COLLEGE SCIENCES DOCTORAL DE LA MER BRETAGNE ET DU LITTORAL

Université de Bretagne Occidentale

THESE DE DOCTORAT DE

L'UNIVERSITE
DE BRETAGNE OCCIDENTALE

ECOLE DOCTORALE N° 598
Sciences de la Mer et du Littoral
Spécialité : *Biologie, Biochimie Cellulaire et Moléculaire*

Par

Emilie CHAGNY

Phylogenetic and biochemical insights into the diversity of TET peptidases in Archaea

Thèse présentée et soutenue à Plouzané, le 6 décembre 2024
Unité de recherche : Unité de Biologie et Ecologie des Ecosystèmes marins Profonds (UMR 6197)

Rapporteurs avant soutenance :

Céline BROCHIER-ARMANET	Professeur – Université Claude Bernard Lyon 1
Yves BOURNE	Directeur de recherche CNRS – Aix Marseille Université

Composition du Jury :

Présidente :	Stéphanie MADEC	Professeur – Université de Bretagne Occidentale
Examinateurs :	Céline BROCHIER-ARMANET	Professeur – Université Claude Bernard Lyon 1
	Yves BOURNE	Directeur de recherche CNRS – Aix Marseille Université
	Raphaël DUTOIT	Responsable d'unité – Institut Labiris
Dir. de thèse :	Didier FLAMENT	Chercheur – IFREMER
Co-dir. de thèse :	Bruno FRANZETTI	Directeur de recherche CNRS – Université Grenoble Alpes

Invité

Eric GIRARD	Directeur de recherche CNRS – Université Grenoble Alpes

Remerciements

J’ai l’impression d’être arrivée à Brest il y a quelques mois à peine, et pourtant cela fait déjà plus de trois ans… Si cette thèse a filé aussi vite, c’est grâce aux nombreuses personnes qui m’ont épaulée pendant ces trois années et que je voudrais remercier ici.

D’abord, je tiens à remercier mes directeurs, Didier et Bruno, qui m’ont fait confiance et m’ont guidée tout au long de ce projet. Je suis reconnaissante de la bonne humeur et de la légèreté qui ont caractérisé toutes nos réunions, toujours sérieuses mais jamais trop, je m’y suis toujours sentie à l’aise. Il a été réconfortant de toujours trouver vos portes ouvertes quand j’avais la moindre question. Merci pour votre encadrement bienveillant.

Merci à mon directeur adoptif, Eric, qui m’a écoutée parler de mes purifs ratées pendant 6 mois alors qu’il n’est pas biochimiste. Merci d’avoir pris part à mon encadrement en apportant ta bienveillance, ton humanité, et ton légendaire pragmatisme. Il fallait bien qu’il y en ait un pour garder les pieds sur terre ! Et bien sûr, merci aussi pour ton œil expert de structuraliste.

Je tiens également à remercier Najwa qui m’a ouvert les portes du monde merveilleux de la phylogénie. Merci pour ton implication dans ce projet, pour ton expertise et pour le temps que tu m’as accordé, mais aussi pour ta gentillesse et ta patience (même lorsque je te demande une 165472^{e} version de la même figure).

Merci à tous les membres du LMEE pour votre accueil chaleureux. J'ai la chance de pouvoir dire que je ne suis jamais venue à reculon, même lors des semaines les plus chargées. Cela a été un vrai plaisir de passer trois ans dans ce laboratoire où les ‘‘grands’’ et les ‘‘petits’’ se mélangent aussi naturellement, que ce soit lors des pauses cafés ou dans les grandes tablées au restaurant Ifremer. Je remercie en particulier les membres de l’équipe Répli pour les discussions stimulantes et pour les coups de pouce au labo. Un énorme merci à Audrey, toujours prête à donner un coup de main ou un bon conseil (et à discuter), et toujours avec le sourire. Tu aimes ce que tu fais et ça se voit, c’est un vrai plaisir de travailler avec toi. Merci aux microbiologistes, et en particulier à Stéphane, Laurent et Erwan, qui ont toujours répondu à mes questions naïves de biochimiste avec bienveillance. Je n’ai pas forcément eu le temps de creuser toutes vos bonnes idées autant que je l’aurais voulu, mais ces échanges ont toujours été stimulants et enrichissants. Merci à Anne-Laure, le petit soleil au bout du couloir, qui chouchoute les doctorants du labo et qui nous rend la vie tellement plus facile.

Vous êtes plutôt mer ou montagne ? J'ai eu la chance de ne pas avoir eu à choisir et d'avoir passé 6 mois au laboratoire ELMA. Je remercie Bruno, Eric, Frank, Domi, Mylene, Viviam et Sandrine pour leur accueil. Merci aussi à Daphna, Lefteris et Guy pour l'accès VIP à la plateforme de microscopie, et pour m'avoir laissé découvrir les coulisses. Quelques bâtiments plus moins, merci à Valérie pour son temps, ses précieux conseils et la relecture de ce manuscrit.

Anna, ma raleuse au grand cœur, merci de ne pas m'avoir détestée alors qu'on venait à peine de se rencontrer et que je t'ai fait repacker une superose un vendredi à 17h (pour rien), et merci pour les nombreux coups de main au labo qui ont suivi. Merci pour les pauses cafés et les apéros. Merci pour ton soutien dans les moments les moins drôles.

J'ai eu la chance d'être très bien entourée à Brest aussi, avec tous les heureux habitants des algécos. Jordan, tu as su dès le départ m'inculquer une culture forte de la pause café, qui ne m'a ensuite plus jamais quittée. Au fait, tu as regardé le chant du loup ? Merci à ma première collègue de bureau, Léa, jamais à court de potins. Merci à Logan d'avoir ensuite pris le relais et de m'avoir supportée ces derniers mois, sans (trop) me juger. Et pardon pour le bruit. Marion I, ta bonne humeur et ton optimisme sont communicatifs, « change pas t'es la meilleure » comme dirait l'autre. Merci Marions I et II pour les innombrables pauses passées dans votre bureau. Merci à Yves, complètement freilauf, qui a toujours une idée de start-up improbable en tête. Tu nous manques beaucoup à Brest. Promis je m'hydrate ! KP, tu es le meilleur compagnon de course. Merci pour les sessions psychanalyse et les infusions mangue-passion. Merci à Claire, ma fan n°1, pour ses arrêts quotidiens dans mon bureau et pour toutes ces pauses café sans café.

Merci Caro et Mathisse pour ces rituels repas du dimanche soir. Merci Marine d'être complètement folle, merci pour les fous rires. Mathisse, les virées en Audi, les lundis poterie, les mardis run (RIP les genoux), les gin to' (bien trop rares ces derniers mois), les pouces au surf et les crêpes ont été de vraies bouffées d'oxygène qui m'ont aidée à finir cette thèse dans les meilleures conditions. Merci aussi d'avoir relu ce manuscrit.

Thomas, tu es un rayon de soleil au quotidien. Merci de me pousser à donner le meilleur de moi-même. Merci pour ma première tendinite et ma première fracture. Merci d'avoir été mon supporter n°1, si j'ai pu vivre ces derniers mois (à peu près) sereinement, c'est surtout grâce à toi. Merci pour tout.

Merci évidemment à ma famille, qui a toujours été là pour moi. Merci Nico pour les afterworks, les restos, et pour ces conversations qu'on n'avait jamais eu l'occasion d'avoir avant. C'était réconfortant d'avoir un petit bout de famille à Grenoble. Merci à mes parents pour leur soutien indéfectible, même si j'ai choisi de partir à l'autre bout de la France. Merci à ma sœur, qui sait maintenant expliquer mieux que moi ce qu'est une archée. Merci d'être ma personne préférée sur Terre.

Enfin, merci à ceux que j'aurais pu oublier de ne pas m'en vouloir.

Table of content

List of abbreviations

AAA+	ATPases associated with diverse cellular activities
ACE	Angiotensin-converting enzyme
AEX	Anion exchange chromatography
AMA	*Archaeoglobus* and methanogenic archaea ATPase
AMC	7-amino-4-methylcoumarin
bp	Base pair
cdc48	Cell division control 48 protein
Cd	Cadmium
chl	Chloramphenicol
CRC	Colorectal cancer
DNA	Deoxyribonucleic acid
DPANN	Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanohaloarchaeota, Nanoarchaeota
EDTA	Ethylene diamine tetraacetic acid
ELMA	Extremophiles and Large Molecular Assemblies
EM	Electron microscopy
ESP	Eukaryotic signature protein
FCB	Fibrobacteres, Chlorobi, Bacteroidetes
gen	Gentamycin
HGT	Horizontal gene transfer
HIC	Hydrophobic interaction chromatography
His-tag	Poly-histidine tag
HMM	Hidden Markov model
HmTET	*Haloarcula marismortui* TET
HoTET	*Hodarchaeales archaeon LC_3* TET
HPLC	High pressure liquid chromatography
IBD	Inflammatory bowel disease
IBS	Irritable bowel syndrome
ICP-MS	Intercoupled plasma mass spectrometry
IMAC	Immobilized metal affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kan	Kanamycin

LB	Lysogeny broth
LBA	Long branch attraction
LUCA	Last universal common ancestor
MAG	Metagenome-assembled genome
MAP	Methionine aminopeptidase
MjTET	*Methanocaldococcus jannaschii* TET
MtTET	*Methanoculleus thermophilus* TET
NCBI	National Center for Biotechnology Information
NC-IUBMB	Nomenclature Committee of the International Union of Biochemistry and Molecular Biology
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PAN	Proteasome-activating nucleotidase
PDB	Protein databank
PDE	Phosphodiesterase
PfTET	*Pyrococcus furiosus* TET
PhTET	*Pyrocuccus horikoshii* TET
pNA	*Para-nitroaniline*
PsTET	*Prometheoarchaeum syntrophicum* TET
RNA	Ribonucleic acid
SAMP	Small archaeal modifier protein
SANS	Small-angle neutron scattering
SDS	Sodium dodecyl sulfate
SEC	Size-exclusion chromatography
SEM	Scanning electron microscopy
TaTET	*Thermosphaera aggregans* TET
TACK	Thaumarchaeota, Aigarchaeota, Crenarchaeota, Korarchaeota
TET	Tetrahedral aminopeptidase
ThTET	*Thorarchaeaceae archaeon MP8T_1* TET
TMA	Trimethylamine
*Ton*Lon	*Thermococcus onnurineus NA1* Lon
TRI	Tricorn aminopeptidase
UFB	Ultra-fast bootstrap
UPS	Ubiquitin-proteasome system
VAT	Valosin-containing protein-like ATPase of Thermoplasma acidophilum

Introduction

1. Hidden in plain sight: the overlooked world of Archaea

1.1. Discovery of Archaea

As far back as Antiquity, Aristotle classified living beings into four groups: mineral, plant, animal and human. Since then, this classification has constantly evolved and undergone profound changes. In 1735, the introduction of the modern nomenclature "*Systema naturæ*" by Carl von Linné marked a major turning point for the discipline. Linné established the binomial system of nomenclature and organised species into a hierarchical system of categories: kingdom, phylum, class, order, family, genus and species. Over time, this classification has been continually enriched in line with progress in scientific techniques, particularly advances in microscopy and genetics. By 1970, the Tree of Life was made up of five domains: Monera, Protista, Plantæ, Fungi and Animalia (**Figure 1**) (Whittaker, 1969).

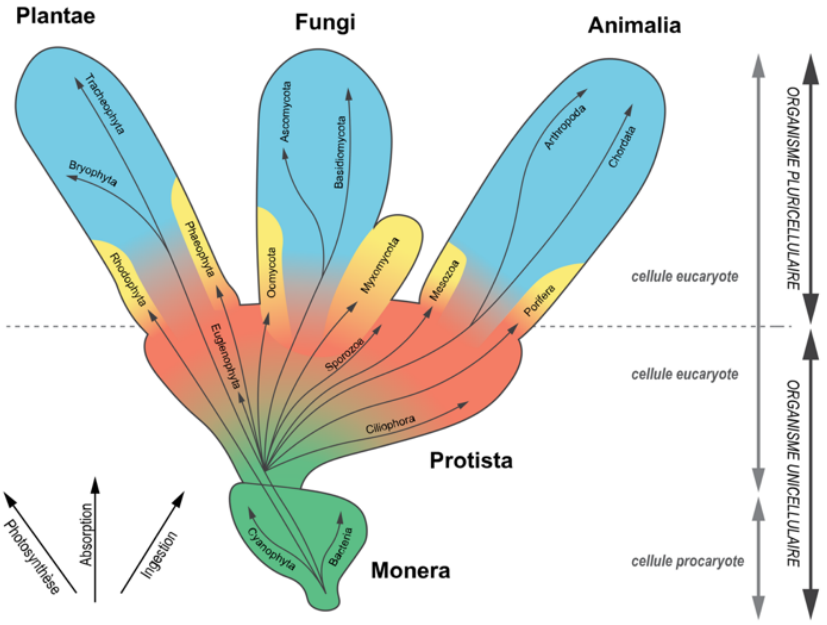


Figure 1: Classification of living organisms proposed by Whittaker in 1969. This five-kingdom classification is based on the complexity level and nutritional mode of organisms.

Subsequently, the development of molecular biology and DNA sequencing techniques led to the advent of phylogenetic classification, based on the notions of evolution and common ancestry. In 1977, Carl Woese and George Edward Fox made a groundbreaking discovery that revolutionized our understanding of organism classification. Analysis of the 16S ribosomal RNA sequence revealed that methanogenic 'bacteria', although similar to other bacteria in size,

shape and cellular organization, are in fact as phylogenetically distant from other bacteria as bacteria and eukaryotes (Fox *et al.*, 1977). This discovery prompted the proposal of a new classification system comprising three domains: Eukarya, Eubacteria and Archaebacteria (Woese & Fox, 1977), later renamed Archaea to avoid any confusion with bacteria (Woese *et al.*, 1990).

1.2. Unique yet familiar: the biochemistry of Archaea

Since their discovery, the isolation and sequencing of numerous archaeal species have unveiled characteristics shared with both bacteria and eukaryotes, alongside distinctive features (**Table 1**). For instance, both archaeal and bacterial cells lack intracellular compartmentalization, with transcription and translation processes presumed to be occurring simultaneously. While this coupling is well-established in bacteria, it is still only hypothesized in archaea. These prokaryotic organisms also share small circular genomes with a high density of genes that can be organized into operons (Koonin, 2009). On the other hand, the archaeal machineries involved in cellular information processing (DNA replication and repair, transcription, translation) are partly homologous to that of eukaryotes (Makarova *et al.*, 2014). Archaea also possess unique features distinguishing them from both bacteria and eukaryotes. For example, archaeal cytoplasmic membranes are composed of isoprenoid chains linked to glycerol-1-phosphate *via* ether bonds, contrasting with the fatty acid esters linked to glycerol-

Table 1: Comparison of selected characteristics of the three domains of Life. Adapted from Dombrowski *et al.*, 2021.

Characteristic	Bacteria	Archaea	Eukarya
Membrane-enclosed nucleus	No	No	Yes
Chromosomal structure	Circular	Circular	Linear
Peptidoglycan in cell wall	Yes	No	No
Membrane lipids	Ester-linked	Ether-linked	Ester-linked
Glycerol	Glycerol-3-phosphate	Glycerol-1-phosphate	Glycerol-3-phosphate
Ribosomes (mass)	70S	70S	80S
Initiator tRNA	formylmethionine	methionine	methionine
Introns	No	No	Yes
Operons	Yes	Yes	No
RNA polymerase	One (4 subunits)	One (8–12 subunits)	Three (12–14 subunits)
Transcription factors required	No	Yes	Yes
TATA box in promoter	No	Yes	Yes

3-phosphate found in bacterial and eukaryotic membranes (Cavicchioli, 2011). Recently, a putative metabolic pathway for the synthesis of 'archaeal' membrane lipids was identified in a Metagenome-Assembled Genome (MAG) of *Ca.* Cloacimonetes, a bacterium of the Fibrobacteres-Chlorobi-Bacteroidetes (FCB) group (Villanueva *et al.*, 2021). This is to date the only instance of archaeal-like membrane outside the archaeal domain.

1.3. Ecology and physiology of Archaea

Initially discovered in extreme environments, considered by human standards, such as around hydrothermal vents where temperatures can reach 350°C (Fiala & Stetter, 1986), archaea were long associated with extremophily. Some archaea exhibit particular metabolic capacities, such as the ability to use inorganic compounds like hydrogen or sulphur as energy sources (Bertoldo & Antranikian, 2006; Garcia *et al.*, 2006), demonstrating a metabolic flexibility that enables them to colonize varied and often hostile ecological niches.

Due to their often stringent and specialized growth requirements, cultivation of archaea in laboratory settings remains challenging. From the 1990s on, most of the detection of unknown archaeal diversity has been achieved through environmental surveys and culture-independent approaches, such as metagenomics and high-throughput sequencing (Baker *et al.*, 2020). These techniques allowed for the detection and characterization of archaea directly from environmental samples, bypassing the need for traditional culturing methods, and have shown that archaea are present in many more environments than initially thought. They are now known to be ubiquitous species and are estimated to account for more than 20% of the Earth's total biomass (DeLong & Pace, 2001). Owing to their incredible metabolic diversity, they play pivotal roles in global biogeochemical cycles of carbon, nitrogen and sulphur (Offre *et al.*, 2013).

From the early 1980s, archaea were discovered in the human microbiota. Initial studies identified *Methanobrevibacter smithii* in human faeces, marking the first evidence of methanogenic archaea in the human gut (Miller & Wolin, 1982). The presence of *Methanobrevibacter* species was then established in the digestive tracts of various mammals, including ruminants (Miller & Wolin, 1986). Subsequent studies highlighted the presence of methanogenic and non-methanogenic archaeal species on the human skin, in the oral and nasal cavities, the lungs and the vagina (Belay *et al.*, 1990; Ferrari *et al.*, 1994; Koskinen *et al.*, 2017; Kulik *et al.*, 2001) (**Figure 2**). Despite the evidence of numerous archaea being associated with a wide range of eukaryotes, including humans, no pathogenic species has

been reported to date. Archaea are known to produce toxins, such as halocins or sulfolobicins (O'Connor & Shand, 2002). They also possess secretion and adhesion systems, recognized as crucial virulence factors (Kachlany *et al.*, 2000; Makarova *et al.*, 2016). Moreover, their genome are highly dynamic due to the occurrence of horizontal gene transfers (HGT), the presence of mobile genetic elements and recombination systems (Wagner *et al.*, 2017). The absence of identified pathogenic species is therefore intriguing, considering these potential mechanisms for virulence and adaptation. Ongoing research endeavors to unravel the ecological and evolutionary factors underlying the apparent lack of pathogenic archaea.

Even if there is no direct evidence of pathogenicity, studies seem to indicate that methanogenic species could be associated with various diseases, such as colorectal cancer and inflammatory bowel disease (Mohammadzadeh *et al.*, 2022). It is also believed that archaea could have an indirect damaging impact on their hosts through the production of metabolites (**Figure 2**). For instance, it has been suggested that *Methanobrevibacter smithii* metabolites could cause cancer by producing oxidative DNA damage and/or inhibiting DNA repair (Dopp *et al.*, 2004). They have also been shown to inhibit the growth of other microbiota species and could therefore have the ability to modulate the gut microbiota (Meyer *et al.*, 2008).

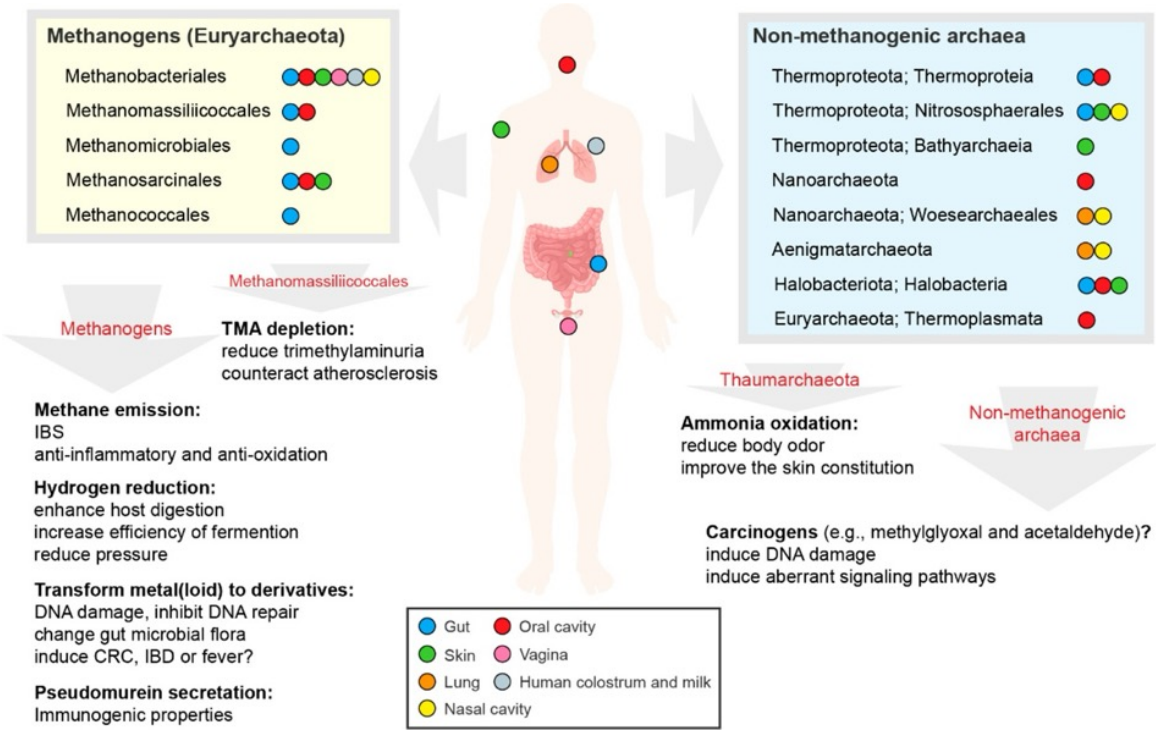


Figure 2: Archaea associated with the human body. Representative metabolites and their potential contributions to human and health and disease are presented. CRC: colorectal cancer. IBD: inflammatory bowel disease. IBS: irritable bowel syndrome. TMA: trimethylamine. Adapted from Cai et Tang, 2022.

1.4. The growing tree of Archaea

In 1994, Gary Olsen stated '...overlooking the Archaea has been equivalent to surveying one square kilometer of the African savanna and missing over 300 elephants' (Olsen, 1994). Today, this comparison has been widely verified, and the initial projections have been largely exceeded. Following the publication of the first complete archaeal genome of *Methanococcus jannashii* in 1996 (Bult *et al.*, 1996), the number of available genomes available is growing exponentially. While in 2011, around a hundred genomes had been published (Brochier-Armanet *et al.*, 2011), the National Center for Biotechnology Information (NCBI) genome database includes almost 23,000 genomes as of July 2024, representing data from both cultured microorganisms and direct environmental sampling. This massive burst in the availability of genomic data has catalyzed a rapid evolution of the phylogenetic tree of archaea since its first version proposed in 1990 (**Figure 3**).

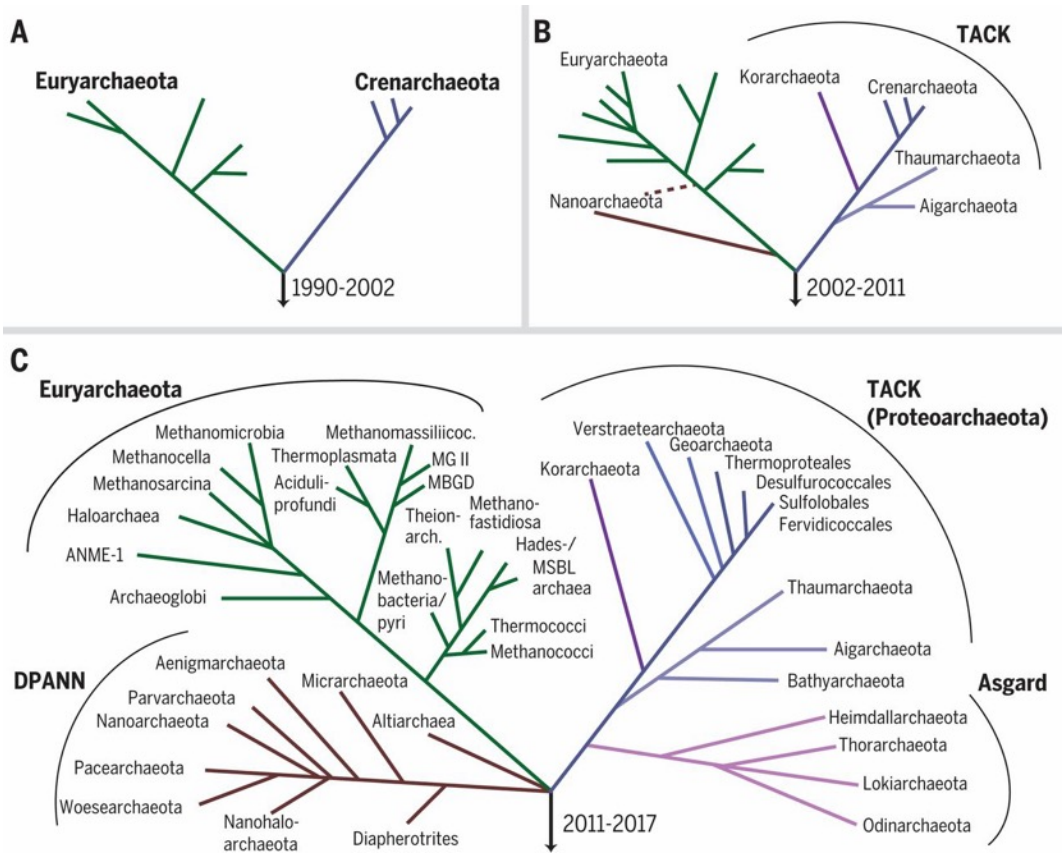


Figure 3: The expanding archaeal diversity. (A) Between 1990 and 2002, only two archaeal phyla were known (Euryarchaeota and Crenarchaeota). (B) Additional phyla were identified between 2002 and 2011, but the phylum-status of Nanoarchaeota was still controversial. (C) Since 2011, various additional archaeal lineages of high taxonomic rank have been discovered, propelled by advances in sequencing technologies and the use of metagenomic and single-cell genomic techniques. Adapted from Spang *et al.*, 2017.

For over a decade, Archaea were thought to be divided between Euryarchaeota and Crenarchaeota. Between 2002 and 2011, several new lineages were proposed, such as Nanoarchaeota (Huber *et al.*, 2002), Korarchaeota (Elkins *et al.*, 2008), Thaumarchaeota (Brochier-Armanet *et al.*, 2008), or Aigarchaeota (Nunoura *et al.*, 2011). In 2011, the Thaum-, Aig-, Cren- and Korarchaeota were shown to form a monophyletic group, the TACK superphylum (Guy & Ettema, 2011). Metagenomic studies subsequently uncovered a plethora of nanosized archaea, which were grouped into the proposed DPANN superphylum, named according to the first taxa identified: Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanohaloarchaeota, Nanoarchaeota (Rinke *et al.*, 2013). DPANNs typically possess compact genomes of approximately 1 Mbp and seem to have undergone extensive HGT events. They are believed to be fast-evolving species with parasitic or symbiotic lifestyle (Dombrowski *et al.*, 2019). Their monophyly remains debated and could result from a long branch attraction (LBA) artifact (Adam *et al.*, 2017). As presented in **Figure 4**, this phenomenon causes distantly related lineages, which have long branches due to high rates of evolution, to be incorrectly inferred as closely related.

Figure 4: Long branch attraction (LBA) artefact in phylogenetic analysis. In cases of trees where some branches are considerably longer than others, methods such as parsimony may prefer the wrong tree where long branches are grouped together. Adapted from Bawono and Heringa, 2014.

Since, numerous additional phyla have been described. Notably, a 2015 metagenomic survey allowed the identification of Lokiarchaeota from sediment samples collected near Loki's Castle, a hydrothermal vent system in the Arctic Ocean (Spang *et al.*, 2015). Lokiarchaeota fall into a new superphyla, named Asgard after Norse gods. In the following years, the Asgard group has been expanded with the addition of new lineages, including Thorarcheota, Odinarchaeota, and Heimdallarchaeota (Seitz *et al.*, 2016; Zaremba-Niedzwiedzka *et al.*, 2017). Notably, phylogenetic and comparative genomic analyses have indicated that this archaeal clade includes the closest archaeal sister lineage of eukaryotes, as detailed below.

1.5. The evolving Tree of Life: insights from Archaea

Since the establishment of Archaea as a separate domain of Life, their relationship with the two other domains has sparked considerable debate.

Eukaryotic cells are characterized by their intracellular compartmentalization, distinguishing them from prokaryotic cells. In 1967, Lynn Sagan first proposed the hypothesis that eukaryotic organelles, such as mitochondria and chloroplasts, originated through endosymbiosis (Sagan, 1967). A viral origin of the eukaryotic nucleus was also proposed by Philip Bell (2001) and later gained popularity with the discovery of large, complex DNA viruses capable of protein synthesis (Scola *et al.*, 2003). Although the endosymbiosis hypothesis is generally accepted, at least for the acquisition of the alphaproteobacterial ancestor of mitochondria (Lane & Martin, 2010), the nature of the host cell remains enigmatic. Some hypotheses state that this cell was a 'proto-eukaryote' similar to modern eukaryotes (Dacks & Field, 2007), whereas others propose it was an archaeon or even a bacterium (López-García & Moreira, 2015; Martin *et al.*, 2015). For a long time, the prevailing view was thus that Archaea and Eukaryotes represent two independent sister lineages in a three-domain (3D) Tree of Life (**Figure 5A**).

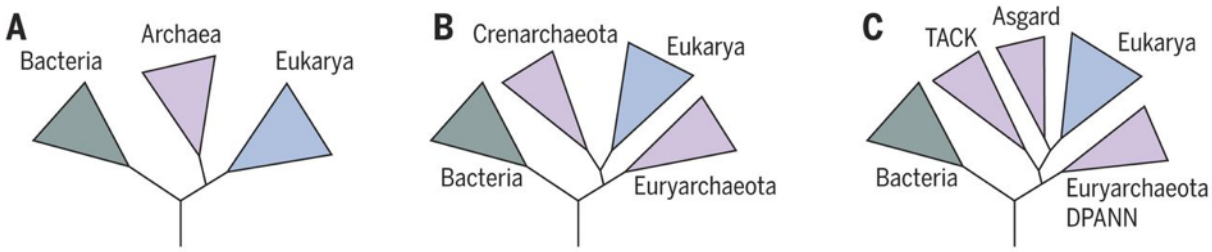


Figure 5: Evolution of the Tree of Life topology. Schematic depiction of the relationship of Archaea, Bacteria, and Eukaryotes according to the (A) Three- and (B) Two-Domain topology, updated with (C) the Asgard archaea. Adapted from Spang *et al.*, 2017.

However, some phylogenetic analyses have challenged this model, leading to alternative scenarios where Eukaryotes evolved from within the Archaea. The development of cultivation-independent genomic approaches and advances in phylogenetic tools resulted in the first two-domain (2D) trees, in which Eukaryotes branched from within the archaeal domain, more precisely close to the TACK superphylum, and Archaea are not monophyletic (**Figure 5B**). This topology was supported by the presence of proteins that were previously assumed to be specific to eukaryotes in TACK genomes, *i.e.,* eukaryotic signature proteins (ESP) (Guy

& Ettema, 2011) (**Figure 6**). Nevertheless, these analyses failed to identify a specific archaeal lineage as being more closely related to Eukaryotes than the others. The position of Eukaryotes among Archaea became clearer with the recent discovery of the Asgard archaea. Phylogenomic analyses placed Asgard as a sister group of Eukaryotes, and genome analyses revealing extended sets of ESPs in these species further supported this topology (Spang *et al.*, 2015; Zaremba-Niedzwiedzka *et al.*, 2017) (**Figure 5C**). Eukaryotes are currently proposed to be a sister lineage to Hodarchaeales, a newly proposed order within Heimdallarchaeia (Eme *et al.*, 2023).

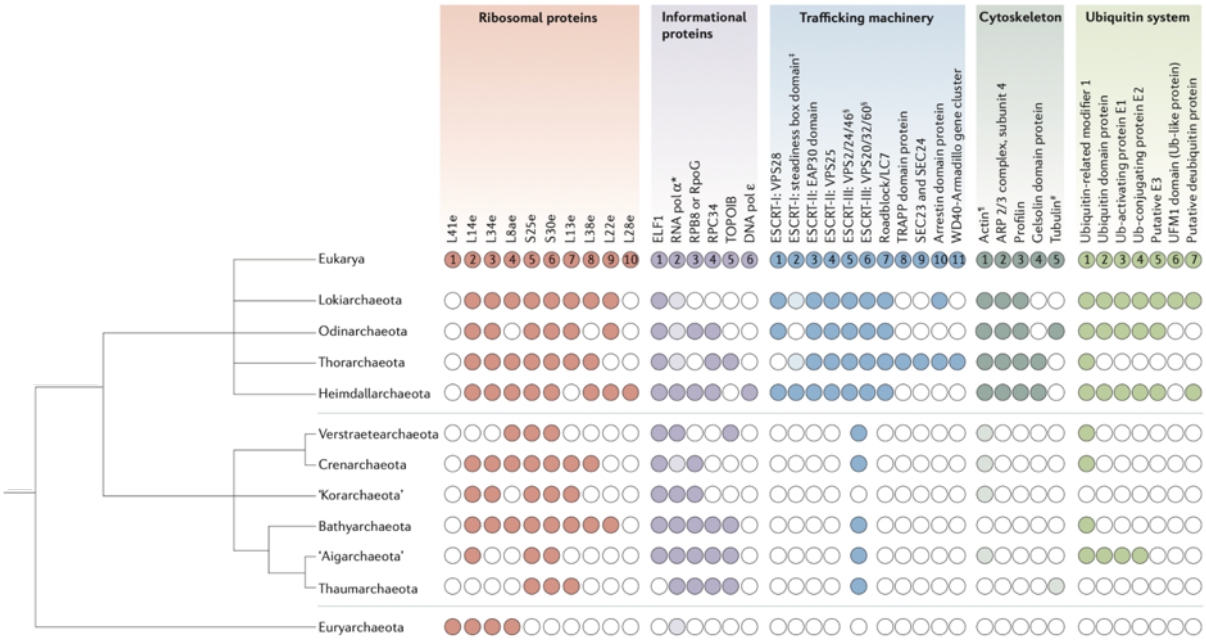


Figure 6: Distribution of eukaryotic signature proteins (ESP) across the archaeal tree. Putative homologues are lightly shaded. Adapted from Eme *et al.*, 2017.

Although the scientific community is increasingly favoring the 2D Tree of Life, an active and vigorous debate continues regarding the validity and implications of both the 2D and 3D topologies, and recent studies have formulated several critiques in opposition to the 2D scenario. For instance, the overrepresentation of Archaea and Asgard species in the datasets could be erroneously favoring 2D topology. Misalignments in multiple sequence alignments and the inclusion of fast-evolving species further exacerbate these biases. LBA artifacts, particularly between fast-evolving archaea and the long bacterial branch, could misplace Eukaryotes within Archaea. Concern has also been raised regarding the potential contamination of Asgard metagenomic data from eukaryotic sources, allegedly resulting in the false grouping of Eukaryotes and Asgard archaeota. Moreover, Asgard proteins show

inconsistent placement in single-protein trees, challenging the robustness of the 2D model. Marker selection also influences tree topology; large proteins tend to support 3D trees, while short proteins often support 2D trees due to insufficient phylogenetic signal. Additionally, individual proteins can significantly affect the overall topology; removing or including specific proteins can transform a tree from a 3D to a 2D topology and conversely (Da Cunha *et al.*, 2017, 2018, 2022).

Recent cultivation of two Asgard archaeon, *Prometheoarchaeum syntrophicum* (Imachi *et al.*, 2020) and *Lokiarchaeum ossiferum* (Rodrigues-Oliveira *et al.*, 2023) tempered the resonance of these criticisms (**Figure 7**). Sequencing of the whole genome of cultured single cells confirmed the previously observed proximity between Asgard species and Eukaryotes, and that the detected ESPs did not stem from eukaryotic contamination. According to recent estimates, around 50% of existing archaeal diversity still flies under the radar (Eloe-Fadrosh *et al.*, 2016). Further genomic data and phylogenetic analyses will undoubtedly provide additional insight on the matter in the years to come.

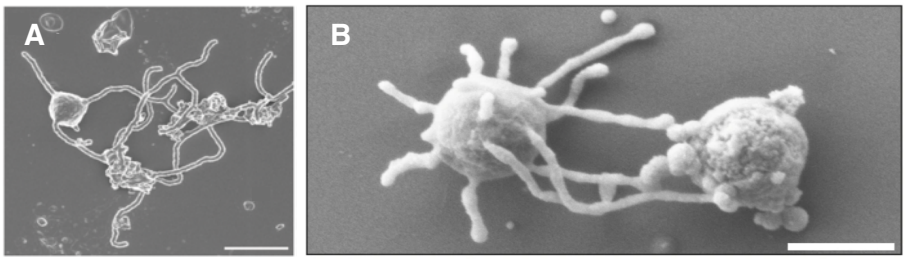


Figure 7: Scanning electron microscopy (SEM) imaging of (A) *Prometheoarchaeum syntrophicum* and (B) *Lokiarchaeum ossiferum*. Scale bars represent 200 nm and 500 nm, respectively. Adapted from Imachi *et al.*, 2020 and Rodrigues-Oliveira *et al.*, 2023.

1.6. Relevance of archaeal models in biochemistry and structural biology

As previously discussed, the evolutionary trajectories of archaeal and eukaryotic lineages are deeply interconnected. According to current hypotheses, the eukaryotic ancestor likely originated within Archaea and subsequently gained complexity over time, eventually resulting in modern eukaryotic organisms (see **section 1.5**). This evolutionary scenario is supported by the strong homologies observed between numerous archaeal and eukaryotic proteins and cellular processes. Histone proteins, which are nucleoid associated proteins involved in genome compaction and organization, are a good illustration of this phenomenon. While absent in Bacteria, all eukaryotic and most archaeal genomes encode histone proteins

(Henneman *et al.*, 2018). In Eukaryotes, there are five classes of histones: H1, H2A, H2B, H3, and H4. They assemble into a histone core, which is constituted of two H2A-H2B dimers and a H3-H4 tetramer. A DNA segment of approximately 150 base pair (bp) is wrapped twice around each octameric core, forming a nucleosome (Luger *et al.*, 1997). Linker histones H1 bind the DNA at the entry and exit points of the core particle, sealing the nucleosome (Bednar *et al.*, 2017). Nucleosomes interact with each other, forming a compact 30 nm fiber. Histone-encoding genes have also been identified in most archaea, with the exception of Crenarchaeota (Henneman *et al.*, 2018). Many archaeal genomes encode more than one histone variant (Sandman *et al.*, 1994), and they bind DNA as dimers (Sandman *et al.*, 1990) or tetramers (Pereira *et al.*, 1997). The structure of archaeal histone is very similar to that of their eukaryotic homologs, altough they do not possess the N-terminal tail found in eukaryotic proteins, which plays a critical role in DNA packaging and regulation through post-translational modifications (Henneman *et al.*, 2018) (**Figure 8**).

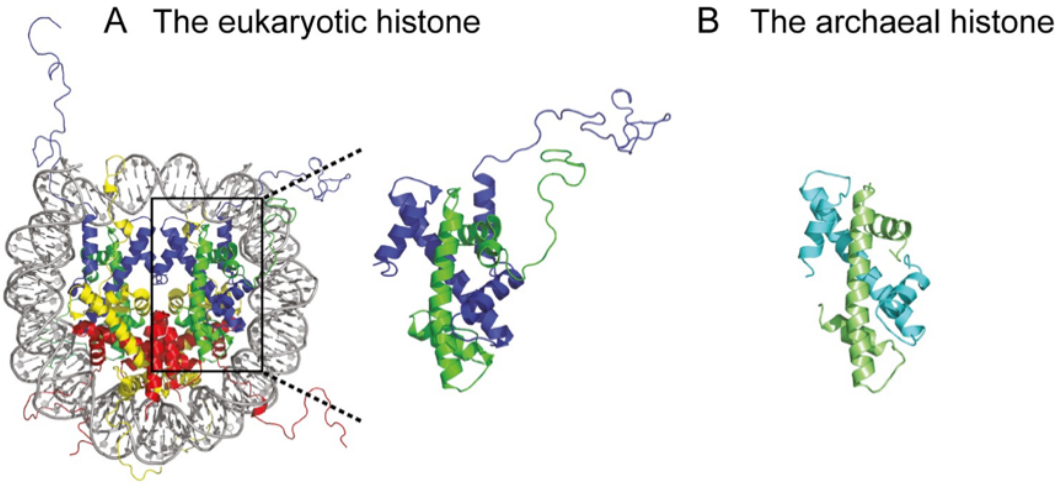


Figure 8: Eukaryotic and archaeal histones are very similar. (A) Eukaryotic nucleosomes are composed of DNA wrapped twice around two H2A-H2B dimers and a H3-H4 tetramer. Yellow, H2A; red, H2B; blue, H3; green, H4. (B) Archaeal histones are homologous to H3 and H4 eukaryotic histones but lack the N-terminal tail. Adapted from Henneman *et al.*, 2018.

When related, archaeal macromolecular edifices are often smaller and simplified version of their eukaryotic counterparts. Although very similar, the cellular machineries of Archaea involve fewer proteins, and their regulatory pathways are usually simpler. For instance, the archaeal 20S proteasome has only two to four different subunits, whereas its eukaryotic counterpart has fourteen (Maupin-Furlow *et al.*, 2006). Consequently, the archaeal complex is much simpler to reconstitute *in vitro*. This makes archaea relevant cellular models for integrative studies, providing valuable insights into their complex eukaryotic homologs. In addition to enabling more straightforward studies, archaeal models can also provide

information regarding the evolutionary history of protein families. As an example, the structural and biochemical characterization of the family D replicative DNA polymerase (PolD) of the archaeon *Pyrococcus abyssi* has shed light on the evolutionary history of the phosphodiesterase (PDE) domain, a fold found in numerous nucleases and therefore a key player in genome maintenance (**Figure 9**). The ancestor of the PDE domain may have already existed in the Last Universal Common Ancestor (LUCA). This ancestral nuclease domain would have acquired different substrate specificities, and was integrated into various DNA-binding proteins to serve distinct functions in genome maintenance, such as DNA recombination (Mre11/Rad50) and DNA proofreading (PolD). While Mre11/Rad50 is conserved across the Tree of Life, the active polymerase-associated PDE domain is only found in Archaea. In Eukaryotes, the PDE domain architecture is preserved in certain DNA polymerases but has lost its nuclease activity, replaced by the DnaQ nuclease fold (Betancurt-Anzola *et al.*, 2023).

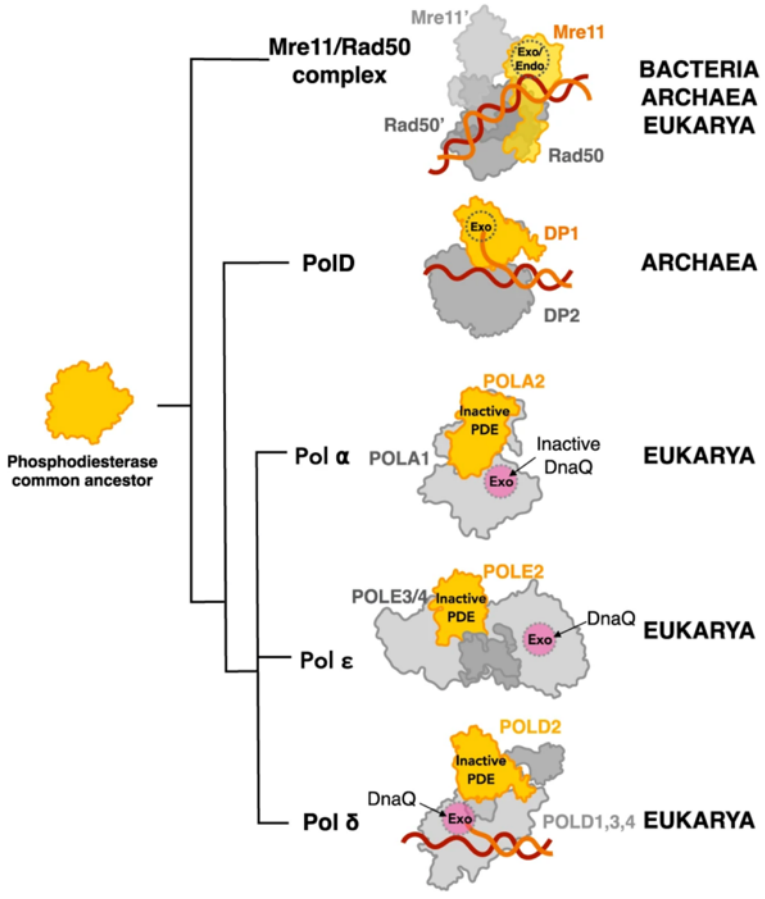


Figure 9: Inferred evolution and diversification of the PDE fold across the three domains of Life. Mre11 and DNA polymerases PDE folds would share a common ancestor. An active PDE motif has been preserved in archaeal PolD and in the universal Mre11. This motif became inactive in eukaryotic replicative DNA polymerases as they evolved to incorporate a DnaQ-like exonuclease domain. Adapted from Betancurt-Anzola *et al.*, 2023.

Finally, as previously mentioned, many species of archaea are extremophilic (see **section 1.3**), and (hyper)thermophilic species are of particular interest. Thanks to intrinsic adaptations of molecular components such as membranes and proteins, these organisms are able to thrive at temperatures above 60°C, and up to 122°C (Takai *et al.*, 2008). At the protein level, these adaptations correspond to changes in the amino acid sequence, resulting in variations of the structure, flexibility, charge, and/or hydrophobicity of the enzymes, enabling them to maintain folding and functionality at high temperatures. These proteins often have an increased hydrophobic core, and contain more salt bridges and ionic interactions that stabilize the structure through electrostatic interactions. Enhanced hydrogen bonding networks and aromatic interactions, such as π-π stacking, also contribute to their thermal stability (Jaenicke & Böhm, 1998). Furthermore, these proteins tend to have fewer thermolabile amino acids such as glutamine and asparagine, which are prone to deamidation at high temperatures (Carré *et al.*, 2022). Their structure is generally more compact and rigid, lowering entropy and increasing stability (Tehei *et al.*, 2005). Beyond enabling a deeper understanding of molecular adaptations in thermophilic organisms, thermostable proteins offer several benefits for biochemistry and structural biology experimentations. Firstly, the protein purification process can be carried out at room temperature. In the case of recombinant protein expression in a non-thermophilic host, a first clarification step consisting in a heat-shock can be used to eliminate the host proteins. Moreover, protein production yields are typically higher (Fujino *et al.*, 2020), which is particularly interesting for structural biology studies requiring large quantities of the protein of interest. Furthermore, in the case of enzyme characterization, activity assays can be controlled through temperature regulation. Enzymatic reactions can thus be either activated or inhibited so that the time scales are compatible with the techniques used (Ibrahim *et al.*, 2017).

All these characteristics make archaea highly relevant models for the biochemical and structural characterization of a protein family, which is precisely the aim of this thesis project. It therefore seemed most appropriate to focus our study of TET peptidases, which will be presented in sections **2.4.2.2** and **3** below, on archaeal representatives. Despite being present in all three domains of the Tree of Life, archaeal TET peptidases have been more extensively studied than their bacterial of eukaryotic counterparts, and have been investigated for over 20 years in the Extremophiles and Large Molecular Assemblies (ELMA) laboratory, at the Structural Biology Institute of Grenoble. During the course of this project, it has been extremely valuable to be able to draw upon the expertise previously acquired on those enzymes, either regarding the purification, functional or structural characterization of these molecular machineries. Despite the studies already conducted on this protein family, the biological

function of these peptidases has not yet been clearly determined. It therefore seemed appropriate to study them in 'simple' cellular models. Moreover, the presence of many extremophiles in this domain of life enabled us to investigate proteins from thermophilic organisms, which considerably facilitated the purification process, as detailed in **chapter 2** . Lastly, the study of archaeal TETs seemed relevant to gain insight on the evolutionary history of those enzymes (see **general discussion**).

2. Proteolysis: roles in cellular health and key players

2.1. The proteome clockwork: precision and dynamics of protein regulation

In 2011, Mora *et al.* estimated that there are 8.75 million living species on Earth, of which only 20% would have been described (Mora *et al.*, 2011). In recent years, culture-independent approaches, such as metagenomics and high-throughput sequencing, have transformed our exploration of biodiversity (Baker *et al.*, 2020; Handelsman, 2004), and the resulting wealth of new genomic data suggest that the existing biodiversity far exceeds this estimate (Wiens, 2023). Even though our current estimates of existing biodiversity are likely far below the actual figures, it is astonishing to contemplate the vast diversity arising from genetic information universally encoded by merely four nitrogenous bases —adenine (A), thymine (T), cytosine (C), and guanine (G). Despite this seeming simplicity, DNA encodes a surprising breadth of genetic information, with the human genome being estimated to comprise between 20,000 and 25,000 genes (Lander *et al.*, 2001).

The translation of these genes into proteins introduces a further layer of complexity. For instance, 100,000 distinct proteins are estimated to be produced from the human genome (Wilhelm *et al.*, 2014). The emergence of sophisticated mechanisms in the course of evolution allowed this tremendous complexity to be generated. For instance, alternative splicing allows a single pre-mRNA transcript to be spliced in different ways to produce multiple mature mRNA variants (**Figure 10**). This process involves the selective inclusion or exclusion of specific exons (coding regions) and introns (non-coding regions) during mRNA processing, enabling the production of multiple protein isoforms form a single gene (Pan *et al.*, 2008).

The proteome, which is defined as the complete set of proteins expressed by an organism, cell, or tissue at a specific time, is highly dynamic and is continually reshaped in response to internal or external stimuli, such as developmental stages and environmental conditions (Goldberg, 2003). This tight regulation of proteome composition can be perceived in the variation of renewal rates from one protein to another, with half-lives ranging from a few minutes to several years. For instance, regulatory proteins such as protein involved in DNA replication or cell cycle progression often have very short half-lives (Fatima *et al.*, 2022; Jenal, 1998; Udvardy, 1996), allowing cells to quickly adapt to changes in their environment by rapidly modulating gene expression (Sears *et al.*, 2000). Conversely, structural proteins, such as collagen, have much longer half-lives, up to several years (Verzijl *et al.*, 2000).

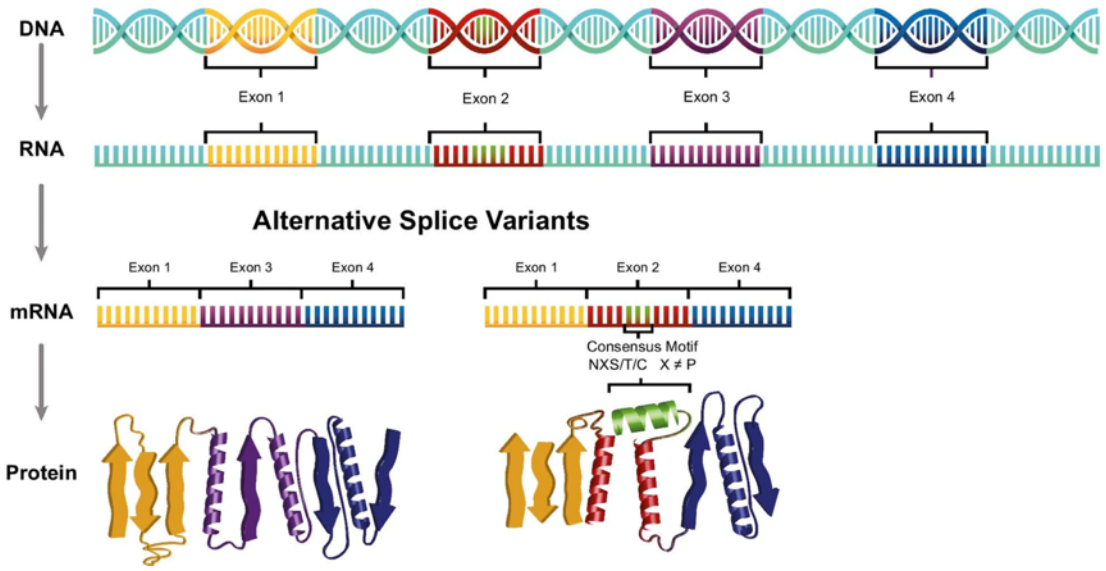


Figure 10: Alternative splicing allows proteome diversifying. In eukaryotes, genes are transcribed into pre-mRNA including both exons and introns. During alternative splicing, different combination of exons are joined together to form distinct mRNA variants, leading to the production of diverse protein isoforms from a single gene. Adapted from Kelly *et al.*, 2021.

Proteome composition is controlled through various regulatory mechanisms. The first level of regulation is exerted during transcription. mRNA synthesis from DNA can be enhanced or repressed by transcription factors and is also modulated by epigenetic modifications, such as DNA methylation or histone modification, which can regulate gene expression without alteration of the DNA sequence (Gibney & Nolan, 2010; Lee & Young, 2013). After mRNA is synthesized, various mechanisms can influence its stability, splicing, transport, and translation efficiency, such as the alternative splicing mechanism previously mentioned. The following step of translation of mRNAs into proteins is also fine-tuned, by translation initiation factors for example (Brito Querido *et al.*, 2024). The resulting proteins can then be subjected to post-translational modifications, such as phosphorylation, ubiquitination, methylation, acetylation, or glycosylation, affecting their activity and stability (Lee *et al.*, 2023). Protein localization can be controlled through signal peptides addressing proteins to specific cellular compartments, or membrane anchors targeting proteins to cell membranes (Bauer *et al.*, 2015). Finally, proteolysis ensures the degradation of damaged or excess proteins, maintaining protein homeostasis (Wolf & Menssen, 2018) but also serving several other crucial biological functions.

While the regulation of proteome composition involves multiple intricate mechanisms, the following sections will be dedicated to the exploration of proteolysis, with a particular focus on its key players in prokaryotes.

2.2. Biological functions of proteolysis

Both specific and non-specific proteolysis serve critical biological functions. As mentioned above, proteolysis is responsible for the degradation of abnormal proteins arising from transcriptional, translational, or folding defects, as well as those damaged by environmental stress or oxidative damage (Ciechanover, 2017). This house-keeping function is vital for maintaining cellular homeostasis by preventing the accumulation of defective proteins, and contributes to amino acid recycling by breaking down these damaged proteins into their constituent amino acids, which can then be reused for the synthesis of new proteins and other essential biomolecules. Additionally, proteolysis is also involved in protein activity control through the regulation of protein levels, activation and maturation. For instance, the renin-angiotensin system, crucial for blood pressure regulation in mammals, relies on the action of two key peptidases (**Figure 11**). Firstly, the renin peptidase cleaves the precursor angiotensinogen, producing the inactive peptide angiotensin I. This peptide is subsequently converted into the active angiotensin II by the angiotensin-converting enzyme (ACE), which is a carboxypeptidase. Angiotensin II is a potent vasoconstrictor involved in maintaining arterial volume and blood pressure (Banegas *et al.*, 2006). This regulatory function of proteolysis is central in numerous cellular processes, such cell cellular differentiation (Hillary & FitzGerald, 2018), apoptosis (Sahoo *et al.*, 2023), virulence (Bertels *et al.*, 2018), stress response (Mackinnon & Stone, 2022), *etc.*

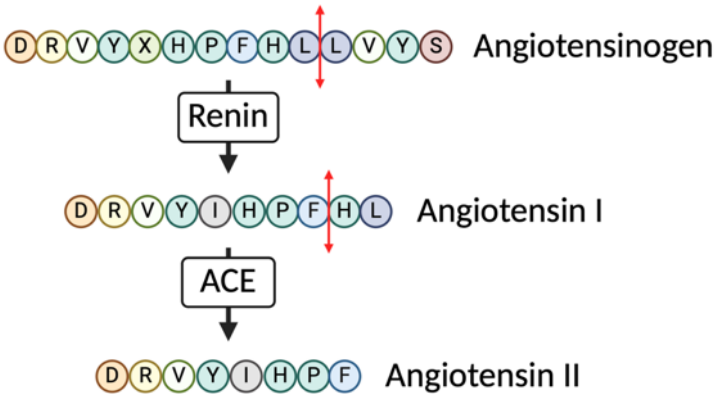


Figure 11: Role of proteolysis in the renin-angiotensin system. Blood pressure regulation in mammals involves the renin-angiotensin system, which produces the active angiotensin II from the inactive precursor angiotensinogen through successive cleavages by renin and ACE peptidases. Cleavage sites are indicated by red arrows. Adapted from Banegas *et al.*, 2006.

2.3. Terminology and nomenclature of proteolytic enzymes

The nomenclature of proteins can be perplexing, with sometimes various terms referring to the same entity. Before going over the key players proteolysis in Archaea, it is first necessary to present and explain the relevant vocabulary.

2.3.1. Panorama of the synonyms

A **protease** is defined as an enzyme degrading proteins by hydrolyzing peptide bonds. Similarly, a **peptidase** hydrolyzes peptide bonds, but it can act on both peptides and proteins. The Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) recommends using this term and suggests abandoning the term protease.

Pre- and suffixes can be added to provide information on the activity of the enzyme (**Figure 12**). An **endopeptidase** is an enzyme cleaving within a polypeptide chain, whereas an **exopeptidase** will be targeting the N-terminal or C-terminal end of the peptide. Exopeptidases that target the N-terminal extremity and cleave one, two or three residues are referred to as **aminopeptidases**, **dipeptidyl-** or **tripeptidyl-peptidases**, respectively. For the C-terminal end, these are referred to as **carboxypeptidase**s, **peptidyl-dipeptidases** or **peptidyl-tripeptidases,** respectively (Rawlings & Salvesen, 2013).

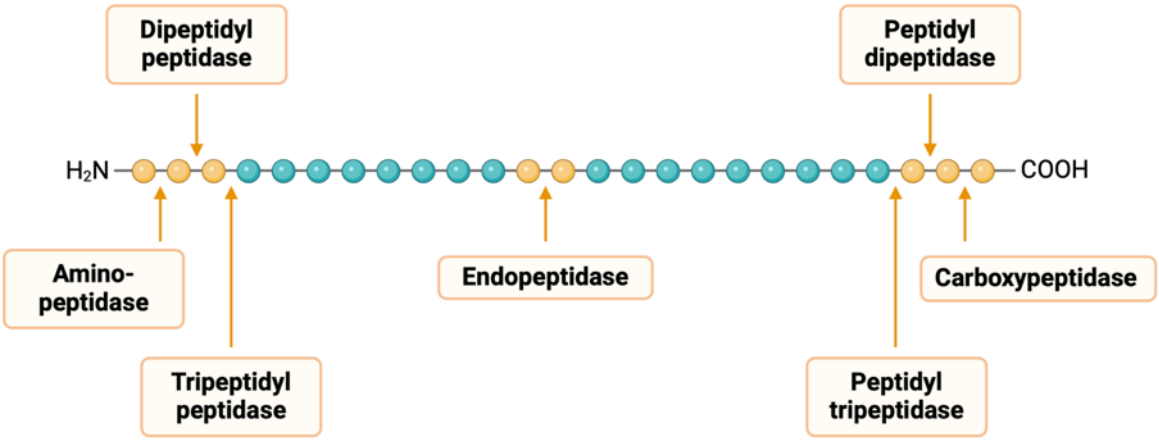


Figure 12: Nomenclature of peptidases according to their cleavage site.

2.3.2. MEROPS classification of peptidases

The MEROPS classification is commonly used to describe and categorize the wide diversity of proteolytic enzymes brought in above. Based on Hartley's classification of proteases according to their catalytic type (Hartley, 1960), Rawling and Barrett (2018) developed this classification, which now integrates the notion of evolutionary families.

Built around a holotype, peptidase families are constructed based on amino acid sequence homology. They correspond to groups of enzymes deriving from a common ancestor and showing evolutionary relationship, either throughout the whole sequence or at least in the part of the sequence responsible for the catalytic activity. The name of a family is constructed from a letter indicating the catalytic type (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) and a number (*e.g.,* M42).

Families can share similarities in their tertiary structures while lacking statistically significant similarities in sequence. These proteins have probably derived from the same ancestral protein but have diverged to such an extent that their evolutionary relationship can no longer be determined by comparison of the primary sequences. Such sets of families are grouped in clans, which are named by a letter indicating the catalytic type of the peptidase followed by an arbitrary second capital letter (*e.g.*, MH).

2.3.3. Active sites nomenclature

The commonly used active site nomenclature system was originally proposed by Schechter and Berger (1967). This system designates several binding sites in the enzyme, each site accommodating an amino acid of the substrate peptidic chain. By convention, the scissile bond is located between the P_1' and P_1' residues of the substrate peptide. The residues of the peptide are annotated P_1 … P_n toward the N-terminal end, and P_1' … P_n' toward the C-terminal end. The enzyme binding sites recognizing these residues are numbered S_1 … S_n and S_1' … S_n' for the P and P' sites, respectively (**Figure 13**).

