Phylogenetic survey of the subtilase family and biochemical characterisation of novel subtilisins derived from halotolerant or halophilic *Bacillaceae*

Inaugural-Dissertation

for the attainment of the title of doctor
in the Faculty of Mathematics and Natural Sciences
at the Heinrich-Heine-University Düsseldorf

presented by

Fabian Falkenberg
from Neuss

Neuss, June 2023

The thesis in hand has been performed at the Institute of Nano- and Biotechnologies, FH Aachen - University of Applied Sciences in cooperation with the Institute of Bio- and Geosciences, IBG-1: Biotechnology, Forschungszentrum Jülich GmbH, from September 2019 until March 2023 under the supervision of Prof. Dr. Petra Siegert, Prof. Dr. Johannes Bongaerts and Prof. Dr. Michael Bott.

Published by permission of the
Faculty of Mathematics and Natural Sciences at
Heinrich Heine University Düsseldorf

Examiner:

1. Prof. Dr. Michael Bott
   Institute of Bio- and Geosciences 1: Biotechnology
   Forschungszentrum Jülich GmbH

2. Prof. Dr. Johannes Bongaerts
   Institute of Nano- and Biotechnologies
   FH Aachen – University of Applied Sciences

Date of the oral examination: 20.10.2023

The results described in this dissertation were published or submitted in the following articles in peer-reviewed journals:

Falkenberg F, Bott M, Bongaerts J & Siegert P (2022) Phylogenetic survey of the subtilase family and a data-mining-based search for new subtilisins from *Bacillaceae*. *Front Microbiol* **13**. DOI: 10.3389/fmicb.2022.1017978

Falkenberg F, Rahba J, Fischer D, Bott M, Bongaerts J & Siegert P (2022). Biochemical characterization of a novel oxidatively stable, halotolerant and high-alkaline subtilisin from *Alkalihalobacillus okhensis* Kh10-101^T. *FEBS Open Bio* **12,** 1729–1746. DOI: 10.1002/2211-5463.13457

Falkenberg F, Voß L, Bott M, Bongaerts J, Siegert P (2023). New robust subtilisins from halotolerant and halophilic *Bacillaceae*. *Appl Microbiol Biotechnol* **107**, 3939–3954. DOI: 10.1007/s00253-023-12553-w

Falkenberg F, Kohn S, Bott M, Bongaerts J, Siegert P (2023). Biochemical characterisation of a novel broad pH spectrum subtilisin from *Fictibacillus arsenicus* DSM 15822^T. *FEBS Open Bio.* DOI: doi.org/10.1002/2211-5463.13701

Falkenberg F, Vonderbank V, Herten J, Kus N Y, Bott M, Bongaerts J, Siegert P (2023). *Metabacillus indicus* DSM 16189 as promising host for recombinant protease production (Manuscript).

The following patents were filed based on the results of this work:

"Wasch- und Reinigungsmittel enthaltend Protease", Applicant: Henkel AG & Co. KGaA; Patent Office: German Patent and Trade Mark Office; Application number: DE 102022213537.7; Application date: 13.12.2022

"Wasch- und Reinigungsmittel enthaltend Protease" Applicant: Henkel AG & Co. KGaA; Patent Office: German Patent and Trade Mark Office; Application number: DE 102022213538.5; Application date: 13.12.2022

Poster presentations:

Falkenberg F, Rahba J, Siegert P, Bongaerts J. Identification and evaluation of novel subtilases from *Bacillaceae*. Annual Conference of the *Vereinigung für Allgemeine und Angewandte Mikrobiologie* (VAAM), Leipzig (2020).

Falkenberg F, Siegert P, Bongaerts J. Exploring the biochemical potential of new subtilisins from halotolerant *Bacillaceae*. Annual Conference of the *Vereinigung für Allgemeine und Angewandte Mikrobiologie* (VAAM), Düsseldorf/digital (2022).

Falkenberg F, Siegert P, Bongaerts J. Exploring new subtilisins from halotolerant *Bacillaceae* by database mining and phylogenetics. International Congress on Biocatalysis, Hamburg (2022).

ElHamdi M, Falkenberg F, Ghorbel S, Bongaerts J, Siegert P, Hmidet N. Two subtilisin-like proteases from *Bacillus zhangzhouensis* MH1: identification, molecular characterisation and heterologous expression in *Bacillus subtilis* DB104. International Biotechnology Days (IDB) of the Tunisian Association of Biotechnology (2022)

Zusammenfassung

Die steigende Nachfrage nach Proteasen für die industrielle Anwendung in verschiedenen Bereichen wie Waschmittel, Kosmetika, Lederverarbeitung, Abwasserbehandlung und Pharmazeutik mit immer neuen Anforderungen an die Eigenschaften von Proteasen zeigt die Notwendigkeit des Bedarfes an neuen Proteasen mit verbesserten Eigenschaften. Um diesem Bedarf gerecht zu werden, wurden in dieser Arbeit neue Protease-Kandidaten aus der Gruppe der Subtilisine, die zu den Subtilasen gehören, identifiziert.

Zunächst wurden die Sequenzen bereits charakterisierter Proteasen aus der Subtilase-Familie untersucht und hinsichtlich ihrer Klassifizierung neu bewertet, da eine Übersicht über diese große Familie seit 1997 (Siezen und Leunissen) nicht mehr aktualisiert wurde und die Zahl der neu charakterisierten Subtilasen stetig zunimmt. Die phylogenetische Analyse des MEROPS-Holotyp-Datensatzes der Subtilase-Familie ergab, dass eine große Anzahl verschiedener Subtilasen neue Gruppen und Untergruppen bilden. Es wurden acht neue Gruppen und 13 neue Untergruppen identifiziert und eingeführt. Darüber hinaus wurde eine aktualisierte Übersicht erstellt, um die neu gefundenen Sequenzen phylogenetisch den bereits bekannten zuordnen zu können.

So wurden in einer datenbankbasierten Suche nach neuen Subtilisin-Sequenzen aus der Familie der *Bacillaceae* rund 1400 Sequenzen ausgewertet. Unter Verwendung des aktualisierten phylogenetischen Kontextes und weiterer bioinformatischer Analysen konnte eine Gruppe von 120 Sequenzen identifiziert werden, die für weitere Untersuchungen ausgewählt wurden. Insbesondere Enzyme aus halophilen oder halotoleranten Organismen bieten ein enormes Potenzial, um industrielle Anforderungen zu erfüllen. Daher wurden die Subtilisin-Gene von *Pontibacillus marinus* DSM 16465^T (SPPM), *Metabacillus indicus* DSM 16189 (SPMI), *Litchfieldia alkalitelluris* DSM 16976^T (SPLA), *Fictibacillus arsenicus* DSM 15822^T (SPFA), *Alkalibacillus haloalkaliphilus* DSM 5271^T (SPAH) und *Halalkalibacter okhensis* Kh10-101^T (SPAO) isoliert und die Enzyme rekombinant mit *Bacillus subtilis* DB104 produziert, anschließend gereinigt und biochemisch charakterisiert. Dabei konnte bestätigt werden, dass die Wildtyp-Subtilisine aus halotoleranten und halophilen Spezies technisch interessantere biochemische Eigenschaften aufweisen als die für die technische Anwendung entwickelten hochalkalischen Subtilisine. Besonders erwähnenswert ist die hohe Stabilität gegenüber SDS, NaCl und H_2O_2 sowie die hohe Aktivität bei alkalischen pH-Werten, die sowohl wissenschaftlich als auch für industrielle Anwendungen interessant ist. Darüber hinaus wurden die oben genannten Organismen auf ihre genetische Zugänglichkeit hin untersucht. Dabei konnte *Metabacillus indicus* DSM 16189 als neuer Wirt etabliert und für die rekombinante Produktion des nativen sekretierten Subtilisins genutzt werden.

Die Untersuchung neuer Subtilisine mit halotolerantem und halophilem Ursprung in dieser Arbeit trägt zum Verständnis der Anpassung extrazellulärer Enzyme und der Beziehung zwischen Aminosäuresequenz und biochemischen Eigenschaften bei. Die gewonnenen Erkenntnisse, insbesondere aus der Struktur-Funktions-Analyse mit Schwerpunkt auf der Oberflächenladungsanalyse, bilden den Ausgangspunkt für weitere enzymtechnische Studien.

Summary

The increasing demand for proteases for industrial applications in various fields such as detergents, cosmetics, leather processing, waste water treatment and pharmaceuticals with ever new requirements for the proteases necessitates the need for new proteases with improved properties. To meet this demand, new protease candidates were searched for in this work in the group of subtilisins, which belong to the subtilases.

First, the sequences of already characterised proteases from the subtilase family were collected and re-evaluated with regard to their classification, since the number of newly characterised subtilases is constantly increasing and the overview of this large family has not been updated since 1997 (Siezen and Leunissen). The phylogenetic analysis of the MEROPS holotype data set of the subtilase family revealed that a large number of different subtilases form new groups and subgroups. Eight new groups and 13 new subgroups were identified and introduced. Furthermore, an updated overview was necessary to be able to assign newly found sequences phylogenetically to the already known ones. Thus, in a database-based search for new subtilisin sequences from the family *Bacillaceae*, about 1400 sequences were evaluated. Using the updated phylogenetic context and further bioinformatic analysis, a group of 120 sequences was identified and selected for further analysis. In particular, enzymes from halophilic or halotolerant organisms offer enormous potential for meeting industrial requirements. Therefore, the subtilisin genes from *Pontibacillus marinus* DSM 16465^T (SPPM), *Metabacillus indicus* DSM 16189 (SPMI), *Litchfieldia alkalitelluris* DSM 16976^T (SPLA), *Fictibacillus arsenicus* DSM 15822^T (SPFA), *Alkalibacillus haloalkaliphilus* DSM 5271^T (SPAH) and *Halalkalibacter okhensis* Kh10-101^T (SPAO) were isolated and the enzymes recombinantly produced with *Bacillus subtilis* DB104, then purified and biochemically characterised. Here it could be confirmed that the wild-type subtilisins from halotolerant and halophilic species had technically more interesting biochemical properties than the previously developed high-alkaline subtilisins. Especially worth mentioning is the high stability towards SDS, NaCl, H_2O_2, as well as the high activity at alkaline pH, which is interesting for both research and industrial applications. In addition, the organisms mentioned above were investigated for their genetic accessibility. Here, *Metabacillus indicus* DSM 16189 was established as a new host and used for the recombinant production of its own secreted subtilisin.

By studying novel subtilisins with halotolerant and halophilic origins, this work contributes to the understanding of the adaptation of extracellular enzymes and the relationship between amino acid sequence and biochemical properties. The results, especially from structure-function analysis with emphasis on surface charge analysis, form the starting point for further enzyme engineering studies.

Abbreviations

ALTP, *Alkaliphilus transvaalensis* protease
aprE, extracellular alkaline protease gene
ASP, *Aeromonas sobria* protease
bpF, bacillopeptidase F
CD, Circular dichroism
GRAS, Generally Regarded as Safe
EC, Enzyme commission
EDTA, ethylenediaminetetraacetic acid
HEPES, 4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid
HMS, high-molecular-mass subtilases
IEF, isoelectric focusing
Isp, Intracellular serine protease/intracellular subtilisin
LB, lysogeny broth
LM, length marker
MALDI-TOF-MS, matrix-assisted laser desorption ionisation-time of flight mass spectrometry
MSA, multiple sequence alignment
NCBI, National Center for Biotechnology Information
NC-IUBMB, Nomenclature Committee of the International Union of Biochemistry and Molecular Biology
OSP, oxidatively stable proteases
PAGE, polyacrylamide gel electrophoresis
PCR, polymerase chain reaction
PDB, protein data bank
pI, isoelectric point
PIS, phylogenetically intermediate subtilisins
PMSF, phenylmethanesulfonyl fluoride
pNA, para-nitroanilide
SDS, sodium dodecyl sulphate
SN, supernatant
SPAH, subtilisin protease *Alkalibacillus haloalkaliphilus*
SPAO, subtilisin protease *Alkalihalobacillus okhensis* (*Halalkalibacter okhensis*)
SPFA, subtilisin protease *Fictibacillus arsenicus*
SPLA, subtilisin protease *Litchfieldia alkalitelluris*
SPMI, subtilisin protease *Metabacillus indicus*
SPPM, subtilisin protease *Pontibacillus marinus*
suc, N-succinyl
Tk-subtilisin, *Thermococcus kodakarensis* subtilisin
T_m, melting temperature
TM-score, template modeling score
TPPII, tripeptidyl peptidase II
TPPS, tripeptidyl peptidase S

Content

1 Introduction

1.1 Peptidases

Peptidases are enzymes that catalyse the hydrolysis of peptide bonds of proteins or shorter peptides, producing peptides or amino acids. They are found in all living organisms, are necessary for their survival, and about 2 % of the genome of each organism encodes for peptidases [1]. The term peptidase is recommended by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) for any protein that catalyses the hydrolysis of peptide bonds [2]. Peptidases are also called proteinases, proteolytic enzymes, or proteases [1]. They belong to the third class of the EC classification system (Enzyme Commission), the hydrolases. Within this class, they form the subclass EC 3.4. and generally cleave the peptide bond of a substrate by the addition of water. Initially, peptidases were classified either according to their specificity or according to the catalytic type [1, 3]. The EC systematic uses a hybrid of the two systems [2]. The catalytic type refers to the chemical groups responsible for catalysis. The six known types are serine, cysteine, asparagine, glutamine, threonine and metallo-peptidases [1]. Serine and threonine peptidases use a hydroxy group and cysteine peptidases a thiol group as a catalytic nucleophile [1]. In aspartic, glutamic and metallo-peptidases, the nucleophile is usually an activated water molecule [1, 4]. However, an additional type of proteolytic enzyme was recognised, which are not hydrolases but lyases that use asparagine as a nucleophile to cleave peptide bonds and would therefore belong to a different EC subclass (EC 4.3) [5]. The catalytic asparagine forms a cyclic structure with its own carbonyl carbon which leads to the cleavage of its own peptide bond [5].

With the exception of asparagine lyases, all peptidases hydrolyse peptides, but the hydrolysis differs by the position of the peptide bond within the substrate and by the amino acid residue near the cleavage site [1]. Therefore, peptidases within this system are called endopeptidases or exopeptidases. The classification within the EC nomenclature of peptidases is listed in Table 1. Endopeptidases hydrolyse at sites within the polypeptide chain away from the N- or C-terminus. In the EC nomenclature, endopeptidases are subdivided into subclasses (EC 3.4.21 to EC 3.4.25) [2]. However, there are also endopeptidases that can only cleave short peptides, which are called oligopeptidases [6]. Exopeptidases hydrolyse the polypeptide chain not more than three residues from the N- or C-terminus. Depending on the preferred hydrolysis site they are named as amino- or carboxypeptidases (EC 3.4.11, EC 3.4.16, EC 3.4.17, EC 3.4.18). Determined by the size of the cleaved fragment they are termed as dipeptidases, dipeptidyl-, tripeptidyl-peptidases or peptidyl-dipeptidases [1]. Dipeptidases (EC 3.4.13) hydrolyse a dipeptide and require that both termini are free. Dipeptidyl-and tripeptidyl peptidases (EC 3.4.14) release an N-terminal di- or tripeptide from its substrate. Peptidyl-dipeptidases (EC 3.4.15) hydrolyse a dipeptide from the C-terminus of their substrate. In addition, there are omega-peptidases

(EC 3.4.19) that catalyse the removal of terminal peptide residues that are substituted, cyclised or linked by isopeptide bonds [7]. However, none of these classification systems represent the evolutionary relationships between peptidases, as a peptidase family can include both endopeptidases and exopeptidases [6].

Table 1 Classification of peptidases by EC system [2]

EC number	Peptidase type
EC 3.4.11	Aminopeptidases
EC 3.4.13	Dipeptidases
EC 3.4.14	Dipeptidyl-and tripeptidyl-peptidases
EC 3.4.15	Peptidyl-dipeptidases
EC 3.4.16	Serine-type carboxypeptidases
EC 3.4.17	Metallocarboxypeptidases
EC 3.4.18	Cysteine-type carboxypeptidases
EC 3.4.19	Omega peptidases
EC 3.4.21	Serine endopeptidases
EC 3.4.22	Cysteine endopeptidases
EC 3.4.23	Asparagine endopeptidases
EC 3.4.24	Metalloendopeptidases
EC 3.4.25	Threonine endopeptidases

The third and latest classification system is based on the 3D-structure and homology. Related peptidase sequences were assigned to families and families with related structures were grouped into clans [8, 9]. Many proteolytic enzymes are multi-domain proteins in which the proteolytic activity is restricted to one domain, which is solely considered when assigning a proteolytic enzyme to a family and clan [10]. This classification system is the basis for the MEROPS database (https://www.ebi.ac.uk/merops/), which contains information on more than 4400 different peptidases as well as peptidase inhibitors and substrates [9]. Within a family, almost all peptidases are restricted to one catalytic type, giving the family its name, which consists of a letter indicating the catalytic type (S, T, C, A, G, M) and a consecutive number. The designation of the individual clans is based on the catalytic mechanism (e.g. S for serine) and another consecutive letter (e.g. SB). The catalytic type within a clan can vary, therefore the letter P indicates a mixed catalytic type. Families with an unknown catalytic type are assigned an U [10]. Asparagine peptide lyases, the only non-hydrolytic proteolytic enzymes, are named with an N [5]. Most peptidases contain only a single peptidase domain, but when a protein contains multiple peptidase domains, they are defined as a complex peptidase, with the identifier beginning with the letter X and following the conventional MEROPS nomenclature for each domain [10]. If there are significant

differences within a clan or family, they are subdivided again [10]. Thus, the hierarchy in the MEROPS database can include up to six levels, in descending order: type, clan, subclan, family, subfamily and peptidase species [6]. A peptidase species represents one and the same enzyme from different organisms and is assigned a unique MEROPS identifier, which is only defined for biochemically characterised proteins [10]. For each protein species, there is a holotype that represents the first biochemically characterised peptidase [10]. Because of the rapidly growing database of peptidase sequences that are uncharacterised and too different from those of any holotype, these proteins are referred to as unassigned homologues [1].

The binding of the amino acids of the active site of a peptidase to its substrate is described in the nomenclature of Schechter and Berger [11], in which the active centre of the enzyme consists of seven subsites ("S") located on either side of the catalytic centre ("Cleavage site"). The positions ("P") of the substrate are counted from the cleavage site and therefore have the same numbering as the subsites they occupy [11], as shown in Figure 1.

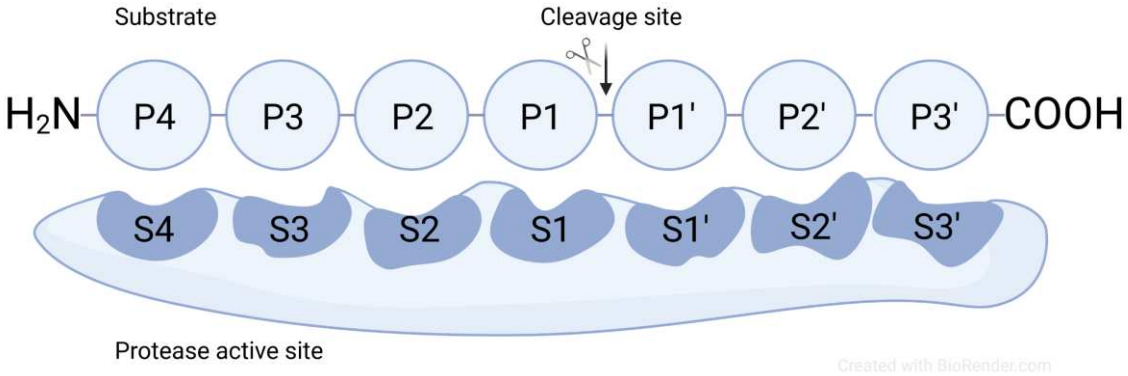


Figure 1 Schematic representation of the Schechter and Berger nomenclature [11].

1.1.1 Importance of proteases

Physiologically, proteases are important for many biological processes, such as digestion of dietary proteins, the blood clotting cascade, recycling of intracellular proteins, processing of antigens, defence of host cells in plants, and activation of a variety of proteins including hormones, enzymes and neurotransmitters, and they even act as virulence factors of pathogens [1]. About 18 % of all proteins in the SwissProt protein sequence database undergo posttranslational proteolytic processing during maturation [1]. Because proteases are involved in so many physiological processes, about 14 % of the five hundred human proteases are being investigated as drug targets [12].

From an industrial perspective, there are many uses for proteases. One of the first applications, as described by Homer in the Illias was cheese making. Although it was not known at the time that it contained proteases, they used fig juice, which contains the

protease ficain, to clot the milk [13]. This process was then adapted to the present day with rennet, animal stomach contents that contain chymosin [14]. Proteases are also used to tenderise meat, clarify beer and improve the taste of cheese and pet food [1]. Another application is, for example, in the medical field, besides being the targets of drugs, the removal of dead skin from wounds [15]. An overview of the different application fields for proteases is shown in Figure 2. However, looking at the amount of proteases used, the largest area of application is as an active ingredient in detergents and cleaning agents, as outlined in the following sections [16].

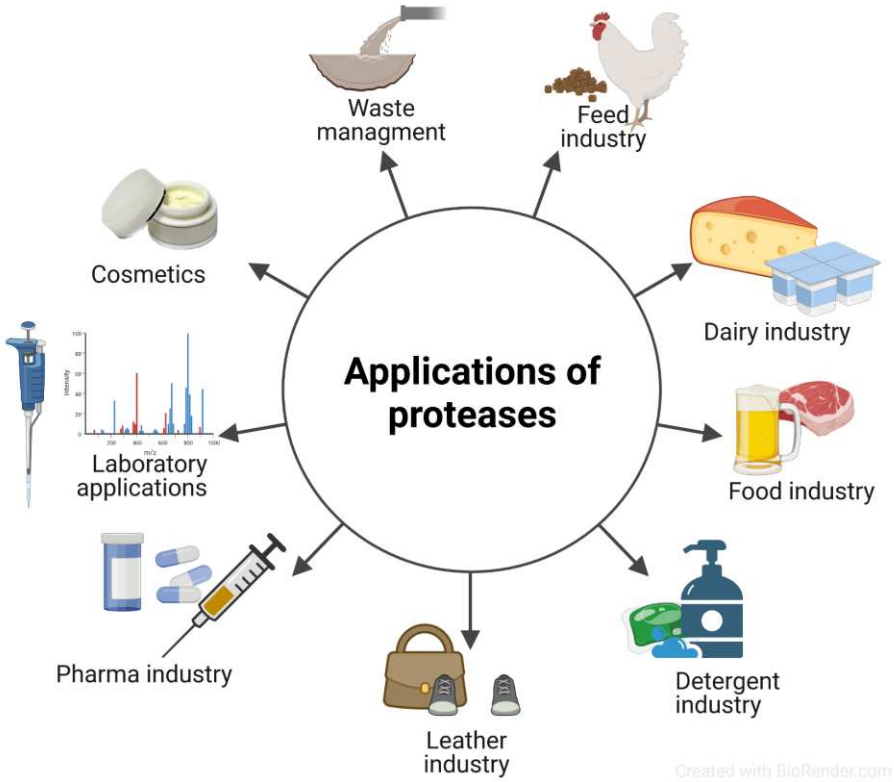


Figure 2 Applications of proteases in industries and sectors.

1.1.2 Serine proteases

The largest peptidase family are serine proteases, which use the hydroxy group of a Ser residue as a nucleophile to hydrolyse the peptide bond in a protein [17]. MEROPS count 53 serine families and 13 clans [18]. The majority of serine proteases are endopeptidases [19]. They are found in animals, plants, fungi, many microbial species, and viruses [17]. Intensively studied representatives are for example trypsin, chymotrypsin, thrombin, or subtilisin [20–23]. They are involved in various processes such as digestion, immune response, signal transduction, fibrinolysis, reproduction, or nutrient supply [20, 21, 24–27].

The catalytic mechanism of serine proteases was first established primarily by kinetic studies of chymotrypsin and involves a nucleophile, a general base and an acid [28, 29]. In the classic trypsin and subtilisin families, this task is fulfilled by the three amino acids serine, histidine and aspartate [29]. This arrangement of amino acid residues is called a

catalytic triad and exhibits similar spatial arrangements [24, 29]. However, different catalytic triads were identified, including a different order of the three amino acids His-Asp-Ser (H-D-S) for clan PA, S-D-H for clan SC, D-H-S for clan SB, S-H-E for clan SN and catalytic diads such as S-L for clan SE [29, 9].

Figure 3 illustrates the catalytic mechanisms for the clan PA, SB and SC, which are termed as the classic serine proteases. Apart from some structural changes and order of the amino acid residues of the catalytic triad, the chemical mechanism is the same [29]. In the active site, the side chain of the serine is connected to the imidazole ring of the histidine via a hydrogen bridge. The amino group of the imidazole ring is in turn connected to the carboxylate group of the aspartate via a hydrogen bridge. The histidine residue positions the serine side chain and polarises its hydroxy group, which can then be easily deprotonated [30]. In the presence of the substrate, it acts as an acceptor for the proton of the hydroxy group of serine and functions as a general base. Aspartic acid favours proton uptake through electrostatic effects and supports the spatial orientation of histidine. The reaction begins with the nucleophilic attack of the hydroxy group of the serine of the catalytic triad on the carbonyl carbon atom of the peptide. Here, an unstable tetrahedral intermediate is formed that has a negative charge on the oxygen atom of the carbonyl group [31]. This charge is stabilised by the interaction with two amino groups of the peptide backbone, which is called the oxyanion hole. In subtilisins, one of the amino groups is provided by the side chain of an asparagine residue [32]. The intermediate collapses, the peptide bond is cleaved and an acyl enzyme is formed with the release of an amino group, assisted by histidine acting as a general acid. A water molecule subsequently attacks the carbonyl group of the enzyme and at the same time the histidine removes a proton from the water molecule. This again produces an unstable tetrahedral intermediate with a negative charge, which is stabilised by the oxyanion hole. The intermediate collapses, a carboxylic acid product is released and then the free peptidase is able to carry out a new catalytic cycle [31].

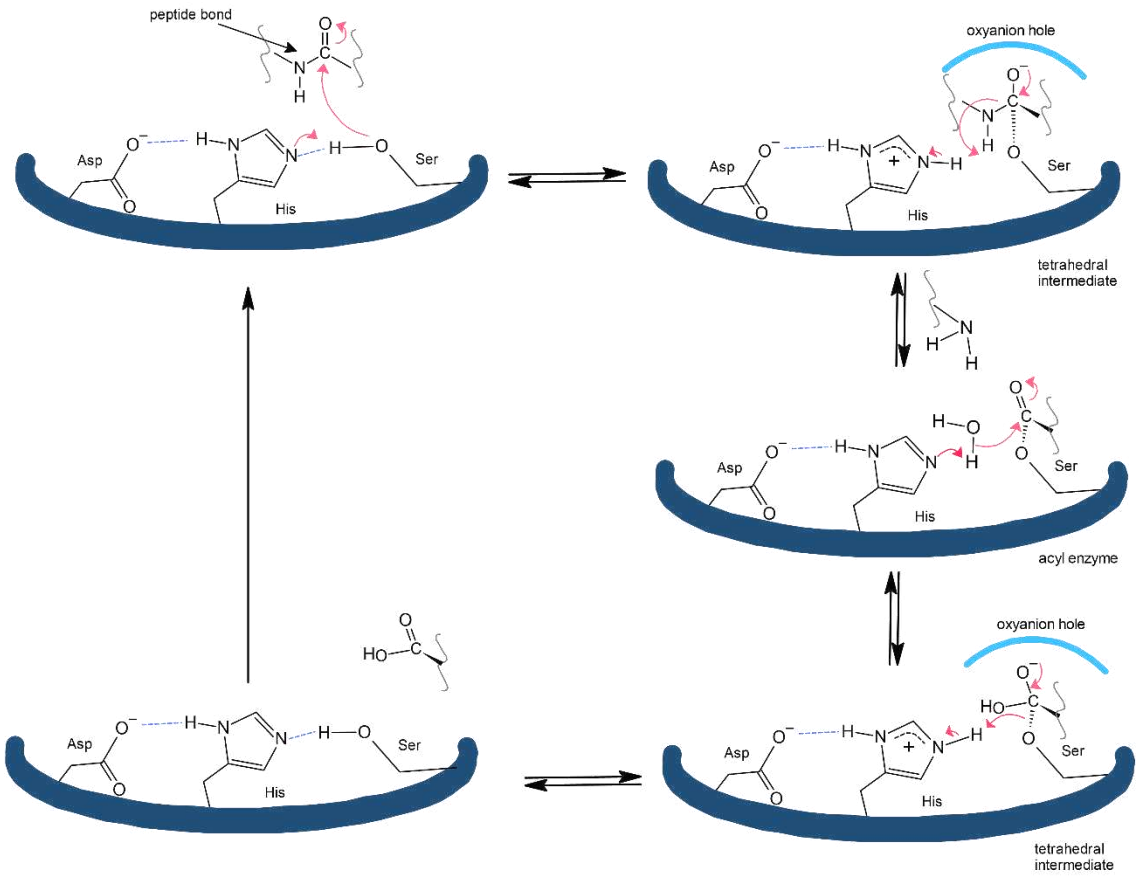


Figure 3 Schematic representation of the active site of classic serine proteases. Shown is the catalytic triad of aspartate (Asp), histidine (His) and serine (Ser) and the binding of a peptide. Blue dotted lines represent hydrogen bonds and red arrows illustrate the moving of electron pairs (based on Hedstrom [30]).

1.1.3 Subtilases

Family S8, also known as subtilases or subtilisin-like proteases, is the third largest family of serine proteases after the family S1 (chymotrypsin family) and S9 (prolyl oligopeptidases family), with regard to the number of sequences as well as the characterised proteases [9]. The family belongs to the clan SB, which includes the S8 family as well as the S53 family (sedolisin family) [19]. The two families of the clan SB differ in their catalytic triad, with the S8 family being termed as the "classic" D-H-S family, as described above, while the catalytic triad is changed to Glu, Asp and Ser for members of the S53 family, which is termed as the ED-S family [33]. The S53 family is very similar to that of the subtilisins and, in addition to the modified catalytic triad, has an aspartate instead of asparagine in the oxyanion binding site [33]. In general, the subtilases are a very diverse family with little sequence homology, often restricted to regions around the catalytic triad [33]. The analysis by Siezen *et al.* revealed another subgroup of the S53 family with a different conserved Asp residue (E-D-S family), where Glu and Asp are located in different sequence regions within the E-D-S family compared to the ED-S family, where Glu and Asp are located in the same sequence region. In addition another subgroup related to the S8 family in which the original Asp residue is replaced by a catalytic Glu

residue (E-H-S family) was found [33]. Variations in catalytic residues are common and lead to differences in catalytic mechanism and substrate specificity [34].

