosteric regulation in proteins. *FEBS Lett* **583:** 1692-1698

Lipscomb WN, Hartsuck JA, Reeke GN, Jr., Quiocho FA, Bethge PH, Ludwig ML, Steitz TA, Muirhead H, Coppola JC (1968) The structure of carboxypeptidase A. VII. The 2.0-angstrom resolution studies of the enzyme and of its complex with glycyltyrosine, and mechanistic deductions. *Brookhaven Symp Biol* **21:** 24-90

López-Otín C, Bond JS (2008) Proteases: multifunctional enzymes in life and disease. *J Biol Chem* **283:** 30433-30437

López-Pelegrín M, Cerdà-Costa N, Cintas-Pedrola A, Arolas JL, Gomis-Rüth FX (2013) Novel Proteolytic Enzyme for Specific Cleavage and use thereof. European Patent Application 13382350.0.

Luthy JA, Praissman M, Finkenstadt WR, Laskowski M, Jr. (1973) Detailed mechanism of interaction of bovine -trypsin with soybean trypsin inhibitor (Kunitz). I. Stopped flow measurements. *J Biol Chem* **248:** 1760-1771

M

Mach L (2002) Biosynthesis of lysosomal proteinases in health and disease. *Biol Chem* **383:** 751-756

Malemud CJ (2006) Matrix metalloproteinases (MMPs) in health and disease: an overview. *Front Biosci* **11:** 1696-1701

Mallorquí-Fernández G, Marrero A, García-Pique S, García-Castellanos R, Gomis-Rüth FX (2004) Staphylococcal methicillin resistance: fine focus on folds and functions. *FEMS Microbiol Lett* **235:** 1-8

Mallya SK, Van Wart HE (1989) Mechanism of inhibition of human neutrophil collagenase by Gold(I) chrysotherapeutic compounds. Interaction at a heavy metal binding site. *J Biol Chem* **264:** 1594-1601

Mansfeld J, Petermann E, Durrschmidt P, Ulbrich-Hofmann R (2005) The propeptide is not required to produce catalytically active neutral protease from Bacillus stearothermophilus. *Protein Expr Purif* **39:** 219-228

Marcyjaniak M, Odintsov SG, Sabala I, Bochtler M (2004) Peptidoglycan amidase MepA is a LAS metallopeptidase. *J Biol Chem* **279:** 43982-43989

Marie-Claire C, Roques BP, Beaumont A (1998) Intramolecular processing of prothermolysin. *J Biol Chem* **273:** 5697-5701

Marrero A, Duquerroy S, Trapani S, Goulas T, Guevara T, Andersen GR, Navaza J, Sottrup-Jensen L, Gomis-Rüth FX (2012) The crystal structure of human alpha2-macroglobulin reveals a unique molecular cage. *Angew Chem Int Ed Engl* **51:** 3340-3344

Marrero A, Mallorquí-Fernández G, Guevara T, García-Castellanos R, Gomis-Rüth FX (2006) Unbound and acylated structures of the MecR1 extracellular antibiotic-sensor domain provide insights into the signal-transduction system that triggers methicillin resistance. *J Mol Biol* **361:** 506-521

Matthews BW (1988) Structural basis of the action of thermolysin and related zinc peptidases. *Acc Chem Res* **21** (9)**:**333-340

Matthews BW, Jansonius JN, Colman PM, Schoenborn BP, Dupourque D (1972) Three-dimensional structure of thermolysin. *Nat New Biol* **238:** 37-41

McGwire BS, Chang KP (1996) Posttranslational regulation of a Leishmania HEXXH metalloprotease (gp63). The effects of site-specific mutagenesis of catalytic, zinc binding, N-glycosylation, and glycosyl phosphatidylinositol addition sites on N-terminal end cleavage, intracellular stability, and extracellular exit. *J Biol Chem* **271:** 7903-7909

McIver K, Kessler E, Ohman DE (1991) Substitution of active-site His-223 in Pseudomonas aeruginosa elastase and expression of the mutated lasB alleles in Escherichia coli show evidence for autoproteolytic processing of proelastase. *J Bacteriol* **173:** 7781-7789