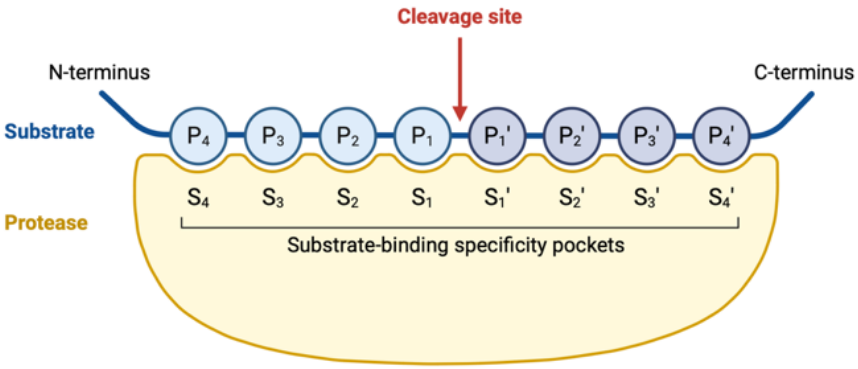


Figure 13: Peptidase active site nomenclature according to Schechter and Berger (1967). The cleavage site, located between the P_1 and P_1' residues of the peptide, is indicated by a red arrow.

2.4. Key players of archaeal proteolysis

Protein destruction in the cytoplasm involves numerous proteases with diverse structures, ranging from small single catalytic units to highly complex macromolecular assemblies containing multiple domains (Bond, 2019). These enzymes can be distinguished depending on whether they are energy-dependent or –independent. Energy-dependent proteases (see **section 2.4.1**) bear an endopeptidase activity and carry out the degradation of proteins into 3 to 30 residues polypeptides (Sauer & Baker, 2011). Following this initial degradation, the generated peptides are then further processed by ATP-independent peptidases (see **section 2.4.2**), releasing free amino acids (Yao & Cohen, 1999). In the following sections, the main protagonists of archaeal cytoplasmic proteolysis will be presented.

2.4.1. AAA+ proteases

The foundation of cellular homeostasis lies in energy-dependent proteases, which achieve the targeted degradation of damaged or specific regulatory proteins. These enzymes share a common architecture, forming large, multimeric, barrel-shaped complexes in which the proteolytic active sites are sequestrated in an inner chamber. ATP-dependent proteases are also referred to as AAA+ proteases, as the access to the catalytic chamber is controlled by regulatory domains or separate regulatory proteins belonging to the ATPases associated with diverse cellular activities (AAA+) superfamily. These ATPases assemble into hexameric rings with a narrow central axial opening preventing the passage of native folded proteins, thereby protecting the cellular environment from uncontrolled degradation (Lupas *et al.*, 1997; Shin *et al.*, 2020).

As depicted in **Figure 14**, the recognition of a target protein by the AAA+ protein is mediated by a degradation tag, or degron. Other parts of the substrate can interact with auxiliary domains of the AAA+ protein, either directly or using adaptor proteins (Sauer & Baker, 2011). ATP binding and hydrolysis trigger conformational changes of the AAA+ ring and fuel the mechanical unfolding of the protein substrate and its subsequent translocation through the pore and into the degradation chamber, where proteolysis occurs (Gottesman *et al.*, 1997).

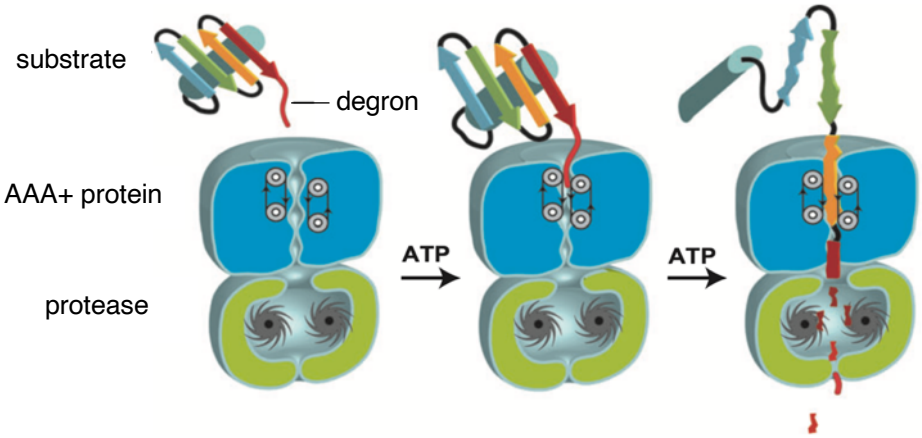


Figure 14: Protein degradation by AAA+ proteases. A degradation tag, or degron, on the substrate is recognized by an AAA+ protein. ATP binding and hydrolysis is used to fuel the mechanic unfolding and translocation of the substrate into the degradation chamber of the associated protease. Adapted from Sauer *et al.*, 2011.

AAA+ proteases can be classified into families according to their ATPase, protease and auxiliary domain(s) sequences (Sauer & Baker, 2011). In eukaryotes, protein homeostasis is ensured by the 26S proteasome (Bard *et al.*, 2018). With the exception of Actimonycetes, bacteria generally lack the proteasome 20S core particle and instead possess several AAA+ proteolytic complexes (Pouch *et al.*, 2000). For instance, four complexes are found in *Escherichia coli*: ClpXP/ClpAP, FtsH, HslUV, and Lon (Goff *et al.*, 1984; Gottesman *et al.*, 1998; Rohrwild *et al.*, 1996; Tomoyasu *et al.*, 1995). In archaea, energy-dependent proteolysis is fulfilled by the 20S proteasome and the Lon protease (Chandu & Nandi, 2004).

2.4.1.1 20S proteasome

As mentioned above, the 20S proteasome is ubiquitous in eukaryotes and archaea but is lacking in most bacterial phyla, apart from Actimonycetes. Despite divergences in the primary sequence, the architecture of the 20S particle is highly conserved, underscoring its fundamental role in protein degradation. The hallmark barrel-like structure consists of four

stacked heptameric rings, with two central β-rings flanked by two outer α-rings, delineating a ~ 700 kDa hollow particle. The proteolytic sites, sequestrated in the inner chamber, are located within the β-rings featuring nucleophilic threonine residues at their N-termini (Groll *et al.*, 1997). It is therefore a threonine peptidase belonging to the T1 MEROPS family. α-subunits interact with each other through their N-termini, forming a gate-like structure that is closed in the free 20S proteasome. This allows the control of the access to the proteolytic chamber, ensuring that only unfolded proteins enter the degradation pathway (**Figure 15A**) (Groll *et al.*, 2000).

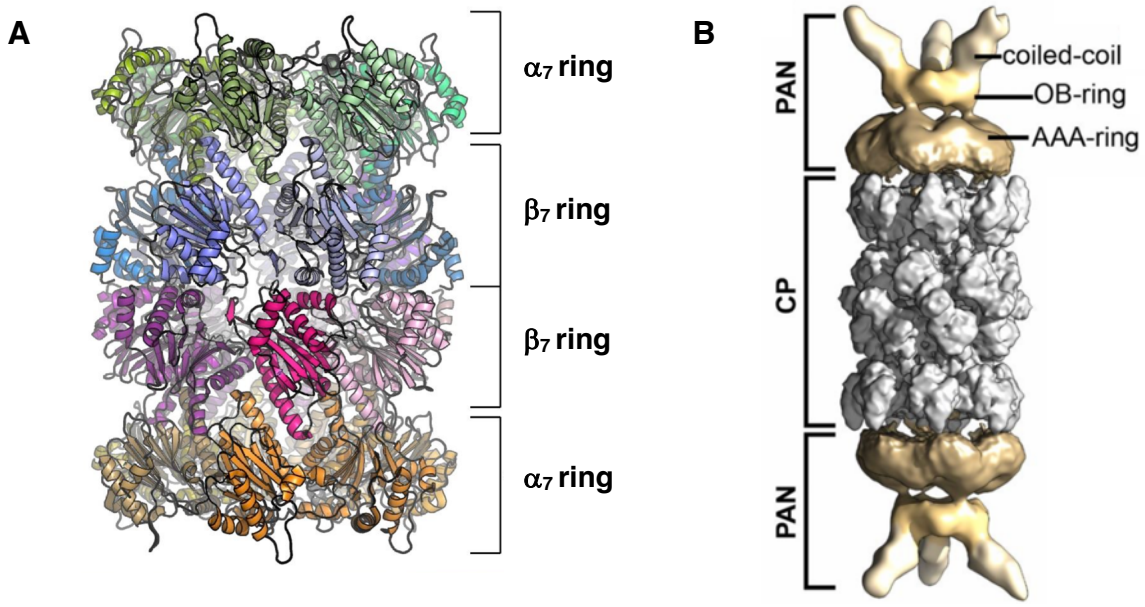


Figure 15: Architecture of the PAN-20S proteasome. (A) Cryo-EM structure of the *Thermoplasma acidophilum* 20S core particle made of four stacked heptameric rings, with the outer rings formed by α-subunits and the inner catalytic rings by β-subunits. Adapted from Dutoit, 2021. (B) Cryo-EM reconstruction of the 20S core particle (CP) capped by two PAN particles. Six PAN monomers associate to form three two-stranded coiled coils, a hexameric oligonucleotide/oligosaccharide binding (OB) ring, and a hexameric AAA ring. Adapted from Majumder *et al.*, 2019.

The opening of the entry channel of the archaeal 20S core particle was initially thought to be controlled by the proteasome-associated nucleotidase (PAN) ATPase, which was found to be crucial for recognizing, unfolding, and translocating substrates into the 20S core for degradation (Benaroudj & Goldberg, 2000). PAN possesses the canonical AAA+ domain, along with an oligonucleotide/oligosaccharide binding (OB) domain exhibiting a chaperone activity and recognizing degrons. Two PAN hexameric rings assemble on each side of the 20S core particle (Majumder *et al.*, 2019) (**Figure 15B**). The interaction of the C-terminus part of PAN with the core particle is sufficient to trigger the opening of the entry channel (Rabl *et al.*, 2008).

However, some archaeal species, such as *Thermoplasma acidophilum*, lack the PAN complex (J. Maupin-Furlow, 2012). In these organisms, activation of the 20S proteasome involves other AAA+ ATPases such as the cell division control protein 48 (Cdc48), the Valosin-containing protein-like ATPase of Thermoplasma acidophilum (VAT) or the Archaeoglobus and Methanogenic Archaea (AMA) ATPase (Barthelme *et al.*, 2014; Djuranovic *et al.*, 2006; Forouzan *et al.*, 2012; Rockel *et al.*, 2002).

Similarly to the ubiquitin system used to address proteins to the 26S proteasome in eukaryotes, archaeal proteins to be degraded are tagged with small archaeal modifier proteins (SAMPs) (Humbard *et al.*, 2010). Sampylation is carried out by the ubiquitin-like activating protein of archaea (UbaA), which is homologous to the eukaryotic E1 enzyme. The C-terminus carboxyl group of the SAMP is covalently bound to the ε-amino group of Lys residues of the target protein (Hepowit *et al.*, 2016). This degron is recognized by the OB domain on PAN.

As highlighted above, the archaeal PAN-20S complex is homologous to the eukaryotic 26S proteasome machinery, and PAN shares 41-45% sequence similarity with human and yeast 19S ATPase subunits (Zwickl *et al.*, 1999). The archaeal macromolecular assembly is nevertheless simpler than its eukaryotic counterpart. Archaea have one single type of α and β subunits, while eukaryotes have fourteen genes coding seven different α subunits and seven different β subunits (Bard *et al.*, 2018; Groll *et al.*, 1997; Löwe *et al.*, 1995; Maupin-Furlow *et al.*, 2006). The archaeal proteasome also lacks the elaborate regulatory complex controlling the proteasome 26S activity. Instead, the ATPase is solely responsible for the proteolytic activity regulation. Combining the compositional simplicity of prokaryotic ATP-dependent proteases along with key architectural features of the 26S proteasome, the archaeal proteasome is hypothetized to be the evolutionary precursor of the eukaryotic proteasome (Majumder *et al.*, 2019). As discussed in **section 1.6** above, archaeal macromolecular assembly studies offer numerous advantages, and characterization efforts carried out on PAN-20S in the past decades have provided valuable insights on the eukaryotic 26S proteasome regarding structural, mechanistic and evolutionary aspects.

2.4.1.2 Lon

Lon was the first ATP-dependent protease to be identified (Swamy & Goldberg, 1981). Initially identified in *E. coli*, it was later shown to be ubiquitous in all domains of life. Lon proteases are homo-oligomeric enzymes forming hexameric ring structures encoded by a single gene. In contrast to the proteasome, it is a homooligomer, with the protease and AAA+ domains being

carried by the same subunit (Rotanova *et al.*, 2006). Lon belongs to the S16 MEROPS family, with a Ser-Lys dyad being responsible for its proteolytic activity (Botos *et al.*, 2004; Polgár, 2005). Three subfamilies A, B, and C can be identified based on sequences and structural features that will be discussed below (Rawlings *et al.*, 2018).

LonA is the most common subfamily and is mainly found in bacteria and eukaryotic organelles (mitochondria, chloroplast, and peroxisome), even though *Methanosarcinae* archaeon are reported to be harnessed with both LonA and LonB (Maupin-Furlow *et al.*, 2005). In addition to the canonical AAA+ module and proteolytic domain, LonA possesses an extended N-terminal region preceeding the AAA+ module (**Figure 16**) (Li *et al.*, 2005; Rotanova *et al.*, 2013). Beyond general protein degradation, LonA has been reported to be involved in the regulation of numerous biological processes, such as cellular division and differentation, DNA replication, pathogenocity or stress response (Boddicker & Jones, 2004; Kuroda, 2006; Schmidt *et al.*, 1994; Tsilibaris *et al.*, 2006). Despite its involvement in many crucial functions, LonA is not essential in *E. coli* under normal growth conditions due to the presence of the ClpAP, ClpXP and HslUV proteolytic complexes fulfilling overlapping functions (Wu *et al.*, 1999).

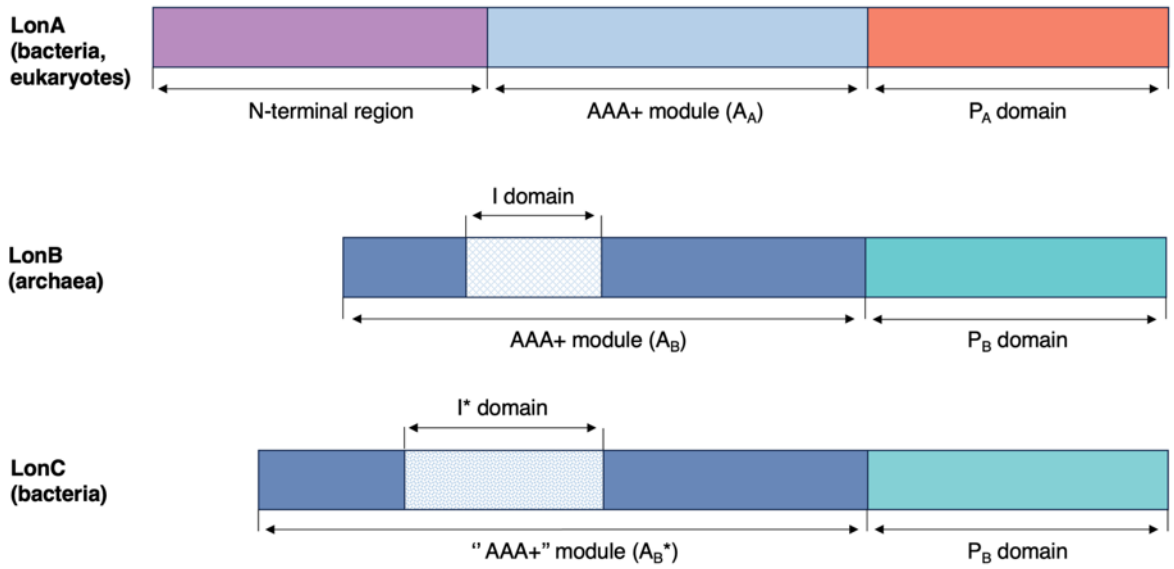


Figure 16: Domain organization of Lon proteases. All three subfamilies possess an AAA+ module and a proteolytic domain. LonA harbours a distinctive N-terminal region. Insertion domains I and I* are found in LonB and LonC, respectively. In the membrane-bound LonB, this insertion forms a transmembrane segment. Adapted from Kudzhaev *et al.*, 2022.

LonB is restricted to archaea and is mostly found in Euryarchaeota and DPANN species. The N-terminal region found in LonA is absent in LonB. Due to the presence of an insertion domain in the AAA+ module forming a transmembrane segment, LonB is a membrane-bound protease (**Figure 16**) (Rotanova *et al.*, 2004; Rotanova & Melnikov, 2008). This protease is expected to be a functional analog of the bacterial membrane-bound ATP protease FtsH, as the latter is absent in archaea. The majority of structural studies have been carried out on LonA proteins, but the crystal structure of *Thermococcus onnurineus NA1* Lon (*Ton*Lon) has been solved at a 2.0 Å resolution by Cha *et al.* (2010), revealing a hexagonal cylinder architecture with a large sequestered chamber accessible through an axial channel (**Figure 17**). The AAA+ domains exhibit different nucleotide states, with tight- and weak-binding conformations, reflecting the dynamic nature of ATP-driven protein unfolding and translocation processes within the protease. Unfortunately, no structural data is available for the membrane-anchoring region, that had to be deleted to generate soluble proteins.

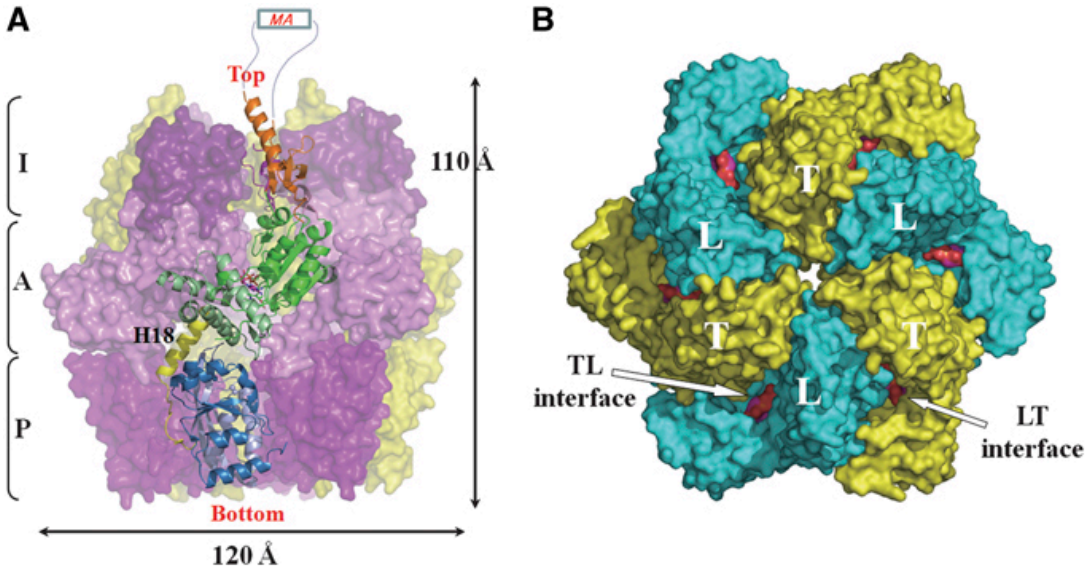


Figure 17: *Thermococcus onnurineus NA1* Lon exhibits a typical AAA+ protease architecture. (A)The hexameric structure of TonLon is shown with five subunits in surface representation and one monomer as a ribbon diagram. TonLon monomers are constituted of a protease domain (P), an AAA+ domain (A) and an insertion domain (I). The putative membrane-anchoring region MA is missing in this structure. (B) Surface representation of TonLon viewed from the top. Each monomer exists in two conformations : one with a tightly bound ADP (T-monomer) and one with a loosely bound ADP (L-monomer). Adapted from Cha *et al.*, 2010.

Proteases featuring a LonB-like proteolytic center and a similar ATPase component have been identified in some thermophilic bacteria. However, due to substitutions in the ATPase region, these enzymes can only bind ATP but not hydrolyze it, and selectively degrade

unfolded proteins in an energy-independent manner (Liao *et al.*, 2012, 2013; Rotanova *et al.*, 2004). These protein form the LonC subfamily, which to date only counts about 10 representatives and remains largely unexplored (**Figure 16**).

Recently, the determination of cryo–electron microscopy (cryoEM) structures of *Yersinia pestis* LonA in the absence and presence of substrate highligthed the existence of open- and closed-ring conformations, providing a better understanding of the protease activation mechanism (Shin *et al.*, 2020). In the absence of substrate, the monomers are arranged in an open left-handed spiral. The proteolytic active site is therefore accessible, but promiscuous protein degradation is prevented by steric hindrance, with a helix blocking the access to the catalytic dyad. Upon substrate binding, this helix unfolds into a loop, liberating the access to the catalytic site. The protease adopts a compact closed conformation in which the proteolytic activity is confined in a buried inner chamber (**Figure 18**). This dynamic switch is crucial for the controlled degradation of target proteins, driven by ATP hydrolysis.

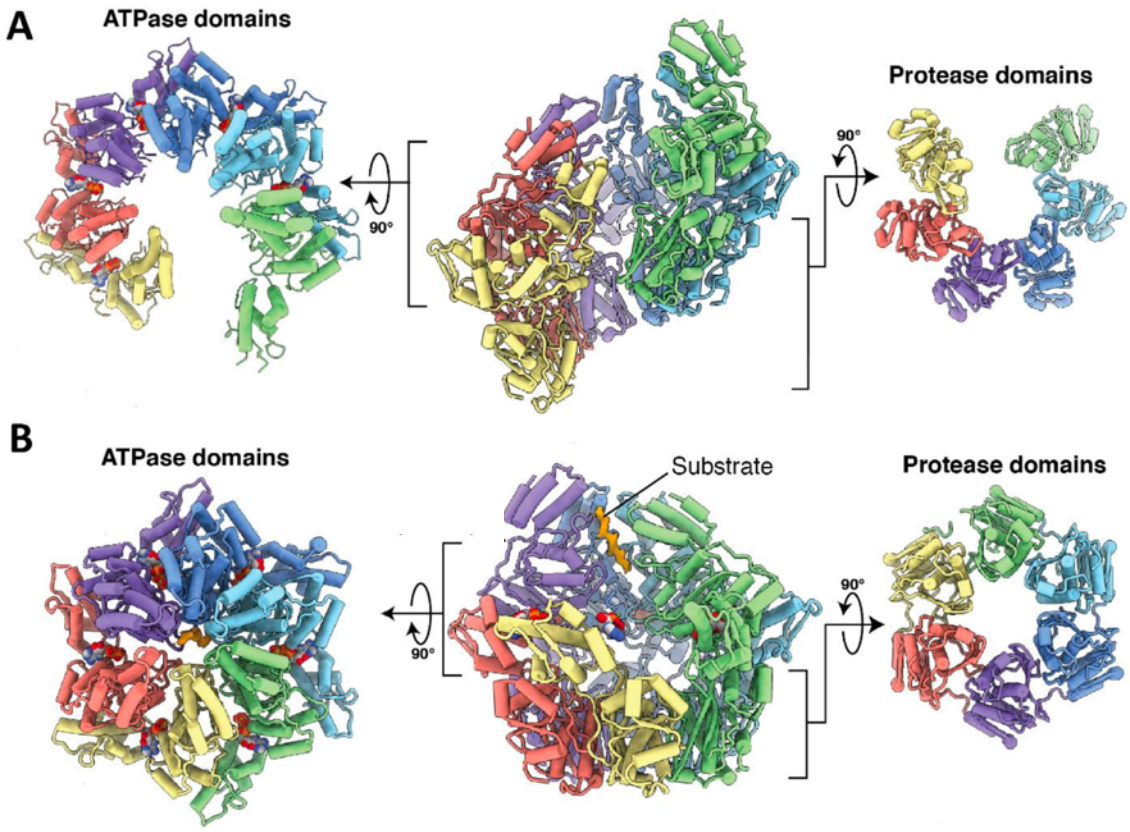


Figure 18: LonA proteolytic activity is controlled by conformational changes. (A) In its idle state, LonA adopts an open spiral conformation. (B) Substrate binding triggers rearrangement into the active closed-ring conformation. Adapted from Shin *et al.*, 2020.

2.4.2. Energy-independent peptidases

Initial protein degradation by the proteolytic complexes presented in the previous sections generates peptides ranging from 3 to 30 residues in length, with an average length of 6 to 9 amino acids (Kisselev *et al.*, 1999). The complete breakdown of those peptides is then achieved by ATP-independent peptidases, releasing free amino acids and thereby contributing to maintain the amino acids cellular pool. In addition to this amino acid recycling function, energy-independent peptidases are also involved in the regulation of numerous cellular processes (see **section 2.2**). There are a myriad of peptidases acting downstream of the proteasome and related proteolytic complexes, and it would be cumbersome to introduce them all. Nevertheless, tetrahedral and tricorn peptidases, commonly referred to as TET and TRI peptidases, stand out for their ability to oligomerize and form large self-compartmentalized assemblies. These complexes have been hypothetized to be functional analogs achieving the final steps of protein degradation in prokaryotes, each organism being presumed to possess either TRI or TET (Borissenko & Groll, 2005).

2.4.2.1 TRI peptidases

Searching for regulatory partners of the proteasome, Tamura *et al.* (1996) identified a large tricorn-shaped proteolytic complex, which was therefore named tricorn peptidase. First identified in the archaeon *Thermoplasma acidophilum*, the presence of TRI has been predicted *in silico* in Sulfolobales, Thermoplasmatales, and Thermoproteales. In bacteria, TRI have been identified in the *Bacteroides*, *Prevotella, Streptomyces, and Vibrio* genera (Ng'ong'a, 2017).

TRI crystallographic structure revealed the assembly of six 120 kDa monomers into a hexameric toroid consisting of two staggered and interlocking trimeric rings (**Figure 19**) (Brandstetter *et al.*, 2001). The ~ 720 kDa hexamer has a diameter of about 160 Å, a thickness of 88 Å, and features a central conical pore with an external and internal diameter of 45 Å and 20 Å, respectively (Brandstetter *et al.*, 2002). The *in vivo* formation of a higher oligomeric state has been reported, with 20 hexamers assembling into an astonishing 14,6 MDa icosahedral capside particle (**Figure 20**) (Walz *et al.*, 1997).

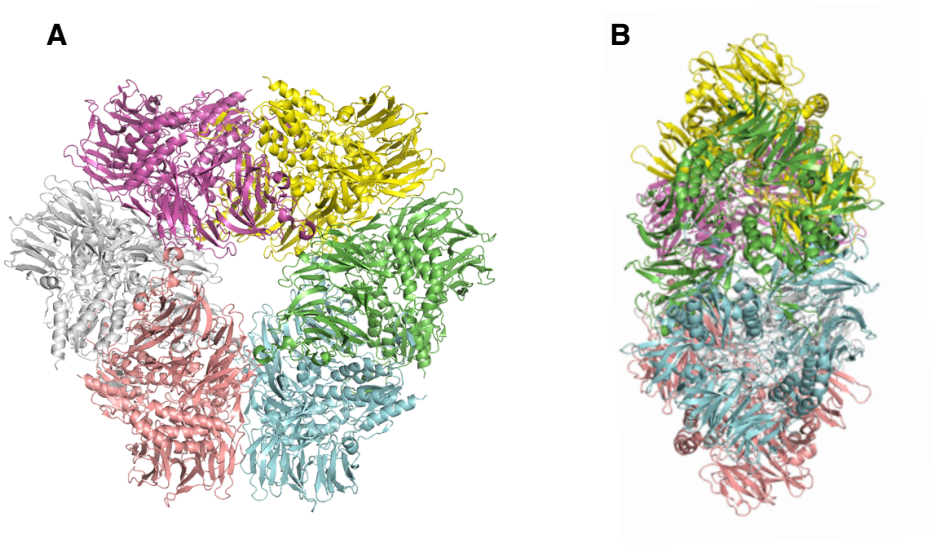


Figure 19: Structure of the tricorn peptidase in ribbon representation (Protein Data Bank (PDB) code 1K32, Brandstetter *et al.*, 2001). Six monomers assemble as two perfectly staggered trimeric rings, forming a ~ 720 kDa hexagonal toroid structure, here viewed from (A) the front or (B) the side.

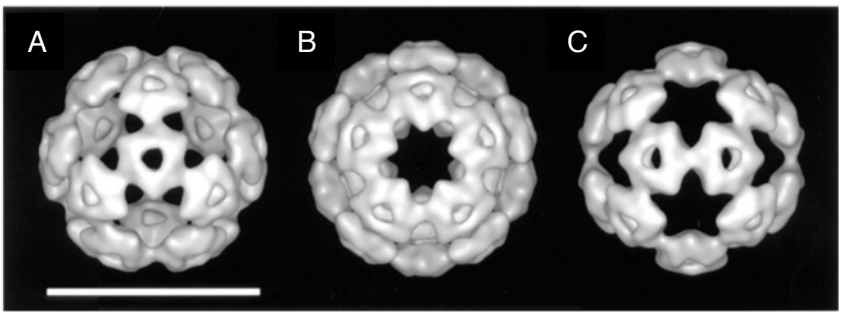


Figure 20: TRI capside structure solved by cryoEM. The 14,6 MDa assembly is composed of 20 hexamers. The scale bar represents 50 nm.

TRI is a serine endopeptidase belonging to the S41 family, according to the MEROPS classification. Its catalytic site is buried in the particle and is composed of four residues: Ser745, His746, Ser965, and Glu1023 (Brandstetter *et al.*, 2002). There is no evidence of direct interaction with the proteasome or other related proteolytic complexes, but functional characterization of TRI revelead that it hydrolyzes peptides up to 30 amino acids long, releasing smaller peptides of 2 to 4 residues. It has therefore been proposed that TRI operates downstream of AAA+ proteases (Tamura *et al.*, 2001). The resulting short peptides are then further processed by three auxiliary aminopeptidases interacting factors F1, F2 and F3. F1 is a 33.5 kDa proline iminopeptidase belonging to the S33 MEROPS family (Tamura *et al.*, 1996). F2 and F3 are 89 kDa homolgous metallopeptidases of the M1 family. F2 is a broad-spectrum peptidase active against neutral, hydrophobic, and basic amino acid substrates. F3 displays a

narrower substrate specificity, targeting acidic residues (Tamura *et al.*, 1998). The complementary substrate specificities of these three peptidases allows complete degradation of all TRI products.

2.4.2.2 TET peptidases

The first description of TET peptidases dates back to the early 2000s. At the ELMA laboratory, Bruno Franzetti's team was then conducting a research project focusing on the adaptation of halophilic archaea to saline stress, specifically searching for chaperone and protease complexes associated with saline stress. During the native purification of *Haloarcula marismortui* proteasome, a 500 kDa hollow tetrahedral complex was observed by negative-stain EM (**Figure 21**). Functional characterization of this 12-subunits complex revealed it to be an ATP-independent broad-spectrum aminopeptidase, which has been named TET for tetrahedral peptidase, in reference to the tricorn-shaped TRI peptidase (Franzetti *et al.*, 2002; Tamura *et al.*, 1996) (see **section 2.4.2.1**).

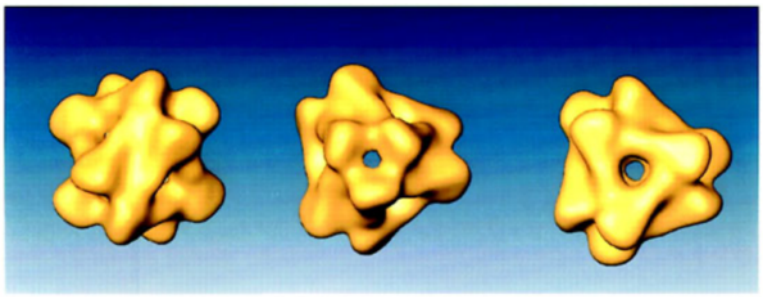


Figure 21: First three-dimensional structure of a TET peptidase. Negative-stain EM images were used to reconstruct a 17 Å resolution structure, revealing the assembly of 12 subunits forming a hollow tetrahedral complex. Adapted from Franzetti *et al.*, 2002.

Following this initial discovery, the study of TET peptidases has been extended to other organisms. As halophilic archaeal proteins require multimolar salt concentrations for proper folding and solubility, recombinant expression in *E. coli* is challenging and often unsuccessful. *H. marismortui* TET was no exception to this rule, so Bruno Franzetti's team therefore oriented its efforts toward the characterization of hyperthemophilic archaeon TETs (see **section 3.3**). As previously discussed in **section 1.6** above, these organisms constitute highly relevant models for structural biology and biochemistry studies. Research on other archaeal, bacterial, and eukaryotic TETs in several laboratories has contributed to deepen the understanding of this enzymatic family. TET peptidases are now known to be belonging to the M18 and M42 MEROPS families. The two families share a common architecture, characterized by the assembly of 12 subunits into the hollow tetrahedral complex displayed above. Each monomer

is composed of two domains: a catalytic domain and a dimerization domain (Chaikuad *et al.*, 2012; Russo & Baumann, 2004). As presented in **Figure 22**, M18 and M42 peptidases are unified by a common catalytic domain fold. However, they can be distinguished by the structure of their dimerization domains, adopting either a butterfly fold for M18 or a PDZ-like fold for M42 members (see **sections 3.1** and **3.2.1**). They also differ by their taxonomic distributions; M18 members are predominantly found in eukaryotes and bacteria, while M42 peptidases are restricted to prokaryotes (Paysan-Lafosse *et al.*, 2023).

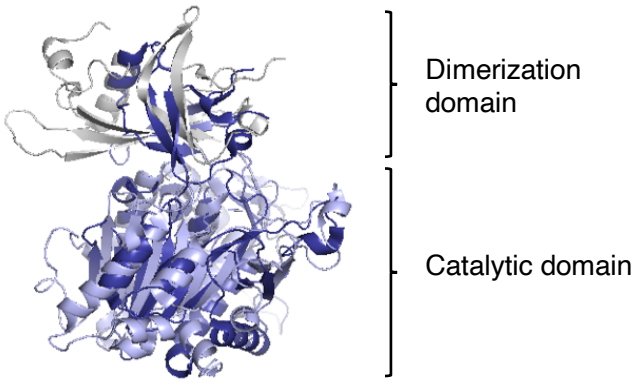


Figure 22: M18 and M42 monomers architecture. Monomers structural alignment of the M18 peptidase Ape1 from *S. cerevisiae* (grey and light blue, PDB code 4R8F) vs. the M42 peptidase PhTET2 from *P. horikoshii* (dark blue, PDB code 1XFO). Both catalytic domains share a common fold but the dimerization domains have distinct structures.

Given that several M42 aminopeptidases have been both functionally and structurally thoroughly characterized, it seemed most appropriate to focus on this family during the course of this PhD project. The pre-existing knowledge on these enzymes has been instrumental in conducting the comprehensive study presented in the following chapters. As the M42 family is central to this project, a detailed overview will be provided in **section 3** below.

Conversely, the aspartyl-aminopeptidase from *Pseudonomonas aeruginosa* is the only characterized prokaryotic M18, and archaeal and bacterial M18 peptidases have been very poorly investigated. Their eukaryotic counterparts received more interest. In particular, the conserved aspartyl-aminopeptidase DNPEP has been structurally and/or functionally in numerous species, including *Homo sapiens* (Chaikuad *et al.*, 2012; Park *et al.*, 2017; Teuscher *et al.*, 2007; Wilk *et al.*, 1998; Zheng *et al.*, 2016). Overexpression of the DNPEP gene has been observed in brain tumors, breast cancer, and colorectal cancer, but its contribution to these pathologies remains poorly understood (Larrinaga *et al.*, 2013; Martínez-Martos *et al.*, 2011; Mayas *et al.*, 2012).

3. M42 aminopeptidases

The work presented in the rest of the manuscript will be focused on archaeal M42 aminopeptidases, often designated as TET peptidases by a stretch of terminology. **In the remainder of this manuscript, both terms will be used to refer to M42 peptidases.**

3.1. The M42 family within the MH clan

According to the MEROPS classification of peptidases, the M42 family is part of the MH clan, along with the M18, M20, and M28 families. This clan gathers zinc-dependent exopeptidases with aminopeptidase, dipeptidase, tripeptidase, or carboxypeptidase activities (Rawlings *et al.*, 2018).

The clan type peptidase is the Ap1 aminopeptidase from *Vibrio proteolyticus* (Uniprot ID Q01693). VpAp1 is a M28 leucyl-aminopeptidase exhibiting a preference for large hydrophobic residues. Basic amino acids and proline can also be cleaved to a lesser extent, but glutamate, aspartate and cystein are not degraded (Chevrier & D'Orchymont, 2013). The extensive characterization of this enzyme allowed the development of biotechnological applications. For instance, it is used for the removal of the N-terminal methionine from recombinant therapeutic proteins, such as the recombinant human methionine-interferon α-2b used in the treatment of chronic hepatitis B and C and certain cancers (Pérez-Sánchez *et al.*, 2011). N-terminal methionine cleavage is a common post-translational modification essential for proper protein maturation. It is often necessary for subsequent modifications, such as N-terminal acetylation, which modulate protein stability, activity and localization. In the case of recombinant protein expression, the N-terminal methionine is often retained, presumably due to the saturation of methionine aminopeptidases (MAP) (Wingfield, 2017). This can result in reduced stability, impaired functionality, and altered localization. In the case of a therapeutic protein, *in vitro* digestion with purified MAP can be used to ensure that the final methionine is removed, guaranteeing the efficacy and safety of the molecule (Miller *et al.*, 1987). The use of VpAp1 MAP is particularly suitable for the human methionine-interferon α-2b, whose N-terminus sequence is Met-Cys-Asp-Leu-Pro; only the final Met will be cleaved since the following Cys and Asp residues cannot be hydrolyzed.

All enzymes of the MH clan share seven conserved catalytic residues defining the active site: His97, Asp99, Asp117, Glu151, Glu152, Asp179, and His256 according to VpAp1 numbering. In each active site, two metal ions are coordinated. The first ion is bound by

Asp117, Glu152, and His256, while the second is coordinated by His97, Asp117, and Asp179. Glu151 acts as the catalytic base and is responsible for the deprotonation of a water molecule bound by the two metal ion (Chevrier *et al.*, 1994). To date, the role of Asp99 is still poorly understood but there is evidence of its interaction with His97 and the second metal ion. Dutoit *et al.* (2020) proposed that this residue could reduce the Lewis acidity of the second metal ion, explaining the distinct role of the two metal sites in the catalytic mecanism (see **section 3.3.3** below).

In addition to sharing the same catalytic amino acids, the families of the MH clan are unified by a common catalytic domain fold. Conversely, their oligomeric state and associated oligomerization domains structure differ greatly (**Figure 23**). M28 peptidases are either monomeric or dimeric. In this family, oligomerization involves four α helices adjacent to the active site. Most enzymes of the M20 family form dimers, but some have been reported to assemble as tetramers, such as *Pseudomonas aeruginosa* carboxypeptidase CG2 (PDB code 1CG2). Monomers assembly is mediated by a dimerization domain distinct from the catalytic

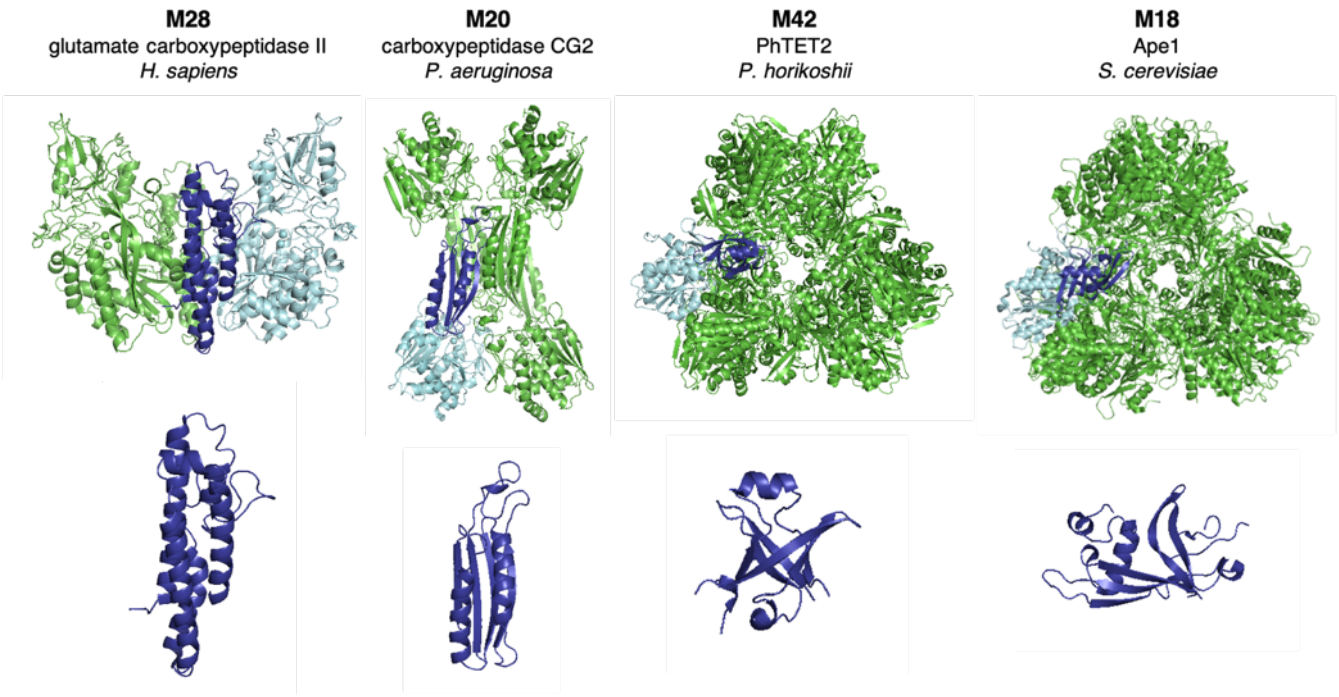


Figure 23: Diversity of the oligomeric states of MH clan enzymes. On the top row, quaternary structures of *H. sapiens* glutamate carboxypeptidase II (PDB code 1Z8L), *P. aeruginosa* carboxypeptidase CG2 (PDB code 1CG2), *P. horikoshii* PhTET2 (PDB code 1Y0R) and *S. cerevisiae* Ape1 (PDB code 4R8F). For each complex, one monomer is colored in blue, with the catalytic and dimerization domains in pale and dark blue, respectively. For greater convenience, each dimerization domain fold is represented alone in the lower row.

domain, consisting in a four-stranded β sheet flanked by two α helices. For tetramer-forming enzymes, dimers interact through the α helices, forming a four-layer (βααβ)-sandwich (Rowsell *et al.*, 1997). Higher oligomeric states are achieved in the M18 and M42 families. As presented in **section 2.4.2.2**, these enzymes are typically forming 12-subunits tetrahedral particles. Despite this common quartenary structure, protomers assembly is mediated by distinct dimerization domain structures, adopting either a butterfly fold for M18 peptidases or a PDZ-like fold for the M42 family. Dimers assemble into dodecamers or even tetracosamers. The structure of these complexes will be more extensively covered in **section 3.2** below.

3.2. The hallmark architecture of TET peptidases

The most remarkable feature of TET aminopeptidases is undoubtedly the distinctive hollow tetrahedral architecture adopted by the dodecameric complex. In the following sections, the structure of the different oligomeric forms adopted by these enzymes will be presented.

3.2.1. Structure of the monomer

M42 peptidase monomer is typically around 340 residues long, with a molecular weight of 37 to 40 kDa. The high-resolution structures obtained for several enzymes have revealed the common structure adopted by M42 peptidase monomers. All subunits consist of two separated domains, a dimerization and a catalytic domain, which have been briefly outlined in previous sections. The dimerization domain adopts a PDZ-like fold made of a mixed six-stranded β-sheet and two α helices. The catalytic domain adopt the fold common to all MH clan enzymes, which consists of a central β-sheet encircled by eight α helices, forming a three-layer αβα sandwich. A β-sheet extension, also present in M18 and M20 families, is involved in dimer formation through interaction with the dimerizaition domain of another subunit (**Figure 24A**).

As presented in **section 3.1** above, the active site itself consists of seven catalytic residues, five of which are involved in the coordination of two metallic ion co-factors. In *Pyrococcus horikoshii* TET2 (PDB code 1Y0R), the catalytic residues are His68, Asp70, Asp182, Glu212, Glu213, Asp235, and His323 (**Figure 25**). The existence of a third metallic site formed by Thr232, Asp254, Glu281, Thr285 and Thr287 has been reported in *Pyrococcus furiosus* TET3 by Colombo *et al.* (2016). This additional site could extend the substrate specificity of the enzyme.

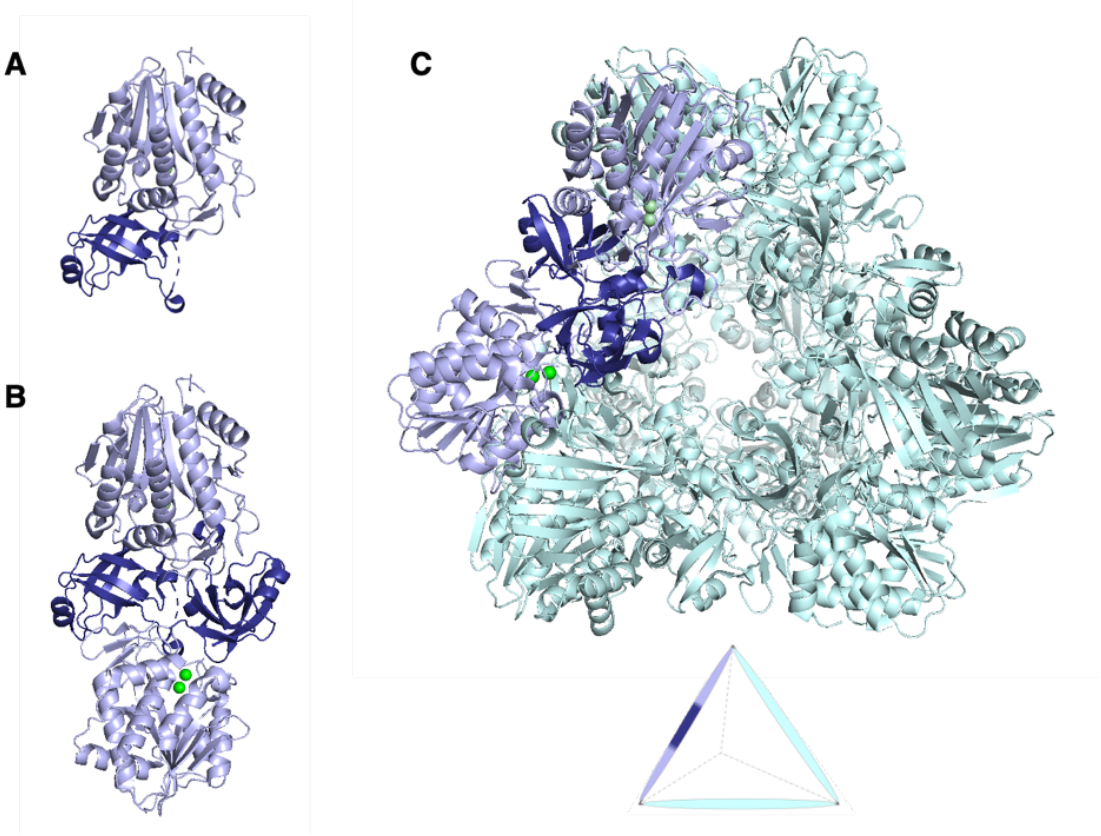


Figure 24: Main oligomeric forms adopted by M42 aminopeptidases. The depicted structures correspond to *P. horikoshii* PhTET2 (PDB code 1Y0R). (A) The monomer is composed of a dimerization domain (dark blue) and a catalytic domain (light blue). (B) Dimers are formed through the interaction of the dimerization domain of one subunit with the catalytic domain of another subunit. (C) Dimers self-assemble as a dodecameric tetrahedral particle.

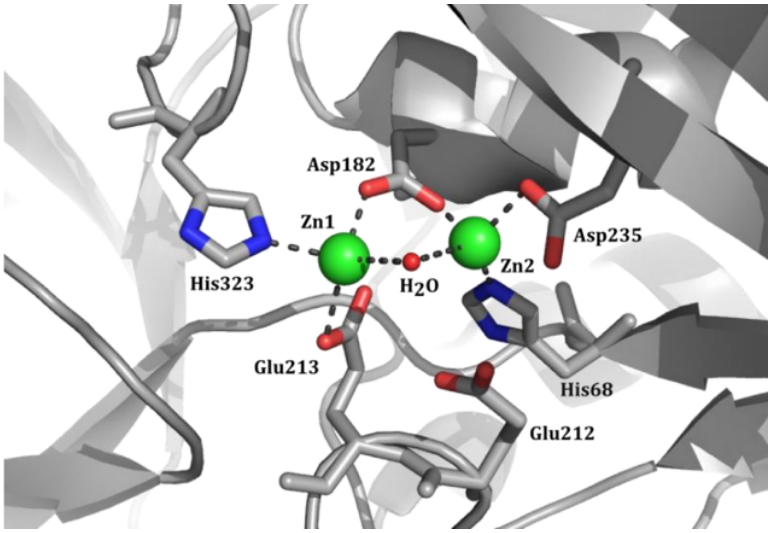


Figure 25: Representation of the active site of *Pyrococcus horikoshii* TET2 aminopeptidase (PDB code 1Y0R). Residues involved in metal ion coordination and catalysis are represented as sticks. The metal cofactors are shown as green spheres. The oxygen of the water molecule is represented by a red sphere. The metal ions Zn_1 and Zn_2 are coordinated by the residues Asp182, Glu213, His323 and His68, Asp182, Asp235, respectively. Glu212 is the catalytic base. A water molecule is bound by the metal ions.

All M42 peptidases share the same active site, but they exhibit distinct preferences for specific amino acid substrates. This specificity could be dictated by the composition of the substrate-binding specificity pocket S_1 that is adjacent to the active site. In PhTET2, it consists of Thr237, Ile238, Glu291, Leu293 and Ile322. The link between S1 pocket composition and substrate specificity will be further discussed in **section 3.3.2** below.

3.2.2. Formation of the dimer, building-block of TET peptidases

A few hydrogen bonds occur between two dimerization domains, but dimers are mainly formed by the interaction of the β-sheet extension of the catalytic domain of one momoner with the dimerization domain of another monomer, forming an extensive hydrogen bonds and salt bridges network (Schoehn *et al.*, 2006) (**Figure 24B**). Due to this large interaction surface, the dimer is highly stable and the existence of this oligomeric form has been observed *in vivo* (Appolaire *et al.*, 2013; Appolaire, Durá, *et al.*, 2014). The dimer is therefore considered to be the building-block of the tetrahedral assembly, and the dodecamer is often described as the assembly of six dimers. This oligomeric form is catalytically active, although less active than the dodecamer (Colombo *et al.*, 2016; Lee *et al.*, 2012).

3.2.3. The tetrahedral dodecamer

The self-assembly of six dimers is driven by the metal ion cofactors (Dutoit *et al.*, 2019). The removal of the metal ions, for instance by treatment with a chelating agent or by formation of metal hydroxides in alkaline conditions often leads to the disassembly, or at least destabilization of the complex (Colombo *et al.*, 2016; Moser *et al.*, 1970; Rosenbaum *et al.*, 2011). Dimers interaction results in the formation of a 12-subunit hollow tetrahedral particle with a molecular weight of about 450 kDa and a diameter approaching 135 Å (**Figure 24C**). The dimers are positioned along the edges of the complex and interact through extensive contacts between the catalytic domains. These interactions have been estimated to be more important for the stability of the dodecamer than the interactions between two monomers (Krissinel & Henrick, 2007). The faces of the tetrahedron are delimited by three dimers creating a central opening (~20 Å diameter), assumed to be the substrate entrance, leading to a wide inner cavity (60 Å diameter). All catalytic sites are oriented toward this chamber. The protease activity is thus confined within the complex, protecting the cell from uncontrolled peptide degradation. The apexes of the tetrahedron, containing the active sites, are formed by three monomers delineating smaller openings (~7 Å) believed to be the product exit ways (Appolaire *et al.*, 2016; Borissenko & Groll, 2005).

Inside the catalytic chamber are located twelve loops, one for each monomer, hypothetized to be involved in substrate trafficking and specificity. Due to their high flexibility and mobility, these parts are either poorly modelled or missing in 3D structures determined so far. Their position in the particle and functional importance for the catalytic activity of TET peptidases have been investigated using Nuclear Magnetic Resonance (NMR) (Gauto *et al.*, 2022). However, the mechanisms of substrate trafficking and processing remain debated and will be further discussed in **section 3.2.5** below.

Appolaire *et al.* (2013) proposed an assembly process for the dodecamer involving hexameric intermediates formed by the interaction of three dimers through their catalytic domains. Closed (tricorn-like form) and open (Z-form) conformations of the hexamer would coexist, with the Z-form conformer being able to self-associate, forming the tetrahedral dodecamer (**Figure 26**). More recently, Macek *et al.* (2017) challenged this assembly process and proposed a stochastic model instead, supported by the observation of oligomeric forms ranging from dimer to dodecamer by increments of two subunits. Interestingly, the existence of octamers was also reported in Appolaire's study. The discrepencies between the two models could be linked to the mutations introduced by Appolaire *et al.* to slow down the oligomerization process, possibly altering dimers interaction. The Macek *et al.* model may also suffer from technical biases, as they used harsh physico-chemical conditions (high EDTA concentrations and low pH) to generate the TET precursors, which could destabilize the particles. Evidence from a Small-Angle Neutron Scattering (SANS) study on TET2-3 heterooligomers aligns with the Appolaire model (Appolaire, Girard, *et al.*, 2014). In this study, no mutations were introduced to slow down the oligomerization process.

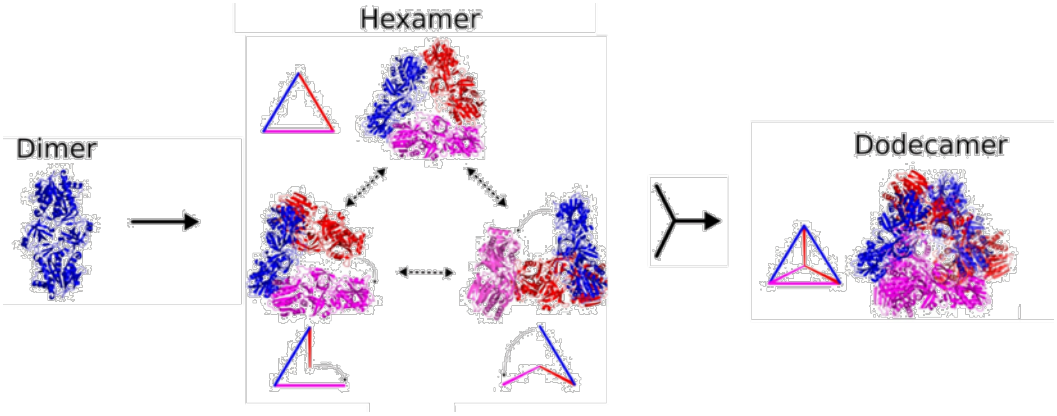


Figure 26: TET dodecamer assembly model. Dimers are the building-blocks of the tetrahedron and associate to form closed hexamers. This closed conformation (tricorn) coexists with two open conformations (Z-form) that can assemble into a dodecameric particle. Adapted from Appolaire *et al.*, 2013.

Of note, the formation of a higher oligomeric form has been reported for *P. horikoshii* PhTET1, with 24 subunits assembling to form a 800 kDa octahedron-shaped tetracosamer (Schoehn *et al.*, 2006). To date, this constitutes the only reported instance of such oligomer.

3.2.4. Heterocomplex formation

The *in vitro* and *in vivo* existence of PhTET2-PhTET3 heterocomplexes have been demonstrated in *P. horikoshii* (Appolaire, Durá, *et al.*, 2014; Appolaire, Girard, *et al.*, 2014). Comparaison of homo- and heterododecamer peptidolytic activites revelead an enhanced efficiency of the heterocomplexes. Interstingly, PhTET1 was not found in these heterooligomers. As PhTET4 was only discovered at a later stage, its involvement in heterocomplexes formation was not assessed.

The occurrence of M42 peptidases heterooligomers displaying increased peptide destruction properties has also been reported in the bacteria *Geobacillus staerothermophilus* and *Symbiobacterium thermophilum* (Kumaki *et al.*, 2011; Stoll *et al.*, 1972). Dutoit (2021) suggested that such complexes could also exist in *Thermotoga maritima*. During the study of the three M42 peptidases of this species, he was unable to obtain *in vitro* homododecamers for TmPep1048 and TmPep1049 proteins. He postulated that these enzymes may only exist in a dodecameric form in heterocomplexes and suggested coexpressing the entire TM_1048-TM_1049-TM_1050 operon to ascertain this hypothesis.

3.2.5. Substrate trafficking and processing within the TET particle

The EM observation of the internal organization of the PhTET1 complex, combined with the analysis of surface properties of archaeal particles (PhTET2 and PhTET3), bacterial (MHJ_0125), and eukaryotic (hDNPEP) particles, have led the proposal of a peptide trafficking model (**Figure 27**).

The TET particle features two types of pores: four located on the faces (~20 Å in diameter) and four at the vertices (~15 Å in diameter). The large pores on each of the four faces of the tetrahedron are considered to be the entry pores, and provide access to four channels that intersect at the center of the particle. The distribution of electrostatic charges within the channel is highly contrasted, and would be guiding the substrate through the particle and toward the catalytic pockets. In the glutamyl-aminopeptidase PhTET1, the electrostatic potential is negative (Asp74, Glu75) at the entrance of the channel. As the substrate

progresses into the particle, the potential becomes positive and then negative again before reaching the intersection of the channels. Lys122 is part of a flexible loop that would aid in directing the substrate toward the catalytic pockets located at the tetrahedron's apices (Schoehn *et al.*, 2006).

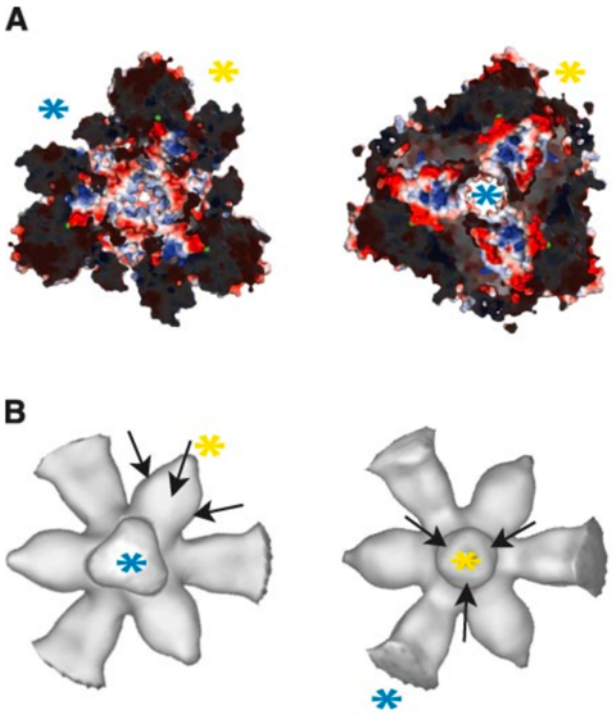


Figure 27: Interior of the PhTET1 particle from *P. horikoshii* (PDB ID: 2WYR). (A) Cross-section of the tetrahedral structure. The surface of the particle is coloured according to the negative (blue) or positive (red) electrostatic potential. Left: seen from an entrance pore. Right: seen from the exit pore. The blue and yellow stars indicate the entry and exit pores, respectively. (B) Schematic representation of the internal channels traversing PhTET1 from the entrance pores (blue stars) to the exit pores (yellow stars). The three arrows indicate the three active sites adjacent to the exit pores.

The loop, missing in high-resolution crystal structures due to its high flexibility, was recently modelled for the first time using magic-angle spinning NMR (Gauto *et al.*, 2022). In PhTET2, it fills close to 30% of the catalytic chamber volume. This flexible element would play a crucial role in substrate trafficking by stabilizing ligands in the active site, and would also be involved in the activation of the catalytic water molecule. Its dynamic nature and electrostatic charge distribution would guide substrates through the wide central chamber and into the catalytic pockets.

These pockets open to the exterior *via* small pores located at the vertices of the complex, which are thought to be the exit pores for the released free amino acids. Another hypothesis proposed for PhTET2 by Borissenko *et al.* (2005) suggests that the products are redirected into the central cavity and expelled through small pores surrounding the large entry pore.

3.3. M42 aminopeptidases activity

3.3.1. Functional flexibility in M42 peptidases

M42 peptidases are aminopeptidases known to sequentially degrade the N-terminal residues of peptides up to 40 amino-acids in length (Franzetti *et al.*, 2002). These enzymes exhibit a non-processive mechanism, releasing the substrate after each cleavage event (Durá *et al.*, 2005, 2009; Schoehn *et al.*, 2006).