In addition, the S8 family is subdivided into the subfamilies S8A and S8B [9]. Furthermore, Siezen and Leunissen categorised subtilases into six groups based on a comparison of the sequence of the catalytic domain and its phylogenetics. These groups include subtilisins, thermitases, proteinase K, lantibiotic peptidases, pyrolysins, and kexins. The kexins are part of the subfamily S8B, whereas the remaining five groups fall under the subfamily S8A [24]. The S8B subfamily is mainly found in eukaryotes and includes a large group of proprotein convertases involved in the activation of peptide hormones, growth factors, viral proteins, etc. [24]. Kexin, the prototype of this subfamily, has a biosynthetic rather than a degradative function and is highly specific for the cleavage after Lys-Arg or Arg-Arg [35]. Based on sequence homology, furins also belong to this subfamily and are the mammalian homologues of kexins and responsible for cholesterol metabolism and maturation of hormone precursors [36]. Subtilases of the S8A subfamily are present in microorganisms, such as archaea, bacteria, fungi, and yeasts, as well as in higher eukaryotes [9]. In bacteria, archaea and lower eukaryotes, they are usually nonspecific proteases and mainly play a role in nutrition by providing peptides and amino acids for cell growth or host invasion [24]. Involvement in host invasion is observed in several entomopathogenic and nematophagous fungi, which are able to destroy the structural integrity of the insect or nematode cuticle during invasion and colonisation [37–39]. However, there are also maturation subtilases in prokaryotes that are involved in the maturation of e.g. lantibiotics, extracellular adhesins such as filamentous haemagglutinin or spore germination enzymes [40–42]. Subtilases are also involved in developmental processes and immune reactions in plants, play a role in the pathogenesis of *Dermatophilus congolensis* or in the metabolic processing of neuropeptides in the fruit fly, *Drosophila melanogaster* [43–45]. With the exception of TPPII (tripeptidyl peptidase II), which cleaves tripeptides from the N-terminus of peptides, the other known members of the S8 family are characterised as endopeptidases [43, 24].

Most subtilases have a multidomain structure consisting of a signal peptide, a propeptide, a peptidase domain, and in some cases a C-terminal extension, with additional features such as sequence repeats, Cys-rich domains as cell surface anchors, or transmembrane segments [24]. Subtilases without signal peptide remain inside the cell and most likely play a role in intracellular maturation processes, as mentioned above. Extracellular subtilases with an additional anchoring domain remain attached to the cell wall, such as Mycosin-1 from *Mycobacterium tuberculosis*, which is produced during infection of macrophages [46]. The N-terminal propeptide is found in all subtilases except the subtilase ASP (*Aeromonas sobria* protease) [47–49]. The propeptide functions during maturation as an intramolecular chaperone that supports the folding of the catalytic domain and serves as an inhibitor that forms a complex with the subtilase [50]. In the

absence of a propeptide, the peptidase domain alone is unable to fold into its native structure and stops at a molten globule-like intermediate [47]. The propeptide gets autoprocessed at its primary cleavage site and interacts with two surface helices and the pocket in the active site of the subtilase, where it binds in a substrate-like manner and competitively inhibits activity [51, 52]. When the first propeptide dissociates from the active site, the free subtilase triggers an exponential cascade that selectively targets other bound propeptides [53].

1.1.4 Subtilisins

Subtilisins are a group of subtilases that are one of the best studied groups of proteases in terms of structure and function due to their properties, such as thermostability, broad pH range, and broad specificity [54]. They are generally isolated from a wide range of bacterial sources of the family of *Bacillaceae*, such as *Bacillus licheniformis*, *Bacillus subtilis* and *Bacillus amyloliquefaciens* [55]. They are not only found in species of the *Bacillaceae* family, but also in other microorganisms, such as *Alkaliphilus transvaalensis* or *Pyrobaculum aerophilum* [56, 57]. The classification by Siezen and Leunissen provides a further division into the subgroups of intracellular subtilisins, true subtilisins, high-alkaline subtilisins, and later by Saeki *et al.* into phylogenetically intermediate subtilisins (PIS) [24, 58]. Intracellular subtilisins have no signal peptide and are found in many different *Bacillus* species and related bacteria, where they form the main component of the degradome, whereas extracellular subtilisins mainly play a role in nutrient supply due to their broad substrate specificity [24, 59, 60]. Typical extracellular subtilisins have a molecular mass of about 27 kDa and are secreted in an about 39 kDa precursor form including the signal peptide, the propeptide and the peptidase domain [61–64]. The structure of the well characterised subtilisin E (S08.036) from *B. subtilis* with its propeptide is shown in Figure 4 [51]. The secondary structure consists of three β-sheets in the form of a highly twisted seven-stranded parallel β-sheet with two antiparallel β-sheets and nine α-helices [51]. Subtilisins are Ca^{2+}-dependent and the structure usually includes a strong and a weak Ca^{2+}-binding site [24]. The binding of Ca^{2+} is essential for the stability and/or activity of the enzyme [64]. While the high-affinity Ca^{2+}-binding site in the protein is always fully occupied, even without the addition of $CaCl_2$, the weak binding site is occupied by a Na^+ or K^+ ion at low $CaCl_2$ concentrations [64]. Both sites are located far from the active site, with the high-affinity Ca^{2+} binding site found near the N-terminus and the other one about 32 Å away [65].

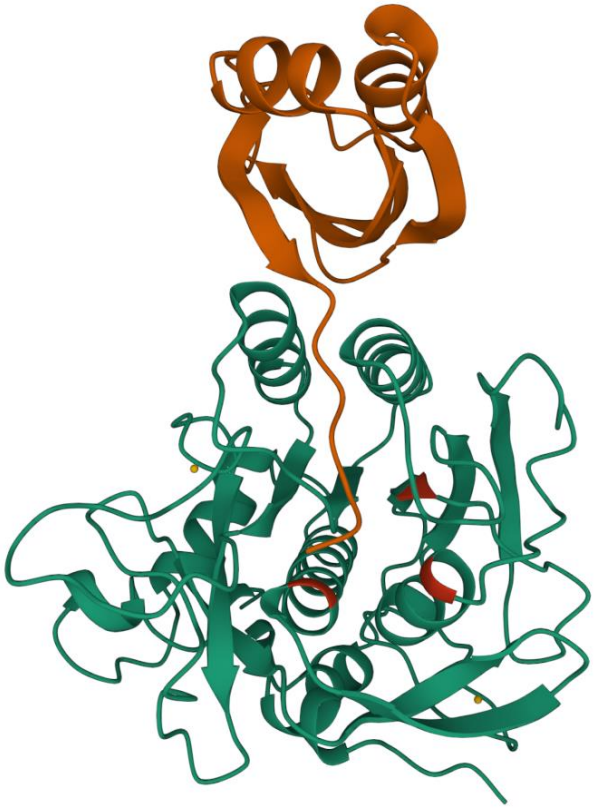


Figure 4 Crystal structure of an autoprocessed Ser221Cys-subtilisin E-propeptide complex (PDB: 1SCJ). Two Ca^{2+} ions are displayed with yellow balls. The catalytic residues Asp^{32}, His^{64}, and $Ser(Cys)^{221}$ are shown in red. The structure of the autoprocessed propeptide is shown in orange [51].

Subtilisins are widely used industrially in various products such as detergents, cosmetics, pharmaceuticals, as well as in processes such as leather processing and wastewater treatment [54]. Sales of alkaline proteases such as subtilisin account for the highest share of the global enzymes market, with 2.2 billion USD in 2019 and an expected size of 2.9 billion USD by the end of 2026 [66, 67]. The idea of using proteases for detergents dates back to the use of pancreatic extracts by Roehm in 1913, but the big breakthrough came in the 1960s with the availability of subtilisins from bacteria [16]. Within the washing liquor they degrade proteinaceous stains, such as milk, egg, blood, grass, and sauces [16]. For use in detergents, subtilisins must meet several properties, such as high catalytic activity and stability at alkaline pH, high temperatures, broad substrate specificity, and stability in the presence of bleaching agents and surfactants [54]. The current trend towards lower washing temperatures in order to save energy and the use of synthetic fibres that cannot tolerate high temperatures has led to the need for subtilisins that are optimal at lower temperatures [54]. The majority of subtilisins used in detergents are isolated from species such as *B. licheniformis* with subtilisin Carlsberg [68], *Lederbergia lenta* (formerly *Bacillus lentus*) with Savinase [69], *Alkalihalobacillus alcaliphilus* (formerly *Bacillus alcaliphilus*) with subtilisin PB92 [70], and *B. amyloliquefaciens* with BPN' [23].

The first generation of subtilisins used for detergents showed high activity at pH 9 - 10 and belonged to the group of true subtilisins [16]. With the discovery of high-alkaline subtilisins with high activity at pH 11 - 12, the second generation of subtilisins for use in detergents was established. A selection of commercial subtilisins and their organism source used in liquid, powder, or in machine dishwashing detergents is listed in Table 2.

Table 2 Commercial protease products for detergent applications

Product	Organism source	Application
Alcalase®	*Bacillus licheniformis*	L
BLAP S	*Lederbergia lenta*	P, D
BLAP X	*Lederbergia lenta*	P
KAP	*Shouchella clausii*	P
Liquanase®	*Bacillus* spec.	L
Polarzyme®	*Bacillus* spec.	L
Properase®	*Bacillus* spec.	D
Purafect™	*Lederbergia lenta*	P
Purafect Ox™	*Lederbergia lenta*	P
Purafect Prime™	*Lederbergia lenta*	L
PUR	*Lederbergia lenta*	L
Savinase®	*Lederbergia lenta*	P

L: liquid detergent, P: powder detergent, D: automatic dishwashing detergent; based on Sahm *et al.* [71]

Since the discovery of subtilisin Carlsberg for the detergent industry, a number of alkaline proteases, mainly from different genera of the *Bacillaceae* family, but also from *Streptomyces*, have been characterised and explored for various biotechnological applications [72]. In these studies, most of these proteases were referred to as subtilases with limited reports on other peptidase families. Most of the published articles dealing with proteases for use in detergents relate to purification and characterisation, but recombinant production of the protease is described in only few cases [72]. For example, Mechri and co-workers reported about subtilisins from *Virgibacillus natechei* and *Anoxybacillus kamchatkensis* M1V, which both showed similar or higher stability and compatibility in detergents compared to commercial proteases [73, 74]. A subtilisin from the psychrophilic *Bacillus pumilus* strain BO1 effectively removed blood stains under cold washing conditions (20°C) [75]. The subtilase from *Thermoactinomyces vulgaris* CDF showed also high compatibility with commercial laundry detergents [76]. In general, only a few selected proteases are suitable for use in liquid detergents, as many have insufficient catalytic activity in such environments. Sufficient cleaning performance against soiling on cotton is determined by the degree of cleaning of the washed fabric under test conditions. Here, the proteases act on substrates bound to the surface of

usually cotton and not on soluble substrates in solutions [16]. Proteinaceous stains on cotton are less accessible to enzymatic degradation due to ageing, heating and oxidation, in contrast to biochemical environments where denaturation of protein substrates normally leads to enhanced enzyme activity [16].

Since most of the subtilisins used are sensitive to oxidising agents, the enzyme is protected from oxidation by bleaching agents in the detergents by site-directed mutagenesis, in which a methionine residue near the active Ser residue is replaced by a non-oxidisable amino acid [67]. Enzyme engineering was also used to lower the optimal temperature for effective cold washing at 10 -20 °C [67, 77]. Various other enzyme engineering approaches were carried out to improve stability and activity with respect to pH, temperature, detergents, and oxidants, resulting in the third generation of subtilisins for use in detergents [77–83].

Due to their activity at alkaline pH values, subtilisins also have potential for applications in the food industry, e.g. for the removal of allergens. Alcalase® (Subtilisin Carlsberg) reduced IgE reactivity in roasted peanut products, reduced fish allergen content, and removed proteins from milk that induce an immunological response in non-breastfed infants [84–86]. In the context of pharmaceutical applications, Chen *et al.* reported an engineered subtilisin variant with high specificity against the rat sarcoma oncoprotein could be used to treat cancer, as this oncoprotein causes one third of human cancers [87]. Nattokinase from *B. subtilis* natto showed fibrinolytic activity and might be used as a potential agent for thrombolytic therapy. In addition it increases the secretion of tissue plasminogen activator from cells [88]. Subtilisins can be also used for peptide synthesis rather than hydrolysis when catalysis takes place in organic solvents, for example [89]. In leather and textile processing, subtilisins can be used instead of harsh chemicals for hair removal [67]. In addition, an application in waste management is possible, as subtilisins can, for example, degrade fish biowaste generated in the fish processing industry and convert it into protein hydrolysates and chitin-containing materials that could be used in other industries [90].

1.1.5 The search for new proteases

Approaches to discover and engineer new and improved enzymes for industrial applications are rapidly growing [91]. In addition to elaborate protein engineering studies, there is the *de novo* design of enzymes, which is still in its infancy due to the limited understanding of structure and function [91]. However, nature, and in particular the almost unlimited microbial biodiversity, harbours great untapped potential for contributing to the enrichment of the repertoire of known and new enzymes [92]. Several methods are available to exploit this potential. The classical approach is the isolation and cultivation of microorganisms from a specific environment with extraction of genomic DNA or purification of proteins and identification of genes, which can be an effective but

time-consuming procedure [91]. However, the development of high-throughput sequencing technologies has made it possible to extract an exponentially growing amount of sequence information from genome sequencing projects [93, 94]. Here, predictive bioinformatic tools made it possible to search for open reading frames and automatically annotate putative enzymes [91]. Since most microorganisms cannot be cultured, metagenomic analysis is possible, in which the genetic information of the entire microbial community in a given environment is examined [94]. Therefore, the characterisation of proteases lags far behind the available sequences in the database [93]. In particular, the identical protein groups database (https://www.ncbi.nlm.nih.gov/ipg) of the National Center for Biotechnology Information (NCBI) contains a single entry for each protein translation found in different NCBI sources and combines annotated coding regions from different databases such as PDB, SwissProt, GenBank and RefSeq [95]. The database displays an excellent source for mining new enzymes [91]. Here, either a keyword search can be performed or the amino acid sequence of a reported enzyme with the desired properties can be used as a template for a BLAST search that yields homologous protein sequences [91]. However, one aspect of such huge number of uncharacterised proteases is the prioritisation of characterisation. Here it is important to look at the microbial origin. As already mentioned, the genus *Bacillus* is particularly important for subtilisins. Since several bacteria within the genus *Bacillus* have been reclassified into genera such as *Alkalihalobacillus*, *Metabacillus*, *Cytobacillus*, etc., in order to divide the known *Bacillus* species into different clades, the entire family of *Bacillaceae* is interesting in this respect [96, 97]. For *Bacillaceae*, multiple subtilase-encoding genes are also more likely to be found than a single gene, showing the large amount of sequences that can be exploited [33]. Most of the industrially used subtilisins originate from mesophilic *Bacillus* species [98–100]. However, enzymes derived from extreme environments gained attention since they have the potential to fulfil several industrial needs which will be outlined in the following [101].

1.1.6 Extremozymes and where to find them

Extreme environments offer the potential to yield novel microbial diversity and enzymes with interesting properties that can withstand harsh conditions [102]. Enzymes isolated from organisms living in extreme environments such as the Antarctic, geothermal hot springs or saline environments are called extremozymes [103–106]. They have evolved to withstand extreme physico-chemical conditions and can be divided into two classes. The first class are enzymes that are adapted to the respective environment, e.g. to extreme temperatures or pressures that reach into the intracellular milieu, so that intracellular and extracellular enzymes are adapted to these conditions [101]. In other environments having for example extreme pH values or salinity, extremophilic microorganisms persist in such environments by homeostatic mechanisms allowing to maintain intracellular conditions that are similar to those of non-extremophiles [101]. Therefore, not all

enzymes from extreme environments are extremozymes, because the intracellular enzymatic functions of the second class are not adapted to the extreme conditions, whereas the extracellular enzymes are [101]. In particular, subtilisins derived from bacteria with a halotolerant or halophilic background provide immense potential for meeting industrial demands, as shown by the increasing number of newly characterised proteases [72, 107, 108]. They have properties such as high tolerance to pH, temperature and osmotic stress and therefore have great practical importance for industrial applications [90, 109–111].

Halophiles are a heterogeneous group of organisms that have the ability to survive or grow best in a highly saline environment [112]. Several classifications for halophilic bacteria have been proposed, but one of the most widely used schemes states that halophilic bacteria are divided into slightly halophilic (when they grow best between 0.2 and 0.5 M salt), moderately halophilic (0.5 - 2.5 M salt) and extremely halophilic (above 2.5 M salt) species [113, 107]. Bacteria that can grow both in the absence of salt and in the presence of relatively high salt concentrations are called halotolerant or extremely halotolerant if growth occurs above 2.5 M salt [114, 107]. They occur in all three domains of life, and hypersaline niches are inhabited by extremely and moderately halophilic archaea and bacteria [115]. The best studied examples of extreme halophiles are the archaeal representatives of the class *Halobacteria*, most of which show optimal growth above 2.5 M salt [116]. Moderately halophilic and halotolerant bacteria are found in a large number of phylogenetic branches, including species belonging to the *Pseudomonadota* (formerly *Proteobacteria* (*Gamma-* and *Alphaproteobacteria*)), *Bacillota* (*Firmicutes*) and *Actinomycetota* (*Actinobacteria*) [117]. Halophily is less widespread in eukaryotes and is mainly found in a few species of green algae, filamentous fungi, yeasts and protozoa [118]. Halophilic organisms are mainly isolated from saline environments, such as salt lakes, saline soils and marine sediments, but also from non-ordinary places such as textile effluents, halophytes, mine wastes and processed foods [107].

Microbial adaptation to high concentrations of soluble salts is facilitated by achieving at least osmotic equilibrium with the surrounding medium in the cytoplasm [107]. The dissolved salts increase the osmotic potential of the environment, which removes water from the cell, leading to cell death [107]. There are two main adaptive mechanisms for maintaining high osmotic pressure within the cell [115]. The first strategy is the "salt-in" strategy, in which the enrichment of inorganic ions such as K^+ serves to balance the osmotic pressure [118]. Second, the accumulation of water-soluble, low-molecular organic compounds called compatible solutes or osmolytes which replace the inorganic ions [118, 119]. The accumulation of compatible solutes occurs through uptake or synthesis, e.g. of sugars, polyols and amino acids [118]. In the event of a salt depletion, the compatible solutes being excreted or degraded and metabolised [118, 120]. In *Bacillus*

species, the cell wall is reinforced with teichuronic acid, a negatively charged layer that surrounds the peptidoglycan cell wall [121]. The negative charges generated by the polymers are thought to repel hydroxy ions and adsorb sodium and hydrogen ions, protecting the cell from high salinity and high pH [122]. Additionally, the adaptation of Bacilli to high salinity and high pH is believed to rely heavily on effective $Na^+(K^+)/H^+$ antiporters [123].

1.1.7 Protein adaptation to extreme environments

The temperature is a key factor to determine the dynamics and catalytic efficiency of an enzyme [124]. An increase in temperature can lead to increased mobility and instability of the proteins, which is why thermophilic enzymes (50 - 80 °C) require higher structural rigidity and a more compact conformation to withstand high temperatures without thermal denaturation [124]. The adaptation of proteins enabling stability at high temperatures is attributed to an increased number of charged residues, disulphide bonds and a hydrophobic core [125]. Furthermore, the interaction between the solvent and the protein plays a key role in protein folding, flexibility, stability and function [124]. With increasing temperature, the entropy of the solvent increases as the hydrogen bonds between the solvent molecules are weakened, resulting in exposed hydrophobic groups being forced inside the protein, which stabilises the protein to a certain extent [126]. The amino acid composition and the surface polarity can therefore lead to different interactions with the solvent and thus to different thermotolerance. A higher hydrophobicity of a protein often leads to a more compact structure and thereby to a higher resistance to heat denaturation [124]. In addition, chaperones are produced that help to prevent heat denaturation or refold the proteins back to their native functional state after heat denaturation [127]. On the other hand, enzymes that are psychrophilic (<15 °C) are more flexible in their structure, which allows them to make conformational changes during catalysis and improve catalytic efficiency at lower temperatures [128].

The pH value has a major influence on enzyme activity, as the pH value affects the ionisation state of the acidic or basic amino acids, which in turn influences the build-up of salt bridges and hydrogen bonds that maintain the three-dimensional shape of the protein. A change in structure can lead to changes in protein function or inactivation of the enzyme. Comparison of the crystal structures and amino acid sequence between the high-alkaline subtilisin M-protease from *Bacillus* sp. KSM-K16 and the true subtilisin Carlsberg from *Bacillus licheniformis* showed that the number of negatively charged amino acids Asp and Glu as well as the positively charged Lys is reduced in the M-protease, while the number of Arg, His, Asn and Gln is increased [129]. This indicates that adaptation to highly alkaline conditions is accompanied by an increase in the isoelectric point (pI). The Arg residues contribute to an increase in hydrogen bonds or ion pairs in the peptidase, as the pK_a of the guanidino group of Arg (ca. 13.8) is higher than that of the ε-amino group of Lys (ca. 10.5) and can form ion pairs with amino acid residues more easily under alkaline

conditions [130, 129]. His, Asn and Gln are uncharged at high alkaline pH and might help to maintain protein solubility in water [129].

Since salt adaptation is based on a high proportion of negative or positive charges on the surface, this is related to adaptation to a high pH value [131]. High salt concentrations lead to a disruption of the local water structure around the protein, reducing the propensity for intermolecular hydrogen bonding, which affects protein solubility, stability and crystallisation. In addition, the surface tension of water increases, pulling the essential water layer away from the protein surface and increasing hydrophobic interactions, leading to protein aggregation and precipitation [132]. Therefore, the charged amino acids play a crucial role in building a hydrate shell around the protein that maintains solubility. Here, the negatively charged amino acids are most beneficial for protein solubility, followed by the positively charged amino acids and finally the charge-neutral amino acids [133]. Thus, halophilic enzymes contain a relatively high proportion of acidic amino acids (Asp and Glu), which leads to a reduction of the theoretical isoelectric point (pI) (4.2 - 6.8) [131]. In contrast to halophilic enzymes, halotolerant enzymes are defined as enzymes that maintain their activity over a wide range of salt concentrations without salt dependence [111]. The comparison of the number of acidic amino acids (Glu and Asp) to the number of basic amino acids (Lys, His, and Arg), referred to as the AB ratio can be used to predict salt adaptation [111, 134]. However, the mechanism behind the adaptation of halotolerant enzymes to moderately high salt concentrations (i.e. 0.5 - 2.5 M NaCl) is poorly understood. Comparisons of the amino acid sequences of a halotolerant subtilisin with homologous non-halotolerant subtilisins revealed that at least eight acidic and polar amino acids in the non-halotolerant group were replaced by non-polar amino acids [131, 135].

The biggest problem of proteases in terms of their adaptability to extreme environments is autoproteolysis. Autoproteolysis refers to the hydrolytic cleavage of the protease itself in a mostly intermolecular mechanism (cleavage of one protease by another) and is coupled to subsequent unfolding [136]. Proteases are protected from autoproteolysis in their native conformation because they are normally more compactly packed than other proteins, but occurs during unfolding because the unfolded molecules are a substrate for the molecules that are still native and active [137]. This is a particular problem for non-specific proteases such as subtilisins and leads to the inactivation and irreversibility of unfolding [138]. Therefore, in addition to structural stability, autoproteolytic stability must also be taken into account for proteases [136].

1.2 Protease production

For the production of recombinant proteins on an industrial scale, different prokaryotic and eukaryotic expression systems are used depending on the target protein [139]. Bacteria in particular are interesting for production because they are easy to handle and tools for genetic manipulation exist for many bacteria [140]. Industrial production must be cost-efficient, with high yields and low efforts for protein purification. The genus *Bacillus* has become one of the most important hosts for the production of subtilisins [141, 142]. In particular, *Bacillus subtilis* is widely used as a cell factory due to its highly efficient protein secretion system, adaptable metabolism and simple and diverse genetic manipulation systems [143]. *B. subtilis* is an aerobic, gram-positive soil bacterium and, with its versatile metabolism, easy to cultivate on cheap substrates, with short fermentation cycles of about 48 hours [143]. These cheap substrates are by-products of agricultural production such as molasses from sugar cane and corn starch for submerged fermentation or various types of bran and cake for solid-state fermentation [144]. Protein secretion is facilitated by the lack of an outer membrane, which is a major advantage as it simplifies separation from the cell biomass which facilitates downstream processing and reduces process costs [143]. Importantly, products synthesized with *B. subtilis* are usually granted the Generally Regarded as Safe (GRAS) status, such as vitamins, inositols, acetoin, hyaluronan, or fermented soybeans [143, 145, 146].

Bacillus species naturally secrete various hydrolytic enzymes such as proteases, amylases and lipases to respond to changing environmental conditions and to hydrolyse complex polymers in the cellular environment and import the resulting monomers as nutrients [147]. There are two main pathways of protein secretion in bacteria, the Sec pathway and the twin-arginine translocation (Tat) pathway [139]. In bacteria, the ubiquitous general secretion pathway Sec is the most important pathway for proteins exported from the cytosol [148]. Proteins secreted via the Sec pathway are synthesised as precursor proteins with an N-terminal signal peptide responsible for targeting the membrane-bound Sec translocase [149]. The actual membrane translocation takes place in an unfolded state of the protein and occurs cotranslationally or posttranslationally [150]. During or immediately after translocation, the N-terminal signal peptide is cleaved off by the signal peptidase [139]. Finally, proteins in the extracellular space can be folded with the help of extracellular chaperones [151]. In the Tat pathway, the protein contains a specific twin arginine signal peptide and is secreted in the folded state, sometimes in an oligomeric form or together with another protein [139]. As mentioned above, the secretion of a target protein into the growth medium simplifies subsequent downstream processing as no cell disruption is required, thus reducing overall production costs [152]. Since translocation is one of the known constraints for protein production of subtilisins, efforts have been made to find the optimal signal peptide [153]. However, this must always be checked for each individual protein and production host [139]. Other possible limitations

are found in transcription, translation, and the subsequent maturation of the protein to its native conformation [143].

Since *Bacillus* species naturally produce proteases, strains that lack the genes for the major extracellular proteolytic activity are used [151]. Other bacterial hosts such as *B. licheniformis*, *S. clausii* or *B. amyloliquefaciens* are extremely important for the production of commercial proteases, as they have a higher enzyme yield than *B. subtilis* [142]. However, in contrast to *B. subtilis*, effective introduction of new DNA, effective chromosome manipulation and achieving high transformation rates have proven to be more challenging in these strains [142]. In contrast to *B. subtilis*, strains such as *B. licheniformis* encode the glyoxylate cycle, which is necessary to use overflow metabolites such as acetate or acetoin as a carbon source [154]. In highly optimised industrial fermentations, the *Bacillus* strains are capable of producing extracellular proteins in concentrations of 20 g/l [147]. However, the high product yields are predominantly achieved only for homologous proteins or for proteins derived from close relatives [139]. Hence, the yield of heterologous proteins is lower or the target proteins are not produced at all [155]. Therefore, different alternative secretory expression system such as *Streptomyces lividans* [156], *Corynebacterium glutamicum* [157] and *Lactococcus lactis* [158] are used as alternatives. In addition, other strains of the *Bacillaceae* family can serve as alternative production hosts, either for homologous or heterologous protein production. One of these alternatives could be the strain *Metabacillus indicus* DSM 16189, which is investigated in this work.

1.2.1 *Metabacillus indicus* DSM 16189

Metabacillus indicus strain DSM 16189 was first described as *Bacillus cibi* JG-30 [159]. However, due to its close genetic relationship to *Bacillus indicus* LMG 22858 it was renamed to *Bacillus indicus* [160]. Phenotypic characterisation of the two strains revealed only differences in the utilisation of mannose and galactose and a slight variation in pigmentation. The 16S rRNA gene sequences derived from the genome show an average similarity of 99.74 %. The nucleotide sequence identity of the genomes, amino acid sequence identity of orthologous proteins and DNA-DNA hybridisation values also exceed the species thresholds, which is why the new classification as *Bacillus indicus* DSM 16189 was necessary [160]. Based on further genome analysis, the genus was subsequently renamed to *Metabacillus indicus* DSM 16189 [96].

M. indicus DSM 16189 is a Gram-variable and halotolerant microorganism [159]. In addition, it is identified as an endospore former. The aerobic rods with a size of 0.6 - 0.8 µm x 1.5 - 3.5 µm have a peritrichous flagellation that makes the organism mobile. *M. indicus* was isolated from the seafood jeotgal, a traditional Korean fermented food. Its colonies are described as smooth and circular with yellow-orange colouration. The optimal pH for growth is between 6.5 and 7.5. It grows in the presence of 0 - 12 % (w/v)

NaCl, with the values for optimal growth being 0 - 1 % (w/v) NaCl. The yellow-orange colouration of the colonies is caused by carotenoids, which are present in vegetative cells as yellow pigment and in the spores as orange pigment [161].

1.3 Aims of this thesis

The search for new subtilisin sequences and their characterisation is particularly interesting for industrial applications such as detergents. Consequently, the first part of this work aimed at finding new subtilisins through a data mining-based method. Since most detergent-tolerant subtilisins are from the *Bacillaceae* family, the focus should be on this family. To classify the data mining results, a comprehensive phylogenetic overview of the subtilisin and superordinate classic subtilase family (D-H-S family) is necessary, as last one was given by Siezen and Leunissen 1997, when much less sequences were available [24]. Publicly available sequence data should be used in comparison to scientifically described proteases and their sequences. In the context of subtilisins, several new candidates should be selected of halophilic or halotolerant origin, as these subtilisins are most likely to meet industrial needs. In the second part of this work, the selected candidate genes should be cloned into a production plasmid by Golden Gate cloning and the plasmid then used to transform *Bacillus subtilis* DB104 via natural competence. Expression, secretion and purification of the gene products should be followed by comprehensive biochemical characterisation, comparison with each other and with previously characterized subtilisins. In addition, they should be evaluated for their technical applications, focusing on their use in detergents. In particular, the biochemical characterisation is necessary to better understand the relationship between the sequence, structure and the biochemical properties. The expected difficulties with respect to activity and stability, especially due to autoproteolysis, should be overcome. Although *Bacillus subtilis* DB104 is an established laboratory production host, limitations in the production of heterologous subtilisins are also very likely. Therefore, in the third part of this work, the wild-type strains, from which the new subtilisin sequences were obtained, should be evaluated as alternative production hosts. For this purpose, genetic accessibility and especially triparental conjugation should be investigated [162].