Milla ME, Gonzales PE, Leonard JD (2006) The TACE zymogen: re-examining the role of the cysteine switch. *Cell Biochem Biophys* **44:** 342-348

Miyoshi S, Wakae H, Tomochika K, Shinoda S (1997) Functional domains of a zinc metalloprotease from Vibrio vulnificus. *J Bacteriol* **179:** 7606-7609

Mock WL, Wang L (1999) Synergistic inhibition of carboxypeptidase A by zinc ion and imidazole. *Biochem Biophys Res Commun* **257:** 239-243

Montecucco C, Schiavo G (1993) Tetanus and botulism neurotoxins: a new group of zinc proteases. *Trends Biochem Sci* **18:** 324-327

Morgunova E, Tuuttila A, Bergmann U, Isupov M, Lindqvist Y, Schneider G, Tryggvason K (1999) Structure of human pro-matrix metalloproteinase-2: activation mechanism revealed. *Science* **284:** 1667-1670

Murphy G, Nagase H (2008) Reappraising metalloproteinases in rheumatoid arthritis and osteoarthritis: destruction or repair? *Nat Clin Pract Rheumatol* **4:** 128-135

Murzin AG (2008) Biochemistry. Metamorphic proteins. *Science* **320:** 1725-1726

N

Nagase H (1997) Activation mechanisms of matrix metalloproteinases. *Biol Chem* **378:** 151-160

Nagase H (2001) Metalloproteases. *Current protocols in protein science / editorial board, John E Coligan [et al]* **Chapter 21:** Unit 21 24

Nagase H, Murphy G (2013) Metalloproteinases in cartilage matrix breakdown: the roles in rheumatoid arthritis and osteoarthritis. In *Proteases: Structure and Function*, Brix K, Stöcker W (eds), pp 433-469. Springer Vienna

Nagase H, Visse R, Murphy G (2006) Structure and function of matrix metalloproteinases and TIMPs. *Cardiovasc Res* **69:** 562-573

Nagase H, Woessner JF, Jr. (1999) Matrix metalloproteinases. *J Biol Chem* **274:** 21491-21494

Nalivaeva NN, Belyaev ND, Zhuravin IA, Turner AJ (2012) The Alzheimer's amyloid-degrading peptidase, neprilysin: can we control it? *Int J Alzheimers Dis* **2012:** 383796

Nesheim ME, Kettner C, Shaw E, Mann KG (1981) Cofactor dependence of factor Xa incorporation into the prothrombinase complex. *J Biol Chem* **256:** 6537-6540

Neurath H, Walsh KA (1976) Role of proteolytic enzymes in biological regulation (a review). *Proc Natl Acad Sci U S A* **73:** 3825-3832

Ng NM, Littler DR, Paton AW, Le Nours J, Rossjohn J, Paton JC, Beddoe T (2013) EcxAB is a founding member of a new family of metalloprotease AB5 toxins with a hybrid cholera-like B subunit. *Structure* **21:** 2003-2013

Nickerson NN, Joag V, McGavin MJ (2008) Rapid autocatalytic activation of the M4 metalloprotease aureolysin is controlled by a conserved N-terminal fungalysin-thermolysin-propeptide domain. *Mol Microbiol* **69:** 1530-1543

Nissinen L, Kahari VM (2014) Matrix metalloproteinases in inflammation. *Biochim Biophys Acta* **1840:** 2571-2580

Noel A, Sounni NE (2013) MMP-mediated collagen remodeling and vessel functions. In *Proteases: Structure and Function,* Brix K, Stöcker W (eds), pp 471-489. Springer Vienna

Novinec M, Lenarcic B, Turk B (2014) Cysteine cathepsin activity regulation by glycosaminoglycans. *Biomed Res Int* **2014:** 309718

O

O'Donohue MJ, Beaumont A (1996) The roles of the prosequence of thermolysin in enzyme inhibition and folding in vitro. *J Biol Chem* **271:** 26477-26481

