ULB FACULTÉ DES SCIENCES

UNIVERSITÉ LIBRE DE BRUXELLES

The Paradigm of Self-compartmentalized M42 Aminopeptidases: Insight into Their Oligomerization, Substrate Specificities, and Physiological Function

Thesis submitted by Raphaël DUTOIT
in fulfilment of the requirements of the PhD Degree in Biochemistry and Molecular and Cell Biology ("Docteur en Biochimie et Biologie Moléculaire et Cellulaire")
Academic year 2020-2021

Supervisor: Professor Louis DROOGMANS

Co-supervisor: Dr. Martine ROOVERS

Thesis jury:

Luc VANHAMME (Université libre de Bruxelles, Chair)
Anna Maria MARINI (Université libre de Bruxelles, Secretary)
Laurence VAN MELDEREN (Université libre de Bruxelles)
Abel GARCIA PINO (Université libre de Bruxelles)
Bruno FRANZETTI (Institut de biologie structurale, Grenoble, France)
Georges FELLER (Université de Liège)

Summary

M42 aminopeptidases are dinuclear enzymes widely found in prokaryotes but completely absent from eukaryotes. They have been proposed to hydrolyze peptides downstream the proteasome or other related proteolytic complexes. Their description relies mainly on the pioneering work on four M42 aminopeptidases from *Pyrococcus horikoshii*. Their quaternary structure consists of twelve subunits adopting a tetrahedral-shaped structure. Such a spatial organization allows the compartmentalization of the active sites which are only accessible to unfolded peptides. The dodecamer assembly results from the self-association of dimers under the control of the metal ion cofactors. Both oligomers have been shown to co-exist *in vivo* and heterododecamers with broadened substrate specificity may even occur. Yet, the molecular determinants behind the dodecamer assembly remain unknown due the lack of a high-resolution structure of a stable dimer. In addition, the bacterial M42 aminopeptidases are still ill-described due to the paucity of structural studies.

This work focuses mainly on the characterization of TmPep1050, an M42 aminopeptidase from *Thermotoga maritima*. As expected, TmPep1050 adopts the genuine tetrahedral-shaped structure with twelve subunits. It also displays a leucyl-aminopeptidase activity requiring Co^{2+} as a cofactor. In addition to its catalytic function, Co^{2+} has a role in the enzyme thermostability and oligomerization. The absence of Co^{2+} provokes the disassembly of active TmPep1050 dodecamers into inactive dimers. The process, however, is reversible since Co^{2+} triggers the self-association of dimers into dodecamers, as shown by native MS. The main achievement of this work is the determination of the first high-resolution structure of a dimer, allowing to better understand the dimer-dodecamer transition. Several structural motifs involved in oligomerization are displaced or highly flexible in the TmPep1050 dimer structure. Furthermore, a loop bringing two catalytic relevant residues is displaced outside the catalytic site. These residues are the catalytic base and a ligand involved in the Co^{2+} binding at the M1 site. The metal ion binding sites have been further investigated to define how they influence the oligomerization of TmPep1050. A mutational study shows that the M1 site strictly controls the dodecamer formation while the M2 site contributes only partly to it. A strictly conserved aspartate residue of the M2 site second shell also plays an important structural role in maintaining the active site integrity. Indeed, its substitution prevents the formation of dodecamer probably due to the lack of stabilization of the active site loop.

The characterization of TmPep1050 supports that bacterial M42 aminopeptidases probably share the quaternary structures and dodecamer assembly with their archaeal counterparts. The dimer structure highlights several structural modifications occurring in the dimer-dodecamer transition. Yet, based on current knowledge, no general rules can be drawn for the role of the M1 and M2 sites in oligomerization. Besides, the physiological function of the M42 aminopeptidases is under-examined albeit the proposed link to the proteasome. In this work, this has been investigated using the *Escherichia coli* M42 aminopeptidases as a model. Yet, no phenotype has been associated to the deletion of their coding genes. Preliminary results have shown that the three enzymes (i) display a redundant substrate specificity, (ii) could be localized partly to the membrane, and (iii) form heterocomplexes. Further experiments are still required to crack the function of these M42 aminopeptidases.

Acknowledgment

None of this work would have been possible without the help of many people. I would like to express my gratitude to them, whichever the reason is. While writing these few lines, I'm remembering various events in the last five years: hours spent in the lab, side activities and funny stories, failure and success, moments of felicity and sadness, and how much connectedness influences our lives. I will probably forget a few names in the following list. To you, guys, I beg your pardon.

First, I would like to warmly thank Louis Droogmans, my thesis supervisor. Louis, I really appreciate your gentle style of guidance and every scientific discussion that we had together. I'm still awed by your never-ending knowledge about so many topics. Besides your scientific qualities, you are a nice person, respected and respectful, with your so peculiar humor. Speaking of which, as a promoter, Louis can always rely on his operator, Martine Roovers. Martine, I'm grateful to have you as my co-supervisor. You've always been there to support me whatever the circumstance. I like your personality, a blend of Dutch manner, limitless curiosity, and unstoppable mirthful laughter.

I also thank Alain Durieux, my boss at Labiris, to have allowed me to follow a PhD at the Université Libre de Bruxelles. I'm beholden to the Commission Communautaire française for funding our research. I would like to thank Laurence Van Melderen and Abel Garcia Pino, my thesis committee. I've enjoyed our yearly discussion and appreciated your external point of view. Thanks again to Abel for helping me to answer to some reviewers' comments for the JBC article.

I express my gratitude to Frank Sobott, from UAntwerp and the University of Leeds, for his interest in our work and the productive collaboration that we had together (I was awestruck by our brainstorming). You are genuinely brilliant and a gentle person. At UAntwerp, I would like to thank Tom and Jeroen for their remarkable work for analyzing TmPep1050 samples by native MS. Thanks again for showing me the facility.

I want also to be thankful to my colleagues. From Louis' lab, I would like to thank André "Papy" and Dany. I know, Papy, I kept you waiting too long but finally you can read the most important page of this thesis. I appreciate how you are balanced, between your scientific rigor and your human qualities. Then, there is Dany, the master of ludicrous "zwanze", always having a good joke in his pockets and teasing people with his provocative thoughts to make them fly off the handle.

At Labiris, I'm grateful to my colleagues. It would take pages to thank them so I've to keep it simple. I feel so tempted to tell stories or praise their unique trait like André our crossword champion, Isabelle's fighting spirit, Geoffray's quiet strength, the passionate Corinne, the flamboyant Nicole, etc. I must particularly thank Nathalie for everything she has done in this project. To all of you, thank you: André, Mélanie, Isabelle, Fabienne, Melody, Geoffray, Amaury, Nathalie A., Martine L., Luc, Corinne, Bertrand, Catherine, Yamina, Nathalie H., Sandra, Camille, Nicole, Martine D., Marc, Françoise, and Evelyne.

There are two colleagues that I would like to especially thank because of their influence on life trajectory. First, Laetitia, I know that there were quarrels and grudges between us in the past. I'm happy that we put an end to our enmity and to see how our relationship has evolved, a real lesson of life. Despite being a genuine princess, I like your honesty, humor, and friendship. Then, Eric, I would

like to thank you as you were my first mentor. You've given me so much knowledge and helped me to shape my scientific mind.

During the past five years, I've been involved in supervising several students in their internship. Directly linked to this work, I would like to thank Manon and Maëva for their participation. I also express my gratitude to Emna who did a part of her PhD internship in our lab. You're a kind person and I really enjoy having you as a friend.

Speaking of friendship, I'm grateful to my friends and family for their support outside the lab. In the inner circle, I would like to warmly thank Melody, you make life more colorful and much more intense. My thanks to the "rikiki" band, the aficionados of card playing, Anne, JM, Claude, Yolande, Cécile, and Jérôme. All my thanks to the Spanish classes, Sandra, Martine, Bernard, Jerry, Ginette, Isabelle, Vincent, and many other amazing people. We spent so many hours weekly practicing and talking about so many topics, a real breath of fresh air. Talking about Spanish lessons, I would like to thank the Colombian connection, Gloria and Carolina, I've learnt so much with you. Big thanks to my travel buddies, Martine, Jess, Matt & Co. we've spent so many good times together. I also thank the last kernel of my fellow buddies from the master internship, Thalie, Vincent, Régis, Aurélie, and Isabelle. In a completely different register, I'm grateful to Danielle who taught me the path to mindfulness. Finally, I would like to thank my family, especially my parents. I won't be who I am without you.

Again, I thank you all.

Table of contents

Abbreviations

AAA+: ATPase associated with diverse cellular activities
Abs: absorbance
ACE: angiotensin converting enzyme
Anbu: ancestral β unit protein
APH: acylpeptide hydrolases
ASU: asymmetric unit
BPH: betaproteobacteria proteasome homolog
CMC: carboxymethylcellulose
CODAS: cerebral, ocular, dental, articular, and skeletal anomalies
Cryo-EM: cryogenic electron microscopy
EDTA: ethylenediaminetetraacetic acid
ESRF: European Synchrotron Radiation Facility
FRT: Flp recombinase recognition site
GsApI: *Geobacillus stearothermophilus* aminopeptidase I
HPLC: high performance liquid chromatography
IPTG: isopropyl β-D-thiogalactoside
L-Leu-*p*NA: L-Leucine-*p*-nitroanilide
MAGUK: membrane associated guanylate kinase
M_r: relative mass
MS: mass spectrometry
mUA: milli unit of absorbance
NMR: nuclear magnetic resonance
OB: oligonucleotide/oligosaccharide binding
OD: optical density
PAN: proteasome-activating nucleotidase
PDB: Protein Data Bank
PDZ: PSD95, DLG, ZO-1
Pf: *Pyrococcus furiosus*
Ph: *Pyrococcus horikoshii*
*p*NA: *p*-nitroaniline
POP: prolyl oligopeptidases
PTS: phosphotransferase system
PUP: prokaryotic ubiquitin-like protein
RBS: ribosome binding site
r.m.s.d.: root mean square deviation
SAMP: small archaeal modifier protein
SANS: small angle neutron scattering
SAXS: small angle X-ray scattering
TaTRI: *Thermoplasma acidophilum* Tricorn peptidase
TET: tetrahedral shape
TLC: thin layer chromatography
T_m: temperature of melting
TRI: Tricorn peptidase
TSA: thermal shift assay
V_e: volume of elution
VpAp1: *Vibrio proteolyticus* Aminopeptidase 1
Xaa: any amino acid

Introduction

A.1 The roles of proteolytic enzymes in Life and diseases

Proteolysis plays an important role in a wide array of cellular functions as diverse as homeostasis, cell cycle, tissue development and morphogenesis, stress response, antigen presentation, transcription regulation, stalled ribosome rescue, immune response, virulence, cell differentiation, autophagy, apoptosis, etc.[1–8]. To achieve such functions, organisms produce an arsenal of proteolytic enzymes to ensure specific and nonspecific hydrolysis of proteins and peptides[9]. For instance, 641 genes coding proteolytic enzymes are found in the genome of *Homo sapiens*[3]. The proteolytic events are tightly regulated in cells via transcription regulation, posttranslational modifications, inhibitor synthesis, and compartmentalization[3]. In mammalian cells, about 80% of proteolysis is achieved by the proteasome, mainly for homeostasis and regulation purposes[1,10]. The proteasome also plays an important role in immunity since the immunoproteasome is responsible for presenting antigens on cell surface[6,11]. Abnormal and deregulated activity of proteolytic enzymes often lead to dramatic consequences due to their cellular functions.

Several human diseases are related to either a deficient or increased proteolytic activity like neurodegenerative diseases, tumoral cell development, inflammatory bowel disease, insulin resistance, CODAS (cerebral, ocular, dental, articular, and skeletal anomalies) syndrome, cardiovascular and lysosomal storage diseases[12–19]. In Alzheimer's, Parkinson's, and Huntington's diseases, the amyloid β, α synuclein, and huntingtin have been reported to form protein aggregates, progressively leading to neuron death[15,20]. Protein aggregation results from decreased proteasome function, probably due to an impaired ubiquitin tagging system and an increased immunoproteasome content in neurons[21,22]. The inhibition of the proteasome function is further hastened by protein aggregates since soluble aggregates are known to bind to the proteasome 20S core particle, preventing any interactions with the 19S subunit or other auxiliary partners[15]. A deregulation of proteasome function also contributes to tumoral cell development. Indeed, an increased proteasome activity alleviates the inhibition of NF-κB pathway in tumoral cells, promoting cell growth and proliferation[12,19]. Matrix metallopeptidases are also involved in a series of diseases[14,16,18]. For instance, an abnormal degradation of the extracellular matrix is observed in inflammatory bowel disease and atherosclerosis[16,18]. The lack of cathepsin K, a lysosomal matrix metallopeptidase, reduces the degradation of bone matrix in osteoclasts, causing lysosomal storage disease[14]. CODAS syndrome is another disease resulting from a defective proteolytic enzyme[17]. In that case, a substitution of a residue in the central pore of LonP1 provokes a reduced protein homeostasis in mitochondria.

Due to their role in many cellular functions, proteolytic enzymes are undoubtedly therapeutic targets[1,19,23–28]. Currently, about 5% of available drugs targets proteases, the most well-known being gliptins, angiotensin converting enzyme (ACE) inhibitors, bortezomib, and antithrombotic agents. Gliptins are used to treat type 2 diabetes as they target DPP4 peptidase[13]. The inhibition of DPP4 increases the amount of incretin hormones in blood, improving the insulin secretion and satiety awareness. ACE inhibitors are definitively a success story in drug development with an annual sale exceeding eleven billion of US dollars in 2017. They are widely used to regulate blood pressure and treat hypertension[23,24,29]. Such drugs inhibit ACE by coordinating its catalytic Zn^{2+} cofactor, preventing the conversion of angiotensin I to angiotensin II and, subsequently, lowering blood pressure[30]. Some

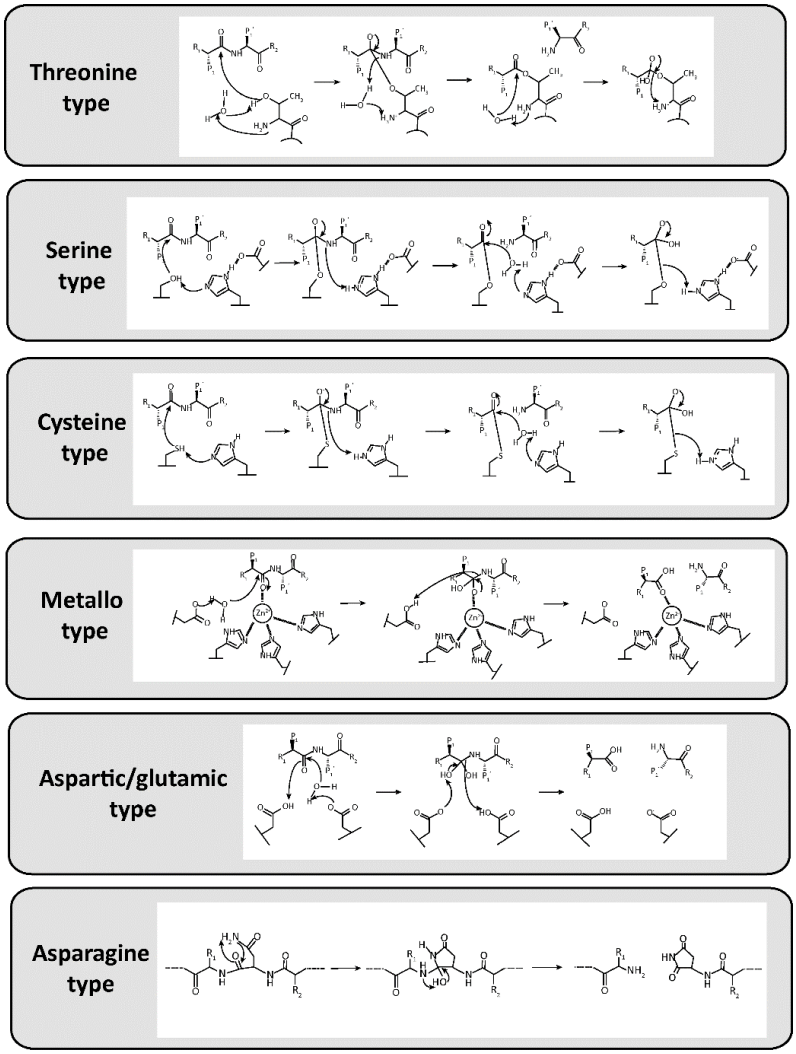


Figure 1 – The catalytic types of proteolytic enzymes. Each type is named according to the catalytic residue or metal ions. The catalytic mechanism is schematized for each type. The metallo-type is exemplified with a mononuclear active site with three histidine residues coordinating the metal ion. Dinuclear active site also exists and different ligand patterns can coordinate the metal ions. The catalytic mechanism of the asparagine type is simplified to the peptide bond breakage leading to the formation of a succinimide. The catalytic asparagine residue is activated depending on its surrounding. Inspired and adapted from Deu *et al.* (2012)[28].

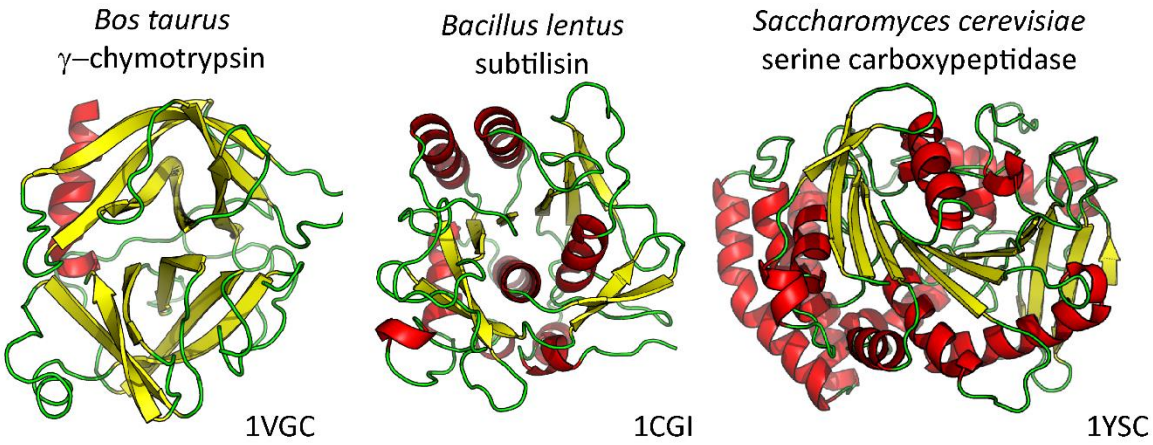


Figure 2 – The structural diversity of serine proteases. As an example, the structures of three serine proteases are schematized in cartoon representation. While the γ-chymotrypsin has a β-barrel fold, the subtilisin and the serine carboxypeptidase adopt a Rossmann fold but are different by their size and substructure organization[63–65].

cardiovascular diseases can also be treated with proteolytic enzyme inhibitors[18,23,24]. For instance, antithrombic agents, which inhibit the factor IIa and factor Xa, are commonly used as anticoagulants, preventing the conversion of fibrinogen to fibrin in blood clotting[31]. Proteasome inhibitors are also developed to act as antitumoral drugs[12,19,23,26]. Notably, bortezomib is a potent proteasome inhibitor commonly used to treat multiple myeloma, triggering apoptosis of tumoral cells[12].

Proteolytic enzymes are also involved in pathogenesis resulting from viral and microbial infections[4,16,32–36]. In bacteria, proteasome and other associated complexes contribute directly to virulence by a tightly and timely degradation of regulatory proteins[4,37]. In addition to be linked to virulence, they are involved in antibiotic resistance and stress response[32,37]. For instance, ClpP and Lon regulate the type III secretion system allowing to inject proteins inside eukaryotic host cells[34,38]. The injected proteins may fulfil various roles like preventing phagocytosis, killing macrophages, and modulating the host immune response[38]. DegP, a periplasmic protease, is involved in oxidative stress response, improving bacteria survival in phagocytes[37,39]. Many other proteolytic enzymes have been described to be important for the virulence of pathogenic bacteria, either in biofilm formation or to degrade host matrix barrier[37,40–45]. Some bacterial pathogens can even hijack host proteases, inducing uncontrolled degradation of extracellular matrix[46]. Proteolytic enzymes could be promising drug targets to develop new strategies to control pathogen infection[25,32,34,39,47]. Consequently, studying such enzymes is important to better understand their cellular functions. The proteasome and other related proteolytic complexes of bacteria have been extensively studied[48–53] although developing specific drugs targeting them remains challenging[33]. Auxiliary peptidases should also be characterized since they define the peptidome (i.e. the complete set of peptides). Peptides are known to be a key player in quorum sensing in bacteria[54]. Furthermore the peptidome could be a molecular mimicry to cheat host immune system[55] or induce an autoimmune response[56].

A.2 Lexicon about proteolytic enzymes

Before introducing the proteolytic complexes (see **section A3**) and their auxiliary peptidases (see **section A.4**), some prerequisites are required to better apprehend the next sections. A classification system has been introduced for proteolytic enzymes, the MEROPS classification (see **section A.2.1**). The different catalytic types will be presented hereafter, as well as the important notion of families and clans of proteolytic enzymes. Another convention has been introduced to describe the catalytic cleft of proteolytic enzymes, which will be used throughout the next sections (see **section A.2.2**).

A.2.1 MEROPS classification

In 1960, Hartley proposed a classification of proteolytic enzymes based on their catalytic type [57]. The catalytic type refers to the chemical mechanism of peptide bond hydrolysis, involving either a catalytic residue or metal ions (see **Figure 1**). Assigning a peptidase to one of the catalytic types is quite simple using generic inhibitors, like a strong chelating agent inhibiting all metallopeptidases. Such a classification, however, is overly simplistic since it does not consider the vast diversity of sequences, structures, other catalytic mechanisms, as well as evolutionary aspects. For instance, chymotrypsin, subtilisin, and serine carboxypeptidase are an example of convergent evolution with respect to a same catalytic type[58]. The three enzymes have distinguishable structural folds but share the common Ser-His-Asp catalytic triad of serine proteases (see **Figure 2**).

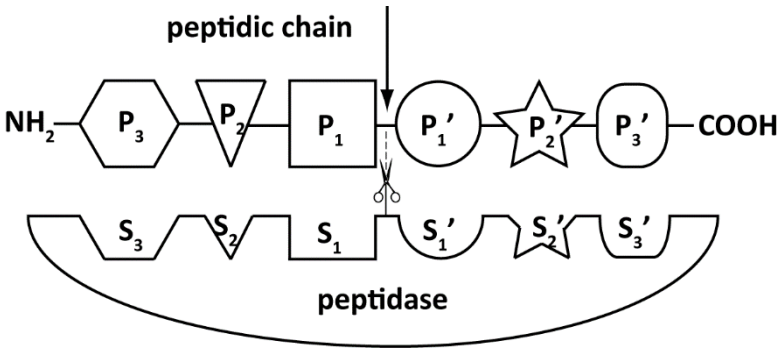


Figure 3 – Schematic representation of a peptidase active site to illustrate the nomenclature of Schechter and Berger[66]. The fictional peptidase has six binding sites, recognizing six amino acids of a peptide chain. Each amino acid is represented by a distinctive shape. The cleavage site is indicated with an arrow, the catalytic residues with a scissor.

Rawlings and Barrett improved Hartley's classification by introducing the notion of evolutionary distinct families[59]. Each family is built around an archetypal enzyme, often biochemically well characterized. Based on amino acid sequence homology, each family member must be significantly related to its archetypal enzyme. A family can also be divided into subfamilies if an ancient divergence is observed among its members. Some sequence-divergent families may also share a common tertiary structure, resulting from common ancestry. Such structurally related families are regrouped in a clan. The family and clan nomenclature has been released as a manually curated database called MEROPS[60]. In 2018, the MEROPS database catalogs about a million of peptidase sequences divided into 268 families and 62 clans[61]. The family name consists of an initial letter referring to the catalytic type and a number assigned arbitrarily. The initial letters are S, T, C, A, G, N, M, P, and U corresponding to serine, threonine, cysteine, aspartic, glutamic, asparagine, metallo-, mixed, and unknown catalytic type, respectively. The amino acids refer to the catalytic residues involved in peptide bond hydrolysis (see **Figure 1**). Of note, the proteases of the asparagine type, often referred to as asparagine lyases, are peculiar self-splicing proteases, like the viral coat proteins and inteins[62]. The mixed type means that the catalytic residue nature may vary between, for instance, a threonine and a serine residue. Nine families are classified as unknown type as their catalytic residues remain to be identified. The clan name consists of a two-letter code, with the first letter referring to the catalytic type, like the family nomenclature, and the second letter attributed arbitrarily.

A.2.2 Nomenclature of peptidase active sites

The **sections A.4-A.5** will refer to the nomenclature used to describe the active sites of proteolytic enzymes. Basically, an active site consists of several binding sites, each site accommodating an amino acid of the peptidic chain. These binding sites are numbered according the nomenclature introduced by Schechter and Berger[66]. The numbering starts from the peptide bond to be cleaved. By convention, the hydrolysis of the peptide bound takes place between S_1 and S_1' binding sites (see **Figure 3**). The amino acids towards the N-terminus occupy the S_1, S_2, S_3, ... binding sites while those towards the C-terminus occupy the S_1', S_2', S_3', … binding sites (see **Figure 3**). Accordingly, the amino acids of the peptidic chain recognized by these binding sites are numbered: P_1, P_2, P_3, … for the S sites and P_1', P_2', P_3', … for the S' sites.

A.3 The proteolytic complexes in prokaryotes

Like in eukaryotes, protein homeostasis is fulfilled by the proteasome complex in archaea and actinomycetes (see **section A.3.1**)[49–51,53,67,68]. The proteasome 20S core particle, however, is absent in many bacterial phyla[69] for which other proteasome associated complexes achieve protein degradation. For instance, *Escherichia coli* possesses four AAA+ proteolytic complexes: HslUV[70], ClpXP/ClpAP[71], Lon[72], and FtsH[73]. The HslUV complex (see **section A.3.2**) is often regarded as the bacteria equivalent of the proteasome 20S[48]. HslV has probably originated from a simplification of the proteasome 20S[74]. Meanwhile, the latter could have evolved from an ancestral protease, Anbu, mainly found in cyanobacteria (see **section A.3.3**)[75]. ClpXP/ClpAP (see **section A.3.4**) and Lon (see **section A.3.5**) are two AAA+ proteolytic complexes having a redundant function to HslUV in *E. coli*[76–78]. The three complexes, however, have different quaternary structures and aim different regulatory

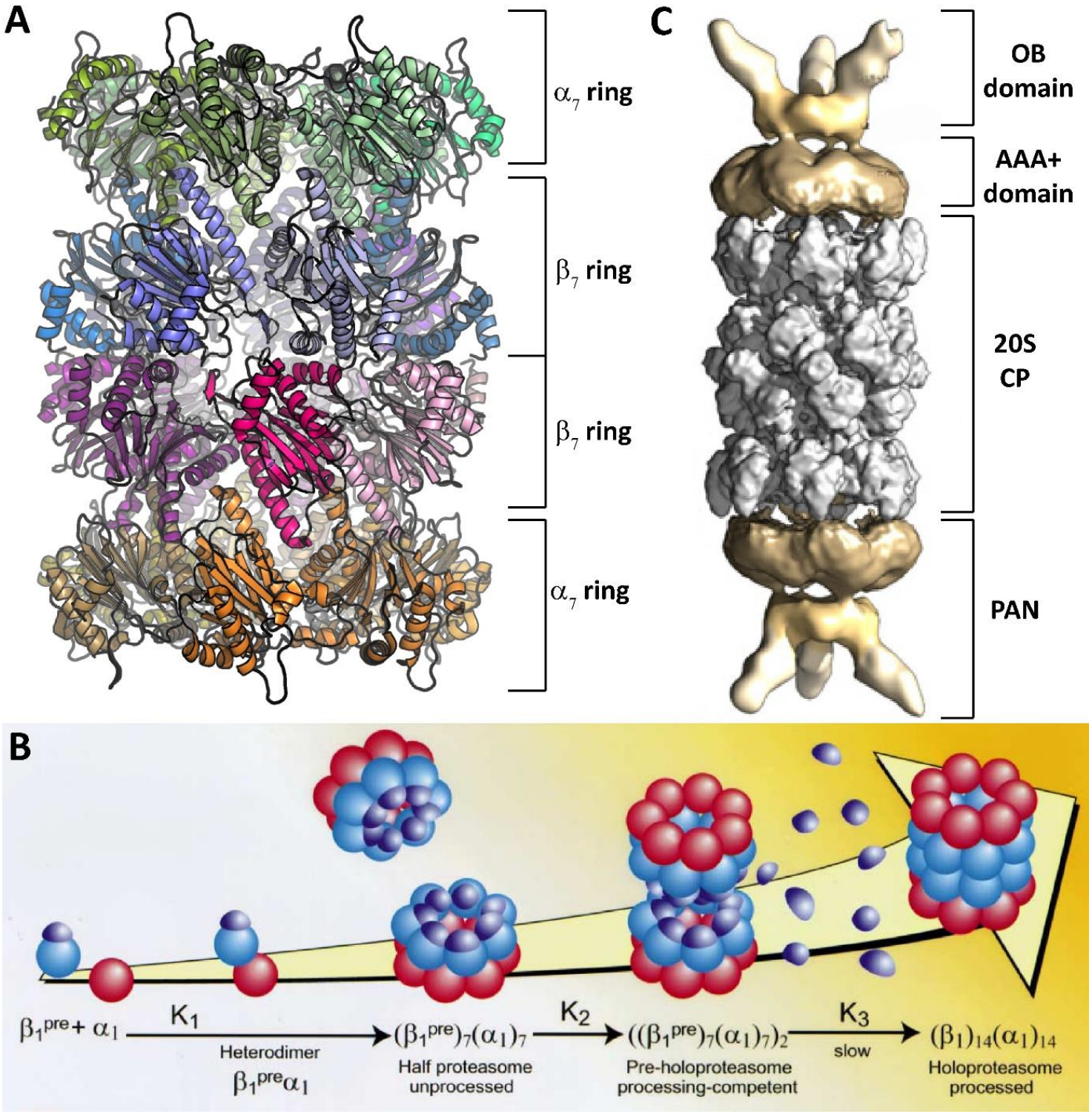


Figure 4 – The prokaryotic proteasome 20S core particle. (A) Quaternary structure of the proteasome 20S core particle made of four stacked heptameric rings. Two middle rings define the catalytic chamber with fourteen β subunits, two rings of α subunits sandwich the catalytic chamber. PDB code: 6BDF[89]. **(B)** The self-assembly of α and β subunits to build a functional proteasome core particle. The α subunits are represented as red spheres, the β subunits as blue spheres, and the β subunit prepeptides as purple drops. Spontaneously, α and β subunits form heterodimer that assemble into half proteasome particles. The formation of the proteasome 20S core particle requires the autoprocessing of prepeptides to allow the interaction between two β rings. Reproduced from Zühl *et al.* (1997)[93]. **(C)** Structure of the proteasome 20 core particle (20S CP) capped by two PANs. The AAA+ and OB domains of PAN are emphasized. Adapted from Majumder *et al.* (2019)[102].

proteins for degradation. FtsH is another AAA+ protease localized in the membrane achieving membrane protein turnover (see **section A.3.6**)[73,79,80]. Finally, HtrA is an exception for proteolytic complexes since it does not require ATP or ATPase partners to degrade periplasmic proteins (see **section A.3.7**)[81].

A.3.1 Proteasome 20S

Proteolysis has a central function in many cellular processes including protein turnover, posttranslational regulation, stress response, cell-cycle regulation, virulence, etc.[4,51–53,67,82]. In these processes, the proteasome 20S core particle is a key actor playing a main role. It is ubiquitously found in all eukaryotes and archaea while being absent in many bacterial phyla, excepting actinomycetes[69]. Although their sequences being divergent, the proteasome 20S of archaea and actinomycetes shares many similarities with that of eukaryotes. They adopt a barrel-shaped structure consisting of four stacked heptameric rings: the two outer rings with fourteen α subunits and two inner rings with fourteen β subunits (see **Figure 4.A**)[67,83–89]. The β subunits are catalytically active while the α subunits control the access to the inner catalytic chamber[50,51,67]. Regarding the diversity of α and β subunits, the proteasome 20S core particle of prokaryotes, however, differs from that of eukaryotes. Archaea and many actinomycetes have one single type of α and β subunits while eukaryotes have fourteen genes coding seven different α subunits and seven different β subunits[51]. In addition, only three of the seven β subunits are catalytically active in the proteasome 20S core particle of eukaryotes[90]. Regarding its catalytic type, the β subunit is a threonine protease of the T1A family[61,91], the catalytic threonine residue being located at the N-terminus of the β subunit mature form. The catalytic mechanism relies on a catalytic triad Thr-Lys-Asp: the triad lysine residue deprotonates the catalytic threonine residue while the aspartate residue modulates the amine charge of the lysine residue[91].

The quaternary structure of the proteasome 20S results from a controlled self-assembly of the α and β subunits[51,85,90,92–94]. The α subunits spontaneously form heptameric rings while the association of β subunits requires the presence of α subunits[93,95]. The assembly process involves the formation of heterodimers prior to heptamerization (see **Figure 4.B**)[93,94,96]. The prepeptide of the β subunit could promote the dimerization and shield the intermediate half proteasome particle[85]. The latter consists of a heptameric α ring stacked over a heptameric β ring. Two half proteasome particles further assemble into the functional barrel through the processing of prepeptides, freeing the N-terminal catalytic threonine residue[94]. In the idle state, the seven N-terminal octapeptides of an α ring completely close the access to the inner catalytic chamber, preventing the entrance of any peptides[94]. The gate opens when auxiliary proteins, also known as proteasome-activating nucleotidases (PANs), bind to the proteasome 20S core particle[97,98]. The C-terminus of PANs has been shown to be sufficient to unlock the door to the catalytic chamber[98].

PAN has the canonical domain of ATPases associated with diverse activities (AAA+) and are known to recruit and unfold proteins tagged for degradation[52,99–102]. In addition to the AAA+ domain, PAN also possesses an oligonucleotide/oligosaccharide binding (OB) domain. The OB domain exhibits a chaperone activity and recognizes degrons[103]. The quaternary structure of PAN is a hexamer ring adopting a spiral staircase motif[102]. The AAA+ domains are oriented towards the proteasome 20S while the OB domains crown the PAN-proteasome 20S complex (see **Figure 4.C**)[102–104]. Like the ubiquitin tagging system in eukaryotes, proteins are marked by small proteins known as small

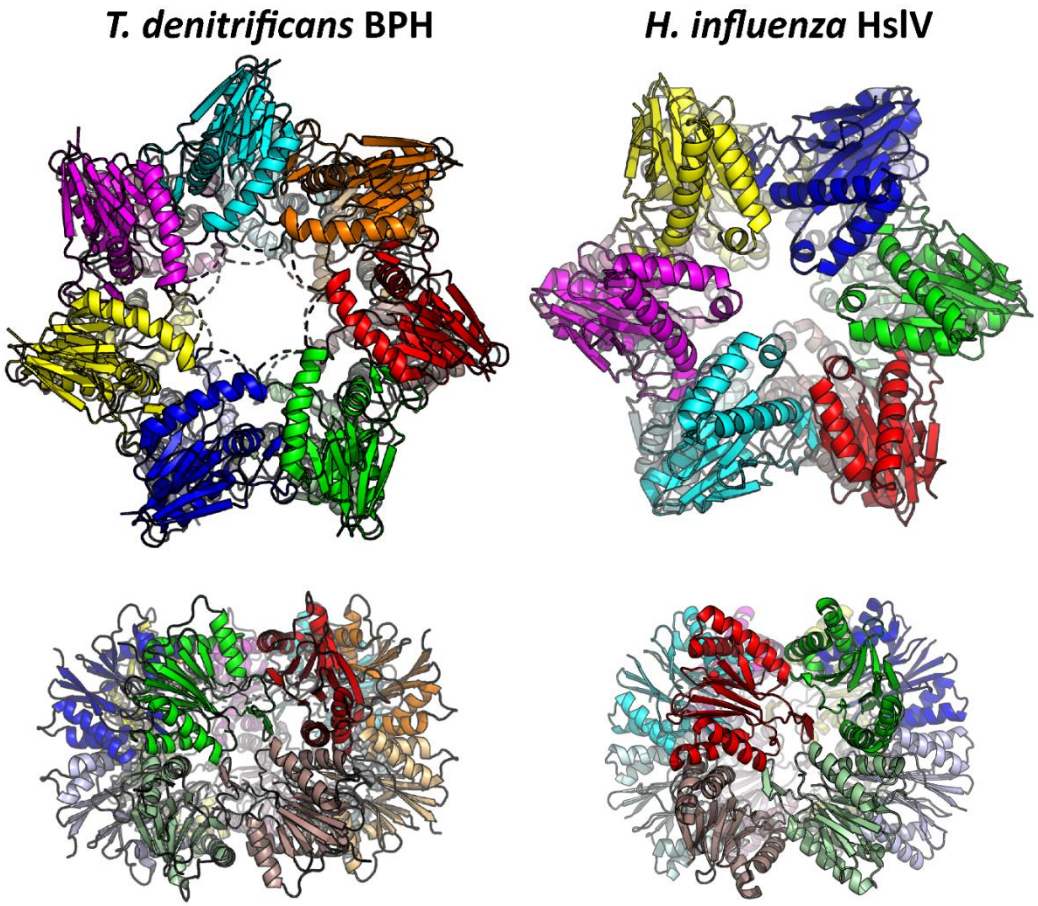


Figure 6 – Structures of BHP from *T. denitrificans* and its structural homolog HslV from *H. influenza*. Above and below: top view and side view, respectively, of BHP and HslV. PDB codes: 1G3K (HslV)[117] and 5OVS (BHP)[122].

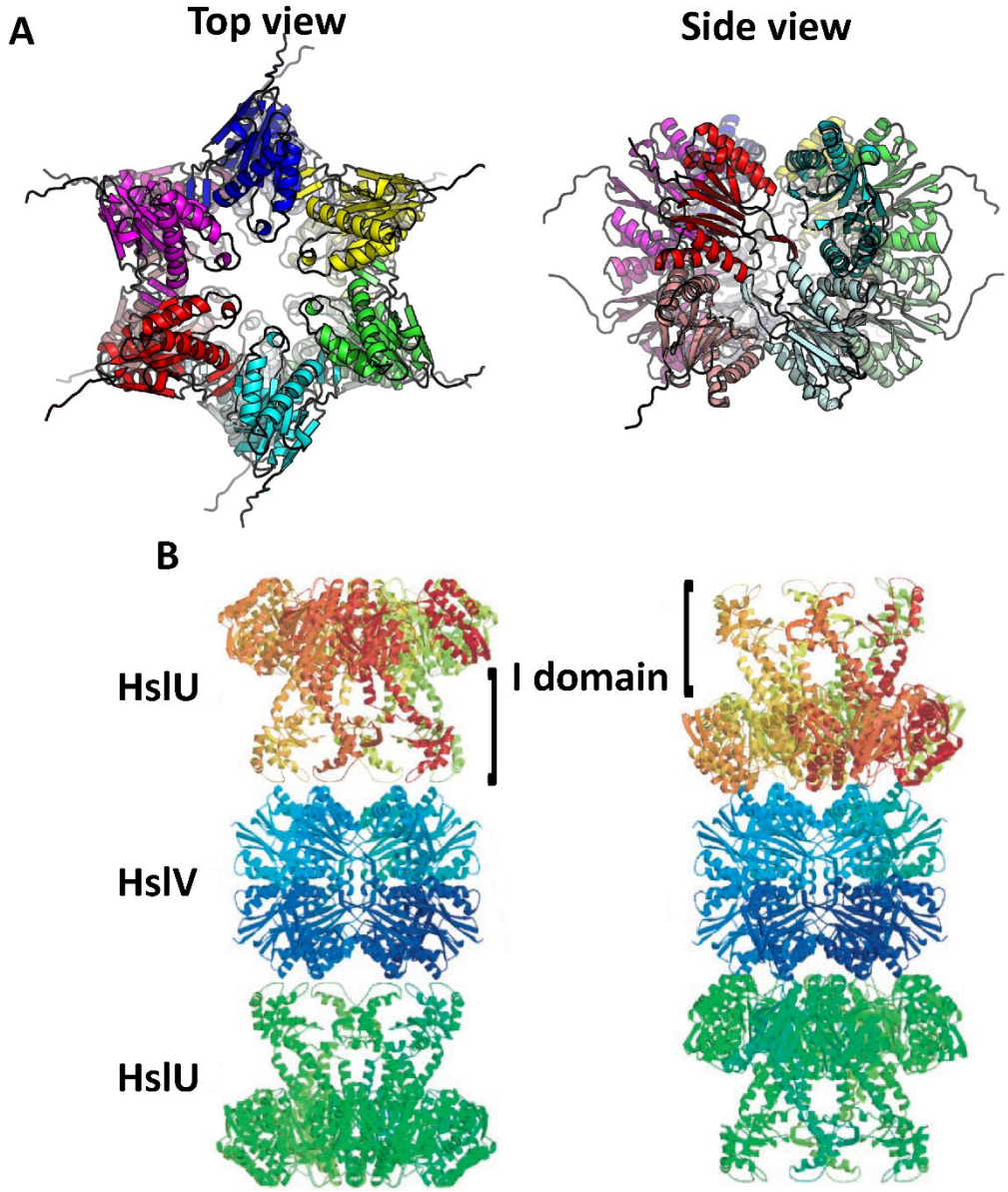


Figure 5 – Structure of the HslUV complex. (A) Structure of HslV dodecamer forming two stacked rings seen from the six-fold (left) and a two-fold (right) symmetry axis. PDB code: 1NED[116]. **(B)** Structure of the HslUV complex with the two binding modes of HslU and HslV represented. The I domain is emphasized with a bracket. Adapted from Song et al. (2000)[118].

archaeal modifier proteins (SAMP) and prokaryotic ubiquitin-like proteins (PUP) in archaea and actinomycetes, respectively[68,105]. SAMP and PUP are small unstructured proteins which are covalently bound to lysine residues of proteins marked for proteolysis[99,106]. In both cases, polysampylation and polypupylation may occur[106,107] but, unlike polyubiquitylation, they could prevent marked proteins to be degraded by the proteasome[107]. In association with PAN, the proteasome 20S carries out protein turnover in an ATP-dependent manner. Although, another interacting factor, PafE, has been described in *Mycobacterium tuberculosis* which favors an ATP-independent proteolysis[108,109]. Its mechanism of action is still ill-described but PafE has been shown to form a dodecameric ring capping both ends of the proteasome 20S core particle.

A.3.2 HslUV

While being divergent with less than 20% of homology, the HslUV complex strikingly resembles to the proteasome 20S. Indeed its catalytic subunit, HslV, is structurally related to the β subunit of the proteasome 20S[110,111]. Their catalytic sites are also conserved with a N-terminal catalytic threonine residue. Furthermore, HslV could be an evolutionary simplification of the proteasome 20S (see **section A.3.3**)[112]. To be active, HslV absolutely requires its AAA+ binding partner, HslU[113,114], which is closely related to ClpA and ClpX (see **section A.3.4**). In *E. coli*, HslUV complements the deletion of *lon* and *clp*P, indicating that the three complexes are redundant[76–78]. Regarding its quaternary structure, HslV is a dodecamer forming a barrel-shaped structure with two stacked hexameric rings (see **Figure 5.A**)[111,115]. In the HslUV complex, both ends of the catalytic HslV barrel are capped by an hexameric ring of HslU (see **Figure 5.B**)[110,116,117]. Two modes of docking between HslV and HslU have been described in distinct structures, with the I domain of HslU pointing either towards or outwards the HslV barrel (see **Figure 5.B**)[116–118]. Whichever the docking mode, the activity of HslUV relies exclusively on ATP, which finely tunes the structure of HslU, modulating the HslV activity[119,120]. Of note, although adopting a barrel-shaped structure, HslUV differs from the proteasome 20S since the former consists of stacked hexameric rings while the latter is a pile of heptameric rings[83,116].

A.3.3 Anbu and BPH

As discussed earlier, the prokaryotic proteasome 20S is found only in archaea and actinomycetes[49]. Many bacterial phyla lacking the core proteasome possess HslV. Despite having a different symmetry, both proteolytic complexes share a common catalytic mechanism for their β subunits. Surprisingly, the proteasome 20S and HslUV are not found in cyanobacteria as well as some proteobacteria[121]. In these phyla, two genuine divergent proteasome-like complexes have been identified: the ancestral β unit protein (Anbu) and the betaproteobacteria proteasome homolog (BPH)[121]. The former is found in genomes of cyanobacteria and α-/β-/γ-proteobacteria while the latter occurs only in β-proteobacteria. Due to its narrow distribution, BPH is considered as a young proteasome that could be a descendant of HslV[121,122]. On the other hand, Anbu could be the common ancestor of HslV and the proteasome 20S[74,112,121] since cyanobacteria constitute an ancient bacterial phylum that has diverged some two billions years ago[123]. Recently, the structures of Anbu and BPH have been reported showing their relationship with the proteasome 20S and HslV[74,75,122,124].

The structure of BPH from *Thiobacillus denitrificans* has revealed that BPHs are tetradecameric complexes consisting of two stacked heptameric rings[122]. Albeit its seven-fold symmetry, the BPH structure is closely related to that of HslV (see **Figure 6**). The inner pore of BHP is

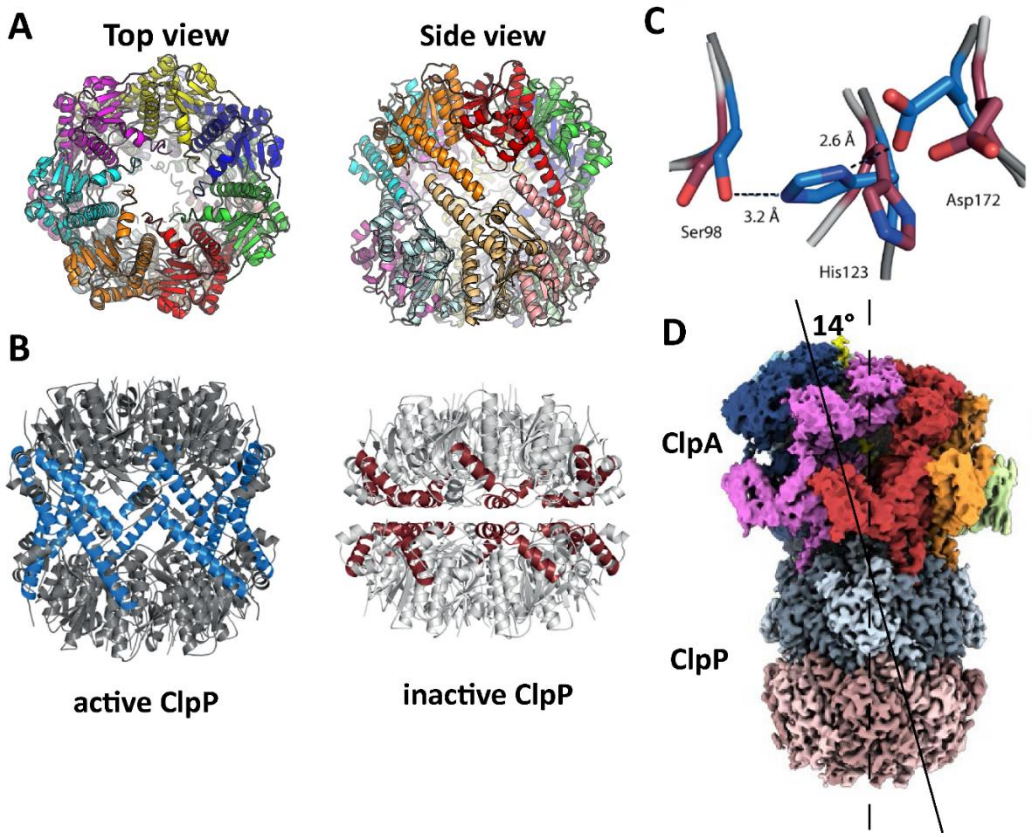


Figure 9 – Structure of ClpP and ClpAP. (A) Quaternary structure of ClpP forming two stacked heptameric rings viewed from the seven-fold (left) and a two-fold (right) symmetry axis. PDB code: 1TYF[132]. **(B)** The structure of active and inactive ClpP. The α helices of the "handle" domains which regulate the activity, are highlighted in blue for the active form and in red for the inactive form. **(C)** Close-up view of the ClpP catalytic triad (Ser-His-Asp). The position of the catalytic residue in active and inactive ClpP are emphasized in blue and red, respectively. Panels B and C reproduced from Gersch *et al.* (2012)[127]. **(D)** Structure of the ClpAP complex, with one ClpA hexameric ring capping the catalytic ClpP barrel. The symmetry axis of ClpA is tilted by 14° from that of ClpP. The ClpA subunits are arranged in a spiral conformation. Panel D reproduced from Lopez *et al.* (2020)[135].

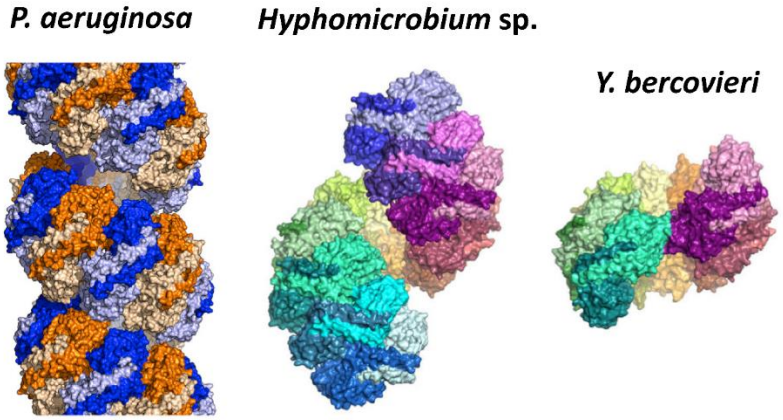


Figure 7 – Quaternary structures of Anbu from *P. aeruginosa*, *Hyphomicrobium* sp., and *Y. bercovieri*. Reproduced from Fuchs *et al.* (2017)[74] and Piasecka *et al.* (2018)[124].

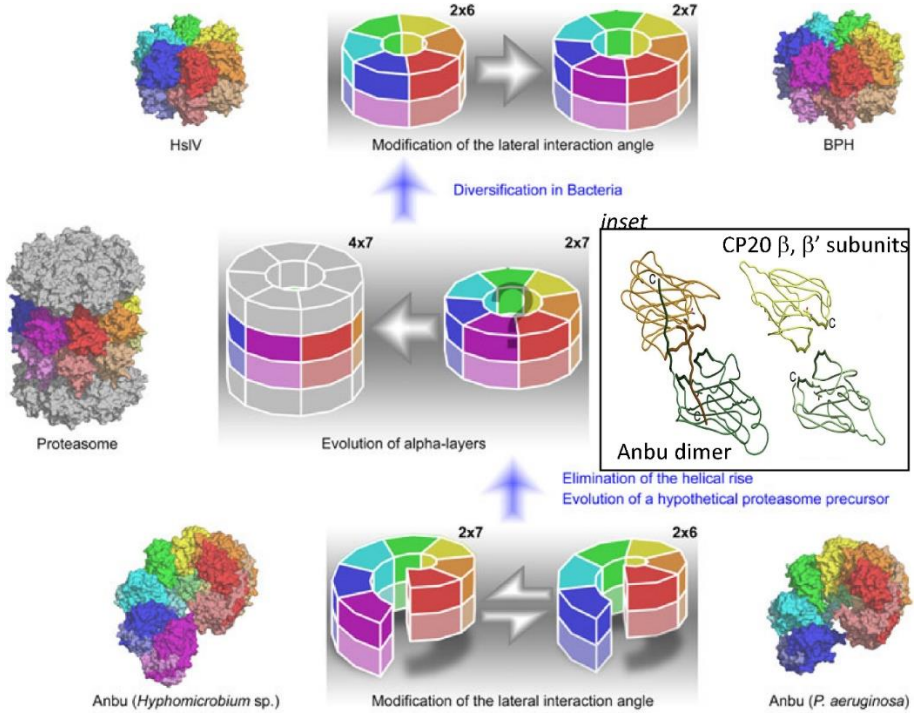


Figure 8 – Evolution history of the proteasome family with the symmetry changes observed in Anbu, the proteasome 20S (CP20), HslV, and BHP. Reproduced from Fuchs and Hartmann (2019)[112]. Inset: side-by-side representation of Anbu dimer and the β and β' subunits of CP20. The C-terminal tail lost during the evolution is highlighted in brown. Reproduced from Vielberg et al. (2018)[75].

larger than that of HslV but has a similar size to that of HslUV. The catalytic Thr-Lys-Asp triad is strictly conserved in BPH and the catalytic threonine residue can form a covalent complex with epoxomicin, a common inhibitor of threonine proteases. BPH appears to be inactive on various substrates and cannot complement Δ*lon* Δ*clp*XP Δ*hsl*UV in *E. coli*, although being constitutively expressed in *T. denitrificans*[122]. To date, three structures of Anbu are available, showing an unique quaternary structure (see **Figure 7**)[74,75,124]. Unlike the proteasome 20S and HslV, Anbu does not adopt a closed ring structure but forms a helical structure made of dimers. The three known structures present some differences regarding their oligomerization and quaternary structure. Anbu of *P. aeruginosa* is a dodecamer[74] while those of *Hyphomicrobium* sp. and *Yersinia bercovieri* are tetradecameric (see **Figure 7**)[75,124]. In addition, Anbu oligomers of *P. aeruginosa* and *Hyphomicrobium* sp. further organize spatially into a continuous helix while that of *Y. bercovieri* is more like a lock-washer. Like BPH, the Anbu subunit has the catalytic Thr-Lys-Asp found in HslV and the proteasome 20S but shows no proteolytic activity[74,75,124].

Interestingly, the closest structural homolog of Anbu is the β subunit of proteasome 20S from *Saccharomyces cerevisiae* and not bacterial HslV, suggesting that Anbu could be the ancestor of the eukaryotic proteasome 20S[75]. The evolutionary transition from Anbu to the proteasome 20S had probably involved the loss of the C-terminal tail of Anbu subunit (see **Figure 8**)[75]. Such an event had allowed the passage from a tetradecameric helix to a double stacked heptameric rings (see **Figure 8**)[112]. HslV had probably evolved from the proteasome 20S by the loss of the α subunit and the duplication of the β subunit[74]. Evolutionary speaking, HslV can be viewed as a simplification of the proteasome 20S since it can directly interact with its ATPase partner, HslU[112]. BHP could have evolved from HslV by reverting to the seven-fold symmetry.

A.3.4 ClpP

ClpP is another proteolytic system involved in protein turnover forming barrel-shaped structure[48,125]. Its quaternary structure consists of two stacked heptameric rings[126]. To achieve protein degradation, ClpP interacts with either ClpA or ClpX, two AAA+ acting as unfoldases[127]. Both ClpA and ClpX form hexameric rings capping both ends of the ClpP barrel[126]. They are known to recognize different degrons[128], but both ClpAP and ClpXP degrade SsrA-tagged substrates[71]. ClpXP is also known to have a regulatory role, degrading RpoH in response to RssB[129]. Despite their genuine proteolytic activity, the ClpAP and ClpXP complexes are dispensable in bacteria[130]. Albeit the striking resemblance of their quaternary structures, ClpP is not related to the proteasome 20S since the former is a serine protease[125]. ClpP is found in almost all bacteria phyla and in eukaryotes, but no homologs have been reported yet in archaea[125]. The function of ClpP homologs in eukaryotes remains ill-described but they are thought to maintain the mitochondrial proteome, increase the survival of different cancer cell types, and cause apoptosis when their activity goes rogue[131].

As explained above, ClpP is a tetradecamer made of two stacked heptameric rings (see **Figure 9.A**)[132]. The contact between two rings is stabilized by the "handle" domain of ClpP subunits. The "handle" domain is also known to control the activity via the bending of a long α helix (see **Figure 9.B**)[127,133]. This bending adjusts the position of catalytic residues to either an idle or active conformation (see **Figure 9.C**). The proteolytic barrel of ClpP is capped on both ends by a hexameric ring of either ClpA or ClpX[126]. Due to the symmetry difference, the AAA+ partner is not centered on

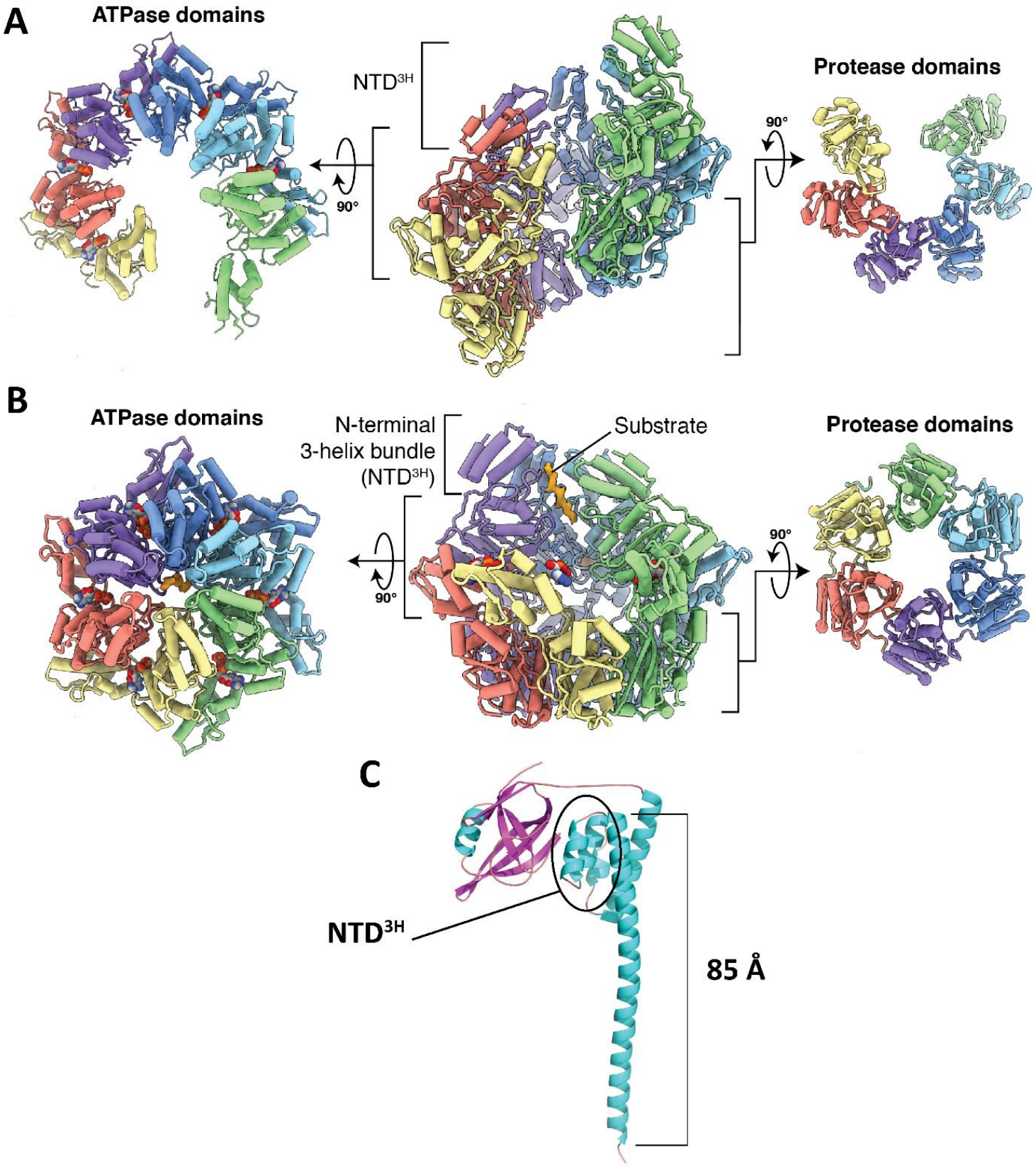


Figure 10 – Structure of LonA protease. (A) Open spiral conformation of Lon hexamer from *Yersina pestis*. NTD^{3H} corresponds to a part of the N-terminal domain (see panel C). **(B)** Closed ring conformation of Lon hexamer from *Y. pestis* in complex with ATP/ADP (represented as spheres), and a penta-alanine peptide as substrate (in orange). Panels A and B reproduced from Shin *et al.* (2020)[155]. **(C)** Structure of the N-terminal domain of Lon from *E. coli*. Reproduced from Li *et al.* (2010)[142]

the seven-fold symmetry axis of ClpP but is rather tilted by 14 to 16 Å (see **Figure 9.D**)[134,135]. Such a structural imbrication imposes a constraint on the relative position of the AAA+ domains, organized a spiral conformation[136]. Consequently, the AAA+ hexameric ring may unfold up to eight residues per ATP consumed.

A.3.5 Lon

Lon proteases are proteolytic complexes found in all kingdoms of Life[137]. They belong to the serine protease S16 family[61] but they have a catalytic dyad Ser-Lys instead of the classical Ser-His-Asp/Glu of the serine protease-type[138]. Three classes of Lon proteases have been described: (i) LonA, ubiquitous to bacteria and eukaryotes, (ii) LonB, found in archaea only, and (iii) LonC, mainly found in Gram negative bacteria[139,140]. Each class is defined by a different subunit architecture. LonA and LonB are both ATP-dependent proteases having a AAA+ domain along with the catalytic domain but they differ on their N-terminal domain. The N-terminal domain of LonA has been reported to be involved in substrate recognition[141,142]. On the other hand, LonB does not possess the N-terminal domain of LonA but has a transmembrane region inserted in the AAA+ domain[139,143]. LonB probably fulfils a function equivalent to that of FtsH since the latter is absent in archaea[139]. LonC has also three domains but its N-terminal and ATPase domains are nonhomologous to those of LonA and LonB[140]. Its N-terminal domain consists of a zinc finger while its ATPase domain is closely related to RecA rather than the AAA+ domain[144].

In *E. coli*, Lon (of the A class) has been reported to achieve about 50% of protein turnover related to abnormal translation termination[145]. Indeed, it recognizes SsrA-tagged proteins resulting from the ribosome rescue system[146]. Lon is also involved in degrading various protein substrates and regulating numerous pathways[137,147]. It recognizes especially aromatic sequences being exposed during protein unfolding[148]. In addition, Lon also plays a major role in the response to amino acid starvation by degrading ribosomal proteins[149]. Albeit its involvement in protein turnover, Lon is dispensable in *E. coli* under normal growth conditions as it is functionally redundant to HslUV, ClpAP, and ClpXP[76]. Regarding its catalytic activity, Lon can degrade in an ATP-dependent manner some substrates but also hydrolyses unfolded substrates without consuming ATP[150,151]. Three conformations have been described for Lon, depending on the availability of ATP and Mg^{2+}. In absence of both cofactors, Lon is completely inactive. Lon displays a high ATPase activity when ATP is abundant while in the presence of Mg^{2+}, the proteolytic activity is maximal[150,152]. Finally, LonA and LonC has been described to bind nonspecifically to chromosomal DNA via, respectively, the AAA+ domain and the catalytic domain[144,149].

The members of the three Lon classes have been reported to be hexamers adopting similar quaternary structures[138,141,143,144,152–157], although no structural data is available regarding the ATPase domain of LonC and the transmembrane region of LonB. Lon can be regarded as a trimer of dimers adopting either a closed hexameric ring[138,144,153–155,157] or an open spiral conformation[141,143,155]. Mg^{2+} has been shown to modulate the self-assembly of dimers into a hexamer in an ATP-independent manner[157,158]. The open spiral conformation corresponds to the idle state of Lon (see **Figure 10.A**)[141,143,155]. The six monomers are spatially organized in a left-handed spiral that resembles to the "lock-washer" structure of Anbu (see **section A.3.3**). The closed hexameric ring conformation is promoted by the presence of ATP and a substrate (see **Figure 10.B**)[152,155]. In that conformation, the catalytic domains are buried in an inner cavity, only accessible via the gate made of the AAA+

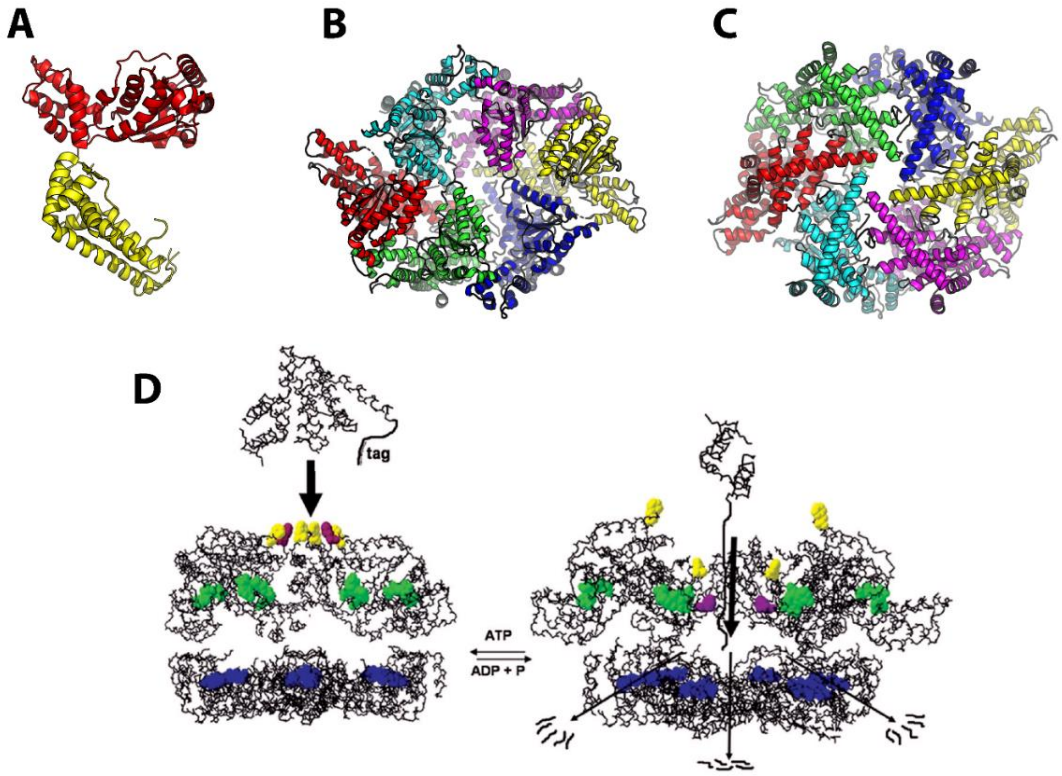


Figure 12 – The cytoplasmic part of FtsH. (A) The structure of subunit with the AAA+ domain highlighted in red and the catalytic domain in yellow. PDB code: 2CEA. **(B)** Quaternary structure of hexameric of cytoplasmic FtsH centered on the central pore with the AAA+ domains facing up. **(C)** Quaternary structure of hexameric of cytoplasmic FtsH centered on the catalytic domain, at the opposite side of the central pore. **(D)** Schematic drawing of the degradation mechanism of FtsH. Panel D reproduced from Bieniossek *et al.* (2006)[168].

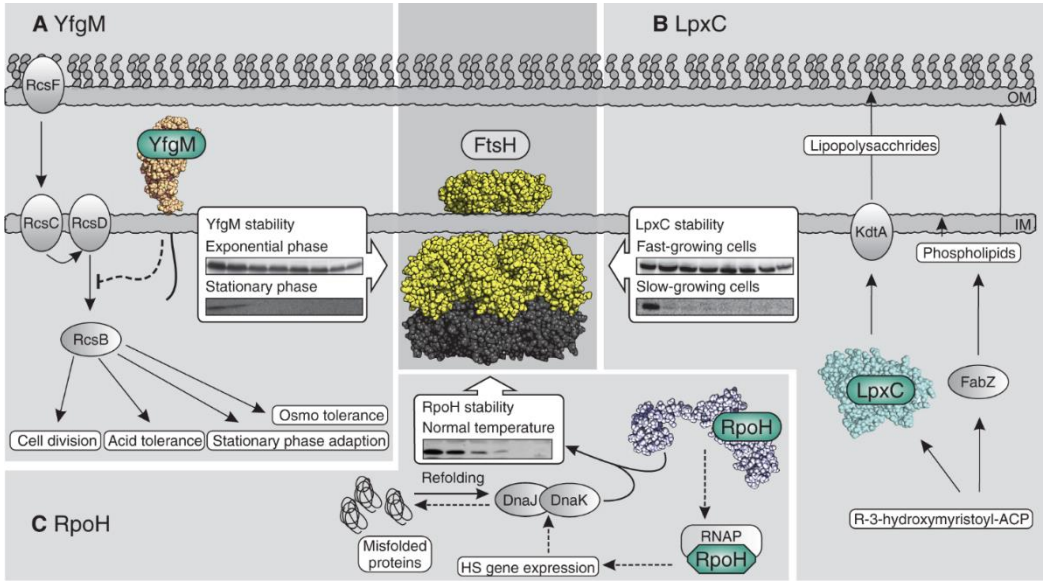


Figure 11 – The membrane anchored FtsH regulating the levels of (A) YfgM, (B) LpxC, and (C) RpoH. Reproduced from Bittner *et al.* (2017)[79].

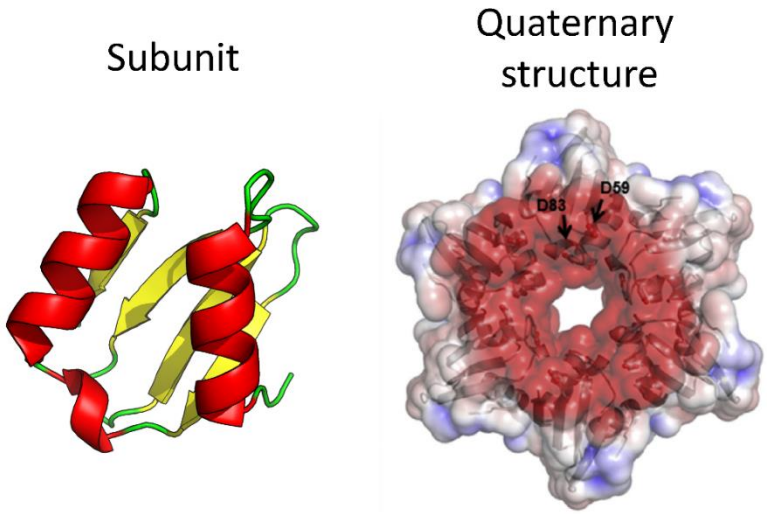


Figure 13 – The periplasmic domain of FtsH. *Left:* the subunit structure of the periplasmic domain. PDB code: 4M8A. *Right:* the surface representation of the quaternary structure of the periplasmic domain. Reproduced from An *et al.* (2018)[166].

domains[138,153]. The N-terminal domain of LonA has also been structurally characterized although separately from the AAA+ and catalytic domains (see **Figure 10.C**)[142,159,160]. Notably, it displays a very long α helix of about 85 Å in length. This oversized α helix has been proposed to function like a "crowbar" to promote the disaggregation of protein aggregates[142].

A.3.6 FtsH

FtsH is the fifth AAA+ proteolytic complexes encountered in bacteria[48]. It is a membrane anchored protein achieving mainly the turnover of membrane proteins[73,79,80]. FtsH plays also a critical role in regulating the level of lipid A as shown in *E. coli* (see **Figure 11**)[161]. The regulation is mediated by the conditional proteolysis of LpxC, the UDP-3-O-(R-3-hydroxymyristoyl)-N-acetylglucosamine deacetylase, catalyzing the first step of lipopolysaccharide synthesis[79,80,161]. The level of LpxC is known to be lowered during the stationary phase via FtsH-mediated proteolysis. FtsH also regulates the level of RpoH (also known as σ^{32}) and YfgM (see **Figure 11**)[79]. In normal growth condition, the chaperonins DnaK and GroEL/S bind to RpoH and send it to FtsH to be degraded. During a heat stress, the chaperonins release their grip on RpoH, inducing the transcription of the heat stress regulon[162]. Meanwhile, YfgM is a transmembrane protein having a cytoplasmic domain and a periplasmic domain[163]. The latter has a putative chaperonin function while the former sequestrates the Rcs phosphorelay regulator, RcsB[163,164]. Under osmotic stress or in stationary phase, YfgM is rapidly hydrolyzed by FtsH, releasing RcsB[163].

Currently, the structure of full-length FtsH remains unknown but its cytoplasmic part and periplasmic part have been structurally characterized[80,165–168]. The cytoplasmic part is divided in three domains (see **Figure 12.A**)[80,165,168]. The first two subdomains, an α/β fold domain and a four-helix bundle, constitute the AAA+ domain, involved in nucleotide binding[167]. The third domain possesses the catalytic site and shows a unique fold for a co-catalytic peptidase. Despite having the "HEXXH" motif, the FtsH catalytic domain fold does not look like zincins[168], which are a superfamily of zinc-dependent peptidases[169]. Consequently, a MEROPS family has been designated for FtsH, the metallopeptidase M41 familiy[61]. It is worth noting that the FtsH subunit carries both the ATPase and the protease functions, unlike ClpP and HslV. Amazingly, the cytoplasmic fragment of FtsH is a hexamer with a quaternary structure reminding HslV (see **Figure 12.B-C**)[165,168]. All the AAA+ domains are oriented on one side of the hexamer and the six catalytic domains on the other side. Protein substrates are believed to enter in the central pore on the side having the AAA+ domains (see **Figure 12.D**)[168]. Interestingly, the periplasmic part also forms a hexamer, strongly suggesting that full-length FtsH is hexameric too (see **Figure 13**)[166]. It consists of a β sheet with four strands surrounded by two α helices and a 3_{10} helix. The function of the periplasmic domain is still unknown but it could be involved in substrate recognition in the perisplasm[166].

A.3.7 HtrA

Proteases of the HtrA-family, belonging to the S1C family, are widespread among all kingdoms of Life[81]. In bacteria, HtrA proteases are localized in the periplasm but, in some species, they are also found in the extracellular space[39,81]. They are involved in the quality control of periplasmic and outer membrane proteins, acting as either chaperones or proteases[81]. The shift between both activities seems to be temperature dependent, the chaperone activity being predominant at low temperature and the protease activity being induced by heat stress[170]. Since there is no ATP in the periplasm or

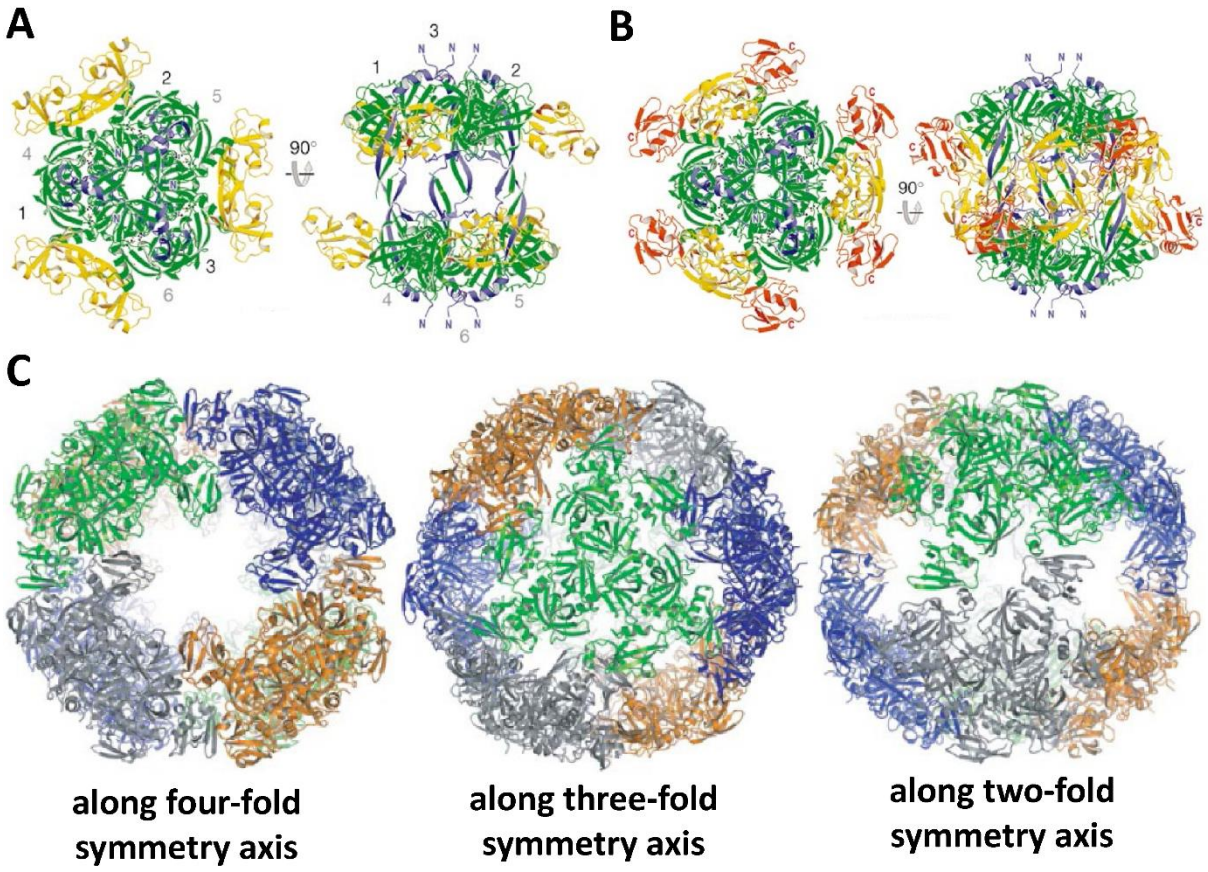


Figure 14 – The quaternary structures of DegP. (A) Structure of hexameric DegP in an "open" conformation. It consists of two stacked trimeric rings. The catalytic domains are highlighted in green, the PDZ-like domains in yellow and red. The distance between two stacked trimers is about 15 Å. **(B)** Structure of hexameric DegP in a "closed" conformation. The two PDZ-like domains form extensive contacts. Panels A and B reproduced from Krojer *et al.* (2002)[173]. **(C)** Structure of tetracosameric DegP viewed from different symmetry axes. Each trimer is highlighted in different colors. Reproduced from Krojer *et al.* (2008)[175].

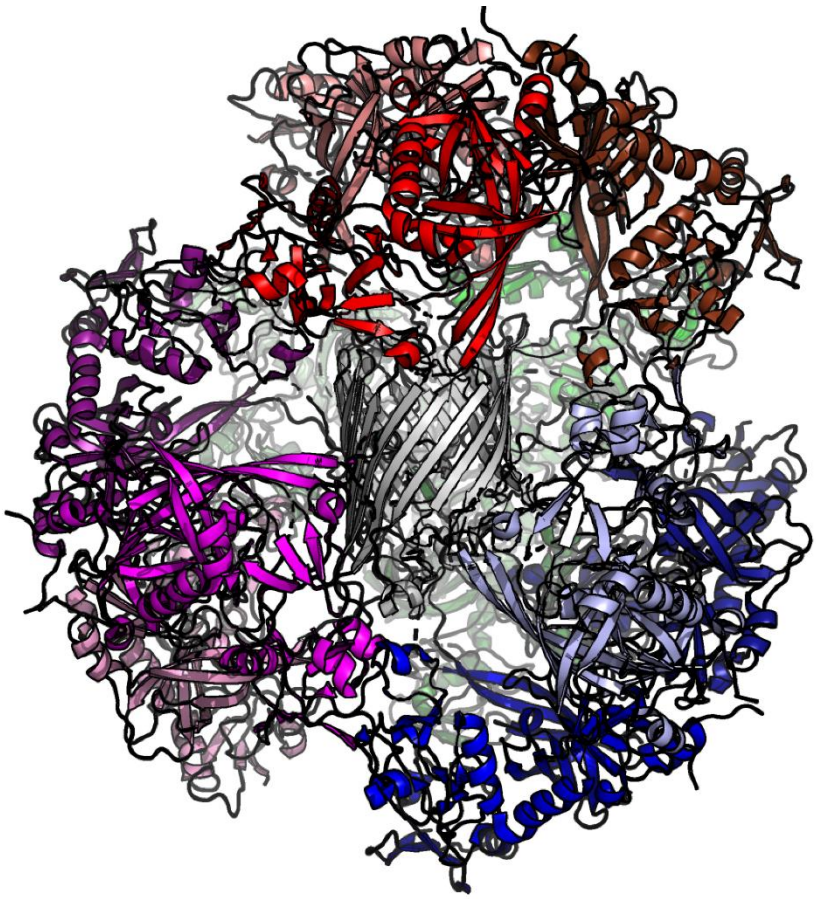


Figure 15 – Structure of dodecameric DegP encapsulating an OmpC molecule. The DegP trimers are emphasized in red, green, blue, and magenta; OmpC in grey. PDB code: 2ZLE[175].

the extracellular space, HtrA proteases are completely devoid of any ATPase activity[171]. Consequently, protein degradation must rely on a different system for protein unfolding. The most well-studied HtrA protease is DegP from *E. coli*. Initially, DegP was characterized as a protease having a limited activity on a few target proteins *in vitro*[170–172]. The first structural study showed that DegP is an hexamer made of two stacked trimeric rings (see **Figure 14.A**)[173]. The DegP subunit consists of three domains: a catalytic domain and two PDZ-like domains. The PDZ-like domains were described to control the distance between two stacked trimers (see **Figure 14.B**) and be involved in substrate binding. The distance between the two rings is about 15 Å, thus only oligopeptides or unfolded proteins can enter in the catalytic tunnel[173].

Hexameric DegP, however, is not regarded as the enzyme active form since a trimeric variant retains the chaperone and proteolytic activities[174]. When incubated with denatured proteins, DegP hexamers have been shown to self-assemble into cage-like superstructures with 12 or 24 subunits (see **Figure 14.C**)[175,176]. To attain such quaternary structures, hexamers must intermediately dissociate into trimers. It has been proposed that the signal promoting the self-assembly is the covalent binding of a degron to the first PDZ-like domain[177]. The tetracosamer of DegP has a molecular weight of 1.13 MDa and has a spherical shape with a diameter of about 200 Å[175,176]. The inner cavity is so wide that it can accommodate a protein of 300 kDa. Thus, DegP is large enough to make contact, via its PDZ-like domains, between the inner and outer membranes in *E. coli*[175]. It could help outer membrane proteins to be folded and transported to the outer membrane, as shown in the structure of dodecameric DegP in complex with OmpC (see **Figure 15**)[175]. In addition, DegP dodecamers and tetracosamers have been shown to display a higher proteolytic activity than hexamers[176]. It is worth noting that *E. coli* possesses two other HtrA proteases, DegQ and DegS. DegQ seems to be also involved in the quality control of periplasmic proteins, forming the same cage-like structures than DegP[178,179]. Meanwhile, DegS forms only trimers and acts as a sensor of unfolded proteins in the periplasm[180]. The binding of unfolded peptides to DegS triggers the degradation of RseA, relieving the suppression of σ^E stress response[180].

A.4 Proteasome-associated peptidases in prokaryotes

The previous sections have introduced the major proteolytic complexes found in prokaryotes. The size of peptides generated during proteolysis ranges from 3 to 30 amino acids in length, with a mean length of six to nine residues[181]. In other words, the proteolytic complexes hydrolyze only 10-15% of peptide bounds in proteins, while auxiliary enzymes must degrade the other 85-90% of peptide bounds[181]. There is a bountiful of auxiliary oligopeptidases that achieve peptide degradation. It would be challenging to present every oligopeptidase families. Nonetheless, some of them have been linked to proteolytic complexes or have a nanocompartmentalized activity. Nanocompartmentalization is often seen as a strategy to avoid uncontrolled degradation of disordered parts of proteins. Several oligopeptidases have been linked to peptide degradation in *E. coli*, especially PepA, PepB, PepD, PepN, and PepQ (see **section A.4.1**)[182]. In archaea, Hsp31 homologs and acylpeptide hydrolases could also be involved in peptide processing downstream the proteasome (see **sections A.4.2-A.4.3**)[183]. Finally, two gigantic self-compartmentalized peptidases, the Tricorn peptidases and TET peptidases, have also been designated to act downstream the proteasome (see **sections A.4.4-A.4.5**)[50,184].

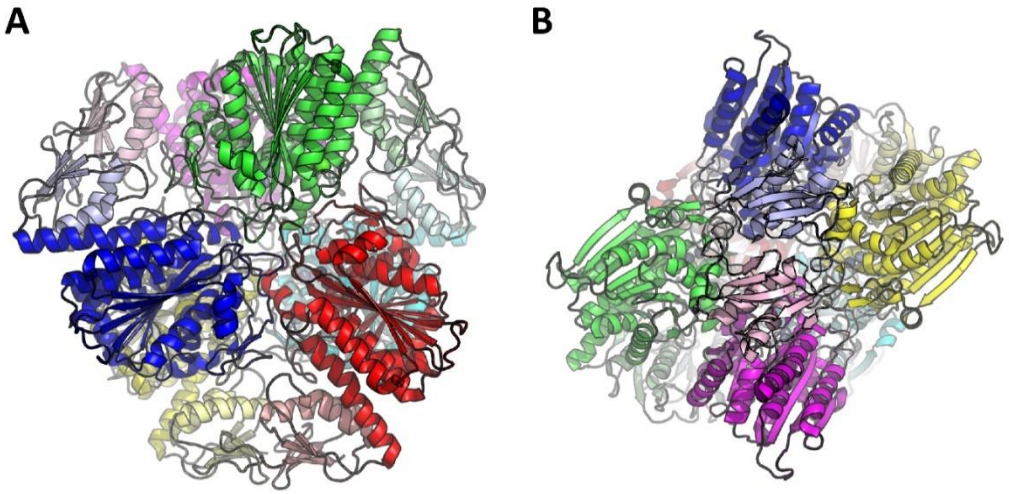


Figure 17 – Structure of *E. coli* PepB. Schematic representation viewed from **(A)** the 3-fold axis and **(B)** a 2-fold axis. The catalytic domains and the dimerization domains are shown in dark and light colors, respectively. PDB code: 6OAD[200].

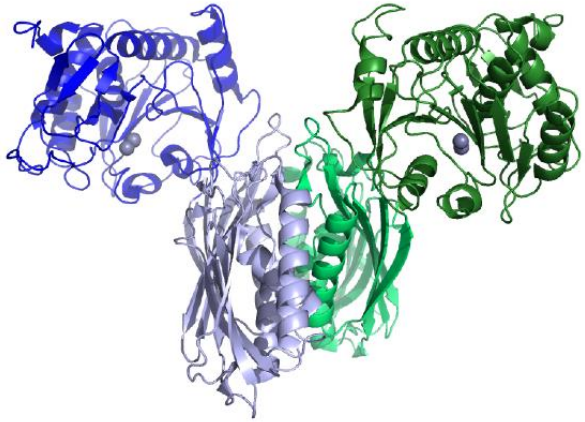


Figure 18 – Structure of *Vibrio alginolyticus* PepD. The two subunits are highlighted in blue and green. The catalytic domains and the dimerization domains are shown in dark and light colors, respectively. PDB code: 3MRU[204].

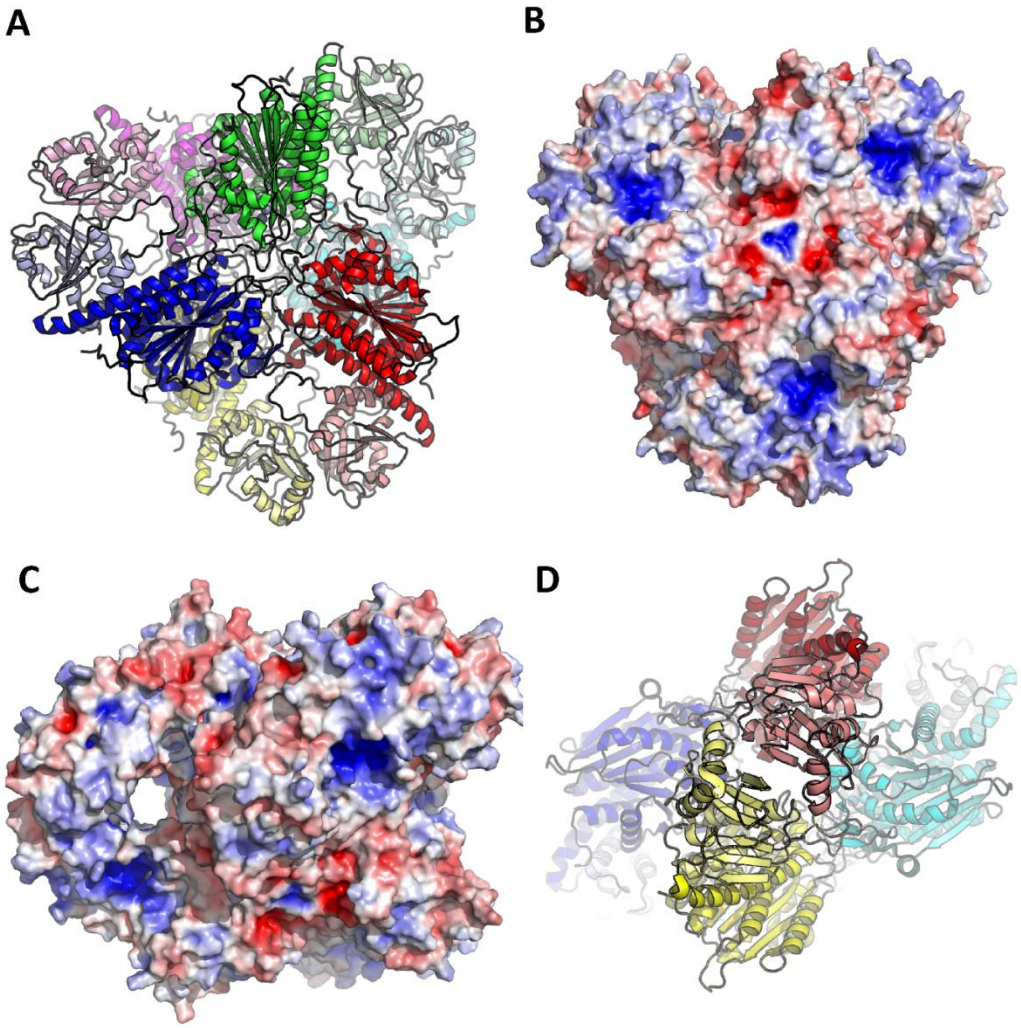


Figure 16 – Structure of *E. coli* PepA. (A) Schematic representation and **(B)** surface representation of the *E. coli* PepA structure viewed from the 3-fold symmetry axis. **(C)** Structure of *E. coli* PepA centered on a channel leading to the catalytic chamber. **(D)** Structure of *E. coli* PepA viewed from a 2-fold symmetry axis. The catalytic domains and the dimerization domains are shown in dark and light colors, respectively. The electronic surface was calculated using APBS[225]. PDB code: 1GYT[198].

A.4.1 Peptidases A, B, D, N, E, P, and Q

Miller and Schwartz (1978) reported that four peptidases (PepA, PepB, PepD, and PepN) are involved in peptide degradation in *E. coli*[182]. The knock-out of their respective genes confers a resistance to polyvaline peptides. In addition, such a Δ*pep*A Δ*pep*B Δ*pep*D Δ*pep*N strain is dramatically impaired in metabolizing peptides[182,185] but it also accumulates peptides during growth, probably produced during protein turnover and signal peptide processing[186]. PepA and PepB belong to the metalloprotease M17 family. They have a broad substrate specificity but PepB can hydrolyze Asp-Xaa dipeptides while PepA cannot[187,188]. PepA has been studied thoroughly due to its moonlighting functions. Indeed, it is involved in the regulation of the *car*AB operon and the site-specific recombination of ColE1[189–192]. ColE1 is a natural plasmid presents in high copy number in *E. coli* thanks to the site-specific *cer* recombination system resolving unstable plasmid multimers into monomers[191]. The *cer* recombination is strictly dependent on two trans-acting factors, ArgR and PepA. The former is a regulatory protein controlling the expression of various regulons involved in arginine metabolism[193]. The role of PepA in the *cer* recombination does not require its aminopeptidase activity[194]. The regulation of the *car*AB operon also involves both ArgR and PepA[190]. The latter ensures a regulatory function on its own as the DNA of *car*P1 operator warps around one PepA molecule, inhibiting the transcription initiation[195,196]. Of note, PepA, as a regulator, plays an important role in the virulence of *Pseudomonas aeruginosa* and *Vibrio cholerae*[45,197].

The structure of PepA is a triangle-shaped hexamer consisting of two stacked trimers (see **Figure 16.A-D**)[198,199]. The triangle length is about 135 Å while its thickness is about 80 Å. The PepA subunit has two domains: (i) a N-terminal domain involved in the dimerization of two trimers and (ii) a C-terminal domain harboring the active site. The six catalytic domains are clustered around the three-fold symmetry axis of the triangle-shaped structure (see **Figure 16.A-B**), forming an inner cavity of about 15 Å of diameter. Three channels give access to the inner cavity. Each channel is delineated by two N-terminal domains and two C-terminal domains (see **Figure 16.C**). Consequently, the quaternary structure of PepA can be regarded as a nanocompartmentalization of activity. Each catalytic domain of a trimer interacts with another catalytic domain of the dyad trimer. The interaction between two trimers also rely on the extensive contacts between the N-terminal domains (see **Figure 16.D**). PepB adopts a hexagon-shaped hexamer with two stacked trimers (see **Figure 17.A-B**)[200]. Like PepA, PepB has a N-terminal domain and a C-terminal domain having the same functions. The N-terminal domain of PepB, however, is shorter than its PepA counterpart. The N-terminal extension of PepA is believed to be involved in protein-DNA interaction[198]. The six C-terminal catalytic domains are also oriented inwards the PepB structure and are involved in stabilizing the interaction between trimers (see **Figure 17.A**). The N-terminal domains assume the dimerization of two trimers, like in PepA (see **Figure 17.B**). It is worth noting that the absence of metal ions in the catalytic site of PepB provokes the displacement of several key-residues important in catalysis via loop disorders[200].