2 Results

2.1 Phylogenetic survey of the subtilase family and a data-mining-based search for new subtilisins from *Bacillaceae*

Fabian Falkenberg, Michael Bott, Johannes Bongaerts and Petra Siegert (2022). Frontiers in Microbiology, DOI: 10.3389/fmicb.2022.1017978

Authors contributions:
FF collected and analysed the data and wrote the original draft.
PS and JB helped with the data analysis.
JB, MB and PS supervised the study and revised the manuscript.

All figures and tables are from this study.

Overall contribution FF: 95 %

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 26 September 2022
DOI 10.3389/fmicb.2022.1017978

OPEN ACCESS

EDITED BY
Javier Pascual,
Darwin Bioprospecting Excellence, Spain

REVIEWED BY
Bing Tang,
Wuhan University, China
Dominik Łagowski,
University of Life Sciences of Lublin, Poland

*CORRESPONDENCE
Petra Siegert
siegert@fh-aachen.de

SPECIALTY SECTION
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

RECEIVED 12 August 2022
ACCEPTED 30 August 2022
PUBLISHED 26 September 2022

CITATION
Falkenberg F, Bott M, Bongaerts J and Siegert P (2022) Phylogenetic survey of the subtilase family and a data-mining-based search for new subtilisins from *Bacillaceae*. *Front. Microbiol.* 13:1017978. doi: 10.3389/fmicb.2022.1017978

Phylogenetic survey of the subtilase family and a data-mining-based search for new subtilisins from *Bacillaceae*

Fabian Falkenberg[1], Michael Bott[2], Johannes Bongaerts[1] and Petra Siegert[1]*

[1]Institute of Nano- and Biotechnologies, Aachen University of Applied Sciences, Jülich, Germany, [2]Institute of Bio- and Geosciences, IBG-1: Biotechnology, Forschungszentrum Jülich, Jülich, Germany

The subtilase family (S8), a member of the clan SB of serine proteases are ubiquitous in all kingdoms of life and fulfil different physiological functions. Subtilases are divided in several groups and especially subtilisins are of interest as they are used in various industrial sectors. Therefore, we searched for new subtilisin sequences of the family *Bacillaceae* using a data mining approach. The obtained 1,400 sequences were phylogenetically classified in the context of the subtilase family. This required an updated comprehensive overview of the different groups within this family. To fill this gap, we conducted a phylogenetic survey of the S8 family with characterised holotypes derived from the MEROPS database. The analysis revealed the presence of eight previously uncharacterised groups and 13 subgroups within the S8 family. The sequences that emerged from the data mining with the set filter parameters were mainly assigned to the subtilisin subgroups of true subtilisins, high-alkaline subtilisins, and phylogenetically intermediate subtilisins and represent an excellent source for new subtilisin candidates.

KEYWORDS

***Bacillaceae*, S8 protease family, subtilisin, data mining, subtilase, phylogenetic analysis**

Introduction

The subtilase family or subtilisin-like proteases defined by the MEROPS database as S8 family is the third largest family of serine proteases, both in terms of number of sequences and characterised peptidases, which are represented in microorganisms (archaea, bacteria, fungi, yeast) as well as in higher eukaryotes (Rawlings et al., 2014). The MEROPS database is a comprehensive source of information for proteases. It uses a hierarchical, structure-based classification, and based on statistically significant amino acid sequence similarities, proteases are grouped into families and clans (Rawlings et al., 2014). Here, the S8 family belongs to the clan SB, which is one of 13 clans of serine

proteases and, in addition to the S8 family, also contains the S53 family (sedolisin family) (Page and Di Cera, 2008). The two families differ in their catalytic mechanism in that within S8 they form a catalytic triad with their active residues in the order Asp, His, Ser, referred to as the "classical" D-H-S family (Siezen et al., 2007). While protein folding of members of the family S53 is very similar to that of subtilisins, the catalytic triad has been altered to Glu, Asp, and Ser, which is referred to as the ED-S family (Siezen et al., 2007). Furthermore, the S8 family is subdivided into the two subfamilies S8A (subtilisin as a type example) and S8B (kexin as a type example) (Rawlings, 2020). In addition, subtilases were classified by Siezen and Leunissen (1997) into six groups based on sequence alignments of the catalytic domain, namely subtilisins, thermitases, proteinase K, lantibiotic peptidases, pyrolysins and kexins, while mentioning that further subdivision may become useful with more available sequences. The kexins form the subfamily S8B, while the other five groups belong to the subfamily S8A. In the MEROPS database, well-characterised specimens are selected and designated as "holotypes" at the subfamily level (Rawlings et al., 2014). There are currently 186 holotypes listed for S8A and 21 for S8B (Rawlings and Bateman, 2021). Uncharacterised homologs of a holotype were assigned to the same MEROPS identifier (Rawlings and Bateman, 2021). Here, only the catalytically active part of the protease (peptidase unit) is considered, and a new holotype is created when a protein is characterised that either has a different specificity than another protein in the subfamily or the same specificity but a different cellular location, has a different architecture, or the sequence in a phylogenetic tree does not cluster with that of an existing holotype with similar specificity (Rawlings and Bateman, 2021).

The known members of the S8 family are endopeptidases, with the exception of TPPII (tripeptidyl peptidase II), which releases tripeptides from the N-terminus of peptides (Siezen and Leunissen, 1997; Renn et al., 1998). In most bacteria, archaea, and lower eukaryotes they are mostly unspecific proteases and are involved in nutrition (Siezen and Leunissen, 1997). They fulfil other functions as well, as they are involved in developmental processes and immune responses in plants (Schaller, 2013), play a role in the metabolism of neuropeptides in *Drosophila melanogaster* (Renn et al., 1998), or are involved in pathogenesis (Garcia-Sanchez et al., 2004). Several subtilases contain a C-terminal extension, relative to the subtilisins, with additional properties such as sequence repeats, Cys-rich domains as cell surface anchors, or transmembrane segments (Siezen and Leunissen, 1997). Except for the subtilase ASP (*Aeromonas sobria* protease), an N-terminal propeptide acts as an intramolecular chaperone during maturation, supporting the folding of the catalytic domain (Zhu et al., 1989; Eder et al., 1993; Kobayashi et al., 2015). Members of the subtilase family find a wide range of applications in industry, such as lactocepins playing an important economic role in the industrial production of cheese and fermented milk (Broadbent and Steele, 2013). Of particular interest within this study are subtilisins, which find applications in several industrial sectors such as in detergents, leather processing, food, wastewater treatment, cosmetics, and pharmaceuticals (Kalisz, 1988; Solanki et al., 2021; Azrin et al., 2022). They are typically isolated from various species of the genus *Bacillus* such as *B. subtilis*, *B. licheniformis*, *Shouchella clausii* (formerly *Bacillus clausii*), etc. (Kalisz, 1988; Christiansen et al., 2003; Maurer, 2004; Joshi et al., 2021). They consist of about 270 amino acids and are secreted *via* the Sec-secretion pathway in a precursor form containing a signal peptide of about 28 amino acids and a propeptide of about 75 amino acids (Markland and Smith, 1971; Power et al., 1986; Siezen et al., 1991; Tjalsma et al., 2000).

Due to the increasing number of genome sequencing projects, the amount of data on uncharacterised proteins is growing exponentially (Rawlings, 2013). This rapidly increasing online database provides an excellent resource for broadening the sequence space of subtilisins. Our research on uncharacterised subtilisin sequences began with the analysis of the MEROPS S8 dataset in a phylogenetic tree containing only characterised proteases. The analysis quickly revealed the presence of previously uncharacterised groups and subgroups within the S8 family and motivated us to update the phylogeny of this family, which was necessary to place the uncharacterised sequences in this context. Sequences from a data mining approach for new subtilisin proteases from the *Bacillaceae* family were then evaluated and placed in the context of the S8 subfamilies and groups.

Materials and methods

Sequence-based phylogenetic analysis

The amino acid sequences of the mature part, referred to as the "peptidase unit" in the MEROPS database, comprising the structural domain of the protein directly responsible for peptidase activity and substrate binding, including the larger insertions compared to other subtilases (Rawlings et al., 2018). Other structural domains, if present, were excluded, such as the signal peptide, the propeptide and C-terminal domains. The sequences were aligned using MAFFT v 7.490 with L-INS-I parameter[1] (Katoh and Standley, 2013; Katoh et al., 2019). The alignment was trimmed using trimAi v1.2 with the "gappyout" parameter[2] (Sánchez et al., 2011). The phylogeny was made using iqtree v1.6.12[3] (Trifinopoulos et al., 2016) with automated ModelFinder (Kalyaanamoorthy et al., 2017) and ultrafast bootstrap (Hoang et al., 2018) options. Phylogenetic trees were

1 https://mafft.cbrc.jp/alignment/server/

2 http://phylemon.bioinfo.cipf.es/

3 http://iqtree.cibiv.univie.ac.at/

displayed and annotated with the iTOL software[4] (Letunic and Bork, 2021).

Data mining

Holotype protein sequences from the S8 family for the analysis were obtained from the MEROPS database[5] (Rawlings et al., 2014). In addition, a selection of sequences of characterised subtilases was chosen from the Protein Data Bank (PDB). Only the mature part of the proteases was used as described above.

To search for uncharacterised subtilisin sequences from *Bacillaceae*, new amino acid sequences for the analysis were obtained from the NCBI Identical protein groups database[6] by searching for "S8 peptidase *Bacillaceae*." To selectively search for subtilisins, a filter was set for peptide sequences with a length of 350–410 amino acids. The resulting dataset was clustered with a identity threshold of 85% by using CD-HIT[7] (Huang et al., 2010). Intracellular proteases were excluded by analysing the sequences with the SignalP 6.0 prediction tool[8] and including only protein with a predicted Sec signal peptide (Teufel et al., 2022). The propeptide was removed manually after alignment with Clustal Omega[9] (Sievers et al., 2011), using the JalView alignment annotation software[10] (Waterhouse et al., 2009). MSA was drawn with ESpript 3.0 using %strict option (percentage of strictly conserved residues per column) for the colouring scheme[11] (Robert and Gouet, 2014).

Bioinformatic analysis

The isoelectric point of a protein was calculated with the sequence manipulation suite v2[12] using pK_a values from DTAselect (Stothard, 2000).

Results and discussion

Phylogenetic tree analysis of the MEROPS S8 holotype dataset

The first part of this study aimed at gaining a comprehensive overview of the S8A family in order to be able to categorise the sequences obtained from the data mining approaches in the second part. Therefore, a phylogenetic tree was first constructed containing only the MEROPS holotype dataset of the S8 peptidase unit and a selection of sequences of biochemically characterised proteases from the PDB with a total of 168 sequences. The additional PDB sequences were added to support some of the subfamilies described in literature. The amino acid sequences of the mature part, referred to as the "peptidase unit" in the MEROPS database, were used as described in methods. Other structural domains, if present, were excluded, such as the signal peptide, the propeptide and C-terminal domains. In uncurated datasets, it becomes difficult to identify N- and C-terminal extensions in a wide range of different sequences. During the alignment curation, trimAI reduced the alignment length to 248 positions as opposed to 2,330 positions without curation. Curation with trimAI, which uses a less stringent algorithm, was preferred to more stringent filtering methods, as these often result in the deletion of positions in the alignment that contain a gap (Tan et al., 2015). However, gaps can contain significant phylogenetic information (Dessimoz and Gil, 2010), not to mention that a sequence dataset may contain an incomplete or incorrect sequence, which can result in a large loss of information if left undetected. According to Tan et al. a less stringent filtering algorithm has little impact on tree accuracy and is a trade-off in terms of computation time saved for phylogenetic tree computation (Tan et al., 2015). An overview of the workflow of the analysis and the methods used is shown in **Figure 1**.

The curated alignment was used to create a maximum likelihood tree, the standard option for constructing phylogenetic trees, which, together with the Bayes method, is widely recognised as the most accurate approach in molecular phylogenetics (Kuhner and Felsenstein, 1994; Dereeper et al., 2008). In addition, it is important to statistically evaluate the reliability of the tree, which is usually done using a bootstrap-based bias correction method that calculates the branch support of the tree by repeating the tree construction (Hoang et al., 2018). However, the different parameters for the alignment, the curation methods, and the different tree generation methods result in different phylogenetic trees, making a detailed comparison of a generated tree with literature data difficult. Since the constructed tree is not rooted, an outgroup must be selected for restructuring the tree, which contains a set of sequences that are outside the ingroup but closely related to it (Sanderson and Shaffer, 2002). **Figure 2** shows the phylogenetic tree of S8A subfamily, which is closely related to the S8B subfamily, which was selected as the outgroup (Siezen and Leunissen, 1997). In this tree, we identified the groups proteinase K, pyrolysins, thermitases, subtilisins, lantibiotic peptidases and kexins as described by Siezen and Leunissen (1997), and several subgroups within these groups. However, our analysis revealed that the S8A proteases form more groups and subgroups than previously

4 https://itol.embl.de/

5 https://www.ebi.ac.uk/merops; published before January 19, 2022.

6 https://www.ncbi.nlm.nih.gov/ipg; published before January 31, 2022.

7 http://weizhong-lab.ucsd.edu/cdhit-web-server

8 https://services.healthtech.dtu.dk/service.php?SignalP-6.0

9 https://www.ebi.ac.uk/Tools/msa/clustalo/

10 https://www.jalview.org/

11 https://espript.ibcp.fr/ESPript/ESPript/

12 https://www.bioinformatics.org/sms2

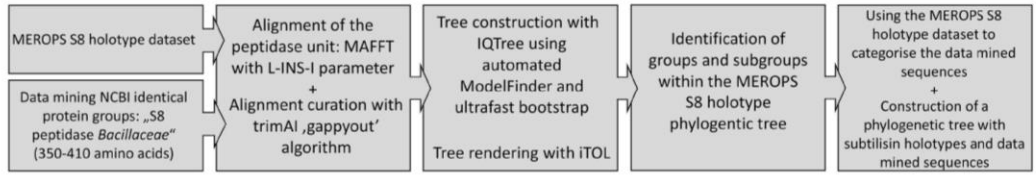


FIGURE 1
Workflow of used data and methods.

described. To account for the diversity resulting from their different positions in the phylogenetic tree, their biochemical properties, biological functions, their structural similarity, and the taxa- and species-specific clusters formed, we propose a revision of the subtilase groups and smaller, better defined subgroups. The naming is based on the criteria mentioned if a connection is recognisable, otherwise they are named according to the protease first described in this group. These groups and subgroups are discussed below in order to place the subtilisins in the context of the subtilases and **Table 1** provides an overview of them. Additionally, each group is shown as a pruned tree (**Supplementary Figures 1–17**).

Pyrolysin group

The pyrolysin group clustered within the phylogenetic tree in 90% of the replicates. Pyrolysins are a heterogenous group of enzymes of diverse origin and low sequence conservation (Siezen and Leunissen, 1997). The average sequence identity of the pyrolysin group present in the phylogenetic tree is 38% (**Figure 2**). Within this branch, lactocepin 1 (S08.116) and lactocepin 3 (S08.019) are described by Siezen and Leunissen as pyrolysins in the gram-positive subgroup (Siezen and Leunissen, 1997; Broadbent and Steele, 2013). However, in the MEROPS phylogenetic tree, together with other sequences of gram-positive bacteria, they form a subgroup of their own, named here cell wall-associated pyrolysins with a bootstrap support of 100% and 51% sequence identity (**Supplementary Figure 1**). Lactocepins are cell envelope-associated endopeptidases of *Lactococci*, which play an important economic role in the industrial production of cheese and fermented milk due to their use as starter bacteria. Rapid growth is ensured by lactocepin, which provides amino acids from milk proteins and ensures autolysis, which is important for cheese ripening (Broadbent and Steele, 2013). They cluster together with other cell wall-associated proteases: S08.020 (Kagawa and Cooney); S08.153 (Pastar et al., 2003); S08.147 (Genay et al., 2009); S08.064 (Bethe et al., 2001); S08.027 (Lawrenson and Sriskandan); S08.138 (Karlsson et al., 2007); S08.118 (Gilbert et al., 1996). An exception is the thermophilic collagenolytic protease from *Geobacillus collagenovorans* MO-1 (S08.142), which contains a collagen-binding segment and is secreted into the culture supernatant without being displayed on the cell surface (Okamoto et al., 2001; Itoi et al., 2006).

Thermicin (S08.029), tengconlysin (S08.135), and the AprX proteases (S08.137) cluster together in the tree in 99% of the replicates (**Supplementary Figure 2**). Thermicin from the extremely thermophilic bacterium *Thermoanaerobacter yonseiensis* KB-1 (Jang et al., 2002) and tengconlysin from *Thermoanaerobacter tengcongensis* (Koma et al., 2007) show both high-temperature optima above 90°C. As mentioned by Jang et al. (2002) thermicin is a novel enzyme that differs from other thermostable proteases and here forms the new subgroup of thermicins. The Bacilli-derived AprX subtilases form another subgroup as previously described, because they lack a signal peptide and exhibit a mesophilic temperature optimum (Valbuzzi et al., 1999; Phrommao et al., 2011). The intracellular subtilase AprX-SK37 from *Virgibacillus* sp. SK37 is a halotolerant, oxidation-stable, and moderately thermophilic alkaline serine protease with properties that could be attractive for various biotechnological applications (Phrommao et al., 2011).

The three proteases CspB (S08.108), CspA (S08.159), and CspC (S08.158) (Clostridial serine proteases) from *Clostridium perfringens* form a new subgroup (Csp pyrolysins) with 45% sequence identity, related to germination and synthesised in the mother cell compartment of spore-forming cells (**Supplementary Figure 3**; Masayama et al., 2006).

Site-1 peptidase (S08.063) from *Cricetulus griseus* is an important processing enzyme of the endoplasmic reticulum/Golgi lumen that acts on sterol regulatory element binding proteins (SREBPs) to regulate cholesterol and fatty acid biosynthesis in addition to other cellular functions (Seidah, 2013b). CP70 (S08.083) from *Flavobacterium balustinum* is a cold-active extracellular protease (Morita et al., 1998); STABLE (S08.096) is a hyperthermostable protease bound to the surface layer of the archaeon *Staphylothermus marinus* and is responsible for the generation of the peptides required in the energy metabolism of the cell (Mayr et al., 1996). Because of their different physiological functions and origins, they will most likely all form new individual subgroups as more homologs are added (Site-1 pyrolysins, CP70 pyrolysins, STABLE pyrolysins). TagA (S08.128) and TagC (S08.127) are

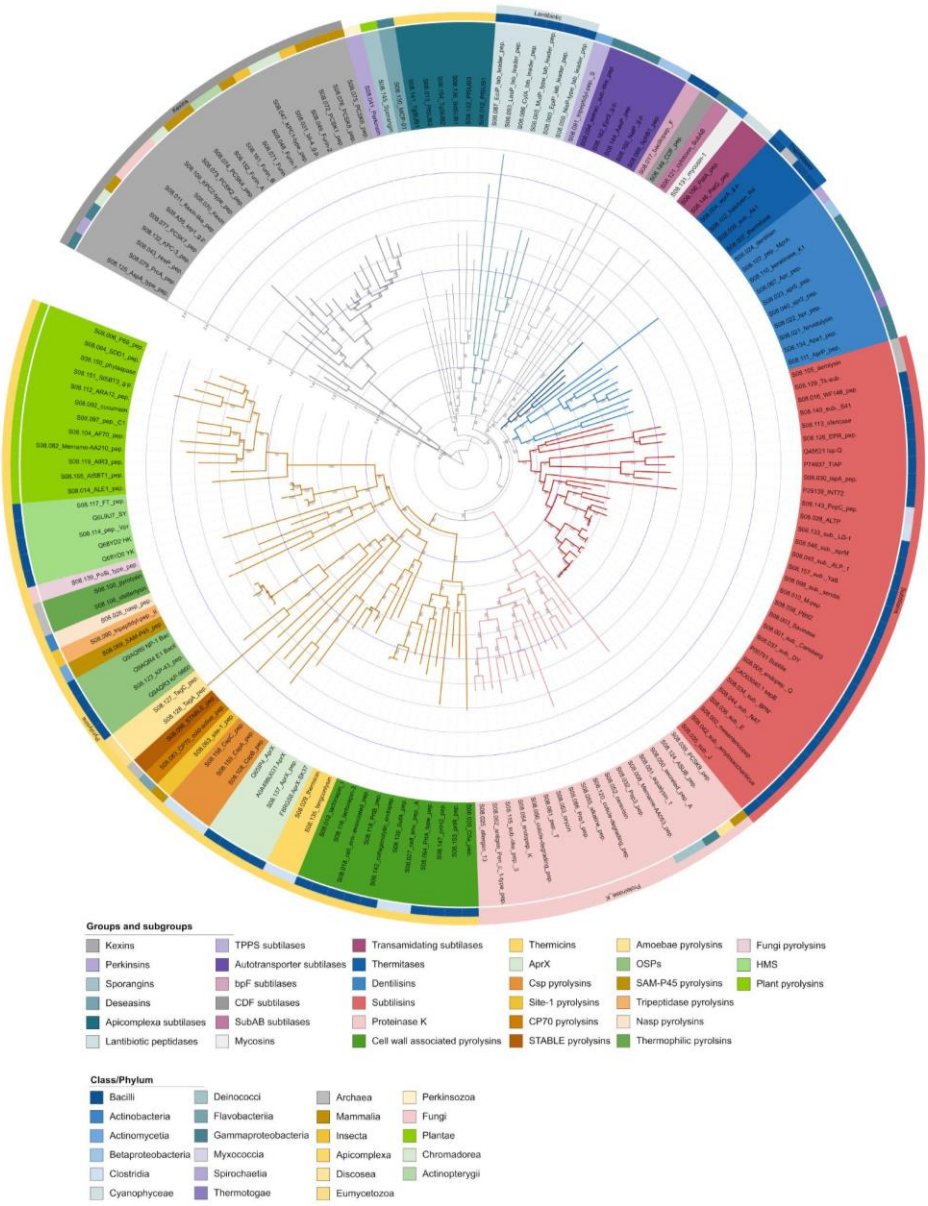


FIGURE 2
MEROPS S8 holotype phylogenetic tree. Phylogenetic relationship of S8 groups and subgroups using the mature protease sequences. The tree was constructed with IQ-TREE by employing the maximum likelihood method with ultrafast bootstrap support (model: LG + I + G4, predicted by Modelfinder, 1,000 replicates). The coloured range within the labels represents the subgroups as shown within the legend. The circles around the phylogenetic tree represent the class or phylum from which the holotype originates, marked by different colours. The kexin proteases S8B were selected as the outgroup. The outer ring represents the group classification after Siezen and Leunissen (1997). For each clade, the numbers above the branches indicate the bootstrap values based on 1,000 repetitions. The tree can be accessed under the following link: https://itol.embl.de/shared/2H14VxXLj30E2.

produced by the amoeba *Dictyostelium discoideum*. While TagA is involved in the differentiation of cell types (Good et al., 2003), TagC is part of a transmembrane protein of the ABC family, which is expressed during the aggregation stage of development (Anjard and Loomis, 2005). TagA and TagC are forming the new subgroup of amoebae pyrolysins with a 100%

TABLE 1 Group and subgroups of the subtilase family.

Group	Subgroup	References
Pyrolysins	Plant pyrolysins	Siezen and Leunissen, 1997
	High-mass subtilases (HMS)	Okuda et al., 2004
	Fungi pyrolysins	Faraco et al., 2005
	Thermophilic pyrolisins	Siezen and Leunissen, 1997
	Nasp pyrolysins	+
	Tripeptidase pyrolysins	Siezen and Leunissen, 1997
	SAM-P45 pyrolysins	+
	Oxidatively stable proteases (OSP)	Saeki et al., 2000
	Amoebae pyrolysins	+
	STABLE pyrolysins	+
	CP70 pyrolysins	+
	Site-1 pyrolysins	+
	Csp pyrolysins	+
	AprX	Phrommao et al., 2011
	Thermicins	+/ Jang et al., 2002
	Cell-wall associated pyrolysins	+
Proteinase K	Gram -	Siezen and Leunissen, 1997
	Fungal	Siezen and Leunissen, 1997
	Yeast	Siezen and Leunissen, 1997
Subtilisins	EPR subtilisins	+
	PopC subtilisins	+
	Extremophilic subtilisins	+
	Archaea subtilisins	+
	True subtilisins	Siezen and Leunissen, 1997
	High-alkaline subtilisins	Siezen and Leunissen, 1997
	Phylogenetically intermediate subtilisins (PIS)	Saeki et al., 2003
Dentilisins		+
Thermitases	Extremophilic thermitases	Siezen and Leunissen, 1997
Transamidating subtilases		Agarwal et al., 2012
Mycosins		Dave et al., 2002
SubAB subtilases		+
CDF subtilases		+
bpF subtilases		+
Autotransporter subtilases		Henderson and Nataro, 2001
TPPS subtilases		+
Lantibiotic peptidases		Siezen and Leunissen, 1997
Apicomplexa subtilases		+
Deseasins		Chen et al., 2007
Sporangins		+
Perkinsins		+
Kexins (S8B)	PC1	Siezen and Leunissen, 1997
	PC2	Siezen and Leunissen, 1997
	Furins	Siezen and Leunissen, 1997
	Yeast Kexins	Siezen and Leunissen, 1997

+Indicates that the group or subgroup was defined within this work.

bootstrap support and 42% sequence identity (**Supplementary Figure 3**).

KP-43 (S08.123) from *Bacillus* strain KSM-KP43 and the sequences within this clade form the subgroup of oxidatively stable serine proteases (OSPs) as described by Saeki et al., with a sequence identity of 92% (Saeki et al., 2000, 2002). However, it should be mentioned that subtilisins not belonging to this subgroup were reported to have higher stability against H_2O_2 (Joshi and Satyanarayana, 2013; Falkenberg et al., 2022). The C-terminal half of KSM-KP43 downstream of the putative catalytic residue, Ser-255, is homologous to the internal segments of TagC (Saeki et al., 2002). While the MEROPS dataset suggests that OSPs are only of bacterial origin, Li et al. (2017) described sequences originating from some species of *Pezizomycotina* fungi.

SAM-P45 (S08.069), a membrane-anchored protease from *Streptomyces albogriseolus* which is considered to be an evolutionary link between primitive bacterial subtilisins and highly diversified eukaryotic proteases, forms its own new subgroup (SAM-P45 pyrolysins) (Suzuki et al., 1997). TPPII, isolated from *Drosophila melanogaster*, has an elongated C-terminus compared to other subtilases and is involved in the metabolism of neuropeptides, which are important signalling molecules in insects and belongs to the tripeptidase pyrolysins subgroup (Siezen and Leunissen, 1997; Renn et al., 1998). The extracellular serine protease Nasp (S08.026) from *Dermatophilus congolensis* is involved in pathogenesis and forms the new subgroup Nasp pyrolysins (Garcia-Sanchez et al., 2004). The two thermophilic proteases from archaea *Thermococcus stetteri* (stetterlysin; S08.106) (Klingeberg et al., 1995) and *Pyrococcus furiosus* (pyrolysin; S08.100), the eponym of the whole group (Blumentals et al., 1990) are both resistant against SDS (1% w/v) and have a high-temperature optimum (85°C, 115°C) forming the thermophilic pyrolysins subgroup (**Supplementary Figure 4**; Siezen and Leunissen, 1997).

PoSI (*P. ostreatus* extracellular protease) (S08.139), a protease from the fungus *Pleurotus ostreatus*, is involved in the activation of other secreted proteases and the post-translational regulation of laccase (Faraco et al., 2005). It shows a high sequence identity with the minor extracellular serine protease from *Bacillus subtilis*, Vpr (S08.114) (31%), and together with other proteases from Ascomycetes and Basidiomycetes forms a separate pyrolysin subgroup (fungi pyrolysins) (**Supplementary Figure 5**; Faraco et al., 2005). A further subdivision of fungal subtilases has been made by others and is beyond the scope of this study (Hu and Leger, 2004; Muszewska et al., 2011; Li et al., 2017).

According to Okuda et al. Vpr (S08.114) belongs to the subgroup of high-molecular-mass subtilisins, which can be divided into at least two classes (**Supplementary Figure 5**; Okuda et al., 2004). One class is less alkaline, its stability depends on Ca^{2+} ions and it is resistant to proteolysis. The other class is strongly alkaline, its stability also depends on Ca^{2+} ions and is

sensitive to proteolysis (Okuda et al., 2004). The subgroup name is slightly misleading, as they do not cluster together with the subtilisins. Therefore, they were named here high-molecular-mass subtilases (HMS). Here, the average sequence identity is 82%.

Plant subtilases are a widely distributed subgroup involved in plant developmental processes and immune responses (Schaller, 2013). The average sequence identity between the investigated sequences is 56% (**Figure 2** and **Supplementary Figure 6**). The first subtilase cloned from plants was the extracellular alkaline protease cucumisin (S08.092) from melon fruit (Kaneda and Tominaga, 1975). The plant subtilases have been divided into seven classes (Xu et al., 2019). A detailed discussion of each of these classes is beyond the scope of this study. Good overviews of the classes and the plant subtilases were provided by Schaller (2013), Taylor and Qiu (2017), and Xu et al. (2019).