The existence of N-deblocking activity, which refers to the ability to degrade N-terminally modified peptides, is debated. According to Franzetti and coworkers, *P. horikoshii* TETs would be devoid of deblocking activity (Appolaire *et al.*, 2016). Nonetheless, several studies have documented N-deblocking activity in various M42 aminopeptidases, suggesting a broader functional capability for these enzymes (Ando *et al.*, 1999; Jia *et al.*, 2011; Kamp *et al.*, 1998; Tsunasawa, 1998).

Additionnaly, TET peptidases could possess secondary, moonlighting activities that extend beyond peptide bond hydrolysis. One such putative activity is glycoside hydrolase function, first reported by Kobayashi *et al.* (1993). In their study, a new glycoside hydrolase, designated CelM, was isolated from *Clostridium thermocellum*. This enzyme was observed to degrade carboxymethylcellulose. Curiously, CelM amino acid sequence did not match any known glycoside hydrolase family in the CAZy database, but was instead classified under the M42 family. Additionally, the sequence of CelM does not contain any tryptophan residue, which is commonly found in vicinity of the active site of glycoside hydrolases (Nakamura *et al.*, 2013). Further investigations into CelM activity produced conflicting results. A study published by Dutoit *et al.* (2012) revealed CelM aminopeptidase activity while failing to replicate the carboxymethylcellulase activity reported by Kobayashi and colleagues. Despite the contentious nature of their glycoside hydrolase activity, M42 aminopeptidases have demonstrated the ability to bind carbohydrates. For example, MHJ_0125 from *Mycoplasma hyopneumoniae* can bind heparin without degrading it (Robinson *et al.*, 2013). Conversely, Franzetti and coworkers never observed any moonlighting activity for any of the TET peptidases they have been studying.

3.3.2. TET peptidases exhibit contrasted substrate specificities

The substrate specificities of TET peptidases are typically determined using synthetic substrates, which consist of amino acids coupled to a chromogenic or fluorogenic reporter

molecule mimicking a peptide bond. This method offers a convenient alternative to the more time-consuming and labor-intensive determination of peptide digestion profiles by high pressure liquid chromatography (HPLC).

As summarized in **Table 2**, the functional characterization of several archaeal and bacterial representatives revealed strongly heterogenous activity spectra. Some enzymes, such as *Haloarcula marismortui* TET or *Geobacillus stearothermophilus* aminopeptidase I, exhibit a broad-spectrum activity (Franzetti *et al.*, 2002; Roncari & Zuber, 1969). In contrast, other M42 peptidases exhibit much narrower susbtrate specificities, targeting either neutral, acidic or basic residues. For instance, *P. horikoshii* peptidases PhTET1, PhTET2, and PhTET3 are classified as glutamyl-, leucyl-, and lysyl-aminopeptidases, respectively (Durá *et al.*, 2005, 2009; Durá & Franzetti, 2013). The fourth TET of this organism, PhTET4, is the only strict glycyl-aminopeptidase described to date in this family (Basbous *et al.*, 2018). Given that only ten archaeal enzymes have been functionally characterized so far, and that most are from closely related species, it is uncertain whether our current knowledge fully accounts for the functional diversity of these enzymes, or if novel substrate specificites remain to be discovered.

Despite these marked preferences observed on short synthethic substrates, the activity spectrum of M42 peptidases seems to be broadening with substrate chain length. For example, the glutamyl-peptidase PhTET1 is able to cleave an Ala-Ala-Ala peptide (Schoehn *et al.*, 2006).

Furthermore, substrate specificities have not been determined in the same way for all enzymes. For instance, TNA1_DAP1 and TNA1_DAP2 are classified as leucyl-aminopeptidases, but their activity has been evaluated on leucine and methionine only (Lee *et al.*, 2012).

The substrate specificity of a M42 aminopeptidase cannot be inferred from its primary sequence. It is conferred by the tree-dimensional folding of the particle, which enables the formation of the specificity pocket S_1. The dimension of this catalytic pocket and the charge properties of the residue delineating the pocket appear to be driving the specificity of the enzyme. Broad-spectrum M42 peptidases are characterized by a large S_1 pocket bordered by hydrophobic residues such as Val236, Ile238, Leu293 and Ile322 in PhTET2 (Borissenko & Groll, 2005; Durá *et al.*, 2005). The lysl aminopeptidase PhTET3 preferentially cleaves lysine, and less effectively arginine. Non-polar amino acids such as leucine, phenylalanine and alanine are hydrolysable to a lesser extent. The S_1 pocket of this enzyme is bordered

Domain	Species	Enzyme	Substrate specificity	Cofacor	PDB	Reference
Archaea	*H. marismortui*	HmTET	Broad spectrum Gln	-	-	(Franzetti et al., 2002)
	P. horikoshii	PhTET1	Glu	Co^{2+}	2CF4	(Durá and Franzetti, 2013; Schoehn et al., 2006)
		PhTET2	Leu	Co^{2+}	1Y0R 1XFO	(Borissenko and Groll, 2005; Durá et al., 2005; Russo and Baumann, 2004)
		PhTET3	Lys	Co^{2+}	2WZN	(Durá et al., 2009)
		PhTET4	Gly	Ni^{2+}	-	(Basbous et al., 2018)
	P. furiosus	PfTET3	Lys	Co^{2+}	4X8I	(Colombo et al., 2016)
	M. jannaschii	MjTET	Broad spectrum, Leu	Co^{2+}	-	(Basbous, 2016)
	D.amylolyticus	APDkam589	Non polar residues	Mn^{2+}	4WWV	(Petrova et al., 2015; Slutskaya et al., 2012)
	T. onnurineus	TNA1_DAP1	Leu	Mn^{2+}	-	(Lee et al., 2007, 2012)
		TNA1_DAP2	Leu	-	-	(Lee et al., 2012)
	Thaumarchaeota	TET	-	-	5DS0	-
Bacteria	*T. maritima*	TmPep1048	not detected	-	1VHO	(Dutoit, 2021)
		TmPep1049	not detected	-	2FVG	(Dutoit, 2021)
		TmPep1050	Leu	Co^{2+}	6NW5	(Dutoit et al., 2020, 2019, 2012)
	E. coli	ypdE	Broad spectrum Non polar residues	Co^{2+}	-	(Dutoit, 2021)
		sgcX	Broad spectrum Non polar residues	-	-	(Dutoit, 2021)
		frvX	Broad spectrum Non polar residues	-	-	(Dutoit, 2021)
	S. pneumoniae	SpPepA	Glu	-	3KL9	(Kim et al., 2010)
	L. lactis	Aminopeptidase A	Glu	Co^{2+}, Zn^{2+}	-	(Bacon et al., 1994; Exterkate and De Veer, 1987)
	S.thermophilum	Pep1079, Pep1080, Pep1081	Lys	Co^{2+}	-	(Kumaki et al., 2011)
	M. hyopneumoniae	MHJ_0125	Glu	Co^{2+}	-	(Robinson et al., 2013)
	G. staerothermophilus	Aminopeptidase I	Broad spectrum	Co^{2+}	-	(Moser et al., 1970; Roncari et al., 1972; Roncari and Zuber, 1969)
	S. flexneri	FrvX	-	-	1YLO	-

Table 2: Overview of characterized M42 aminopeptidases. When available, the determined substrate specificity, metallic cofactor and structure PDB identifier are given.

by Asp262, Thr240, Thr295 and Thr297, the presence of Asp262 allowing to accommodate the positive side chain of lysines and arginines (Durá *et al.*, 2009). Accordingly, Chaikuad *et al.* (2012) suggested that the conserved Lys374 residue could be a marker of the aspartyl-aminopeptidase activity of hDNPEP, PaAP, PfM18AAP and bDNPEP. In PhTET1 and SpPepA, both classified as glutamyl-aminopeptidases, this lysine is substituted by an arginine, which also carries a positively charged side chain.

Currently, predicting the substrate specificity of TET peptidases based on primary sequence signatures or well-defined structural features remains challenging; therefore, functional characterization of these enzymes is still required (Basbous *et al.*, 2018; Durá *et al.*, 2005, 2009; Durá & Franzetti, 2013; Schoehn *et al.*, 2006). However, as more TET peptidases are characterized, the accumulation of comprehensive functional and structural data may eventually facilitate the identification of markers for substrate specificity and allow for accurate predictions.

3.3.3. At the molecular scale: catalytic mechanism and metallic cofactor

The catalytic mechanism of TET peptidases, common to all enzymes in the MH clan, is shown in **Figure 28**, and described using the protease cleavage site nomenclature introduced by Schechter and Berger (1967) presented in **section 2.3.3** above.

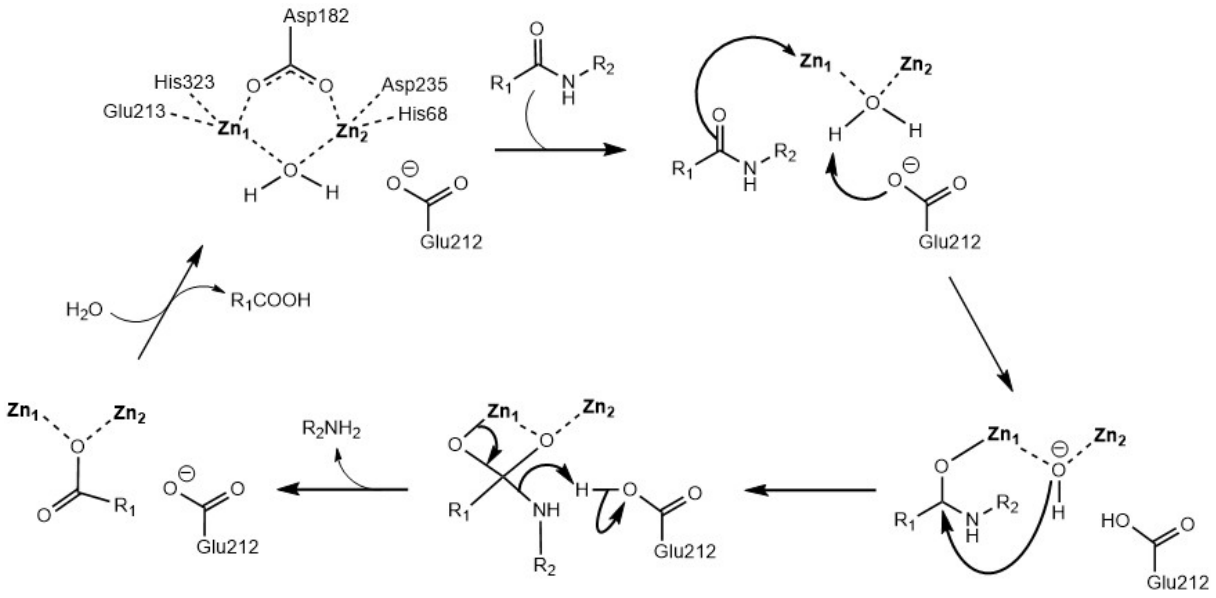


Figure 28: Proposed catalytic mechanism for TET aminopeptidases. The numbering of the catalytic residues represented corresponds to *Pyrococcus horikoshii* PhTET2 (PDB code 1Y0R). The ligands of the two zinc ions are shown in the first panel only (upper left corner).

The carboxyl oxygen of the P_1 residue is bound by Zn_1. Concomitantly, the catalytic base Glu212 deprotonates the water molecule bound by the two metal ions. The hydroxide ion formed then attacks the carbonyl group of the peptidic bond of the substrate. Glu212, in its protonated form, acts as an acid and yields its proton to the nitrogen of the amide group of the P'_1 residue of the substrate, resulting in the rupture of the peptidic bond. Finally, a water molecule allows the product to be released and the catalytic site to be regenerated (Auld, 2004).

The two metal ion cofactors coordinated in the active site are critical for the activity of the enzyme. As presented in **Table 2**, the metal cofactor preferences of M42 peptidases were investigated in several biochemical studies, revealing that most of these metallopeptidases are cobalt-activated enzymes (Ando *et al.*, 1999; Bacon *et al.*, 1994; Colombo *et al.*, 2016; Durá *et al.*, 2005, 2009; Dutoit *et al.*, 2019; Kumaki *et al.*, 2011; Robinson *et al.*, 2013). Nevertheless, some enzymes do not follow this general rule. For example, PhTET4 and APDkam589 are preferentially activated by nickel and manganese, respectively (Basbous *et al.*, 2018; Slutskaya *et al.*, 2012). As illustrated by the work of Colombo *et al.* (2016), the effect of divalent metals are temperature- and concentration-dependent. In the case of PfTET3, cobalt acts as an inhibitor at 20°C and 50°C, but as an activator at 85°C. Additionally, the activation is more than twice as significant at a concentration of 0.1 mM compared to 1 mM.

Although cobalt is preferred by most enzymes, zinc is often modeled in the active site of many M42 peptidase structures. This can be explained by experimental biases, such as the heterologous production of proteins in *E. coli*. The bacterial metallome is estimated to be approximately eight times richer in zinc than in cobalt (Barton *et al.*, 2007). The abundance of zinc and the lower availability of cobalt can thus explain the presence of zinc in the active sites. Rosenbaum *et al.* (2011) measured the metal ion content of PhTET2 recombinantly produced in *E. coli.* Even though this enzyme has been characterized as cobalt-activated, zinc represented the majority of the metal bound, followed by calcium and iron.

3.4. Beyond their structure: M42 peptidases in a biological context

3.4.1. The elusive distribution and phylogeny of M42 peptidases

M42 peptidases are reported to be restricted to prokaryotes, being present in most archaeal phyla but less prevalent in bacteria (Rawlings & Bateman, 2019). Initial characterization of archaeal TETs concentrated on hyperthermophilic species within the order Thermococcales, which facilitated extensive structural biology studies (see **section 3.3**). As in the case of the

archaeon *P. horikoshii,* up to four homologs are found in these species (Basbous *et al.*, 2018; Durá *et al.*, 2005, 2009; Durá & Franzetti, 2013). Conversely, the Halobacteriales archaeon *H. marismortui* only possesses a single TET. Little is known about the copy number per organism in other archaeal lineages, leaving significant gaps in our understanding of this enzymatic family. Moreover, due to a low primary sequence conservation (Appolaire *et al.*, 2016), along with frequent misannotations as cellulases or endoglucanases (Dutoit *et al.*, 2012; Slutskaya *et al.*, 2012), the identification of M42 peptidases in genomic or proteomic databases remains challenging, hindering the assessment of their prevalence and distribution across Archaea.

A preliminary phylogenetic study conducted as part of Alexandre Appolaire's PhD thesis and in collaboration with Simonetta Gribaldo suggests the existence of two distinct lineages with either archaeal or bacterial origins, and indicates possible HGTs between these lineages (Appolaire, 2014). Nevertheless, the evolutionary history of M42 peptidases remains enigmatic and would benefit from a more in-depth investigation.

Overall, in contrast to the thorough study of their structure, as evidenced by the in-depth description provided in the previous sections, key aspects such as the prevalence, ecological and taxonomic distributions, and evolution of M42 peptidases have been scarcely investigated. These topics represent significant opportunities for further research, which could lead to a more comprehensive understanding of these enzymes.

3.4.2. TET peptidases in biological systems : cellular localization and potential roles

TET peptidases do not contain any signal peptide indicative of membrane translocation or secretion, and are believed to be intracellular peptidases. In archaea, this putative cellular localization is supported by the absence of the three M42 peptidases of *Thermococcus barophilus,* encoded by the TERMP_00689, TERMP_00728, and TERMP_01137 genes, in its secretome (Batour *et al.*, 2023). However, contrary to long-held beliefs, proteins can be exported without a classical N-terminal signal peptide through leaderless secretion pathways (Bendtsen *et al.*, 2004; Maffei *et al.*, 2017). Several studies identified membrane-bound or surface-presented M42 aminopeptidases. For instance, Roncari and Zuber (1969) reported that GsApI remains tightly bound to the membrane after cell disruption. In *M. hyopneumoniae*, MHJ_0125 is mainly located in the intracellular compartment, but is also found bound to the membrane or secreted into the extracellular environment (Robinson *et al.*, 2013).

Genetic knock-down experiments, consisting in abolishing the expression of a specific gene to study the resulting effects on the organism, are typically used to determine the function on a gene. To this day, no genetic knock-down experiments have ever been performed on M42 peptidases, and their cellular function are still undetermined. Even though the physiological functions of membrane-bound or secreted M42 peptidases remain obscure, several hypotheses have been formulated concerning the intracellular role of these enzymes.

As suggested in **section 2.4.2** above, these peptidases could be among the ATP-independent peptidases intervening downstream of the proteasome or other related proteolytic complexes. They would therefore be involved in the non-specific degradation of peptides generated by AAA+ proteases, participating in proteins turnover and amino acid homeostasis (Borissenko & Groll, 2005; Franzetti *et al.*, 2002). In accordance with this hypothesis, the overexpression of two M42 aminopeptidases has been reported in the archaeon *Thermococcus kodakarensis* under oxidative and heat stresses (Jia *et al.*, 2011).

They could also be responsible of the degradation of exogenous or endogenous peptides for metabolic or energetic purposes, contributing to metabolic pathways such as carbon fixation or amino acid fermentation (Aklujkar *et al.*, 2014; Lloyd *et al.*, 2013; Ward *et al.*, 2002). In aerobic conditions, glutamate can be used to fuel the Krebs cycle by being converted to α-ketoglutarate through transamination and oxidative deamination. The resulting α-ketoglutarate can then enter the Krebs cycle, contributing to energy production (Plaitakis *et al.*, 2017). In *Pyrococcus furisous,* the overexpression of a gene encoding a PhTET1 homologue has been observed when cells are cultured in a peptide-enriched medium, corroborating this hypothesis (Schut *et al.*, 2003).

Finally, the narrow substrate specificities observed for some M42 aminopeptidases may be indicative of selective, targeted degradation of specific proteins. They could intervene in regulatory processes such as protein or peptide maturation.

Among the different hypotheses put forward, it is plausible that multiple may hold true. Given the vast diversity of M42 peptidases, they likely fulfill a variety of roles. Further research is required to elucidate their physiological function(s) and to determine the full spectrum of their activities.

Objectives

As outlined in the introductory section, while TET peptidases have been extensively structurally characterized, particularly in archaeal species of the Thermococcales order, our understanding of the biological significance of these proteic complexes remains limited. The work presented here aims at expanding our knowledge on TETs by adopting a broader perspective, exploring them at the scale of the entire archaeal domain. We also seek to move beyond structural approaches to investigate their roles as integral components of the cellular machinery and explore their biological functions.

The general objective of this work is to establish a comprehensive overview of TET peptidases within archaeal genomes, addressing the following key questions:

- Is it possible to define conserved sequence domains, sites, motives or markers that can serve as reliable markers for large-scale identification of TET peptidases?
- What is the phylogenetic diversity and distribution of TET peptidases across Archaea?
- Does current knowledge fully capture the functional diversity of these enzymes, or could further exploration uncover new substrate specificities?
- Are the number of copy per genome and the substrate specificities of TET peptidases associated with environmental adaptations or metabolic capacities of archaeal species?
- What are the evolutionary relationships within archaeal TET peptidases, and between archaeal and bacterial TET peptidases?

To address these questions, the first chapter of this thesis will describe the development of markers for the identification of TET peptidases, which will be used for extensive screening of archaeal genomes. This approach will lay the foundation for exploring the functional diversity of TET peptidases, which has so far only been assessed in a limited set of enzymes within a narrow taxonomic range. The second and third chapters of this manuscript will cover the production, purification and functional characterization of selected enzymes. Finally, by integrating biochemical and phylogenetic data, the findings of this study will be reframed within a broader biological context to explore the evolutionary history and biological significance of TET peptidases.

Chapter 1

From identification to classification: tracing TET peptidases in Archaea

1. Foreword

M42 peptidases are widely distributed accross many prokaryotic species. However, despite studies of a few bacterial and archaeal proteins (see **Table 2** above), our understanding of the abundance, taxonomic distribution and diversity of these enzymes remains limited.

Crystallographic structure resolution of several prokaryotic TETs revealed a high degree of homology (Borissenko & Groll, 2005; Colombo *et al.*, 2016; Durá *et al.*, 2005, 2009; Dutoit *et al.*, 2019; Kim *et al.*, 2010; Petrova *et al.*, 2015). All structurally characterized M42 peptidases form hollow tetrahedral particles (400-500 kDa) composed of 12 subunits (~40 kDa). Despite this very strong structural homology, TET peptidases exhibit low primary sequence conservation (Appolaire *et al.*, 2016). This variability, coupled with frequent misannotations as cellulases or endoglucanases (Dutoit *et al.*, 2012; Slutskaya *et al.*, 2012), makes it challenging to identify TETs in genomic or proteomic databases, hindering a deeper understanding of this enzymatic family.

One of the objectives of this thesis was to provide a comprehensive assessment of M42 peptidases distribution in archaea. Consequently, the first stage of the project involved establishing identification markers that would enable fast and robust identification of these enzymes in archaeal genomes. The approach used to establish these identification criteria, based on the pioneering work of Alexandre Appolaire (2014), will be detailed in this chapter. Briefly, since all the enzymes in this family share a highly conserved architecture, the structural elements involved in the formation of the unique tetrahedral edifice were used as key determinants.

These structure-based identification criteria were then used to screen archaeal genomes for M42 peptidases, prodiving a global perspective on their distribution in these organisms. The results of this exaustive analysis will be presented in this chapter.

Finally, phylogenetic analysis of the retrieved sequences allowed us to propose a classification for archaeal TET peptidases. This classification served as a foundation for guiding subsequent characterization of new enzymes, ensuring that the diversity of TET peptidases in archaea is comprehensively represented.

2. Material and methods

2.1. Identification criteria refinement

To refine the identification criteria previously proposed in Alexandre Appolaire's PhD thesis (2014), a dataset of 195 M42 archaeal aminopeptidases was constructed by retrieving all archaeal M42 peptidases sequences from the MEROPS database (v12.0) and eliminating truncated entries comprising less than 260 amino acids (Rawlings *et al.*, 2018). Then, as we estimated TET4 homologues to be under-represented in this dataset, we used the pattern LD[AE][EL]EKKED (canonical for the TET4 group) and ran a PHI-BLAST search against the National Center for Biotechnology Information (NCBI) nr database restricted to Archaea (Zhang *et al.*, 1998). The 15 more divergent sequences were identified using the T-Coffee trim tool (v11.0.8) and were added to the initial set of sequences (Notredame *et al.*, 2000).

Finally, the 210 resulting sequences were aligned with T-Coffee using default parameters. The alignment was visualized using AliView (v1.28) (Larsson, 2014).

Additionnaly, M42 peptidase structures were visualized and aligned using PyMOL (The PyMOL Molecular Graphics System, v2.5.2 Schrödinger, LLC). PDB identifiers are provided in **Table 2**.

2.2. Detection of TET peptidases in archaeal genomes

The 210 sequence alignment was used to build an HMM profile using the HMMBUILD tool from the HMMER suite (v3.3.2) (Johnson *et al.*, 2010).

The rest of the work presented in this section was carried out by Najwa Taib from the Biologie Evolutive de la Cellule Microbienne laboratory (Institut Pasteur, Paris).

Briefly, a local database containing 4,016 archaeal genomes representative of all major phyla available in public databases as of January 2022 was constructed. The generated HMM profile was used to carry out homology-based searches using HMMSEARCH. All hits were retrieved, aligned using MAFFT (v7.481) with the option -auto (Katoh & Standley, 2013) and filtered upon the presence of the conserved motifs characterizing TET peptidases (*i.e.,* residues Gly44, His62, Asp64, Gly77, Gly85, Gly86, Asp173, Glu205, Glu206, Glu211,

Asp228, and His307 according to PhTET1 numbering, PDB code 2WYR). This resulted in 2,030 archaeal TET peptidase homologue sequences.

The set of homologous sequences was aligned using MAFFT with the option -auto, and the resulting alignment was trimmed using BMGE (v1.12) with the options (-m BLOSUM30 -b 1 -w 1 -h 0.95) (Criscuolo & Gribaldo, 2010).

Finally, a maximum likelihood tree was inferred using IQ-TREE (v2.0.6) with the model LG+F+R10 (Minh *et al.*, 2020), selected by ModelFinder according to BIC criteria (Kalyaanamoorthy *et al.*, 2017). All phylogenies were annotated using IToL (Letunic & Bork, 2019).

3. Results

3.1. Structure-based M42 peptidase identification criteria

During his PhD thesis, Alexandre Appolaire studied the structure of M42 peptidases, focusing on their oligomerisation process and functional regulation (2014). Part of his work involved identifying conserved structural motifs to help pinpoint TET peptidases in archaeal genomes. Since the identification criteria used in this project are largely based on those established by Alexandre, I will first present his original determinants before discussing the modifications I introduced.

3.1.1. Appolaire's original identification criteria

To identify the conserved structural elements involved in the formation of the canonical tetrahedral structure of TET peptidases, Appolaire compared the structures of fifteen TET homologues identified using the structural homology search tool Dali (Holm & Rosenström, 2010). His study highlighted five conserved features shared by all tetrahedral dodecameric proteins, here described according to PhTET1 sequence numbering (PDB 2WYR).

Firstly, the presence of a **dimerization domain** seemed to be essential for monomers assembly and for the formation of the tetrahedral particle. To illustrate this, Appolaire pointed out the monomeric M28 peptidase AAP (PDB 1RTQ). While the catalytic domain of AAP closely resembles that of TET peptidases, it completely lacks the dimerization domain present in all TETs (**Figure 29**). Based on these observations, he proposed that the presence of this domain would be a marker of oligomerization in M42 peptidases.

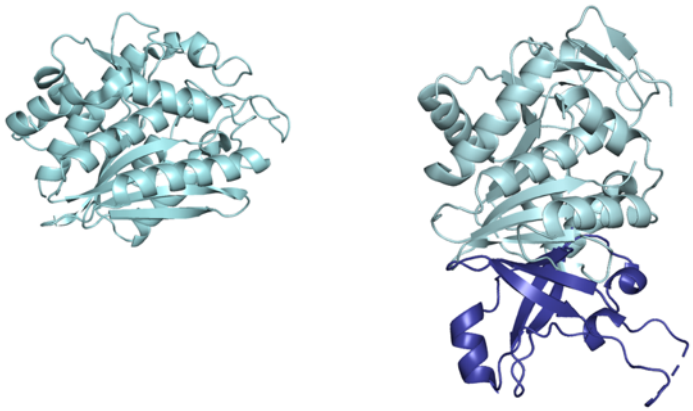


Figure 29: Structure comparison of AAP (left, PDB 1RTQ) and PhTET1 (right, PDB 2WYR). AAP lacks the dimerization domain represented in dark blue.

However, the mere presence of this domain did not seem to guarantee tetrahedral particle formation. This is exemplified by the Q11Z05_CYTH3 aminopeptidase from *Cytophaga hutchinsonii* (PDB 3CPX), which assembles as a trimer rather than forming a tetrahedral structure. A comparison with dodecameric TET peptidases revealed that Q11Z05_CYTH3 dimerization domain lacks two critical motifs—**IDIGAXXXE (I132 to E140) and [RK]303Y304**— found in all TETs at the contact interfaces between monomers. Appolaire proposed that these motifs would be essential for the oligomerization of TET peptidases.

By studying the interactions between dimers forming the TET dodecamer, Appolaire also identified the dodecamerization interface, distinct from the dimerization interface. This interface involves few interactions and is not highly conserved in sequence, though it shows strong structural similarities. Comparison of dodecameric TET peptidase structures with TmPep1050 aminopeptidase from *Thermotoga maritima* (PDB 3ISX), which had been purified as a dimer only (Dutoit *et al.*, 2012), seemed to indicate that the **α5 helix**—specifically, the presence of **residues with long, charged side chains**—was crucial for the dodecamer formation.

Furthermore, sequence analysis of dodecameric TET peptidases revealed several highly conserved glycine residues near the catalytic site—**Gly19, 67, 85, 86, 208, and 211**—which were hypothetized to provide the necessary flexibility for proper folding around the active site.

Finally, the presence of the seven conserved catalytic residues presented in **introduction section 3.1**—**His62, Asp64, Asp173, Glu205, Glu206, Asp/Glu228, and His307**—was considered to be essential for both the catalytic activity and oligomerization of TET peptidases.

3.1.2. Identification criteria refinement

When establishing the initial identification criteria, Appolaire relied on a few proteins only due to the limited number of characterized M42 peptidases available at the time. Since then, new structures have been resolved, and many additional sequences have been annotated as putative M42 peptidases. These new data have been used to modify the original identification criteria, ensuring that our study adequately captures the diversity of TET peptidases.

To refine the identification criteria, a more comprehensive dataset was constructed by retrieving 195 full-length sequences annotated as M42 peptidases from the MEROPS

database. Among the characterized M42 peptidases, PhTET4 clearly stands out, notably due to the absence of the IDIGAXXXE and [RK]Y motifs identified by Appolaire in the dimerization domain of TETs. Instead, a LD[AE][EL]EKKED motif is found in PhTET4 and homologues. To incorporate this diversity into our study, which was poorly represented in MEROPS, a PHI-BLAST search was conducted against the NCBI nr database restricted to archaea using the LD[AE][EL]EKKED pattern. The fifteen more divergent sequences were retained. The resulting 210 sequences were aligned, and the resulting alignment was used to refine the identification criteria.

3.1.2.1 Dimerization domain and dimerization interface

The IDIGAXXXE and [RK]Y motifs used by Appolaire to describe the dimerization interface were not strictly conserved in our extended dataset. It seemed difficult to define a pattern broad enough to cover the variability observed in this region while remaining discriminating (**Figure 30**). Moreover, these motifs are absent in PhTET4, which has nevertheless been determined to be dodecameric in the ELMA laboratory (Basbous *et al.*, 2018). It was therefore decided to only consider the presence of the dimerization domain as the main criterion for dimerization.

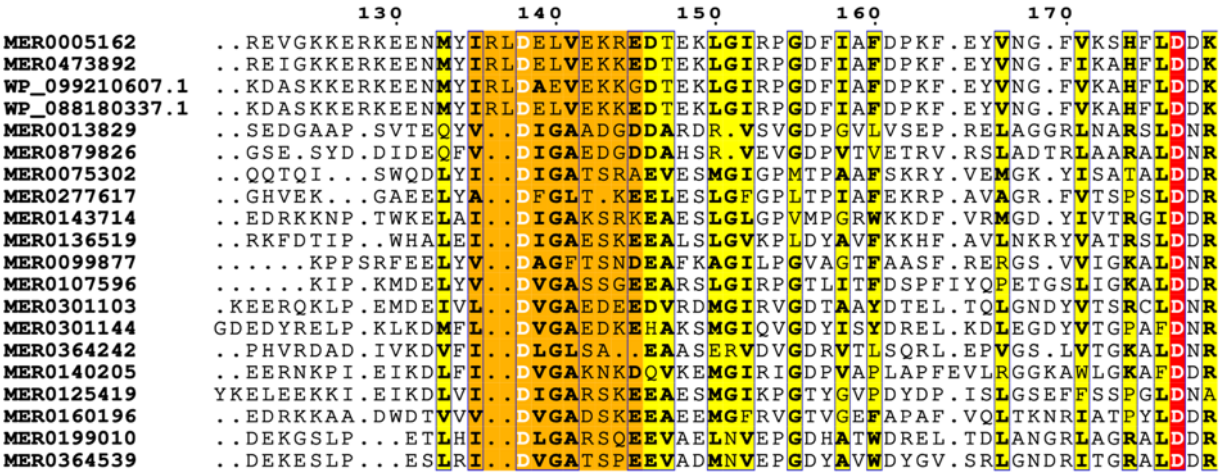


Figure 30: Multipe sequence alignement of 20 of the sequences used for the identification criteria refinement. The IDIGAXXXE motif identified by Appolaire in the dimerization domain, here highlighted in orange, is not strictly conserved and is completely lacking in PhTET4 homologues (top four sequences).

3.1.2.2 Dodecamerization interface

To describe the dodecamerization interface, Appolaire used the example of TmPep1050 aminopeptidase from *Thermotoga maritima*, which had only been purified as a dimer despite having a sequence very similar to that of dodecameric TETs (Dutoit *et al.*, 2012). The absence

of charged residues in the α5 helix of TmPep1050 was pointed out to be responsible for this lower oligomeric state.

However, it was later discovered that the oligomeric state of TmPep1050 was altered by the presence of a His-tag used for its purification (Dutoit, 2021). The untagged protein was successfully purified as a dodecamer, even though it still lacked the charged residues in the α5 helix that Appolaire identified as crucial for dodecamerization (Dutoit *et al.*, 2019). These findings led me to revise the relevance of the presence of these charged residues as markers for the oligomerization of TET peptidases.

3.1.2.3 Glycine residues

In Appolaire's limited dataset, seven glycine residues were identified as conserved. However, in our extended dataset of 210 sequences, some of these residues were no longer conserved, while other nearby glycines showed higher conservation. Consequently, it was decided to look for the conservation of the following limited set of glycine residues: Gly44, 77, 85, 86, and 211.

3.1.3. Synthesis of retained identification criteria

Ultimately, the following criteria were retained for M42 peptidase identification:

- presence of a dimerization domain
- conservation of the catalytic residues His62, Asp64, Asp173, Glu205, Glu206, Asp/Glu228, and His307
- conservation of the glycine residues Gly44, 77, 85, 86, and 211

These criteria were subsequently applied for archaeal genomes screening, as presented in **section 3.2** below. Filtering based on the presence of the catalytic residues enabled to retrieve sequences from all MH clan families (M18, M20, M28, and M42). Further filtering based on the conserved glycines residues allowed to eliminate M18, M20, and M28 sequences, effectively segregating M42 peptidases. All sequences meeting these two criteria also contained the expected dimerization domain. Consequently, the catalytic residues and conserved glycine residues would be sufficient to identify M42 peptidases and could be used alone (**Figure 31**).

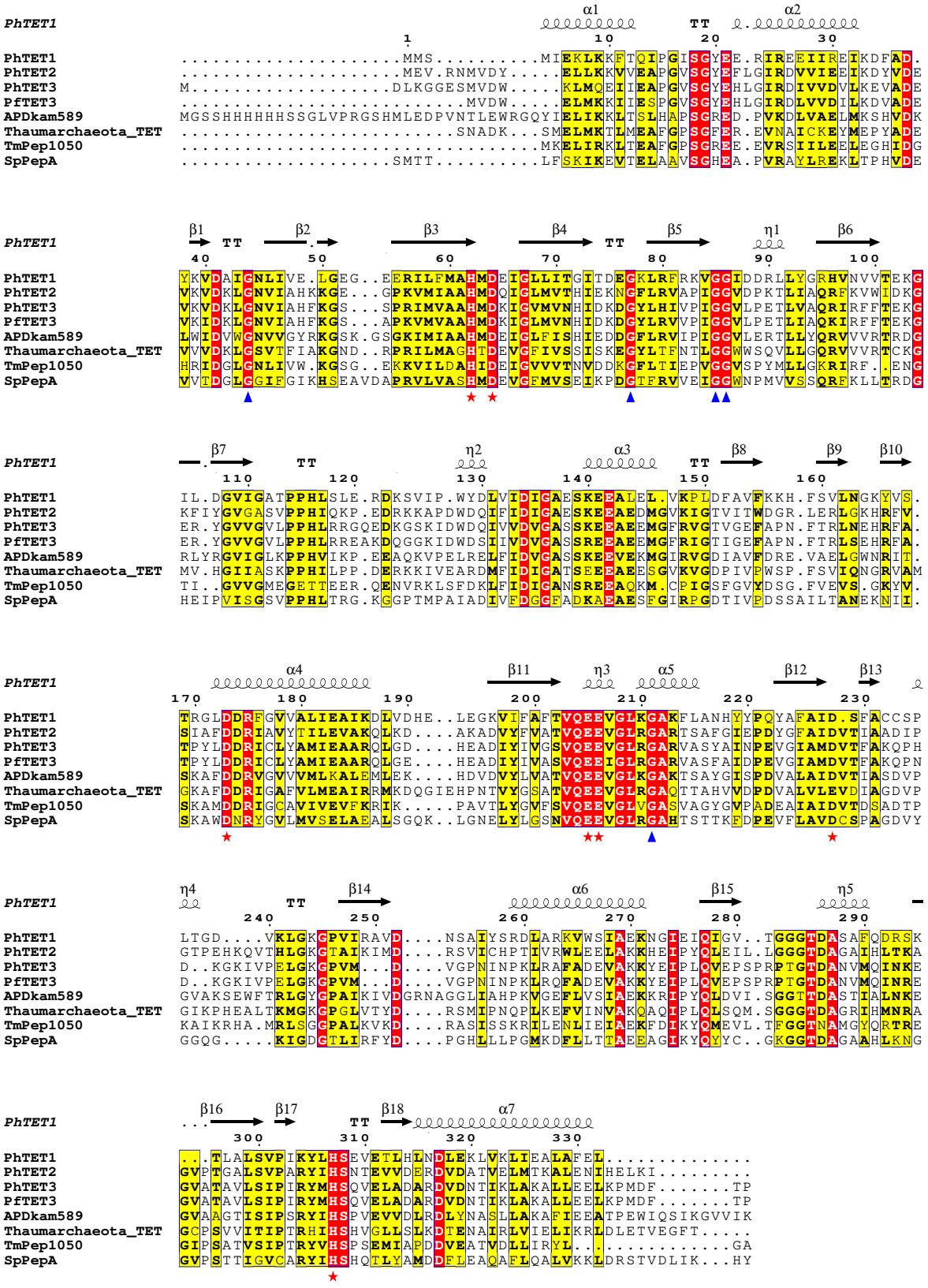


Figure 31: Structure-based determinants for identifying M42 peptidases. The represented secondary structure corresponds to PhTET1 (PDB code 2WYR). Three criteria were retained for the identification of TET aminopeptidases: conserved catalytic residues H62, D64, D173, E205, E206, [DE]228, H307 (red stars), conserved glycine residues 44, 77, 85, 86, and 211 (blue triangles) and presence of the dimerization domain (residues 65 to 154). Residue numbering according to PhTET1 sequence.

3.2. M42 peptidases in Archaea

3.2.1. Comprehensive screening for M42 peptidases in archaeal genomes

To comprehensively assess the distribution of M42 peptidases in Archaea, a hidden Markov model (HMM) profile was constructed using 210 archaeal M42 sequences obtained from the MEROPS and NCBI nr databases. This profile was used for an extensive homology search of M42 peptidases against a large database comprising 4,016 archaeal genomes, covering all currently available diversity including uncultured phyla. All positive hits were retrieved and filtered based on the presence of conserved catalytic sites and glycine residues presented in **section 3.1.3** above, resulting in 2,030 archaeal M42 homologues. M42 peptidases were identified in 39.7% (1,595) of genomes, with copy numbers ranging from one to five per genome.

M42 homologues were prevalent in Euryarchaeota and Asgard but exhibited uneven to sporadic distribution across the TACK and DPANN superphyla, respectively. Intriguingly, some M42 peptidase sequences from Asgard displayed a unique insertion of approximately 20 residues, which has never been detected in any M42 peptidase described to date. This distinctive insertion, located just after the dimerization interface, was specific to Asgard species and exhibits a strong charge contrast. A similar but shorter insertion (9-10 residues) was detected in a few Thermoplasmata species (**Figure 32**). The potential roles of this insertion will be discussed in **chapter 3 section 3.1** below.

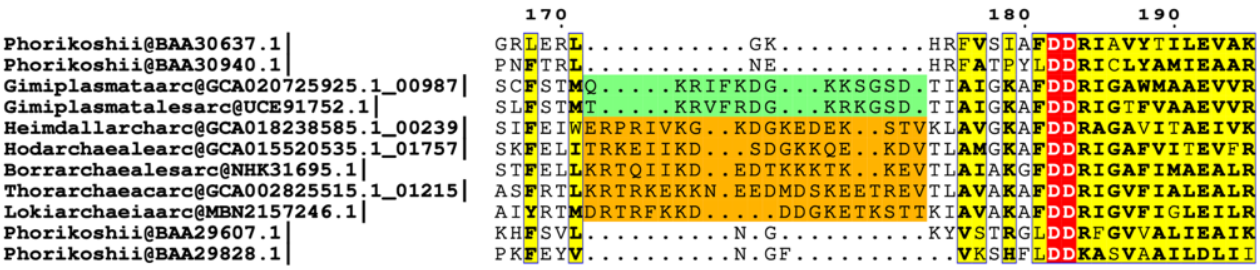


Figure 32: Multiple sequence alignment of 11 archaeal M42 peptidases illustrating a previously unreported insertion in Asgard and Thermoplasmata sequences. The unique ~20 residue insertion, specific to some Asgard species, is highlighted in orange. A shorter but similar insertion (9-10 residues), highlighted in green, was identified in certain Thermoplasmata sequences.

3.2.2. Archaeal M42 peptidases phylogeny

To explore the relationships among archaeal M42 peptidases, we inferred a maximum likelihood tree using the 2,030 archaeal M42 sequences (**Figure 33**). According to the topology of the tree, several distinct families can be delineated.

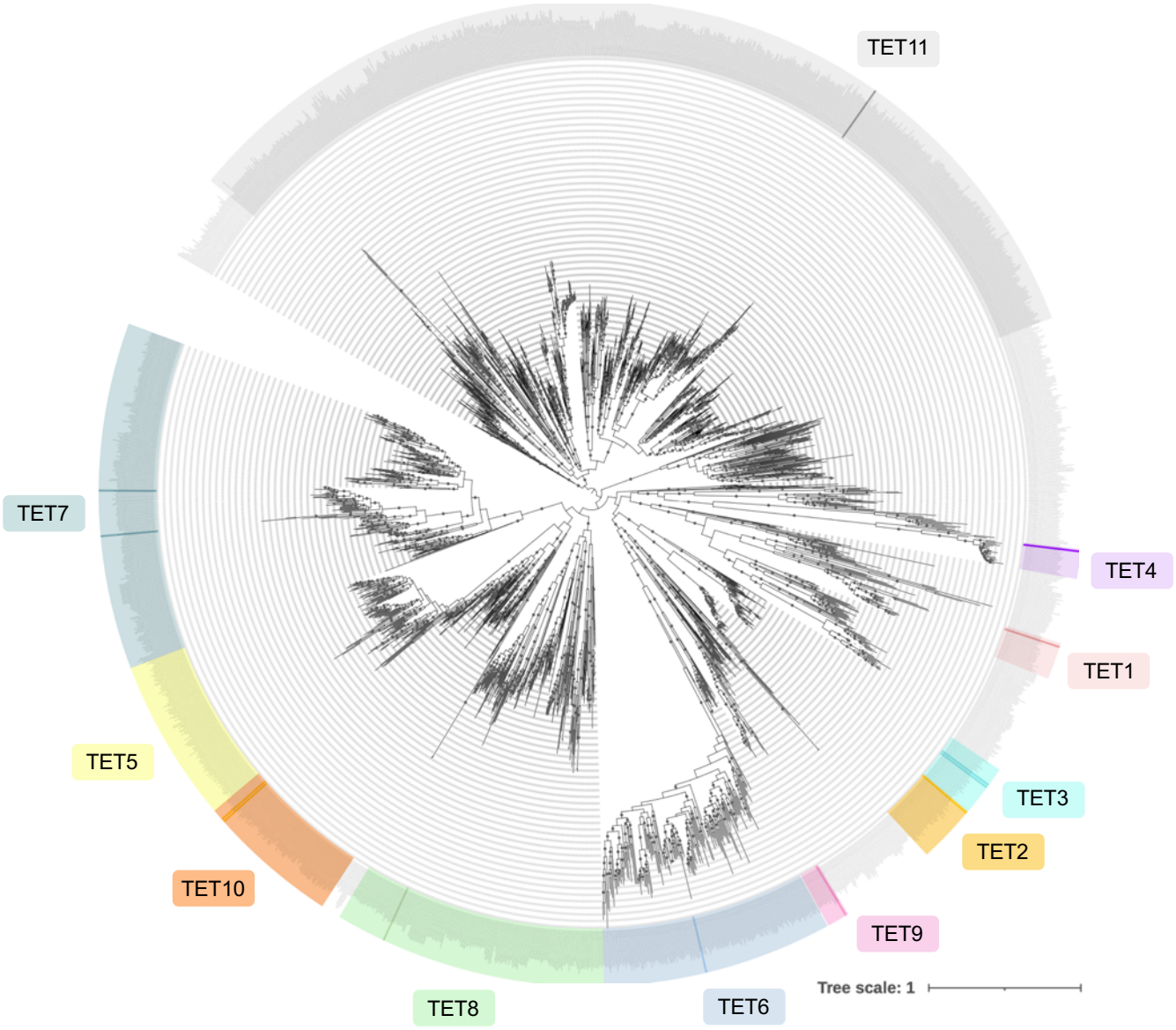


Figure 33: Phylogeny of archaeal M42 peptidase homologues and delineation of TET families in Archaea. Maximum-likelihood phylogeny obtained from an alignment of 2,030 sequences and 337 amino acid positions. The scale bar represents the average number of substitutions per site. Circles at the branches indicate ultra-fast bootstrap values ≥ 90 %. TET families were delineated based on the taxonomic distribution and the topology of the tree (*i.e.,* branch lengths, node supports). Darker bars on the outer circle represent enzymes characterized either before or during this study.

The TET1, TET2, TET3, and TET4 groups were already identified prior to this study, with several characterized representatives from species such as *Pyrococcus horikoshii*, *Pyrococcus furiosus* and *Thermococcus onnurineus* (see **Table 2**) (Basbous *et al.*, 2018; Colombo *et al.*, 2016; Durá *et al.*, 2005, 2009; Durá & Franzetti, 2013; H. S. Lee *et al.*, 2007; Y. G. Lee *et al.*, 2012; Schoehn *et al.*, 2006). These groups form four distinct and supported monophyletic clades, with ultra-fast bootstrap (UFB) values of 98%, 100%, 100%, and 100% respectively. TET2 and TET3 appear as sister clades, suggesting a recent duplication event,

which is further supported by studies showing the *in vitro* and *in vivo* formation of PhTET2-PhTET3 heterocomplexes in *P. horikoshii* (Appolaire, Durá, *et al.*, 2014; Appolaire, Girard, *et al.*, 2014). Interestingly, while TET2, TET3, and TET4 contain exclusively sequences from Thermococcales, TET1 also includes three sequences from *Geoglobus*, suggesting a HGT from Thermococcales to Archaeoglobales (see **Appendix 1**).

To delineate new families, the topology of the phylogenetic tree—considering branch length, support, monophyly, and taxonomic distribution—was analyzed. This approach led to the identification of seven new families, which were named TET5 to TET11 in accordance with the existing nomenclature (**Figure 33**). Some of these newly identified families were specific to taxonomic groups: TET5 and TET6 were exclusively found in the Halobacteriales, TET9 in Asgard, and TET10 in TACK. In contrast, the TET7, TET8, and TET11 families exhibitied a broader distribution across Archaea, with TET11 being found in all superphyla. The taxonomic distribution of the twelve TET families will be explored in more detail in the **general discussion** section. The previously characterized HmTET from *Haloarcula marismortui* (Franzetti *et al.*, 2002) and APDkam589 from *Desulfurococcus amylolyticus* (Petrova *et al.*, 2015; Slutskaya *et al.*, 2012) fall into the TET6 and TET10 families, respectively, and were accordingly renamed HmTET6 and DaTET10. Notably, all sequences featuring the novel insertion described in **section 3.2.1** above were found in the TET7 group.

Collectively, these eleven families account for 1,619 sequences, the remaining 411 sequences could not be affiliated to any family due to poorly supported branching (**Figure 33**).

3.2.3. Sequence selection for functional and structural characterization

A striking observation from this phylogenetic analysis is that previous extensive characterization of archaeal TET peptidases, which focused mainly on a few species within the Thermococcales order (TET1-4 groups) (Basbous *et al.*, 2018; Colombo *et al.*, 2016; Durá *et al.*, 2005, 2009; Durá & Franzetti, 2013; Franzetti *et al.*, 2002), remained marginal with regard to the real taxonomic distribution and overall diversity in archaeal genomes. These findings highlighted the need for broader protein characterization to explore the functional diversity of TET peptidases in Archaea.

To address this, thirteen sequences were selected for recombinant protein expression, purification, and characterization, ensuring representation from all the TET peptidase families delineated in our analysis, while also capturing sequences from taxonomically diverse archaeal

species. Given that the TET1 to TET4 groups have already been extensively studied, efforts were directed toward the other lesser-known or entirely unexplored families. Careful consideration was given to ensuring taxonomic diversity, including representatives from the Euryarchaeota, TACK, DPANN, and Asgard phyla. Whenever possible, sequences were chosen from cultured and thermophilic organisms to provide ecological context related to their lifestyle and environment and facilitate purification. Three sequences featuring the insertion unique to Asgard sequences were selected (*i.e.,* HoTET7, PsTET7, and ThTET7) to investigate the role of this novel extension. The selected candidates are presented in **Table 3**, and the corresponding sequences are provided in **Appendix 2**.

Table 3: Selected candidate proteins for further characterization. The choosen proteins are listed below, along with the corresponding species and taxonomy.

TET	Species	Taxonomy
HvTET5	*Haloferax volcanii*	Methanotecta; Halobacteria; Halobacteriales; Haloferacaceae; *Haloferax*;
HvTET6	*Haloferax volcanii*	Methanotecta; Halobacteria; Halobacteriales; Haloferacaceae; *Haloferax*;
HoTET7	*Ca. Hodarchaeales archaeon LC_3*	Asgardarchaeota; Heimdallarchaeia; Hodarchaeales; LC-3; LC-3;
PsTET7	*Ca. Prometheoarchaeum syntrophicum*	Asgardarchaeota; Lokiarchaeia; CR-4; AMARA-1; *Prometheoarchaeum*
ThTET7	*Ca. Thorarchaeota archaeon MP8T_1*	Asgardarchaeota; Thorarchaeia; Thorarchaeales; Thorarchaeaceae; MP8T-1;
HoTET8	*Ca. Hodarchaeales archaeon LC_3*	Asgardarchaeota; Heimdallarchaeia; Hodarchaeales; LC-3; LC-3; LC-3 sp001940645;
PsTET8	*Ca. Prometheoarchaeum syntrophicum*	Asgardarchaeota; Lokiarchaeia; CR-4; AMARA-1; *Prometheoarchaeum*;
PsTET9	*Ca. Prometheoarchaeum syntrophicum*	Asgardarchaeota; Lokiarchaeia; CR-4; AMARA-1; *Prometheoarchaeum*;
TaTET10	*Thermosphaera aggregans*	TACK; Thermoproteia/Crenarchaeota; Sulfolobales; Desulfurococcaceae; *Thermosphaera*;
AlTET11	*Altarchaeia archaeon ex4484_2*	DPANN; Altarchaeia; IMC4; QMZM01; EX4484-2;
MfTET11	*Methanothermus fervidus*	Methanomada; Methanobacteria; Methanobacteriales; Methanothermaceae; *Methanothermus*;
MtTET11	*Methanoculleus thermophilus*	Methanotecta; Methanomicrobia; Methanomicrobiales; Methanoculleaceae; *Methanoculleus*;
MkTET	*Methanopyrus kandleri*	Methanomada; Methanopyri; Methanopyrales; Methanopyraceae; *Methanopyrus*;

The cloning, recombinant expression and purification of these proteins will be covered in **chapter 2**. Their functional and structural characterization will be presented in **chapter 3**.

4. Discussion and future perspectives

The refinement of TET petidase identification criteria presented here represents a significant advancement over the initial criteria established by Alexandre Appolaire. Due to the limited dataset he was working with, Appolaire's original criteria were highly stringent. While effective for a specific subset of TET peptidases, these criteria limited the discovery of the full diversity present in archaeal genomes. New studies on TET peptidases published since his initial work revealed a greater diversity than previously recognized, prompting us to broaden the identification criteria. This adjustment was crucial for identifying diverse TET families that do not strictly conform to the original rigid motifs but still maintain the key functional characteristics of TET peptidases. The refined criteria now allow for a more inclusive and comprehensive survey of TET peptidases, reflecting the diversity of these enzymes more accurately.

However, the current criteria will likely require further revision as more genomes continue to be sequenced and our understanding of TET molecular machineries deepens. This iterative process of refinement is essential to ensure that those criteria remain relevant and effective.

Despite these improvements, it is important to acknowledge the potential limitations of our methodology. In particular, the TET4 group, which has a distinctive signature that distinguishes it from other families (see **section 3.1.2.1**), was over-represented in the HMM profile used for sequence identification. We enriched our initial dataset with sequences from this group to ensure that this diversity was well represented, with PhTET4 homologs accounting for 11% of the sequences used to construct the HMM profile. However, the results of archaeal genomes screening indicate that we overestimated the prevalence of TET4 sequences, which are in fact found in Thermococcales exclusively and account for 1% only of identified M42 peptidases.

While this study aimed at exploring the diversity of archaeal TET peptidases as extensively as possible, practical limitations necessitated the selection of a restricted set of sequences for recombinant expression, purification, and characterization. The selected sequences were chosen to cover all the identified TET families as well as a wide taxonomic range of archaeal species. However, given the vast diversity within some of the larger families, particularly TET11 and TET-other, further characterization of additional members from these families would be highly valuable. The TET-other group, in particular, might benefit from subdivision into more refined subgroups based on further phylogenetic analysis and functional

data. Detailed studies focusing on these subgroups could provide deeper insights into the diversity of TET peptidases.

Ultimately, the work presented in this chapter underscores the dynamic and evolving nature of enzyme classification and the ongoing need to adapt our approaches as new data becomes available. While the refined criteria developed here offer a more inclusive tool for identifying TET peptidases, they are not without their limitations. Continued efforts should be directed toward refining these approaches, incorporating new sequences, and characterizing new peptidases. By doing so, future studies will provide a deeper understanding of TET peptidase diversity, functional roles in archaeal biology, and ecological and evolutionary significance.

5. Résumé du chapitre en français

Depuis leur découverte en 2002, un peu plus d'une vingtaine de peptidases TET ont été fonctionnellement et structuralement caractérisées chez les archées. Ces études ont révélé une grande conservation de l'architecture caractéristique des peptidases TET, tout en mettant en évidence une diversité fonctionnelle surprenante. Bien que ces travaux aient permis une description détaillée de certains représentants, la prévalence et la diversité des peptidases TETs à l'échelle globale des archées restent mal comprises.

Ce chapitre présente une étude exhaustive de la distribution des peptidases M42 chez les archées. Pour ce faire, des critères d'identification basés sur des éléments structuraux conservés ont été mis en place. Ces critères, initialement établis par Alexandre Appolaire (2014), ont été modifiés pour mieux correspondre à la diversité de ces enzymes mise à jour par les études récentes. En collaboration avec Najwa Taib du laboratoire de Biologie Evolutive de la Cellule Microbienne (Institut Pasteur, Paris), un ensemble de 210 séquences de peptidases M42 a été utilisé pour construire un modèle de Markov (Hidden Markov Model). Ce modèle a ensuite été utilisé pour mener une recherche par similarité à grande échelle dans les génomes archéens, permettant d'identifier 2 030 homologues de peptidase M42.

Une analyse phylogénétique approfondie des séquences ainsi identifiées a révélé l'existance de plusieurs familles de peptidases TET. En plus des familles TET1 à TET4 déjà mises à jour lors de précédentes études, sept nouvelles familles ont été identifiées et nommées TET5 à TET11, en accord avec la nomenclature existante. De façon surprenante, les familles TET1 à TET4, bien que largement étudiées auparavant, se sont avérées minoritaires et restreintes à l'ordre des Thermococcales. Ces résultats soulignent la nécessité de caractériser plus largement les peptidases TET, en particulier au sein des groupes TET5 à TET11, encore peu explorés.

Treize séquences ont donc été sélectionnées pour être produites sous forme de protéines recombinantes, purifiées et caractérisées. Ces séquences couvrent toutes les familles de peptidases TET nouvellement définies ainsi que la diversité taxonomique des archées. La caractérisation de ces nouvelles enzymes devrait ainsi permettre d'approfondir nos connaissances des peptidases TET, offrant une vision plus complète de leur diversité et de leur évolution au sein des archées.

Chapter 2

Considerations for the production and purification of TET peptidases: challenges and lessons learned

1. Foreword

Previous PhD theses conducted on M42 aminopeptidases in the ELMA laboratory were devoted to the characterization of one or two enzymes. Given the large number of proteins targeted for this project, it seemed sensible to adopt a standardized production and purification approach using identical, or at least similar, protocols for all the proteins. This strategy seemed all the more relevant considering that so far, with minor adaptations, the techniques and protocols employed for the purification and characterization of the various TET aminopeptidases described in the laboratory have remained largely consistent.

The use of a poly-histidine tag (His-tag), which generally facilitates purification and increases the purification yield, has been reported for several M42 peptidases (Dutoit *et al.*, 2012; Kim *et al.*, 2010; Kumaki *et al.*, 2011; Lee *et al.*, 2007; Robinson *et al.*, 2013; Zheng *et al.*, 2005). In particular, this approach has been successfully used for the study of *Methanocaldococcus jannaschii* TET in our laboratory (Basbous, 2016). For this project, the use of a His-tag seemed particularly advantageous to ease, streamline and speed up the purification step, which can be very labour-intensive and time-consuming.

Consequently, after a step of heterologous expression optimization, the following generic purification protocol was adopted: (i) clarification by heat-shock for thermostable proteins followed by (ii) immobilized metal affinity chromatography (IMAC), (iii) anion-exchange (AEX) chromatography, and (iv) size-exclusion chromatography (SEC).

Unfortunately, this standardized protocol turned out to be unsuitable for most of our targets. As highlighted by the wide diversity of TET peptidases revealed by archaeal genome screening (see **chapter 1**), it is actually difficult to apply a single protocol for the purification of all enzymes within this family. Moreover, as previously pointed out by Dutoit and colleagues (Dutoit, 2021), the use of a His-tag can be problematic in the case of metallopeptidases such as TETs.

In the subsequent chapter, the initial approach used for the production and purification of the thirteen new TET peptidases to be characterized, along with the associated identified difficulties, will be presented. The different strategies considered to overcome these challenges will then be discussed. Finally, the protocols that enabled successful purification of novel M42 aminopeptidases will be described.

2. Material and methods

2.1. Bacterial strains and general information

E. coli Rosetta2(DE3)pLysS, BL21(DE3)pLysS and ArcticExpress(DE3)RIL chemically competent cells were used for recombinant expression. These cells were grown in lysogeny broth (LB) media. When used, kanamycin (kan), chloramphenicol (chl) and gentamycin (gen) concentrations were 30 μg/mL, 34 μg/mL and 20 μg/mL respectively. *E. coli* chemical transformations were done following manufacturer's instructions. For sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), protein samples were mixed with 3X Loading Buffer (50 mM Tris-HCl, 8 M urea, 2 M thiourea, 75 mM DTT, 3% SDS, 0.05% bromophenol blue, pH 6.8), heated to 100 °C for 4 min, and loaded on 12% Criterion™ XT Bis-Tris Protein gels (BioRad). Protein molecular weights were estimated relatively to Precision Plus Protein All Blue Standards (Biorad).

2.2. Cloning

The open reading frames of the selected TET genes were codon-optimized for *E. coli* and synthesized by Twist Bioscience. For N-terminal tagged protein expression, synthetic genes were digested using *NdeI* and *BamHI* restriction enzymes and inserted into the pET28a(+) multiple cloning site in frame with a thrombin-cleavable N-terminal His6-tag. For untagged protein expression, genes were digested with *NdeI* and *XhoI* restriction enzymes and cloned into the pET41c(+) vector. Cloning accuracy was confirmed by Sanger sequencing (Eurofins).

2.3. Heterologous expression in *E. coli*

2.3.1. Expression optimization

3 mL of LB were inoculated at a 1:100 ration with overnight culture and cultured at 37 °C (for Rosetta2 and BL21 cells) or 30°C (for ArcticExpress cells), 140 rpm until the OD_{600} reached 0.6. Unless otherwise indicated, protein overexpression was induced with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at the selected temperature for the desired duration.

Cells were harvested by centrifugation at 8,000 × *g* for 15 min at 4 °C, and pellets were stored at -80 °C. Cells were resuspended in lysis buffer (50 mM Tris-HCl, 150 mM NaCl, 0.1% Triton ×100, pH 8.0) supplemented with 0.05 mg/mL lysozyme, 0.01 mL/mL $MgSO_4$ 2M, 1 mg/mL Pefabloc SC, 0.05 mg/mL DNase, 0.2 mg/mL RNase, and were disrupted on ice in a

Vibra-Cell sonifier (35% amplitude with three on/off cycles of 10 s each). The cell lysates were centrifugated at 16,000 × *g*, 4 °C for 15 min and the supernatant and pellet were separated. A fraction of the supernatant was heat-shocked at 70°C for 15 min and centrifugated at 16,000 × *g*, 4 °C for 15 min. Once again, the supernatant and pellet were separated. Samples were deposited on SDS-PAGE gels in order to determine the solubility and thermostability of the proteins of interest.