Oefner C, D'Arcy A, Hennig M, Winkler FK, Dale GE (2000) Structure of human neutral endopeptidase (Neprilysin) complexed with phosphoramidon. *J Mol Biol* **296:** 341-349

Overall CM, López-Otín C (2002) Strategies for MMP inhibition in cancer: innovations for the post-trial era. *Nat Rev Cancer* **2:** 657-672

P, Q

Pannifer AD, Wong TY, Schwarzenbacher R, Renatus M, Petosa C, Bienkowska J, Lacy DB, Collier RJ, Park S, Leppla SH, Hanna P, Liddington RC (2001) Crystal structure of the anthrax lethal factor. *Nature* **414:** 229-233

Petersen PE, Ogawa H (2005) Strengthening the prevention of periodontal disease: the WHO approach. *J Periodontol* **76:** 2187-2193

Potempa J, Pike RN (2005) Bacterial peptidases. *Contrib Microbiol* **12:** 132-180

Pryor EE, Jr., Horanyi PS, Clark KM, Fedoriw N, Connelly SM, Koszelak-Rosenblum M, Zhu G, Malkowski MG, Wiener MC, Dumont ME (2013) Structure of the integral membrane protein CAAX protease Ste24p. *Science* **339:** 1600-1604

Pungercar JR, Caglic D, Sajid M, Dolinar M, Vasiljeva O, Pozgan U, Turk D, Bogyo M, Turk V, Turk B (2009) Autocatalytic processing of procathepsin B is triggered by proenzyme activity. *FEBS J* **276:** 660-668

Quigley A, Dong YY, Pike AC, Dong L, Shrestha L, Berridge G, Stansfeld PJ, Sansom MS, Edwards AM, Bountra C, von Delft F, Bullock AN, Burgess-Brown NA, Carpenter EP (2013) The structural basis of ZMPSTE24-dependent laminopathies. *Science* **339:** 1604-1607

Quirós PM, Ramsay AJ, López-Otín C (2013) New roles for OMA1 metalloprotease: From mitochondrial proteostasis to metabolic homeostasis. *Adipocyte* **2:** 7-11

R

Ra HJ, Parks WC (2007) Control of matrix metalloproteinase catalytic activity. *Matrix Biol* **26:** 587-596

Rawlings ND (2010) Peptidase inhibitors in the MEROPS database. *Biochimie* **92:** 1463-1483

Rawlings ND, Barrett AJ (1993) Evolutionary families of peptidases. *Biochem J* **290 (Pt 1):** 205-218

Rawlings ND, Barrett AJ (1999) MEROPS: the peptidase database. *Nucleic Acids Res* **27:** 325-331

Rawlings ND, Barrett AJ (2013) Chapter 77 - Introduction: Metallopeptidases and their clans. In *Handbook of Proteolytic Enzymes*, Salvesen NDR (ed), pp 325-370. Academic Press

Rawlings ND, Barrett AJ, Bateman A (2011) Asparagine peptide lyases: a seventh catalytic type of proteolytic enzymes. *J Biol Chem* **286:** 38321-38328

Rawlings ND, Barrett AJ, Bateman A (2014a) Using the MEROPS Database for Proteolytic Enzymes and Their Inhibitors and Substrates. *Curr Protoc Bioinformatics.* **48:** 1.25.1-1.25.33.

Rawlings ND, Tolle DP, Barrett AJ (2004) Evolutionary families of peptidase inhibitors. *Biochem J* **378:** 705-716

Rawlings ND, Waller M, Barrett AJ, Bateman A (2014b) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* **42:** D503-509

Rawson RB (2013) The site-2 protease. *Biochim Biophys Acta* **1828:** 2801-2807

Reyda S, Jacob E, Zwilling R, Stöcker W (1999) cDNA cloning, bacterial expression, in vitro renaturation and affinity purification of the zinc endopeptidase astacin. *Biochem J* **344 Pt 3:** 851-857

Riedl SJ, Salvesen GS (2007) The apoptosome: signalling platform of cell death. *Nat Rev Mol Cell Biol* **8:** 405-413