PepD has first been identified as a carnosinase in *E. coli* and *Staphylococcus typhimurium*, providing a source of histidine[201,202]. Although PepD is not restricted to Xaa-His bond cleavage and exhibits a broad substrate specificity but with a low affinity[201,203]. As expected from its membership to the metalloprotease M20 family[61] (see also **section A.5.1.1**), PepD is a dimer forming a peculiar V-shaped structure (see **Figure 18**) as shown for the *Vibrio alginolyticus* enzyme[204]. Its subunit consists of two domains: (i) a catalytic domain with the typical α/β fold of the MH clan members, and (ii) a

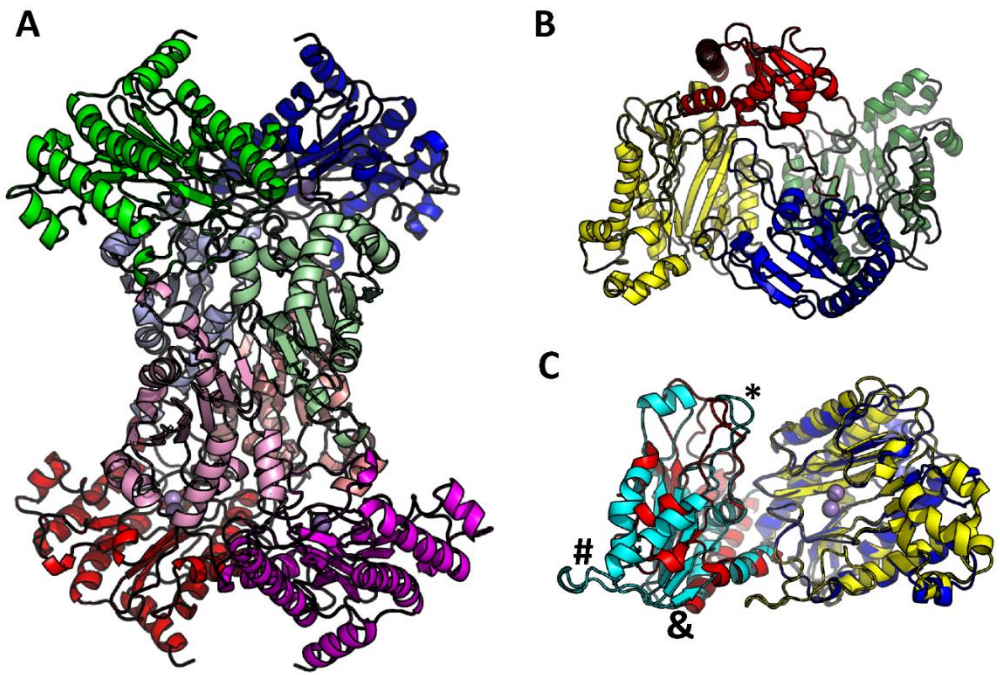


Figure 21 – Structures of *E. coli* PepP and PepQ. (A) Structure of PepP tetramer, with the subunits highlighted in green, blue, magenta, and red. The catalytic domains and the dimerization domains are shown in dark and light colors, respectively. PDB code: 1A16[222]. **(B)** Structure of PepQ dimer with the catalytic domains colored in yellow and green and the dimerization domains colored in blue and red. PDB code: 4QR8[221]. **(C)** Structural alignment of PepP vs. PepQ. The catalytic domain and dimerization domain of PepP are colored in blue and cyan, respectively. The catalytic domain and dimerization domain of PepQ are colored in yellow and red, respectively. The long loop involved in dimerization is marked by an asterisk (*), the extra β strand of PepP by an ampersand (&), and the two elongated α helices of PepP by an octothorpe (#).

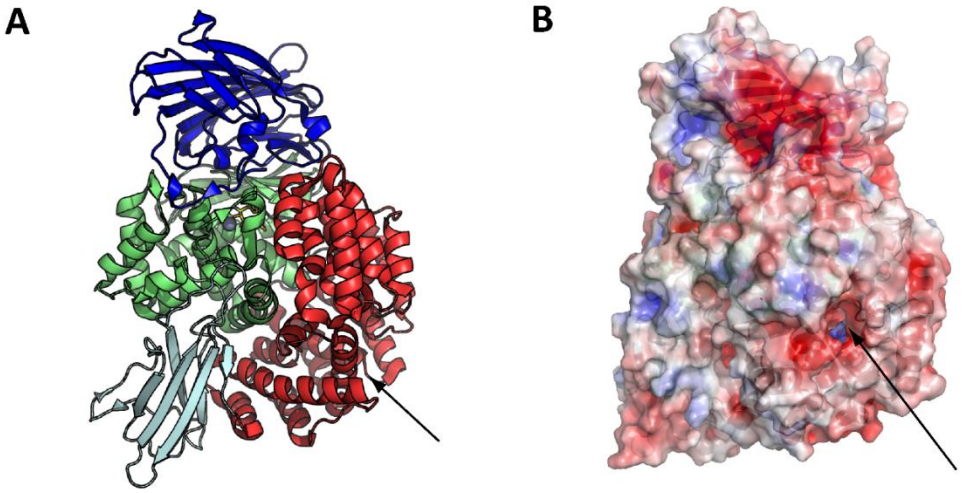


Figure 19 – Structure of *E. coli* PepN. (A) Schematic representation and **(B)** surface representation. The arrows indicate the access pore to the catalytic site. The catalytic domain is shown in green, the three other domains are colored in cyan, blue and red. The electronic surface was calculated using APBS[225]. PDB code: 2HPT[212].

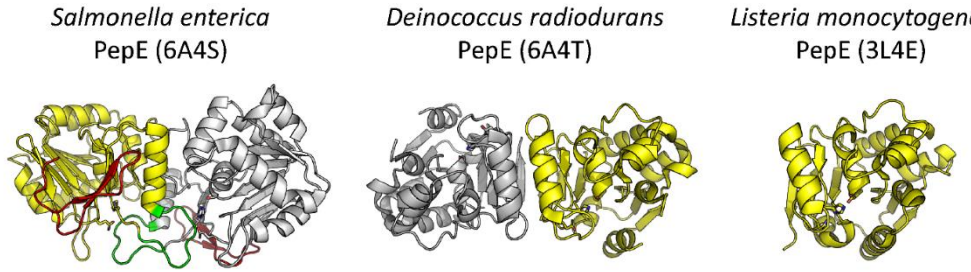


Figure 20 – Structures of PepE from *Salmonella enterica*, *Deinococcus radiodurans*, and *Listeria monocytogenes*. For dimeric enzymes, the second subunit is colored in grey. The β hairpin tail and loop linked to aminopeptidase activity are colored in red and green, respectively.

dimerization/lid domain with an eight-strand β sheet on one side and four α helices on the other side. In addition to its role in dimer formation, the dimerization/lid domain could control the access of the substrate to the catalytic chamber[204].

PepN belongs to the metalloprotease M1 family, which is widely distributed in the whole tree of Life[61]. It has been proposed to fulfil peptide degradation downstream proteasome-related complexes in *E. coli*[205,206]. The alanyl-aminopeptidase activity in *E. coli* has mainly been attributed to PepN[207] but it accommodates preferentially basic residues in its S1 pocket[205,206,208,209]. Meanwhile, PepN does not recognize peptides with an acidic residue at the P1 position[205,208,209]. The deletion of *pep*N is not lethal but it results in a decreased hydrolysis of some synthetic peptides[206,210]. Consequently, peptide turnover probably relies on redundant peptidases in *E. coli* as observed by Miller and Schwartz (1978)[182]. PepN seems to be involved in maturing peptidyl nucleosides such as blasticidin S in *Streptomyces griseochromogenes*[211]. Regarding its structure, PepN does not oligomerize but its catalytic cleft is compartmentalized. It has four domains, one of which being the catalytic domain (see **Figure 19.A**)[208,212]. The latter adopts a thermolysin-fold typical of gluzincins[213]. The catalytic domain of PepN is sandwiched by two domains, restricting the access to the catalytic cleft. The molecular mechanism behind a controlled substrate entry, however, is not yet fully understood but the substrate seems to enter the catalytic cleft via a pore of about 8-10 Å in size (see **Figure 19.B**)[208]. In addition, the catalytic domain could alternate between an active closed conformation and an inactive open conformation depending on substrate binding[208,212]. Of note, the structure of PepN is closely related to that of the aminopeptidase F3, an auxiliary peptidase interacting with the Tricorn peptidase (see **section A.4.4**)[212,214].

Three other peptidases, PepE, PepP, and PepQ, could be complementary to the four peptidases described above. PepE is a strict aspartyl-aminopeptidase releasing N-terminal aspartate residue from di- and tripeptide[215]. It belongs to the serine protease S51 family[61], exhibiting an unique fold completely unrelated to the subtilisin-like, trypsin-like, and serine carboxypeptidase-like folds[216–218]. PepE homologs, however, are rather structurally heterogeneous regarding their oligomerization state and catalytic triad (see **Figure 20**)[217]. Indeed, both monomers and dimers are encountered in the S51 family. In addition, three different configurations of catalytic triad have been reported so far: Ser-His-Glu, Ser-His-Asp, and Ser-His-Asn. All monomeric PepE homologs display the Ser-His-Asn triad (see **Figure 20**), resulting in a loss of aminopeptidase activity[217,218]. Even dimers show an activity variability, depending on the presence of a β hairpin tail and loop capping the active site (see **Figure 20**). PepE homologs without these structural motifs act as an esterase rather than an aminopeptidase[217]. PepP and PepQ are both prolyl-aminopeptidases cleaving Xaa-His bond[219,220]. PepQ hydrolyzes dipeptides while PepP degrades longer peptides. Both enzymes are structurally related since they belong to the metalloprotease M24 family[61]. Their catalytic domains adopt the common Pita-bread fold of the M24 family (see **Figure 21.A-B**)[221–224]. PepP and PepQ, however, show structural differences regarding their dimerization domains and oligomerization states. The dimerization domains of PepP and PepQ are related, showing the same dimerization interface. Indeed, a long loop interacts with its dyad counterparts capping the catalytic domain (see **Figure 21.C**). The dimerization domain of PepP, however, has an additional β strand in the central β sheet and two elongated α helices that promote the formation of tetramers (see **Figure 21.C**).

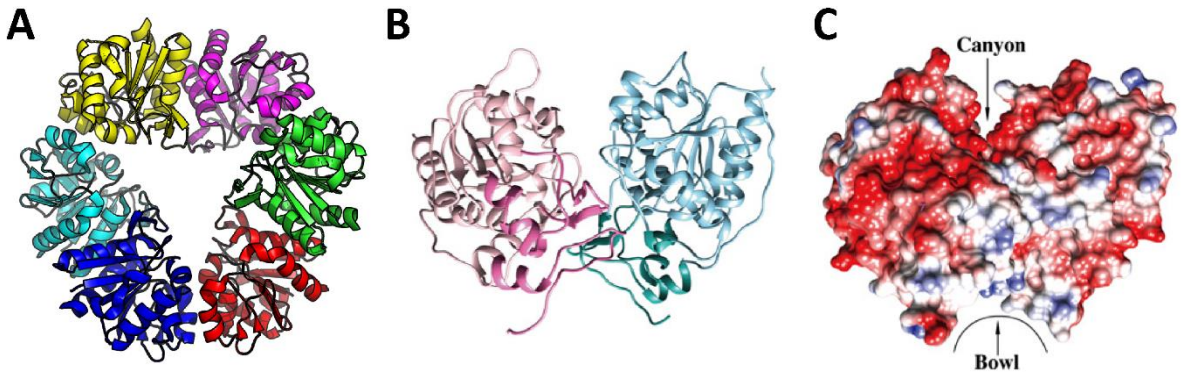


Figure 22 – The Hsp31 peptidases. (A) Structure of PH1704 hexamer with the six subunits highlighted in red, green, blue, yellow, magenta, and cyan. PDB code: 1G2I[228]. **(B)** Schematic representation and **(C)** surface representation of the Hsp31 structure from *S. aureus*. The two subunits are colored in pink and cyan, with the P domain highlighted in dark color. Reproduced from Kim *et al.* (2017)[234].

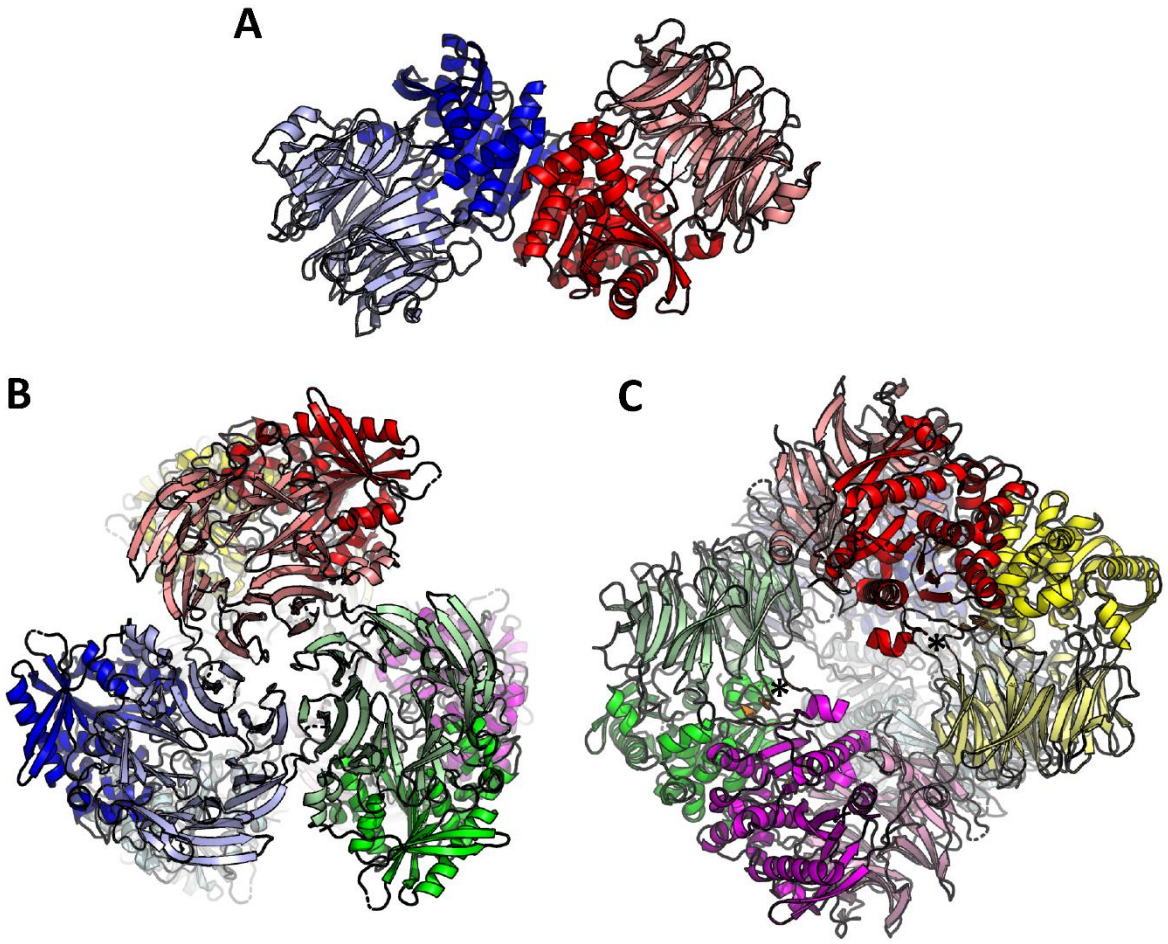


Figure 23 – The acylpeptide hydrolases of the S9C family. (A) Structure of APH from *A. pernix*, with the subunits highlighted in blue and red. The catalytic domain and the β7 propeller domain are shown in dark and light colors, respectively. PDB code: 1VE6[241]. **(B)** Structure of APH from *P. horikoshii* viewed from the 3-fold symmetry axis, with the six subunits highlighted in blue, red, green, yellow, magenta, and cyan. The catalytic domain and the β7 propeller domain are shown in dark and light colors, respectively. PDB code: 4HXE[243]. **(C)** Structure of APH from *P. horikoshii* viewed from a 2-fold symmetry axis, corresponding to an access pore to the central catalytic chamber. The loop extension of the β7 propeller domain is highlighted in orange and marked by an asterisk (*).

A.4.2 Hsp31 peptidases

Cystein proteases of the C56 family could also be involved in peptide degradation[183,226]. The C56 family occurs in all kingdoms of Life but its members have moonlighting functions. Archaeal enzymes are described as proteases since PfpI and PH1704 of *P. furiosus* and *P. horikoshii*, respectively, can degrade casein and gelatin[227–229]. Despite their proteolytic activity, they are more likely oligopeptidases since they cannot hydrolyze peptides of more than 20 amino acids in length[230]. Both enzymes are homologous to Hsp31, a heat shock protein found in bacteria and fungi[231]. In *E. coli*, Hsp31 is a moonlight enzyme behaving as a chaperone[231,232], an aminopeptidase[226], and a glyoxylase III[233]. The combination of chaperone and aminopeptidase activities is not surprising since other proteases fulfil both functions, such as DegP[175]. The genuine glyoxylase activity of Hsp31 is glutathione-independent and has been proven to be required in glycated proteins[233,234]. Albeit being homologs, the oligomerization states of archaeal PfpI and PH1704 differ from that of Hsp31. Indeed, PfpI and PH1704 have been reported to form dimer, trimer, and hexamer[227,228]. The structure of PH1704 hexamer, the most active oligomeric form, has been solved showing a toroidal quaternary structure with a central hole limiting the access to the active sites (see **Figure 22.A**)[228]. Meanwhile, Hsp31 forms dimers only, as shown in *E. coli* and *S. aureus* (see **Figure 22.B**)[234,235]. Its subunit consists of two domains, A and P, both important for catalysis and oligomerization. Since being absent in PfpI, the domain P has been proposed to hinder the hexamer formation[229,235]. The quaternary structure of Hsp31, however, displays two structural features, a “canyon” and a “bowl”, probably involved in protein-protein interactions (see **Figure 22.C**). Such interactions could be important for Hsp31 activity[231,235] and intervene in the interaction of Hsp31 with ClpAP, GroEL, and the elongation factor EF-Tu in *E. coli*[226].

A.4.3 Acylpeptide hydrolases and prolyl oligopeptidases

Acylpeptide hydrolases (APH) and prolyl oligopeptidases (POP) are serine proteases of the S9 family[61]. While the latter is widely distributed in the whole tree of Life, the former is found mainly in eukaryotes and archaea. Consequently, they have been classified into two subfamilies: APH is part of the S9C subfamily and POP belongs to the S9A subfamily. APH and POP have been proposed to degrade peptides downstream the proteasome due to their broad substrate specificities[236]. APH hydrolyzes N-terminally blocked peptides, like acylated and formylated peptides, releasing acyl amino acids[237,238]. Archaeal enzymes were thought to be strict APH having an exopeptidase activity[238] but they have been shown to hydrolyze unblocked peptides and act as endopeptidases as well[239,240]. Several structures of archaeal APH are readily available showing diverse oligomeric states despite sharing a common subunit structure[241–243]. Indeed, APH of *Aeropyrum pernix* is a dimer (see **Figure 23.A**) while that of *Pyrococcus horikoshii* is a hexamer (see **Figure 23.B**). The difference of oligomeric states could be linked to nanocompartmentalization seen in the *P. horikoshii* enzyme. In dimeric APH, the access to the catalytic site is probably controlled via an open/closed conformation, each subunit being independently active[242]. In *P. horikoshii* APH, the six subunits are organized to form three permanent pores leading to a central catalytic chamber towards which point the six active sites (see **Figure 23.B-C**)[243]. The typical APH subunit consists of two domains, a seven-blade β propeller and an α/β catalytic domain (see **Figure 23.A-C**). The former is probably involved in substrate specificity, guiding substrates to the catalytic site. The catalytic domain displays the Ser-His-Asp catalytic triad

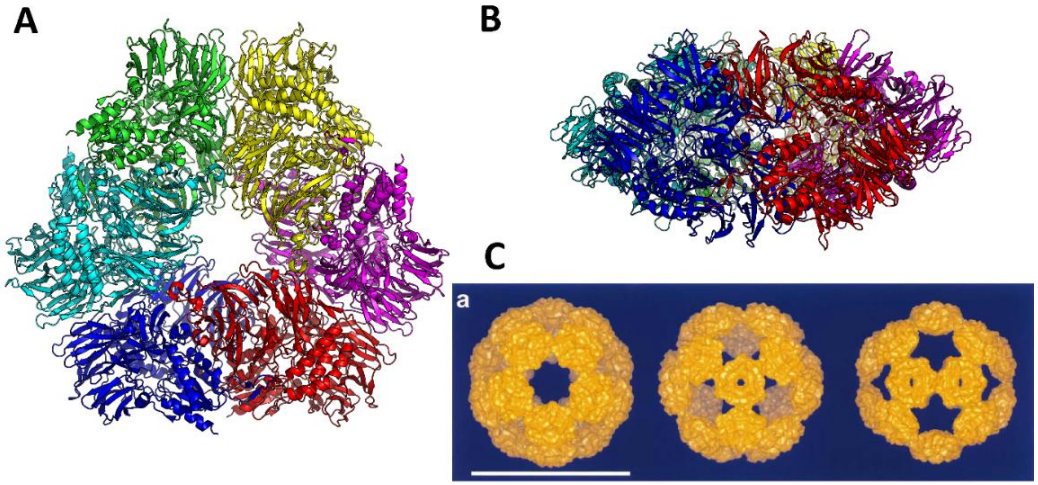


Figure 25 – The quaternary structure of Tricorn peptidases. (A) Structure of TaTRI hexamer (PDB code 1K32) viewed from the three-fold symmetry axis. **(B)** Side-view of the structure of TaTRI hexamer viewed from a two-fold symmetry axis. **(C)** Representations of TaTRI capsid structure solved by CryoEM viewed from a five-fold (left), three-fold (center), and two-fold (right) symmetry axis. Panel C was reproduced from Walz *et al.* (1999)[257].

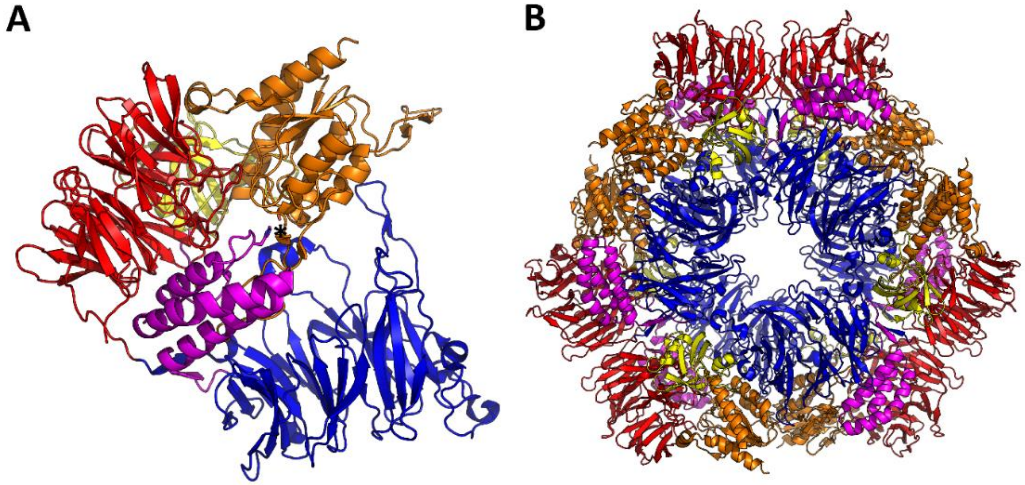


Figure 26 – Spatial organization of TaTRI domains. (A) Structure of TaTRI subunit (PDB code 1K32) with its five domains differently highlighted: the β6 propeller domain (in red), β7 propeller domain (in blue), the core domain C1 (in magenta), the PDZ domain (in yellow), and the core domain C2 (in orange). The catalytic site is marked by an asterisk. **(B)** Structure of TaTRI hexamer, viewed from the three-fold symmetry axis.

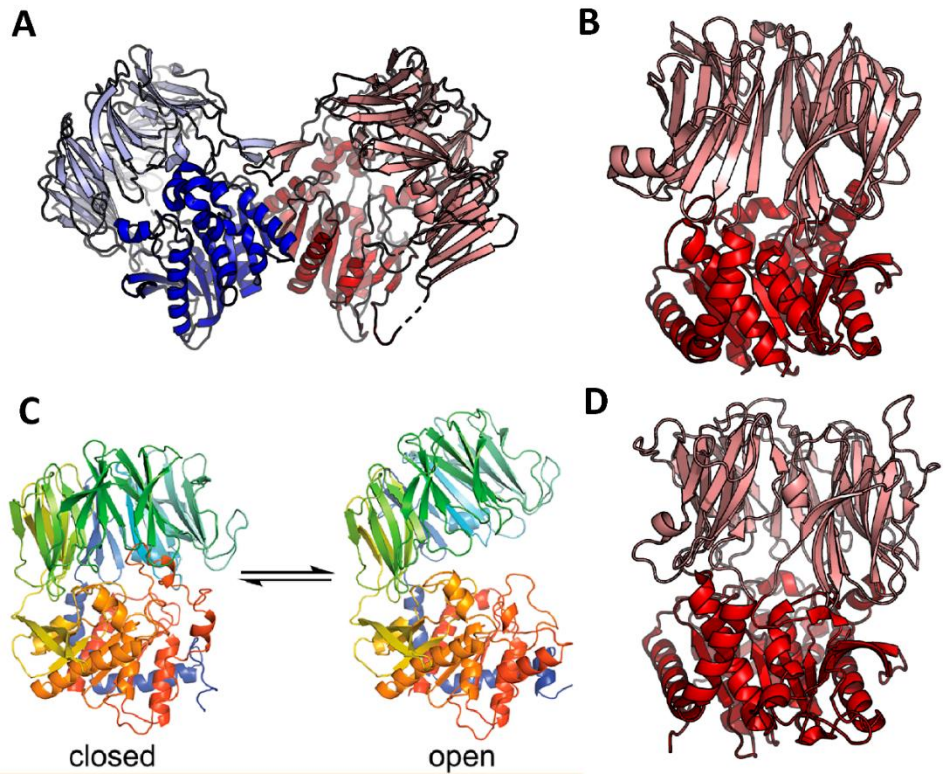


Figure 24 – The prolyl oligopeptidases of the S9A family. (A) Structure of prolyl tripeptidyl aminopeptidase from *Porphyromonas gingivalis*, with two subunits highlighted in blue and red. The catalytic domains and the β7 propeller domains are shown in dark and light colors, respectively. PDB code: 2D5L[244]. **(B)** Structure of the prolyl oligopeptidase from *P. furiosus*. The catalytic domain and the β7 propeller domain are shown in dark and light colors, respectively. PDB code: 5T88[245]. **(C)** The open/closed conformation of the prolyl oligopeptidase from *P. furiosus*. Reproduced from Ellis-Guardiola *et al.* (2019)[245]. **(D)** Structure of the oligopeptidase B from *Leishmania major*. The catalytic domain and the β7 propeller domain are shown in dark and light colors, respectively. PDB code: 2XE4[249].

and ensures dimerization via the formation of a sixteen-strand β sheet[241,242]. In *P. horikoshii* APH, the dimerization involves an adaption of the β propeller which consists of a long loop of the β propeller interacting with the catalytic domain of the dyad (see **Figure 23.C**). The trimerization of dimers relies on extensive interactions between β7 propeller domains (see **Figure 23.B**).

POP is structurally related to APH, except that only monomers and dimers have been described. The POP subunit also possesses two domains, a seven-blade β propeller and an α/β catalytic domain (see **Figure 24.A-B**)[244–246]. The movement of the two domains between an open and a closed conformations regulates the access to the catalytic site (see **Figure 24.C**)[245]. As suggested by its name, POP hydrolyzes preferentially Xaa-Pro or Pro-Xaa bonds[247]. Albeit being homologous to POP, the oligopeptidase B has a different substrate specificity[247]. The oligopeptidase B recognize only basic residues at the P1 position[248]. Its structure is closely related to monomeric POP (see **Figure 24.D**)[249]. Of note, no homolog of oligopeptidase B is found in archaea[47].

A.4.4 Tricorn peptidases

Tricorn peptidases (TRI) are another example of activity nanocompartmentalization through oligomerization in prokaryotes. They are ATP-independent carboxypeptidases belonging to the serine protease S41 family, according to MEROPS classification[61]. The TRI of *Thermoplasma acidophilum* (TaTRI) has extensively been studied and has been proposed to act downstream the proteasome[50]. A Tricorn peptidase hydrolyzes peptides of up to 30 amino acids in length into shorter ones of 2-4 amino acids[250]. The shorter peptides are further processed by three interacting auxiliary aminopeptidases, F1, F2, and F3[250–252]. TRI recognizes preferentially aliphatic, aromatic, and basic residues in the S1 pocket, although being able to completely degrade insulin B[250,252,253]. The most striking feature of TRI is unquestionably their quaternary structure. TaTRI is a 720 kDa-hexamer adopting a hexagon-like shaped toroid consisting of two perfectly staggered trimeric rings (see **Figure 25.A-B**)[250,254]. The toroid has a diameter of about 160 Å and a thickness of 88 Å. The central pore is conical shaped with an external diameter of 45 Å and an internal diameter of 20 Å. Despite the three-fold symmetry, TaTRI hexamer is an assembly of three dimers having an extensive interaction surface involving several domains[254,255]. The peculiar quaternary structure has given the name to TRI, referring to a tricorn hat. A higher oligomeric form has also been reported for TaTRI *in vivo* with twenty hexamers assembled into a virus-like icosahedron (see **Figure 25.C**)[250,256,257].

A TRI subunit consists of five domains: a β6 propeller domain, a β7 propeller domain, two core domains C1 and C2 forming the catalytic domain, and a PDZ domain between C1 and C2 (see **Figure 26.A**)[254,255]. In the hexamer structure, the six β7 propeller domains are oriented inwards the central pores while the β6 propeller domains are exposed at the toroid external surface (see **Figure 26.B**). The β7 propeller domains partly ensure the oligomerization by interacting with neighbor PDZ domains. They also control the access of substrate to the catalytic site[255,258]. The β6 propeller domains are involved in the egress of hydrolysis products[255,258] and probably dock the auxiliary aminopeptidases F1, F2, and F3[214,259]. In addition to its role in the TRI oligomerization, the PDZ domain could also play an important role in substrate recognition[255,258]. The catalytic serine residue is localized in the C2 domain while the histidine residue of the catalytic triad is within the C1 domain[254]. An aspartate residue brought by the C2 domain of the neighboring dyad subunit has also been described to be important for the specificity of the S1 pocket[254,255,258].

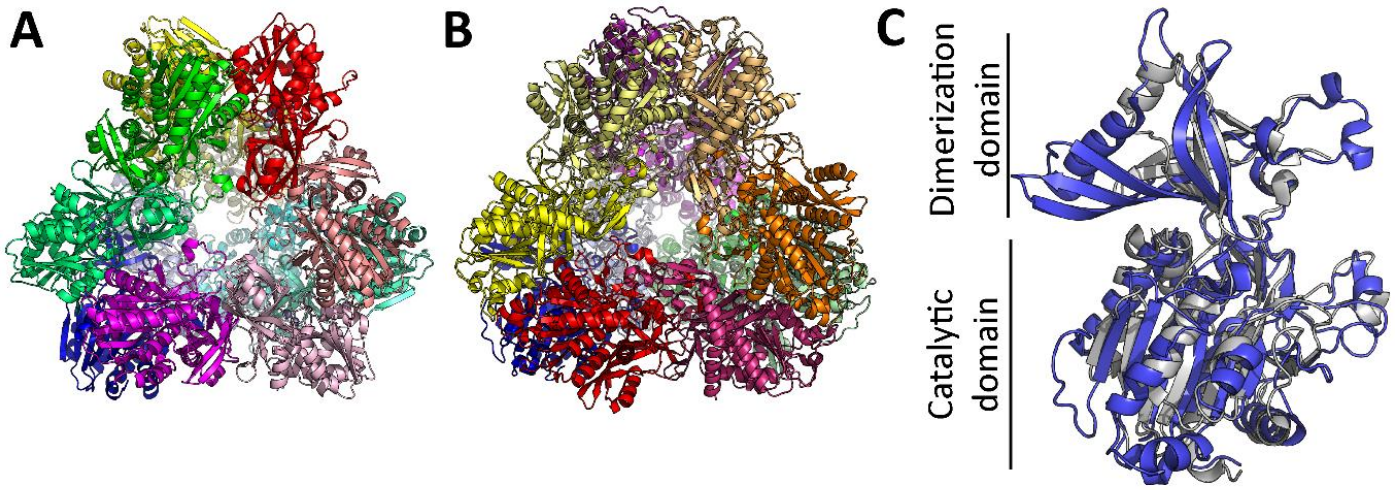


Figure 27 – Structure of TET aminopeptidases. (A) Structure of the M42 aminopeptidase PhTET2 of *Pyrococcus horikoshii* (PDB code 1XFO). **(B)** Structure of the M18 aminopeptidase PaAP of *P. aeruginosa* (PDB code 4NJR). **(C)** Structural alignment of PhTET2 subunit (white) vs. PaAP subunit (blue).

The aminopeptidases F1, F2, and F3 achieve the complete peptide processing downstream TRI, thanks to their complementary substrate specificities. Indeed, the aminopeptidase F1 is a prolyl-oligopeptidase belonging to the serine protease S33 family[251,252,259,260]. The aminopeptidase F2 is a tyrosyl-/arginyl-aminopeptidase while the aminopeptidase F3 has a clear preference towards acidic residue[214,252]. Both aminopeptidases F2 and F3 are related and belong to the metalloprotease M1 family[214]. Albeit their original quaternary structure, TRI are not widespread among prokaryotes. In Bacteria, TRI are found in Vibrio, Bacteriodes, Prevotella, and Streptomyces genera while, in Archaea, TRI occur in Sulfolobales, Thermoplasmatales, and Thermoproteales[261]. The function of TRI could be fulfilled by TET aminopeptidases in prokaryotes lacking any TRI coding gene in their genome[184,262].

A.4.5 TET aminopeptidases

Although being evolutionary distinct, the M18 and M42 families are closely related since their members are structurally related[61]. Indeed, their quaternary structure consists of twelve subunits forming a tetrahedron-shaped particle (see **Figure 27.A-B**). The M18 and M42 aminopeptidases have been nicknamed "TET aminopeptidases" after such a peculiar trait[262–264]. The subunit structure contains two domains: a catalytic domain and a dimerization domain (see **Figure 27.C**)[184,265–267]. The catalytic domain structure is conserved between the M18 and M42 families. Their dimerization domains, however, adopt two different folds: the PDZ-like domain of M42 aminopeptidases and the butterfly-fold domain of M18 aminopeptidases. Of note, the latter is still related to the former as the butterfly-fold resembles to an extended PDZ-like domain[266,267]. The major difference between the M18 and M42 family is their phylogenic distribution (see **section A.5.1.1**). While the M18 family is widely distributed in all kingdoms of Life, the M42 family is ubiquitous to prokaryotes.

The M42 aminopeptidases have been proposed to work in concert with proteolytic complexes in prokaryotes[184,262,263,268]. Since this PhD thesis focuses on M42 aminopeptidases, they will be extensively described in the next introductory section (see **section A.5**). The knowledge about prokaryotic M18 aminopeptidases is extremely scarce, the aspartyl-aminopeptidase of *Pseudomonas aeruginosa* being the unique characterized member[266]. Eukaryotic M18 aminopeptidases have been better studied, especially a highly conserved aspartyl-aminopeptidase[269–271]. In mammals, the M18 aspartyl-aminopeptidase DNPEP, is particularly abundant in the brain, neuroendocrine tissues, and neural retina[269,272,273]. DNPEP has been reported to convert angiotensin II to angiotensin III, potentially playing an important role in the renin angiotensin system that regulates blood pressure[269]. It has also been shown to interact with a chloride transporter, improve albumin uptake, and stabilize the cytoskeleton[274]. Furthermore, the gene coding DNPEP is overexpressed in brain tumors, breast cancer, and colorectal cancer[275–277], but its role in tumoral tissues is not yet understood. The structures of *Homo sapiens* and *Bos taurus* DNPEP have been solved, showing the tetrahedral architecture of the twelve subunits like M42 aminopeptidases[267,273]. In the human parasite *Plasmodium falciparum*, DNPEP plays an important role in hemoglobin catabolism and binds to several erythrocyte membrane proteins[278–280]. As expected, the *P. falciparum* DNPEP adopts the same quaternary structure as its mammalian counterparts[281]. Of note, the knockout of DNPEP is lethal in *P. falciparum*[280] while it reduces the virulence of *Toxoplasma gondii*[282].

In addition to DNPEP, fungi possess a second M18 aminopeptidase. They have been thoroughly characterized in *Saccharomyces cerevisiae*, especially the leucyl-aminopeptidase Ape1p. This enzyme is transported as an inactive pre-protein from the cytosol to the vacuole independently

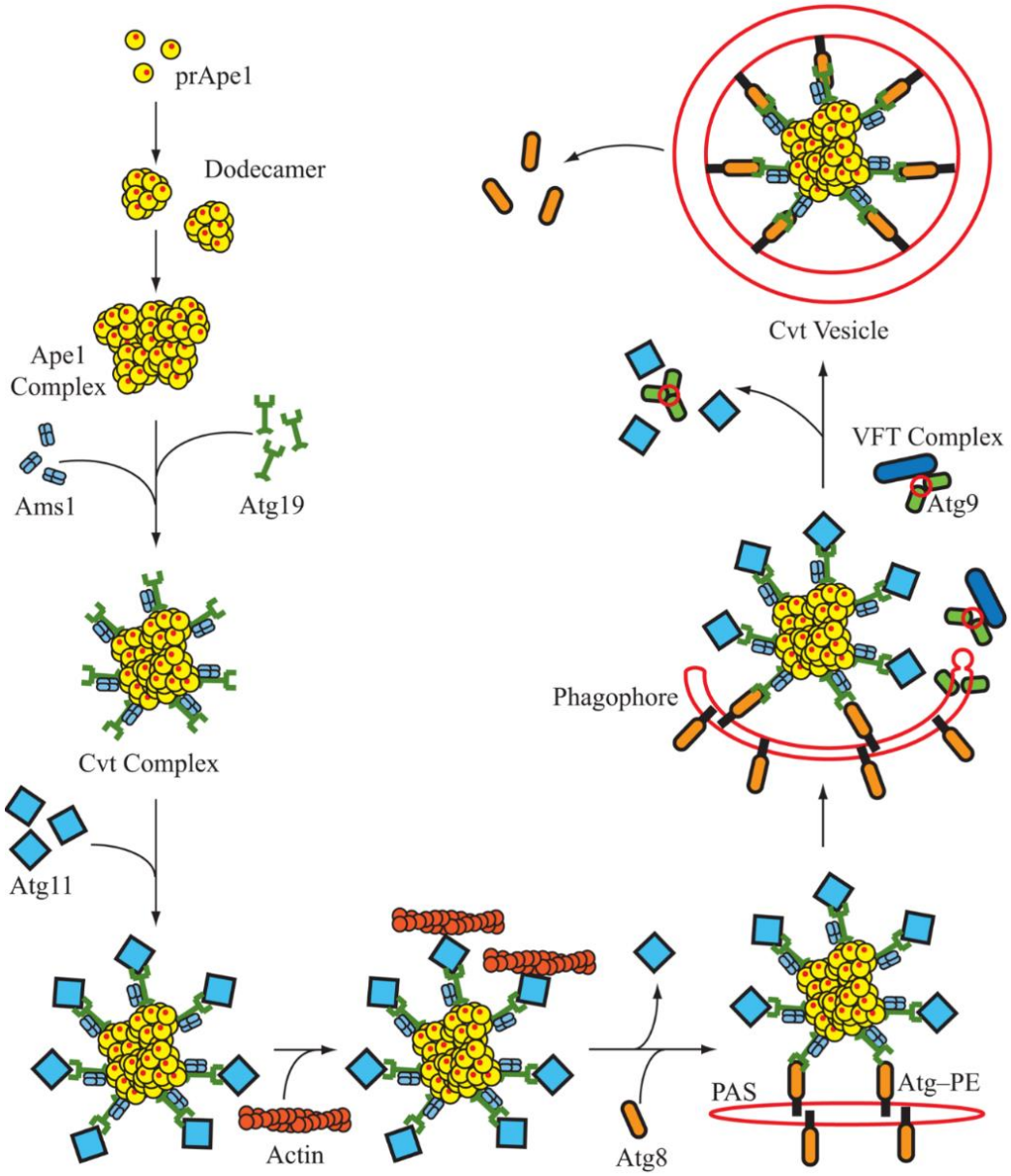


Figure 28 – The vacuolar transport of Ape1 via the Cvt pathway. Pre-Ape1p oligomerizes into inactive dodecamers in the cytosol. Atg19 recognizes the signal peptide of Pre-Ape1p and promotes its aggregation into Cvt complexes. The recruitment of Atg11 and Atg8 are required to stabilize the Cvt complexes and to form lipid bilayer vesicles. Reproduced from Lynch-Day and Klionsky (2010)[285].

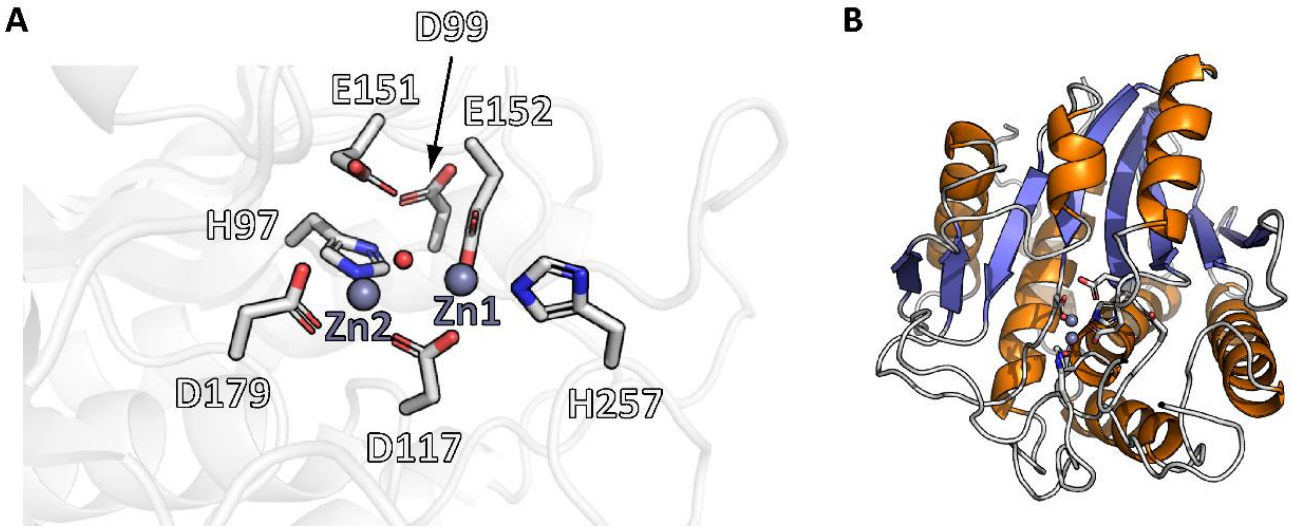


Figure 29 – The structure of the aminopeptidase Ap1 of *V. proteolyticus*. (A) Close-up view of the catalytic site with the seven conserved residues found in all MH clan members highlighted as sticks. The two Zn^{2+} are represented as grey spheres and annotated as Zn1 and Zn2. The water molecule involved in peptide bond hydrolysis is represented as a red sphere. **(B)** Schematic representation of VpAp1 structure with α helices colored in orange and β strands in blue. PDB code: 1AMP[301].

of the Golgi secretory pathway[283,284]. It relies on the Cvt pathway (cytosol-to-vacuole transport) which consists in the formation of phagophores using the same machinery than macroautophagy (see **Figure 28**)[285,286]. Atg19p recognizes the signal peptide of Ape1p triggering the recruitment of Atg11p and Atg8p, both required to stabilize the lipid bilayer vesicle[285–290]. It is worth noting that Ape1p has a scaffold function in the Cvt vesicles which are not formed in its absence[287]. In addition, the Cvt vesicles are smaller and more uniform (about 150 nm of diameter) than autophagosomes (about 300-900 nm)[291,292]. Ape1p gains its function when it is delivered into the vacuole and its signal peptide is cleaved by the Prb1p[293]. The structure of Ape1p has revealed the importance of the signal peptide in the controlled aggregation of Ape1p-Atg19p complex[294]. Two other hydrolases, the α-mannosidase Ams1p and the DNPEP homolog Ape4p are known to hitchhike Cvt vesicles by interacting with Ape1p[295,296].

A.5 M42 aminopeptidases

This introductory section is a literature review about the M42 aminopeptidases, the topic of this PhD thesis. Their biochemical characteristics (see **section A.5.2**), structural peculiarities (see **section A.5.3**), and putative function (see **section A.5.4**) will be covered thoroughly in the next pages. Before getting into the heart of the matter, the MH clan will be introduced to show the diversity of its members. The phylogeny of the M42 family will also be discussed with a cross reference to the Tricorn peptidases (see **section A.4.4**) and the M18 aminopeptidases (see **section A.4.5**). Their occurrence in the three kingdoms of Life will also be discussed in detail along with some genetic aspects which will shed light on their diversity.

A.5.1 M42 family

A.5.1.1 MH clan

The MH clan regroups four families of proteolytic enzymes, namely M18, M20, M28, and M42. Its members exhibit a panel of activities: aminopeptidase, carboxypeptidase, dipeptidase, and tripeptidase[297]. In addition, some of them are moonlighting enzymes having non-peptidolytic activities like DapE, an aspartyl aminopeptidase acting also as an N-succinyl-L,L-diaminopimelate desuccinylase[298,299]. Despite being evolutionary distinct based on their sequence homology, the four families of the MH clan have a shared sequence motif with seven highly conserved residues[300]. These residues are localized in the active site, as observed in the structure of the aminopeptidase Ap1 of *Vibrio proteolyticus* (VpAp1). In VpAp1, the seven conserved residues are His-97, Asp-99, Asp-117, Glu-151, Glu-152, Asp-179, and His-256 (see **Figure 29.A**). Five of them are involved in the coordination of two metal ions. The first metal ion binding site is composed of Asp-117, Glu-152, and His-256 while the second is made of His-97, Asp-117, and Asp-179[301]. Asp-117 is shared between both binding site as it bridges both metal ions. Glu-151 is the catalytic base required for deprotonating the water molecule prior to hydrolysis[302,303]. Finally, while not being catalytically involved, Asp-99 may play an important role as it forms a kind of catalytic triad with His-97 and the second metal ion[304–307]. It has been proposed that Asp-99 could reduce the Lewis acidity of the metal ion via its interaction with His-97. As a result, the two metal ions could have a different Lewis acid strength, explaining their distinct role in the catalytic mechanism (see **section A.5.2.2**).

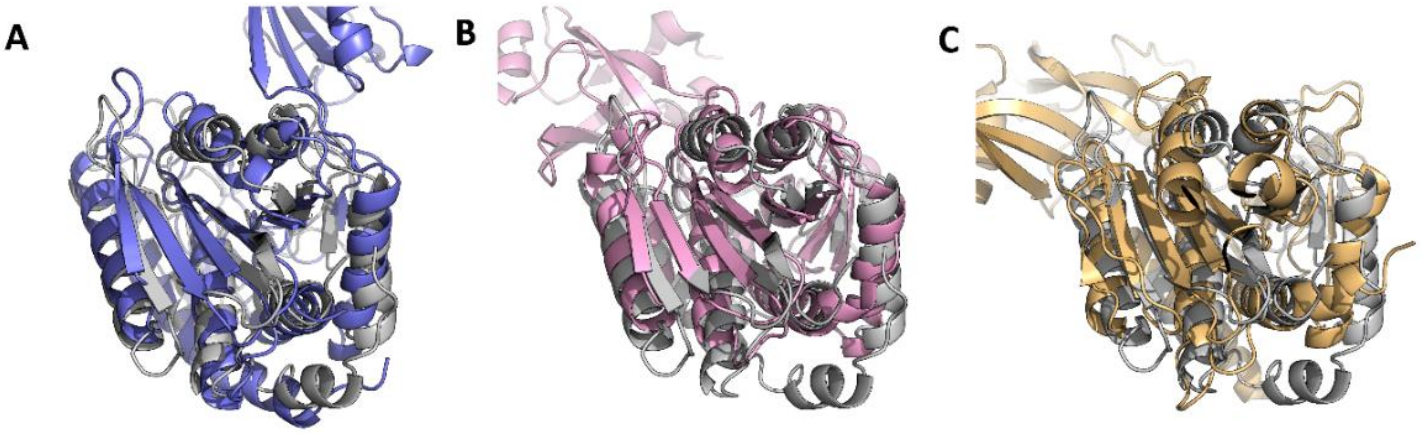


Figure 30 – The catalytic domain fold conserved in the MH clan. (A) Structural alignment of VpAp1 (in gray, PDB code: 1AMP) vs. the carboxypeptidase G2 from *P. aeruginosa* (blue, PDB code: 1CG2). **(B)** Structural alignment of VpAp1 vs. PhTET2 from *P. horikoshii* (pink, PDB code: 1Y0Y). **(C)** Structural alignment of VpAp1 vs. hDNPEP from *H. sapiens* (gold, PDB code: 4DYO). The carboxypeptidase G2, PhTET2, and DNPEP belong to the M20, M42, and M18 families, respectively.

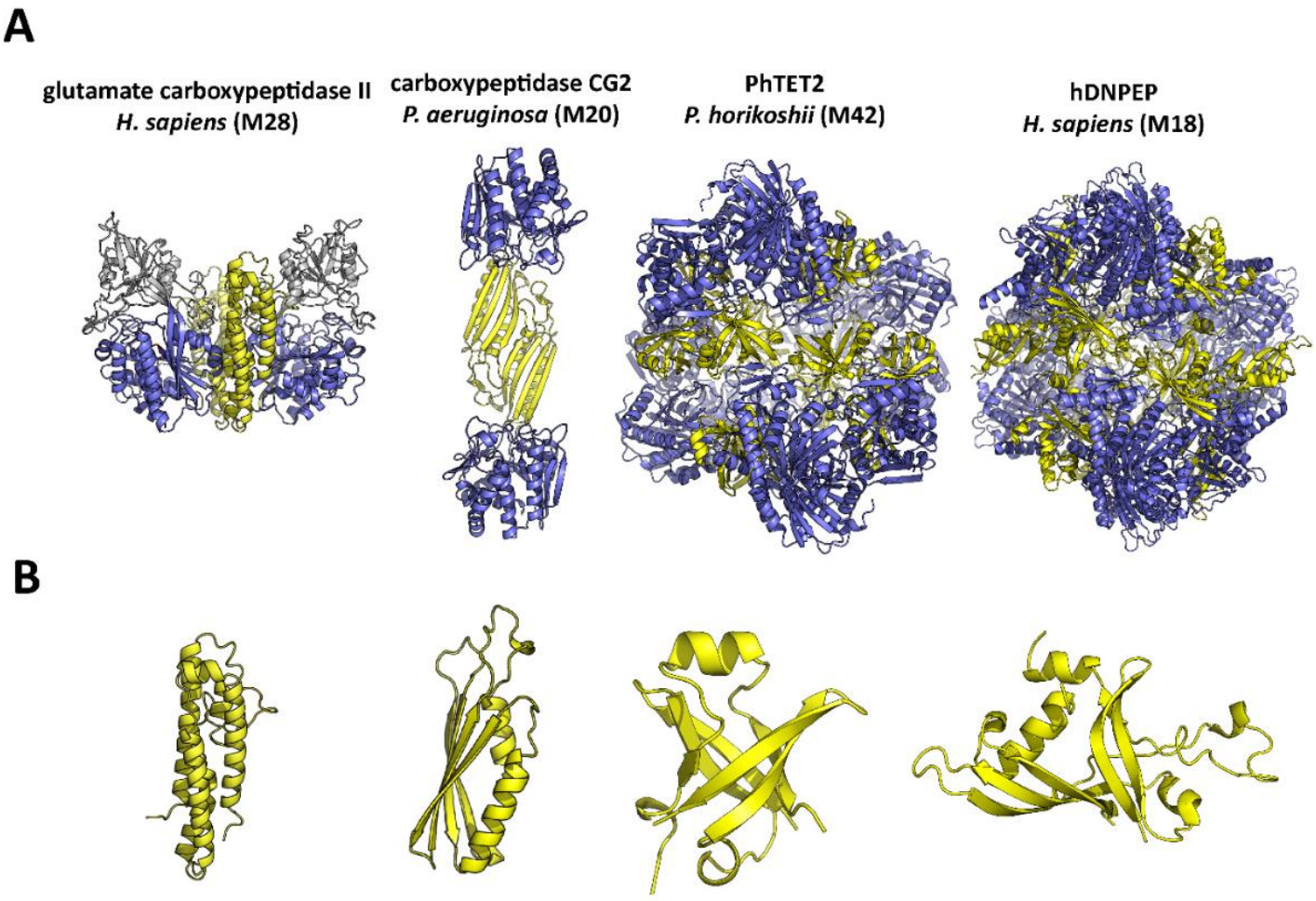


Figure 31 – The oligomerization states of MH clan enzymes. (A) The quaternary structures of the glutamate carboxypeptidase II from *H. sapiens* (PDB code 1Z8L), the carboxypeptidase G2 from *P. aeruginosa* (PDB code 1CG2), PhTET2 from *P. horikoshii* (PDB code 1Y0Y), and hDNPEP from *H. sapiens* (PDB code 4DYO) are shown as examples to illustrate the oligomerization states found in the M28C, M20, M42, and M18 families, respectively. The catalytic domains are colored in blue, the dimerization domains in yellow. The glutamate carboxypeptidase II possesses a third domain (in grey) having a putative receptor function[310]. **(B)** The dimerization domain folds of the enzymes illustrated in the upper panel.

VpAp1 has been thoroughly studied biochemically and structurally[308]. As a result, it has been proposed to be an MH clan type enzyme. It belongs to the M28 family and displays a leucyl-aminopeptidase activity. VpAp1 is a monomeric enzyme secreted by *V. proteolyticus*, probably to degrade peptides for nutrients. Its structure adopts a common α/β fold with a central β sheet made of eight β strands surrounded by eight α helices (see **Figure 29.B**). Surprisingly, all MH clan members possess a catalytic domain with the same fold as VpAp1 even though the four families are evolutionary distinct (see **Figure 30**). The four families, however, differ greatly on their oligomerization state and how their quaternary structure is achieved. Monomeric enzymes are reported in the M28A and M28E subfamilies, like the aminopeptidase S from *Streptomyces griseus* and VpAp1[301,309]. The M28C subfamily seems to include dimeric enzymes, like the glutamate carboxypeptidases II and III from *Homo sapiens*[310,311]. In these two enzymes, the dimerization domain consists of four α helices juxtaposed to the catalytic domain so that the subunits interact back to back (see **Figure 31.A**). In the M18, M20, and M42 families, the dimerization domains are totally different from that of the M28C subfamily as they are clearly separated from the catalytic domain (see **Figure 31.A**). The M18, M20, and M42 peptidases, however, have dimerization domains adopting different folds (see **Figure 31.B**).

The dimerization domain of the M20 peptidases consists of a β sheet made of four β strands flanked by two α helices. The interaction between two subunits relies on hydrophobic interactions between two helices and the formation of a continuous β sheet between both dimerization domain (see **Figure 31.A**). The vast majority of M20 peptidases forms dimers, such as the carboxypeptidase G2 from *Pseudomonas* sp., PepT from *Salmonella typhimurium*, the carnosinase CN2 from *Mus musculus*, the L-carbamoylase from *Geobacillus stearothermophilus*, and the bacterial DapE[299,312–315]. HmrA from *Staphylococcus aureus*, belonging to the M20 family, has been reported as tetramer where two dimers interact via the eight α helices of their dimerization domains[316]. As explained in **section A.4.5**, the M18 and M42 aminopeptidases are dodecameric enzymes adopting a tetrahedral-shaped structure. They possess a dimerization domain having two distinct, but related, folds (see **Figure 31.B**): a butterfly fold for the M18 family and a PDZ-like fold for the M42 family[184,267]. In both cases, dimers self-assemble into dodecamers or even in tetracosamers (24 subunits). The structures of the M42 aminopeptidases will be described in **section A.5.3**.

A.5.1.2 M42 family phylogeny

The M42 family has the singularity to be ubiquitous to prokaryotes while the M18, M20, and M28 families are distributed in all kingdoms of Life. According to Pfam database (accession number of the M42 family: PF05343), M42 aminopeptidases are found in 2,242 out of 6,948 reference bacterial proteomes[317]. They occur mainly in Firmicutes, Proteobacteria, Bacteriodetes, Actinobacteria, Deinococcus-Thermus, Cyanobacteria, Thermotogae, Planctomycetes, and Chloroflexi phyla. In Archaea, M42 aminopeptidases are present in 238 out of 295 reference proteomes, i.e. all archaeal phyla except Methanomicrobia and Thermoplasmata classes. The phylogeny of the M42 family does not show a clear distinction between bacterial and archaeal sequences (see **Figure 32**). A phylogenetic branch of some archaeal enzymes clearly diverges while the others are related to bacterial enzymes. Two distinct lineages, however, have been proposed based on whether the origin is coming from Archaea or Bacteria[318]. In addition, lateral gene transfer might have occurred between Bacteria and Archaea species during the evolution. The number of genes coding M42

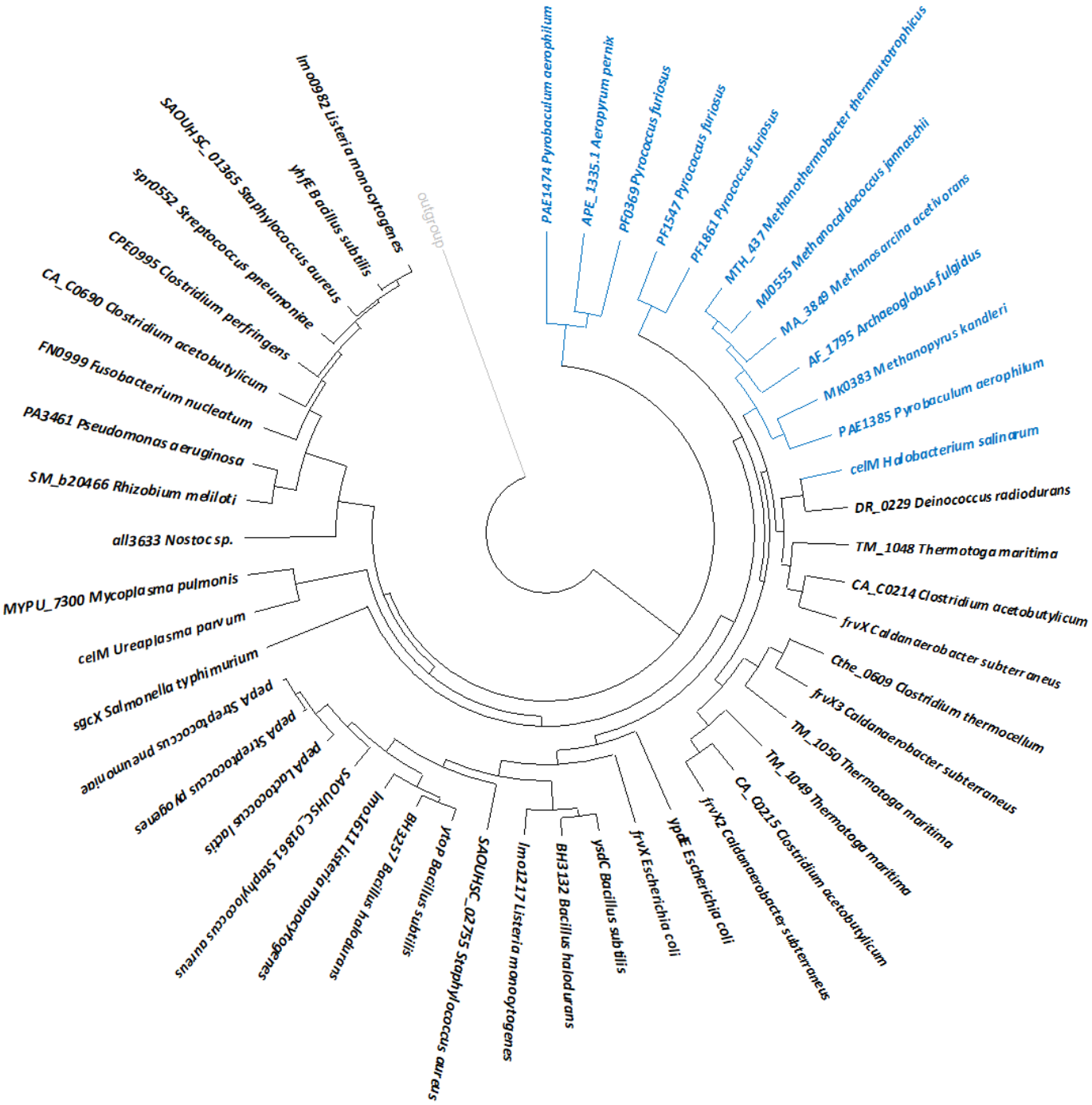


Figure 32 – A reduced phylogenetic tree of the M42 aminopeptidases. The phylogenetic tree is based on 48 sequences used to define the consensus sequence of the Pfam PF05343 family. It has been built by using the Maximum Likelihood method and RelTime method run in MEGA X[320–322]. The archaeal sequences are highlighted in blue.

aminopeptidases may also vary from one to four copies. For instance, in Archaea, *Haloarcula marismortui* has only one gene coding an M42 aminopeptidase[262] while *Pyrococcocus horikoshii* has four genes[318,319]. The same multiplicity is also observed in Bacteria, like in *Escherichia coli* and *Bacillus subtilis* having three copies of gene coding M42 aminopeptidase located in different operons. Moreover, in some bacteria, the different copies can be organized in a single operon, like in *Thermotoga maritima*. The copy number of genes coding M42 aminopeptidases reflects the difference of substrate specificities. This point will be discussed later in **section A.5.2**.

Due to their wide distribution among prokaryotes, the M42 family has been proposed to achieve peptide degradation downstream the proteasome and related complexes[184,262]. Tricorn proteases have been assumed to fulfil the same function in species lacking an M42 aminopeptidase coding gene. The latter hypothesis, however, seems unlikely as tricorn proteases are found only in the Streptomycetaceae family and in Thermoplasmatales and Sulfolobales orders for Bacteria and Archaea, respectively[323]. It is worth noting the distinct distribution of the M18 and M42 families. The phylogenetic distribution of the M18 family is less widespread in prokaryotes than the M42 family. Indeed, it is found mainly in Firmicutes, Proteobacteria, and Actinomycetes while being almost absent from Archaea, except in the Thermoplasmata class. The main difference between both family is that M18 aminopeptidases are widely distributed among Eukaryota, occurring in 809 out of 1,230 reference proteomes[317]. Hence, taking the M18 and M42 families together makes the TET peptidases good candidates to achieve peptide degradation in association with the proteasome since many organisms possess either an M18 or M42 aminopeptidase.

A.5.2 Activity of M42 aminopeptidases

A.5.2.1 Substrate specificity

The members of the M42 family are aminopeptidases, i.e. exopeptidases degrading peptides from their N-terminus and releasing one amino acid at a time. They are not processive meaning that the peptide is released each time a peptide bond is hydrolyzed[324–326]. Regarding to the substrate length, the M42 aminopeptidases could degrade peptides of up to 32 amino acids in length, as shown for the TET peptidase from *H. marismortui*[262]. The processing of long peptides, however, is slower than that of shorter substrates. Still, an efficient hydrolysis is observed for peptides of up to 12 amino acids in length[262,324]. Synthetic substrates are commonly used to determine the substrate specificity of aminopeptidases, mostly an amino acid coupled with *p*-nitroaniline or 4-methylcoumarin (the amid bond mimicking a peptide bond). Some M42 aminopeptidases, however, are inactive with such synthetic substrates but display hydrolytic activity against dipeptides[326].

The substrate specificity of an M42 aminopeptidase will depend on the S_1 pocket that accommodates the N-terminal residue of the substrate. A variety of charge distributions in the S_1 pocket have been reported for several M42 aminopeptidases. A broad substrate specificity has been reported for two M42 aminopeptidases, the TET peptidase from *H. marismortui* and YpdE from *E. coli*[262,327]. The broad substrate specificity has been linked to the gene copy number as *H. marismortuis* has only one gene coding a M42 aminopeptidase[263]. This hypothesis, however, cannot be generalized since *E. coli* possesses three different M42 aminopeptidases, including YpdE. That topic will be further debated in **section B.5** (page 99) where the substrate specificity of the three *E. coli* enzymes is described. The other characterized M42 aminopeptidases display a much narrower substrate

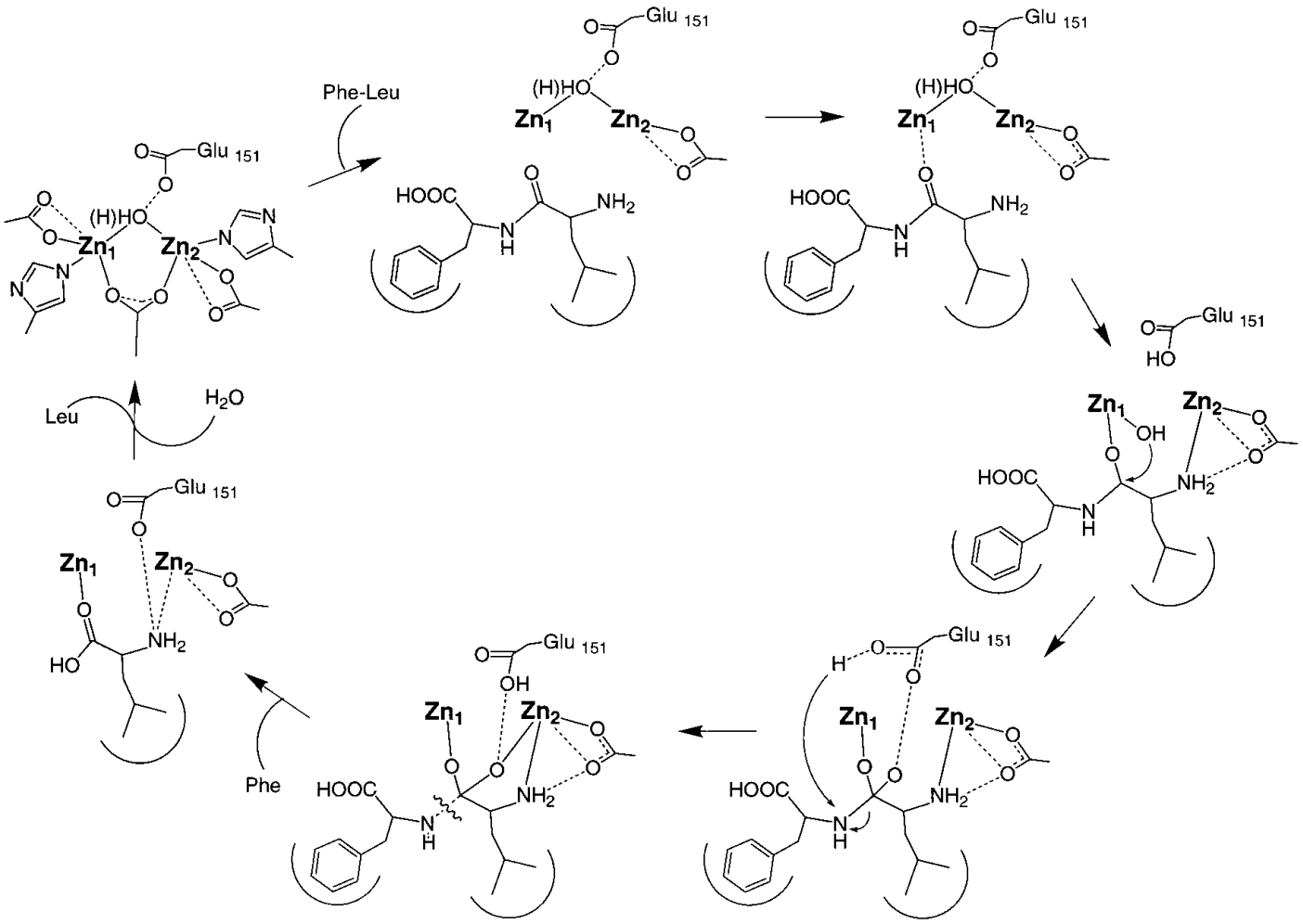


Figure 33 – Schematic representation of the catalytic mechanism of VpAp1. The ligands of the two zinc ions are only shown in the first panel (left corner). Reproduced from Stamper *et al.* (2001)[307].

specificity than the TET peptidase of *H. marismortui*. Some M42 aminopeptidases have a glutamyl-aminopeptidase activity, degrading preferentially peptides with a glutamate or an aspartate residue at the P_1 position[328–331]. The lysyl-aminopeptidase activity has also been reported for M42 aminopeptidases, with the degradation of peptides with a lysine or an arginine residue at the P_1 position[325,332,333]. Other M42 aminopeptidases display a leucyl-aminopeptidase activity, cleaving peptides with a leucine, methionine, valine, or alanine at the P_1 position[323,324,334–338]. Recently, a glycyl-aminopeptidase activity has been reported for PhTET4 from *P. horikoshii*[319]. This finding is quite interesting as *P. horikoshii* possesses four M42 aminpeptidases having each a different substrate specificity[319,324,325,329,338]. The combination of these enzymes could virtually degrade any peptide, except the Xaa-Pro bond requiring an Xaa-Pro peptidase activity (not yet discovered in the M42 family). This point will be further discussed in **section A.5.4.2** about the existence of heterocomplexes. It is worth noting that, despite their marked substrate preferences, many M42 aminopeptidases can still degrade peptide with an unfavored P_1 residue. For instance, PhTET1, which is a glutamyl-aminopeptidase, can hydrolyze an Ala-Ala peptide[326,339].

A.5.2.2 Proposed catalytic mechanism

As discussed earlier, M42 aminopeptidases are metalloenzymes having two metal ions involved in catalysis. Due to their two metal ion centers, they are also referred as dinuclear or co-catalytic enzymes. The catalytic mechanism of M42 aminopeptidases has not yet been studied extensively but it may be similar to that of VpAp1[265], due to the relatedness of their catalytic domains. It is even generally accepted that all MH clan enzymes have a similar catalytic mechanism to that of VpAp1[340]. The catalytic mechanism of VpAp1 is rather well understood due to a bountiful of studies about this enzyme[301–307,341–362]. Stamper *et al.* (2001) proposed a catalytic mechanism for VpAp1 involving two metal ions in the substrate binding and the activation of a water molecule[307]. This catalytic mechanism was extensively reviewed by Richard Holz[363] and is presented in **Figure 33**. The two bound Zn^{2+} are coordinated in a distorted tetrahedral geometry[301]. The first metal binding site (M1 site)* consists of Asp-117, Glu-152, and His-259, coordinating the Zn_1. The second site (M2 site)* is made of His-97, Asp-117, and Asp-179. The water molecule involved in the hydrolysis constitutes the fourth ligand in the tetrahedral coordination. The two zinc ions, however, are biochemically distinct and fulfil different roles[347,348]. Their roles have been defined thanks to the structures of VpAp1 obtained in complex with four inhibitors: *p*-iodo-D-phenylalanine hydroxamate, 1-butaneboronic acid, L-leucine phosphonic acid, and bestatin[303,304,307,358]. Of note, the second inhibitor mimics the Michaelian complex while the third mimics the tetrahedral transition-state intermediate.

The first step is the binding of the incoming peptide substrate via the interaction between the carbonyl oxygen of the P_1 residue and Zn_1. Concomitantly, the water molecule is deprotonated by the catalytic base, Glu-151, with the assistance of Zn_2. The environment of Zn_2 is thought to be so negatively charged that its charge is neutralized[304,307,363]. Asp-99 may play an important role as it interacts with His-97, forcing the imidazole ring to be deprotonated. As a result, due to charge repulsion, the hydroxide ion is transferred to Zn_1, which stands close to the peptide bond to be hydrolyzed. Meanwhile, Zn_2 can establish an interaction with the N-terminus of the substrate. The

* Author's note: The author draws reader's attention to the definition of the M1 and M2 sites. Both abbreviations will be used extensively hereafter.

Enzyme name	Species	Substrate specificity	Activating metal ions	Inhibiting metal ions	PDB code	M1 site	M2 site	References
PhTET1	*P. horikoshii*	Glutamyl-aminopeptidase	Co^{2+}	?	2CF4	Co^{2+} (T4)	Co^{2+} (T4)	326,329,339,366
PhTET2	*P. horikoshii*	Leucyl-aminopeptidase	Co^{2+} + Zn^{2+} > Co^{2+}	Mn^{2+}, Zn^{2+}	1XFO 1Y0Y	Zn^{2+} (T4) Zn^{2+} (TP5)	Zn^{2+} (T4) Zn^{2+} (TP5)	184,265,324,338,367
PhTET3	*P. horikoshii*	Lysyl-aminopeptidase	Co^{2+}	Zn^{2+}, Ca^{2+}, Cd^{2+}, Mg^{2+}, Mn^{2+}, Cu^{2+}	2WZN	Zn^{2+} (TP5)	Zn^{2+} (TP5)	325,367
PhTET4	*P. horikoshii*	Glycyl-aminopeptidase	Ni^{2+} > Co^{2+} > Mn^{2+}	Ca^{2+}	?	?	?	319
PfTET3	*P. furiosus*	Lysyl-aminopeptidase	Co^{2+} > Zn^{2+} + Co^{2+}	Cd^{2+}, Cu^{2+}, Ca^{2+}, Mg^{2+}, Fe^{2+}, Zn^{2+}, Mn^{2+}, Ni^{2+}	4X8I	Co^{2+} (TP5)	Zn^{2+} (O6)	333,365
PepA	*Streptococcus pneumoniae*	Glutamyl-aminopeptidase	?	?	3KL9	Zn^{2+} (TP5)	Zn^{2+} (T4)	330
ADPkam589	*D. kamchatkensis*	Leucyl-aminopeptidase	Mn^{2+} > Mg^{2+} > Zn^{2+}	Co^{2+}, Ca^{2+}, Cu^{2+}	4WWV	none	none	337,378
Pep1079, Pep1080, Pep1081	*Symbiobacterium thermophilum*	Lysyl-aminopeptidase	Co^{2+}	?	?	?	?	332
MHJ_0125	*Mycoplasma hyopneumoniae*	Glutamyl-aminopeptidase	Co^{2+}	Mn^{2+}, Cu^{2+}	?	?	?	331
YpdE	*E. coli*	broad	Co^{2+} > Ni^{2+} > Mn^{2+} > Cu^{2+}	Fe^{2+}, Ca^{2+}, Zn^{2+}, Mg^{2+}	?	?	?	327
TNA1_DAP1	*Thermococcus onnurineus*	Leucyl-aminopeptidase	Mn^{2+} > Co^{2+} = Ba^{2+} = Mg^{2+}	Cu^{2+}, Ni^{2+}, Zn^{2+}	?	?	?	335,336
TNA1_DAP2	*T. onnurineus*	Leucyl-aminopeptidase	Ba^{2+} = Co^{2+} = Mg^{2+}	Cu^{2+}, Zn^{2+}	?	?	?	336
Aminopeptidase A	*L. lactis*	Glutamyl-aminopeptidase	Co^{2+}, Zn^{2+}	Mn^{2+}, Ca^{2+}, Cu^{2+}	?	?	?	328,370
Aminopeptidase I	*Geobacillus stearothermophilus*	broad	Co^{2+}	Zn^{2+}, Mn^{2+}, Cd^{2+}, Ca^{2+}, Cu^{2+}, Mg^{2+}, Ni^{2+}, Hg^{2+}	?	?	?	369,371,372
YsdC	*B. subtilis*	Leucyl-aminopeptidase			1VHE	none	Zn^{2+} (O6)	none
FrvX	*S. flexneri*	?	?	?	1YLO	Zn^{2+} (TP5)	Zn^{2+} (T4)	none
TET	*Thaumarchaeota* archaeon	?	?	?	5DS0	Co^{2+} (TP5)	Co^{2+} (TP5)	none

Table 1 – The metal ion cofactors of M42 aminopeptidases. Are listed the M42 aminopeptidases which activity and structure have been characterized. The coordination geometry is given for the M1 and M2 sites: tetrahedral (T4), trigonal pyramidal (TP5), and octahedral (O6).

second step is the formation of the transition-state intermediate following the nucleophilic attack of the hydroxide ion on the carbonyl. Due to the formation of a *gem*-diol (*i.e.* two alcohol functions on the same carbon), the transition-state intermediate adopts a tetrahedral geometry. This intermediate is stabilized via its interactions with Zn_1, Zn_2, and Glu-151. Both Zn^{2+} are hexacoordinated with a distorted octahedral geometry[307,358,362]. The third step is the cleavage of the scissile C-N bond and the release of reaction products. During this step, Glu-151 acts as a proton donor, ceding a proton to the penultimate nitrogen of the amine bond. The tetrahedral coordination geometry of both metal ions is restored as a water molecule replaces the products being released.

Having two bound metal ions is not a strict prerequisite for the activity of VpAp1. Indeed, VpAp1 keeps about 80 % of its activity with only Zn_1 bound[304,344,347]. A similar observation has been reported for the methionyl aminopeptidase from *E. coli*, an unrelated dinuclear enzyme from the M24 family[364]. Regarding the M42 family, the catalytic mechanism of M42 aminopeptidases could differ from that of VpAp1, mainly on the role of the metal ions. Russo and Baumann (2004) proposed the role of Zn_1 and Zn_2 could be swapped in PhTET2 from *P. horikoshii*[265]. Indeed, in the structure of PhTET2, the M2 site has a greater metal ion binding affinity than the M1 site. Colombo *et al.* (2016) reported the same observation for PfTET3 from *P. furiosus*[365]. In that case of study, PfTET3 retained 75% of its activity after losing the metal ion from the M1 site. Such a behavior contrasts with VpAp1 for which the M1 site has the highest affinity and must be occupied for peptide hydrolysis. Nevertheless, the structure of PhTET2 in complex with amastatin suggests that the substrate binding is the same as that of VpAp1[184]. Finally, the oligomerization of M42 aminopeptidases seems to be driven by their metal ion cofactors (see **section A.5.3.3**)[365].

A.5.2.3 Metal ion cofactor

As seen in the previous section, two metal ions are involved in peptide bond hydrolysis by M42 aminopeptidases. Cobalt ions have been shown to activate many M42 aminopeptidases, suggesting cobalt could be their favored cofactor[325,327,328,331–333,335,339,365–372]. Other divalent metal ions can have various effect on these enzymes, often inhibitory (see **Table 1**). In some exceptions, another divalent metal ion is preferred to Co^{2+}, like Ni^{2+} and Mg^{2+}/Mn^{2+} activating PhTET4 and APDkam589, respectively[319,337]. In the available structures of M42 aminopeptidases, both Co^{2+} and Zn^{2+} are observed in the active sites, even in enzymes activated by Co^{2+} and inhibited by Zn^{2+}, like PhTET3[325]. This discrepancy is quite surprising but three key concepts must be kept in mind.

Firstly, the Irving-Williams series $Mn^{2+} < Fe^{2+} < Co^{2+} < Ni^{2+} < Cu^{2+} > Zn^{2+}$ dictates the stability between a metal ion and its ligands[373,374]. The most stable complex is encountered with Cu^{2+} while the less stable happens with Mn^{2+}. Depending on oxidoreduction conditions, Zn^{2+} may form more stable complex with ligands than Cu^{2+}, which is typically the case in the cytoplasm[375,376]. The Irving-Williams series, however, is modulated by (i) the availability of metal ions *in vivo*, (ii) the nature of ligands, and (iii) the coordination geometry of metal ion binding sites found in proteins. Indeed, there are more Mg^{2+} binding proteins than Zn^{2+} binding proteins because Mg^{2+} is far more abundant in cell than Zn^{2+} (mM vs. µM range)[377]. Mg^{2+} and Mn^{2+} are more likely bound by carboxylate functions while Cu^{2+} favors interactions with amine groups, such as the imidazole ring. The other divalent metal ions are “borderline” as they interact with both carboxylate and amine groups. The coordination geometry also influences the binding of metal ions. For instance, Cu^{2+} prefers a square planar while Zn^{2+} a tetrahedral geometry (see **Figure 34**). Even though Co^{2+} adopts preferentially an octahedral

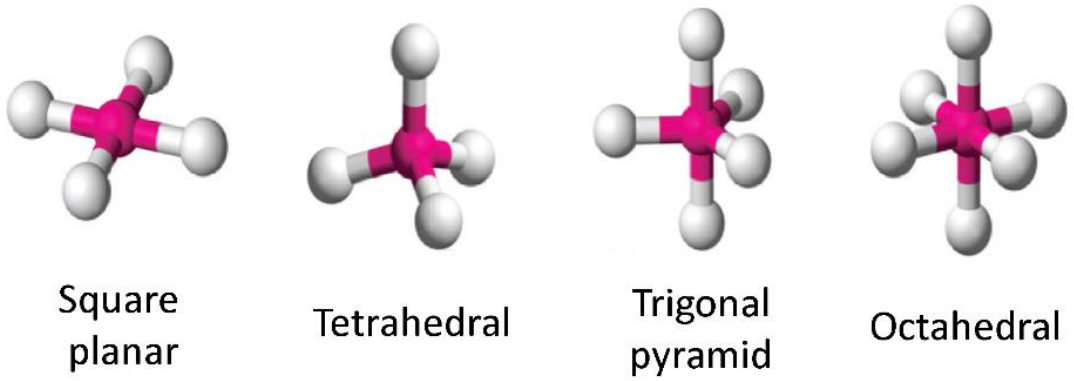


Figure 34 – Some common coordination geometries of divalent metal ions. Adapted from Yao *et al.* (2015)[379].

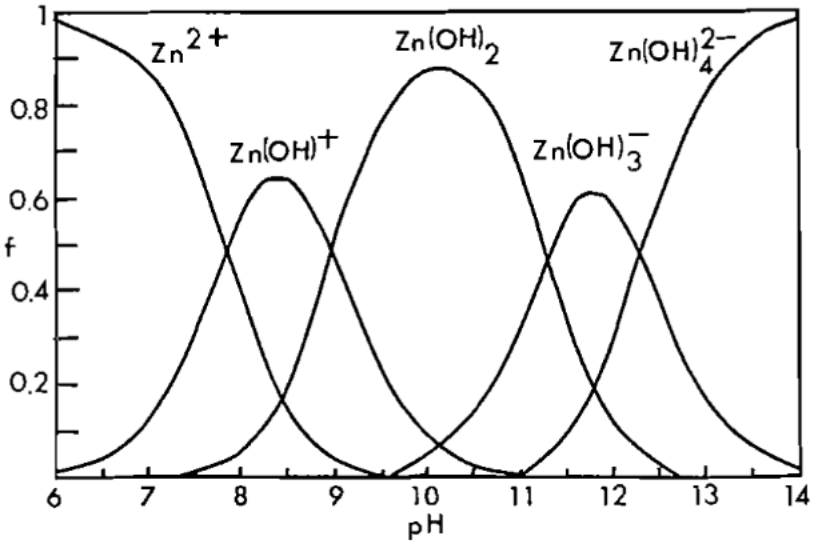


Figure 35 – Fraction *f* of Zn(II) existing in solution as Zn^{2+}, $Zn(OH)^{+}$, $Zn(OH)_2$, $Zn(OH)_3^{-}$, and $Zn(OH)_4^{-}$ vs. pH. Reproduced from Reichle *et al.* (1975)[383].

Enzyme form		L-Ala-*p*NA			L-Val-*p*NA			L-Leu-*p*NA		
		k_{cat}	K_m	$k_{cat}/K_m \times 10^{-4}$	k_{cat}	K_m	$k_{cat}/K_m \times 10^{-4}$	k_{cat}	K_m	$k_{cat}/K_m \times 10^{-4}$
Site 1	Site 2	(min^{-1})	(μM)	(M^{-1} min^{-1})	(min^{-1})	(μM)	(M^{-1} min^{-1})	(min^{-1})	(μM)	(M^{-1} min^{-1})
Zn^{2+}	Zn^{2+}	17.3	1000	1.7	10.6	25	42.4	4720	23	20520
Zn^{2+}	Co^{2+}	21.2	1250	1.7	10.5	22	47.7	2166	14	15470
Zn^{2+}	Ni^{2+}	27.0	286	9.4	14.6	20	73.0	4200	23	18260
Zn^{2+}	Cu^{2+}	14.2	740	1.9	12.4	29	42.8	3270	13	25150
Co^{2+}	Co^{2+}	44.8	435	10.3	7.3	11	66.4	1360	4.0	34000
Co^{2+}	Zn^{2+}	52.2	870	6.0	9.5	7.0	135	1990	8.0	24880
Co^{2+}	Ni^{2+}	85.6	333	25.7	9.8	11	89.1	942	5.0	18840
Co^{2+}	Cu^{2+}	41.3	80	51.6	5.9	1.0	590	706	4.0	17650
Ni^{2+}	Ni^{2+}	53.6	137	39.1	5.3	0.9	589	784	10	7840
Ni^{2+}	Zn^{2+}	312.0	213	146.0	14.8	0.4	3700	248	16	1550
Ni^{2+}	Co^{2+}	93.3	122	76.5	8.1	0.9	900	713	6.0	11880
Ni^{2+}	Cu^{2+}	14.4	114	12.6	4.1	0.8	513	271	13	2090
Cu^{2+}	Cu^{2+}	5.4	36	15.0	2.6	1.0	260	240	13	1850
Cu^{2+}	Zn^{2+}	61.5	25	246.0	10.5	3.0	350	557	15	3710
Cu^{2+}	Co^{2+}	27.6	39	70.8	3.9	1.0	390	447	6.0	7450
Cu^{2+}	Ni^{2+}	15.3	64	23.9	3.9	0.6	650	523	21	2490

Table 2 – Kinetic parameters of VpAp1 with different combinations of metal ions. Adapted from Lowther and Matthews (2002)[385] with the data from Bayliss and Prescott (1986)[346].

geometry, it can still accommodate to the tetrahedral constraints. Based on the structures of M42 aminopeptidases, both Co^{2+} and Zn^{2+} are found in tetrahedral, trigonal pyramidal, and octahedral geometries (see **Table 1**).

Secondly, all structures of M42 aminopeptidases have been solved using recombinant proteins produced in *E. coli*. To my opinion, overproduction probably introduces a bias in the metal ion binding. One can easily imagine that overproduced aminopeptidase molecules may titrate the cellular metal ion pool with, according to Irving-Williams series, Zn^{2+} being preferentially bound. Under native physiological conditions, the metal ion loading of M42 aminopeptidases would probably depend on metal ion availability. Indeed, cells have a limited amount of metal ions, believed to be lower than the number of metalloprotein molecules. Therefore, there is a strong competition in metal ion binding between metalloproteins, so that an enzyme can bind an unfavorable metal ion cofactor. For instance, the *E. coli* methionyl aminopeptidase binds only one Fe^{2+} under native physiological conditions while the maximal activity is achieved *in vitro* with two Co^{2+} [380]. Furthermore, the cellular metal ion pool varies from a prokaryote to another. The metallome of *E. coli* contains about 7.5 µM zinc, 4.6 µM iron, and 0.15 µM manganese. In comparison, the metallome of *P. furiosus* has about 5.4 µM zinc, 0.4 µM iron, and 0.15 µM manganese while that of *Methanococcus jannaschii* contains trace amount of zinc, 11 µM iron, and 22.5 µM manganese[381]. One may suspect that the metal ion content of a recombinant metalloprotein will reflect the host metal ion pool. Rosenbaum *et al.* (2011) measured the metal ion content of PhTET2 recombinantly produced in *E. coli* [382]. In this case of study, the enzyme had about 1.2 mol of zinc, 0.5 mol of calcium, and 0.3 mol of iron per mol of protein.

Finally, the inhibition of M42 aminopeptidases by Zn^{2+} could be an unfortunate artefact derived from *in vitro* experimental conditions. In fact, the inhibition is probably linked to the formation of zinc hydroxides at pH 6.5 and above (see **Figure 35**)[383]. Zinc hydroxides are potent inhibitors of metallopeptidases, as shown for the bovine carboxypeptidase[384]. In many studies on M42 aminopeptidases, activity assays were achieved in a pH range of 7-8, meaning that an inhibition by zinc hydroxides cannot be excluded[324,325,335,365,366,369]. Another worthy consideration is the metal ion exchangeability and its impact on activity and substrate specificity, which has barely been studied for M42 aminopeptidases. It might be interesting to address this issue because the nature of bound metal ions modulates the activity of VpAp1[344–346]. Using amino acid *p*-nitroanilide derivatives, VpAp1 displays the highest activity on L-Leu-*p*NA when both M1 and M2 sites are occupied with Co^{2+} (see **Table 2**). Unexpectedly, its catalytic efficiency is higher against L-Val-*p*NA than L-Leu-*p*NA when Ni^{2+} and Zn^{2+} occupy the M1 and M2 sites, respectively.

A.5.2.4 Does an M42 aminopeptidase moonlight?

Several studies have reported that some M42 aminopeptidases display a deblocking aminopeptidase activity, i.e. they are able to degrade N-terminally modified peptides. For instance, PhTET1 can degrade formyl-, acetyl-, and pyroglutamyl[†]-peptides, even though the hydrolysis rate is much lower than that of an unblocked peptide[339]. Indeed, PhTET1 degrades Ala-Ala-Ala and Ac-Ala-Ala-Ala with k_{cat} of 8.71 s^{-1} and 0.034 s^{-1}, respectively. Deblocking aminopeptidase activity has also been reported for PhTET2, PhTET3, PfTET3, APDkam589, TkDAP1 and TkDAP2 from *Thermococcus kodakarensis*, and

[†] Pyroglutamate = glutamate whose lateral chain has cyclized with the amine group.

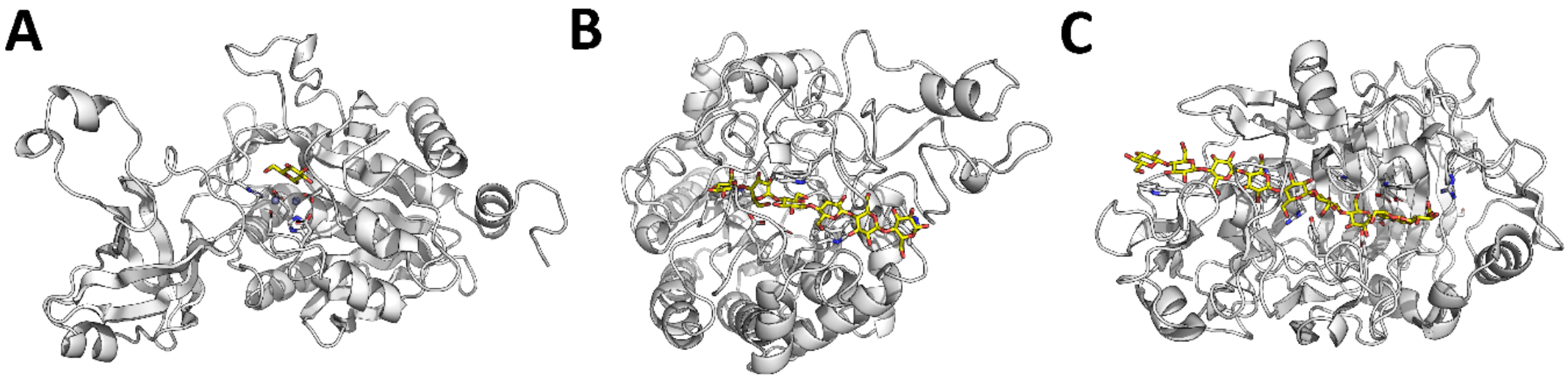


Figure 36 – Model of carbohydrate binding by the M42 aminopeptidase of *Bacillus* sp. and comparison with inverting and retaining glycoside hydrolases. (A) Model of *Bacillus* sp. build using 1VHE coordinates with a docked glucose (in yellow) in the active site[396]. **(B)** The structure of Cel6B from *Thermobifida fusca* (PDB code 4B4F), an inverting glycoside hydrolase, in complex with cellohexaose (in yellow)[399]. Key residues are shown as sticks: the two catalytic aspartate residues and several residues interacting with the substrate. **(C)** The structure of Cel7A from *Hypocrea jecorina* (PDB code 4C4C), a retaining glycoside hydrolase, in complex with cellononaose (in yellow)[398]. Key residues are shown as sticks: the two catalytic glutamate residues and several residues interacting with the substrate. Several tryptophan residues line the substrate in Cel6B and Cel7A.

TNA1_DAP1 from *T. onnurineus*[334,335,337,367,386,387]. Of note, PfTET3 has been commercialized to be used for N-terminal protein sequencing (Edman degradation). Franzetti and coworkers, however, reported that in no circumstances a deblocking aminopeptidase activity was found in PhTET1, PhTET2, and PhTET3[324–326]. Several studies have also shown a lack of deblocking activity in other M42 aminopeptidases[262,323,370,388]. Hence, whether M42 aminopeptidases have a deblocking activity or not remains controversial and argy-bargied.