2.3.2. Large-volume cultures

Once the optimal culture conditions were determined for each protein, large-volume cultures were realized. 1.5 L of LB was inoculated at a 1:100 ration with overnight culture and cultured at 37 °C (for Rosetta2) or 30°C (for ArcticExpress cells), 140 rpm until the OD_{600} reached 0.6. Protein overexpression was induced with IPTG at the selected temperature for the desired duration. Cells were harvested by centrifugation at 8,000 × *g* for 45 min at 4 °C, and pellets were stored at -80 °C. Cells were resuspended in lysis buffer (50 mM Tris-HCl, 150 mM NaCl, 0.1% Triton ×100, pH 8.0) supplemented with 0.05 mg/mL lysozyme, 0.01 mL/mL $MgSO_4$ 2M, 1 mg/mL Pefabloc SC, 0.05 mg/mL DNase, 0.2 mg/mL RNase, and were disrupted on ice in a Vibra-Cell sonifier (35% amplitude with five on/off cycles of 30 s each). For thermostable proteins, the lysate was heated at 70° C for 15 min. Insoluble particles were pelleted by centrifugation (16,000 × *g* for 30 min at 4 °C) and the cleared extract was filtered at 0.45 µm and 0.22 µm.

2.4. Protein purification

The recombinant proteins were purified from the soluble fractions to near homogeneity using various combinations of affinity, anion exchange and gel filtration chromatography. The following columns, all supplied by Cytiva, were used: HiTrap Chelating HP 5 mL, DEAE sepharose CL-6B resin packed in a XK16 column, ResourceQ 6 mL, Superose 6 Increase 10/300 GL, Superdex 200 10/300 GL. Purification strategies will be presented below, and detailed protocols are available in **Appendix 3**. Protein concentrations were determined following the method of Bradford (1976) using bovine serum albumin as the standard.

2.5. Native PAGE electrophoresis

Protein samples were mixed with one volume of loading buffer (20 mM Tris-HCl, 62.5% glycerol, pH 6.9) just before analysis. Polyacrylamide gels containing 8%

acrylamide/bisacrylamide, 20 mM NaCl, 50 mM Tris-HCl, pH 8.8, were run at 4 °C in a Tris/glycine buffer (25 mM Tris-HCl, 192 mM glycine, pH 7.5).

2.6. Negative-stain electron microscopy

4 μL of purified protein samples (0.1 mg/mL) were absorbed onto the clean side of a carbon film on mica, stained, and transferred to a 400-mesh copper grid. Images were taken under low dose conditions (<10 $e^-/Å^2$) with defocus values between 1.2 and 2.5 µm on a Tecnai 12 LaB6 electron microscope at 120 kV accelerating voltage using CCD Camera Gatan Orius 1000.

2.7. AlphaFold model predictions

AlphaFold model predictions were calculated using the AlphaFold3 server (https://golgi.sandbox.google.com accessed on 17 May 2024).

3. Results

3.1. Production and purification of His-tagged TETs

3.1.1. Heterologous expression optimization

For N-terminally His-tagged protein expression, synthetic genes were cloned in the pET28a(+) vector. For each protein, expression tests were carried out by modulating the expression cells and the temperature. The following conditions were tested:

- Rosetta2(DE3)pLysS cells at 37 °C for 3 h 30 min / at 20 °C for 16 h
- BL21(DE3)pLysS cells at 37 °C for 3 h 30 min / at 20°C for 16 h
- ArcticExpress(DE3)RIL cells at 14 °C for 24 h

The resistance of the proteins to a 15 min heat-shock at 70°C was also assessed. The conditions yielding the highest expression levels and solubility were retained (**Table 4**). In most cases, expression in Rosetta2(DE3)pLysS cells at 20°C gave the best results, producing a satisfactory quantity of soluble protein. For proteins with low solubility, ArcticExpress(DE3)RIL cells were found to be the most effective, but the quantity of soluble protein produced remained low. For HvTET6, no overexpression was obtained in any of the conditions tested despite proper cloning.

Table 4: Selected expression conditions for His-tagged protein production. The quantity and solubility of overexpressed proteins were evaluated by SDS-PAGE.

Protein	Expression cells	Expression conditions	Heat-shock resistant
HvTET5	ArcticExpress(DE3)RIL	24 h at 14°C	no
HvTET6	no expression		
HoTET7	ArcticExpress(DE3)RIL	24 h at 14°C	no
PsTET7	Rosetta2(DE3)pLysS	16 h at 20°C	no
ThTET7	Rosetta2(DE3)pLysS	16 h at 20°C	yes
HoTET8	Rosetta2(DE3)pLysS	16 h at 20°C	yes
PsTET8	Rosetta2(DE3)pLysS	16 h at 20°C	no
PsTET9	Rosetta2(DE3)pLysS	16 h at 20°C	no
TaTET10	Rosetta2(DE3)pLysS	16 h at 20°C	yes
AlTET11	ArcticExpress(DE3)RIL	24 h at 14°C	no
MfTET11	ArcticExpress(DE3)RIL	24 h at 14°C	no
MtTET11	Rosetta2(DE3)pLysS	16 h at 20°C	no
MkTET	Rosetta2(DE3)pLysS	16 h at 20°C	no

3.1.2. Development of the purification of His-tagged TETs

Based on the TET purification protocols in use at the ELMA laboratory, the following strategy was initially implemented:

However, this approach only enabled the successful purification of ThTET7 (**Figure 34**). For all other proteins, instability and aggregation issues were encountered; precipitation occurred either suddenly during or gradually after the IMAC step, ultimately resulting in the loss of all soluble protein. Given that IMAC appeared to be the central issue, I explored the potential effects of this chromatographic technique and of the use of a His-tag on TETs, and on metalloproteins more generally. This purification strategy may prove problematic for metal-dependent enzymes for several reasons that will be discussed below.

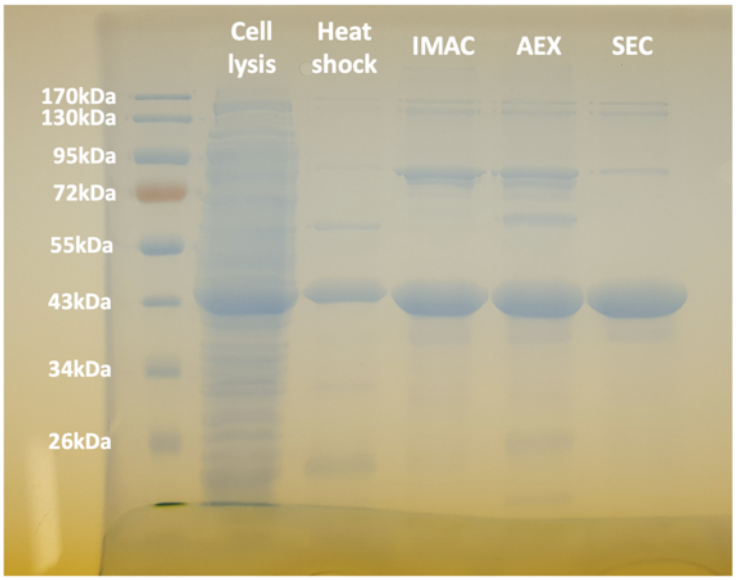


Figure 34: SDS-PAGE gel recapitulating the purification of ThTET7. Samples were collected after each purification step. ThTET7 monomer migrated at an apparent molecular weight of ~43 kDa.

3.1.3. IMAC purification: potential pitfalls for metalloproteins

Due to its cost-effectiveness, ease of use, and exceptional robustness, IMAC is widely used for His-tagged protein purification. IMAC columns are typically packed with a resin, often agarose or sepharose, that is functionalized with metal ions like nickel, cobalt, or copper. These metal ions coordinate with the histidine residues of the His-tag, allowing for selective binding.

During elution, imidazole is introduced as a competitive ligand, binding to the metal ions on the column and displacing the His-tagged proteins.

IMAC offers high affinity and specificity, allowing efficient protein capture even in complex mixtures. Additionnally, this technique performs well under various conditions, including both native and denaturing environments, as well as oxidizing and reducing conditions. However, when working with metalloproteins, there are a few aspects to consider before diving in blindly.

First, while the addition of a His-tag can enhance the expression and solubility of a protein, it may have detrimental effects on metal-dependent proteins. IMAC purification of His-tagged proteins relies on the affinity between the imidazole group carried by the side chain of histidine residues and metal ions. As a result, the His-tag can compete with the native metal-binding site of the protein for metal ion binding, thereby altering protein structure or enzymatic activity (Ayoub *et al.*, 2023; Freydank *et al.*, 2008; Majorek *et al.*, 2014; Zhao & Huang, 2016). Additionally, it is important to note that the imidazole ring is not catalytically inert and can exhibit esterase activity. For instance, Schoonen *et al.* (2017) reported that green fluorescent protein (GFP) showed esterase activity after the introduction of a His-tag. While this is not a concern for TETs, it could be problematic when working with enzymes possessing an esterase activity.

Secondly, the IMAC column itself can impact metalloproteins. Metal ions from the column may leach out and replace the native metal ions bound to the metalloprotein. Block *et al.* (2009) pinpointed this phenomenon by expressing a His-tagged zinc-finger containing transcription factor in *E. coli* and purifing the protein using both nickel- and zinc-based IMAC in parallel. They then determined the metal ion content of both preparations using intercoupled plasma mass spectrometry (ICP-MS), revealing that significant exchange of metal ions occured between the charged IMAC ligand and the zinc fingers. Conversely, the ion exchange could also occur in the opposite direction, with the colomn stripping the protein of its essential metal ions.

Finally, the imidazole used during the elution step can also interfere with metal-dependent proteins. Imidazole competes with the histidine residues of the His-tag for binding to the metal ions on the IMAC resin. This competition is essential for eluting the bound protein from the column. However, the high concentration of imidazole needed for efficient elution may affect the metalloprotein. If the metal binding affinitiy of imidazole is higher than that of the

protein, it can chelate these ions, potentially stripping them from the metalloprotein and disrupting its structure or function (Block *et al.*, 2009).

Overall, while IMAC is a highly effective and widely used method for purifying His-tagged proteins, it presents specific challenges when applied to metalloproteins. Therefore, careful consideration and optimization are crucial in this context. Although this approach has been successfully applied to the purification of several enzymes in the TET family, oligomeric state and catalytic activity alteration has already been reported for His-tagged *Thermotoga maritima* TETs (Dutoit, 2021).

3.1.4. Strategies to restore metal cofactor binding

On my end, the presence of a His-tag and the use of IMAC did not cause any difficulties for the purification of ThTET7, that was obtained as an active dodecameric protein. For the other recombinant proteins I attempted to purify, the binding of the essential metal cofactors seemed to be impaired, resulting in protein instability and aggregation (**Figure 35**). Considering the crucial role of these metallic cofactors in the TET edifice, metal depletion could be causing the disassembly of the dodecameric complex into less stable dimeric or even monomeric forms, ultimately leading to protein precipitation. I therefore explored several approaches aiming at restoring proper metal cofactor binding, hoping to promote stable dodecamer formation.

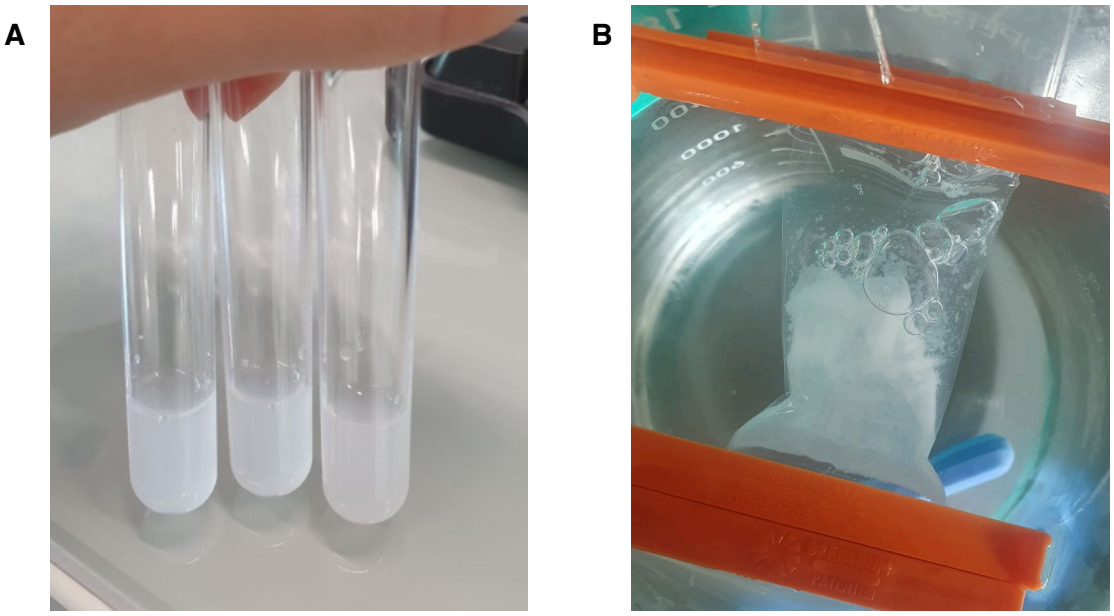


Figure 35: Precipitation of His-tagged TETs. Precipitation and aggregation occurred at various stages during the purification of His-tagged TETs, including (A) during IMAC or (B) post-IMAC dialysis.

3.1.4.1 Cobalt-based IMAC

Firstly, assuming that the difficulties encountered could stem from metal exchange between the IMAC column and the proteins, it could be sensible to charge the IMAC ligand with the appropriate metal. In our case, the columns were usually loaded with nickel. Considering that cobalt seems to be preferred by most TETs, I attempted to perform IMAC with a cobalt-charged resin; the column was stripped of all nickel ions using ethylene diamine tetraacetic acid (EDTA) and then recharged with cobalt ions, rendering potential metal exchanges between the column and the proteins inconsequential. Although no precipitation was observed, the small quantity of soluble protein recovered was insufficient for thorough purification. It is therefore difficult to conclude on the effectiveness of this method.

3.1.4.2 Metal boosting

As previously discussed, multiple oligomeric forms of TETs coexist. This is illustrated by the presence of several peaks during anion-exchange and size-exclusion chromatography. The main forms recovered during recombinant protein purification are the dimer and the dodecamer, with variable proportions; sometimes the dimer may be the predominant oligomer. Given that the dimer-dodecamer transition is induced by metal ions, metal supplementation appears to be a promising approach to promote the formation of dodecamer.

This strategy has already been used by Dutoit and colleagues for the purification of TmPep1050; all buffers were supplemented with 1 mM $CoCl_2$ (Dutoit *et al.*, 2019). At the ELMA laboratory, preliminary results were obtained concerning the metal-induced re-dodecamerization of dimers. Specifically, recombinant PhTET3 eluted from a strong AEX column in two distinct peaks, which were identified to be dimeric and dodecameric forms *via* SEC. The dimeric fraction was treated by overnight incubation with 0.2 mM $ZnCl_2$. Subsequent SEC analysis confirmed the successful reformation of dodecamers.

I attempted to apply this strategy during PsTET9 purification, where only the dimeric form was recovered after SEC. The protein was treated overnight with 1 mM of various metal salts, including $CaCl_2$, $CoCl_2$, $MgCl_2$, $MnCl_2$, $NiCl_2$, and $ZnCl_2$, before being analyzed by native PAGE. Unfortunately, the native PAGE protocol employed had been optimized for another protein, and our protein of interest did not migrate under the conditions used. Further optimization would have been necessary to obtain conclusive results, especially regarding the pH of the migration buffer.

A variation of this strategy, involving a milder introduction of the metal through dialysis, was also experimented. During the purification of PsTET7, most of the protein of interest did not bind to the weak AEX column and eluted in the flow-through. Based on previous results obtained with other TET proteins, these were likely light oligomeric forms, possibly monomers. Consequently, the flow-through was collected and dialysed against a buffer containing 0.2 mM $ZnCl_2$. The sample was then re-applied on the weak AEX column. While a portion PsTET7 was still recovered in the flow-through, some of the protein bound to the column and could then be eluted with a saline gradient, and further purified using subsequently strong AEX and SEC columns. According to the SEC chromatogram, the great majority of the purified protein was dodecameric. The dialysis with 0.2 mM $ZnCl_2$ therefore seems to have enabled re-dodecamerization of PsTET7. However, we cannot exclude the possibility that the quantity of protein injected on the weak AEX column was just too important; the binding capacity of the resin had perhaps just been exceeded. The same result would maybe have been obtained by reinjecting the flow-through directly on the column without dialysis. It would have been interesting to precisely evaluate the effects of $ZnCl_2$ treatment by re-injecting the flow-through on the column with or without metal dialysis in parallel. Unfortunately, due to time constraints and limited protein material, this experiment could not be conducted.

Finally, it was also attempted to supplement the lysis buffer with 0.2 mM $ZnCl_2$ to improve the solubility of stubborn proteins by promoting the formation of dodecamer right from the start of the purification, rather than during the final steps (**Figure 36**). Nonetheless, the qualitative SDS-PAGE analysis did not reveal any improvement of proteins solubility.

Overall, metal supplementation appears as a promising strategy to promote dodecamer formation, thereby improving protein solubility, stability, and purification yields. Despite some encouraging outcomes, none of the approaches presented here produced clear conclusive results. It would undoubtedly be worthwile looking further into this matter and trying to optimize the suggested protocols. For example, fine-tuning the metal concentration is definitely crucial to achieve more consistent and conclusive results, as this parameter is crucial for controlling the oligomeric state of TET peptidases.

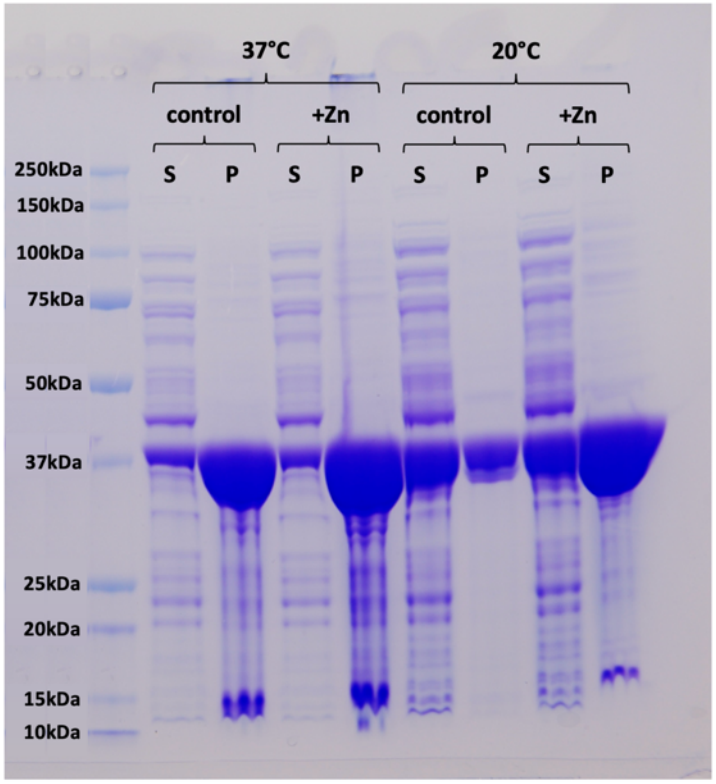


Figure 36: Effect of $ZnCl_2$ supplementation in the lysis buffer on PsTET8 solubility. Recombinant PsTET8 was expressed in Rosetta2(DE3)pLysS cells at either 37°C or 20°C. Cells were lysed in a buffer with or without 0.2 mM $ZnCl_2$. Following lysis, supernatants (S) and pellets (P) were separated by centrifugation and analyzed by SDS-PAGE. PsTET8 monomer migrated at an apparent molecular weight of ~42 kDa.

3.1.4.3 Alternatives to IMAC

Given the detrimental impact that the IMAC column or the imidazole used for elution can have, I then decided to bypass the IMAC step altogether. However, the samples were too complex after cell lysis to be directly applied to the strong AEX column, which is too resolutive for such crude mixtures, necessitating a clarification step first.

For thermostable proteins, a 15 min heat-shock at 70°C was used to eliminate most mesophilic host proteins. For non-thermostable proteins, ammonium sulfate precipitation was used, followed by a weak AEX column. Ammonium sulfate precipitation is a method used to purify and concentrate proteins by exploiting their differential solubility in high salt concentrations. When ammonium sulfate is added to a protein solution, the salt ions compete with proteins for water molecule (salting-out effect). The increased ionic strength enhances hydrophobic interactions, reducing the solubility of proteins and causing them to precipitate. By carefully adjusting the concentration of ammonium sulfate, specific proteins can be

selectively precipitated based on their unique solubility properties, allowing for their separation from other components in the mixture.

These clarification steps allowed to obtain samples pure enough for subsequent purification on the strong AEX column, followed by the final SEC column. This strategy allowed successful purification of dodecameric PsTET7 and HoTET8 enzymes (**Figure 37**). Nevertheless, only HoTET8 showed activity during the subsequent functional characterization.

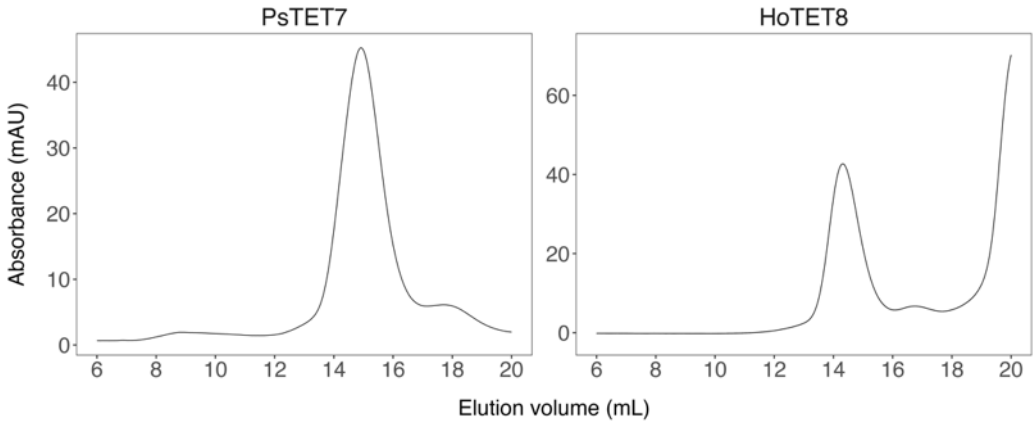


Figure 37: Size exclusion chromatography purification of His-tagged PsTET7 and HoTET8. Both proteins were purified using a Superose 6 Increase 10/300 GL column and eluted as high molecular mass complexes of c. 450 kDa.

3.1.5. Concluding remarks

IMAC purification of His-tagged proteins has already been successfully used on several instances for TET peptidases, and it initially appeared to be a convenient and suitable approach for our project, which aimed at purifying thirteen new enzymes of this family. However, this strategy only allowed the recovery of one functional dodecameric enzyme (*i.e.,* ThTET7). By bypassing the IMAC column, we were able to the purify of one additional active His-tagged protein (*i.e.,* HoTET8). Despite multiple attempts to enhance metal ion binding, the presence of the His-tag seemed to profoundly alter TET proteins, rendering them insoluble of inactive. Consequently, it was decided to fall back on untagged protein purification.

3.2. Production and purification of untagged TETs

The N-terminal His-tag introduced by the pET28a+ vector could have been removed *via* thrombin digestion. However, to avoid issues such as incomplete protease cleavage or retention of extra amino acids from the cleavage site in the fusion protein sequence, the genes

encoding the protein of interest were instead re-cloned into a different expression vector enabling the production of untagged proteins.

Given that the decision to resort to untagged protein purification was taken at the end of the second year of my PhD thesis, it seemed unrealistic to continue working with thirteen different proteins. We therefore narrowed our focus on a reduced set of targets, prioritizing the proteins most relevant for our study. Firstly, as active dodecameric proteins were obtained for His-tagged ThTET7 and HoTET8, untagged versions of these proteins were not produced. Secondly, the challenges in recombinantly expressing and purifying halophilic proteins like HvTET5 and HvTET6, which require multimolar salt concentrations for proper folding and solubility, led us to exclude them from our study. Additionally, AlTET11 and MfTET11 were ruled out because of their high isoelectric points (8.76 and 8.84, respectively), which would have required replacing the anion exchange (AEX) column used in all purification protocols with a cation exchange (CEX) I did not have at my disposal. Since these proteins belong to the same group as MtTET11, which could be purified using columns already available, it was decided to keep MtTET11 as the only representative of the TET11 group. Finally, we decided to focus on proteins assigned to a family and to discontinue our work on MkTET, which belonged to the broader and less defined TET-others group.

3.2.1. Heterologous expression optimization

Synthetic genes coding for HoTET7, PsTET7, PsTET8, PsTET9, TaTET10, and MtTET11 were cloned in the pET41c(+) vector. As described above for His-tagged protein production, expression and thermostability tests were carried out for each protein. The conditions yielding the highest expression levels and solubility were retained (**Table 5**). Similarly, the best results were obtained using Rosetta2(DE3)pLysS cells at 20°C for most proteins.

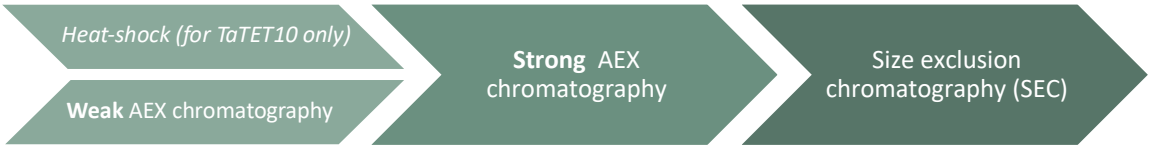
Table 5: Selected expression conditions for untagged protein production. The quantity and solubility of overexpressed proteins were evaluated by SDS-PAGE.

Protein	Expression cells	Expression conditions	Heat-shock resistant
HoTET7	insoluble		
PsTET7	Rosetta2(DE3)pLysS	16 h at 20°C	no
PsTET8	insoluble		
PsTET9	Rosetta2(DE3)pLysS	16 h at 20°C	no
TaTET10	Rosetta2(DE3)pLysS	16 h at 20°C	yes
MtTET11	Rosetta2(DE3)pLysS	16 h at 20°C	no

Since none of the conditions tested yielded soluble HoTET7 or PsTET8 proteins, we tried adding $ZnCl_2$ to the lysis buffer (as described in **section 3.1.4.2**) or varying the concentration of IPTG used to induce overexpression (0.01 / 0.1 / 1 / 5 mM). Although reducing the IPTG concentration slightly increased the solubility of PsTET8, the quantity of soluble protein remained too low for effective purification. Consequently, neither HoTET7 nor PsTET8 could be purified.

3.2.2. Development of the purification of untagged TETs

The following purification strategy was implemented. For thermostable TaTET10, a 15 min heat-shock at 70°C was used as a clarification step. For other non-thermostable proteins, we stopped using ammonium sulfate precipitation because this stage led to significant protein losses, and lysates were directly injected on a weak AEX column. These clarification steps were followed by strong AEX and SEC chromatographies.



This strategy allowed successful purification of dodecameric PsTET7, PsTET9, TaTET10, and MtTET11 that were all determined to be active during subsequent functional characterization. In contrast, the His-tagged versions of these proteins either formed dodecamers but were inactive (PsTET7), were active but only dimeric (TaTET10), or could not be purified at all, confirming that the His-tag was indeed responsible of the encountered issues.

3.3. Light at the end of the tunnel - overview of successful purifications

Overall, the following proteins were successfully purified using various combinations of affinity, anion exchange and gel filtration chromatography: His-tagged ThTET7 and HoTET8, and untagged PsTET7, PsTET9, TaTET10, and MtTET11. An overview of the purification strategies that were used is presented in **Table 6**, and detailed protocols are provided in **Appendix 3**.

During the final gel filtration chromatography step, all proteins eluted as well-separated high molecular mass complexes corresponding to particles of a molecular mass of *c.* 450 kDa. SDS-PAGE analysis of the elution peaks consistently showed a predominant band around 40

Enzyme	ThTET7	PsTET7	HoTET8	PsTET9	TaTET10	MtTET11
Tag	His-tag	-	His-tag	-	-	-
Purification protocol	• 15 min **heat-shock** at 70°C • **IMAC** Linear elution gradient 10 mM to 500 mM imidazole (elution peak at 400 mM) • **Strong AEX** Linear elution gradient 20 mM to 500 mM NaCl (elution peak at 300 mM) • **SEC**	• **Weak AEX** Linear elution gradient 20 mM to 600 mM NaCl (elution peak at 140 mM) • **Strong AEX** Linear elution gradient 50 mM to 500 mM NaCl (elution peak at 160 mM) • **SEC**	• 15 min **heat-shock** at 70°C • **Strong AEX** Linear elution gradient 75 mM to 300 mM NaCl (elution peak at 140 mM) • **SEC**	• **Weak AEX** Linear elution gradient 20 mM to 600 mM NaCl (elution peak at 170 mM) • **Strong AEX** Linear elution gradient 50 mM to 500 mM NaCl (elution peak at 200 mM) • **SEC**	• 15 min **heat-shock** at 70°C • **Strong AEX** Linear elution gradient 50 mM to 1 M NaCl (inverse purification: protein not bound to column) • **SEC**	• **Weak AEX** Linear elution gradient 20 mM to 600 mM NaCl (elution peak at 170 mM) • **Strong AEX** Linear elution gradient 50 mM to 500 mM NaCl (elution peak at 240 mM) • **SEC**
Purification yield	9.0 mg protein / L culture	2.7 mg protein / L culture	0.5 mg protein / L culture	5.5 mg protein / L culture	1.5 mg protein / L culture	1.2 mg protein / L culture

Table 6: Protocols used for the purification of six new M42 peptidases. All proteins were produced in *E. coli* using Rosetta2(DE3)pLysS. Protein expression was induced at 20°C for 16 h using 1 mM IPTG. Detailed protocols are provided in **Appendix 3**.

kDa, indicating the formation of homo-dodecameric complexes. These findings were further validated by negative-stain electron microscopy observations, which revealed homogeneous populations of tetrahedral particles for most enzymes (**Figure 38A**). Lower purity of PsTET9 and TaTET10 samples precluding satisfactory negative-staining imaging, structure predictions were generated using AlphaFold3 (**Figure 38B**). The resulting models displayed the expected hollow tetrahedral edifices with high confidence scores (ipTM 0.88 and 0.91), further supporting PsTET9 and TaTET10 ability to form high molecular weight assemblies. Alignment of the AlphaFold models with the crystallographic structure of the dodecameric PhTET2 (PDB code 1Y0R) yielded RMSD values of 0.783 Å for PsTET9 and 0.534 Å for TaTET10, indicating a robust superposition.

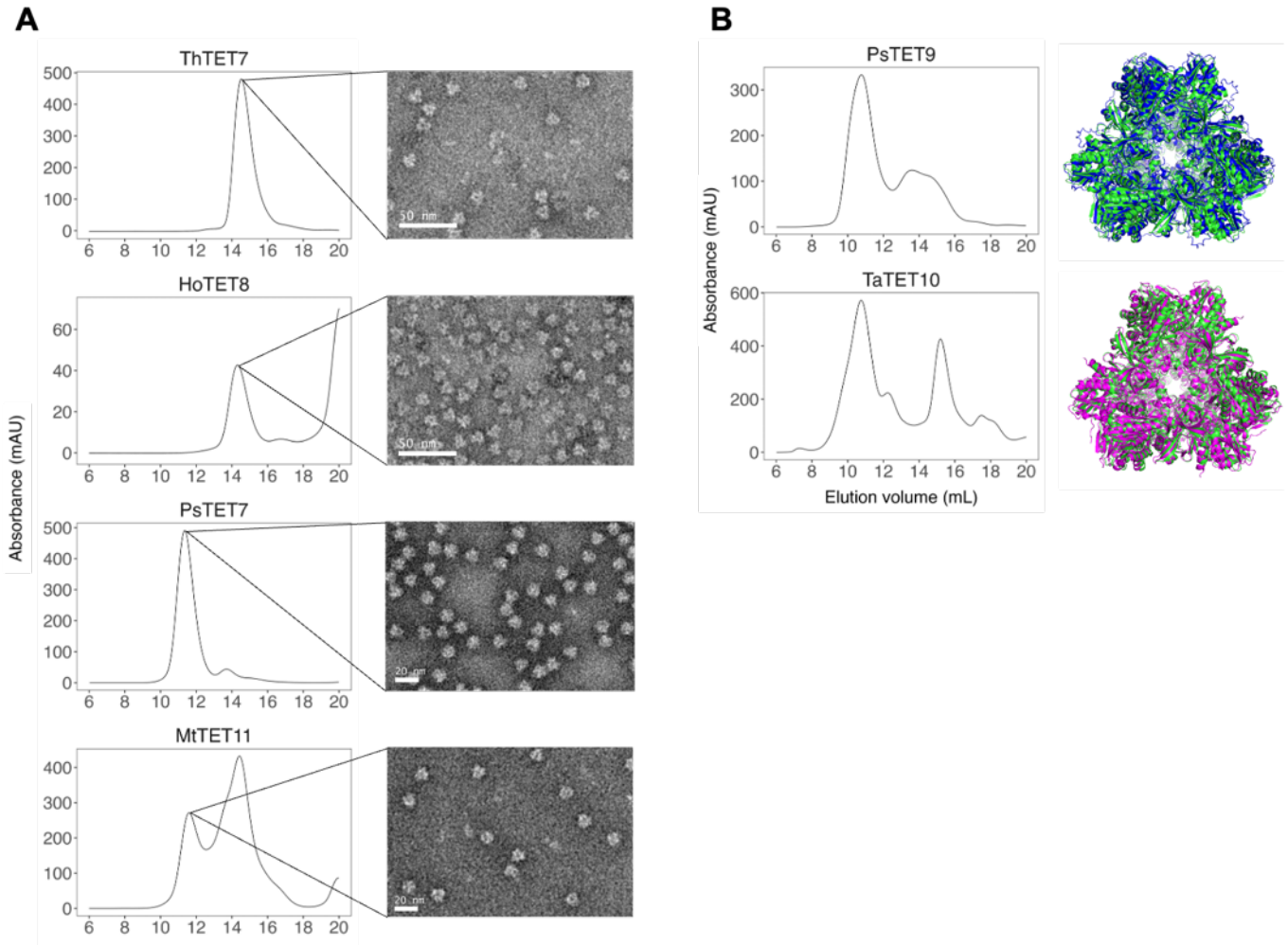


Figure 38: Oligomerization state of purified TETs. Size exclusion chromatography purification of ThTET7 and HoTET8 was carried out using a Superose 6 Increase 10/300 GL column. Other proteins were purified on a Superdex 200 10/300 GL column. All proteins eluted as high molecular mass complexes of c. 450 kDa. (a) Negative-stain electron microscopy observation of ThTET7, HoTET8, PsTET7, and MtTET11 revealed homogenous populations of hollow tetrahedral particles. (b) Typical TET structures were predicted for PsTET9 (blue) and TaTET10 (pink) using AlphaFold3. Models were aligned with PhTET2 structure (PDB code 1Y0R, in green).

4. Discussion and future perspectives

Ultimately, six proteins were successfully purified out of the thirteen initially targeted. Despite the difficulties encountered during the production and purification processes, we managed to obtain a set of purified proteins which, combined with pre-existing studies, covers almost all families defined during the phylogenetic analysis presented in **chapter 1**; only the TET5 and TET-others groups will be left unexplored.

While our first intent was to use a generic protocol for all proteins, each purification required considerable optimization. Given the vast diversity of TET peptidases revealed through the screening of archaeal genomes (see **chapter 1**) and considering the challenges associated with the production and purification of TET peptidases discussed in this chapter, it appears unlikely that a universal protocol could be successfully applied to all proteins. In particular, while the use of a His-tag and IMAC purification can be highly effective, it may also introduce particular challenges and complications with respect to TET peptidases. As perfectly illustrated in this chapter, each protein may respond differently to this purification method; some will tolerate it very well, but others may have their functional integrity or structural stability impaired.

It is therefore difficult to draw any general conclusion regarding the use of a His-tag for these proteins. It seems sensible to test this approach for each protein individually on a small-scale. Specifically, the stability of the tagged protein after being subjected to an IMAC column should be evaluated.

If one wishes to use affinity purification for TET peptidases, it may be preferable to use a different type of tag. For instance, the Strep-tag seems to be a promising alternative (Schmidt & Skerra, 2007). This approach was, in fact, adopted by Dutoit and colleagues for the purification of *E. coli* M42 peptidases after they encoutered difficulties with His-tagged enzymes from *T. maritima* (Dutoit, 2021). It should be mentioned that the resins used for purifying Strep-tagged proteins are more expensive, more sensitive to regeneration processes, limiting their re-use, and have a lower binding capacity compared to those used for His-tag purification (Pina *et al.*, 2014). Nevertheless, since no metals are involved in Strep-tag purification, the potential issues associated with His-tag, as discussed in **section 3.1.3** above, are effectively avoided.

Whenever possible, it would be advisable to avoid using any tag, as it sometimes can alter the protein's natural structure, stability, or function, potentially affecting its biological activity or interactions with other molecules. For instance, it seems dispensable in the case of thermostable proteins for which an initial heat-shock clarification step already allows to eliminate most of the protein contaminants.

Although more time-consuming and labor-intensive, untagged purification is feasible for non-thermostable proteins, as demonstrated in **section 3.2.2** above. Due to constraints in time, protein material, and availability of chromatography columns, the presented purification protocols were not fully optimized and would benefit from further refinement. In particular, the salt concentrations used for the equilibration, washing, and elution steps in AEX chromatographies could be optimized to improve the separation of the proteins of interest from contaminants. In addition, it could be preferable to avoid using two types of AEX columns (weak and strong) sequentially, as both bind similar charged species. This redundancy can result in little to no improvement in purity, thereby reducing purification process efficiency. Instead, combining AEX with an orthogonal purification method, such as hydrophobic interaction chromatography (HIC), could enhance purification effectiveness. Finally, due to its weak chelating properties for divalent metal ions, it might be prudent to avoid using Tris buffer when working with metal-dependant (Allen *et al.*, 1967; Hanlon *et al.*, 1966). This chelation can reduce the availability of metal ions, potentially inhibiting the activity of the protein by disrupting its active site or structural stability. Good's buffers, such as HEPES or PIPES, which have minimal metal-binding affinity, could be preferable alternatives to Tris.

5. Résumé du chapitre en français

Ce chapitre présente les défis liés à la production et à la purification de nouvelles aminopeptidases TET de la famille M42. Etant donné le nombre important de protéines visées dans le cadre de ce projet, une approche standardisée de production et de purification a initialement été adoptée, utilisant un protocole générique qui devait pouvoir être facilement adapté pour chaque protéine. Cette stratégie se justifiait en raison de la cohérence des techniques et protocoles employés pour étudier diverses aminopeptidases TET au laboratoire ELMA. En particulier, l'utilisation d'un tag histidine et de la chromatographie par affinité pour les métaux immobilisés (IMAC), méthode ayant déjà été utilisée avec succès pour la purification de plusieurs peptidases M42, devait faciliter et accélérer le processus de purification. Cependant, cette méthode n'a permis d'obtenir qu'une seule enzyme fonctionnelle, ThTET7. Pour toutes les autres protéines, des problèmes d'instabilité et d'agrégation ont été rencontrés.

Les défis rencontrés lors de l'utilisation de ce protocole standardisé et les stratégies développées pour surmonter ces difficultés sont exposés dans ce chapitre. Il permet notamment de détailler comment la présence d'un tag histidine et l'utilisation de l'IMAC peuvent affecter la stabilité et l'activité des métalloprotéines en perturbant la liaison des cofacteurs métalliques, qui sont essentiels pour ces protéines. Différentes stratégies mises en place, visant à rétablir la liaison des cofacteurs métalliques, sont présentées. Cependant, seule HoTET8 ayant ainsi pû être purifiée, le recours à une purification sans tag a été envisagée comme l'alternative la plus convaincante.

Ce changement de stratégie, intervenu tardivement au début de la troisième année de thèse, a nécessité une réduction du nombre de protéines cibles. Les deux protéines déjà purifiées n'ont pas été reproduites et les protéines présentant des défis de purification importants ont été écartées. Seules les protéines les plus pertinentes de chaque famille, à savoir PsTET7, PsTET9, TaTET10 et MtTET11, ont été retenues pour une purification sans tag. Les protocoles utilisés pour ces protéines sont détaillés dans le chapitre.

En conclusion, ce chapitre souligne l'importance d'adapter les stratégies de purification aux caractéristiques spécifiques de chaque enzyme. Des recommandations générales pour la purification des peptidases TETs sont également proposées.

Chapter 3

Expanding the understanding of TET peptidases: functional characterization of six novel enzymes

1. Foreword

Previous functional characterization of a few archaeal peptidases has already demonstrated significant diversity in TET functional characteristics, particularly in terms of substrate specificity (see **introduction section 3.3.2**) (Basbous *et al.*, 2018; Durá *et al.*, 2005, 2009; Durá & Franzetti, 2013; Franzetti *et al.*, 2002; Slutskaya *et al.*, 2012).

However, the comprehensive identification of TET peptidases in archaeal genomes presented in **chapter 1** uncovered a much broader and unexpected diversity, revealing that our current understanding of these enzymes is incomplete and does not capture the full extent of their diversity. In light of these results, we wanted to investigate whether the diversity observed through phylogenetic analysis correlates with a corresponding functional diversity, and to explore the potential existence of novel unreported substrate specificities. For this purpose, thirteen proteins were selected, covering all TET peptidase families delineated in our analysis and spanning a range of taxonomically diverse archaeal species (see **chapter 1 section 3.2.3**).

To this day, the factors determining the substrate specificity of TET peptidases remain poorly understood. Attempts to deduce substrate specificity on the basis of sequence similarity are hampered by the lack of clear sequence signatures that correlate with experimentally determined function, and even structural analyses have yielded limited insights. For the most thoroughly characterized enzymes, the combination of biochemical and structural data points to the criticial role of the specificity pocket adjacent to the catalytic site. The dimensions of this pocket and the charge characteristics of its defining residues could be driving the substrate specificity of the enzyme (see **introduction section 3.3.2**).

Despite these promising indications, *in silico* prediction of substrate specificities for uncharacterized enzymes remains out of reach, leaving experimental biochemical approaches as the only mean to elucidate the functional properties of these peptidases. Accordingly, the following chapter will cover the functional characterization of the six novel TET peptidases that were successfully purified: PsTET7, ThTET7, HoTET8, PsTET9, TaTET10, and MtTET11 (see **chapter 2**). Specifically, the determination of their cleavage specificities and optimal enzymatic conditions (*i.e.,* temperature, pH, and metallic cofactors) will be presented.

2. Material and methods

TET peptidase hydrolytic activities on synthetic chromogenic and fluorogenic substrates were assayed using aminoacyl-para-nitroaniline (pNA) and aminoacyl-7-amino-4-methylcoumarin (AMC) conjugates ordered from Bachem. Substrates were solubilized in 100% dimethylsulfoxide (DMSO) to a final concentration of 20 mM.

2.1. Enzymatic characterization general protocol

All assays described below were carried out according to the following standard procedure (Durá *et al.*, 2005). Reactions were initiated by addition of 2 to 10 μg/mL of enzyme to a pre-warmed mixture containing 2.5 mM of the synthetic substrate in 50 mM buffer (pH 5,5 – 11), 150 mM KCl, and 1 mM XCl_2 (X = Ca, Co, Fe, Mg, Mn, Ni or Zn) in a total volume of 60 µL. To avoid water evaporation, the total volume was covered by 25 µL of mineral oil. Incubations were performed for 3 min to 1 h, reactions were stopped by the addition of 60 µL of 0.1 M acetic acid, and samples were placed on ice. After centrifugation at 6,000 × *g* for 3 min, liberated pNA or AMC quantities were quantified by OD_{405} or fluorescence (excitation and emission wavelengths 360 nm and 460 nm, respectively) measurement in a Synergy HT microplate reader (BioTek) (**Figure 39**). Three replicates and two enzyme blanks were assayed for each experimental point.

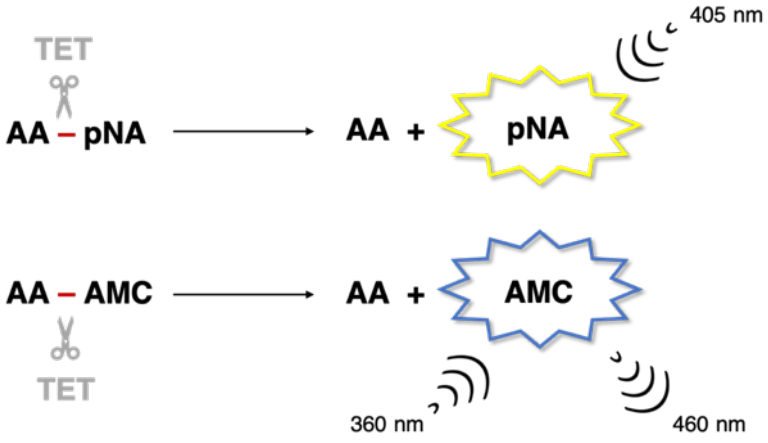


Figure 39: Principle of aminopeptidase activity assays on monoacyl-pNA/AMC conjugates. Synthetic substrates consisting of amino acids linked by a peptide bond (in red) to the chromophore pNA or the fluorophore AMC were used.

Enzyme concentrations and incubation durations were adjusted for each peptidase to produce a robust signal for accurate measurement (**Table 7**).

Table 7: Enzyme concentration and incubation duration used for each enzyme.

TET	Enzyme concentration	Incubation duration
PsTET7	10 μg/mL	60 min
ThTET7	10 μg/mL	3 min
HoTET8	2 μg/mL	3 min
PsTET9	10 μg/mL	20 min
TaTET10	10 μg/mL	15 min
MtTET11	8 μg/mL	10 min

2.2. Effect of temperature, metal cations, and pH on TET peptidase activities

The optimal temperature, pH and metallic cofactor were determined using the substrate on which maximum activity was measured.

The effect of temperature on M42 peptidase activities was evaluated between 20 and 100°C. Assays were conducted as previously described in presence of 50 mM HEPES, 150 mM KCl, and 1 mM $CoCl_2$, pH 7.5. To prevent enzyme denaturation and to ensure stable enzymatic activity**,** optimal pH and metallic cofactor were then established 10°C below the determined optimal temperature.

The effect of metal cations on M42 peptidase activities was assessed using 1 mM (0.1 mM for TaTET10) of XCl_2 metal (X = Ca, Co, Fe, Mg, Mn, Ni or Zn) with 50 mM HEPES, 150 mM KCl, pH 7.5.

The influence of pH was studied in presence of 1 mM $CoCl_2$ (0.1 mM for TaTET10) using the following buffers: MES (pH 5.5 to 6.5), HEPES (pH 7.0 to 8.0), CHES (pH 8.5 to 9.5), and CAPS (pH 10.0 to 11.0).

2.3. Substrate specificity determination

For each peptidase, substrate specificity was determined using optimal metal cofactor and pH. Incubations were performed 10°C below the established optimal temperatures.

2.4. Effect of metal cation concentration on TaTET10 and MtTET11 activities

TaTET10 and MtTET11 activities were determined on Met-pNA in presence of either 0.1 mM or 1 mM of XCl_2 metal (X = Ca, Co, Fe, Mg, Mn, Ni or Zn). Assays were conducted as previously described, using 50 mM HEPES, 150 mM KCl, at pH 7.5 and 80°C for TaTET10, and at pH 8 and 50°C for MtTET11.

3. Results

3.1. Archaeal TET peptidases exhibit remarkable functional versatility

Cleavage specificities were determined using chromogenic (pNA conjugated) and fluorogenic (AMC conjugated) aminoacyl substrates. These synthetic substrates are commonly used to study the activity of proteolytic enzymes as they enable straightforward and quantitative enzyme activity assays. Upon cleavage by a peptidase, they release either pNA, which produces a yellow color measurable at 405 nm, or AMC, which fluoresces and can be detected with high sensitivity. The flexibility in conjugating different amino acids to pNA or AMC enables the investigation of aminopeptidases substrate specificities.

As presented in **Figure 40**, observed activity spectra were heterogeneous and could be divided into two main groups. The first group includes HoTET8 and PsTET9, which exhibited selective activities and can be classified as specialized enzymes. Analogous to the previously described PhTET1 peptidase (Durá & Franzetti, 2013), they specifically targeted acidic amino acids. Interestingly, PsTET9 maximum activity was measured on Glu-pNA, whereas no hydrolysis could be detected on Asp-pNA despite the similarity of these substrates. The same substrate specificity has already been reported for the MHJ_0125 glutamyl-aminopeptidase of *Mycoplasma hyopneumiae* (Robinson *et al.*, 2013). No activity profiles resembling PhTET3, which targets basic residues (Durá *et al.*, 2009), or PhTET4, which specifically cleaves glycine (Basbous *et al.*, 2018), were identified.

Conversely, PsTET7, ThTET7, MtTET11, and TaTET10 exhibited broad-spectrum activity and can be classified as generalist enzymes. These enzymes were found to preferentially cleave hydrophobic residues, with PsTET7 and ThTET7 displaying the broadest specificities. Optimal amidolytic activities were observed with Ile-pNA, Leu-pNA, Met-pNA, and Met-pNA, respectively. Similarities with the cleavage profiles of PhTET2 (Durá *et al.*, 2005) or MjTET11 (Basbous, 2016) can be outlined, but this study is the first description of methionyl and isoleucyl aminopeptidases in the TET family. Notably, none of the newly characterized enzymes exhibited an activity profile similar to HmTET6, which is capable of cleaving a remarkably broad range of neutral, basic, and acidic residues (Franzetti *et al.*, 2002).

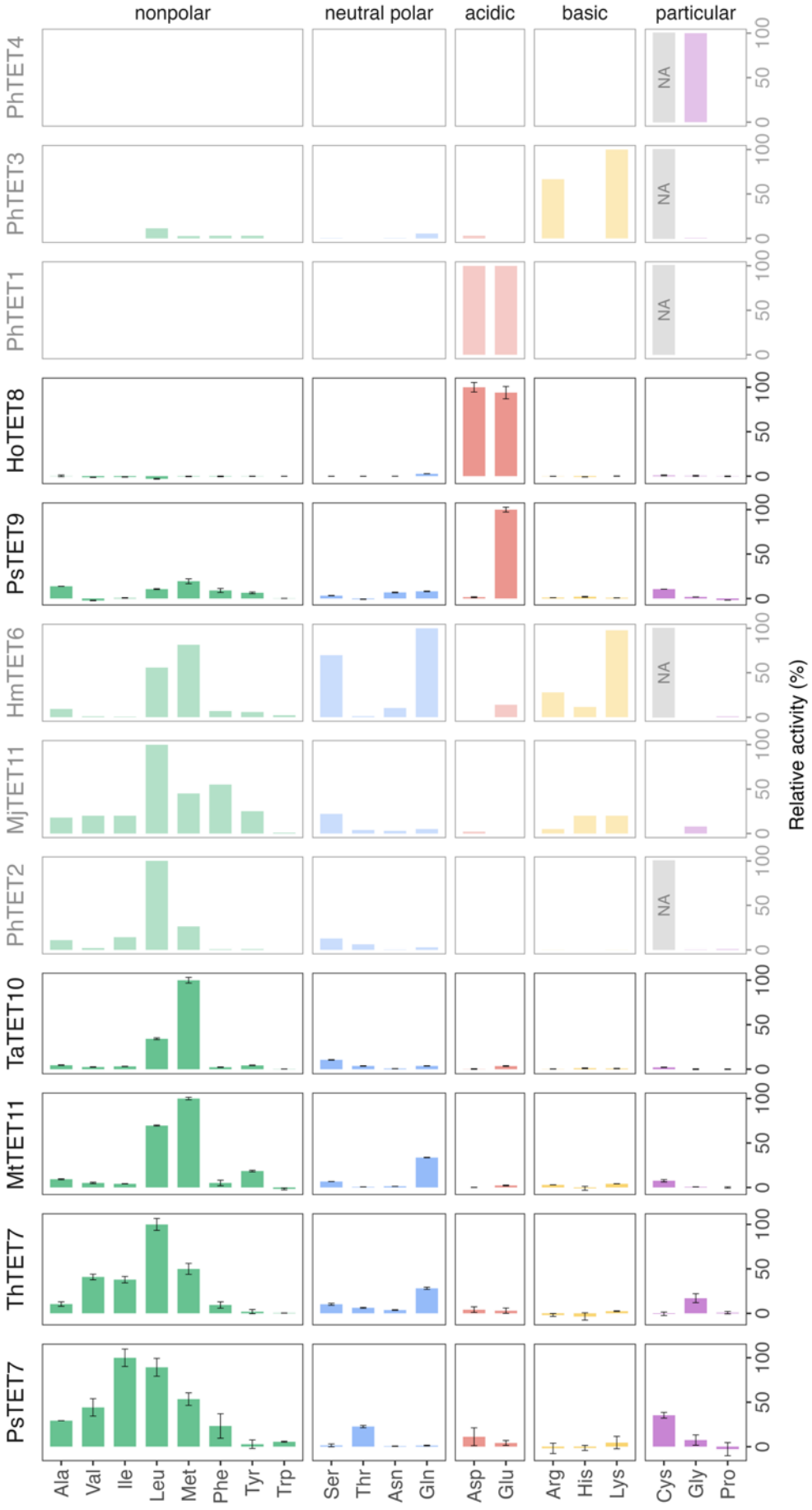


Figure 40: Characterized TET aminopeptidases exhibit diverse substrate specificities. Cleavage specificities were assayed using synthetic chromogenic and fluorogenic substrates. For each enzyme, activities are expressed as percentage of the maximum activity observed, which was attributed a value of 100%. Activity profiles of enzymes characterized prior to this study (*i.e.*, PhTET1, PhTET2, PhTET3, PhTET4, HmTET6, and MjTET11) are also presented. NA: not assessed. Error bars indicate ±s.d. with n=3.

Interestingly, the two enzymes featuring the novel insertion identified in Asgard sequences (*i.e.,* PsTET7 and ThTET7) displayed similar but broader spectrum of activity compared to PhTET2, MtTET11, and TaTET10, effectively cleaving all hydrophobic aliphatic amino acids. According to AlphaFold3 structure predictions (**Appendix 4**), these insertions would form protruding two-stranded β-sheets that partially obstruct the pores located on the faces of the tetrahedral particle, which are believed to be the entry points for substrates (Schoehn *et al.*, 2006). The presence of these insertions does not appear to affect the oligomerization of PsTET7 and ThTET7 since both enzymes were found to form typical hollow tetrahedral particles (see **chapter 2 section 3.3**). Nonetheless, these protrusions could be involved in substrate recognition and selection, conferring broader specificities to PsTET7 and ThTET7. Alternatively, they could be mediating interaction with a partner protein. To fully elucidate the role of this novel insertion, further functional and structural studies on representative members of this family should be prioritized.

Overall, the results obtained for these new enzymes, combined with pre-existing data, underscore the remarkable functional diversity of TET peptidases, which display highly contrasting substrate specificities. Two distinct activity profiles emerge: generalist enzymes, which effectively cleave a broad range of substrates, particularly hydrophobic residues, and specialized enzymes, which exhibit narrower substrate preferences, targeting specific amino acids with high selectivity.

3.2. Effects of temperature, pH, and metal ions on TET peptidase activities

TET enzymatic behaviors were investigated across a temperature range of 20°C to 100°C (**Figure 41**). The studied peptidases exhibited markedly contrasting profiles, with maximum activities measured for temperatures ranging from 40°C to 90°C. Specifically, PsTET7 and PsTET9 showed peak activity within the mesophilic range, while MtTET11, ThTET7, TaTET10, and HoTET8 exhibited a preference for higher temperatures, with optimal temperatures for enzymatic activities of 60°C, 70°C, 90°C, and 90°C, respectively.

Of note, during the study of the effect of temperature on PsTET7 activity, two activity peaks were consistently observed: one at 40°C and another at 80°C. Given that the native organism of this enzyme, *Prometheoarchaeum syntrophicum*, has been shown to grow optimally at 20°C and only grows weakly at 40°C (Imachi *et al.*, 2020), we questioned the validity of the second peak observed in our experiments. An additional experiment was thus

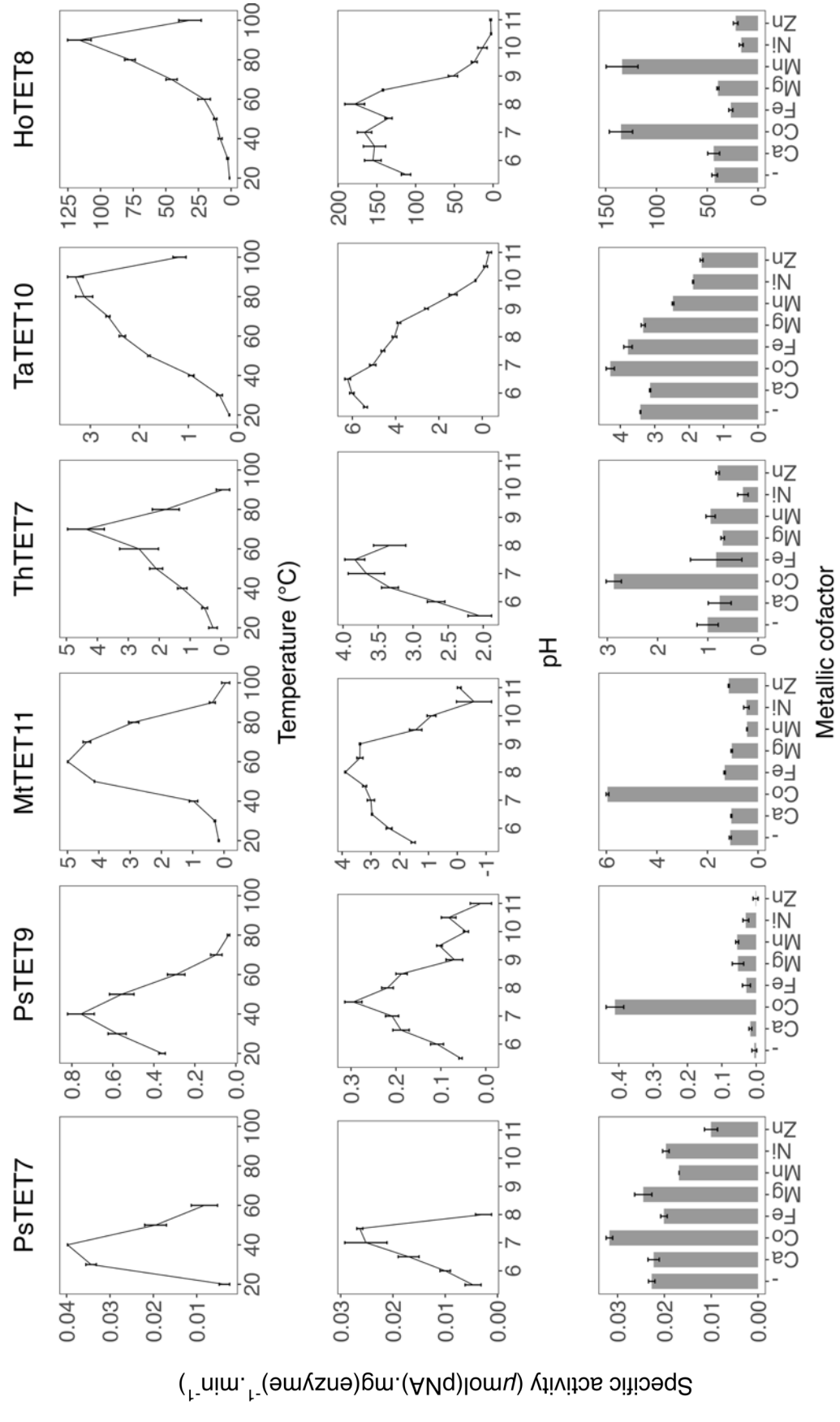


Figure 41: Temperature, pH, and metal ions influence on TET enzymatic activities. For each enzyme, optimal conditions regarding temperature (top), pH (middle), and divalent cations (bottom) were determined. Error bars indicate ±s.d. with n=3.

conducted to investigate this phenomenon. PsTET7 was incubated at 80°C for varying durations from 0 to 60 minutes before being tested for activity on Ile-pNA at 30°C, pH 7.5 in presence of 1 mM $CoCl_2$. The results of this experiment, conducted in triplicate and presented in **Figure 42**, show that the enzyme's activity decreased by 75% after just 5 minutes of incubation at 80°C. These findings suggest that the activity peak observed at 80°C does not represent a true maximum of activity but rather an artifact caused by enzyme denaturation, which likely exposes the active site, combined with increased molecular agitation at high temperatures, leading to a transient spike in activity. Consequently, we decided to ignore this peak and to consider the PsTET7 optimal activity to be at 40°C, which is in agreement with the optimal temperature observed for the other characterized enzyme of *P. syntrophicum* PsTET9 (**Figure 41**).

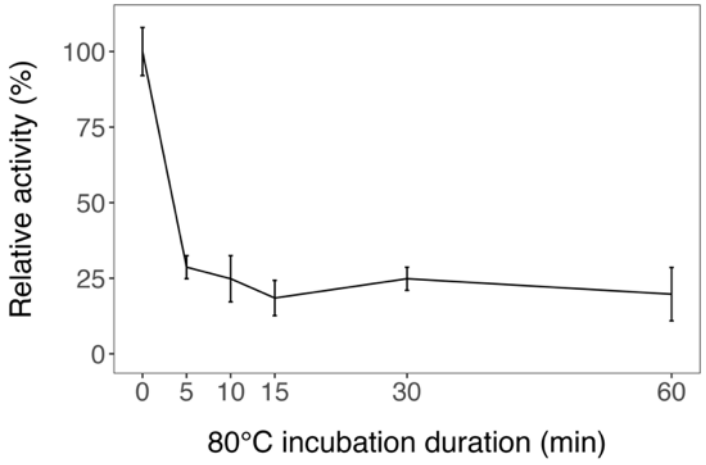


Figure 42: Residual activity of PsTET7 after incubation at 80°C. The enzyme was incubated at 80°C for 0 to 60 min before being used for activity assays against Ile-pNA at 30°C, pH 7.5 in presence of 1 mM $CoCl_2$. Error bars indicate ±s.d. with n=3.