Rosenblum G, Meroueh S, Toth M, Fisher JF, Fridman R, Mobashery S, Sagi I (2007) Molecular structures and dynamics of the stepwise activation mechanism of a matrix metalloproteinase zymogen: challenging the cysteine switch dogma. *J Am Chem Soc* **129:** 13566-13574

Rosing J, Tans G, Govers-Riemslag JW, Zwaal RF, Hemker HC (1980) The role of phospholipids and factor Va in the prothrombinase complex. *J Biol Chem* **255:** 274-283

Rusmili MR, Yee TT, Mustafa MR, Hodgson WC, Othman I (2014) Proteomic characterization and comparison of Malaysian Bungarus candidus and Bungarus fasciatus venoms. *J Proteomics* **110C:** 129-144

S

Sabotic J, Kos J (2012) Microbial and fungal protease inhibitors--current and potential applications. *Appl Microbiol Biotechnol* **93:** 1351-1375

Saito T, Ishihara K, Kato T, Okuda K (1997) Cloning, expression, and sequencing of a protease gene from Bacteroides forsythus ATCC 43037 in Escherichia coli. *Infect Immun* **65:** 4888-4891

Sakoh M, Ito K, Akiyama Y (2005) Proteolytic activity of HtpX, a membrane-bound and stress-controlled protease from Escherichia coli. *J Biol Chem* **280:** 33305-33310

Schechter I, Berger A (1967) On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun* **27:** 157-162

Schechter NM, Eng GY, Selwood T, McCaslin DR (1995) Structural changes associated with the spontaneous inactivation of the serine proteinase human tryptase. *Biochemistry* **34:** 10628-10638

Schlagenhauf E, Etges R, Metcalf P (1998) The crystal structure of the Leishmania major surface proteinase leishmanolysin (gp63). *Structure* **6:** 1035-1046

Schuijt TJ, Bakhtiari K, Daffre S, Deponte K, Wielders SJ, Marquart JA, Hovius JW, van der Poll T, Fikrig E, Bunce MW, Camire RM, Nicolaes GA, Meijers JC, van 't Veer C (2013) Factor Xa activation of factor V is of paramount importance in initiating the coagulation system: lessons from a tick salivary protein. *Circulation* **128:** 254-266

Seemuller E, Lupas A, Stock D, Lowe J, Huber R, Baumeister W (1995) Proteasome from Thermoplasma acidophilum: a threonine protease. *Science* **268:** 579-582

Sharma A (2010) Virulence mechanisms of Tannerella forsythia. *Periodontol 2000* **54:** 106-116

Shen A (2010) Allosteric regulation of protease activity by small molecules. *Mol Biosyst* **6:** 1431-1443

Sobhanifar S, Prehna G, Strynadka NCJ (2013) Chapter 277 - BlaR1 and MecR1 gene products of Staphylococcus aureus. In *Handbook of Proteolytic Enzymes*, Salvesen NDR (ed), pp 1237-1242. Academic Press

Socransky SS, Haffajee AD, Cugini MA, Smith C, Kent RL, Jr. (1998) Microbial complexes in subgingival plaque. *J Clin Periodontol* **25:** 134-144

Sommerhoff CP, Bode W, Matschiner G, Bergner A, Fritz H (2000) The human mast cell tryptase tetramer: a fascinating riddle solved by structure. *Biochim Biophys Acta* **1477:** 75-89

Springman EB, Angleton EL, Birkedal-Hansen H, Van Wart HE (1990) Multiple modes of activation of latent human fibroblast collagenase: evidence for the role of a Cys73 active-site zinc complex in latency and a "cysteine switch" mechanism for activation. *Proc Natl Acad Sci U S A* **87:** 364-368

Srinivasan S, Romagnoli M, Bohm A, Sonenshein GE (2014) N-glycosylation regulates ADAM8 processing and activation. *J Biol Chem* **289:** 33676-33688

Stöcker W, Bode W (1995) Structural features of a superfamily of zinc-endopeptidases: the metzincins. *Curr Opin Struct Biol* **5:** 383-390