M42 aminopeptidases could moonlight, i.e. achieving other activities unrelated to peptide bond hydrolysis. A putative glycoside hydrolase activity has been reported for some M42 aminopeptidases, albeit being controversial. Kobayashi *et al.* (1993) reported a new glycoside hydrolase, CelM, from *Clostridium thermocellum*[389]. The authors showed that CelM was able to degrade carboxymethylcellulose (CMC) but, weirdly, not acid-swollen cellulose or chromogenic cello-oligosaccharides. CelM contains no tryptophan, a residue commonly found in glycoside hydrolase catalytic grooves[390–392]. Another peculiarity, the amino acid sequence of CelM does not match to any glycoside hydrolase family according to CAZy database‡ [393,394] but is rather classified in the M42 family[61]. While confirming its CMCase activity, Cottrell *et al.* (2005) failed to measure any activity on L-Glu-*p*NA. Contrariwise, the aminopeptidase activity of CelM has been revealed in another study while no cellulolytic activity has been measured[323]. Indeed, CelM is a leucyl-aminopeptidase activated by Co^{2+}, just like PhTET2, hydrolyzing L-Leu-*p*NA. In this study, the authors failed to report any CMCase activity under the same experimental condition used in Kobayashi *et al.* (1993). Nonetheless, by cross-checking both studies, CelM hydrolyzes L-Leu-*p*NA at a higher rate than CMC, with specific activities of 320 and 1.3 µmol min^{-1} per µmol of enzyme, respectively[323,389].

In three other studies, CMCase activity has been reported for the M42 aminopeptidases of *Brevibacillus agri*, *Bacillus* sp., and *Geobacillus denitrificans*[395–397]. In contrast, the absence of CMCase activity has been shown for PhTET1 and ADPkam589[337,339]. Sharma *et al.* (2019) modelled the structure of two aminopeptidases of *Bacillus* sp. and docked several monosaccharides[396]. While a glucose can fit in the catalytic site, the hydrolysis of a β-1,4 glycoside bond seems to be unlikely because it would be at more than 5 Å from catalytic residues (see **Figure 36**). In addition, the model clearly denotes from both inverting and retaining glycoside hydrolases[398,399]. Maiti *et al.* (2017) reported that an M42 aminopeptidase of *B. agri* could degrade CMC at a maximal rate of 0.197 µmol min^{-1} per µmol of enzyme. Ma *et al.* (2020) showed that Cel9, an M42 aminopeptidase of *G. denitrificans*, hydrolyzed CMC at a maximal rate of 0.001 µmol min^{-1} per µmol of enzyme[397]. The substrate specificity of Cel9 is astonishing since the enzyme has been shown to degrade CMC, xylan, barley glucan, avicel, and laminarin[397]. The hydrolysis of these substrates means that Cel9 could hydrolyze β-1,3; β-1,4; and β-1,6 glycosidic bonds. Such a broad substrate specificity has only been reported in the glycoside hydrolase family 131 which members are found exclusively in fungi[400,401].

The activity reported in Maiti *et al.* (2017) and Ma *et al.* (2020) must be interpreted with caution as the reported CMCase activities are extremely low compared to genuine cellulases. In our hands, the low detection limit of CMCase assay is about 0.004 µmol min^{-1} per µmol of enzyme, following the IUPAC guidelines[323,402]. Previously, we characterized a cellulase, Ps_Cel5A, from *Pseudomonas stutzeri*, which is not considered as highly cellulolytic enzyme[403]. As a benchmark,

‡ CAZy database is a repertoire of all known glycoside hydrolase families curated by Henrissat and coworkers. Available online: http://www.cazy.org/

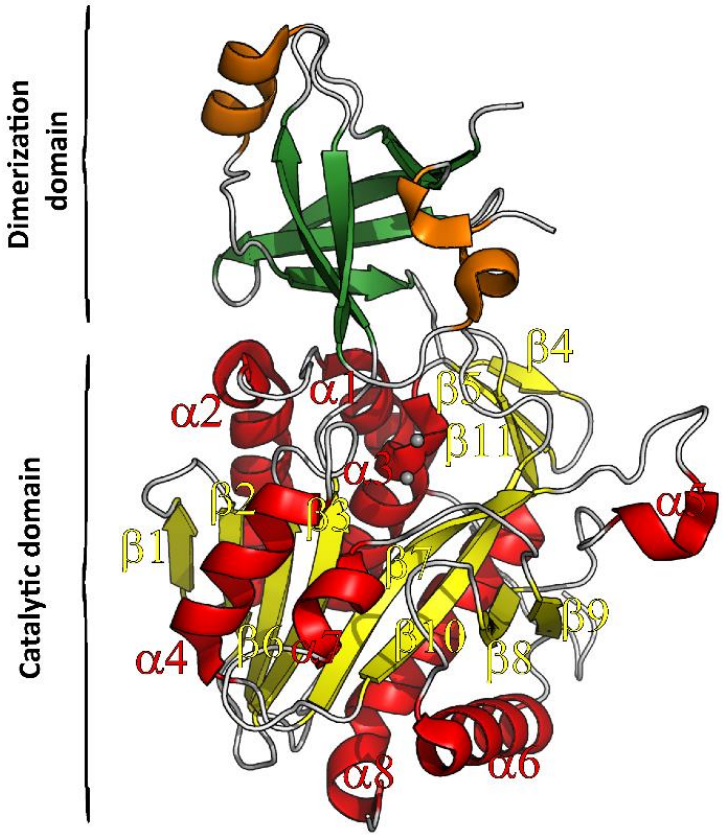


Figure 37 – Schematic representation of the PhTET2 subunit structure (PDB code: 1XFO). The α helices and β strands of the catalytic domains are colored in red and yellow, respectively. The α helices and β strands of the dimerization domains are colored in orange and green, respectively. The two Zn^{2+} found in the catalytic are represented as grey spheres.

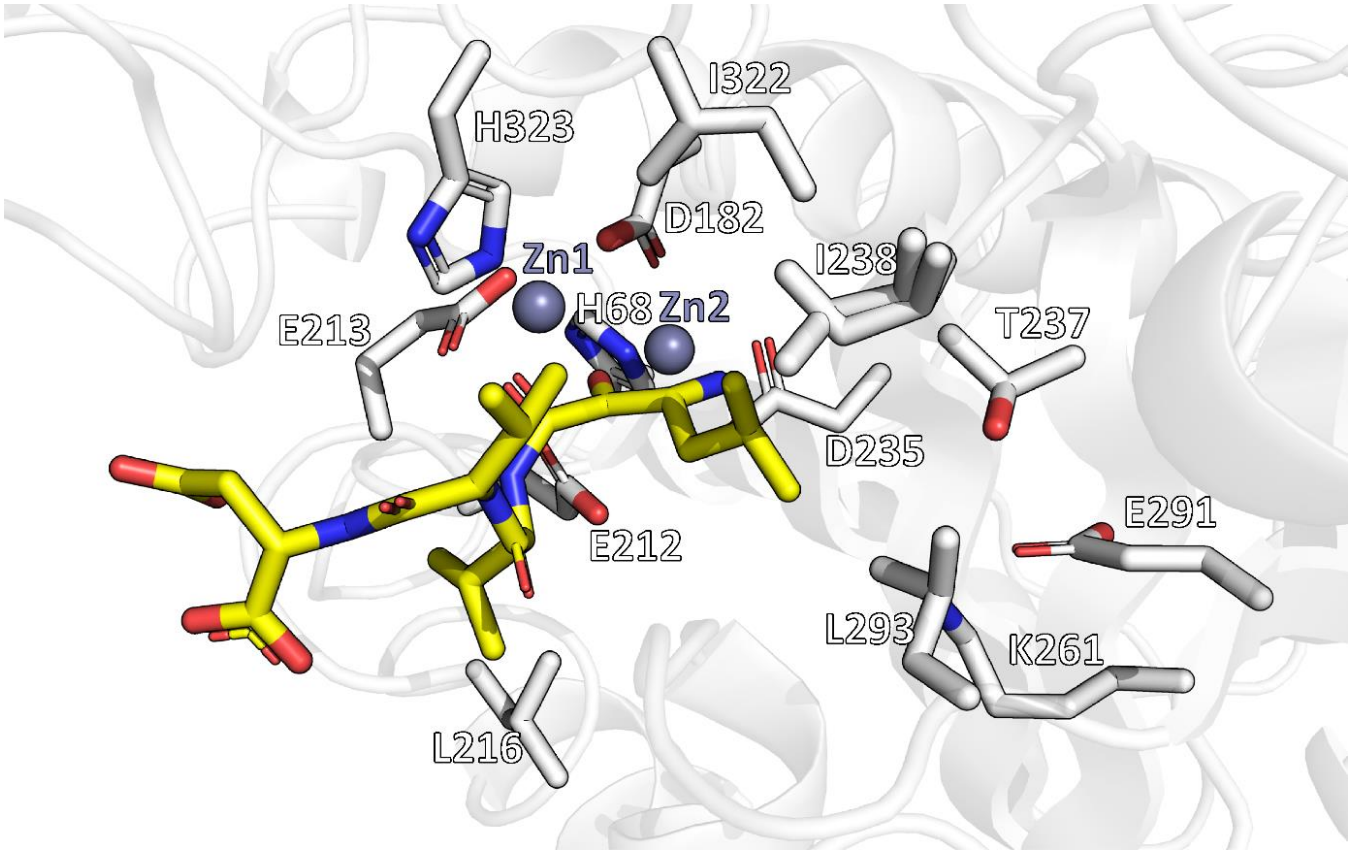


Figure 38 – Schematic representation of the catalytic domain of PhTET2. The residues involved in metal ion binding and the catalytic glutamate residue are highlighted as well as the residues of the S_1 and S_1' pockets. Amastatin is colored in yellow while the two Zn^{2+} are represented as gray spheres. To draw this figure, the published coordinates (PDB code: 1Y0Y) have been modified to place correctly amastatin in the experimental electronic density.

Ps_Cel5A hydrolyzes CMC at a maximal rate of 840 µmol min^{-1} per µmol of enzyme. Moreover, a polysaccharide like laminarin could hardly fit in the active site of an M42 aminopeptidase, being too exiguous to fit branched carbohydrates. Consequently, the cellulolytic activity observed in some M42 aminopeptidases seems fortuitous and artefactual, probably relying on exogenous nucleophiles, such as azide[404]. M42 aminopeptidases, however, could bind carbohydrates as MHJ_0125 of *M. hyopneumoniae* can bind heparin without degrading it[331]. Another side function of MHJ_0125 is being involved in the plasminogen activation. Somehow, it promotes the conversion of inactive plasminogen to active plasmin by the plasminogen activator in a porcine model. Intriguingly, MHJ_0125 does not degrade neither the plasminogen nor the plasminogen activator[331].

A.5.3 Structure of M42 aminopeptidases

As already mentioned in the former **sections A.4.5 and A.5.1.1**, the most remarkable trait of M42 aminopeptidases is their quaternary structure. Along with M18 aminopeptidases, they adopt a peculiar tetrahedron-shaped structure with twelve subunits. Such a spatial organization is important to compartmentalize the activity, avoiding uncontrolled peptide degradation[184,263]. Before describing in depth this quaternary structure in **section A.5**, the subunit structure will be explained as some of its structural features intervene in oligomerization.

A.5.3.1 Subunit structure

The structure of PhTET2 has been the first structure of an M42 aminopeptidase solved at high resolution[184,265]. Its subunit is made of two distinct domains, a catalytic domain and a dimerization domain (see **Figure 37**). The PhTET2 subunit fold is observed in all structurally characterized M42 aminopeptidases[184,265,325,326,330,365,378].

A. Catalytic domain structure

As discussed in **section A.5.1.1**, the catalytic domain adopts the α/β globular structure shared by all MH clan members. Indeed, it shows a significant structural homology with VpAp1 (see **Figure 30**, page 24). This domain consists of a central β sheet surrounded by eight helices (see **Figure 37**). The central β sheet is made of eight strands with β7 and β10 strands being twisted like in the VpAp1 structure. An extension of three β strands is grafted on the β10 strand. This β sheet extension is directly involved in dimerization as it interacts with the dimerization domain of the neighbor subunit. Interestingly, such an extension is also observed in M18 and M20 peptidases, suggesting that it is an ancestral adaptation promoting oligomerization. As expected, the active site of M42 aminopeptidases resembles to that of VpAp1 with the metal ion ligands and the catalytic residue being strictly conserved. In the PhTET2 structure, the M1 site consists of Asp-182, Glu-213, and His-323 while the M2 site is made of His-68, Asp-182, and Asp-235 (see **Figure 38**). Of note, Asp-182 is shared between both sites as it bridges both Zn^{2+}. The position of Asp-182 side chain requires important geometry constraints, especially the Cis-peptide bond between Asp-182 and Asp-183. Such a Cis-peptide bond is observed in all MH clan members.

Borissenko and Groll (2005) solved the structure of PhTET2 in complex with amastatin[184]. This natural peptide analog from *Streptomyces* sp. is a known inhibitor of dinuclear aminopeptidases via the non-hydrolysable α-hydroxyl-β-leucine found at its N-terminus[405–409]. In PhTET2, amastatin binds to the two Zn^{2+} of the active site, mimicking substrate binding (see **Figure 38**). The position of the N-terminal residue has allowed to define the S_1 pocket of PhTET2[184]. The S_1 pocket consists of Thr-237,

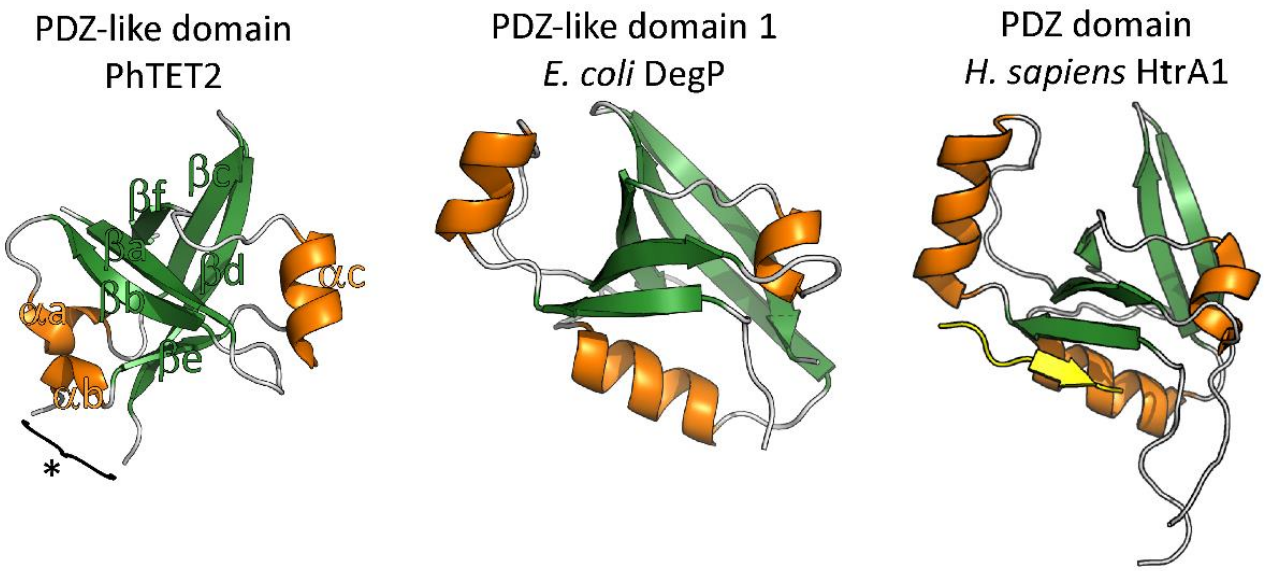


Figure 39 – Comparison of the PDZ-like domain of PhTET2 (PDB code: 1XFO) with the PDZ domains of *E. coli* DegP (PDB code: 3CS0) and *H. sapiens* HtrA1 (PDB code: 2JOA). α helices and β strands are colored in orange and green, respectively. The gap in the PDZ-like domain of PhTET2 is shown by an asterisk (*). The target peptide recognized by HtrA1 is highlighted in yellow.

Ile-238, Lys-261, Glu-291, Leu-293, and Ile-322. Borissenko and Groll (2005) also identified the S_1' pocket to which take part the backbone of Val-214 - Gly-215 and the side of Leu-216. As seen in **section A.5.2.1**, the substrate specificity is defined by the composition of the S_1 pocket. For PhTET2, the S_1 pocket is considered as hydrophobic, despite the presence of a lysine residue and a glutamate residue[184]. This hydrophobicity correlates to the leucyl-aminopeptidase activity of PhTET2. Accordingly, an arginine residue is found in the S_1 pocket of PhTET1, matching with its preference towards negatively charged residues[326]. Furthermore, for the lysyl-aminopeptidase PhTET3, two glutamate residues are lined at the bottom of the S_1 pocket[325]. A third metal ion could be bound in the S_1 pocket of lysyl-aminopeptidases, broadening their substrate specificity to negatively charged residue. Indeed, Colombo *et al.* (2016) reported a third metal ion center in the S_1 pocket of PfTET3[365]. PfTET3 was shown to degrade L-Glu-*p*NA in addition to L-Lys-*p*NA but only in presence of a Co^{2+} excess.

B. Dimerization domain structure

PDZ domains are protein modules involved in peptide recognition, protein-protein and protein-lipid interactions[410]. They mediate protein-protein interactions by interacting with either the C-terminus or an unfolded internal peptide motif or another PDZ domain of an interacting partner. In *Homo sapiens*, 270 PDZ domains have been identified in more than 150 proteins, many of them participating in heterocomplexes[411]. Originally, they have been identified as repeated domains of 80-100 residues found in metazoan membrane-associated guanylate kinases (MAGUK)[412]. The name "PDZ" even originated from three MAGUK: PSD95, DLG, and ZO-1[413]. Despite their name, MAGUK act as scaffolding proteins only since their kinase domain is inactive[414]. They play critical roles in tissue development and differentiation in metazoans. Their PDZ domains mediate interactions between membrane proteins and the cytoskeleton. Different families of PDZ domains have been reported, especially in proteins unrelated to MAGUK[415]. To mark the difference of sequence and structure, the PDZ domains of MAGUK are often referred as canonical[410]. They are found only in metazoans while non-canonical PDZ domains are present in the whole tree of Life. Recently, twelve PDZ domain families have been identified from analysis of 1,400 prokaryotic and fungal genomes[416]. In prokaryotes, the combination of a PDZ domain and a protease occurs in 88 % of PDZ domain-containing proteins.

The PDZ-like domain of M42 aminopeptidases looks like that of the HtrA family (see **Figure 39**). The resemblance, however, is structurally rough as the succession of secondary structures is different between PDZ domains of M42 aminopeptidases and HtrA. In bacteria, HtrA, also known as DegP, is a serine protease found in the periplasm. As mentioned in **section A.3.7**, DegP possesses two PDZ domains, one involved in DegP oligomerization and the other one in sensing unfolded proteins[175,177,417]. The peptide binding site has been identified for the mitochondrial HtrA1 from *H. sapiens*: the target peptide takes part to the β sheet of the PDZ domain while being sandwiched by an α helix[415,418]. The canonical structure of a PDZ domain of the HtrA family consists of a β sheet of 5-6 strands surrounded by three α helices. The PDZ-like domain of M42 aminopeptidases is also made of a β sheet of 6 strands and three helices. Although, the position of two helices differs from HtrA PDZ domains (see **Figure 39**). The first helix (αb), sandwiching the target peptide, appears to be shorter in M42 aminopeptidases. The second helix (αa) is not delocalized like in the HtrA PDZ domains

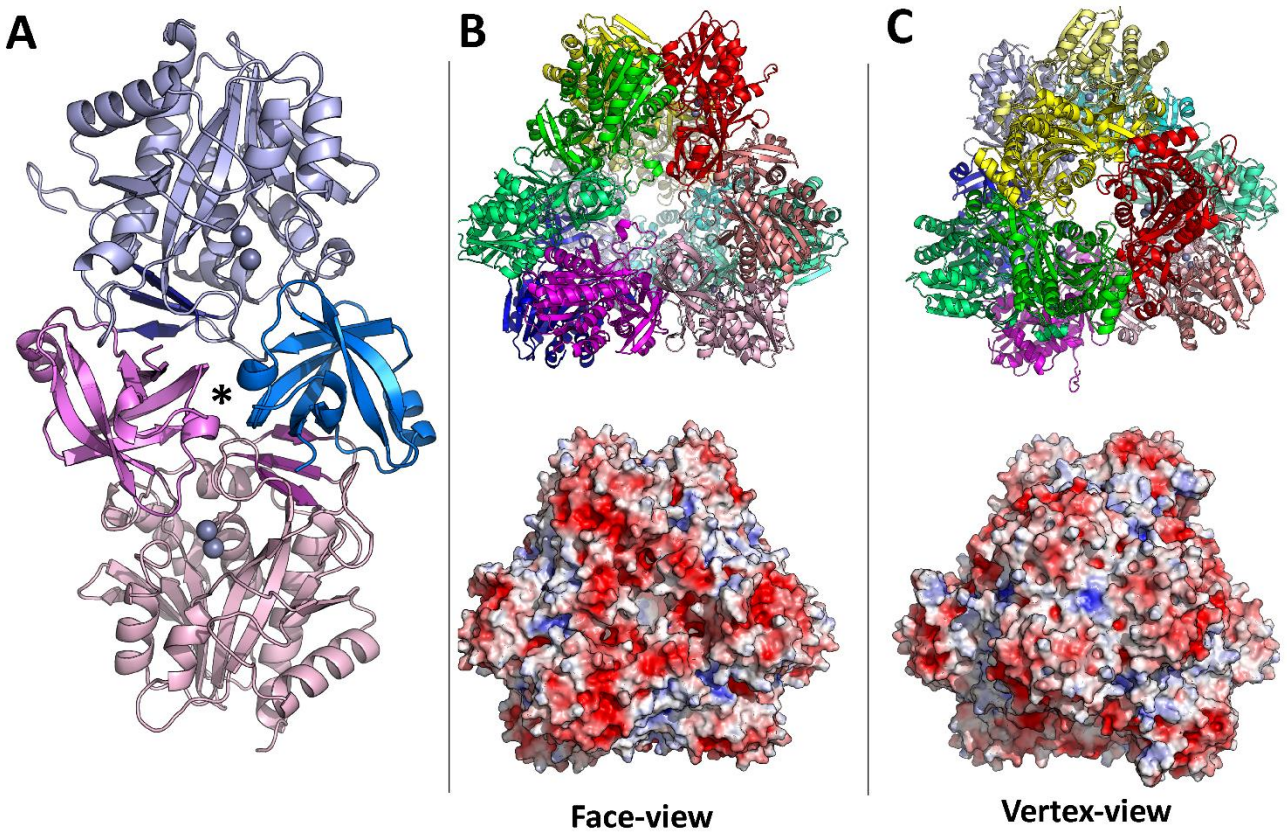


Figure 40 – Schematic representation of the quaternary structure of M42 aminopeptidases. The structure of PhTET2 is chosen as a representative model (PDB code: 1XFO). **(A)** The dimer structure taken from the dodecamer structure. The two subunits are distinguished in pink and blue hues. The catalytic domains are colored in light pink and light blue. The dimerization domains are colored in magenta and blue. The β sheet extensions are highlighted in purple and dark blue. The interaction between αa-βb hinges of the dimerization domains is marked by an asterisk (*). Zn^{2+} are represented as grey spheres. **(B)** Above: Schematic representation of the dodecamer structure seen from a face of the tetrahedron. The three dimers delineating the tetrahedron face are colored in green, red, and pink hues. Below: Surface representation of the dodecamer structure showing the entrance at the center of the tetrahedron face. **(C)** Above: Schematic representation of the dodecamer structure seen from a vertex of the tetrahedron. The three subunits found at the vertex are colored in green, red, and yellow. Below: Surface representation of the dodecamer structure showing the exit at the intersection between the three subunits. The electronic surface was calculated using Advanced Poisson-Boltzmann Solver (APBS)[225].

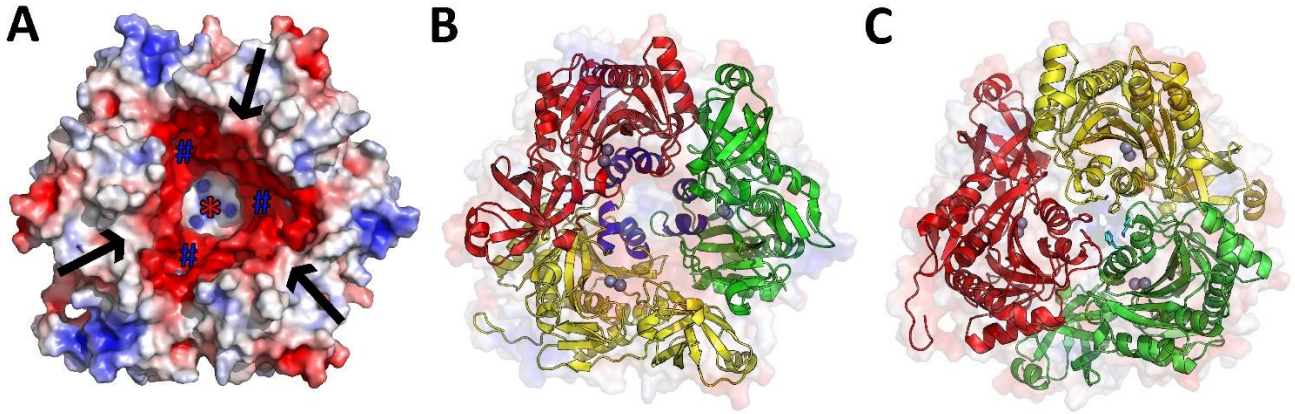


Figure 41 – Spatial organization of three subunits at a vertex of the tetrahedron. The structure of PhTET2 is chosen as a representative model (PDB code: 1XFO). **(A)** Surface representation of three subunits found at a vertex of the tetrahedron. The vertex is oriented upside down as seen from the center of the inner cavity. The arrows indicate the access points from where peptides enter the inner cavity. The positions of the catalytic sites are indicated by blue octothorpes (#). The exit is indicated by a red asterisk (*). **(B)** Schematic representation of the same tetrahedron vertex with the three monomers colored in green, red, and yellow. The six α helices forming the exit tunnel are highlighted in blue. These helices correspond to α4 and α7 helices of PhTET2 as represented in **Figure 37**. Zn^{2+} are represented as grey spheres. **(C)** Schematic representation of the same tetrahedron vertex as seen from the outside. The three monomers are colored in green, red, and yellow. The side chains of three phenylalanine residues are highlighted as they close the exit. Three lysine residues could also limit the access of the exit channel. For B and C, the surface is overlaid in transparency. The electronic surface was calculated using APBS[225].

but is found in close vicinity of the β sheet. Intriguingly, in all structurally characterized M42 aminopeptidases, a gap has been reported in the PDZ-like domain between the βd strand and αb helix (see **Figure 39**)[184,265,325,326,330,365,378]. This segment has been described to be too flexible so that no electron density can be seen in the X-ray data collections.

A.5.3.2 Quaternary structure

M42 aminopeptidases adopt a peculiar tetrahedron-shaped structure made of twelve subunits. Such a quaternary structure has been reported for all structurally characterized M42 aminopeptidases[184,262,265,319,325,326,330,334,365,378]. It is even commonly referred as the catalytically active oligomer. In fact, the dodecamer is often regarded as the self-assembly of six dimers. The dimer is supposed to be the second stable oligomer encountered in M42 aminopeptidases[184,382,419]. Several studies have reported that dimers are less active than dodecamers[336,365,419]. The dimer-dodecamer transition seems to be driven by the metal ion cofactors[184,365,382,420]. This transition will be further discussed in the next section. The dimer structure, however, is not yet known but is expected to be close to that found in the dodecamer[419]. As explained in the previous section, two subunits can interact through their dimerization domains. The PDZ-like domain of one subunit makes several hydrogen bonds and salt bridges with the catalytic domain of the other monomer, more specifically the β sheet extension of three β strands (see **Figure 40.A**). Few hydrogen bonds occur between two dimerization domains, mainly at the hinges between the βb strand and αa helix in PhTET2. Hence, the putative peptide binding site of the PDZ-like domain remains vacant, even in the dodecamer. Hypothetically, it could bind an external peptide acting as a degradation signal or it could promote transient interactions between dimers in the path to dodecamer formation.

The dodecamer consists of six dimers spatially organized under a tetrahedron geometry (see **Figure 40.B-C**). Each dimer is positioned along an edge of the tetrahedron. Hence, each face of the tetrahedron is delineated by three dimers while three subunits are found at each vertex of the tetrahedron. A central entrance lies at the center of each face leading to a wide inner cavity (see **Figure 40.B**). The entrances have been described to be the access gate through which enters the substrate. The entrance diameter is about 18-20 Å, allowing the entry of folded peptides but not proteins[184,326]. The inner cavity has a diameter of about 60 Å[184]. The dimers are organized in such a manner that the twelve catalytic sites are oriented inward this cavity (see **Figure 41.A**). Thus, the hydrolytic activity is compartmentalized in a wide catalytic chamber with a restricted access via the four entrances. Such a nano-compartment avoids uncontrolled peptide degradation. Amino acids generated during peptide hydrolysis are believed to exit via the vertices of the tetrahedron. Indeed, at each vertex, three subunits form an exit channel delineated by the six α helices (see **Figure 41.B**). The exit channel ends with a small opening whose diameter varies from 3.5 to 15 Å[184,325,326]. In several structures, this opening appears to be closed by three residues, but a conformation change of these residues could open the exit door (see **Figure 41.C**)[325].

In the dodecamer structure, dimers are connected through extensive contacts between their catalytic domains, mainly at the vertices of the tetrahedron (see **Figure 40.C** and **Figure 41.B-C**). According to PDBe PISA analysis of the PhTET2 structure, the interaction between dimers contributes more to the stability of the complex than the interaction between two subunits in a dimer[421]. Indeed, the interaction between two catalytic domains contributes to complex stabilization at the level of $\Delta^{i}G$ of -16.5 kcal mol^{-1}. In comparison, the formation of a dimer increases the free energy of interface

formation by -7.0 kcal mol^{-1}. Of note, each metal ion in the active site contributes greatly to the complex stabilization, with a $\Delta^{i}G$ of about -33 kcal mol^{-1}. The surface of interaction between two catalytic domains is about 1,200 $Å^2$ (8% of subunit surface), involving 35 residues of each subunit. The stability of dodecamers have been demonstrated for several M42 aminopeptidases in different experimental conditions including extreme ranges of temperature, salt, pressure, denaturant, and pH[319,324,325,328,333,339,369,371,382,419,422,423].

A.5.3.3 Dodecamer assembly

In addition to their catalytic role, the metal ion cofactors seem to be the driving force of the oligomerization of M42 aminopeptidases. Several studies have reported that dodecamers can dissociate into dimers or even monomers under particular experimental conditions[184,365,382,419,420,424]. These conditions are often linked to a loss of metal ion cofactors. As discussed in the previous section, metal ions contribute to the dodecamer stability. Indeed, for several M42 aminopeptidases, it has been demonstrated that the apo-form tends to be either thermolabile or pH sensitive[333,371,372,382,424]. The addition of metal ions to the apo-enzyme usually restores the activity and stability. The removal of metal ions can be achieved by dialyzing the enzyme in presence of a chelating agent, such as EDTA or 1,10-phenanthroline[333,365,371,372,382]. After such a treatment, the apo-enzyme may remain dodecameric, as shown for the aminopeptidase I of *G. stearothermophilus* (GsApI)[372]. In their study, Roncari *et al.* (1972) observed that the apo-GsApI dissociated into species of lower molecular weight upon heating. Temperature seems to induce structural rearrangement in the dodecamer structure, as shown for ADPkam589[425]. Dodecamers can also disassemble after a treatment at a basic pH, usually higher than 8.5[382,420,426]. Rosenbaum *et al.* (2011) showed that a treatment at pH 11 provokes the dissociation of about 50% of PhTET3 dodecamers[382]. Such a phenomenon is probably linked to the formation of metal ion hydroxides (see **Figure 35**, page 28). A high dose of Zn^{2+} at acidic pH also leads to deoligomerization. For PhTET2, dodecamers dissociate into dimers or even monomers in presence of 0.5 M Zn^{2+} at pH < 4 [184,427]. Several studies have shown that the dissociation can be reverted by adding the right metal ion cofactors[382,419,420,423,426]. Thanks to the reversibility of the dimer-dodecamer transition, it has been possible to study the interaction between dimers on the path to the dodecamer formation[382,419,420,426].

Rosenbaum *et al.* (2011) highlighted several transition oligomers for PhTET3[382]. In their study, the apo-form is mainly a dimer, but tetramers and hexamers were also observed in minority. Adding Co^{2+} promoted the dimer-dodecamer transition with a third transition oligomer being observed, octamers. In addition, the authors reported that PhTET3 dimers are thermolabile and could either dissociate into monomers or partly lose their fold. Appolaire *et al.* (2013) also observed tetramers, hexamers, and octamers as transition oligomers of PhTET2[419]. The authors adopted another strategy to force the dissociation of PhTET2 dodecamers. They substituted five residues of the catalytic domain involved in the interaction interface at the tetrahedron vertex. Remarkably, such a variant of PhTET2 is a dimer that slowly self-assembles into dodecamers over the time. Due to this peculiar property, the transition oligomers tend to accumulate with the following proportions: 17.1% dimers, 19.5% tetramers, 46.4% hexamers, 8.3% octamers, and 3.8% dodecamers. The structure of PhTET2 dimer was studied by SAXS. Its structure, however, does not adopt a different fold compared to that of a dimer in the dodecamer structure. Appolaire *et al.* (2013) also studied the structures of transition oligomers by electron microscopy. Three different quaternary structures have been reported: a Z-

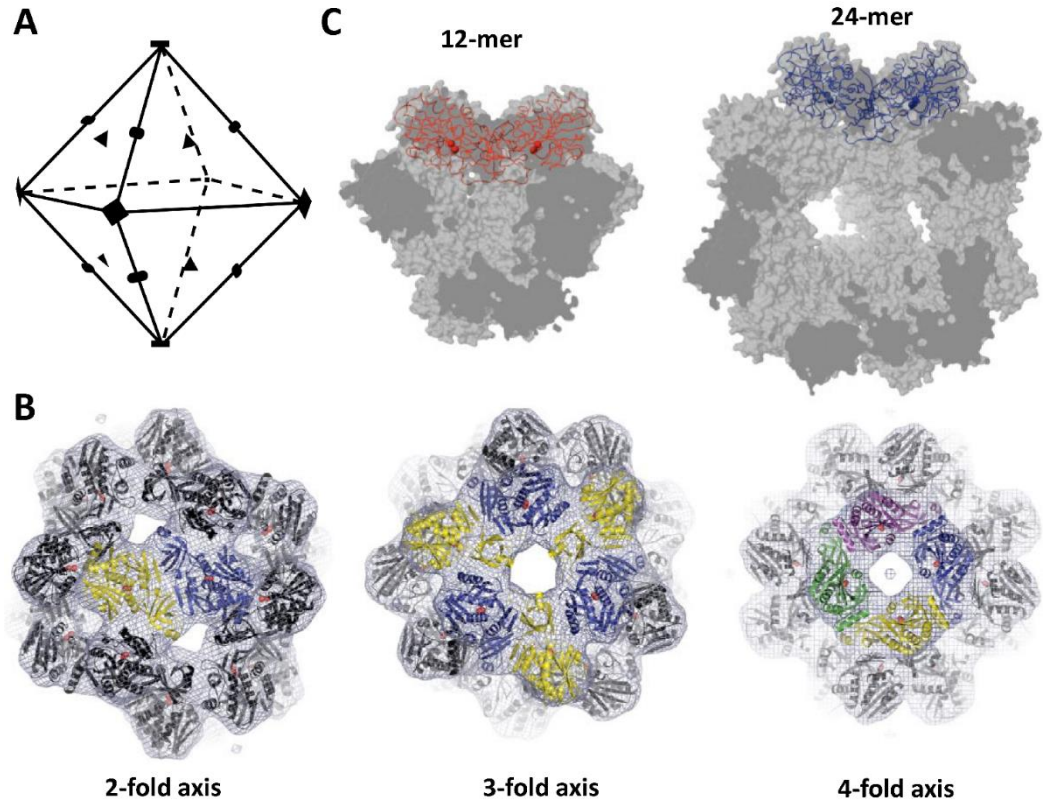


Figure 43 – The octahedron-shaped tetracosamer of PhTET1. (A) Scheme of an octahedron with the three types of symmetry axes. 2-fold, 3-fold, and 4-fold axes are marked with ovals, triangles, and square, respectively. **(B)** The electron density of tetracosameric PhTET1 seen from each symmetry axis. A model, build using the coordinates of the dodecamer structure, is fitted inside the electron density. **(C)** Representation of the distortion of the dimer in the tetrahedron-shaped dodecamer (in red) and the octahedron-shaped tetracosamer (in blue). Figure designed using artworks from Schoehn *et al.* (2006)[326].

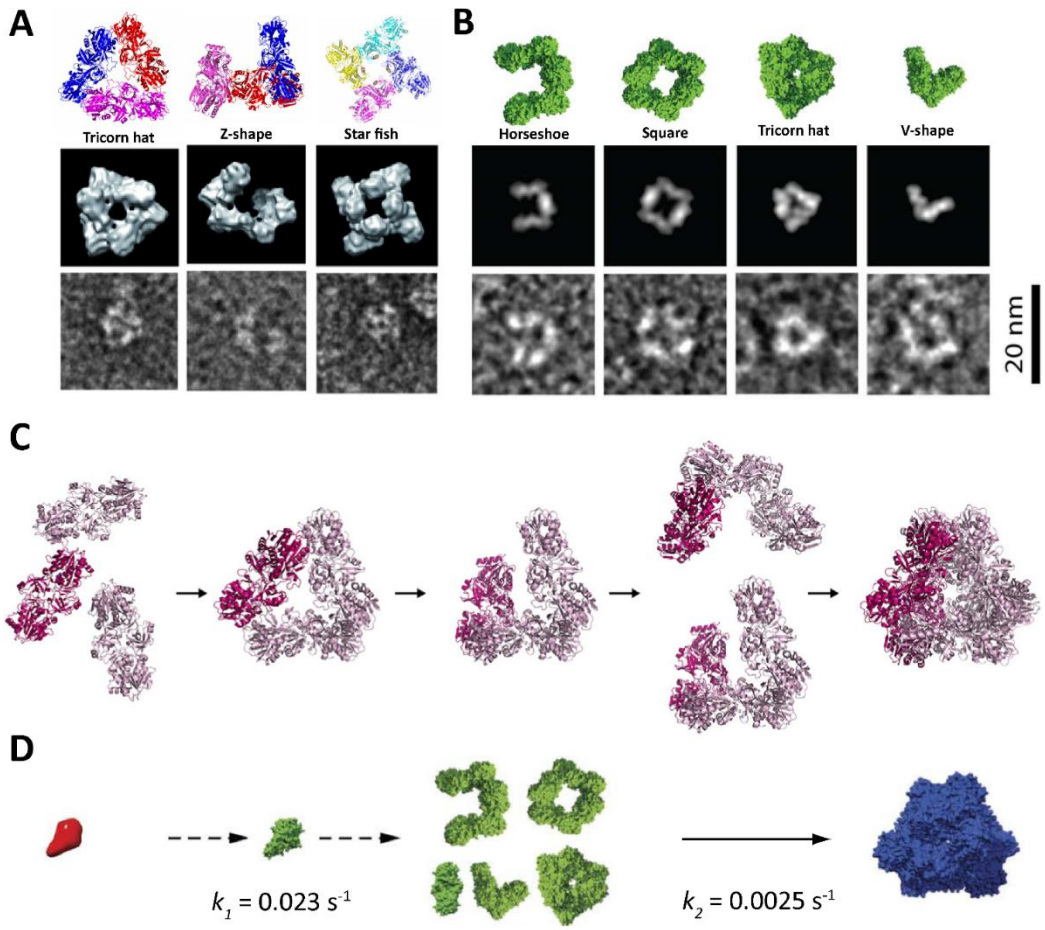


Figure 42 – The dodecamer assembly of PhTET2. (A) The structure of transition oligomers observed in Appolaire *et al.* (2013)[419]. **(B)** The structure of transition oligomers observed in Macek *et al.* (2017)[420]. **(C)** The dodecamer assembly model of Appolaire *et al.* (2016)[263]. Three dimers associate into either a tricorn hat hexamer or a Z-shape hexamer. The dodecamer is formed by two interacting Z-shape hexamers. **(D)** The stochastic dodecamer assembly model of Macek *et al.* (2017)[420]. The dodecamer is formed by either incremental addition of dimers or interaction between intermediate oligomers. Figure designed using artworks from Appolaire *et al.* (2013), Appolaire *et al.* (2016), and Macek *et al.* (2017)[263, 419, 420].

shaped hexamer, a tricorn hat hexamer, and a four legged star fish octamer (see **Figure 42.A**). While the hexamer structures are compatible with the dodecamer, the octamer structure suggests that it could be an intermediate oligomer in the formation of a tetracosamer (24-mer). This higher form of oligomerization has been reported for PhTET1 (see next section). The Z-shaped hexamer has been observed in the formation of heterododecamer made of PhTET2 and PhTET3 subunits (see **section A.5.4.2**)[426].

Appolaire *et al.* (2016) proposed a dodecamer assembly process (see **Figure 42.C**)[263]. In this model, dimers do not interact randomly but rather form hexamers. The two Z-shape hexamers further interact to form a dodecamer by adapting their "dimeric arms". Still, a recent study has challenged this model in favor to the random association of dimers. Macek *et al.* (2017) also studied the dodecamer assembly of PhTET2 by combining native MS, NMR, and cryo-EM data[420]. In their study, PhTET2 dodecamers were dissociated completely to monomers by adding Zn^{2+} in acidic condition. After such a treatment, the PhTET2 monomers are unstructured, lacking stable secondary structures. Monomers were refolded by removing Zn^{2+} excess and increasing the pH back to neutral. During the refolding, several intermediates were observed, ranging from dimer to dodecamer by increments of two subunits. Various structures of transition oligomers were observed: a V-shape tetramer, a tricorn hat hexamer, a horseshoe hexamer and a square octamer (see **Figure 42.B**). The decamer appeared to have the same architecture than the dodecamer but with some missing density. Based on their results, Macek *et al.* (2017) proposed a stochastic model where PhTET2 dimers interact randomly (see **Figure 42.D**). The discrepancy between the two models could be due to an altered dimer-to-dimer interaction in the PhTET2 variant used in Appolaire *et al.* (2013). Another interesting fact about oligomerization is the distinct role of the two metal ions in the active site. Colombo *et al.* (2016) showed that the metal ion in the M2 site is more tightly bound than its counterpart in the M1 site for PfTET3[365]. When the metal ion of the M1 site is chelated, PfTET3 retains 75% of its activity but partly dissociates into monomers (60% dodecamers and 40% monomers). When both metal ions are removed, the dodecamer completely dissociates into monomers. Hence, the M2 metal ion is catalytically active and is enough for the dodecamer formation while the M1 metal ion probably stabilizes the complex.

A.5.3.4 Miscellaneous quaternary structures

As explained in the previous section, two stable oligomeric forms, dimers and dodecamers, are encountered for the M42 aminopeptidases. Both oligomers co-exist *in vivo* as shown in *P. horikoshii*[268,419]. A third edifice has been reported for PhTET1, an octahedron-shaped tetracosamer[326,428]. An octahedron possesses twelve edges with a 2-fold symmetry axe, eight faces with a 3-fold symmetry axe, and six vertices with a four-fold symmetry axe (see **Figure 43.A**). In this quaternary structure, a dimer is found along each edge (see **Figure 43.B**). Thus, three dimers delineate each face, like in the tetrahedron-shape dodecamer. Meanwhile, each vertex of the octahedron is defined by four subunits (see **Figure 43.B**). Consequently, the size of the tetracosamer exit pores is increased to 25 Å of diameter instead of 15 Å in the dodecamer. In that case, the exit pores are bigger than the entrances found at the faces (20 Å). According to Schoehn *et al.* (2006), dimers must be distorted to accommodate the octahedral geometry constraints[326]. In the tetracosamer, one subunit of the dimer must be rotated by 14° compared to the center of mass of the dimer in the PhTET1 dodecamer (see **Figure 43.C**). The octahedron-shaped structure, however,

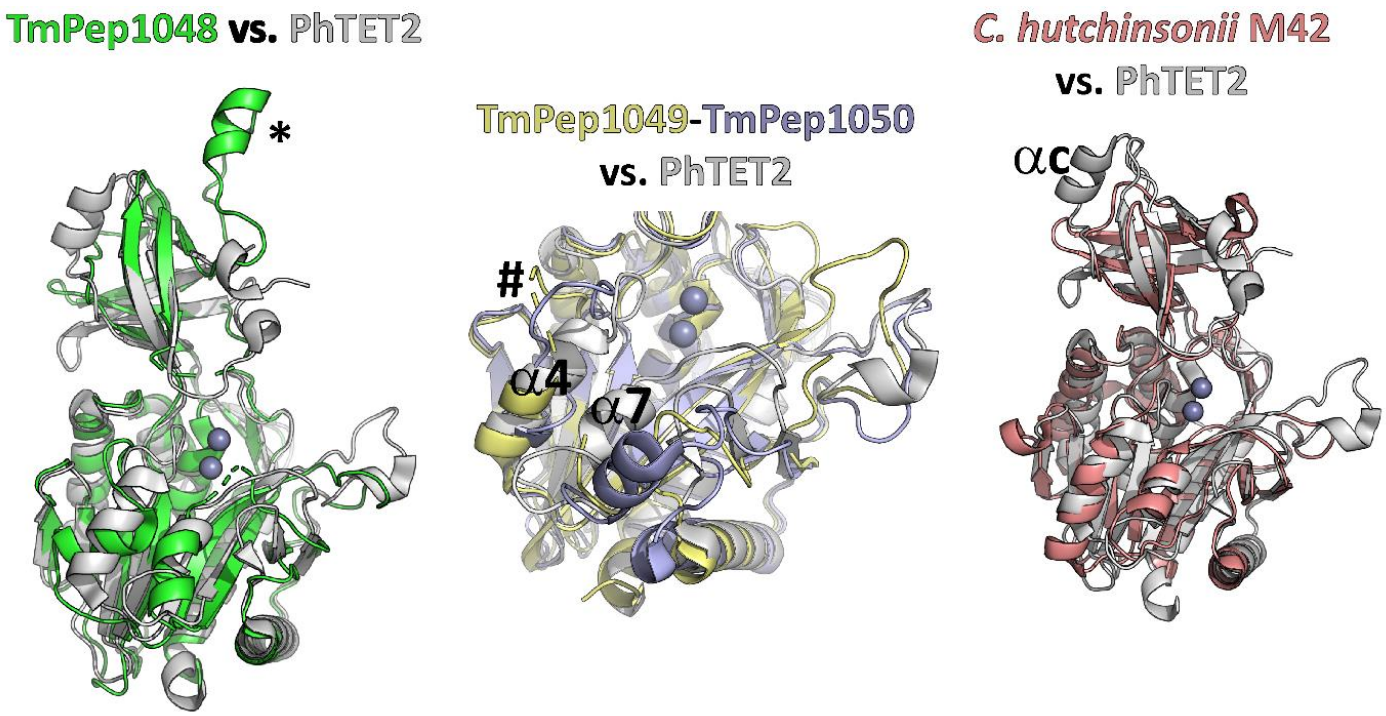


Figure 44 – Structural comparison of PhTET2 (PDB code: 1XFO, in grey) vs. TmPep1048 (PDB code: 1VHO, in green), TmPep1049 (PDB code: 2FVG, in yellow), TmPep1050 (PDB code: 3ISX, in blue), and *C. hutchinsonii* M42 aminopeptidase (PDB code: 3CPX, in pink). The αc, α4, and α7 helices are annotated by their names. The additional decentered α helix of the dimerization domain of TmPep1048 is marked by an asterisk (*). The loop bringing the catalytic base into the active site is marked by an octothorpe (#). The two Zn^{2+} of PhTET2 are represented as grey spheres.

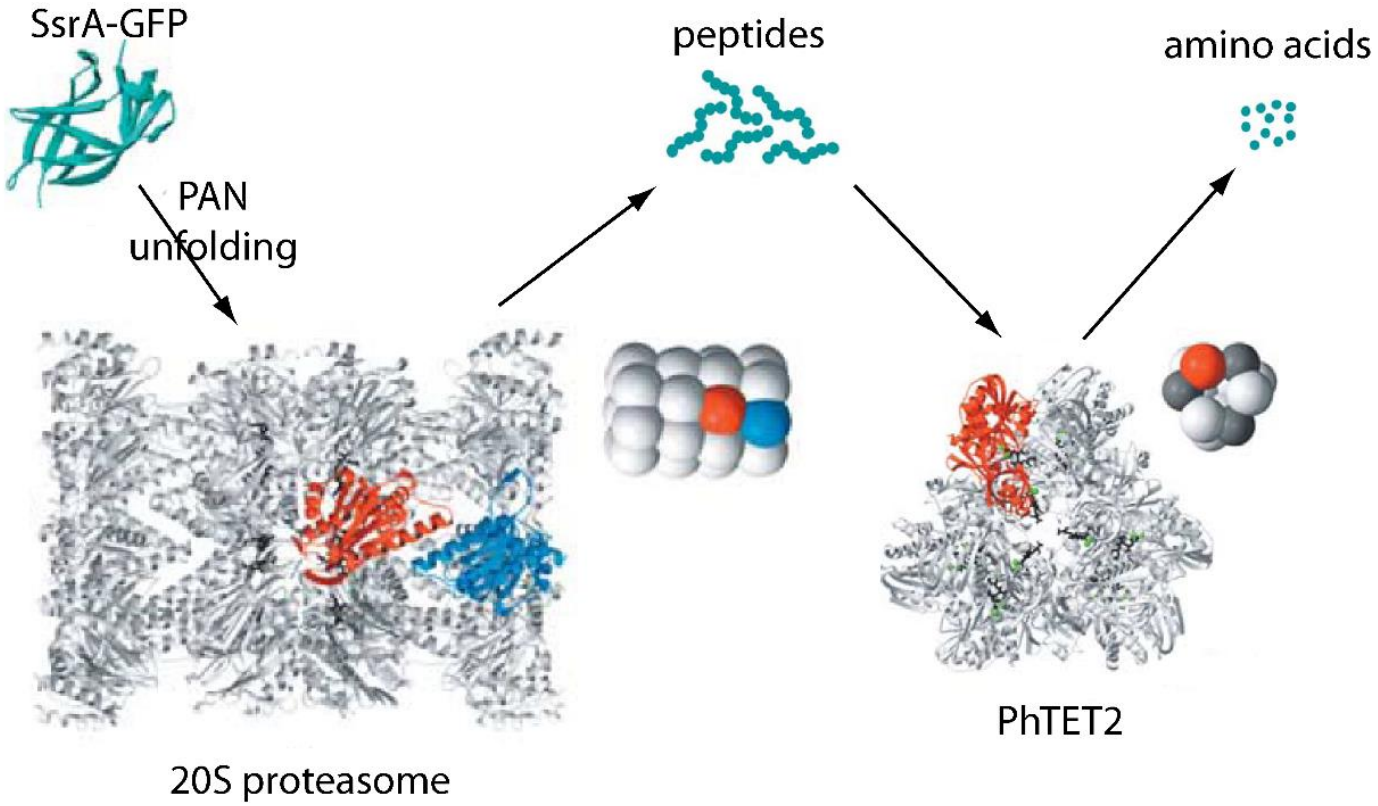


Figure 45 – Reconstruction of a protein degradation complex involving PAN and proteasome 20S from *M. jannashii*, and PhTET2. Figure adapted from Borissenko and Groll (2005)[184].

remains a singularity among M42 aminopeptidases as PhTET1 is currently the single example. Of note, Appolaire *et al.* (2013) postulated that transition octamers of PhTET2 could be tetracosamer precursors[419].

Several studies have reported different oligomeric states, other than dimers and dodecamers, for various M42 aminopeptidases[328,333–336,367,370]. To my opinion, these degrees of oligomerization, ranging from monomer to octadecamer, must be interpreted with caution. Indeed, the molecular weight of M42 aminopeptidases is usually determined by size exclusion chromatography. The protein elution profile greatly depends on properties of the gel matrix being used[429]. Principally, the size separation range of the gel matrix can be critical to determine the oligomeric state of high molecular weight proteins (typically 400-500 kDa for a dodecameric M42 aminopeptidase). In addition, the structure of M42 aminopeptidases is not globular, influencing gyration radii[429]. For instance, a dimer will look bigger than its real size. Still, some unpublished data are worth to be reported[318]. The structure of TmPep1048 from *T. maritima* (PDB code: 1VHO) appears as a monomer, despite its high structural homology with the PhTET2 subunit[184,265]. The structure of a monomer is in full contradiction with NMR data in Macek *et al.* (2017), showing the unfolded structure of PhTET2 monomers. The dimerization domain of TmPep1048 displays an unusually decentered α helix (see **Figure 44**). The active site is not correctly folded, with the loop bringing the catalytic base 8 Å away from the catalytic center. Intriguingly, the two other M42 aminopeptidases from *T. maritima* have a structure differing from that of PhTET2, for instance. Based on available structural data in the Protein Data Bank, TmPep1049 and TmPep1050 are monomeric and dimeric, respectively. Their catalytic domain seems to be "exploded", with dramatic movement of the α4 and α7 helices (see **Figure 44**). A fourth bizarre structure is that of an M42 aminopeptidase of *Cytophaga hutchinsonii* which forms, apparently, trimers. Its catalytic domain seems to be correctly folded with two iron ions found in the active site. The dimerization domain, however, appears to lack the αc helix (see **Figure 44**). This helix does not seem to be involved in the oligomerization of PhTET2, being exposed at the surface of the dodecamer. Hence, the structure of a trimeric M42 aminopeptidase remains puzzling. These anomalies will be discussed further in **sections B.1** and **B.4** (page 41 and page 98).

A.5.4 Physiological function

A.5.4.1 Proteasome link

M42 aminopeptidases adopt the peculiar tetrahedron-shaped structure where peptide hydrolysis occurs inside a vast catalytic chamber. They can degrade peptides of up to 32 amino acids in length. Based on these criteria, Franzetti *et al.* (2002) proposed that the M42 aminopeptidases could be linked to the proteasome or other related proteolytic complexes[262]. Borissenko and Groll (2005) reconstituted *in vitro* an entire protein degradation system[184]. The authors combined the proteasome-activating nucleotidase and the proteasome 20S of *Methanococcus jannashii* with PhTET2 (see **Figure 45**). The three complexes allowed the complete destruction of SsrA tagged GFP into amino acids. The link between the proteasome and the M42 aminopeptidases, however, has not yet been proved *in vivo* despite several indications. Indeed, in *T. kodakarensis*, two M42 aminopeptidases are overproduced under oxidative and heat stresses[334]. Coincidently, proteolysis is also induced under these stresses in prokaryotes[430,431]. Furthermore, some M42 aminopeptidases form heterocomplexes with the advantage of a broader substrate specificity and a synergetic peptide

hydrolysis. Such heterocomplexes could destroy any kind of peptides generated by the proteasome or other related proteolytic complexes[268].

A.5.4.2 Heterocomplexes

Franzetti and coworkers published key studies about the occurrence of heterocomplexes in *P. horikoshii*[268,419,426]. Heterocomplexes made of PhTET2 and PhTET3 subunits have been reported in *P. horikoshii*. Unexpectedly, PhTET1 was not found in these heterocomplexes and even only dimers were observed for PhTET1 *in vivo*[268]. Hence, the function of PhTET1 could be different from PhTET2 and PhTET3 or the dodecamer assembly could be regulated in another way. The PhTET2-PhTET3 heterocomplexes can degrade Val-Leu-Lys-pNA more efficiently than homo-dodecamers, showing a broader specificity of the heterocomplexes and a synergy in peptide hydrolysis. The PhTET2-PhTET3 heterocomplexes were reconstituted *in vitro* by combining de-oligomerized PhTET2 and PhTET3[426]. Two different heterocomplexes were characterized by SANS showing two possible PhTET2-to-PhTET3 subunit ratios. One heterocomplex is composed of six subunits of each enzyme while the other is made of eight PhTET2 subunits and four PhTET3 subunits.

Several studies support the existence of heterocomplexes in other microorganisms than *P. horikoshii*[332,372,423,424,432]. The aminopeptidase I of *G. stearothermophilus*, GsApI, displays leucyl-, glutamyl-, and glycyl-aminopeptidase activities[369]. Its apparent broad specificity is due the existence heterocomplexes. Stoll *et al.* (1972) identified two types of subunit in GsApI, named α and β[423]. Three different heterocomplexes of GsApI were purified from *G. stearomthermophilus*, having different α-to-β ratios: six α subunits and six β subunits, eight α subunits and four β subunits, and ten α subunits and two β subunits. Homocomplexes of GsApI were obtained *in vitro* after de-oligomerization and reassociation. The β subunits failed to reassemble into dodecamers while the α subunits did[423]. The α homododecamer, however, did not display any glutamyl-aminopeptidase activity, indicating this specificity is linked to the β subunit[432]. The inability of the GsApI β subunit to form dodecamers is not surprising since some M42 aminopeptidases are monomeric or dimeric, like the enzymes of *T. maritima* (see **section A.5.3.4**). Heterocomplexes have also been reported in *Symbiobacterium thermophilum*[332]. In this bacteria, three genes coding M42 aminopeptidases are organized as an operon. Taken separately, Pep1079, Pep1080, and Pep1081 are poorly active but, when combined, a significant arginyl-aminopeptidase activity was demonstrated. Kumaki *et al.* (2011) also showed by surface plasmonic resonance that Pep1079 strongly interacts with Pep1080[332].

A.5.4.3 Cellular localization

The M42 aminopeptidases are believed to be intracellular enzymes, which is consistent with their supposed link to the proteasome[262,263,268,318,382,419]. Their N-terminal sequences do not contain any secretion signal peptide that could indicate a translocation to the membrane or the periplasm. The absence of a secretion signal peptide, however, is not an irrefutable proof of protein localization. Indeed, some proteins lacking a signal peptide can be secreted in *E. coli* and *B. subtilis*[433]. It is known that type 1, 3, 4, 6, and 7 secretion systems can export proteins lacking a secretion signal peptide[434]. Several studies have reported that M42 aminopeptidases could be associated to the membrane or even present at the cell surface[328,331,368–370,388,435]. For instance, GsApI seems to stick firmly to the membrane after cell disruption, necessitating a mechanical treatment to retrieve the enzyme[368,369]. Even with such a treatment, a fraction of GsApI remains associated to the membrane and can be

retrieved only with an alkaline extraction at pH 9.6[388]. Interestingly, GsApI purified from the membrane is a dodecamer displaying an additional activity towards positively charge residues. M42 aminopeptidases have also been reported to be partly associated to the membrane in *Lactococcus lactis*[328,370,435]. An intriguing case of study is MHJ_0125, an M42 aminopeptidase of *M. hyopneumoniae*[331]. While the majority of MHJ_0125 is intracellular, a fraction is bound to the membrane. In their experimental conditions, Robinson *et al.* (2013) observed that MHJ_0125 is the most abundant protein on cell surface. The authors also showed the presence of extracellular MHJ_0125 without recording any significant cell lysis. The physiological role of membrane bound M42 aminopeptidases is completely unknown. The PDZ-like domain could play a role in membrane anchoring as some canonical PDZ domains are involved in lipid binding[410].

Objectives

The M42 aminopeptidases of *P. horikoshii* have been well studied uncovering their peculiar tetrahedral-shaped structure, the oligomeric state transition induced by their metal ion cofactor, and the existence of heterocomplexes with broadened substrate specificity. Their bacterial counterparts, however, are ill-described due to the scarcity of structural data. For instance, it is currently unknown if the metal ion cofactor modulates their oligomeric states in a similar way to that of archaeal enzymes. Inasmuch as the oligomeric state transition is genuine, the structural modifications, that occur during the passage from dimers to dodecamers, are still poorly understood due to the lack of a high-resolution structure of a dimer. Furthermore, the physiological function of M42 aminopeptidases is under-examined although being postulated to hydrolyze peptides downstream the proteasome. M42 aminopeptidase heterocomplexes have been identified in *P. horikoshii* and *G. stearothermophilus*. Since they display an extended substrate specificity, they may act as "peptidasomes" but their occurrence in other microorganisms is not clearly established.

Considering the unanswered questions, this work will focus on the characterization of bacterial M42 aminopeptidases. Two bacterial models have been chosen: *T. maritima* and *E. coli*, each having three genes coding M42 aminopeptidases in their genome. The choice of *T. maritima* is motivated by a previous study and advanced progress in obtaining the structure of TmPep1050. Concretely, TmPep1050 will be recombinantly produced in *E. coli* and purified to homogeneity using conventional purification techniques. Purified proteins will be characterized by different methods: activity assays, metal ion quantification, thermal shift assays, molecular size determination by native MS, and X-ray crystallography. The main objectives in studying TmPep1050 will be to obtain stable dimers and dodecamers, to understand the dimer-dodecamer transition, and to solve the structure of both oligomers. Ultimately it would allow to better understand the molecular determinants behind oligomerization. The two other *T. maritima* M42 aminopeptidases, TmPep1048 and TmPep1049, will be studied following the same methodology as that of TmPep1050. Both enzymes could have a substrate specificity complementary to TmPep1050. Thus, comparing the structure of their substrate binding pockets will provide valuable information about substrate recognition.

Regarding *E. coli* as a model, its main advantage is the availability of proven genetic tools and transcriptomic data. Consequently, studying the M42 aminopeptidases of *E. coli* could reveal their physiological functions. Their genes will be deleted, and phenotypes will be sought in the light of transcriptomic studies. The three aminopeptidases will also be characterized according to the same approach than that used for *T. maritima* enzymes. They are expected to have different substrate specificities, like the *P. horikoshii* M42 aminopeptidases, to complete peptide hydrolysis in their assumed role linked to the proteasome. Their cellular localization will also be investigated since several M42 aminopeptidases have been reported to be, at least partly, associated to the membrane. Finally, the formation of heterocomplexes will be studied using the *E. coli* enzymes as a model. Their existence would mean that such "peptidasomes" are widespread among prokaryotes.

Results and discussion

B.1 How metal cofactors drive dimer-dodecamer transition of the M42 aminopeptidase TmPep1050 of *Thermotoga maritima*

B.1.1 Background

The quaternary structure of M42 aminopeptidases is dependent on divalent metal ion availability. The metal ion cofactor is known to trigger the transition of dimers or monomers to dodecamers[268,382,420,423]. De-oligomerization of dodecamers is achieved by removing metal ions under certain experimental conditions[184,365,382,420]. Several aspects of the dodecamer assembly process remain either unclear or ambiguous. The identity of the physiological relevant building block is still debated, as both monomers and dimers have been reported in de-oligomerization experiments[184,325,365,382,420]. On the one hand, PhTET2 and PfTET3 can de-oligomerize into monomers under chelating conditions[365,420]. While PhTET2 monomers have been shown to be unfolded, PfTET3 monomers keep about 30% of relative activity. On the other hand, no monomers of PhTET1, PhTET2, and PhTET3 have been observed *in vivo* in *P. horikoshii*[268]. Subsequently, one may suspect that monomers could not be physiologically relevant.

The two metal ions found in the active site seem to have distinct roles in the oligomerization. The metal ion of the M1 site has been shown to stabilize the quaternary structure of PfTET3 dodecamers[365]. Its absence, however, led to a partial dissociation of 40% of PfTET3 dodecamers. The metal ion of the M2 site seems to be enough to maintain the dodecameric structure. The M2 site has also been described to have a higher affinity for metal ions than the M1 site[265,365]. Consequently, the metal ion of the M2 site has been proposed to be catalytically active rather than that of the M1 site, challenging the catalytic mechanism of VpAp1. Thus, dimers (or monomers) are expected to exist with a loaded M2 site and be active. Appolaire *et al.* (2013) showed that PhTET2 dimers are active on dipeptides only[419]. Therefore, the dodecamer formation seems to be required to process longer peptides. Yet, the distinct roles of the M1 and M2 sites remain ill-described with just one study published about that topic.

Another caveat is the lack of high-resolution structure of a dimer. According to the current literature, no major structural rearrangement is supposed to happen during dimer-dodecamer transition. Indeed, the subunit structure of a PhTET2 dimer is believed to be the same as that of a PhTET2 dodecamer[419,420]. The current hypothesis, however, relies only on a low resolution structure of PhTET2 dimer obtained by SAXS. Thus, a high-resolution structure of a dimer would show whether any structural rearrangement occurs during dodecamer assembly. Few structures of M42 aminopeptidase monomers or dimers are available in the Protein Data Bank but it would be hazardous to draw any conclusion from them due to either the lack of biochemical characterization or the presence of a hexa-histidine tag[263]. The latter could introduce an oligomerization artefact since it coordinates metal ions, probably interfering in metal ion binding within the catalytic site.

Previously, we characterized TmPep1050, an M42 aminopeptidase of *T. maritima*[323]. In our study, due to the presence of a hexa-histidine tag, TmPep1050 was purified as a dimer displaying a weak leucyl-aminopeptidase activity. We decided to study the native enzyme devoid of any purification tag. Preliminary results showed that native TmPep1050 is a dodecamer exhibiting a much

higher activity than the histidine tagged version. This first hint prompted us to further characterize TmPep1050 which could be a relevant model to study the dimer-dodecamer transition of M42 aminopeptidases. We succeeded in solving the structure of both the dodecamer and the dimer of TmPep1050. Consequently, several structural rearrangements were identified, revealing how the metal ions intervene in stabilizing the dodecamer structure. The work has been published in the *Journal of Biological Chemistry* (JBC) in October 2019.

B.1.2 Methodology

Basically, we used a conventional approach to study the dimer-dodecamer transition of TmPep1050. Our methodological approach has been described in length in a second article (see **section B.2**, page 68). In short, the enzyme was produced recombinantly in *E. coli* and purified to homogeneity without using an affinity tag. The molecular weight of recombinant TmPep1050 was determined by size exclusion chromatography and native MS. The specific activity was determined using amino acid *p*-nitroanilide derivatives, whose amide bound mimics a peptide bound. Several crystallization conditions were identified using various crystallization kits. Two conditions were optimized to get crystals suitable for X-ray diffraction. For data collection, crystals were diffracted at BM30a (ESRF synchrotron) and Proxima 2 (SOLEIL synchrotron) beamlines. The structures were solved classically by molecular replacement. The apo-enzyme can be prepared in three different ways using either malic acid or 1,10-phenanthroline as chelating agent or an extended dialysis. In the article, we used only the results obtained with the malic acid treated TmPep1050 because one of the crystallization conditions contains malic acid. The dimer-dodecamer transition was studied by coupling different methods: size-exclusion chromatography, cobalt quantification, thermal-shift assays, and native MS. Stable dimers of TmPep1050 were obtained by simply heating the apo-enzyme. The dimer was subsequently characterized using the same methods as for the dodecamer. Two variants were engineered by single-point mutagenesis to study the role of the M1 and M2 sites. Both variants were characterized by native MS and X-ray crystallography.

B.1.3 Article summary and discussion

Two structures were solved for TmPep1050 dodecamers: the apo-form and the complex with Zn^{2+} and Co^{2+} (hereafter named apo-TmPep1050$_{12\text{-mer}}$ and TmPep1050$_{12\text{-mer}}$, respectively). Apo-TmPep1050$_{12\text{-mer}}$ probably resulted from malic acid used in the crystallization condition, which is considered as a mild chelating agent[436]. Observing a dodecameric structure for apo-TmPep1050$_{12\text{-mer}}$ seemed unusual as metal ion chelation often leads to the dodecamer disassembly[365,382]. Despite the role of metal ions in the oligomerization, the dodecamer disassembly may require an energy input, especially for thermophilic enzymes. A similar observation was made by Roncari *et al.* (1972) for GsApI: the disassembly of the apo-GsApI was achieved only after a heat treatment[372]. In addition, the structure of ADPkam589 has been reported as an apo-form dodecamer[378]. Besides, the structure of apo-TmPep1050$_{12\text{-mer}}$ appeared to be nearly identical to that of TmPep1050$_{12\text{-mer}}$. Some slight differences, however, were observed in their catalytic sites. As expected, the quaternary structure of TmPep1050$_{12\text{-mer}}$ adopts the same genuine structure as other characterized M42 aminopeptidases. It consists of six dimers spatially organized to form a tetrahedron-shaped particle. Nevertheless, the structure of TmPep1050$_{12\text{-mer}}$ unraveled a new structural feature in the PDZ-like domain. As explained in **section A.5.3.1** (page 31), a gap has been reported in the PDZ-like domain for all M42 aminopeptidase structures published so far. In PhTET2, this gap is located between the βd strand and

the αb helix (see **Figure 39**, page 31). In the TmPep1050$_{12\text{-mer}}$ structure, the α4 helix filled that gap and could have two roles. The α4 helices are positioned in such a manner that the entrance (13 Å of diameter) found at each tetrahedron face are narrower than that of other M42 aminopeptidases (18-20 Å of diameter). Consequently, an open/closed conformation of the entrances could happen as the α4 helix seems to be flexible. Thus, we inferred a first role for this helix: the control of the inner chamber accessibility. Furthermore, at the dimer level, the α4 helix of one subunit was found delineating the catalytic pocket of the other subunit. Subsequently, it could fulfil a second role in substrate binding. During the revision of our article, Gauto *et al.* (2019) observed the α4 helix in the PDZ-like domain of PhTET2[427]. In their study, the authors reported the structure of PhTET2 in solution using NMR and cryo-EM. Interestingly, almost no NMR signal was obtained at room temperature for this segment of the PhTET2 dimerization domain while the backbone could be seen in the cryo-EM density. Such an observation strongly supports our hypothesis on the function of the α4 helix.

The structures of apo-TmPep1050$_{12\text{-mer}}$ and TmPep1050$_{12\text{-mer}}$ also provided some hints about metal ion binding and the catalytic mechanism of M42 aminopeptidases. The comparison of both structures showed that the metal ions impose some constraints on their ligand residues, bringing them closer to the reaction center. In the TmPep1050$_{12\text{-mer}}$ structure, the water molecule thought being involved in peptide bond hydrolysis is positioned asymmetrically between both metal ions. In other M42 aminopeptidase structures or in VpAp1, this water molecule is symmetric to the metal ions. Our data suggests that Zn^{2+} of the M2 site could activate the water molecule, which is in accordance with the proposed catalytic mechanism of VpAp1 (see **section A.5.2.2**, page 26). Activity assays showed that TmPep1050 is a strict leucyl-aminopeptidase activated by Co^{2+}. Compared to our previous study, we observed a 550-fold increase of k_{cat} against L-Leu-*p*NA. Co^{2+} is bound with an apparent dissociation constant of 50 µM. The most remarkable trait of TmPep1050 is its thermostability, with half-lives at 95°C and 75°C of 24 hours and 20 days, respectively. When metal ions are removed by extended dialysis against malic acid, the half-life at 75°C dropped to only one hour. This dramatic loss of activity was accompanied with the disassembly of dodecamers into dimers. The phenomenon was further explored to confirm that the dimer-dodecamer transition of TmPep1050 is driven by Co^{2+}. Adding Co^{2+} to dimers restored the activity and thermostability of TmPep1050. In the process, the ratio of dodecamer-to-dimer increased with the concentration of Co^{2+} accordingly. Native MS experiments confirmed that the dimer-dodecamer transition is dependent on Co^{2+}. Moreover, several transition oligomers were identified: tetramer, hexamer, and octamer. As no decamer was observed, our study does not support the stochastic model proposed by Macek *et al.* (2017)[420]. Still, due to the scarcity of data, it would be hazardous to exclude or acknowledge any hypothesis (stochastic, hexamer-based, or hybrid).

The structure of the TmPep1050 dimer (hereafter named TmPep1050$_{2\text{-mer}}$) was solved at high resolution. The overall structure is highly similar to that of TmPep1050$_{12\text{-mer}}$ although with some important differences. Especially, the active site of TmPep1050$_{2\text{-mer}}$ undergoes major rearrangements following the dodecamer disassembly. More particularly, the catalytic residue, Glu-197, and a metal ion ligand of the M1 site, Glu-198, are displaced far away from the catalytic center. The α carbons of Glu-197 and Glu-198 are at, respectively, 6 Å and 7 Å from their normal position in the dodecamer structure. Their side chains are even oriented outward the active site. Two other metal ion ligands, Asp-220 of the M2 site and His-307 of the M1 site, have their side chain displaced compared to the active site in the TmPep1050$_{12\text{-mer}}$ structure. Besides these rearrangements in the active site, the α8

and α10 helices, corresponding to the α4 and α7 helices in PhTET2 (see **Figure 37**, page 30), are so disordered in the dimer that they cannot be seen in the electron density. A native MS experiment showed that TmPep1050$_{2\text{-mer}}$ fragmented preferentially within the α8 and α10 helices, proving their high flexibility. These helices have been reported to be part of the exit channel. Furthermore, two loops involved in dimer-to-dimer interaction are misplaced in the TmPep1050$_{2\text{-mer}}$ structure. To demonstrate their role in oligomerization, three key residue – Lys-232, Arg-233, and Arg-249 – were substituted to glutamate residues. The resulting variant lost the ability to form dodecamer, as only dimers and tetramers were observed.

Finally, we investigated the role of the M1 and M2 sites in oligomerization. To achieve that, His-60 and His-307 were substituted to alanine residues independently. His-60 and His-307 are involved in metal ion binding in the M2 and M1 sites, respectively. The substitution of His-60 provoked an alteration in the dodecamer formation as both dimers and dodecamers were observed for this variant. This result was expected as it correlates with the role of the M2 site proposed by Colombo *et al.* (2016)[365]. Unexpectedly, the substitution of His-307 completely abolished the dodecamer formation as the variant is a strict dimer. Thus, the oligomerization of TmPep1050 is strictly controlled by the M1 site. Conversely, PfTET3 tends to disassemble partly into monomers when the M1 site metal ion was removed. The complete disassembly of dodecamers was achieved only when both sites were depleted of their metal ions[365]. Consequently, whether the oligomerization is strictly controlled by the M1 site remains uneasy to settle.

For TmPep1050, the metal ion binding affinity of the M1 site could not be determined experimentally since no biphasic dose-response curve was observed during the Co^{2+} binding assays. Based on the structural data, the M2 site seems to have a higher affinity than the M1 site for metal ion binding. Indeed, Zn^{2+} is found in the M2 site of TmPep1050 and, according to the Irving-Williams series, zinc has a higher affinity than cobalt to occupy a metal ion binding site. In addition, the M2 site of TmPep1050 is occupied at 87% while the M1 site is occupied at 34%. Thus, our structural data are in accordance with Colombo *et al.* (2016) but our results with TmPep1050$_{H60A}$ and TmPep1050$_{H307A}$ suggest the contrary. One should be cautious when interpreting structural data since X-ray radiation damage may occur. When collecting a data set at a wavelength near the metal absorption edge, the metal ion will absorb more energy and may oxidize, resulting in an anomalous signal loss[437,438]. Furthermore, the dimer and dodecamer could have different binding affinities for the M1 and M2 sites. Henceforth, other experiments will be required to disambiguate their respective roles in the oligomerization. The Co^{2+} binding assays also showed that TmPep1050 could bind more than two Co^{2+}, suggesting a potential third site like in PfTET3[365].

B.1.4 The article with supplementary data

JBC ARTICLE

How metal cofactors drive dimer–dodecamer transition of the M42 aminopeptidase TmPep1050 of *Thermotoga maritima*

Received for publication, May 9, 2019, and in revised form, September 24, 2019 Published, Papers in Press, October 14, 2019, DOI 10.1074/jbc.RA119.009281

Raphaël Dutoit[‡§1], Tom Van Gompel[¶], Nathalie Brandt[§], Dany Van Elder[‡], Jeroen Van Dyck[¶], Frank Sobott[¶‖], and Louis Droogmans[‡]

From the [‡]Laboratory of Microbiology, Department of Molecular Biology, Université Libre de Bruxelles, rue des Professeurs Jeener et Brachet 12, B6041 Charleroi, Belgium, [§]Labiris Institut de Recherche, avenue Emile Gryzon 1, B1070 Brussels, Belgium, [¶]Biomolecular and Analytical Mass Spectrometry, Department of Chemistry, Universiteit van Antwerpen, Groenenborgerlaan 171, B2020 Antwerpen, Belgium, and [‖]Astbury Centre for Structural and Molecular Biology, University of Leeds, LS2 9JT Leeds, United Kingdom

Edited by Karen G. Fleming

The M42 aminopeptidases are dinuclear aminopeptidases displaying a peculiar tetrahedron-shaped structure with 12 subunits. Their quaternary structure results from the self-assembly of six dimers controlled by their divalent metal ion cofactors. The oligomeric-state transition remains debated despite the structural characterization of several archaeal M42 aminopeptidases. The main bottleneck is the lack of dimer structures, hindering the understanding of structural changes occurring during the oligomerization process. We present the first dimer structure of an M42 aminopeptidase, TmPep1050 of *Thermotoga maritima*, along with the dodecamer structure. The comparison of both structures has allowed us to describe how the metal ion cofactors modulate the active-site fold and, subsequently, affect the interaction interface between dimers. A mutational study shows that the M1 site strictly controls dodecamer formation. The dodecamer structure of TmPep1050 also reveals that a part of the dimerization domain delimits the catalytic pocket and could participate in substrate binding.

Cells possess an arsenal of proteolytic enzymes to ensure specific and nonspecific hydrolysis of proteins and peptides (for instance, up to 2% of the human genome encodes proteases) (1). Proteolysis plays an important role in a wide array of cellular functions as diverse as homeostasis, stress responses, cell-cycle regulation, quorum sensing, stalled ribosome rescue, immune response, virulence, etc. (2–6). About 80% of protein degradation is achieved by the proteasome 26S in eukaryotes (3). Protein turnover is carried out by either the proteasome 20S in archaea and actinomycetes or several proteolytic complexes, notably HslUV, ClpAP, ClpXP, Lon, and FtsH, in other prokaryotes (4, 7). The proteasome and its related complexes generate peptides 6–9 amino acids in length (8) that are further processed by a set of endopeptidases and exopeptidases. Several peptidases have been proposed to act downstream of the proteasome: the tricorn protease (9), thimet oligopeptidase (10), tripeptidyl peptidase II (11), aminopeptidase N (12), and TET[2] aminopeptidases (13). The latter peptidases are widely distributed in all kingdoms of life (14). They adopt a peculiar tetrahedron-shaped structure compartmenting the active sites in a buried catalytic chamber (15). In *Pyrococcus horikoshii*, four TET aminopeptidases, PhTET1, PhTET2, PhTET3, and PhTET4, have been described, and each of them displays a different substrate specificity, aspartyl, leucyl, lysyl, and glycyl aminopeptidase activity, respectively (15–18). Remarkably, heterocomplexes, made of PhTET2 and PhTET3, have been reported, leading to the assumption of the existence of a peptidasome particle (19, 20).

According to MEROPS classification, the TET aminopeptidases are found in the M18 and M42 families (14, 21). Both families belong to the MH clan, encompassing metallopeptidases sharing a common α/β catalytic domain. The archetypal MH clan enzyme is the *Vibrio proteolyticus* aminopeptidase 1 (22). The M18 family is widely distributed in all domains of life (14), whereas the M42 family is unique to prokaryotes (23). Several structures of archaeal M42 aminopeptidases have been studied (13, 15, 17, 18, 24–26), but only one structure has been reported for bacteria (27). Franzetti *et al.* (25) described the first TET aminopeptidase structure of the *Haloarcula marismortui* M42 aminopeptidase, consisting of 12 subunits adopting a tetrahedron-shaped quaternary structure. High-resolution structures of *P. horikoshii* PhTET2 have revealed that the M42 aminopeptidase subunit is composed of an α/β catalytic domain and a PDZ-like dimerization domain (13, 15). A similar quaternary structure has been reported for several M18 aminopeptidases from prokaryotes and eukaryotes (28–31), although their dimerization domain adopts a butterfly fold instead of a PDZ-like fold. The quaternary structure is often seen as the assembly of six dimers with a dimer positioned on each edge of the tetrahedron (13). At the center of the tetrahedron faces are four

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

This article contains Figs. S1–S9 and Tables S1–S4.

The atomic coordinates and structure factors (codes 4P6Y, 5NE6, 5NE7, 5NE8, and 6NW5) have been deposited in the Protein Data Bank (http://wwpdb.org/).

[1] To whom correspondence should be addressed. Tel.: 32-2-526-72-83; E-mail: rdutoit@ulb.ac.be.

[2] The abbreviations used are: TET, tetrahedral shape; r.m.s.d., root-mean-square deviation; *p*NA, *p*-nitroanilide; LAP, leucyl aminopeptidase; *k*, specific activity; k_{cat}, turnover number; T_m, melting temperature; V_e, elution volume; [Co^{2+}], Co^{2+} concentration; Tm, *T. maritima*; Ph, *P. horikoshii*; Sp, *Streptococcus pneumoniae*; CV, column volumes; PDB, Protein Data Bank; MR, molecular replacement; Pf, P.furiosus.

Cobalt-driven dimer–dodecamer transition of TmPep1050

gates leading to a wide inner cavity. Four exit pores are located at the tetrahedron vertices through which amino acids, generated during peptide hydrolysis, are released.

The catalytic site is characterized by (i) two catalytic residues, (ii) five residues forming the metal ion–binding sites M1 and M2, and (iii) two divalent metal ions (usually Zn^{2+}, Co^{2+}, and Mg^{2+}). The metal ions have been described to bind the substrate, to facilitate nucleophile generation (water molecule deprotonation), and to stabilize the transition state of peptide hydrolysis (32). In addition to their catalytic roles, the metal ions could also control the TET aminopeptidase oligomerization. Such a structural role has been reported for three M42 aminopeptidases, PhTET2 and PhTET3 from *P. horikoshii* (13, 33, 34), and PfTET3 from *Pyrococcus furiosus* (24). Under chelating conditions, dodecamers disassemble into either dimers (PhTET3) or monomers (PhTET2 and PfTET3). Colombo *et al.* (24) inferred the role of the M1 and M2 sites in the oligomerization: the presence of a metal ion in the M2 site is required for the oligomerization, whereas the M1 site controls protein flexibility. According to Macek *et al.* (33), the dissociation is reversible as dodecamers are formed through random association of dimers. The understanding of the dodecamer formation mechanism, however, suffers from the lack of dimer (or monomer) structure. A low-resolution structure of PhTET2 dimer has been reported, but no significant structural change has been pinpointed (35). The current knowledge on M42 aminopeptidases relies on a set of archaeal enzymes, whereas their bacterial counterparts remain ill-described.

In this work, we focus on the bacterial M42 aminopeptidase model, TmPep1050 from *Thermotoga maritima*, previously characterized as a cobalt-activated leucyl aminopeptidase (23). We present the first high-resolution dimer structure of an M42 aminopeptidase alongside the dodecamer structure. The dimer structure clearly showed the cofactor role in the active-site fold that reverberates the interaction interfaces between dimers. In addition, the thermostability of TmPep1050 depends on the oligomeric state. The dodecamer assembly/disassembly was further studied by native MS, revealing the intermediate oligomeric states as well as the flexible parts in the dimer structure. Finally, we inferred the role of the M1 and M2 sites through the structural study of TmPep1050 variants. Our results showed that the M1 site strictly controls the dodecamer assembly. Based on the current knowledge, we should avoid drawing general dogma on the TET aminopeptidases due to the lack of studies on metal ion binding behavior and structures in complex with substrates and inhibitors.

Results and discussion

TmPep1050 has a genuine TET aminopeptidase structure

The *Tm_1050* open reading frame (ORF) was overexpressed in *Escherichia coli*, and the recombinant enzyme was purified to homogeneity through three chromatographic steps. Size-exclusion chromatography showed that the purified enzyme, subsequently named $TmPep1050_{12\text{-mer}}$, had an apparent molecular weight of 330 ± 15 kDa (S.E. with $n = 12$; molecular weight of a monomer, 36.0 kDa). Native MS analysis demonstrated that $TmPep1050_{12\text{-mer}}$ is a dodecamer with a measured mass of 435,323 ± 64 Da (theorical mass, 432,783 Da). Crystals were obtained in two crystallization conditions (see Table S1). X-ray fluorescence scanning detected Co^{2+} and Zn^{2+} in crystals grown in 0.18 M triammonium citrate, 40% PEG3350, pH 7.5, whereas no trace of metal ions was found in crystals grown in 2.1 M malic acid, pH 6.75 (see Fig. S1). The absence of metal ions could be due to the use of malic acid, a weak chelating agent of divalent metal ions (K_d, 10^{-3} M for Co^{2+} binding) (36), in the crystallization buffer. Under such conditions, $TmPep1050_{12\text{-mer}}$ was barely active, although dodecameric oligomerization remained unaffected at room temperature (see below). A similar statement has been reported for PhTET2 (13): when cocrystallized with *o*-phenanthroline, PhTET2 remained dodecameric even if the two Zn^{2+} were chelated. The apoenzyme structure (apo-$TmPep1050_{12\text{-mer}}$) was determined by molecular replacement using the coordinates of YpdE, an uncharacterized M42 aminopeptidase of *Shigella flexneri*. The metal-bound complex structure was determined by MR and single-wavelength anomalous diffraction using apo-$TmPep1050_{12\text{-mer}}$ coordinates.

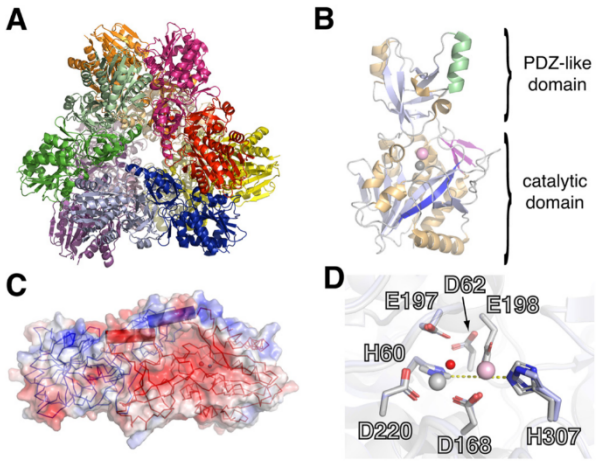


Figure 1. The $TmPep1050_{12\text{-mer}}$ structure. *A*, schematic representation of $TmPep1050_{12\text{-mer}}$ quaternary structure (PDB code 6NW5) centered on one of the tetrahedron faces. As the asymmetric unit contains four monomers, the quaternary structure was reconstituted by PDBePisa. *B*, schematic representation of $TmPep1050_{12\text{-mer}}$ subunit. The α4 helix, β16 strand, and β sheet extension (β10, β11, and β17) are colored in *green*, *dark blue*, and *purple*, respectively. Zn^{2+} and Co^{2+} are displayed as *gray* and *pink spheres*, respectively. *C*, schematic representation of subunits A and B (*red* and *blue*, respectively) composing a dimer in the $TmPep1050_{12\text{-mer}}$ structure. The subunit A catalytic pocket is indicated by an *arrow*, the α4 helices are schematized as *cylinders*. *D*, structural alignment of the five metal ion–binding residues and the two catalytic residues of $TmPep1050_{12\text{-mer}}$ (in *light gray*) and apo-$TmPep1050_{12\text{-mer}}$ (PDB code 4P6Y; in *light blue*). Zn^{2+}, Co^{2+}, and H_2O are represented as *gray*, *pink*, and *red spheres*, respectively. The distance between His-60 and His-307 Nϵ2 is about 6.7 Å in the $TmPep1050_{12\text{-mer}}$ structure. In the apo-$TmPep1050_{12\text{-mer}}$ structure, this distance is increased to about 8.2 Å.

The $TmPep1050_{12\text{-mer}}$ quaternary structure consists of 12 subunits adopting a tetrahedron-shaped architecture (see Fig. 1*A*) like other available structures of M42 aminopeptidases (13, 17, 18, 24, 26, 27). The tetrahedron-shaped architecture is often described as the self-assembly of six dimers in such a manner that a dimer lies on each tetrahedron edge. Four entrances are

located at the center of the faces, and four exit channels are located at the vertices. The entrances and exits lead to an inner cavity with the 12 catalytic sites oriented inward, compartmenting the active sites. The interaction between dimers is maintained through a polar and hydrophobic interaction network at the vertices (see Fig. S2*A*) and nine salt bridges between adjacent subunits (see Fig. S2*B*). The residues forming these salt bridges are, however, not conserved among all M42 aminopeptidases. The apo-TmPep1050$_{12\text{-mer}}$ structure appears to be nearly identical to the TmPep1050$_{12\text{-mer}}$ structure obtained with bound metal ions (structural alignment r.m.s.d., 0.235Å). There are, however, several differences observed in the catalytic site, and they are described below.

The α4 helix of the PDZ-like domain delimits the TmPep1050$_{12\text{-mer}}$ catalytic pocket

The TmPep1050$_{12\text{-mer}}$ subunit is composed of a catalytic domain and a dimerization domain (Fig. 1*B*). The catalytic domain adopts an α/β globular structure similar to that of *V. proteolyticus* aminopeptidase Ap1 (see Fig. S3) except that the β sheet displays an extension of three antiparallel strands (β10, β11, and β17) connected to the β16 strand. The β sheet extension is conserved among the M42 family (13, 15, 21) and interacts with the neighbor subunit dimerization domain (13, 27). The TmPep1050$_{12\text{-mer}}$ dimerization domain adopts the typical fold of a PDZ-like domain (13). In all structurally characterized M42 aminopeptidases, the α4 helix of the PDZ-like domain is highly flexible (13, 15, 17, 18, 24, 26, 27). The α4 helix, however, was modeled in TmPep1050$_{12\text{-mer}}$ structure, probably due to its stabilization in the crystal. In the quaternary structure, the α4 helices are positioned in such a manner that the entrances are restricted to 13 Å (see Fig. S4), which is smaller than those reported for PhTET1, PhTET2, and SpPepA (13, 18, 27). The carboxylate functions of three glutamate residues (Glu-110, Glu-114, and Glu-117) are oriented toward the entrance. In addition, the TmPep1050$_{12\text{-mer}}$ active sites appear to be more buried and delimited than those of other characterized M42 aminopeptidases as the α4 helix of one monomer delimits the active-site pocket of the other monomer at the dimer level (see Figs. 1*C* and S5*A*). Hence, the α4 helix flexibility could indicate a possible open/closed conformation of the enzyme. It is worth noticing that an α helix of the butterfly-fold dimerization domain of the M18 aminopeptidases could also be highly flexible (30, 31) or, to the contrary, more stable (28, 29). Such an α helix, flexible or not, seems to be a conserved structural feature among TET aminopeptidases. In the human M18 aminopeptidase structure, this α helix is directly involved in substrate binding (28). A similar function could be expected for the α4 helix in M42 aminopeptidases.

Seven residues conserved in the MH clan define the TmPep1050$_{12\text{-mer}}$ active site

The M42 aminopeptidase catalytic site consists of two divalent metal ions (M1 and M2) bound by five strictly conserved residues and two catalytic residues. In the TmPep1050$_{12\text{-mer}}$ catalytic site, the M1 site consists of Asp-168, Glu-198, and His-307 residues bound to a Co^{2+} ion, whereas the M2 site consists of His-60, Asp-168, and Asp-220 residues bound to a Zn^{2+} ion (see Fig. 1*D*). In the apo-TmPep1050$_{12\text{-mer}}$ structure, these residues are correctly positioned despite the absence of metal ions. The distance between the imidazole rings of His-60 and His-307 is, however, increased by 1.5 Å, whereas the carboxylate of Glu-198 is displaced by 0.4 Å away from the M1 site center compared with the TmPep1050$_{12\text{-mer}}$ structure with its metallic cofactors. Thus, the presence of metal ions seems to pull the side chains of metal-binding residues closer, especially those of the M1 site (see Fig. 1*D*). The two metal ions also coordinate a water molecule involved in peptide bond hydrolysis. The water molecule is asymmetrically positioned, being closer to the M2 site rather than the M1 site (see Fig. S5*B*). In the monomer D, the distance between Zn^{2+} and the oxygen atom is so short that only a hydroxide ion could be modeled instead of a water molecule, suggesting that the M2 site stabilizes the hydroxide ion prior to the nucleophilic attack. The TmPep1050$_{12\text{-mer}}$ structure contrasts with *V. proteolyticus* Ap1, PhTET1, and PhTET2 where the water molecule is positioned symmetrically (see Fig. S5*B*). In the TmPep1050$_{12\text{-mer}}$ catalytic site, the acid/base catalyst is Glu-197, which is conserved in all MH clan members. Indeed, the mutation of Glu-197 to glutamine completely abolished the activity (specific activity of L-Leu-*p*NA of less than 0.1 s^{-1}), whereas the oligomeric state remained unaffected. The seventh conserved residue is Asp-62, which has been described as a modulator of Lewis acid strength of M2 in *V. proteolyticus* Ap1 (37). The TmPep1050$_{12\text{-mer}}$ S1-binding pocket is similar to that of PhTET2 because six of seven residues are conserved (see Fig. S5*C*), which is in accordance with their substrate specificities as both enzymes are leucyl aminopeptidases.

Table 1
TmPep1050$_{12\text{-mer}}$ aminopeptidase activity against L-aminoacyl-pNA derivatives
k, specific activity; ND, not detectable ($k < 0.05\ s^{-1}$).

Substrates	*k*
	s^{-1}
L-Leu-*p*NA	118.3 ± 3.6
L-Ile-*p*NA	42.9 ± 1.6
L-Val-*p*NA	7.2 ± 0.8
L-Met-*p*NA	6.3 ± 0.2
L-Phe-*p*NA	0.9 ± 0.02
L-Ala-*p*NA	0.1 ± 0.01
L-Glu-*p*NA	ND
Gly-*p*NA	ND
L-His-*p*NA	ND
L-Lys-*p*NA	ND
L-Pro-*p*NA	ND
Ac-Leu-*p*NA	ND

TmPep1050$_{12\text{-mer}}$ is a cobalt-activated leucyl aminopeptidase

TmPep1050$_{12\text{-mer}}$ aminopeptidase activity was assayed with various L-aminoacyl-*p*-nitroanilide (*p*NA) derivatives. The substrate specificity is mainly toward nonpolar aliphatic L-aminoacyl-*p*NA, with a clear preference for L-Leu-*p*NA (Table 1). TmPep1050$_{12\text{-mer}}$ activity is maximal at pH between 7.0 and 7.8 and up to 90 °C (see Figs. S6 and S7). Its kinetic parameters, k_{cat} and K_m, were determined for L-Leu-*p*NA, L-Ile-*p*NA, and L-Met-*p*NA (Table 2). These values differ greatly from our previous results obtained with His-tagged TmPep1050 (23), indi-

Cobalt-driven dimer–dodecamer transition of TmPep1050

Table 2
TmPep1050$_{12\text{-mer}}$ kinetic parameters
The ratio k_{cat}/K_m represents the catalytic efficiency.