The enzymatic activities of TET peptidases were also investigated at pH values from 5.5 to 11.0. All assayed enzymes exhibited maximum amidolytic activity at neutral pH values (**Figure 41**). Notably, MtTET11 and HoTET8 exhibited higher resilience to alkaline environments, whereas aggregation of ThTET7 and PsTET7 enzymes was observed above pH 8, revealing their instability under these conditions. TaTET10 showed increased tolerance to acidic conditions, maintaining 87% of its optimal activity at pH 5.5.

The effects of several divalent cations on the enzymatic activities were also examined (**Figure 41**). Consistent with prior characterizations of archaeal and bacterial TET peptidases (Colombo *et al.*, 2016; Durá *et al.*, 2005, 2009; Durá & Franzetti, 2013; Dutoit *et al.*, 2019),

maximum stimulatory effect was observed with Co^{2+} ions for all enzymes. However, TaTET10 was inhibited by all assayed metallic ions at a 1 mM concentration, and stimulatory effect of Co^{2+} was restored at a reduced concentration of 0.1 mM (see **section 3.3**). These findings are in agreement with previous reports of concentration-dependent activation of TET peptidases by metallic ions (Colombo *et al.*, 2016). Interestingly, while the APDkam589 peptidase from *Desulfurococcus amylolyticus*, here renamed DaTET10, is known to be activated by Mg^{2+} and Mn^{2+} ions (Slutskaya *et al.*, 2012), HoTET8 stands out as the first characterized enzyme in this family to exhibit equal activation by Co^{2+} and Mn^{2+} ions.

3.3. Concentration-dependent activation of TET peptidases by metal cofactors

Following the observed inhibition of TaTET10 by 1 mM divalent ions (see **section 3.2**), further tests were performed at both 0.1 mM and 1 mM concentrations. Due to time constraints and limited protein material, this additional analysis was only extended to MtTET11 (**Figure 43**).

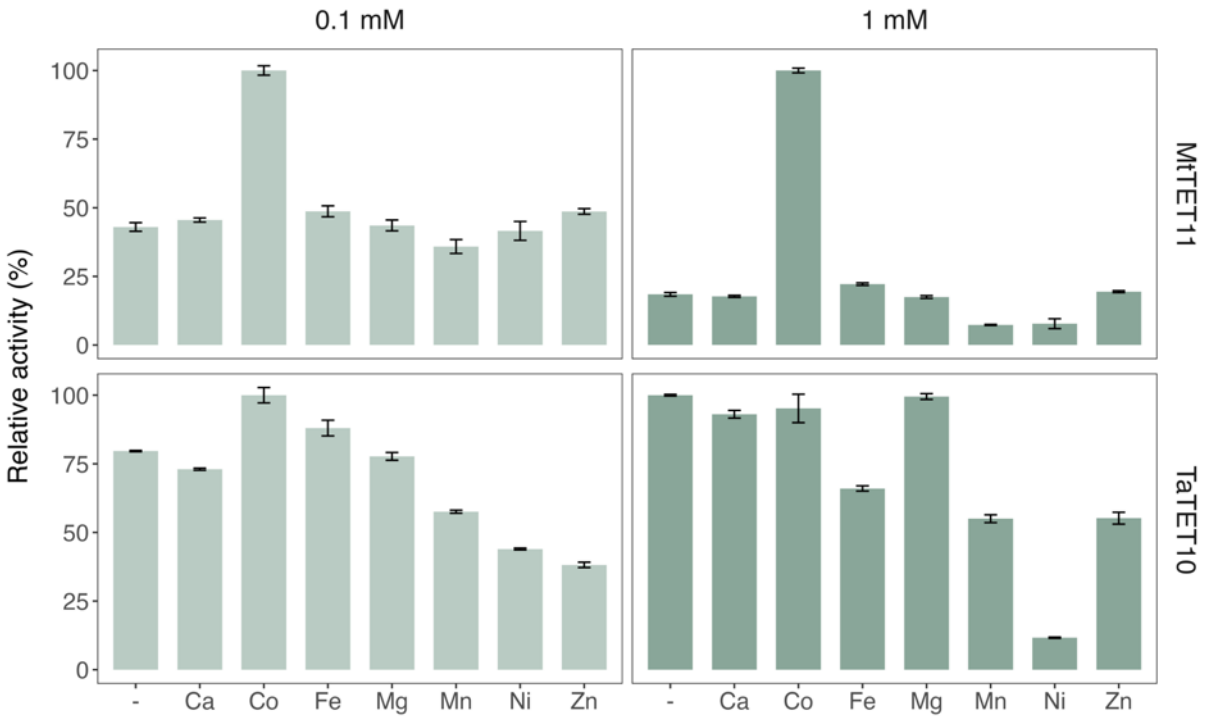


Figure 43: Effect of various metal ions at 0.1 or 1 mM concentrations on TaTET10 and MtTET11 activities. For each condition, activities are expressed as percentage of the maximum activity observed, which was attributed a value of 100%. Error bars indicate ±s.d. with n=3.

At lower concentrations, maximal activation of TaTET10 was observed with Co^{2+} ions. Interestingly, Fe^{2+} also had a weak activating effect, whereas it had a strong inhibitory effect at 1 mM. Regarding MtTET1, maximum stimulatory effect was observed with Co^{2+} ions in both conditions, though this effect diminished at lower concentrations. Notably, the activating effect of Co^{2+} was less pronounced for TaTET10 compared to MtTET11.

According to previous studies, a 1 mM concentration of metal ions stimulates the activity for many TET peptidases; however, inhibition at this concentration is not unique to TaTET10, as other enzymes are similarly affected. For instance, Colombo *et al.* (2016) investigated the influence of metal ion concentration on PfTET3 activity, and observed that at 85°C, 0.1 mM Co^{2+} strongly activated the enzyme, whereas 1 mM resulted in only weak activation.

3.4. Impact of metallic cofactor concentration on TET peptidase substrate specifity

One of the key findings from the study by Colombo *et al.* (2016) is the existence of a third metal-binding site M_3 in PfTET3 aminopeptidase, identified through a combination of X-ray crystallography and anomalous X-ray scattering experiments. This site is coordinated by the side chains of residues Glu281, Asp254, and Thr232, and is located in the enzyme's specificity pocket.

The authors also found that the M_1, M_2, and M_3 sites have distinct binding affinities for metal ions, therefore enabling to selectively control the occupation of these sites by carefully adjusting Co^{2+} concentrations. They took advantage of this property to favor Co^{2+} binding at the M_3 site, while keeping M_1 and M_2 predominantly occupied by Zn^{2+}. They were thus able to demonstrate that Co^{2+} binding at M_3 altered the electrostatic properties of the specificity pocket, extending the substrate specificity of the enzyme to include the degradation of glutamate, which is normally not processed by the lysyl-aminopeptidase.

To date, this third metallic site has only been identified in PfTET3, and it remains to be determined whether the enzymes characterized in this project possess this additional site. Nonetheless, we investigated the effects of metal cofactor concentration on TET peptidase substrate specificities by determining the activity spectra of MtTET11 in presence of either 0.1 mM or 1 mM $CoCl_2$ (**Figure 44**).

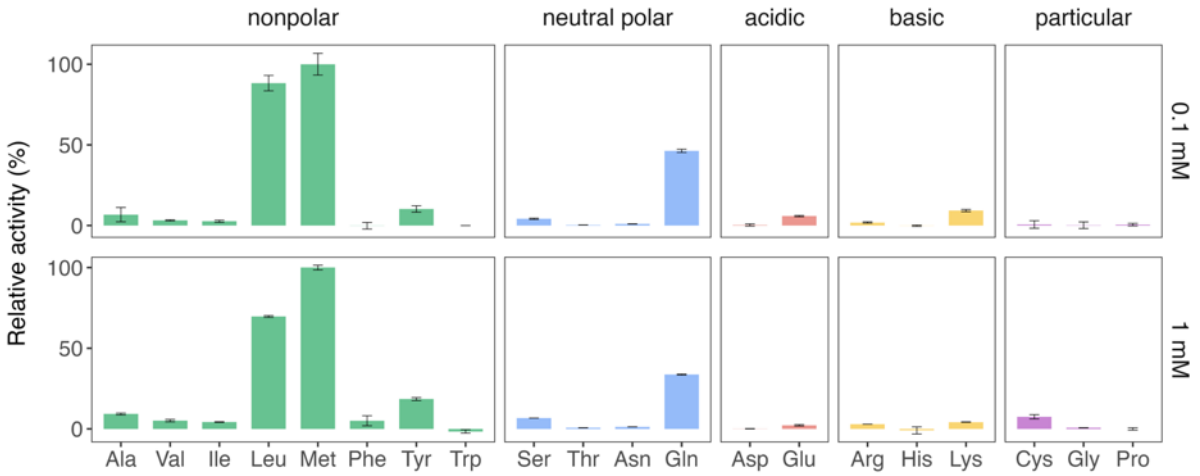


Figure 44: **Impact of metallic cofactor concentration on MtTET11 substrate specificity.** MtTET11 substrate specificity was determined in presence of either 0.1 mM or 1 mM $CoCl_2$. For each condition, activities are expressed as percentage of the maximum activity observed, which was attributed a value of 100%. Error bars indicate ±s.d. with n=3.

In our case, no impact of metal cofactor concentration on the substrate specificity of the enzyme could be observed. While the enzyme was found to be approximately four times more active in presence of 1 mM $CoCl_2$ than with 0.1 mM, the enzyme's activity spectrum did not appear to be affected by this variation in metal ion concentration. These results can naturally not be generalized for all enzymes, and it would have been necessary to repeat the experiment with the five other peptidases. It could also have been insightfull to test a wider range of metallic ion concentrations. However, limited time prevented a more in-depth exploration of this matter.

3.5. Cd-ninhydrin assay

In addition to the activity assays on synthetic substrates presented above, I also developed the use of the ninhydrin assay, a colorimetric method allowing to detect free amino acids generated by the exopeptidase activity of TETs. The free amino groups react with ninhydrin–a chemical reagent that specifically targets primary amines–generating a colored complex known as Ruhemann's purple. The intensity of the color, measured at 570 nm using a spectrophotometer, is directly proportional to the concentration of free amino groups in the solution (Moore & Stein, 1948).

Since ninhydrin reacts with all primary amines, the N-terminal extremity of undegraded peptides will also be targeted, resulting in background signal. To mitigate this issue, Doi *et al.* (1981) proposed a modification of the conventional ninhydrin method, incorporating cadmium (Cd) to the ninhydrin reagent. The Cd-ninhydrin complex more selectively reacts with free

amino acids, reducing the background signal. This reaction produces a red-colored compound, which can detected at 505 nm.

This assay provides a simple but effective tool to study the activity of peptidases, and can be very convenient as a preliminary method. However, it does not provide information on the enzyme's substrate specificity, as it only measures the release of free amino acids without distinguishing between different substrates. Additionally, the Cd-ninhydrin reagent contains high concentrations of cadmium, which is a carcinogenic, mutagenic, and reproductive toxin, and should be avoided for routine use. This method has therefore only been used for occasional tests, and the results will not be discussed in detail here. Nevertheless, given its potential utility in specific contexts, the developed protocol is provided in **Appendix 5**.

4. Discussion and perspectives

Sampling of the previously unexplored families for functional characterization revealed the existence of highly contrasted substrates specificities, bringing the functional versatility of TET peptidases to light (**Figure 40**). Temperature, pH, and metal cofactor optima determined for MtTET11, TaTET10, PsTET7, and PsTET9 are consistent with the optimal growth conditions of *Methanoculleus thermophilus*, *Thermosphaera aggregans*, and *Candidatus Prometheoarchaeum syntrophicum* (Huber *et al.*, 1998; Imachi *et al.*, 2020; Rivard & Smith, 1982; Zabel *et al.*, 1985). The physiology of the uncultivated species *Candidatus Hodarchaeales archaeon LC_3* and *Candidatus Thorarchaeota archaeon MP8T_1* remain unknown, and data concerning the physico-chemical conditions at their sampling sites are lacking. Regardless, significant differences between the sampling site conditions and the species' native environment are commonly observed for metagenomic samples (Cangelosi & Meschke, 2014; Rose *et al.*, 2014). Nevertheless, the optimal growth temperatures of Asgard species have been estimated in two parallel studies using either genome-derived features or the optimal GDP-binding temperature of the EF-1α translation elongation factor (Eme *et al.*, 2023; Lu *et al.*, 2024). Interestingly, the high optimal temperatures obtained for ThTET7 and HoTET8 (70°C and 90°C, respectively) are in disagreement with the predicted mesophilic lifestyles of Thorarchaeales and Hodarchaeales (**Figure 41**).

It is worth noting that the observed optima may be affected by experimental biais. In particular, the activity assays were based on endpoint measurements rather than continuous monitoring of substrate degradation, which raises a few potential issues. One concern is ensuring that, within the experimental conditions, the enzyme did not fully degrade the substrate by the time the reaction was stopped. In this case, this problem can be ruled out, as I confirmed that all signals obtained were lower than what would have been measured in the case of complete substrate digestion. One should also keep in mind that the incubation times and enzyme concentrations had to be adjusted for each enzyme to ensure detectable and reliable signals; the incubation times ranged from 3 to 60 minutes, and the enzyme concentrations from 2 to 10 μg/mL. Our objective was not to determine kinetic parameters such as Km or Vmax, and it may seem that varying incubation times is less critical since we were not measuring initial velocity.

However, it has been shown that both the duration of the reaction and the enzyme concentration can influence the temperature optima. For instance, Almeida *et al.* (2019)

reported a 5°C shift in the optimal temperature of *Spodoptera frugiperda* β-glucosidase (from 42°C to 37°C) when extending the assay time from 20 to 60 minutes. This can be explained by the fact that the optimal temperature is often close to the point where the enzyme begins to denature. Thus, while the enzyme may be more active at 42°C, it may not be able to sustain this temperature for extended periods. To mitigate this issue, we chose to determine the optimal pH, metal cofactors, and substrate specificities 10°C below the determined optimum temperatures. It would have been valuable to complement our study with half-life experiments to further assess enzyme stability.

Additionally, the authors reported that the optimal temperature varied with enzyme concentration, being 37°C at 85 nM and 42°C at 280 nM. As far as possible, it would be advisable to use consistent concentrations for all enzymes, or at least keep concentrations within the same order of magnitude. In our case, we ensured that enzyme concentrations were kept between 2 and 10 μg/mL. Given the varying levels of activity among our enzymes, we could not reduce this range any further.

Despite these potential experimental biais, that could mainly be affecting the determined optimal temperatures, the obtained results support the delination of TET families presented in **chapter 1 section 3.2.2**. If the families had been delimited on the basis of tree topology alone, without considering taxonomy, the TET6 and TET9 groups could have been merged into a single family. However, given their very different taxonomic distributions (TET6 in Halobacteriales, TET9 in Asgard), we chose to divide them into two distinct families. Biochemical data supported this classification, showing that the HmTET6 and PsTET9 enzymes have radically different substrate specificities (**Figure 40**).

Although additional enzyme characterizations are needed to confirm this trend, especially for the broader groups, enzymes within the same family seem to share similar substrate specificities. For instance, TaTET10 has been identified as a generalist enzyme predominantly targeting hydrophobic residues, which is consistent with the substrate specificity of the previously characterized DaTET10 peptidase (Slutskaya *et al.*, 2012). Similarly, the activity spectrum of the newly characterized MtTET11 aligns with that of MjTET11, despite these enzymes being far apart in the phylogenetic tree. The same holds true for PhTET3 and PfTET3 enzymes, which also display consistent substrate specificities (**Figure 33**). These results highlight the strength of the hybrid approach used in this study, combining phylogeny and biochemistry, to achieve a more accurate delineation of enzyme families.

5. Résumé du chapitre en français

Ce dernier chapitre présente la caractérisation fonctionnelle de six nouvelles peptidases de la famille des TET (*i.e.,* PsTET7, ThTET7, HoTET8, PsTET9, TaTET10 et MtTET11). Des tests d'activités enzymatiques ont été réalisés sur des substrats chromogéniques et fluorogéniques pour déterminer leurs spécificités de clivage, ainsi que leurs conditions optimales en termes de température, de pH et de cofacteurs métalliques.

Concernant la spécificité de substrat de ces enzymes, des spectres d'activités très contrastés ont été observés, mettant en évidence la grande polyvalence fonctionnelle des peptidases TETs. Deux principaux profils d'activité peuvent être distingués. D'une part, PsTET9 et HoTET8 se sont révélées très sélectives, ciblant spécificiquement les acides aminés chargés négativement. D'autre part, MtTET11, PsTET7, TaTET10 et ThTET7 se sont distinguées par leur spectres d'activité plus large, ciblant préférentiellement les résidus hydrophobes. Bien que des spécificités de substrat similaires aient déjà été identifiées chez d'autres peptidases TET caractérisées avant cette étude, des variations subtiles mais significatives dans les préférences de substrats sont observées. Ces résultats soulignent la forte diversité fonctionnelle de ces enzymes.

Les conditions optimales de température, pH et ions métalliques ont également été déterminées. Les activités maximales ont été enregistrées sur une plage de températures allant de 40°C à 90°C, avec des comportements qui varient selon les enzymes ; PsTET7 et PsTET9 peuvent être qualifiées de mésophiles, tandis que les optima de MtTET11, ThTET7, TaTET10 et HoTET8 ont été obtenus pour des températures plus élevées. Concernant le pH, la majorité des enzymes ont montré une activité optimale à des valeurs neutres. MtTET11 et HoTET8 ont démontré une meilleure tolérance aux environnements alcalins, tandis que ThTET7 et PsTET7 ont montré des signes d'instabilité au-dessus de pH 8. Enfin, comme la majorité des peptidases TET caractérisées à ce jour, les six enzymes se sont révélées être activées par le cobalt. Des expériences supplémentaires ont également permis d'étudier l'effet de la concentration en ions métalliques sur l'activation et la spécificité de substrat des TETs.

Finalement, les données biochimiques s'alignent avec ceux obtenus pour les peptidases TET précédemment caractérisées et confirment la délimitation des familles initialement établie sur la base de l'analyse phylogénétique, soulignant ainsi la solidité de l'approche combinée employée dans cette étude.

General discussion

During this thesis, the implementation of structure-based identification criteria for high-throughput screening of M42 peptidases allowed for a comprehensive assessment of their prevalence and diversity in Archaea, shedding light on an unsuspected wealth and revealing that prior studies only captured a narrow glimpse of TETs diversity. Phylogenetic analysis of archaeal TET peptidases allowed to delineate eleven families, of which seven were newly identified in the course of this study (see **chapter 1 section 3.2.2**). Biochemical characterization of representatives from the previously uncharted TET7 to TET11 groups, combined with previous studies, provided a broad overview of the functional diversity of TET peptidases (see **chapter 3**) (**Figure 45**).

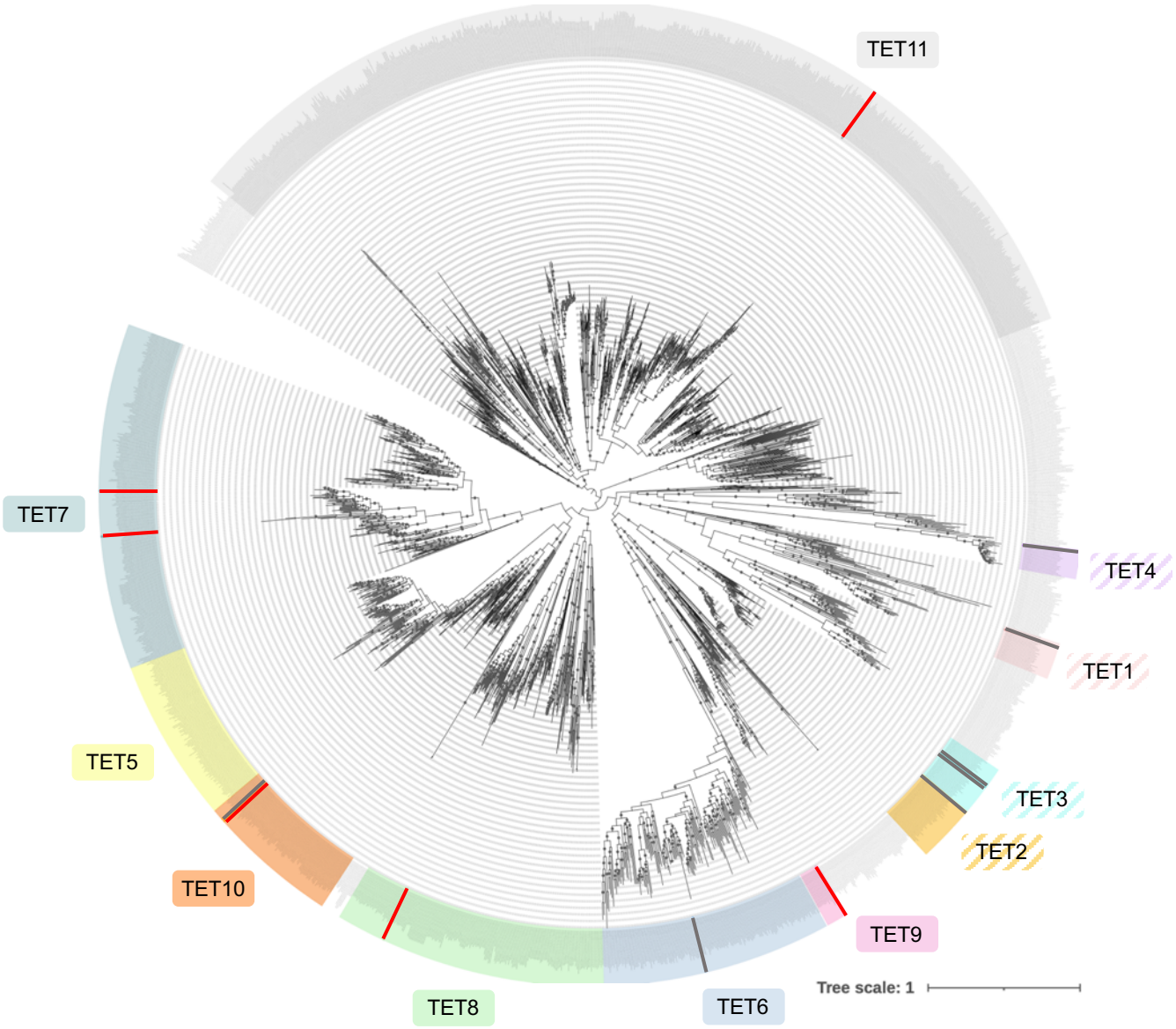


Figure 45: Contribution of this study to the comprehensive understanding of TET peptidases in Archaea. This work allowed to uncover the full diversity of archaeal TET peptidases, and to functionally characterize six new enzymes representative of this diversity. Enzymes characterized before or during this study are indicated in grey and red, respectively.

To gain insight of the significance of the striking diversity observed, a comprehensive examination of the taxonomic distribution of each defined family across the archaeal tree was conducted (**Figure 46**). TET11 emerged as the most prevalent group, distributed across the entire Archaea tree, and consistently present in methanogenic species. TET7 and TET8 groups span two superphyla and are found in species from the Asgard and TACK groups. These sister clades also possess unique TET families, with TET9 and TET10 being found exclusively in Heimdallarchaeia and Crenarchaeota, respectively. The remaining groups exhibit more confined distributions; TET1, TET2, TET3, and TET4 groups appear to be restricted to Thermocci species (with the exception of a few TET1 sequences found in Archaeoglobales), whereas TET5 and TET6 members were exclusively detected in Halobacteria. While being less widespread than the TET11 family, TET-other members are found in species spanning the entire archaeal tree. Apart from a notable absence in the vast majority of methanogenic species, no clear distribution pattern can be identified for this group.

Remarkably, previous studies on archaeal TET peptidases were largely focused on the characterization of TET1, TET2, TET3, and TET4 family members, which turn out to be a unique case restricted to Thermococci species. The TET11 family, which is the most prevalent and spans the entire archaeal tree, was completely overlooked until now.

Significant heterogeneity is observed in the number of TET peptidases per organism. Organisms with a single TET enzyme tend to possess a member of the TET11 group, which likely comprises broad-spectrum activity peptidases (*e.g.,* MtTET11 from *M. thermophilus* and MjTET11 from *Methanocaldococcus jannaschii* are broad-spectrum methionyl-aminopeptidases, see **Figure 40**). Conversely, narrow substrate specificity appears to occur in species with multiple TETs only, such as *P. horikoshii*, which possesses four TET peptidases: PhTET2 functions as a broad-spectrum leucyl-aminopeptidase, while PhTET1, PhTET3, and PhTET4 specifically target acidic, basic, and glycine residues, respectively (Basbous *et al.*, 2018; Durá *et al.*, 2005, 2009; Durá & Franzetti, 2013). Similarly, specialized enzymes belonging to the TET8 and TET9 groups have been identified in the genomes of *Ca. Hodarchaeales archaeon LC_*3 and *Ca. P. syntrophicum* containing two and three putative TET peptidase genes, respectively. Multiplicity is also observed in Halobacteriales, which typically harbor both a TET5 and a TET6. HmTET6, the only characterized enzyme of the TET6 group, displays broad-spectrum activity (Franzetti *et al.*, 2002). Although no enzyme from the TET5 group has been characterized to date, it could be hypothesized that this group exhibits

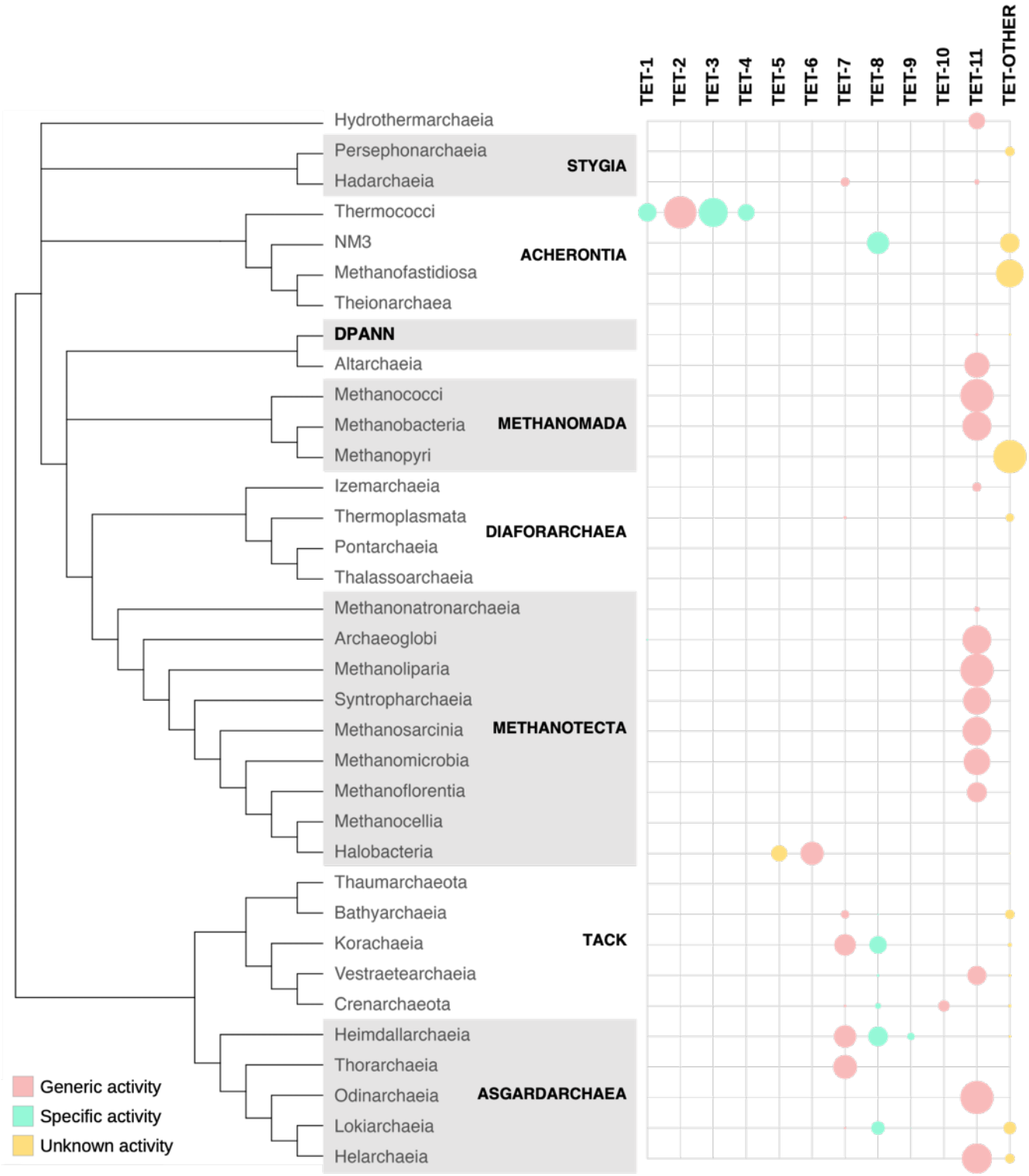


Figure 46: Phylogenetic distribution of the TET families in Archaea. Distribution of the different TET families homologs on a schematic reference phylogeny of Archaea based on Garcia *et al.* (2022). The sizes of the circles vary between 0% and 100% and indicate the percentage of genomes where a family is found. Circles are colored according to the activity spectrum of the characterized representatives of each family: pink for generic activities, green for specific activities, and yellow for undetermined activities.

narrower activity profiles. Accordingly, since TET-other group members are found in species possessing either a single TET or both a TET-other and an additional specialized TET, it can be hypothesized that enzymes of this group function as generalists.

Interestingly, the four TETs of *P. horikoshii* exhibit complementary activity spectra, suggesting that they function in concert to achieve complete peptide hydrolysis (Basbous *et al.*, 2018; Durá *et al.*, 2005, 2009; Durá & Franzetti, 2013). Similarly, PsTET7 and PsTET9 have distinct, non-redundant activities. However, further study of PsTET8, the third TET from *Ca. P. syntrophicum*, would be necessary to determine if a similar complementarity also exists in this species. Unfortunately, we did not manage to purify this enzyme. Synergic specific activities of multiple TET peptidases have also been reported for the bacteria *Geobacillus stearothermophilus* (Roncari & Zuber, 1969) and *Symbiobacterium thermophilum* (Kumaki *et al.*, 2011). In contrast, preliminary studies on the three TETs of *Escherichia coli* revealed redundant broad activity spectra (Dutoit, 2021). Additional characterization of TET peptidases from other organisms would be needed to provide additional insight.

To delve deeper into the origin of the unsuspected diversity of TET peptidases, we extended our analysis to bacterial homologues. We first balanced taxonomic sampling by reducing the number of archaeal genomes (593 Archaea and 401 Bacteria). Using the same methodology as for our initial screening of archaeal genomes (see **chapter 1 section 2.2**), we retrieved 339 archaeal and 187 bacterial homologues. A maximum likelihood tree was inferred using the model LG+R10 (**Figure 47**). The resulting phylogeny revealed a complex evolutionary history, shaped by multiple HGTs intra and inter domains, and several duplications. Two distinct groups can be delineated: the first group primarily contains archaeal sequences, spanning the full diversity of Archaea, with representatives from the TET1, TET4, and TET11 families. Interestingly, sequences belonging to Bacteria, mainly Elusimicrobia, Thermotogae, Firmicutes, and Proteobacteria branch within this group, indicating several independent HGTs between archaea and these bacteria. The second group consists of a mixture of archaeal, mainly from the TACK and Asgard superphyla, and bacterial sequences, encompassing the remaining TET families.

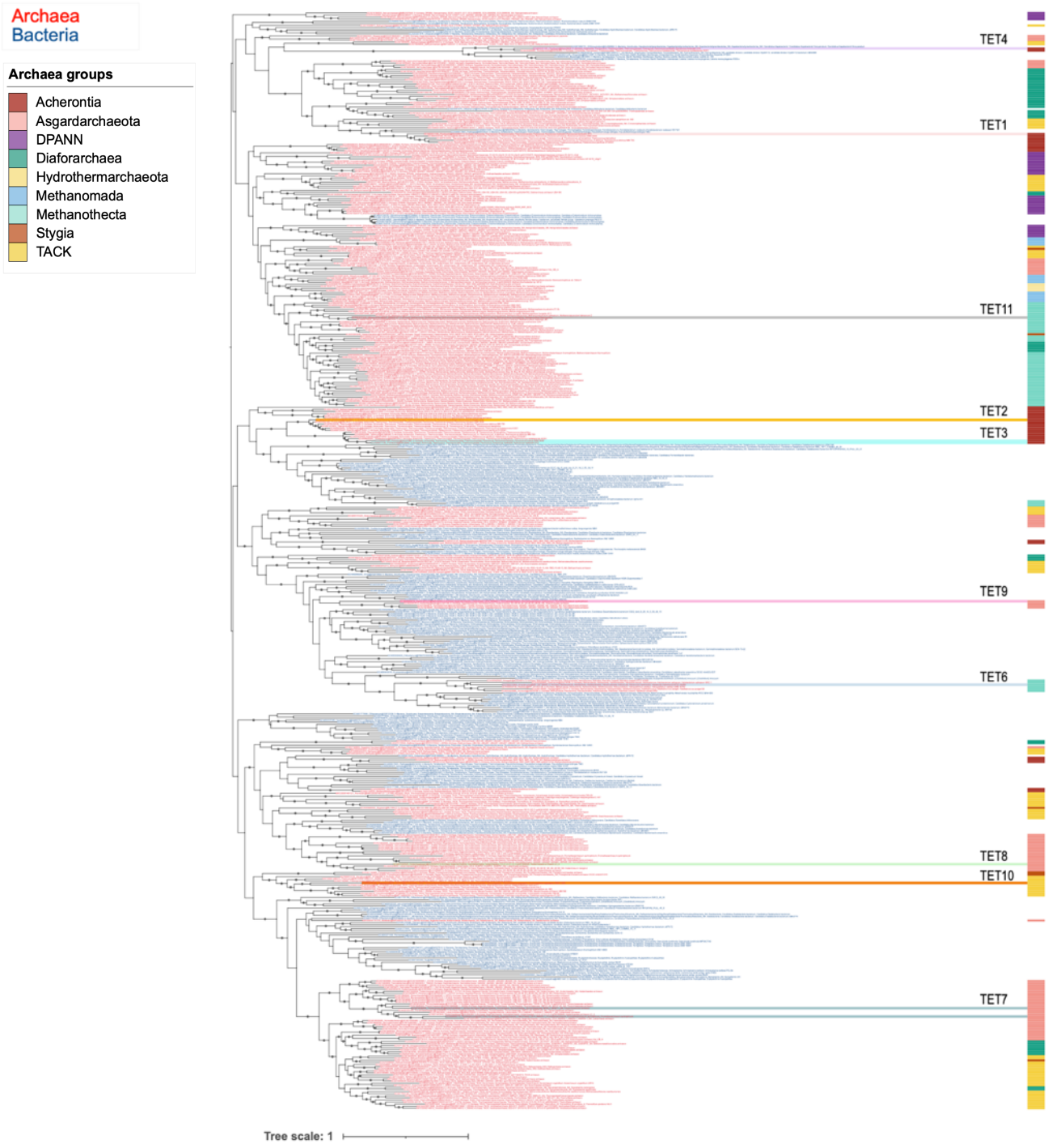


Figure 47: Phylogeny of archaeal and bacterial M42 peptidase homologues. Maximum-likelihood phylogeny obtained from an alignment of 526 sequences and 337 amino acid positions. The scale bar represents the average number of substitutions per site. Circles at the branches indicate ultra-fast bootstrap values ≥ 90 %. Archaeal and bacterial sequences are indicated in red and blue, respectively. Archaeal characterized sequences are highlighted. Archaeal taxonomic distribution is represented on the right.

The phylogenetic analysis indicates that TET multiplicity arose independently multiple times. This phenomenon is not exclusively attributable to HGTs, as evidenced by the emergence of the TET1 and TET4 groups by duplication. It does not appear to stem from environmental adaptation either, as no correlation was identified between TET distribution and specific biotopes. For example, Archaeoglobales, Thermococcales, Methanococcales, and Desulfurococcales have been identified as primary colonizers of deep-sea hydrothermal vents (Huber *et al.*, 2002; Nercessian *et al.*, 2003; Takai *et al.*, 2004). Despite sharing the same ecological niche, these organisms exhibit markedly different TET distribution patterns.

On the other hand, the number and degree of specificity of TET peptidases present in an organism might be correlated with the metabolic capabilities of the species. Indeed, multiple TETs are found in Hadarchaea, Thermococcales, Halobacteriales, Crenarchaeota (*Vulcanisaeta* genus), Heimdallarchaeota, Korarchaeota, and Bathyarchaeota species, for which a heterotrophic or mixotrophic lifestyle involving protein and peptide degradation has been proposed (Baker et al., 2016; Bertoldo & Antranikian, 2006; Bulzu et al., 2019; Cai et al., 2020; Elkins et al., 2008; Feng et al., 2019; Hou et al., 2023; Itoh, 2002; Khomyakova et al., 2023; Lee et al., 2008; Liu et al., 2021; McKay et al., 2019; Oren, 2006; Pillot et al., 2021). It can thus be hypothesized that TETs play a metabolic role in the degradation of environmental peptides used as carbon source. The presence of multiple TETs in these organisms may allow for a more efficient utilization of organic matter in their environment. To investigate this potential function, it could be enlightening to study the expression profiles of TET peptidases in Thermococcales in different metabolic contexts. While these species are heterotrophic, thriving on organic substrates, certain species have also demonstrated the ability to fix CO_2 under hydrogen-rich, sulfur-depleted conditions (Le Guellec *et al.*, 2021). It would be particularly insightful to explore whether TET peptidase expression is differentially regulated in response to these distinct metabolic states.

However, TET peptidases are also found as single copies in autotrophic species such as methanogens, which do not depend on the degradation of exogenous peptides, hinting at an alternative physiological role for these enzymes. As initially proposed by Franzetti *et al.*, TETs may participate in protein homeostasis and amino acid recycling by processing peptides downstream of the proteasome and other related proteolytic complexes (Borissenko & Groll, 2005; Franzetti *et al.*, 2002).

The biological function of TET peptidases remains an open question. In an attempt to identify co-located genes that might be functionally related or co-regulated, the genomic neighborhood was examined; however, no significant genomic conservation or noteworthy syntenic blocks were identified (data not shown). To this day, the preliminary results from the triple knock-out experiment of *E. coli* TET peptidases remain the only genetic studies reported on these enzymes. No growth defects were observed in minimal media or under various stress conditions, suggesting that the genes encoding TET peptidases are not essential in this organism (Dutoit, 2021). Similarly, a transposon sequencing experiment conducted on the archaeon *Methanococcus maripaludis* indicates that the single TET peptidase in this species is also non-essential (Sarmiento *et al.*, 2013).

These findings could be interpreted in two different ways: TET peptidases may not be essential in these species, or there could be functional analogs compensating for their absence. Previous studies suggested complementarity between M42 and M18 or TRI peptidases (Borissenko & Groll, 2005; Dutoit, 2021; Dutoit *et al.*, 2012). To investigate these theories, we used the PFAM domains PF14684 and PF02127 to search for TRI and M18 homologues in our local database of archaeal genomes (**Appendix 6**). Our results challenge these hypotheses; M18 and TRI homologues were only sparsely detected, primarily in species possessing M42 peptidases. Furthermore, several lineages (*e.g.,* Theionarchaea, Pontarchaeia, Thalassoarchaeia, Methanocellia, Thaumarchaeota) lack all three peptidase families. This points to a more complex relationship and suggests the existence of other functional analogs yet to be identified. Notably, although M18 peptidases are typically described in the literature as being restricted to eukaryotes and bacteria, homologs have also been identified in archaea, and were even found to be more abundant than TRI peptidases. Still, M18 peptidases remain relatively rare in these organisms compared to M42 peptidases.

Conclusion and perspectives

In this study, we adopted a hybrid approach, combining structural biology, phylogeny, and biochemistry, aiming for a more global understanding of TET peptidases. We sought to broaden our perspective on these peptidases, leveraging structural insights to delve deeper into their biological significance.

Building on Alexandre Appolaire's pioneering work (2014), the development of structure-based identification criteria allowed for a comprehensive detection of TET peptidases in archaeal genomes, shedding light on an unsuspected diversity. This extensive screening, coupled with phylogenetic analysis, allowed us to outline an eleven-family classification for archaeal TETs, including four already established and seven new families.

Considering that previous studies only scratched the surface of the diversity and taxonomic distribution of TET peptidases in Archaea, we selected thirteen sequences to capture this diversity as extensively as possible. Production and purification of some of these enzymes proved challenging, providing an opportunity to reflect on the difficulties of working with metalloenzymes and to suggest ways to prevent similar issues in future studies.

We successfully produced and purified six peptidases and proceeded with their functional characterization. Heterogeneous activity spectra were observed, underscoring the functional versatility of TET peptidases. These contrasting profiles were categorized into two groups, distinguishing between generalist and specialized enzymes. With more time at our disposal, additional experiments could have expanded the presented functional characterization. In particular, conducting activity assays on longer peptides coupled with HPLC analysis would have been insightful. Although more time-consuming and costly than assays using synthetic substrates, this would have allowed us to verify whether the broadening of PhTET1 substrate specificity observed with increasing peptide chain length also applied to the six newly characterized enzymes.

Importantly, the results of this biochemical study are in agreement with our proposed classification of archaeal TETs, as enzymes within the same family exhibited similar substrate specificities. Nonetheless, additional enzyme characterizations would be required to fully validate this classification. In particular, no representatives of the TET5 and TET-other groups have yet been characterized. Expanding the taxonomic scope to include enzymes from DPANN species, which could not be produced in this study, would also be crucial for confirming the relevance of our classification across the entire archaeal tree.

In light of these new data on TET peptidase activity, and by extending our analysis to bacterial TETs, we gained valuable insights into the evolutionary history of this enzymatic family. The diversity of archaeal TETs likely originated from an ancestral enzyme within the TET11 subgroup, which probably exhibited broad substrate specificity. Over time, multiple HGTs occurred both within Archaea and between Archaea and Bacteria, along with several duplication events, eventually giving rise to the eleven families identified in this study and allowing the emergence of more specialized substrate specificities. To obtain a complete understanding of the evolutionary dynamics of TET peptidases, similar comprehensive studies should be conducted on bacterial and eukaryotic enzymes, including both M18 and M42 peptidase families. Given the expansive nature of such a study and the significant number of enzymes involved, this project would be challenging and highly time-consuming.

Our comprehensive approach enabled us to investigate the biological role of TET peptidases. In addition to their commonly proposed function as ATP-independent peptidases involved in intracellular proteolysis, acting downstream of the proteasome, our findings suggest that these enzymes could also play a metabolic role, participating in the degradation of environmental peptides. Genetic studies would be required to establish the precise physiological role of these enzymes. *Methanocaldococcus jannaschii* might be a suitable candidate for this type of study, as genetic tools are available for this species (Susanti *et al.*, 2019), and it possesses a single TET peptidase belonging to the TET11 family, thought to be the ancestral archaeal family. It would be valuable to conduct genetic studies in species with different trophic types. In addition to the autotroph *M. jannaschii*, it could be particularly interesting to work on the heterotroph *Thermococcus barophilus*, for which the LMEE laboratory developed a genetic tool (Birien *et al.*, 2018; Thiel *et al.*, 2014). While working on *T. barophilus* would be more laborious as it possesses three TETs, it would provide a unique opportunity to investigate the effects of a triple knock-out and various knock-out combinations, offering deeper insights into the functional roles of these enzymes.

To fully leverage the findings of this study, it would be beneficial to combine them with structural studies. In particular, obtaining the structures of PsTET7 and ThTET7 enzymes, which display an insertion newly identified in some Asgard sequences, would help elucidate the role of this insertion. To this end, cryo-EM grids were prepared for both enzymes, and data were collected at 2.3–2.5 Å resolution for the two TETs. However, due to time constraints, the density maps could not be further refined to obtain the final cryo-EM structures.

Such structural studies could also provide valuable insights into what drives the substrate specificity of these enzymes. Now that multiple enzymes with the same substrate specificity have been identified, structural comparisons might reveal specific features critical for substrate recognition and binding. Ultimately, this could allow us to identify sequence signatures enabling the prediction of TET substrate specificities.

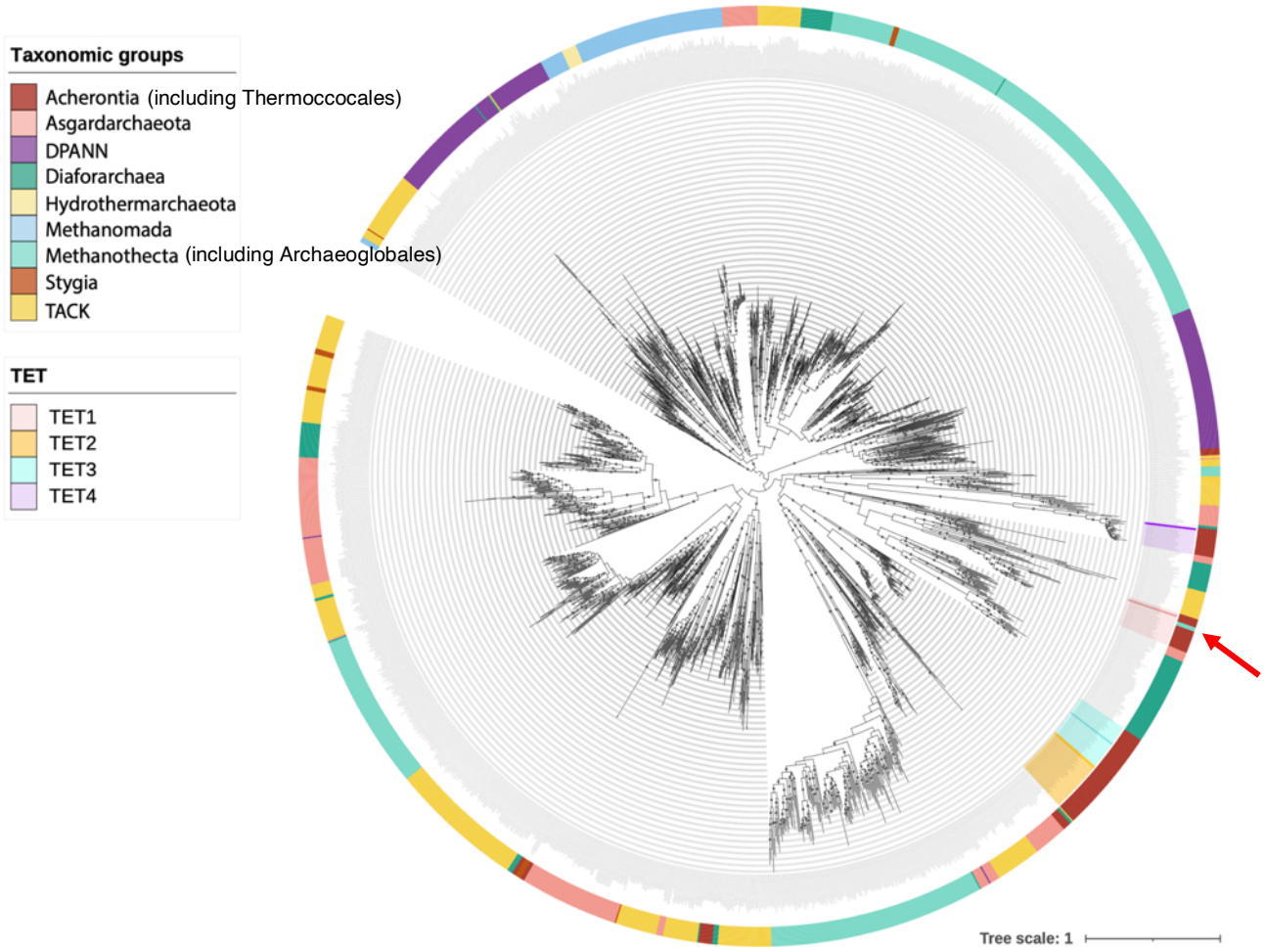
The innovative and comprehensive strategy presented here could significantly expand the potential biotechnological applications of TET peptidases. The ELMA laboratory is already actively working on the industrial valorization of these enzymes. Preliminary studies have demonstrated that the synergistic use of TETs can eliminate toxic peptides, such as gluten, or produce a wide variety of peptides from hydrolyzates. This technology addresses the needs of a rapidly growing market that utilizes enzymatic hydrolysis to produce biostimulants for the nutrition and health sectors. The ELMA laboratory has already filed three European patents covering two specific peptidases and the synergistic action process (EP 3498831 / 3724328 / 3498108). A collaboration has also been initiated with the LEMAR laboratory at IFREMER Brest as part of the NutriZym project. The project aims to produce biostimulants for the aquaculture sector using industrial bio-waste. In this context, processing the protein fraction of a microalgal biomass with TET peptidases could generate a broader diversity of peptides compared to current hydrolysis protocols, conferring biostimulant properties to the hydrolyzate (antiviral, antibacterial, growth-promoting effects, *etc.*).

In this study, we uncovered a vast and previously unsuspected diversity of TET peptidases that remains largely untapped in terms of biotechnological applications. Further exploration of this diversity could significantly expand the range of enzymes available for industrial use.

Overall, this work represents a significant step toward uniting structural and functional insights, offering a more integrated understanding of TET peptidases. The findings of this study will be published in a scientific article, a draft of which is provided in **Appendix 7**.

Appendices

Appendix 1: Phylogeny of archaeal M42 peptidase homologues indicates the occurrence of a HGT from Thermococcales to Archaeoglobales. Maximum-likelihood phylogeny obtained from an alignment of 2,030 sequences and 337 amino acid positions. The scale bar represents the average number of substitutions per site. Circles at the branches indicate ultra-fast bootstrap values ≥ 90 %. Taxonomic distribution is represented on the outer circle. TET2, TET3, and TET4 contain exclusively sequences from Thermococcales, TET1 also includes three sequences from *Geoglobus* (indicated by a red arrow), suggesting a HGT from Thermococcales to Archaeoglobales.



Appendix 2: Sequences of the proteins retained for functional and structural characterization

>HvTET5

MDDDRRAFLEDLLTTPSPSGYEVAGQRVWVDYVSQFADDVTVDDYGNAVAVHEGTGEGP
EIAFTGHADQIGYIVRDIDDDGFVRIGPIGGADRTVSKGQHVTVHGDDGDVAGVIGQTAIHLR
DVGSEEYDDLEEQFVDIGATSKGDAKDHVEVGDPVTVEARVRDLAGDRVAANGMDNRVGT
WSAAEGLRAAVEADVDATVYAVSTVQEEVGVQGAKMVGYDLDPDAMVAVDVTHATDNPD
VPGKRKGPVELGEGPVVSRGSANHPNVVALARDAAGEADIGVQLQAAGIRTGTDADAFYTS
RSGIPSLNIGIPNRYMHTPVEVVSTSDLDDVAALLGSMAALAGDVESFGVEL

>HvTET6

MEFDFDRLKQLTETSGVPGYEDRIRALVREDLEATTDSVRVDGMGNVVGTIEGESDYEVAV
AAHMDEIGFMVRHVNDEGFLQLDALGGWDPRVLKAQRVTVHAEDEDLTGVIGSVPPHTLTE
EQKQKEPKVEDVYVDLGLPAETVEETVSVGDLVTMEQTTVRMGNHVTGKAIDDRVCLFAML
EAAKRIENPEVTIHFAATVQEEVGLRGAQALGVDLDPDLALGLDTTVANDVPGFDPADHVTK
LGAGAGIKLKDSSAIANPKVHRRLRAVAEDDDIPYQMELLPAGGTDTGGFQNSYGAKPVGAI
SMPTRYLHTVTESVHEDDVVAYIDLLTAFLESETGEFDYTL

>HoTET7

MIFLSQNTINTKFLEEICNAFGPSGFETQVQTIIHNYGKQFADEVLFDRLGSVIFKRGSNGPKI
MLAGHSDEIGFIISEIVESGFLKISNIGGWWSQTLLNTQLLIKPFKGEEKIIGIITAKPPHILSSEE
RNKVVELKQMFVDIGCNSKEEVEALGLRIGDPAIPYGSFRTMKRTRKEKKDGKSEEREVNLA
VSKAFDDRIGAFIVTEVLRRLFEENIDHPNTIYSVSSTQEEIGLRGARTSAQMIQPDIGFALDV
DISGDVPGTTGIVQKMGQGVSISAMDGSMIPNPIFRKFVIEVAEEEKIKWQPAFLPAGGTDA
GIIHLTGIGAPSIFIGVPTRHIHSHHSMLDLDDVQQAVNLLIAVIKRLDKEKLKEFITL

>PsTET7

MADDKLDLKLLETLTNAFGPSGHENEVQKITRDYGQKFADDVLYDRTGSVIFKYGKSGPKIM
LAGHADEIGYIISSIGKKGFLKISNIGGIFPTVQLGQEILIRPFKGGEDIIGIIQSAFPGAIKDMKEV
KPLDQLLVDIGCNSEKEVEALGIKVGDPAVPYATYRSIKRKRLIKDDNGKEEEKDVHLVVAKA
FDDRIGVFIALEIIRRLSEEKSNPPNIIYSVSTVQEEIGCRGARTAAQLLQPDIGMSLDVTVSGD
VPGTKNADQKMGDGVVIHAMDNSMMANPKFRKFAIKIAEENGIKWQMGFLNRGGTDAGSI
HLTGAGVPSLFIGIATRYVHSHHSLLDLEDVEGAIQLVIAMLKKLDQKTVESFTTL

>ThTET7

MADMKFDLKLLETLCNAFGPSGHEHEVQKIVQEYGKKYADEVLQDGLGSVIFRKGEGGPKI
MLAGHVDEIGFVITEIKKDGYLAFHQLGGWWDQTLLTQEVVISPYEGGEKIIGVIGAPPPHILP
PEARTKVVTKEKMFIDIGCSSADEVKDLGIRVGDPAVPHAIFRTMKRTRKEKKDDDKEAKEE
TREVTLGVAKAFDDRIGVFIILEALKRISENNISLPNEVYFVSTTQEEVGLRGARTASQKIKPDI
GFALDVDISGDSPGAEGLVQKMGKGVSISAGDGSMIPNPRLRKFVIKLAEEKEIKYQPAFLKA
GGTDAGIIHLAGMGAPSLFLGIPTRYIHSHHGMLDFDDVENAIQLLVEVLQKLDEKTVESFTA
L

>HoTET8

MIFLSQNTINTKFLEEICNAFGPSGFETQVQTIIHNYGKQFADEVLFDRLGSVIFKRGSNGPKI
MLAGHSDEIGFIISEIVESGFLKISNIGGWWSQTLLNTQLLIKPFKGEEKIIGIITAKPPHILSSEE
RNKVVELKQMFVDIGCNSKEEVEALGLRIGDPAIPYGSFRTMKRTRKEKKDGKSEEREVNLA
VSKAFDDRIGAFIVTEVLRRLFEENIDHPNTIYSVSSTQEEIGLRGARTSAQMIQPDIGFALDV
DISGDVPGTTGIVQKMGQGVSISAMDGSMIPNPIFRKFVIEVAEEEKIKWQPAFLPAGGTDA
GIIHLTGIGAPSIFIGVPTRHIHSHHSMLDLDDVQQAVNLLIAVIKRLDKEKLKEFITL

>PsTET8

MSESNLLENVVEIQRKLSTLIGVSGHENDVSAFIFNELKQHADNVWMDKMGNVLGIKKGTDP
DGLRIMIDAHMDEVGLLISYIEKNGFLRFVPLGGIDKRLYPGSDIKIQTKSGKTISGIIGMNPPHI
TDPKLRDISPDHANLFIDIGAKDEEEVKNLGIIVGNRAVLDGRFEYNPDIGGGFMRGRAFDDR
TGCNVALQVSKLINDMEPIPNTILYSFTVCEEVGGRGVPAATDGLNPDIGIALENTIAADVPGV
PLNKQITQLEHGPAFSVADRRTLYHEKLLEIFKMRAEELGYSWQYKQPAFGGTNAGLWHTM
HKGIPSGCISVPSRYIHSPIAMIKISDILATINTLLAILTKPIEI

>PsTET9

MSEKIENIEKKPRIINEKFLEKLVNAPSPTGFEQPAQSVYREYLNDIADKIKTDVMGNVDAVLN
LEGNPRIALMAHVDEVGLQVKYIDEKGFIRFHMLGGVDAHLTPGNRVRILTGSKGKILGVIGK
KAIHLQKPEERKNVVKLDQQFIDIGATSREEAVKELGIEIGDPIIFEHSYAPLGKGDLVVSRCF
DDKIGTFIIAEVIKSLKGTKFEGEVHAVATTQEEIGTRGAITSTYSVNPQVGIAYDVTFATDTPD
MKESDIGIVKMGGGPVIVRGPNINPILFNLFVETAKELGIPYQILATPRATGNDARSIQISRSGV
ATGVIAIPNRYMHSMSEIVNLNDVNMIVELTVAVIQKITKEMSFIPQ

>TaTET10

MSWREELLLLIKELTQLPGPSGYEHRVRDKVVELVKPYADRLWIDAWGNVIAVKKGKTSDR
RLMLAAHMDEIGLFVSHIEDDGFLRAIPIGGVLERTLLYQRVIVVTRTGRMIRGVIGLKPPHVIK
PEEAQKVPELRELFIDIGASSKEEVEKLGVRNGDIVVFDRDVSELSGNRITGKAFDDRVGVA
VLIKALEMMGQPEVDAYFVFTVQEEVGLKGARTSAYGITPDVALAVDVTIASDVPGVAKSEW
FTRLGRGPAIKIVDGRNATGLIAHPKVVDFLVTLAEKHKIPYQLDVVPGGTTDASIIALNKDGV
PAGTVSIPSRYIHSPVEVLDLEDVYNASKLAAAFASEATEEWIKSLKGMVIK

>MtTET11

MVKELLKKLSDAHGVTSREGNIRDIVRAELAGVVDEFREDKMGNLIAIKRGDGFSIMLAAHM
DEIGLMVQYIDEKGFIRVVPLGGWFGPVLYCQRVILHGTKGPVPGVIGAKPPHVMKEEERRK
EIKIEDMFIDVGAASAEEVKNLGIEIGTPITIDREYRELAGTRVTGKALDNRVGVAMLIRALQQA
DSPHTLYAVFTVQEELGLKGAKVSAYSLNPDCAIATDVTIPGDHPGIDKKDASVEMGKGPVL
VLVSASGRGLMADPRMTAWLRETAEKNNIPYQLEVGTGGNTDATIIHLERGGIPSIPFSIPAR
YIHSPVEVVDTADIEAGVRLLVEALKSKPAL

>MfTET11

MKKLIKKLTETPGISGFEEKIRELIKNEIEDYVDEVKYDSLGNLIAVKKGGDKKVMLAAHMDEI
GLMVKHINKKGFVKFSKIGGINDQMLLNQTVKIHGEKTVTGVIGSKPPHRMKPKERKRVISYE
NMFIDIGVKNEKEAKKLISVGDPITFEGNFEMLSDKIFKAKALDNRLGCLVMIEVLKKVDSSATI
YGVGTVQEEVGLKGARTSAYQINPDLAIALDVTIAGDHPGIKKEDAPVELGKGPVIVLTDASG
RGIITHPKVRKLLITTAEKENIEYQLEVSEGGTTDATAIHLTRAGIPSGVLSVPTRYIHTPVSVA
HLDDVKNCIKLLTKALEKYLEVIEN

>AITET11

MNLIEELCRASGVSGFEKDVASVMKKQLQESCEKVEEDSFGNVIARKGKGEKKIMLAAHMD
EIGLMVKHINEKGFISFIKIGGIDDRILLNQRVIIKSREGDVPGIIGSKPPHLQKDDERKKVIKHG
DMFIDIGAKDRKDADKKIAVGDPILFEANYGSLNKKLFYGKAIDNRIGCYVMLKVMEKLPRDIK
STIYAVATAQEEVGLKGARVSAFKLEPDYAFAIDTTIAGDTPQIKENESDLKVGEGPAITITEA
SGRGVVTHPKLRDLLIKTAKKHKIPYQVDVLEGGMTDAAIIYMTRAGVPSGVISIPCRYIHSSS
GVFSIDDVNNSIKLLVNALKEFK

>MkTET

MGTVETLTEHLRELVGRVAPPGWEDEVREYVEATLEKYCDDVHVDTLGNVIGTIEGSEYEV
MVAAHMDEVGFIVKSIDKNGFIRFAKLGGIDDRILPGSRVIIVNSEGEKVPGVVGTKPPHIQEP
KDRRKVPKHKDLFIDIGASDREEAEELVSVGDVGVLAGEFVELVGSRVNGRGLDDKIGVAVL
LALAERLADLDGDHPTFYLVGTVQEEVGLKGAKTSAFEVYPDGAVVLDTAVAGDVPGVKEA
ELKLGKGPAITVVDASGRGLITHPKVRKLLIDTAEELDIPYQLEVGEGGTTDATAIHLTRGGVP
TGVVGIPTRYLHSPAEVLDLEDAKHALELVVEVVQRFPDYVPR

Appendix 3: Purification protocols

ThTET7

After cell lysis, incubation at 70 °C for 15 min and clarification, the resulting supernatant was supplemented with imidazole (final concentration 10 mM) and loaded on a HiTrap Chelating HP 5 mL column (Cityva) equilibrated with 50 mM Tris, 150 mM NaCl, 10 mM imidazole, pH 8.0. Bound proteins were eluted with a linear gradient of imidazole (10 to 500 mM). Fractions corresponding to the elution peak at 400 mM imidazole were pooled, dialysed against 50 mM Tris, 20 mM NaCl, pH 8.0 and loaded on a ResourceQ column (Cytiva) equilibrated with the same buffer. Elution was achieved by a linear NaCl gradient (20 to 500 mM) and fractions containing protein of similar mass (37-39 kDa) according to SDS-PAGE were combined and concentrated using an Amicon Ultra-15 ultrafiltration unit (Millipore) with a 30 kDa cutoff. The protein was utlimately loaded on a Superose 6 Increase 10/300 GL column (Cytiva) in 50 mM Tris, 150 mM NaCl, pH 8.0. Fractions from the elution peak corresponding to a molecular mass around 450 kDa were pooled and subsequently concentrated using an an Amicon Ultra-15 ultrafiltration unit (Millipore) with a 30 kDa cutoff.