Proteinase K group

The alkaline proteinase secreted into the culture medium by the mould *Tritirachium album* Limber is commonly known as proteinase K (S08.054) and is the type example for this group (Ebeling et al., 1974). It can be used to synthesise peptides (Ageitos et al., 2013), and besides peptide bonds, it can also hydrolyse esters (Borhan et al., 1996). In contrast to subtilisins, which contain no cysteine residues, proteinase K contains five Cys residues, four which form two disulfide bridges (Betzel et al., 1990). Because of its remaining activity at higher temperatures (> 60°C) in the presence of urea, 0.5% (w/v) SDS, or 1% (w/v) Triton X100, proteinase K is used for the degradation of proteins and in the preparation of nucleic acids (Sweeney and Walker, 1993; Goldenberger et al., 1995). Most of the proteinase K holotypes found in the phylogenetic tree derive from fungi (**Figure 1**). Worlflow of used data and methods.

There, the fungal proteinase K subgroup is separated from the other proteinase K members and may play an important role in the evolution of pathogenicity, as several entomopathogenic and nematophagous fungi have been characterised as having the ability to destroy the structural integrity of insect or nematode cuticle during invasion and colonisation. Therefore, they are also referred to as cuticle-degrading proteases (S08.120, S08.056) (Leger et al., 1987; Tunlid and Jansson, 1991; Li et al., 2010). For many saprophytes, the subtilases as broad-spectrum proteases play a role in nutrition acquisition, such as digesting proteins to release peptides and amino acids (Gunkel and Gassen, 1989; Hu and Leger, 2004). Further phylogenetic analysis by Li et al. of 138 fungal proteinase K genes revealed a subdivision into five distinct classes (Li et al., 2017). The fungal proteinase K-like proteases are separated from the bacterial ones including aqualysin (S08.051) from the thermophilic bacterium *Thermus aquaticus* (Sakaguchi), the *Amoebozoa* protease ASUB (S08.124) from *Acanthamoeba healyi* (Kong et al., 2000), and the proprotein convertase PCSK9 (S08.039) from *Mus musculus* (Seidah, 2013a; **Supplementary Figure 7**). The average sequence identity within the proteinase K family in the phylogenetic tree is 48%.

Thermitase/subtilisin group

Thermitase, the type enzyme for this group, is an extracellular, thermostable protease of the thermophilic microorganism *Thermoactinomyces vulgaris* (Frömmel et al., 1978). For an in-depth review of this protease see Betzel (2004). The other three enzymes of the thermitase-type WprA (S08.004) (Margot and Karamata, 1996), halolysin (S08.102) (Kamekura et al., 1996), and Subtilisin AK1 (S08.009) (Toogood et al., 2000) were also identified by Siezen and Leunissen (1997) and forming the subgroup of extremophilic thermitases. While halolysin derives from the halophilic archaeon *Haloferax mediterranei* (Kamekura et al., 1996), all other members come from Bacilli (see **Supplementary Figure 8** and **Figure 2**). **Figure 2** (light blue) shows a more distinctive group with two clades. Here, Siezen and Leunissen identified bpr (S08.022) of *Dichelobacter nodosus* as a thermitase and AprP of *Pseudomonas sp.* KFCC 10818 as a subtilisin (Lilley et al., 1992; Jang et al., 1996; Siezen and Leunissen, 1997). However, the bootstrap value for these two clades is only 18%, which is why they are treated as an intermediate subgroup between thermitases and subtilisins. This intermediate new subgroup is named here as dentilisins, since this protease was already described in 1990 (Que and Kuramitsu, 1990). In general, thermitases are co-located in the clade with subtilisins, highlighting their similarity.

The subtilisin Carlsberg (S08.001) is the type example of the subtilisins and the entire S8 family and belongs to the subgroup of true subtilisins, along with the high-alkaline subtilisins, the intracellular subtilisins, and the phylogenetically intermediate subtilisins (PIS) (Smith et al., 1966; Siezen and Leunissen, 1997; Saeki et al., 2003). Extracellular subtilisins play an important role in nutrition, whereas intracellular subtilisins (Isp), such as IspA (S08.030), play a role in protein turnover and processing during sporulation or are involved in the heat shock response (Reysset and Millet, 1972; Koide et al., 1986). As shown in **Figure 2** all holotypes are derived from microorganisms, mainly from Bacilli, while aerolysin (S08.105) (Völkl et al., 1994), Tk-subtilisin (*Thermococcus kodakaraensis* subtilisin) (S08.129) (Kannan et al., 2001), PopC, (S08.143) (Rolbetzki et al., 2008), and ALTP (*Alkaliphilus transvaalensis* protease) (S08.028) (Kobayashi et al., 2007) derive from Archaea, Myxococci, and Clostridia, respectively. However, *Bacillus* as the most prominent source of subtilisins spawned alkaline proteases such as subtilisin Carlsberg, BPN', and Savinase, which have their major application as detergent enzymes

with excellent properties, including high stability toward extreme temperatures, pH, organic solvents, detergents, and oxidising compounds (Kalisz, 1988; Contesini et al., 2018). Besides the application in detergents, subtilisins are applied for example in leather processing, food, wastewater treatment, and cosmetics (Kalisz, 1988; Solanki et al., 2021). These subtilisin holotype sequences will be used for the classification of the new sequences provided by a data mining approach.

Various diverse groups

Several holotypes form a clade together within the phylogenetic tree (**Figure 2** and **Supplementary Figure 9**). Due to their different origins, their different biological functions, and their low bootstrap value (40%), they most likely form individual groups. Bacillopeptidase F (bpF) (S08.017) from *B. subtilis* is a cell envelope protein and contributes to nutrition in the soil environment (bpF subtilases) (Hageman). While Siezen and Leunissen grouped bpF within the pyrolysins, it is separated in this study, which could be due to the fact that Siezen and Leunissen only analysed the amino acids around the catalytically active ones (Siezen and Leunissen, 1997). CDF (S08.149) from Thermoactinomyces sp. CDF is a protease located on the surface of the spore coat (CDF subtilases) (Cheng et al., 2009). Cytotoxin SubAB (S08.121) is a toxin from Escherichia coli with two subunits, where subunit B binds to the surface receptor of target cells and subunit A, the enzymatically active moiety, is responsible for cytotoxicity and has a very narrow substrate specificity (SubAB subtilases) (Yahiro et al.).

Mycosin-1 (S08.131) from *Mycobacterium tuberculosis* is an extracellular protein that is membrane- and cell wall-associated and is expressed after infection of macrophages, forming the group of mycosins (Dave et al., 2002). PatA (S08.156) and PatG (S08.146) from *Prochloron didemni* are involved in the maturation of cyanobactins in *Cyanobacteria* and form the group of transamidating subtilases (Lee et al., 2009; Agarwal et al., 2012).

A separated clade can be also observed for the new group of tripeptidyl peptidase subtilases (TPPS) and the group of autotransporter subtilases (AT) (Henderson and Nataro, 2001; **Supplementary Figure 10**). Like tripeptidyl peptidase II (TPPII, S08.090) from the pyrolysin group, tripeptidyl peptidase S (TPPS) (S08.091) from Streptomyces lividans is an exopeptidase that cleaves tripeptide units from oligopeptides or polypeptides and probably forms its own group (Butler, 2013). The extracellular *Serratia* serine protease (SSP) (S08.094) from *Serratia marcescens* (Yanagida et al., 1986) was grouped by Siezen and Leunissen (1997) to gram-negative pyrolysins. However, SSP together with EprS (S08.162) (Kida et al., 2013), AasP (autotransported serine protease A) (S08.1449) (Ali et al., 2008), NalP (Neisserial autotransporter lipoprotein) (S08.160) (Turner et al., 2002), and SphB1 (S08.068) (Coutte et al., 2001), are forming the group of autotransporter subtilases in gram-negative bacteria with a low sequence identity of 35%.

An additional clade with 94% bootstrap support is formed by perkinsin (S08.041) from the protist *Perkinsus marinus*, which is an enzyme of unknown function but may be involved in cell invasion of the eastern oyster *Crassostrea virginica* (Brown and Reece, 2003; **Supplementary Figure 11**). Sporangin (S08.145) from the alga *Chlamydomonas reinhardtii* is localised to the flagella of daughter cells within the sporangial cell wall and is released into the culture medium where it is involved in the digestion of the sporangial cell wall (Kubo et al., 2009). Perkinsin and sporangin are forming two new groups (perkinsins, sporangins). MCP-01 (S08.130), the extracellular cold-adapted protease from the deep-sea bacterium *Pseudoalteromonas* sp. SM9913 forms the group of deseasins secreted mainly by bacteria in deep-sea or lake sediments (Chen et al., 2007; Zhao et al., 2008). It is a multidomain protein with a collagen-binding domain at its C-terminus that exhibits collagenolytic activity and therefore plays an important role in the degradation of particulate organic nitrogen from deep-sea sediments (Zhao et al., 2008). The proteases produced by the parasites in the phylum Apicomplexa are forming a new group (Apicomplexa subtilases) with TgSub1 (Toxoplasma gondii) (S08.141) (Miller et al., 2001), PfSUB2 (Plasmodium falciparum) (S08.013) (Hackett et al., 1999), TgSUB2 (S08.154) (Miller et al., 2003), BdSUB1 (Babesia divergens) (S08.136) (Montero et al., 2006), PfSUB3 (S08.122) (Withers-Martinez et al., 2004), and PfSUB1 (S08.012) (Withers-Martinez et al., 2004), with a sequence identity of 34%. These proteases are involved in the host-cell invasion (Silmon de Monerri et al.).

Lantibiotic peptidase group

The lantibiotic peptidase group within S8A subfamily comprises highly specialised enzymes for cleavage of leader peptides from precursors of the antimicrobial peptides (lantibiotics) (Sahl et al., 1995; **Supplementary Figure 12**). ElkP (S08.095) and PepP (S08.85) from *Staphylococcus epidermidis* are not included in the dataset because only sequence fragments were available. Lantibiotic peptidases are found intracellularly, extracellularly and membrane-anchored (Bierbaum et al.). The sequences included in the phylogenetic tree show 31% sequence identity.

Kexin subfamily (S8B)

Kexin the type example for the subfamily S8B and was first identified in *Saccharomyces cerevisiae*. It can process the yeast precursors of alpha-mating factor and killer toxin and plays a significant role in post-translational modification in eukaryotes (Rogers et al., 1979). For a review see Fuller (2013).

AspA (S08.125) is one of the two subtilases which is lacking a propeptide and stands apart within the S8B family (**Figure 2**; Mellergaard, 1983; Kobayashi et al., 2015). Siezen and Leunissen (1997) also mentioned that AspA is a more distant member. As mentioned above, the S8B subfamily was used as an outgroup for the phylogenetic tree as it is closely related to the S8A subfamily (Siezen and Leunissen, 1997). Within the phylogenetic tree, it forms a clearly defined clade comprising all 21 holotype sequences of the MEROPS S8B subfamily, with an average sequence identity of 53% (**Supplementary Figure 13**). The clustering with a bootstrap support of 89% supports that kexins are a distinct subfamily (S8B) within the subtilase subset (Siezen and Leunissen, 1997). Kexins are also divided into at least four subgroups: PC1, PC2, furins and yeast kexins, but the subdivision will not be discussed further here as the focus is on the S8A subfamily (Siezen and Leunissen, 1997).

Data mining and phylogenetic tree analysis of subtilisins from *Bacillaceae*

Due to the increasing number of genome sequencing projects, the amount of data on uncharacterised proteins is growing exponentially (Rawlings, 2013). Many genomes encode multiple secreted proteases and many proteases can be found in different species (Takimura et al., 2007). The great potential of the data mining approach becomes clear when looking at the huge number of 247.897 hits (January 31, 2022) that were found when searching for S8 peptidases within the NCBI identical protein groups database. Within this second part of our study, we focused on subtilisins derived from Bacillaceae because they are right now the most relevant industrial proteases (Maurer, 2004; Azrin et al., 2022).

To search for new subtilisin sequences from *Bacillaceae*, the database search was performed as described above and in **Figure 1**. The search yielded 1,424 sequences with the set values. With the length specification of 350–410 amino acids, sequences typical of AprX, lantibiotic peptidases, kexins, OSP, and HMS are excluded, while typical thermitases, intracellular subtilisins, proteinase K, and high/true/PIS subtilisins can still be found. The size exclusion was set to reduce the number of sequences (18.881 without size exclusion) and was chosen because typical subtilisin sequences derived from *Bacillaceae* are around 380 amino acids long, including the signal peptide and the propeptide (Markland and Smith, 1971; Power et al., 1986; Siezen et al., 1991; Tjalsma et al., 2000). Without the size exclusion, many additional new subtilases from *Bacillaceae* could probably be found. CD-HIT clustering with an identity threshold of 85% yielded 375 clusters. For each cluster, one representative was used for further analysis. The number of sequences within one cluster is displayed as a bar chart around the phylogenetic tree in **Figure 3**. Signal peptide analysis identified 135 sequences without a signal peptide, reducing the dataset to 240 sequences, as we are only interested in extracellular proteases for further analysis and potential biochemical characterisation. The remaining sequences were aligned and the propeptide was manually removed as described above. Sequences that could be directly visually classified as thermitases after alignment were discarded, leaving a sequence set of 120 sequences within the sequence space of subtilisins as shown in **Figure 3** (**Supplementary Table 1**). Sequences from the first phylogenetic tree comprising all 168 MEROPS holotypes, which build the subfamilies of subtilisins, were used again and aligned with the 120 sequences from the data mining approach. The sequence alignment was refined with TrimAl, which reduced the alignment length to 260 positions in contrast to 448 positions without refinement. Here, the two archaea subtilisins were used as an outgroup to reroot the tree. **Figure 3** shows that all sequences derived from *Bacillaceae* in the data mining set represent the three main subgroups within the subtilisins, the true subtilisins, the high-alkaline subtilisins, and the phylogenetically intermediate subtilisins.

In addition to the above mentioned three subgroups, the following proteases form additional subgroups (**Supplementary Figure 14**): Aerolysin (S08.105) and Tk-subtilisin (S08.129) from the hyperthermophilic archaea *Pyrobaculum aerophilum* (Völkl et al., 1994) and *Thermococcus kodakaraensis* (Kannan et al., 2001) are forming an own new subgroup (Archaea subtilisins) with a sequence identity of 47%. Subtilisin S41 (S08.140) a psychrophilic protease from antarctic *Bacillus* TA41 (Almog et al., 2009), WF146 (S08.016), a thermophilic protease from *Bacillus sp.* WF146 (Wu et al., 2004), and Sfericase (S08.113), a psychrophilic protease from *Lysinibacillus sphaericus*, are forming a new subgroup named here extremophilic subtilisins with a sequence identity of 70%. The fact that no similar sequences from the group of extremophilic subtilisins were found in the data mining search could be due to the fact that all three representatives are larger than 410 amino acids, which also applies to EPR (S08.126), an extracellular protease from *B. subtilis* involved in cell motility (Dixit et al., 2002). PopC (S08.143) is involved in the cell signalling cascade for forming *Myxococcus xanthus* cells into fruiting bodies and sporulation (Rolbetzki et al., 2008). EPR and PopC are forming two new individual subgroups (EPR subtilisins, PopC subtilisins). The intracellular subtilisins form their own known subgroup with a sequence identity of 72% (Siezen and Leunissen, 1997). Since all sequences without signal peptides were excluded from the data mining set, no sequences are clustered with the holotypes. In general, subtilisins are mainly found in Bacilli and none in fungi (Siezen et al., 2007; Muszewska et al., 2011). In the following, all amino acid positions refer to the BPN' numbering.

The subgroup of **true subtilisins** includes subtilisin Carlsberg (S08.001) from *Bacillus licheniformis* (Linderstrøm-Lang and Ottesen, 1947; Güntelberg and Ottesen, 1952; Smith et al., 1966), which toghether with BPN' (S08.032) from *Bacillus*

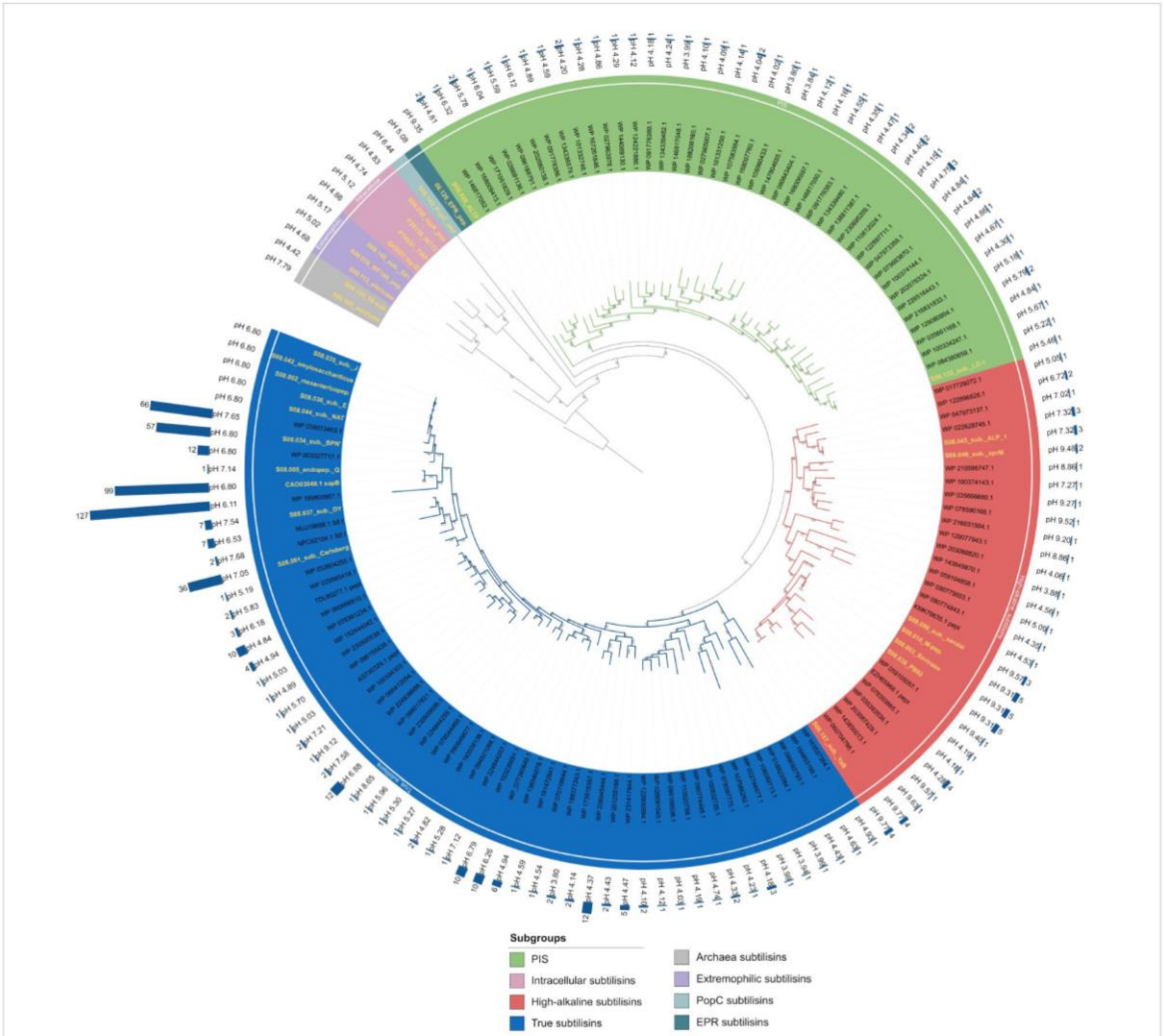


FIGURE 3
Phylogenetic tree of 120 novel subtilisins from *Bacillaceae* identified by a database search. The tree was constructed with IQ-TREE using the maximum likelihood method with ultrafast bootstrap support (model: LG + I + G4 predicted by Modelfinder, 1,000 replicates). The coloured area inside and outside the labels represents the subgroups, as indicated in the legend. The theoretical isoelectric point is given for each sequence in the outer circle. Aerolysin was chosen as the outgroup. Archaea subtilisins, extremophilic subtilisins, EPR – and PopC subgroup were added as additional holotypes. All holotypes are highlighted with a yellow text colour. The bar graphs represent the cluster size of the data mining search. For each group, the numbers above the branches indicate the bootstrap values based on 1000 replications. The tree can be accessed at the following link: https://itol.embl.de/shared/2H14VxXLj30E2.

amyloliquefaciens were the first two subtilisins to be studied in detail (Matsubara et al., 1965; Smith et al., 1966). Their group is supported by a 91% bootstrap value within the phylogenetic tree (**Figure 3** and **Supplementary Figure 17**). Interestingly the data mining search revealed the most similar sequences within the subgroup around these holotypes, as indicated by the bar chart with cluster sizes up to 127 sequences. Several newly found sequences are phylogenetically more distinct from any known holotype, which suggests that these sequences could have other biochemical characteristics and may form new classes. The calculated isoelectric point of the sequences within the true subtilisins is on average rather acidic to neutral. The representatives of this subgroup characterised so far are more sensitive and less active under high-alkaline conditions compared to the high-alkaline subtilisins (Nakamura et al., 1973; Maeda et al., 2001).

The subgroup of **phylogenetically intermediate subtilisins** (PIS) was introduced by Saeki et al. with the biochemical

characterisation of the subtilisin LD1 (S08.133) from the alkaliphilic *Bacillus* sp. KSM-LD1. Due to its properties and the phylogenetic position, LD1 forms a subgroup at an intermediate position between true subtilisins and high-alkaline subtilisins (Saeki et al., 2003). LD1 has a C-terminal extension of 29 amino acids, suggesting an association with the cell surface of *Bacillus* sp. KSM-LD1 (Saeki et al., 2003). Interestingly, the sequences WP_100334247.1 from *Bacillus alkalisoli* and WP_084380659.1 from *Sutcliffiella cohnii* have a C-terminal extension like LD1 (Spanka and Fritze, 1993; Liu et al., 2019). LD1 and other PIS have multiple amino acid insertions compared to BPN', but this does not affect substrate specificity toward synthetic substrates (Saeki et al., 2003). The protease ALTP (S08.028) from the anaerobic and extremely alkaliphilic *Alkaliphilus transvaalensis* is the first high-alkaline protease reported from a strict anaerobe (Kobayashi et al., 2007). ALTP is 66% identical to LD1 and, according to Kobayashi et al., it is in an intermediate position between the true and the highly alkaline subtilisins (Kobayashi et al., 2007). This assignment is supported by the phylogenetic tree (**Figure 2**) constructed in our study. In the phylogenetic tree with the newly mined database sequences (**Figure 3** and **Supplementary Figure 15**), ALTP is separated from the *Bacillaceae*-derived phylogenetically intermediate subtilisins, as it is derived from the bacterial class *Clostridia*. ALTP has solely an alkaline isoelectric point, while the other sequences within this subgroup all have an acidic pI (**Figure 3**).

The subgroup of **high-alkaline subtilisins** was discovered in the 1980s and originates from alkaliphilic *Bacilli* (Ito et al., 1998; Maurer, 2004). Since the first discovery of protease no. 221, an increasing number of high-alkaline subtilisins have been characterised (Nakamura et al., 1973). Alkaline subtilisins, such as Savinase, are much more stable in an alkaline environment than true subtilisins such as BPN' or subtilisin Carlsberg and can be used to adapt to harsh industrial conditions, especially in modern detergents (Maurer, 2004). Within the phylogenetic tree they are forming a distinct subgroup with a branch support of 93% (**Figure 3** and **Supplementary Figure 16**). ALP-1 (S08.045) from *Bacillus* sp. NKS-21 (Yamagata et al., 1995), WP_017729072.1 from *Halalkalibacterium ligniniphilum* (Zhu et al., 2014; Joshi et al., 2021), WP_122896828.1 from *Alteribacter keqinensis* (Liu et al., 2022), WP_047973137.1 from *Bacillus* sp. LL01 (Vilo et al., 2015), and WP_022628745.1 from *Alkalihalophilus marmarensis* (Denizci et al., 2010; Joshi et al., 2021) form a more separated clade, as they lack the four amino acid deletion around position 160, in contrast to the other high-alkaline proteases. This position corresponds to a loop near the P1 binding site (Wells et al., 1987; Betzel et al., 1992). They form another class of the ALP-1-type subtilisins, as mentioned by Yamagata et al. (2002). Additionally, the theoretical isoelectric point of these proteins is neutral to acidic in contrast to the majority of the other high-alkaline proteases. High-alkaline proteases adapt to higher alkaline conditions by an altered surface charge at higher pHs, as indicated by an increased pI value caused by a higher number of Arg and a decreased number of Lys residues (Masui et al., 1998). The substrate specificity of ALP-1 toward the B-chain of insulin differs from that of other alkaline subtilisins, but is similar to that of neutrophilic subtilisins, which may be related to the deletion of four amino acids around position 160 (Tsuchida et al., 1986; Yamagata et al., 1995). For ALP-1, an enzyme engineering study identified amino acids in the C-terminal region that increased stability 120-fold under alkaline conditions after replacement (Yamagata et al., 2002). Some of the high-alkaline proteases including Savinase have an extra proline at position 131, which provides extra active-site rigidity compared with other subtilisins (Betzel et al., 1996). Recently, we reportet about SPAO from *Alkalihalobacillus okhensis* Kh10-101^T, which showed high stability against hydrogen peroxide and NaCl concentrations up to 5.0 M (Falkenberg et al., 2022). SPAO can be assigned here to the holotype subtilisin sendai (S08.098) (**Figure 3**).

The average sequence identity between the sequences within the three subgroups of high-alkaline, PIS, and true subtilisins was calculated to be 67, 72, and 66%, respectively (sequences from **Figure 3**). The identity between true and high-alkaline subtilisins is 58%, between true and PIS 57%, and between high-alkaline subtilisins and PIS 55%.

A detailed investigation of all insertions and deletions within the three subgroups PIS, high-alkaline, and true subtilisins showed that the four amino acid deletion in the clade of aprM (S08.046) (Takami et al., 1990; Masui et al., 1994) is between Ser^{161} and Thr^{174}, while for the other high-alkaline proteases, except for the ALP-1 clade mentioned above, the deletion is between Gly^{160} and Thr^{174} (**Supplementary Figure 18**). Interestingly, all high-alkaline subtilisins have a deletion of one amino acid at positions 37 and 57 (**Supplementary Figure 18**). The loop of amino acids 50–59 is known to be one of the most variable parts of subtilisin structures (van der Laan et al., 1992). Therefore, deletions within this loop could be detected in several sequences within the three subfamilies. Several PIS sequences have a double insertion between positions 42 and 43 in common. All sequences within the PIS subgroup share the insertion between positions 159 and 160, while high-alkaline subtilisins have a deletion of four amino acids around this position. Position 160 is localised in a loop that, as mentioned above, takes part in the conformation of the P1 pocket and might be involved in the P1 preference, and the recognition of steric conformation (Yamagata et al., 1995). Additionally, shorter loops can increase the stability of an enzyme (Gavrilov et al., 2015). In general, all insertions or deletions are located at the surface of the protease, which could be due to the fact that the overall structure within the subtilisins is highly conserved (Goddette et al., 1992).

For a distinct further subdivision of true subtilisins, high-alkaline subtilisins and PIS into classes, supporting biochemical data might be necessary. However, based on the phylogenetic

tree, the deletion and insertion analysis, and the isoelectric point, a further subdivision into classes is most likely.

Conclusion

Phylogenetic studies of the S8 family within the MEROPS holotype dataset revealed a large number of different subtilases forming new groups and subgroups. In addition to the known groups of proteinase K, pyrolysins, kexins, subtilisins, thermitases and lantibiotic peptidases, the analysis revealed new groups or subgroups within the S8A subfamily depending on their position in the phylogenetic tree, their biochemical properties or their origin. This analysis was used in the second part of this study to categorise 120 newly identified predicted S8 protease sequences derived from *Bacillaceae*. They were found to represent the three main subgroups within the subtilisins, the true subtilisins, the high-alkaline subtilisins, and the phylogenetically intermediate subtilisins. However, without the specified filter parameters for data mining, more new subtilases outside the group of subtilisins from *Bacillaceae* could probably be found. In the absence of experimental characterisation for most of the found subtilisin sequences, a subdivision needs further experimental studies, because with bioinformatic analysis alone, a prediction of their biological and biochemical properties is possible only to a limited extent. For the newly found enzymes it is thus possible that they possess unique specificities and are of high interest for biotechnological applications.

Data availability statement

The original contributions presented in this study are included in the article/**Supplementary material**, further inquiries can be directed to the corresponding author.

Author contributions

FF collected and analyzed the data and wrote the original draft. JB, MB, and PS supervised the study and revised the manuscript. All authors contributed to the final manuscript.