Substrates	K	k_{cat}	k_{cat}/K
	μM	s^{-1}	$s^{-1}\ M^{-1}$
L-Leu-*p*NA	1750 ± 250	138 ± 12	7.9×10^4
L-Ile-*p*NA	1100 ± 100	53 ± 4	4.8×10^4
L-Met-*p*NA	1750 ± 100	16 ± 1	9.1×10^3

cating that the polyhistidine tag could interfere in the binding of divalent metal ions to the M1 and M2 sites. Indeed, we reported a k_{cat} of 0.25 s^{-1} for His-tagged TmPep1050, even with a Co^{2+}-to-enzyme ratio of 500.

Dialysis of TmPep1050$_{12\text{-mer}}$ against 2.1 M malic acid, pH 7.0 (one of the crystallization conditions), resulted in a loss of activity (dropping to 1 s^{-1}), whereas its oligomeric state remained unaffected as determined by gel filtration with an apparent molecular weight of 300 ± 10 kDa (S.E. with $n = 3$). The addition of Co^{2+} partly restored TmPep1050$_{12\text{-mer}}$ LAP activity, whereas Zn^{2+} had no effect on LAP activity, although it was found in the active site (see Table 3). Similar observations have been reported for other M42 aminopeptidases: although their structures contain Zn^{2+}, PhTET2, PhTET3, PfTET3, and SpPepA are activated by Co^{2+} (13, 17, 24, 27, 38). Other metal ions had no significant effect on TmPep1050$_{12\text{-mer}}$ LAP activity (see Table 3).

To quantify its affinity for Co^{2+}, metal ion–depleted TmPep1050$_{12\text{-mer}}$ was incubated with an increasing amount of Co^{2+}, ranging from 0 to 2560 μM. After 24-h incubation at 75 °C, Co^{2+} binding was quantified with an Amplex UltraRed fluorescent probe. In parallel, LAP activity was measured to follow the reactivation of metal-depleted enzyme (see Fig. 2). Co^{2+} content of metal ion–depleted enzyme was less than one atom per 10 monomers (1.5 μM Co^{2+} for 20 μM proteins). TmPep1050$_{12\text{-mer}}$ can bind Co^{2+} with an apparent association constant of 50 ± 5 μM.

The thermostability and oligomerization state of TmPep1050 are Co^{2+}-dependent

As TmPep1050$_{12\text{-mer}}$ displays a LAP activity up to 90 °C, we expected TmPep1050$_{12\text{-mer}}$ to be a highly thermostable enzyme. Its thermostability was determined at 75 and 95 °C: at 95 °C the half-life of TmPep1050$_{12\text{-mer}}$ is about 24 h, whereas at 75 °C it is about 20 days. TmPep1050$_{12\text{-mer}}$ is highly thermostable, probably due to many ionic and hydrophobic interactions between dimers maintaining the whole quaternary structure in dodecamers. Shorter half-lives (a few hours) have been reported for various M42 aminopeptidases (17, 38–41).

Remarkably, in the absence of its metal ion cofactors, the half-life at 75 °C is dramatically reduced to merely 1 h. Moreover, under these conditions, TmPep1050 dodecamers dissociated to dimers (namely TmPep1050$_{2\text{-mer}}$) after 2-h incubation at 75 °C as an apparent molecular weight of 52 ± 3 kDa (S.E., $n = 12$) was measured by gel filtration. Thermal shift assays were conducted to further characterize the role of cobalt ions in TmPep1050$_{12\text{-mer}}$ thermostability. Cobalt-loaded TmPep1050$_{12\text{-mer}}$ had a T_m of about 97 °C, confirming its remarkable thermostability. After dialysis against 2.1 M malic acid, pH 7.0 (as mentioned above), the T_m dropped to 91 °C; meanwhile Co^{2+} addition restored its thermostability (see Fig. 3).

Table 3
Effect of divalent metal ions on TmPep1050$_{12\text{-mer}}$ LAP aminopeptidase activity
Metal ion–depleted TmPep1050$_{12\text{-mer}}$ was diluted at 20 μM in 50 mM MOPS, 0.5 M $(NH_4)_2SO_4$, pH 7.2, supplemented with 1.28 mM of various metal chlorides. After 24-h incubation at 75 °C, specific activities (k) were measured with L-Leu-*p*NA as substrate.

Metal ion	k
	s^{-1}
Co^{2+}	40.1 ± 1.1
Zn^{2+}	1.9 ± 0.02
Mn^{2+}	1.3 ± 0.02
Mg^{2+}	1.4 ± 0.01
Ni^{2+}	1.1 ± 0.03
Cu^{2+}	1.4 ± 0.01
Ca^{2+}	1.6 ± 0.01

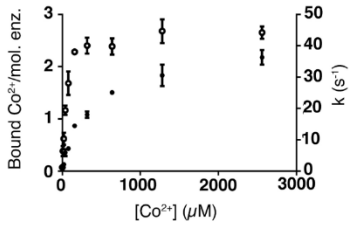


Figure 2. Evolution of bound Co^{2+} per molecule of enzyme (*mol. enz.*) (*open circles*) and LAP activity (*k*; *closed circles*) in response to an increasing Co^{2+} concentration ([Co^{2+}]) ranging from 0 to 2560 μM. *Error bars* represent S.E. with $n = 3$.

The link between the metal cofactors and thermostability has been reported for several M42 aminopeptidases: *Geobacillus stearothermophilus* aminopeptidase I (42), *Thermococcus onnurineus* deblocking aminopeptidase (43), and PhTET3 (34). The dodecamer dissociation has often been achieved in harsh conditions such as EDTA treatment (24, 34, 42) or in acidic buffer (pH < 4) (33). Dimer formation has been reported for PhTET3 (34) and PhTET2 (13), whereas a further breakdown of dimers into monomers has been achieved for PfTET3 (24) and PhTET2 (33). Neither of these studies have, however, described the dodecamer dissociation in physiologically compatible conditions. Although our results focused on TmPep1050$_{12\text{-mer}}$ treated with malic acid, the same dissociation of dodecamers into dimers was achieved with an extensive dialysis against 50 mM MOPS, 0.5 M $(NH_4)_2SO_4$, pH 7.2, followed by a heat treatment at 75 °C (see Table S2). Therefore, our data tend to support a dimer–dodecamer equilibrium depending on metal cofactor availability.

The oligomeric-state transition is reversible and Co^{2+}-driven

Thermal shift assays suggested that dodecamer dissociation could be reversible, but the different oligomeric states occurring during the association/dissociation process had to be identified. For that purpose, we opted for native MS. After a buffer exchange in ammonium acetate, TmPep1050$_{12\text{-mer}}$ retained its oligomeric state (molecular weight, 300 kDa according to gel filtration) but had a LAP activity reduced by two-thirds, suggesting that metal cofactors were partly lost. Native MS showed that, under these conditions, TmPep1050 was mainly a mixture

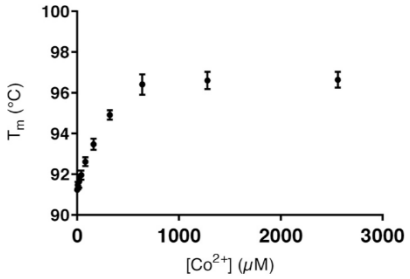


Figure 3. Thermal shift assays of Co^{2+} effect on TmPep1050$_{12\text{-mer}}$ thermostability. Metal ion–depleted TmPep1050$_{12\text{-mer}}$ was diluted to 20 μM in 50 mM MOPS, 0.5 M $(NH_4)_2SO_4$, pH 7.2, supplemented with [Co^{2+}] ranging from 0 to 2560 μM. After 24-h incubation at 75 °C, the thermostability was determined by measuring T_m. *Error bars* represent S.E. with $n = 3$.

of dimers and dodecamers (see Fig. 4*A*). MS/MS analysis of dodecamer +46 peak revealed two masses for dissociated monomer of 36,076.2 ± 1.7 and 36,136.9 ± 4.4 Da (see Fig. S8). The first mass is almost equal to the theoretical molecular weight of 36,065 Da, whereas the second mass could correspond to a monomer with one cobalt ion bound.

When samples were loaded with Co^{2+} and heated at 75 °C for 30 min prior to MS analysis, peaks corresponding to dimers tended to disappear in favor of dodecamer peaks (see Fig. 4*B*). MS analysis of Co^{2+}-reloaded TmPep1050$_{12\text{-mer}}$ showed that monomer had a mass about 176 Da higher than the theoretical molecular weight. Such difference could be explained by the presence of at least two cobalt ions per subunit. In addition, MS experiments showed the existence of intermediate oligomeric states: tetramers, hexamers, and octamers (see Fig. 4*C*). The self-assembly pathway of M42 aminopeptidases is still poorly understood. Appolaire *et al.* (35) proposed that dimers self-assemble into dodecamers via intermediate hexamers for PhTET2. Macek *et al.* (33) debated this theory arguing that dimers self-assemble randomly to form tetra-, hexa-, octa-, and decamer intermediates for PhTET2. Our native MS data suggest that a dodecamer could result from the association of either two hexamers or a tetramer and an octamer. Such a transition of oligomeric states could occur and be controlled *in vivo* due to the low availability of divalent metal ions.

Native MS analysis of TmPep1050$_{2\text{-mer}}$ (Co^{2+}-depleted and heat-treated) revealed peaks corresponding to dimers only (see Fig. 4*D*). The addition of Co^{2+} combined with a heat treatment at 75 °C completely changed the mass spectra. Indeed, dimer peaks became almost undetectable, whereas dodecamers were the most abundant oligomer in the presence of a 6× excess of cobalt (see Fig. 4*D*). In comparison, only partial reassociation was reported for PhTET3 (34) and PhTET2 (35). A second experiment was conducted to quantify the different oligomers occurring during the reassociation. TmPep1050$_{2\text{-mer}}$ was incubated with an increasing concentration of Co^{2+} at 75 °C for 30 min, and the oligomers were identified by size-exclusion chromatography (see Fig. 5*A*). The intermediate oligomers were not detected due to their low abundance (as seen in MS experiments). As expected, the ratio of dodecamers to dimers increased according to the Co^{2+} concentration (see Fig. 5, *A* and *B*).

TmPep1050$_{2\text{-mer}}$ structure highlights the structural changes triggered by metal ion loss

We showed that cobalt ions are important for activity, thermostability, and oligomerization of TmPep1050. Native MS experiments demonstrated the reversibility of dodecamer dissociation and how it depends on metal cofactors. As dimers are unable to form dodecamers in the absence of cobalt, structural changes must occur at oligomerization interfaces. To understand these changes, the structure of TmPep1050$_{2\text{-mer}}$ was solved by X-ray crystallography.

The overall shape of TmPep1050$_{2\text{-mer}}$ does not differ from a dimer in the TmPep1050$_{12\text{-mer}}$ structure (see Fig. 6*A*). However, several structural dissimilarities having a dramatic impact on the oligomerization capability are observed. Two segments are too flexible to be modeled in the TmPep1050$_{2\text{-mer}}$ structure: Gly-203–Gly-208 and Phe-279–Glu-292 corresponding to the α8 and α10 helices, respectively. To rule out a potential crystallographic artifact linked to protein stacking in the crystal, TmPep1050$_{2\text{-mer}}$ backbone flexibility was probed by collision-induced dissociation. Mass spectra showed that proteins were fragmented at two preferential regions, Gly-203–Pro-212 and Asn-283–Thr-290, indicating the soundness of the crystallographic data (see Fig. 6*B*). At each vertex of the tetrahedron-shaped dodecamer, the α8 and α10 helices of three adjacent subunits form the exit tunnel via an interaction network between Tyr-209 and Arg-289 (see Fig. S2*A*). This interaction network has been described to be highly important for oligomerization (21, 35). In the TmPep1050$_{2\text{-mer}}$ structure, interaction between Tyr-209 and Arg-289 is prevented as Tyr-209 is completely buried.

Structural alignment between dimer subunit and dodecamer subunit shows that the Gln-196–Val-202, Lys-229–Ala-235, and Lys-247–Ser-254 segments diverge greatly (see Fig. 6*A*). High B-factors were observed for these segments, indicating that they are highly flexible (see Fig. S9). The Gln-196–Val-202 segment contains two conserved residues of the catalytic site, Glu-197 and Glu-198, which are the catalytic general base and a metal ion–binding residue of the M1 site, respectively. In the TmPep1050$_{2\text{-mer}}$ structure, the Glu-197–Gly-200 loop is so disordered that Glu-197 and Glu-198 side chains point outward from the catalytic site, which explains why the dimers are far less active than the dodecamers (see Fig. 6*C*). Thus, metal ion binding in the M1 site is strongly impaired in TmPep1050$_{2\text{-mer}}$. The displacement of Glu-197 and Glu-198 probably has an impact on His-60 and Asp-62 as their predicted pK_a values increase from 8.5 to 10.1 and from 2.6 to 5.0, respectively. His-60 plays a pivotal role in an H-bond network in the dodecamer subunit, interacting with Asp-62, Asp-168, Asp-169, Glu-197, Glu-198, and Asp-220. In the TmPep1050$_{2\text{-mer}}$ structure, the whole H-bond network is disrupted. In the dodecamer structure, the Gln-196–Val-202 loop probably imposes the fold of the subsequent α8 helix via an entangled network of H-bonds and polar interactions. In addition, the Gln-196–Val-202 segment is closely connected to the α10 helix.

The Lys-229–Ala-235 and Lys-247–Ser-254 segments are important for oligomerization as Lys-232, Arg-233, and Arg-

Cobalt-driven dimer– dodecamer transition of TmPep1050

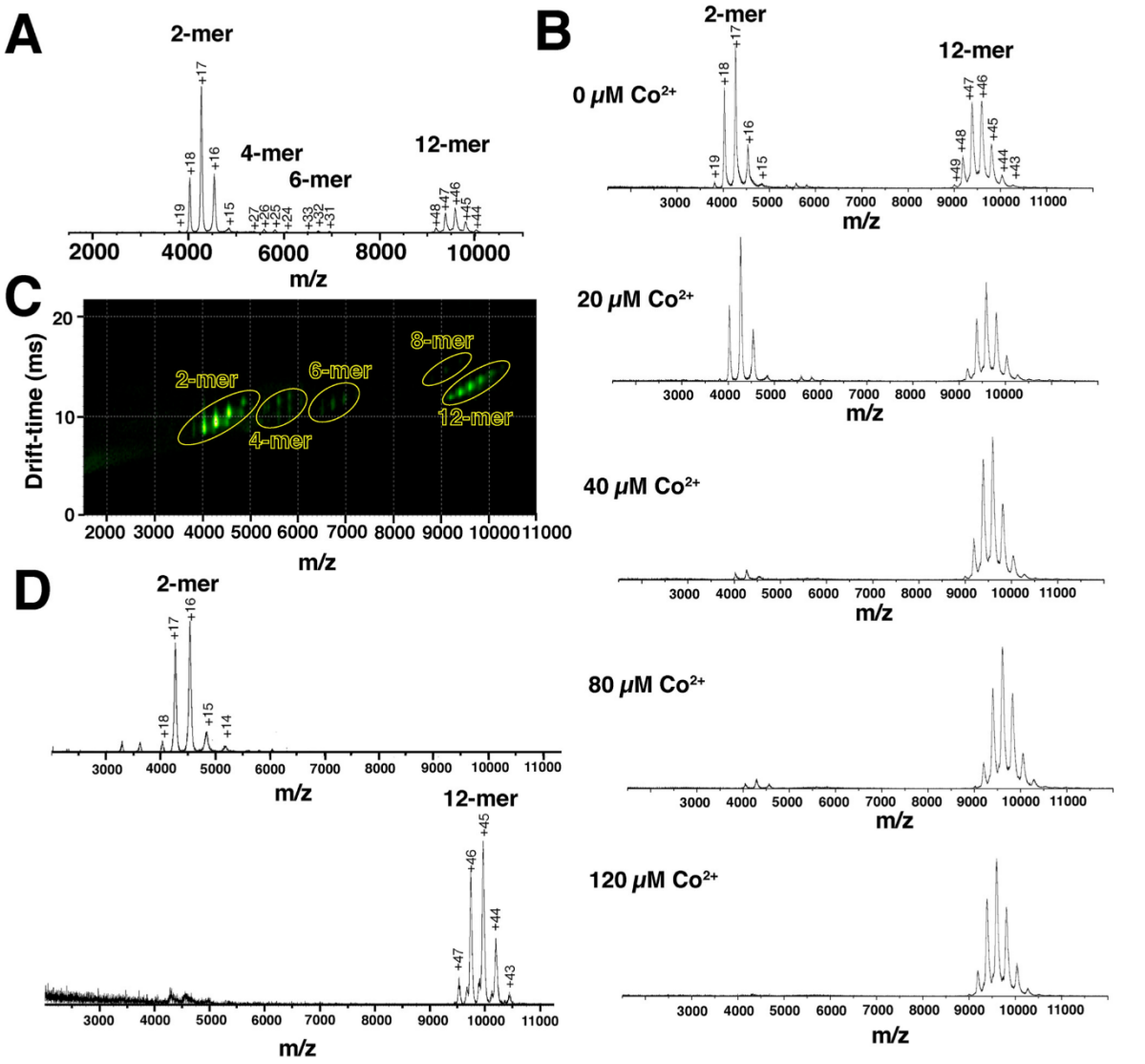


Figure 4. Oligomeric state determination by native MS. *A*, mass spectrum of TmPep1050$_{12\text{-mer}}$ at a concentration of 50 μM (trap collision energy set to 10 V). *B*, mass spectra of TmPep1050$_{12\text{-mer}}$ in the presence of an increasing Co^{2+} concentration. 20 μM enzyme was incubated for 30 min at 75 °C with Co^{2+} at a concentration ranging from 0 to 120 μM. *C*, ion-mobility mass spectra of TmPep1050$_{12\text{-mer}}$ (50 μM). *D*, mass spectra of TmPep1050$_{2\text{-mer}}$ (*upper*) and its reassociation into dodecamers (*lower*). For the reassociation experiment, 20 μM enzyme was incubated with 120 μM Co^{2+} for 30 min at 75 °C.

249 are involved in the formation of salt bridges in the dodecamer (see Fig. S1*B*). Systematic mutagenesis of these residues was set up to support the role of salt bridges in the dodecamer structure. Single point mutations of Lys-232, Arg-233, and Arg-249 to alanine residues did not impact the oligomerization (data not shown). However, the triple mutation to either alanine or glutamate residues (TmPep1050$_{K232A/R233A/R249A}$ and TmPep1050$_{K232E/R233E/R249E}$, respectively) greatly disturbed the formation of dodecamers. Indeed, the oligomeric state of TmPep1050$_{K232A/R233A/R249A}$ is mainly dimeric as shown by gel filtration (see Fig. 7). It also forms dodecamers and tetramers representing about 32 and 20% of the purified sample, respectively. For TmPep1050$_{K232E/R233E/R249E}$, the mutations have a dramatic impact on the oligomerization as the dimeric fraction represents about 88% of the purified sample (see Fig. 7). TmPep1050$_{K232A/R233A/R249A}$ has a reduced specific activity on L-Leu-*p*NA of 37.2 $\pm$ 0.9 s^{-1}, whereas TmPep1050$_{K232E/R233E/R249E}$ is barely active with a specific activity of 0.15 $\pm$ 0.01 s^{-1}. Our results strongly support the role of Lys-232, Arg-233, and Arg-249 to stabilize the dodecameric structure. Nevertheless, one should avoid generalizing such a conclusion to the whole M42 family as these residues are not conserved. Appolaire *et al.* (35) reported the destabilization of PhTET2 by mutating five residues involved in interdimer interactions. However, in their study, the destabilization was only transitory as PhTET2 dimers slowly reassembled into dodecamers over time (35).

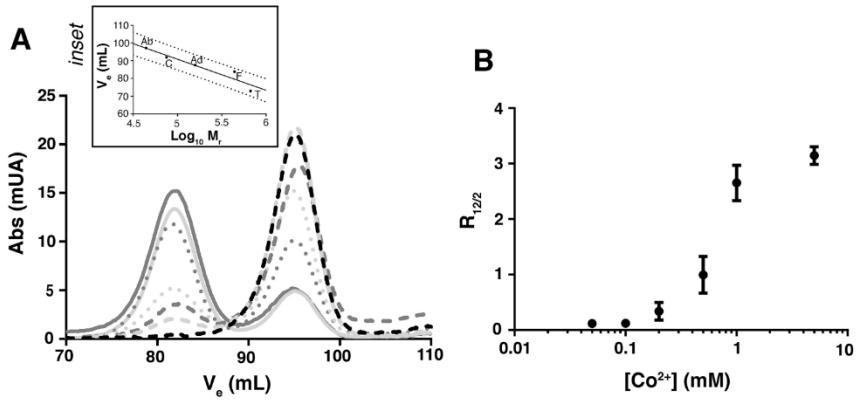


Figure 5. Reassociation of dimers into dodecamers in response to Co^{2+} concentration. $TmPep1050_{2\text{-mer}}$ was incubated with Co^{2+} at a concentration ranging from 0 to 5 mM for 30 min at 75 °C. Oligomers were detected and quantified by gel filtration on a Superdex 200 column (volume of 120 ml). *A*, gel-filtration chromatogram of $TmPep1050_{2\text{-mer}}$ after incubation with increasing Co^{2+} concentrations of 0 (*black dashes*), 0.05 (*light gray dashes*), 0.1 (*gray dashes*), 0.2 (*light gray dots*), 0.5 (*gray dots*), 1 (*light gray line*), and 5 mM (*gray line*). The peak at elution volume (V_e) of 95.0 ml corresponds to dimers, whereas the peak at V_e of 81.8 ml corresponds to dodecamers. *Inset*, the calibration of the Superdex 200 column with albumin (*Ab*), conalbumin (*C*), aldolase (*Ad*), ferritin (*F*), and thyroglobulin (*T*) as standards. The correlation between V_e (ml) and the logarithm of the relative mass (M_r) is linear with R^2 of 0.91. The 95% confidence intervals of the linear regression are shown in *dots*. *B*, ratio of dodecamers to dimers ($R_{12/2}$) after incubating $TmPep1050_{2\text{-mer}}$ with Co^{2+}. Ratios were calculated based on the peak areas. *Abs*, absorbance. *mUA*, milli unit of absorbance. *Error bars* represent S.E. with $n = 3$.

The metal ion cofactors are directly involved in structural changes during dodecamer dissociation

The $TmPep1050_{2\text{-mer}}$ structure shows how the loss of metal ions induces dodecamer dissociation. The structural role of either M1 or M2 sites, however, was not so obvious as Glu-198 (M1 site) interacts with His-60 (M2 site). Therefore, systematic mutagenesis of His-60 and His-307 (M1 site) to alanine residues was undertaken; the resulting variants were named $TmPep1050_{H60A}$ and $TmPep1050_{H307A}$, respectively. The variants were barely active using L-Leu-*p*NA as substrate (specific activity of less than 0.1 s^{-1}). Their molecular states were determined by native MS: $TmPep1050_{H60A}$ was a mixture of dimers and dodecamers, whereas $TmPep1050_{H307A}$ formed only dimers (see Fig. 8*A*). Size-exclusion chromatography of $TmPep1050_{H60A}$ showed that the dodecamer-to-dimer ratio was about 0.75.

As $TmPep1050_{H307A}$ only exists as a dimer, the M1 site is probably the most important feature leading to $TmPep1050_{12\text{-mer}}$ formation. The structures of $TmPep1050_{H60A}$ (dimer only) and $TmPep1050_{H307A}$ were solved by X-ray crystallography. The H307A mutation provokes the same structural dissimilarities as observed in $TmPep1050_{2\text{-mer}}$ (see Figs. 8*B* and S9). The α8 and α10 helices could not be modeled in the $TmPep1050_{H307A}$ structure, whereas the α9 helix is elongated by three residues. Glu-197 and Glu-198 residues are also oriented outward from the catalytic site. The M2 site presents the same spatial arrangement as in the dimer, indicating that only the apo form was crystallized. Our data strongly support the structural role of the metal cofactors in addition to their catalytic role. Colombo *et al.* (24) drew the same conclusion with PfTET3 and hypothesized that the metal ion bound to the M1 site could stabilize the oligomerization.

The presence of Co^{2+} in the M2 site is not strictly required for oligomerization as $TmPep1050_{H60A}$ is able to form a dodecamer. Still, the H60A mutation impeded cobalt binding in the M1 site, probably due to H-bond network disruption. The $TmPep1050_{H60A}$ dimer structure afforded a better understanding of conformational changes in the Gln-196–Tyr-209 region. Glu-197–Ile-201 forms a wide loop instead of a tight U-turn. The α8 helix is completely unwound in the $TmPep1050_{H60A}$ dimer (see Fig. 8*B*).

Mutagenesis of His-60 and His-307, which are involved in metal ion binding in the M2 and M1 sites, respectively, led to an interesting observation. In M42 aminopeptidases, the M1 site has been described to have a lower affinity to metal ion than the M2 site (15, 24). According to Colombo *et al.* (24), the removal of Co^{2+} from the M1 site led to a partial dissociation of PfTET3 dodecamer into monomer. The dodecamers were fully dissociated when the second metal ion in the M2 site was removed. Surprisingly, the H307A mutation abolished the dodecamer formation for $TmPep1050_{H307A}$, whereas a partial dissociation was observed for $TmPep1050_{H60A}$. Our data strongly suggest that the M1 site strictly controls the TmPep1050 oligomerization. Further studies would be required to understand how M1 and M2 sites finely tune the oligomerization according to their affinity and selectivity.

Experimental procedures

Cloning and mutagenesis of Tm_1050

The *Tm_1050* ORF was amplified from BspHI-digested TmCD00089984 plasmid (Joint Center for Structural Genomics) using *Pfu* DNA polymerase (Thermo Fisher Scientific) and primers ocej419 and ocej420 (see Table S3). The PCR product was inserted into the pBAD vector (Thermo Fisher Scientific) by homologous recombination in *E. coli* according to the seamless ligation cloning extract (SLiCE) protocol (44), giving rise to pCEC43. Briefly, the insertion of the PCR product is allowed via two 30-bp sequences, homologous to the insertion site of the pBAD vector, flanking the gene of interest. The homologous recombination is mediated by using a cell extract of PPY strain, expressing the λ prophage Red recombination system. Site-directed mutagenesis was carried out following the single-primer

Cobalt-driven dimer–dodecamer transition of TmPep1050

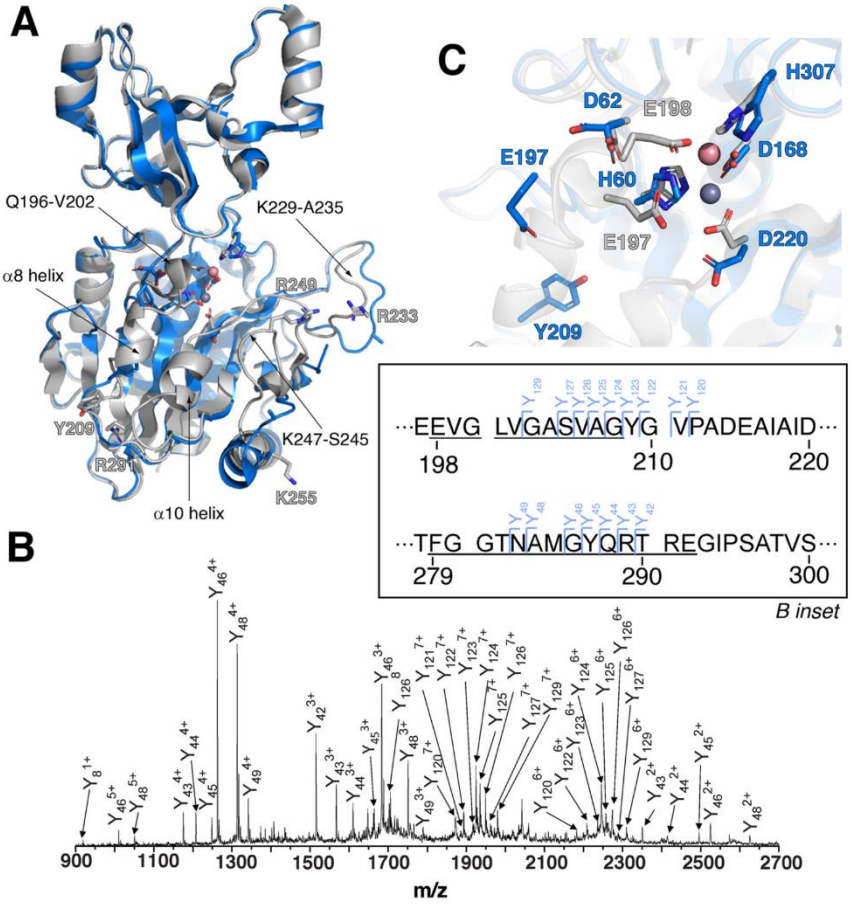


Figure 6. The TmPep1050$_{2\text{-mer}}$ structure. *A*, structural alignment of TmPep1050$_{2\text{-mer}}$ subunit (PDB code 5NE6; *blue*) *versus* TmPep1050$_{12\text{-mer}}$ subunit (PDB code 6NW5; *light gray*). *Arrows* point to the structural dissimilarities. The metal-binding residues, catalytic residues, and residues involved in the oligomerization are shown in *stick* representation. *B*, MS fragmentation pattern of TmPep1050$_{2\text{-mer}}$ under 125-V collision trap energy. In the *inset*, preferential fragmentations are annotated in *blue*. The sequences *underlined* are the unmodeled flexible parts of the TmPep1050$_{2\text{-mer}}$ structure. *C*, close-up of the TmPep1050$_{2\text{-mer}}$ (*blue*) active site *versus* the TmPep1050$_{12\text{-mer}}$ active site (*light gray*). The side chains of the metal-binding residues and the catalytic residues are represented as *sticks*. The metal cofactors of TmPep1050$_{12\text{-mer}}$, Co^{2+} and Zn^{2+}, are shown as *pink* and *gray spheres*, respectively.

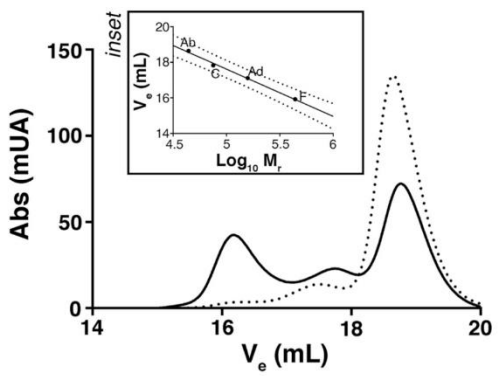


Figure 7. Oligomeric states of TmPep1050$_{K232A/R233A/R249A}$ (plain line) and TmPep1050$_{K232E/R233E/R249E}$ (dots) determined by gel filtration on Superose 6 column. The peaks at V_e of 16.2, 17.1, and 18.3 ml correspond to dodecamers (347 kDa), tetramers (151 kDa), and dimers (54 kDa), respectively. *Inset*, the calibration of the Superose 6 column with albumin (*Ab*), conalbumin (*C*), aldolase (*Ad*), and ferritin (*F*) as standards. The correlation between V_e (ml) and the logarithm of the relative mass (M_r) is linear with R^2 of 0.99. The 95% confidence intervals of the linear regression are shown in *dots*. *Abs*, absorbance. *mUA*, milli unit of absorbance.

reactions in parallel (SPRINP) protocol (45) except for the two vectors used for TmPep1050$_{K232A/R233A/R249A}$ and TmPep1050$_{K232E/R233E/R249E}$ production. In that case, two synthetic genes harboring the desired mutations (GeneArt, Thermo Fisher Scientific) were introduced into the pBAD vector by homologous recombination. The primers used to generate TmPep1050 variants are listed in Table S1. All genetic constructs were verified by sequencing (Genetic Service Facility, University of Antwerp) and are listed in Table S4. The *E. coli* MC1061 strain (46) was used for cloning and expression. Cells were grown on LB broth in the presence of 100 μg/ml ampicillin for positive selection.

Production and purification of recombinant enzymes

Cultures and protein extracts were prepared following previously published procedures (23) with two modifications: (i) cells from a 1-liter culture were disrupted in 40 ml of 50 mM MOPS, 1 mM $CoCl_2$, pH 7.2, and (ii) protein extracts were heated at 70 °C for 10 min. The purification consisted of three chromatographic steps. The first step was anion-exchange

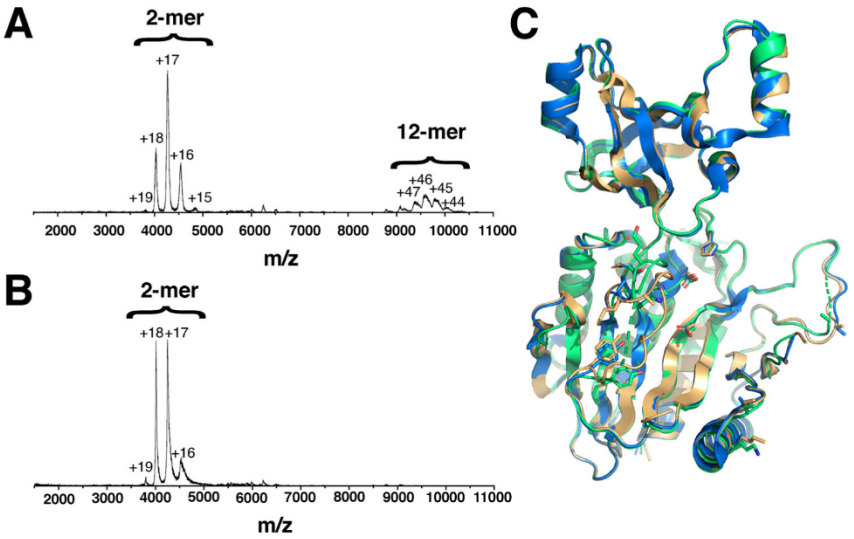


Figure 8. *A*, native mass spectrum of TmPep1050$_{H60A}$ in 100 mM ammonium acetate. Dimers and dodecamers were the two major species observed. *B*, native mass spectrum of TmPep1050$_{H307A}$ in 100 mM ammonium acetate. Dimers were the main oligomeric form observed. *C*, structural alignment of TmPep1050$_{2\text{-mer}}$ (*blue*), TmPep1050$_{H60A}$ structure (*gold*), and TmPep1050$_{H307A}$ (*green*). r.m.s.d. values are 0.122 and 0.468 for TmPep1050$_{H60A}$ and TmPep1050$_{H307A}$, respectively.

chromatography on Source 15Q resin (GE Healthcare, Tricorn 10/150 column) in 50 mM MOPS, 1 mM $CoCl_2$, pH 7.2. Elution was performed with a gradient step from 0 to 0.5 M NaCl for five column volumes (CV). Fractions containing the protein of interest (2 CV) were pooled, and $(NH_4)_2SO_4$ powder was added to a concentration of 1.5 M $(NH_4)_2SO_4$. The second chromatographic step was hydrophobic-interaction chromatography on Source 15Phe resin (GE Healthcare, XK 16/20 column) in 50 mM MOPS, 1.5 M $(NH_4)_2SO_4$, 1 mM $CoCl_2$, pH 7.2. Elution was performed with a gradient step from 1.5 to 0 M $(NH_4)_2SO_4$ for 5 CV. Fractions (1.5 CV) containing the protein of interest were pooled and concentrated using an Amicon Ultra-15 ultrafiltration unit with 30-kDa cutoff (Merck Millipore). The third step consisted of gel filtration on Superdex 200 resin (GE Healthcare, XK 16/70 column) in 50 mM MOPS, 0.5 M $(NH_4)_2SO_4$, 1 mM $CoCl_2$, pH 7.2. The concentration of $(NH_4)_2SO_4$ had to be maintained at 0.5 M to avoid protein precipitation. Purified proteins were finally concentrated using an Amicon Ultra-15 ultrafiltration unit with 30-kDa cutoff. The presence and purity of the recombinant enzymes were checked throughout the purification procedure by SDS-PAGE and enzymatic assays with L-Leu-*p*NA as substrate (see below). Proteins were quantified by measuring the absorbance at 280 nm and applying the extinction coefficient of 18,910 M^{-1} cm^{-1}. This purification protocol allowed the purification of about 10 mg of TmPep1050 from 1 liter of culture. The concentrations of purified TmPep1050 are given in Table S1. Molecular weights were determined by gel filtration on Superdex 200 resin (GE Healthcare, XK 16/70 column) and Superpose 6 10/300 GL (GE Healthcare) using 50 mM MOPS, 0.5 M $(NH_4)_2SO_4$, pH 7.2, as running buffer. The gel-filtration columns were calibrated using both a high-molecular-weight gel filtration calibration kit (GE Healthcare) and gel-filtration standard (Bio-Rad).

Enzymatic assays

Aminoacyl-*p*NA substrates were purchased from Bachem AG. Aminopeptidase (EC 3.4.11.1) activity was assayed as described previously (23) except that enzymatic reactions were stopped by adding 1 ml of 20% acetic acid to 1 ml of reaction mixture (200 μl of substrate in 10% methanol, 790 μl of 50 mM MOPS, pH 7.2, and 10 μl of enzyme at a concentration ranging from 10 to 50 nM depending on the substrate). For the determination of substrate specificity of TmPep1050$_{12\text{-mer}}$, aminopeptidase assays were carried out at 75 °C with an enzyme concentration of 30 nM and the appropriate amino acid-*p*NA substrate. All substrates were used at 2.5 mM except L-Gly-*p*NA (1.25 mM), L-Phe-*p*NA (0.75 mM), and L-His-*p*NA and L-Glu-*p*NA (0.5 mM). The effects of metal ions, pH, and temperature on the activity were studied as described previously (23). For the determination of kinetic parameters, assays were performed at 75 °C with an enzyme concentration of 10 nM for L-Leu-*p*NA and L-Ile-*p*NA and an enzyme concentration of 50 nM for L-Met-*p*NA. Reaction mixtures were supplemented with 250 μM $CoCl_2$. Kinetic parameters (k_{cat}, K_m, and k_{cat}/K_m) were determined from the initial reaction rates using Lineweaver–Burk linearization of the Michaelis–Menten equation. Activation energies were calculated from the slope of the trend line obtained by plotting the logarithm of the specific activity *versus* the inverse of the temperature. For thermostability assays, TmPep1050$_{12\text{-mer}}$ was diluted to 1 μM in 50 mM MOPS, 0.5 M $(NH_4)_2SO_4$, 1 mM $CoCl_2$, pH 7.2, and it was incubated at either 75 or 95 °C. At various time intervals, the activity was measured at the incubation temperature (75 or 95 °C) by diluting the enzyme to 10 nM in 1 ml of reaction mixture containing 2.5 mM L-Leu-*p*NA as substrate.

Cobalt binding assays

A 1 mM TmPep1050$_{12\text{-mer}}$ sample was diluted in 10 volumes of 2.1 M malic acid, pH 7.0, and concentrated back to 1 volume

Cobalt-driven dimer–dodecamer transition of TmPep1050

Table 4
Data collection and refinement statistics
Values in parentheses are for the highest-resolution shell. ESRF, European Synchrotron Radiation Facility; ASU, asymmetric unit.

	Apo-TmPep1050$_{12\text{-mer}}$	TmPep1050$_{12\text{-mer}}$	TmPep1050$_{2\text{-mer}}$	TmPep1050$_{H60A}$	TmPep1050$_{H307A}$
Data collection					
Temperature (K)	100	100	100	100	100
Radiation source	ESRF BM30a	SOLEIL Proxima 2	SOLEIL Proxima 2	SOLEIL Proxima 2	SOLEIL Proxima 2
Wavelength (Å)	0.9797	0.9801	0.9800	0.9801	0.9801
Detector	ADSC QUANTUM 315r	DECTRIS EIGER X 9M	DECTRIS EIGER X 9M	DECTRIS EIGER X 9M	DECTRIS EIGER X 9M
Rotation range (°)	0.37	0.10	0.20	0.10	0.10
Exposure time (s)	20	0.025	0.025	0.025	0.025
Space group	P 1	H 3	C 2 2 21	C 2 2 21	P 1 21 1
Unit cell parameters					
α, β, γ (°)	114.46, 91.71, 105.69	90.00, 90.00, 120.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 110.51, 90.00
a, b, c (Å)	114.26, 114.57, 114.04	131.15, 131.15, 285.61	42.55, 114.71, 267.69	42.63, 114.22, 267.96	42.79, 138.65, 61.25
Resolution (Å)	44.05–2.20 (2.28–2.20)	47.60–1.70 (1.74–1.70)	48.25–2.00 (2.05–2.00)	43.46–1.84 (1.89–1.84)	40.11–1.75 (1.79–1.75)
Unique reflections	237,152	201,316	45,086	57,222	67,094
R_{merge} (%)	8.9 (39.2)	8.2 (67.0)	9.5 (69.1)	8.7 (60.6)	5.6 (56.3)
Redundancy	3.2 (2.2)	10.34 (10.41)	13.2 (13.0)	12.0 (9.8)	6.7 (6.7)
$\langle I/\sigma \rangle$	8.56 (2.22)	15.81 (2.81)	16.81 (3.46)	16.52 (2.63)	18.04 (2.38)
Completeness (%)	93.5 (84.8)	99.8 (97.6)	99.9 (99.2)	99.4 (91.7)	99.5 (97.2)
$CC_{1/2}$ (%)	99.4 (81.1)	99.9 (87.1)	99.8 (90.0)	99.9 (90.0)	99.9 (88.1)
Refinement					
Resolution	44.05–2.20	47.60–1.70	48.25–2.00	43.46–1.84	40.11–1.75
Reflections	237,090	201,316	45,076	57,213	67,094
R_{free} set test count	11,854	9,715	2,254	2,862	3,363
R_{work}/R_{free}	0.212/0.247	0.143/0.164	0.166/0.203	0.167/0.195	0.165/0.185
Protein molecules per ASU	12	4	2	2	2
V_M (Å^3/Da)	2.98	3.27	2.26	2.26	2.36
Solvent content (%)	58.7	62.4	45.6	45.6	47.9
Protein/solvent atoms	29,969/2,223	10,759/1,474	4,610/362	4,730/500	4,621/335
r.m.s.d. bond lengths (Å)	0.009	0.019	0.004	0.004	0.012
r.m.s.d. bond angles (°)	1.254	1.680	0.647	0.672	1.263
Average B-factors (Å^2)	37.0	24.0	37.8	33.2	28.9
Favored/disallowed Ramachandran φ/ψ (%)	95.71/0.46	95.35/0.34	94.86/0.17	94.95/0.00	93.64/0.52
Twin law					$h, -k, -h-l$
PDB code	4P6Y	6NW5	5NE6	5NE7	5NE8

using an Amicon Ultra-15 ultrafiltration unit with a 30-kDa cutoff. The malic acid–treated sample was dialyzed four times against 100 volumes of 50 mM MOPS, 0.5 M $(NH_4)_2SO_4$, pH 7.2 using SnakeSkin™ dialysis tubing with a 3.5-kDa cutoff (Thermo Fisher Scientific). To monitor metal ion removal, the specific activity was measured with L-Leu-*p*NA as substrate. Cobalt binding assays were performed by incubating 100 μl of 20 μM cobalt-depleted TmPep1050$_{12\text{-mer}}$ with $CoCl_2$ at a concentration ranging from 0 to 2560 μM in 50 mM MOPS, 0.5 M $(NH_4)_2SO_4$, pH 7.2, for 24 h at 75 °C. After incubation, the specific activity was measured using 10 nM enzyme and 2.5 mM L-Leu-*p*NA as substrate without added cobalt in the reaction mixture. Other metal ions were tested following this procedure to identify the metal cofactor of TmPep1050$_{12\text{-mer}}$.

In parallel, cobalt concentration was determined using Amplex UltraRed (Thermo Fisher Scientific), a fluorescent probe that specifically binds cobalt at high pH. The protocol established by Tsai and Lin (47) was adapted to fit a 384-well black microplate (Corning). Fluorescence was measured on a SpectraMax 5 (Molecular Devices) with excitation wavelength set on 495 nm and emission wavelength set on 570 nm. Prior to cobalt ion quantification, samples were diluted twice in 50 mM MOPS, 0.5 M $(NH_4)_2SO_4$, pH 7.2. In addition to total cobalt concentration, unbound cobalt was quantified after filtering samples using an Amicon Ultra-0.5 ultrafiltration unit with a 30-kDa cutoff. The affinity constant K_d was determined from Scatchard plot data representation.

To study the reassociation of dimers into dodecamers, 100 μl of 50 μM TmPep1050$_{2\text{-mer}}$ was incubated with Co^{2+} at a concentration ranging from 0 to 5 mM for 30 min at 75 °C. The oligomers were detected and quantified by size-exclusion chromatography using Superdex 200 resin (GE Healthcare Life Sciences, XK16/20 column). 50 mM MOPS, 0.5 M $(NH_4)_2SO_4$, pH 7.2, buffer was used for this assay.

Thermal shift assay

SYPRO Orange™ (Thermo Fisher Scientific) was diluted 1:125 in 50 mM MOPS, 0.5 M $(NH_4)_2SO_4$, pH 7.2. The fluorescent probe was mixed with protein samples conditioned in 50 mM MOPS, 0.5 M $(NH_4)_2SO_4$, pH 7.2, and Co^{2+} at a concentration ranging from 0 to 2560 μM. The working SYPRO Orange dilution was 1:1000, and the protein concentration was 20 μM for a reaction volume of 20 μl. Thermal shift assays were performed in a 96-well plate on a StepOnePlus™ Real-Time PCR System (Thermo Fisher Scientific). Fluorescence curves were treated with StepOne™ software.

Native MS

Samples for native MS were transferred into 20 mM ammonium acetate, pH 7.2. This was done using Zeba 7-kDa desalting columns (Thermo Fisher Scientific). If further desalting was needed, Bio-Spin P-6 gel columns (Bio-Rad) were used. The protein samples were diluted in 100 mM ammonium acetate, pH 7.2, to a working concentration of 5 μM unless stated otherwise. In house–prepared borosilicate gold-coated needles filled with 2–3 μl of sample were used to introduce the protein into the gas phase using nanoelectrospray ionization. The spectra were recorded in positive ion mode on a traveling-wave ion

mobility Q-TOF instrument (Synapt G2 HDMS, Waters). Different settings were tuned to optimize sample measurement. The most important settings applied during the measurements, unless stated differently in the figure legends, were 20-V sampling cone, 10-V trap collision energy, and pressures of 8.30 e^{-3} and 5.31 e^{-2} millibar for the source and trap, respectively.

Crystallization

$TmPep1050_{12\text{-}mer}$, $TmPep1050_{2\text{-}mer}$, $TmPep1050_{H60A}$, and $TmPep1050_{H307A}$ were crystallized using the hanging-drop vapor diffusion method at 292 K in EasyXtal Tool plates (Qiagen). Drops contained 2 μl of recombinant enzyme mixed with 2 μl of well buffer. Crystallization conditions are described in Table S1. One cycle of seeding was necessary to get monocrystals of each species.

Structure determination and analysis

For apo-$TmPep1050_{12\text{-}mer}$, diffraction data were collected on the FIP-BM30a beamline at the European Synchrotron Radiation Facility (Grenoble, France) (48, 49). For $TmPep1050_{12\text{-}mer}$, $TmPep1050_{2\text{-}mer}$, $TmPep1050_{H60A}$, and $TmPep1050_{H307A}$, diffraction data were collected on Proxima 2 beamline at SOLEIL (Saint-Aubin, France). The data collection and refinement statistics are presented in Table 4. Diffraction data were processed using the XDS program package (50). Molecular replacement and model building were performed using PHENIX software package v.1.10.1–2155 (51). The initial solution of $TmPep1050_{12\text{-}mer}$ was determined by molecular replacement with MR-Rosetta using the coordinates of YpdE of *S. flexneri* (PDB code 1YLO) as the search model (translation-function Z-score value, 23.9; log-likelihood gain, 5150). For $TmPep1050_{2\text{-}mer}$, $TmPep1050_{H60A}$, and $TmPep1050_{H307A}$, molecular replacement was achieved with Phaser-MR using the coordinates of $TmPep1050_{12\text{-}mer}$ as search model. The models were built using phenix.autobuild. Iterative manual building was done in Coot (52). Multiple rounds of refinement were performed using phenix.refine. Model stereochemical quality was assessed using MolProbity (53). Protein structures were analyzed with PDBePisa (54), Arpeggio (55), Rosetta pK_a protocol (56, 57), Advanced Poisson–Boltzmann Solver (APBS) (58), and PyMOL Molecular Graphics System version 2.2 (Schrödinger, LLC).

Author contributions—R. D., F. S., and L. D. conceptualization; R. D. and N. B. formal analysis; R. D. and N. B. validation; R. D., T. V. G., and J. V. D. investigation; R. D. and T. V. G. visualization; R. D. and N. B. methodology; R. D., J. V. D., and L. D. writing-original draft; R. D. and L. D. writing-review and editing; D. V. E. resources; D. V. E. software; F. S. and L. D. supervision; F. S. and L. D. project administration.

Acknowledgments—We thank Martine Roovers for proofreading this paper and giving constructive comments. Access to BM30A beamline (European Synchrotron Radiation Facility) was supported by Fonds National de la Recherche Scientifique under Contract IISN 4.4503.11F. Access to Proxima 2 beamline (SOLEIL synchrotron) was within Block Allocation Groups 20151139 and 20171555.

References

1. Puente, X. S., Sánchez, L. M., Overall, C. M., and López-Otín, C. (2003) Human and mouse proteases: a comparative genomic approach. *Nat. Rev. Genet.* **4,** 544–558 CrossRef Medline
2. Butler, S. M., Festa, R. A., Pearce, M. J., and Darwin, K. H. (2006) Self-compartmentalized bacterial proteases and pathogenesis. *Mol. Microbiol.* **60,** 553–562 CrossRef Medline
3. Collins, G. A., and Goldberg, A. L. (2017) The logic of the 26S proteasome. *Cell* **169,** 792–806 CrossRef Medline
4. Gur, E., Biran, D., and Ron, E. Z. (2011) Regulated proteolysis in Gram-negative bacteria—how and when? *Nat. Rev. Microbiol.* **9,** 839–848 CrossRef Medline
5. Janssen, B. D., and Hayes, C. S. (2012) The tmRNA ribosome-rescue system. *Adv. Protein Chem. Struct. Biol.* **86,** 151–191 CrossRef Medline
6. Konovalova, A., Søgaard-Andersen, L., and Kroos, L. (2014) Regulated proteolysis in bacterial development. *FEMS Microbiol. Rev.* **38,** 493–522 CrossRef Medline
7. Baumeister, W., Walz, J., Zühl, F., and Seemüller, E. (1998) The proteasome: paradigm of a self-compartmentalizing protease. *Cell* **92,** 367–380 CrossRef Medline
8. Kisselev, A. F., Akopian, T. N., Woo, K. M., and Goldberg, A. L. (1999) The sizes of peptides generated from protein by mammalian 26 and 20 S proteasomes: implications for understanding the degradative mechanism and antigen presentation. *J. Biol. Chem.* **274,** 3363–3371 CrossRef Medline
9. Brandstetter, H., Kim, J.-S., Groll, M., and Huber, R. (2001) Crystal structure of the tricorn protease reveals a protein disassembly line. *Nature* **414,** 466–470 CrossRef Medline
10. Saric, T., Graef, C. I., and Goldberg, A. L. (2004) Pathway for degradation of peptides generated by proteasomes: a key role for thimet oligopeptidase and other metallopeptidases. *J. Biol. Chem.* **279,** 46723–46732 CrossRef Medline
11. Chuang, C. K., Rockel, B., Seyit, G., Walian, P. J., Schönegge, A.-M., Peters, J., Zwart, P. H., Baumeister, W., and Jap, B. K. (2010) Hybrid molecular structure of the giant protease tripeptidyl peptidase II. *Nat. Struct. Mol. Biol.* **17,** 990–996 CrossRef Medline
12. Ito, K., Nakajima, Y., Onohara, Y., Takeo, M., Nakashima, K., Matsubara, F., Ito, T., and Yoshimoto, T. (2006) Crystal structure of aminopeptidase N (proteobacteria alanyl aminopeptidase) from *Escherichia coli* and conformational change of methionine 260 involved in substrate recognition. *J. Biol. Chem.* **281,** 33664–33676 CrossRef Medline
13. Borissenko, L., and Groll, M. (2005) Crystal structure of TET protease reveals complementary protein degradation pathways in prokaryotes. *J. Mol. Biol.* **346,** 1207–1219 CrossRef Medline
14. Rawlings, N. D., Barrett, A. J., Thomas, P. D., Huang, X., Bateman, A., and Finn, R. D. (2018) The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* **46,** D624–D632 CrossRef Medline
15. Russo, S., and Baumann, U. (2004) Crystal structure of a dodecameric tetrahedral-shaped aminopeptidase. *J. Biol. Chem.* **279,** 51275–51281 CrossRef Medline
16. Basbous, H., Appolaire, A., Girard, E., and Franzetti, B. (2018) Characterization of a glycyl-specific TET aminopeptidase complex from *Pyrococcus horikoshii*. *J. Bacteriol.* **200,** e00059-18 CrossRef Medline
17. Durá, M. A., Rosenbaum, E., Larabi, A., Gabel, F., Vellieux, F. M., and Franzetti, B. (2009) The structural and biochemical characterizations of a novel TET peptidase complex from *Pyrococcus horikoshii* reveal an integrated peptide degradation system in hyperthermophilic Archaea: characterization of *P. horikoshii* TET3 peptidase. *Mol. Microbiol.* **72,** 26–40 CrossRef Medline
18. Schoehn, G., Vellieux, F. M., Asunción Durá, M., Receveur-Bréchot, V., Fabry, C. M., Ruigrok, R. W., Ebel, C., Roussel, A., and Franzetti, B. (2006) An archaeal peptidase assembles into two different quaternary structures: a tetrahedron and a giant octahedron. *J. Biol. Chem.* **281,** 36327–36337 CrossRef Medline
19. Appolaire, A., Girard, E., Colombo, M., Durá, M. A., Moulin, M., Härtlein, M., Franzetti, B., and Gabel, F. (2014) Small-angle neutron scattering re-

veals the assembly mode and oligomeric architecture of TET, a large, dodecameric aminopeptidase. *Acta Crystallogr. D Biol. Crystallogr.* **70,** 2983–2993 CrossRef Medline

20. Appolaire, A., Durá, M. A., Ferruit, M., Andrieu, J.-P., Godfroy, A., Gribaldo, S., and Franzetti, B. (2014) The TET2 and TET3 aminopeptidases from *Pyrococcus horikoshii* form a hetero-subunit peptidasome with enhanced peptide destruction properties: TET aminopeptidase multi-subunit complex. *Mol. Microbiol.* **94,** 803–814 CrossRef Medline
21. Appolaire, A., Colombo, M., Basbous, H., Gabel, F., Girard, E., and Franzetti, B. (2016) TET peptidases: a family of tetrahedral complexes conserved in prokaryotes. *Biochimie* **122,** 188–196 CrossRef Medline
22. Chevrier, B., Schalk, C., D'Orchymont, H., Rondeau, J.-M., Moras, D., and Tarnus, C. (1994) Crystal structure of *Aeromonas proteolytica* aminopeptidase: a prototypical member of the co-catalytic zinc enzyme family. *Structure* **2,** 283–291 CrossRef Medline
23. Dutoit, R., Brandt, N., Legrain, C., and Bauvois, C. (2012) Functional characterization of two M42 aminopeptidases erroneously annotated as cellulases. *PLoS One* **7,** e50639 CrossRef Medline
24. Colombo, M., Girard, E., and Franzetti, B. (2016) Tuned by metals: the TET peptidase activity is controlled by 3 metal binding sites. *Sci. Rep.* **6,** 20876 CrossRef Medline
25. Franzetti, B., Schoehn, G., Hernandez, J. F., Jaquinod, M., Ruigrok, R. W., and Zaccai, G. (2002) Tetrahedral aminopeptidase: a novel large protease complex from archaea. *EMBO J.* **21,** 2132–2138 CrossRef Medline
26. Petrova, T. E., Slutskaya, E. S., Boyko, K. M., Sokolova, O. S., Rakitina, T. V., Korzhenevskiy, D. A., Gorbacheva, M. A., Bezsudnova, E. Y., and Popov, V. O. (2015) Structure of the dodecamer of the aminopeptidase APDkam598 from the archaeon *Desulfurococcus kamchatkensis*. *Acta Crystallogr. F Struct. Biol. Commun.* **71,** 277–285 CrossRef Medline
27. Kim, D., San, B. H., Moh, S. H., Park, H., Kim, D. Y., Lee, S., and Kim, K. K. (2010) Structural basis for the substrate specificity of PepA from *Streptococcus pneumoniae*, a dodecameric tetrahedral protease. *Biochem. Biophys. Res. Commun.* **391,** 431–436 CrossRef Medline
28. Chaikuad, A., Pilka, E. S., De Riso, A., von Delft, F., Kavanagh, K. L., Vénien-Bryan, C., Oppermann, U., and Yue, W. W. (2012) Structure of human aspartyl aminopeptidase complexed with substrate analogue: insight into catalytic mechanism, substrate specificity and M18 peptidase family. *BMC Struct. Biol.* **12,** 14 CrossRef Medline
29. Nguyen, D. D., Pandian, R., Kim, D., Ha, S. C., Yoon, H.-J., Kim, K. S., Yun, K. H., Kim, J.-H., and Kim, K. K. (2014) Structural and kinetic bases for the metal preference of the M18 aminopeptidase from *Pseudomonas aeruginosa*. *Biochem. Biophys. Res. Commun.* **447,** 101–107 CrossRef Medline
30. Sivaraman, K. K., Oellig, C. A., Huynh, K., Atkinson, S. C., Poreba, M., Perugini, M. A., Trenholme, K. R., Gardiner, D. L., Salvesen, G., Drag, M., Dalton, J. P., Whisstock, J. C., and McGowan, S. (2012) X-ray crystal structure and specificity of the *Plasmodium falciparum* malaria aminopeptidase PfM18AAP. *J. Mol. Biol.* **422,** 495–507 CrossRef Medline
31. Su, M.-Y., Peng, W.-H., Ho, M.-R., Su, S.-C., Chang, Y.-C., Chen, G.-C., and Chang, C.-I. (2015) Structure of yeast Ape1 and its role in autophagic vesicle formation. *Autophagy* **11,** 1580–1593 CrossRef Medline
32. Holz, R. C. (2002) The aminopeptidase from *Aeromonas proteolytica*: structure and mechanism of co-catalytic metal centers involved in peptide hydrolysis. *Coord. Chem. Rev.* **232,** 5–26 CrossRef
33. Macek, P., Kerfah, R., Boeri Erba, E., Crublet, E., Moriscot, C., Schoehn, G., Amero, C., and Boisbouvier, J. (2017) Unraveling self-assembly pathways of the 468-kDa proteolytic machine TET2. *Sci. Adv.* **3,** e1601601 CrossRef Medline
34. Rosenbaum, E., Ferruit, M., Durá, M. A., and Franzetti, B. (2011) Studies on the parameters controlling the stability of the TET peptidase superstructure from *Pyrococcus horikoshii* revealed a crucial role of pH and catalytic metals in the oligomerization process. *Biochim. Biophys. Acta* **1814,** 1289–1294 CrossRef Medline
35. Appolaire, A., Rosenbaum, E., Durá, M. A., Colombo, M., Marty, V., Savoye, M. N., Godfroy, A., Schoehn, G., Girard, E., Gabel, F., and Franzetti, B. (2013) *Pyrococcus horikoshii* TET2 peptidase assembling process and associated functional regulation. *J. Biol. Chem.* **288,** 22542–22554 CrossRef Medline
36. Cannan, R. K., and Kibrick, A. (1938) Complex formation between carboxylic acids and divalent metal cations. *J. Am. Chem. Soc.* **60,** 2314–2320 CrossRef
37. Stamper, C., Bennett, B., Edwards, T., Holz, R. C., Ringe, D., and Petsko, G. (2001) Inhibition of the aminopeptidase from *Aeromonas proteolytica* by L-leucinephosphonic acid. Spectroscopic and crystallographic characterization of the transition state of peptide hydrolysis. *Biochemistry* **40,** 7035–7046 CrossRef Medline
38. Durá, M. A., Receveur-Brechot, V., Andrieu, J.-P., Ebel, C., Schoehn, G., Roussel, A., and Franzetti, B. (2005) Characterization of a TET-like aminopeptidase complex from the hyperthermophilic archaeon *Pyrococcus horikoshii*. *Biochemistry* **44,** 3477–3486 CrossRef Medline
39. Ando, S., Ishikawa, K., Ishida, H., Kawarabayasi, Y., Kikuchi, H., and Kosugi, Y. (1999) Thermostable aminopeptidase from *Pyrococcus horikoshii*. *FEBS Lett.* **447,** 25–28 CrossRef Medline
40. Mori, K., and Ishikawa, K. (2005) New deblocking aminopeptidases from *Pyrococcus horikoshii*. *Biosci. Biotechnol. Biochem.* **69,** 1854–1860 CrossRef Medline
41. Roncari, G., Stoll, E., and Zuber, H. (1976) Thermophilic aminopeptidase I. *Methods Enzymol.* **45,** 522–530 CrossRef Medline
42. Moser, P., Roncari, G., and Zuber, H. (1970) Thermophilic aminopeptidases from *Bac. stearothermophilus*: II. Aminopeptidase I (AP I): physicochemical properties; thermostability and activation; formation of the apoenzyme and subunits. *Int. J. Protein Res.* **2,** 191–207 Medline
43. Lee, H. S., Cho, Y., Kim, Y. J., Nam, K., Lee, J.-H., and Kang, S. G. (2007) Biochemical characterization of deblocking aminopeptidase from hyperthermophilic archaeon *Thermococcus onnurineus* NA1. *J. Biosci. Bioeng.* **104,** 188–194 CrossRef Medline
44. Zhang, Y., Werling, U., and Edelmann, W. (2012) SLiCE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Res.* **40,** e55-e55 CrossRef Medline
45. Edelheit, O., Hanukoglu, A., and Hanukoglu, I. (2009) Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies. *BMC Biotechnol.* **9,** 61 CrossRef Medline
46. Casadaban, M. J., and Cohen, S. N. (1980) Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. *J. Mol. Biol.* **138,** 179–207 CrossRef Medline
47. Tsai, C.-Y., and Lin, Y.-W. (2013) A highly selective and sensitive fluorescence assay for determination of copper(II) and cobalt(II) ions in environmental water and toner samples. *Analyst* **138,** 1232–1238 CrossRef Medline
48. Ferrer, J.-L. (2001) Automated data processing on beamline FIP (BM30A) at ESRF. *Acta Crystallogr. D Biol. Crystallogr.* **57,** 1752–1753 CrossRef Medline
49. Roth, M., Carpentier, P., Kaïkati, O., Joly, J., Charrault, P., Pirocchi, M., Kahn, R., Fanchon, E., Jacquamet, L., Borel, F., Bertoni, A., Israel-Gouy, P., and Ferrer, J.-L. (2002) FIP: a highly automated beamline for multiwavelength anomalous diffraction experiments. *Acta Crystallogr. D Biol. Crystallogr.* **58,** 805–814 CrossRef Medline
50. Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 125–132 CrossRef Medline
51. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 213–221 CrossRef Medline
52. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 486–501 CrossRef Medline
53. 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 CrossRef Medline
54. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372,** 774–797 CrossRef Medline

55. Jubb, H. C., Higueruelo, A. P., Ochoa-Montaño, B., Pitt, W. R., Ascher, D. B., and Blundell, T. L. (2017) Arpeggio: a web server for calculating and visualising interatomic interactions in protein structures. *J. Mol. Biol.* **429,** 365–371 CrossRef Medline
56. Kilambi, K. P., and Gray, J. J. (2012) Rapid calculation of protein pKa values using Rosetta. *Biophys. J.* **103,** 587–595 CrossRef Medline
57. Lyskov, S., Chou, F.-C., Conchúir, S. Ó., Der, B. S., Drew, K., Kuroda, D., Xu, J., Weitzner, B. D., Renfrew, P. D., Sripakdeevong, P., Borgo, B., Havranek, J. J., Kuhlman, B., Kortemme, T., Bonneau, R., *et al.* (2013) Serverification of molecular modeling applications: the Rosetta Online Server That Includes Everyone (ROSIE). *PLoS One* **8,** e63906 CrossRef Medline
58. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **98,** 10037–10041 CrossRef Medline

How metal cofactors drive dimer-dodecamer transition of the M42 aminopeptidase TmPep1050 of *Thermotoga maritima*

Raphaël Dutoit[1,2] *, Tom Van Gompel[3], Nathalie Brandt[2], Dany Van Elder[1], Jeroen Van Dyck[3], Frank Sobott[3,4], Louis Droogmans[1]

[1] Laboratory of Microbiology, Department of Molecular Biology, Université Libre de Bruxelles, rue des Professeurs Jeener et Brachet 12, B6041 Charleroi, Belgium
[2] Labiris Institut de Recherche, avenue Emile Gryzon 1, B1070 Brussels, Belgium
[3] Biomolecular & Analytical Mass Spectrometry, Department of Chemistry, Universiteit van Antwerpen, Groenenborgerlaan 171, B2020 Antwerpen, Belgium
[4] Astbury Centre for Structural and Molecular Biology, University of Leeds, LS2 9JT, Leeds, United Kingdom

Running title: *Cobalt-driven dimer-dodecamer transition of TmPep1050*

* To whom correspondence should be addressed: Raphaël Dutoit, rdutoit@ulb.ac.be, Tel. +32 2 526 72 83

Supporting information

Figure S1 X-ray fluorescence of TmPep1050$_{12\text{-mer}}$.
Figure S2 Interaction interface in the TmPep1050$_{12\text{-mer}}$ structure.
Figure S3 Structural alignment of TmPep1050$_{12\text{-mer}}$ subunit A (light grey, PDB ID 6NW5) with *V. proteolyticus* aminopeptidase Ap1 (light blue, PDB ID 1AMP).
Figure S4 The quaternary structures of TmPep1050$_{12\text{-mer}}$ (6NW5), PhTET2 (1Y0R), PhTET1 (2WYR), and SpPepA (3KL9).
Figure S5 Structural comparison of TmPep1050$_{12\text{-mer}}$ with other M42 aminopeptidases.
Figure S6 Influence of pH on TmPep1050$_{12\text{-mer}}$ LAP activity.
Figure S7 Influence of temperature on TmPep1050$_{12\text{-mer}}$ LAP activity.
Figure S8 MS/MS analysis of TmPep1050$_{12\text{-mer}}$.
Figure S9 B-factors plot
Table S1 Crystallization conditions.
Table S2 Effect of Co^{2+} removal on the LAP activity (k) and thermostability (T_m) of TmPep1050$_{12\text{-mer}}$.
Table S3 List of the oligonucleotides used for *Tm_1050* ORF cloning and mutagenesis.
Table S4 List of the plasmids used for expressing TmPep1050$_{12\text{-mer}}$, TmPep1050$_{H60A}$, and TmPep1050$_{H307A.}$

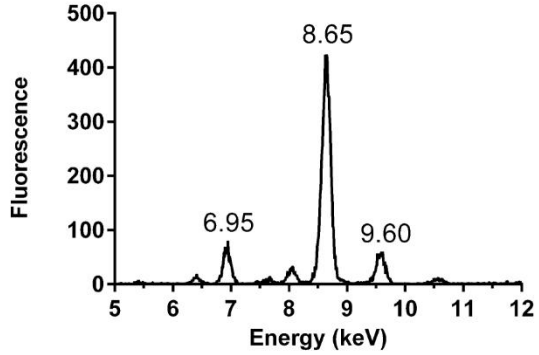


Figure S1 X-ray fluorescence of TmPep1050$_{12\text{-mer}}$. The energy peak at 6.95 keV corresponds to Co^{2+} (K_α = 6.931 keV) while the energy peaks at 8.65 keV and 9.60 keV correspond to Zn^{2+} (K_α = 8.637 keV and K_β = 9.570 keV).

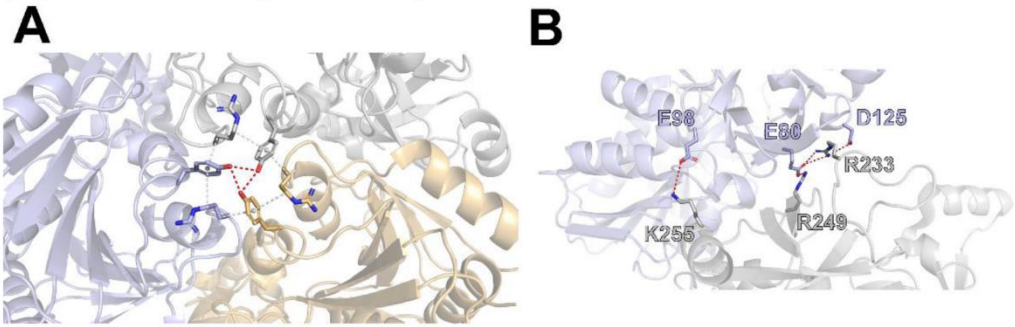


Figure S2 Interaction interface in the TmPep1050$_{12\text{-mer}}$ structure. **A** Polar (red dash) and hydrophobic (white dash) interaction network involving Y209 and R291 residues of three contiguous subunits located at a tetrahedron vertex. **B** Salt bridges between two contiguous subunits of two different dimers involving R233, R249, and K255 of one subunit and E80, E98, and D125 of the other subunit.

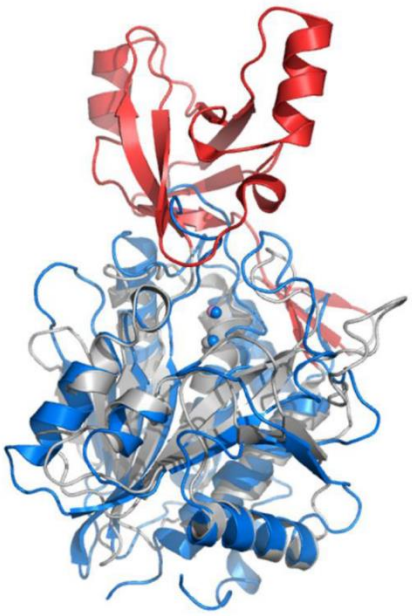


Figure S3 Structural alignment of TmPep1050$_{12\text{-mer}}$ subunit A (light grey, PDB ID 6NW5) with *V. proteolyticus* aminopeptidase Ap1 (light blue, PDB ID 1AMP). RMSD = 2.521. The structural motifs defining the M42 aminopeptidases are shown in red: the PDZ-like domain and β sheet extension (β10, β11, β17).

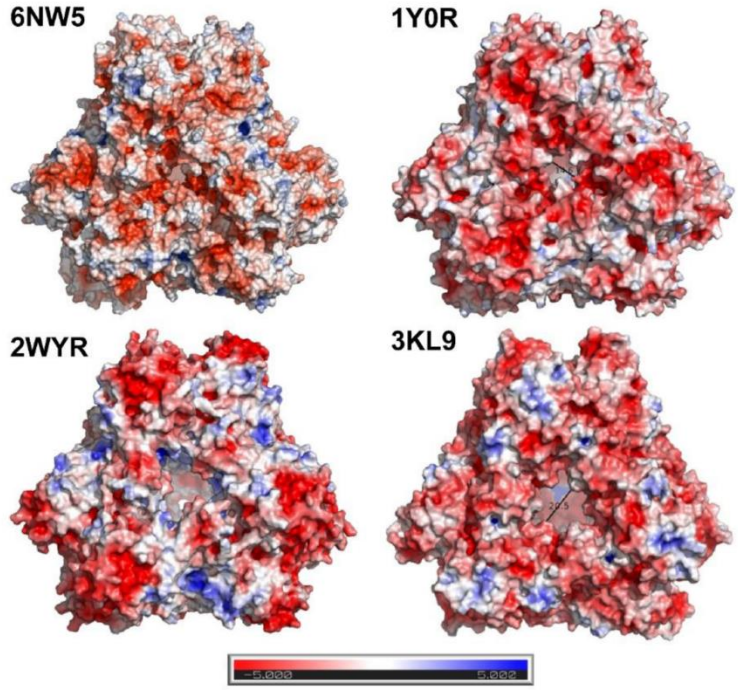


Figure S4 The quaternary structures of TmPep1050$_{12\text{-mer}}$ (6NW5), PhTET2 (1Y0R), PhTET1 (2WYR), and SpPepA (3KL9). The structures are centred on the entrance. The surface electrostatic potentials were calculated using APBS. PhTET2 has an entrance of about 15 Å of diameter while PhTET1 and SpPepA have an entrance diameter of about 20 Å. For TmPep1050$_{12\text{-mer}}$, the entrances are funnel-shaped with an outside diameter of 20 Å and an inside diameter of 11 Å.

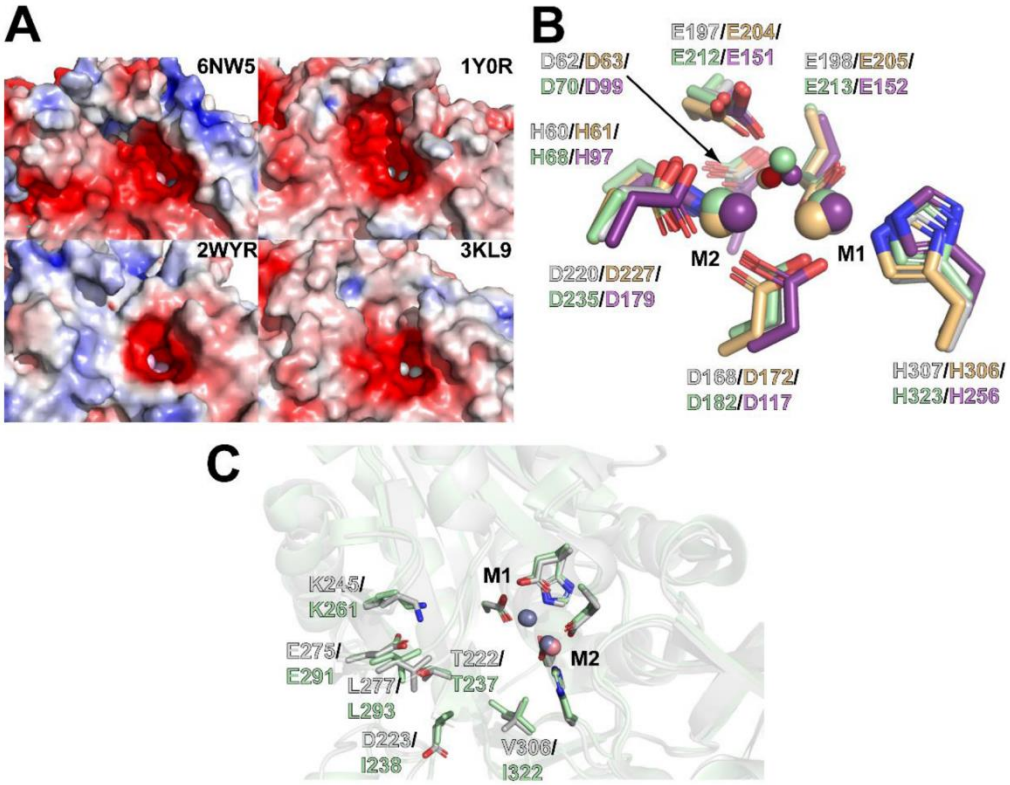


Figure S5 Structural comparison of TmPep1050$_{12\text{-mer}}$ with other M42 aminopeptidases. **A** Catalytic pocket surface electrostatic potential of TmPep1050$_{12\text{-mer}}$ (6NW5), PhTET2 (1Y0R), SpPepA (3KL9), and PhTET1 (2WYR). Surface potentials were calculated using APBS. **B** Structural alignment of the metal binding residues and the catalytic residues of TmPep1050$_{12\text{-mer}}$ (light grey, PDB ID 6NW5), PhTET1 (gold, PDB ID 2WYR), PhTET2 (light green, PDB ID 1Y0R), and *V. proteolyticus* aminopeptidase Ap1 (purple, PDB ID 1AMP). The distances between metal ions are 3.9 Å, 3.4 Å, 3.8 Å, and 3.5 Å, respectively. The distances between either water molecule or hydroxide ion or arsenic ion and M1 / M2 are 2.8 Å / 2.1 Å, 2.3 Å / 2.1 Å, 3.1 Å / 3.0 Å, and 2.2 Å / 2.3 Å, respectively. **C** Structural alignment of the TmPep1050$_{12\text{-mer}}$ catalytic pocket vs. the PhTET2 catalytic pocket. The catalytic base, the metal binding residues, and S1 pocket residues are displayed as sticks. The PhTET2 S1 pocket consists of T237, I238, K261, E291, L293, G296, and I322 corresponding to T222, D223, K245, E275, L277, G280, and V306, respectively, in TmPep1050$_{12\text{-mer}}$.

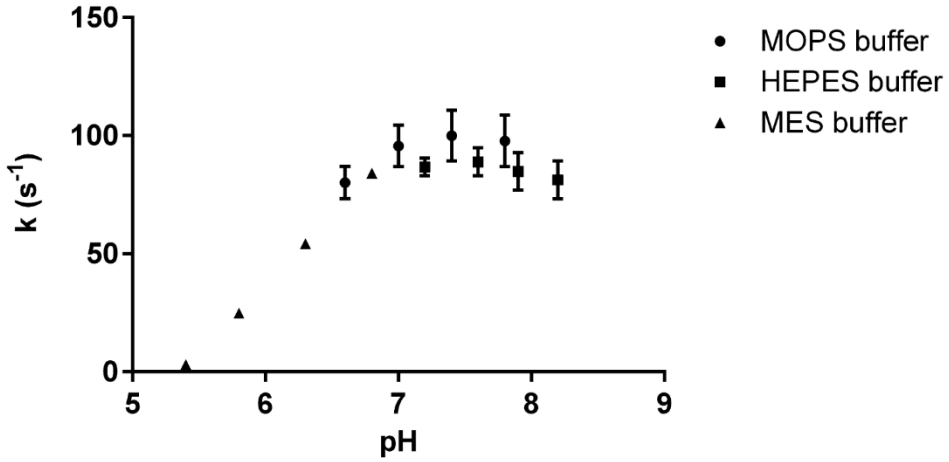


Figure S6 Influence of pH on TmPep1050$_{12\text{-mer}}$ LAP activity. Specific activity (k) vs. pH in MES (triangles), MOPS (circles), and HEPES (squares) buffers. LAP activity was measured using 2.5 mM L-Leu-*p*NA as substrate and 10 nM enzyme.

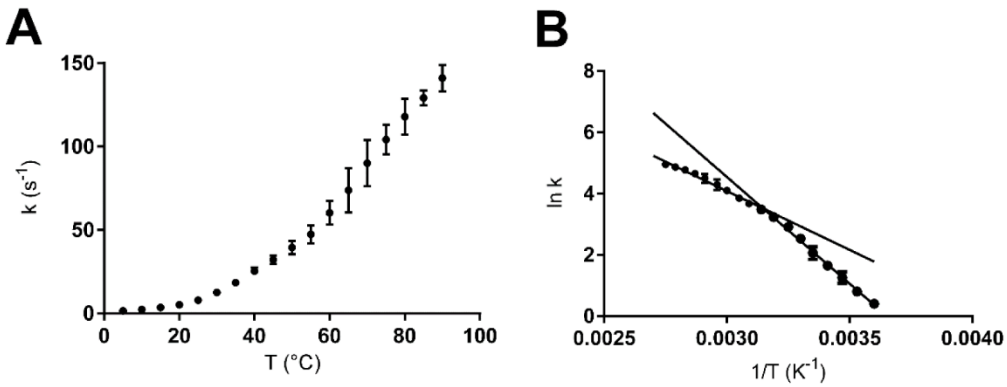


Figure S7 Influence of temperature on TmPep1050$_{12\text{-mer}}$ LAP activity. **A** Specific activity (k) vs. temperature. LAP activity was measured using 2.5 mM L-Leu-*p*NA as substrate and 10 nM enzyme. **B** Arrhenius plot: $\Delta G^{\ddagger}$ = 840 ± 20 J mol^{-1} (R^2 = 0.985), $\Delta G^{\ddagger}$ = 460 ± 20 J mol^{-1} (R^2 = 0.947), temperature transition = 320 K.

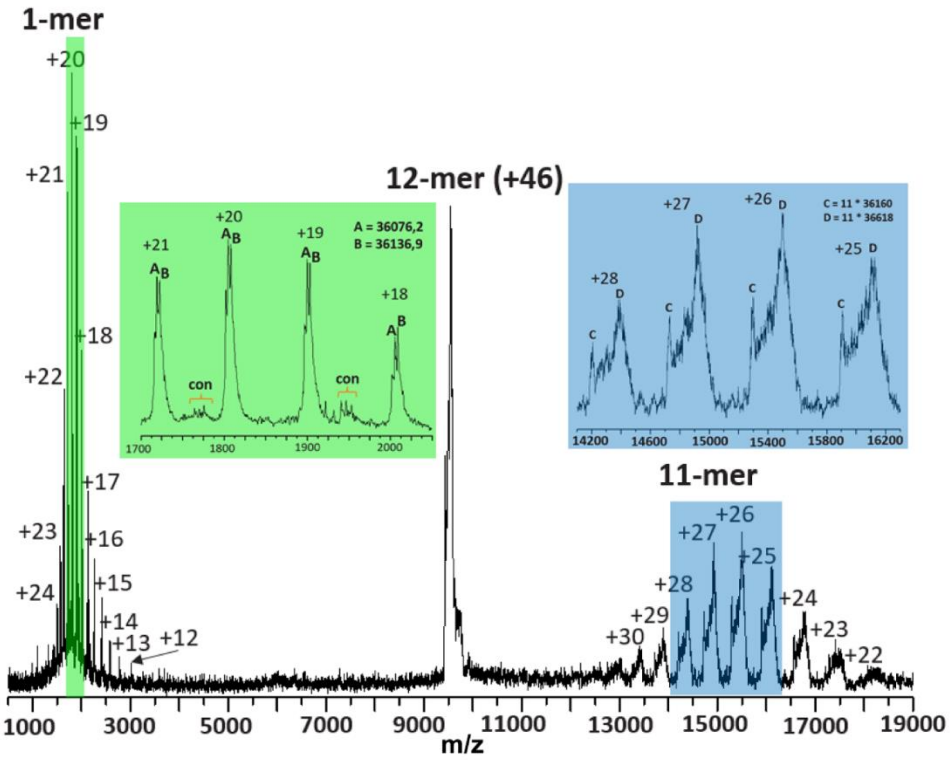


Figure S8 MS/MS analysis of TmPep1050$_{12\text{-mer}}$. The dodecamer +46 peak was selected in the quadrupole of the Synapt G2 mass spectrometer under a collision energy of 200 V. Monomer peaks are highlighted in green and 11-mer peaks in blue. Close-ups are shown as insets in their respective colors. Two masses were measured for the monomer: 36076.2 Da and 36136.9 Da.

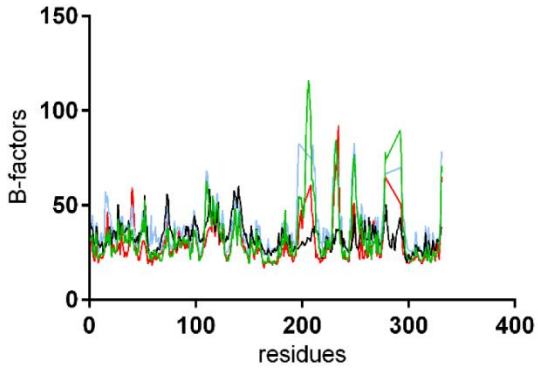


Figure S9 B-factors plot of apo-TmPep1050$_{12\text{-mer}}$ (black), TmPep1050$_{2\text{-mer}}$ (light blue), TmPep1050$_{H60A}$ (green), and TmPep1050$_{H307A}$ (red).

	apo-TmPep1050$_{12\text{-mer}}$	TmPep1050$_{12\text{-mer}}$	TmPep1050$_{2\text{-mer}}$	TmPep1050$_{H60A}$	TmPep1050$_{H307A}$
Crystallization buffer	2.1 M malic acid pH6.75	0.18 M tri-ammonium citrate, 40% PEG3350 pH7.5	0.1 M sodium citrate, 10% PEG3350 pH6.0	0.1 M sodium citrate, 5 % PEG3350 pH4.5	0.1 M sodium citrate, 10 % PEG3350 pH4.5
Protein concentration	390 µM	1 mM	230 µM	230 µM	230 µM
Cryoprotectant soaking	-	-	0.1 M sodium citrate, 5% PEG3350, 20% glycerol pH 5.2	0.1 M sodium citrate, 5% PEG3350, 20% glycerol pH 5.2	0.1 M sodium citrate, 5% PEG3350, 20% glycerol pH 5.2

Table S1 Crystallization conditions.

	k (s^{-1})	T_m (°C)
TmPep1050$_{12\text{-mer}}$ w/ Co^{2+}	118	96.6
Malic acid-treated TmPep1050$_{12\text{-mer}}$	0.3	91.2
Dialyzed TmPep1050$_{12\text{-mer}}$	0.9	92.6

Table S2 Effect of Co^{2+} removal on the LAP activity (k) and thermostability (T_m) of TmPep1050$_{12\text{-mer}}$. The malic acid treatment is described in Results and Discussion section. Similar results were obtained when TmPep1050$_{12\text{-mer}}$ was extensively dialyzed against 50 mM MOPS, 0.5 M $(NH_4)_2SO_4$ pH7.2. LAP activity was determined using 2.5 mM L-Leu-*p*NA as substrate. T_m were measured by thermal shift assay.

Oligonucleotide name	Sequence	Description
ocej419	5' – TTTAACTTTAAGAAG GAGATATACATACCCATG AAGGAACTGATCAGAAAG CTG	Amplification of *Tm_1050* ORF with 30 pb homology to arabinose promoter for homologous recombination.
ocej420	5' – ATCCGCCAAAACAGC CAAGCTGGAGACCGTTTA CGCCCCCAGATACCTGATG AG	Amplification of *Tm_1050* ORF with 30 pb homology to pBAD terminator for homologous recombination.
ocej687	5' – GAAAAAGGTGATACT GGACGCTgccATAGATGAG ATAGGTGTTGTCG	*Tm_1050* H60 codon mutation, forward primer complementary to ocej688.
ocej688	5' – CGACAACACCTATCT CATCTATggcAGCGTCCAG TATCACCTTTTTC	*Tm_1050* H60 codon mutation, reverse primer complementary to ocej687.
ocej689	5' – GTCTATTCCCACACG ATACGTTgccTCACCCAGT GAGATGATCGC	*Tm_1050* H307 codon mutation, forward primer complementary to ocej690.
ocej690	5' – GCGATCATCTCACTG GGTGAggcAACGTATCGT GTGGGAATAGAC	*Tm_1050* H307 codon mutation, reverse primer complementary to ocej689.
ocej691	5' – CTACGGTGTTTTCAG TGTTCAGcaaGAAGTGGG ACTGGTCGGTG	*Tm_1050* E197 codon mutation, forward primer complementary to ocej692.

ocej692	5' – CACCGACCAGTCCCA CTTCttgCTGAACACTGAA AACACCGTAG	*Tm_1050* E197 codon mutation, reverse primer complementary to ocej691.
ocej693	5' – GACACTCCGAAGGCC ATCAAGgcaCACGCAATGA GGCTCTCCG	*Tm_1050* R233 codon mutation, forward primer complementary to ocej694.
ocej694	5' – CGGAGAGCCTCATTG CGTGtgcCTTGATGGCCTTC GGAGTGTC	*Tm_1050* R233 codon mutation, reverse primer complementary to ocej693.
ocej695	5' – CAGACACTCCGAAGG CCATCgcgAGACACGCAAT GAGGCTCTCCG	*Tm_1050* K232 codon mutation, forward primer complementary to ocej696.
ocej696	5' – CGGAGAGCCTCATTGC GTGTCTcgcGATGGCCTTCGG AGTGTCTG	*Tm_1050* K232 codon mutation, reverse primer complementary to ocej695.
ocej699	5' – CCCGCTCTGAAAGTGA AAGACgcgGCATCGATCAG CAGCAAACG	*Tm_1050* R249 codon mutation, forward primer complementary to ocej700.
ocej700	5' – CGTTTGCTGCTGATCGA TGCcgcGTCTTTCACTTTCAG AGCGGG	*Tm_1050* R249 codon mutation, reverse primer complementary to ocej699.
ocej814	5' – Gtttaactttaagaaggagatataca tacccATGAAAGAGCTGATT CGTAAACTGACC	Amplification of synthetic *Tm_1050* gene harbouring K232, R233 and R249 mutations with 30 pb homology to arabinose promoter for homologous recombination.
ocej815	5' – catccgccaaaacagccaagctgga gaccgtTTATGCACCCAGATA GCGAATCAGC	Amplification of synthetic *Tm_1050* gene harbouring K232, R233 and R249 mutations with 30 pb homology to pBAD terminator for homologous recombination.

Table S3 List of the oligonucleotides used for *Tm_1050* ORF cloning and mutagenesis.

Plasmid name	Description
pCEC43	pBAD vector harbouring *Tm_1050* ORF for $TmPep1050_{12\text{-}mer}$ expression.
pCEC153	pCEC43 with mutagenized H60 codon for $TmPep1050_{H60A}$ expression.
pCEC154	pCEC43 with mutagenized H307 codon for $TmPep1050_{H307A}$ expression.
pCEC155	pCEC43 with mutagenized E197 codon for $TmPep1050_{E197Q}$ expression.

pCEC156	pCEC43 with mutagenized R233 codon for TmPep1050$_{R233A}$ expression.
pCEC157	pCEC43 with mutagenized K232 codon for TmPep1050$_{K232A}$ expression.
pCEC159	pCEC43 with mutagenized R249 codon for TmPep1050$_{R249A}$ expression.
pCEC194	pBAD vector harbouring synthetic *Tm_1050* gene with K232E, R233E and R249E for TmPep1050$_{K232E/R233E/R249E}$ expression.
pCEC195	pBAD vector harbouring synthetic *Tm_1050* gene with K232A, R233A and R249A for TmPep1050$_{K232A/R233A/R249A}$ expression.

Table S4 List of the plasmids used for expressing TmPep1050$_{12\text{-mer}}$ and its mutants.

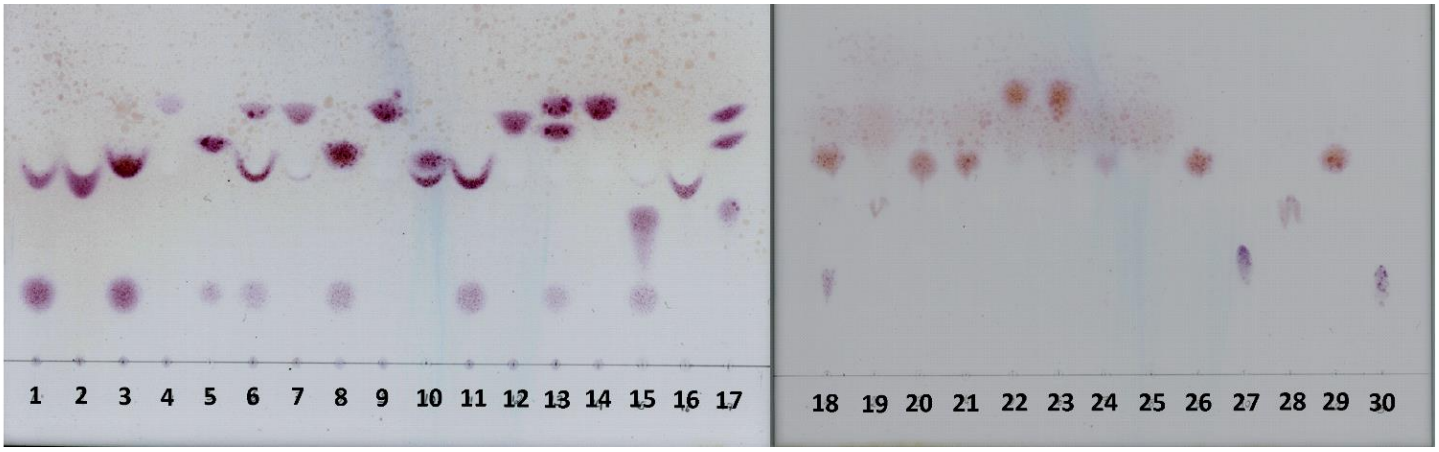


Figure 46 – Substrate specificity of TmPep1050 against Xaa-Ala dipeptides. 5 mM Xaa-Ala was incubated with 100 nM TmPep1050 in 50 mM MOPS pH 7.2 during 30 min. at 75°C. The hydrolysis products were analyzed by TLC using either 40:10:2 isopropanol:water:formic acid (lanes 1 to 17) or 50:30:10:10 tert-butanol:acetone:ammonia:water (lanes 18-30) as solvent. Amino acid spots were revealed using ninhydrin staining. For each peptide, two reactions were conducted in presence or in absence of enzyme. The two samples were deposited side-by-side, beginning from the left with the reaction in presence of enzyme. Lanes 1-2: Ala-Ala; lanes 3-4: Ile-Ala; lanes 6-7: Tyr-Ala; lanes 8-9: Leu-Ala; lanes 11-12: Met-Ala; lanes 13-14: Phe-Ala; lanes 15-16: His-Ala; lanes 18-19: Arg-Ala; lanes 20-21: Asp-Ala; lanes 22-23: Glu-Ala; lanes 24-25: Lys-Ala. Amino acid standards were deposited as: lane 5: Ile (upper spot) and Ala; lane 10: Leu (upper spot) and Tyr; lane 17: Phe (upper spot), Met, His (lower spot); lane 26: Ala; lane 27: Arg; lane 28: Asp; lane 29: Glu; lane 30: Lys.