HoTET8

After cell lysis, incubation at 70 °C for 15 min and clarification, the resulting supernatant was diluted to a final NaCl concentration of 75 mM and loaded on a ResourceQ column (Cytiva) equilibrated with 50 mM Tris, 75 mM NaCl, pH 8.0. Elution was achieved by a linear NaCl gradient (75 to 300 mM) and fractions containing protein of similar mass (37-39 kDa) according to SDS-PAGE were combined and concentrated using an Amicon Ultra-15 ultrafiltration unit (Millipore) with a 30 kDa cutoff. The protein was then loaded on a Superose 6 Increase 10/300 GL column (Cytiva) in 50 mM Tris, 150 mM NaCl, pH 8.0. Fractions from the elution peak corresponding to a molecular mass around 450 kDa were pooled and subsequently concentrated using an Amicon Ultra-15 ultrafiltration unit (Millipore) with a 30 kDa cutoff.

TaTET10

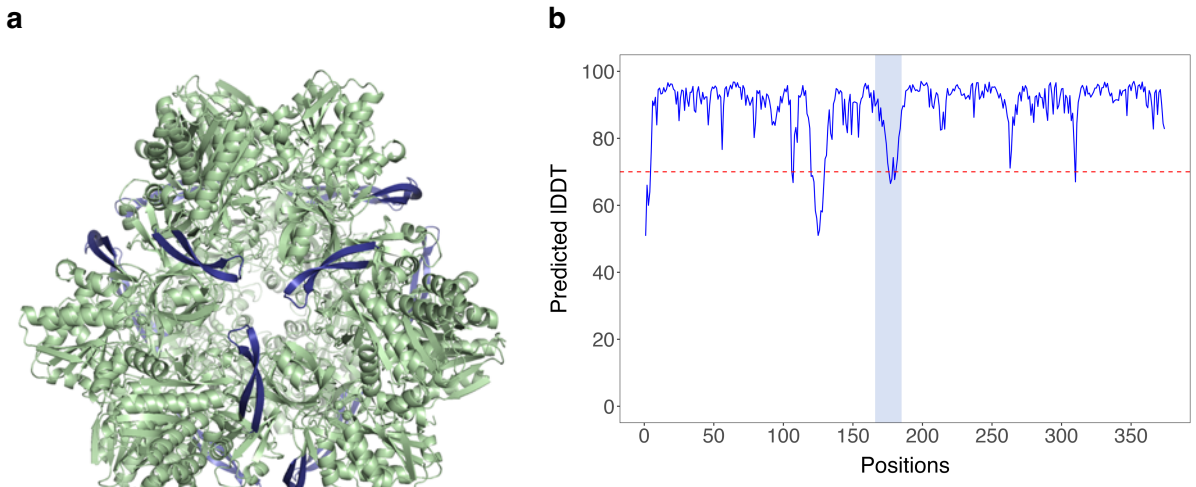
After cell lysis, incubation at 70 °C for 15 min and clarification, the resulting supernatant was dialysed against 50 mM Tris, 50 mM NaCl, pH 8.0 and loaded on a ResourceQ column (Cytiva) equilibrated with the same buffer. Elution was achieved by a linear NaCl gradient (50 mM to 1 M) and fractions containing protein of similar mass (37-39 kDa) according to SDS-PAGE were combined and concentrated using an Amicon Ultra-15 ultrafiltration unit (Millipore) with a 30 kDa cutoff. The protein was then loaded on a Superdex 200 10/300 GL column (Cytiva) in 50

mM Tris, 150 mM NaCl, pH 8.0. Fractions from the elution peak corresponding to a molecular mass around 450 kDa were pooled and subsequently concentrated using an Amicon Ultra-15 ultrafiltration unit with a 30 kDa cutoff.

PsTET7, PsTET9, and MtTET11

After cell lysis and clarification, the resulting supernatant was dialysed against 50 mM Tris, 20 mM NaCl, pH 8.0 and loaded on a DEAE sepharose CL-6B resin (Cytiva, XK16/20 column) equilibrated with the same buffer. Elution was achieved by a linear NaCl gradient (20 to 600 mM) and fractions containing protein of similar mass (37-39 kDa) according to SDS-PAGE were combined, dialysed against 50 mM Tris, 50 mM NaCl, pH 8.0 and loaded on a ResourceQ column (Cytiva) equilibrated with the same buffer. Elution was achieved by a linear NaCl gradient (50 to 500 mM) and fractions containing the protein of interest were pooled and concentrated using an Amicon Ultra-15 ultrafiltration unit (Millipore) with a 30 kDa cutoff. The protein was then loaded on a Superdex 200 10/300 GL column (Cytiva) in 50 mM Tris, 150 mM NaCl, pH 8.0. Fractions from the elution peak corresponding to a molecular mass around 450 kDa were combined and subsequently concentrated using an Amicon Ultra-15 ultrafiltration unit with a 30 kDa cutoff.

Appendix 4: (a) PsTET7 AlphaFold3 model (ipTM score 0.87) featuring the novel insertion identified in some Asgard sequences, here colored in blue. **(b) Associated plDDT plot per residue.** The dashed red line indicates the confidence threshold (plDDT > 70), above which predicted structures are generally considered reliable. The region of the novel insertion (between residues 166 and 185) is highlighted in blue.



Appendix 5: Cd-ninhydrin assay protocol

In my case, activity assays were conducted on whey hydrolyzate as substrate, using the same protocol as described for pNA/AMC substrates (see **chapter 3 section 2**), before revelation with the Cd-ninhydrin reagent.

Preparation of the Cd-ninhydrin reagent:

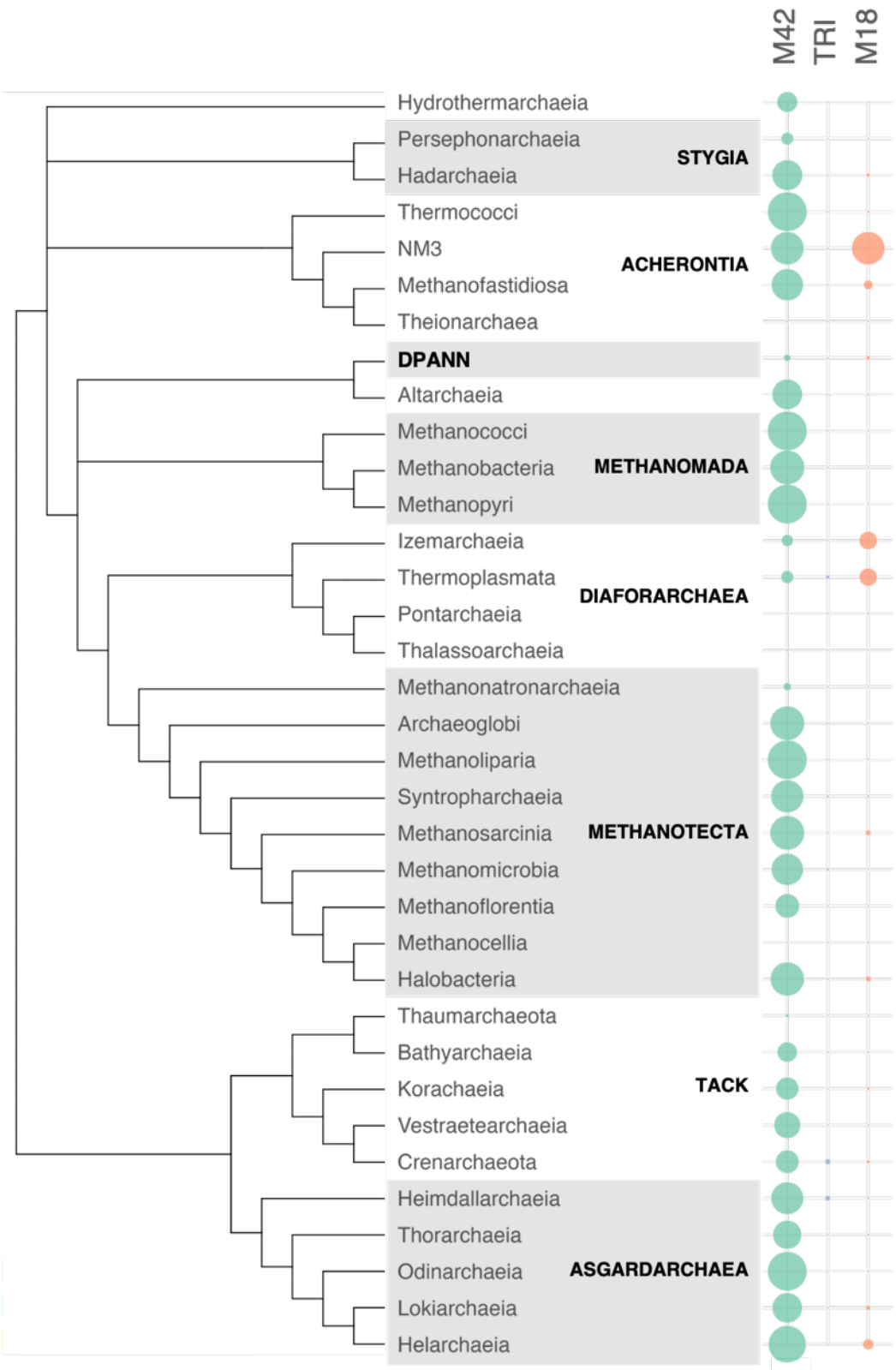
- Dissolve 80 mg of ninhydrin in 8 mL of ethanol.
- Add 1 mL of acetic acid to the solution.
- Separately, dissolve 100 mg of $CdCl_2$ in 0.1 mL of distilled water.
- Slowly add the $CdCl_2$ solution to the ninhydrin solution.
- Mix thoroughly until homogenous.

The prepared Cd-ninhydrin reagent can be stored at room temperature for up to one month.

Sample revelation:

- Mix the Cd-ninhydrin reagent with the sample at a 2:1 ratio (reagent:sample).
- Incubate the reaction mixture at 84°C for 5 minutes.
- Immediately place the tubes on ice to stop the reaction.
- Centrifuge 13,000 rpm for 5 minutes to remove any precipitate.
- Measure the absorbance at 505 nm.

Appendix 6: Phylogenetic distribution of M42, M18, and TRI peptidases in Archaea. Distribution of the M42, M18, and TRI homologs on a schematic reference phylogeny of Archaea based on Garcia *et al.* (2022). The sizes of the circles vary between 0% and 100% and indicate the percentage of genomes where a family is found.



Appendix 7: Scientific article draft

Bioprospecting of environmental archaeal genomes using a structure-based approach reveals novel types of M42 peptidases with contrasted enzymatic properties

Emilie Chagny[1], Najwa Taib[2,3], Daphna Fenel[4], Eric Girard[4], Simonetta Gribaldo[2], Didier Flament[1]*, Bruno Franzetti[4]*

[1] Université de Brest, Ifremer, CNRS, UMR6197 Biologie et Ecologie des Ecosystèmes marins Profonds (BEEP), Plouzané, France

[2] Institut Pasteur, Université Paris Cité, Evolutionary Biology of the Microbial Cell Laboratory, Paris, France

[3] Institut Pasteur, Université Paris Cité, Bioinformatics and Biostatistics Hub, Paris, France

[4] Université Grenoble Alpes, CNRS, CEA, IBS, Grenoble, France

*Corresponding authors: didier.flament@ifremer.fr, bruno.franzetti@ibs.fr

Introduction

In all living cells, proteolysis is a central process serving critical biological functions such as protein quality control, cell-cycle regulation, stress response, virulence, immune response, etc[1–6]. To maintain cellular integrity, proteolytic enzymes must be tightly regulated. Numerous proteases have been identified to self-compartmentalize; the proteolytic activity is confined in an inner chamber, thereby protecting the cellular environment from uncontrolled degradation[7]. While most cytosolic compartmentalized proteases adopt a barrel-shaped architecture, TET peptidases form unique edifices with a distinctive tetrahedral shape.

According to MEROPS classification, TET aminopeptidases are found in the M18 and M42 families. Both families belong to the MH clan, which also encompasses the M20 and M28 families[8]. M18 members are predominantly found in eukaryotes and bacteria, while M42 peptidases are restricted to prokaryotes[9]. The first TET complex was identified in the archaeon *Haloarcula marismortui*[10] and following this initial discovery, several high-resolution structures revealed a conserved architecture, characterized by the assembly of twelve subunits into a hollow tetrahedral complex. Dimers, which are the building block of the dodecamer, are positioned along the edges of the complex. The faces of the tetrahedron are delimited by three dimers creating a central opening (~20 Å diameter), assumed to be the substrate entrance, leading to a wide inner cavity (60 Å diameter). All catalytic sites, situated in the apexes of the complex, are oriented toward this chamber[11–14]. The active site comprises seven catalytic residues, conserved across all enzymes in the MH clan[15], with five of these residues coordinating two metallic ion cofactors[12,16–18].

Although TET peptidases are found in all three domains of Life, archaeal complexes of the M42 family have been the focus of particularly extensive research. Initial studies concentrated on hyperthermophilic species within the order Thermococcales, which possess adapted, robust proteins facilitating structural biology experiments. Functional characterization of the four TETs of *Pyrococcus horikoshii* reavealed for types with distinct substrate specificities: PhTET1, PhTET2, PhTET3, and PhTET4 are glutamyl-, leucyl-, lysyl-, and glycyl-specific aminopeptidases, respectively[13,19–22]. Beyond their utility in structural biology, extremophilic archaeal proteins hold significant potential for biotechnological applications, as they can withstand harsh conditions such as high temperatures or salinity. Three European patents covering two specific TET peptidases and a synergistic action process have already been

filed[23–25]. However, archaea are not limited to extreme environments, and are found in a wide range of habitats, making the study of archaeal proteins valuable for understanding environmental adaptation as well.

Even though archaeal TET peptidases have been extensively structurally characterized, particularly in species of the Thermococcales order, their biological significance remains poorly understood. TETs have been hypothesized to be involved in intracellular proteolysis, acting downstream of the proteasome[10,11]; however, no clear evidence supporting this putative biological function has been presented. Furthermore, the prevalence, ecological and taxonomic distributions of these enzymes are still poorly investigated. Notably, while four homologs were identified in *P. horikoshii,* only one was found in *H. marismortuii,* and little is known about the copy number per organism in other archaeal lineages. With only a few archaeal enzymes thoroughly functionally characterized so far, most of them from closely related species, it is uncertain whether our current knowledge fully accounts for the functional diversity of these enzymes[10,13,16,19–22,26].

In this work, we introduce structure-based identification, enabling high-throughput screening for M42 TET peptidases. By applying these criteria to 4,016 archaeal genomes, the first comprehensive analysis of TET distribution in Archaea was conducted, uncovering an unsuspected diversity. Through phylogenetic analysis, a classification of archaeal TETs in eleven families is proposed. Biochemical characterization of six new enzymes from previously undescribed families and covering a wide taxonomic range of archaeal species offers an extensive overview of the functional diversity of TETs. Finally, by combining biochemical and phylogenetic data, the evolutionary history of these peptidases is addressed, offering new insights into their potential biological roles.

Material and methods

M42 peptidases identification in bacterial and archaeal genomes

To study the taxonomic distribution and the evolution of the M42 peptidase family in Archaea, we assembled a large database containing 4,016 archaeal genomes and 401 bacterial genomes

representatives of all major phyla available in public databases as of January 2022 (**Supplementary Table XX**).

For homology searches, we built a specific HMM profile for the M42 archaeal peptidase family. For this, we used the MEROPS database (v12.0)[8] and retrieved all archaeal M42 peptidase sequences longer than 260 amino acids (195 sequences). We estimated TET4 homologues to be under-represented in this dataset so in parallel, the LD[AE][EL]EKKED pattern, canonical for the TET4 group, was used to search the National Center for Biotechnology Information (NCBI) nr database restricted to Archaea using PHI-BLAST (default parameters)[27]. We retrieved 43 hits and used the T-Coffee trim tool (v11.0.8)[28] to identify the 15 more divergent sequences, which were added to the initial set. Finally, the 210 resulting sequences were aligned with T-Coffee using default parameters, and the alignment was used to build an HMM profile using the HMMBUILD tool from the HMMER suite (v3.3.2)[29].

This profile was used to carry out homology-based searches against our local Archaea database using HMMSEARCH. All hits were retrieved, aligned using MAFFT (v7.481, with the option -auto)[30] and filtered upon the presence of the conserved motifs characterizing M42 peptidases (*i.e.*, residues Gly44, His62, Asp64, Gly77, Gly85, Gly86, Asp173, Gl205, Glu206, Gly211, Asp/Glu228 and His307 according to PhTET1 numbering, PDB code 2WYR). This resulted in 2030 archaeal M42 peptidase homologues (**Supplementary Table XX**). These sequences were aligned using MAFFT (with the option -auto), and the resulting alignment was trimmed using BMGE (v1.12, with the options -m BLOSUM30 -b 1 -w 1 -h 0.95)[31]. Finally, a maximum likelihood phylogeny was inferred using IQ-TREE (v2.0.6)[32] with the model LG+F+R10 selected by ModelFinder according to BIC criteria[33].

To investigate the origin and evolution of archaeal M42 peptidase, we extracted a reduced and taxonomically balanced database covering both archaea and bacteria from our local database (593 Archaea and 401 Bacteria, see **Supplementary Table XX**). Homology searches, alignment, and filtering steps were conducted as described above, yielding in 339 archaeal and 187 bacterial M42 peptidase homologues (**Supplementary Table XX**). The 526 sequences were aligned using MAFFT (with the option -auto), and trimmed using BMGE (with the options -m BLOSUM30 -b 1 -w 1 -h 0.95). A maximum likelihood phylogenetic tree was inferred using

IQ-TREE and the model LG+R10 selected by ModelFinder according to BIC criteria (**Supplementary Fig. 5**). All phylogenies were annotated using IToL[34].

Finally, we used HMMSEARCH (with the option -cut_nc) and the PFAM domains PF14684, PF02127, and PF05299 to search for peptidases tricorn, M18, and M61, respectively, in our local database of Archaea (**Supplementary Table XX**).

Bacterial strains and general information

Escherichia coli DH5α and Rosetta 2(DE3)pLysS chemically competent cells were used for cloning and recombinant expression, respectively. Cells were grown in lysogeny broth (LB) media in a rotary shaker at 37°C (or 20°C when specified), 140 rpm. When used, final concentrations of kanamycin and chloramphenicol were 30 μg/mL and 34 μg/mL, respectively.

For SDS-PAGE analysis, protein samples were mixed with loading buffer (50 mM Tris-HCl, 8 M urea, 2 M thiourea, 75 mM DTT, 3% SDS, 0.05% bromophenol blue, pH 6.8) in a 1:3 ratio, heated to 100°C for 4 min, and loaded on 12% Criterion™ XT Bis-Tris Protein gels (BioRad). Protein bands were visualized by staining with InstantBlue (Expedon). Molecular weights were estimated relative to Precision Plus Protein All Blue Prestained Standards (Biorad).

Expression and purification

The open reading frames of the selected genes were optimized for *E. coli* codon usage and synthesized by Twist Bioscience. For tagged protein expression, synthetic genes were digested with *NdeI* and *BamHI* restriction enzymes and inserted into the pET28a(+) vector, in frame with a thrombin-cleavable N-terminal His6-tag. For untagged protein expression, genes were digested with *NdeI* and *XhoI* restriction enzymes and cloned into the pET41c(+) vector. Cloning accuracy was assessed by Sanger sequencing (Eurofins).

The resulting recombinant plasmids were used for transformation of *E. coli* Rosetta 2(DE3)pLysS cells according to standard procedures[35]. Overnight cultures were diluted 1:100 and grown at 37°C, 140 rpm until OD_{600} reached 0.6. Protein overexpression was induced with 1 mM of isopropyl-β-D-thiogalactopyranoside (IPTG) for 16 h at 20°C. Cells were harvested

by centrifugation at 8,000 × *g* for 45 min at 4°C, and pellets were stored at -80°C. Cells were resuspended in lysis buffer (50 mM Tris-HCl, 150 mM NaCl, 0.1% Triton ×100, pH 8.0) supplemented with 0.05 mg/mL lysozyme, 0.01 mL/mL $MgSO_4$ 2M, 1 mg/mL Pefabloc SC, 0.05 mg/mL DNase, 0.2 mg/mL RNase, and were disrupted on ice in a Vibra-Cell sonifier (35% amplitude with five on/off cycles of 30 s each). For thermostable proteins, the lysate was heated at 70°C for 15 min. Insoluble particles were pelleted by centrifugation (16,000 × *g* for 30 min at 4°C) and the cleared extract was filtered at 0.45 µm and 0.22 µm. The recombinant proteins were purified from the soluble fractions to near homogeneity using various combinations of affinity, anion exchange and gel filtration chromatography (see supplementary information). Ultimately, purified proteins were loaded on a Superose 6 Increase 10/300 GL or Superdex 200 10/300 GL size exclusion column (Cytiva) equilibrated with 50 mM Tris-HCl, 150 mM NaCl, pH 8.0. A high molecular weight calibration kit (Cytiva) was used as a standard to assess the oligomeric state of purified TETs. Protein concentrations were determined following the method of Bradford[36] using bovine serum albumin as the standard.

Negative-stain electron microscopy

4 µL of purified protein samples (0.1 mg/mL) were absorbed onto the clean side of a carbon film on mica, stained, and transferred to a 400-mesh copper grid. Images were taken under low dose conditions (<10 $e^-/Å^2$) with defocus values between 1.2 and 2.5 µm on a Tecnai 12 LaB6 electron microscope at 120 kV accelerating voltage using CCD Camera Gatan Orius 1000.

AlphaFold model predictions

AlphaFold model predictions were calculated using the AlphaFold3 server (https://alphafoldserver.com/ accessed on May 17th, 2024).

Enzymatic characterization general protocol

M42 peptidase hydrolytic activities on synthetic chromogenic and fluorogenic substrates were assayed using aminoacyl-para-nitroaniline (pNA) and aminoacyl-7-amino-4-methylcoumarin (AMC) conjugates ordered from Bachem. Substrates were solubilized in 100% dimethylsulfoxide (DMSO) to a final concentration of 20 mM. All assays described below were carried out according to the following standard procedure[19]. Reactions were initiated by addition of 2 to 10 µg/mL of enzyme to a pre-warmed mixture containing 2.5 mM of the

synthetic substrate in 50 mM buffer (pH 5,5 – 11), 150 mM KCl, and 1 mM XCl_2 (X = Ca, Co, Fe, Mg, Mn, Ni or Zn) in a total volume of 60 µL. To avoid water evaporation, the total volume was covered by 25 µL of mineral oil. Incubations were performed for 3 min to 1 h, reactions were stopped by the addition of 60 µL of 0.1 M acetic acid, and samples were placed on ice. After centrifugation at 6,000 × *g* for 3 min, liberated pNA or AMC quantities were quantified by OD_{405} or fluorescence (excitation and emission wavelengths 360 nm and 460 nm, respectively) measurement in a Synergy HT microplate reader (BioTek). Three replicates and two enzyme blanks were assayed for each experimental point.

Enzyme concentrations and incubation durations were adjusted for each peptidase to produce a robust signal for accurate measurement.

TET	Enzyme concentration	Incubation duration
HoTETb	2 µg/mL	3 min
MtTET	8 µg/mL	10 min
PsTETa	10 µg/mL	60 min
PsTETc	10 µg/mL	20 min
TaTET	10 µg/mL	15 min
ThTET	10 µg/mL	3 min

Effect of temperature, metal cations, and pH on M42 peptidase activities

For each enzyme, optimal temperature, pH and metallic cofactor were determined using the substrate on which maximum activity was measured. The effect of temperature on M42 peptidase activities was evaluated between 20 and 100°C. Assays were conducted as previously described in presence of 50 mM HEPES, 150 mM KCl, and 1 mM $CoCl_2$, pH 7.5. To prevent enzyme denaturation and to ensure stable enzymatic activity**,** optimal pH and metallic cofactor were established 10°C below the determined optimal temperature. The effect of metal cations on M42 peptidase activities was assessed using 1 mM (0.1 mM for TaTET) of XCl_2 metal (X = Ca, Co, Fe, Mg, Mn, Ni or Zn) with 50 mM HEPES, 150 mM KCl, pH 7.5. The influence of pH was studied in presence of 1 mM $CoCl_2$ (0.1 mM for TaTET) using the following buffers: MES (pH 5.5 to 6.5), HEPES (pH 7.0 to 8.0), CHES (pH 8.5 to 9.5), and CAPS (pH 10.0 to 11.0).

Substrate specificity determination

For each peptidase, substrate specificity was determined using optimal metal cofactor and pH. Incubation was performed 10°C the established optimal temperature.

Results

1. Structural determinants for the identification of M42 aminopeptidases

Despite a very strong structural homology, M42 peptidases show low primary sequence conservation (30-40% similarity)[14]. This variation, coupled with frequent misannotations as cellulases or endoglucanases[26,37], makes it challenging to identify M42 peptidases in genomic or proteomic databases, hindering the comprehensive assessment of their distribution. Using the structural elements involved in the formation of their unique tetrahedral architecture as key determinants, we identified markers to better delineate M42 aminopeptidases (**Fig. 1**). First, the presence of the seven conserved catalytic residues (His62, Asp64, Asp173, Glu205, Glu206, Asp/Glu228, and His307 in PhTET1, PDB code 2WYR) is essential[12]. These amino acids coordinate two metal ions that are central to the catalytic activity and structural stability of the complex[16–18]. These residues being shared by all MH clan enzymes, additional criteria are needed for accurate M42 peptidase segregation. Sequence and structure comparison of M18, M20, M28 and M42 enzymes highlighted the unique presence of five glycine residues (Gly44, 77, 85, 86, 211 in PhTET1) in M42 peptidases that confer the required flexibility for proper protein folding in the periphery of the active site. Taken together, these two criteria provide a robust method for screening M42 aminopeptidases in genomic and proteomic databases.

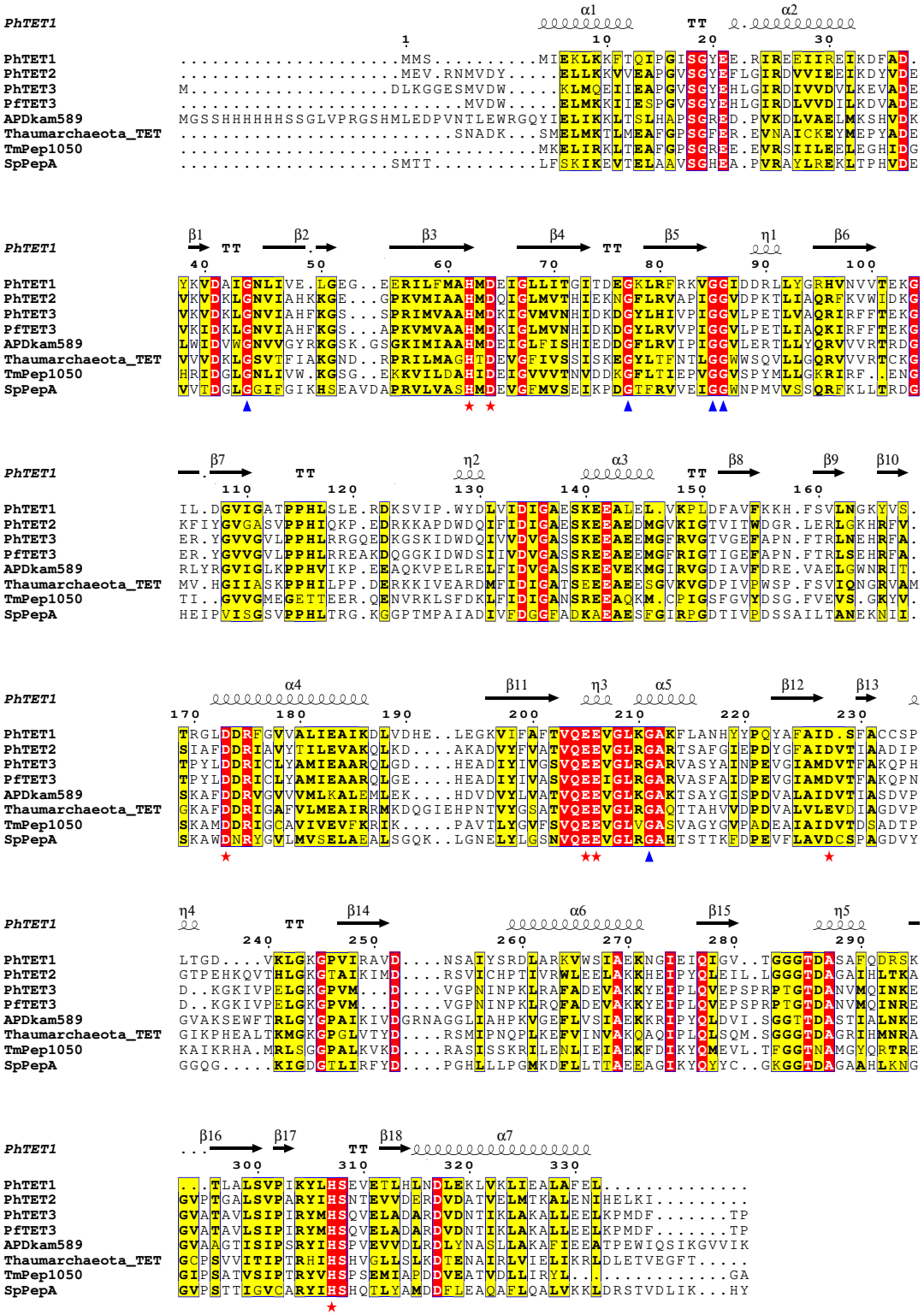


Fig. 1: Structure-based determinants for identifying M42 peptidases. *Multiple sequence alignment of structurally and functionally characterized M42 aminopeptidases. The represented secondary structure corresponds to PhTET1 (PDB code 2WYR). Two criteria were retained for the identification of TET aminopeptidases: conserved catalytic residues H62, D64, D173, E205, E206, [DE]228, H307 (red stars), and conserved glycine residues 44, 77, 85, 86, and 211 (blue triangles). Residue numbering according to PhTET1 sequence.*

2. Beyond the tip of the iceberg: unveiling the hidden diversity of TET peptidases in archaea

This study focuses on M42 peptidases, often designated as TET peptidases by a stretch of terminology. To ensure clarity and consistency with previous studies, the term 'TET peptidases' will be used exclusively for the remainder of this article.

An HMM profile was built using 210 archaeal TET sequences gathered from the MEROPS and NCBI nr databases. This profile was used to conduct an exhaustive homology search of TET peptidases against a large database covering all currently available diversity, including uncultured phyla, and containing 4,016 genomes of Archaea. All positive hits were retrieved and filtered upon the presence of the conserved catalytic site and glycine residues mentioned above, resulting in 2,030 archaeal TET homologues (**Supplementary Table XX**). TET peptidases were found in 39.7% of archaeal genomes (1,595) with varying numbers of TET copies per species, ranging from one to five per genome. While TET homologues are widely present in Euryarchaeota and Asgard, their distribution is uneven to sporadic in the TACK and the DPANN superphyla, respectively (Supp table XXX). Intriguingly, some TET peptidase sequences from the Asgard display a specific pattern consisting in an insertion of approximately 20 residues, which has never been detected in any TET peptidase described to date. This peculiar insertion, located just after the dimerization interface, was found in 72 Asgard species only and exhibits a strong charge contrast. A similar but shorter insertion (9-10 residues) was also observed in 24 Thermoplasmata species (**Supplementary Fig. 1**).

To understand how archaeal TET peptidases are related to each other, we inferred a maximum likelihood tree using the 2,030 archaeal M42 sequences (**Fig. 2**). The TET1, TET2, TET3, and TET4 groups were already identified prior to this study, with several characterized representatives from species such as *Pyrococcus horikoshii*, *Pyrococcus furiosus* and *Thermococcus onnurineus* species[13,16,19–22,38,39]. In TET phylogeny, these groups form four distinct and supported monophyletic clades (with UFB values of 98%, 100%, 100%, and 100% respectively). TET2 and TET3 are sister clades, indicating a recent duplication event. Interestingly, prior studies showed the *in vitro* and *in vivo* formation of PhTET2-PhTET3 heterocomplexes in *P. horikoshii*[17,40]. While TET2, TET3, and TET4 contain exclusively sequences from Thermococcales, TET1 also includes three sequences from *Geoglobus*, indicating a horizontal gene transfer (HGT) from Thermococcales to Archaeoglobales (**Fig. 2**).

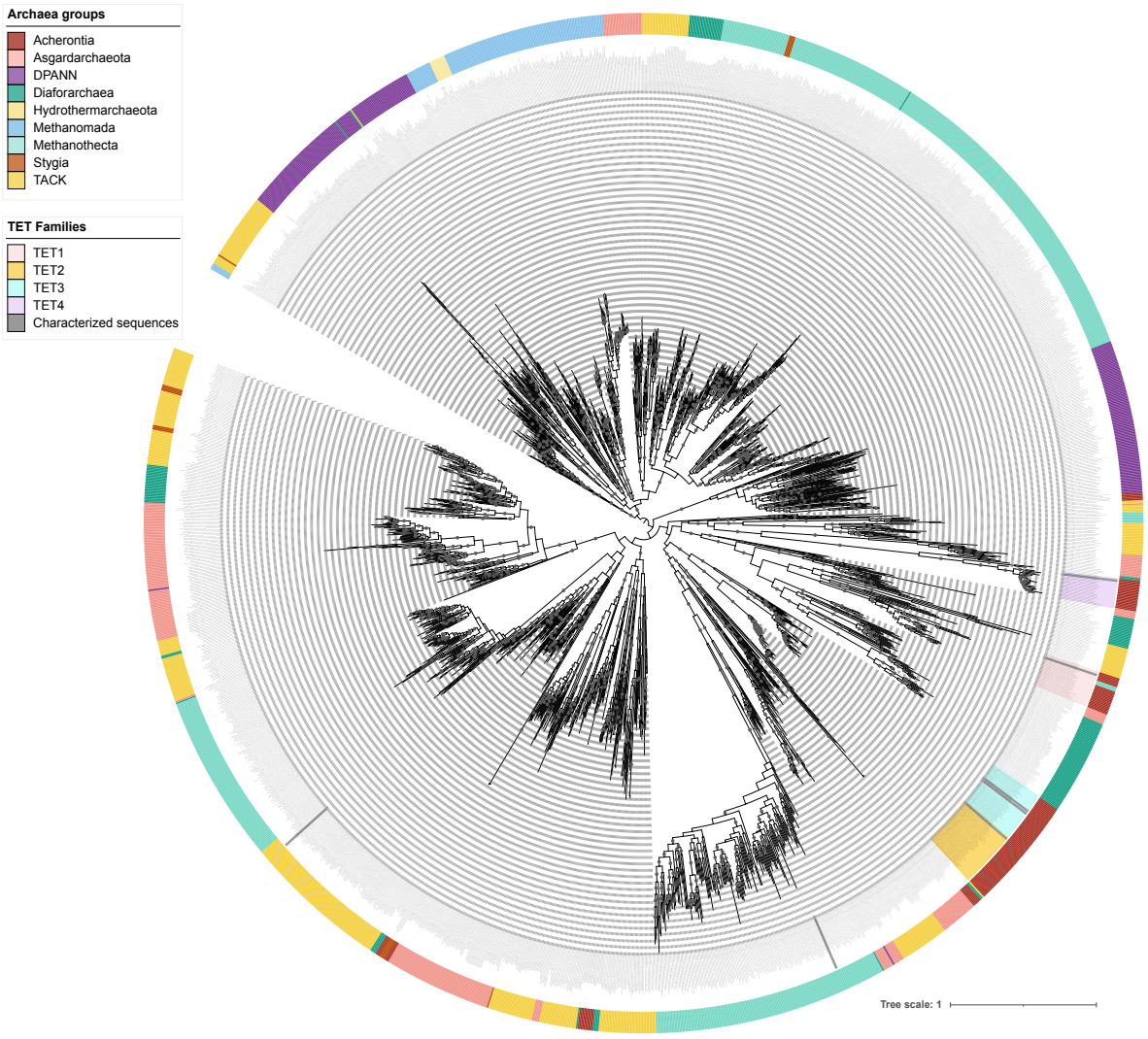


Fig. 2: Phylogeny of archaeal TET peptidase homologues. *Maximum-likelihood phylogeny obtained from an alignment of 2,030 sequences and 337 amino acid positions. The scale bar represents the average number of substitutions per site. Circles at the branches indicate ultra-fast bootstrap values >= 90 %. TET1 to TET4 families were delineated based on the taxonomic distribution and the topology of the tree. Gray bars on the inner circle indicate enzymes characterized prior to this study. Archaeal taxonomic groups are represented on the outer circle.*

Prior studies on the four enzymes of *P. horikoshii* reavealed distinct substrate specificities, already unveiling the functional versatility of TET peptidases: PhTET2 functions as a broad-spectrum leucyl-aminopeptidase[19], while PhTET1, PhTET3, and PhTET4 specifically target acidic, basic, and glycine residues, respectively[13,20,21]. However, a striking observation from the present analysis is that previous characterization of a few archaeal TET peptidases, primarily from TET1 to TET4 groups, remained marginal with regard to the real taxonomic distribution and overall diversity of TET peptidases in archaeal genomes[10,13,16,19–21]. These findings

highlight the need for broader protein characterization to fully explore their functional diversity in Archaea. To this aim, we selected thirteen sequences from nine different species (1 TACK, 1 DPANN, 3 Asgardarchaeota, 2 Methanotecta, and 2 Methanomada) for recombinant protein expression and purification (**Table 1**, sequences provided in **Supplementary Table XX**). Notably, three Asgard sequences featuring the newly identified insertion were selected to investigate the role of this extension.

***Table 1: Candidate proteins for further characterization**. Thirteen proteins spanning the phylogenetic tree and representing a broad range of taxonomically diverse species were selected.*

TET	Species	Taxonomy
HvTETa	*Haloferax volcanii*	Methanotecta; Halobacteria; Halobacteriales; Haloferacaceae; Haloferax;
HvTETb		
HoTETa	*Ca. Hodarchaeales archaeon LC_3*	Asgardarchaeota; Heimdallarchaeia; Hodarchaeales; LC-3; LC-3;
HoTETb		
PsTETa	*Ca. Prometheoarchaeum syntrophicum*	Asgardarchaeota; Lokiarchaeia; CR-4; AMARA-1; Prometheoarchaeum;
PsTETb		
PsTETc		
ThTET	*Ca. Thorarchaeota archaeon MP8T-1*	Asgardarchaeota; Thorarchaeia; Thorarchaeales; Thorarchaeaceae; MP8T-1;
TaTET	*Thermosphaera aggregans*	TACK; Thermoproteia/Crenarchaeota; Sulfolobales; Desulfurococcaceae; Thermosphaera;
MfTET	*Methanothermus fervidus*	Methanomada; Methanobacteria; Methanobacteriales; Methanothermaceae; Methanothermus;
MtTET	*Methanoculleus thermophilus*	Methanotecta; Methanomicrobia; Methanomicrobiales; Methanoculleaceae; Methanoculleus;
MkTET	*Methanopyrus kandleri*	Methanomada; Methanopyri; Methanopyrales; Methanopyraceae; Methanopyrus;
AlTET	*Altarchaeia archaeon ex4484_2*	DPANN; Altarchaeia; IMC4; QMZM01; EX4484-2;

3. Archaeal TET peptidases exhibit contrasting substrate specificities

Of the thirteen initially targeted proteins, six were successfully produced and purified to near homogeneity from *E. coli* extracts using a combination of affinity, anion exchange, and gel filtration chromatography (*i.e.,* HoTETb, PsTETa, PsTETc, ThTET, TaTET, and MtTET). During the final gel filtration chromatography step, all proteins eluted as well-separated high molecular mass complexes corresponding to particles of a molecular mass of *c.* 450 kDa

(**Supplementary Fig. 2a**). SDS-PAGE analysis of the elution peaks consistently showed a major band around 40 kDa, indicating the formation of homo-dodecameric complexes (data not shown). These findings were further validated by negative-stain electron microscopy observations, which revealed homogeneous populations of tetrahedral particles for most enzymes (**Supplementary Fig. 2b**). Lower purity of PsTETc and TaTET samples precluding satisfactory negative-staining imaging (data not shown), structure predictions were generated using AlphaFold3 (**Supplementary Fig. 2c**). The resulting models displayed the expected hollow tetrahedral edifices with high confidence scores (ipTM 0.88 and 0.91), further supporting PsTETc and TaTET ability to form high molecular weight assemblies. Alignment of the AlphaFold models with the crystallographic structure of the dodecameric PhTET2 (PDB code 1Y0R) yielded RMSD values of 0.783 Å for PsTETc and 0.534 Å for TaTET, indicating a robust superposition.

Previous functional characterization of various Thermococcales TETs revealed wide disparities in terms of substrate specificity. However, structural analyses of these different enzymatic edifices, based on primary sequence signatures or well-defined structural properties, fail to predict these different specificities[13,19–22]. To investigate whether the diversity here observed through phylogenetic analysis reflects a functional diversity, cleavage specificities were thus studied using chromogenic [para-nitroaniline (pNA) conjugated] and fluorogenic [7-amino-4-methylcoumarin (AMC) conjugated] aminoacyl substrates. The observed activity spectra were heterogenous and can be divided into two main groups. The first group includes PsTETa, ThTET, MtTET, and TaTET, which exhibited broad-spectrum activities and can be classified as generalist enzymes. These enzymes were found to preferentially cleave hydrophobic residues, with PsTETa and ThTET displaying broader specificities. Optimal amidolytic activities were observed with Ile-pNA, Leu-pNA, Met-pNA, and Met-pNA, respectively. Similarities with the cleavage profile of PhTET2[19] can be outlined, but this study is the first description of methionyl and isoleucyl aminopeptidases in the TET family (**Fig. 3**). Conversely, PsTETc and HoTETb exhibited more selective activities and can be classified as specialized enzymes. Analogous to the previously described PhTET1 peptidase, they specifically targeted acidic amino acids[20]. Interestingly, PsTETc maximum activity was measured on Glu-pNA, whereas no hydrolysis could be detected on Asp-pNA despite the similarity of these substrates (**Fig. 3**). The same substrate specificity has already been reported for the MHJ_0125 glutamyl-aminopeptidase of *Mycoplasma hyopneumiae*[41].

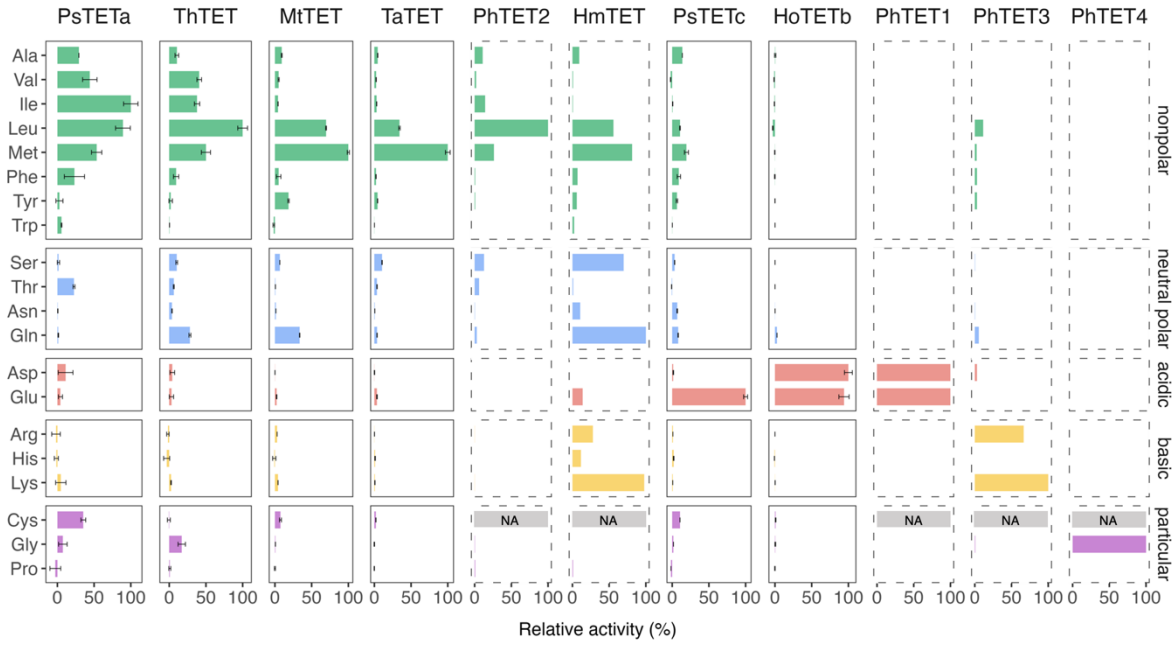


Fig. 3: Characterized TET aminopeptidases exhibit diverse substrate specificities. *Cleavage specificities were assayed using synthetic chromogenic and fluorogenic substrates. For each enzyme, activities are expressed as percentage of the maximum activity observed, which was attributed a value of 100%. Enzymes characterized prior to this study are indicated by dashed lines. Error bars indicate ±s.d. with n=3. NA: not assessed.*

4. Effects of temperature, pH, and metal ions on TET aminopeptidase activities

TET enzymatic behaviors were investigated across a temperature range of 20°C to 100°C and at pH values from 5.5 to 11.0 (**Fig. 4**). The studied peptidases exhibited markedly contrasting profiles, with maximum activities measured for temperatures ranging from 40°C to 90°C. Specifically, PsTETa and PsTETc showed peak activity within the mesophilic range, while MtTET, ThTET, TaTET, and HoTETb exhibited a preference for higher temperatures, with optimal temperatures for enzymatic activities of 60°C, 70°C, 90°C, and 90°C, respectively. For all assayed enzymes, maximum amidolytic activity was measured at neutral pH values. Notably, MtTET and HoTETb demonstrated higher resilience to alkaline environments, whereas aggregation of ThTET and PsTETa enzymes was observed above pH 8.0, revealing their instability under these conditions. TaTET showed increased tolerance to acidic conditions, maintaining 87% of its optimal activity at pH 5.5.

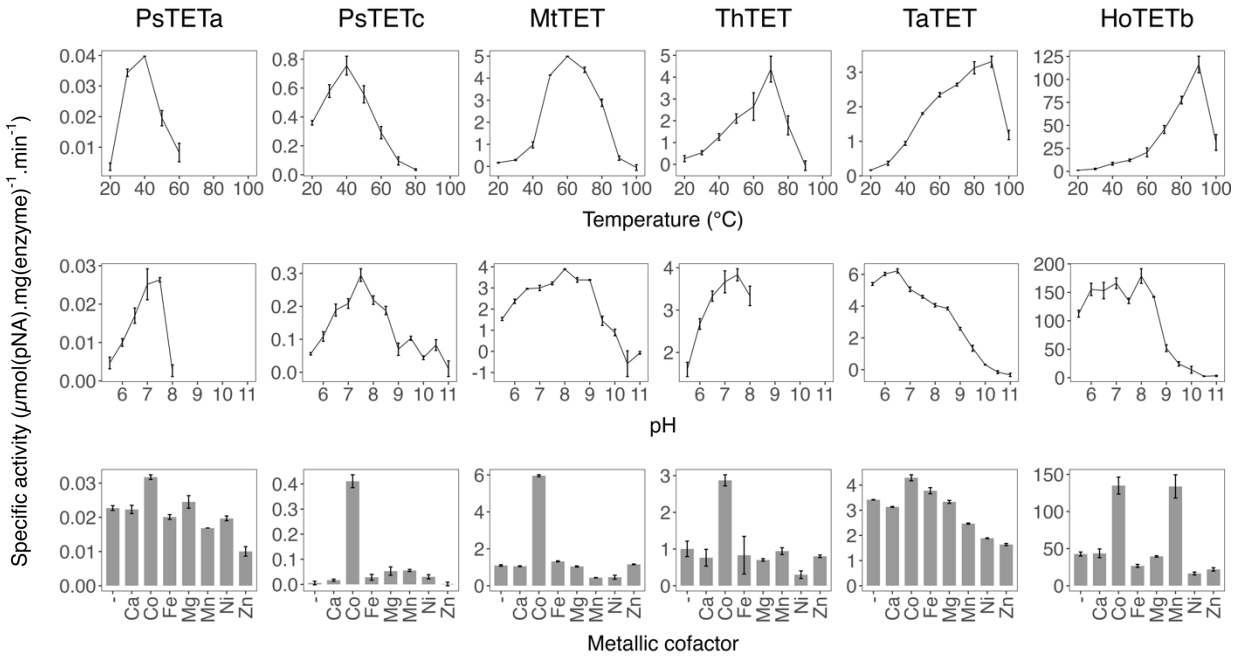


Fig. 4: Temperature, pH and metal ions influence on TET enzymatic activities. *For each enzyme, optimal conditions regarding temperature (top), pH (middle) and divalent cations (bottom) were determined. Error bars indicate ±s.d. with n=3.*

The effects of several divalent cations on the enzymatic activities were also examined (**Fig. 4**). Consistent with prior characterizations of archaeal and bacterial TET peptidases[13,16,19,20,42], maximum stimulatory effect was observed with Co^{2+} ions for all enzymes. However, TaTET was inhibited by all assayed metallic ions at a 1 mM concentration, and stimulatory effect of Co^{2+} was restored at a reduced concentration of 0.1 mM (data not shown). These findings are in agreement with previous reports of concentration-dependent activation of TET peptidases by metallic ions[16]. Interestingly, while the APDkam589 peptidase from *Desulfurococcus kamchatkensis* is known to be activated by Mg^{2+} and Mn^{2+} ions[26], HoTETb stands out as the first characterized enzyme in this family to exhibit equal activation by Co^{2+} and Mn^{2+} ions.

Temperature, pH and metal cofactor optima determined for MtTET, TaTET, PsTETa and PsTETc are consistent with the optimal growth conditions of *Methanoculleus thermophilus*[43,44], *Thermosphaera aggregans*[45], and *Candidatus Prometheoarchaeum syntrophicum*[46]. The physiologies of the uncultivated species *Candidatus Hodarchaeales archaeon LC_3* and *Candidatus Thorarchaeota archaeon MP8T_1* remain unknown, and data concerning the

physico-chemical conditions at their sampling sites are lacking. Nevertheless, the optimal growth temperatures of Asgard species have been estimated in two parallel studies using either genome-derived features[47] or the optimal GDP-binding temperature of the EF-1α translation elongation factor[48]. Interestingly, the high optimal temperatures obtained for ThTET and HoTETb (70°C and 90°C, respectively) are in disagreement with the predicted mesophilic lifestyles of Thorarchaeales and Hodarchaeales (**Fig. 4**).

5. Archaeal TET peptidases can be delineated into several families

The phylogenetic analysis presented in this study uncovered a previously underappreciated wealth of TET peptidases in archaea (**Fig. 2**). Considering that the pre-existing TET1 to TET4 families covered only a small fraction of this diversity, new families were defined using the topology of the phylogenetic tree, *i.e.,* branch length, branch supports, monophyly, and taxonomic distribution. The delineation of these new families was further refined according to the distinct biochemical properties of characterized enzymes (**Fig. 3**). Specifically, beyond their distinct taxonomic distribution, contrasted substrate specificities led us to separate the TET6 and TET9 groups. Conversely, the broad TET11 family was maintained as a single family due to the similarities between the substrate specificities of the enzymes of *Methanocaldococcus jannaschii* (Atalah *et al.,* in preparation) and *Methanoculleus* thermophilus, along with consistent taxonomic distribution. Additionally, TaTET has been identified as a generalist enzyme predominantly targeting hydrophobic residues, which is consistent with the substrate specificity of the previously characterized APDkam589 peptidase belonging to the same family[26]. Consequently, seven new families were described across the tree, which were named TET5 to TET11 in accordance with the existing nomenclature (**Supplementary Fig. 3**). Notably, all sequences featuring the novel insertion described above were found in the TET7 group. Collectively, these eleven families account for 1,619 sequences, the remaining 411 sequences were not affiliated to any family due to poorly supported branching or lack of characterized representatives.

To gain insight into the significance of TET peptidase diversity, a comprehensive examination of the taxonomic distribution of each identified group across the tree of Archaea was conducted (**Fig. 5**).

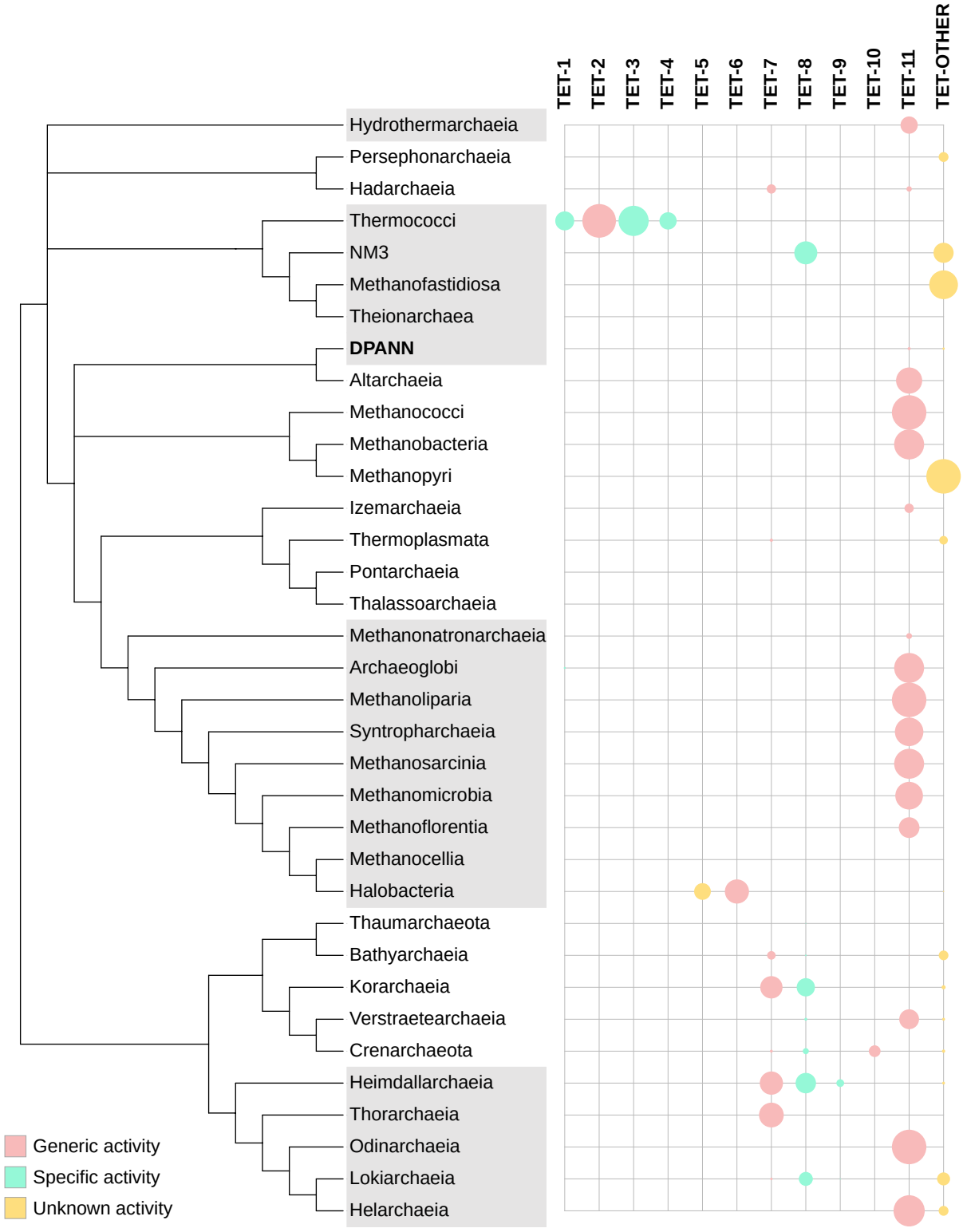


Fig. 5: Phylogenetic distribution of the TET families in archaea. *Distribution of the different TET families homologs on a schematic reference phylogeny of Archaea based on Garcia et al[49]. The sizes of the circles vary between 0% and 100% and indicate the percentage of genomes where a family is found. Circles are colored according to the activity spectrum of the characterized representatives of each family: pink for generic activities, green for specific activities, and yellow for undetermined activities.*

TET11 emerged as the most prevalent group, distributed across the entire Archaea tree, and consistently present in methanogenic species. TET7 and TET8 groups span two superphyla and are found in species from the Asgard and TACK groups. These sister clades also possess unique TET families, with TET9 and TET10 being found exclusively in Heimdallarchaeia and Crenarchaeota, respectively. The remaining groups exhibit more confined distributions; TET1, TET2, TET3, and TET4 groups appear to be restricted to Thermocci species (with the exception of a few TET1 sequences found in Archaeoglobales), whereas TET5 and TET6 members were exclusively detected in Halobacteria. Sequences that were not affiliated to any family were regrouped in the TET-other category. While being less widespread than the TET11 family, enzymes of this group are found in species spanning the entire Archaea tree. Apart from a notable absence in the vast majority of methanogenic species, no clear distribution pattern can be identified for this group.

Discussion

In this study, the implementation of structure-based identification criteria for high-throughput screening of M42 peptidases allowed for a comprehensive assessment of their prevalence and diversity in archaea, shedding light on an unsuspected wealth. Phylogenetic analysis and characterization of archaeal TET peptidases identified eleven families, seven of which had never been investigated before (**Supplementary Fig. 3**). Strikingly, previous studies on archaeal TET peptidases have focused on the extensive characterization of TET1, TET2, TET3, and TET4 families, which are now understood to be a unique case restricted to Thermococci species. The TET11 family, which is the most prevalent, was completely overlooked until now (**Fig. 5**).

This approach also revealed a previously unreported ~20-residue insertion in some Asgard TET peptidases, located immediately after the dimerization domain essential for TET particle assembly (**Fig. 1**). This insertion does not appear to disrupt oligomerization, since both PsTETa and ThTET still form typical hollow tetrahedral particles (**Supplementary Fig. 2a and b**). AlphaFold3 structure predictions suggest that these insertions would form protruding two-stranded β-sheets that may partially obstruct the pores on the faces of the tetrahedral particle, which are believed to be the entry points for substrates[22] (**Supplementary Fig. 4**). Considering that PsTETa and ThTET exhibit a similar but broader substrate specificity compared to

PhTET2, MtTET and TaTET, this insertion may be involved in substrate recognition and recruitment. Alternatively, this insertion located on the surface of the TET particle could be mediating interaction with a partner protein. To fully elucidate the role of this novel insertion, further functional and structural studies on representative members of this family should be prioritized.

Significant heterogeneity is observed in the number of TET peptidases per organism. Organisms with a single TET enzyme tend to possess a member of the TET11 group, which likely comprises broad-spectrum activity peptidases (*e.g.,* MtTET from *M. thermophilus* is a broad-spectrum methionyl-aminopeptidase, **Fig. 3**). Conversely, narrow substrate specificity appears to occur in species with multiple TETs only, such as *P. horikoshii*, which possesses four TET peptidases: PhTET2 functions as a broad-spectrum leucyl-aminopeptidase[19], while PhTET1, PhTET3, and PhTET4 specifically target acidic, basic, and glycine residues, respectively[13,20,21]. Similarly, specialized enzymes belonging to the TET8 and TET9 groups have been identified in the genomes of *Ca. Hodarchaeales archaeon LC_3* and *Ca. P. syntrophicum* containing two and three putative TET peptidase genes, respectively. Multiplicity is also observed in Halobacteriales, which typically harbor both a TET5 and a TET6. HmTET6, the only characterized enzyme of the TET6 group, displays broad-spectrum activity. Although no enzyme from the TET5 group has been characterized to date, it could be hypothesized that this group exhibits narrower activity profiles. Accordingly, since TET-other group members are found in species possessing either a single TET or both a TET-other and an additional specialized TET, it can be hypothesized that enzymes of this group function as generalists.