Funding

This work was supported by the FH Aachen University of Applied Sciences within the framework of the program for the promotion of young scientists.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.1017978/full#supplementary-material

References

Agarwal, V., Pierce, E., McIntosh, J., Schmidt, E. W., and Nair, S. K. (2012). Structures of cyanobactin maturation enzymes define a family of transamidating proteases. *Chem. Biol.* 19, 1411–1422. doi: 10.1016/j.chembiol.2012.09.012

Ageitos, J. M., Baker, P. J., Sugahara, M., and Numata, K. (2013). Proteinase K-catalyzed synthesis of linear and star oligo(L-phenylalanine) conjugates. *Biomacromolecules* 14, 3635–3642. doi: 10.1021/bm4009974

Ali, T., Oldfield, N. J., Wooldridge, K. G., Turner, D. P., and Ala'Aldeen, D. A. A. (2008). Functional characterization of AasP, a maturation protease autotransporter protein of *Actinobacillus pleuropneumoniae*. *Infect. Immun.* 76, 5608–5614. doi: 10.1128/IAI.00085-08

Almog, O., González, A., Godin, N., Leeuw, M., de Mekel, M. J., Klein, D., et al. (2009). The crystal structures of the psychrophilic subtilisin S41 and the mesophilic subtilisin Sph reveal the same calcium-loaded state. *Proteins* 74, 489–496. doi: 10.1002/prot.22175

Anjard, C., and Loomis, W. F. (2005). Peptide signaling during terminal differentiation of Dictyostelium. *Proc. Natl. Acad. Sci. U.S.A.* 102, 7607–7611. doi: 10.1073/pnas.0501820102

Azrin, N. A. M., Ali, M. S. M., Rahman, R. N. Z. R. A., Oslan, S. N., and Noor, N. D. M. (2022). Versatility of subtilisin: A review on structure, characteristics, and applications. *Biotechnol. Appl. Biochem.* [Epub ahead of print]. doi: 10.1002/bab.2309

Bethe, G., Nau, R., Wellmer, A., Hakenbeck, R., Reinert, R. R., Heinz, H. P., et al. (2001). The cell wall-associated serine protease PrtA: A highly conserved virulence factor of *Streptococcus pneumoniae*. *FEMS Microbiol. Lett.* 205, 99–104. doi: 10.1111/j.1574-6968.2001.tb10931.x

Betzel, C. (2004). "Thermitase," in *Handbook of proteolytic enzymes*, eds A. J. Barrett, J. F. Woessner, and N. D. Rawlings (London: Elsevier Academic Press), 3167–3169. doi: 10.1016/B978-0-12-382219-2.00696-7

Betzel, C., Klupsch, S., Branner, S., and Wilson, K. S. (1996). Crystal structures of the alkaline proteases savinase and esperase from *Bacillus lentus*. *Adv. Exp. Med. Biol.* 379, 49–61. doi: 10.1007/978-1-4613-0319-0_7

Betzel, C., Klupsch, S., Papendorf, G., Hastrup, S., Branner, S., and Wilson, K. S. (1992). Crystal structure of the alkaline proteinase Savinase from *Bacillus lentus* at 1.4 A resolution. *J. Mol. Biol.* 223, 427–445. doi: 10.1016/0022-2836(92)90662-4

Betzel, C., Teplyakov, A. V., Harutyunyan, E. H., Saenger, W., and Wilson, K. S. (1990). Thermitase and proteinase K: A comparison of the refined three-dimensional structures of the native enzymes. *Protein Eng.* 3, 161–172. doi: 10.1093/protein/3.3.161

Bierbaum, G., Jack, R. W., and Sahl, H.-G. (2013). "Lantibiotic leader peptidases," in *Handbook of proteolytic enzymes*, eds A. J. Barrett, J. F. Woessner, and N. D. Rawlings (London: Elsevier Academic Press), 3220–3222. doi: 10.1016/B978-0-12-382219-2.00709-2

Blumentals, I. I., Robinson, A. S., and Kelly, R. M. (1990). Characterization of sodium dodecyl sulfate-resistant proteolytic activity in the hyperthermophilic archaebacterium *Pyrococcus furiosus*. *Appl. Environ. Microbiol.* 56, 1992–1998. doi: 10.1128/aem.56.7.1992-1998.1990

Borhan, B., Hammock, B., Seifert, J., and Wilson, B. W. (1996). Methyl and phenyl esters and thioesters of carboxylic acids as surrogate substrates for microassay of proteinase K esterase activity. *Anal. Bioanal. Chem.* 354, 490–492. doi: 10.1007/s0021663540490

Broadbent, J. R., and Steele, J. L. (2013). "Lactocepin: The cell envelopeassociated endopeptidase of lactococci," in *Handbook of proteolytic enzymes*, eds A. J. Barrett, J. F. Woessner, and N. D. Rawlings (London: Elsevier Academic Press.), 3188–3195. doi: 10.1016/B978-0-12-382219-2.00703-1

Brown, G. D., and Reece, K. S. (2003). Isolation and characterization of serine protease gene(s) from *Perkinsus marinus*. *Dis. Aquatic Organisms* 57, 117–126. doi: 10.3354/dao057117

Butler, M. J. (2013). "Tripeptidyl-peptidase S," in *Handbook Of proteolytic enzymes*, eds A. J. Barrett, J. F. Woessner, and N. D. Rawlings (London: Elsevier Academic Press), 3222–3223. doi: 10.1016/B978-0-12-382219-2.00710-9

Chen, X.-L., Xie, B.-B., Lu, J.-T., He, H.-L., and Zhang, Y. (2007). A novel type of subtilase from the psychrotolerant bacterium *Pseudoalteromonas* sp. SM9913: Catalytic and structural properties of deseasin MCP-01. *Microbiology* 153, 2116–2125. doi: 10.1099/mic.0.2007/006056-0

Cheng, G., Zhao, P., Tang, X.-F., and Tang, B. (2009). Identification and characterization of a novel spore-associated subtilase from *Thermoactinomyces* sp. CDF. *Microbiology* 155, 3661–3672. doi: 10.1099/mic.0.031336-0

Christiansen, T., Michaelsen, S., Wümpelmann, M., and Nielsen, J. (2003). Production of savinase and population viability of *Bacillus clausii* during high-cell-density fed-batch cultivations. *Biotechnol. Bioeng.* 83, 344–352. doi: 10.1002/bit.10675

Contesini, F. J., Melo, RRd, and Sato, H. H. (2018). An overview of *Bacillus* proteases: From production to application. *Crit. Rev. Biotechnol.* 38, 321–334. doi: 10.1080/07388551.2017.1354354

Coutte, L., Antoine, R., Drobecq, H., Locht, C., and Jacob-Dubuisson, F. (2001). Subtilisin-like autotransporter serves as maturation protease in a bacterial secretion pathway. *EMBO J.* 20, 5040–5048. doi: 10.1093/emboj/20.18.5040

Dave, J. A., van Gey Pittius, N. C., Beyers, A. D., Ehlers, M. R. W., and Brown, G. D. (2002). Mycosin-1, a subtilisin-like serine protease of *Mycobacterium tuberculosis*, is cell wall-associated and expressed during infection of macrophages. *BMC Microbiol.* 2:30. doi: 10.1186/1471-2180-2-30

Denizci, A. A., Kazan, D., and Erarslan, A. (2010). *Bacillus marmarensis* sp. nov., an alkaliphilic, protease-producing bacterium isolated from mushroom compost. *Int. J. Syst. Evo. Microbiol.* 60, 1590–1594. doi: 10.1099/ijs.0.012369-0

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., et al. (2008). Phylogeny.fr: Robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* 36:W465–W469. doi: 10.1093/nar/gkn180

Dessimoz, C., and Gil, M. (2010). Phylogenetic assessment of alignments reveals neglected tree signal in gaps. *Genome Biol.* 11:R37. doi: 10.1186/gb-2010-11-4-r37

Dixit, M., Murudkar, C. S., and Rao, K. K. (2002). Epr is transcribed from a final sigma(D) promoter and is involved in swarming of *Bacillus subtilis*. *J. Bacteriol.* 184, 596–599. doi: 10.1128/JB.184.2.596-599.2002

Ebeling, W., Hennrich, N., Klockow, M., Metz, H., Orth, H. D., and Lang, H. (1974). Proteinase K from *Tritirachium* album Limber. *Eur. J. Biochem.* 47, 91–97. doi: 10.1111/j.1432-1033.1974.tb03671.x

Eder, J., Rheinnecker, M., and Fersht, A. R. (1993). Folding of subtilisin BPN': Role of the pro-sequence. *J. Mol. Biol.* 233, 293–304. doi: 10.1006/jmbi.1993.1507

Falkenberg, F., Rahba, J., Fischer, D., Bott, M., Bongaerts, J., and Siegert, P. (2022). Biochemical characterization of a novel oxidatively stable, halotolerant and high-alkaline subtilisin from Alkalihalobacillus okhensis Kh10-101^{T}. *FEBS Open Bio.* [Epub ahead of print]. doi: 10.1002/2211-5463.13457

Faraco, V., Palmieri, G., Festa, G., Monti, M., Sannia, G., and Giardina, P. (2005). A new subfamily of fungal subtilases: Structural and functional analysis of a *Pleurotus ostreatus* member. *Microbiology* 151, 457–466. doi: 10.1099/mic.0.27441-0

Frömmel, C., Hausdorf, G., Höhne, W. E., Behnke, U., and Ruttloff, H. (1978). Characterization of a protease from *Thermoactinomyces vulgaris* (thermitase). *Acta Biol. Medica Germanica* 37, 1193–1204.

Fuller, R. S. (2013). "Kexin," in *Handbook of proteolytic enzymes*, eds A. J. Barrett, J. F. Woessner, and N. D. Rawlings (London: Elsevier Academic Press), 3270–3277. doi: 10.1016/B978-0-12-382219-2.00722-5

Garcia-Sanchez, A., Cerrato, R., Larrasa, J., Ambrose, N. C., Parra, A., Alonso, J. M., et al. (2004). Characterisation of an extracellular serine protease gene (nasp gene) from *Dermatophilus congolensis*. *FEMS Microbiol. Lett.* 231, 53–57. doi: 10.1016/S0378-1097(03)00958-3

Gavrilov, Y., Dagan, S., and Levy, Y. (2015). Shortening a loop can increase protein native state entropy. *Proteins* 83, 2137–2146. doi: 10.1002/prot.24926

Genay, M., Sadat, L., Gagnaire, V., and Lortal, S. (2009). prtH2, not prtH, is the ubiquitous cell wall proteinase gene in *Lactobacillus helveticus*. *Appl. Environ. Microbiol.* 75, 3238–3249. doi: 10.1128/AEM.02395-08

Gilbert, C., Atlan, D., Blanc, B., Portailer, R., Germond, J. E., Lapierre, L., et al. (1996). A new cell surface proteinase: Sequencing and analysis of the prtB gene from *Lactobacillus delbruekii* subsp. bulgaricus. *J. Bacteriol.* 178, 3059–3065. doi: 10.1128/jb.178.11.3059-3065.1996

Goddette, D. W., Paech, C., Yang, S. S., Mielenz, J. R., Bystroff, C., Wilke, M. E., et al. (1992). The crystal structure of the *Bacillus* lentus alkaline protease, subtilisin BL, at 1.4 A resolution. *J. Mol. Biol.* 228, 580–595. doi: 10.1016/0022-2836(92)90843-9

Goldenberger, D., Perschil, I., Ritzler, M., and Altwegg, M. (1995). A simple "universal" DNA extraction procedure using SDS and proteinase K is compatible with direct PCR amplification. *PCR Methods Appl.* 4, 368–370. doi: 10.1101/gr.4.6.368

Good, J. R., Cabral, M., Sharma, S., Yang, J., van Driessche, N., Shaw, C. A., et al. (2003). TagA, a putative serine protease/ABC transporter of *Dictyostelium* that is required for cell fate determination at the onset of development. *Development* 130, 2953–2965. doi: 10.1242/dev.00523

Gunkel, F. A., and Gassen, H. G. (1989). Proteinase K from *Tritirachium* album Limber. Characterization of the chromosomal gene and expression of the cDNA in *Escherichia coli*. *Eur. J. Biochem.* 179, 185–194. doi: 10.1111/j.1432-1033.1989.tb14539.x

Güntelberg, A. V., and Ottesen, M. (1952). Preparation of crystals containing the plakalbumin-forming enzyme from *Bacillus subtilis*. *Nature* 170:802. doi: 10.1038/170802a0

Hackett, F., Sajid, M., Withers-Martinez, C., Grainger, M., and Blackman, M. J. (1999). PfSUB-2: A second subtilisin-like protein in *Plasmodium falciparum* merozoites. *Mol. Biochem. Parasitol.* 103, 183–195. doi: 10.1016/S0166-6851(99)00122-X

Hageman, J. H. (2013). "Bacillopeptidase F," in *Handbook Of proteolytic enzymes*, eds A. J. Barrett, J. F. Woessner, and N. D. Rawlings (London: Elsevier Academic Press), 3170–3171. doi: 10.1016/B978-0-12-382219-2.00697-9

Henderson, I. R., and Nataro, J. P. (2001). Virulence functions of autotransporter proteins. *Infect. Immun.* 69, 1231–1243. doi: 10.1128/IAI.69.3.1231-1243.2001

Hoang, D. T., Chernomor, O., Haeseler, A., von Minh, B. Q., and Le Vinh, S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522. doi: 10.1093/molbev/msx281

Hu, G., and Leger, R. J. S. (2004). A phylogenomic approach to reconstructing the diversification of serine proteases in fungi. *J. Evol. Biol.* 17, 1204–1214. doi: 10.1111/j.1420-9101.2004.00786.x

Huang, Y., Niu, B., Gao, Y., Fu, L., and Li, W. (2010). CD-HIT Suite: A web server for clustering and comparing biological sequences. *Bioinformatics* 26, 680–682. doi: 10.1093/bioinformatics/btq003

Ito, S., Kobayashi, T., Ara, K., Ozaki, K., Kawai, S., and Hatada, Y. (1998). Alkaline detergent enzymes from alkaliphiles: Enzymatic properties, genetics, and structures. *Extremophiles* 2, 185–190. doi: 10.1007/s007920050059

Itoi, Y., Horinaka, M., Tsujimoto, Y., Matsui, H., and Watanabe, K. (2006). Characteristic features in the structure and collagen-binding ability of a thermophilic collagenolytic protease from the thermophile *Geobacillus* collagenovorans MO-1. *J. Bacteriol.* 188, 6572–6579. doi: 10.1128/JB.00767-06

Jang, H. J., Kim, B. C., Pyun, Y. R., and Kim, Y. S. (2002). A novel subtilisin-like serine protease from *Thermoanaerobacter yonseiensis* KB-1: Its cloning,

expression, and biochemical properties. *Extremophiles* 6, 233–243. doi: 10.1007/s00792-001-0248-1

Jang, W. H., Kim, E. K., Lee, H. B., Chung, J. H., and Yoo, O. J. (1996). Characterization of an alkaline serine protease from an alkaline-resistant *Pseudomonas* sp.: Cloning and expression of the protease gene in *Escherichia coli*. *Biotechnol. Lett.* 18, 57–62. doi: 10.1007/BF00137811

Joshi, A., Thite, S., Karodi, P., Joseph, N., and Lodha, T. (2021). *Alkalihalobacterium elongatum* gen. nov. sp. nov.: An antibiotic-producing bacterium isolated from lonar lake and reclassification of the genus *Alkalihalobacillus* into seven novel genera. *Front. Microbiol.* 12:722369. doi: 10.3389/fmicb.2021.722369

Joshi, S., and Satyanarayana, T. (2013). Characteristics and applications of a recombinant alkaline serine protease from a novel bacterium *Bacillus lehensis*. *Bioresou. Technol.* 131, 76–85. doi: 10.1016/j.biortech.2012.12.124

Kagawa, T. F., and Cooney, J. C. (2013). "C5a Peptidase," in *Handbook Of proteolytic enzymes*, eds A. J. Barrett, J. F. Woessner, and N. D. Rawlings (London: Elsevier Academic Press), 3202–3208. doi: 10.1016/B978-0-12-382219-2.00705-5

Kalisz, H. M. (1988). Microbial proteinases. *Adv. Biochem. Eng. Biotechnol.* 36, 1–65. doi: 10.1007/BFb0047944

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285

Kamekura, M., Seno, Y., and Dyall-Smith, M. (1996). Halolysin R4, a serine proteinase from the halophilic archaeon *Haloferax mediterranei*; gene cloning, expression and structural studies. *Biochim. Biophysica acta* 1294, 159–167. doi: 10.1016/0167-4838(96)00016-7

Kaneda, M., and Tominaga, N. (1975). Isolation and characterization of a proteinase from the sarcocarp of melon fruit. *J. Biochem.* 78, 1287–1296. doi: 10.1093/oxfordjournals.jbchem.a131026

Kannan, Y., Koga, Y., Inoue, Y., Haruki, M., Takagi, M., Imanaka, T., et al. (2001). Active subtilisin-like protease from a hyperthermophilic archaeon in a form with a putative prosequence. *Appl. Environ. Microbiol.* 67, 2445–2452. doi: 10.1128/AEM.67.6.2445-2452.2001

Karlsson, C., Andersson, M.-L., Collin, M., Schmidtchen, A., Björck, L., and Frick, I.-M. (2007). SufA–a novel subtilisin-like serine proteinase of *Finegoldia magna*. *Microbiology* 153, 4208–4218. doi: 10.1099/mic.0.2007/010322-0

Katoh, K., Rozewicki, J., and Yamada, K. D. (2019). MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinf.* 20, 1160–1166. doi: 10.1093/bib/bbx108

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kida, Y., Taira, J., Yamamoto, T., Higashimoto, Y., and Kuwano, K. (2013). EprS, an autotransporter protein of *Pseudomonas aeruginosa*, possessing serine protease activity induces inflammatory responses through protease-activated receptors. *Cell. Microbiol.* 15, 1168–1181. doi: 10.1111/cmi.12106

Klingeberg, M., Galunsky, B., Sjoholm, C., Kasche, V., and Antranikian, G. (1995). Purification and properties of a highly thermostable, sodium dodecyl sulfate-resistant and stereospecific proteinase from the extremely *Thermophilic Archaeon Thermococcus stetteri*. *Appl. Environ. Microbiol.* 61, 3098–3104. doi: 10.1128/aem.61.8.3098-3104.1995

Kobayashi, H., Yoshida, T., Miyakawa, T., Tashiro, M., Okamoto, K., Yamanaka, H., et al. (2015). Structural basis for action of the external chaperone for a propeptide-deficient serine protease from *Aeromonas sobria*. *J. Biol. Chem.* 290, 11130–11143. doi: 10.1074/jbc.M114.622852

Kobayashi, T., Lu, J., Li, Z., Hung, V. S., Kurata, A., Hatada, Y., et al. (2007). Extremely high alkaline protease from a deep-subsurface bacterium, *Alkaliphilus transvaalensis*. *Appl. Microbiol. Biotechnol.* 75, 71–80. doi: 10.1007/s00253-006-0800-0

Koide, Y., Nakamura, A., Uozumi, T., and Beppu, T. (1986). Cloning and sequencing of the major intracellular serine protease gene of *Bacillus subtilis*. *J. Bacteriol.* 167, 110–116. doi: 10.1128/jb.167.1.110-116.1986

Koma, D., Yamanaka, H., Moriyoshi, K., Ohmoto, T., and Sakai, K. (2007). Overexpression and characterization of thermostable serine protease in *Escherichia coli* encoded by the ORF TTE0824 from *Thermoanaerobacter tengcongensis*. *Extremophiles* 11, 769–779. doi: 10.1007/s00792-007-0103-0

Kong, H.-H., Kim, T.-H., and Chung, D.-I. (2000). Purification and characterization of a secretory serine proteinase of *Acanthamoeba healyi* isolated from Gae. *J. Parasitol.* 86, 12–17. doi: 10.1645/0022-3395(2000)086[0012:PACOAS]2.0.CO;2

Kubo, T., Kaida, S., Abe, J., Saito, T., Fukuzawa, H., and Matsuda, Y. (2009). The Chlamydomonas hatching enzyme, sporangin, is expressed in specific phases of the cell cycle and is localized to the flagella of daughter cells within the sporangial cell wall. *Plant Cell Physiol.* 50, 572–583. doi: 10.1093/pcp/pcp016

Kuhner, M. K., and Felsenstein, J. (1994). A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Mol. Biol. Evol.* 11, 459–468.

Lawrenson, R. A., and Sriskandan, S. (2013). "Cell envelope proteinase A (*Streptococcus*)," in *Handbook Of proteolytic enzymes*, eds A. J. Barrett, J. F. Woessner, and N. D. Rawlings (London: Elsevier Academic Press), 3195–3202. doi: 10.1016/B978-0-12-382219-2.00704-3

Lee, J., McIntosh, J., Hathaway, B. J., and Schmidt, E. W. (2009). Using marine natural products to discover a protease that catalyzes peptide macrocyclization of diverse substrates. *J. Am. Chem. Soc.* 131, 2122–2124. doi: 10.1021/ja8092168

Leger, R., Charnley, A. K., and Cooper, R. M. (1987). Characterization of cuticle-degrading proteases produced by the entomopathogen *Metarhizium anisopliae*. *Arch. Biochem. Biophys.* 253, 221–232. doi: 10.1016/0003-9861(87)90655-2

Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49:W293–W296. doi: 10.1093/nar/gkab301

Li, J., Gu, F., Wu, R., Yang, J., and Zhang, K.-Q. (2017). Phylogenomic evolutionary surveys of subtilase superfamily genes in fungi. *Sci. Rep.* 7:45456. doi: 10.1038/srep45456

Li, J., Yu, L., Yang, J., Dong, L., Tian, B., Yu, Z., et al. (2010). New insights into the evolution of subtilisin-like serine protease genes in Pezizomycotina. *BMC Evol. Biol.* 10:68. doi: 10.1186/1471-2148-10-68

Lilley, G. G., Stewart, D. J., and Kortt, A. A. (1992). Amino acid and DNA sequences of an extracellular basic protease of *Dichelobacter nodosus* show that it is a member of the subtilisin family of proteases. *Eur. J. Biochem.* 210, 13–21. doi: 10.1111/j.1432-1033.1992.tb17385.x

Linderstrøm-Lang, K., and Ottesen, M. (1947). A new protein from ovalbumin. *Nature* 159:807. doi: 10.1038/159807a0

Liu, G.-H., Narsing Rao, M. P., Dong, Z.-Y., Wang, J.-P., Chen, Z., Liu, B., et al. (2019). Two novel alkaliphiles, *Bacillus alkalisoli* sp. nov., and Bacillus solitudinis sp. nov., isolated from saline-alkali soil. *Extremophiles* 23, 759–764. doi: 10.1007/s00792-019-01127-2

Liu, J., Zhang, X., Cao, H., Guo, L., Zhao, B., Zhang, X., et al. (2022). *Alteribacter keqinensis* sp. nov., a moderately halophilic bacterium isolated from a soda lake. *Int. J. Syst. Evol. Microbiol.* 72. doi: 10.1099/ijsem.0.005351

Maeda, H., Mizutani, O., Yamagata, Y., Ichishima, E., and Nakajima, T. (2001). Alkaline-resistance model of subtilisin ALP I, a novel alkaline subtilisin. *J. Biochem.* 129, 675–682. doi: 10.1093/oxfordjournals.jbchem.a002906

Margot, P., and Karamata, D. (1996). The wprA gene of *Bacillus subtilis* 168, expressed during exponential growth, encodes a cell-wall-associated protease. *Microbiology* 142, 3437–3444. doi: 10.1099/13500872-142-12-3437

Markland, F. S., and Smith, E. L. (1971). "16 Subtilisins: Primary structure, chemical and physical properties," in *The Enzymes*, ed. P. D. Boyer (New York, NY: Academic Press), 3, 561–608. doi: 10.1016/S1874-6047(08)60407-2

Masayama, A., Hamasaki, K., Urakami, K., Shimamoto, S., Kato, S., Makino, S., et al. (2006). Expression of germination-related enzymes, CspA, CspB, CspC, SleC, and SleM, of Clostridium perfringens S40 in the mother cell compartment of sporulating cells. *Genes Genetic Syst.* 81, 227–234. doi: 10.1266/ggs.81.227

Masui, A., Fujiwara, N., and Imanaka, T. (1994). Stabilization and rational design of serine protease AprM under highly alkaline and high-temperature conditions. *Appl. Environ. Microbiol.* 60, 3579–3584. doi: 10.1128/aem.60.10.3579-3584.1994

Masui, A., Fujiwara, N., Yamamoto, K., Takagi, M., and Imanaka, T. (1998). Rational design for stabilization and optimum pH shift of serine protease AprN. *J. Ferment. Bioeng.* 85, 30–36. doi: 10.1016/S0922-338X(97)80349-2

Matsubara, H., Kasper, C. B., Brown, D. M., and Smith, E. L. (1965). Subtilisin BPN'. *J. Biol. Chem.* 240, 1125–1130. doi: 10.1016/S0021-9258(18)97548-4

Maurer, K.-H. (2004). Detergent proteases. *Curr. Opin. Biotechnol.* 15, 330–334. doi: 10.1016/j.copbio.2004.06.005

Mayr, J., Lupas, A., Kellermann, J., Eckerskorn, C., Baumeister, W., and Peters, J. (1996). A hyperthermostable protease of the subtilisin family bound to the surface layer of the archaeon *Staphylothermus marinus*. *Curr. Biol.* 6, 739–749. doi: 10.1016/S0960-9822(09)00455-2

Mellergaard, S. (1983). Purification and characterization of a new proteolytic enzyme produced by *Aeromonas salmonicida*. *J. Appl. Bacteriol.* 54, 289–294. doi: 10.1111/j.1365-2672.1983.tb02619.x

Miller, S. A., Binder, E. M., Blackman, M. J., Carruthers, V. B., and Kim, K. (2001). A conserved subtilisin-like protein TgSUB1 in microneme organelles

of *Toxoplasma gondii*. *J. Biol. Chem.* 276, 45341–45348. doi: 10.1074/jbc.M106665200

Miller, S. A., Thathy, V., Ajioka, J. W., Blackman, M. J., and Kim, K. (2003). TgSUB2 is a *Toxoplasma gondii* rhoptry organelle processing proteinase. *Mol. Microbiol.* 49, 883–894. doi: 10.1046/j.1365-2958.2003.03604.x

Montero, E., Gonzalez, L. M., Rodriguez, M., Oksov, Y., Blackman, M. J., and Lobo, C. A. (2006). A conserved subtilisin protease identified in *Babesia divergens* merozoites. *J. Biol. Chem.* 281, 35717–35726. doi: 10.1074/jbc.M604344200

Morita, Y., Hasan, Q., Sakaguchi, T., Murakami, Y., Yokoyama, K., and Tamiya, E. (1998). Properties of a cold-active protease from psychrotrophic *Flavobacterium balustinum* P104. *Appl. Microbiol. Biotechnol.* 50, 669–675. doi: 10.1007/s002530051349

Muszewska, A., Taylor, J. W., Szczesny, P., and Grynberg, M. (2011). Independent subtilases expansions in fungi associated with animals. *Mol. Biol. Evol.* 28, 3395–3404. doi: 10.1093/molbev/msr176

Nakamura, K., Matsushima, A., and Horikoshi, K. (1973). The state of amino acid residues in alkaline protease produced by *Bacillus* No. 221. *Agricul. Biol. Chem.* 37, 1261–1267. doi: 10.1080/00021369.1973.10860834

Okamoto, M., Yonejima, Y., Tsujimoto, Y., Suzuki, Y., and Watanabe, K. (2001). A thermostable collagenolytic protease with a very large molecular mass produced by thermophilic *Bacillus* sp. strain MO-1. *Appl. Microbiol. Biotechnol.* 57, 103–108. doi: 10.1007/s002530100731

Okuda, M., Sumitomo, N., Takimura, Y., Ogawa, A., Saeki, K., Kawai, S., et al. (2004). A new subtilisin family: Nucleotide and deduced amino acid sequences of new high-molecular-mass alkaline proteases from *Bacillus* spp. *Extremophiles* 8, 229–235. doi: 10.1007/s00792-004-0381-8

Page, M. J., and Di Cera, E. (2008). Serine peptidases: Classification, structure and function. *Cell. Mol. Life Sci.* 65, 1220–1236. doi: 10.1007/s00018-008-7565-9

Pastar, I., Tonic, I., Golic, N., Kojic, M., van Kranenburg, R., Kleerebezem, M., et al. (2003). Identification and genetic characterization of a novel proteinase, PrtR, from the human isolate *Lactobacillus rhamnosus* BGT10. *Appl. Environ. Microbiol.* 69, 5802–5811. doi: 10.1128/AEM.69.10.5802-5811.2003

Phrommao, E., Yongsawatdigul, J., Rodtong, S., and Yamabhai, M. (2011). A novel subtilase with NaCl-activated and oxidant-stable activity from *Virgibacillus* sp. SK37. *BMC Biotechnol.* 11:65. doi: 10.1186/1472-6750-11-65

Power, S. D., Adams, R. M., and Wells, J. A. (1986). Secretion and autoproteolytic maturation of subtilisin. *Proc. Natl. Acad. Sci. U.S.A.* 83, 3096–3100. doi: 10.1073/pnas.83.10.3096

Que, X. C., and Kuramitsu, H. K. (1990). Isolation and characterization of the *Treponema denticola* prtA gene coding for chymotrypsin like protease activity and detection of a closely linked gene encoding PZ-PLGPA-hydrolyzing activity. *Infect. Immun.* 58, 4099–4105. doi: 10.1128/iai.58.12.4099-4105.1990

Rawlings, N. D. (2013). Identification and prioritization of novel uncharacterized peptidases for biochemical characterization. *Database* 2013:bat022. doi: 10.1093/database/bat022

Rawlings, N. D. (2020). Twenty-five years of nomenclature and classification of proteolytic enzymes.. *Biochim. Biophys. Acta Proteins Proteom* 1868:140345. doi: 10.1016/j.bbapap.2019.140345

Rawlings, N. D., Barrett, A. J., Thomas, P. D., Huang, X., Bateman, A., and Finn, R. D. (2018). The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* 46:D624–D632. doi: 10.1093/nar/gkx1134

Rawlings, N. D., and Bateman, A. (2021). How to use the MEROPS database and website to help understand peptidase specificity. *Protein Sci.* 30, 83–92. doi: 10.1002/pro.3948

Rawlings, N. D., Waller, M., Barrett, A. J., and Bateman, A. (2014). MEROPS: The database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* 42:D503–D509. doi: 10.1093/nar/gkt953

Renn, S. C., Tomkinson, B., and Taghert, P. H. (1998). Characterization and cloning of tripeptidyl peptidase II from the fruit fly, *Drosophila melanogaster*. *J. Biol. Chem.* 273, 19173–19182. doi: 10.1074/jbc.273.30.19173

Reysset, G., and Millet, J. (1972). Characterization of an intracellular protease in *B. subtillus* during sporulation. *Biochem. Biophys. Res. Commun.* 49, 328–334. doi: 10.1016/0006-291X(72)90414-7

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* 42:W320–W324. doi: 10.1093/nar/gku316

Rogers, D. T., Saville, D., and Bussey, H. (1979). Saccharomyces cerevisiae killer expression mutant kex2 has altered secretory proteins and glycoproteins. *Biochem. Biophys. Res. Commun.* 90, 187–193. doi: 10.1016/0006-291X(79)91607-3

Rolbetzki, A., Ammon, M., Jakovljevic, V., Konovalova, A., and Søgaard-Andersen, L. (2008). Regulated secretion of a protease activates intercellular signaling during fruiting body formation in *M. xanthus*. *Dev. Cell* 15, 627–634. doi: 10.1016/j.devcel.2008.08.002

Saeki, K., Hitomi, J., Okuda, M., Hatada, Y., Kageyama, Y., Takaiwa, M., et al. (2002). A novel species of alkaliphilic *Bacillus* that produces an oxidatively stable alkaline serine protease. *Extremophiles* 6, 65–72. doi: 10.1007/s007920100224

Saeki, K., Magallones, M. V., Takimura, Y., Hatada, Y., Kobayashi, T., Kawai, S., et al. (2003). Nucleotide and deduced amino acid sequences of a new subtilisin from an alkaliphilic *Bacillus* isolate. *Curr. Microbiol.* 47, 337–340. doi: 10.1007/s00284-002-4018-9