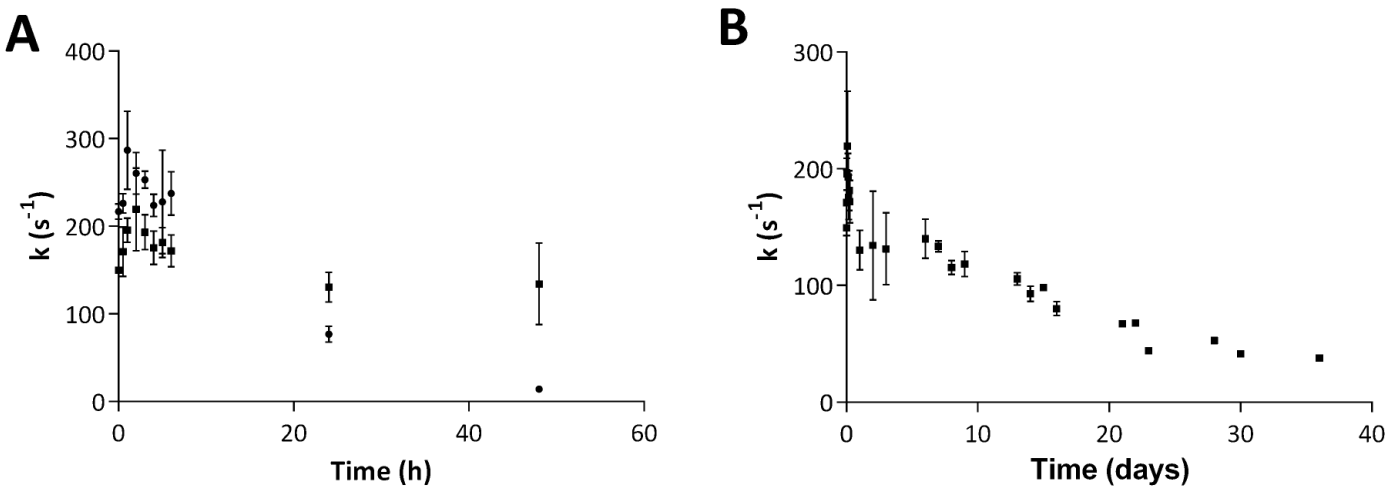


Figure 47 – Thermostability of TmPep1050 at 75°C and 95°C. The thermostability was monitored by measuring the activity (k) using L-Leu-*p*NA as substrate. k is expressed as µM of released *p*NA per second and per µM of TmPep1050. **(A)** Thermostability of TmPep1050 incubated at 75°C (squares) and 95°C (circles) over a period of 48 hours. **(B)** Thermostability of TmPep1050 incubated at 75°C over a period of 36 days.

B.1.5 Unpublished related results

Several experiments were not included in the article to avoid overcrowding the results and weakening the main message. Nonetheless, they are worth to be mentioned here as they bring a complementary insight. As shown in the article, TmPep1050 exhibits a high leucyl-aminopeptidase activity using amino acid *p*-nitroanilide derivatives. It degrades preferentially peptides beginning with, in order of preference, Leu, Ile, Val, and Met. No significant activity was observed with Glu-*p*NA or Lys-*p*NA. The substrate specificity of TmPep1050, however, appeared to be broader when assayed on dipeptides (see **Figure 46**). As expected, Leu-Ala, Ile-Ala, and Met-Ala peptides were completely degraded but TmPep1050 also hydrolyzed almost completely His-Ala peptide. To a lesser extent, TmPep1050 showed an activity against Ala-Ala, Tyr-Ala, Phe-Ala, Arg-Ala, and Lys-Ala peptides. Thus, the S_1 pocket appeared to accommodate amino acids with either an aromatic or a positively charged side chain. Nevertheless, TmPep1050 was found inactive on both Glu-Ala and Asp-Ala peptides. Somehow the *p*-nitroanilide group prevents the hydrolysis of some derivatives. Such a phenomenon has been reported for PhTET1, which cannot hydrolyze amino acid *p*-nitroanilide derivatives[326]. Consequently, a broader specificity could be expected for all M42 aminopeptidases. The activity of $TmPep1050_{12\text{-}mer}$ was also assayed on Met-Lys-Arg-Pro-Pro-Gly-Phe-Ser-Pro-Phe-Arg peptide (Met-Lys-bradykinin). This peptide was used previously to discriminate the activity of PhTET2 dimer and dodecamer in Appolaire *et al.* (2013)[419]. The PhTET2 dimer was unable to hydrolyze such a peptide, albeit both oligomers were shown to hydrolyze L-Leu-*p*NA with almost the same catalytic efficiency. The kinetic parameters of TmPep1050 were determined using Met-Lys-bradykinin by monitoring the release of methionine by HPLC. $TmPep1050_{12\text{-}mer}$ hydrolyzed Met-Lys-bradykinin with a k_{cat} of 2.6 ± 0.2 s^{-1} and a K_M of 3.1 ± 0.7 mM. Since $TmPep1050_{2\text{-}mer}$ was proven to be completely inactive, the activity assay on Met-Lys-bradykinin was not included in the article.

In the article, we presented the thermostability of TmPep1050, notably its half-life at 75°C and 95°C. During the experiment, the activity was enhanced by the temperature within the first hours of incubation (see **Figure 47**). This activity boost has also been reported for GsApI[369]. Anecdotally, TmPep1050 is extremely stable at 25°C as no activity loss was recorded after 700 days of incubation. Regarding Co^{2+} binding assays, we used a fluorescent probe, Amplex UltraRed, which reacts specifically with Co^{2+} at pH 9. During the article reviewing, a reviewer asked to determine the dissociation constant for Co^{2+} binding using either isothermal titration calorimetry or inductively coupled plasma mass spectrometry. In collaboration with George Feller (Uliège), we tried to measure Co^{2+} binding by ITC but failed to get an interpretable signal at 75°C.

To study the role of active site residues, six variants of TmPep1050 were produced, although only two of them were described in this article: $TmPep1050_{H60A}$ and $TmPep1050_{H307A}$. A third variant with both His-60 and His-307 substituted with alanine residues ($TmPep1050_{H60A\ H307A}$) was described in a methodological paper (see **section B.2**, page 68). A fourth variant with Asp-62 being substituted with an alanine residue ($TmPep1050_{D62A}$) was exploited in an article published in Proteins (see **section B.3**, page 83). The two remaining variants were not included in any article since their degree of oligomerization remained unaffected by the substitution. The first one has the catalytic base Glu-197 substituted with a glutamine residue. This variant was found completely inactive on L-Leu-*p*NA despite being a thermostable dodecamer (T_m of 97.2°C determined by thermal shift assay). The second one has Asp-168, the aspartate residue bridging both metal ions, substituted with an alanine residue. This variant was also inactive on L-Leu-*p*NA while still being a dodecamer. The Asp168Ala

substitution, however, had a dramatic impact on the variant thermostability, with a T_m of 55.2°C. We have no explanation for that result as we did not carry out further experiment with this variant.

B.2 X-Ray crystallogaphy to study oligomeric state transition of the *Thermotoga maritima* M42 aminopeptidase TmPep1050

B.2.1 Background

Following the publication of our article in JBC, an editor invited us to submit a methodology article in the Journal of Visualize Experiments (JoVE). Subsequently, several methods used in the precedent article were consolidated into a workflow for studying oligomeric state transition applied to TmPep1050. The proposed methodology covers various fields: gene cloning, protein purification, activity assays, dimer preparation, crystallogenesis, data indexation, and molecular replacement. It can be viewed as an accompanying article to the work published in JBC. Without being cutting-edge, this methodology is transposable in any laboratory as it does not require peculiar devices, except an access to a radiation facility (home source or synchrotron). The methods associated to protein crystallography could be insightful for non-crystallographers. Especially, the different steps are presented along with their related software. Software used in protein crystallography have become user-friendly thanks to the developers' impressive work. Beyond its use to study an M42 aminopeptidase, the methodology could be applied to any protein which oligomerization degree depends on metal ion availability.

Some unreleased data were added to illustrate the methodology, especially the structure of $TmPep1050_{H60A\ H307A}$ (with His-60 and His-307 substituted with alanine residues). This variant could indicate if dimers can further dissociate into monomers when metal ion binding is totally prevented. As discussed earlier, PfTET3 dodecamers were shown to fully dissociate into monomers when metal ions were depleted with EDTA[365]. The existence of PhTET2 monomer after metal ion removal, however, is not clearly established due to a discrepancy between two studies. By applying the same treatment (i.e. 0.5 M Zn^{2+} at pH < 4.5), Borissenko and Groll (2005) and Macek *et al.* (2017) reported contradictory results, with PhTET2 dodecamers dissociating into dimers and monomers, respectively[184,420]. On the other hand, Rosenbaum *et al.* (2011) showed that PhTET3 could dissociate into dimers after metal ion depletion[382]. Perhaps a further dissociation of PhTET3 dimers occurred after a prolonged heat-treatment of two hours at 90°C in presence of EDTA. The authors suggested that dimers could either unfold or dissociate into monomers. Of note, $TmPep1050_{2\text{-mer}}$ can be incubated for 24 hours at 75°C without any sign of dissociation being observed, as shown in our article published in JBC (see **section B.1.4**, page 45). Yet, a further dissociation of dimers into monomers could still occur since trace amount of metal ions was observed in dimers. Consequently, studying the oligomeric state of $TmPep1050_{H60A\ H307A}$ could demonstrate whether a complete dissociation into monomers may happen. Applying the methodology developed for studying the dimer-dodecamer transition of TmPep1050, we observed that $TmPep1050_{H60A\ H307A}$ is an inactive dimer presenting the same structural modifications as those of $TmPep1050_{2\text{-mer}}$. The methodology and its associated results were published in JoVE in May 2020.

B.2.2 Methodology

The methodology relies on a set of methods straight from the biochemistry toolbox. Each method is described step-by-step in the article. The pBAD vector was used for recombinant production in *E. coli*. The promotor of the *ara*BAD operon was preferred to that of T7 expression system as gene expression can be finely tuned with arabinose, allowing a moderate recombinant protein production. Gene cloning was performed by homologous recombination according to the SLiCE protocol (see section C.1.1, page 111)[439]. To generate TmPep1050 variants, site-directed mutagenesis was carried out using the SPRINP protocol[440]. Recombinant protein purification was achieved via a four-step protocol consisting of a heat-treatment, an anionic exchange chromatography, a hydrophobic interaction chromatography, and a size-exclusion chromatography. A heat-treatment was allowed as a pre-purification step since TmPep1050 is stable at high temperature. The last step is critical and must be undertaken with care considering (i) the chromatographic medium, (ii) the column size, (iii) the size and concentration of the sample, and (iv) the running conditions[429]. The aminopeptidase activity was assayed by monitoring the release of *p*NA resulting from hydrolysis of L-Xaa-*p*NA derivatives, the amide bound between the amino acid and *p*NA mimicking the peptide bound[441]. The sensitivity of this spectrophotometric method allows to determine *p*NA release rate as low as 0.4 µmol min^{-1}. Apo-enzyme can be prepared by incubating TmPep1050 with 2.1 M malic acid pH7.0, as described in our article published in JBC. This procedure, however, presents a drawback: concentrating the sample using an ultrafiltration device takes several hours due to a high viscosity. To speed up apo-enzyme preparation, 1,10-phenanthroline is a good alternative giving identical results to the procedure with malic acid. The activity was monitored to check the process efficiency.

The crystallogenesis of TmPep1050 is rather straightforward, relying on a screening using commercial kits to identify crystallization conditions. The identified crystallization conditions for either $TmPep1050_{12\text{-}mer}$ or $TmPep1050_{2\text{-}mer}$ were optimized by varying pH vs. precipitation agent concentration as a matrix fashion. Size and shape of crystals were improved using microseeding. The hanging drop method was used throughout the whole process of TmPep1050 crystallization (see section C.1.8, page 114). X-ray diffraction experiments were performed at synchrotron facilities. The oscillation range, i.e. the number of frames taken per degree of rotation, was a critical parameter to allow data indexation since the unit cell can be large in one direction (up to 285 Å for $TmPep1050_{12\text{-}mer}$). Usually, an oscillation range of 0.1° is enough to discriminate diffraction spots. Data sets were indexed using XDS applying the recommendations of Karplus and Diederichs (2015)[442,443]. Phaser and Phenix software were used for molecular replacement and model building, respectively[444,445]. The models were refined manually and automatically using Coot and Phenix software[445,446].

B.2.3 Article summary and discussion

To illustrate the methodology described in the article, we presented the characterization of $TmPep1050_{H60A\ H307A}$. This variant was found to be an inactive dimer, even in presence of Co^{2+}. Thus, a complete dissociation seems to be unlikely for TmPep1050 since stable monomers were not observed. As expected, $TmPep1050_{H60A\ H307A}$ crystallized in conditions similar to $TmPep1050_{2\text{-}mer}$. A data set was obtained at 2.37 Å of resolution. According to its indexation, the space group of $TmPep1050_{H60A\ H307A}$ crystal was potentially $C222_1$ with a likelihood of 0.711 according to Pointless. Of note, this space group was observed for $TmPep1050_{2\text{-}mer}$ crystals. After indexation in $C222_1$, the data set was analyzed by Xtriage, revealing a pseudo-merohedral twinning. Twinning results from

crystal growth anomaly, like two crystals growing in such a manner that some of their lattice directions are parallel[447,448]. Since twinning is considered unlikely in high symmetry space groups, the data set was indexed again in the space group $P2_1$, having a likelihood of 0.149[449]. Both indexations were tried in molecular replacement, but a solution was found only for the indexation in the space group $P2_1$, strongly supporting the existence of a pseudo-merohedral twinning. The data set indexed in that space group was subsequently used to solve the structure of $TmPep1050_{H60A\ H307A}$.

As expected from its oligomerization state, the $TmPep1050_{H60A\ H307A}$ structure is highly similar to that of $TmPep1050_{2\text{-}mer}$. Indeed, the structural modifications inherent to the degree of oligomerization are observed in $TmPep1050_{H60A\ H307A}$: a high flexibility of the α8 and α10 helices, the Gln-196–Val202 loop displaced outwards the catalytic site, and two misplaced segments involved in oligomerization (Lys-229–Ala-235 and Lys-247–Ser-254). Although being similar, two slight differences were observed between $TmPep1050_{H60A\ H307A}$ and $TmPep1050_{2\text{-}mer}$, highlighting the structural role of His-60 and His-307. Firstly, the carboxylate of Asp-168 is rotated by 40° in the variant compared to the wild-type dimer, indicating that both histidine residues impose the position of Asp-168 side chain. Secondly, the side chain of Asp-62 is rotated by about 98° in $TmPep1050_{H60A\ H307A}$, with the carboxylate oriented towards Glu-18 instead of the catalytic site. Our structural data suggests that Asp-62 is related to His-60 and His-307 via a H-bond network. The role of Asp-62 will be discussed in the third article (see **section B.3**, page 83).

B.2.4 The article

jove Journal of Visualized Experiments www.jove.com

Video Article

X-Ray Crystallography to Study the Oligomeric State Transition of the *Thermotoga maritima* M42 Aminopeptidase TmPep1050

Raphael Dutoit[1,2], Nathalie Brandt[2], Dany Van Elder[1], Louis Droogmans[1]

[1]Laboratory of Microbiology, Department of Molecular Biology, Université Libre de Bruxelles

[2]Labiris Institut de Recherche

Correspondence to: Raphael Dutoit at Raphael.Dutoit@ulb.ac.be

URL: https://www.jove.com/video/61288
DOI: doi:10.3791/61288

Keywords: protein oligomerization, M42 aminopeptidase, oligomeric state transition, metalloenzyme, protein purification, protein crystallization, X-ray crystallography, data processing, molecular replacement

Date Published: 5/11/2020

Citation: Dutoit, R., Brandt, N., Van Elder, D., Droogmans, L. X-Ray Crystallography to Study the Oligomeric State Transition of the *Thermotoga maritima* M42 Aminopeptidase TmPep1050. *J. Vis. Exp.* (), e61288, doi:10.3791/61288 (2020).

Abstract

The M42 aminopeptidases form functionally active complexes made of 12 subunits. Their assembly process appears to be regulated by their metal ion cofactors triggering a dimer-dodecamer transition. Upon metal ion binding, several structural modifications occur in the active site and at the interaction interface, shaping dimers to promote the self-assembly. To observe such modifications, stable oligomers must be isolated prior to structural study. Reported here is a method that allows the purification of stable dodecamers and dimers of TmPep1050, an M42 aminopeptidase of *T. maritima*, and their structure determination by X-ray crystallography. Dimers were prepared from dodecamers by removing metal ions with a chelating agent. Without their cofactor, dodecamers became less stable and were fully dissociated upon heating. The oligomeric structures were solved by the straightforward molecular replacement approach. To illustrate the methodology, the structure of a TmPep1050 variant, totally impaired in metal ion binding, is presented showing no further breakdown of dimers to monomers.

Introduction

Oligomerization is a predominant process that dictates the biological functions of many proteins. In *Escherichia coli*, it is estimated that only 35% of proteins are monomeric[1]. Some proteins, called morpheeins, can even adopt several oligomeric states with subunits having distinct structure in each oligomeric state[2]. The transition between their oligomeric states is often a mean to regulate protein activity as each oligomeric state may have a different specific activity or function. Several examples of morpheeins have been well-documented in literature, notably the porphobilinogen synthase[3], HPr kinase/phosphatase[4], Lon protease[5], lactate dehydrogenase[6], glyceraldehyde-3-phosphate dehydrogenase[7], pyruvate kinase[8], citrate synthase[9], and ribonuclease A[10]. Recently, we described the M42 aminopeptidase TmPep1050, another example of enzyme with morpheein-like behavior, whose activity depends on its oligomeric states[11]. The transition between its oligomeric states is mediated by its metallic cofactors which induce several structural modifications of the subunits.

The M42 aminopeptidase family belongs to the MH clan[12,13], and is widely distributed among Bacteria and Archaea[14]. The M42 aminopeptidases are genuine dinuclear enzymes degrading peptides up to 35 amino acid residues in length[15]. They adopt a peculiar tetrahedron-shaped structure made of 12 subunits with their active sites oriented towards an inner cavity. Such an arrangement is often described as a nano-compartmentalization of the activity to avoid uncontrolled proteolysis. The physiological function of the M42 aminopeptidases may be associated with the proteasome, hydrolyzing peptides resulting from protein degradation[16,17]. *Pyrococcus horikoshii* possesses four M42 aminopeptidases, each presenting distinct but complementary specificities[18,19,20,21]. Singularly, heterocomplexes made of two different types of subunits have been described in *P. horikoshii*, suggesting the existence of peptidasome complexes[22,23].

Several structures of M42 aminopeptidases have been described in the literature[11,16,18,19,20,24,25,26]. The subunit is composed of two distinct domains, a catalytic domain and a dimerization domain. The catalytic domain adopts a common α/β fold conserved in the whole MH clan, the archetypal catalytic domain being the aminopeptidase Ap1 of *Vibrio proteolyticus*[27]. The dimerization domain adopts a PDZ-like fold[16] and may have, in addition to its role in the oligomerization, a role in controlling substrate access and binding in the inner cavity[11]. As the basic building block is a dimer, the dodecamer is often described as the association of six dimers, each dimer being positioned at each edge of the tetrahedron[16]. The oligomerization of the M42 aminopeptidases relies on the availability of its metal cofactors. Divalent metal ions, often Zn^{2+} and Co^{2+}, are catalytically involved in the peptide binding and hydrolysis. They are found in two distinct binding sites, namely M1 and M2 sites. The two metal ions also drive and finely tune the oligomerization as demonstrated for PhTET2, PhTET3, PfTET3, and TmPep1050[11,24,28,29]. When the metal cofactors are depleted, the dodecamer disassembles into dimers, like in PhTET2, PhTET3, and TmPep1050[11,16,28], or even monomers, like in PhTET2 and PfTET3[24,29].

Presented here is a protocol used for studying the structures of TmPep1050 oligomers. This protocol is a set of common methods including protein purification, proteolytic activity screening, crystallization, X-ray diffraction, and molecular replacement. Subtleties inherent to dealing with metalloenzymes, protein oligomerization, protein crystallization and molecular replacement are emphasized. A case of study is also presented to show whether TmPep1050 dodecamers may further dissociate into monomers or not. To address this question, a TmPep1050 variant,

TmPep1050$_{H60A\ H307A}$, has been studied whose metal binding sites are impaired by mutating His-60 (M2 site) and His-307 (M1 site) to Ala residues. This protocol may be accommodated to study other M42 aminopeptidases or any metalloenzymes with morpheein-like behavior.

Protocol

1. Production and purification of recombinant TmPep1050

NOTE: Hereafter are described the cloning procedure and purification of wild-type TmPep1050 adapted from a previous study[11]. Alternatively, the cloning can be done using a synthetic gene. To generate TmPep1050 variants, site directed mutagenesis can be performed following, for instance, the single-primer reactions in parallel protocol (SPRINP) method[30]. The purification protocol can be used for TmPep1050 variants. The use of His-tag should be avoided as it interferes with metal ion binding.

1. Expression vector design
    1. Acquire genomic DNA of *Thermotoga maritima* MSB8 (ATCC 43589) or TmCD00089984 (Joint Center for Structural Genomics).
    2. Amplify TM_1050 open reading frame (ORF) using either genomic DNA or template plasmid, a high-fidelity DNA polymerase, and the following primers: ocej419 (5'- TTTAACTTTAAGAAGGAGATATACATACCCATGAAGGAACTGATCAGAAAGCTG) and ocej420 (5'- ATCCGCCAAAACAGCCAAGCTGGAGACCGTTTACGCCCCCAGATACCTGATGAG). Run the polymerase chain reaction (PCR) screening according to the following scheme: 5 min at 95 °C, 30 cycles of 3 steps (30 s at 95 °C, 30 s at 55 °C, 90 s at 72 °C), and 10 min at 72 °C as the final step.
    3. Clone the PCR fragment into a suitable expression vector (**Table of Materials**) by homologous recombination (**Figure 1**) in *E. coli* according to the SLiCE protocol[31]. To 50 ng of linearized vector, add PCR fragment in a 10:1 molar ratio of fragment to vector, 1 µL of PPY strain extract, 50 mM Tris-HCl pH 7.5, 100 mM $MgCl_2$, 10 mM ATP, and 10 mM dithiothreitol (DTT) for a reaction volume of 10 µL. Incubate for 1 h at 37 °C.
    4. Transform chemically competent *E. coli* XL1 blue strain (or any *rec*A$^-$ suitable strain) with 1 µL of recombination reaction. Plate the cells on LB medium containing 100 µg/mL ampicillin. Incubate the plates overnight at 37 °C.
    5. Pick colonies on fresh LB plates with 100 µg/mL ampicillin. Incubate the plates at 37 °C for a least 8 h.
    6. Screen for positive candidates by colony PCR using suitable primer pair (5'- ATGCCATAGCATTTTTATCC and 5'- ATTTAATCTGTATCAGGC if using the recommended vector listed in **Table of Materials**). With a microtip end, scratch a picked colony and transfer the cells to 20 µL of reaction mix containing 0.5 µM of each primer and 10 µL of a commercial Taq DNA polymerase mix.
    7. Run the PCR screening according to the following scheme: 5 min at 95 °C as the denaturation step, 30 cycles of 3 steps (30 s at 95 °C, 30 s at 55 °C, 90 s at 72 °C), and 10 min at 72 °C as the final step.
    NOTE: The PCR reactions can be stored overnight in the PCR machine at 12 °C.
    8. Load 10 µL of each PCR reaction on a 0.8% agarose gel prepared in Tris-acetate-EDTA (TAE) buffer. Run the electrophoresis for 25 min at 100 V.
    NOTE: A 1.1 kbp amplicon is expected.
    9. Extract plasmids from candidates using a commercial kit (**Table of Materials**) and sequence them using the same primer pair used in step 1.1.6.

2. Cell culture
NOTE: When a suitable candidate has been identified by sequencing, the clone can be directly used as the expression if using the recommended vector (**Table of Materials**). In that case, the expression is controlled by the arabinose-inducible P_{BAD} promoter[32].
    1. Inoculate 10 mL of LB medium containing 100 µg/mL ampicillin with the candidate and incubate the preculture overnight at 37 °C under orbital shaking. Add 5 mL of the preculture to 1 L of LB medium with 100 µg/mL ampicillin. Mind to respect an air to liquid ratio of 3.
    2. Let cells grow at 37 °C under orbital shaking. Monitor the optical density at 660 nm (OD_{660}).
    3. When OD_{660} has reached 0.5−0.6, rapidly cool the culture for 5 min on ice and transfer it to an incubator set to 18 °C.
    4. Add 0.2 g/L arabinose to induce gene expression and incubate for 12−18 h at 18 °C.
    5. Harvest cells by centrifuging the culture at 6,000 x *g* for 30 min at 4 °C. Discard the supernatant and wash cells with 100 mL of 0.9% (w/v) NaCl.
    6. Centrifuge again at 6,000 x *g* for 15 min at 4 °C and discard the supernatant.
    NOTE: Cell pellets can be used directly for protein extraction or stored at -80 °C.

3. Protein purification
    1. Resuspend the cell pellets in 40 mL of 50 mM MOPS, 1 mM $CoCl_2$, pH 7.2. Add 1 µL of 25 U/µL DNA/RNA endonuclease and one tablet of protease inhibitor cocktail that does not contain EDTA. Sonicate the suspension in pulse mode under cooling for 30 min.
    2. Centrifuge the crude extract at 20,000 x *g* for 30 min at 4 °C. Collect the supernatant and heat it in a water bath at 70 °C for 10 min.
    3. Centrifuge the denatured cell extract at 20,000 x *g* for 30 min at 4 °C and collect the supernatant for purification.
    4. Use suitable anion exchange resin (**Table of Materials**) packed in a column of ~15 mL of volume. Refer to manufacturer's recommendations for the working flow rate and column pressure limit. Equilibrate the resin with 50 mM MOPS, 1 mM $CoCl_2$, pH 7.2.
    5. Load the supernatant collected from step 1.3.3 into the column. Monitor the absorbance of the eluate at 280 nm. When it has reached the baseline, proceed to the elution.
    6. Apply a gradient from 0 to 0.5 M NaCl in 50 mM MOPS, 1 mM $CoCl_2$, pH 7.2 for 5 column volumes (CV). Wait till the conductivity is stabilized and the absorbance has reached the baseline.
    7. Apply a final gradient from 0.5 to 1 M NaCl in 50 mM MOPS, 1 mM $CoCl_2$, pH 7.2 for 1 CV.
    8. Analyze some fractions (see **Figure 2A** for guidance) by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).
    NOTE: TmPep1050 appears as a 36 kDa band after Coomassie staining. Alternatively, the presence of TmPep1050 can be confirmed by activity assay (see section 2.1). At this step, fractions can be stored at 4 °C overnight.
    9. Pool the fractions containing TmPep1050 and add finely ground powder of $(NH_4)_2SO_4$ to obtain a concentration of 1.5 M $(NH_4)_2SO_4$. Mix gently by inverting the tube upside down till complete dissolution.

10. Use hydrophobic interaction resin (**Table of Materials**) packed in a column of ~30 mL of volume. Refer to manufacturer's recommendations for the working flow rate and column pressure limit. Equilibrate the resin with 50 mM MOPS, 1.5 M $(NH_4)_2SO_4$, 1 mM $CoCl_2$, pH 7.2.
11. Load the sample onto the column and monitor the absorbance of the eluate at 280 nm. When the absorbance has reached the baseline, elute bound proteins by applying a gradient from 1.5 M to 0 M $(NH_4)_2SO_4$ in 50 mM MOPS, 1 mM $CoCl_2$, pH 7.2 for 5 CV.
12. Analyze some fractions (see **Figure 2B** for guidance) by SDS-PAGE.
NOTE: TmPep1050 appears as a 36 kDa band after Coomassie staining. Alternatively, the presence of TmPep1050 can be confirmed by activity assay (see section 2.1). At this step, fractions can be stored at 4 °C overnight.
13. Pool the fractions containing TmPep1050 and concentrate to 2 mL using ultrafiltration units with 30 kDa cutoff (**Table of Materials**). Proceed to section 1.4 to determine the molecular weight.

4. Size exclusion chromatography
    1. Use size exclusion resin (**Table of Materials**) packed in a column of ~120 mL of volume. Refer to manufacturer's recommendations for the working flow rate and column pressure limit. Equilibrate the resin with 50 mM MOPS, 0.5 M $(NH_4)_2SO_4$, 1 mM $CoCl_2$, pH 7.2.
    2. Load the sample onto the column and monitor the absorbance of the eluate at 280 nm. Fractionate from the column dead volume (~0.33 CV) until the end of the elution (1 CV).
    3. Measure the elution volume for each observed peak.
    NOTE: For guidance, dodecameric TmPep1050 elutes at ~82 mL (**Figure 3A**) under current experimental conditions, while dimeric TmPep1050, such as the $TmPep1050_{H60A\ H307A}$ variant, elutes at ~95 mL (**Figure 3B**). Some TmPep1050 may adopt both oligomeric forms, such as $TmPep1050_{H60A}$ (**Figure 3C**).
    4. Analyze fractions corresponding to the maxima and tails of observed peaks using SDS-PAGE.
    NOTE: TmPep1050 appears as a 36 kDa band after Coomassie staining.
    5. Pool the fractions of each peak and concentrate using ultrafiltration units with 30 kDa cutoff (**Table of Materials**) to obtain a concentration of ~300 µM.
    6. Measure the absorbance at 280 nm on a nano-volume spectrophotometer and calculate the concentration using the molecular extinction coefficient of 18,910 M^{-1} cm^{-1}.
    7. Store the purified protein at -18 °C.
    8. To determine the molecular weight, calibrate the size exclusion chromatography (SEC) column using molecular weight standards (**Table of Materials**). Analyze the standards using 50 mM MOPS, 0.5 M $(NH_4)_2SO_4$, 1 mM $CoCl_2$ pH 7.2 as the running buffer.

2. Activity assay and apo-enzyme preparation

NOTE: Originally, the apo-enzyme was prepared by diluting 1 volume of TmPep1050 in 10 volumes of 2.1 M malic acid pH 7.0 and concentrating back to 1 volume prior to dialysis[11]. Below is presented an alternative procedure using 1,10-phenanthroline, a metal ion chelator. This procedure reduces protein loss and gives the same results than the previously published method.

1. Activity assay
    1. Prepare a stock solution of 100 mM L-Leucine-*p*-nitroanilide (**Table of Materials**) in methanol.
    2. Add 25 µL of 100 mM L-Leucine-*p*-nitroanilide in 965 µL of 50 mM MOPS, 250 µM $CoCl_2$, pH 7.2, 10% methanol. Preincubate the reaction mix at 75 °C in a dry bath.
    3. Dilute the enzyme in 50 mM MOPS pH 7.2 to a concentration of 1 µM. Add 10 µL to the reaction mix, vortex, and incubate at 75 °C either until it has turned yellowish or for 1 h.
    4. Stop the reaction by adding 1 mL of 20% acetic acid. Vortex well and let it cool down to room temperature.
    5. Transfer the reaction mix in a spectrophotometer cell. Read the absorbance at 410 nm against a negative control (incubated reaction mix without enzyme).

2. Apo-enzyme preparation
    1. Prepare a stock solution of 1 M 1,10-phenanthroline in ethanol. Add 10 µL of 1,10-phenanthroline stock solution to 890 µL of 50 mM MOPS, 0.5 M $(NH_4)_2SO_4$, pH 7.2. Add 100 µL of purified TmPep1050 (300 µM−1 mM concentration).
    2. Check the activity loss using the activity assay described in section 2.1 without adding $CoCl_2$ in the reaction mix.
    3. Transfer the sample in a dialysis tube. Dialyze against 200 mL of 50 mM MOPS, 0.5 M $(NH_4)_2SO_4$, pH 7.2 at 4 °C. Exchange thrice the dialysate with fresh buffer during the 48 h dialysis.
    4. Collect the sample from the dialysis tube and concentrate back to 100 µL using ultrafiltration units with 30 kDa cutoff (**Table of Materials**). Check the concentration by reading the absorbance at 280 nm using a nano-volume spectrophotometer.

3. Dimer preparation
    1. Dilute the apo-enzyme to a concentration of 1 µM in 50 mM MOPS, 0.5 M $(NH_4)_2SO_4$, pH 7.2. Incubate for 2 h at 75 °C in a dry bath, then let the sample cool down to room temperature.
    2. Concentrate the sample to an enzyme concentration of at least 50 µM. Check the molecular weight by SEC (see section 1.4). The elution peak must shift from ~82 mL to ~95 mL (under current experimental conditions).

3. TmPep1050 crystallization

NOTE: Protein crystallization remains an empirical science as it is a multifactorial phenomenon[33]. While some parameters can be identified and controlled (such as temperature, pH, precipitation agent concentration), others may influence elusively the crystallization (such as protein and chemical purity, proteolysis, sample history). Nowadays protein crystallization is tackled in a rational and systematic manner thanks to a bunch of

commercial crystallization screening conditions and automation. The optimization of a crystallization condition, however, relies mostly on a trial-and-error approach. Hereafter are described a blueprint for crystallizing proteins and several tips for optimizing the crystallization conditions.

1. Crystallization screening
   NOTE: Using commercial crystallization kits, crystals of dodecameric TmPep1050 have been obtained in 2.2 M DL-malic acid pH 7.0, 0.1 M Bis-Tris propane pH 7.0 and 0.18 M tri-ammonium citrate, 20% polyethylene glycol (PEG) 3350. Crystals of dimeric TmPep1050 have been obtained in 0.1 M sodium citrate pH 5.6, 0.2 M ammonium acetate, 30% PEG4000. Crystals of dodecamers appear within a week and reach their full size in a month. Crystals of dimers usually appear within 24 h and grow to full size in a week.
   1. Acquire several commercial crystallization kits (see **Table of Materials** for examples).
   2. Set up crystallization plates (**Table of Materials**) for the hanging drop method. Fill the wells with 500 µL of each solution of a crystallization screening kit.
   3. For each well, set up a crystallization support. On the support, deposit a 1 µL drop of purified protein (usually ~10 mg/mL).
   4. Immediately pipet 1 µL of crystallization solution from the well. Add it carefully to the protein drop and mix gently by pipetting upside down thrice. The drop must remain semispherical without any bubbles.
   5. Screw the support on top of the corresponding well. Repeat the operation for the whole kit.
   6. After setting up the plates, observe each drop with a binocular. Refer to the crystallization kit user guide for interpretation (clear drop, phase separation, precipitate, needles, etc.).
   7. Incubate the plates at 20 °C. Check the plates once per day during the first week and once per week afterwards.
   8. Score each well using the score sheet and the user guide provided with the crystallization kits.

2. Crystallization optimization
   NOTE: The initial crystallization conditions of dodecameric TmPep1050 have been optimized to 2.1 M DL-malic acid pH 6.75 and 0.18 M tri-ammonium citrate pH 7.5, 40% (w/v) PEG3350 while the crystallization condition of dimeric TmPep1050 has been shifted to 0.1 M sodium citrate pH 6.0, 10% (w/v) PEG3350. One cycle of seeding has been necessary to improve crystallinity. Hereafter is described how the crystallization of $TmPep1050_{H60A\ H307A}$ variant has been optimized.
   1. Prepare stock solutions of 0.5 M sodium citrate buffer at different pH (4.5, 5.2, and 6.0), and 50% (w/v) PEG3350 solution.
   2. Set up a crystallization plate as a matrix of pH vs. precipitation agent (see **Figure 4A**).
   3. Incubate the plate at 20 °C. Observe each well with a binocular once per day during a week.
   4. Score each well according to crystal size and shape. Select a condition giving crystals of at least 50 µm. Proceed to microseeding.

3. Microseeding
   NOTE: Microseeding is a powerful method to improve the shape, size and crystallinity of protein crystals[34]. A faster seeding approach is streak seeding using a cat whisker. See **Figure 4** as an example how crystallization optimization and microseeding have improved the crystal shape and size for $TmPep1050_{H60A\ H307A}$.
   1. Prepare the seeds using the selected well in step 3.2.4.
   2. Increase the volume of a drop containing crystals to 10 µL by adding crystallization solution from the well. Pipet the drop and add 90 µL of crystallization solution from the well. Vortex thoroughly and keep the seeds on ice.
   3. Prepare several dilutions of the seeds: 1x, 10x, 25x, and 100x. Vortex well the seeds before pipetting. Keep the dilutions on ice.
   4. For each seed dilution, set up a crystallization plate as a matrix of pH vs. precipitation agent (see **Figure 4B**). Use the stock solutions prepared in step 3.2.1.
   5. When making the drop, add 0.2 µL of seeds for a 2 µL drop. Incubate the plate at 20 °C. Observe each well with a binocular once per day during a week.
      NOTE: Crystal size distribution and shape must be improved, see **Figure 4C** as an example for $TmPep1050_{H60A\ H307A}$. For further uses, seeds can be stored at -20 °C.

4. X-ray diffraction

1. Crystal picking
   NOTE: Sample preparation depends on the X-ray source facility (home facility vs. synchrotron). Use storage devices (vials and vial-holder basket) accordingly. The addition of cryoprotectant (such as glycerol) may be required depending on salt/precipitation agent concentration. For TmPep1050 crystals, a cryoprotectant is not necessary as PEG or buffer concentration is high enough to avoid water crystals.
   1. Prepare a bath filled with liquid nitrogen, plunge any vials or basket used for sample handling.
   2. Set up sample picking loops of different sizes: 100, 150, and 200 µm (**Table of Materials**). Choose the sizes according to the crystal size.
   3. Using a binocular, check the drop containing crystals and spot isolated crystals (the easiest to pick). With a loop, gently pick a crystal from the bottom. Immediately plunge the loop in liquid nitrogen and place the loop in a suitable vial.

2. Data collection
   NOTE: Data collection may vary greatly depending on the X-ray source (home facility vs. synchrotron) and detector sensitivity. The collection strategy may also differ greatly from a sample to another, depending on the resolution, spot intensity, space group, etc. The topic has been extensively reviewed by Dauter[35].
   1. Mount the loop carrying the crystal on the goniometer head of the diffractometer.
   2. Tune the goniometer head along XYZ axes to align the crystal with the X-ray beam path.
   3. Set the wavelength to 0.98 Å and move the detector to get 2 Å of resolution.
   4. Start a short data collection by acquiring images in at least two different crystal orientations. Take 10 images (1 image per 0.1°) at 0° and 90°.
   5. Check the collected images with suitable software (e.g., ADXV, XDS-Viewer or Albula Viewer). Determine the spot intensity and the highest resolution where spots are seen. Check also the monocrystallinity and spot separation.

6. Eventually, repeat steps 4.2.3−4.2.5 by changing the detector position for higher or lower resolution and the exposure time in accordance to the observation.
7. Start data collection around 360° with 1 image taken per 0.1°. Remember to set the detector position and exposure time optimally.

5. Indexation, molecular replacement and model building

1. Indexation
   NOTE: Indexation is a method for measuring diffraction spots intensity, giving the amplitudes of structure factors[36]. Four software packages are commonly used for processing collected images: Mosflm[37], HKL2000[38], DIALS[39], and XDS[40]. The latter has been used for indexing the data sets obtained from TmPep1050 crystal diffraction.
   1. Install XDS package and XDSME[41]. If processing HDF5 files, install XDS Neggia plugin (available on Dectris website). For more information, visit XDS wiki web page https://strucbio.biologie.uni-konstanz.de/xdswiki/index.php/Main_Page and XDSME web page https://github.com/legrandp/xdsme.
   2. Before processing data, create a folder from where XDS will be run. Locate the path to the images.
   3. To run XDSME, type **xdsme /path_to_images/image.extension** in a terminal window.
   4. After XDS has ended the job, check the CORRECT.LP file. Note the probability of the space group determination, data completeness, the highest resolution, crystal mosaicity, and data quality. Check also XDS_pointless.log to obtain the likelihood of space groups.
      NOTE: See **Figure 5** as an example of output.
   5. Rerun XDSME with different space group solutions proposed by XDS in separate folder to avoid overwriting the previous process. Type **xdsme -s space_group_name -c "unit_cell_parameters" /path_to_images/image.extension** (e.g., **xdsme -s P21 -c "43.295 137.812 61.118 90.000 110.716 90.000**).
   6. Check the CORRECT.LP files and choose the best solution based on data statistics.
   7. Run XSCALE by typing **xscale.py XDS_ASCII.HKL**. Run XDSCONV by typing **xdsconv.py XSCALE.HKL ccp4**.
      NOTE: In some cases, XDSME fails to identify the space group or fails to cut the resolution range properly or generates weird data statistics. If such a problem is encountered, it is worth to run XDS natively. Several parameters must be introduced in the XDS.INP initiation file (see XDS wiki page). When using XDS, the likelihood of possible space groups can be checked by using Pointless, part of CCP4 package[42]. To cut the data set resolution, R_{meas} < 60% and I/σ ~2 are commonly accepted to determine the highest resolution[43]. The molecular replacement and model refinement, however, can be improved by extending the resolution to I/σ ~0.5−1.5 and $CC_{1/2}$ down to 0.2−0.4[44].

2. Molecular replacement
   NOTE: Experimental data give access to the amplitude of structural factors but, without knowing the phase, they are useless. The phase can be determined experimentally by different methods relying on an anomalous signal (from a heavy atom, for instance)[45]. Molecular replacement is another method for determining the phase without an anomalous scattering atom[46,47]. This method uses the coordinates of a related molecule to find and improve the phase iteratively. We use Phaser[48] in Phenix GUI[49] for molecular replacement.
   1. Prepare the starting model for molecular replacement using 4P6Y coordinates. From the pdb file, extract the monomer A and truncate its aminoacids in alanine using the PDB file editor in Phenix (under the **Model tools** tab).
   2. Run Xtriage in Phenix (under the **Data analysis** tab) with the reflection file generated by XDSCONV (5.1.9) and the sequence as inputs.
   3. Check the log file from Xtriage. Note the completeness, the number of subunits in the asymmetric unit, the anisotropy, the presence of ice rings, and twinning occurrence.
   4. Run Phaser-MR in Phenix (under the **Molecular replacement** tab) for molecular replacement using the reflection file, the sequence and the starting 4P6Y model truncated in poly alanine (step 5.2.1).
   5. Upon completion, check if a model has been found and the score of the molecular replacement. A translation factor Z-score (TFZ) of at least 8 indicates that the solution is definitively correct.

3. Model building
   NOTE: After determining the phase by molecular replacement, the model must be built and refined. This protocol uses Phenix GUI[49] for automatic building and iterative refinements, and Coot[50] for manual structure building and refinement.
   1. After molecular replacement using Phase-MR in Phenix, select **Run Autobuild**. All the required files will be automatically added. Simply press **Run** to start autobuild.
   2. Upon completion, check the model in Coot. Build and refine the model manually according the electron density map in Coot.
   3. Refine the manually curated model in Phenix (in the **Refinement** tab) using the model, the sequence, and the diffraction data as inputs. Refer to Phenix help to choose the right strategy.
   4. After refinement, check the results: R_{free} and R_{work} must decrease, Molprobity[51] indicators must be respected, and outliers with low real-space correlation must be limited.
   5. Repeat steps 5.3.2−5.3.4 until the best refined model is generated.
   6. Run Molprobity on the server: http://molprobity.biochem.duke.edu/. Check any outliers identified by Molprobity.
   7. Eventually repeat steps 5.3.2−5.3.6 until the best refined model is obtained.

Representative Results

To study a possible dodecamer dissociation into monomers in TmPep1050, the His-60 and His-307 codons were replaced by alanine codon using a synthetic gene. This gene was then cloned in pBAD vector for expression and purification of the corresponding TmPep1050 variant subsequently named $TmPep1050_{H60A\ H307A}$. Size exclusion chromatography (**Figure 3B**) showed that the purified protein had an apparent molecular weight of 56 kDa (molecular weight of the monomer being 36.0 kDa). A similar apparent molecular weight, 52 kDa, has been reported for TmPep1050 dimer[11]. Hence, the oligomeric state of $TmPep1050_{H60A\ H307A}$ could be inferred as dimeric. Regarding its specific activity,

TmPep1050$_{H60A\ H307A}$ was completely inactive on L-Leu-*p*NA as substrate, even in the presence of cobalt ions. This result strongly suggests that the variants cannot bind any metal ions.

The crystallization condition of TmPep1050$_{H60A\ H307A}$ was optimized by varying pH vs. PEG concentration (**Figure 4**) around the condition of the dimer (i.e., 0.1 M sodium citrate pH 6.0 10% PEG3350). The best crystals of TmPep1050$_{H60A\ H307A}$ were obtained in 0.1 M sodium citrate pH 5.2 20% PEG3350 with one cycle of microseeding for improving monocrystallinity. A complete data set was collected at Proxima 2 beamline (SOLEIL synchrotron) at a resolution 2.36 Å (**Table 1**). Data indexation showed that the space group of the TmPep1050$_{H60A\ H307A}$ crystal is $C222_1$ but XDS proposed another solution, the *m*P space group (see **Figure 5**). According to Pointless, the likelihood of $C222_1$ and $P2_1$ space groups were 0.711 and 0.149, respectively. According to the data quality analysis, two monomers are found in the asymmetric unit. The analysis by Xtriage revealed that the data set is probably twinned but twinning in $C222_1$ space group is unlikely[52]. Twinning results from crystal growth anomaly where several definite domains have some of their lattice directions parallel to each other[53]. Twinning may also result from a higher crystal symmetry, indicating an erroneous data indexation. Hence, a pseudo-merohedral twin may exist so that a $P2_1$ crystal lattice looks like a $C222_1$. The data set was subsequently indexed in space group $P2_1$ and tested in molecular replacement. Xtriage analysis of the data set indexed in $P2_1$ revealed a pseudo-merohedral twin following a twin law *h, -k, -h-l*.

Using the coordinates of a monomer from dodecameric TmPep1050 (PDB code 4P6Y), a molecular replacement solution was found for the data set indexed in $P2_1$ only, with a TFZ score of 28.9. Therefore, the diffraction data were treated as a twinned data set for model building. To minimize the bias of molecular replacement, a first model was built by using phenix.autobuild[54,55]. The structure of TmPep1050$_{H60A\ H307A}$ was completed after several cycles of automated and manual refinement in Phenix and Coot (**Table 1** and **Figure 6A**). The structure confirmed the oligomeric state with an interface surface of 1,710 Å^2 between both monomers and a Δ^iG of -16.2 kcal mol^{-1} as calculated by PDBe Pisa[56]. In comparison, the interface surface and Δ^iG of dimeric TmPep1050$_{2\text{-mer}}$ is 1,673 Å^2 and -16.7 kcal mol^{-1}, respectively.

The structure of TmPep1050$_{H60A\ H307A}$ is highly similar to the wild-type dimer structure with RMS of 0.774 Å upon alignment. Importantly, the same structural modifications are observed in both structure: high flexibility of the α8 and α10 helices, disordered active site Gln-196–Val-202, and the displacement of Lys-229–Ala-235 and Lys-247–Ser 254. These modifications were correlated previously with the hindrance of dodecamer formation in absence of its metal cofactor[11]. The two mutations of His-60 and His-307, however, had a slight effect on the side chains of Asp-168 and Asp-62. They appeared to be locked in a conformation different from the wild-type dimer (**Figure 6B**). The Asp-168 carboxylate is rotated by 40° due to the absence of His-60 and His-307. Hence both histidine residues are important for positioning the Asp-168 carboxylate correctly for bridging the two metal ions. The Asp-62 side chain is oriented towards Glu-18 carboxylate, outside the catalytic site. Asp-62 may have an important role in catalysis as it is assumed to modulate the pK_a of His-60 and, thus, influence metal ion binding in the M2 site. In addition, it could be implicated structurally in the stabilization of the catalytic site upon metal ion binding, favoring the formation of the dodecamer.

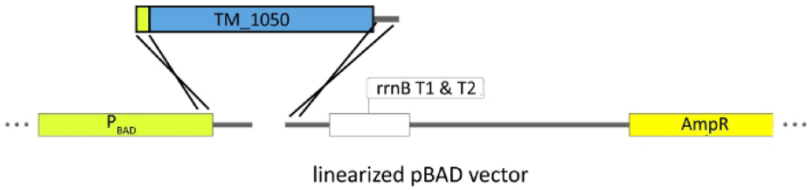


Figure 1: Schematic representation of TM_1050 ORF cloning into pBAD vector by homologous recombination.
The ORF is flanked by two 30 bp sequences homologous to the promoter BAD end and the sequence upstream *Pme*I restriction site. Please click here to view a larger version of this figure.

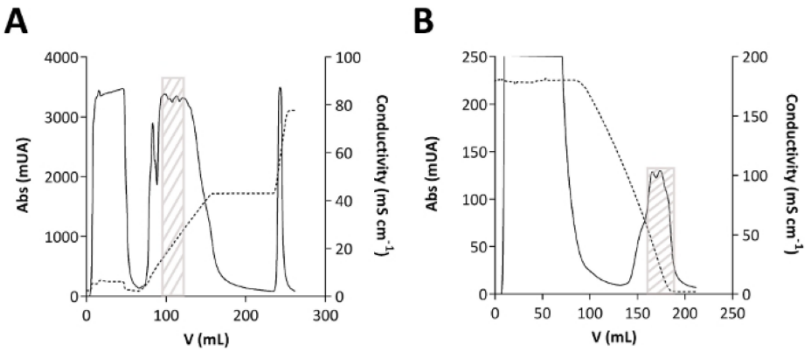


Figure 2: Chromatograms of TmPep1050 purification.
(**A**) Anion exchange chromatography. (**B**) Hydrophobic interaction chromatography. The absorbance (Abs), expressed in milliunits of absorbance (mUA), is shown in plain line. The conductivity, expressed in mS cm^{-1}, is shown in dashed line. The grey box indicates where TmPep1050 eluates on the chromatograms. Please click here to view a larger version of this figure.

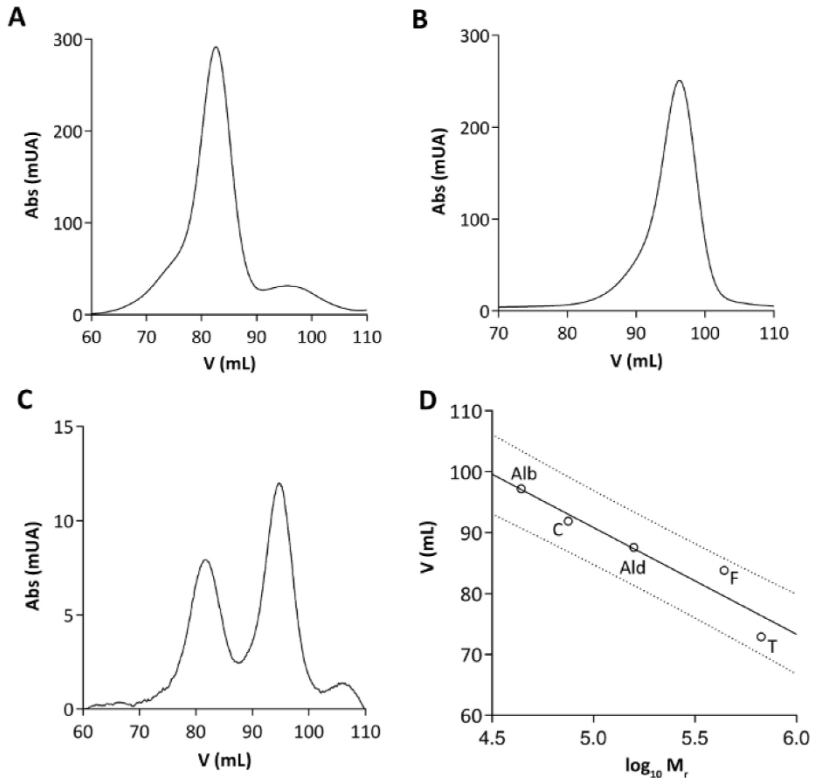


Figure 3: Size exclusion chromatography of (A) TmPep1050 dodecamer, (B) TmPep1050$_{H60A\ H307A}$, and (C) TmPep1050$_{H60A}$. Samples were analyzed using SEC resin packed in a 120 mL column. Absorbance (Abs) is expressed in milliunits of absorbance (mUA). (**D**) Calibration of the SEC column using thyroglobulin (T), ferritin (F), aldolase (Ald), conalbumin (C), and albumin (Alb) as standards. The correlation between the logarithm of the relative mass and the elution volume is linear, with a R^2 of 0.91. The 95% confidence intervals are represented as dots. Please click here to view a larger version of this figure.

jove Journal of Visualized Experiments www.jove.com

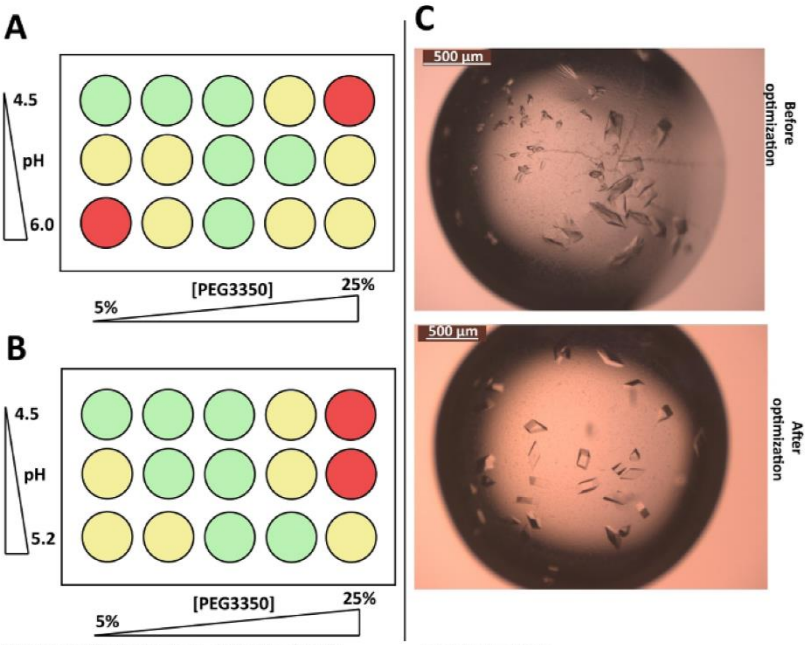


Figure 4: Optimization of TmPep1050$_{H60A\ H307A}$ crystallization.
(**A**) The first optimization strategy consists of varying pH (between 4.5 and 6.0) vs. PEG3350 concentration (between 5% and 25%). The crystallization plate is schematized, and the wells are color coded: red for precipitate, yellow for polycrystals, and green for monocrystals. (**B**) The second optimization strategy includes the use of seeds diluted 25x with a narrower variation of pH vs. PEG3350. (**C**) Crystal shape and size before (upper image) and after (lower image) crystallization optimization and microseeding. Please click here to view a larger version of this figure.

	LATTICE-CHARACTER	BRAVAIS-LATTICE	QUALITY OF FIT	UNIT CELL CONSTANTS (ANGSTROEM & DEGREES) a	b	c	alpha	beta	gamma	REINDEXING TRANSFORMATION									
*	31	aP	0.0	43.2	61.1	137.8	90.0	90.0	69.3	-1	0	0	0	0	1	0	0	0	0 -1 0
*	44	aP	0.7	43.2	61.1	137.8	90.0	90.0	110.7	1	0	0	0	-1	-1	0	0	0	0 -1 0
*	29	mC	1.0	43.2	114.4	137.8	90.0	90.0	90.0	-1	0	0	0	-1	-2	0	0	0	0 1 0
*	34	mP	1.7	43.2	137.8	61.1	90.0	110.7	90.0	-1	0	0	0	0	0	1	0	1	1 0 0
*	39	mC	1.8	114.4	43.2	137.8	90.0	90.0	90.0	1	2	0	0	-1	0	0	0	0	0 1 0
*	38	oC	2.1	43.2	114.4	137.8	90.0	90.0	90.0	-1	0	0	0	-1	-2	0	0	0	0 1 0
	35	mP	251.1	61.1	43.2	137.8	90.0	90.0	110.7	0	-1	0	0	-1	0	0	0	0	0 -1 0

SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= -3.0 AS FUNCTION OF RESOLUTION

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR COMPARED expected	I/SIGMA	R-meas	CC(1/2)	Anomal Corr	SigAno	Nano
7.20	6916	566	569	99.5%	15.4%	15.6%	6914	16.27	16.1%	98.8*	-61	0.389	382
5.12	11789	951	951	100.0%	17.2%	16.1%	11789	15.45	18.0%	98.5*	-45	0.524	754
4.19	13849	1200	1200	100.0%	18.2%	16.0%	13849	14.95	19.0%	97.9*	-30	0.639	1002
3.63	17555	1386	1386	100.0%	20.2%	17.2%	17555	14.00	21.1%	98.2*	-28	0.693	1201
3.25	21012	1570	1570	100.0%	21.7%	19.6%	21012	11.98	22.6%	98.6*	-26	0.625	1369
2.97	23980	1718	1718	100.0%	23.2%	25.5%	23980	9.24	24.1%	98.3*	-27	0.531	1518
2.75	25165	1854	1854	100.0%	28.8%	38.2%	25165	6.27	29.5%	97.0*	-19	0.403	1666
2.57	26007	1999	1999	100.0%	40.3%	55.8%	26007	4.18	41.9%	97.4*	-22	0.456	1799
2.43	25612	2052	2081	98.6%	48.9%	93.3%	25596	2.62	50.9%	92.4*	-29	0.374	1829
total	171885	13296	13328	99.8%	19.5%	19.2%	171867	9.12	20.3%	99.0*	-29	0.529	11520

NUMBER OF REFLECTIONS IN SELECTED SUBSET OF IMAGES 176795
NUMBER OF REJECTED MISFITS 4812
NUMBER OF SYSTEMATIC ABSENT REFLECTIONS 98
NUMBER OF ACCEPTED OBSERVATIONS 171885
NUMBER OF UNIQUE ACCEPTED REFLECTIONS 13296

SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= -3.0 AS FUNCTION OF RESOLUTION

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR COMPARED expected	I/SIGMA	R-meas	CC(1/2)	Anomal Corr	SigAno	Nano
7.04	7597	1058	1070	98.9%	9.4%	10.6%	7595	18.00	10.1%	99.4*	-49	0.409	981
5.01	12628	1860	1868	99.6%	11.1%	11.0%	12628	16.21	12.0%	98.4*	-32	0.584	1725
4.10	14550	2346	2364	99.2%	12.0%	10.9%	14550	15.38	13.1%	97.7*	-21	0.736	2104
3.55	18699	2808	2817	99.7%	13.9%	12.1%	18699	13.83	15.1%	97.9*	-24	0.757	2603
3.18	22384	3181	3190	99.7%	15.7%	14.5%	22383	11.38	17.0%	98.3*	-20	0.710	3044
2.91	25451	3477	3482	99.9%	19.0%	20.2%	25450	8.34	20.5%	98.0*	-19	0.623	3379
2.69	26408	3821	3831	99.7%	25.7%	32.2%	26408	5.28	27.8%	96.0*	-16	0.568	3663
2.52	27171	4086	4091	99.9%	36.7%	49.6%	27171	3.39	39.9%	92.2*	-14	0.563	3896
2.37	28434	4265	4358	97.9%	51.3%	82.8%	28359	2.12	55.6%	84.1*	-20	0.480	3967
total	183322	26902	27071	99.4%	13.6%	14.0%	183243	8.64	14.7%	99.2*	-21	0.606	25362

NUMBER OF REFLECTIONS IN SELECTED SUBSET OF IMAGES 189041
NUMBER OF REJECTED MISFITS 5617
NUMBER OF SYSTEMATIC ABSENT REFLECTIONS 102
NUMBER OF ACCEPTED OBSERVATIONS 183322
NUMBER OF UNIQUE ACCEPTED REFLECTIONS 26902

Figure 5: Excerpts from the log output CORRECT.LP of TmPep1050$_{H60A\ H307A}$ data indexation by XDS.
Upper panel, the possible Bravais lattices, the most likely being *m*C, *m*P, and *o*C. Middle panel, overall statistics of data indexed in C222$_1$ space group. Lower panel, overall statistics of data indexed in P2$_1$ space group. Please click here to view a larger version of this figure.

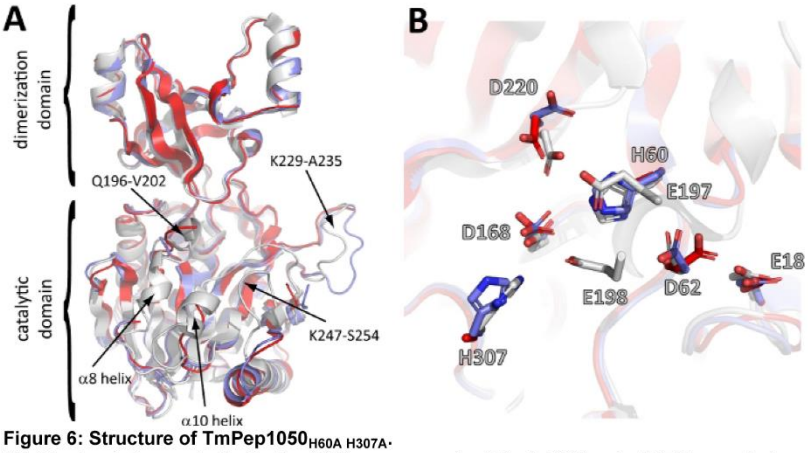


Figure 6: Structure of TmPep1050$_{H60A\ H307A}$.
(**A**) Structural alignment of a TmPep1050$_{H60A\ H307A}$ subunit (red, PDB code 5NE9) vs. a dodecamer subunit (white, PDB code 6NW5) and a dimer subunit (blue, PDB code 5NE6). Arrows indicate the structural dissimilarities between dodecamers and dimers. (**B**) Close-up of the TmPep1050$_{H60A\ H307A}$ active site (red) compared to the active site of TmPep1050 dimer (blue) and dodecamer (white). Please click here to view a larger version of this figure.

	TmPep1050$_{H60A\ H307A}$
Data collection	
Temperature (K)	100
Radiation source	SOLEIL Proxima 2
Wavelength (Å)	0.9801
Detector	Dectris Eiger X 9M
Oscillation range (°)	0.1
Exposure time (s)	0.025
Space group	P 1 21 1
Unit cell parameters	
α, β, γ (°)	90.00, 110.69, 90.00
a, b, c (Å)	43.24, 137.79, 61.11
Resolution	43.99 – 2.37 (2.52-2.37)
Unique reflections	26.902
Rmerge (%)	0.14
Redundancy	6,815
<I/σ>	8.64 (2.12)
Completeness (%)	99.6 (97.9)
$CC_{1/2}$ (%)	99.2 (84.1)
Refinement	
Resolution	43.99 – 2.37
Reflections	26.9
R_{free} set test count	1345
R_{work}/R_{free}	0.206/0.234
Protein molecules per ASU	2
V_M (Å^3/Da)	2.37
Solvent content (%)	49.0
Protein/solvent atoms	4,559/96
r.m.s.d. bond lengths (Å)	0.31
r.m.s.d. bond angles (°)	0.51
Average B-factors (Å^2)	57.0
Favored/disallowed Ramachandran φ/ψ (%)	95.02 / 0.17
Twin law	h, -k, -h-l
PDB code	5NE9

Table 1: Data collection and refinement statistics. Values in parentheses are for the highest-resolution shell.

Discussion

The protocol described herein allows understanding the dimer-dodecamer transition of TmPep1050 at the structural level. The methodology was experienced previously for determining the structure of both TmPep1050 oligomers[11]. The most challenging step was to find conditions promoting the dissociation of dodecamers into stable dimers. Such conditions had to be mild enough to permit the reassociation of dimers into dodecamers when the metal ion cofactor was added. The separation of oligomers was also a critical step as it conditions the structural studies and further biochemical characterization (e.g., studying the dodecamer reassociation in various doses of Co^{2+}). The molecular replacement, a proven method for phase determination, was used to solve the structures of TmPep1050 oligomers and its variants. The proposed protocol may be adapted to study other metallo-enzymes whose oligomerization states depend on the availability of their metal cofactors.

To illustrate the protocol, a case of study was presented, TmPep1050$_{H60A\ H307A}$ whose metal binding sites were impaired by mutating His-60 and His-307 to alanine. These residues bind Co^{2+} at the M2 and M1 sites, respectively. Interfering in metal binding could have perturbed the oligomerization state and led to a complete dissociation into monomers. Evidences of such a phenomenon have been reported for PhTET2 and PfTET3, two M42 aminopeptidases from *P. horikoshii* and *P. furiosus*, respectively[24,29]. TmPep1050$_{H60A\ H307A}$ did not behave as expected as

this variant formed dimers only. Its structure showed the same modifications as the wild-type dimer but with two small exceptions. Indeed, the side chains of Asp-168 and Asp-62 appeared to be locked in an unconventional orientation preventing the stabilization of the active site. Their orientation seemed to be imposed by His-60 and His-307 as such modifications were not observed in the single point mutation variants.

Disclosures

The authors have nothing to disclose.

Acknowledgments

We thank Martine Roovers for proofreading this paper and giving constructive comments. Access to Proxima 2 beamline (SOLEIL synchrotron) was within Block Allocation Groups 20151139.

References

1. Levy, E. D., Teichmann, S. A. Structural, Evolutionary, and Assembly Principles of Protein Oligomerization. *Progress in Molecular Biology* and *Translational Science*. **117**, 25–51 (2013).
2. Selwood, T., Jaffe, E. K. Dynamic dissociating homo-oligomers and the control of protein function. *Archives of Biochemistry* and *Biophysics*. **519** (2), 131–143 (2012).
3. Jaffe, E. K. The Remarkable Character of Porphobilinogen Synthase. *Accounts of Chemical Research*. **49** (11), 2509–2517 (2016).
4. Ramström, H. et al. Properties and Regulation of the Bifunctional Enzyme HPr Kinase/Phosphatase in *Bacillus subtilis*. *Journal of Biological Chemistry*. **278** (2), 1174–1185 (2003).
5. Rudyak, S. G., Brenowitz, M., Shrader, T. E. Mg2+-Linked Oligomerization Modulates the Catalytic Activiy of the Lon (La) Protease from *Mycobacterium smegmatis*. *Biochemistry*. **40** (31), 9317–9323 (2001).
6. Yamamoto, S., Storey, K. B. Dissociation-Association of lactate dehydrogenase Isozymes: Influences on the formation of tetramers vs. dimers of M4-LDH and H4-LDH. *International Journal of Biochemistry*. **20** (11), 1261–1265 (1988).
7. Sirover, M. A. Structural analysis of glyceraldehyde-3-phosphate dehydrogenase functional diversity. *The International Journal of Biochemistry & Cell Biology*. **57**, 20–26 (2014).
8. Gupta, V., Bamezai, R. N. K. Human pyruvate kinase M2: A multifunctional protein: Multifunctional Human PKM2. *Protein Science*. **19** (11), 2031–2044 (2010).
9. Wiegand, G., Remington, S. J. Citrate synthase: Structure, Control, and Mechanism. *Annual Review of Biophysics and Biophysical Chemistry*. **15**, 97–117 (1986).
10. Libonati, M., Gotte, G. Oligomerization of bovine ribonuclease A: structural and functional features of its multimers. *Biochemical Journal*. **380** (2), 311–327 (2004).
11. Dutoit, R. et al. How metal cofactors drive dimer–dodecamer transition of the M42 aminopeptidase TmPep1050 of *Thermotoga maritima*. *Journal of Biological Chemistry*. **294** (47), 17777–17789 (2019).
12. Rawlings, N. D. et al. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Research*. **46** (D1), D624–D632 (2018).
13. Neuwald, A. F., Liu, J. S., Lipman, D. J., Lawrence, C. E. Extracting protein alignment models from the sequence database. *Nucleic Acids Research*. **25** (9), 1665–1677 (1997).
14. Dutoit, R., Brandt, N., Legrain, C., Bauvois, C. Functional Characterization of Two M42 Aminopeptidases Erroneously Annotated as Cellulases. *PLoS ONE*. **7** (11), e50639 (2012).
15. Franzetti, B. et al. Tetrahedral aminopeptidase: a novel large protease complex from archaea. *The EMBO Journal*. **21** (9), 2132–2138 (2002).
16. Borissenko, L., Groll, M. Crystal Structure of TET Protease Reveals Complementary Protein Degradation Pathways in Prokaryotes. *Journal of Molecular Biology*. **346** (5), 1207–1219 (2005).
17. Appolaire, A. et al. TET peptidases: A family of tetrahedral complexes conserved in prokaryotes. *Biochimie*. **122**, 188–196 (2016).
18. Russo, S., Baumann, U. Crystal Structure of a Dodecameric Tetrahedral-shaped Aminopeptidase. *Journal of Biological Chemistry*. **279** (49), 51275–51281 (2004).
19. Schoehn, G. et al. An Archaeal Peptidase Assembles into Two Different Quaternary Structures: *A tetrahedron* and *a giant octahedron*. *Journal of Biological Chemistry*. **281** (47), 36327–36337 (2006).
20. Durá, M. A. et al. The structural and biochemical characterizations of a novel TET peptidase complex from Pyrococcus horikoshii reveal an integrated peptide degradation system in hyperthermophilic Archaea: Characterization of P. horikoshii TET3 peptidase. *Molecular Microbiology*. **72** (1), 26–40 (2009).
21. Basbous, H., Appolaire, A., Girard, E., Franzetti, B. Characterization of a Glycyl-Specific TET Aminopeptidase Complex from *Pyrococcus horikoshii*. *Journal of Bacteriology*. **200** (17), e00059-18 (2018).
22. Appolaire, A. et al. Small-angle neutron scattering reveals the assembly mode and oligomeric architecture of TET, a large, dodecameric aminopeptidase. *Acta Crystallographica Section D Biological Crystallography*. **70** (11), 2983–2993 (2014).
23. Appolaire, A. et al. The TET2 and TET3 aminopeptidases from *P yrococcus horikoshii* form a hetero-subunit peptidasome with enhanced peptide destruction properties: TET aminopeptidase multi-subunit complex. *Molecular Microbiology*. **94** (4), 803–814 (2014).
24. Colombo, M., Girard, E., Franzetti, B. Tuned by metals: the TET peptidase activity is controlled by 3 metal binding sites. *Scientific Reports*. **6** (1), 20876 (2016).
25. Petrova, T. E. et al. Structure of the dodecamer of the aminopeptidase APDkam598 from the archaeon *Desulfurococcus kamchatkensis*. *Acta Crystallographica Section F Structural Biology Communications*. **71** (3), 277–285 (2015).
26. Kim, D. et al. Structural basis for the substrate specificity of PepA from Streptococcus pneumoniae, a dodecameric tetrahedral protease. *Biochemical* and *Biophysical Research Communications*. **391** (1), 431–436 (2010).
27. Chevrier, B. et al. Crystal structure of Aeromonas proteolytica aminopeptidase: a prototypical member of the co-catalytic zinc enzyme family. *Structure*. **2**, 283–291 (1994).

28. Rosenbaum, E., Ferruit, M., Durá, M. A., Franzetti, B. Studies on the parameters controlling the stability of the TET peptidase superstructure from Pyrococcus horikoshii revealed a crucial role of pH and catalytic metals in the oligomerization process. *Biochimica et Biophysica Acta (BBA) - Proteins* and *Proteomics.* **1814** (10), 1289–1294 (2011).
29. Macek, P. et al. Unraveling self-assembly pathways of the 468-kDa proteolytic machine TET2. *Science Advances.* **3** (4), e1601601 (2017).
30. Edelheit, O., Hanukoglu, A., Hanukoglu, I. Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies. *BMC Biotechnology.* **9** (1), 61 (2009).
31. Zhang, Y., Werling, U., Edelmann, W. SLiCE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Research.* **40** (8), e55–e55 (2012).
32. Schleif, R. AraC protein, regulation of the l-arabinose operon in *Escherichia coli*, and the light switch mechanism of AraC action. *FEMS Microbiology Reviews.* **34** (5), 779–796 (2010).
33. McPherson, A., Gavira, J. A. Introduction to protein crystallization. *Acta Crystallographica Section F Structural Biology Communications.* **70** (1), 2–20 (2014).
34. Bergfors, T. Seeds to crystals. *Journal of Structural Biology.* **142** (1), 66–76 (2003).
35. Dauter, Z. Collection of X-Ray Diffraction Data from Macromolecular Crystals. *Protein Crystallography.* **1607**, 165–184 (2017).
36. Powell, H. R. X-ray data processing. *Bioscience Reports.* **37** (5), BSR20170227 (2017).
37. Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R., Leslie, A. G. W. *iMOSFLM* : a new graphical interface for diffraction-image processing with *MOSFLM. Acta Crystallographica Section D Biological Crystallography.* **67** (4), 271–281 (2011).
38. Otwinowski, Z., Minor, W. [20] Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology.* **276**, 307–326 (1997).
39. Clabbers, M. T. B., Gruene, T., Parkhurst, J. M., Abrahams, J. P., Waterman, D. G. Electron diffraction data processing with *DIALS. Acta Crystallographica Section D Structural Biology.* **74** (6), 506–518 (2018).
40. Kabsch, W. XDS. *Acta Crystallographica Section D Biological Crystallography.* **66** (2), 125–132 (2010).
41. Legrand, P. *legrandp/xdsme: March 2019 version working with the latest XDS version (Jan 26, 2018).* Zenodo. (2019).
42. Evans, P. Scaling and assessment of data quality. *Acta Crystallographica Section D Biological Crystallography.* **62** (1), 72–82 (2006).
43. Wlodawer, A., Minor, W., Dauter, Z., Jaskolski, M. Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures: Protein crystallography for non-crystallographers. *FEBS Journal.* **275** (1), 1–21 (2008).
44. Karplus, P. A., Diederichs, K. Assessing and maximizing data quality in macromolecular crystallography. *Current Opinion in Structural Biology.* **34**, 60–68 (2015).
45. Taylor, G. L. Introduction to phasing. *Acta Crystallographica Section D Biological Crystallography.* **66** (4), 325–338 (2010).
46. Rossmann, M. G., Blow, D. M. The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallographica.* **15**, 24–31 (1962).
47. Rossmann, M. G. The molecular replacement method. *Acta Crystallographica Section A Foundations of Crystallography.* **46** (2), 73–82 (1990).
48. McCoy, A. J. et al. *Phaser* crystallographic software. *Journal of Applied Crystallography.* **40** (4), 658–674 (2007).
49. Adams, P. D. et al. *PHENIX*: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography.* **66** (2), 213–221 (2010).
50. Emsley, P., Lohkamp, B., Scott, W. G., Cowtan, K. Features and development of *Coot. Acta Crystallographica Section D Biological Crystallography.* **66** (4), 486–501 (2010).
51. Chen, V. B. et al. *MolProbity*: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D Biological Crystallography.* **66** (1), 12–21 (2010).
52. Zwart, P. H., Grosse-Kunstleve, R. W., Lebedev, A. A., Murshudov, G. N., Adams, P. D. Surprises and pitfalls arising from (pseudo)symmetry. *Acta Crystallographica Section D Biological Crystallography.* **64** (1), 99–107 (2008).
53. Yeates, T. O. Detecting and overcoming crystal twinning. *Methods in Enzymology.* **276**, 344–358 (1997).
54. Terwilliger, T. C. Using prime-and-switch phasing to reduce model bias in molecular replacement. *Acta Crystallographica Section D Biological Crystallography.* **60** (12), 2144–2149 (2004).
55. Terwilliger, T. C. et al. Iterative model building, structure refinement and density modification with the *PHENIX AutoBuild* wizard. *Acta Crystallographica Section D Biological Crystallography.* **64** (1), 61–69 (2008).
56. Krissinel, E., Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology.* **372** (3), 774–797 (2007).

B.3 M42 aminopeptidase catalytic site: the structural and functional role of a strictly conserved aspartate residue

B.3.1 Background

As discussed in the Introduction (see **section A.5.1.1**, page 23), the members of the MH clan share seven highly conserved residues located within their active site. Five of them are involved in the metal ion binding, forming the M1 and M2 sites. The metal ion centers play a critical role for the activity but also for the oligomerization of M42 aminopeptidases[184,365,372,382,420,424]. We showed that two histidine residues are especially crucial for the binding of Co^{2+} and, subsequently, the oligomerization of TmPep1050 (see **sections B.1-B.2**). The two other conserved residues are a glutamate residue and an aspartate residue. According to literature on VpAp1, the former acts as the catalytic base while the latter probably modulates the Lewis acid strength of the metal ion bound in the M2 site[302,304,307]. The catalytic role of the conserved glutamate residue has been proven for M42 aminopeptidases. Indeed, its substitution with either an alanine or glutamine residue yields to an inactive enzyme for PhTET1[366]. The same impact was observed for TmPep1050 as the Glu197Gln substitution drastically impaired the activity (see **sections B.1.4-B.1.5**). The catalytic base, however, does not play a role in the oligomerization of TmPep1050 despite being part of a H-bond network involving the seven conserved residues. The role of the conserved aspartate residue remains ill-described. According to the hypothesis drawn from the structure of VpAp1, this residue interacts with the histidine residue of the M2 site, forcing its imidazole ring to be deprotonated[304,307]. As a consequence, the Lewis acid strength of the M2 metal ion is decreased, favoring the hydroxide ion transfer to the M1 metal ion (see also **section A.4.2.3**, page 26). Nothing is known about the implication of the conserved aspartate residue in metal ion binding or its structural role in the catalytic site. Studying these roles in M42 aminopeptidases was particularly relevant since their oligomerization is dependent on its metal ion cofactors[365,382,420]. In addition, S. Russo reported in her PhD thesis that Asp-70 (corresponding to Asp-99 in VpAp1 and Asp-62 in TmPep1050) plays a critical role in the activity of PhTET2[338].

TmPep1050 could be a relevant model for studying the role of the conserved aspartate residue close to the M2 site. Indeed, after the residue substitution, the degree of oligomerization could tell whether it is involved in metal ion binding or not. According to the oligomers being observed, its role as a second shell residue could be linked to the M1 or M2 site. We showed that the substitution of Asp-62 completely abolished the activity of TmPep1050. Moreover, the dodecamer formation appeared to be dramatically impaired as only dimers were observed. The structural study of a variant highlighted the structural role of Asp-62 in maintaining an important loop correctly positioned in the active site. Albeit not interacting directly with His-307, Asp-62 could influence metal ion binding in the M1 site. The work has been published in Proteins: Structure, Function, and Bioinformatics in July 2020.

B.3.2 Methodology

Basically, we applied the methodology described earlier (see **sections B.1.2-B2**) to study the role of Asp-62 in TmPep1050. Asp-62 was substituted with either an alanine or asparagine or glutamate residue by site-directed mutagenesis. The variants were purified using the same protocol as for

purifying TmPep1050$_{12\text{-mer}}$. Their oligomeric state was determined by size-exclusion chromatography and native MS. Their activity was assayed using L-Leu-*p*NA as substrate in presence of Co^{2+}. Metal ion binding was examined by measuring for the three variants (i) the influence of Co^{2+} on thermostability and (ii) the apparent association constant for Co^{2+}. Ultimately, the structure of TmPep1050$_{D62A}$ was solved at 1.5 Å by X-ray crystallography.

B.3.3 Article summary and discussion

The activity of the three variants was completely abolished regardless the substitution nature. In addition, the Asp-62 substitution led to an impaired dodecamer formation since only dimers were observed. The activity loss was expected in accordance with the supposed role of Asp-62 in catalysis. Its impact on oligomerization, however, was unexpected since it interacts with His-60 (in the M2 site) but not with His-307 (in the M1 site). As a reminder of our previous study (see **section B.1**), the substitution of His-307 with an alanine residue completely prevented the dodecamer formation while that of His-60 partly interfered in oligomerization. Thus, the Asp-62 variants behaved as if the M1 site was crippled. They were further characterized to determine their property to bind Co^{2+}. The thermostability of the Asp-62 variants was monitored by TSA and compared to that of the wild-type dimer. The Asp62Ala and Asp62Glu substitutions had a destabilizing effect as these variants are less thermostable than the wild-type dimer. The thermostability of Asp-62 variants increased with the concentration of Co^{2+}, indicating the binding of metal ions. Of note, their T_m remained lower than T_m of TmPep1050$_{12\text{-mer}}$ even at Co^{2+}-to-protein molar ratio of 50. Especially, TmPep1050$_{D62A}$ and TmPep1050$_{D62E}$ had a T_m of about 91°C compared to a T_m of 97°C observed for the dodecamer. Co^{2+} binding assays were conducted to determine the amount of bound Co^{2+} per molecule of enzyme. Each variant was shown to bind only one cobalt atom per molecule of enzyme with an apparent dissociation constant ranging from 70 to 87 µM. In comparison, wild-type TmPep1050 binds at least two cobalt atoms with an apparent dissociation constant of 50 µM. Our data suggested that the Asp-62 substitution strongly impaired one of the metal ion binding sites. According to the molecular weight of Asp-62 variants, it could be the M1 site being affected.

The structure of TmPep1050$_{D62A}$ was solved at 1.5 Å by X-ray crystallography, confirming its oligomeric state. As expected, it displays the same structural adaptation as those of TmPep1050$_{2\text{-mer}}$. Especially, the Gln-196 – Val-202 loop is displaced outwards the active site in TmPep1050$_{D62A}$, like in the wild-type dimer. The α8 and α10 helices are also so disordered that they cannot be modelled. Globally, the structure of TmPep1050$_{D62A}$ does not differ from the wild-type dimer except for the position of His-60 sidechain. Indeed, it adopts two alternate rotamers, m90° and p-80°, the former being observed in the wild-type enzyme structure. Thus, Asp-62 seems to be important in positioning His-60 side chain correctly. As the dodecamer formation is abolished, Asp-62 must have a structural role on the M1 site fold. In the dodecamer structure, the carboxylate of Asp-62 interacts with the backbone nitrogen of Glu-197 and Glu-198. Its absence probably prevents the correct positioning of Gln-196 – Val-202 loop. Consequently, the M1 site is affected as Glu-198 cannot participate in metal ion binding. To fulfil such a structural role, the length and charge are crucial at the position 62. Substituting Asp-62 with an asparagine or glutamate residue also impaired the dodecamer formation. In fact, three residues impose structural constraint on Asp-62 sidechain, Ser-15, Glu-18, and His-60. A longer sidechain at position 62 will fail to be sandwiched between Ser-15 and His-60. The charge is probably important for the correct positioning due to charge repulsion with Glu-18. It would be interesting to study the structural role of the conserved aspartate residue in other MH clan enzymes,

especially monomeric enzymes. If the destabilization is also observed for such enzymes, it will support its structural role thorough the whole MH clan. If not, it will indicate that the destabilization of the active site, as observed for TmPep1050$_{D62A}$, could be an adaptation to oligomerization.

B.3.4 The article with supplementary data

Received: 31 March 2020 | Revised: 6 June 2020 | Accepted: 2 July 2020
DOI: 10.1002/prot.25982

RESEARCH ARTICLE

PROTEINS WILEY

M42 aminopeptidase catalytic site: the structural and functional role of a strictly conserved aspartate residue

Raphaël Dutoit[1,2] | Nathalie Brandt[2] | Tom Van Gompel[3] | Dany Van Elder[1] | Jeroen Van Dyck[3] | Frank Sobott[3,4] | Louis Droogmans[1]

[1]Laboratory of Microbiology, Department of Molecular Biology, Université Libre de Bruxelles, Brussels, Belgium

[2]Labiris Institut de Recherche, Brussels, Belgium

[3]Biomolecular & Analytical Mass Spectrometry, Department of Chemistry, Universiteit van Antwerpen, Antwerpen, Belgium

[4]Astbury Centre for Structural and Molecular Biology, University of Leeds, Leeds, UK

Correspondence
Raphaël Dutoit, Laboratory of Microbiology, Department of Molecular Biology, Université Libre de Bruxelles, 1 avenue Emile Gryzon, IU BE1070 Brussels, Belgium.
Email: rdutoit@ulb.ac.be

Funding information
Fonds National de la Recherche Scientifique, Grant/Award Number: IISN 4.4503.11F

Peer Review
The peer review history for this article is available at https://publons.com/publon/10.1002/prot.25982.

Abstract

The M42 aminopeptidases are a family of dinuclear aminopeptidases widely distributed in Prokaryotes. They are potentially associated to the proteasome, achieving complete peptide destruction. Their most peculiar characteristic is their quaternary structure, a tetrahedron-shaped particle made of twelve subunits. The catalytic site of M42 aminopeptidases is defined by seven conserved residues. Five of them are involved in metal ion binding which is important to maintain both the activity and the oligomeric state. The sixth conserved residue, a glutamate, is the catalytic base deprotonating the water molecule during peptide bond hydrolysis. The seventh residue is an aspartate whose function remains poorly understood. This aspartate residue, however, must have a critical role as it is strictly conserved in all MH clan enzymes. It forms some kind of catalytic triad with the histidine residue and the metal ion of the M2 binding site. We assess its role in TmPep1050, an M42 aminopeptidase of *Thermotoga maritima*, through a mutational approach. Asp-62 was substituted with alanine, asparagine, or glutamate residue. The Asp-62 substitutions completely abolished TmPep1050 activity and impeded dodecamer formation. They also interfered with metal ion binding as only one cobalt ion is bound per subunit instead of two. The structure of Asp62Ala variant was solved at 1.5 Å showing how the substitution has an impact on the active site fold. We propose a structural role for Asp-62, helping to stabilize a crucial loop in the active site and to position correctly the catalytic base and a metal ion ligand of the M1 site.

KEYWORDS

catalytic site fold, cobalt ion, M42 aminopeptidase, metal ion binding, MH clan, oligomerization, strictly conserved aspartate, *Thermotoga maritima*

1 | INTRODUCTION

The M42 aminopeptidases are dinuclear enzymes proposed to achieve peptide degradation downstream the proteasome in Bacteria and Archaea.[1,2] In *Pyrococcus horikoshii*, four M42 aminopeptidases (PhTET1, PhTET2, PhTET3, and PhTET4) have been described, each having a different, but complementary, substrate specificity.[3-6] PhTET2 and PhTET3 have even been shown to form heterocomplexes, suggesting that peptidasome particles may exist.[7,8] The M42 aminopeptidases are characterized by a genuine quaternary structure made of twelve subunits organized spatially as a tetrahedron.[1,3-6,9-13] The association of the twelve subunits is often described as the assembly of six dimers, each dimer being located at a tetrahedron edge.[1] Four gates are found at the middle of the tetrahedron faces, leading to a wide inner chamber. The gate size (between 12 and 20 Å) probably restricts the access to the inner chamber to unfolded peptides only.[1,10,13,14]

The 12 catalytic sites are oriented inward the chamber, compartmentalizing the activity. The amino acids, generated after peptide hydrolysis, exit the catalytic sites through four channels located at the tetrahedron vertexes. The oligomerization of M42 aminopeptidases is controlled by their metal ion cofactors, as shown for PhTET2, PhTET3, PfTET3 of *Pyrococcus furiosus*, and TmPep1050 of *Thermotoga maritima*.[1,12,13,15,16] Indeed, active dodecamers disassemble into inactive dimers upon the depletion of metal ions. The disassembly is although reversible as adding metal ions to dimers restores the dodecamer formation and activity.

The M42 family, ubiquitous to Bacteria and Archaea,[17] belongs to the MH clan alongside the M18, M20, and M28 families.[18] All MH clan enzymes share a common α/β globular fold for their catalytic domain, the archetypal model being the aminopeptidase Ap1 of *Vibrio proteolyticus*.[19] The typical subunit of M42 aminopeptidases possesses a PDZ-like dimerization domain in addition to the catalytic domain.[1] Seven residues define the catalytic site of MH clan enzymes.[18,20] Five of them are involved in the binding of two metal ion cofactors. The first metal ion binding site, M1, consists of three conserved residues: a histidine, a glutamate, and an aspartate. The latter is shared with the second metal ion binding site, M2. Two other residues define the M2 site: a conserved histidine and a glutamate/aspartate. These five residues impose the tetrahedral coordination geometry to the two bound metal ions. The implication of both metal ions in the catalytic mechanism has been thoroughly studied in *V. proteolyticus* aminopeptidase Ap1.[21-23] Indeed, they interact with the free N-terminal amine and the first carbonyl of the peptide substrate. The metal ion of the M2 site has been proposed to assist the water molecule deprotonation. The hydroxide is then transferred to the metal ion of the M1 site for peptide bond hydrolysis. In M42 aminopeptidases, the two metal ions are also involved in the dimer–dodecamer transition. We previously showed that the metal ion of the M1 site strictly controls the oligomerization of TmPep1050 while that of the M2 site could have a stabilizer role.[13] Intriguingly, their roles are swapped in PfTET3,[12] suggesting that other factors may influence oligomerization.

In addition to the five metal ion ligands, two other residues, a glutamate and an aspartate, define MH clan enzymes. These two residues are not involved directly in metal ion binding but are strictly conserved in the whole clan. The glutamate residue acts as a general base deprotonating the water molecule during the peptide bond hydrolysis.[24] It may also play a role in decreasing the Lewis acidity of the metal ion of the M2 site. As proposed for *V. proteolyticus* aminopeptidase Ap1, the catalytic base Glu-151 is within hydrogen bond distance to interact with His-97.[21,25] The aspartate residue is not directly involved in hydrolysis but may have a role in decreasing the Lewis acidity of the metal ion of the M2 site.[26] In *V. proteolyticus* aminopeptidase Ap1, Asp-99 has been proposed to form some kind of catalytic triad (hereafter mentioned as Asp-His-metal) with His-97 and Zn^{2+} in the M2 binding site.[21,23,25,27] Indeed, a strong hydrogen bond links the $O^{\delta 1}$ atom of Asp-99 with the N^{δ} atom of His-97, forcing the imidazole ring to be deprotonated. As a result, the Lewis acidity of Zn^{2+} would be decreased sufficiently to allow the transfer of the hydroxide to the metal ion of the M1 site, which is closer to the peptide bond to be hydrolyzed.

Still, how the aspartate residue participates in peptide bond hydrolysis remains intricate. Its role has only been hypothesized for *V. proteolyticus* aminopeptidase Ap1. To our knowledge, nothing has been well established about its implication in metal ion binding or active site stabilization for MH clan enzymes. In this work, we propose to study the Asp-His-metal triad of TmPep1050, an M42 aminopeptidase from *T. maritima*. TmPep1050 is a leucyl-aminopeptidase whose activity depends on its oligomeric states: inactive dimer and active dodecamer.[13] The transition between dimers and dodecamers is regulated by its metal ion cofactor, Co^{2+}. We present the characterization of three TmPep1050 variants for which Asp-62, equivalent to Asp-99 of *V. proteolyticus* aminopeptidase Ap1, was substituted with alanine, asparagine, or glutamate residue. All substitutions completely abolished the activity and dodecamer formation of TmPep1050. The Asp-62 variants were still able to bind one cobalt ion per subunit. As dodecamer formation relies on the M1 site, we propose that the M2 site remained functional despite the substitution. The structure of $TmPep1050_{D62A}$ was solved at 1.5 Å resolution and, by comparing with the dimer and dodecamer structures, a structural role was inferred for Asp-62.

2 | MATERIALS AND METHODS

2.1 | Mutagenesis of *Tm_1050*

Site-directed mutagenesis was performed following the single-primer reactions in parallel protocol (SPRINP)[28] using the pCEC43 as template vector. This vector allows the production of TmPep1050 under the regulation of P_{BAD} promoter in *Escherichia coli*.[13] The primers used to generate the Asp-62 variants of TmPep1050 are described in Table S1. Phusion DNA polymerase (ThermoFisher Scientific) was used for DNA polymerization reactions. Cloning was carried out in the *E. coli* strain MC1061.[29] The genetic constructs were checked by sequencing (Genetic Service Facility, University of Antwerp).

2.2 | Production and purification

The Asp-62 variants of TmPep1050 were produced and purified to homogeneity according to the protocol described elsewhere.[13,30] It consists of four steps: (a) a heat treatment of crude cell extract at 70°C for 15 minutes, (b) an anion-exchange chromatography using Source 15Q resin (GE Healthcare Life Sciences), (c) an hydrophobic interaction chromatography using Source 15Phe resin (GE Healthcare Life Sciences), and (d) a size-exclusion chromatography using a Superdex200 column (GE Healthcare Life Sciences, XK16/20 column of 120-mL volume). During the last purification step, an elution peak was observed at 95 mL for the three Asp-62 variants. The purified Asp-62 variants were concentrated to 250 μM using an Amicon Ultra-15 ultrafiltration unit with 30-kDa cutoff (Merck Millipore). The protein concentration was estimated by measuring the absorbance at 280 nm and applying the mass extinction coefficient of 18 910 M^{-1} cm^{-1}.

For molecular weight determination, 50 μM of purified protein was loaded on a Superdex200 column (GE Healthcare Life Sciences, XK16/20 column of 120-mL volume) using 50 mM MOPS, 0.5 M $(NH_4)_2SO_4$, 1 mM $CoCl_2$, and pH 7.2 as running buffer. The size exclusion column was calibrated using Gel-filtration standard (Biorad) and high-molecular weight gel filtration calibration kit (GE Healthcare Life Sciences) under the same running conditions.

2.3 | Mass spectrometry

Sample preparation and native mass spectrometry (MS) analysis were performed as previously described.[13] $TmPep1050_{D62A}$ was conditioned in 20 mM Ammonium acetate, pH 7.2 using Zeba 7-kDa desalting columns (Thermo Fisher Scientific). Prior to MS analysis, the sample was diluted to 5 μM in 20 mM ammonium acetate, pH 7.2. The spectrum was recorded in positive ion mode on a traveling-wave ion mobility Q-TOF instrument (Synapt G2 HDMS, Waters).