Stöcker W, Gomis-Rüth FX (2013) Astacins: proteases in development and tissue differentiation. In *Proteases: Structure and Function*, Brix K, Stöcker W (eds), pp 235-263. Springer Vienna

Stöcker W, Grams F, Baumann U, Reinemer P, Gomis-Rüth FX, McKay DB, Bode W (1995) The metzincins--topological and sequential relations between the astacins, adamalysins, serralysins, and matrixins (collagenases) define a superfamily of zinc-peptidases. *Protein Sci* **4:** 823-840

Swaminathan S (2011) Molecular structures and functional relationships in clostridial neurotoxins. *FEBS J* **278:** 4467-4485

T

Tajima N, Kawai F, Park SY, Tame JR (2010) A novel intein-like autoproteolytic mechanism in autotransporter proteins. *J Mol Biol* **402:** 645-656

Tallant C, García-Castellanos R, Baumann U, Gomis-Rüth FX (2010a) On the relevance of the Met-turn methionine in metzincins. *J Biol Chem* **285:** 13951-13957

Tallant C, García-Castellanos R, Seco J, Baumann U, Gomis-Rüth FX (2006) Molecular analysis of ulilysin, the structural prototype of a new family of metzincin metalloproteases. *J Biol Chem* **281:** 17920-17928

Tallant C, Marrero A, Gomis-Rüth FX (2010b) Matrix metalloproteinases: fold and function of their catalytic domains. *Biochim Biophys Acta* **1803:** 20-28

Tam A, Schmidt WK, Michaelis S (2001) The multispanning membrane protein Ste24p catalyzes CAAX proteolysis and NH2-terminal processing of the yeast a-factor precursor. *J Biol Chem* **276:** 46798-46806

Tholander F, Muroya A, Roques BP, Fournie-Zaluski MC, Thunnissen MM, Haeggstrom JZ (2008) Structure-based dissection of the active site chemistry of leukotriene A4 hydrolase: implications for M1 aminopeptidases and inhibitor design. *Chem Biol* **15:** 920-929

Toma S, Campagnoli S, De Gregoriis E, Gianna R, Margarit I, Zamai M, Grandi G (1989) Effect of Glu-143 and His-231 substitutions on the catalytic activity and secretion of Bacillus subtilis neutral protease. *Protein Eng* **2:** 359-364

Toth M, Bernardo MM, Gervasi DC, Soloway PD, Wang Z, Bigg HF, Overall CM, DeClerck YA, Tschesche H, Cher ML, Brown S, Mobashery S, Fridman R (2000) Tissue inhibitor of metalloproteinase (TIMP)-2 acts synergistically with synthetic matrix metalloproteinase (MMP) inhibitors but not with TIMP-4 to enhance the (Membrane type 1)-MMP-dependent activation of pro-MMP-2. *J Biol Chem* **275:** 41415-41423

Trame CB, Chang Y, Axelrod HL, Eberhardt RY, Coggill P, Punta M, Rawlings ND (2014) New mini- zincin structures provide a minimal scaffold for members of this metallopeptidase superfamily. *BMC Bioinformatics* **15:** 1

Trillo-Muyo S, Martínez-Rodríguez S, Arolas JL, Gomis-Rüth FX (2013) Mechanism of action of a Janus-faced single-domain protein inhibitor simultaneously targeting two peptidase classes. *Chem Sci* **4:** 791-797

Turk B (2006) Targeting proteases: successes, failures and future prospects. *Nat Rev Drug Discov* **5:** 785-799

V

Van den Burg B, Eijsink V (2013) Chapter 111 - Thermolysin and related bacillus metallopeptidases. In *Handbook of Proteolytic Enzymes,* Salvesen NDR (ed), pp 540-553. Academic Press

Van Wart HE, Birkedal-Hansen H (1990) The cysteine switch: a principle of regulation of metalloproteinase activity with potential applicability to the entire matrix metalloproteinase gene family. *Proc Natl Acad Sci U S A* **87:** 5578-5582