Saeki, K., Okuda, M., Hatada, Y., Kobayashi, T., Ito, S., Takami, H., et al. (2000). Novel oxidatively stable subtilisin-like serine proteases from alkaliphilic *Bacillus* spp.: Enzymatic properties, sequences, and evolutionary relationships. *Biochem. Biophys. Res. Commun.* 279, 313–319. doi: 10.1006/bbrc.2000.3931

Sahl, H. G., Jack, R. W., and Bierbaum, G. (1995). Biosynthesis and biological activities of lantibiotics with unique post-translational modifications. *Eur. J. Biochem.* 230, 827–853. doi: 10.1111/j.1432-1033.1995.tb20627.x

Sakaguchi, M. (2013). "Aqualysin I," in *Handbook Of proteolytic enzymes*, eds A. J. Barrett, J. F. Woessner, and N. D. Rawlings (London: Elsevier Academic Press), 3174–3176. doi: 10.1016/B978-0-12-382219-2.00699-2

Sánchez, R., Serra, F., Tárraga, J., Medina, I., Carbonell, J., and Pulido, L. (2011). Phylemon 2.0: A suite of web-tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing. *Nucleic Acids Res.* 39:W470–W474. doi: 10.1093/nar/gkr408

Sanderson, M. J., and Shaffer, H. B. (2002). Troubleshooting Molecular Phylogenetic Analyses. *Annu. Rev. Ecol. Syst.* 33, 49–72. doi: 10.1146/annurev.ecolsys.33.010802.150509

Schaller, A. (2013). "Plant Subtilisins," in *Handbook Of proteolytic enzymes*, eds A. J. Barrett, J. F. Woessner, and N. D. Rawlings (London: Elsevier Academic Press), 3247–3254. doi: 10.1016/B978-0-12-382219-2.00717-1

Seidah, N. G. (2013a). "Proprotein Convertase PCSK9," in *Handbook Of proteolytic enzymes*, eds A. J. Barrett, J. F. Woessner, and N. D. Rawlings (London: Elsevier Academic Press), 3315–3322. doi: 10.1016/B978-0-12-382219-2.00732-8

Seidah, N. G. (2013b). "Site-1 Protease," in *Handbook Of proteolytic enzymes*, eds A. J. Barrett, J. F. Woessner, and N. D. Rawlings (London: Elsevier Academic Press), 3265–3270. doi: 10.1016/B978-0-12-382219-2.00721-3

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7:539. doi: 10.1038/msb.2011.75

Siezen, R. J., and Leunissen, J. A. (1997). Subtilases: The superfamily of subtilisin-like serine proteases. *Protein Sci.* 6, 501–523. doi: 10.1002/pro.5560060301

Siezen, R. J., Renckens, B., and Boekhorst, J. (2007). Evolution of prokaryotic subtilases: Genome-wide analysis reveals novel subfamilies with different catalytic residues. *Proteins* 67, 681–694. doi: 10.1002/prot.21290

Siezen, R. J., Vos, W. M., de Leunissen, J. A., and Dijkstra, B. W. (1991). Homology modelling and protein engineering strategy of subtilases, the family of subtilisin-like serine proteinases. *Protein Eng.* 4, 719–737. doi: 10.1093/protein/4.7.719

Silmon de Monerri, N. C., Ruecker, A., and Blackman, M. J. (2013). "Plasmodium Subtilisins," in *Handbook Of proteolytic enzymes*, eds A. J. Barrett, J. F. Woessner, and N. D. Rawlings (London: Elsevier Academic Press), 3260–3265. doi: 10.1016/B978-0-12-382219-2.00720-1

Smith, E. L., Markland, F. S., Kasper, C. B., DeLange, R. J., Landon, M., and Evans, W. H. (1966). The complete amino acid sequence of two types of subtilisin, BPN' and carlsberg. *J. Biol. Chem.* 241, 5974–5976. doi: 10.1016/S0021-9258(18)96365-9

Solanki, P., Putatunda, C., Kumar, A., Bhatia, R., and Walia, A. (2021). Microbial proteases: Ubiquitous enzymes with innumerable uses. *3 Biotech* 11:428. doi: 10.1007/s13205-021-02928-z

Spanka, R., and Fritze, D. (1993). *Bacillus* cohnii sp. nov., a new, obligately alkaliphilic, oval-spore-forming *Bacillus* species with ornithine and aspartic acid instead of diaminopimelic acid in the cell wall. *Int. J. Syst. Bacteriol.* 43, 150–156. doi: 10.1099/00207713-43-1-150

Stothard, P. (2000). The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *BioTechniques* 28, 1102–1104. doi: 10.2144/00286ir01

Suzuki, M., Taguchi, S., Yamada, S., Kojima, S., Miura, K. I., and Momose, H. (1997). A novel member of the subtilisin-like protease family from *Streptomyces albogriseolus*. *J. Bacteriol.* 179, 430–438. doi: 10.1128/jb.179.2.430-438.1997

Sweeney, P. J., and Walker, J. M. (1993). Proteinase K (EC 3.4.21.14). *Methods Mol. Biol.* 16, 305–311. doi: 10.1385/0-89603-234-5:305

Takami, H., Akiba, T., and Horikoshi, K. (1990). Characterization of an alkaline protease from *Bacillus* sp. no. AH-101. *Appl. Microbiol. Biotechnol.* 33, 519–523. doi: 10.1007/BF00172544

Takimura, Y., Saito, K., Okuda, M., Kageyama, Y., Saeki, K., Ozaki, K., et al. (2007). Alkaliphilic *Bacillus* sp. strain KSM-LD1 contains a record number of subtilisin-like serine proteases genes. *Appl. Microbiol. Biotechnol.* 76, 395–405. doi: 10.1007/s00253-007-1022-9

Tan, G., Muffato, M., Ledergerber, C., Herrero, J., Goldman, N., Gil, M., et al. (2015). Current methods for automated filtering of multiple sequence alignments frequently worsen single-gene phylogenetic inference. *Syst. Biol.* 64, 778–791. doi: 10.1093/sysbio/syv033

Taylor, A., and Qiu, Y.-L. (2017). Evolutionary history of subtilases in land plants and their involvement in symbiotic interactions. *Mol. Plant Microbe Interact.* 30, 489–501. doi: 10.1094/MPMI-10-16-0218-R

Teufel, F., Almagro Armenteros, J. J., Johansen, A. R., Gíslason, M. H., Pihl, S. I., Tsirigos, K. D., et al. (2022). SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat. Biotechnol.* 40, 1023–1025. doi: 10.1038/s41587-021-01156-3

Tjalsma, H., Bolhuis, A., Jongbloed, J. D., Bron, S., and van Dijl, J. M. (2000). Signal peptide-dependent protein transport in *Bacillus* subtilis: A genome-based survey of the secretome. *Microbiol. Mol. Biol. Rev.* 64, 515–547. doi: 10.1128/MMBR.64.3.515-547.2000

Toogood, H. S., Smith, C. A., Baker, E. N., and Daniel, R. M. (2000). Purification and characterization of Ak.1 protease, a thermostable subtilisin with a disulphide bond in the substrate-binding cleft. *Biochem. J.* 350, 321–328. doi: 10.1042/bj3500321

Trifinopoulos, J., Nguyen, L.-T., Haeseler, A., and von Minh, B. Q. (2016). W-IQ-TREE: A fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* 44:W232–W235. doi: 10.1093/nar/gkw256

Tsuchida, O., Yamagata, Y., Ishizuka, T., Arai, T., Yamada, J.-I., Takeuchi, M., et al. (1986). An alkaline proteinase of an alkalophilic *Bacillus* sp. *Curr. Microbiol.* 14, 7–12. doi: 10.1007/BF01568094

Tunlid, A., and Jansson, S. (1991). Proteases and their involvement in the infection and immobilization of nematodes by the nematophagous fungus *Arthrobotrys oligospora*. *Appl. Environ. Microbiol.* 57, 2868–2872. doi: 10.1128/aem.57.10.2868-2872.1991

Turner, D. P. J., Wooldridge, K. G., and Ala'Aldeen, D. A. A. (2002). Autotransported serine protease A of *Neisseria meningitidis*: An immunogenic, surface-exposed outer membrane, and secreted protein. *Infect. Immun.* 70, 4447–4461. doi: 10.1128/IAI.70.8.4447-4461.2002

Valbuzzi, A., Ferrari, E., and Albertini, A. M. (1999). A novel member of the subtilisin-like protease family from *Bacillus* subtilis. *Microbiology* 145, 3121–3127. doi: 10.1099/00221287-145-11-3121

van der Laan, J. M., Teplyakov, A. V., Kelders, H., Kalk, K. H., Misset, O., Mulleners, L. J., et al. (1992). Crystal structure of the high-alkaline serine protease PB92 from *Bacillus alcalophilus*. *Protein Eng.* 5, 405–411. doi: 10.1093/protein/5.5.405

Vilo, C., Galetovic, A., Araya, J. E., Gómez-Silva, B., and Dong, Q. (2015). Draft genome sequence of a *Bacillus* bacterium from the atacama desert wetlands metagenome. *Genome Announcements* 3:e955–e915. doi: 10.1128/genomeA.00955-15

Völkl, P., Markiewicz, P., Stetter, K. O., and Miller, J. H. (1994). The sequence of a subtilisin-type protease (aerolysin) from the hyperthermophilic archaeum *Pyrobaculum aerophilum* reveals sites important to thermostability. *Protein Sci.* 3, 1329–1340. doi: 10.1002/pro.5560030819

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview Version 2–a multiple sequence alignment editor and analysis workbench. *Bioinf.* 25, 1189–1191. doi: 10.1093/bioinformatics/btp033

Wells, J. A., Powers, D. B., Bott, R. R., Graycar, T. P., and Estell, D. A. (1987). Designing substrate specificity by protein engineering of electrostatic interactions. *Proc. Natl. Acad. Sci. U.S.A.* 84, 1219–1223. doi: 10.1073/pnas.84.5.1219

Withers-Martinez, C., Jean, L., and Blackman, M. J. (2004). Subtilisin-like proteases of the malaria parasite. *Mol. Microbiol.* 53, 55–63. doi: 10.1111/j.1365-2958.2004.04144.x

Wu, J., Bian, Y., Tang, B., Chen, X., Shen, P., and Peng, Z. (2004). Cloning and analysis of WF146 protease, a novel thermophilic subtilisin-like protease with four inserted surface loops. *FEMS Microbiol. Lett.* 230, 251–258. doi: 10.1016/S0378-1097(03)00914-5

Xu, Y., Wang, S., Li, L., Sahu, S. K., Petersen, M., Liu, X., et al. (2019). Molecular evidence for origin, diversification and ancient gene duplication of plant subtilases (SBTs). *Sci. Rep.* 9:12485. doi: 10.1038/s41598-019-48664-6

Yahiro, K., Moss, J., and Noda, M. (2013). "*Subtilase Cytotoxin* (SubAB)," in *Handbook Of proteolytic enzymes*, eds A. J. Barrett, J. F. Woessner, and N. D. Rawlings (London: Elsevier Academic Press), 3155–3161. doi: 10.1016/B978-0-12-382219-2.00694-3

Yamagata, Y., Maeda, H., Nakajima, T., and Ichishima, E. (2002). The molecular surface of proteolytic enzymes has an important role in stability of the enzymatic activity in extraordinary environments. *Eur. J. Biochem.* 269, 4577–4585. doi: 10.1046/j.1432-1033.2002.03153.x

Yamagata, Y., Sato, T., Hanzawa, S., and Ichishima, E. (1995). The structure of subtilisin ALP I from alkalophilic *Bacillus* sp. NKS-21. *Curr. Microbiol.* 30, 201–209. doi: 10.1007/BF00293634

Yanagida, N., Uozumi, T., and Beppu, T. (1986). Specific excretion of *Serratia marcescens* protease through the outer membrane of *Escherichia coli*. *J. Bacteriol.* 166, 937–944. doi: 10.1128/jb.166.3.937-944.1986

Zhao, G.-Y., Chen, X.-L., Zhao, H.-L., Xie, B.-B., Zhou, B.-C., and Zhang, Y.-Z. (2008). Hydrolysis of insoluble collagen by deseasin MCP-01 from deep-sea *Pseudoalteromonas* sp. SM9913: Collagenolytic characters, collagen-binding ability of C-terminal polycystic kidney disease domain, and implication for its novel role in deep-sea sedimentary particulate organic nitrogen degradation. *J. Biol. Chem.* 283, 36100–36107. doi: 10.1074/jbc.M804438200

Zhu, D., Tanabe, S.-H., Xie, C., Honda, D., Sun, J., and Ai, L. (2014). *Bacillus ligniniphilus* sp. nov., an alkaliphilic and halotolerant bacterium isolated from sediments of the South China Sea. *Int. J. Syst. Evol. Microbiol.* 64, 1712–1717. doi: 10.1099/ijs.0.058610-0

Zhu, X. L., Ohta, Y., Jordan, F., and Inouye, M. (1989). Pro-sequence of subtilisin can guide the refolding of denatured subtilisin in an intermolecular process. *Nature* 339, 483–484. doi: 10.1038/339483a0

2.2 Biochemical characterisation of a novel oxidatively stable, halotolerant, and high-alkaline subtilisin from *Alkalihalobacillus okhensis* Kh10-101^T

Fabian Falkenberg, Jade Rahba, David Fischer, Michael Bott, Johannes Bongaerts and Petra Siegert (2022) FEBS Open Bio, DOI: 10.1002/2211-5463.13457

Authors contributions:
FF, JB, and PS conceived and designed the experiments.
FF conducted all experiments except those of DF and JR, collected and analysed the data.
DF and JR carried out the cloning and pre-experiments.
FF wrote the original draft.
FF, JB, MB, and PS revised the manuscript.

All figures and tables are from this study.

Overall contribution FF: 90 %

FEBS openbio

FEBS PRESS science publishing by scientists

Biochemical characterization of a novel oxidatively stable, halotolerant, and high-alkaline subtilisin from *Alkalihalobacillus okhensis* Kh10-101^T

Fabian Falkenberg[1], Jade Rahba[1], David Fischer[1], Michael Bott[2], Johannes Bongaerts[1] and Petra Siegert[1]

1 Institute of Nano- and Biotechnologies, Aachen University of Applied Sciences, Jülich, Germany
2 Institute of Bio- and Geosciences, IBG-1: Biotechnology, Forschungszentrum Jülich, Germany

Keywords
Alkalihalobacillus okhensis; detergent protease; halotolerant protease; high-alkaline subtilisin; oxidative stable protease

Correspondence
P. Siegert, Institute of Nano- and Biotechnologies, Aachen University of Applied Sciences, 52428 Jülich, Germany
Tel: +49 241 6009 53124
E-mail: siegert@fh-aachen.de
Website: https://www.fh-aachen.de/forschung/inb

(Received 20 May 2022, revised 13 June 2022, accepted 20 June 2022)

doi:10.1002/2211-5463.13457

Halophilic and halotolerant microorganisms represent a promising source of salt-tolerant enzymes suitable for various biotechnological applications where high salt concentrations would otherwise limit enzymatic activity. Considering the current growing enzyme market and the need for more efficient and new biocatalysts, the present study aimed at the characterization of a high-alkaline subtilisin from *Alkalihalobacillus okhensis* Kh10-101^T. The protease gene was cloned and expressed in *Bacillus subtilis* DB104. The recombinant protease SPAO with 269 amino acids belongs to the subfamily of high-alkaline subtilisins. The biochemical characteristics of purified SPAO were analyzed in comparison with subtilisin Carlsberg, Savinase, and BPN'. SPAO, a monomer with a molecular mass of 27.1 kDa, was active over a wide range of pH 6.0–12.0 and temperature 20–80 °C, optimally at pH 9.0–9.5 and 55 °C. The protease is highly oxidatively stable to hydrogen peroxide and retained 58% of residual activity when incubated at 10 °C with 5% (v/v) H_2O_2 for 1 h while stimulated at 1% (v/v) H_2O_2. Furthermore, SPAO was very stable and active at NaCl concentrations up to 5.0 M. This study demonstrates the potential of SPAO for biotechnological applications in the future.

Proteases are among the most commercially valuable enzymes, with subtilisins or alkaline proteases from microbial sources accounting for the largest market share [1,2]. They have been extensively studied in terms of their biological function to gain insights into the mechanism of enzyme catalysis and the structure–function relationship of proteins, and because of their significant applications in various industries [1]. Subtilisins belong to the group of subtilases, which is one of the largest families of serine peptidases, and are classified as S8 according to the MEROPS database [3]. Furthermore, subtilisins are further classified as true subtilisins, high-alkaline subtilisin, intracellular subtilisin, and phylogenetically intermediate subtilisins (PIS) [4,5]. They are ubiquitously distributed in various organisms, including bacteria, archaea, eukaryotes, yeasts, and viruses [4]. However, *Bacillus* as the most prominent source spawned alkaline proteases such as

Abbreviations
aa, amino acid; *aprE*, extracellular alkaline protease gene; CHCA, α-Cyano-4-hydroxycinnamic acid; EDTA, ethylenediaminetetraacetic acid; IEF, isoelectric focussing; LB, lysogeny broth; LM, length marker; MALDI-TOF-MS, matrix-assisted laser desorption/ionization -time- of- flight mass spectrometry; MSA, multiple sequence alignment; MWCO, molecular weight cut-off; PAGE, polyacrylamide gel electrophoresis; PCR, polymerase chain reaction; PDB, protein data bank; pI, isoelectric point; PMSF, phenylmethylsulfonyl fluoride; pNA, para-nitroanilide; SDS, sodium dodecyl sulfphate; SPAO, subtilisin protease *A.* okhensis; suc, N-succinyl; TCA, trichloroacetic acid; tet, tetracycline; TFA, trifluoroacetic acid; T_m, melting temperature.

g

subtilisin Carlsberg, BPN', and Savinase, with their major application as detergent enzymes with excellent properties including high stability toward extreme temperatures, pH, organic solvents, detergents, and oxidizing compounds [6,7]. Besides application in detergents, subtilisins find applications, for example, in leather processing, food, wastewater treatment, and cosmetics [6,8].

The extracellular subtilisins of microorganisms are mainly involved in nutrient supply, and their properties are thought to depend entirely on the host and its adaptability to the immediate environment [6]. Since the size of the microbial world is almost unlimited, and many different microbial sources can be exploited, this biodiversity holds great potential for enriching the repertoire of known enzymes with new and high-performing enzymes. Enzymes isolated from extremophilic microorganisms such as thermophilic, psychrophilic, and especially halotolerant or halophilic organisms offer enormous potential to meet industrial needs, as evidenced by the increasing number of newly characterized subtilisins [9-11]. In addition to classical methods of screening microorganisms with new interesting proteases from various environments, different molecular biology techniques, such as metagenomic analysis, directed evolution, and site-directed mutagenesis have been used to gain or engineer numerous proteases with improved or novel properties [9]. Beside these labor-intensive methods, genome sequencing and automated annotation are adding potential sequences to the rapidly growing online database and provide an alternative approach to search for candidate protease genes for industrial applications.

Alkalihalobacillus okhensis Kh10-101^T is a gram-positive, strictly aerobic, rod-like bacterium isolated by Nowlan et al. [12] from an Indian saltpan near the port of Okha. The strain was first classified as *Bacillus okhensis* and in 2020 reclassified into *Alkalihalobacillus okhensis* by Patel and Gupta [13]. As described by Krishna et al. [14], the genome of *A. okhensis* encoded almost 40 different proteases with members of the serine protease family. Since *A. okhensis* is described as a moderate halophile and an alkaliphile with optimal growth conditions of pH 10 and 5% NaCl, the extracellular proteases derived from this organism may be of potential industrial importance [14]. In this research, the gene for subtilisin WP_034632645.1 was cloned, overexpressed in *B. subtilis* DB104, and purified by ion-exchange chromatography. This is the first report on the biochemical characterization of the recombinant subtilisin protease of *A. okhensis* (SPAO) and includes a comparison with the commercially applied subtilisins Carlsberg, Savinase, and BPN'.

Materials and methods

Reagents and enzymes

Polymerase chain reactions (PCR) were performed with Phusion® Hot Start II High-Fidelity Green Master Mix from Thermo Fisher Scientific GmbH (Karlsruhe, Germany). Oligonucleotides were synthesized by Eurofins Genomics GmbH (Ebersberg, Germany). Restriction enzymes, T4 DNA Ligase, and GeneRuler™ 1 kb DNA ladder were purchased from Thermo Fisher Scientific GmbH. Peptide protease substrates were purchased from BACHEM (Bubendorf, Switzerland). Azocasein, α-CHCA (alpha-cyano-4-hydroxycinnamic acid), subtilisin Carlsberg, and Savinase were purchased from Sigma-Aldrich (Schnelldorf, Germany). BPN' was from DuPont (Wilmington, NC, USA). MALDI-TOF MS protein standards were purchased from LaserBio Labs (Valbonne, France). Lysozyme from chicken and materials for isoelectric focusing (IEF) were purchased from SERVA (Heidelberg, Germany). Centrifugal spin columns and PMSF (phenylmethylsulfonyl fluoride) were purchased from Avantor VWR (Radnor, PA, USA). Molecular weight marker for use with SDS/PAGE (sodium dodecyl sulfate–polyacrylamide gel electrophoresis) was purchased from Bio-Rad (Hercules, CA, USA). All other chemicals were acquired from Carl Roth (Karlsruhe, Germany).

Bioinformatic analysis

To illuminate the sequence similarity between different well-known characterized subtilisins and the protease SPAO, a multiple sequence alignment (MSA) was performed using the mature subtilisin/serine protease sequences from various *Bacillus* strains. The SPAO sequence was blasted by employing the blastp suite of NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The signal peptide and propeptide sequences were excluded before MSA, and phylogenetic tree construction was performed via Phylogeny.fr (http://www.phylogeny.fr/index.cgi) using the 'One-Click' option [15]. MSA for analysis with ESPript 3.0 was performed with CLUSTAL OMEGA (https://www.ebi.ac.uk/Tools/msa/clustalo/) [16,17]. Phylogenetic trees were displayed with the ITOL software (https://itol.embl.de/) [18]. MSA was drawn with ESPript 3.0 using %strict option (percentage of strictly conserved residues per column) for the coloring scheme (https://espript.ibcp.fr/ESPript/ESPript/). For homology modeling, the functional amino acid sequence of SPAO without its signal peptide and propeptide was used. A structure prediction was performed through the Iterative Threading Assembly Refinement (I-TASSER) server (https://zhanggroup.org/I-TASSER/) [19]. The homology model was visualized with Mol* Viewer (https://www.rcsb.org/3d-view) [20]. Swiss-PdbViewer (http://www.expasy.org/spdbv/) was used to determine

FEBS Open Bio **12** (2022) 1729–1746 © 2022 The Authors. *FEBS Open Bio* published by John Wiley & Sons Ltd on behalf of Federation of European Biochemical Societies.

surface-exposed residues and to calculate the electrostatic potential with standard settings using the Poisson–Boltzmann Equation [21]. The theoretical pI and molecular mass of the functional domain were calculated with https://web.expasy.org/compute_pi/. The identification of the signal peptide was performed with the SIGNALP6.0 software https://services.healthtech.dtu.dk/service.php?SignalP-6.0 [22].

Strains and growth conditions

Alkalihalobacillus okhensis Kh10-101^T (DSM 23308) was purchased from the German collection of microorganisms and cell cultures GmbH (DSMZ) and cultivated according to their recommendations in medium 830, pH 9.5, at 35 °C. An overnight culture was used for genomic DNA preparation with the InnuSPEED Bacteria/Fungi DNA Kit (Analytik Jena™, Jena, Germany) according to the manufacturer's recommendations. *Bacillus subtilis* DB104 was used as host for cloning and protein production [23] and cultivated in lysogeny broth (LB) medium (10 g·L^{-1} tryptone, 5 g·L^{-1} yeast extract, 10 g·L^{-1} NaCl, pH 7.0; Carl Roth).

Plasmid construction and cloning

A pBC16-based expression plasmid (Acct. No. U32369.1) was used for recombinant protease production with *B. subtilis* DB104 [24]. The pBC16 derivative pFF-RED was obtained by exchanging the *mob* region by an expression cassette comprising the subtilisin Carlsberg promoter from *B. paralicheniformis* ATCC 9945a followed by the gene *eforRED* encoding a red chromoprotein (acc. no. ACD13196.1). This marker gene is flanked by BbsI restriction sites allowing cloning the gene of interest via Golden Gate cloning.

The DNA sequences encoding the protease signal peptide, the propeptide, and the mature domain were amplified from the genomic DNA using the Phusion® Hot Start II High-Fidelity polymerase according to the manufacturer's recommendations. The NCBI reference sequence NZ_JRJU01000039.1 was used to design primers for the *aprE* gene (extracellular alkaline protease) encoding the protein WP_034632645.1 [12]. To remove an internal BbsI restriction site within the propeptide, two PCRs were performed with primers introducing a silent point mutation at the internal BbsI restriction site. For PCR 1, the forward primer (5′-AAAGAAGACGGAATGAAAAAGTTATTTACGAAAGTAGTTGCC-3′) and the reverse primer (5′-GGTTAAAAATACTAACCTCAATATCTTCCTCGATGAAAGCAATAG-3′) were used. For PCR 2, the forward primer (5′-CTATTGCTTTCATCGAGGAAGATATTGAGGTTAGTATTTTTAACC-3′) and the reverse primer (5′-AAAGAAGACCCGTTATCTTGTAGCAGCTTCGGCATTAACAAG-3′) were used. The two PCR fragments of 342 and 851 bp were combined within an overlap extension PCR using the two outer primer pairs introducing two BbsI restriction sites and corresponding overhangs to the cloning site of pFF-RED.

The PCR product was cloned via Golden Gate cloning into the BbsI site of pFF-RED [25]. The product (pFF003) was subsequently used to transform *B. subtilis* DB104 naturally competent cells as described elsewhere [26].

PCR success was verified by agarose gel electrophoresis. After transformation, clones were spread onto LB agar plates (1.5% (w/v) agar) supplemented with 20 µg·mL^{-1} tetracycline (tet_{20}) and 2.5% (w/v) skim milk powder. The clones were grown at 37 °C overnight. Colonies with clear halo zones indicated protease activity and were selected as positive clones. The clones were subsequently cultivated overnight at 37 °C and 200 r.p.m. in 10 mL LB_{tet20} medium. Plasmids from each protease-positive clone were isolated by GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific), and the desired cloning result was confirmed by double-restriction digestion and plasmid DNA Sanger sequencing (Eurofins Genomics).

Recombinant protease production

Production of the protease by *Bacillus subtilis* DB104 was carried out by inoculating 10 mL of LB_{tet20} medium with a freshly plated clone and cultivated over 8 h at 180 r.p.m. and 37 °C. Subsequently, an overnight culture with 50 mL of the preculture medium (Tables S1 and S3) was inoculated with 100 µL of the over-the-day culture. After cultivation overnight at 37 °C and 180 r.p.m., the bioreactors were inoculated to an optical density at 600 nm (OD600) of 0.25. The fermentation was performed in a DASGIP® parallel reactor system (DASGIP, Jülich, Germany) with four 1-L reactors. The air supply was performed with an L-sparger and a volume flow of 0.5 vvm. The oxygen saturation was set at 30% and regulated by the stirrer speed (max. 1500 r.p.m., min. 350 r.p.m.). The pH was adjusted to 7.4 and regulated by adding either 4 M NaOH or 20% (v/v) H_2SO_4. The reaction temperature was set at 37 °C. For cultivation, a high protein content medium with soy peptone was used (Tables S2 and S3). After 12 h of cultivation, a glucose feed (1.5 g·mL^{-1}) was started with 1 mL·h^{-1} for 7 h, then 1.5 mL·h^{-1} for 9 h and 1.1 mL·h^{-1} until the end of the fermentation. After 10 h, a polypropylene glycol 2000 (PPG) feed (0.5 mol·L^{-1}) was established for 10 h with 0.25 mL·h^{-1}. The fermentation was performed over 48 h, and the supernatant was harvested by centrifugation at 3000 ***g*** for 20 min. For storage at 4 °C, 10% (v/v) of propylene glycol was added to the supernatant. The protease production was confirmed by a proteolytic activity assay using suc-AAPF-pNA as substrate and by SDS/PAGE.

Enzyme purification

One hundred and forty millilitre of the cell-free supernatant was divided into portions of 15 mL to which 30 mL ice-cold 96% (v/v) ethanol was added, and the mixture was incubated overnight at −20 °C. After incubation, the samples were centrifuged for 20 min at 13 000 ***g*** and 4 °C. The supernatant was discarded, and the pellets were washed with 15 mL ice-cold 96% (v/v) EtOH by vortexing. The samples were centrifuged again as described above, and the supernatant was removed. The pellets were resuspended in 7 mL of running buffer (10 mM HEPES/NaOH buffer, pH 8.0). The concentrated sample was then centrifuged as described above before being applied to a 50 mL HiPrep 26/10 desalting column coupled to an Äkta Avant 25 (Cytiva Europe, Freiburg, Germany). The protein was eluted with running buffer. The desalting process was monitored by measuring the conductivity and the absorbance at 280 nm. The peak protein fractions without salt were collected and used for ion-exchange chromatography.

According to the theoretical isoelectric points and the pH of the buffer used, a cation exchanger column (22 mL S-Sepharose FF, GE Healthcare, Chicago, IL, USA) was used using the Äkta Avant 25 device (Cytiva Europe GmbH, Freiburg, Germany). The column was equilibrated with running buffer. The desalted protein sample was applied to the ion exchanger column and washed with two column volumes of running buffer. The protease was eluted with a linear salt gradient using the elution buffer (10 mM HEPES/NaOH, pH 8.0, 1 M NaCl) and was collected in fractions of 5 mL. Protein elution was monitored by measuring the absorbance at 280 nm. The protein-containing fractions were subsequently used for further analysis. For long-term storage, the purified proteases were stored at −80 °C with 10% (v/v) of glycerol.