2.4 | Enzymatic and cobalt binding assays

The specific activity of Asp-62 variants was assayed using L-leucine-*p*-nitroanilide (L-Leu-*p*NA, Bachem AG) as substrate. About 10 μL of 200 μM enzyme was added to 990 μL of reaction mix containing 2.5 mM L-Leu-*p*NA in 50 mM MOPS, 10% methanol, 1 mM $CoCl_2$, and pH 7.2. The reaction was performed at 75°C during 1 hour and stopped by adding 1 mL of 20% acetic acid. Released *p*-nitroaniline was quantified by measuring the absorbance at 410 nm. Prior to cobalt binding assays, apo-enzyme was prepared as previously described.[30] About 200 μL of 250 μM enzyme was diluted in 1.8 mL of 50 mM MOPS, 0.5 M $(NH_4)_2SO_4$, 10 mM 1,10-phenanthroline, pH 7.2. The sample was then dialyzed four times against 200 mL of 50 mM MOPS, 0.5 M $(NH_4)_2SO_4$, and pH 7.2. After the dialysis, the sample was concentrated to 100 μL using an Amicon Ultra-15 ultrafiltration unit with 30-kDa cutoff (Merck Millipore). Protein concentration was determined as described above. To study the metal binding, apo-enzyme was diluted to 20 μM in either 50 mM MOPS, 0.5 M $(NH_4)_2SO_4$, pH 7.2 or 100 mM sodium citrate, 5% PEG3350, pH 5.2. $CoCl_2$ was added at a concentration ranging from 0 to 1280 μM. The reaction volume was 140 μL and the samples were incubated at 50°C for 24 hours. The thermostability of the samples was determined by Thermal Shift Assay using SyproOrange (Thermofisher Scientific) as previously described.[13] Bound Co^{2+} was quantified using Amplex UltraRed fluorescent probe (ThermoFisher Scientific) as previously described.[13]

2.5 | X-ray crystallography

$TmPep1050_{D62A}$ was crystallized in 0.1 M citric acid, 5% PEG3350, pH 5.2 using the hanging drop diffusion method. Drops containing 2 μL of 250 μM $TmPep1050_{D62A}$ and 2 μL of crystallization reagent were set up in EasyXtal Tool plates (Qiagen). Fully grown crystals were obtained after a week at 292 K. Crystal shape and size were improved by microseeding. Diffraction data were collected on the FIP-BM30a beamline at the European Synchrotron Research Facility (Grenoble, France).[31,32] Data collection and refinement statistics are presented in Table 1. XDS program package[33] was used for indexing diffraction data. The phase was determined by molecular replacement using Phaser[34] with the monomer A coordinates of TmPep1050 dodecamer structure (PDB code 4P6Y). The model was build using *phenix.autobuild* in PHENIX software package.[35] The model was iteratively refined manually with Coot[36] and automatically with *phenix.refine*. Model stereochemical quality was assessed using MolProbity.[37]

TABLE 1 Data collection and refinement statistics (values in parentheses are for the highest-resolution shell)

	$TmPep1050_{D62A}$
Data collection	
Temperature (K)	100
Radiation source	ESRF BM30A
Wavelength	0.9797
Detector	ADSC QUANTUM 315r
Oscillation range (°)	0.5
Space group	C 2 2 2_1
Unit cell parameters	
α, β, γ (°)	90.00, 90.00, 90.00
a, b, c (Å)	42.18, 113.96, 267.23
Resolution	48.00 – 1.50 (1.58-1.50)
Unique reflections	103,939
R_{merge} (%)	9.1 (49.3)
Redundancy	5.8
$<I/\sigma>$	14.80 (3.34)
Completeness (%)	99.7 (89.4)
$CC_{1/2}$ (%)	99.8 (87.6)
Refinement	
Resolution	48.00 – 1.50
Reflections	103,177
R_{free} set test count	5,158
R_{work}/R_{free}	0.183/0.215
Protein molecules per ASU	2
V_M (Å^3/Da)	2.27
Solvent content (%)	45.79
Protein/solvent atoms	4,950/820
r.m.s.d. bond lengths (Å)	0.36
r.m.s.d. bond angles (°)	0.55
Average B-factors (Å^2)	20.0
Favored/disallowed Ramachandran φ/ψ (%)	97.71/0.00
PDB code	5L6Z

Protein structures were analyzed with PDBe Pisa,[38] Arpeggio,[39] and PyMOL Molecular Graphics System version 2.2 (Schrödinger, LLC).

3 | RESULTS AND DISCUSSION

3.1 | The role of Asp-62 in the activity and oligomerization of TmPep1050

Asp-62, equivalent to Asp-99 of *V. proteolyticus* aminopeptidase Ap1, was substituted with alanine, asparagine, or glutamate residue. The resulting variants were named TmPep1050$_{D62A}$, TmPep1050$_{D62N}$, and TmPep1050$_{D62E}$, respectively. The variants were recombinantly produced in *E. coli* and purified to homogeneity. Their molecular weights were determined by size-exclusion chromatography (Figure S1). The substitution of Asp-62 had a dramatic impact on the TmPep1050 oligomerization state as only dimers were observed even in the presence of Co^{2+}. The oligomerization state of TmPep1050$_{D62A}$ was confirmed by native MS showing dimers mainly (Figure 1). MS experiments also highlighted a minor tetrameric form. Minor intermediate oligomers (tetramer, hexamer, and octamer) were reported previously for the wild-type enzyme.[13] These oligomers were inferred as intermediate forms occurring during the dimer-dodecamer transition. For TmPep1050$_{D62A}$, only tetramers were observed as intermediates, supporting the impaired oligomerization of the Asp-62 substitution. The three variants were barely active when assayed on L-Leu-*p*NA in comparison to the dodecameric wild-type enzyme (Table 2).

The loss of activity resulting from Asp-62 substitution was expected according to the putative function of Asp-99 in *V. proteolyticus* Ap1. The negative charge and the side chain length are critical for the activity. Intriguingly, the Asp-62 substitution prevents the self-assembly of dimers into dodecamers. Such an observation contrasts with the known role of the two metal ion binding sites in TmPep1050 oligomerization. Indeed, we previously demonstrated that the M1 binding site strictly controls the oligomerization as the His307Ala substitution prevented the self-assembly of dimers.[13] The substitution of His-60, a residue of the M2 binding site, perturbed the oligomerization to a lesser extent as both dimers and dodecamers were observed. Since Asp-62 interacts with His-60, its substitution would have had a rather small impact on oligomerization, like the His-60 substitution. Unexpectedly, the Asp-62 variants behaved as if the M1 binding site had been impaired as only dimers were observed. Therefore, the Asp-62 substitution could somehow interfere with the metal ion binding in both M1 and M2 binding sites.

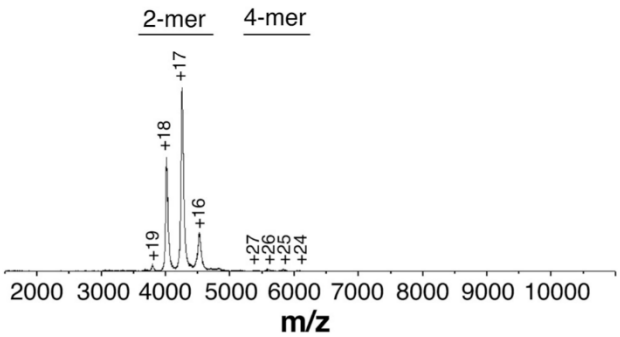


FIGURE 1 Native mass spectrum of TmPep1050$_{D62A}$. Dimers are the major observed species. The sample was conditioned in 20 mM ammonium acetate and diluted to 5 μM prior to native MS analysis

3.2 | Asp-62 substitution partly impaired the Co^{2+} binding

To show the potential impact of the Asp-62 substitution on Co^{2+} binding, thermal stability of Asp-62 variants was measured by thermal shift assay in presence of various concentrations of Co^{2+}. Using such a method, we previously showed that TmPep1050 dimers depleted of cobalt were more prone to thermal denaturation than TmPep1050 dodecamers (hereafter named TmPep1050$_{12\text{-mer}}$), with melting temperatures of 91°C and 97°C, respectively. The addition of Co^{2+} restored the thermal stability of TmPep1050 dimers.[13] Prior to the binding assays, the Asp-62 variants were treated with 1,10-phenanthroline to remove any bound Co^{2+}. Then 20 μM of each apo-enzyme was incubated in presence of Co^{2+} at a concentration ranging from 0 to 1280 μM for 24 hours. The melting temperatures were subsequently determined to monitor an eventual cobalt binding. Apo-TmPep1050$_{D62N}$ had a melting temperature of 89.1°C ± 0.4°C, close to that of the wild-type dimer. Apo-TmPep1050$_{D62A}$ and apo-TmPep1050$_{D62E}$ had a melting temperature of 86°C ± 0.1°C and 83.2°C ± 0.2°C, respectively, indicating that the Asp62Ala and Asp62Glu substitutions destabilized partly the structure. As the thermal stability increased with the cobalt concentration (Figure 2), the three variants were found to bind cobalt ions. Of note, none of them exhibited a thermostability comparable to the dodecamer upon Co^{2+} binding. Indeed, the maximal melting temperature of TmPep1050$_{D62N}$, TmPep1050$_{D62E}$, and TmPep1050$_{D62A}$ was 95.1°C, 91.1°C, and 90.7°C, respectively, for a Co^{2+}-to-protein molar ratio of 50.

Bound Co^{2+} was quantified to define how many metal ion binding sites are occupied in each variant. 20 μM of apo-enzyme was incubated with varying Co^{2+} concentration ranging from 0 to 1280 μM. The concentration of bound Co^{2+} was determined at the equilibrium

TABLE 2 Aminopeptidase activity of Asp-62 variants against L-Leu-*p*NA

Enzyme	*k*
TmPep105012-mer	7098 ± 216
TmPep1050D62A	0.084 ± 0.003
TmPep1050D62N	0.087 ± 0.005
TmPep1050D62E	0.036 ± 0.002

Notes: Specific activity (*k*) is expressed as μmol of released *p*-nitroaniline per minute per μmol of enzyme. The specific activities of the variants are compared with wild-type TmPep1050 dodecamer (TmPep1050$_{12\text{-mer}}$). SE is given for each value with n = 3.

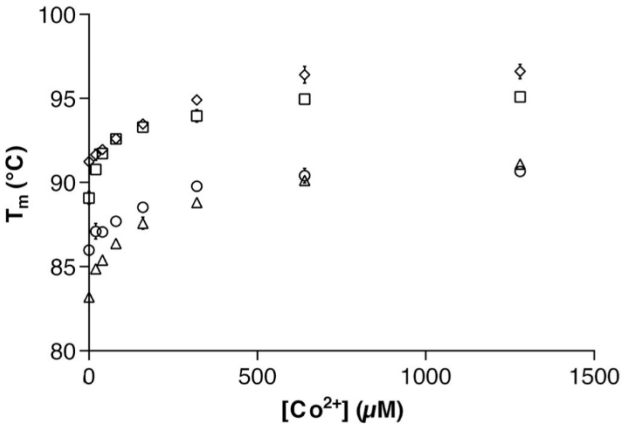


FIGURE 2 Thermal shift assay showing Co^{2+} binding effect on Asp-62 variant thermostability. The thermostability of $TmPep1050_{D62A}$ (circle), $TmPep1050_{D62N}$ (square), and $TmPep1050_{D62E}$ (triangle) was determined by measuring T_m at each Co^{2+} concentration. From our previous study, the thermostability of wild-type TmPep1050 dimer is presented as rhombi.[13] Error bars represent SE with n = 3

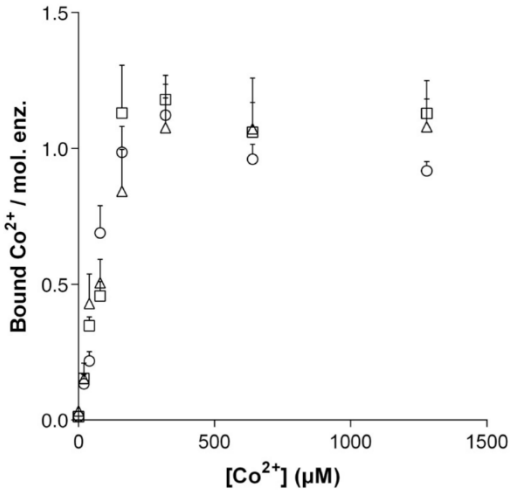


FIGURE 3 Co^{2+} binding of apo-$TmPep1050_{D62A}$ (circle), apo-$TmPep1050_{D62N}$ (square), and apo-$TmPep1050_{D62E}$ (triangle). The results are expressed as the number of bound Co^{2+} per molecule of enzyme (mol. enz.) in response to a varying concentration of Co^{2+} ranging from 0 to 1280 μM. Error bars represent SE with n = 3

using Amplex Ultra Red fluorescent probe for quantification. The results are shown in Figure 3. $TmPep1050_{D62A}$, $TmPep1050_{D62N}$, and $TmPep1050_{D62E}$ were able to bind Co^{2+} with an apparent association constant of 70 ± 12, 86 ± 2, and 87 ± 13 μM, respectively. Only one cobalt ion was bound per monomer (maximal bound Co^{2+} being 1.3), suggesting that one of the two metal binding sites is impaired in the Asp-62 variants. In comparison, wild-type TmPep1050 could bind two cobalt ions with an apparent association constant of 50 μM.[13] For the Asp-62 variants, the cobalt quantification did not allow to determine whether both M1 and M2 sites were partially occupied or one of them remained vacant. The second hypothesis is the likeliest based on (a) the dimeric state of Aps-62 variants and (b) the M1 site controlling the dimer–dodecamer transition. Thus, the M2 site was inferred to be functional even when Asp-62 was substituted. Our results contrast with what has been reported for the human copper/zinc superoxide dismutase. This enzyme displays an Asp-His-metal triad analogous to that of the MH clan enzymes. The substitution of the aspartate residue with an asparagine or alanine residue completely abolished both the activity and the metal ion binding.[40]

3.3 | The structure of $TmPep1050_{D62A}$ reveals how Asp-62 maintains the structural fold of the active site

To understand how Asp-62 could influence Co^{2+} binding in the M1 site, we solved the structure of $TmPep1050_{D62A}$ at 1.5 Å. The crystallization condition was 0.1 M citric acid, 5% PEG3350, pH 5.2. Unfortunately, X-ray fluorescence scanning of crystals did not detect any trace of metal ion, probably due to cobalt chelation by citric acid.[41] Co^{2+} binding assays were undertaken for $TmPep1050_{D62A}$ to demonstrate the interference of citric acid in metal ion binding. $TmPep1050_{D62A}$ was incubated with Co^{2+} at a concentration ranging from 0 to 320 μM according to the crystallization condition. Thermal shift assay showed that the melting temperature, 84.6°C, was invariant whichever the Co^{2+} concentration (see Table S2). Furthermore, the concentration of bound Co^{2+} remained under the detection limit, that is, less than 1 μmol Co^{2+} per 20 μmol enzyme).

Based on the crystal space group and unit cell parameters (see Table 1), $TmPep1050_{D62A}$ is a dimer according to PDBe Pisa. The interaction interface between the two subunits is similar to that of TmPep1050 dimer.[13] In addition, the absence of a 3-fold symmetry strongly supports the dimeric state observed in the crystal unit cell. Such a symmetry has been reported for several structures of dodecameric M42 aminopeptidases.[1,3-5,11,13] When compared with $TmPep1050_{12\text{-mer}}$, the subunit structure of $TmPep1050_{D62A}$ presents the same structural dissimilarities as TmPep1050 dimer (Figure 4A and Figure S2).[13] Two segments, Lys-229-Ala-235 and Lys-247-Ser-254, diverge greatly between both structures. For instance, the backbone is displaced by 3 and 11 Å for Lys-232 and Arg-249, respectively. Both segments were described as critical in the oligomerization as they are involved in the interaction between dimers. The Asp-62 substitution also affects the catalytic pocket fold as the α8 and α10 helices are too flexible to be modeled. Furthermore, the Gln-196-Val-202 loop is displaced outward the catalytic site, explaining the lack of activity of $TmPep1050_{D62A}$ (Figure 5).

Because of the Gln-196-Val-202 loop displacement, Glu-197 and Glu-198 residues are not correctly positioned to fulfill their function (Figure 6). The former is the catalytic base, assisting the deprotonation of a water molecule during the first step of the peptide bond hydrolysis. The latter is involved in the binding of a cobalt ion at the M1 site, along with His-307 and Asp-168. In the $TmPep1050_{D62A}$ structure, the

His-307 side chain adopts a m-70° rotamer, compared with the m170° rotamer observed in the dodecamer structure. Consequently, the imidazole ring is tilted by about 28° between both structures. Asp-168 belongs to both M1 and M2 sites as it bridges the two cobalt ions. The Asp-62 substitution did not have any effect on the Asp-168 position. Regarding the M2 site, the side chains of Asp-220 and His-60 are oriented differently in $TmPep1050_{D62A}$ compared with $TmPep1050_{12\text{-mer}}$. The Asp-220 side chain adopts an m-20° rotamer instead of a t0° rotamer found in the dodecamer structure. Due to side chain orientation change, the carboxylate center is displaced by 2.4 Å and interacts with the carbonyl of Ile-221. In the monomer A of $TmPep1050_{D62A}$, the imidazole ring of His-60 is correctly positioned with a p-80° rotamer but is rotated by 180°. Consequently, $N^{\delta1}$ and $N^{\varepsilon2}$ of

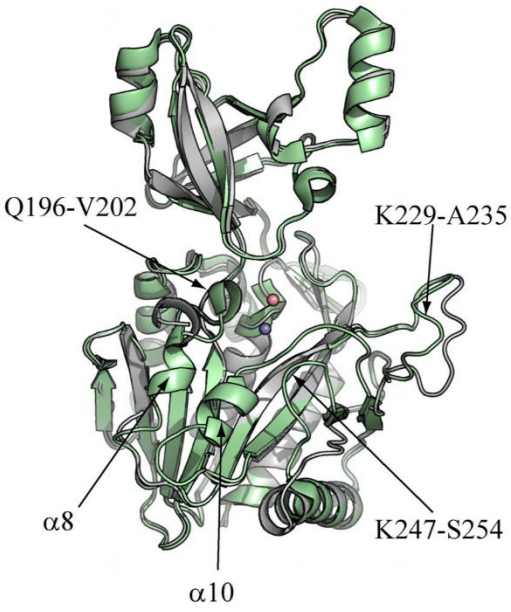


FIGURE 4 Structural alignment of $TmPep1050_{D62A}$ subunit (PDB code 5L6Z, light grey) with $TmPep1050_{12\text{-mer}}$ subunit (PDB code 6NW5, green). The structural dissimilarities are indicated with arrows. Zn^{2+} and Co^{2+} are represented as gray and pink spheres, respectively. Structural alignment r.m.s.d. value is 0.504 Å

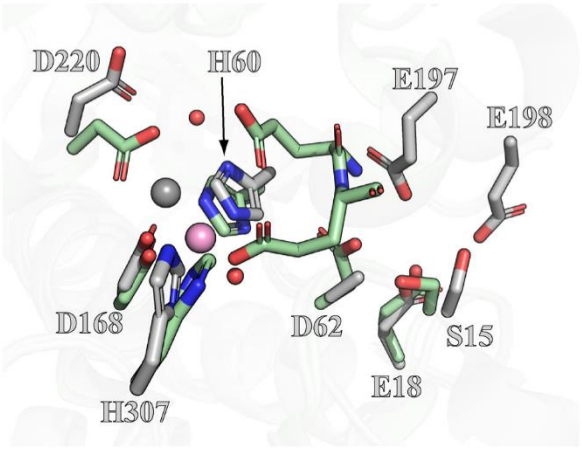


FIGURE 6 Close-up view of the active sites of $TmPep1050_{D62A}$ (light gray) and $TmPep1050_{12\text{-mer}}$ dodecamer (green). The water molecules interacting with His-60 in the $TmPep1050_{D62A}$ structure are showed as red spheres. Zn^{2+} and Co^{2+} are represented as gray and pink spheres, respectively

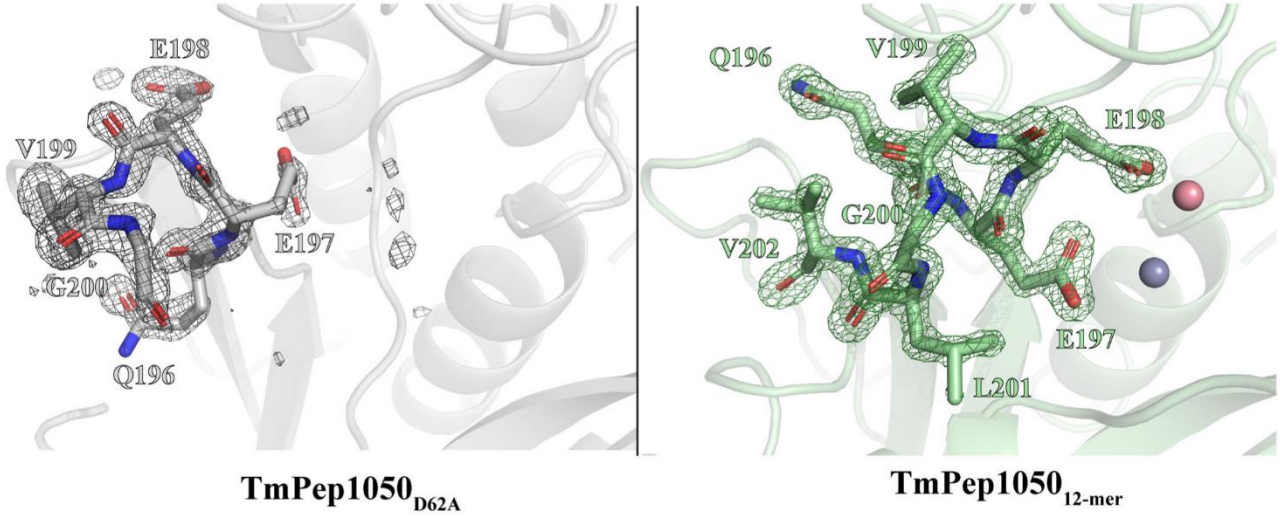


FIGURE 5 The Gln-196-Val202 loop displacement in the $TmPep1050_{D62A}$ structure. The $TmPep1050_{D62A}$ subunit (left, light grey) is compared with $TmPep1050_{12\text{-mer}}$ subunit (right, green). The loop displacement is justified by omit maps represented as dark gray and dark green meshes for $TmPep1050_{D62A}$ and $TmPep1050_{12\text{-mer}}$, respectively. Omit maps were calculated using Phenix by omitting Gln-196-Val202. Zn^{2+} and Co^{2+} are represented as grey and pink spheres, respectively

His-60 are found interacting with two water molecules which are not observed in the dodecamer structure. In the monomer B, His-60 side chain adopts two alternate rotamers, m90° and p-80°, with an occupation of 0.57 and 0.43, respectively.

Excepting the alternate rotamers of His-60, the TmPep1050 dimer structure shows the same dissimilarities in the active site as $TmPep1050_{D62A}$. Despite being structurally similar, TmPep1050 dimers are still able to bind two cobalt ions and revert to dodecamers while $TmPep1050_{D62A}$ cannot. Consequently, the Asp-62 substitution must have a structural role for stabilizing the metal-bound form. In the dodecamer structure, $O^{\delta1}$ and $O^{\delta2}$ of Asp-62 are in tight H-bond interaction with the backbone nitrogen of Glu-197 and Glu-198, respectively (Figure 5B). Thus, the Asp62Ala substitution probably prevents this interaction and, consequently, could not stabilize the Gln-196-Val-202 loop. It has a dramatic impact on the M1 site as Glu-198 cannot be correctly positioned to allow the binding of a cobalt ion. The side chain length and charge of residue 62 are critical since $TmPep1050_{D62N}$ and $TmPep1050_{D62E}$ are also inactive dimers. The position of the Asp-62 carboxylate is probably imposed by H-bond interactions with Ser-15 and His-60. In the $TmPep1050_{D62A}$ structure, Ser-15 hydroxyl is rotated by 107.2° compared with wild-type dimer and dodecamer. In addition, charge repulsion may occur between Glu-18 and Asp-62. Placing a glutamate residue at position 62 would not allow the correct alignment of the carboxylate with Ser-15 and His-60. In the case of an asparagine residue at position 62, the amide group would not sustain the charge repulsion from Glu-18. Furthermore, the amine group would clash with either the Ser-15 or His-60 due to short H-H distance.

4 | CONCLUSION

The carboxylate-histidine-metal ion triad is commonly encountered in metalloenzymes. Christianson and Alexander (1989) reported that at least 35% of histidine residues bound to a metal ion interacts with either a glutamate or an aspartate residue.[26] The supposed role of the carboxylate is to modulate the basicity of the neighbor histidine residue, affecting the Lewis acidity of bound metal ion. It may also have an impact of metal ion binding affinity and nucleophilicity of the water molecule bound to the metal ion.[42,43] The MH clan enzymes display such a carboxylate-histidine-metal ion triad at the M2 site. Both the aspartate and histidine residues of the triad are strictly conserved in the whole clan. A function of the aspartate residue has been postulated for the MH clan archetypal enzyme, the *V. proteolyticus* aminopeptidase Ap1.[21-23,25,27] Asp-99 has been described to decrease Lewis acidity of the M2 site Zn^{2+} via a strong H-bond between Asp-99 and His-97. Consequently, the interaction between the nucleophilic hydroxide and the M2 site Zn^{2+} is disfavored and may transitorily alter the coordination geometry of the metal ion.

In this work, we showed that the aspartate residue of the Asp-His-metal triad could have another function for the M42 aminopeptidases. Using TmPep1050 as a case of study, Asp-62, equivalent to Asp-99 of *V. proteolyticus* aminopeptidase Ap1, was substituted with alanine, asparagine, or glutamate residue. This substitution completely abolished the activity, as expected, but also interfered in the dodecamer formation. We suppose that the Asp-62 substitution affects the M1 binding site as the oligomerization of TmPep1050 strictly relies on the M1. The structure of $TmPep1050_{D62A}$ presented the same structural dissimilarities as observed for the wild-type dimer. Notably, the Gln-196-Val-202 loop is disordered, preventing the correct positioning of Glu-197 and Glu-198 (the catalytic base and a ligand of the M1 site). Asp-62 has probably a structural role for stabilizing this active site loop since its carboxylate function is in close interaction with the backbone nitrogen of Glu-197 and Glu-198. Nevertheless, we could not ascertain the structural role of Asp-62 to all MH clan enzymes, as no structure of equivalent variant has been reported in other families. It would be interesting to know whether substituting the aspartate residue would destabilize the active site of a monomeric enzyme like *V. proteolyticus* aminopeptidase Ap1. If the destabilization is also observed for such an enzyme, it will support the structural role of this strictly conserved aspartate residue thorough the whole MH clan. If not, it will indicate that the destabilization of the active site as observed for $TmPep1050_{D62A}$ could be an adaptation to oligomerization. Nevertheless, both hypotheses agree with the dependency of M42 aminopeptidase oligomerization on the active site fold.

ACKNOWLEDGMENTS

We thank Martine Roovers for proof reading this article and giving constructive comments. Access to BM30A beamline (European Synchrotron Radiation Facility) was supported by Fonds National de la Recherche Scientifique under Contract IISN 4.4503.11F.

ORCID

Raphaël Dutoit https://orcid.org/0000-0002-7139-8402

REFERENCES

1. Borissenko L, Groll M. Crystal structure of TET protease reveals complementary protein degradation pathways in prokaryotes. *J Mol Biol*. 2005;346(5):1207-1219. https://doi.org/10.1016/j.jmb.2004.12.056.
2. Appolaire A, Colombo M, Basbous H, Gabel F, Girard E, Franzetti B. TET peptidases: a family of tetrahedral complexes conserved in prokaryotes. *Biochimie*. 2016;122:188-196. https://doi.org/10.1016/j.biochi.2015.11.001.
3. Russo S, Baumann U. Crystal structure of a dodecameric tetrahedral-shaped aminopeptidase. *J Biol Chem*. 2004;279(49):51275-51281. https://doi.org/10.1074/jbc.M409455200.
4. Schoehn G, Vellieux FMD, Asunción Durá M, et al. An archaeal peptidase assembles into two different quaternary structures: a tetrahedron and a giant octahedron. *J Biol Chem*. 2006;281(47):36327-36337. https://doi.org/10.1074/jbc.M604417200.
5. Durá MA, Rosenbaum E, Larabi A, Gabel F, Vellieux FMD, Franzetti B. The structural and biochemical characterizations of a novel TET peptidase complex from *Pyrococcus horikoshii* reveal an integrated peptide degradation system in hyperthermophilic Archaea: characterization of *P. horikoshii* TET3 peptidase. *Mol Microbiol*. 2009;72(1):26-40. https://doi.org/10.1111/j.1365-2958.2009.06600.x.
6. Basbous H, Appolaire A, Girard E, Franzetti B. characterization of a glycyl-specific TET aminopeptidase complex from *Pyrococcus horikoshii*. *J Bacteriol*. 2018;200(17):e00059-18. https://doi.org/10.1128/JB.00059-18.

7. Appolaire A, Girard E, Colombo M, et al. Small-angle neutron scattering reveals the assembly mode and oligomeric architecture of TET, a large, dodecameric aminopeptidase. *Acta Crystallogr Sect D Biol Crystallogr*. 2014;70(11):2983-2993. https://doi.org/10.1107/S1399004714018446.
8. Appolaire A, Durá MA, Ferruit M, et al. The TET2 and TET3 aminopeptidases from *P yrococcus horikoshii* form a hetero-subunit peptidasome with enhanced peptide destruction properties: TET aminopeptidase multi-subunit complex. *Mol Microbiol*. 2014;94(4):803-814. https://doi.org/10.1111/mmi.12775.
9. Franzetti B, Schoehn G, Hernandez J-F, Jaquinod M, Ruigrok RWH, Zaccai G. Tetrahedral aminopeptidase: a novel large protease complex from archaea. *EMBO J*. 2002;21(9):2132-2138. https://doi.org/10.1093/emboj/21.9.2132.
10. Kim D, San BH, Moh SH, et al. Structural basis for the substrate specificity of PepA from *Streptococcus pneumoniae*, a dodecameric tetrahedral protease. *Biochem Biophys Res Commun*. 2010;391(1):431-436. https://doi.org/10.1016/j.bbrc.2009.11.075.
11. Petrova TE, Slutskaya ES, Boyko KM, Sokolova OS, Rakitina TV, Korzhenevskiy DA, Gorbacheva MA, Bezsudnova EY, Popov VO Structure of the dodecamer of the aminopeptidase APDkam598 from the archaeon *Desulfurococcus kamchatkensis*. Acta Crystallogr Sect F Struct Biol Commun 2015;71(3):277–285. doi:https://doi.org/10.1107/S2053230X15000783
12. Colombo M, Girard E, Franzetti B. Tuned by metals: the TET peptidase activity is controlled by 3 metal binding sites. *Sci Rep*. 2016;6(1):20876. https://doi.org/10.1038/srep20876.
13. Dutoit R, Van Gompel T, Brandt N, et al. How metal cofactors drive dimer–dodecamer transition of the M42 aminopeptidase TmPep1050 of *Thermotoga maritima*. *J Biol Chem*. 2019;294(47):17777-17789. https://doi.org/10.1074/jbc.RA119.009281.
14. Gauto DF, Estrozi LF, Schwieters CD, et al. Integrated NMR and cryo-EM atomic-resolution structure determination of a half-megadalton enzyme complex. *Nat Commun*. 2019;10(1):2697. https://doi.org/10.1038/s41467-019-10490-9.
15. Rosenbaum E, Ferruit M, Durá MA, Franzetti B. Studies on the parameters controlling the stability of the TET peptidase superstructure from *Pyrococcus horikoshii* revealed a crucial role of pH and catalytic metals in the oligomerization process. *Biochim Biophys Acta*. 2011;1814(10):1289-1294. https://doi.org/10.1016/j.bbapap.2010.11.008.
16. Macek P, Kerfah R, Erba EB, et al. Unraveling self-assembly pathways of the 468-kDa proteolytic machine TET2. *Sci Adv*. 2017;3(4):e1601601. https://doi.org/10.1126/sciadv.1601601.
17. Dutoit R, Brandt N, Legrain C, Bauvois C. Functional characterization of two M42 aminopeptidases erroneously annotated as cellulases. *PLoS One*. 2012;7(11):e50639. https://doi.org/10.1371/journal.pone.0050639.
18. Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A, Finn RD. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res*. 2018;46(D1):D624-D632. https://doi.org/10.1093/nar/gkx1134.
19. Chevrier B, Schalk C, D'Orchymont H, Rondeau J-M, Moras D, Tarnus C. Crystal structure of *Aeromonas proteolytica* aminopeptidase: a prototypical member of the co-catalytic zinc enzyme family. *Structure*. 1994;2:283-291. https://doi.org/10.1016/s0969-2126(00)00030-7.
20. Neuwald AF, Liu JS, Lipman DJ, Lawrence CE. Extracting protein alignment models from the sequence database. *Nucleic Acids Res*. 1997;25(9):1665-1677. https://doi.org/10.1093/nar/25.9.1665.
21. Stamper C, Bennett B, Edwards T, Holz RC, Ringe D, Petsko G. Inhibition of the aminopeptidase from *Aeromonas proteolytica* by L-leucinephosphonic acid. Spectroscopic and crystallographic characterization of the transition state of peptide hydrolysis. *Biochemistry*. 2001;40(24):7035-7046. https://doi.org/10.1021/bi0100891.
22. Holz RC. The aminopeptidase from *Aeromonas proteolytica*: structure and mechanism of co-catalytic metal centers involved in peptide hydrolysis. *Coord Chem Rev*. 2002;232(1–2):5-26. https://doi.org/10.1016/S0010-8545(01)00470-2.
23. Schürer G, Lanig H, Clark T. Aeromonas proteolytica aminopeptidase: an investigation of the mode of action using a quantum mechanical/molecular mechanical approach. *Biochemistry*. 2004;43(18):5414-5427. https://doi.org/10.1021/bi0340191.
24. Bzymek KP, Holz RC. The catalytic role of glutamate 151 in the leucine aminopeptidase from *Aeromonas proteolytica*. *J Biol Chem*. 2004;279(30):31018-31025. https://doi.org/10.1074/jbc.M404035200.
25. De Paola CC, Bennett B, Holz RC, Ringe D, Petsko GA. 1-Butaneboronic acid binding to *Aeromonas proteolytica* aminopeptidase: a case of arrested development. *Biochemistry*. 1999;38(28):9048-9053. https://doi.org/10.1021/bi9900572.
26. Christianson DW, Alexander RS. Carboxylate-histidine-zinc interactions in protein structure and function. *J Am Chem Soc*. 1989;111(16):6412-6419. https://doi.org/10.1021/ja00198a065.
27. Desmarais W, Bienvenue DL, Bzymek KP, Petsko GA, Ringe D, Holz RC. The high-resolution structures of the neutral and the low pH crystals of aminopeptidase from *Aeromonas proteolytica*. *J Biol Inorg Chem*. 2006;11(4):398-408. https://doi.org/10.1007/s00775-006-0093-x.
28. Edelheit O, Hanukoglu A, Hanukoglu I. Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies. *BMC Biotechnol*. 2009;9(1):61. https://doi.org/10.1186/1472-6750-9-61.
29. Casadaban MJ, Cohen SN. Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. *J Mol Biol*. 1980;138(2):179-207. https://doi.org/10.1016/0022-2836(80)90283-1.
30. Dutoit R, Brandt N, Elder DV, Droogmans L. X-ray crystallography to study the oligomeric state transition of the *Thermotoga maritima* M42 aminopeptidase TmPep1050. *J Vis Exp*. 2020;(159):e61288. https://doi.org/10.3791/61288.
31. Ferrer J-L. Automated data processing on beamline FIP (BM30A) at ESRF. *Acta Crystallogr Sect D Biol Crystallogr*. 2001;57(11):1752-1753. https://doi.org/10.1107/S0907444901013385.
32. Roth M, Carpentier P, Kaïkati O, et al. FIP: a highly automated beamline for multiwavelength anomalous diffraction experiments. *Acta Crystallogr Sect D Biol Crystallogr*. 2002;58(5):805-814. https://doi.org/10.1107/S0907444902003943.
33. Kabsch W. XDS. *Acta Crystallogr Sect D Biol Crystallogr*. 2010;66(2):125-132. https://doi.org/10.1107/S0907444909047337.
34. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. *Phaser* crystallographic software. *J Appl Cryst*. 2007;40(4):658-674. https://doi.org/10.1107/S0021889807021206.
35. Adams PD, Afonine PV, Bunkóczi G, et al. *PHENIX*: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr Sect D Biol Crystallogr*. 2010;66(2):213-221. https://doi.org/10.1107/S0907444909052925.
36. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of *Coot*. *Acta Crystallogr Sect D Biol Crystallogr*. 2010;66(4):486-501. https://doi.org/10.1107/S0907444910007493.
37. Williams CJ, Headd JJ, Moriarty NW, et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci*. 2018;27(1):293-315. https://doi.org/10.1002/pro.3330.
38. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol*. 2007;372(3):774-797. https://doi.org/10.1016/j.jmb.2007.05.022.
39. Jubb HC, Higueruelo AP, Ochoa-Montaño B, Pitt WR, Ascher DB, Blundell TL. Arpeggio: a web server for calculating and visualising interatomic interactions in protein structures. *J Mol Biol*. 2017;429(3):365-371. https://doi.org/10.1016/j.jmb.2016.12.004.
40. Banci L, Bertini I, Cabelli DE, Hallewell RA, Tung JW, Viezzoli MS. A characterization of copper/zinc superoxide dismutase mutants at position 124 zinc-deficient proteins. *Eur J Biochem*. 1991;196(1):123-128. https://doi.org/10.1111/j.1432-1033.1991.tb15794.x.

41. Wyrzykowski D, Chmurzyński L. Thermodynamics of citrate complexation with Mn^{2+}, Co^{2+}, Ni^{2+} and Zn^{2+} ions. *J Therm Anal Calorim*. 2010;102(1):61-64. https://doi.org/10.1007/s10973-009-0523-4.
42. El Yazal J, Pang Y-P. Proton dissociation energies of zinc-coordinated hydroxamic acids and their relative affinities for zinc: insight into design of inhibitors of zinc-containing proteinases. *J Phys Chem B*. 2000;104(27):6499-6504. https://doi.org/10.1021/jp0012707.
43. Lin LC. Factors governing the protonation state of Zn-bound histidine in proteins: a DFT/CDM study. *J Am Chem Soc*. 2004;126(8):2602-2612. https://doi.org/10.1021/ja038827r.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Dutoit R, Brandt N, Van Gompel T, et al. M42 aminopeptidase catalytic site: the structural and functional role of a strictly conserved aspartate residue. *Proteins*. 2020;1–9. https://doi.org/10.1002/prot.25982

Oligonucleotide name	Sequence	Description
ocej462	5'-GATACTGGACGCTCACATAaatGAGATAGGTGTTGTCGTCAC	Asp-62 codon mutation to Asn codon, forward primer complementary to ocej463
ocej463	5'-GTGACGACAACACCTATCTCattTATGTGAGCGTCCAGTATC	Asp-62 codon mutation to Asn codon, reverse primer complementary to ocej462
ocej464	5'-GATACTGGACGCTCACATAgctGAGATAGGTGTTGTCGTCAC	Asp-62 codon mutation to Ala codon, forward primer complementary to ocej465
ocej465	5'-GTGACGACAACACCTATCTCagcTATGTGAGCGTCCAGTATC	Asp-62 codon mutation to Ala codon, reverse primer complementary to ocej464
ocej820	5'-TGATACTGGACGCTcacATAGAAGAGATAGGTGTTGTCGTCACA	Asp-62 codon mutation to Glu codon, forward primer complementary to ocej821
ocej821	5'-TGTGACGACAACACCTATCTCTTCTATgtgAGCGTCCAGTATCA	Asp-62 codon mutation to Glu codon, reverse primer complementary to ocej820

Table S1 List of the oligonucleotides used for Tm_1050 ORF mutagenesis.

[$CoCl_2$] (µM)	T_m (°C)
0	84.73 ± 0.04
20	84.63 ± 0.08
40	84.53 ± 0.15
80	84.70 ± 0.12
160	84.73 ± 0.04
320	84.00 ± 0.12

Table S2 Thermal shift assay of Co^{2+} effect on $TmPep1050_{D62A}$ thermostability in 0.1 M sodium citrate, 5% PEG3350, pH 5.2. Standard error is given for each value with n = 3.

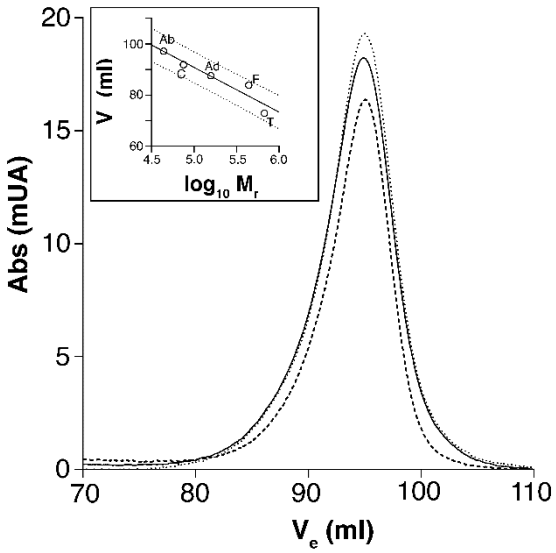


Figure S1 Size exclusion chromatography of $TmPep1050_{D62A}$ (plain line), $TmPep1050_{D62N}$ (dot line), and $TmPep1050_{D62E}$ (dashed line). An elution peak was observed at 95 ml for the three variants using a Superdex 200 column (V = 120 ml). *Inset:* The column was calibrated using albumin (Ab), conalbumin (C), aldolase (Ad), ferritin (F), and thyroglobulin (T) as standards. The correlation between the logarithm of the relative mass and the elution volume is linear, with a R^2 of 0.91. The 95% confidence intervals are represented as dots.

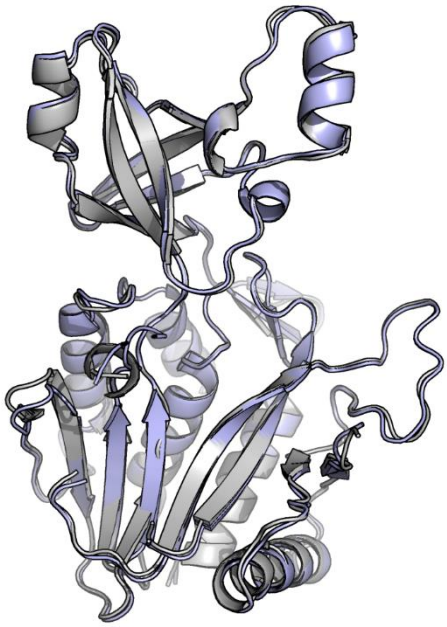


Figure S2 Structural alignment of TmPep1050D62A subunit (PDB code 5L6Z, light grey with TmPep1050 dimer subunit (PDB code 5NE6, light blue). Alignment r.m.s.d. value is 0.470 Å.

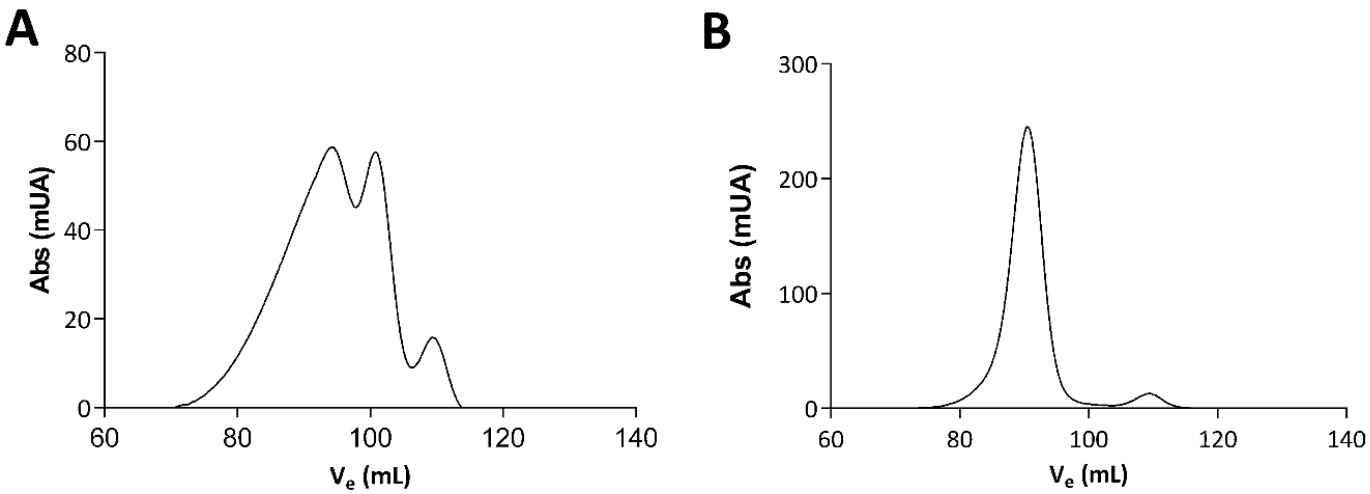


Figure 48 – Determination of oligomeric states of (A) TmPep1048 and (B) TmPep1049. The oligomeric states were determined by size-exclusion chromatography using a Superdex200 XK16/70 column. Abs, absorbance. mUA, milli unit of absorbance. V_e, volume of elution.

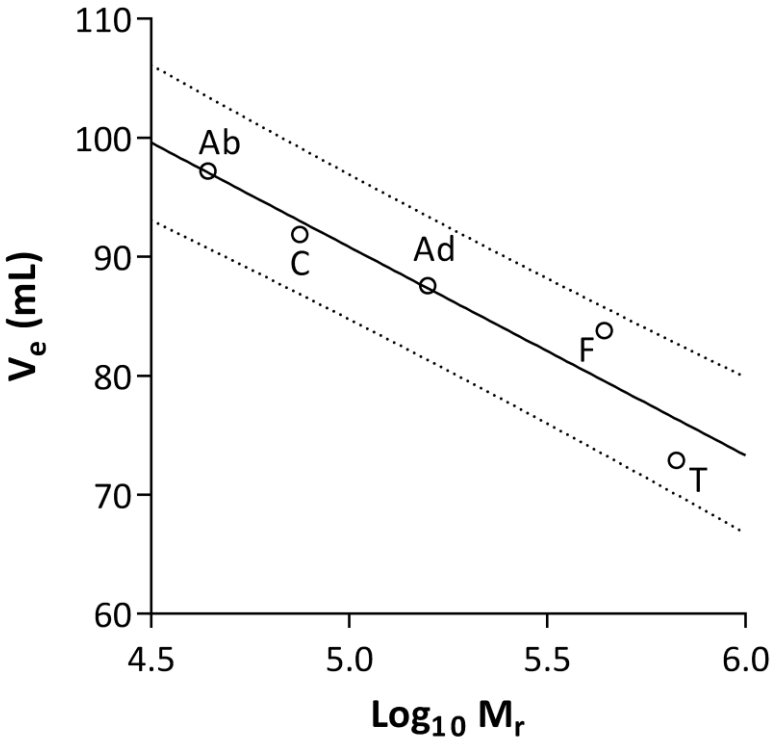


Figure 49 – Calibration of the Superdex200 XK16/70 column used to determine the relative mass of TmPep1048 and TmPep1049. Albumin (Ab), Conalbumin (C), Aldolase (Ad), Ferritin (F), and Thyroglobulin (T) were used as standards. The correlation between the elution volume (V_e) and the relative mass (M_r) is linear with R^2 of 0.91. The 95% confidence intervals of the linear regression are shown as dots.

B.4 Characterization of TmPep1048 and TmPep1049

B.4.1 Background

The genome of *T. maritima* possesses three genes coding M42 aminopeptidases, *TM_1048*, *TM_1049*, and *TM_1050*, the three ORFs being organized as an operon. As discussed in the introduction (see **section A.5.3.4**, page 34), the structures of enzymes corresponding to these genes were deposited in the PDB as part of a structural genomic project[450]. Based on structural data (PDB codes: 1VHO, 2FVG, and 3ISX), TmPep1048 and TmPep1049 are monomers while TmPep1050 is a dimer. Although, we showed that TmPep1050 is a dodecameric leucyl-aminopeptidase activated by Co^{2+} (see **section B.1**, page 41). The discrepancy between our results and the structural genomic project could be due to an artefact linked to the histidine tag used in the high-throughput structure determination pipeline. Thus, we could reasonably suspect that the structures of TmPep1048 and TmPep1049 could be erroneous like for TmPep1050.

B.4.2 Methodology

To study the oligomeric state of TmPep1048 and TmPep1049, both proteins were recombinantly produced in *E. coli* and purified according to the protocol developed to study TmPep1050 (see **section B.2.4**, page 71). In addition, both proteins were produced as Strep-tag fusion. The Strep-tag consists of an eight-residue peptide (Trp-Ser-His-Pro-Gln-Phe-Glu-Lys) which displays a near covalent affinity to streptavidin[451]. It allows to purify Strep-tag fusion proteins by affinity chromatography. The specific activity of TmPep1048 and TmPep1049 was determined using Xaa-*p*NA derivatives and Xaa-Ala dipeptides.

B.4.3 Results and discussion

TmPep1048 and TmPep1049 were purified either as Strep-tag fusion or devoid of any purification tag. In both cases, size-exclusion chromatography showed that TmPep1048 and TmPep1049 are not dodecameric. Indeed, two major peaks were observed for TmPep1048 at elution volumes of 94.3 and 100.8 ml (**Figure 48.A**). The first peak corresponds to a dimer with an apparent relative mass of 68.4 kDa while the second peak corresponds to a species of only 16.2 kDa (the TmPep1048 monomer relative mass being 36.6 kDa). The species of low relative mass could be unstructured monomers since (i) the sample was concentrated using a filter of 30 kDa cut-off and (ii) no degradation product was observed by SDS-PAGE. Of note, the first peak broadness suggested that other oligomers could exist. The elution profile of TmPep1049 was quite different from TmPep1048 as only an evenly distributed peak was observed at an elution volume of 90.6 ml (**Figure 48.B**). According to the column calibration (**Figure 49**), it corresponds to a tetramer with an apparent relative mass of 132.7 kDa (the TmPep1049 monomer relative mass being 37.3 kDa). Notwithstanding not being dodecamers, TmPep1048 and TmPep1049 were assayed on various Xaa-*p*NA derivatives but no significant activity was observed (i.e. specific activity less than 0.05 µmol of *p*NA s^{-1} per µmol of enzyme). As some M42 aminopeptidases are unable to hydrolyze chromogenic substrates[326], their activity was also assayed using Xaa-Ala dipeptides. One µM of enzyme was incubated with 5 mM of dipeptides for one hour at 75°C. The hydrolysis products were then analyzed by TLC, but no significant activity was detected on none of the substrates.

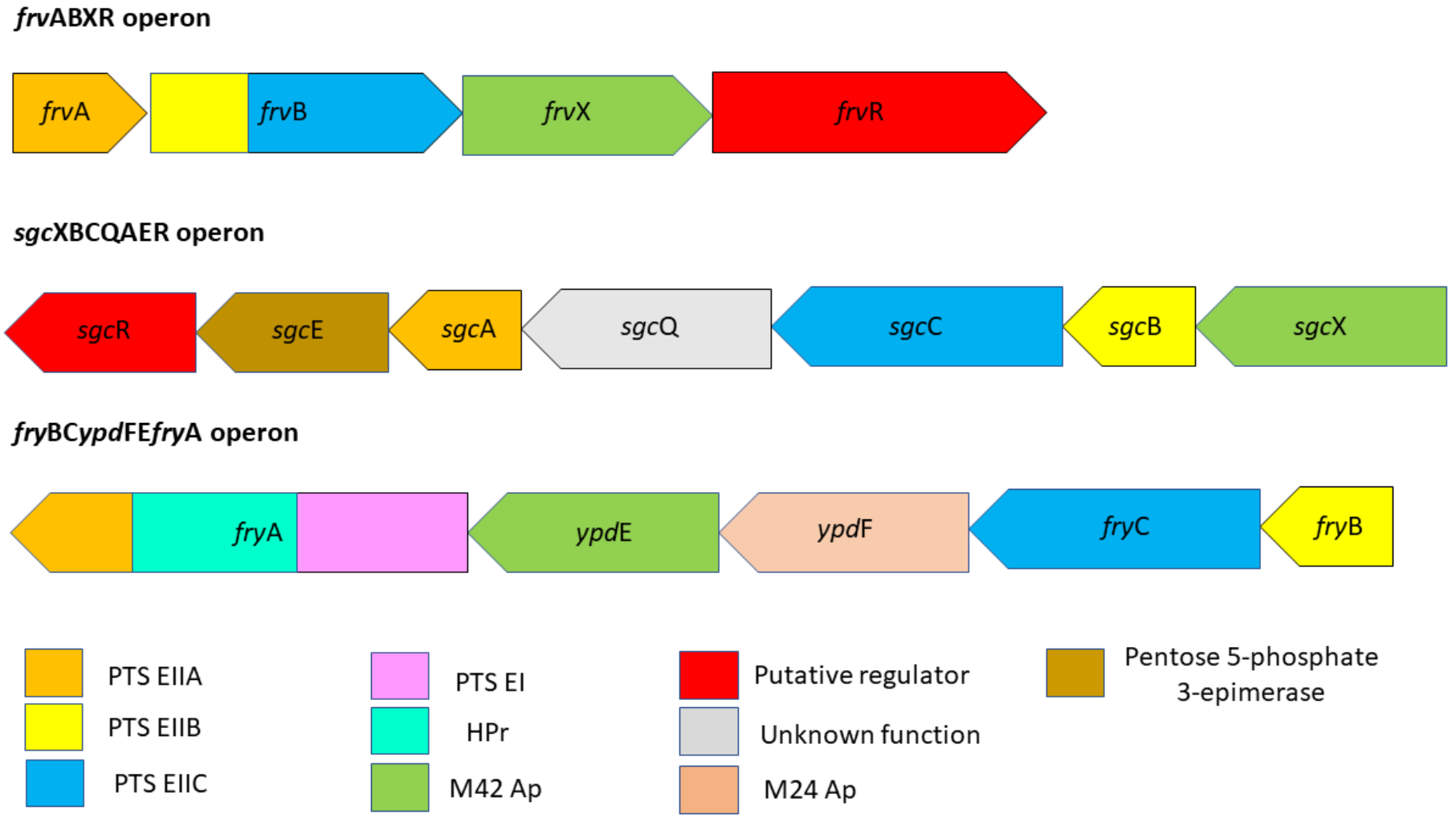


Figure 50 – The *frv*ABXR, *sgc*XBCQAER, and *fry*BC*ypd*FE*fry*A operons of *E. coli*. Each ORF is color-coded according to their function. Ap, aminopeptidase. EI, enzyme I. EII, enzyme II. HPr, phosphocarrier protein.

Our preliminary data showed that TmPep1048 and TmPep1049 are not monomers despite being annotated so based on structural data. Both proteins, however, failed to form active dodecamers under the tested experimental conditions. A plausible explanation could be that TmPep1048 and TmPep1049 are involved in heterocomplexes either together or with TmPep1050. Such a hypothesis is supported by the study of the M42 aminopeptidase GsApI. Indeed, GsApI is a heterododecamer composed of two different subunits, Iα and Iβ[372,423,432]. Several stoichiometries have been shown to exist *in vivo* in *Geobacillus stearothermophilus*: ten Iα for two Iβ, eight Iα for four Iβ, and six Iα for six Iβ[432]. Interestingly, Stoll *et al.* (1972) reported that the Iα subunit can form homododecamers while Iβ subunits fail to reassociate[423]. The Iβ subunit, however, has been reported to be important for the aspartyl-aminopeptidase activity of GsApI. Thus, it would be worth to co-express the whole *TM_1048-TM_1049-TM_1050* operon to show the occurrence of heterocomplexes.

B.5 Characterization of M42 aminopeptidases from *E. coli*

B.5.1 Background

Albeit the clear progress in biochemically characterizing M42 aminopeptidases, their physiological function remains poorly described due to the paucity of *in vivo* studies. Borissenko and Groll (2005) proposed that M42 aminopeptidases could be linked to the proteasome, achieving the ultimate degradation of peptides released by the proteasome or other degradation machines[184]. The existence of heterododecamers has also been reported for M42 aminopeptidases from *P. horikoshii*[268,426] and *G. stearothermophilus*[372,423,424,432]. Such heterododecamers have a broad substrate specificity recognizing almost any type of amino acid at the P_1 position, excepting proline. In addition, an activity synergy has been reported between the distinct subunits in heterocomplexes[268,332]. In analogy to the proteasome, Appolaire *et al.* (2014) even proposed to name such M42 aminopeptidase heterocomplexes "peptidasome". *E. coli* would be an interesting model for studying the function of M42 aminopeptidases due to availability of genetic tools and transcriptomic studies. Especially, it would be worth to show the existence of heterocomplexes having broad substrate specificity like those of *P. horikoshii* and *G. stearothermophilus*.

The genome of *E. coli* possesses three genes coding M42 aminopeptidases, *frv*X, *sgc*X, and *ypd*E, each gene being in a different operon. Surprisingly, the three operons are potentially linked to carbohydrate metabolism (**Figure 50**). Indeed, *frv*X is found in the *frv*ABXR operon where *frv*A and *frv*B code, respectively, an enzyme IIA and an enzyme IIBC of a putative phosphotransferase system (PTS) while *frv*R could be a regulatory gene. Likewise, the *sgc*XBCQAER operon has three genes coding PTS enzyme IIB (*sgc*B), PTS enzyme IIC (*sgc*C), and PTS enzyme IIA (*sgc*A). In addition to the PTS, *sgc*E codes a pentose 5-phosphate 3-epimerase while *sgc*R could be a regulatory gene. The *sgc*Q gene product, however, is a putative nucleoside triphosphatase. Finally, the *fry*BC*ypd*FE*fry*A operon has been proposed to be linked to the peptidoglycan metabolism[452]. It could allow the production of a complete PTS system since *fry*A codes a hybrid protein with a PTS enzyme I, a phosphocarrier protein (HPr) domain and a PTS enzyme IIA domain while *fry*B and *fry*C code a PTS enzyme IIB and a PTS enzyme IIC, respectively. In addition, *ypd*F codes a prolyl-aminopeptidase of the M24 family[327,453].

PTS systems are involved in carbohydrate transport and its regulation[454]. Each PTS system is believed to transport one carbohydrate or closely chemically related carbohydrates. A complete PTS

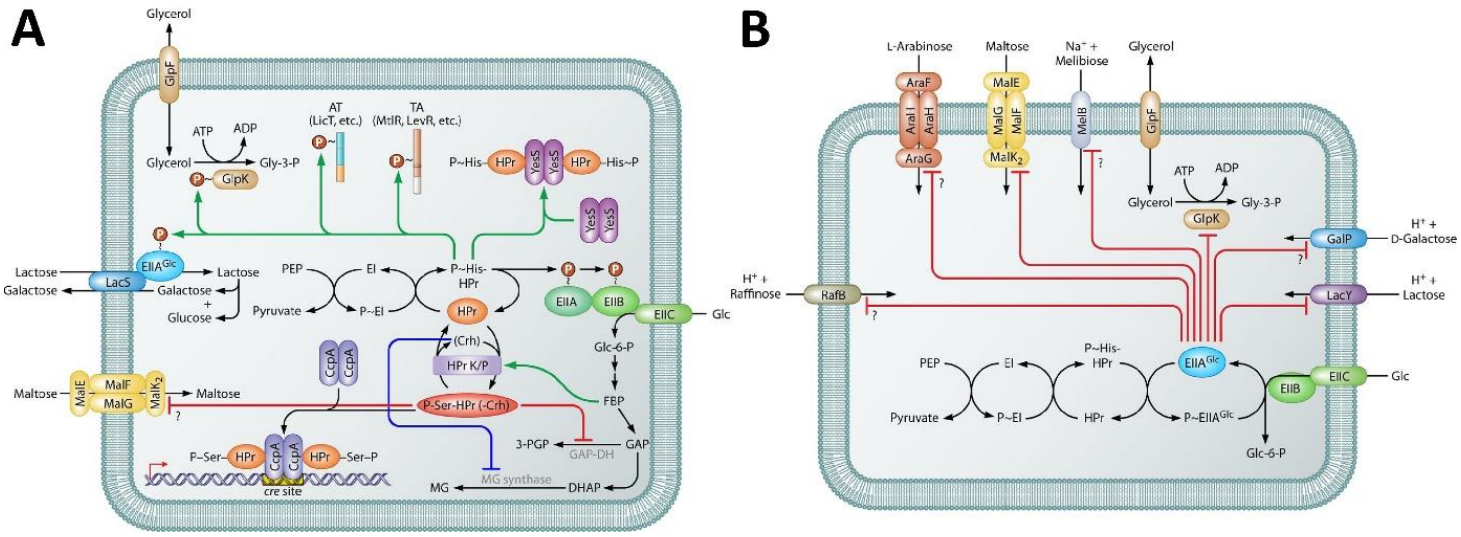


Figure 51 – Overview of the phosphotransferase system for glucose uptake in bacteria. (A) Glucose uptake by its dedicated PTS and the associated HPr regulon. **(B)** The regulation of the uptake of other carbohydrates by the enzyme IIA (EIIA) of the glucose PTS. Reproduced from Deutscher *et al.* (2014)[454].

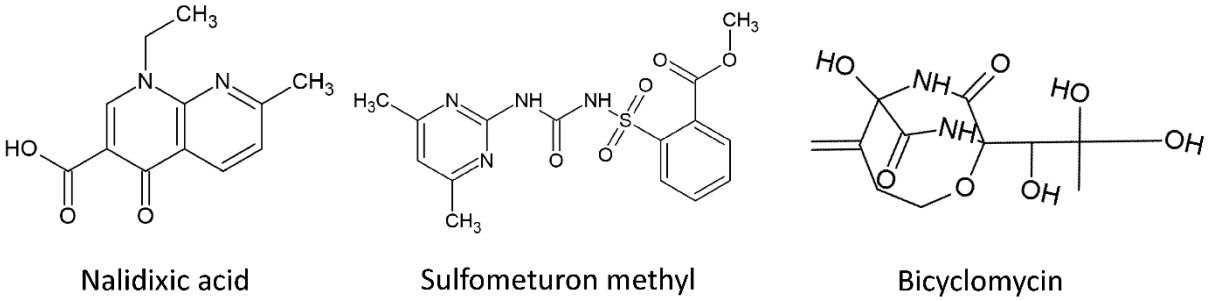


Figure 52 – Chemical compounds inducing the expression of M42 aminopeptidase coding genes in *E. coli*.

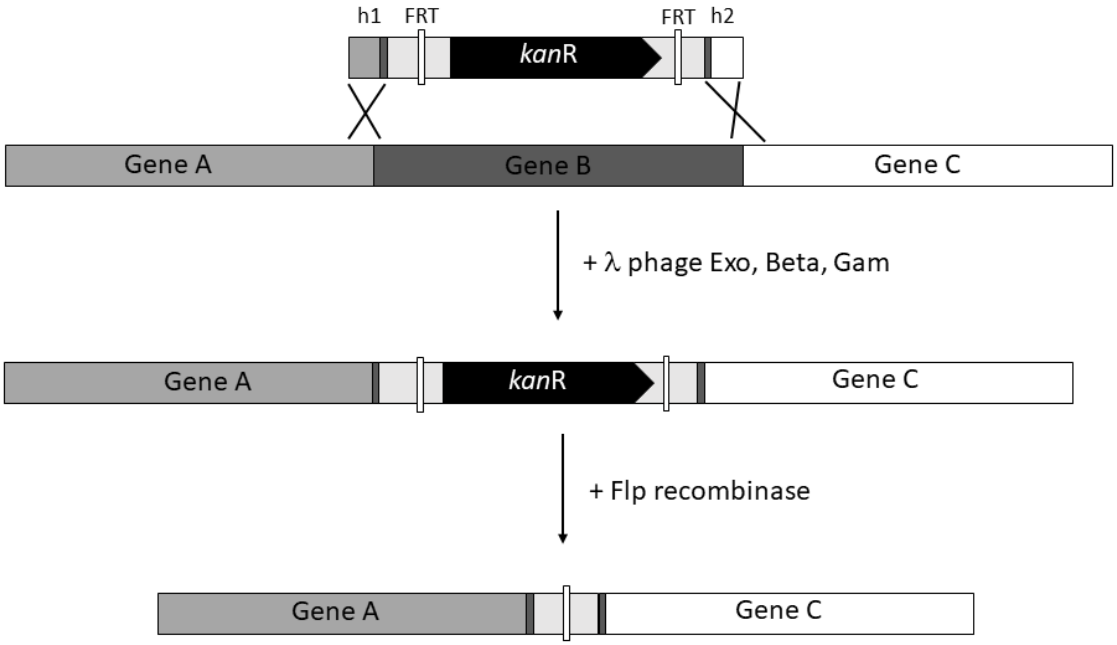


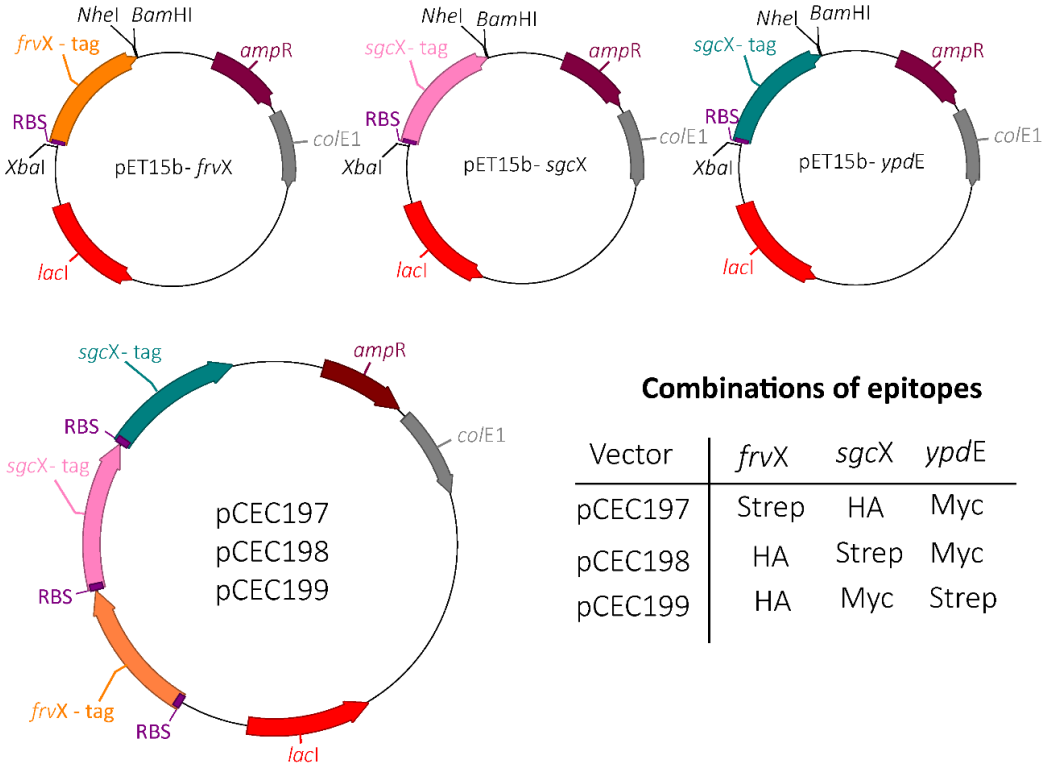
Figure 53 – Gene deletion in *E. coli*. The deletion cassette consists of the kanamycin resistance marker *kan*R flanked by two FRT sequences and two 50-pb sequences homologous to upstream (h1) and downstream (h2) genomic sequence of the targeted Gene B. The integration is mediated by the λ phage homologous recombination system (Exo, Beta, Gam). The selection marker can be excised using the *S. cerevisiae* Flp recombinase recognizing specifically the FRT sequences. The λ phage homologous recombination system and *S. cerevisiae* Flp recombinase are brought by two distinct thermosensitive plasmids. Adapted from Baba *et al.* (2006)[486].

system is usually composed of a PTS enzyme I, HPr, and up to four PTS enzyme II[455]. The different functional units can be either individual proteins or multidomain proteins. Carbohydrates are transported through the membrane PTS enzyme IIC. During their transportation, they are phosphorylated via a phosphorylation cascade involving phosphoenolpyruvate, as a phosphate donor, the PTS enzyme I, HPr, and the PTS enzyme IIA and IIB (**Figure 51.A-B**). The phosphorylated carbohydrate then enters the Embden-Meyerhof-Parnas, pentose phosphate, and Entner-Doudouroff pathways. HPr plays an important regulatory role since it phosphorylates various carbohydrate metabolism regulators (**Figure 51.A**)[454,456]. In addition, when phosphorylated by HprK, HPr can act as a transcription regulator in complex with CcpA[457,458]. The PTS enzyme IIA has a regulatory function by interacting mainly with various carbohydrate transporters (**Figure 51.B**)[454].

The expression of *frv*X, *sgc*X, and *ypd*E seems to be upregulated in different conditions. The different chemicals cited hereafter are represented in **Figure 52**. The expression of *frv*ABXR operon is upregulated in response to either an oxidative stress[459,460] or a nalidixic acid treatment[461] or an exposure to sulfometuron methyl[462]. Nalidixic acid is a synthetic quinolone that inhibits DNA replication and the gyrase subunit A[463,464]. Meanwhile, sulfometuron methyl is a herbicide known to inhibit the synthesis of branched amino acids by acting on the acetolactate synthase[465]. The expression of *fry*BC*ypd*FE*fry*A operon responds to the passage from planktonic to sessile lifestyle[466] as well as to a bicyclomycin treatment[467]. Bicyclomycin is a natural antibiotic (cyclized Ile-Leu peptide) produced in *Streptomyces* and *Pseudomonas* sp. that targets the transcription terminator factor Rho[468,469]. Finally, the expression of *sgc*XBCQAER operon is upregulated when cells reach the stationary phase or form biofilm[470–474] as well as in response to a shift of carbohydrate source[475,476]. Of note its expression is also repressed by SlyA, a regulatory protein of the MarR family[477]. MarR is a repressor of the *mar*RAB and *mar*CD operons involved in multiple antibiotic resistance in *E. coli*[478]. MarR has a winged-helix DNA binding domain[479] and binds the palindromic operator sequence *mar*O as a dimer[478,480]. The repression is lifted when MarR binds phenolic compounds like salicylic acid and 2,4-dinitrophenol[481]. Several MarR homologs have been described acting as repressors and activators regulating genes involved in antibiotic resistance, response to oxidative stress, catabolism of phenolic compounds, and virulence[482]. SlyA has been shown to be important for virulence of pathogenic bacteria, especially for their survival in macrophages[483–485]. Regarding the characterization of *E. coli* M42 aminopeptidases, only YpdE has been studied so far. It has been shown to be an aminopeptidase with broad specificity hydrolyzing peptides up to 18 residues long[327]. Furthermore, YpdE is probably a dodecamer based on the structure of a homolog from *S. flexneri* sharing 95% of identity with YpdE (see **Figure 53**).

B.5.2 Methodology

The preliminary study of the M42 aminopeptidases of *E. coli* was conducted following two axes. Firstly, their role in *E. coli* fitness to stresses was assessed using a strain lacking *frv*X, *sgc*X, and *ypd*E. To engineer such a strain, the three genes were deleted according to the method of gene knock-out used to generate the Keio collection[486,487]. Briefly, this method relies on the integration of a deletion cassette mediated by the homologous recombination system of bacteriophage λ (see **Figure 54**). The deletion cassette consists of an antibiotic resistance gene, like *kan*R, flanked by two 50-pb sequences homologous to sequences upstream and downstream the targeted gene. It contains also two Flp recombinase recognition target (FRT) sites to allow to recycle the antibiotic marker. Two heat



Combinations of epitopes

Vector	*frv*X	*sgc*X	*ypd*E
pCEC197	Strep	HA	Myc
pCEC198	HA	Strep	Myc
pCEC199	HA	Myc	Strep

Figure 56 – Cloning strategy to engineer vectors to co-express *frv*X, *sgc*X, *ypd*E. Each gene was cloned in pET15b vector with different epitopes. Polycistronic vectors were generated by excising the gene using *Xba*I and *Bam*HI and inserting it into acceptor vector between *Nhe*I and *Bam*HI (*Xba*I and *Nhe*I restriction sites being compatible). The epitopes are: Strep-tag (Ser-Ala-Trp-Ser-His-Pro-Gln-Phe-Glu-Lys), HA-tag (Tyr-Pro-Tyr-Asp-Val-Pro-Asp-Tyr-Ala), and Myc-tag (Glu-Gln-Lys-Leu-Ile-Ser-Glu-Glu-Asp-Leu).

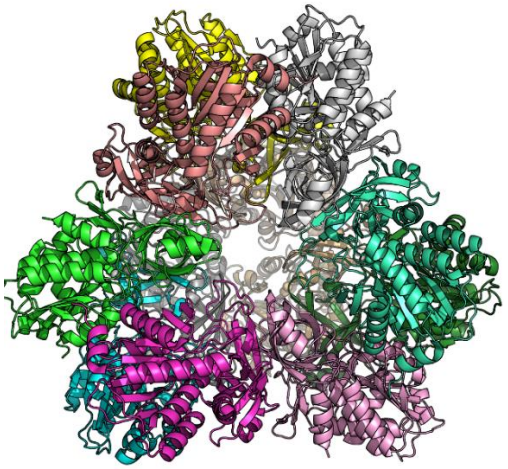


Figure 54 – Structure of YpdE of *S. flexneri* (PDB code 1YLO).

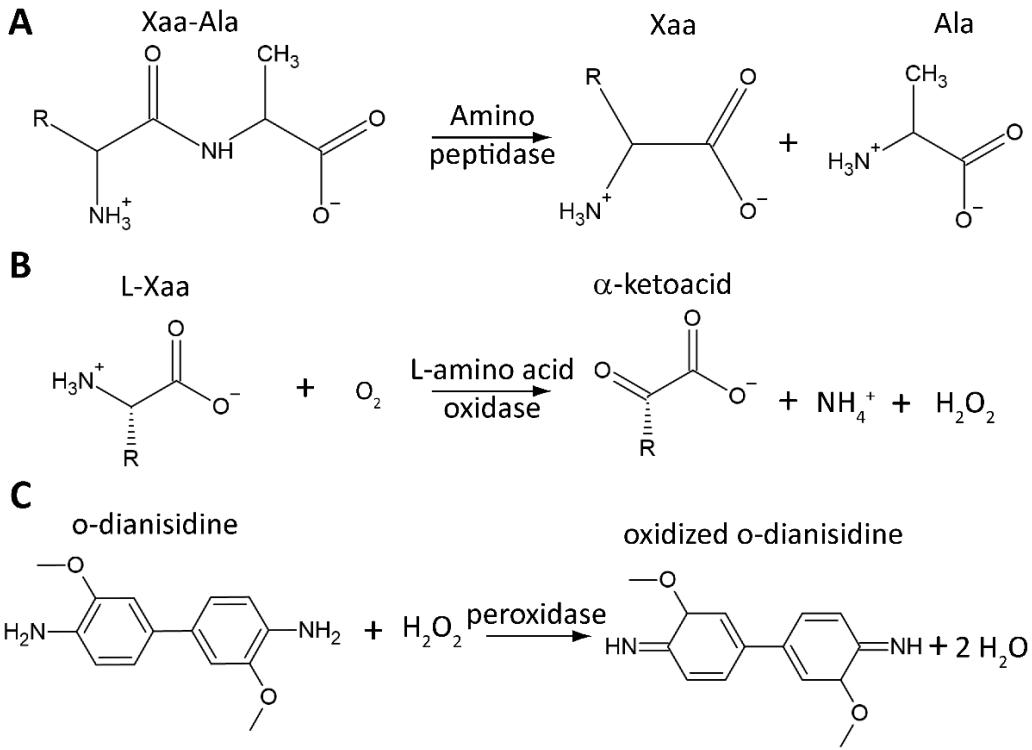


Figure 55 – Principle of coupling (A) aminopeptidase activity with (B) L-amino acid oxidase and (C) peroxidase reactions.

sensitive vectors are used sequentially to express the homologous recombination system of bacteriophage λ and the *S. cerevisiae* Flp recombinase.

The second axis consisted in studying the specific activity of FrvX, SgcX, and YpdE. The three proteins were produced as Strep-tag fusions to facilitate protein purification by affinity chromatography. Their activity was measured using various Xaa-*p*Na and Xaa-Ala substrates. The hydrolysis of dipeptides was monitored by coupling the reaction with the L-amino acid oxidase and peroxidase, as described elsewhere[419]. Basically, the L-amino acid oxidase oxidizes L-amino acids released during peptide hydrolysis (see **Figure 55**). During this reaction, hydrogen peroxide is generated and is used by the peroxidase to oxidize *o*-dianisidine, allowing to follow peptide hydrolysis spectrophotometrically. Crystallogenesis was also undertaken with FrvX, SgcX, and YpdE according to the method used for TmPep1050 (see **section B.2**, page 68). In parallel, the putative formation of heterocomplexes was studied by co-expressing *frv*X, *sgc*X, and *ypd*E. To engineer a suitable vector for co-expression, each gene was cloned separately into the pET15b vector in fusion with an epitope sequence. The latter allows to C-terminally tag the recombinant protein. Three epitopes were used: Strep-tag (for affinity chromatography), Myc-tag (derived from c-Myc), and HA-tag (derived from human influenza hemagglutinin). The polycistronic vectors were generated by the cut and paste approach using suitable restriction enzymes (see **Figure 56**)[488]. Briefly, a donor vector was restricted using *Xba*I and *Bam*HI to excise one of the three gene with its ribosome binding site (RBS). This fragment was then inserted into an acceptor vector between the *Nhe*I and *Bam*HI restriction sites. Ultimately, three vectors were engineered to co-produce FrvX, SgcX, and YpdE with different combinations of epitopes (see **Figure 56**). To detect heterocomplexes, the Strep-tagged proteins were purified using streptavidin and potential binding partners were immunodetected by Western blot using anti-Myc and anti-HA antibodies.

B.5.3 Results and discussion

The first axis of this preliminary study was to show if *frv*X, *sgc*X, and *ypd*E are essential genes for *E. coli* under stress conditions. The deletion of the three ORFs was achieved using the *E. coli* strain JW 3869 from the Keio collection, which is Δ*frv*X[486]. The *kan*R cassette integration and excision were verified by PCR to ensure the correct knock-out. Ultimately, the deletion of *frv*X, *sgc*X, and *ypd*E was confirmed by sequencing. The *E. coli* M42 aminopeptidases are not essential as Δ*frv*X Δ*sgc*X Δ*ypd*E strain did not show any growth defect on minimal M9 medium. The growth of Δ*frv*X Δ*sgc*X Δ*ypd*E strain was tested under several stresses: osmotic stress (addition of 0.7 M NaCl), heat stress (recovery from an incubation of 15 min. at 42°C), cold stress (recovery from a 2 hour-incubation at 15°C), oxidative stress (addition of 10 mM H_2O_2), and glucose-acetate carbon shift. None of these conditions induced a reduced fitness of Δ*frv*X Δ*sgc*X Δ*ypd*E strain compared to the wild-type strain MG1655, even though the M42 aminopeptidase coding genes are overexpressed in such conditions.

The second axis of this preliminary investigation was to study the specific activity of *E. coli* M42 aminopeptidases. FrvX, SgcX, and YpdE were recombinantly produced with a C-terminal Strep-tag to allow their purification by affinity chromatography. Their oligomeric states were determined by size-exclusion chromatography. While YpdE is mainly dodecameric, FrvX and SgcX form dodecamers and tetracosamers (see **Figure 57**). The occurrence of tetracosamers is not common in M42 aminopeptidases since it has been reported for PhTET1 only[326]. Furthermore, a third species of about 1.9 - 2.0 MDa was observed for FrvX and SgcX which could astonishingly correspond to a

Substrates	Specific activity (uA min^{-1} μM^{-1})		
	FrvX	SgcX	YpdE
Gly-Ala	n.a.	n.a.	0.022
Ala-Ala	0.011	n.a.	0.057
Val-Ala	0.005	n.a.	n.a.
Leu-Ala	0.247	0.014	0.049
Ile-Ala	0.086	0.004	0.028
Phe-Ala	0.145	0.008	0.015
Tyr-Ala	0.277	0.020	0.057
Trp-Ala	n.a.	n.a.	0.069
Ser-Ala	n.a.	n.a.	0.036
Thr-Ala	n.a.	n.a.	0.038
Met-Ala	n.a.	n.a.	0.021
Asp-Ala	n.a.	n.a.	n.a.
Glu-Ala	n.a.	n.a.	n.a.
His-Ala	n.a.	n.a.	n.a.
Lys-Ala	n.a.	n.a.	n.a.
Arg-Ala	n.a.	n.a.	n.a.

Table 3 – Specific activity of FrvX, SgcX, and YpdE on Xaa-Ala peptides. Activity assays were performed in 50 mM MOPS, 1 mM $CoCl_2$ pH 7.2 buffer with 5 mM Xaa-Ala and 1 µM of enzyme. Released amino acids were quantified by coupling amino acid oxidation with *o*-dianisine oxidation catalyzed by L-amino acid oxidase and peroxidase, respectively. Specific activities are expressed as absorbance unit per minute per µM of enzyme. uA, unit of absorbance at 440 nm. n.a., no activity.

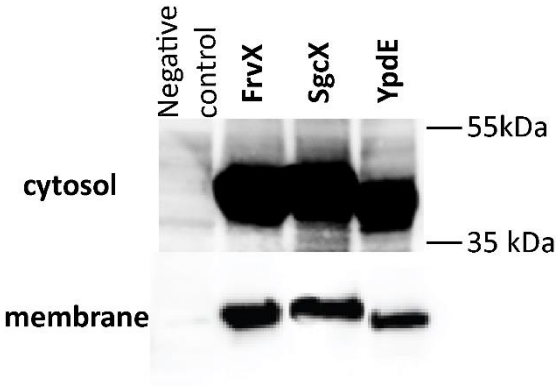


Figure 58 – Localization of FrvX, SgcX, and YpdE produced as Strep-tag fusions in *E. coli* by Western blot. Strep-tagged proteins were detected using HRP-streptavidin.

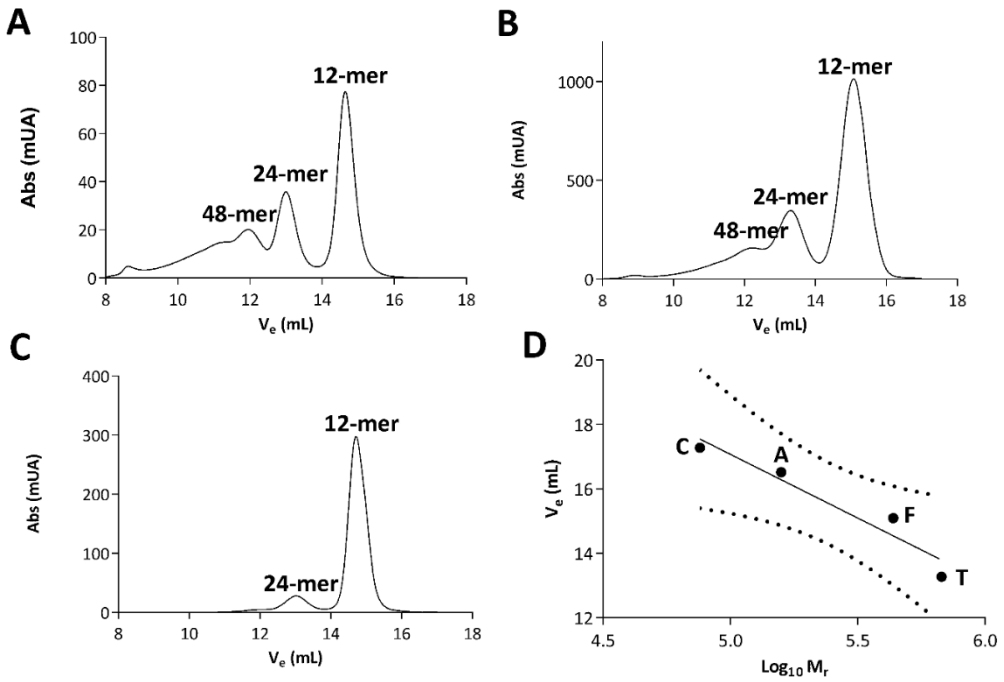


Figure 57 – Determination of oligomeric states of (A) FrvX, (B) SgcX, and (C) YpdE. Oligomeric states were determined by size exclusion chromatography using Superose 6 10/300 GL. **(D)** Superose 6 column calibration with conalbumin (C), aldolase (A), ferritin (F), and thyroglobulin (T) as standards. The correlation between V_e (ml) and the logarithm of the relative mass (M_r) is linear with R^2 of 0.93. The 95% confidence intervals of the linear regression are shown in dots. Abs, absorbance. mUA, milli unit of absorbance.

putative octotetracontamer. Of note, this third species was not eluted in the void volume of the size-exclusion chromatography column (being of 4 MDa). FrvX, SgcX, and YpdE did not show any significant activity on Xaa-*p*NA substrates but displayed some activity against Xaa-Ala dipeptides (see **Table 3**). FrvX appeared to fit in the S_1 pocket bulky aromatic or aliphatic amino acids such as tyrosine, phenylalanine, leucine, and isoleucine. While being less active than FrvX, SgcX displayed the same specificity towards aromatic and aliphatic residues. Of note, dodecamers and tetracosamers had almost the same hydrolysis rate, excluding any loss of activity linked to a high oligomerization degree. In comparison, YpdE had a broad substrate specificity but it appeared to exclude negatively and positively charged residues in the S_1 pocket. Our results are in contradiction with the previous characterization of YpdE. Indeed, Zheng *et al.* (2005) reported that YpdE recognized any type of residue, excepting proline, and it could hydrolyze peptide bonds like Asp-Ala and Lys-Ala[327]. Intriguingly, the three M42 aminopeptidases of *E. coli* did not show complementary specific activities which casts doubt on their auxiliary function linked to the proteasome or associated proteolytic complexes. According to the current hypothesis (see **section A.5.2.1**, page 25), M42 aminopeptidases work in concert to achieve a complete peptide hydrolysis thanks to their complementary or even synergetic specific activity as showed in *P. horikoshii*[268,319,324,325,329] and *G. stearothermophilus*[369,432].

The use of a C-terminal Strep-tag allowed to localize FrvX, SgcX, and YpdE using streptactin. The three M42 aminopeptidases were mainly localized in the cytosol while being absent from the extracellular medium and the periplasm. Surprisingly, about 30 % of Strep-tagged proteins were localized in the membrane fraction, albeit the absence of signal peptide for membrane translocation (see **Figure 58**). The membrane localization of FrvX, SgcX, and YpdE could be an artefact linked to overexpression, but such a localization has been reported for several M42 aminopeptidases[328,331,368,370,388,435]. Finally, FrvX, SgcX, and YpdE were co-produced in *E. coli* with different C-terminal epitopes for immunodetection. The co-production of SgcX-Strep, FrvX-HA, and YpdE-Myc is presented hereafter as a representative case of study. A similar result was obtained with YpdE-Strep while FrvX-Strep has not been tested yet. SgcX was purified using immobilized streptavidin and the oligomerization states were determined by size-exclusion chromatography. When co-produced with FrvX and YpdE, SgcX formed mainly dodecamers while its higher oligomeric states were almost absent (see **Figure 59.A**). The presence of FrvX and YpdE was then checked by immunodetection using antibodies raised against the HA and Myc tags. Both proteins were detected alongside SgcX-Strep, hinting the formation of M42 aminopeptidase heterocomplexes in *E. coli* (see **Figure 59.B**). Thermal shift assays were performed to dovetail the immunodetection of heterocomplexes. Taken separately, FrvX-Strep, SgcX-Strep, and YpdE-Strep homododecamers had a T_m of 72.3°C, 71.3°C, and 80.4°C, respectively. The heterocomplex, however, displayed a melting curve with two plateaux and, subsequently two T_m: 72.0°C and 89.0°C. The first T_m could correspond to thermal denaturation of SgcX homododecamers or FrvX-SgcX heterocomplexes. Surprisingly, the second T_m was much higher than those of homododecamers and it could result from a synergetic stabilization when the three proteins are associated in heterocomplexes. Our preliminary results strongly support that the existence of heterocomplexes could be widespread among prokaryotes, corroborating the previous studies on *P. horikoshii* and *G. stearothermophilus* heterocomplexes [268,372,423,426,432].

A further characterization of *E. coli* M42 aminopeptidases is required to better understand the assembly of complexes and their physiological function. The lack of phenotype in Δ*frv*X Δ*sgc*X

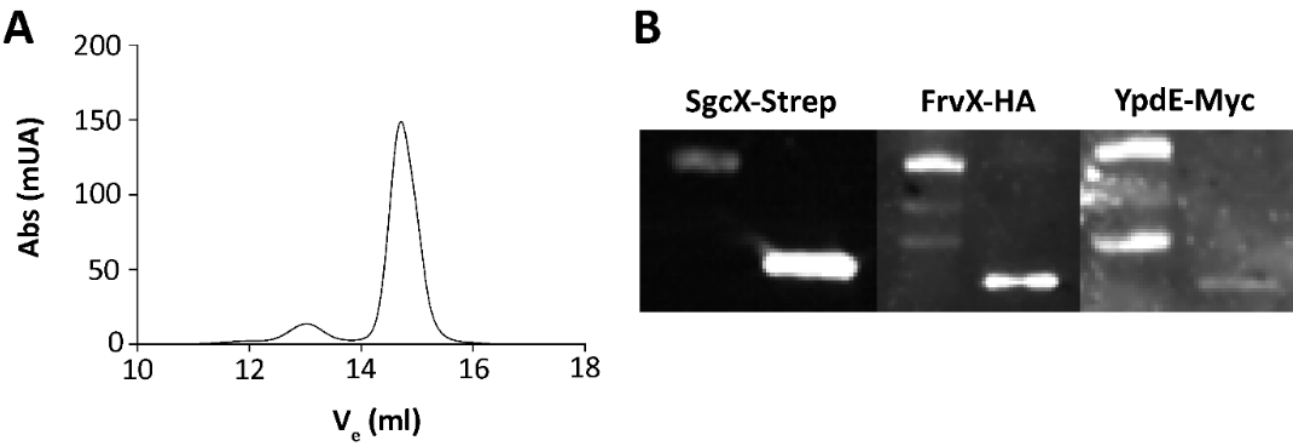


Figure 59 – Characterization of FrvX-SgcX-YpdE heterocomplexes. (A) Determination of the oligomeric states of SgcX-Strep co-produced with FrvX-HA and YpdE-Myc using Superose 6 10/300 GL. **(B)** Immunodetection of SgcX-Strep, FrvX-HA, and YpdE-Myc by Western blot using, respectively, HRP-streptavidin, anti-HA, and anti-Myc antibodies. Left lane: molecular weight ladder, right lane: FrvX-SgcX-YpdE heterocomplexes. Anti-HA and anti-Myc antibodies were controlled using pure SgcX-Strep (data not shown).

Δ*ypd*E strain probably results from redundant aminopeptidases as shown in *Lactococcus lactis*[489]. It would be interesting to study to the possible link between M42 aminopeptidases with carbohydrate metabolism in *E. coli*. They could potentially be involved in degrading peptidoglycans[452]. The formation of tetracosamers and octotetracontamers must also be characterized more thoroughly to settle whether these oligomers are physiologically relevant or not. The structural study of FrvX, SgcX, and YpdE could be particularly relevant to map the S_1 pocket accepting bulky aromatic residues. Several crystallization conditions were identified, and a complete data set at 3.5 Å was obtained for YpdE. Finally, the formation of heterocomplexes must be better understood, especially the subunit stoichiometry and how the activity could be modulated. Again, the physiological relevance must be ascertained since overproduction could bias protein interactions.

Conclusions

Due to their peculiar quaternary structure, TET peptidases are thought to be involved in peptide degradation downstream the proteasome[184,262]. Indeed, they adopt a singular tetrahedron-shaped structure with twelve subunits which are organized to compartmentalize the twelve active sites in an inner chamber. The access to the inner chamber is strictly controlled by the size and surface charge of entrance pores. The TET peptidases from *P. horikoshii* have been characterized extensively thanks to several pioneer works which have revealed several of their key-features[184,265,268,319,324–326,329,382,419,420,422,426,427]. The tetrahedral-shaped structure has been observed in the four *P. horikoshii* TET peptidases, each of them having a distinct substrate specificity[184,265,319,324–326,329]. In addition to the dodecameric form, tetracosamers may occur as shown for PhTET1[326], yet no physiological clue has been gathered about their existence. Conversely, the existence of dimers has been proven *in vivo* prompting a potential activity regulation through oligomerization[268,419]. Indeed, *in vitro* experiments have shown that (i) PhTET2 dimers display a reduced activity against long peptides compared to dodecamers[419] and (ii) the dimer-dodecamer transition is mediated by the metal ion cofactor availability[268,382,419]. The metal ion cofactor seems to play a crucial role since dodecamers tend to disassemble into dimers when metal ions are chelated. This phenomenon is not irreversible as metal ions promote the dodecamer assembly. During the dimer-dodecamer transition, different intermediates (tetramers, hexamers, octamers, and decamers) have been emphasized for PhTET2 and PhTET3[419,420,426]. Finally, the last important landmark about the *P. horikoshii* TET peptidases is the existence of heterocomplexes, notably dodecamers made of PhTET2 and PhTET3 subunits[268,419]. Such heterocomplexes have been shown to occur *in vivo* and to display a synergic activity with a broader substrate specificity than homododecamers of PhTET2 and PhTET3.