Veillard F, Sztukowska M, Mizgalska D, Ksiazek M, Houston J, Potempa B, Enghild JJ, Thogersen IB, Gomis-Rüth FX, Nguyen KA, Potempa J (2013) Inhibition of gingipains by their profragments as the mechanism protecting Porphyromonas gingivalis against premature activation of secreted proteases. *Biochim Biophys Acta* **1830:** 4218-4228

Verma RP (2012) Hydroxamic acids as matrix metalloproteinase inhibitors. *EXS* **103:** 137-176

Visse R, Nagase H (2003) Matrix metalloproteinases and tissue inhibitors of metalloproteinases: structure, function, and biochemistry. *Circ Res* **92:** 827-839

W

Waltersperger S, Widmer C, Wang M, Baumann U (2010) Crystal structure of archaemetzincin AmzA from Methanopyrus kandleri at 1.5 A resolution. *Proteins* **78:** 2720-2723

Wang Y, Ha Y (2007) Open-cap conformation of intramembrane protease GlpG. *Proc Natl Acad Sci U S A* **104:** 2098-2102

Whisstock JC, Bottomley SP (2006) Molecular gymnastics: serpin structure, folding and misfolding. *Curr Opin Struct Biol* **16:** 761-768

Wiederanders B, Kaulmann G, Schilling K (2003) Functions of propeptide parts in cysteine proteases. *Curr Protein Pept Sci* **4:** 309-326

Wilke MS, Hills TL, Zhang HZ, Chambers HF, Strynadka NC (2004) Crystal structures of the Apo and penicillin-acylated forms of the BlaR1 beta-lactam sensor of Staphylococcus aureus. *J Biol Chem* **279:** 47278-47287

Woessner JF, Nagase H (2000) *Matrix Metalloproteinases and TIMPs*, New Edition edn.: Oxford University Press.

Wu JW, Chen XL (2011) Extracellular metalloproteases from bacteria. *Appl Microbiol Biotechnol* **92:** 253-262

X, Y, Z

Xu Q, Gohler AK, Kosfeld A, Carlton D, Chiu HJ, Klock HE, Knuth MW, Miller MD, Elsliger MA, Deacon AM, Godzik A, Lesley SA, Jahreis K, Wilson IA (2012) The structure of Mlc titration factor A (MtfA/YeeI) reveals a prototypical zinc metallopeptidase related to anthrax lethal factor. *J Bacteriol* **194:** 2987-2999

Ye S, Goldsmith EJ (2001) Serpins and other covalent protease inhibitors. *Curr Opin Struct Biol* **11:** 740-745

Yiallouros I, Kappelhoff R, Schilling O, Wegmann F, Helms MW, Auge A, Brachtendorf G, Berkhoff EG, Beermann B, Hinz HJ, Konig S, Peter-Katalinic J, Stöcker W (2002) Activation mechanism of pro-astacin: role of the pro-peptide, tryptic and autoproteolytic cleavage and importance of precise amino-terminal processing. *J Mol Biol* **324:** 237-246

Yu AC, Worrall LJ, Strynadka NC (2012) Structural insight into the bacterial mucinase StcE essential to adhesion and immune evasion during enterohemorrhagic E. coli infection. *Structure* **20:** 707-717

Zakharova E, Horvath MP, Goldenberg DP (2009) Structure of a serine protease poised to resynthesize a peptide bond. *Proc Natl Acad Sci U S A* **106:** 11034-11039

Zelenski NG, Rawson RB, Brown MS, Goldstein JL (1999) Membrane topology of S2P, a protein required for intramembranous cleavage of sterol regulatory element-binding proteins. *J Biol Chem* **274:** 21973-21980

Zhang HZ, Hackbarth CJ, Chansky KM, Chambers HF (2001) A proteolytic transmembrane signaling pathway and resistance to beta-lactams in staphylococci. *Science* **291:** 1962-1965

Zhu XL, Ohta Y, Jordan F, Inouye M (1989) Pro-sequence of subtilisin can guide the refolding of denatured subtilisin in an intermolecular process. *Nature* **339:** 483-484