Enzyme activity assay

The hydrolytic activity of proteases was determined with the tetrapeptide substrate *N*-succinyl-Ala-Ala-Pro-Phe-p-nitroanilide (suc-AAPF-pNA) at 30 °C in 100 mM Tris/HCl buffer, pH 8.6, containing 0.1% (w/v) Brij®35 [27]. The substrate was prepared as a 110 mM stock solution in dimethyl sulfoxide and diluted 1 : 100 in the reaction mix. The amount of released *p*-nitroaniline ($\varepsilon_{410\ nm}$ = 8.48 $mM^{-1} \cdot cm^{-1}$) was determined by measuring the absorbance at 410 nm for 5 min every 30 s [27]. The reaction was performed either in 1 mL cuvette format (Ultrospec 2100 pro, Amersham Biosciences, Little Chalfont, UK) or in microtiter format (Infinite 200Pro, Tecan, Männedorf, Switzerland) with a reaction volume of 1 mL and 250 µL, respectively. One unit (U) of enzyme activity was defined as the amount of enzyme that produced 1 µmol of *p*-nitroaniline per minute under the assay conditions.

Protease activity was determined using azocasein as a substrate based on Brock et al. [28]. The partial hydrolysis of the substrate by the proteases releases dye-labeled smaller peptides that are no longer precipitable. The dye-labeled peptides were quantified at a wavelength of 440 nm. For the assay, a fresh substrate solution of 2% (w/v) azocasein in 100 mM Tris/HCl, pH 8.6, was prepared. The reaction was performed in 125 µL substrate solution with 75 µL of the protease sample in an appropriate dilution. The mixture was incubated at 37 °C and 300 r.p.m. for exact 30 min. The reaction was stopped by adding 600 µL of 20% (w/v) trichloroacetic acid (TCA) and kept at room temperature for 15 min, followed by centrifugation at 12 500 ***g*** for 5 min. Then, 600 µL of the supernatant was mixed with 700 µL of 1 M NaOH, and the absorbance at 440 nm (0.5 ± 0.1) was measured. One unit (U) of activity was defined as the amount of enzyme required to increase the corresponding absorbance value by 0.01 units per minute under the conditions described above.

Protein measurement, electrophoresis, and analytical methods

Protein concentration was determined by measuring the absorbance ratio 590/450 nm using Roti® Nanoquant (Carl Roth) with bovine serum albumin fraction V (Carl Roth) as a standard based on the method of Bradford [29].

The molecular mass of the purified proteases was analyzed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) using an Axima confidence (Shimadzu Europe, Duisburg, Germany) in linear positive mode with pulsed extraction optimized for the theoretical molecular mass. Data were analyzed with MMASS [30]. Purified protease was precipitated by mixing 1 : 2 with 20% (w/v) TCA. The precipitate was resuspended with 50 µL dH_2O/0.1% (v/v) trifluoroacetic acid (TFA) and 2 µL of 2 M NaOH. Samples were diluted 1 : 10 with α-cyano-4-hydroxycinnamic acid (CHCA; Sigma-Aldrich) as matrix. One microlitre of the dilution was applied onto the MALDI-TOF-MS target plate and air-dried, and then, again 1 µL was applied. As mass standards, trypsinogen and bovine serum albumin from Laser-Bio (Valbonne, France) were used.

SDS/PAGE was performed using an 8–20% (v/v) resolving gel and a 6% (v/v) stacking gel as described by Miller et al. [31]. For sample preparation, 50 µL of 20% (w/v) TCA was added to 25 µL of the protein sample. The mixture was incubated for 2 min on ice and then centrifuged for 5 min at 12 000 ***g***. The supernatant was carefully discarded, and 50 µL of 2 × reducing SDS sample buffer after Laemmli and 50 µL of 0.1 M NaOH were added to the precipitated protein. The sample was subsequently boiled for 10 min at 95 °C. The electrophoresis was performed for 30 min at 300 V in Tris-glycine/SDS (Laemmli) buffer, pH 8.6. The gel was stained with Roti® Blue quick (Carl Roth) for 2 h, destained in dH_2O for 30 min, and photographed with the UVP® GelStudio touch device (Analytic Jena, Jena, Germany).

FEBS Open Bio **12** (2022) 1729–1746 © 2022 The Authors. *FEBS Open Bio* published by John Wiley & Sons Ltd on behalf of Federation of European Biochemical Societies.

For IEF-PAGE, purified SPAO and the reference proteases were rebuffered in 10 mM HEPES/NaOH, pH 7.0, using centrifugal spin columns with a molecular weight cutoff (MWCO) of 3 kDa. The SERVAGel™ IEF 3–10 gels were used according to the manufacturer's recommendations.

Effect of sodium dodecyl sulfate, hydrogen peroxide, and phenylmethylsulfonyl fluoride (PMSF) on enzyme activity and stability

To examine the effect of surfactant and oxidizing agents, hydrogen peroxide (1 and 5% (v/v)) and SDS (1 and 5% (w/v)) were added to the enzyme solution in 10 mM HEPES/NaOH, pH 8.0, and incubated for 1 h at 10 °C. The low incubation temperature was chosen to prevent autoproteolysis. The remaining activity was measured in the standard suc-AAPF-pNA activity assay. Residual activity of the proteases incubated in buffer with no additives was set as 100%.

The effect of 1 mM of the protease inhibitor PMSF was investigated by incubating the proteases in 10 mM HEPES/NaOH, pH 8.0, for 30 min on ice. Residual activity of the proteases incubated in buffer with no additions was set as 100%.

Effect of NaCl, EDTA, and $CaCl_2$ on enzyme activity and stability

The effect of NaCl on proteolytic activity was measured under standard reaction conditions for suc-AAPF-pNA with NaCl (0–5 M) in the reaction buffer. The influence of NaCl on enzyme stability was investigated by incubating the proteases in 10 mM HEPES/NaOH, pH 8.0, with NaCl (0–5 M) at 20 °C for 2 h. The % residual activities measured before incubation were set as 100%.

For the investigation of the effect of ethylenediaminetetraacetic acid (EDTA) and $CaCl_2$, the purified proteases were rebuffered in centrifugal spin columns (3 kDa MWCO). 10 mM HEPES/NaOH, pH 8.0, was added twice to the retentate after centrifugation at 4 °C and 12 000 ***g***. The rebuffered proteases (20 $\mu g \cdot mL^{-1}$) were incubated for 12 h at 4 °C with and without 20 mM EDTA in duplicates. After incubation, the reaction mixtures were diluted properly for analysis in a modified suc-AAPF-pNA assay. In this case, the diluted protease sample was incubated with reaction buffer containing $CaCl_2$ (0–25 mM) for 5 min before adding the substrate solution (suc-AAPF-pNA, 1.1 mM final).

Substrate spectrum

The substrate specificity of the proteases was determined using the synthetic peptide-4-nitroanilide substrates suc-Tyr-Val-Ala-Asp-pNA (YVAD), suc-Phe-Ala-Ala-Phe-pNA (FAAF), suc-Ala-Ala-Ala-pNA (AAA), suc-Ala-Ala-Val-Ala-pNA (AAVA), suc-Ala-Leu-Pro-Phe-pNA (ALPF), suc-Ala-Gly-Pro-Phe-pNA (AGPF), suc-Ala-Ala-Pro-Phe-pNA (AAPF), suc-Ala-Ala-Pro-Leu-pNA (AAPL), suc-Thr-Val-Ala-Ala-pNA (TVAA), and suc-Ala-Gly-Pro-Pro-pNA (AGPP) dissolved in dimethyl sulfoxide to a final concentration of 0.34 mM in the assay. Kinetic experiments were carried out as described above.

Effects of temperature and pH on enzyme activity and stability

To monitor thermal protein unfolding and to determine the melting point of the proteases, the environmentally sensitive fluorescent dye SYPRO™ Orange (Thermo Fisher Scientific GmbH) was used. During the unfolding process at higher temperatures, hydrophobic residues get exposed causing an increase in the SYPRO™ Orange fluorescence, which is monitored (Ex/Em = 470/550 nm) [32]. The unfolding kinetics were performed in a qPCR cycler (qTower3G, Analytic Jena) in at least triplicates. The detection mixture contained 5 × SYPRO™ Orange solution in 10 mM HEPES/NaOH buffer, pH 8.0, with 3 mM PMSF and purified and rebuffered proteases (35–320 $\mu g \cdot \mu L^{-1}$). The purified proteases were rebuffered in centrifugal spin columns (3 kDa MWCO). 10 mM HEPES/NaOH, pH 8.0, was added twice to the retentate after centrifugation at 4 °C and 12 000 ***g***. The thermocycler block was heated from 25 to 95 °C in steps of 2 °C with a 2-min holding time measuring fluorescence intensities at each step. The data analysis was performed using QPCRSOFT 4.0 (Analytic Jena), and apparent T_m values were calculated for each protease as the inflection point of the melting curve.

The temperature optimum was assayed between 20 and 90 °C in 5 °C steps with the suc-AAPF-pNA assay as described above. Before every measurement, the cuvette holder and the buffer were preheated at least for 5 min at the desired temperature. The pH of the reaction buffer was adapted until 80 °C. For each measurement, a new blank was used to exclude thermal instability of the substrate. The temperature stability was determined by measuring the residual activity after incubating the enzymes in 10 mM HEPES/NaOH, pH 8.0, at 20 and 50 °C for 3 h. The % residual activities were followed by measuring the residual activities every 20 min under standard reaction conditions for the suc-AAPF-pNA assay.

The optimal pH of the proteases was determined at 30 °C in 0.1 M Tris/maleate buffer (pH 5.0–7.0), 0.1 M Tris/HCl (pH 7.0–9.0) and 0.1 M glycine/NaOH (pH 9.0–12.5) under standard reaction conditions for suc-AAPF-pNA substrate. The effect of pH on enzyme stability was assayed by preincubating enzymes in said buffers for 4 h at 4 °C to prevent autoproteolysis. The % residual activities were measured under standard reaction conditions for the suc-AAPF-pNA assay.

Results and discussion

Cloning and expression of *aprE_A. okhensis* in *B. subtilis* DB104

The gene sequence of *aprE_A. okhensis* for SPAO was amplified from the halophilic strain *A. okhensis* Kh10-101^T as described in Materials and methods. A fragment of 1148 bp containing the coding region for the signal peptide, the propeptide, and the mature part of the protease was obtained, cloned into the vector pFF-RED, and transferred into *B. subtilis* DB104. Transformants with a clearing zone on LB agar plates supplemented with 2.5% (w/v) skim milk powder were analyzed by plasmid preparation, and pFF003 coding for SPAO was identified by restriction digestion-based molecular screening. Sequence data obtained by the Sanger sequencing method showed that the inserted sequence was identical to the nucleotide sequence of the *aprE_A. okhensis* gene available in GenBank, except for the elimination of the internal BbsI site in the coding region of the propeptide.

Bioinformatic analysis and homology modeling

The *aprE* gene from *A. okhensis* Kh10-101^T comprises 1149 bp encoding a protein of 382 amino acids (aa). Most subtilases have a multi-domain structure consisting of a signal peptide (for translocation), a propeptide (for maturation by autoproteolytic cleavage), a protease domain, and frequently one or more additional domains [33]. The signal peptide prediction revealed a Sec signal peptide with a cleavage site between amino acids 27 and 28 with a probability of 96.9% [22]. The propeptide was identified by multiple sequence alignment to comprise amino acids 29–113. Thus, three functional domains were identified with a 27-aa signal peptide, a 86-aa propeptide, and a 269-aa mature endopeptidase S8 domain (Fig. 2). The analysis of the mature protease sequence revealed a molecular mass of 27.14 kDa and a calculated pI of 9.6. The catalytic triad consisted of Asp^{32}, His^{62}, and Ser^{215} (numbers based on the mature protease sequence).

The predicted amino acid sequence of the mature part of SPAO was aligned and compared with well-characterized proteases of the three subtilisin families (true, high-alkaline, phylogenetic intermediate) retrieved from MEROPS [3] and the UniProt database [34] resulting in a phylogenetic tree (Fig. 1). Additionally, a BlastP search was performed. The mature part of SPAO showed the highest sequence similarity with a subtilisin from *Alkalihalobacillus alcalophilus* (91.45%), which corresponds to subtilisin Sendai from *Bacillus sp.* G-825-6 [35]. Furthermore, it showed high similarity to the subtilisin of *Alkalihalobacillus pseudoalkaliphilus* (90.33%), subtilisin Savinase from *Lederbergia lenta* (formerly *Bacillus lentus*; 82.44%) [36], and subtilisin PB92 from *Alkalihalobacillus alcalophilus* (82.84%) [37]. Moreover, SPAO showed a more distant relationship to the well-characterized true subtilisins BPN' (56.13%) [38] and subtilisin Carlsberg (57.84%) [39]. In the phylogenetic tree, SPAO is therefore clearly distinct from the phylogenetic intermediate subtilisins and the true subtilisins, indicating its affiliation with the highly alkaline subtilisin group. Therefore, for comparative biochemical characterization, the two true subtilisins BPN' and subtilisin Carlsberg and the high-alkaline subtilisin Savinase were used. The MSA for those three enzymes and SPAO is shown in Fig. 2. The sequence similarity between SPAO and

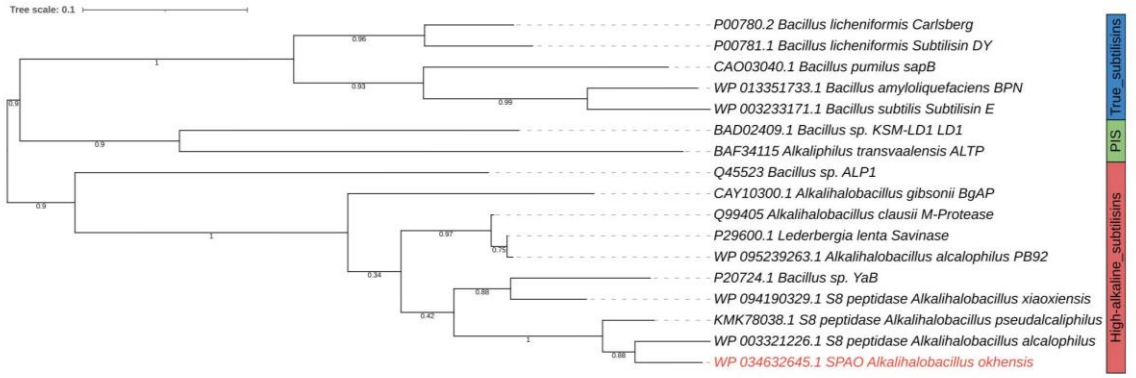


Fig. 1. Evolutionary phylogenetic tree of various subtilisins from different species of the family *Bacillaceae*. Maximum-likelihood phylogenetic analysis of the mature protease domains was performed using the Phylogeny.fr server. Numbers at nodes indicate support for the internal branches within the tree obtained by approximate likelihood ratio test (SH-like aLRT).

FEBS Open Bio **12** (2022) 1729–1746 © 2022 The Authors. *FEBS Open Bio* published by John Wiley & Sons Ltd on behalf of Federation of European Biochemical Societies.

Savinase is high as mentioned above and shows typical amino acid residues for high-alkaline subtilisins, which were identified by MSA of true subtilisins, PIS, and highly alkaline subtilisins, as described by Yamagata et al. [40].

The 3D structure identification of an enzyme is of great interest in order to identify potential key residues essential for the specific activity, halotolerance, etc. These key residues and their interactions also play a role in potential enzyme engineering efforts. Therefore, homology modeling-based structural analysis of recombinant SPAO was also performed. The predicted 3D structure of mature SPAO calculated with the I-TASSER server [19] is shown in Fig. 3 with a C-score of 1.55. The C-score measurement determines the quality of the resulting models in the range [−5.2], where a C-score with a higher value means a model with high confidence [41]. The highest structural similarity (99.4%) was displayed to *Lederbergia lenta* subtilisin Savinase (PDB: 1C9J) with a template modeling score of 0.994 and 0.46 Å root-mean-square deviation. In general, subtilisins show a great similarity in their molecular structure [42]. *In silico* analysis of the model for probable metal-binding sites suggested that SPAO harbors two potential Ca^{2+} binding sites within its functional domain (site 1: Asp^{40}, Leu^{73}, Val^{79}, Ile^{77}; site 2: Ala^{163}, Ala^{168}). Using this model to analyze the surface-exposed residues with the Swiss-PdbViewer at a threshold of 20% of the accessible surface revealed that six Arg residues and one Lys residue are exposed at the surface, which is important for salt and pH adaptation, as will be explained later [21]. In addition, the electrostatic potential of SPAO was evaluated using the Swiss-PdbViewer and showed a mainly positively charged backside and a negatively charged region around the active site at pH 7.0, which helps to interpret the adaption to high salt concentrations as discussed later (Fig. 4).

Recombinant protease production and purification

SPAO was produced by *B. subtilis* DB104 pFF003 at 1-L scale using a DASGIP parallel fermentation device as described in Materials and methods. The culture supernatant showed an activity of 103 U·mL^{-1} (AAPF) and a protein concentration of 0.6 mg·mL^{-1}. It was used for a three-step purification process that involved ethanol precipitation, desalting, and ion-exchange chromatography. Successful production and apparent homogeneity of SPAO were confirmed via SDS/PAGE (Fig. 5). The purified SPAO had a molecular mass of about 27 kDa on SDS/PAGE, which correlates to the theoretical mass of 27.1 kDa (Fig. 5). Additionally, the size of 27.1 kDa was confirmed by MALDI-TOF MS analysis (Fig. S1). The purified protease had a specific activity of 139 U·mg^{-1} for the AAPF substrate and 528 U·mg^{-1} for azocasein. Similar activities after production with *B. subtilis* and purification could be achieved for subtilisin E with 486 U·mg^{-1} for azocasein under the same experimental conditions [43]. The analysis of the purified and rebuffered SPAO for its isoelectric point revealed a pI of approx. 9.8, which is close to the predicted pI of 9.6 (Fig. 5). Subtilisin Carlsberg showed a pI of 8.0 and Savinase of 9.8 (data not shown). High-alkaline subtilisins have a high pI in common [44].

Effects of temperature and pH on enzyme activity and stability

Stability measurements of proteolytic enzymes are difficult due to potential autoproteolytic cleavage during unfolding. To monitor thermal protein unfolding rather than autoproteolysis and to obtain an estimate of the conformational stability of SPAO, the denaturation curve of the enzyme inhibited by PMSF was measured. The inhibition with 1 mM PMSF was tested before, resulting in a complete loss of activity (Table 1). Figure 6 shows the normalized fluorescent signal curves (relative to the maximal fluorescence measured) for SPAO in comparison with the well-characterized proteases BPN' and subtilisin Carlsberg. SPAO showed a T_m value of 53.0 °C, which was lower than the T_m values of BPN′ (58.5 °C) and subtilisin Carlsberg (64.0 °C). As expected, the proteases having a higher temperature optimum (see below; Fig. 7) display also a higher melting point, suggesting higher structural integrity at elevated temperatures. In applications such as detergents, the trend is toward more cold-active enzymes, which, however, lack stability at higher temperatures [32,45]. The high-alkaline subtilisin BgAP from *Bacillus gibsonii* showed a melting point of 52.5 °C in the thermal shift assay comparable to SPAO, while several rounds of sequence saturation mutagenesis increased thermal stability to 58 °C with an additional reduction in the temperature optimum [32].

The effect of temperature on the enzyme activity was studied in a temperature range of 20–90 °C at pH 8.6 (standard suc-AAPF-pNA assay), as shown in Fig. 7. SPAO showed a temperature optimum of 55 °C, while BPN', Savinase, and subtilisin Carlsberg exhibited temperature optima of 55, 60, and 65 °C, respectively. The lower temperature optimum of SPAO correlates with temperature growth rates of the

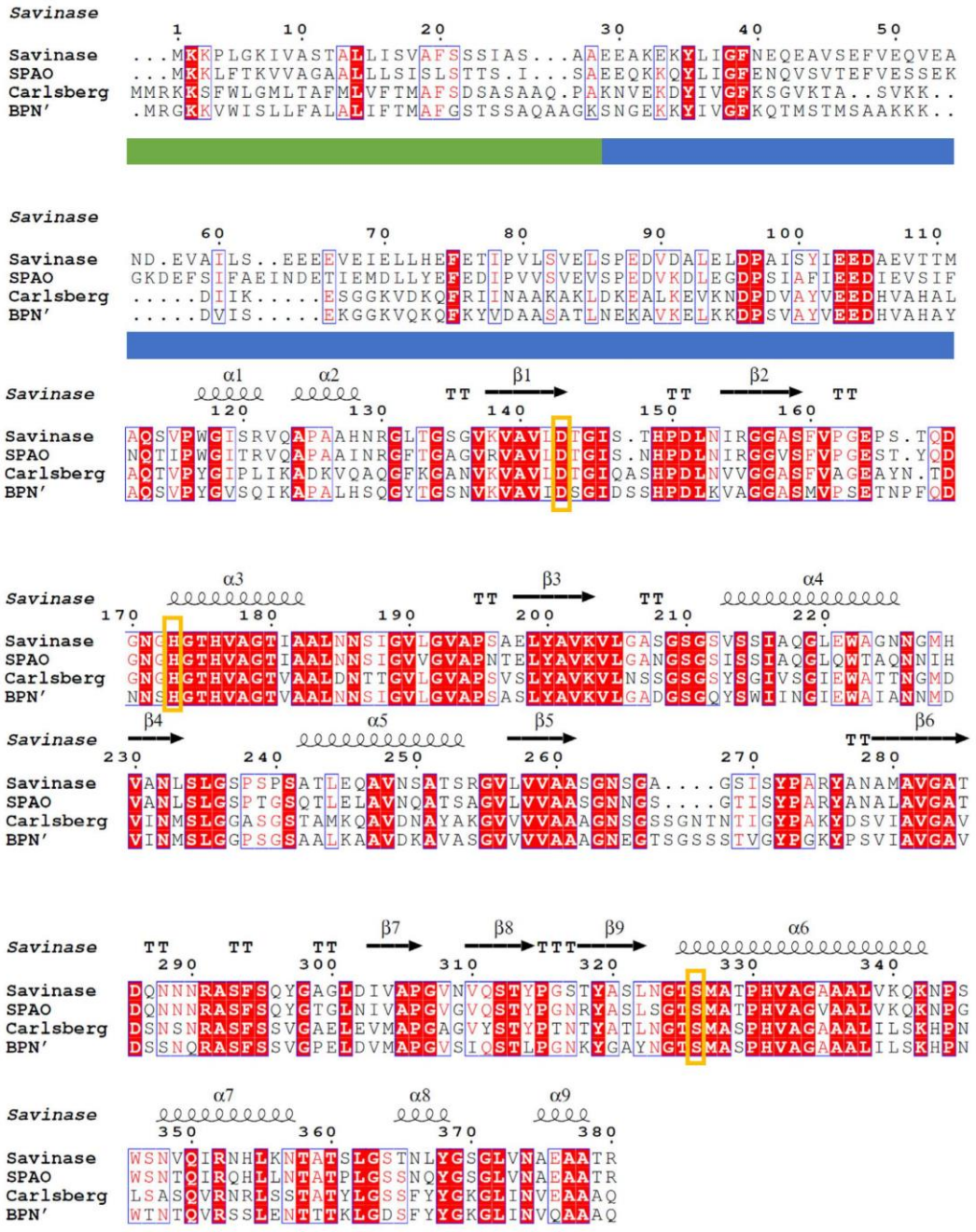


Fig. 2. Multiple sequence alignment (MSA) of SPAO with Savinase (WP_094423791.1), subtilisin Carlsberg (WP_020450819.1), and BPN' (WP_013351733.1). The alignment was calculated by CLUSTAL OMEGA and drawn using ESPript 3.0 and Savinase (PDB: 1C9J) as a template. Solid green and blue bars indicate the signal peptide sequence and propeptide of SPAO. Secondary structure elements are presented on top (helices with squiggles, β-strands with arrows, and turns with TT letters). Orange boxes show residues comprising the catalytic triad (Asp^{145}, His^{175}, Ser^{328}; SPAO numbering).

FEBS Open Bio **12** (2022) 1729–1746 © 2022 The Authors. *FEBS Open Bio* published by John Wiley & Sons Ltd on behalf of Federation of European Biochemical Societies.

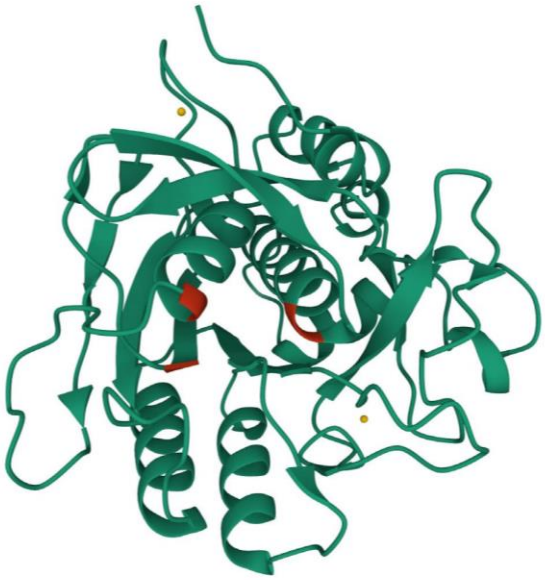


Fig. 3. Homology model of mature SPAO using I-TASSER software. *In silico* metal-binding analysis predicted the existence of two Ca^{2+} binding sites (yellow balls). The catalytic residues Asp^{32}, His^{62}, and Ser^{215} are shown in red.

bacterial origin (25–40 °C) [12]. For subtilisin Carlsberg, the higher temperature optimum coincides with the growth optimum of mesophilic *Bacillus licheniformis*. A serine protease isolated from the thermophilic *Geobacillus toebii* LBT 77 elevated an even higher optimum of 95 °C [46], while most other serine proteases from *Bacilli* have a temperature optimum between 50 and 70 °C [47-50].

Figure 8 shows the loss of protease activity during incubation for 4 h at 20 °C. SPAO showed a similar loss of activity as Savinase and subtilisin Carlsberg, with almost no activity remaining after 4 h, while BPN' retained about one-third of its activity after this time. As expected, the loss of activity is more pronounced for all proteases when incubated at 50 °C with residual activities of less than 25% (data not shown). Phrommao et al. [51] reported high stability at 20 °C but complete loss of activity after 2 h of incubation at 50 °C for the alkaline serine protease from *Virgibacillus sp.* SK37. The subtilisin from *Bacillus halodurans* C-125 demonstrated high stability with 94% remaining activity after 1 h of incubation at 50 °C [52]. A comparable loss of activity for subtilisin Carlsberg was also observed by others [53,54]. Higher thermal stability could be achieved by the addition of stabilizers such as polyols, which hindered the unfolding process and therefore reduced autoproteolysis [55,56].

The effects of pH on the activity of the enzymes toward the substrate suc-AAPF-pNA were studied in a pH range of 5.0–12.0 at 30 °C, as shown in Fig. 9. The highest activity of the protease SPAO was observed at pH 9.0–9.5, whereas at pH 6.5 and pH 12, only 7% and 60% of the maximal activity were measured, respectively. For SPAO, a stronger effect of the buffer system was observed, since a high activity rise after a buffer system change was observed. Savinase, BPN', and subtilisin Carlsberg also exhibited an optimal activity at pH 9.0. The acquired data for the reference proteases are congruent to the data found in literature [36,57]. The pH optimum of SPAO at pH 9.0 and its working range until pH 12.0 draw attention to its great potential toward various industrial applications, similar to other alkaline proteases reported before [51,52,58]. Thus, SPAO and Savinase are not only highly alkaline subtilisins based on their amino acid sequence, but also differ clearly from BPN' and subtilisin Carlsberg with increased activity at pH 12.0. High-alkaline proteases adapt to higher alkaline conditions by an altered surface charge at higher pHs, expressed by an increased pI value of the enzyme with a higher number of Arg and a decreased number of Lys [57]. This is also true for SPAO, especially when

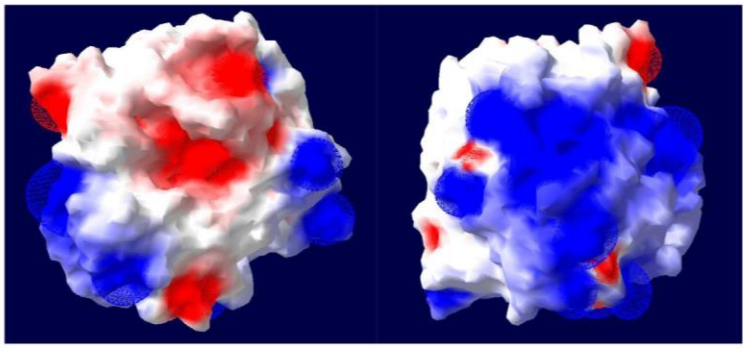


Fig. 4. Protein surface electrostatic potential calculations for the structural model of SPAO. (Left) front view of the active site; (right) back of the active site. Electrostatic potential at pH 7.0 is shown as red (negative) and blue (positive) and was calculated using the Swiss-PdbViewer.

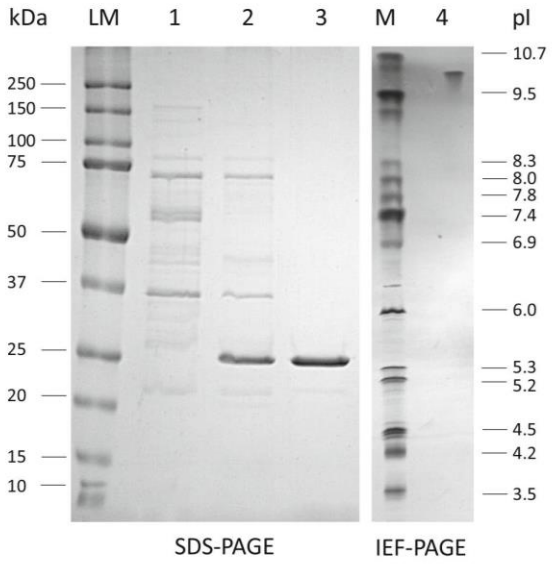


Fig. 5. SDS/PAGE and IEF-PAGE analysis of recombinant SPAO. Samples were electrophoresed using an 8–20% SDS/polyacrylamide gel. As protein standards, the LM mixture Bio-Rad Precision Plus Dual Color was used. Culture supernatant of *B. subtilis* DB104 carrying pFF-RED (lane 1); culture supernatant of *B. subtilis* DB104 carrying pFF003 producing SPAO (lane 2); and SPAO after purification (lane 3). Purified and desalted SPAO on an IEF gel SERVALYT™ PRECOTES™ wide range pH 3–10 (lane 4) and marker (M; SERVA IEF marker 3–10).