Despite the deep knowledge gathered on *P. horikoshii* TET peptidases, drawing general conclusions remains challenging and several issues remain to be addressed. Firstly, few structures of M42 aminopeptidases have been described for other phyla than Euryarchaeota[262,330,378]. Notably, PepA from *S. pneumoniae* remains the sole bacterial M42 aminopeptidase to be structurally characterized[330]. In addition, some M42 aminopeptidases have been reported to have a lower degree of oligomerization than dodecamers[328,334–336,370]. Secondly, M42 aminopeptidases have been linked to protein turnover, achieving the final degradation of peptides upstream the proteasome 20S or other related proteolytic complexes[184,262]. Yet, no direct proof of this cellular function has been reported *in vivo*. Several other enzymes may also fulfil the same role, such as TRI, APH, Hsp31, PepA, and PepN[50,182,183,205,206,236]. Thirdly, the dimer-dodecamer transition is not fully understood due to the lack of any dimer structure. Dodecamers are known to disassemble into dimers when metal ions are chelated but the subunit structure of a dimer is thought to be the same as that of a dodecamer[184,382,419]. Dimers may even further dissociate into monomers although the monomer structure has been described to be so flexible that monomers are prone to rapidly self-associate into dimers[365,420]. Moreover, the two metal ion binding sites probably have a distinct role in the dodecamer formation. The M2 site has been proposed to strictly control the oligomerization while the M1 site stabilizes the dodecamer structure[365]. Such a hypothesis implies that the metal ion binding affinity of the M2 site is greater than that of the M1 site, unlike VpAp1, and should consequently challenge the canonical catalytic mechanism[265,365]. Finally, M42 aminopeptidases seem to form heterocomplexes, as observed in *P. horikoshii*[268,426] and *G. stearothermophilus*[423,432]. Such complexes may act as peptide degradation machines[263]. Their occurrence and broad substrate specificity remain to be further characterized in other microorganisms.

The aim of this work was to investigate the dimer-dodecamer transition, their physiological functions, and the potential formation of heterocomplexes by studying several bacterial M42 aminopeptidases. TmPep1050 was chosen as a model to study its quaternary structures and the role of metal ions in oligomerization. Previously, it was described as a dimeric enzyme displaying a weak leucyl-aminopeptidase activity[323]. Its unusual oligomeric state could be linked to the histidine tag used to purify the recombinant enzyme. Indeed, such a purification tag may have artefactual effects on oligomerization, stabilization, and activity[490,491]. Henceforth, for this study, we decided to work with untagged TmPep1050. In that case, we showed that TmPep1050 forms stable dodecamers much more active than the tagged dimer, with a k_{cat} increasing from 0.25 to 138 s^{-1}. According to our structural data, $TmPep1050_{12\text{-mer}}$ adopts the genuine tetrahedral-shaped structure of TET peptidases. Two new structural features were highlighted: (i) the peculiar interaction interface between dimers and (ii) the implication of the PDZ-like domain in defining the catalytic chamber. Several salt bridges were shown to play a critical role in maintaining the quaternary structure, involving residues either located at the access pore or buried in the exit channel. These residues are not conserved among other structurally characterized M42 aminopeptidases, the interaction between dimers relying more on close contact at the tetrahedron vertices[184,325,326,330,419]. Regarding the PDZ-like domain, a loop has been described to be disordered since it was not modelled in all structures of M42 aminopeptidases published so far[184,265,325,326,330,365,378]. In the $TmPep1050_{12\text{-mer}}$ structure, this fragment of the PDZ-domain was clearly defined in electron density as an α helix (α4). It is localized next to the access pore in such a manner that three α4 helices form a kind of funnel, restricting the pore size to 13 Å. Subsequently, the pore diameter appears to be smaller than those reported for other M42 aminopeptidases, ranging from 16 Å to 20 Å[184,325,326,330]. Our observation, however, corroborates with the recently published structure of PhTET2 solved by combining NMR and cryo-EM data[427]. At the dimer level, the α4 helix of a subunit is also ideally placed to participate, especially Arg-115, in defining the substrate binding pocket of the other subunit.

Studying the structure of $TmPep1050_{12\text{-mer}}$ revealed the role of metal ion cofactors in governing enzyme stability and oligomerization. In presence of Co^{2+}, $TmPep1050_{12\text{-mer}}$ was extremely stable, having a half-life of about twenty days at 75°C. Meanwhile, when Co^{2+} was depleted, its thermostability was lost and, after two hours, the dodecamers dissociated into dimers. The dissociation, however, is reversible since dimers self-assembled into dodecamers upon adding Co^{2+} as shown by native MS. Our data strongly support the role of metal ion cofactors as a driving force of the oligomerization state transition as observed in PhTET2[184,419,420], PhTET3[382], PfTET3[365], and GsApl[424]. The structural modifications, underwent during dimer-dodecamer transition, were identified thanks to the structure of $TmPep1050_{2\text{-mer}}$. Two α helices (α8 and α10) were shown to be highly flexible in the dimer subunit since no electronic density that could match them was observed. In addition, two fragments involved in the interdimer interactions are misplaced in $TmPep1050_{2\text{-mer}}$, preventing the formation of crucial salt-bridges between dimers. The active site of $TmPep1050_{2\text{-mer}}$ also undergoes deep modifications, especially the displacement of an important loop outside the active site. Consequently, Glu-197 and Glu-198 are not correctly positioned to fulfil their function as, respectively, a catalytic base and a ligand of the M1 site. In the dodecamer structure, this active site loop is closely connected to the α8 and α10 helices via an entangled H-bond network. It is worth noting that the structure of $TmPep1050_{2\text{-mer}}$ is currently the sole structure of a M42 aminopeptidase dimer to be published.

The role of the metal ion cofactors was further investigated by substituting His-60 and His-307, two metal ion ligands of the M2 and M1 sites, respectively. The substitutions influenced differently the TmPep1050 oligomerization state. When His-60 was substituted with an alanine residue, both dodecamers and dimers were observed, suggesting that the M2 site contributes partly to the dodecamer formation. The substitution of His-307 with an alanine residue, however, had a more pronounced impact on oligomerization since only dimers were observed with this variant. Consequently, the dimer-dodecamer transition of TmPep1050 seems to rely mainly on the M1 site. Moreover, the structures of $TmPep1050_{H60A}$ dimer and $TmPep1050_{H307A}$ resemble greatly to that of $TmPep1050_{2\text{-}mer}$, showing the same structural modifications inherent to the dimeric state. Our observations differ from what Colombo *et al.* (2016) reported for PfTET3. Indeed, the role of the M1 and M2 sites in oligomerization is reversed in PfTET3. In that case of study, the M1 site contributes to stabilizing the dodecamer formation while the complete dissociation of dodecamer is achieved when no metal ion occupies both the M1 and M2 sites[365]. As discussed earlier, since the binding affinity of the M2 site is greater than that of the M1 site, the M2 is probably involved in the deprotonation of the water molecule prior to hydrolysis instead of the M1. TmPep1050, however, could follow the same catalytic mechanism as described for VpAp1 with the M1 involved in the water molecule deprotonation. In addition, a further dissociation of dimers into monomers has been reported for PhTET2 and PfTET3[365,420]. To prevent any metal ion binding, a variant having both His-60 and His-307 substituted with an alanine residue was characterized and shown to remain dimeric. Thus, for TmPep1050, the dimer formation does not seem to rely on metal ion.

Finally, the role of Asp-62 was also investigated in TmPep1050 since this residue could be important in catalysis. Indeed, according to the hypothesis drawn for VpAp1[304,307,363], Asp-62 interacts with His-60, forcing the imidazole ring to be deprotonated. Consequently, the Lewis acid strength of the M2 is decreased, favoring the transfer of the hydroxide ion to the M1. Due to its position in the second shell of the M2 site, Asp-62 could also influence the binding of the M2. Its substitution, however, resulted in the formation of inactive dimers only, suggesting that the M1 site was impaired too. Although, the Asp-62 variants were shown to still bind one cobalt ion per subunit. Native MS and crystallographic analyses confirmed that $TmPep1050_{D62A}$ is a genuine dimer. Because of the oligomeric state of $TmPep1050_{D62A}$, the M2 site was inferred to be functional. Regarding the M1 site, Asp-62 probably fulfils a structural role as it interacts tightly with the backbone nitrogen of Glu-197 and Glu-198. It would be interesting to know whether substituting the aspartate residue equivalent to Asp-62 would destabilize the active site of a monomeric enzyme like VpAp1. If the destabilization is also observed for such an enzyme, it will support its structural role thorough the whole MH clan. If not, it will indicate that the destabilization of the active site as observed for $TmPep1050_{D62A}$ could be an adaptation to oligomerization.

To investigate the physiological functions of M42 aminopeptidases, *E. coli* was chosen as a model due to the availability of genetic tools and transcriptomic data. Its genome possesses three ORFs coding M42 aminopeptidases (FrvX, SgcX, and YpdE), each being part of an operon linked to carbohydrate metabolism. Indeed, the three operons include a complete PTSII system, but the transported carbohydrates are unknown. We failed to link any phenotype to the deletion of *frv*X, *sgc*X, and *ypd*E in the tested conditions. The absence of phenotype is not surprising since *E. coli* possesses an arsenal of proteolytic complexes and peptidases (see **sections A3-A.4**). The biochemical characterization of the three *E. coli* M42 aminopeptidases gave interesting preliminary results. Firstly,

when recombinantly produced, FrvX, SgcX, and YpdE were localized mainly in the cytoplasm but also in the membrane fraction. Several studies have reported that M42 aminopeptidases are associated to the inner membrane despite the absence of any secretion signal peptide or transmembrane motif[328,331,368–370,388,435]. Yet, the localization to the membrane of *E. coli* M42 aminopeptidase must be demonstrated in more physiological conditions since overexpression could have biased it. Secondly, the oligomeric states of FrvX, SgcX, and YpdE were determined *in vitro*, each forming dodecamers and tetracosamers. FrvX and SgcX could also exist as octotetracontamers, although such an oligomer could have resulted from unspecific interactions between dodecamers or tetracosamers. Nonetheless, our data suggest that tetracosamers could be common among M42 aminopeptidases, as previously shown for PhTET1[326]. Thirdly, FrvX, SgcX, and YpdE share a similar substrate specificity with a clear preference to aliphatic and aromatic residues at the P_1 position. They are completely devoid of any glutamyl-/lysyl-aminopeptidase activity. The absence of complementary specificity between FrvX, SgcX, and YpdE differs from the TET peptidases of *P. horikoshii*. Indeed, the four enzymes have different but complementary substrate specificities[319,324,325,329]. In heterocomplexes, they could virtually degrade any peptide, which agrees with their putative function linked to the proteasome. What is true for *P. horikoshii* does not seem to apply to *E. coli*, prompting the requirement to characterize other combinations of M42 aminopeptidases from other prokaryotes. It is worth noting that the GsApI heterocomplex displays a broad substrate specificity like the PhTET2-PhTET3 heterocomplex[268,369,432]. Finally, the formation of heterododecamers made of FrvX, SgcX, and YpdE was observed, revealing an increased thermostability of the heterododecamers compared to homododecamers. Further characterization is required to uncover the stoichiometry of heterododecamers and the potential synergy of different subunits on activity.

Several perspectives can be foreseen in the wake of this work. The function of the metal ion cofactors must be further studied since they are involved in catalysis and oligomerization. Many M42 aminopeptidases have been reported to be fully active in presence of Co^{2+} [324,325,327,328,331,333,339,366,367,369–372], albeit the low abundance of Co^{2+} in *E. coli* and *P. furiosus* cells[381]. The metallome of *T. maritima* has not yet been determined but the Co^{2+} content is also expected to be low. Indeed, the Co^{2+} concentration does not exceed 20 µg/l in groundwater samples from Vulcano island (i.e. where *T. maritima* was isolated) while Zn^{2+} is quite abundant (more than 1 mg/l)[492]. Given the low abundance of intracellular Co^{2+} and the binding affinity in the micromolar range[365], M42 aminopeptidases are expected to bind other divalent metal ions. Yet, the binding affinity to these metal ions must still be determined as well as the subsequent impact of metal ion binding on oligomerization and substrate specificity. Of note, VpAp1 has been reported to accommodate various divalent metal ions in the M1 and M2 sites, modulating its substrate specificity and kinetic parameters of peptide hydrolysis (see **Table 2**, page 28)[346]. We showed that the M1 site of TmPep1050 strictly controls the dimer-dodecamer transition, but our observation contradicts the model established for PfTET3. Consequently, the binding affinity of the M1 and M2 sites must be determined for other M42 aminopeptidases to understand their respective function in the catalytic mechanism and oligomerization. Still regarding the role of metal ions in oligomerization, it would be interesting to study M18 aminopeptidases to determine whether their oligomerization state relies on metal ion availability.

The PDZ-like domain of M42 aminopeptidases could have other functions than dimerization. PDZ domains are often involved in protein-peptide, protein-protein, and protein-lipid interactions[410].

The putative binding site involved in those interactions, however, remains completely free in the M42 aminopeptidase PDZ-like domain. Such an interaction site could recognize unfolded substrates modulating the activity or even promoting oligomerization. Since M42 aminopeptidases are localized partly at the membrane, the PDZ-like domain could also fulfil a role to anchor the enzyme to either a binding partner or directly the lipid bilayer. A membrane localization is not so weird as proteolytic events occur at its vicinity. Indeed, FtsH is involved in the turnover of membrane proteins *in situ*[73,79,80] and DegP ensures as similar role in the periplasm[81]. Another putative binding partner could be PTS enzyme IIC (carbohydrate transporters) since the genes coding M42 aminopeptidases are found in operons harboring genes coding PTS enzymes II in *E. coli*. Although, the link to PTS enzymes is not true in all microorganisms as, for instance, the genes coding M42 aminopeptidases are organized in a single operon in *T. maritima*.

Studying M42 aminopeptidase heterocomplexes may also bring valuable information about their function. We showed that TmPep1048 and TmPep1049 are not dodecameric like TmPep1050 and form rather inactive dimers and tetramers, respectively. Thus, it would be interesting to verify the existence of an heterocomplex made of TmPep1048, TmPep1049, and TmPep1050. Indeed, the lack of activity of TmPep1048 and TmPep1049 could be due to their oligomerization state. As part of a heterocomplex, one may expect to discover their substrate specificity, like in GsApI. The latter is a heterododecamer consisting of two subunits, α and β[423]. While the α subunit can form homododecamers, the β subunit forms dimers only[423]. The β subunit, however, has been shown to be responsible for the glutamyl-aminopeptidase activity[432]. As mentioned earlier, another example of heterododecamer to be studied is the FrvX-SgcX-YpdE complex. In that case, the stoichiometry and substrate specificity remain to be determined. Both models, however, rely on recombinant expression that could bias the stoichiometry or promote artefactual interactions between the different subunits. GsAPI could be an interesting model to study M42 aminopeptidase heterocomplexes. Indeed, it can be purified in near-physiological conditions because of its abundance in *G. stearothermophilus*[369]. In addition, three heterocomplexes of different stoichiometries can be isolated: (i) ten α and two β subunits, (ii) eight α and four β subunits, and (iii) six α and six β subunits[423]. Studying their structures would be an important achievement to understand the cooperativity between the α and β subunits as well as their structural organization.

Appendices

Oligonucleotide name	Sequence	Description
ocej419	5’ – TTTAACTTTAAGAAGGAGATATACA TACCCATGAAGGAACTGATCAGAAAGCT G	Amplification of *Tm_1050* ORF with 30 pb homology to arabinose promoter for homologous recombination
ocej420	5’ – ATCCGCCAAAACAGCCAAGCTGGAG ACCGTTTACGCCCCCAGATACCTGATGAG	Amplification of *Tm_1050* ORF with 30 pb homology to pBAD terminator for homologous recombination
ocej434	5’ – TTTAACTTTAAGAAGGAGATATACAT ACCCATGAAGATGGAAACCGGAAAACTC T	Amplification of *Tm_1048* ORF with 30 pb homology to arabinose promoter for homologous recombination
ocej435	5’ – ATCCGCCAAAACAGCCAAGCTGGAG ACCGTTCACACCTCCAGTTCTACCGCG	Amplification of *Tm_1048* ORF with 30 pb homology to pBAD terminator for homologous recombination
ocej436	5’ – TTTAACTTTAAGAAGGAGATATACAT ACCCATGTATCTCAAAGAGCTTTCGATGA T	Amplification of *Tm_1049* ORF with 30 pb homology to arabinose promoter for homologous recombination
ocej437	5’ – ATCCGCCAAAACAGCCAAGCTGGAG ACCGTTCATGAGACCACCTCCACGATTTT	Amplification of *Tm_1049* ORF with 30 pb homology to pBAD terminator for homologous recombination
ocej462	5’ – GATACTGGACGCTCACATAaatGAG ATAGGTGTTGTCGTCAC	*Tm_1050* D62 codon mutation to Asn codon, forward primer complementary to ocej463
ocej463	5’ – GTGACGACAACACCTATCTCattTAT GTGAGCGTCCAGTATC	*Tm_1050* D62 codon mutation to Asn codon, forward primer complementary to ocej462
ocej464	5’ – GATACTGGACGCTCACATAgctGAG ATAGGTGTTGTCGTCAC	*Tm_1050* D62 codon mutation to Ala codon, forward primer complementary to ocej465
ocej465	5’ – GTGACGACAACACCTATCTCagcTAT GTGAGCGTCCAGTATC	*Tm_1050* D62 codon mutation to Ala codon, forward primer complementary to ocej464
ocej530	5’ – AATGGCTAGCTGGAGCCACCCGCAG TTCGAAAAAGGCGCCAAGATGGAAACCG GAAAACTCTTGA	Amplification of *Tm_1048* ORF with 30 pb homology to pASK-IBA5plus for homologous recombination
ocej531	5’ – CAATGTGCGCCATTTTTCACTTCACAG GTCAAGCTTTCACACCTCCAGTTCTACCGC GAC	Amplification of *Tm_1048* ORF with 30 pb homology to pASK-IBA5plus for homologous recombination
ocej532	5’ – AATGGCTAGCTGGAGCCACCCGCAGT TCGAAAAAGGCGCCTATCTCAAAGAGCTTT CGATGATG	Amplification of *Tm_1049* ORF with 30 pb homology to pASK-IBA5plus for homologous recombination
ocej533	5’ – CAATGTGCGCCATTTTTCACTTCACAG GTCAAGCTTTCATGAGACCACCTCCACGAT TTTTCC	Amplification of *Tm_1049* ORF with 30 pb homology to pASK-IBA5plus for homologous recombination
ocej687	5’ – GAAAAAGGTGATACTGGACGCTgccA TAGATGAGATAGGTGTTGTCG	*Tm_1050* H60 codon mutation, forward primer complementary to ocej688
ocej688	5’ – CGACAACACCTATCTCATCTATggcAG CGTCCAGTATCACCTTTTTC	*Tm_1050* H60 codon mutation, reverse primer complementary to ocej687
ocej689	5’ – GTCTATTCCCACACGATACGTTgccTC ACCCAGTGAGATGATCGC	*Tm_1050* H307 codon mutation, forward primer complementary to ocej690
ocej690	5’ – GCGATCATCTCACTGGGTGAggcAAC GTATCGTGTGGGAATAGAC	*Tm_1050* H307 codon mutation, reverse primer complementary to ocej689
ocej691	5’ – CTACGGTGTTTTCAGTGTTCAGcaaGA AGTGGGACTGGTCGGTG	*Tm_1050* E197 codon mutation, forward primer complementary to ocej692
ocej692	5’ – CACCGACCAGTCCCACTTCttgCTGAAC ACTGAAAACACCGTAG	*Tm_1050* E197 codon mutation, reverse primer complementary to ocej691
ocej693	5’ – GACACTCCGAAGGCCATCAAGgcaCAC GCAATGAGGCTCTCCG	*Tm_1050* R233 codon mutation, forward primer complementary to ocej694
ocej694	5’ – CGGAGAGCCTCATTGCGTGtgcCTTGA TGGCCTTCGGAGTGTC	*Tm_1050* R233 codon mutation, reverse primer complementary to ocej693
ocej695	5’ – CAGACACTCCGAAGGCCATCgcgAGAC ACGCAATGAGGCTCTCCG	*Tm_1050* K232 codon mutation, forward primer complementary to ocej696

C.1 Material and methods

C.1.1 Cloning and mutagenesis

TM_1050 open reading frames (ORF) was amplified from *Bsp*HI-digested TmCD00089984 plasmid (Joint Center for Structural Genomics) using *Pfu* DNA polymerase (ThermoFischer Scientific) and ocej419 and ocej420 primers (see **Table 4**). *TM_1048* and *TM_1049* ORFs were amplified from *T. maritima* strain MSB8 genomic DNA using ocej434/ocej435 and ocej436/ocej437 primer pairs, respectively (see **Table 4**). The PCR products were inserted into the pBAD vector (ThermoFisher Scientific) by homologous recombination in *E. coli* according to the SLiCE protocol[439]. Briefly, the insertion of the PCR product is allowed via two 30 bp sequences, homologous to the insertion site of the pBAD vector, flanking the gene of interest. The homologous recombination is mediated by using a cell extract of *E. coli* strain PPY, expressing the λ prophage Red recombination system. *TM_1048* and *TM_1049* ORFs were also cloned as Strep-tag fusion into pASK-IBA5plus vector (IBA Life Sciences) using ocej530/ocej531 and ocej532/ocej533 primer pairs, respectively. To co-express *frv*X, *sgc*X, and *ypd*E, the ORFs were amplified from *E. coli* strain MG1655 genomic DNA using Phusion Master Mix (ThermoFisher Scientific) and the primers listed in **Table 4**. The PCR products of *frv*X amplification were digested by *Nco*I and *Bam*HI while the others were digested by *Bsp*HI and *Bam*HI. The restricted fragments were ligated into *Nco*I/*Bam*HI-digested pET15b vector (Merck Millipore) using T4 DNA ligase. To create the polycistronic vectors, the donor vectors were digested by *Xba*I and *Bam*HI. After purification, the restricted fragments were ligated into the acceptors vectors digested by *Nhe*I and *Bam*HI according to a strategy described elsewhere[488].

Site-directed mutagenesis was carried out following the SPRINP protocol[440] except for the two vectors used for TmPep1050$_{K232A/R233A/R249A}$ and TmPep1050$_{K232E/R233E/R249E}$ production. In that case, two synthetic genes harboring the desired mutations (GeneArt – ThermoFisher Scientific) were introduced into the pBAD vector by homologous recombination. The primers used to generate TmPep1050 variants are listed in **Table 4**. All genetic constructs were verified by sequencing (Genetic Service Facility, University of Antwerp) and are listed in **Table 5**. The *E. coli* strain MC1061[493] was used for cloning. PCR fragments and plasmid DNA were purified using GeneJet PCR Purification kit and GeneJet Plasmid Miniprep kit (ThermoFisher Scientific), respectively. For expression, the *E. coli* strains MC1061 and BL21 (Merck Millipore) were used for, respectively, pBAD- and pET15b-derived vectors Cells were grown on LB broth in presence of 100 µg/ml ampicillin for positive selection.

C.1.2 Production and purification of recombinant enzymes

For the purification of TmPep1050 and its variants, cultures and protein extracts were prepared following previously published procedures[323] with two modifications: (i) cells from 1 l culture were disrupted in 40 ml of 50 mM MOPS, 1 mM $CoCl_2$ pH 7.2 and (ii) protein extracts were heated at 70°C for 10 minutes. The purification consisted in three chromatographic steps. The first step was an anion exchange chromatography on Source 15Q resin (GE Healthcare Life Sciences, Tricorn 10/150 column) in 50 mM MOPS, 1 mM $CoCl_2$ pH 7.2. Elution was performed with a gradient step from 0 to 0.5 M NaCl for five column volumes (CV). Fractions containing the protein of interest (2 CV) were pooled and $(NH_4)_2SO_4$ powder was added to a concentration of 1.5 M $(NH_4)_2SO_4$. The second chromatographic step was a hydrophobic interaction chromatography on Source 15Phe resin (GE Healthcare Life Sciences, XK 16/20 column) in 50 mM MOPS, 1.5 M $(NH_4)_2SO_4$, 1 mM $CoCl_2$ pH 7.2. Elution was performed with a gradient step from 1.5 to 0 M $(NH_4)_2SO_4$ for five CV. Fractions (1.5 CV)

ocej696	5’ – CGGAGAGCCTCATTGCGTGTCTcgcGATGGCCTTCGGAGTGTCTG	*Tm_1050* K232 codon mutation, reverse primer complementary to ocej695
ocej699	5’ – CCCGCTCTGAAAGTGAAAGACgcgGCATCGATCAGCAGCAAACG	*Tm_1050* R249 codon mutation, forward primer complementary to ocej700
ocej700	5’ – CGTTTGCTGCTGATCGATGCcgcGTCTTTCACTTTCAGAGCGGG	*Tm_1050* R249 codon mutation, reverse primer complementary to ocej699
ocej796	5’ – ATTATTCCATGGACATTGAGTTACTGCAACAG	Amplification of *frv*X ORF with *Nco*I restriction site for cloning into pET15b
ocej797	5’ – AATAATGGATCCACAACAGCTAGCTTATTTTTCAAACTGTGGATGGCTCCACGCGCTATCCACCTGACGGAACTGGC	Amplification of *frv*X ORF with Strep-tag fusion and *Nhe*I and *Bam*HI restriction sites for cloning into pET15b
ocej798	5’ – AATAATGGATCCACAACAGCTAGCTTATGCATAATCAGGCACATCATAAGGATAATCCACCTGACGGAACTGGC	Amplification of *frv*X ORF with HA-tag fusion and *Nhe*I and *Bam*HI restriction sites for cloning into pET15b
ocej799	5’ – ATTATTTCATGACATTTTCTGTGCAGGAAAC	Amplification of *sgc*X ORF with *Bsp*HI restriction site for cloning into pET15b
ocej800	5’ – AATAATGGATCCACAACAGCTAGCTTATTTTTCAAACTGTGGATGGCTCCACGCGCTTAATGGATGTGCCTCTTGTGTA	Amplification of *sgc*X ORF with Strep-tag fusion and *Nhe*I and *Bam*HI restriction sites for cloning into pET15b
ocej801	5’ – AATAATGGATCCACAACAGCTAGCTTATGCATAATCAGGCACATCATAAGGATATAATGGATGTGCCTCTTGTGTA	Amplification of *sgc*X ORF with HA-tag fusion and *Nhe*I and *Bam*HI restriction sites for cloning into pET15b
ocej802	5’ – TATTATGGATCCACAACAGCTAGCTTAAAGATCCTCCTCACTTATGAGTTTCTGCTCTAATGGATGTGCCTCTTGTGTA	Amplification of *sgc*X ORF with *myc*-tag fusion and *Nhe*I and *Bam*HI restriction sites for cloning into pET15b
ocej803	5’ – ATTATTTCATGAATTTATCGCTATTAAAAGCG	Amplification of *ypd*E ORF with *Bsp*HI restriction site for cloning into pET15b
ocej804	5’ – AATAATGGATCCACAACAGCTAGCTTATTTTTCAAACTGTGGATGGCTCCACGCGCTTCTGAAATCCGTCAGTTGAAC	Amplification of *ypd*E ORF with Strep-tag fusion and *Nhe*I and *Bam*HI restriction sites for cloning into pET15b
ocej805	5’ – TATTATGGATCCACAACAGCTAGCTTAAAGATCCTCCTCACTTATGAGTTTCTGCTCTCTGAAATCCGTCAGTTGAAC	Amplification of *ypd*E ORF with *myc*-tag fusion and *Nhe*I and *Bam*HI restriction sites for cloning into pET15b
ocej814	5’ – GTTTAACTTTAAGAAGGAGATATACATACCCATGAAAGAGCTGATTCGTAAACTGACC	Amplification of synthetic *Tm_1050* gene harbouring K232, R233 and R249 mutations with 30 pb homology to arabinose promoter for homologous recombination
ocej815	5’ – CATCCGCCAAAACAGCCAAGCTGGAGACCGTTTATGCACCCAGATAGCGAATCAGC	Amplification of synthetic *Tm_1050* gene harbouring K232, R233 and R249 mutations with 30 pb homology to pBAD terminator for homologous recombination
ocej820	5’- TGATACTGGACGCTcacATAGAAGAGATAGGTGTTGTCGTCACA	*Tm_1050* D62 codon mutation to Glu codon, forward primer complementary to ocej821
ocej821	5’- TGTGACGACAACACCTATCTCTTCTATgtgAGCGTCCAGTATCA	*Tm_1050* D62 codon mutation to Glu codon, forward primer complementary to ocej820

Table 4 – List of the nucleotides used in cloning and mutagenesis.

containing the protein of interest were pooled and concentrated using an Amicon Ultra-15 ultrafiltration unit with 30 kDa cut-off (Merck Millipore). The third step consisted of a gel filtration on Superdex 200 resin (GE Healthcare Life Sciences, XK 16/70 column) in 50 mM MOPS, 0.5 M $(NH_4)_2SO_4$, 1 mM $CoCl_2$ pH 7.2. The concentration of $(NH_4)_2SO_4$ had to be maintained at 0.5 M to avoid protein precipitation. Purified proteins were finally concentrated using an Amicon Ultra-15 ultrafiltration unit with 30 kDa cut-off. The presence and purity of the recombinant enzymes were checked throughout the purification procedure by SDS-PAGE and enzymatic assays with L-Leu-*p*NA as substrate (see **section C.1.3**). Proteins were quantified by measuring the absorbance at 280 nm and applying the extinction coefficient of 18,910 M^{-1} cm^{-1}. This purification protocol allowed the purification of about 10 mg of TmPep1050 from 1 l of culture. Molecular weights were determined by gel filtration on Superdex 200 and Superdex 75 resins (GE Healthcare Life Sciences, XK 16/70 column) and Superose 6 10/300 GL (GE Healthcare Life Sciences) using 50 mM MOPS 0.5 M $(NH_4)_2SO_4$ pH 7.2 as running buffer. The gel filtration columns were calibrated using both HMW gel filtration calibration kit (GE Healthcare Life Sciences) and gel filtration standard (Bio-Rad). The same procedure was applied to purify TmPep1048 and TmPep1049.

TmPep1048, TmPep1049, FrvX, SgcX, and YpdE were purified as Strep-tag fusions. Cells were grown in 1 l of LB broth supplemented with 100 µg/ml ampicillin at 37°C. When OD_{660nm} reached 0.5, cultures were cooled on ice for 5 minutes and transferred in an incubator at 18°C. After 15 minutes, cultures were induced by adding 1 mM IPTG for pET15b-derived vectors or 200 µg/l tetracyclin for pASK-IBA5plus-derived vectors. Induction was carried out at 18°C for 18 h. Cells were harvested by centrifuging 30 min. at 5,500 x g and disrupted in 40 ml of 50 mM MOPS, 150 mM NaCl pH 7.2 buffer containing 25 U benzonase (Merck Millipore) and one tab of Complete EDTA-free protease inhibitors (Roche). The cell extract was centrifuged at 20,000 x g for 30 min. and the supernatant was collected for purification. The first purification step was performed using a 15-ml column packed with Strep-tactin Superflow resin (IBA Life Sciences) conditioned in 50 mM MOPS, 150 mM NaCl pH 7.2. Strep-tagged proteins were eluted with 50 mM MOPS, 150 mM NaCl, 2.5 mM desthiobiotin pH 7.2. The second purification step consisted of a size exclusion chromatography using a Superose 6 10/300 GL (GE Healthcare Life Sciences) using 50 mM MOPS, 150 mM NaCl pH 7.2 as running buffer. Purified proteins were finally concentrated using an Amicon Ultra-15 ultrafiltration unit with 30 kDa cut-off. The presence and purity of the recombinant enzymes were checked throughout the purification procedure by SDS-PAGE. Proteins were quantified by measuring the absorbance at 280 nm and applying the extinction coefficient of 25,440 ; 22,920 ; and 20,970 M^{-1} cm^{-1} for FrvX, SgcX, and YpdE, respectively. The same procedure was used to purify heterocomplexes made of FrvX, SgcX, and YpdE.

C.1.3 Activity assays with aminoacyl-*p*NA

Aminoacyl-*p*-nitroanilide (*p*NA) substrates were purchased from Bachem AG. Aminopeptidase (EC 3.4.11.1) activity was assayed as previously described[323] except that enzymatic reactions were stopped by adding 1 ml of 20 % acetic acid to 1 ml of reaction mix (100 µl of substrate in 10 % methanol, 890 µl of 50 mM MOPS pH 7.2, and 10 µl of enzyme at a final concentration ranging from 10 nM to 2 µM). To determine the substrate specificity of $TmPep1050_{12\text{-mer}}$, aminopeptidase assays were carried out at 75°C with an enzyme concentration of 30 nM and the appropriate amino acid-*p*-NA substrate. For TmPep1048, TmPep1049, FrvX, SgcX, and YpdE, aminopeptidase assays were carried out with an enzyme concentration of 1 µM. All substrates were used at 2.5 mM, except L-Gly-*p*NA (1.25 mM), L-Phe-*p*NA (0.75 mM), L-His-*p*NA and L-Glu-*p*NA (0.5 mM). The effect of metal ions,

Plasmid name	Description
pCEC43	pBAD vector harbouring *Tm_1050* ORF for TmPep1050$_{12\text{-mer}}$ expression
pCEC48	pBAD vector harbouring *Tm_1048* ORF for TmPep1048 expression
pCEC49	pBAD vector harbouring *Tm_1049* ORF for TmPep1049 expression
pCEC75	pASK-IBA5plus harbouring TM_1048 ORF in fusion with Strep-tag
pCEC76	pASK-IBA5plus harbouring TM_1049 ORF in fusion with Strep-tag
pCEC84	pCEC43 with mutagenized D62 codon for TmPep1050$_{D62N}$ expression
pCEC85	pCEC43 with mutagenized D62 codon for TmPep1050$_{D62A}$ expression
pCEC153	pCEC43 with mutagenized H60 codon for TmPep1050$_{H60A}$ expression
pCEC154	pCEC43 with mutagenized H307 codon for TmPep1050$_{H307A}$ expression
pCEC155	pCEC43 with mutagenized E197 codon for TmPep1050$_{E197Q}$ expression
pCEC156	pCEC43 with mutagenized R233 codon for TmPep1050$_{R233A}$ expression
pCEC157	pCEC43 with mutagenized K232 codon for TmPep1050$_{K232A}$ expression
pCEC159	pCEC43 with mutagenized R249 codon for TmPep1050$_{R249A}$ expression
pCEC165	pCEC43 with mutagenized H60 and H307 codons for TmPep1050$_{H60A\ H307A}$ expression
pCEC184	pET15b vector harbouring *frv*X ORF in fusion with Strep-tag
pCEC185	pET15b vector harbouring *frv*X ORF in fusion with HA-tag
pCEC186	pET15b vector harbouring *sgc*X ORF in fusion with Strep-tag
pCEC187	pET15b vector harbouring *sgc*X ORF in fusion with HA-tag
pCEC188	pET15b vector harbouring *sgc*X ORF in fusion with *myc*-tag
pCEC189	pET15b vector harbouring *ypd*E ORF in fusion with Strep-tag
pCEC190	pET15b vector harbouring *ypd*E ORF in fusion with *myc*-tag
pCEC194	pBAD vector harbouring synthetic *Tm_1050* gene with K232E, R233E and R249E for TmPep1050$_{K232E/R233E/R249E}$ expression
pCEC195	pBAD vector harbouring synthetic *Tm_1050* gene with K232A, R233A and R249A for TmPep1050$_{K232A/R233A/R249A}$ expression
pCEC196	pCEC43 with mutagenized D62 codon for TmPep1050$_{D62E}$ expression
pCEC197	pET15b vector harbouring *frv*X-Strep, *sgc*X-HA, and *ypd*E-*myc* for co-expression
pCEC198	pET15b vector harbouring *frv*X-HA, *sgc*X-Strep, and *ypd*E-*myc* for co-expression
pCEC199	pET15b vector harbouring *frv*X-HA, *sgc*X-*myc*, and *ypd*E-Strep for co-expression

Table 5 – List of plasmids used in this thesis.

pH and temperature on the activity was studied as previously described[323]. To determine the kinetic parameters of TmPep1050$_{12\text{-mer}}$, assays were performed at 75°C with an enzyme concentration of 10 nM for L-Leu-*p*NA and L-Ile-*p*NA and an enzyme concentration of 50 nM for L-Met-*p*NA. Reaction mixes were supplemented with 250 µM $CoCl_2$. Kinetic parameters (k_{cat}, K_m and k_{cat}/K_m) were determined from the initial reaction rates, using Lineweaver-Burk linearization of the Michaelis-Menten equation. Activation energies were calculated from the slope of the trend line obtained by plotting the logarithm of the specific activity vs. the inverse of the temperature. For thermostability assays, TmPep1050$_{12\text{-mer}}$ was diluted to 1 µM in 50 mM MOPS, 1.5 M $(NH_4)_2SO_4$, 1 mM $CoCl_2$ pH 7.2, and it was incubated at either 75°C or 95°C. At various time-intervals, the activity was measured at the incubation temperature (75°C or 95°C) using by diluting the enzyme to 10 nM in 1 ml of reaction mix containing 2.5 mM L-Leu-*p*NA as substrate.

C.1.4 Activity assays with peptides

The specific activity of TmPep1048, TmPep1049, and TmPep1050 was assayed against various Xaa-Ala peptides (Bachem AG). Protein samples were diluted to a concentration ranging from 10 nM to 1 µM in 20 µl of 50 mM MOPS, 1 mM $CoCl_2$ pH 7.2 buffer containing 5 mM Xaa-Ala. Assays were performed at 75°C for one hour. Hydrolysis products were analyzed by Thin Layer Chromatography (TLC). 3 x 2 µl of each sample was spotted onto a TLC silica gel 60 plate (Merck Millipore). TLC was conducted in two solvents, depending on the chemical properties of released amino acids. The amino acids with a hydrophobic or polar uncharged side chain were separated with 40:10:2 isopropanol:water:formic acid as solvent. The amino acids with a charged side chain were separated with 50:30:10:10 *tert*-butanol:acetone:ammonia:water as solvent. Amino acids were revealed by spraying the plates with 0.35% ninhydrin solution in ethanol and heating at 80°C.

The specific activity of FrvX, SgcX, and YpdE was assayed against various Xaa-Ala peptides according to a previously described method[419]. Protein samples were diluted to 1 µM in 40 µl of 50 mM MOPS, 1 mM $CoCl_2$ pH 7.2 buffer containing 5 mM Xaa-Ala, 0.1 mg/ml *o*-dianisidine, 20 U/ml horseradish peroxidase (Sigma-Aldrich), and 3.5 U/ml L-amino acid oxidase (Sigma-Aldrich). Assays were performed at 37°C and the absorbance was recorded at 440 nm using a Cary-60 spectrophotometer (Agilent).

The hydrolysis of Met-Lys-bradykinin peptide (Bachem AG) was assayed for TmPep1050$_{12\text{-mer}}$. The protein sample was diluted to 1 µM in 100 µl of 50 mM MOPS, 1 mM $CoCl_2$ pH 7.2 buffer containing Met-Lys-bradykinin at a concentration ranging from 0 to 5 mM. Assays were performed at 75°C for 10 minutes and samples were taken at different time-intervals. Released amino acids were derivatized using AccQ Tag Chemistry kit (Waters) and analyzed by HPLC on a Waters e2695 HPLC system. Amino acids were separated on AccQ Tag Amino Acid Analysis Column (Waters) according to the manufacturer's conditions. Derivatized amino acids were detected by fluorescence with excitation wavelength set on 250 nm and emission wavelength set on 395 nm.

C.1.5 Cobalt binding assays

Protein samples were diluted to 100 µM in either 2.1 M malic acid pH 7.0 or 50 mM MOPS, 0.5 M $(NH_4)_2SO_4$, 10 mM 1,10-phenanthroline pH 7.2. Sample were dialyzed four times against 100 volumes of 50 mM MOPS, 0.5 M $(NH_4)_2SO_4$ pH 7.2 using SnakeSkin dialysis tubing with a 3.5 kDa cut-off (ThermoFisher Scientific). To monitor metal ion removal, the specific activity was measured with L-Leu-*p*NA as substrate. Samples were concentrated back to their original concentration using Amicon

Ultra-0.5 ultrafiltration unit with a 30 kDa cut-off. Cobalt binding assays were performed by incubating 100 µL of 20 µM cobalt-depleted enzyme with $CoCl_2$ at a concentration ranging from 0 to 2560 µM in 50 mM MOPS, 0.5 M $(NH_4)_2SO_4$ pH 7.2 for 24 H at 75°C for wild-type TmPep1050 or at 50°C for the Asp-62 variants. For TmPep1050$_{12\text{-mer}}$, the specific activity was measured using 10 nM of enzyme and 2.5 mM L-Leu-*p*NA as substrate without added cobalt in the reaction mix. Other metal ions were tested following this procedure to identify the metal cofactor of TmPep1050$_{12\text{-mer}}$.

In parallel, cobalt concentration was determined using Amplex UltraRed (ThermoFisher Scientific), a fluorescent probe that binds specifically cobalt at high pH. The protocol established by Tsai & Lin (2013) was adapted to fit a 384-well black microplate (Corning)[494]. Fluorescence was measured on a SpectraMax 5 (Molecular Device) with excitation wavelength set on 495 nm and emission wavelength set on 570 nm. Prior to cobalt ion quantification, samples were diluted twice in 50 mM MOPS, 0.5 M $(NH_4)_2SO_4$ pH 7.2 working buffer. In addition to total cobalt concentration, unbound cobalt was quantified after filtering samples using an Amicon Ultra-0.5 ultrafiltration unit with a 30 kDa cut-off. The affinity constant K_d was determined from Scatchard plot data representation.

To study the reassociation of dimers into dodecamers, 100 µl of 50 µM TmPep1050$_{2\text{-mer}}$ was incubated with Co^{2+} at a concentration ranging from 0 to 5 mM for 30 min. at 75°C. The oligomers were detected and quantified by size exclusion chromatography using Superdex200 resin (GE Healthcare Life Sciences, XK16/20 column). 50 mM MOPS 0.5 M $(NH_4)_2SO_4$ pH 7.2 buffer was used for this assay.

C.1.6 Thermal shift assay

SyproOrange (ThermoFisher Scientific) was diluted 1:125 in 50 mM MOPS, 0.5 M $(NH_4)_2SO_4$ pH 7.2. The fluorescence probe was mixed with protein samples conditioned in 50 mM MOPS, 0.5 M $(NH_4)_2SO_4$ pH 7.2 and Co^{2+} at a concentration ranging from 0 to 2560 µM. The working SyproOrange dilution was 1:1000 and the protein concentration was 20 µM for a reaction volume of 20 µl. Thermal shift assays were performed in 96-well plate on a StepOnePlus Real-Time PCR System (ThermoFisher Scientific). Fluorescence curves were treated with StepOne Software.

C.1.7 Native mass spectrometry

Samples for native mass spectrometry were conditioned in 20 mM ammonium acetate (AmAc) pH 7.2 using Zeba 7 kDaA desalting columns (ThermoFisher Scientific). If further desalting was needed, Bio-spin P-6 gel columns (Bio-Rad) were used. The protein samples were diluted in 100 mM AmAc pH 7.2 to working concentration of 5 µM unless stated otherwise. In house prepared borosilicate gold coated needles filed with 2-3 µl sample were used to introduce the protein into the gas phase using nano electrospray ionization. The spectra were recorded in positive ion mode on a traveling-wave ion mobility Q-TOF instrument (Synapt G2 HDMS, Waters). Different settings were tuned to optimize sample measurement. The most important settings applied during the measurements, unless stated differently at the figure legend, were 20 Volts sampling cone, 10 Volts trap collision energy, and pressures 8.30 e^{-3} mbar and 5.31 e^{-2} mbar of source and trap respectively.

C.1.8 Crystallization

TmPep1050$_{12\text{-mer}}$, TmPep1050$_{2\text{-mer}}$, TmPep1050$_{H60A}$, TmPep1050$_{H307A}$, TmPep1050$_{H60A\ H307A}$, and TmPep1050$_{D62A}$ were crystallized using the hanging-drop vapor diffusion method, at 292 K in EasyXtal Tool plates (Qiagen). Drops contained 2 µl of recombinant enzyme mixed with 2 µl of well buffer.

	Crystallization buffer	Protein concentration	Cryoprotectant soaking
Apo-TmPep1050$_{12\text{-mer}}$	2.1 M malic acid pH 6.75	390 µM	None
TmPep1050$_{12\text{-mer}}$	0.18 M tri-ammonium citrate, 40 % PEG3350 pH 7.5	1 mM	None
TmPep1050$_{2\text{-mer}}$	0.1 M sodium citrate, 10% PEG3350 pH 6.0	230 µM	0.1 M sodium citrate, 5% PEG3350, 20% glycerol pH 5.2
TmPep1050$_{H60A}$	0.1 M sodium citrate, 5 % PEG3350 pH 4.5	230 µM	0.1 M sodium citrate, 5% PEG3350, 20% glycerol pH 5.2
TmPep1050$_{H307A}$	0.1 M sodium citrate, 10 % PEG3350 pH 4.5	230 µM	0.1 M sodium citrate, 5% PEG3350, 20% glycerol pH 5.2
TmPep1050$_{H60A\ H307A}$	0.1 M sodium citrate, 20 % PEG3350 pH 5.2	320 µM	None
TmPep1050$_{D62A}$	0.1 M sodium citrate, 5 % PEG3350 pH 5.2	250 µM	None

Table 6 – Crystallization conditions.

Crystallization conditions are described in **Table 6**. One cycle of seeding was necessary to get monocrystals of each species.

C.1.9 Structure determination and analysis

For apo-TmPep1050$_{12\text{-mer}}$ and TmPep1050$_{D62A}$, diffraction data were collected on the FIP-BM30a beamline at ESRF (Grenoble, France)[495,496]. For TmPep1050$_{12\text{-mer}}$, TmPep1050$_{2\text{-mer}}$, TmPep1050$_{H60A}$, TmPep1050$_{H307A}$, and TmPep1050$_{H60A\ H307A}$, diffraction data were collected on Proxima 2 beamline at SOLEIL (Saint-Aubin, France). The data collection and refinement statistics are presented in **Table 7**. Diffraction data were processed using the XDS program package[442,443]. Molecular replacement and model building were performed using PHENIX software package[445]. The initial solution of apo-TmPep1050$_{12\text{-mer}}$ was determined by molecular replacement with MR-Rosetta using the coordinates of YpdE of *Shigella flexneri* (PDB ID: 1YLO) as the search model (TFZ-value = 23.9, log-likelihood gain = 5,150). The structure of TmPep1050$_{12\text{-mer}}$ was solved by MR-SAD using ShelX and Phaser[444,497]. For TmPep1050$_{2\text{-mer}}$, TmPep1050$_{H60A}$, TmPep1050$_{H307A}$, TmPep1050$_{H60A\ H307A}$, and TmPep1050$_{D62A}$, molecular replacement was achieved with Phaser-MR using the coordinates of apo-TmPep1050$_{12\text{-mer}}$ as search model. The models were built using phenix.autobuild. Iterative manual building was done in COOT[446]. Multiple rounds of refinement were performed using phenix.refine. Model stereochemical quality was assessed using MolProbity[498,499]. Protein structures were analyzed with PDBe Pisa[421], Arpeggio[500], Rosetta pKa protocol[501,502], APBS[225], and PyMOL Molecular Graphics System version 2.2 (Schrödinger, LLC).

C.1.10 Gene deletion in *E. coli*

E. coli strains JW3869, JW5776, and JW2381 from the Keio collection[486] were used for the deletion of *frv*X, *sgc*X, and *ypd*E. The deletion was performed using the strategy and the primers described elsewhere[486,487]. The deletion cassettes were amplified using Phusion Master Mix (ThermoFisher Scientific) and pKIKO*lacZ*Km as template[503]. Strains were transformed with pREDIA vector[504] to allow λ Red recombination expression necessary for gene deletion. Strains with pREDIA were grown in SOB medium supplemented with 50 µg/ml ampicillin at 30°C. At OD_{600} 0.2, 0.2% arabinose was added, and cultures were transferred at 37°C. Cells were harvested when OD_{600nm} reached 0.6 and treated for electroporation as previously described[505]. Electrocompetent cells were electroporated with 200-400 ng of deletion cassettes using a BTX ECM830 electroporator. Cells were resuspended in 1 ml of SOB and incubated for three hours at 37°C before being plated onto LB agar supplemented with 50 µg/ml kanamycin. Deletions were verified by PCR as described previously[486]. To curate pREDIA vector, cells were streaked on LB agar and grown at 42°C. Isolated colonies were then tested on LB agar supplemented with 50 µg/ml ampicillin to check the pREDIA loss. The integrated kanamycin cassette was excised using pCP20 vector allowing the production of Flp recombinase[506]. The cassette loss was verified on LB agar supplemented with 50 µg/ml kanamycin. pCP20 vector was curated as described for pREDIA vector.

C.1.11 FrvX-SgcX-YpdE heterocomplexes detection

The polycistronic vectors pCEC197, pCEC198, and pCEC199 (see **Table 5**) were used for co-expression of *frv*X, *sgc*X, and *ypd*E. Culture conditions and the purification of Strep-tag fusions are described above (see **section C.1.2**). After separation by SDS–PAGE, proteins were electroblotted onto a Hybond nitrocellulose membrane (GE Healthcare) with XcellII module (ThermoFisher) and with

	Apo-TmPep $1050_{12\text{-mer}}$	TmPep $1050_{12\text{-mer}}$	TmPep $1050_{2\text{-mer}}$	TmPep 1050_{H60A}	TmPep 1050_{H307A}	TmPep $1050_{H60A\ H307A}$	Tmpep 1050_{D62A}
Data collection							
Temperature (K)	100	100	100	100	100	100	100
Radiation source	ESRF BM30a	Soleil Proxima 2	Soleil Proxima 2	Soleil Proxima 2	Soleil Proxima 2	Soleil Proxima 2	ESRF BM30a
Wavelength (Å)	0.9797	0.9801	0.9800	0.9800	0.9800	0.9801	0.9797
Detector	ADSC QUANTUM 315r	DECTRIS EIGER X 9M	DECTRIS EIGER X 9M	DECTRIS EIGER X 9M	DECTRIS EIGER X 9M	DECTRIS EIGER X 9M	ADSC QUANTUM 315r
Rotation range (°)	0.37	0.10	0.20	0.10	0.10	0.10	0.50
Exposure time (s)	20	0.025	0.025	0.025	0.025	0.025	10
Space group	P1	H3	C 2 2 2_1	C 2 2 2_1	P 1 2_1 1	P 1 2_1 1	C 2 2 2_1
Unit cell parameters							
α, β, γ (°)	114.46, 91.71, 105.69	90.00, 90.00, 120.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 110.51, 90.00	90.00, 110.69, 90.00	90.00, 90.00, 90.00
a, b, c (Å)	114.26, 114.57, 114.04	131.15, 131.15, 285.61	42.55, 114.71, 267.69	42.63, 114.22, 267.96	42.79, 138.65, 61.25	43.24, 137.79, 61.11	42.18, 113.96, 267.23
Resolution (Å)	44.05-2.20 (2.28-2.20)	47.60-1.70 (1.74-1.70)	48.25-2.00 (2.05-2.00)	43.46-1.84 (1.89-1.84)	40.11-1.75 (1.79-1.75)	43.99-2.37 (2.52-2.37)	48.00-1.50 (1.58-150)
Unique reflections	237,152	201,316	45,086	57,222	67,094	26,902	103,939
R_{merge} (%)	8.9 (39.2)	8.2 (67.0)	9.5 (69.1)	8.7 (60.6)	5.6 (56.3)	14.0	9.0
Redundancy	3.2 (2.2)	10.3 (10.4)	13.2 (13.0)	12.0 (9.8)	6.7 (6.7)	6.8	5.8
<I/σ>	8.56 (2.22)	15.81 (2.81)	16.81 (3.46)	16.52 (2.63)	18.04 (2.38)	8.64 (2.12)	14.80 (3.34)
Completeness	93.5 (84.8)	99.8 (97.6)	99.9 (99.2)	99.4 (91.7)	99.5 (97.2)	99.6 (97.9)	99.7 (89.4)
$CC_{1/2}$	99.4 (81.1)	99.9 (87.1)	99.8 (90.0)	99.9 (90.0)	99.9 (88.1)	99.2 (84.1)	99.8 (87.6)
Refinement							
Resolution (Å)	44.05-2.20	47.60-1.70	48.25-2.00	43.46-1.84	40.11-1.75	43.99-2.37	48.00-1.50
Reflections	237,090	201,316	45,076	57,213	67,094	26,900	103,177
R_{free} set test count	11,854	9,715	2,254	2,862	3,363	1,345	5,158
R_{work} / R_{free}	0.212/ 0.247	0.143/ 0.164	0.166/ 0.203	0.167/ 0.195	0.165/ 0.185	0.206/ 0.234	0.183/ 0.215
Protein molecules per ASU	12	4	2	2	2	2	2
V_M (Å^3/Da)	2.98	3.27	2.26	2.26	2.36	2.37	2.27
Solvent content (%)	58.7	62.4	45.6	45.6	47.9	49.0	45.8
Protein/solvent atoms	29,969/ 2,223	10,759/ 1,474	4,610/ 362	4,730/ 500	4,621/ 335	4,559/ 96	4,950/ 820
r.m.s.d bond length (Å)	0.009	0.019	0.004	0.004	0.012	0.310	0.360
r.m.s.d bond angles (°)	1.254	1.680	0.647	0.672	1.263	0.510	0.550
Average B-factors (Å^2)	37.0	24.0	37.8	33.2	28.9	57.0	20.0
Favored/disallowed Ramachandran φ/ϕ (%)	95.71/0.00	95.35/0.34	94.86/0.17	94.95/0.00	93.64/0.52	95.02/0.17	97.71/0.00
Twin law	none	none	none	none	h, -k, -h-l	h, -k, -h-l	none
PDB code	4P6Y	6NW5	5NE6	5NE7	5NE8	5NE9	5L6Z

Table 7 – Data collection and refinement statistics. Values in parentheses are for the highest-resolution shell. ESRF: European Synchrotron Radiation Facility, r.m.s.d.: root mean square deviation, ASU: asymmetric unit.

NuPage Transfer Buffer (ThermoFisher). Tagged proteins were immunodetected using Streptavidin-HRP (ThermoFisher), anti-HA tag antibody (Abcam), and anti-Myc tag antibody (Abcam).

References

1. Ciechanover A. Intracellular protein degradation: from a vague idea thru the lysosome and the ubiquitin–proteasome system and onto human diseases and drug targeting. *Cell Death Differ*. 2005;12(9):1178-1190. doi:10.1038/sj.cdd.4401692
2. López-Otín C, Bond JS. Proteases: Multifunctional Enzymes in Life and Disease. *J Biol Chem*. 2008;283(45):30433-30437. doi:10.1074/jbc.R800035200
3. Bond JS. Proteases: History, discovery, and roles in health and disease. *J Biol Chem*. 2019;294(5):1643-1651. doi:10.1074/jbc.TM118.004156
4. Konovalova A, Søgaard-Andersen L, Kroos L. Regulated proteolysis in bacterial development. *FEMS Microbiology Reviews*. 2014;38(3):493-522. doi:10.1111/1574-6976.12050
5. Livneh I, Cohen-Kaplan V, Cohen-Rosenzweig C, Avni N, Ciechanover A. The life cycle of the 26S proteasome: from birth, through regulation and function, and onto its death. *Cell Res*. 2016;26(8):869-885. doi:10.1038/cr.2016.86
6. Ferrington DA, Gregerson DS. Immunoproteasomes. In: *Progress in Molecular Biology and Translational Science*. Vol 109. Elsevier; 2012:75-112. doi:10.1016/B978-0-12-397863-9.00003-1
7. Janssen BD, Hayes CS. The tmRNA ribosome-rescue system. In: *Advances in Protein Chemistry and Structural Biology*. Vol 86. Elsevier; 2012:151-191. doi:10.1016/B978-0-12-386497-0.00005-0
8. Jang HH. Regulation of Protein Degradation by Proteasomes in Cancer. *J Cancer Prev*. 2018;23(4):153-161. doi:10.15430/JCP.2018.23.4.153
9. Klein T, Eckhard U, Dufour A, Solis N, Overall CM. Proteolytic Cleavage—Mechanisms, Function, and "Omic" Approaches for a Near-Ubiquitous Posttranslational Modification. *Chem Rev*. 2018;118(3):1137-1168. doi:10.1021/acs.chemrev.7b00120
10. Collins GA, Goldberg AL. The Logic of the 26S Proteasome. *Cell*. 2017;169(5):792-806. doi:10.1016/j.cell.2017.04.023
11. Vigneron N, Abi Habib J, Van den Eynde BJ. Learning from the Proteasome How To Fine-Tune Cancer Immunotherapy. *Trends in Cancer*. 2017;3(10):726-741. doi:10.1016/j.trecan.2017.07.007
12. Chauhan D, Singh A, Brahmandam M, et al. Combination of proteasome inhibitors bortezomib and NPI-0052 trigger in vivo synergistic cytotoxicity in multiple myeloma. 2008;111(3):11.
13. Röhrborn D. DPP4 in diabetes. *Front Immunol*. 2015;6. doi:10.3389/fimmu.2015.00386
14. Lacombe J, Karsenty G, Ferron M. Regulation of lysosome biogenesis and functions in osteoclasts. *Cell Cycle*. 2013;12(17):2744-2752. doi:10.4161/cc.25825
15. Thibaudeau TA, Anderson RT, Smith DM. A common mechanism of proteasome impairment by neurodegenerative disease-associated oligomers. *Nat Commun*. 2018;9(1):1097. doi:10.1038/s41467-018-03509-0
16. Antalis TM, Shea-Donohue T, Vogel SN, Sears C, Fasano A. Mechanisms of Disease: protease functions in intestinal mucosal pathobiology. *Nat Rev Gastroenterol Hepatol*. 2007;4(7):393-402. doi:10.1038/ncpgasthep0846
17. Peter B, Waddington CL, Oláhová M, et al. Defective mitochondrial protease LonP1 can cause classical mitochondrial disease. *Human Molecular Genetics*. 2018;27(10):1743-1753. doi:10.1093/hmg/ddy080
18. Hua Y, Nair S. Proteases in cardiometabolic diseases: Pathophysiology, molecular mechanisms and clinical applications. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*. 2015;1852(2):195-208. doi:10.1016/j.bbadis.2014.04.032
19. Manasanch EE, Orlowski RZ. Proteasome inhibitors in cancer therapy. *Nat Rev Clin Oncol*. 2017;14(7):417-433. doi:10.1038/nrclinonc.2016.206
20. Ciechanover A, Brundin P. The Ubiquitin Proteasome System in Neurodegenerative Diseases: Sometimes the Chicken, Sometimes the Egg. *Neuron*. 2003;40(2):427-446.
21. Mishto M, Bellavista E, Santoro A, et al. Immunoproteasome and LMP2 polymorphism in aged and Alzheimer's disease brains. *Neurobiology of Aging*. 2006;27(1):54-66. doi:10.1016/j.neurobiolaging.2004.12.004
22. Oddo S. The ubiquitin-proteasome system in Alzheimer's disease. *J Cellular Mol Med*. 2008;12(2):363-373. doi:10.1111/j.1582-4934.2008.00276.x
23. Drag M, Salvesen GS. Emerging principles in protease-based drug discovery. *Nat Rev Drug Discov*. 2010;9(9):690-701. doi:10.1038/nrd3053
24. Turk B. Targeting proteases: successes, failures and future prospects. *Nat Rev Drug Discov*. 2006;5(9):785-799. doi:10.1038/nrd2092
25. Cudic M, Fields G. Extracellular Proteases as Targets for Drug Development. *CPPS*. 2009;10(4):297-307. doi:10.2174/138920309788922207
26. Adams J. The proteasome: a suitable antineoplastic target. *Nat Rev Cancer*. 2004;4(5):349-360. doi:10.1038/nrc1361
27. Haselbach D, Schrader J, Lambrecht F, Henneberg F, Chari A, Stark H. Long-range allosteric regulation of the human 26S proteasome by 20S proteasome-targeting cancer drugs. *Nat Commun*. 2017;8(1):15578. doi:10.1038/ncomms15578

28. Deu E, Verdoes M, Bogyo M. New approaches for dissecting protease functions to improve probe development and drug discovery. *Nat Struct Mol Biol*. 2012;19(1):9-16. doi:10.1038/nsmb.2203
29. Cushman DW, Ondetti MA. Design of angiotensin converting enzyme inhibitors. *Nat Med*. 1999;5(10):1110-1112. doi:10.1038/13423
30. Natesh R, Schwager SLU, Evans HR, Sturrock ED, Acharya KR. Structural Details on the Binding of Antihypertensive Drugs Captopril and Enalaprilat to Human Testicular Angiotensin I-Converting Enzyme. *Biochemistry*. 2004;43(27):8718-8724. doi:10.1021/bi049480n
31. Mega JL, Simon T. Pharmacology of antithrombotic drugs: an assessment of oral antiplatelet and anticoagulant treatments. *The Lancet*. 2015;386(9990):281-291. doi:10.1016/S0140-6736(15)60243-4
32. Culp E, Wright GD. Bacterial proteases, untapped antimicrobial drug targets. *J Antibiot*. 2017;70(4):366-377. doi:10.1038/ja.2016.138
33. Bibo-Verdugo B, Jiang Z, Caffrey CR, O'Donoghue AJ. Targeting proteasomes in infectious organisms to combat disease. *FEBS J*. 2017;284(10):1503-1517. doi:10.1111/febs.14029
34. Butler SM, Festa RA, Pearce MJ, Darwin KH. Self-compartmentalized bacterial proteases and pathogenesis. *Molecular Microbiology*. 2006;60(3):553-562. doi:10.1111/j.1365-2958.2006.05128.x
35. Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*. 2020;181(2):281-292.e6. doi:10.1016/j.cell.2020.02.058
36. Huang L, Chen C. Understanding HIV-1 protease autoprocessing for novel therapeutic development. *Future Medicinal Chemistry*. 2013;5(11):1215-1229. doi:10.4155/fmc.13.89
37. Frees D, Brøndsted L, Ingmer H. Bacterial Proteases and Virulence. In: *Regulated Proteolysis in Microorganisms*. Vol 66. Subcellular Biochemistry. Springer Netherlands; 2013:161-192. doi:10.1007/978-94-007-5940-4_7
38. Wagner S, Grin I, Malmsheimer S, Singh N, Torres-Vargas CE, Westerhausen S. Bacterial type III secretion systems: a complex device for the delivery of bacterial effector proteins into eukaryotic host cells. *FEMS Microbiology Letters*. 2018;365(19). doi:10.1093/femsle/fny201
39. Backert S, Bernegger S, Skórko-Glonek J, Wessler S. Extracellular HtrA serine proteases: An emerging new strategy in bacterial pathogenesis. *Cellular Microbiology*. 2018;20(6):e12845. doi:10.1111/cmi.12845
40. Esoda CN, Kuehn MJ. *Pseudomonas aeruginosa* Leucine Aminopeptidase Influences Early Biofilm Composition and Structure via Vesicle-Associated Antibiofilm Activity. *Bio*. 2019;10(6):e02548-19, /mbio/10/6/mBio.02548-19.atom. doi:10.1128/mBio.02548-19
41. Cheng C, Wang X, Dong Z, et al. Aminopeptidase T of M29 Family Acts as A Novel Intracellular Virulence Factor for *Listeria monocytogenes* Infection. *Sci Rep*. 2015;5(1):17370. doi:10.1038/srep17370
42. Luckett JCA, Darch O, Watters C, et al. A Novel Virulence Strategy for *Pseudomonas aeruginosa* Mediated by an Autotransporter with Arginine-Specific Aminopeptidase Activity. *PLoS Pathog*. 2012;8(8):e1002854. doi:10.1371/journal.ppat.1002854
43. Carroll RK, Robison TM, Rivera FE, et al. Identification of an intracellular M17 family leucine aminopeptidase that is required for virulence in *Staphylococcus aureus*. *Microbes and Infection*. 2012;14(11):989-999. doi:10.1016/j.micinf.2012.04.013
44. Tamura Y, Suzuki S, Sawada T. Role of elastase as a virulence factor in experimental *Pseudomonas aeruginosa* infection in mice. *Microbial Pathogenesis*. 1992;12(3):237-244. doi:10.1016/0882-4010(92)90058-V
45. Woolwine SC, Wozniak DJ. Identification of an *Escherichia coli pep*A Homolog and Its Involvement in Suppression of the *alg*B Phenotype in Mucoid *Pseudomonas aeruginosa*. *J Bacteriol*. 1999;181(1):107-116. doi:10.1128/JB.181.1.107-116.1999
46. Tomlin H, Piccinini AM. A complex interplay between the extracellular matrix and the innate immune response to microbial pathogens. *Immunology*. 2018;155(2):186-201. doi:10.1111/imm.12972
47. Coetzer THT, Goldring JPD, Huson LEJ. Oligopeptidase B: A processing peptidase involved in pathogenesis. *Biochimie*. 2008;90(2):336-344. doi:10.1016/j.biochi.2007.10.011
48. Sauer RT, Baker TA. AAA+ Proteases: ATP-Fueled Machines of Protein Destruction. *Annu Rev Biochem*. 2011;80(1):587-612. doi:10.1146/annurev-biochem-060408-172623
49. Humbard MA, Maupin-Furlow JA. Prokaryotic Proteasomes: Nanocompartments of Degradation. *J Mol Microbiol Biotechnol*. 2013;23(4-5):321-334. doi:10.1159/000351348
50. Groll M, Bochtler M, Brandstetter H, Clausen T, Huber R. Molecular Machines for Protein Degradation. *ChemBioChem*. 2005;6(2):222-256. doi:10.1002/cbic.200400313
51. Becker SH, Darwin KH. Bacterial Proteasomes: Mechanistic and Functional Insights. *Microbiol Mol Biol Rev*. 2017;81(1):e00036-16, e00036-16. doi:10.1128/MMBR.00036-16
52. Müller AU, Weber-Ban E. The Bacterial Proteasome at the Core of Diverse Degradation Pathways. *Front Mol Biosci*. 2019;6:23. doi:10.3389/fmolb.2019.00023
53. Becker SH, Li H, Heran Darwin K. Biology and Biochemistry of Bacterial Proteasomes. In: *Macromolecular Protein Complexes II: Structure and Function*. Vol 93. Subcellular Biochemistry. Springer International Publishing; 2019:339-358. doi:10.1007/978-3-030-28151-9_11

54. Debunne N, Verbeke F, Janssens Y, Wynendaele E, De Spiegeleer B. Chromatography of Quorum Sensing Peptides: An Important Functional Class of the Bacterial Peptidome. *Chromatographia*. 2018;81(1):25-40. doi:10.1007/s10337-017-3411-2
55. Trost B, Kusalik A, Lucchese G, Kanduc D. Bacterial peptides are intensively present throughout the human proteome. *Self/Nonself*. 2010;1(1):71-74. doi:10.4161/self.1.1.9588
56. Cusick MF, Libbey JE, Fujinami RS. Molecular Mimicry as a Mechanism of Autoimmune Disease. *Clinic Rev Allerg Immunol*. 2012;42(1):102-111. doi:10.1007/s12016-011-8294-7
57. Hartley BS. Proteolytic Enzymes. *Annual Review of Biochemistry*. 1960;29:45-72.
58. Perona JJ, Craik CS. Structural basis of substrate specificity in the serine proteases. *Protein Science*. 1995;4(3):337-360. doi:10.1002/pro.5560040301
59. Rawlings ND, Barrett AJ. Evolutionary families of peptidases. *Biochemical Journal*. 1993;290(1):205-218. doi:10.1042/bj2900205
60. Rawlings ND, Barrett AJ. MEROPS: the peptidase database. *Nucleic Acids Research*. 1999;27(1):325-331.
61. Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A, Finn RD. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Research*. 2018;46(D1):D624-D632. doi:10.1093/nar/gkx1134
62. Rawlings ND, Barrett AJ, Bateman A. Asparagine Peptide Lyases: A SEVENTH CATALYTIC TYPE OF PROTEOLYTIC ENZYMES. *J Biol Chem*. 2011;286(44):38321-38328. doi:10.1074/jbc.M111.260026
63. Endrizzi JA, Breddam K, Remington J. 2.8 Å Structure of Yeast Serine Carboxypeptidase. *Biochemistry*. 1994;33:11106-11120.
64. Kuhn P, Knapp M, Soltis SM, Ganshaw G, Thoene M, Bott R. The 0.78 Å Structure of a Serine Protease: *Bacillus lentus* Subtilisin. *Biochemistry*. 1998;37(39):13446-13452. doi:10.1021/bi9813983
65. Stoll VS, Eger BT, Hynes RC, Martichonok V, Jones JB, Pai EF. Differences in Binding Modes of Enantiomers of 1-Acetamido Boronic Acid Based Protease Inhibitors: Crystal Structures of γ-Chymotrypsin and Subtilisin Carlsberg Complexes. *Biochemistry*. 1998;37(2):451-462. doi:10.1021/bi971166o
66. Schechter I, Berger A. On the size of the active site in proteases. I. Papain. *Biochemical and Biophysical Research Communications*. 1967;27(2):157-162. doi:10.1016/S0006-291X(67)80055-X
67. Baumeister W, Walz J, Zühl F, Seemüller E. The Proteasome: Paradigm of a Self-Compartmentalizing Protease. *Cell*. 1998;92(3):367-380. doi:10.1016/S0092-8674(00)80929-0
68. Maupin-Furlow JA. Archaeal Proteasomes and Sampylation. In: *Regulated Proteolysis in Microorganisms*. Vol 66. Subcellular Biochemistry. Springer Netherlands; 2013:297-327. doi:10.1007/978-94-007-5940-4_11
69. Volker C, Lupas AN. Molecular Evolution of Proteasomes. In: *The Proteasome — Ubiquitin Protein Degradation Pathway*. Vol 268. Current Topics in Microbiology and Immunology. Springer Berlin Heidelberg; 2002:1-22. doi:10.1007/978-3-642-59414-4_1
70. Rohrwild M, Coux O, Huang HC, et al. HslV-HslU: A novel ATP-dependent protease complex in *Escherichia coli* related to the eukaryotic proteasome. *Proceedings of the National Academy of Sciences*. 1996;93(12):5808-5813. doi:10.1073/pnas.93.12.5808
71. Gottesman S, Roche E, Zhou Y, Sauer RT. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes & Development*. 1998;12(9):1338-1347. doi:10.1101/gad.12.9.1338
72. Goff SA, Casson LP, Goldberg AL. Heat shock regulatory gene htpR influences rates of protein degradation and expression of the ion gene in *Escherichia coli*. *Proc NatL Acad Sci USA*. Published online 1984:6.
73. Tomoyasu T, Gamerl J, Bukaul B, et al. *Escherichia coli* FtsH is a membrane-bound, ATP- dependent protease which degrades the heat-shock transcription factor G32. *The EMBO Journal*. 1995;14(11):2551-2560.
74. Fuchs ACD, Alva V, Maldoner L, Albrecht R, Hartmann MD, Martin J. The Architecture of the Anbu Complex Reflects an Evolutionary Intermediate at the Origin of the Proteasome System. *Structure*. 2017;25(6):834-845.e5. doi:10.1016/j.str.2017.04.005
75. Vielberg M-T, Bauer VC, Groll M. On the Trails of the Proteasome Fold: Structural and Functional Analysis of the Ancestral β-Subunit Protein Anbu. *Journal of Molecular Biology*. 2018;430:628-640.
76. Wu W-F, Zhou Y, Gottesman S. Redundant *In Vivo* Proteolytic Activities of *Escherichia coli* Lon and the ClpYQ (HslUV) Protease. *J Bacteriol*. 1999;181(12):3681-3687. doi:10.1128/JB.181.12.3681-3687.1999
77. Kanemori M, Nishihara K, Yanagi H, Yura T. Synergistic Roles of HslVU and Other ATP-Dependent Proteases in Controlling *In Vivo* Turnover of σ^{32} and Abnormal Proteins in *Escherichia coli*. *J BACTERIOL*. 1997;179:7.
78. Chang C-Y, Hu H-T, Tsai C-H, Wu W-F. The degradation of RcsA by ClpYQ(HslUV) protease in *Escherichia coli*. *Microbiological Research*. 2016;184:42-50. doi:10.1016/j.micres.2016.01.001
79. Bittner L-M, Arends J, Narberhaus F. When, how and why? Regulated proteolysis by the essential FtsH protease in *Escherichia coli*. *Biological Chemistry*. 2017;398(5-6):625-635. doi:10.1515/hsz-2016-0302
80. Langklotz S, Baumann U, Narberhaus F. Structure and function of the bacterial AAA protease FtsH. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*. 2012;1823(1):40-48. doi:10.1016/j.bbamcr.2011.08.015

81. Clausen T, Southan C, Ehrmann M. The HtrA Family of Proteases: Implications for Protein Composition and Cell Fate. *Molecular Cell*. 2002;10:443-455.
82. Gur E, Biran D, Ron EZ. Regulated proteolysis in Gram-negative bacteria — how and when? *Nature Reviews Microbiology*. 2011;9(12):839-848. doi:10.1038/nrmicro2669
83. Groll M, Ditzel L, Löwe J, Stock D, Bochtler M, Bartunikt HD. Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature*. 1997;386(3):463-471.
84. Tamura T, Nagy I, Lupas A, et al. The first characterization of a eubacterial proteasome: the 20S complex of Rhodococcus. *Current Biology*. 1995;5(7):766-774. doi:10.1016/S0960-9822(95)00153-9
85. Kwon YD, Nagy I, Adams PD, Baumeister W, Jap BK. Crystal Structures of the *Rhodococcus* Proteasome with and without its Pro-peptides: Implications for the Role of the Pro-peptide in Proteasome Assembly. *Journal of Molecular Biology*. 2004;335(1):233-245. doi:10.1016/j.jmb.2003.08.029
86. Lowe J, Stock D, Jap B, Zwickl P, Baumeister W, Huber R. Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *Science*. 1995;268(5210):533-539. doi:10.1126/science.7725097
87. Bard JAM, Goodall EA, Greene ER, Jonsson E, Dong KC, Martin A. Structure and Function of the 26S Proteasome. *Annu Rev Biochem*. 2018;87(1):697-724. doi:10.1146/annurev-biochem-062917-011931
88. Hu G, Lin G, Wang M, et al. Structure of the *Mycobacterium tuberculosis* proteasome and mechanism of inhibition by a peptidyl boronate: Mtb proteasome structure. *Molecular Microbiology*. 2006;59(5):1417-1428. doi:10.1111/j.1365-2958.2005.05036.x
89. Campbell MG, Veesler D, Cheng A, Potter CS, Carragher B. 2.8 Å resolution reconstruction of the T*hermoplasma acidophilum* 20S proteasome using cryo-electron microscopy. *eLife*. 2015;4:e06380.
90. Heinemeyer W, Fischer M, Krimmer T, Stachon U, Wolf DH. The Active Sites of the Eukaryotic 20 S Proteasome and Their Involvement in Subunit Precursor Processing. *J Biol Chem*. 1997;272(40):25200-25209. doi:10.1074/jbc.272.40.25200
91. Huber EM, Heinemeyer W, Li X, Arendt CS, Hochstrasser M, Groll M. A unified mechanism for proteolysis and autocatalytic activation in the 20S proteasome. *Nat Commun*. 2016;7(1):10900. doi:10.1038/ncomms10900
92. Kunjappu MJ, Hochstrasser M. Assembly of the 20S proteasome. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*. 2014;1843(1):2-12. doi:10.1016/j.bbamcr.2013.03.008
93. Zühl F, Seemüller E, Golbik R, Baumeister W. Dissecting the assembly pathway of the 20S proteasome. *FEBS Letters*. 1997;418(1-2):189-194. doi:10.1016/S0014-5793(97)01370-7
94. Li D, Li H, Wang T, Pan H, Lin G, Li H. Structural basis for the assembly and gate closure mechanisms of the *Mycobacterium tuberculosis* 20S proteasome. *EMBO J*. 2010;29(12):2037-2047. doi:10.1038/emboj.2010.95
95. Zwickl P, Kleinz J, Baumeister W. Critical elements in proteasome assembly. *Nature Structural Biology*. 1994;1(11):765-770.
96. Sharon M, Witt S, Glasmacher E, Baumeister W, Robinson CV. Mass Spectrometry Reveals the Missing Links in the Assembly Pathway of the Bacterial 20 S Proteasome. *J Biol Chem*. 2007;282(25):18448-18457. doi:10.1074/jbc.M701534200
97. Latham MP, Sekhar A, Kay LE. Understanding the mechanism of proteasome 20S core particle gating. *Proceedings of the National Academy of Sciences*. 2014;111(15):5532-5537. doi:10.1073/pnas.1322079111
98. Rabl J, Smith DM, Yu Y, Chang S-C, Goldberg AL, Cheng Y. Mechanism of Gate Opening in the 20S Proteasome by the Proteasomal ATPases. *Molecular Cell*. 2008;30(3):360-368. doi:10.1016/j.molcel.2008.03.004
99. Striebel F, Hunkeler M, Summer H, Weber-Ban E. The mycobacterial Mpa–proteasome unfolds and degrades pupylated substrates by engaging Pup's N-terminus. *EMBO J*. 2010;29(7):1262-1271. doi:10.1038/emboj.2010.23
100. Darwin KH, Lin G, Chen Z, Li H, Nathan CF. Characterization of a *Mycobacterium tuberculosis* proteasomal ATPase homologue: Mycobacterium tuberculosis AAA ATPase. *Molecular Microbiology*. 2004;55(2):561-571. doi:10.1111/j.1365-2958.2004.04403.x
101. Wolf S, Nagy I, Lupas A, et al. Characterization of ARC, a divergent member of the AAA ATPase family from *Rhodococcus erythropolis*. *Journal of Molecular Biology*. 1998;277(1):13-25. doi:10.1006/jmbi.1997.1589
102. Majumder P, Rudack T, Beck F, et al. Cryo-EM structures of the archaeal PAN-proteasome reveal an around-the-ring ATPase cycle. *Proc Natl Acad Sci USA*. 2019;116(2):534-539. doi:10.1073/pnas.1817752116
103. Djuranovic S, Hartmann MD, Habeck M, et al. Structure and Activity of the N-Terminal Substrate Recognition Domains in Proteasomal ATPases. *Molecular Cell*. 2009;34(5):580-590. doi:10.1016/j.molcel.2009.04.030
104. Wang T, Li H, Lin G, et al. Structural Insights on the *Mycobacterium tuberculosis* Proteasomal ATPase Mpa. *Structure*. 2009;17(10):1377-1385. doi:10.1016/j.str.2009.08.010
105. Barandun J, Delley CL, Weber-Ban E. The pupylation pathway and its role in mycobacteria. *BMC Biol*. 2012;10(1):95. doi:10.1186/1741-7007-10-95
106. Humbard MA, Miranda HV, Lim J-M, et al. Ubiquitin-like small archaeal modifier proteins (SAMPs) in *Haloferax volcanii*. *Nature*. 2010;463(7277):54-60. doi:10.1038/nature08659

107. Alhuwaider AAH, Truscott KN, Dougan DA. Pupylation of PafA or Pup inhibits components of the Pup-Proteasome System. *FEBS Lett*. 2018;592(1):15-23. doi:10.1002/1873-3468.12930
108. Bai L, Hu K, Wang T, Jastrab JB, Darwin KH, Li H. Structural analysis of the dodecameric proteasome activator PafE in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA*. 2016;113(14):E1983-E1992. doi:10.1073/pnas.1512094113
109. Hu K, Jastrab JB, Zhang S, et al. Proteasome substrate capture and gate opening by the accessory factor PafE from *Mycobacterium tuberculosis*. *J Biol Chem*. 2018;293(13):4713-4723. doi:10.1074/jbc.RA117.001471
110. Rohrwild M, Pfeifer G, Santarius U, et al. The ATP-dependent HslVU protease from *Escherichia* coli is a four-ring structure resembling the proteasome. *Nature Structural Biology*. 1997;4(2):133-139.
111. Bochtler M, Ditzel L, Groll M, Huber R. Crystal structure of heat shock locus V (HslV) from *Escherichia coli*. *Proceedings of the National Academy of Sciences*. 1997;94(12):6070-6074. doi:10.1073/pnas.94.12.6070
112. Fuchs ACD, Hartmann MD. On the Origins of Symmetry and Modularity in the Proteasome Family: Symmetry Transitions are Pivotal in the Evolution and Functional Diversification of Self-Compartmentalizing Proteases. *BioEssays*. 2019;41(5):1800237. doi:10.1002/bies.201800237
113. Seol JH, Yoo SJ, Shin DH, et al. The Heat-Shock Protein HslVU from *Escherichia Coli* is a Protein-Activated ATPase as well as an ATP-Dependent Proteinase. *Eur J Biochem*. 1997;247(3):1143-1150. doi:10.1111/j.1432-1033.1997.01143.x
114. Ramachandran R, Hartmann C, Song HK, Huber R, Bochtler M. Functional interactions of HslV (ClpQ) with the ATPase HslU (ClpY). *Proceedings of the National Academy of Sciences*. 2002;99(11):7396-7401. doi:10.1073/pnas.102188799
115. Song HK, Bochtler M, Azim MK, Hartmann C, Huber R, Ramachandran R. Isolation and characterization of the prokaryotic proteasome homolog HslVU (ClpQY) from *Thermotoga maritima* and the crystal structure of HslV. *Biophysical Chemistry*. 2002;100(1-3):437-452. doi:10.1016/S0301-4622(02)00297-1
116. Bochtler M, Hartmann C, Song HK, Bourenkov GP, Bartunik HD, Huber R. The structures of HslU and the ATP-dependent protease HslU–HslV. *Nature*. 2000;403(6771):800-805. doi:10.1038/35001629
117. Sousa MC, Trame CB, Tsuruta H, Wilbanks SM, Reddy VS, McKay DB. Crystal and Solution Structures of an HslUV Protease–Chaperone Complex. *Cell*. 2000;103(4):633-643. doi:10.1016/S0092-8674(00)00166-5
118. Song HK, Hartmann C, Ramachandran R, et al. Mutational studies on HslU and its docking mode with HslV. *Proceedings of the National Academy of Sciences*. 2000;97(26):14103-14108. doi:10.1073/pnas.250491797
119. Wang J, Song JJ, Seong IS, et al. Nucleotide-Dependent Conformational Changes in a Protease-Associated ATPase HslU. *Structure*. 2001;9:1107-1116.
120. Park E, Lee JW, Yoo HM, et al. Structural Alteration in the Pore Motif of the Bacterial 20S Proteasome Homolog HslV Leads to Uncontrolled Protein Degradation. *Journal of Molecular Biology*. 2013;425(16):2940-2954. doi:10.1016/j.jmb.2013.05.011
121. Valas RE, Bourne PE. Rethinking Proteasome Evolution: Two Novel Bacterial Proteasomes. *J Mol Evol*. 2008;66(5):494-504. doi:10.1007/s00239-008-9075-7
122. Fuchs ACD, Maldoner L, Hipp K, Hartmann MD, Martin J. Structural characterization of the bacterial proteasome homolog BPH reveals a tetradecameric double-ring complex with unique inner cavity properties. *J Biol Chem*. 2018;293(3):920-930. doi:10.1074/jbc.M117.815258
123. Demoulin CF, Lara YJ, Cornet L, et al. Cyanobacteria evolution: Insight from the fossil record. *Free Radical Biology and Medicine*. 2019;140:206-223. doi:10.1016/j.freeradbiomed.2019.05.007
124. Piasecka A, Czapinska H, Vielberg M-T, et al. The *Y. bercovieri* Anbu crystal structure sheds light on the evolution of highly (pseudo)symmetric multimers. *Journal of Molecular Biology*. 2018;430(5):611-627. doi:10.1016/j.jmb.2017.11.016
125. Yu AYH, Houry WA. ClpP: A distinctive family of cylindrical energy-dependent serine proteases. *FEBS Letters*. 2007;581(19):3749-3757. doi:10.1016/j.febslet.2007.04.076
126. Beuron F, Maurizi MR, Belnap DM, et al. At Sixes and Sevens: Characterization of the Symmetry Mismatch of the ClpAP Chaperone-Assisted Protease. *Journal of Structural Biology*. 1998;123(3):248-259. doi:10.1006/jsbi.1998.4039
127. Gersch M, List A, Groll M, Sieber SA. Insights into Structural Network Responsible for Oligomerization and Activity of Bacterial Virulence Regulator Caseinolytic Protease P (ClpP) Protein. *J Biol Chem*. 2012;287(12):9484-9494. doi:10.1074/jbc.M111.336222
128. Hoskins JR, Yanagihara K, Mizuuchi K, Wickner S. ClpAP and ClpXP degrade proteins with tags located in the interior of the primary sequence. *Proceedings of the National Academy of Sciences*. 2002;99(17):11037-11042. doi:10.1073/pnas.172378899
129. Zhou Y, Gottesman S, Hoskins JR, Maurizi MR, Wickner S. The RssB response regulator directly targets sS for degradation by ClpXP. *Genes & Development*. 2001;15:627-637.
130. Stahlhut SG, Alqarzaee AA, Jensen C, et al. The ClpXP protease is dispensable for degradation of unfolded proteins in *Staphylococcus aureus*. *Sci Rep*. 2017;7(1):11739. doi:10.1038/s41598-017-12122-y

131. Wong KS, Mabanglo MF, Seraphim TV, et al. Acyldepsipeptide Analogs Dysregulate Human Mitochondrial ClpP Protease Activity and Cause Apoptotic Cell Death. *Cell Chemical Biology*. 2018;25(8):1017-1030.e9. doi:10.1016/j.chembiol.2018.05.014
132. Wang J, Hartling JA, Flanagan JM. The Structure of ClpP at 2.3 Å Resolution Suggests a Model for ATP-Dependent Proteolysis. *Cell*. 1997;91(4):447-456. doi:10.1016/S0092-8674(00)80431-6
133. Zhang J, Ye F, Lan L, Jiang H, Luo C, Yang C-G. Structural Switching of *Staphylococcus aureus* Clp Protease: A KEY TO UNDERSTANDING PROTEASE DYNAMICS. *J Biol Chem*. 2011;286(43):37590-37601. doi:10.1074/jbc.M111.277848
134. Gatsogiannis C, Balogh D, Merino F, Sieber S, Raunser S. *Cryo-EM Structure of the ClpXP Protein Degradation Machinery*. Molecular Biology; 2019. doi:10.1101/638692
135. Lopez KE, Rizo AN, Tse E, et al. Conformational plasticity of the ClpAP AAA+ protease couples protein unfolding and proteolysis. *Nat Struct Mol Biol*. 2020;27(5):406-416. doi:10.1038/s41594-020-0409-5
136. Fei X, Bell TA, Jenni S, et al. Structures of the ATP-fueled ClpXP proteolytic machine bound to protein substrate. *eLife*. 2020;9:e52774. doi:10.7554/eLife.52774
137. Lee I, Suzuki CK. Functional mechanics of the ATP-dependent Lon protease- lessons from endogenous protein and synthetic peptide substrates. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*. 2008;1784(5):727-735. doi:10.1016/j.bbapap.2008.02.010
138. Botos I, Melnikov EE, Cherry S, et al. The Catalytic Domain of *Escherichia coli* Lon Protease Has a Unique Fold and a Ser-Lys Dyad in the Active Site. *J Biol Chem*. 2004;279(9):8140-8148. doi:10.1074/jbc.M312243200
139. Rotanova TV, Melnikov EE, Khalatova AG, et al. Classification of ATP-dependent proteases Lon and comparison of the active sites of their proteolytic domains. *Eur J Biochem*. 2004;271(23-24):4865-4871. doi:10.1111/j.1432-1033.2004.04452.x
140. Liao J-H, Kuo C-I, Huang Y-Y, et al. A Lon-Like Protease with No ATP-Powered Unfolding Activity. *PLoS ONE*. 2012;7(7):e40226. doi:10.1371/journal.pone.0040226
141. Botos I, Lountos GT, Wu W, et al. Cryo-EM structure of substrate-free *E. coli* Lon protease provides insights into the dynamics of Lon machinery. *Current Research in Structural Biology*. 2019;1:13-20. doi:10.1016/j.crstbi.2019.10.001
142. Li M, Gustchina A, Rasulova FS, et al. Structure of the N-terminal fragment of *Escherichia coli* Lon protease. *Acta Crystallogr D Biol Crystallogr*. 2010;66(8):865-873. doi:10.1107/S0907444910019554
143. Duman RE, Löwe J. Crystal Structures of *Bacillus subtilis* Lon Protease. *Journal of Molecular Biology*. 2010;401(4):653-670. doi:10.1016/j.jmb.2010.06.030
144. Inoue M, Fukui K, Fujii Y, et al. The Lon protease-like domain in the bacterial RecA paralog RadA is required for DNA binding and repair. *J Biol Chem*. 2017;292(23):9801-9814. doi:10.1074/jbc.M116.770180
145. Kowit D, Goldberg L. Intermediate Steps in the Degradation of a Specific Abnormal Protein in *Escherichia coli*. *Journal of Biological Chemistry*. 1977;252(23):8350-8357.
146. Choy JS, Aung LL, Karzai AW. Lon Protease Degrades Transfer-Messenger RNA-Tagged Proteins. *JB*. 2007;189(18):6564-6571. doi:10.1128/JB.00860-07
147. Van Melderen L, Aertsen A. Regulation and quality control by Lon-dependent proteolysis. *Research in Microbiology*. 2009;160(9):645-651. doi:10.1016/j.resmic.2009.08.021
148. Gur E, Sauer RT. Recognition of misfolded proteins by Lon, a AAA+ protease. *Genes & Development*. 2008;22(16):2267-2277. doi:10.1101/gad.1670908
149. Nomura K, Kato J, Takiguchi N, Ohtake H, Kuroda A. Effects of Inorganic Polyphosphate on the Proteolytic and DNA-binding Activities of Lon in *Escherichia coli*. *J Biol Chem*. 2004;279(33):34406-34410. doi:10.1074/jbc.M404725200
150. Gur E. The Lon AAA+ Protease. In: *Regulated Proteolysis in Microorganisms*. Vol 66. Subcellular Biochemistry. Springer Netherlands; 2013:35-51. doi:10.1007/978-94-007-5940-4_2
151. Van Melderen L, Thi MHD, Lecchi P, Gottesman S, Couturier M, Maurizi MR. ATP-dependent Degradation of CcdA by Lon Protease: EFFECTS OF SECONDARY STRUCTURE AND HETEROLOGOUS SUBUNIT INTERACTIONS. *J Biol Chem*. 1996;271(44):27730-27738. doi:10.1074/jbc.271.44.27730
152. Su S-C, Lin C-C, Tai H-C, et al. Structural Basis for the Magnesium-Dependent Activation and Hexamerization of the Lon AAA+ Protease. *Structure*. 2016;24(5):676-686. doi:10.1016/j.str.2016.03.003
153. Cha S-S, An YJ, Lee CR, et al. Crystal structure of Lon protease: molecular architecture of gated entry to a sequestered degradation chamber. *EMBO J*. 2010;29(20):3520-3530. doi:10.1038/emboj.2010.226
154. Botos I, Melnikov EE, Cherry S, et al. Atomic-resolution Crystal Structure of the Proteolytic Domain of *Archaeoglobus fulgidus* Lon Reveals the Conformational Variability in the Active Sites of Lon Proteases. *Journal of Molecular Biology*. 2005;351(1):144-157. doi:10.1016/j.jmb.2005.06.008
155. Shin M, Puchades C, Asmita A, et al. Structural basis for distinct operational modes and protease activation in AAA+ protease Lon. *Sci Adv*. 2020;6(21):eaba8404. doi:10.1126/sciadv.aba8404