Interestingly, the four TETs of *P. horikoshii* exhibit complementary activity spectra, suggesting that they function in concert to achieve complete peptide hydrolysis[13,19–21]. Similarly, PsTETa and PsTETc have distinct, non-redundant activities. However, further study of the third TET from *Ca. P. syntrophicum* would be necessary to determine if a similar complementarity also exists in this species. Synergic specific activities of multiple TET peptidases have also been reported for the bacteria *Geobacillus stearothermophilus*[50] and *Symbiobacterium thermophilum*[51]. In contrast, preliminary studies on the three TETs of *Escherichia coli* revealed redundant broad activity spectra[52]. Additional characterization of TET peptidases from other organisms would be needed to provide additional insight. Still, current evidence suggests that the presence of multiple TETs does not necessarily imply complementary substrate specificities.

To investigate the origin of the unsuspected diversity of TET peptidases, bacterial homologues were included in the phylogenetic analysis. We retrieved 339 archaeal and 187 bacterial homologues. In Bacteria, TET homologues were identified in 30% of genomes (121) and displayed a patchy distribution across phyla such as Thermotogae, Proteobacteria, Firmicutes, Chloroflexi, Deinococcota, Atribacteria, Bipolaricaulota, and Verrumicrobia. We inferred a maximum likelihood tree (**Supplementary Table XX, Supplementary Fig. 5**), which revealed a complex evolutionary history, shaped by multiple HGT intra and inter domains and several duplications. Two distinct groups can be delineated: the first group primarily contains archaeal sequences, spanning the full diversity of Archaea, with representatives from the TET1, TET4, and TET11 families. Interestingly, sequences belonging to Bacteria, mainly Elusimicrobia, Thermotogae, Firmicutes, and Proteobacteria branch within this group indicating several independent HGT between archaea and these bacteria. The second group consists of a mixture of archaeal, mainly from the TACK and Asgard superphyla, and bacterial sequences, encompassing the remaining TET families.

While the majority of the TET groups are restricted to few taxonomic groups (*e.g.*, TET1-4 in Thermococci, TET5-6 in Halobacteria), the TET11 group is widespread in Archaea and is found in all superphyla (**Fig. 5**), suggesting that this family emerged at the base of Archaea. According to MtTET (**Fig. 3**) and MjTET (Atalah *et al.,* in preparation) characterization, TET11 members would display broad-spectrum activities. The ancestral origin of TET11, coupled with its enzymatic activity, might suggest that this family was the first to appear in Archaea, followed by several duplications and horizontal transfers giving rise to multiple families with different activities.

Regardless, the phylogenetic analysis indicates that TET multiplicity arose independently multiple times. This phenomenon is not exclusively attributable to HGTs, as evidenced by the emergence of the TET1 and TET4 groups by duplication. It does not appear to stem from environmental adaptation either, as no correlation was identified between TET distribution and specific biotopes. For example, although Archaeoglobales, Thermococcales, Methanococcales, and Desulfurococcales all share the same ecological niche as primary colonizers of deep-sea hydrothermal vents[53–55], these organisms exhibit markedly different TET distribution patterns.

On the other hand, the number and degree of specificity of TET peptidases present in an organism may correlate with its metabolic capabilities. Multiple TETs are found in

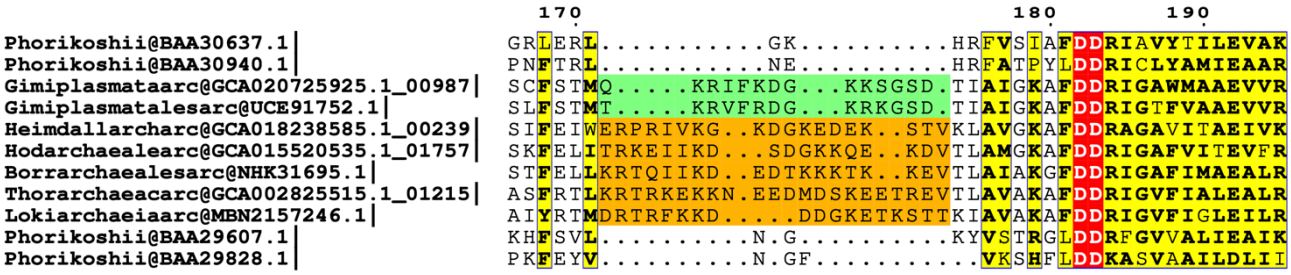
heterotrophic and mixotrophic Hadarchaea[56], Thermococcales[57–59], Halobacteriales[60], Crenarchaeota[61] (*Vulcanisaeta* genus), Heimdallarchaeota[62–64], Korarchaeota[65,66], and Bathyarchaeota[67–69] species. This suggests that TETs may play a metabolic role in the degradation of environmental peptides used as carbon sources, enabling efficient organic matter utilization. In contrast, TET peptidases are typically found in single copy in autotrophic species such as methanogens, which do not depend on the degradation of exogenous peptides, hinting at an alternative physiological role for these enzymes. As initially proposed by Franzetti *et al.*, TETs may participate in protein homeostasis and amino acid recycling by processing peptides downstream of the proteasome and other related proteolytic complexes[10,11].

The biological function of TET peptidases remains an open question. In an attempt to identify co-located genes that might be functionally related or co-regulated, the genomic neighborhood was examined; however, no significant genomic conservation or noteworthy syntenic blocks were identified (data not shown). To this day, the preliminary results from the triple knock-out experiment of *E. coli* TET peptidases remain the only genetic studies reported on these enzymes. No growth defects were observed in minimal media or under various stress conditions, suggesting that the genes encoding TET peptidases are not essential in this organism[52]. Similarly, a transposon sequencing experiment conducted on the archaeon *Methanococcus maripaludis* indicate that the single TET peptidase in this species is also non-essential[70]. These findings could be interpreted in two different ways: TET peptidases may not be essential in these species, or there could be functional analogs compensating for their absence. Previous studies suggested complementarity between M42 and M18 or TRI peptidases[11,37,52]. To investigate these theories, we used the PFAM domains PF14684 and PF02127 to search for TRI and M18 homologues in our local database of archaeal genomes (**Supplementary Table XX**). Our results challenge these hypotheses; M18 and TRI homologues were only sparsely detected, primarily in species possessing M42 peptidases. Furthermore, several lineages (*e.g.,* Theionarchaea, Pontarchaeia, Thalassoarchaeia, Methanocellia, Thaumarchaeota) lack all three peptidase families. This points to a more complex relationship and suggests the existence of other functional analogs yet to be identified.

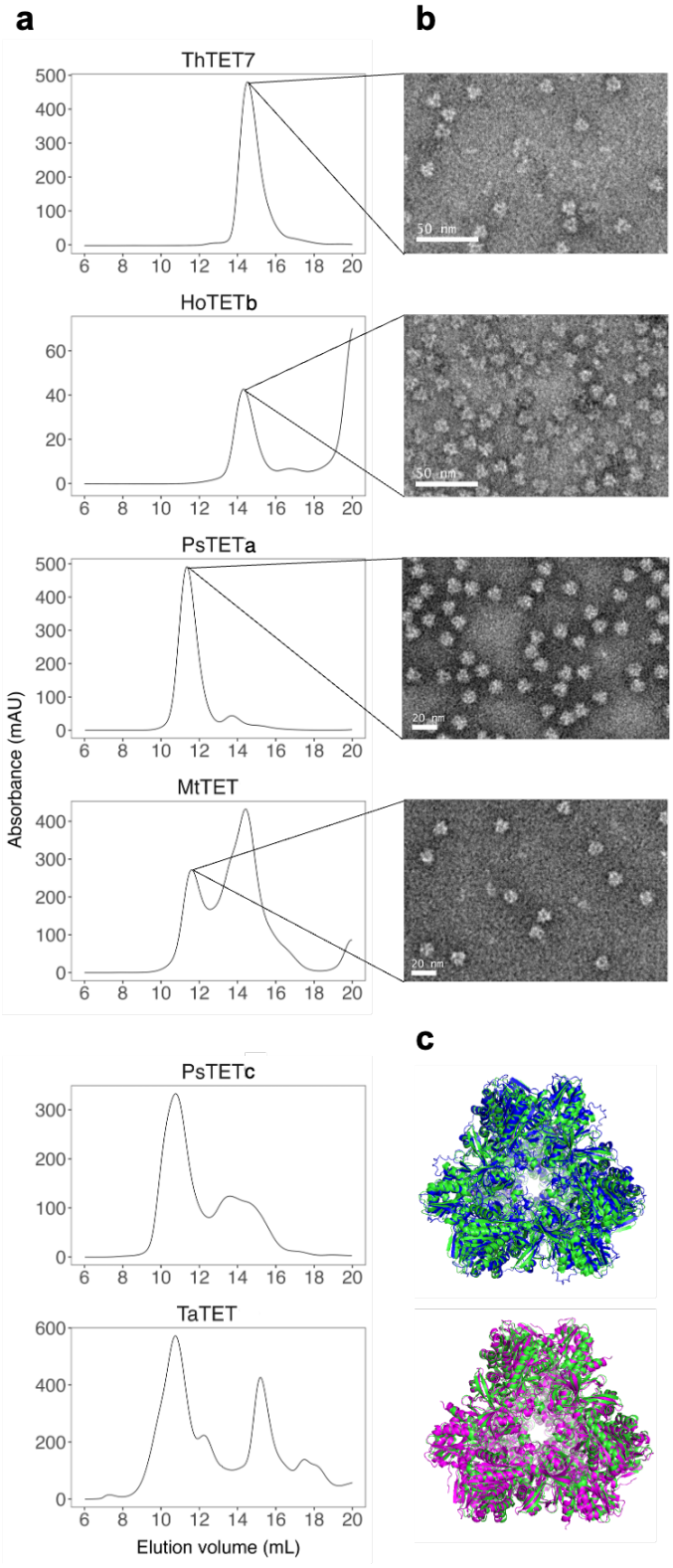
Further genetic studies would be needed to establish the exact physiological role of these enzymes. *Methanocaldococcus jannaschii* might be a good candidate for this type of study; genetic tools are available for this species[71] and it possesses a single TET peptidase belonging to the TET11 family, thought to be the archaeal ancestral family.

Ultimately, the hybrid approach adopted in this study—integrating structural biology, phylogeny, and biochemistry—revealed an unsuspected diversity of TET peptidases. This innovative strategy shed light a complex evolutionary history to light, uncovering an ancient subgroup of archaeal enzymes that had so far gone unnoticed, yet appears to represent the general case among archaea. Additionally, this approach allowed for the classification of archaeal TET peptidases into eleven distinct families. Characterized representatives from these families revealed contrasting and biochemical properties, underscoring the potential value of this approach for biotechnological applications. For instance, archaeal TET peptidases, used individually or combined in mixtures, can increase the diversity of bioactive peptides in hydrolyzates derived from natural biomass[23–25]. These functionalized hydrolyzates have potential applications in nutrition, health, and cosmetics. Further research on the newly discovered classes of archaeal TET peptidases could expand the range of enzymes available for these industrial applications.

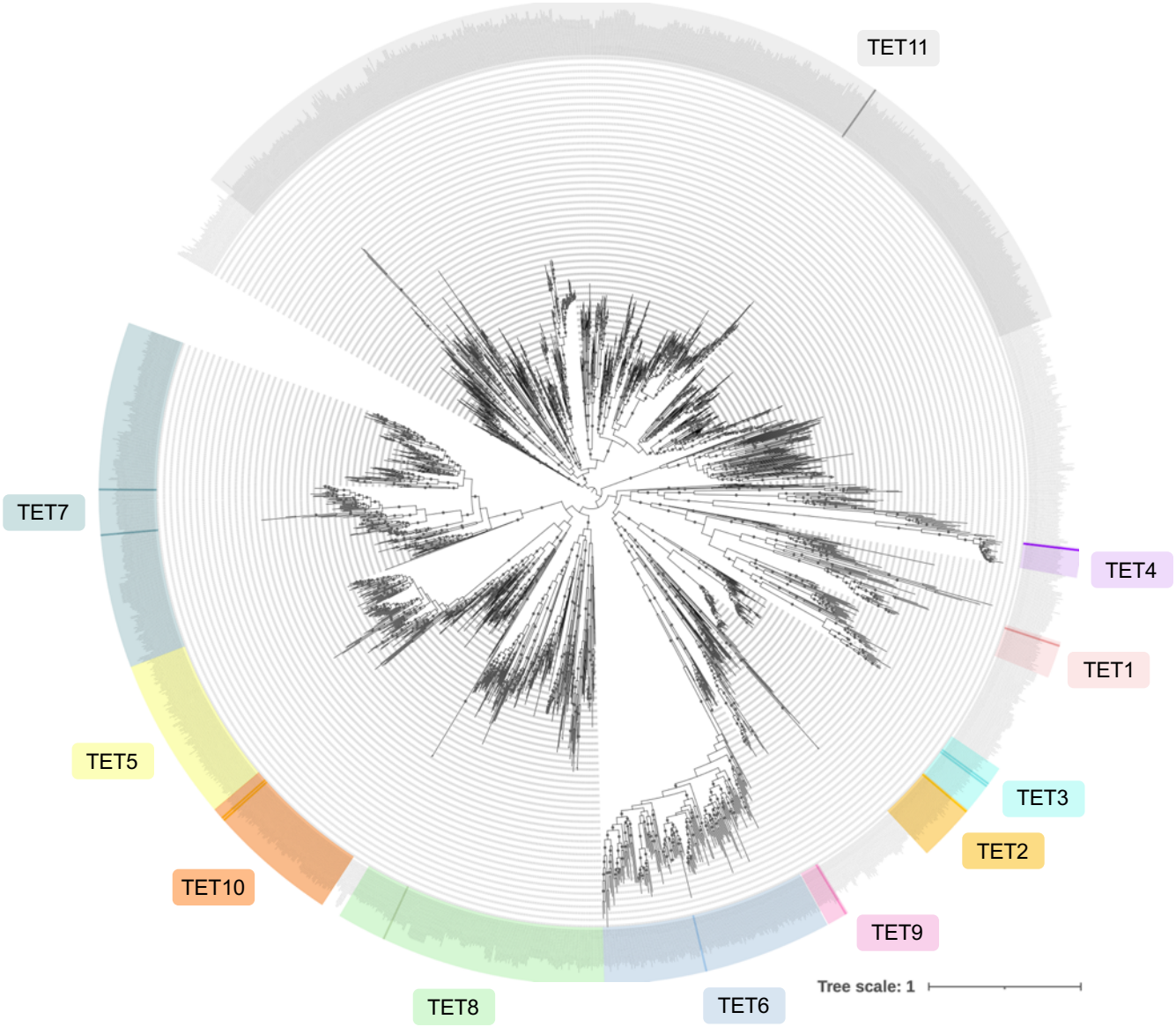
Supplementary figures



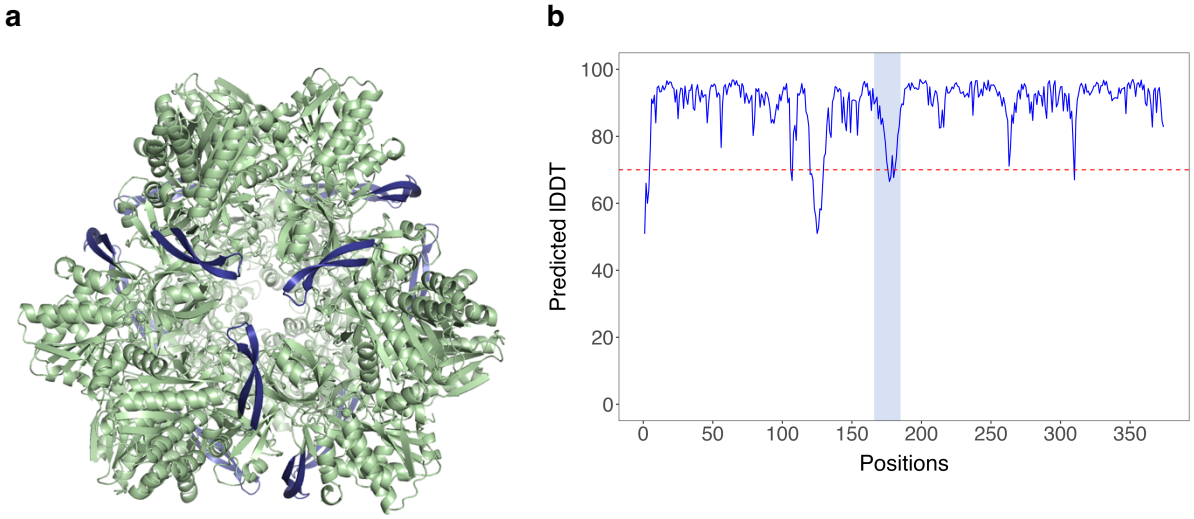
Supplementary Fig. 1: Multiple sequence alignment of 11 archaeal M42 peptidases illustrating a previously unreported insertion in Asgard and Thermoplasmata sequences. *The unique ~20 residue insertion, specific to some Asgard species, is highlighted in orange. A shorter but similar insertion (9-10 residues), highlighted in green, was identified in certain Thermoplasmata sequences.*



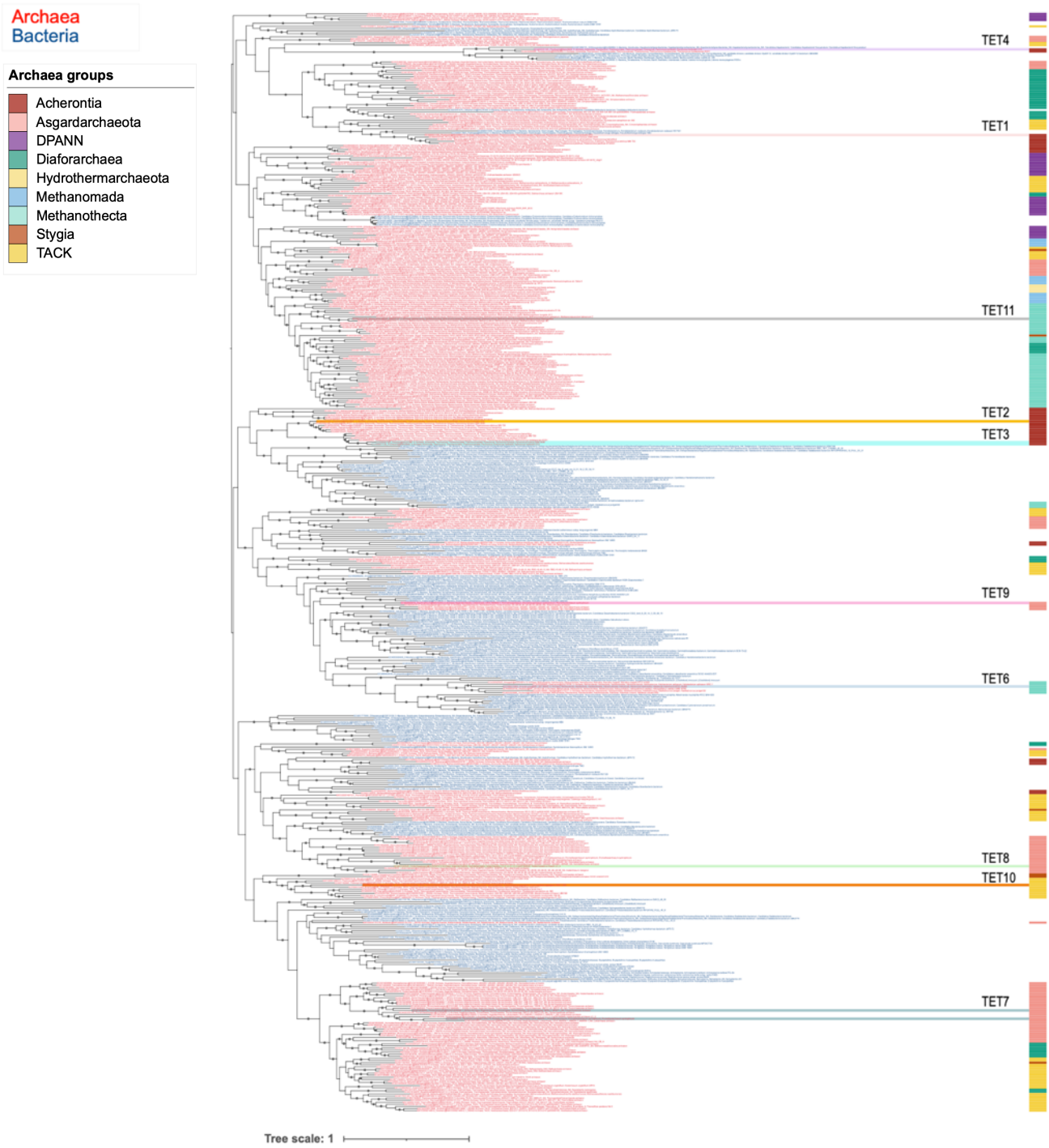
Supplementary Fig. 2: New TET peptidases purification. *(a) Size exclusion chromatography purification. ThTET and HoTETb were purified using a Superose 6 Increase 10/300 GL column. Other proteins were purified on a Superdex 200 10/300 GL column. All proteins eluted as high molecular mass complexes of c. 450 kDa. (b) Negative-stain electron microscopy observation of purified TETs. Homogenous populations of hollow tetrahedral particles were observed. (c) Structural alignment of PsTETc (blue) and TaTET (pink) AlphaFold3 models vs. PhTET2 (PDB code 1Y0R, in green). Typical TET structures were predicted for PsTETc and TaTET.*



Supplementary Fig. 3: Delineation of TET peptidases families in Archaea. *Maximum-likelihood phylogeny obtained from an alignment of 2,030 sequences and 337 amino acid positions. The scale bar represents the average number of substitutions per site. Circles at the branches indicate ultra-fast bootstrap values >= 90 %. TET families were delineated based on the taxonomic distribution, the topology of the tree (i.e., branch lengths, node supports) and substrate specifities. Darker bars on the outer circle represent enzymes characterized either before or during this study.*



Supplementary Fig. 4: (a) PsTETa AlphaFold 3 model *(ipTM score 0.87) featuring the novel insertion identified in some Asgard sequences, here colored in blue.* ***(b) Associated plDDT plot per residue.*** *The dashed red line indicates the confidence threshold (plDDT > 70), above which predicted structures are generally considered reliable. The region of the novel insertion (between residues 166 and 185) is highlighted in blue.*



Supplementary Fig. 5: Phylogeny of archaeal and bacterial M42 peptidase homologues. *Maximum-likelihood phylogeny obtained from an alignment of 526 sequences and 337 amino acid positions. The scale bar represents the average number of substitutions per site. Circles at the branches indicate ultra-fast bootstrap values ≥ 90 %. Archaeal and bacterial sequences are indicated in red and blue, respectively. Archaeal characterized sequences are highlighted. Archaeal taxonomic distribution is represented on the right.*

Supplementary information

ThTET purification protocol

After cell lysis, incubation at 70 °C for 15 min and clarification, the resulting supernatant was supplemented with imidazole (final concentration 10 mM) and loaded on a HiTrap Chelating HP 5 mL column (Cityva) equilibrated with 50 mM Tris, 150 mM NaCl, 10 mM imidazole, pH 8.0. Bound proteins were eluted with a linear gradient of imidazole (10 to 500 mM). Fractions corresponding to the elution peak at 400 mM imidazole were pooled, dialysed against 50 mM Tris, 20 mM NaCl, pH 8.0 and loaded on a ResourceQ column (Cytiva) equilibrated with the same buffer. Elution was achieved by a linear NaCl gradient (20 to 500 mM) and fractions containing protein of similar mass (37-39 kDa) according to SDS-PAGE were combined and concentrated using an Amicon Ultra-15 ultrafiltration unit (Millipore) with a 30 kDa cutoff. The protein was utlimately loaded on a Superose 6 Increase 10/300 GL column (Cytiva) in 50 mM Tris, 150 mM NaCl, pH 8.0. Fractions from the elution peak corresponding to a molecular mass around 450 kDa were pooled and subsequently concentrated using an an Amicon Ultra-15 ultrafiltration unit (Millipore) with a 30 kDa cutoff.

HoTETb purification protocol

After cell lysis, incubation at 70 °C for 15 min and clarification, the resulting supernatant was diluted to a final NaCl concentration of 75 mM and loaded on a ResourceQ column (Cytiva) equilibrated with 50 mM Tris, 75 mM NaCl, pH 8.0. Elution was achieved by a linear NaCl gradient (75 to 300 mM) and fractions containing protein of similar mass (37-39 kDa) according to SDS-PAGE were combined and concentrated using an Amicon Ultra-15 ultrafiltration unit (Millipore) with a 30 kDa cutoff. The protein was then loaded on a Superose 6 Increase 10/300 GL column (Cytiva) in 50 mM Tris, 150 mM NaCl, pH 8.0. Fractions from the elution peak corresponding to a molecular mass around 450 kDa were pooled and subsequently concentrated using an Amicon Ultra-15 ultrafiltration unit (Millipore) with a 30 kDa cutoff.

TaTET purification protocol

After cell lysis, incubation at 70 °C for 15 min and clarification, the resulting supernatant was dialysed against 50 mM Tris, 50 mM NaCl, pH 8.0 and loaded on a ResourceQ column (Cytiva) equilibrated with the same buffer. Elution was achieved by a linear NaCl gradient (50 mM to 1

M) and fractions containing protein of similar mass (37-39 kDa) according to SDS-PAGE were combined and concentrated using an Amicon Ultra-15 ultrafiltration unit (Millipore) with a 30 kDa cutoff. The protein was then loaded on a Superdex 200 10/300 GL column (Cytiva) in 50 mM Tris, 150 mM NaCl, pH 8.0. Fractions from the elution peak corresponding to a molecular mass around 450 kDa were pooled and subsequently concentrated using an Amicon Ultra-15 ultrafiltration unit with a 30kDa cutoff.

PsTETa, PsTETc, and MtTET purification protocol

After cell lysis and clarification, the resulting supernatant was dialysed against 50 mM Tris, 20 mM NaCl, pH 8.0 and loaded on a DEAE sepharose CL-6B resin (Cytiva, XK16/20 column) equilibrated with the same buffer. Elution was achieved by a linear NaCl gradient (20 to 600 mM) and fractions containing protein of similar mass (37-39 kDa) according to SDS-PAGE were combined, dialysed against 50 mM Tris, 50 mM NaCl, pH 8.0 and loaded on a ResourceQ column (Cytiva) equilibrated with the same buffer. Elution was achieved by a linear NaCl gradient (50 to 500 mM) and fractions containing the protein of interest were pooled and concentrated using an Amicon Ultra-15 ultrafiltration unit (Millipore) with a 30 kDa cutoff. The protein was then loaded on a Superdex 200 10/300 GL column (Cytiva) in 50 mM Tris, 150 mM NaCl, pH 8.0. Fractions from the elution peak corresponding to a molecular mass around 450 kDa were combined and subsequently concentrated using an Amicon Ultra-15 ultrafiltration unit with a 30kDa cutoff.

References

1. Mackinnon, E. & Stone, S. L. The Ubiquitin Proteasome System and Nutrient Stress Response. *Front. Plant Sci.* **13**, 867419 (2022).

2. Bertels, F. *et al.* Dissecting HIV Virulence: Heritability of Setpoint Viral Load, CD4+ T-Cell Decline, and Per-Parasite Pathogenicity. *Mol. Biol. Evol.* **35**, 27–37 (2018).

3. Koepp, D. M. Cell Cycle Regulation by Protein Degradation. in *Cell Cycle Control* (eds. Noguchi, E. & Gadaleta, M. C.) vol. 1170 61–73 (Springer New York, New York, NY, 2014).

4. Mahmoud, S. A. & Chien, P. Regulated proteolysis in Bacteria. *Annu. Rev. Biochem.* **87**, 677–696 (2018).

5. Ciechanover, A. Intracellular protein degradation: From a vague idea thru the lysosome and the ubiquitin-proteasome system and onto human diseases and drug targeting. *Best Practice & Research Clinical Haematology* **30**, 341–355 (2017).

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

7. Lupas, A., Flanagan, J. M., Tamura, T. & Baumeister, W. Self-compartmentalizing proteases. *Trends Biochem. Sci.* **22**, 399–404 (1997).

8. 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. *Nucl. Acids Res.* **46**, 624–632 (2018).

9. Paysan-Lafosse, T. *et al.* InterPro in 2022. *Nucl. Acids Res.* **51**, D418–D427 (2023).

10. Franzetti, B. *et al.* Tetrahedral aminopeptidase: a novel large protease complex from archaea. *EMBO J.* **21**, 2132–2138 (2002).

11. Borissenko, L. & Groll, M. Crystal structure of TET protease reveals complementary protein degradation pathways in prokaryotes. *J. Mol. Biol.* **346**, 1207–1219 (2005).

12. Russo, S. & Baumann, U. Crystal Structure of a Dodecameric tetrahedral-shaped aminopeptidase. *J. Biol. Chem.* **279**, 51275–51281 (2004).

13. 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. *Mol. Microbiol.* **72**, 26–40 (2009).

14. Appolaire, A. *et al.* TET peptidases: a family of tetrahedral complexes conserved in prokaryotes. *Biochimie* **122**, 188–196 (2016).

15. Rawlings, N. D. & Barrett, A. J. Introduction: Metallopeptidases and Their Clans. in *Handbook of Proteolytic Enzymes* 325–370 (Elsevier, 2013). doi:10.1016/B978-0-12-382219-2.00077-6.

16. Colombo, M., Girard, E. & Franzetti, B. Tuned by metals: the TET peptidase activity is controlled by 3 metal binding sites. *Sci. Rep.* **6**, 20876–20888 (2016).

17. Appolaire, A. *et al.* Small-angle neutron scattering reveals the assembly mode and oligomeric architecture of TET, a large, dodecameric aminopeptidase. *Acta Cryst. D* **70**, 2983–2993 (2014).

18. 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. *Biochim.*

Biophys. Acta Proteins Proteom. **1814**, 1289–1294 (2011).

19. Durá, M. A. *et al.* Characterization of a TET-like aminopeptidase complex from the hyperthermophilic archaeon Pyrococcus horikoshii. *Biochemistry* **44**, 3477–3486 (2005).

20. Durá, M. A. & Franzetti, B. PhTET1 aminopeptidase. in *Handbook of Proteolytic Enzymes* vol. 2 1638–1645 (Elsevier, 2013).

21. Basbous, H., Appolaire, A., Girard, E. & Franzetti, B. Characterization of a glycyl-specific TET aminopeptidase complex from Pyrococcus horikoshii. *J. Bacteriol.* **200**, 1–11 (2018).

22. Schoehn, G. *et al.* An archaeal peptidase assembles into two different quaternary structures. *J. Biol. Chem.* **281**, 36327–36337 (2006).

23. Franzetti, B., Girard, E. & Appolaire, A. Use of a Combination of Tet Exoproteases Obtained from Extremophilic Microorganisms for Hydrolyzing Polypeptides. (2019).

24. Franzetti, B., Girard, E., Appolaire, A. & Basbous, H. Peptidase and Its Uses. (2019).

25. Franzetti, B., Girard, E., Appolaire, A. & Basbous, H. Aminopeptidase and its uses. (2019).

26. Slutskaya, E. S. *et al.* Characterization of a novel M42 aminopeptidase from crenarchaeon Desulfurococcus kamchatkensis. *Dokl. Biochem. Biophys.* **442**, 30–32 (2012).

27. Zhang, Z. *et al.* Protein sequence similarity searches using patterns as seeds. *Nucl. Acids Res.* **26**, 3986–3990 (1998).

28. Notredame, C., Higgins, D. G. & Heringa, J. T-coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217 (2000).

29. Johnson, L. S., Eddy, S. R. & Portugaly, E. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* **11**, 431–438 (2010).

30. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

31. Criscuolo, A. & Gribaldo, S. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* **10**, 210–230 (2010).

32. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).

33. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).

34. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucl. Acids Res.* **47**, 256–259 (2019).

35. Sambrook, J. & Russell, D. W. *Molecular Cloning: A Laboratory Manual*. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y, 2001).

36. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254 (1976).

37. Dutoit, R., Brandt, N., Legrain, C. & Bauvois, C. Functional characterization of two M42 aminopeptidases erroneously annotated as cellulases. *PLoS ONE* **7**, 1–8 (2012).

38. Lee, H. S. *et al.* Biochemical characterization of deblocking aminopeptidase from hyperthermophilic archaeon Thermococcus onnurineus NA1. *J. Biosci. Bioeng.* **104**, 188–194 (2007).

39. Lee, Y. G., Leem, S.-H., Chung, Y.-H. & Kim, S. I. Characterization of thermostable deblocking aminopeptidases of archaeon Thermococcus onnurineus NA1 by proteomic and biochemical approaches. *J. Microbiol.* **50**, 792–797 (2012).

40. Appolaire, A. *et al.* The TET2 and TET3 aminopeptidases from Pyrococcus horikoshii form a hetero-subunit peptidasome with enhanced peptide destruction properties. *Mol. Microbiol.* **94**, 803–814 (2014).

41. Robinson, M. W. *et al.* MHJ_0125 is an M42 glutamyl aminopeptidase that moonlights as a multifunctional adhesin on the surface of *Mycoplasma hyopneumoniae*. *Open Biol.* **3**, 1–15 (2013).

42. Dutoit, R. *et al.* How metal cofactors drive dimer–dodecamer transition of the M42 aminopeptidase TmPep1050 of Thermotoga maritima. *J. Biol. Chem.* **294**, 17777–17789 (2019).

43. Rivard, C. J. & Smith, P. H. Isolation and Characterization of a Thermophilic Marine Methanogenic Bacterium, Methanogenium thermophilicum sp. nov. *Int. J. Syst. Bacteriol.* **32**, 430–436 (1982).

44. Zabel, H.-P., König, H. & Winter, J. Emended Description of Methanogenium thermophilicum, Rivard and Smith, and Assignment of New Isolates to this Species. *Syst. Appl. Microbiol.* **6**, 72–78 (1985).

45. Huber, R., Dyba, D., Huber, H., Burggraf, S. & Rachel, R. Sulfur-inhibited Thermosphaera aggregans sp. nov., a new genus of hyperthermophilic archaea isolated after its prediction from environmentally derived 16S rRNA sequences. *IJSB* **48**, 31–38 (1998).

46. Imachi, H. *et al.* Isolation of an archaeon at the prokaryote–eukaryote interface. *Nature* **577**, 519–525 (2020).

47. Eme, L. *et al.* Inference and reconstruction of the heimdallarchaeial ancestry of eukaryotes. *Nature* **618**, 992–999 (2023).

48. Lu, Z. *et al.* Evolution of optimal growth temperature in Asgard archaea inferred from the temperature dependence of GDP binding to EF-1A. *Nat. Commun.* **15**, 515–521 (2024).

49. Garcia, P. S., Gribaldo, S. & Borrel, G. Diversity and Evolution of Methane-Related Pathways in Archaea. *Annu. Rev. Microbiol.* **76**, 727–755 (2022).

50. Roncari, G. & Zuber, H. THERMOPHILIC AMINOPEPTIDASES FROM BACILLUS STEAROTHERMOPHILUS. I. ISOLATION, SPECIFICITY, AND GENERAL PROPERTIES OF THE THERMOSTABLE AMINOPEPTIDASE I *. *Int. J. Protein Res.* **1**, 45–61 (1969).

51. Kumaki, Y. *et al.* Family M42 aminopeptidase from the syntrophic bacterium Symbiobacterium thermophilum: characterization using recombinant protein. *J. Biosci. Bioeng.* **111**, 134–139 (2011).

52. Dutoit, R. The paradigm of self-compartmentalized M42 aminopeptidases: insight into their oligomerization, substrate specificities, and physiological function. (Université libre de Bruxelles, Faculté des Sciences – Biologie Moléculaire, 2021).

53. Huber, J. A., Butterfield, D. A. & Baross, J. A. Temporal Changes in Archaeal Diversity and Chemistry in a Mid-Ocean Ridge Subseafloor Habitat. *Appl. Environ. Microbiol.* **68**, 1585–1594 (2002).

54. Nercessian, O., Reysenbach, A., Prieur, D. & Jeanthon, C. Archaeal diversity associated with *in situ* samplers deployed on hydrothermal vents on the East Pacific Rise (13°N). *Environ. Microbiol.* **5**, 492–502 (2003).

55. Takai, K. *et al.* Spatial Distribution of Marine Crenarchaeota Group I in the Vicinity of Deep-Sea Hydrothermal Systems. *Appl Environ Microbiol* **70**, 2404–2413 (2004).

56. Baker, B. J. *et al.* Genomic inference of the metabolism of cosmopolitan subsurface Archaea, Hadesarchaea. *Nat. Microbiol.* **1**, 16002–16008 (2016).

57. Lee, H. S. *et al.* The Complete Genome Sequence of *Thermococcus onnurineus* NA1 Reveals a Mixed Heterotrophic and Carboxydotrophic Metabolism. *J. Bacteriol.* **190**, 7491–7499 (2008).

58. Pillot, G. *et al.* Identification of enriched hyperthermophilic microbial communities from a deep-sea hydrothermal vent chimney under electrolithoautotrophic culture conditions. *Sci. Rep.* **11**, 14782–14793 (2021).

59. Bertoldo, C. & Antranikian, G. The Order Thermococcales. in *The Prokaryotes* (eds. Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H. & Stackebrandt, E.) 69–81 (Springer New York, New York, NY, 2006). doi:10.1007/0-387-30743-5_5.

60. Oren, A. The Order Halobacteriales. in *The Prokaryotes* (eds. Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H. & Stackebrandt, E.) 113–164 (Springer New York, New York, NY, 2006). doi:10.1007/0-387-30743-5_8.

61. Itoh, T. Vulcanisaeta distributa gen. nov., sp. nov., and Vulcanisaeta souniana sp. nov., novel hyperthermophilic, rod-shaped crenarchaeotes isolated from hot springs in Japan. *Int. J. Syst. Evol. Microbiol.* **52**, 1097–1104 (2002).

62. Cai, M. *et al.* Diverse Asgard archaea including the novel phylum Gerdarchaeota participate in organic matter degradation. *Sci. China Life Sci.* **63**, 886–897 (2020).

63. Liu, Y. *et al.* Expanded diversity of Asgard archaea and their relationships with eukaryotes. *Nature* **593**, 553–557 (2021).

64. Bulzu, P.-A. *et al.* Casting light on Asgardarchaeota metabolism in a sunlit microoxic niche. *Nat. Microbiol.* **4**, 1129–1137 (2019).

65. McKay, L. J. *et al.* Co-occurring genomic capacity for anaerobic methane and dissimilatory sulfur metabolisms discovered in the Korarchaeota. *Nat. Microbiol.* **4**, 614–622 (2019).

66. Elkins, J. G. *et al.* A korarchaeal genome reveals insights into the evolution of the Archaea. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 8102–8107 (2008).

67. Feng, X., Wang, Y., Zubin, R. & Wang, F. Core Metabolic Features and Hot Origin of Bathyarchaeota. *Engineering* **5**, 498–504 (2019).

68. Khomyakova, M. A., Merkel, A. Y., Mamiy, D. D., Klyukina, A. A. & Slobodkin, A. I. Phenotypic and genomic characterization of Bathyarchaeum tardum gen. nov., sp. nov., a cultivated representative of the archaeal class Bathyarchaeia. *Front. Microbiol.* **14**, 1214631–1214645 (2023).

69. Hou, J. *et al.* Taxonomic and carbon metabolic diversification of Bathyarchaeia during its coevolution history with early Earth surface environment. *Sci. Adv.* **9**, 1–17 (2023).

70. Sarmiento, F., Mrázek, J. & Whitman, W. B. Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon *Methanococcus maripaludis*. *PNAS* **110**, 4726–4731 (2013).

71. Susanti, D., Frazier, M. C. & Mukhopadhyay, B. A genetic system for Methanocaldococcus jannaschii: an evolutionary deeply rooted hyperthermophilic methanarchaeon. *Front. Microbiol.* **10**, (2019).

References

Adam, P. S., Borrel, G., Brochier-Armanet, C., & Gribaldo, S. (2017). The growing tree of archaea : New perspectives on their diversity, evolution and ecology. *The ISME Journal*, *11*(11), Article 11. https://doi.org/10.1038/ismej.2017.122

Aklujkar, M., Risso, C., Smith, J., Beaulieu, D., Dubay, R., Giloteaux, L., DiBurro, K., & Holmes, D. (2014). Anaerobic degradation of aromatic amino acids by the hyperthermophilic archaeon Ferroglobus placidus. *Microbiology*, *160*(12), 2694-2709. https://doi.org/10.1099/mic.0.083261-0

Allen, D. E., Baker, D. J., & Gillard, R. D. (1967). Metal Complexing by Tris Buffer. *Nature*, *214*(5091), 906-907. https://doi.org/10.1038/214906a0

Almeida, V. M., & Marana, S. R. (2019). Optimum temperature may be a misleading parameter in enzyme characterization and application. *PLOS ONE*, *14*(2), 1-8. https://doi.org/10.1371/journal.pone.0212977

Ando, S., Ishikawa, K., Ishida, H., Kawarabayasi, Y., Kikuchi, H., & Kosugi, Y. (1999). Thermostable aminopeptidase from Pyrococcus horikoshii. *FEBS Letters*, *447*(1), Article 1. https://doi.org/10.1016/S0014-5793(99)00257-4

Appolaire, A. (2014). *Etude des grands assemblages protéolytiques de la famille TET : processus d'oligomérisation et régulation fonctionnelle associée* [Université de Grenoble]. https://theses.fr/2014GRENV062

Appolaire, A., Colombo, M., Basbous, H., Gabel, F., Girard, E., & Franzetti, B. (2016). TET peptidases : A family of tetrahedral complexes conserved in prokaryotes. *Biochimie*, *122*, 188-196. https://doi.org/10.1016/j.biochi.2015.11.001

Appolaire, A., Durá, M. A., Ferruit, M., Andrieu, J.-P., Godfroy, A., Gribaldo, S., & Franzetti, B. (2014). The TET2 and TET3 aminopeptidases from Pyrococcus horikoshii form a hetero-subunit peptidasome with enhanced peptide destruction properties. *Molecular Microbiology*, *94*(4), Article 4. https://doi.org/10.1111/mmi.12775

Appolaire, A., Girard, E., Colombo, M., Durá, M. A., Moulin, M., Härtlein, M., Franzetti, B., & Gabel, F. (2014). 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), Article 11. https://doi.org/10.1107/S1399004714018446

Appolaire, A., Rosenbaum, E., Durá, M. A., Colombo, M., Marty, V., Savoye, M. N., Godfroy, A., Schoehn, G., Girard, E., Gabel, F., & Franzetti, B. (2013). Pyrococcus horikoshii TET2 peptidase assembling process and associated functional regulation. *Journal of Biological Chemistry*, *288*(31), Article 31. https://doi.org/10.1074/jbc.M113.450189

Auld, D. S. (2004). Catalytic mechanisms for metallopeptidases. In *Handbook of Proteolytic Enzymes* (Vol. 1, p. 370-396). Elsevier. https://doi.org/10.1016/B978-0-12-079611-3.50076-8

Ayoub, N., Roth, P., Ucurum, Z., Fotiadis, D., & Hirschi, S. (2023). Structural and biochemical insights into His-tag-induced higher-order oligomerization of membrane proteins by cryo-EM and size exclusion chromatography. *Journal of Structural Biology*, *215*(1), 107924. https://doi.org/10.1016/j.jsb.2022.107924

Bacon, C. L., Jennings, P. V., Ni Fhaolain, I., & O'Cuinn, G. (1994). Purification and characterisation of an aminopeptidase A from cytoplasm of Lactococcus lactis subsp. Cremoris AM2. *International Dairy Journal*, *4*(6), 503-519. https://doi.org/10.1016/0958-6946(94)90022-1

Baker, B. J., De Anda, V., Seitz, K. W., Dombrowski, N., Santoro, A. E., & Lloyd, K. G. (2020). Diversity, ecology and evolution of Archaea. *Nature Microbiology*, *5*(7), Article 7. https://doi.org/10.1038/s41564-020-0715-z

Baker, B. J., Saw, J. H., Lind, A. E., Lazar, C. S., Hinrichs, K.-U., Teske, A. P., & Ettema, T. J. G. (2016). Genomic inference of the metabolism of cosmopolitan subsurface Archaea, Hadesarchaea. *Nature Microbiology*, *1*(3), 16002-16008. https://doi.org/10.1038/nmicrobiol.2016.2

Banegas, I., Prieto, I., Vives, F., Alba, F., De Gasparo, M., Segarra, A. B., Hermoso, F., Durán, R., & Ramírez, M. (2006). Review : Brain Aminopeptidases and Hypertension. *Journal of the Renin-Angiotensin-Aldosterone System*, *7*(3), 129-134. https://doi.org/10.3317/jraas.2006.021

Bard, J. A. M., Goodall, E. A., Greene, E. R., Jonsson, E., Dong, K. C., & Martin, A. (2018). Structure and Function of the 26S Proteasome. *Annual Review of Biochemistry*, *87*(1), 697-724. https://doi.org/10.1146/annurev-biochem-062917-011931

Barthelme, D., Chen, J. Z., Grabenstatter, J., Baker, T. A., & Sauer, R. T. (2014). Architecture and assembly of the archaeal Cdc48·20S proteasome. *Proceedings of the National Academy of Sciences*, *111*(17). https://doi.org/10.1073/pnas.1404823111

Barton, L. L., Goulhen, F., Bruschi, M., Woodards, N. A., Plunkett, R. M., & Rietmeijer, F. J. M. (2007). The bacterial metallome : Composition and stability with specific reference to the anaerobic bacterium Desulfovibrio desulfuricans. *BioMetals*, *20*(3-4), 291-302. https://doi.org/10.1007/s10534-006-9059-2

Basbous, H. (2016). *Etudes structurales et propriétés enzymatiques de deux nouvelles aminopeptidases TETs auto-compartimentées chez les archées* [Université Grenoble Alpes]. https://theses.fr/2016GREAV016

Basbous, H., Appolaire, A., Girard, E., & Franzetti, B. (2018). Characterization of a glycyl-specific TET aminopeptidase complex from Pyrococcus horikoshii. *Journal of Bacteriology*, *200*(17), Article 17. https://doi.org/10.1128/JB.00059-18

Batour, M., Laurent, S., Moalic, Y., Chamieh, H., Taha, S., & Jebbar, M. (2023). The secretome of *Thermococcus barophilus* in the presence of carbohydrates and the potential role of the TrmBL4 regulator. *Environmental Microbiology Reports*, *15*(6), 530-544. https://doi.org/10.1111/1758-2229.13186

Bauer, N. C., Doetsch, P. W., & Corbett, A. H. (2015). Mechanisms Regulating Protein Localization. *Traffic*, *16*(10), 1039-1061. https://doi.org/10.1111/tra.12310

Bednar, J., Garcia-Saez, I., Boopathi, R., Cutter, A. R., Papai, G., Reymer, A., Syed, S. H., Lone, I. N., Tonchev, O., Crucifix, C., Menoni, H., Papin, C., Skoufias, D. A., Kurumizaka, H., Lavery, R., Hamiche, A., Hayes, J. J., Schultz, P., Angelov, D., … Dimitrov, S. (2017). Structure and Dynamics of a 197 bp Nucleosome in Complex with Linker Histone H1. *Molecular Cell*, *66*(3), 384-397.e8. https://doi.org/10.1016/j.molcel.2017.04.012

Belay, N., Mukhopadhyay, B., Conway De Macario, E., Galask, R., & Daniels, L. (1990). Methanogenic bacteria in human vaginal samples. *Journal of Clinical Microbiology*, *28*(7), 1666-1668. https://doi.org/10.1128/jcm.28.7.1666-1668.1990

Bell, P. J. (2001). Viral Eukaryogenesis : Was the Ancestor of the Nucleus a Complex DNA Virus? *Journal of Molecular Evolution*, *53*(3), 251-256. https://doi.org/10.1007/s002390010215

Benaroudj, N., & Goldberg, A. L. (2000). PAN, the proteasome-activating nucleotidase from archaebacteria, is a protein-unfolding molecular chaperone. *Nature Cell Biology*, *2*(11), 833-839. https://doi.org/10.1038/35041081

Bendtsen, J. D., Jensen, L. J., Blom, N., Von Heijne, G., & Brunak, S. (2004). Feature-based prediction of non-classical and leaderless protein secretion. *Protein Engineering Design and Selection*, *17*(4), 349-356. https://doi.org/10.1093/protein/gzh037

Bertels, F., Marzel, A., Leventhal, G., Mitov, V., Fellay, J., Günthard, H. F., Böni, J., Yerly, S., Klimkait, T., Aubert, V., Battegay, M., Rauch, A., Cavassini, M., Calmy, A., Bernasconi, E., Schmid, P., Scherrer, A. U., Müller, V., Bonhoeffer, S., … the Swiss HIV Cohort Study. (2018). Dissecting HIV Virulence : Heritability of Setpoint Viral Load, CD4+ T-Cell Decline, and Per-Parasite Pathogenicity. *Molecular Biology and Evolution*, *35*(1), 27-37. https://doi.org/10.1093/molbev/msx246

Bertoldo, C., & Antranikian, G. (2006). The Order Thermococcales. In M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, & E. Stackebrandt (Éds.), *The Prokaryotes* (p. 69-81). Springer New York. https://doi.org/10.1007/0-387-30743-5_5

Betancurt-Anzola, L., Martínez-Carranza, M., Delarue, M., Zatopek, K. M., Gardner, A. F., & Sauguet, L. (2023). Molecular basis for proofreading by the unique exonuclease domain of Family-D DNA polymerases. *Nature Communications*, *14*(1), 8306. https://doi.org/10.1038/s41467-023-44125-x

Birien, T., Thiel, A., Henneke, G., Flament, D., Moalic, Y., & Jebbar, M. (2018). Development of an Effective 6-Methylpurine Counterselection Marker for Genetic Manipulation in Thermococcus barophilus. *Genes*, *9*(2), 77. https://doi.org/10.3390/genes9020077

Block, H., Maertens, B., Spriestersbach, A., Brinker, N., Kubicek, J., Fabis, R., Labahn, J., & Schäfer, F. (2009). Chapter 27 Immobilized-Metal Affinity Chromatography (IMAC). In *Methods in Enzymology* (Vol. 463, p. 439-473). Elsevier. https://doi.org/10.1016/S0076-6879(09)63027-5

Boddicker, J. D., & Jones, B. D. (2004). Lon Protease Activity Causes Down-Regulation of *Salmonella* Pathogenicity Island 1 Invasion Gene Expression after Infection of Epithelial Cells. *Infection and Immunity*, *72*(4), 2002-2013. https://doi.org/10.1128/IAI.72.4.2002-2013.2004

Bond, J. S. (2019). Proteases : History, discovery, and roles in health and disease. *Journal of Biological Chemistry*, *294*(5), 1643-1651. https://doi.org/10.1074/jbc.TM118.004156

Borissenko, L., & Groll, M. (2005). Crystal structure of TET protease reveals complementary protein degradation pathways in prokaryotes. *Journal of Molecular Biology*, *346*(5), Article 5. https://doi.org/10.1016/j.jmb.2004.12.056

Botos, I., Melnikov, E. E., Cherry, S., Tropea, J. E., Khalatova, A. G., Rasulova, F., Dauter, Z., Maurizi, M. R., Rotanova, T. V., Wlodawer, A., & Gustchina, A. (2004). The Catalytic Domain of Escherichia coli Lon Protease Has a Unique Fold and a Ser-Lys Dyad in the Active Site. *Journal of Biological Chemistry*, *279*(9), 8140-8148. https://doi.org/10.1074/jbc.M312243200

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, *72*(1-2), 248-254. https://doi.org/10.1016/0003-2697(76)90527-3

Brandstetter, H., Kim, J.-S., Groll, M., Göttig, P., & Huber, R. (2002). Structural Basis for the Processive Protein Degradation by Tricorn Protease. *Biological Chemistry*, *383*(7-8). https://doi.org/10.1515/BC.2002.127

Brandstetter, H., Kim, J.-S., Groll, M., & Huber, R. (2001). *Crystal structure of the tricorn protease reveals a protein disassembly line* (6862). *414*(6862), Article 6862. https://doi.org/10.1038/35106609

Brito Querido, J., Díaz-López, I., & Ramakrishnan, V. (2024). The molecular basis of translation initiation and its regulation in eukaryotes. *Nature Reviews Molecular Cell Biology*, *25*(3), 168-186. https://doi.org/10.1038/s41580-023-00624-9

Brochier-Armanet, C., Boussau, B., Gribaldo, S., & Forterre, P. (2008). Mesophilic crenarchaeota : Proposal for a third archaeal phylum, the Thaumarchaeota. *Nature Reviews Microbiology*, *6*(3), 245-252. https://doi.org/10.1038/nrmicro1852

Brochier-Armanet, C., Forterre, P., & Gribaldo, S. (2011). Phylogeny and evolution of the Archaea : One hundred genomes later. *Current Opinion in Microbiology*, *14*(3), 274-281. https://doi.org/10.1016/j.mib.2011.04.015

Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J.-F., Adams, M. D., Reich, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K. G., Merrick, J. M., … Venter, J. C. (1996). Complete Genome Sequence of the Methanogenic Archaeon, ***Methanococcus jannaschii*** . *Science*, *273*(5278), 1058-1073. https://doi.org/10.1126/science.273.5278.1058

Bulzu, P.-A., Andrei, A.-Ş., Salcher, M. M., Mehrshad, M., Inoue, K., Kandori, H., Beja, O., Ghai, R., & Banciu, H. L. (2019). Casting light on Asgardarchaeota metabolism in a sunlit microoxic niche. *Nature Microbiology*, *4*(7), 1129-1137. https://doi.org/10.1038/s41564-019-0404-y

Cai, M., Liu, Y., Yin, X., Zhou, Z., Friedrich, M. W., Richter-Heitmann, T., Nimzyk, R., Kulkarni, A., Wang, X., Li, W., Pan, J., Yang, Y., Gu, J.-D., & Li, M. (2020). Diverse Asgard archaea including the novel phylum Gerdarchaeota participate in organic matter degradation. *Science China Life Sciences*, *63*(6), 886-897. https://doi.org/10.1007/s11427-020-1679-1

Cangelosi, G. A., & Meschke, J. S. (2014). Dead or Alive : Molecular Assessment of Microbial Viability. *Applied and Environmental Microbiology*, *80*(19), 5884-5891. https://doi.org/10.1128/AEM.01763-14

Carré, L., Zaccai, G., Delfosse, X., Girard, E., & Franzetti, B. (2022). Relevance of Earth-Bound Extremophiles in the Search for Extraterrestrial Life. *Astrobiology*, *22*(3), 322-367. https://doi.org/10.1089/ast.2021.0033

Cavicchioli, R. (2011). Archaea—Timeline of the third domain. *Nature Reviews Microbiology*, *9*(1), 51-61. https://doi.org/10.1038/nrmicro2482

Cha, S.-S., An, Y. J., Lee, C. R., Lee, H. S., Kim, Y.-G., Kim, S. J., Kwon, K. K., De Donatis, G. M., Lee, J.-H., Maurizi, M. R., & Kang, S. G. (2010). Crystal structure of Lon protease : Molecular architecture of gated entry to a sequestered degradation chamber. *The EMBO Journal*, *29*(20), 3520-3530. https://doi.org/10.1038/emboj.2010.226

Chaikuad, A., Pilka, E. S., De Riso, A., von Delft, F., Kavanagh, K. L., Vénien-Bryan, C., Oppermann, U., & Yue, W. W. (2012). Structure of human aspartyl aminopeptidase complexed with substrate analogue : Insight into catalytic mechanism, substrate specificity and M18 peptidase family. *BMC Structural Biology*, *12*(1), Article 1. https://doi.org/10.1186/1472-6807-12-14

Chandu, D., & Nandi, D. (2004). Comparative genomics and functional roles of the ATP-dependent proteases Lon and Clp during cytosolic protein degradation. *Research in Microbiology*, *155*(9), 710-719. https://doi.org/10.1016/j.resmic.2004.06.003

Chevrier, B., Schalk, C., D'Orchymont, H., Rondeau, J.-M., Moras, D., & Tarnus, C. (1994). Crystal structure of Aeromonas proteolytica aminopeptidase : A prototypical member of the co-catalytic zinc enzyme family. *Structure*, *2*(4), 283-291. https://doi.org/10.1016/S0969-2126(00)00030-7

Ciechanover, A. (2017). Intracellular protein degradation : From a vague idea thru the lysosome and the ubiquitin-proteasome system and onto human diseases and drug targeting. *Best Practice & Research Clinical Haematology*, *30*(4), 341-355. https://doi.org/10.1016/j.beha.2017.09.001

Colombo, M., Girard, E., & Franzetti, B. (2016). Tuned by metals : The TET peptidase activity is controlled by 3 metal binding sites. *Scientific Reports*, *6*(1), Article 1. https://doi.org/10.1038/srep20876

Criscuolo, A., & Gribaldo, S. (2010). BMGE (Block Mapping and Gathering with Entropy) : A new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evolutionary Biology*, *10*(1), 210-230. https://doi.org/10.1186/1471-2148-10-210

Da Cunha, V., Gaïa, M., & Forterre, P. (2022). The expanding Asgard archaea and their elusive relationships with Eukarya. *mLife*, *1*(1), 3-12. https://doi.org/10.1002/mlf2.12012

Da Cunha, V., Gaia, M., Gadelle, D., Nasir, A., & Forterre, P. (2017). Lokiarchaea are close relatives of Euryarchaeota, not bridging the gap between prokaryotes and eukaryotes. *PLOS Genetics*, *13*(6), e1006810. https://doi.org/10.1371/journal.pgen.1006810

Da Cunha, V., Gaia, M., Nasir, A., & Forterre, P. (2018). Asgard archaea do not close the debate about the universal tree of life topology. *PLOS Genetics*, *14*(3), e1007215. https://doi.org/10.1371/journal.pgen.1007215

Dacks, J. B., & Field, M. C. (2007). Evolution of the eukaryotic membrane-trafficking system : Origin, tempo and mode. *Journal of Cell Science*, *120*(17), 2977-2985. https://doi.org/10.1242/jcs.013250

DeLong, E. F., & Pace, N. R. (2001). Environmental diversity of bacteria and archaea. *Systematic Biology*, *50*(4), 470-478.

Djuranovic, S., Rockel, B., Lupas, A. N., & Martin, J. (2006). Characterization of AMA, a new AAA protein from Archaeoglobus and methanogenic archaea. *Journal of Structural Biology*, *156*(1), 130-138. https://doi.org/10.1016/j.jsb.2006.03.010

Doi, E., Shibata, D., & Matoba, T. (1981). Modified colorimetric ninhydrin methods for peptidase assay. *Analytical Biochemistry*, *118*(1), 173-184. https://doi.org/10.1016/0003-2697(81)90175-5

Dombrowski, N., Lee, J.-H., Williams, T. A., Offre, P., & Spang, A. (2019). Genomic diversity, lifestyles and evolutionary origins of DPANN archaea. *FEMS Microbiology Letters*, *366*(2). https://doi.org/10.1093/femsle/fnz008

Dopp, E., Hartmann, L. M., Florea, A.-M., Von Recklinghausen, U., Pieper, R., Shokouhi, B., Rettenmeier, A. W., Hirner, A. V., & Obe, G. (2004). Uptake of inorganic and organic derivatives of arsenic associated with induced cytotoxic and genotoxic effects in Chinese hamster ovary (CHO) cells. *Toxicology and Applied Pharmacology*, *201*(2), 156-165. https://doi.org/10.1016/j.taap.2004.05.017

Durá, M. A., & Franzetti, B. (2013). PhTET1 aminopeptidase. In *Handbook of Proteolytic Enzymes* (Vol. 2, p. 1638-1645). Elsevier. https://doi.org/10.1016/B978-0-12-382219-2.00372-0

Durá, M. A., Receveur-Brechot, V., Andrieu, J.-P., Ebel, C., Schoehn, G., Roussel, A., & Franzetti, B. (2005). Characterization of a TET-like aminopeptidase complex from the hyperthermophilic archaeon Pyrococcus horikoshii. *Biochemistry*, *44*(9), Article 9. https://doi.org/10.1021/bi047736j

Durá, M. A., Rosenbaum, E., Larabi, A., Gabel, F., Vellieux, F. M. D., & 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. *Molecular Microbiology*, *72*(1), Article 1. https://doi.org/10.1111/j.1365-2958.2009.06600.x

Dutoit, R. (2021). *The paradigm of self-compartmentalized M42 aminopeptidases : Insight into their oligomerization, substrate specificities, and physiological function* [Université libre de Bruxelles, Faculté des Sciences – Biologie Moléculaire]. http://hdl.handle.net/2013/ULB-DIPOT:oai:dipot.ulb.ac.be:2013/314337

Dutoit, R., Brandt, N., Legrain, C., & Bauvois, C. (2012). Functional characterization of two M42 aminopeptidases erroneously annotated as cellulases. *PLoS ONE*, *7*(11), Article 11. https://doi.org/10.1371/journal.pone.0050639

Dutoit, R., Brandt, N., Van Gompel, T., Van Elder, D., Van Dyck, J., Sobott, F., & Droogmans, L. (2020). M42 aminopeptidase catalytic site : The structural and functional role of a strictly conserved aspartate residue. *Proteins: Structure, Function, and Bioinformatics*, *88*(12), Article 12. https://doi.org/10.1002/prot.25982

Dutoit, R., Van Gompel, T., Brandt, N., Van Elder, D., Van Dyck, J., Sobott, F., & Droogmans, L. (2019). How metal cofactors drive dimer–dodecamer transition of the M42 aminopeptidase TmPep1050 of Thermotoga maritima. *Journal of Biological Chemistry*, *294*(47), Article 47. https://doi.org/10.1074/jbc.RA119.009281

Elkins, J. G., Podar, M., Graham, D. E., Makarova, K. S., Wolf, Y., Randau, L., Hedlund, B. P., Brochier-Armanet, C., Kunin, V., Anderson, I., Lapidus, A., Goltsman, E., Barry, K., Koonin, E. V., Hugenholtz, P., Kyrpides, N., Wanner, G., Richardson, P., Keller, M., & Stetter, K. O. (2008). A korarchaeal genome reveals insights into the evolution of the Archaea. *Proceedings of the National Academy of Sciences*, *105*(23), 8102-8107. https://doi.org/10.1073/pnas.0801980105

Eloe-Fadrosh, E. A., Ivanova, N. N., Woyke, T., & Kyrpides, N. C. (2016). Metagenomics uncovers gaps in amplicon-based detection of microbial diversity. *Nature Microbiology*, *1*(4), 15032. https://doi.org/10.1038/nmicrobiol.2015.32

Eme, L., Tamarit, D., Caceres, E. F., Stairs, C. W., De Anda, V., Schön, M. E., Seitz, K. W., Dombrowski, N., Lewis, W. H., Homa, F., Saw, J. H., Lombard, J., Nunoura, T., Li, W.-J., Hua, Z.-S., Chen, L.-X., Banfield, J. F., John, E. S., Reysenbach, A.-L., … Ettema, T. J. G. (2023). Inference and reconstruction of the heimdallarchaeial ancestry of eukaryotes. *Nature*, *618*(7967), 992-999. https://doi.org/10.1038/s41586-023-06186-2

Fatima, N. I., Fazili, K. M., & Bhat, N. H. (2022). Proteolysis dependent cell cycle regulation in Caulobacter crescentus. *Cell Division*, *17*(1), 3. https://doi.org/10.1186/s13008-022-00078-z

Feng, X., Wang, Y., Zubin, R., & Wang, F. (2019). Core Metabolic Features and Hot Origin of Bathyarchaeota. *Engineering*, *5*(3), 498-504. https://doi.org/10.1016/j.eng.2019.01.011

Ferrari, A., Brusa, T., Rutili, A., Canzi, E., & Biavati, B. (1994). Isolation and characterization of Methanobrevibacter oralis sp. Nov. *Current Microbiology*, *29*(1), 7-12. https://doi.org/10.1007/BF01570184

Fiala, G., & Stetter, K. O. (1986). Pyrococcus furiosus sp. Nov. Represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100□C. *Archives of Microbiology*, *145*(1), 56-61. https://doi.org/10.1007/BF00413027

Forouzan, D., Ammelburg, M., Hobel, C. F., Ströh, L. J., Sessler, N., Martin, J., & Lupas, A. N. (2012). The Archaeal Proteasome Is Regulated by a Network of AAA ATPases. *Journal of Biological Chemistry*, *287*(46), 39254-39262. https://doi.org/10.1074/jbc.M112.386458

Fox, G. E., Magrum, L. J., Balch, W. E., Wolfe, R. S., & Woese, C. R. (1977). Classification of methanogenic bacteria by 16S ribosomal RNA characterization. *Proceedings of the National Academy of Sciences*, *74*(10), 4537-4541. https://doi.org/10.1073/pnas.74.10.4537

Franzetti, B., Schoehn, G., Hernandez, J. F., Jaquinod, M., Ruigrok, R. W. H., & Zaccai, G. (2002). Tetrahedral aminopeptidase : A novel large protease complex from archaea. *The EMBO Journal*, *21*(9), Article 9. https://doi.org/10.1093/emboj/21.9.2132

Freydank, A., Brandt, W., & Dräger, B. (2008). Protein structure modeling indicates hexahistidine-tag interference with enzyme activity. *Proteins: Structure, Function, and Bioinformatics*, *72*(1), 173-183. https://doi.org/10.1002/prot.21905

Fujino, Y., Goda, S., Suematsu, Y., & Doi, K. (2020). Development of a new gene expression vector for Thermus thermophilus using a silica-inducible promoter. *Microbial Cell Factories*, *19*(1), 126-137. https://doi.org/10.1186/s12934-020-01385-2

Garcia, J.-L., Ollivier, B., & Whitman, W. B. (2006). The Order Methanomicrobiales. In M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, & E. Stackebrandt (Éds.), *The Prokaryotes* (p. 208-230). Springer New York. https://doi.org/10.1007/0-387-30743-5_10

Garcia, P. S., Gribaldo, S., & Borrel, G. (2022). Diversity and Evolution of Methane-Related Pathways in Archaea. Annual Review of Microbiology, 76(1), 727-755. https://doi.org/10.1146/annurev-micro-041020-024935

Gauto, D. F., Macek, P., Malinverni, D., Fraga, H., Paloni, M., Sučec, I., Hessel, A., Bustamante, J. P., Barducci, A., & Schanda, P. (2022). Functional control of a 0.5 MDa TET aminopeptidase by a flexible loop revealed by MAS NMR. *Nature Communications*, *13*(1), 1927. https://doi.org/10.1038/s41467-022-29423-0

Gibney, E. R., & Nolan, C. M. (2010). Epigenetics and gene expression. *Heredity*, *105*(1), 4-13. https://doi.org/10.1038/hdy.2010.54

Goff, S. A., Casson, L. P., & Goldberg, A. L. (1984). Heat shock regulatory gene htpR influences rates of protein degradation and expression of the lon gene in Escherichia coli. *Proceedings of the National Academy of Sciences*, *81*(21), 6647-6651. https://doi.org/10.1073/pnas.81.21.6647

Goldberg, A. L. (2003). Protein degradation and protection against misfolded or damaged proteins. *Nature*, *426*(6968), 895-899. https://doi.org/10.1038/nature02263

Gottesman, S., Maurizi, M. R., & Wickner, S. (1997). Regulatory Subunits of Energy-Dependent Proteases. *Cell*, *91*(4), 435-438. https://doi.org/10.1016/S0092-8674(00)80428-6

Gottesman, S., Roche, E., Zhou, Y., & Sauer, R. T. (1998). The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes & Development*, *12*(9), 1338-1347. https://doi.org/10.1101/gad.12.9.1338

Groll, M., Ditzel, L., Löwe, J., Stock, D., Bochtler, M., Bartunik, H. D., & Huber, R. (1997). Structure of 20S proteasome from yeast at 2.4Å resolution. *Nature*, *386*(6624), 463-471. https://doi.org/10.1038/386463a0

Groll, M., Glickman, M. H., Finley, D., Bajorek, M., Köhler, A., Moroder, L., Rubin, D. M., & Huber, R. (2000). A gated channel into the proteasome core particle. *Nature Structural Biology*, *7*(11), 1062-1067. https://doi.org/10.1038/80992

Guy, L., & Ettema, T. J. G. (2011). The archaeal 'TACK' superphylum and the origin of eukaryotes. *Trends in Microbiology*, *19*(12), 580-587. https://doi.org/10.1016/j.tim.2011.09.002

Handelsman, J. (2004). Metagenomics : Application of Genomics to Uncultured Microorganisms. *Microbiology and Molecular Biology Reviews*, *68*(4), 669-685. https://doi.org/10.1128/MMBR.68.4.669-685.2004

Hanlon, D. P., Watt, D. S., & Westhead, E. W. (1966). The interaction of divalent metal ions with tris buffer in dilute solution. *Analytical Biochemistry*, *16*(2), 225-233. https://doi.org/10.1016/0003-2697(66)90150-3

Hartley, B. S. (1960). Proteolytic enzymes. *Annual Review of Biochemistry*, *29*(1), 45-72. https://doi.org/10.1146/annurev.bi.29.070160.000401

Henneman, B., Van Emmerik, C., Van Ingen, H., & Dame, R. T. (2018). Structure and function of archaeal histones. *PLOS Genetics*, *14*(9), e1007582. https://doi.org/10.1371/journal.pgen.1007582

Hepowit, N. L., De Vera, I. M. S., Cao, S., Fu, X., Wu, Y., Uthandi, S., Chavarria, N. E., Englert, M., Su, D., Söll, D., Kojetin, D. J., & Maupin-Furlow, J. A. (2016). Mechanistic insight into protein modification and sulfur mobilization activities of noncanonical E1 and associated ubiquitin-like proteins of Archaea. *The FEBS Journal*, *283*(19), 3567-3586. https://doi.org/10.1111/febs.13819

Hillary, R. F., & FitzGerald, U. (2018). A lifetime of stress : ATF6 in development and homeostasis. *Journal of Biomedical Science*, *25*(1), 48. https://doi.org/10.1186/s12929-018-0453-1

Holm, L., & Rosenström, P. (2010). Dali server : Conservation mapping in 3D. *Nucleic Acids Research*, *38*(suppl_2), W545-W549. https://doi.org/10.1093/nar/gkq366

Hou, J., Wang, Y., Zhu, P., Yang, N., Liang, L., Yu, T., Niu, M., Konhauser, K., Woodcroft, B. J., & Wang, F. (2023). Taxonomic and carbon metabolic diversification of Bathyarchaeia during its coevolution history with early Earth surface environment. *Science Advances*, *9*(27), 1-17. https://doi.org/10.1126/sciadv.adf5069

Huber, H., Hohn, M. J., Rachel, R., Fuchs, T., Wimmer, V. C., & Stetter, K. O. (2002). A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature*, *417*(6884), 63-67. https://doi.org/10.1038/417063a

Huber, J. A., Butterfield, D. A., & Baross, J. A. (2002). Temporal Changes in Archaeal Diversity and Chemistry in a Mid-Ocean Ridge Subseafloor Habitat. *Applied and Environmental Microbiology*, *68*(4), 1585-1594. https://doi.org/10.1128/AEM.68.4.1585-1594.2002

Huber, R., Dyba, D., Huber, H., Burggraf, S., & Rachel, R. (1998). Sulfur-inhibited Thermosphaera aggregans sp. Nov., a new genus of hyperthermophilic archaea isolated after its prediction from environmentally derived 16S rRNA sequences. *International Journal of Systematic Bacteriology*, *48*(1), 31-38. https://doi.org/10.1099/00207713-48-1-31

Humbard, M. A., Miranda, H. V., Lim, J.-M., Krause, D. J., Pritz, J. R., Zhou, G., Chen, S., Wells, L., & Maupin-Furlow, J. A. (2010). Ubiquitin-like small archaeal modifier proteins (SAMPs) in Haloferax volcanii. *Nature*, *463*(7277), 54-60. https://doi.org/10.1038/nature08659

Ibrahim, Z., Martel, A., Moulin, M., Kim, H. S., Härtlein, M., Franzetti, B., & Gabel, F. (2017). Time-resolved neutron scattering provides new insight into protein substrate processing by a AAA+ unfoldase. *Scientific Reports*, *7*(1), 40948. https://doi.org/10.1038/srep40948

Imachi, H., Nobu, M. K., Nakahara, N., Morono, Y., Ogawara, M., Takaki, Y., Takano, Y., Uematsu, K., Ikuta, T., Ito, M., Matsui, Y., Miyazaki, M., Murata, K., Saito, Y., Sakai, S., Song, C., Tasumi, E., Yamanaka, Y., Yamaguchi, T., … Takai, K. (2020). Isolation of an archaeon at the prokaryote–eukaryote interface. *Nature*, *577*(7791), 519-525. https://doi.org/10.1038/s41586-019-1916-6

Itoh, T. (2002). Vulcanisaeta distributa gen. Nov., sp. Nov., and Vulcanisaeta souniana sp. Nov., novel hyperthermophilic, rod-shaped crenarchaeotes isolated from hot springs in Japan. *INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY*, *52*(4), 1097-1104. https://doi.org/10.1099/ijs.0.02152-0

Jaenicke, R., & Böhm, G. (1998). The stability of proteins in extreme environments. *Current Opinion in Structural Biology*, *8*(6), 738-748. https://doi.org/10.1016/S0959-440X(98)80094-8

Jenal, U. (1998). An essential protease involved in bacterial cell-cycle control. *The EMBO Journal*, *17*(19), 5658-5669. https://doi.org/10.1093/emboj/17.19.5658

Jia, B., Lee, S., Pham, B. P., Kwack, J. M., Jin, H., Li, J., Wang, Y., & Cheong, G.-W. (2011). Biochemical Characterization of Deblocking Aminopeptidases from the Hyperthermophilic Archaeon *Thermococcus kodakarensis* KOD1. *Bioscience, Biotechnology, and Biochemistry*, *75*(6), 1160-1166. https://doi.org/10.1271/bbb.110114

Johnson, L. S., Eddy, S. R., & Portugaly, E. (2010). Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics*, *11*(1), 431-438. https://doi.org/10.1186/1471-2105-11-431

Kachlany, S. C., Planet, P. J., Bhattacharjee, M. K., Kollia, E., DeSalle, R., Fine, D. H., & Figurski, D. H. (2000). Nonspecific Adherence by *Actinobacillus actinomycetemcomitans* Requires Genes Widespread in *Bacteria* and *Archaea*. *Journal of Bacteriology*, *182*(21), 6169-6176. https://doi.org/10.1128/JB.182.21.6169-6176.2000

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder : Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587-589. https://doi.org/10.1038/nmeth.4285

Kamp, R. M., Tsunasawa, S., & Hirano, H. (1998). Application of new deblocking aminopeptidase from Pyrococcus furiosis for microsequence analysis of blocked proteins. *Journal of Protein Chemistry*, *17*(6), 512-513.

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7 : Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30*(4), 772-780. https://doi.org/10.1093/molbev/mst010

Khomyakova, M. A., Merkel, A. Y., Mamiy, D. D., Klyukina, A. A., & Slobodkin, A. I. (2023). Phenotypic and genomic characterization of Bathyarchaeum tardum gen. Nov., sp. Nov., a cultivated representative of the archaeal class Bathyarchaeia. *Frontiers in Microbiology*, *14*, 1214631-1214645. https://doi.org/10.3389/fmicb.2023.1214631

Kim, D., San, B. H., Moh, S. H., Park, H., Kim, D. Y., Lee, S., & Kim, K. K. (2010). Structural basis for the substrate specificity of PepA from Streptococcus pneumoniae, a dodecameric tetrahedral protease. *Biochemical and Biophysical Research Communications*, *391*(1), 431-436. https://doi.org/10.1016/j.bbrc.2009.11.075

Kisselev, A. F., Akopian, T. N., Woo, K. M., & Goldberg, A. L. (1999). The Sizes of Peptides Generated from Protein by Mammalian 26 and 20 S Proteasomes. *Journal of Biological Chemistry*, *274*(6), 3363-3371. https://doi.org/10.1074/jbc.274.6.3363

Kobayashi, T., Romaniec, M. P. M., Barker, P. J., Gerngross, U. T., & Demain, A. L. (1993). Nucleotide sequence of gene celM encoding a new endoglucanase (CelM) of Clostridium thermocellum and purification of the enzyme. *Journal of Fermentation and Bioengineering*, *76*(4), 251-256. https://doi.org/10.1016/0922-338X(93)90189-F

Koonin, E. V. (2009). Evolution of genome architecture. *The International Journal of Biochemistry & Cell Biology*, *41*(2), 298-306. https://doi.org/10.1016/j.biocel.2008.09.015

Koskinen, K., Pausan, M. R., Perras, A. K., Beck, M., Bang, C., Mora, M., Schilhabel, A., Schmitz, R., & Moissl-Eichinger, C. (2017). First Insights into the Diverse Human Archaeome : Specific Detection of Archaea in the Gastrointestinal Tract, Lung, and Nose and on Skin. *mBio*, *8*(6), e00824-17. https://doi.org/10.1128/mBio.00824-17

Krissinel, E., & Henrick, K. (2007). Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology*, *372*(3), 774-797. https://doi.org/10.1016/j.jmb.2007.05.022

Kulik, E. M., Sandmeier, H., Hinni, K., & Meyer, J. (2001). Identification of archaeal rDNA from subgingival dental plaque by PCR amplification and sequence analysis. *FEMS Microbiology Letters*, *196*(2), 129-133. https://doi.org/10.1111/j.1574-6968.2001.tb10553.x

Kumaki, Y., Ogawa, M., Hirano, T., Yoshikawa, K., Iwasawa, N., Yagi, T., Hakamata, W., Oku, T., & Nishio, T. (2011). Family M42 aminopeptidase from the syntrophic bacterium Symbiobacterium thermophilum : Characterization using recombinant protein. *Journal of Bioscience and Bioengineering*, *111*(2), Article 2. https://doi.org/10.1016/j.jbiosc.2010.09.012

Kuroda, A. (2006). A Polyphosphate-Lon Protease Complex in the Adaptation of *Escherichia coli* to Amino Acid Starvation. *Bioscience, Biotechnology, and Biochemistry*, *70*(2), 325-331. https://doi.org/10.1271/bbb.70.325

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., … Morgan, M. J. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860-921. https://doi.org/10.1038/35057062

Lane, N., & Martin, W. (2010). The energetics of genome complexity. *Nature*, *467*(7318), 929-934. https://doi.org/10.1038/nature09486

Larrinaga, G., Perez, I., Ariz, U., Sanz, B., Beitia, M., Errarte, P., Etxezarraga, C., Candenas, M. L., Pinto, F. M., & López, J. I. (2013). Clinical impact of aspartyl aminopeptidase expression and activity in colorectal cancer. *Translational Research*, *162*(5), 297-308. https://doi.org/10.1016/j.trsl.2013.07.010

Larsson, A. (2014). AliView : A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, *30*(22), 3276-3278. https://doi.org/10.1093/bioinformatics/btu531

Le Guellec, S., Leroy, E., Courtine, D., Godfroy, A., & Roussel, E. G. (2021). H2-dependent formate production by hyperthermophilic *Thermococcales* : An alternative to sulfur reduction for reducing-equivalents disposal. *The ISME Journal*, *15*(12), 3423-3436. https://doi.org/10.1038/s41396-021-01020-x

Lee, H. S., Cho, Y., Kim, Y. J., Nam, K., Lee, J.-H., & Kang, S. G. (2007). Biochemical characterization of deblocking aminopeptidase from hyperthermophilic archaeon Thermococcus onnurineus NA1. *Journal of Bioscience and Bioengineering*, *104*(3), 188-194. https://doi.org/10.1263/jbb.104.188

Lee, H. S., Kang, S. G., Bae, S. S., Lim, J. K., Cho, Y., Kim, Y. J., Jeon, J. H., Cha, S.-S., Kwon, K. K., Kim, H.-T., Park, C.-J., Lee, H.-W., Kim, S. I., Chun, J., Colwell, R. R., Kim, S.-J., & Lee, J.-H. (2008). The Complete Genome Sequence of *Thermococcus onnurineus* NA1 Reveals a Mixed Heterotrophic and Carboxydotrophic Metabolism. *Journal of Bacteriology*, *190*(22), 7491-7499. https://doi.org/10.1128/JB.00746-08

Lee, J. M., Hammarén, H. M., Savitski, M. M., & Baek, S. H. (2023). Control of protein stability by post-translational modifications. *Nature Communications*, *14*(1), 201. https://doi.org/10.1038/s41467-023-35795-8

Lee, T. I., & Young, R. A. (2013). Transcriptional Regulation and Its Misregulation in Disease. *Cell*, *152*(6), 1237-1251. https://doi.org/10.1016/j.cell.2013.02.014

Lee, Y. G., Leem, S.-H., Chung, Y.-H., & Kim, S. I. (2012). Characterization of thermostable deblocking aminopeptidases of archaeon Thermococcus onnurineus NA1 by proteomic and biochemical approaches. *Journal of Microbiology*, *50*(5), 792-797. https://doi.org/10.1007/s12275-012-2461-2

Letunic, I., & Bork, P. (2019). Interactive Tree Of Life (iTOL) v4 : Recent updates and new developments. *Nucleic Acids Research*, *47*(W1), 256-259. https://doi.org/10.1093/nar/gkz239

Li, M., Rasulova, F., Melnikov, E. E., Rotanova, T. V., Gustchina, A., Maurizi, M. R., & Wlodawer, A. (2005). Crystal structure of the N-terminal domain of *E. coli* Lon protease. *Protein Science*, *14*(11), 2895-2900. https://doi.org/10.1110/ps.051736805

Liao, J.-H., Ihara, K., Kuo, C.-I., Huang, K.-F., Wakatsuki, S., Wu, S.-H., & Chang, C.-I. (2013). Structures of an ATP-independent Lon-like protease and its complexes with covalent inhibitors. *Acta Crystallographica Section D Biological Crystallography*, *69*(8), 1395-1402. https://doi.org/10.1107/S0907444913008214

Liao, J.-H., Kuo, C.-I., Huang, Y.-Y., Lin, Y.-C., Lin, Y.-C., Yang, C.-Y., Wu, W.-L., Chang, W.-H., Liaw, Y.-C., Lin, L.-H., Chang, C.-I., & Wu, S.-H. (2012). A Lon-Like Protease with No ATP-Powered Unfolding Activity. *PLoS ONE*, *7*(7), e40226. https://doi.org/10.1371/journal.pone.0040226

Liu, Y., Makarova, K. S., Huang, W.-C., Wolf, Y. I., Nikolskaya, A. N., Zhang, X., Cai, M., Zhang, C.-J., Xu, W., Luo, Z., Cheng, L., Koonin, E. V., & Li, M. (2021). Expanded diversity of Asgard archaea and their relationships with eukaryotes. *Nature*, *593*(7860), 553-557. https://doi.org/10.1038/s41586-021-03494-3

Lloyd, K. G., Schreiber, L., Petersen, D. G., Kjeldsen, K. U., Lever, M. A., Steen, A. D., Stepanauskas, R., Richter, M., Kleindienst, S., Lenk, S., Schramm, A., & Jørgensen, B. B. (2013). Predominant archaea in marine sediments degrade detrital proteins. *Nature*, *496*(7444), 215-218. https://doi.org/10.1038/nature12033

López-García, P., & Moreira, D. (2015). Open Questions on the Origin of Eukaryotes. *Trends in Ecology & Evolution*, *30*(11), 697-708. https://doi.org/10.1016/j.tree.2015.09.005

Löwe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W., & Huber, R. (1995). Crystal Structure of the 20 *S* Proteasome from the Archaeon *T. acidophilum* at 3.4 Å Resolution. *Science*, *268*(5210), 533-539. https://doi.org/10.1126/science.7725097

Lu, Z., Xia, R., Zhang, S., Pan, J., Liu, Y., Wolf, Y. I., Koonin, E. V., & Li, M. (2024). Evolution of optimal growth temperature in Asgard archaea inferred from the temperature dependence of GDP binding to EF-1A. *Nature Communications*, *15*(1), 515-521. https://doi.org/10.1038/s41467-024-44806-1

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, *389*(6648), 251-260. https://doi.org/10.1038/38444

Lupas, A., Flanagan, J. M., Tamura, T., & Baumeister, W. (1997). Self-compartmentalizing proteases. *Trends in Biochemical Sciences*, *22*(10), 399-404. https://doi.org/10.1016/S0968-0004(97)01117-1

Macek, P., Kerfah, R., Erba, E. B., Crublet, E., Moriscot, C., Schoehn, G., Amero, C., & Boisbouvier, J. (2017). Unraveling self-assembly pathways of the 468-kDa proteolytic machine TET2. *Science Advances*, *3*(4), e1601601. https://doi.org/10.1126/sciadv.1601601

Mackinnon, E., & Stone, S. L. (2022). The Ubiquitin Proteasome System and Nutrient Stress Response. *Frontiers in Plant Science*, *13*, 867419. https://doi.org/10.3389/fpls.2022.867419

Maffei, B., Francetic, O., & Subtil, A. (2017). Tracking Proteins Secreted by Bacteria : What's in the Toolbox? *Frontiers in Cellular and Infection Microbiology*, *7*, 221. https://doi.org/10.3389/fcimb.2017.00221

Majorek, K. A., Kuhn, M. L., Chruszcz, M., Anderson, W. F., & Minor, W. (2014). Double trouble—Buffer selection and H is-tag presence may be responsible for nonreproducibility of biomedical experiments. *Protein Science*, *23*(10), 1359-1368. https://doi.org/10.1002/pro.2520

Majumder, P., Rudack, T., Beck, F., Danev, R., Pfeifer, G., Nagy, I., & Baumeister, W. (2019). Cryo-EM structures of the archaeal PAN-proteasome reveal an around-the-ring ATPase cycle. *Proceedings of the National Academy of Sciences*, *116*(2), 534-539. https://doi.org/10.1073/pnas.1817752116

Makarova, K. S., Koonin, E. V., & Albers, S.-V. (2016). Diversity and Evolution of Type IV pili Systems in Archaea. *Frontiers in Microbiology*, *7*. https://doi.org/10.3389/fmicb.2016.00667

Makarova, K. S., Krupovic, M., & Koonin, E. V. (2014). Evolution of replicative DNA polymerases in archaea and their contributions to the eukaryotic replication machinery. *Frontiers in Microbiology*, *5*. https://doi.org/10.3389/fmicb.2014.00354

Martin, W. F., Garg, S., & Zimorski, V. (2015). Endosymbiotic theories for eukaryote origin. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1678), 20140330. https://doi.org/10.1098/rstb.2014.0330

Martínez-Martos, J. M., Carrera-González, M. D. P., Dueñas, B., Mayas, M. D., García, M. J., & Ramírez-Expósito, M. J. (2011). Renin angiotensin system-regulating aminopeptidase activities in serum of pre- and postmenopausal women with breast cancer. *The Breast*, *20*(5), 444-447. https://doi.org/10.1016/j.breast.2011.04.008

Maupin-Furlow, J. (2012). Proteasomes and protein conjugation across domains of life. *Nature Reviews Microbiology*, *10*(2), 100-111. https://doi.org/10.1038/nrmicro2696

Maupin-Furlow, J. A., Gil, M. A., Humbard, M. A., Kirkland, P. A., Li, W., Reuter, C. J., & Wright, A. J. (2005). Archaeal proteasomes and other regulatory proteases. *Current Opinion in Microbiology*, *8*(6), 720-728. https://doi.org/10.1016/j.mib.2005.10.005

Maupin-Furlow, J. A., Humbard, M. A., Kirkland, P. A., Li, W., Reuter, C. J., Wright, A. J., & Zhou, G. (2006). Proteasomes from Structure to Function : Perspectives from Archaea. In *Current Topics in Developmental Biology* (Vol. 75, p. 125-169). Elsevier. https://doi.org/10.1016/S0070-2153(06)75005-0

Mayas, M. D., Ramírez-Expósito, M. J., Carrera, M. P., Cobo, M., & Martínez-Martos, J. M. (2012). Renin-angiotensin system-regulating aminopeptidases in tumor growth of rat C6 gliomas implanted at the subcutaneous region. *Anticancer Research*, *32*(9), 3675-3682.

McKay, L. J., Dlakić, M., Fields, M. W., Delmont, T. O., Eren, A. M., Jay, Z. J., Klingelsmith, K. B., Rusch, D. B., & Inskeep, W. P. (2019). Co-occurring genomic capacity for anaerobic methane and dissimilatory sulfur metabolisms discovered in the Korarchaeota. *Nature Microbiology*, *4*(4), 614-622. https://doi.org/10.1038/s41564-019-0362-4

Meyer, J., Michalke, K., Kouril, T., & Hensel, R. (2008). Volatilisation of metals and metalloids : An inherent feature of methanoarchaea? *Systematic and Applied Microbiology*, *31*(2), 81-87. https://doi.org/10.1016/j.syapm.2008.02.001

Miller, C. G., Strauch, K. L., Kukral, A. M., Miller, J. L., Wingfield, P. T., Mazzei, G. J., Werlen, R. C., Graber, P., & Movva, N. R. (1987). N-terminal methionine-specific peptidase in Salmonella typhimurium. *Proceedings of the National Academy of Sciences*, *84*(9), 2718-2722. https://doi.org/10.1073/pnas.84.9.2718

Miller, T. L., & Wolin, M. J. (1982). Enumeration of Methanobrevibacter smithii in human feces. *Archives of Microbiology*, *131*(1), 14-18. https://doi.org/10.1007/BF00451492

Miller, T. L., & Wolin, M. J. (1986). Methanogens in human and animal intestinal Tracts. *Systematic and Applied Microbiology*, *7*(2-3), 223-229. https://doi.org/10.1016/S0723-2020(86)80010-8

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2 : New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, *37*(5), 1530-1534. https://doi.org/10.1093/molbev/msaa015

Mohammadzadeh, R., Mahnert, A., Duller, S., & Moissl-Eichinger, C. (2022). Archaeal key-residents within the human microbiome : Characteristics, interactions and involvement in health and disease. *Current Opinion in Microbiology*, *67*, 102146. https://doi.org/10.1016/j.mib.2022.102146

Moore, S., & Stein, W. H. (1948). Photometric ninhydrin method for use in the chromatography of amino acids. *Journal of Biological Chemistry*, *176*(1), 367-388. https://doi.org/10.1016/S0021-9258(18)51034-6

Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B., & Worm, B. (2011). How Many Species Are There on Earth and in the Ocean? *PLoS Biology*, *9*(8), e1001127. https://doi.org/10.1371/journal.pbio.1001127

Moser, P., Roncari, G., & Zuber, H. (1970). THERMOPHILIC AMINOPEPTIDASES FROM BAC.STEAROTHERMOPHILUS : II. Aminopeptidasc I (AP I): physico-chemical properties; thermostability and activation; formation of the apoenzyme and subunits. *International Journal of Protein Research*, *2*(1-4), 191-207. https://doi.org/10.1111/j.1399-3011.1970.tb01677.x

Nakamura, A., Tsukada, T., Auer, S., Furuta, T., Wada, M., Koivula, A., Igarashi, K., & Samejima, M. (2013). The Tryptophan Residue at the Active Site Tunnel Entrance of Trichoderma reesei Cellobiohydrolase Cel7A Is Important for Initiation of Degradation of Crystalline Cellulose. *Journal of Biological Chemistry*, *288*(19), 13503-13510. https://doi.org/10.1074/jbc.M113.452623

Nercessian, O., Reysenbach, A., Prieur, D., & Jeanthon, C. (2003). Archaeal diversity associated with *in situ* samplers deployed on hydrothermal vents on the East Pacific Rise (13°N). *Environmental Microbiology*, *5*(6), 492-502. https://doi.org/10.1046/j.1462-2920.2003.00437.x

Ng'ong'a, F. (2017). In Silico Analysis of Occurrence of Tricorn Protease and Its Homologs. *Computational Biology and Bioinformatics*, *5*(3), 27. https://doi.org/10.11648/j.cbb.20170503.11

Notredame, C., Higgings, D. G., & Heringa, J. (2000). *T-Coffee : A novel method for fast and accurate multiple sequence alignment* (302). *302*, Article 302.

Nunoura, T., Takaki, Y., Kakuta, J., Nishi, S., Sugahara, J., Kazama, H., Chee, G.-J., Hattori, M., Kanai, A., Atomi, H., Takai, K., & Takami, H. (2011). Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic Acids Research*, *39*(8), 3204-3223. https://doi.org/10.1093/nar/gkq1228

O'Connor, E., & Shand, R. (2002). Halocins and sulfolobicins : The emerging story of archaeal protein and peptide antibiotics. *Journal of Industrial Microbiology & Biotechnology*, *28*(1), 23-31. https://doi.org/10.1038/sj/jim/7000190

Offre, P., Spang, A., & Schleper, C. (2013). Archaea in Biogeochemical Cycles. *Annual Review of Microbiology*, *67*(1), 437-457. https://doi.org/10.1146/annurev-micro-092412-155614

Olsen, G. J. (1994). Archaea, Archaea, every where. *Nature*, *371*(6499), 657-658. https://doi.org/10.1038/371657a0

Oren, A. (2006). The Order Halobacteriales. In M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, & E. Stackebrandt (Éds.), *The Prokaryotes* (p. 113-164). Springer New York. https://doi.org/10.1007/0-387-30743-5_8

Pan, Q., Shai, O., Lee, L. J., Frey, B. J., & Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics*, *40*(12), 1413-1415. https://doi.org/10.1038/ng.259

Park, S.-Y., Scranton, M. A., Stajich, J. E., Yee, A., & Walling, L. L. (2017). Chlorophyte aspartyl aminopeptidases : Ancient origins, expanded families, new locations, and secondary functions. *PLOS ONE*, *12*(10), 1-26. https://doi.org/10.1371/journal.pone.0185492

Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B. L., Salazar, G. A., Bileschi, M. L., Bork, P., Bridge, A., Colwell, L., Gough, J., Haft, D. H., Letunić, I., Marchler-Bauer, A., Mi, H., Natale, D. A., Orengo, C. A., Pandurangan, A. P., Rivoire, C., … Bateman, A. (2023). InterPro in 2022. *Nucleic Acids Research*, *51*(D1), D418-D427. https://doi.org/10.1093/nar/gkac993

Pereira, S. L., Grayling, R. A., Lurz, R., & Reeve, J. N. (1997). Archaeal nucleosomes. *Proceedings of the National Academy of Sciences*, *94*(23), 12633-12637. https://doi.org/10.1073/pnas.94.23.12633

Pérez-Sánchez, G., Leal-Guadarrama, L. I., Trelles, I., Pérez, N. O., & Medina-Rivero, E. (2011). High-level production of a recombinant Vibrio proteolyticus leucine aminopeptidase and its use for N-terminal methionine excision from interferon alpha-2b. *Process Biochemistry*, *46*(9), 1825-1830. https://doi.org/10.1016/j.procbio.2011.06.015

Petrova, T. E., Slutskaya, E. S., Boyko, K. M., Sokolova, O. S., Rakitina, T. V., Korzhenevskiy, D. A., Gorbacheva, M. A., Bezsudnova, E. Y., & Popov, V. O. (2015). Structure of the dodecamer of the aminopeptidase APDkam598 from the archaeon Desulfurococcus kamchatkensis. *Acta Crystallographica Section F Structural Biology Communications*, *71*(3), Article 3. https://doi.org/10.1107/S2053230X15000783

Pillot, G., Amin Ali, O., Davidson, S., Shintu, L., Godfroy, A., Combet-Blanc, Y., Bonin, P., & Liebgott, P.-P. (2021). Identification of enriched hyperthermophilic microbial communities from a deep-sea hydrothermal vent chimney under electrolithoautotrophic culture conditions. *Scientific Reports*, *11*(1), 14782-14793. https://doi.org/10.1038/s41598-021-94135-2

Pina, A. S., Batalha, Í. L., & Roque, A. C. A. (2014). Affinity Tags in Protein Purification and Peptide Enrichment : An Overview. In N. E. Labrou (Éd.), *Protein Downstream Processing* (Vol. 1129, p. 147-168). Humana Press. https://doi.org/10.1007/978-1-62703-977-2_14

Plaitakis, A., Kalef-Ezra, E., Kotzamani, D., Zaganas, I., & Spanaki, C. (2017). The Glutamate Dehydrogenase Pathway and Its Roles in Cell and Tissue Biology in Health and Disease. *Biology*, *6*(1), 11. https://doi.org/10.3390/biology6010011

Polgár, L. (2005). The catalytic triad of serine peptidases. *Cellular and Molecular Life Sciences*, *62*(19-20), 2161-2172. https://doi.org/10.1007/s00018-005-5160-x

Pouch, M., Cournoyer, B., & Baumeister, W. (2000). Characterization of the 20S proteasome from the actinomycete *Frankia*. *Molecular Microbiology*, *35*(2), 368-377. https://doi.org/10.1046/j.1365-2958.2000.01703.x

Rabl, J., Smith, D. M., Yu, Y., Chang, S.-C., Goldberg, A. L., & Cheng, Y. (2008). Mechanism of Gate Opening in the 20S Proteasome by the Proteasomal ATPases. *Molecular Cell*, *30*(3), 360-368. https://doi.org/10.1016/j.molcel.2008.03.004

Rawlings, N. D., Barrett, A. J., Thomas, P. D., Huang, X., Bateman, A., & 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 Research*, *46*(1), Article 1. https://doi.org/10.1093/nar/gkx1134

Rawlings, N. D., & Bateman, A. (2019). Origins of peptidases. *Biochimie*, *166*, 4-18. https://doi.org/10.1016/j.biochi.2019.07.026

Rawlings, N. D., & Salvesen, G. (2013). *Handbook of proteolytic enzymes* (3rd ed). Academic Press.

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J.-F., Darling, A., Malfatti, S., Swan, B. K., Gies, E. A., Dodsworth, J. A., Hedlund, B. P., Tsiamis, G., Sievert, S. M., Liu, W.-T., Eisen, J. A., Hallam, S. J., Kyrpides, N. C., Stepanauskas, R., … Woyke, T. (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature*, *499*(7459), 431-437. https://doi.org/10.1038/nature12352

Rivard, C. J., & Smith, P. H. (1982). Isolation and Characterization of a Thermophilic Marine Methanogenic Bacterium, Methanogenium thermophilicum sp. Nov. *International Journal of Systematic Bacteriology*, *32*(4), 430-436. https://doi.org/10.1099/00207713-32-4-430

Robinson, M. W., Buchtmann, K. A., Jenkins, C., Tacchi, J. L., Raymond, B. B. A., To, J., Roy Chowdhury, P., Woolley, L. K., Labbate, M., Turnbull, L., Whitchurch, C. B., Padula, M. P., & Djordjevic, S. P. (2013). MHJ_0125 is an M42 glutamyl aminopeptidase that moonlights as a multifunctional adhesin on the surface of *Mycoplasma hyopneumoniae*. *Open Biology*, *3*(4), 1-15. https://doi.org/10.1098/rsob.130017

Rockel, B., Jakana, J., Chiu, W., & Baumeister, W. (2002). Electron cryo-microscopy of VAT, the archaeal p97/CDC48 homologue from Thermoplasma acidophilum. Edited by D. Rees. *Journal of Molecular Biology*, *317*(5), 673-681. https://doi.org/10.1006/jmbi.2002.5448

Rodrigues-Oliveira, T., Wollweber, F., Ponce-Toledo, R. I., Xu, J., Rittmann, S. K.-M. R., Klingl, A., Pilhofer, M., & Schleper, C. (2023). Actin cytoskeleton and complex cell architecture in an Asgard archaeon. *Nature*, *613*(7943), 332-339. https://doi.org/10.1038/s41586-022-05550-y

Rohrwild, M., Coux, O., Huang, H. C., Moerschell, R. P., Yoo, S. J., Seol, J. H., Chung, C. H., & Goldberg, A. L. (1996). HslV-HslU : A novel ATP-dependent protease complex in Escherichia coli related to the eukaryotic proteasome. *Proceedings of the National Academy of Sciences*, *93*(12), 5808-5813. https://doi.org/10.1073/pnas.93.12.5808

Roncari, G., & Zuber, H. (1969). THERMOPHILIC AMINOPEPTIDASES FROM BACILLUS STEAROTHERMOPHILUS. I. ISOLATION, SPECIFICITY, AND GENERAL PROPERTIES OF THE THERMOSTABLE AMINOPEPTIDASE I *. *International Journal of Protein Research*, *1*(1-4), 45-61. https://doi.org/10.1111/j.1399-3011.1969.tb01625.x

Rose, M., Landman, D., & Quale, J. (2014). Are community environmental surfaces near hospitals reservoirs for gram-negative nosocomial pathogens? *American Journal of Infection Control*, *42*(4), 346-348. https://doi.org/10.1016/j.ajic.2013.12.025

Rosenbaum, E., Ferruit, M., Durá, M. A., & 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. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1814*(10), Article 10. https://doi.org/10.1016/j.bbapap.2010.11.008

Rotanova, T. V., Botos, I., Melnikov, E. E., Rasulova, F., Gustchina, A., Maurizi, M. R., & Wlodawer, A. (2006). Slicing a protease : Structural features of the ATP-dependent Lon proteases gleaned from investigations of isolated domains. *Protein Science*, *15*(8), 1815-1828. https://doi.org/10.1110/ps.052069306

Rotanova, T. V., Dergousova, N. I., & Morozkin, A. D. (2013). Unique structural organization of ATP-dependent LonA proteases. *Russian Journal of Bioorganic Chemistry*, *39*(3), 268-282. https://doi.org/10.1134/S1068162013030114

Rotanova, T. V., & Melnikov, E. E. (2008). The ATP-dependent proteases and proteolytic complexes involved into intracellular protein degradation. *Biochemistry (Moscow) Supplement Series B: Biomedical Chemistry*, *2*(3), 245-257. https://doi.org/10.1134/S1990750808030049

Rotanova, T. V., Melnikov, E. E., Khalatova, A. G., Makhovskaya, O. V., Botos, I., Wlodawer, A., & Gustchina, A. (2004). Classification of ATP-dependent proteases Lon and comparison of the active sites of their proteolytic domains. *European Journal of Biochemistry*, *271*(23-24), 4865-4871. https://doi.org/10.1111/j.1432-1033.2004.04452.x

Rowsell, S., Pauptit, R. A., Tucker, A. D., Melton, R. G., Blow, D. M., & Brick, P. (1997). Crystal structure of carboxypeptidase G2, a bacterial enzyme with applications in cancer therapy. *Structure*, *5*(3), 337-347. https://doi.org/10.1016/S0969-2126(97)00191-3

Russo, S., & Baumann, U. (2004). Crystal Structure of a Dodecameric tetrahedral-shaped aminopeptidase. *Journal of Biological Chemistry*, *279*(49), 51275-51281. https://doi.org/10.1074/jbc.M409455200

Sagan, L. (1967). On the origin of mitosing cells. *Journal of Theoretical Biology*, *14*(3), 225-IN6. https://doi.org/10.1016/0022-5193(67)90079-3

Sahoo, G., Samal, D., Khandayataray, P., & Murthy, M. K. (2023). A Review on Caspases : Key Regulators of Biological Activities and Apoptosis. *Molecular Neurobiology*, *60*(10), 5805-5837. https://doi.org/10.1007/s12035-023-03433-5

Sandman, K., Grayling, R. A., Dobrinski, B., Lurz, R., & Reeve, J. N. (1994). Growth-phase-dependent synthesis of histones in the archaeon Methanothermus fervidus. *Proceedings of the National Academy of Sciences*, *91*(26), 12624-12628. https://doi.org/10.1073/pnas.91.26.12624

Sandman, K., Krzycki, J. A., Dobrinski, B., Lurz, R., & Reeve, J. N. (1990). HMf, a DNA-binding protein isolated from the hyperthermophilic archaeon Methanothermus fervidus, is most closely related to histones. *Proceedings of the National Academy of Sciences*, *87*(15), 5788-5791. https://doi.org/10.1073/pnas.87.15.5788

Sarmiento, F., Mrázek, J., & Whitman, W. B. (2013). Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon *Methanococcus maripaludis*. *Proceedings of the National Academy of Sciences*, *110*(12), 4726-4731. https://doi.org/10.1073/pnas.1220225110

Sauer, R. T., & Baker, T. A. (2011). AAA+ Proteases : ATP-Fueled Machines of Protein Destruction. *Annual Review of Biochemistry*, *80*(1), 587-612. https://doi.org/10.1146/annurev-biochem-060408-172623

Schechter, I., & Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochemical and Biophysical Research Communications*, *27*(2), 157-162. https://doi.org/10.1016/S0006-291X(67)80055-X

Schmidt, R., Decatur, A. L., Rather, P. N., Moran, C. P., & Losick, R. (1994). Bacillus subtilis lon protease prevents inappropriate transcription of genes under the control of the sporulation transcription factor sigma G. *Journal of Bacteriology*, *176*(21), 6528-6537. https://doi.org/10.1128/jb.176.21.6528-6537.1994

Schmidt, T. G., & Skerra, A. (2007). The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nature Protocols*, *2*(6), 1528-1535. https://doi.org/10.1038/nprot.2007.209

Schoehn, G., Vellieux, F. M. D., Asunción Durá, M., Receveur-Bréchot, V., Fabry, C. M. S., Ruigrok, R. W. H., Ebel, C., Roussel, A., & Franzetti, B. (2006). An archaeal peptidase assembles into two different quaternary structures. *Journal of Biological Chemistry*, *281*(47), Article 47. https://doi.org/10.1074/jbc.M604417200

Schut, G. J., Brehm, S. D., Datta, S., & Adams, M. W. W. (2003). Whole-Genome DNA Microarray Analysis of a Hyperthermophile and an Archaeon : *Pyrococcus furiosus* Grown on Carbohydrates or Peptides. *Journal of Bacteriology*, *185*(13), 3935-3947. https://doi.org/10.1128/JB.185.13.3935-3947.2003

Scola, B. L., Audic, S., Robert, C., Jungang, L., De Lamballerie, X., Drancourt, M., Birtles, R., Claverie, J.-M., & Raoult, D. (2003). A Giant Virus in Amoebae. *Science*, *299*(5615), 2033-2033. https://doi.org/10.1126/science.1081867

Sears, R., Nuckolls, F., Haura, E., Taya, Y., Tamai, K., & Nevins, J. R. (2000). Multiple Ras-dependent phosphorylation pathways regulate Myc protein stability. *Genes & Development*, *14*(19), 2501-2514. https://doi.org/10.1101/gad.836800

Seitz, K. W., Lazar, C. S., Hinrichs, K.-U., Teske, A. P., & Baker, B. J. (2016). Genomic reconstruction of a novel, deeply branched sediment archaeal phylum with pathways for acetogenesis and sulfur reduction. *The ISME Journal*, *10*(7), 1696-1705. https://doi.org/10.1038/ismej.2015.233

Shin, M., Puchades, C., Asmita, A., Puri, N., Adjei, E., Wiseman, R. L., Karzai, A. W., & Lander, G. C. (2020). Structural basis for distinct operational modes and protease activation in AAA+ protease Lon. *Science Advances*, *6*(21), eaba8404. https://doi.org/10.1126/sciadv.aba8404

Slutskaya, E. S., Bezsudnova, E. Yu., Mardanov, A. V., Gumerov, V. M., Rakitina, T. V., Popov, V. O., & Lipkin, V. M. (2012). Characterization of a novel M42 aminopeptidase from crenarchaeon Desulfurococcus kamchatkensis. *Dokl. Biochem. Biophys.*, *442*(1), Article 1. https://doi.org/10.1134/S1607672912010097

Spang, A., Saw, J. H., Jørgensen, S. L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A. E., van Eijk, R., Schleper, C., Guy, L., & Ettema, T. J. G. (2015). Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature*, *521*(7551), 173-179. https://doi.org/10.1038/nature14447

Stoll, E., Hermodson, M. A., Ericsson, L. H., & Zuber, H. (1972). Subunit structure of the thermophilic aminopeptidase I from Bacillus stearothermophilus. *Biochemistry*, *11*(25), 4731-4735. https://doi.org/10.1021/bi00775a015

Susanti, D., Frazier, M. C., & Mukhopadhyay, B. (2019). A genetic system for Methanocaldococcus jannaschii : An evolutionary deeply rooted hyperthermophilic methanarchaeon. *Frontiers in Microbiology*, *10*(1256), Article 1256. https://doi.org/10.3389/fmicb.2019.01256

Swamy, K. H. S., & Goldberg, A. L. (1981). E. coli contains eight soluble proteolytic activities, one being ATP dependent. *Nature*, *292*(5824), 652-654. https://doi.org/10.1038/292652a0

Takai, K., Nakamura, K., Toki, T., Tsunogai, U., Miyazaki, M., Miyazaki, J., Hirayama, H., Nakagawa, S., Nunoura, T., & Horikoshi, K. (2008). Cell proliferation at 122°C and isotopically heavy CH_4 production by a hyperthermophilic methanogen under high-pressure cultivation. *Proceedings of the National Academy of Sciences*, *105*(31), 10949-10954. https://doi.org/10.1073/pnas.0712334105

Takai, K., Oida, H., Suzuki, Y., Hirayama, H., Nakagawa, S., Nunoura, T., Inagaki, F., Nealson, K. H., & Horikoshi, K. (2004). Spatial Distribution of Marine Crenarchaeota Group I in the Vicinity of Deep-Sea Hydrothermal Systems. *Applied and Environmental Microbiology*, *70*(4), 2404-2413. https://doi.org/10.1128/AEM.70.4.2404-2413.2004

Tamura, N., Lottspeich, F., Baumeister, W., & Tamura, T. (1998). The role of tricorn protease and its aminopeptidase-interacting factors in cellular protein degradation. *Cell*, *95*(5), Article 5. https://doi.org/10.1016/S0092-8674(00)81634-7

Tamura, N., Pfeifer, G., Baumeister, W., & Tamura, T. (2001). Tricorn protease in bacteria : Characterization of the enzyme from Streptomyces coelicolor. *Biological Chemistry*, *382*(3), Article 3. https://doi.org/10.1515/BC.2001.055

Tamura, T., Tamura, N., Cejka, Z., Hegerl, R., Lottspeich, F., & Baumeister, W. (1996). Tricorn Protease—The Core of a Modular Proteolytic System. *Science*, *274*(5291), 1385-1389. https://doi.org/10.1126/science.274.5291.1385

Tamura, T., Tamura, N., Lottspeich, F., & Baumeister, W. (1996). Tricorn protease (TRI) interacting factor 1 from *Thermoplasma acidophilum* is a proline iminopeptidase. *FEBS Letters*, *398*(1), 101-105. https://doi.org/10.1016/S0014-5793(96)01163-5

Tehei, M., Madern, D., Franzetti, B., & Zaccai, G. (2005). Neutron Scattering Reveals the Dynamic Basis of Protein Adaptation to Extreme Temperature. *Journal of Biological Chemistry*, *280*(49), 40974-40979. https://doi.org/10.1074/jbc.M508417200

Teuscher, F., Lowther, J., Skinner-Adams, T. S., Spielmann, T., Dixon, M. W. A., Stack, C. M., Donnelly, S., Mucha, A., Kafarski, P., Vassiliou, S., Gardiner, D. L., Dalton, J. P., & Trenholme, K. R. (2007). The M18 Aspartyl Aminopeptidase of the Human Malaria Parasite Plasmodium falciparum. *Journal of Biological Chemistry*, *282*(42), 30817-30826. https://doi.org/10.1074/jbc.M704938200

Thiel, A., Michoud, G., Moalic, Y., Flament, D., & Jebbar, M. (2014). Genetic Manipulations of the Hyperthermophilic Piezophilic Archaeon Thermococcus barophilus. *Applied and Environmental Microbiology*, *80*(7), 2299-2306. https://doi.org/10.1128/AEM.00084-14

Tomoyasu, T., Gamer, J., Bukau, B., Kanemori, M., Mori, H., Rutman, A. J., Oppenheim, A. B., Yura, T., Yamanaka, K., & Niki, H. (1995). Escherichia coli FtsH is a membrane-bound, ATP-dependent protease which degrades the heat-shock transcription factor sigma 32. *The EMBO Journal*, *14*(11), 2551-2560. https://doi.org/10.1002/j.1460-2075.1995.tb07253.x

Tsilibaris, V., Maenhaut-Michel, G., & Van Melderen, L. (2006). Biological roles of the Lon ATP-dependent protease. *Research in Microbiology*, *157*(8), 701-713. https://doi.org/10.1016/j.resmic.2006.05.004

Tsunasawa, S. (1998). Purification and application of a novel N-terminal deblocking aminopeptidase (DAP) from Pyrococcus furiosus. *Journal of Protein Chemistry*, *17*(6), 521-522.

Udvardy, A. (1996). The Role of Controlled Proteolysis in Cell-Cycle Regulation. *European Journal of Biochemistry*, *240*(2), 307-313. https://doi.org/10.1111/j.1432-1033.1996.0307h.x

Verzijl, N., DeGroot, J., Thorpe, S. R., Bank, R. A., Shaw, J. N., Lyons, T. J., Bijlsma, J. W. J., Lafeber, F. P. J. G., Baynes, J. W., & TeKoppele, J. M. (2000). Effect of Collagen Turnover on the Accumulation of Advanced Glycation End Products. *Journal of Biological Chemistry*, *275*(50), 39027-39031. https://doi.org/10.1074/jbc.M006700200

Villanueva, L., Bastiaan Von Meijenfeldt, F. A., Westbye, A. B., Yadav, S., Hopmans, E. C., Dutilh, B. E., & Sinninghe Damsté, J. S. (2021). Bridging the membrane lipid divide : Bacteria of the FCB group superphylum have the potential to synthesize archaeal ether lipids. *The ISME Journal*, *15*(1), 168-182. https://doi.org/10.1038/s41396-020-00772-2

Wagner, A., Whitaker, R. J., Krause, D. J., Heilers, J.-H., Van Wolferen, M., Van Der Does, C., & Albers, S.-V. (2017). Mechanisms of gene flow in archaea. *Nature Reviews Microbiology*, *15*(8), 492-501. https://doi.org/10.1038/nrmicro.2017.41

Walz, J., Tamura, T., Tamura, N., Grimm, R., Baumeister, W., & Koster, A. J. (1997). Tricorn Protease Exists as an Icosahedral Supermolecule In Vivo. *Molecular Cell*, *1*(1), 59-65. https://doi.org/10.1016/S1097-2765(00)80007-6

Ward, D. E., Shockley, K. R., Chang, L. S., Levy, R. D., Michel, J. K., Conners, S. B., & Kelly, R. M. (2002). Proteolysis in hyperthermophilic microorganisms. *Archaea*, *1*(1), 63-74. https://doi.org/10.1155/2002/503191

Whittaker, R. H. (1969). New Concepts of Kingdoms of Organisms : Evolutionary relations are better represented by new classifications than by the traditional two kingdoms. *Science*, *163*(3863), 150-160. https://doi.org/10.1126/science.163.3863.150

Wiens, J. J. (2023). How many species are there on Earth? Progress and problems. *PLOS Biology*, *21*(11), e3002388. https://doi.org/10.1371/journal.pbio.3002388

Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A. M., Lieberenz, M., Savitski, M. M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H., Mathieson, T., Lemeer, S., Schnatbaum, K., Reimer, U., Wenschuh, H., Mollenhauer, M., Slotta-Huspenina, J., Boese, J.-H., Bantscheff, M., … Kuster, B. (2014). Mass-spectrometry-based draft of the human proteome. *Nature*, *509*(7502), 582-587. https://doi.org/10.1038/nature13319

Wilk, S., Wilk, E., & Magnusson, R. P. (1998). Purification, Characterization, and Cloning of a Cytosolic Aspartyl Aminopeptidase. *Journal of Biological Chemistry*, *273*(26), 15961-15970. https://doi.org/10.1074/jbc.273.26.15961

Wingfield, P. T. (2017). N-Terminal Methionine Processing. *Current Protocols in Protein Science*, *88*(1). https://doi.org/10.1002/cpps.29

Woese, C. R., & Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain : The primary kingdoms. *Proceedings of the National Academy of Sciences*, *74*(11), 5088-5090. https://doi.org/10.1073/pnas.74.11.5088

Woese, C. R., Kandler, O., & Wheelis, M. L. (1990). Towards a natural system of organisms : Proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences*, *87*(12), 4576-4579. https://doi.org/10.1073/pnas.87.12.4576

Wolf, D. H., & Menssen, R. (2018). Mechanisms of cell regulation – proteolysis, the big surprise. *FEBS Letters*, *592*(15), 2515-2524. https://doi.org/10.1002/1873-3468.13109

Wu, W.-F., Zhou, Y., & Gottesman, S. (1999). Redundant In Vivo Proteolytic Activities of *Escherichia coli* Lon and the ClpYQ (HslUV) Protease. *Journal of Bacteriology*, *181*(12), 3681-3687. https://doi.org/10.1128/JB.181.12.3681-3687.1999

Yao, T., & Cohen, R. E. (1999). Giant proteases : Beyond the proteasome. *Current Biology*, *9*(15), 551-553. https://doi.org/10.1016/S0960-9822(99)80352-2

Zabel, H.-P., König, H., & Winter, J. (1985). Emended Description of Methanogenium thermophilicum, Rivard and Smith, and Assignment of New Isolates to this Species. *Systematic and Applied Microbiology*, *6*(1), 72-78. https://doi.org/10.1016/S0723-2020(85)80014-X

Zaremba-Niedzwiedzka, K., Caceres, E. F., Saw, J. H., Bäckström, D., Juzokaite, L., Vancaester, E., Seitz, K. W., Anantharaman, K., Starnawski, P., Kjeldsen, K. U., Stott, M. B., Nunoura, T., Banfield, J. F., Schramm, A., Baker, B. J., Spang, A., & Ettema, T. J. G. (2017). Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature*, *541*(7637), 353-358. https://doi.org/10.1038/nature21031

Zhang, Z., Schäffer, A. A., Miller, W., Madden, T. L., Lipman, D. J., Koonin, E. V., & Altschul, S. F. (1998). Protein sequence similarity searches using patterns as seeds. *Nucleic Acids Research*, *26*(17), 3986-3990. https://doi.org/10.1093/nar/26.17.3986

Zhao, D., & Huang, Z. (2016). Effect of His-Tag on Expression, Purification, and Structure of Zinc Finger Protein, ZNF191(243-368). *Bioinorganic Chemistry and Applications*, *2016*, 1-6. https://doi.org/10.1155/2016/8206854

Zheng, J., Cheng, Z., Jia, H., & Zheng, Y. (2016). Characterization of aspartyl aminopeptidase from Toxoplasma gondii. *Scientific Reports*, *6*(1), 34448. https://doi.org/10.1038/srep34448

Zheng, Y., Roberts, R. J., Kasif, S., & Guan, C. (2005). Characterization of Two New Aminopeptidases in *Escherichia coli*. *Journal of Bacteriology*, *187*(11), 3671-3677. https://doi.org/10.1128/JB.187.11.3671-3677.2005

Zwickl, P., Ng, D., Woo, K. M., Goldberg, A. L., & Klenk, H.-P. (1999). An Archaebacterial ATPase, Homologous to ATPases in the Eukaryotic 26 S Proteasome, Activates Protein Breakdown by 20 S Proteasomes. *Journal of Biological Chemistry*, *274*(37), 26008-26014. https://doi.org/10.1074/jbc.274.37.26008

COLLEGE SCIENCES DOCTORAL DE LA MER BRETAGNE ET DU LITTORAL

Université de Bretagne Occidentale

Titre : Etude de la diversité des peptidases TET chez les Archées à travers des approches phylogénétiques et biochimiques

Mots clés : Peptidases TET, Archées, protéolyse, analyse de séquence, enzymologie

Résumé : Les peptidases TET sont des aminopeptidases géantes présentes dans les trois domaines du vivant. Elles sont supposées intervenir dans la protéolyse intracellulaire, en aval du protéasome. Des études structurales ont montré que ces enzymes forment des complexes tétraédriques dodécamériques creux. Chez les Archées, les peptidases TET ont principalement été étudiées chez des espèces de l'ordre des Thermococcales, révélant quatre types distincts de TET. Cependant, la répartition et la diversité fonctionnelle de ces peptidases dans les autres lignées archéennes restent mal connue, et leur rôle biologique précis est toujours indéterminé.
Dans cette étude, des critères d'identification basés sur des éléments structuraux conservés sont définis, permettant l'identification des TETs à grande échelle. En utilisant ces critères pour le screening de 4016 génomes d'archées, la première analyse complète de la distribution des TETs chez ces organismes est présentée, mettant à jour une diversité insoupçonnée. À travers une analyse phylogénétique, une classification des TETs archéennes en onze familles est proposée.
La caractérisation biochimique de six nouvelles enzymes, issues de familles jusqu'ici non explorées et représentant une large diversité taxonomique, offre ainsi un aperçu complet de la diversité fonctionnelle des TETs.
Enfin, la combinaison des résultats des analyses phylogénétiques et biochimiques éclaire l'histoire évolutive des TETs et apporte de nouvelles perspectives sur leurs potentielles fonctions biologiques.

Title: Phylogenetic and biochemical insights into the diversity of TET peptidases in Archaea

Keys words: TET peptidases, Archaea, proteolysis, sequence analysis, enzymology

Abstract: TET peptidases are giant aminopeptidases ubiquitous accross all domains life, proposed to be involved in intracellular proteolysis, acting downstream of the proteasome. Structural studies have shown that these enzymes form hollow dodecameric tetrahedral complexes. In Archaea, TET peptidases have primarily been studied in Thermococcales species, revealing four distinct types of TETs. However, the prevalence and functional diversity of these peptidases in other archaeal lineages remain poorly understood, and their precise biological role is still unclear.
In this work, structure-based identification criteria are introduced, enabling high-throughput screening for TETs. By applying these criteria to 4,016 archaeal genomes, the first comprehensive analysis of TET distribution in Archaea was conducted, uncovering an unsuspected diversity. Through phylogenetic analysis, a classification of archaeal TETs in eleven families is proposed.
Biochemical characterization of six new enzymes from previously undescribed families and covering a wide taxonomic range of archaeal species offers an extensive overview of the functional diversity of TETs.
Finally, by combining biochemical and phylogenetic data, the evolutionary history of these peptidases is addressed, offering new insights into their potential biological roles.