156. Im YJ, Na Y, Kang GB, et al. The Active Site of a Lon Protease from *Methanococcus jannaschii* Distinctly Differs from the Canonical Catalytic Dyad of Lon Proteases. *J Biol Chem*. 2004;279(51):53451-53457. doi:10.1074/jbc.M410437200
157. Park S-C, Jia B, Yang J-K, Jeon Y-J, Chung CH, Cheong G-W. Oligomeric Structure of the ATP-dependent Protease La (Lon) of. *Molecules and Cells*. 2006;21(1):129-134.
158. Rudyak SG, Brenowitz M, Shrader TE. Mg^{2+} -Linked Oligomerization Modulates the Catalytic Activiy of the Lon (La) Protease from *Mycobacterium smegmatis*. *Biochemistry*. 2001;40(31):9317-9323. doi:10.1021/bi0102508
159. Li M, Rasulova F, Melnikov EE, et al. Crystal structure of the N-terminal domain of *E. coli* Lon protease: CRYSTAL STRUCTURE OF THE N-TERMINAL DOMAIN OF *E. COLI* LON PROTEASE. *Protein Science*. 2005;14(11):2895-2900. doi:10.1110/ps.051736805
160. Chen X, Zhang S, Bi F, et al. Crystal structure of the N domain of Lon protease from *Mycobacterium avium* complex. *Protein Science*. 2019:pro.3687. doi:10.1002/pro.3687
161. Ogura T, Inoue K, Tatsuta T, et al. Balanced biosynthesis of major membrane components through regulated degradation of the committed enzyme of lipid A biosynthesis by the AAA protease FtsH (HflB) in *Escherichia coli*. *Mol Microbiol*. 1999;31(3):833-844. doi:10.1046/j.1365-2958.1999.01221.x
162. Guisbert E, Herman C, Lu CZ, Gross CA. A chaperone network controls the heat shock response in *E. coli*. *Genes & Development*. 2004;18(22):2812-2821. doi:10.1101/gad.1219204
163. Bittner L-M, Westphal K, Narberhaus F. Conditional Proteolysis of the Membrane Protein YfgM by the FtsH Protease Depends on a Novel N-terminal Degron. *J Biol Chem*. 2015;290(31):19367-19378. doi:10.1074/jbc.M115.648550
164. Filippova EV, Zemaitaitis B, Aung T, Wolfe AJ, Anderson WF. Structural Basis for DNA Recognition by the Two-Component Response Regulator RcsB. *mBio*. 2018;9(1):e01993-17. doi:10.1128/mBio.01993-17
165. Bieniossek C, Niederhauser B, Baumann UM. The crystal structure of apo-FtsH reveals domain movements necessary for substrate unfolding and translocation. *Proceedings of the National Academy of Sciences*. 2009;106(51):21579-21584. doi:10.1073/pnas.0910708106
166. An JY, Sharif H, Kang GB, et al. Structural insights into the oligomerization of FtsH periplasmic domain from *Thermotoga maritima*. *Biochemical and Biophysical Research Communications*. 2018;495(1):1201-1207. doi:10.1016/j.bbrc.2017.11.158
167. Krzywda S, Brzozowski AM, Verma C, Karata K, Ogura T, Wilkinson AJ. The Crystal Structure of the AAA Domain of the ATP-Dependent Protease FtsH of *Escherichia coli* at 1.5 Å Resolution. *Structure*. 2002;10(8):1073-1083. doi:10.1016/S0969-2126(02)00806-7
168. Bieniossek C, Schalch T, Bumann M, Meister M, Meier R, Baumann U. The molecular architecture of the metalloprotease FtsH. *Proceedings of the National Academy of Sciences*. 2006;103(9):3066-3071. doi:10.1073/pnas.0600031103
169. Bode W, Gomis-Rüth F-X, Stöckler W. 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 Letters*. 1993;331(1-2):134-140. doi:10.1016/0014-5793(93)80312-I
170. Spiess C, Beil A, Ehrmann M. A Temperature-Dependent Switch from Chaperone to Protease in a Widely Conserved Heat Shock Protein. *Cell*. 1999;97(3):339-347. doi:10.1016/S0092-8674(00)80743-6
171. Huber D, Bukau B. DegP: a Protein "Death Star." *Structure*. 2008;16(7):989-990. doi:10.1016/j.str.2008.06.004
172. Jones CH, Dexter P, Evans AK, Liu C, Hultgren SJ, Hruby DE. *Escherichia coli* DegP Protease Cleaves between Paired Hydrophobic Residues in a Natural Substrate: the PapA Pilin. *JB*. 2002;184(20):5762-5771. doi:10.1128/JB.184.20.5762-5771.2002
173. Krojer T, Garrido-Franco M, Huber R, Ehrmann M, Clausen T. Crystal structure of DegP (HtrA) reveals a new protease-chaperone machine. *Nature*. 2002;416(6879):455-459. doi:10.1038/416455a
174. Jomaa A, Damjanovic D, Leong V, Ghirlando R, Iwanczyk J, Ortega J. The Inner Cavity of *Escherichia coli* DegP Protein Is Not Essential for Molecular Chaperone and Proteolytic Activity. *JB*. 2007;189(3):706-716. doi:10.1128/JB.01334-06
175. Krojer T, Sawa J, Schäfer E, Saibil H, Ehrmann M, Clausen T. Structural basis for the regulated protease and chaperone function of DegP. *Nature*. 2008;453:885-890.
176. Jiang J, Zhang X, Chen Y, et al. Activation of DegP chaperone-protease via formation of large cage-like oligomers upon binding to substrate proteins. *Proceedings of the National Academy of Sciences*. 2008;105(33):11939-11944. doi:10.1073/pnas.0805464105
177. Kim S, Grant RA, Sauer RT. Covalent Linkage of Distinct Substrate Degrons Controls Assembly and Disassembly of DegP Proteolytic Cages. *Cell*. 2011;145(1):67-78. doi:10.1016/j.cell.2011.02.024
178. Bai X, Pan X, Wang X, et al. Characterization of the Structure and Function of *Escherichia coli* DegQ as a Representative of the DegQ-like Proteases of Bacterial HtrA Family Proteins. *Structure*. 2011;19(9):1328-1337. doi:10.1016/j.str.2011.06.013

179. Sawa J, Malet H, Krojer T, Canellas F, Ehrmann M, Clausen T. Molecular Adaptation of the DegQ Protease to Exert Protein Quality Control in the Bacterial Cell Envelope. *J Biol Chem*. 2011;286(35):30680-30690. doi:10.1074/jbc.M111.243832
180. Wilken C, Kitzing K, Kurzbauer R, Ehrmann M, Clausen T. Crystal Structure of the DegS Stress Sensor: How a PDZ Domain Recognizes Misfolded Protein and Activates a Protease. *Cell*. 2004;117:483-494.
181. Kisselev AF, Akopian TN, Woo KM, Goldberg AL. The Sizes of Peptides Generated from Protein by Mammalian 26 and 20 S Proteasomes: IMPLICATIONS FOR UNDERSTANDING THE DEGRADATIVE MECHANISM AND ANTIGEN PRESENTATION. *Journal of Biological Chemistry*. 1999;274(6):3363-3371. doi:10.1074/jbc.274.6.3363
182. Miller CG, Schwartz G. Peptidase-deficient mutants of *Escherichia coli*. *Journal of Bacteriology*. 1978;135(2):603-611. doi:10.1128/JB.135.2.603-611.1978
183. Maupin-Furlow JA. Proteolytic systems of archaea: slicing, dicing, and mincing in the extreme. *Emerging Topics in Life Sciences*. 2018;2(4):561-580. doi:10.1042/ETLS20180025
184. Borissenko L, Groll M. Crystal Structure of TET Protease Reveals Complementary Protein Degradation Pathways in Prokaryotes. *Journal of Molecular Biology*. 2005;346(5):1207-1219. doi:10.1016/j.jmb.2004.12.056
185. Hayashi M, Tabata K, Yagasaki M, Yonetani Y. Effect of multidrug-efflux transporter genes on dipeptide resistance and overproduction in *Escherichia coli*: Dipeptide transporters. *FEMS Microbiology Letters*. 2010;304(1):12-19. doi:10.1111/j.1574-6968.2009.01879.x
186. Yen C, Green L, Miller CG. Peptide accumulation during growth of peptidase deficient mutants. *Journal of Molecular Biology*. 1980;143(1):35-48. doi:10.1016/0022-2836(80)90123-0
187. Bhosale M, Pande S, Kumar A, Kairamkonda S, Nandi D. Characterization of two M17 family members in *Escherichia coli*, Peptidase A and Peptidase B. *Biochemical and Biophysical Research Communications*. 2010;395(1):76-81. doi:10.1016/j.bbrc.2010.03.142
188. Vogt V. Purification and Properties of an Aminopeptidase from *Escherichia coli*. *Journal of Biological Chemistry*. 1970;245(18):4760-4769.
189. Roovers M, Charlier D, Feller A, et al. CarP, a novel gene regulating the transcription of the carbamoylphosphate synthetase operon of *Escherichia coli*. *Journal of Molecular Biology*. 1988;204(4):857-865. doi:10.1016/0022-2836(88)90046-0
190. Charlier D, Nguyen Le Minh P, Roovers M. Regulation of carbamoylphosphate synthesis in *Escherichia coli*: an amazing metabolite at the crossroad of arginine and pyrimidine biosynthesis. *Amino Acids*. 2018;50(12):1647-1661. doi:10.1007/s00726-018-2654-z
191. Stirling CJ, Colloms SD, Collins JF, Szatmari G, Sherratt DJ. *xer*B, an *Escherichia coli* gene required for plasmid ColE1 site-specific recombination, is identical to *pep*A, encoding aminopeptidase A, a protein with substantial similarity to bovine lens leucine aminopeptidase. *The EMBO Journal*. 1989;8(5):1623-1627. doi:10.1002/j.1460-2075.1989.tb03547.x
192. Charlier D, Gh GH, Kholti A, Gigot D, Piérard A, Glansdorff N. *car*P, Involved in Pyrimidine Regulation of the *Escherichia coli* Carbamoylphosphate Synthetase Operon Encodes a Sequence-specific DNA-binding Protein Identical to XerB and PepA, also Required for Resolution of ColEI Multimers. *Journal of Molecular Biology*. 1995;250:392-406.
193. Caldara M, Charlier D, Cunin R. The arginine regulon of *Escherichia coli*: whole-system transcriptome analysis discovers new genes and provides an integrated view of arginine regulation. *Microbiology*. 2006;152(11):3343-3354. doi:10.1099/mic.0.29088-0
194. McCulloch R, Burke ME, Sherratt DJ. Peptidase activity of *Escherichia coli* aminopeptidase A is not required for its role in Xer site-specific recombination. *Mol Microbiol*. 1994;12(2):241-251. doi:10.1111/j.1365-2958.1994.tb01013.x
195. Minh PNL, Devroede N, Massant J, Maes D, Charlier D. Insights into the architecture and stoichiometry of *Escherichia coli* PepA*DNA complexes involved in transcriptional control and site-specific DNA recombination by atomic force microscopy. *Nucleic Acids Research*. 2009;37(5):1463-1476. doi:10.1093/nar/gkn1078
196. Nguyen Le Minh P, Nadal M, Charlier D. The trigger enzyme PepA (aminopeptidase A) of *Escherichia coli*, a transcriptional repressor that generates positive supercoiling. *FEBS Lett*. 2016;590(12):1816-1825. doi:10.1002/1873-3468.12224
197. Behari J, Stagon L, Calderwood SB. pepA, a Gene Mediating pH Regulation of Virulence Genes in *Vibrio cholerae*. *J Bacteriol*. 2001;183(1):178-188. doi:10.1128/JB.183.1.178-188.2001
198. Strater N, Sherratt DJ, Colloms SD. X-ray structure of aminopeptidase A from *Escherichia coli* and a model for the nucleoprotein complex in Xer site-specific recombination. *The EMBO Journal*. 1999;18(16):4513-4522. doi:10.1093/emboj/18.16.4513
199. Kale A, Pijning T, Sonke T, Dijkstra BW, Thunnissen A-MWH. Crystal Structure of the Leucine Aminopeptidase from *Pseudomonas putida* Reveals the Molecular Basis for its Enantioselectivity and Broad Substrate Specificity. *Journal of Molecular Biology*. 2010;398(5):703-714. doi:10.1016/j.jmb.2010.03.042

200. Minasov G, Lam MR, Rosas-Lemus M, et al. Comparison of metal-bound and unbound structures of aminopeptidase B proteins from *Escherichia coli* and *Yersina pestis*. *Protein Science*. 2020;29(7):1618-1628. doi:10.1002/pro.3876
201. Kirsh M, Dembinski D, Hartman P, Miller CG. *Salmonella typhimurium* Peptidase Active on Carnosine. *Journal of Bacteriology*. 1978;134(2):361-374.
202. Klein J, Henrich B, Plapp R. Cloning and Expression of the *pep*D Gene of *Escherichia coli*. *Microbiology*. 1986;132(8):2337-2343. doi:10.1099/00221287-132-8-2337
203. Henrich B, Monnerjahn U, Plapp R. Peptidase D gene (*pep*D) of *Escherichia coli* K-12: nucleotide sequence, transcript mapping, and comparison with other peptidase genes. *Journal of Bacteriology*. 1990;172(8):4641-4651. doi:10.1128/JB.172.8.4641-4651.1990
204. Chang C-Y, Hsieh Y-C, Wang T-Y, et al. Crystal Structure and Mutational Analysis of Aminoacylhistidine Dipeptidase from *Vibrio alginolyticus* Reveal a New Architecture of M20 Metallopeptidases. *J Biol Chem*. 2010;285(50):39500-39510. doi:10.1074/jbc.M110.139683
205. Chandu D, Kumar A, Nandi D. PepN, the Major Suc-LLVY-AMC-hydrolyzing Enzyme in *Escherichia coli* , Displays Functional Similarity with Downstream Processing Enzymes in Archaea and Eukarya: IMPLICATIONS IN CYTOSOLIC PROTEIN DEGRADATION. *J Biol Chem*. 2003;278(8):5548-5556. doi:10.1074/jbc.M207926200
206. Chandu D, Nandi D. PepN is the major aminopeptidase in *Escherichia coli*: insights on substrate specificity and role during sodium-salicylate-induced stress. *Microbiology*. 2003;149(12):3437-3447. doi:10.1099/mic.0.26518-0
207. Lazdunski C, Busuttil J, Lazdunski A. Purification and Properties of a Periplasmic Aminoendopeptidase from *Escherichia coli*. *Eur J Biochem*. 1975;60(2):363-369. doi:10.1111/j.1432-1033.1975.tb21011.x
208. Addlagatta A, Gay L, Matthews BW. Structural Basis for the Unusual Specificity of *Escherichia coli* Aminopeptidase N. *Biochemistry*. 2008;47(19):5303-5311. doi:10.1021/bi7022333
209. Chavagnat F, Casey MG, Meyer J. Purification, Characterization, Gene Cloning, Sequencing, and Overexpression of Aminopeptidase N from *Streptococcus thermophilus*. *Appl Environ Microbiol*. 1999;65(7):3001-3007. doi:10.1128/AEM.65.7.3001-3007.1999
210. Latil M, Murgier M, Lazdunski A, Lazdunski C. Isolation and genetic mapping of *Escherichia coli* aminopeptidase mutants. *Molec Gen Genet*. 1976;148(1):43-47. doi:10.1007/BF00268544
211. Yu G, Li L, Liu X, et al. The standalone aminopeptidase PepN catalyzes the maturation of blasticidin S from leucylblasticidin S. *Sci Rep*. 2015;5(1):17641. doi:10.1038/srep17641
212. Addlagatta A, Gay L, Matthews BW. Structure of aminopeptidase N from *Escherichia coli* suggests a compartmentalized, gated active site. *Proceedings of the National Academy of Sciences*. 2006;103(36):13339-13344. doi:10.1073/pnas.0606167103
213. Holmes MA, Tronrud DE, Matthews BW. Structural Analysis of the Inhibition of Thermolysin by an Active-Site-Directed Irreversible Inhibitor? *Biochemistry*. 1983;22:236-240.
214. Kyrieleis OJP, Goettig P, Kiefersauer R, Huber R, Brandstetter H. Crystal Structures of the Tricorn Interacting Factor F3 from *Thermoplasma acidophilum*, a Zinc Aminopeptidase in Three Different Conformations. *Journal of Molecular Biology*. 2005;349(4):787-800. doi:10.1016/j.jmb.2005.03.070
215. Lassy RAL, Miller CG. Peptidase E, a Peptidase Specific for N-Terminal Aspartic Dipeptides, Is a Serine Hydrolase. *J Bacteriol*. 2000;182(9):2536-2543. doi:10.1128/JB.182.9.2536-2543.2000
216. Hakansson K, Wang AH-J, Miller CG. The structure of aspartyl dipeptidase reveals a unique fold with a Ser-His-Glu catalytic triad. *Proceedings of the National Academy of Sciences*. 2000;97(26):14097-14102. doi:10.1073/pnas.260376797
217. Yadav P, Goyal VD, Chandravanshi K, et al. Catalytic triad heterogeneity in S51 peptidase family: Structural basis for functional variability. *Proteins*. 2019;87(8):679-692. doi:10.1002/prot.25693
218. Yadav P, Goyal VD, Gaur NK, Kumar A, Gokhale SM, Makde RD. Structure of Asp-bound peptidase E from *Salmonella enterica*: Active site at dimer interface illuminates Asp recognition. *FEBS Lett*. 2018;592(19):3346-3354. doi:10.1002/1873-3468.13247
219. Matos J, Nardi M, Kumura H, Monnet V. Genetic Characterization of pepP, Which Encodes an Aminopeptidase P Whose Deficiency Does Not Affect *Lactococcus lactis* Growth in Milk, Unlike Deficiency of the X-Prolyl Dipeptidyl Aminopeptidase. *Appl Environ Microbiol*. 1998;64(11):4591-4595. doi:10.1128/AEM.64.11.4591-4595.1998
220. Park M-S, Hill CM, Li Y, et al. Catalytic properties of the PepQ prolidase from *Escherichia coli*. *Archives of Biochemistry and Biophysics*. 2004;429(2):224-230. doi:10.1016/j.abb.2004.06.022
221. Weaver J, Watts T, Li P, Rye HS. Structural Basis of Substrate Selectivity of *E. coli* Prolidase. *PLoS ONE*. 2014;9(10):e111531. doi:10.1371/journal.pone.0111531
222. Wilce MCJ, Bond CS, Dixon NE, et al. Structure and mechanism of a proline-specific aminopeptidase from *Escherichia coli*. *Proceedings of the National Academy of Sciences*. 1998;95(7):3472-3477. doi:10.1073/pnas.95.7.3472

223. Graham SC, Bond CS, Freeman HC, Guss JM. Structural and Functional Implications of Metal Ion Selection in Aminopeptidase P, a Metalloprotease with a Dinuclear Metal Center. *Biochemistry*. 2005;44(42):13820-13836. doi:10.1021/bi0512849
224. Peng C-T, Liu L, Li C-C, et al. Structure–Function Relationship of Aminopeptidase P from *Pseudomonas aeruginosa*. *Front Microbiol*. 2017;8:2385. doi:10.3389/fmicb.2017.02385
225. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences*. 2001;98(18):10037-10041. doi:10.1073/pnas.181342398
226. Malki A, Caldas T, Abdallah J, et al. Peptidase Activity of the *Escherichia coli* Hsp31 Chaperone. *J Biol Chem*. 2005;280(15):14420-14426. doi:10.1074/jbc.M408296200
227. Halio SB, Bauer MW, Mukund S, Adams M, Kelly RM. Purification and Characterization of Two Functional Forms of Intracellular Protease PfpI from the Hyperthermophilic Archaeon *Pyrococcus furiosus*. *Applied and environmental microbiology*. 1997;63(1):289-295. doi:10.1128/AEM.63.1.289-295.1997
228. Du X, Choi I-G, Kim R, et al. Crystal structure of an intracellular protease from *Pyrococcus horikoshii* at 2-Å resolution. *Proceedings of the National Academy of Sciences*. 2000;97(26):14079-14084. doi:10.1073/pnas.260503597
229. Zhan D, Bai A, Yu L, Han W, Feng Y. Characterization of the PH1704 Protease from *Pyrococcus horikoshii* OT3 and the Critical Functions of Tyr120. *PLoS ONE*. 2014;9(9):e103902. doi:10.1371/journal.pone.0103902
230. Hicks PM, Chang LS, Kelly RM. *Pyrococcus furiosus* Protease I. In: *Handbook of Proteolytic Enzymes*. Elsevier; 2013:2449-2451. doi:10.1016/B978-0-12-382219-2.00547-0
231. Sastry MSR, Quigley PM, Hol WGJ, Baneyx F. The linker-loop region of *Escherichia coli* chaperone Hsp31 functions as a gate that modulates high-affinity substrate binding at elevated temperatures. *Proceedings of the National Academy of Sciences*. 2004;101(23):8587-8592. doi:10.1073/pnas.0403033101
232. Mujacic M, Baneyx F. Regulation of *Escherichia coli hch*A, a stress-inducible gene encoding molecular chaperone Hsp31. *Mol Microbiol*. 2006;60(6):1576-1589. doi:10.1111/j.1365-2958.2006.05207.x
233. Subedi KP, Choi D, Kim I, Min B, Park C. Hsp31 of *Escherichia coli* K-12 is glyoxalase III: Hsp31 is glyoxalase III. *Molecular Microbiology*. 2011;81(4):926-936. doi:10.1111/j.1365-2958.2011.07736.x
234. Kim HJ, Lee K-Y, Kwon A-R, Lee B-J. Structural and functional studies of SAV0551 from *Staphylococcus aureus* as a chaperone and glyoxalase III. *Bioscience Reports*. 2017;37(6):BSR20171106. doi:10.1042/BSR20171106
235. Quigley PM, Korotkov K, Baneyx F, Hol WGJ. The 1.6-Å crystal structure of the class of chaperones represented by *Escherichia coli* Hsp31 reveals a putative catalytic triad. *Proceedings of the National Academy of Sciences*. 2003;100(6):3137-3142. doi:10.1073/pnas.0530312100
236. Barrett AJ, Rawlings ND. Oligopeptidases, and the Emergence of the Prolyl Oligopeptidase Family. *Biological Chemistry Hoppe-Seyler*. 1992;373(2):353-360. doi:10.1515/bchm3.1992.373.2.353
237. Jones WM, Scaloni A, Manning JM. Acylaminoacyl-peptidase. *Methods in Enzymology*. 1994;244:227-231.
238. Ishikawa K, Ishida H, Koyama Y, et al. Acylamino Acid-releasing Enzyme from the Thermophilic Archaeon *Pyrococcus horikoshii*. *J Biol Chem*. 1998;273(28):17726-17731. doi:10.1074/jbc.273.28.17726
239. Kiss AL, Hornung B, Rádi K, et al. The Acylaminoacyl Peptidase from *Aeropyrum pernix* K1 Thought to Be an Exopeptidase Displays Endopeptidase Activity. *Journal of Molecular Biology*. 2007;368(2):509-520. doi:10.1016/j.jmb.2007.02.025
240. Szeltner Z, Kiss AL, Domokos K, Harmat V, Náray-Szabó G, Polgár L. Characterization of a novel acylaminoacyl peptidase with hexameric structure and endopeptidase activity. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*. 2009;1794(8):1204-1210. doi:10.1016/j.bbapap.2009.03.004
241. Bartlam M, Wang G, Yang H, et al. Crystal Structure of an Acylpeptide Hydrolase/Esterase from *Aeropyrum pernix* K1. *Structure*. 2004;12(8):1481-1488. doi:10.1016/j.str.2004.05.019
242. Bosch J, Tamura T, Tamura N, Baumeister W, Essen L-O. The β-propeller domain of the trilobed protease from *Pyrococcus furiosus* reveals an open Velcro topology. *Acta Crystallogr D Biol Crystallogr*. 2007;63(2):179-187. doi:10.1107/S0907444906045471
243. Menyhárd DK, Kiss-Szemán A, Tichy-Rács É, et al. A Self-compartmentalizing Hexamer Serine Protease from *Pyrococcus horikoshii*: SUBSTRATE SELECTION ACHIEVED THROUGH MULTIMERIZATION. *J Biol Chem*. 2013;288(24):17884-17894. doi:10.1074/jbc.M113.451534
244. Ito K, Nakajima Y, Xu Y, et al. Crystal Structure and Mechanism of Tripeptidyl Activity of Prolyl Tripeptidyl Aminopeptidase from *Porphyromonas gingivalis*. *Journal of Molecular Biology*. 2006;362(2):228-240. doi:10.1016/j.jmb.2006.06.083
245. Ellis-Guardiola K, Rui H, Beckner RL, et al. Crystal Structure and Conformational Dynamics of *Pyrococcus furiosus* Prolyl Oligopeptidase. *Biochemistry*. 2019;58(12):1616-1626. doi:10.1021/acs.biochem.9b00031
246. Fülöp V, Böcskei Z, Polgár L. Prolyl Oligopeptidase: An Unusual β-Propeller Domain Regulates Proteolysis. *Cell*. 1998;94:161-170.

247. Polgár L. The prolyl oligopeptidase family. *CMLS, Cell Mol Life Sci*. 2002;59(2):349-362. doi:10.1007/s00018-002-8427-5
248. Tsuru D, Yoshimoto T. Oligopeptidase B: Protease II from *Escherichia coli*. *Methods in Enzymology*. 1994;244:201-215.
249. McLuskey K, Paterson NG, Bland ND, Isaacs NW, Mottram JC. Crystal Structure of *Leishmania major* Oligopeptidase B Gives Insight into the Enzymatic Properties of a Trypanosomatid Virulence Factor. *J Biol Chem*. 2010;285(50):39249-39259. doi:10.1074/jbc.M110.156679
250. Tamura T, Tamura N, Cejka Z, Hegerl R, Lottspeich F, Baumeister W. Tricorn Protease--The Core of a Modular Proteolytic System. *Science*. 1996;274(5291):1385-1389. doi:10.1126/science.274.5291.1385
251. Tamura T, Tamura N, Lottspeich F, Baumeister W. Tricorn protease (TRI) interacting factor 1 from *Thermoplasma acidophilum* is a proline iminopeptidase. *FEBS Letters*. 1996;398(1):101-105. doi:10.1016/S0014-5793(96)01163-5
252. Tamura N, Lottspeich F, Baumeister W, Tamura T. The Role of Tricorn Protease and Its Aminopeptidase-Interacting Factors in Cellular Protein Degradation. *Cell*. 1998;95(5):637-648. doi:10.1016/S0092-8674(00)81634-7
253. Tamura N, Pfeifer G, Baumeister W, Tamura T. Tricorn Protease in Bacteria: Characterization of the Enzyme from *Streptomyces coelicolor*. *Biological Chemistry*. 2001;382(3). doi:10.1515/BC.2001.055
254. Brandstetter H, Kim J-S, Groll M, Huber R. Crystal structure of the tricorn protease reveals a protein disassembly line. *Nature*. 2001;414(6862):466-470. doi:10.1038/35106609
255. Brandstetter H, Kim J-S, Groll M, Göttig P, Huber R. Structural Basis for the Processive Protein Degradation by Tricorn Protease. *Biological Chemistry*. 2002;383(7-8). doi:10.1515/BC.2002.127
256. Walz J, Tamura T, Tamura N, Grimm R, Baumeister W, Koster AJ. Tricorn Protease Exists as an Icosahedral Supermolecule In Vivo. *Molecular Cell*. 1997;1(1):59-65. doi:10.1016/S1097-2765(00)80007-6
257. Walz J, Koster AJ, Tamura T, Baumeister W. Capsids of Tricorn Protease Studied by Electron Cryomicroscopy. *Journal of Structural Biology*. 1999;128(1):65-68. doi:10.1006/jsbi.1999.4169
258. Kim J-S, Groll M, Musiol H-J, et al. Navigation Inside a Protease: Substrate Selection and Product Exit in the Tricorn Protease from *Thermoplasma acidophilum*. *Journal of Molecular Biology*. 2002;324(5):1041-1050. doi:10.1016/S0022-2836(02)01153-1
259. Goettig P. Structures of the tricorn-interacting aminopeptidase F1 with different ligands explain its catalytic mechanism. *The EMBO Journal*. 2002;21(20):5343-5352. doi:10.1093/emboj/cdf552
260. Goettig P, Brandstetter H, Groll M, et al. X-ray Snapshots of Peptide Processing in Mutants of Tricorn-interacting Factor F1 from *Thermoplasma acidophilum*. *J Biol Chem*. 2005;280(39):33387-33396. doi:10.1074/jbc.M505030200
261. Ng'ong'a F, Nyanjom S, Wamunyokoli F. *In Silico* Analysis of Occurrence of Tricorn Protease and Its Homologs. *CBB*. 2017;5(3):27. doi:10.11648/j.cbb.20170503.11
262. Franzetti B, Schoehn G, Hernandez J-F, Jaquinod M, Ruigrok RWH, Zaccai G. Tetrahedral aminopeptidase: a novel large protease complex from archaea. *The EMBO Journal*. 2002;21(9):2132-2138. doi:10.1093/emboj/21.9.2132
263. Appolaire A, Colombo M, Basbous H, Gabel F, Girard E, Franzetti B. TET peptidases: A family of tetrahedral complexes conserved in prokaryotes. *Biochimie*. 2016;122:188-196. doi:10.1016/j.biochi.2015.11.001
264. Basbous H. Etudes structurales et propriétés enzymatiques de deux nouvelles aminopeptidases TETs auto-compartimentées chez les archées. Published online 2016.
265. Russo S, Baumann U. Crystal Structure of a Dodecameric Tetrahedral-shaped Aminopeptidase. *Journal of Biological Chemistry*. 2004;279(49):51275-51281. doi:10.1074/jbc.M409455200
266. Nguyen DD, Pandian R, Kim D, et al. Structural and kinetic bases for the metal preference of the M18 aminopeptidase from *Pseudomonas aeruginosa*. *Biochemical and Biophysical Research Communications*. 2014;447(1):101-107. doi:10.1016/j.bbrc.2014.03.109
267. Chaikuad A, Pilka ES, De Riso A, et al. Structure of human aspartyl aminopeptidase complexed with substrate analogue: insight into catalytic mechanism, substrate specificity and M18 peptidase family. *BMC Structural Biology*. 2012;12(1):14. doi:10.1186/1472-6807-12-14
268. Appolaire A, Durá MA, Ferruit M, et al. The TET2 and TET3 aminopeptidases from *Pyrococcus horikoshii* form a hetero-subunit peptidasome with enhanced peptide destruction properties: TET aminopeptidase multi-subunit complex. *Molecular Microbiology*. 2014;94(4):803-814. doi:10.1111/mmi.12775
269. Wilk S, Wilk E, Magnusson RP. Purification, Characterization, and Cloning of a Cytosolic Aspartyl Aminopeptidase. *J Biol Chem*. 1998;273(26):15961-15970. doi:10.1074/jbc.273.26.15961
270. Park S-Y, Scranton MA, Stajich JE, Yee A, Walling LL. Chlorophyte aspartyl aminopeptidases: Ancient origins, expanded families, new locations, and secondary functions. *PLoS ONE*. 2017;12(10):e0185492. doi:10.1371/journal.pone.0185492

271. Yokoyama R, Kawasaki H, Hirano H. Identification of yeast aspartyl aminopeptidase gene by purifying and characterizing its product from yeast cells. *FEBS Journal*. 2006;273(1):192-198. doi:10.1111/j.1742-4658.2005.05057.x
272. Cai WW, Wang L, Chen Y. Aspartyl Aminopeptidase, Encoded by an Evolutionarily Conserved Syntenic Gene, Is Colocalized with Its Cluster in Secretory Granules of Pancreatic Islet Cells. *Bioscience, Biotechnology, and Biochemistry*. 2010;74(10):2050-2055. doi:10.1271/bbb.100349
273. Chen Y, Farquhar ER, Chance MR, Palczewski K, Kiser PD. Insights into Substrate Specificity and Metal Activation of Mammalian Tetrahedral Aspartyl Aminopeptidase. *J Biol Chem*. 2012;287(16):13356-13370. doi:10.1074/jbc.M112.347518
274. Lee A, Slattery C, Nikolic-Paterson DJ, et al. Chloride channel ClC-5 binds to aspartyl aminopeptidase to regulate renal albumin endocytosis. *American Journal of Physiology-Renal Physiology*. 2015;308(7):F784-F792. doi:10.1152/ajprenal.00322.2014
275. Martínez-Martos JM, Carrera-González M del P, Dueñas B, Mayas MD, García MJ, Ramírez-Expósito MJ. Renin angiotensin system-regulating aminopeptidase activities in serum of pre- and postmenopausal women with breast cancer. *The Breast*. 2011;20(5):444-447. doi:10.1016/j.breast.2011.04.008
276. Mayas MD, Ramírez-Expósito MJ, Carrera MP, Cobo M, Martínez-Martos JM. Renin-angiotensin System-regulating Aminopeptidases in Tumor Growth of Rat C6 Gliomas Implanted at the Subcutaneous Region. *ANTICANCER RESEARCH*. Published online 2012:8.
277. Larrinaga G, Perez I, Ariz U, et al. Clinical impact of aspartyl aminopeptidase expression and activity in colorectal cancer. *Translational Research*. 2013;162(5):297-308. doi:10.1016/j.trsl.2013.07.010
278. Dalal S, Klemba M. Roles for Two Aminopeptidases in Vacuolar Hemoglobin Catabolism in *Plasmodium falciparum*. *J Biol Chem*. 2007;282(49):35978-35987. doi:10.1074/jbc.M703643200
279. Lauterbach SB, Coetzer TL. The M18 aspartyl aminopeptidase of *Plasmodium falciparum* binds to human erythrocyte spectrin in vitro. *Malar J*. 2008;7(1):161. doi:10.1186/1475-2875-7-161
280. Teuscher F, Lowther J, Skinner-Adams TS, et al. The M18 Aspartyl Aminopeptidase of the Human Malaria Parasite *Plasmodium falciparum*. *J Biol Chem*. 2007;282(42):30817-30826. doi:10.1074/jbc.M704938200
281. Sivaraman KK, Oellig CA, Huynh K, et al. X-ray Crystal Structure and Specificity of the *Plasmodium falciparum* Malaria Aminopeptidase PfM18AAP. *Journal of Molecular Biology*. 2012;422(4):495-507. doi:10.1016/j.jmb.2012.06.006
282. Zheng J, Cheng Z, Jia H, Zheng Y. Characterization of aspartyl aminopeptidase from *Toxoplasma gondii*. *Sci Rep*. 2016;6(1):34448. doi:10.1038/srep34448
283. Cueva R, García-Alvarez N, Suárez-Rendueles P. Yeast vacuolar aminopeptidase yscI: Isolation and regulation of the APE1 (LAP4) structural gene. *FEBS Letters*. 1989;259(1):125-129. doi:10.1016/0014-5793(89)81510-8
284. Klionsky DJ, Cueva R, Yaver DS. Aminopeptidase I of *Saccharomyces cerevisiae* is localized to the vacuole independent of the secretory pathway. *The Journal of Cell Biology*. 1992;119(2):287-299. doi:10.1083/jcb.119.2.287
285. Lynch-Day MA, Klionsky DJ. The Cvt pathway as a model for selective autophagy. *FEBS Letters*. 2010;584(7):1359-1366. doi:10.1016/j.febslet.2010.02.013
286. Kim J, Scott SV, Oda MN, Klionsky DJ. Transport of a Large Oligomeric Protein by the Cytoplasm to Vacuole Protein Targeting Pathway. *Journal of Cell Biology*. 1997;137(3):609-618. doi:10.1083/jcb.137.3.609
287. Shintani T, Klionsky DJ. Cargo Proteins Facilitate the Formation of Transport Vesicles in the Cytoplasm to Vacuole Targeting Pathway. *J Biol Chem*. 2004;279(29):29889-29894. doi:10.1074/jbc.M404399200
288. Shintani T, Huang W-P, Stromhaug PE, Klionsky DJ. Mechanism of Cargo Selection in the Cytoplasm to Vacuole Targeting Pathway. *Developmental Cell*. 2002;3(6):825-837. doi:10.1016/S1534-5807(02)00373-8
289. Scott SV, Guan J, Hutchins MU, Kim J, Klionsky DJ. Cvt19 Is a Receptor for the Cytoplasm-to-Vacuole Targeting Pathway. *Molecular Cell*. 2001;7(6):1131-1141. doi:10.1016/S1097-2765(01)00263-5
290. Chang C-Y, Huang W-P. Atg19 Mediates a Dual Interaction Cargo Sorting Mechanism in Selective Autophagy. *Molecular Biology of the Cell*. 2007;18:11.
291. Noda T, Kim J, Huang W-P, et al. Apg9p/Cvt7p Is an Integral Membrane Protein Required for Transport Vesicle Formation in the Cvt and Autophagy Pathways. *Journal of Cell Biology*. 2000;148(3):465-480. doi:10.1083/jcb.148.3.465
292. Baba M, Takeshige K, Baba N, Ohsumi Y. Ultrastructural Analysis of the Autophagic Process in Yeast: Detection of Autophagosomes and Their Characterization. *The Journal of Cell Biology*. 1994;124(6):903-913.
293. Seguí-Real B, Martinez M, Sandoval IV. Yeast aminopeptidase I is post-translationally sorted from the cytosol to the vacuole by a mechanism mediated by its bipartite N-terminal extension. *The EMBO Journal*. 1995;14(22):5476-5484. doi:10.1002/j.1460-2075.1995.tb00234.x
294. Yamasaki A, Watanabe Y, Adachi W, et al. Structural Basis for Receptor-Mediated Selective Autophagy of Aminopeptidase I Aggregates. *Cell Reports*. 2016;16:19-27.

295. Hutchins MU, Klionsky DJ. Vacuolar Localization of Oligomeric α-Mannosidase Requires the Cytoplasm to Vacuole Targeting and Autophagy Pathway Components in *Saccharomyces cerevisiae*. *J Biol Chem*. 2001;276(23):20491-20498. doi:10.1074/jbc.M101150200
296. Yuga M, Gomi K, Klionsky DJ, Shintani T. Aspartyl Aminopeptidase Is Imported from the Cytoplasm to the Vacuole by Selective Autophagy in *Saccharomyces cerevisiae*. *J Biol Chem*. 2011;286(15):13704-13713. doi:10.1074/jbc.M110.173906
297. Rawlings ND, Barrett AJ. Introduction: Metallopeptidases and Their Clans. In: *Handbook of Proteolytic Enzymes*. Elsevier; 2013:325-370. doi:10.1016/B978-0-12-382219-2.00077-6
298. Broder DH, Miller CG. DapE Can Function as an Aspartyl Peptidase in the Presence of Mn^{2+}. *Journal of Bacteriology*. 2003;185(16):4748-4754.
299. Nocek BP, Gillner DM, Fan Y, Holz RC, Joachimiak A. Structural Basis for Catalysis by the Mono- and Dimetalated Forms of the dapE-Encoded N-succinyl-l,l-Diaminopimelic Acid Desuccinylase. *Journal of Molecular Biology*. 2010;397(3):617-626. doi:10.1016/j.jmb.2010.01.062
300. Neuwald AF, Liu JS, Lipman DJ, Lawrence CE. Extracting protein alignment models from the sequence database. *Nucleic Acids Research*. 1997;25(9):1665-1677. doi:10.1093/nar/25.9.1665
301. Chevrier B, Schalk C, D'Orchymont H, Rondeau J-M, Moras D, Tarnus C. Crystal structure of *Aeromonas proteolytica* aminopeptidase: a prototypical member of the co-catalytic zinc enzyme family. *Structure*. 1994;2:283-291.
302. Bzymek KP, Holz RC. The Catalytic Role of Glutamate 151 in the Leucine Aminopeptidase from *Aeromonas proteolytica*. *J Biol Chem*. 2004;279(30):31018-31025. doi:10.1074/jbc.M404035200
303. Chevrier B, D'orchymont H, Schalk C, Tarnus C, Moras D. The Structure of the *Aeromonas proteolytica* Aminopeptidase Complexed with a Hydroxamate Inhibitor. Involvement in Catalysis of Glu151 and Two Zinc Ions of the Co-Catalytic Unit. *Eur J Biochem*. 1996;237(2):393-398. doi:10.1111/j.1432-1033.1996.0393k.x
304. De Paola CC, Bennett B, Holz RC, Ringe D, Petsko GA. 1-Butaneboronic Acid Binding to *Aeromonas proteolytica* Aminopeptidase: A Case of Arrested Development . *Biochemistry*. 1999;38(28):9048-9053. doi:10.1021/bi9900572
305. Desmarais W, Bienvenue DL, Bzymek KP, Petsko GA, Ringe D, Holz RC. The high-resolution structures of the neutral and the low pH crystals of aminopeptidase from *Aeromonas proteolytica*. *J Biol Inorg Chem*. 2006;11(4):398-408. doi:10.1007/s00775-006-0093-x
306. Schürer G, Lanig H, Clark T. *Aeromonas proteolytica* Aminopeptidase: An Investigation of the Mode of Action Using a Quantum Mechanical/Molecular Mechanical Approach. *Biochemistry*. 2004;43(18):5414-5427. doi:10.1021/bi0340191
307. Stamper C, Bennett B, Edwards T, Holz RC, Ringe D, Petsko G. Inhibition of the Aminopeptidase from *Aeromonas proteolytica* by L-Leucinephosphonic Acid. Spectroscopic and Crystallographic Characterization of the Transition State of Peptide Hydrolysis. *Biochemistry*. 2001;40(24):7035-7046. doi:10.1021/bi0100891
308. Chevrier B, D'Orchymont H. *Vibrio* Aminopeptidase. In: *Handbook of Proteolytic Enzymes*. Elsevier; 2013:1627-1630. doi:10.1016/B978-0-12-382219-2.00369-0
309. Greenblatt HM, Almog O, Maras B, et al. *Streptomyces griseus* aminopeptidase: X-ray crystallographic structure at 1.75 Å resolution. *Journal of Molecular Biology*. 1997;265(5):620-636. doi:10.1006/jmbi.1996.0729
310. Davis MI, Bennett MJ, Thomas LM, Bjorkman PJ. Crystal structure of prostate-specific membrane antigen, a tumor marker and peptidase. *Proceedings of the National Academy of Sciences*. 2005;102(17):5981_5986.
311. Hlouchova K, Barinka C, Konvalinka J, Lubkowski J. Structural insight into the evolutionary and pharmacologic homology of glutamate carboxypeptidases II and III. *FEBS Journal*. 2009;276(16):4448-4462. doi:10.1111/j.1742-4658.2009.07152.x
312. Håkansson K, Miller CG. Structure of peptidase T from *Salmonella typhimurium*: Structure of peptidase T. *European Journal of Biochemistry*. 2002;269(2):443-450. doi:10.1046/j.0014-2956.2001.02665.x
313. Martinez-Rodriguez S, Garcia-Pino A, Las Heras-Vazquez FJ, et al. Mutational and Structural Analysis of L-N-Carbamoylase Reveals New Insights into a Peptidase M20/M25/M40 Family Member. *Journal of Bacteriology*. 2012;194(21):5759-5768. doi:10.1128/JB.01056-12
314. Rowsell S, Pauptit RA, Tucker AD, Melton RG, Blow DM, Brick P. Crystal structure of carboxypeptidase G2, a bacterial enzyme with applications in cancer therapy. *Structure*. 1997;5(3):337-347. doi:10.1016/S0969-2126(97)00191-3
315. Unno H, Yamashita T, Ujita S, et al. Structural Basis for Substrate Recognition and Hydrolysis by Mouse Carnosinase CN2. *J Biol Chem*. 2008;283(40):27289-27299. doi:10.1074/jbc.M801657200
316. Botelho TO, Guevara T, Marrero A, et al. Structural and Functional Analyses Reveal That *Staphylococcus aureus* Antibiotic Resistance Factor HmrA Is a Zinc-dependent Endopeptidase. *J Biol Chem*. 2011;286(29):25697-25709. doi:10.1074/jbc.M111.247437
317. El-Gebali S, Mistry J, Bateman A, et al. The Pfam protein families database in 2019. *Nucleic Acids Research*. 2019;47(D1):D427-D432. doi:10.1093/nar/gky995

318. Appolaire A. Etude des grands assemblages protéolytiques de la famille TET: processus d'oligomérisation et régulation fonctionnelle associée. Published online 2014. https://tel.archives-ouvertes.fr/tel-01314147
319. Basbous H, Appolaire A, Girard E, Franzetti B. Characterization of a Glycyl-Specific TET Aminopeptidase Complex from *Pyrococcus horikoshii*. *Journal of Bacteriology*. 2018;200(17):e00059-18. doi:10.1128/JB.00059-18
320. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. *Bioinformatics*. 1992;8(3):275-282. doi:10.1093/bioinformatics/8.3.275
321. Kumar KSD, Gurusaran M, Satheesh SN, et al. *Online_DPI*: a web server to calculate the diffraction precision index for a protein structure. *Journal of Applied Crystallography*. 2015;48(3):939-942. doi:10.1107/S1600576715006287
322. Tamura K, Battistuzzi FU, Billing-Ross P, Murillo O, Filipski A, Kumar S. Estimating divergence times in large molecular phylogenies. *Proceedings of the National Academy of Sciences*. 2012;109(47):19333-19338. doi:10.1073/pnas.1213199109
323. Dutoit R, Brandt N, Legrain C, Bauvois C. Functional Characterization of Two M42 Aminopeptidases Erroneously Annotated as Cellulases. *PLoS ONE*. 2012;7(11):e50639. doi:10.1371/journal.pone.0050639
324. Durá MA, Receveur-Brechot V, Andrieu J-P, et al. Characterization of a TET-like Aminopeptidase Complex from the Hyperthermophilic Archaeon *Pyrococcus horikoshii*. *Biochemistry*. 2005;44(9):3477-3486. doi:10.1021/bi047736j
325. Durá MA, Rosenbaum E, Larabi A, Gabel F, Vellieux FMD, Franzetti B. The structural and biochemical characterizations of a novel TET peptidase complex from *Pyrococcus horikoshii* reveal an integrated peptide degradation system in hyperthermophilic Archaea: Characterization of *P. horikoshii* TET3 peptidase. *Molecular Microbiology*. 2009;72(1):26-40. doi:10.1111/j.1365-2958.2009.06600.x
326. Schoehn G, Vellieux FMD, Asunción Durá M, et al. An Archaeal Peptidase Assembles into Two Different Quaternary Structures: *A TETRAHEDRON AND A GIANT OCTAHEDRON*. *Journal of Biological Chemistry*. 2006;281(47):36327-36337. doi:10.1074/jbc.M604417200
327. Zheng Y, Roberts RJ, Kasif S, Guan C. Characterization of Two New Aminopeptidases in *Escherichia coli*. *Journal of Bacteriology*. 2005;187(11):3671-3677. doi:10.1128/JB.187.11.3671-3677.2005
328. Bacon CL, Jennings PV, Ni Fhaolain I, O'Cuinn G. Purification and characterisation of an aminopeptidase A from cytoplasm of *Lactococcus lactis* subsp. *cremoris* AM2. *International Dairy Journal*. 1994;4(6):503-519. doi:10.1016/0958-6946(94)90022-1
329. Durá MA, Franzetti B. PhTET1 Aminopeptidase. In: *Handbook of Proteolytic Enzymes*. Elsevier; 2013:1638-1645. doi:10.1016/B978-0-12-382219-2.00372-0
330. Kim D, San BH, Moh SH, et al. Structural basis for the substrate specificity of PepA from *Streptococcus pneumoniae*, a dodecameric tetrahedral protease. *Biochemical and Biophysical Research Communications*. 2010;391(1):431-436. doi:10.1016/j.bbrc.2009.11.075
331. Robinson MW, Buchtmann KA, Jenkins C, et al. MHJ_0125 is an M42 glutamyl aminopeptidase that moonlights as a multifunctional adhesin on the surface of *Mycoplasma hyopneumoniae*. *Open Biol*. 2013;3(4):130017. doi:10.1098/rsob.130017
332. Kumaki Y, Ogawa M, Hirano T, et al. Family M42 aminopeptidase from the syntrophic bacterium *Symbiobacterium thermophilum*: Characterization using recombinant protein. *Journal of Bioscience and Bioengineering*. 2011;111(2):134-139. doi:10.1016/j.jbiosc.2010.09.012
333. Story SV, Shah C, Jenney FE, Adams MWW. Characterization of a Novel Zinc-Containing, Lysine-Specific Aminopeptidase from the Hyperthermophilic Archaeon *Pyrococcus furiosus*. *Journal of Bacteriology*. 2005;187(6):2077-2083. doi:10.1128/JB.187.6.2077-2083.2005
334. Jia B, Lee S, Pham BP, et al. Biochemical Characterization of Deblocking Aminopeptidases from the Hyperthermophilic Archaeon *Thermococcus kodakarensis* KOD1. *Bioscience, Biotechnology, and Biochemistry*. 2011;75(6):1160-1166. doi:10.1271/bbb.110114
335. Lee HS, Cho Y, Kim YJ, Nam K, Lee J-H, Kang SG. Biochemical characterization of deblocking aminopeptidase from hyperthermophilic archaeon *Thermococcus onnurineus* NA1. *Journal of Bioscience and Bioengineering*. 2007;104(3):188-194. doi:10.1263/jbb.104.188
336. Lee YG, Leem S-H, Chung Y-H, Kim SI. Characterization of thermostable deblocking aminopeptidases of archaeon *Thermococcus onnurineus* NA1 by proteomic and biochemical approaches. *J Microbiol*. 2012;50(5):792-797. doi:10.1007/s12275-012-2461-2
337. Slutskaya ES, Bezsudnova EYu, Mardanov AV, et al. Characterization of a novel M42 aminopeptidase from crenarchaeon *Desulfurococcus kamchatkensis*. *Dokl Biochem Biophys*. 2012;442(1):30-32. doi:10.1134/S1607672912010097
338. Russo S. Structural and Functional Studies on the Dodecameric Tetrahedral-shaped Aminopeptidase FrvX from *Pyrococcus horikoshii*. Published online 2005. http://biblio.unibe.ch/download/eldiss/05russo_s.pdf
339. Ando S, Ishikawa K, Ishida H, Kawarabayasi Y, Kikuchi H, Kosugi Y. Thermostable aminopeptidase from *Pyrococcus horikoshii*. *FEBS Letters*. 1999;447(1):25-28. doi:10.1016/S0014-5793(99)00257-4

340. Auld DS. Catalytic Mechanisms for Metallopeptidases. In: *Handbook of Proteolytic Enzymes*. Elsevier; 2013:370-396. doi:10.1016/B978-0-12-382219-2.00078-8
341. Prescott JM, Wilkes SH, Wagner FW, Wilson KJ. *Aeromonas* Aminopeptidase: Improved Isolation and Some Physical Properties. *Journal of Biological Chemistry*. 1971;246(6):1756-1764.
342. Wagner FW, Wilkes SH, Prescott JM. Specificity of *Aeromonas* Aminopeptidase toward Amino Acid Amides and Dipeptides. *Journal of Biological Chemistry*. 1972;247(4):1208-1210.
343. Wilkes SH, Bayliss ME, Prescott JM. Specificity of *Aeromonas* Aminopeptidase toward Oligopeptides and Polypeptides. *Eur J Biochem*. 1973;34(3):459-466. doi:10.1111/j.1432-1033.1973.tb02780.x
344. Prescott JM, Wagner FW, Holmquist B, Vallee BL. One Hundred Fold Increased Activity of *Aeromonas* Aminopeptidase by Sequential Substitutions with Ni(II) or Cu(II) Followed by Zinc. *Biochemical and Biophysical Research Communications*. 1983;114(2):646-652.
345. Prescott JM, Wagner FW, Holmquist B, Vallee BL. Spectral and kinetic studies of metal-substituted *Aeromonas* aminopeptidase: nonidentical, interacting metal-binding sites. *Biochemistry*. 1985;24(20):5350-5356. doi:10.1021/bi00341a012
346. Bayliss ME, Prescott JM. Modified Activity of *Aeromonas* Aminopeptidase: Metal Ion Substitutions and Role of Substrates. *Biochemistry*. 1986;25:8113-8117.
347. Bennett B, Holz RC. EPR Studies on the Mono- and Dicobalt(II)-Substituted Forms of the Aminopeptidase from *Aeromonas proteolytica* . Insight into the Catalytic Mechanism of Dinuclear Hydrolases. *J Am Chem Soc*. 1997;119(8):1923-1933. doi:10.1021/ja963021v
348. Bennett B, Holz RC. Spectroscopically Distinct Cobalt(II) Sites in Heterodimetallic Forms of the Aminopeptidase from *Aeromonas proteolytica*: Characterization of Substrate Binding. *Biochemistry*. 1997;36(32):9837-9846. doi:10.1021/bi970735p
349. Bennett B, Holz RC. Inhibition of the Aminopeptidase from *Aeromonas proteolytica* by L-Leucinephosphonic Acid, a Transition State Analogue of Peptide Hydrolysis. *J Am Chem Soc*. 1998;120(46):12139-12140. doi:10.1021/ja9824167
350. Holz RC, Bennett B, Chen G, Ming L-J. Proton NMR Spectroscopy as a Probe of Dinuclear Copper(II) Active Sites in Metalloproteins. Characterization of the Hyperactive Copper(II)-Substituted Aminopeptidase from *Aeromonas proteolytica*. *J Am Chem Soc*. 1998;120(25):6329-6335. doi:10.1021/ja971589d
351. Ustynyuk L, Bennett B, Edwards T, Holz RC. Inhibition of the Aminopeptidase from *Aeromonas proteolytica* by Aliphatic Alcohols. Characterization of the Hydrophobic Substrate Recognition Site. *Biochemistry*. 1999;38(35):11433-11439. doi:10.1021/bi991090r
352. Huntington KM, Bienvenue DL, Wei Y, Bennett B, Holz RC, Pei D. Slow-Binding Inhibition of the Aminopeptidase from *Aeromonas proteolytica* by Peptide Thiols: Synthesis and Spectroscopic Characterization. *Biochemistry*. 1999;38(47):15587-15596. doi:10.1021/bi991283e
353. Bienvenue DL, Bennett B, Holz RC. Inhibition of the aminopeptidase from *Aeromonas proteolytica* by l-leucinethiol: kinetic and spectroscopic characterization of a slow, tight-binding inhibitor–enzyme complex. *Journal of Inorganic Biochemistry*. 2000;78(1):43-54. doi:10.1016/S0162-0134(99)00203-2
354. Bienvenue DL, Mathew RS, Ringe D, Holz RC. The aminopeptidase from *Aeromonas proteolytica* can function as an esterase. *J Biol Inorg Chem*. 2002;7(1-2):129-135. doi:10.1007/s007750100280
355. Bzymek KP, D'Souza VM, Chen G, Campbell H, Mitchell A, Holz RC. Function of the signal peptide and N- and C-terminal propeptides in the leucine aminopeptidase from *Aeromonas proteolytica*. *Protein Expression and Purification*. 2004;37(2):294-305. doi:10.1016/j.pep.2004.05.004
356. Bzymek KP, Moulin A, Swierczek SI, et al. Kinetic, Spectroscopic, and X-ray Crystallographic Characterization of the Functional E151H Aminopeptidase from *Aeromonas proteolytica*. *Biochemistry*. 2005;44(36):12030-12040. doi:10.1021/bi0505823
357. Munih P, Moulin A, Stamper CC, et al. X-ray crystallographic characterization of the Co(II)-substituted Tris-bound form of the aminopeptidase from *Aeromonas proteolytica*. *Journal of Inorganic Biochemistry*. 2007;101(8):1099-1107. doi:10.1016/j.jinorgbio.2007.03.010
358. Stamper CC, Bienvenue DL, Bennett B, Ringe D, Petsko GA, Holz RC. Spectroscopic and X-ray Crystallographic Characterization of Bestatin Bound to the Aminopeptidase from *Aeromonas (Vibrio) proteolytica*. *Biochemistry*. 2004;43(30):9620-9628. doi:10.1021/bi049126p
359. Kumar A, Periyannan GR, Narayanan B, Kittell AW, Kim J-J, Bennett B. Experimental evidence for a metallohydrolase mechanism in which the nucleophile is not delivered by a metal ion: EPR spectrokinetic and structural studies of aminopeptidase from *Vibrio proteolyticus*. *Biochemical Journal*. 2007;403(3):527-536. doi:10.1042/BJ20061591
360. Hartley M, Bennett B. Heterologous expression and purification of *Vibrio proteolyticus* (*Aeromonas proteolytica*) aminopeptidase: A rapid protocol. *Protein Expression and Purification*. 2009;66(1):91-101. doi:10.1016/j.pep.2009.02.011

361. Hanaya K, Suetsugu M, Saijo S, Yamato I, Aoki S. Potent inhibition of dinuclear zinc(II) peptidase, an aminopeptidase from *Aeromonas proteolytica*, by 8-quinolinol derivatives: inhibitor design based on Zn2+ fluorophores, kinetic, and X-ray crystallographic study. *J Biol Inorg Chem*. 2012;17(4):517-529. doi:10.1007/s00775-012-0873-4
362. Chen G, Edwards T, D'souz VM, Holz RC. Mechanistic Studies on the Aminopeptidase from *Aeromonas proteolytica* : A Two-Metal Ion Mechanism for Peptide Hydrolysis. *Biochemistry*. 1997;36(14):4278-4286. doi:10.1021/bi9618676
363. Holz RC. The aminopeptidase from *Aeromonas proteolytica*: structure and mechanism of co-catalytic metal centers involved in peptide hydrolysis. *Coordination Chemistry Reviews*. 2002;232(1-2):5-26. doi:10.1016/S0010-8545(01)00470-2
364. Cosper NJ, D'souza VM, Scott RA, Holz RC. Structural Evidence That the Methionyl Aminopeptidase from *Escherichia coli* Is a Mononuclear Metalloprotease. *Biochemistry*. 2001;40(44):13302-13309. doi:10.1021/bi010837m
365. Colombo M, Girard E, Franzetti B. Tuned by metals: the TET peptidase activity is controlled by 3 metal binding sites. *Scientific Reports*. 2016;6(1). doi:10.1038/srep20876
366. Onoe S, Ando S, Ataka M, Ishikawa K. Active Site of Deblocking Aminopeptidase from *Pyrococcus horikoshii*. *Biochemical and Biophysical Research Communications*. 2002;290(3):994-997. doi:10.1006/bbrc.2001.6327
367. Mori K, Ishikawa K. New Deblocking Aminopeptidases from *Pyrococcus horikoshii*. *Bioscience, Biotechnology, and Biochemistry*. 2005;69(10):1854-1860. doi:10.1271/bbb.69.1854
368. Zuber H, Roncari G. Thermophilic and Mesophilic Aminopeptidases from *Bacillus stearothermophilus*. *Angew Chem Int Ed Engl*. 1967;6(10):880-881. doi:10.1002/anie.196708801
369. Roncari G, Zuber H. THERMOPHILIC AMINOPEPTIDASES FROM BACILLUS STEAROTHERMOPHILUS. I. ISOLATION, SPECIFICITY, AND GENERAL PROPERTIES OF THE THERMOSTABLE AMINOPEPTIDASE I. *International Journal of Protein Research*. 1968;1(1-4):45-61. doi:10.1111/j.1399-3011.1969.tb01625.x
370. Exterkate FA, de Veer GJCM. Purification and Some Properties of a Membrane-Bound Aminopeptidase A from *Streptococcus cremoris*. *Applied and Environmental Microbiology*. 1987;53(3):577-583. doi:10.1128/AEM.53.3.577-583.1987
371. Moser P, Roncari G, Zuber H. THERMOPHILIC AMINOPEPTIDASES FROM BAC.STEAROTHERMOPHILUS: II. Aminopeptidase I (AP I): physico-chemical properties; thermostability and activation; formation of the apoenzyme and subunits. *International Journal of Protein Research*. 1970;2(1-4):191-207. doi:10.1111/j.1399-3011.1970.tb01677.x
372. Roncari G, Zuber H, Wyttenbach A. THERMOPHILIC AMINOPEPTIDASES FROM BAC. STEAROTHERMOPHILUS.: III. Determination of the Cobalt and Zinc Content in Aminopeptidase I by Neutron Activation Analysis. *International Journal of Peptide and Protein Research*. 1972;4(4):267-271. doi:10.1111/j.1399-3011.1972.tb03428.x
373. Irving H, Williams RJP. Order of Stability of Metal Complexes. *Nature*. 1948;162(4123):746-747.
374. Williams RJP. METAL IONS IN BIOLOGICAL SYSTEMS. *Biological Reviews*. 1953;28(4):381-412. doi:10.1111/j.1469-185X.1953.tb01384.x
375. Johnson DA, Nelson PG. Factors Determining the Ligand Field Stabilization Energies of the Hexaaqua 2+ Complexes of the First Transition Series and the Irving-Williams Order. *Inorg Chem*. 1995;34(22):5666-5671. doi:10.1021/ic00126a041
376. Foster AW, Osman D, Robinson NJ. Metal Preferences and Metallation. *J Biol Chem*. 2014;289(41):28095-28103. doi:10.1074/jbc.R114.588145
377. Waldron KJ, Robinson NJ. How do bacterial cells ensure that metalloproteins get the correct metal? *Nat Rev Microbiol*. 2009;7(1):25-35. doi:10.1038/nrmicro2057
378. Petrova TE, Slutskaya ES, Boyko KM, et al. Structure of the dodecamer of the aminopeptidase APDkam598 from the archaeon *Desulfurococcus kamchatkensis*. *Acta Crystallographica Section F Structural Biology Communications*. 2015;71(3):277-285. doi:10.1107/S2053230X15000783
379. Yao S, Flight RM, Rouchka EC, Moseley HNB. A less-biased analysis of metalloproteins reveals novel zinc coordination geometries: Novel Zinc Coordination Geometries Revealed. *Proteins*. 2015;83(8):1470-1487. doi:10.1002/prot.24834
380. D'souza VM, Holz RC. The Methionyl Aminopeptidase from *Escherichia coli* Can Function as an Iron(II) Enzyme. *Biochemistry*. 1999;38(34):11079-11085. doi:10.1021/bi990872h
381. Cameron V, House CH, Brantley SL. A First Analysis of Metallome Biosignatures of Hyperthermophilic Archaea. *Archaea*. 2012;2012:1-12. doi:10.1155/2012/789278
382. Rosenbaum E, Ferruit M, Durá MA, Franzetti B. Studies on the parameters controlling the stability of the TET peptidase superstructure from *Pyrococcus horikoshii* revealed a crucial role of pH and catalytic metals in the oligomerization process. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*. 2011;1814(10):1289-1294. doi:10.1016/j.bbapap.2010.11.008

383. Reichle RA, McCurdy KG, Hepler LG. Zinc Hydroxide: Solubility Product and Hydroxy-complex Stability Constants from 12.5–75 °C. *Can J Chem*. 1975;53(24):3841-3845. doi:10.1139/v75-556
384. Larsen KS, Auld DS. Characterization of an inhibitory metal binding site in carboxypeptidase A. *Biochemistry*. 1991;30(10):2613-2618. doi:10.1021/bi00224a007
385. Lowther WT, Matthews BW. Metalloaminopeptidases: Common Functional Themes in Disparate Structural Surroundings. *Chem Rev*. 2002;102(12):4581-4608. doi:10.1021/cr0101757
386. Tsunasawa S. Purification and application of a novel N-terminal deblocking aminopeptidase (DAP) from *Pyrococcus furiosus*. *J Protein Chem*. 1998;17(6):521-522.
387. Kamp RM, Tsunasawa S, Hirano H. Application of new deblocking aminopeptidase from *Pyrococcus furiosus* for microsequence analysis of blocked proteins. *J Protein Chem*. 1998;17(6):512-513.
388. Balerna M, Zuber H. THERMOPHILIC AMINOPEPTIDASE FROM BACILLUS STEAROTHERMOPHILUS.: IV. Aminopeptidases (APIm, APIII) in mesophilic (37°C) cultures of the thermophilic bacterium *B. stearothermophilus*. *International Journal of Peptide and Protein Research*. 1974;6(6):499-514. doi:10.1111/j.1399-3011.1974.tb02408.x
389. Kobayashi T, Romaniec MPM, Barker P, Gerngross U, Demain A. Nucleotide Sequence of Gene celM Encoding a New Endoglucanase (CelM) of *Clostridium thermocellum* and Purification of the Enzyme. *Journal of Fermentation and Bioengineering*. 1993;76(4):251-256.
390. Lo Leggio L, Larsen S. The 1.62 Å structure of *Thermoascus aurantiacus* endoglucanase: completing the structural picture of subfamilies in glycoside hydrolase family 5. *FEBS Letters*. 2002;523(1-3):103-108. doi:10.1016/S0014-5793(02)02954-X
391. Nakamura A, Tsukada T, Auer S, et al. The Tryptophan Residue at the Active Site Tunnel Entrance of *Trichoderma reesei* Cellobiohydrolase Cel7A Is Important for Initiation of Degradation of Crystalline Cellulose. *J Biol Chem*. 2013;288(19):13503-13510. doi:10.1074/jbc.M113.452623
392. Zhang X, Wang S, Wu X, et al. Subsite-specific contributions of different aromatic residues in the active site architecture of glycoside hydrolase family 12. *Sci Rep*. 2015;5(1):18357. doi:10.1038/srep18357
393. Henrissat B, Davies GJ. Structural and sequence-based classification of glycoside hydrolases. *Current Opinion in Structural Biology*. 1997;7:637-644.
394. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*. 2014;42(D1):D490-D495. doi:10.1093/nar/gkt1178
395. Maiti S, Samanta T, Sahoo S, Roy S. The Dual Carboxymethyl Cellulase and Gelatinase Activities of a Newly Isolated Protein from *Brevibacillus agri* ST15c10 Confer Reciprocal Regulations in Substrate Utilization. *J Mol Microbiol Biotechnol*. 2017;27(6):319-331. doi:10.1159/000479109
396. Sharma D, Sharma P, Dev K, Sourirajan A. Endoglucanase gene of M42 aminopeptidase/endoglucanase family from thermophilic *Bacillus* sp. PW1 and PW2 isolated from Tattapani hot spring, Himachal Pradesh, India. *J Genet Eng Biotechnol*. 2019;17(1):4. doi:10.1186/s43141-019-0001-8
397. Ma L, Zhao Y, Meng L, et al. Isolation of Thermostable Lignocellulosic Bacteria From Chicken Manure Compost and a M42 Family Endocellulase Cloning From *Geobacillus thermodenitrificans* Y7. *Front Microbiol*. 2020;11:281. doi:10.3389/fmicb.2020.00281
398. Knott BC, Haddad Momeni M, Crowley MF, et al. The Mechanism of Cellulose Hydrolysis by a Two-Step, Retaining Cellobiohydrolase Elucidated by Structural and Transition Path Sampling Studies. *J Am Chem Soc*. 2014;136(1):321-329. doi:10.1021/ja410291u
399. Sandgren M, Wu M, Karkehabadi S, et al. The Structure of a Bacterial Cellobiohydrolase: The Catalytic Core of the *Thermobifida fusca* Family GH6 Cellobiohydrolase Cel6B. *Journal of Molecular Biology*. 2013;425(3):622-635. doi:10.1016/j.jmb.2012.11.039
400. Lafond M, Navarro D, Haon M, Couturier M, Berrin J-G. Characterization of a Broad-Specificity β-Glucanase Acting on β-(1,3)-, β-(1,4)-, and β-(1,6)-Glucans That Defines a New Glycoside Hydrolase Family. *Appl Environ Microbiol*. 2012;78(24):8540-8546. doi:10.1128/AEM.02572-12
401. Miyazaki T, Yoshida M, Tamura M, et al. Crystal structure of the N-terminal domain of a glycoside hydrolase family 131 protein from *Coprinopsis cinerea*. *FEBS Letters*. 2013;587(14):2193-2198. doi:10.1016/j.febslet.2013.05.041
402. Ghose TK. MEASUREMENT OF CELLULASE ACTIVITIES. *Pure and Applied Chemistry*. 1987;59(2):257-268.
403. Dutoit R, Delsaute M, Collet L, et al. Crystal structure determination of *Pseudomonas stutzeri* A1501 endoglucanase Cel5A: the search for a molecular basis for glycosynthesis in GH5_5 enzymes. *Acta Crystallogr D Struct Biol*. 2019;75(6):605-615. doi:10.1107/S2059798319007113
404. Moracci M, Trincone A, Perugino G, Ciaramella M, Rossi M. Restoration of the Activity of Active-Site Mutants of the Hyperthermophilic β-Glycosidase from *Sulfolobus solfataricus*: Dependence of the Mechanism on the Action of External Nucleophiles. *Biochemistry*. 1998;37(49):17262-17270. doi:10.1021/bi981855f
405. Wilkes SH, Prescott JM. The Slow, Tight Binding of Bestatin and Amastatin to Aminopeptidases. *Journal of Biological Chemistry*. 1985;260(24):13154-13162.

406. Wong AHM, Zhou D, Rini JM. The X-ray Crystal Structure of Human Aminopeptidase N Reveals a Novel Dimer and the Basis for Peptide Processing. *J Biol Chem*. 2012;287(44):36804-36813. doi:10.1074/jbc.M112.398842
407. Yang Y, Liu C, Lin Y-L, Li F. Structural Insights into Central Hypertension Regulation by Human Aminopeptidase A. *J Biol Chem*. 2013;288(35):25638-25645. doi:10.1074/jbc.M113.494955
408. Harding JW, Felix D. The effects of the aminopeptidase inhibitors amastatin and bestatin on angiotensin-evoked neuronal activity in rat brain. *Brain Research*. 1987;424(2):299-304. doi:10.1016/0006-8993(87)91474-0
409. Rich DH, Moon BJ, Harbeson S. Inhibition of Aminopeptidases by Amastatin and Bestatin Derivatives. Effect of Inhibitor Structure on Slow-Binding Processes1p2. *Journal of Medicinal Chemistry*. 1984;27(4):417-422.
410. Ivarsson Y. Plasticity of PDZ domains in ligand recognition and signaling. *FEBS Letters*. 2012;586(17):2638-2647. doi:10.1016/j.febslet.2012.04.015
411. Luck K, Charbonnier S, Travé G. The emerging contribution of sequence context to the specificity of protein interactions mediated by PDZ domains. *FEBS Letters*. 2012;586(17):2648-2661. doi:10.1016/j.febslet.2012.03.056
412. Sheng M, Sala C. PDZ Domains and the Organization of Supramolecular Complexes. *Annu Rev Neurosci*. 2001;24(1):1-29. doi:10.1146/annurev.neuro.24.1.1
413. Kennedy M. The origin of PDZ (DHR, UF) domain. *Trends in Biochemical Sciences*. 1995;20(9):350-350.
414. Funke L, Dakoji S, Bredt DS. MEMBRANE-ASSOCIATED GUANYLATE KINASES REGULATE ADHESION AND PLASTICITY AT CELL JUNCTIONS. *Annu Rev Biochem*. 2005;74(1):219-245. doi:10.1146/annurev.biochem.74.082803.133339
415. Ernst A, Appleton BA, Ivarsson Y, et al. A Structural Portrait of the PDZ Domain Family. *Journal of Molecular Biology*. 2014;426(21):3509-3519. doi:10.1016/j.jmb.2014.08.012
416. Muley VY, Akhter Y, Galande S. PDZ Domains Across the Microbial World: Molecular Link to the Proteases, Stress Response, and Protein Synthesis. *Genome Biology and Evolution*. 2019;11(3):644-659. doi:10.1093/gbe/evz023
417. Iwanczyk J, Damjanovic D, Kooistra J, et al. Role of the PDZ Domains in *Escherichia coli* DegP Protein. *JB*. 2007;189(8):3176-3186. doi:10.1128/JB.01788-06
418. Runyon ST, Zhang Y, Appleton BA, et al. Structural and functional analysis of the PDZ domains of human HtrA1 and HtrA3. *Protein Sci*. 2007;16(11):2454-2471. doi:10.1110/ps.073049407
419. Appolaire A, Rosenbaum E, Durá MA, et al. *Pyrococcus horikoshii* TET2 Peptidase Assembling Process and Associated Functional Regulation. *Journal of Biological Chemistry*. 2013;288(31):22542-22554. doi:10.1074/jbc.M113.450189
420. Macek P, Kerfah R, Erba EB, et al. Unraveling self-assembly pathways of the 468-kDa proteolytic machine TET2. *Science Advances*. 2017;3(4):e1601601. doi:10.1126/sciadv.1601601
421. Krissinel E, Henrick K. Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology*. 2007;372(3):774-797. doi:10.1016/j.jmb.2007.05.022
422. Rosenbaum E, Gabel F, Durá MA, et al. Effects of hydrostatic pressure on the quaternary structure and enzymatic activity of a large peptidase complex from *Pyrococcus horikoshii*. *Archives of Biochemistry and Biophysics*. 2012;517(2):104-110. doi:10.1016/j.abb.2011.07.017
423. Stoll E, Hermodson MA, Ericsson LH, Zuber H. Subunit structure of the thermophilic aminopeptidase I from *Bacillus stearothermophilus*. *Biochemistry*. 1972;11(25):4731-4735. doi:10.1021/bi00775a015
424. Stoll E, Zuber H. Interconversion of the different hybrids of aminopeptidase I. *FEBS Letters*. 1974;40(1):210-212. doi:10.1016/0014-5793(74)80930-0
425. Slutskaya E, Artemova N, Kleymenov S, Petrova T, Popov V. Heat-induced conformational changes of TET peptidase from crenarchaeon *Desulfurococcus kamchatkensis*. *Eur Biophys J*. 2015;44(8):667-675. doi:10.1007/s00249-015-1064-3
426. Appolaire A, Girard E, Colombo M, et al. Small-angle neutron scattering reveals the assembly mode and oligomeric architecture of TET, a large, dodecameric aminopeptidase. *Acta Crystallographica Section D Biological Crystallography*. 2014;70(11):2983-2993. doi:10.1107/S1399004714018446
427. Gauto DF, Estrozi LF, Schwieters CD, et al. Integrated NMR and cryo-EM atomic-resolution structure determination of a half-megadalton enzyme complex. *Nat Commun*. 2019;10(1):2697. doi:10.1038/s41467-019-10490-9
428. Porciero S, Receveur-Bréchot V, Mori K, Franzetti B, Roussel A. Expression, purification, crystallization and preliminary crystallographic analysis of a deblocking aminopeptidase from *Pyrococcus horikoshii*. *Acta Crystallogr F Struct Biol Cryst Commun*. 2005;61(2):239-242. doi:10.1107/S1744309105001910
429. Burgess RR. A brief practical review of size exclusion chromatography: Rules of thumb, limitations, and troubleshooting. *Protein Expression and Purification*. 2018;150:81-85. doi:10.1016/j.pep.2018.05.007
430. Guo MS, Gross CA. Stress-Induced Remodeling of the Bacterial Proteome. *Current Biology*. 2014;24(10):R424-R434. doi:10.1016/j.cub.2014.03.023

431. Meyer AS, Baker TA. Proteolysis in the *Escherichia coli* heat shock response: a player at many levels. *Current Opinion in Microbiology*. 2011;14(2):194-199. doi:10.1016/j.mib.2011.02.001
432. Stoll E, Ericsson LH, Zuber H. The Function of the Two Subunits of Thermophilic Aminopeptidase I. *Proceedings of the National Academy of Sciences*. 1973;70(12):3781-3784. doi:10.1073/pnas.70.12.3781
433. Bendtsen JD, Kiemer L, Fausbøll A, Brunak S. Non-classical protein secretion in bacteria. *BMC Microbiol*. 2005;5(1):58. doi:10.1186/1471-2180-5-58
434. Maffei B, Francetic O, Subtil A. Tracking Proteins Secreted by Bacteria: What's in the Toolbox? *Front Cell Infect Microbiol*. 2017;7:221. doi:10.3389/fcimb.2017.00221
435. Exterkate FA. Location of Peptidases Outside and Inside the Membrane of *Streptococcus cremoris*. *Applied and Environmental Microbiology*. 1984;47(1):177-183. doi:10.1128/AEM.47.1.177-183.1984
436. Cannan RK, Kibrick A. Complex Formation between Carboxylic Acids and Divalent Metal Cations. *Journal of the American Chemical Society*. 1938;60(10):2314-2320. doi:10.1021/ja01277a012
437. Garman EF. Radiation damage in macromolecular crystallography: what is it and why should we care? *Acta Crystallogr D Biol Crystallogr*. 2010;66(4):339-351. doi:10.1107/S0907444910008656
438. Taberman H. Radiation Damage in Macromolecular Crystallography—An Experimentalist's View. *Crystals*. 2018;8(4):157. doi:10.3390/cryst8040157
439. Zhang Y, Werling U, Edelmann W. SLiCE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Research*. 2012;40(8):e55-e55. doi:10.1093/nar/gkr1288
440. Edelheit O, Hanukoglu A, Hanukoglu I. Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies. *BMC Biotechnology*. 2009;9(1):61. doi:10.1186/1472-6750-9-61
441. Tuppy H, Wiesbauer U, Wintersberger E. Aminosäure-*p*-nitroanilide als Substrate für Aminopeptidasen und andere proteolytische Fermente. *Hoppe-Seyler´s Zeitschrift für physiologische Chemie*. 1962;329(Jahresband):278-288. doi:10.1515/bchm2.1962.329.1.278
442. Kabsch W. *XDS*. *Acta Crystallographica Section D Biological Crystallography*. 2010;66(2):125-132. doi:10.1107/S0907444909047337
443. Karplus PA, Diederichs K. Assessing and maximizing data quality in macromolecular crystallography. *Current Opinion in Structural Biology*. 2015;34:60-68. doi:10.1016/j.sbi.2015.07.003
444. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. *Phaser* crystallographic software. *J Appl Crystallogr*. 2007;40(4):658-674. doi:10.1107/S0021889807021206
445. Adams PD, Afonine PV, Bunkóczi G, et al. *PHENIX*: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography*. 2010;66(2):213-221. doi:10.1107/S0907444909052925
446. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of *Coot*. *Acta Crystallographica Section D Biological Crystallography*. 2010;66(4):486-501. doi:10.1107/S0907444910007493
447. Yeates TO. [22] Detecting and overcoming crystal twinning. In: *Methods in Enzymology*. Vol 276. Elsevier; 1997:344-358. doi:10.1016/S0076-6879(97)76068-3
448. Campeotto I, Lebedev A, Schreurs AMM, et al. Pathological macromolecular crystallographic data affected by twinning, partial-disorder and exhibiting multiple lattices for testing of data processing and refinement tools. *Sci Rep*. 2018;8(1):14876. doi:10.1038/s41598-018-32962-6
449. Zwart PH, Grosse-Kunstleve RW, Lebedev AA, Murshudov GN, Adams PD. Surprises and pitfalls arising from (pseudo)symmetry. *Acta Crystallogr D Biol Crystallogr*. 2008;64(1):99-107. doi:10.1107/S090744490705531X
450. Lesley SA, Kuhn P, Godzik A, et al. Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. *Proceedings of the National Academy of Sciences*. 2002;99(18):11664-11669. doi:10.1073/pnas.142413399
451. Schmidt TG, Skerra A. The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nat Protoc*. 2007;2(6):1528-1535. doi:10.1038/nprot.2007.209
452. Tchieu JH, Norris V, Edwards JS, Saier MH. The Complete Phosphotransferase System in *Escherichia coli*. *J Mol Microbio Biotechnol*. 2001;3(3):329-346.
453. Are VN, Kumar A, Goyal VD, et al. Structures and activities of widely conserved small prokaryotic Aminopeptidases P clarify classification of M24B peptidases. *Proteins*. 2018:prot.25641. doi:10.1002/prot.25641
454. Deutscher J, Aké FMD, Derkaoui M, et al. The Bacterial Phosphoenolpyruvate:Carbohydrate Phosphotransferase System: Regulation by Protein Phosphorylation and Phosphorylation-Dependent Protein-Protein Interactions. *Microbiol Mol Biol Rev*. 2014;78(2):231-256. doi:10.1128/MMBR.00001-14
455. Deutscher J, Francke C, Postma PW. How Phosphotransferase System-Related Protein Phosphorylation Regulates Carbohydrate Metabolism in Bacteria. *MMBR*. 2006;70(4):939-1031. doi:10.1128/MMBR.00024-06
456. Poncet S, Soret M, Mervelet P, Deutscher J, Noirot P. Transcriptional Activator YesS Is Stimulated by Histidine-phosphorylated HPr of the *Bacillus subtilis* Phosphotransferase System. *J Biol Chem*. 2009;284(41):28188-28197. doi:10.1074/jbc.M109.046334

457. Schumacher MA, Allen GS, Diel M, Seidel G, Hillen W, Brennan RG. Structural Basis for Allosteric Control of the Transcription Regulator CcpA by the Phosphoprotein HPr-Ser46-P. *Cell*. 2004;118(6):731-741. doi:10.1016/j.cell.2004.08.027
458. Kravanja M, Engelmann R, Dossonnet V, et al. The *hpr*K gene of *Enterococcus faecalis* encodes a novel bifunctional enzyme: the HPr kinase/phosphatase. *Mol Microbiol*. 1999;31(1):59-66. doi:10.1046/j.1365-2958.1999.01146.x
459. Pomposiello PJ, Koutsolioutsou A, Carrasco D, Demple B. SoxRS-Regulated Expression and Genetic Analysis of the *ygg*X Gene of *Escherichia coli*. *JB*. 2003;185(22):6624-6632. doi:10.1128/JB.185.22.6624-6632.2003
460. Blanchard JL, Wholey W-Y, Conlon EM, Pomposiello PJ. Rapid Changes in Gene Expression Dynamics in Response to Superoxide Reveal SoxRS-Dependent and Independent Transcriptional Networks. *PLoS ONE*. 2007;2(11):e1186. doi:10.1371/journal.pone.0001186
461. Van Dyk TK, DeRose EJ, Gonye GE. LuxArray, a High-Density, Genomewide Transcription Analysis of *Escherichia coli* Using Bioluminescent Reporter Strains. *J Bacteriol*. 2001;183(19):5496-5505. doi:10.1128/JB.183.19.5496-5505.2001
462. Van Dyk TK, Ayers BL, Morgan RW, Larossa RA. Constricted Flux through the Branched-Chain Amino Acid Biosynthetic Enzyme Acetolactate Synthase Triggers Elevated Expression of Genes Regulated by rpoS and Internal Acidification. *J Bacteriol*. 1998;180(4):785-792. doi:10.1128/JB.180.4.785-792.1998
463. Goss WA, Deitz WH, Cook TM. Mechanism of Action of Nalidixic Acid on *Escherichia coli*. *Journal of Bacteriology*. 1965;89(4):1068-1074.
464. Sugino A, Peebles CL, Kreuzer KN, Cozzarelli NR. Mechanism of action of nalidixic acid: Purification of *Escherichia coli nal*A gene product and its relationship to DNA gyrase and a novel nicking-closing enzyme. *Proceedings of the National Academy of Sciences*. 1977;74(11):4767-4771. doi:10.1073/pnas.74.11.4767
465. LaRossa RA, Schloss JV. The Sulfonylurea Herbicide Sulfometuron Methyll s an Extremely Potent and Selective Inhibitor of Acetolactate Synthase in *Salmonella typhimurium*. *Journal of Biological Chemistry*. 1984;259(14):8753-8757.
466. Domka J, Lee J, Bansal T, Wood TK. Temporal gene-expression in *Escherichia coli* K-12 biofilms. *Environ Microbiol*. 2007;9(2):332-346. doi:10.1111/j.1462-2920.2006.01143.x
467. Cardinale CJ, Washburn RS, Tadigotla VR, Brown LM, Gottesman ME, Nudler E. Termination Factor Rho and Its Cofactors NusA and NusG Silence Foreign DNA in *E. coli*. *Science*. 2008;320(5878):935-938. doi:10.1126/science.1152763
468. Vior NM, Lacret R, Chandra G, Dorai-Raj S, Trick M, Truman AW. Discovery and Biosynthesis of the Antibiotic Bicyclomycin in Distantly Related Bacterial Classes. *Appl Environ Microbiol*. 2018;84(9):e02828-17, /aem/84/9/e02828-17.atom. doi:10.1128/AEM.02828-17
469. Witwinowski J, Moutiez M, Coupet M, et al. Study of bicyclomycin biosynthesis in *Streptomyces cinnamoneus* by genetic and biochemical approaches. *Sci Rep*. 2019;9(1):20226. doi:10.1038/s41598-019-56747-7
470. Ito A, Taniuchi A, May T, Kawata K, Okabe S. Increased Antibiotic Resistance of *Escherichia coli* in Mature Biofilms. *AEM*. 2009;75(12):4093-4100. doi:10.1128/AEM.02949-08
471. Ito A, May T, Kawata K, Okabe S. Significance of *rpoS* during maturation of *Escherichia coli* biofilms. *Biotechnol Bioeng*. 2008;99(6):1462-1471. doi:10.1002/bit.21695
472. Ito A, May T, Taniuchi A, Kawata K, Okabe S. Localized expression profiles of *rpoS* in *Escherichia coli* biofilms. *Biotechnol Bioeng*. 2009;103(5):975-983. doi:10.1002/bit.22305
473. Sangurdekar DP, Srienc F, Khodursky AB. A classification based framework for quantitative description of large-scale microarray data. *Genome Biol*. 2006;7(4):R32. doi:10.1186/gb-2006-7-4-r32
474. Cho B-K, Zengler K, Qiu Y, et al. The transcription unit architecture of the *Escherichia coli* genome. *Nat Biotechnol*. 2009;27(11):1043-1049. doi:10.1038/nbt.1582
475. Kao KC, Tran LM, Liao JC. A Global Regulatory Role of Gluconeogenic Genes in *Escherichia coli* Revealed by Transcriptome Network Analysis. *J Biol Chem*. 2005;280(43):36079-36087. doi:10.1074/jbc.M508202200
476. Liu M, Durfee T, Cabrera JE, Zhao K, Jin DJ, Blattner FR. Global Transcriptional Programs Reveal a Carbon Source Foraging Strategy by *Escherichia coli*. *J Biol Chem*. 2005;280(16):15921-15927. doi:10.1074/jbc.M414050200
477. Curran TD, Abacha F, Hibberd SP, Rolfe MD, Lacey MM, Green J. Identification of new members of the *Escherichia coli* K-12 MG1655 SlyA regulon. *Microbiology*. 2017;163(3):400-409. doi:10.1099/mic.0.000423
478. Cohen SP, Hächler H, Levy SB. Genetic and functional analysis of the multiple antibiotic resistance (mar) locus in Escherichia coli. *Journal of Bacteriology*. 1993;175(5):1484-1492. doi:10.1128/JB.175.5.1484-1492.1993
479. Alekshun MN, Levy SB, Mealy TR, Seaton BA, Head JF. The crystal structure of MarR, a regulator of multiple antibiotic resistance, at 2.3 Å resolution. *Nature Structural Biology*. 2001;8(8):710-714.
480. Seoane AS, Levy SB. Characterization of MarR, the repressor of the multiple antibiotic resistance (mar) operon in *Escherichia coli*. *Journal of bacteriology*. 1995;177(12):3414-3419. doi:10.1128/JB.177.12.3414-3419.1995

481. Cohen SP, Levy SB, Foulds J, Rosner JL. Salicylate induction of antibiotic resistance in *Escherichia coli*: activation of the *mar* operon and a mar-independent pathway. *Journal of Bacteriology*. 1993;175(24):7856-7862. doi:10.1128/JB.175.24.7856-7862.1993
482. Perera IC, Grove A. Molecular Mechanisms of Ligand-Mediated Attenuation of DNA Binding by MarR Family Transcriptional Regulators. *Journal of Molecular Cell Biology*. 2010;2(5):243-254. doi:10.1093/jmcb/mjq021
483. Buchmeier N, Bossie S, Chen CY, Fang FC, Guiney DG, Libby SJ. SlyA, a transcriptional regulator of *Salmonella typhimurium*, is required for resistance to oxidative stress and is expressed in the intracellular environment of macrophages. *Infection and immunity*. 1997;65(9):3725-3730. doi:10.1128/IAI.65.9.3725-3730.1997
484. Libby SJ, Goebel W, Ludwig A, et al. A cytolysin encoded by *Salmonella* is required for survival within macrophages. *Proceedings of the National Academy of Sciences*. 1994;91(2):489-493. doi:10.1073/pnas.91.2.489
485. Haque MM, Kabir MS, Aini LQ, Hirata H, Tsuyumu S. SlyA, a MarR Family Transcriptional Regulator, Is Essential for Virulence in *Dickeya dadantii* 3937. *JB*. 2009;191(17):5409-5418. doi:10.1128/JB.00240-09
486. Baba T, Ara T, Hasegawa M, et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol*. 2006;2(1). doi:10.1038/msb4100050
487. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences*. 2000;97(12):6640-6645. doi:10.1073/pnas.120163297
488. Busso D, Peleg Y, Heidebrecht T, et al. Expression of protein complexes using multiple *Escherichia coli* protein co-expression systems: A benchmarking study. *Journal of Structural Biology*. 2011;175(2):159-170. doi:10.1016/j.jsb.2011.03.004
489. I'Anson KJA, Movahedi S, Griffin HG, Gasson MJ, Mulholland F. A non-essential glutamyl aminopeptidase is required for optimal growth of Lactococcus lactis MG1363 in milk. *Microbiology*. 1995;141(11):2873-2881. doi:10.1099/13500872-141-11-2873
490. Zhao D, Huang Z. Effect of His-Tag on Expression, Purification, and Structure of Zinc Finger Protein, ZNF191(243-368). *Bioinorganic Chemistry and Applications*. 2016;2016:1-6. doi:10.1155/2016/8206854
491. Kutyshenko VP, Mikoulinskaia GV, Chernyshov SV, Yegorov AY, Prokhorov DA, Uversky VN. Effect of C-terminal His-tag and purification routine on the activity and structure of the metalloenzyme, l-alanyl-d-glutamate peptidase of the bacteriophage T5. *International Journal of Biological Macromolecules*. 2019;124:810-818. doi:10.1016/j.ijbiomac.2018.11.219
492. Aiuppa A, Dongarrà G, Capasso G, Allard P. Trace elements in the thermal groundwaters of Vulcano Island (Sicily). *Journal of Volcanology and Geothermal Research*. 2000;98(1-4):189-207. doi:10.1016/S0377-0273(99)00156-0
493. Casadaban MJ, Cohen SN. Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. *Journal of Molecular Biology*. 1980;138(2):179-207. doi:10.1016/0022-2836(80)90283-1
494. Tsai C-Y, Lin Y-W. A highly selective and sensitive fluorescence assay for determination of copper(ii) and cobalt(ii) ions in environmental water and toner samples. *The Analyst*. 2013;138(4):1232. doi:10.1039/c2an36290b
495. Roth M, Carpentier P, Kaïkati O, et al. FIP: a highly automated beamline for multiwavelength anomalous diffraction experiments. *Acta Crystallographica Section D Biological Crystallography*. 2002;58(5):805-814. doi:10.1107/S0907444902003943
496. Ferrer J-L. Automated data processing on beamline FIP (BM30A) at ESRF. *Acta Crystallographica Section D Biological Crystallography*. 2001;57(11):1752-1753. doi:10.1107/S0907444901013385
497. Sheldrick GM. A short history of *SHELX*. *Acta Crystallogr A Found Crystallogr*. 2008;64(1):112-122. doi:10.1107/S0108767307043930
498. Williams CJ, Headd JJ, Moriarty NW, et al. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Science*. 2018;27(1):293-315. doi:10.1002/pro.3330
499. Chen VB, Arendall WB, Headd JJ, et al. *MolProbity*: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D Biological Crystallography*. 2010;66(1):12-21. doi:10.1107/S0907444909042073
500. Jubb HC, Higueruelo AP, Ochoa-Montaño B, Pitt WR, Ascher DB, Blundell TL. Arpeggio: A Web Server for Calculating and Visualising Interatomic Interactions in Protein Structures. *Journal of Molecular Biology*. 2017;429(3):365-371. doi:10.1016/j.jmb.2016.12.004
501. Kilambi KP, Gray JJ. Rapid Calculation of Protein pKa Values Using Rosetta. *Biophysical Journal*. 2012;103(3):587-595. doi:10.1016/j.bpj.2012.06.044
502. Lyskov S, Chou F-C, Conchúir SÓ, et al. Serverification of Molecular Modeling Applications: The Rosetta Online Server That Includes Everyone (ROSIE). *PLoS ONE*. 2013;8(5):e63906. doi:10.1371/journal.pone.0063906

503. Sabri S, Steen JA, Bongers M, Nielsen LK, Vickers CE. Knock-in/Knock-out (KIKO) vectors for rapid integration of large DNA sequences, including whole metabolic pathways, onto the *Escherichia coli* chromosome at well-characterised loci. *Microb Cell Fact*. 2013;12(1):60. doi:10.1186/1475-2859-12-60
504. Yang J, Sun B, Huang H, et al. High-Efficiency Scarless Genetic Modification in *Escherichia coli* by Using Lambda Red Recombination and I-SceI Cleavage. *Appl Environ Microbiol*. 2014;80(13):3826-3834. doi:10.1128/AEM.00313-14
505. Dower WJ, Miller JF, Ragsdale CW. High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Research*. 1988;16(13):6127-6145. doi:10.1093/nar/16.13.6127
506. Cherepanov PP, Wackernagel W. Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene*. 1995;158(1):9-14. doi:10.1016/0378-1119(95)00193-A

Summary

M42 aminopeptidases are dinuclear enzymes widely found in prokaryotes but completely absent from eukaryotes. They have been proposed to hydrolyze peptides downstream the proteasome or other related proteolytic complexes. Their description relies mainly on the pioneering work on four M42 aminopeptidases from *Pyrococcus horikoshii*. Their quaternary structure consists of twelve subunits adopting a tetrahedral-shaped structure. Such a spatial organization allows the compartmentalization of the active sites which are only accessible to unfolded peptides. The dodecamer assembly results from the self-association of dimers under the control of the metal ion cofactors. Both oligomers have been shown to co-exist *in vivo* and heterododecamers with broadened substrate specificity may even occur. Yet, the molecular determinants behind the dodecamer assembly remain unknown due the lack of a high-resolution structure of a stable dimer. In addition, the bacterial M42 aminopeptidases are still ill-described due to the paucity of structural studies.

This work focuses mainly on the characterization of TmPep1050, an M42 aminopeptidase from *Thermotoga maritima*. As expected, TmPep1050 adopts the genuine tetrahedral-shaped structure with twelve subunits. It also displays a leucyl-aminopeptidase activity requiring Co^{2+} as a cofactor. In addition to its catalytic function, Co^{2+} has a role in the enzyme thermostability and oligomerization. The absence of Co^{2+} provokes the disassembly of active TmPep1050 dodecamers into inactive dimers. The process, however, is reversible since Co^{2+} triggers the self-association of dimers into dodecamers, as shown by native MS. The main achievement of this work is the determination of the first high-resolution structure of a dimer, allowing to better understand the dimer-dodecamer transition. Several structural motifs involved in oligomerization are displaced or highly flexible in the TmPep1050 dimer structure. Furthermore, a loop bringing two catalytic relevant residues is displaced outside the catalytic site. These residues are the catalytic base and a ligand involved in the Co^{2+} binding at the M1 site. The metal ion binding sites have been further investigated to define how they influence the oligomerization of TmPep1050. A mutational study shows that the M1 site strictly controls the dodecamer formation while the M2 site contributes only partly to it. A strictly conserved aspartate residue of the M2 site second shell also plays an important structural role in maintaining the active site integrity. Indeed, its substitution prevents the formation of dodecamer probably due to the lack of stabilization of the active site loop.

The characterization of TmPep1050 supports that bacterial M42 aminopeptidases probably share the quaternary structures and dodecamer assembly with their archaeal counterparts. The dimer structure highlights several structural modifications occurring in the dimer-dodecamer transition. Yet, based on current knowledge, no general rules can be drawn for the role of the M1 and M2 sites in oligomerization. Besides, the physiological function of the M42 aminopeptidases is under-examined albeit the proposed link to the proteasome. In this work, this has been investigated using the *Escherichia coli* M42 aminopeptidases as a model. Yet, no phenotype has been associated to the deletion of their coding genes. Preliminary results have shown that the three enzymes (i) display a redundant substrate specificity, (ii) could be localized partly to the membrane, and (iii) form heterocomplexes. Further experiments are still required to crack the function of these M42 aminopeptidases.