Table 1. Influence of H_2O_2, SDS, and PMSF on enzyme activity. The purified proteases were incubated with 1 and 5% (w/v) SDS; 1 and 5% (v/v) H_2O_2; and 1 mM PMSF at 10 °C in 10 mM HEPES/NaOH pH 7.0 for 1 h. Residual activity of the proteases incubated in buffer with no additions was considered as 100%. All experiments were performed at least in triplicates, and data are shown as mean values ± SD.

	Residual protease activity [%]				
Protease	1% SDS	5% SDS	1% H_2O_2	5% H_2O_2	1 mM PMSF
SPAO	0 ± 0	0 ± 0	108 ± 4	58 ± 3	0 ± 0
Subtilisin Carlsberg	325 ± 6	189 ± 6	71 ± 9	27 ± 2	0 ± 0
Savinase	138 ± 4	113 ± 6	64 ± 3	8 ± 0	0 ± 0
BPN′	205 ± 3	164 ± 7	81 ± 5	11 ± 0	0 ± 0

considering only the surface-exposed residues. Six of the nine Arg residues and one of the three Lys residues of SPAO are surface-exposed (Fig. 3, Table 2).

The protease SPAO showed good stability at pH 6–7 and pH 11–12 and lost more than 50% of its activity at pH 9.0, which is also the pH optimum of the protease and might be due to higher autoproteolysis, while BPN', Savinase, and subtilisin Carlsberg showed almost no remaining protease activity after 4 h (Fig. S2). Subtilisin Sendai was stable at pH 12, and it maintained 80% of its activity after 6 h at 30 °C [40]. The alkaline serine protease from *Geobacillus toebii* LBT 77 lost no activity between pH 8 and 13 after 12-h incubation at 60 °C [46]. Likewise, ALTP from *Alkaliphilus transvaalensis* or the serine protease from *Bacillus* sp. NPST-AK15 showed comparable stability between pH 5 and 11 after incubation for 10 min at 50 °C and 2 h at room temperature, respectively [47,59].

Effect of SDS and H_2O_2 on enzyme activity

The effect of SDS on SPAO and the reference proteases is reported in Table 1. SPAO is highly sensitive against 1% and 5% SDS (w/v) with no residual activity after 1 h at 10 °C. In contrast to that, the reference proteases showed high stability against SDS and were even more active than without. The anionic nature of SDS allows interactions between SDS and amino acid residues leading to the unfolding of the protein and loss of enzyme activity. For some subtilisins, SDS may not unfold the protein but instead helps to achieve a favorable protein conformation that stimulates activity, a behavior that was observed here for the reference proteases and has also been described in the literature. Bhatt and Singh [60] reported an activity of 275% for the alkaline serine protease from the newly isolated haloalkaliphilic *Bacillus lehensis* JO-26 after incubation with 1% SDS for 30 min compared to the activity without SDS. Joshi and Satyanarayana [61] reported for the *Bacillus lehensis* BLAP protease an activity of 99% after incubation with 1% SDS and an increased activity 160% after incubation with 2% SDS. The protease isolated from *Geobacillus toebii* LBT 77 displayed an activity of 120% after incubation with 1% SDS for 1 h at 55 °C [46]. An alkaline protease from *Bacillus licheniformis* RP1 lost 27% activity after incubation with 0.5% SDS [62]. The alkaline protease from *Bacillus clausii* I-52 displayed a remaining activity of 73% after incubation with 5% SDS [63]. Therefore, it is quite unusual for a high-alkaline subtilisin to lose complete activity after incubation with SDS. Another example is the intracellular subtilase AprX-SK37 from *Virgibacillus* sp. SK37, which showed a complete loss of activity when incubated with 0.5% SDS for 30 min at room temperature [51].

Subtilisin Carlsberg was investigated for its stability against 1% SDS for 1 h at 30 °C by Tanaka et al. [54]. In comparison with the activity increase measured in our work, they showed a complete activity loss after

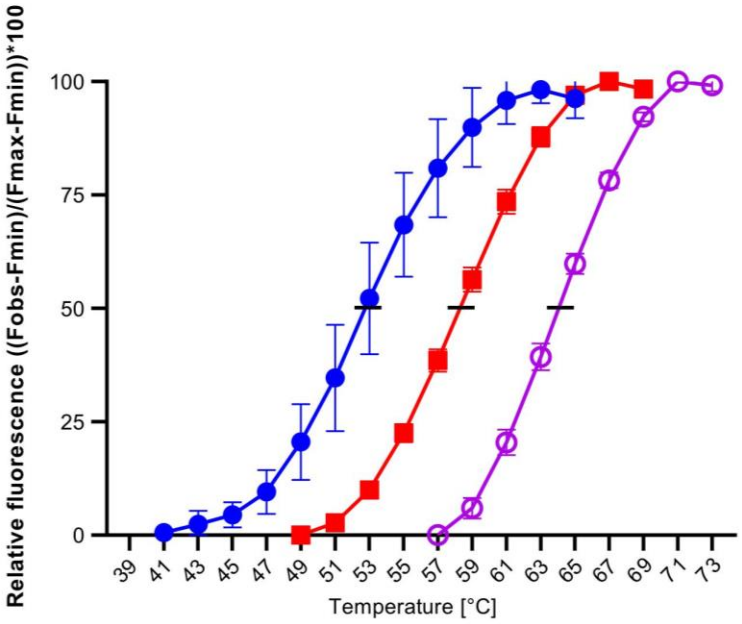


Fig. 6. Melting curves of purified SPAO in comparison with BPN' and subtilisin Carlsberg. The effect of temperature on the stability of the enzyme using SYPRO® Orange as a fluorescence probe, based on the changes in fluorescence emission intensity (Ex/Em = 470/550 nm; 5 × SYPRO® Orange, 10 mM HEPES/NaOH, pH 8.0, 3 mM PMSF), is shown as normalized denaturation curves of the thermal shift assay for the proteases SPAO (closed circles), BPN' (squares), and subtilisin Carlsberg (open circles). The inflection point corresponds to the melting temperature (T_m), at which 50% of the protein is unfolded (–). The experiment was performed in triplicates, and data are plotted as mean values ± SD.

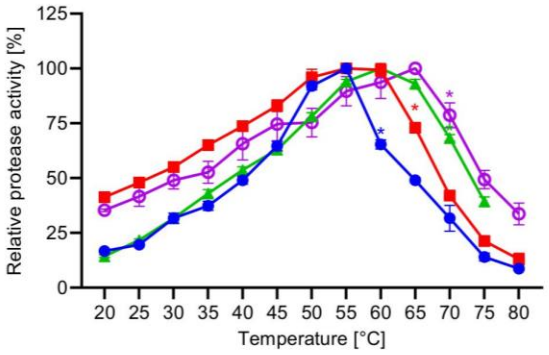


Fig. 7. Effect of temperature on the activity of purified SPAO, BPN', Savinase, and subtilisin Carlsberg. The activities of the proteases were determined by assaying protease activity at temperatures between 20 and 90 °C with suc-AAPF-pNA assay. The maximum activity for each protease was considered as 100% activity; SPAO (closed circles; 1076 U·mg^{-1}), BPN' (squares; 1007 U·mg^{-1}), Savinase (triangles; 2264 U·mg^{-1}), and subtilisin Carlsberg (open circles; 2767 U·mg^{-1}). *The enzyme was not stable for the intended 5 min. The experiment was performed in triplicates, and data are plotted as mean values ± SD.

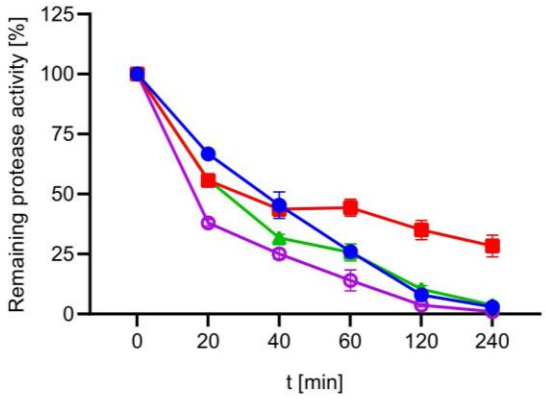


Fig. 8. Temperature stability of purified SPAO, BPN', Savinase, and subtilisin Carlsberg. Stability was investigated at 20 °C in 10 mM HEPES/NaOH buffer, pH 8.0. The activity was measured with the suc-AAPF-pNA assay in reaction buffer at 30 °C. The activity at 0 min was considered as 100% activity; SPAO (closed circles; 325 U·mg^{-1}), BPN' (squares; 343 U·mg^{-1}), Savinase (triangles; 367 U·mg^{-1}), and subtilisin Carlsberg (open circles; 635 U·mg^{-1}). The experiment was performed in triplicates, and data are plotted as mean values ± SD.

60 min. Looking at the temperature stability of subtilisin Carlsberg in Fig. 8, this result might be due to autoproteolysis rather than instability to SDS. Our results are consistent with a previous report that subtilisin Carlsberg was structurally unaffected upon interaction with SDS micelles [64].

The response of SPAO to H_2O_2 is also shown in Table 1. After a 1-h treatment with 1% (v/v) H_2O_2, subtilisin Carlsberg, Savinase, and BPN' displayed a loss of activity of 29%, 36%, and 19%, respectively. In contrast, SPAO showed a slight increase in

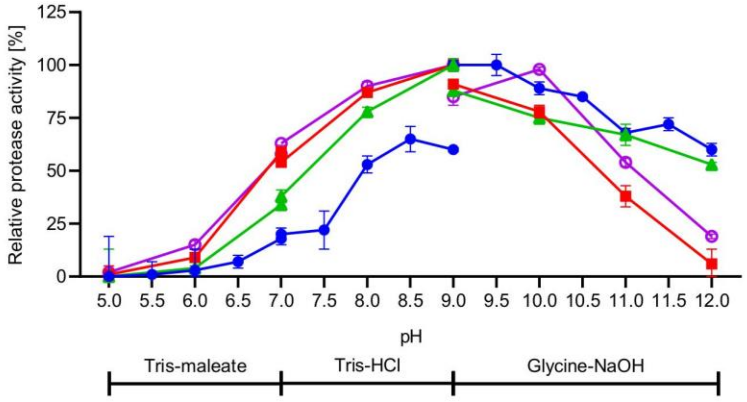


Fig. 9. Effect of pH on the activity of purified SPAO, BPN', Savinase, and subtilisin Carlsberg. The activity was measured with the suc-AAPF-pNA assay at 30 °C in the pH range of 5.0–12.0. The maximum activity for each protease was considered as 100% activity; SPAO (closed circles; 358 U·mg^{-1}), BPN' (squares; 604 U·mg^{-1}), Savinase (triangles; 705 U·mg^{-1}), and subtilisin Carlsberg (open circles; 1193 U·mg^{-1}). The experiments were performed in triplicates, and data are plotted as mean values ± SD.

Table 2. Physicochemical parameters of the proteases.

Protease	Experimental pI	Arg	Asp	Glu	His	Lys	AB ratio[a]
SPAO	9.80	9	4	4	6	3	0.4
Savinase	9.80	8	5	5	7	5	0.5
subtilisin Carlsberg	8.00	3	9	5	5	10	0.8
BPN'	7.80[b]	2	10	5	6	11	0.8

[a]The AB ratio [AB = (Glu + Asp)/(Lys + His + Arg)] was calculated as described in R85].; [b]pI of BPN' according to Matsubara et al. [38].

activity of 8% at 1% (v/v) H_2O_2. After a 1-h treatment with 5% (v/v) H_2O_2, the reference enzymes lost up to 92% activity (Savinase), while SPAO lost only 42% of its activity. SPAO is thus a highly oxidatively stable protease, especially in contrast to subtilisin Carlsberg, Savinase, and BPN'. H_2O_2 likely oxidizes a conserved methionine residue adjacent to the catalytic serine residue to its corresponding sulfoxide, which may lead to the inactivation of the enzyme [65]. The effect on the catalytic efficiency can be attributed to the unfavorable orientation of the sulfoxide oxygen of the oxidized methionine residue toward the oxyanion hole, thus destabilizing the tetrahedral intermediate formed with the carbonyl group of the peptide to be hydrolyzed [66,67]. Engineering the enzyme by replacing the methionine residue 222 (BPN' numbering) has been shown to increase the resistance to oxidants [68-71]. Nonaka et al. [72] showed that the digestion pattern of β-casein cleaved by oxidized proteases differs from that cleaved by unoxidized enzymes, suggesting that oxidation of the methionine is not a fatal modification but alters the substrate specificity. This altered substrate specificity could explain the slight increase in activity observed for SPAO, which possesses also Met^{216} next to the catalytic Ser^{215}.

In previous reports on the stability of alkaline proteases toward oxidants, a protease from *Bacillus clausii* I-52 was also found to have an increase in activity by 14% and 16% at 1% and 5% H_2O_2, respectively [63]. The protease from the alkaliphilic *Bacillus* sp. NPST-AK15 displayed a comparable increase of 2% at 1% H_2O_2 while losing 6% activity at 5% H_2O_2 [47]. Besides the two proteases mentioned above, the two serine proteases BM1 and BM2 derived from *Bacillus mojavensis* lost 62% and 60% activity by treatment with 5% H_2O_2 for 1 h at 30 °C, while a subtilase from *Thermoactinomyces vulgaris* strain CDF lost 90% activity after 1 h at 40 °C with 5% H_2O_2 [73,74]. Rekik et al. [75] reported for a protease from *Bacillus safensis* RH12 (SAPRH) an activity of 160% with surprisingly high H_2O_2 concentrations of 15% (v/v), which was also observed for the reference proteases they used: *Bacillus pumilus* (SAPB) with 109% activity and Alcalase 2.5 L, type DX with 150% activity. In contrast, the subtilase KP-43 from the group of oxidatively stable proteases (OSP) lost its ability to hydrolyze after only 30-min incubation with 3% H_2O_2 [72]. However, comparisons with literature data are often difficult because the buffers, temperatures, pH values, H_2O_2 concentration, and substrate for the activity studies are different.

Proteases can be classified by their sensitivity to various inhibitors [76]. Incubation of SPAO and the three reference proteases with 1 mM PMSF led to a complete inhibition of all proteases (Table 1).

Effects of NaCl, chelating agents (EDTA), and Ca^{2+} on enzyme activity and stability

Members of the subtilase superfamily are calcium-dependent, and the binding of Ca^{2+} is essential for

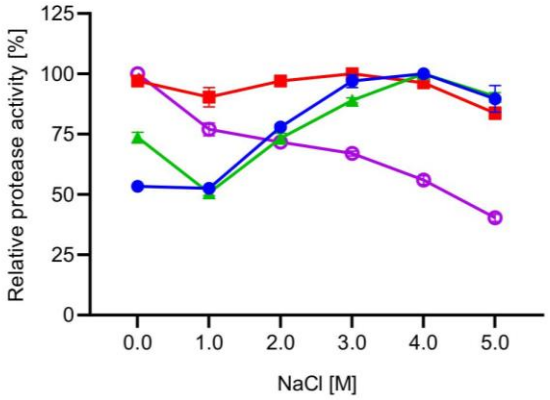


Fig. 10. Effect of NaCl on the activity of the purified SPAO, BPN', Savinase, and subtilisin Carlsberg. The activity was measured in standard buffer (pH 8.6) for suc-AAPF-pNA assay at 30 °C with different NaCl concentrations of 0–5 M. The maximum activity for each protease was considered as 100% activity; SPAO (closed circles; 1205 U·mg^{-1}), BPN' (squares; 560 U·mg^{-1}), Savinase (triangles; 757 U·mg^{-1}), and subtilisin Carlsberg (open circles; 846 U·mg^{-1}). The experiment was performed in triplicates, and data are given as mean values ± SD.

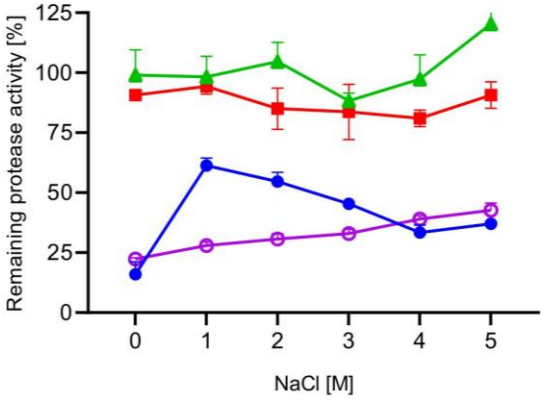


Fig. 11. Stability against NaCl of purified SPAO, BPN', Savinase, and subtilisin Carlsberg. Stability was tested in 10 mM HEPES/NaOH buffer, pH 8.0, with different NaCl concentrations (0–5 M) after 2 h at 20 °C. The activity was measured with the suc-AAPF-pNA assay in standard buffer at pH 8.6. Activity at 0 h for each NaCl concentration was considered as 100% activity; SPAO (circles), BPN' (squares), Savinase (triangles), and subtilisin Carlsberg (open circles). The experiment was performed in triplicates, and data are given as mean values ± SD.

enzyme stability and/or activity [77]. They usually contain two Ca^{2+} binding sites, the first being a strong binding site and the second a weak binding site [4].

The occupancy of these sites depends on the calcium ion concentration in the solution. The first site is always fully occupied even without the addition of $CaCl_2$ in the solution, while the second site is occupied by an Na^+ or K^+ ion at low $CaCl_2$ concentrations [77]. The *in silico* identification of the Ca^{2+} binding sites in SPAO revealed two Ca^{2+} binding sites as known for the three reference proteases [36]. Therefore, the effect of Ca^{2+} on the activity of the proteases was studied. All proteases, except subtilisin Carlsberg, lost more activity by incubating with EDTA before the supplementation with $CaCl_2$ than without EDTA (data not shown). However, for BPN', there is almost no difference in activity between incubation with or without EDTA. For SPAO and Savinase, it was not possible to fully recover activity after the addition of $CaCl_2$. In general, Ca^{2+} seems to be firmly bound, as the proteases were dissolved in a Ca^{2+}-free buffer and it was not necessary to add Ca^{2+} for their activity. At a concentration of 10 mM $CaCl_2$ and above, the activity of all proteases decreased. Similar observations of the inhibitory effect of Ca^{2+} were also made for the alkaline protease from the haloalkaliphilic bacterium sp. AH-6 [78]. The failure to recover activity could be due to a general loss of activity due to higher protein instability during incubation with EDTA. The binding of Ca^{2+} has a stabilizing effect on the protease by reducing molecular flexibility, which reduces thermal denaturation and autolysis [77]. No effect on the activity of a serine protease from *B. clausii* GMBAE 42 after incubation with EDTA and a slight inhibition by Ca^{2+} was also found by Kazan et al. [49]. Others could show a strong activity decrease after the incubation with EDTA [58,79,80]. However, most studies found a comparable decline in activity as seen in this work [46,81,82]. The findings suggest a Ca^{2+} dependency for SPAO and a good stability.

The effect of NaCl on the activity of the proteases was evaluated in the activity assay using the substrate suc-AAPF-pNA and different concentrations of NaCl (0–5 M) in standard reaction buffer (pH 8.6). The proteolytic activity is shown in Fig. 10. SPAO showed increased activity with rising NaCl concentrations up to the maximum at 4 M NaCl, followed by a 10% decrease at 5 M NaCl. Savinase showed the highest activity at 3 and 4 M NaCl. BPN' displayed high activity from 0 to 4 M NaCl, which was reduced to 77% at 5 M NaCl. The stability of SPAO and the three reference proteases was examined by incubation with NaCl concentration between 0 and 5 M NaCl in 10 mM HEPES/NaOH, pH 8.0, at 20 °C for 2 h (Fig. 11). The results show that Savinase and BPN' were stable with and without NaCl. SPAO and subtilisin Carlsberg were stabilized by NaCl. However, SPAO showed stability above 50% only with 1 M

Table 3. Substrate specificities of the proteases against 10 synthetic substrates (suc-XXXX-pNA). Kinetic experiments were carried out in 0.1 M Tris/HCl buffer, pH 8.6, and 0.1% (w/v) Brij®35 over 5 min at 30 °C with 17 mM of a substrate. The experiment was performed in triplicates, and the standard deviation was < 5%. The enzyme activity against AAPF refers to 100% relative activity.

	Relative activity [%]									
Protease	FAAF	AAA	AAVA	ALPF	AGPF	AAPF	TVAA	YVAD	AGPP	AAPL
SPAO	753	12	21	127	107	100 (42 U·mg^{-1})	3	3	216	34
Subtilisin Carlsberg	57	0	2	60	90	100 (570 U·mg^{-1})	1	1	147	104
Savinase	605	8	22	117	96	100 (180 U·mg^{-1})	5	5	144	12
BPN'	96	0	6	106	96	100 (181 U·mg^{-1})	0	0	61	67

NaCl while losing stability with increasing NaCl concentrations.

In contrast to halophilic enzymes, halotolerant enzymes are active over a broad range of NaCl concentrations and retain their activity even in the absence of NaCl [83]. Unlike the mechanism of halophilic enzymes, the mechanism of halotolerance is not yet fully elucidated [79]. Halophilic proteins have a predominance of negatively charged residues on the solvent-exposed surface, which attract water molecules and thereby keep the enzymes hydrated [84]. Therefore, a forecast to salt adaptation could be made based on the ratio of the acidic amino acids Glu and Asp to the basic amino acids Lys, His, and Arg (AB ratio) and a more acidic isoelectric point [84,85]. While SPAO and Savinase share a higher pI, the AB ratio is lower than for BPN' and subtilisin Carlsberg, which indicates a poor adaptation to salt (Table 2). However, the high Arg content of SPAO and Savinase, which favors adaptation to highly alkaline conditions, results in a predominantly positive charge on the protein surface in SPAO (Fig. 4), and in general, a high proportion of negative or positive charges on the surface of the enzyme improves salt adaptation [86]. BPN' has five of the seven amino acid positions identified by Takenaka et al. [86] as favorable for salt adaptation and therefore, unlike subtilisin Carlsberg, shows good adaptation to high salt concentrations.

Proteolytic activity on synthetic peptides

In general, data for proteinase specificity under comparable experimental conditions are limited in the literature. Usually, kinetic data for the hydrolysis of synthetic substrates are collected. To analyze the substrate specificity of SPAO and the three reference proteases, 10 synthetic three or four amino acid peptide-4-nitroanilide substrates were chosen, which are typical subtilisin substrates [87]. However, suc-AAA-pNA is a typical elastase substrate [88]. As shown in Table 3, SPAO showed the highest specific activity for suc-FAAF-pNA and a very low specificity for suc-TVAA-pNA and suc-YVAD-pNA under the selected conditions, which is in good agreement with the results obtained for the typical subtilisins Savinase, subtilisin Carlsberg, and BPN'.

In general, subtilisins show a broad substrate specificity and often display a preference for large hydrophobic groups at the P1 position, the first position N-terminal to the cleavage site (nomenclature of Schechter and Berger [89]), here to the 4-nitroanilide (P1'), which can also be observed in this experiment [90]. However, the protease SPAO and Savinase are able to hydrolyze the substrate if alanine is at the P1 and P2 position, but with higher efficiency, if one of the positions is alanine and an amino acid with a larger hydrophobic group. In literature, data for the hydrolysis of synthetic substrates are collected under different conditions, which makes the results difficult for comparison and interpretation. Kazan et al. [49] reported an alkaline serine protease from *B. clausii* GMBAE 42 with high specificity for suc-AAPF-pNA. Georgieva et al. [87] reported for Savinase also the highest specificity for suc-FAAF-pNA by comparing it with Esperase with a similar preference. Proteinase K showed the highest specificity for suc-AGPF-pNA [91]. For the elastase-specific substrate suc-AAA-pNA, SPAO showed low activity in contrast to the alkaline elastase YaB from *Bacillus* strain YaB [92]. In the context of this experiment and also in line with the high sequence similarity to Savinase (82.16%), SPAO can be considered a typical subtilisin.

Conclusion

In this study, we recombinantly expressed SPAO, a novel high-alkaline subtilisin isolated from *Alkalihalobacillus okhensis* Kh10-101^T. SPAO was effectively produced and secreted by *B. subtilis* DB104. After purification, biochemical characterization revealed a highly oxidatively stable protease with increased activity upon incubation with 1% H_2O_2 and residual

FEBS Open Bio **12** (2022) 1729–1746 © 2022 The Authors. *FEBS Open Bio* published by John Wiley & Sons Ltd on behalf of Federation of European Biochemical Societies.

activity of 58% with 5% H_2O_2. The high Arg content and the predominantly positive charge on the protein surface allow SPAO to be highly active at pH 12.0 and at NaCl concentrations of up to 5 M. The optimal temperature and pH were 55 °C and pH 9.0–10.5, respectively. With its biochemical properties, SPAO shows potential for industrial applications to be evaluated in the future.

Acknowledgements

This work was supported by the FH Aachen University of Applied Sciences within the framework of the program for the promotion of young scientists. Open Access funding enabled and organized by Projekt DEAL. WOA Institution: N/A Consortia Name : Projekt DEAL

Conflict of interest

The authors declare no conflict of interest.

Author contributions

FF, JB, and PS conceived and designed the experiments. FF collected and analyzed the data. DF and JR carried out the cloning and pre-experiments. FF wrote the original draft. FF, JB, MB, and PS revised the manuscript. All authors contributed to the final manuscript.

Data availability statement

The data that support the findings of this study are available from the corresponding author [siegert@fh-aachen.de] upon reasonable request.

References

1 Gupta R, Beg QK, Lorenz P. Bacterial alkaline proteases: molecular approaches and industrial applications. *Appl Microbiol Biotechnol*. 2002;**59**: 15–32.
2 Naveed M, Nadeem F, Mehmood T, Bilal M, Anwar Z, Amjad F. Protease—a versatile and ecofriendly biocatalyst with multi-industrial applications: an updated review. *Catal Lett*. 2021;**151**:307–23.
3 Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A, Finn RD. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res*. 2018;**46**: D624–32.
4 Siezen RJ, Leunissen JA. Subtilases: the superfamily of subtilisin-like serine proteases. *Protein Sci*. 1997;**6**: 501–23.
5 Saeki K, Magallones MV, Takimura Y, Hatada Y, Kobayashi T, Kawai S, et al. Nucleotide and deduced amino acid sequences of a new subtilisin from an alkaliphilic *Bacillus* isolate. *Curr Microbiol*. 2003;**47**:337–40.
6 Kalisz HM. Microbial proteinases. *Adv Biochem Eng Biotechnol*. 1988;**36**:1–65.
7 Contesini FJ, Melo RR, Sato HH. An overview of *Bacillus* proteases: from production to application. *Crit Rev Biotechnol*. 2018;**38**:321–34.
8 Solanki P, Putatunda C, Kumar A, Bhatia R, Walia A. Microbial proteases: ubiquitous enzymes with innumerable uses. *3 Biotech*. 2021;**11**:428.
9 Salwan R, Sharma V. Trends in extracellular serine proteases of bacteria as detergent bioadditive: alternate and environmental friendly tool for detergent industry. *Arch Microbiol*. 2019;**201**:863–77.
10 Alberto Cira-Chávez L, Guevara-Luna J, Yadira Soto-Padilla M, Román-Ponce B, Soledad Vásquez-Murrieta M, Isabel Estrada-Alvarado M. Kinetics of halophilic enzymes. In: Rajendran L, Fernandez C, editors. *Kinetics of enzymatic synthesis*. London: IntechOpen; 2018. p. 3–26.
11 Coker JA. Extremophiles and biotechnology: current uses and prospects. *F1000Res*. 2016;**5**:396.
12 Nowlan B, Dodia MS, Singh SP, Patel BKC. *Bacillus okhensis* sp. nov., a halotolerant and alkalitolerant bacterium from an Indian saltpan. *Int J Syst Evol Microbiol*. 2006;**56**:1073–7.
13 Patel S, Gupta RS. A phylogenomic and comparative genomic framework for resolving the polyphyly of the genus *Bacillus*: proposal for six new genera of *Bacillus* species, *Peribacillus* gen. Nov., *Cytobacillus* gen. Nov., *Mesobacillus* gen. Nov., *Neobacillus* gen. Nov., *Metabacillus* gen. Nov. and *Alkalihalobacillus* gen. Nov. *Int J Syst Evol Microbiol*. 2020;**70**:406–38.
14 Krishna PS, Raghunathan S, Prakash JSS. Comparative genome analysis of *Alkalihalobacillus okhensis* Kh10-101 T reveals insights into adaptive mechanisms for halo-alkali tolerance. *3 Biotech*. 2021;**11**:392.
15 Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, et al. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res*. 2008;**36**:W465–9.
16 Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res*. 2014;**42**:W320–4.
17 Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal omega. *Mol Syst Biol*. 2011;**7**:539.

18 Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res*. 2016;**44**:W242–5.
19 Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER suite: protein structure and function prediction. *Nat Methods*. 2015;**12**:7–8.
20 Sehnal D, Bittrich S, Deshpande M, Svobodová R, Berka K, Bazgier V, et al. Mol* viewer: modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Res*. 2021;**49**: W431–7.
21 Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*. 1997;**18**:2714–23.
22 Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat Biotechnol*. 2022.
23 Kawamura F, Doi RH. Construction of a *Bacillus subtilis* double mutant deficient in extracellular alkaline and neutral proteases. *J Bacteriol*. 1984;**160**:442–4.
24 Bernhard K, Schrempf H, Goebel W. Bacteriocin and antibiotic resistance plasmids in *Bacillus cereus* and *Bacillus subtilis*. *J Bacteriol*. 1978;**133**:897–903.
25 Engler C, Kandzia R, Marillonnet S. A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE*. 2008;**3**:e3647.
26 Vojcic L, Despotovic D, Martinez R, Maurer K-H, Schwaneberg U. An efficient transformation method for *Bacillus subtilis* DB104. *Appl Microbiol Biotechnol*. 2012;**94**:487–93.
27 DelMar EG, Largman C, Brodrick JW, Geokas MC. A sensitive new substrate for chymotrypsin. *Anal Biochem*. 1979;**99**:316–20.
28 Brock FM, Forsberg CW, Buchanan-Smith JG. Proteolytic activity of rumen microorganisms and effects of proteinase inhibitors. *Appl Environ Microbiol*. 1982;**44**:561–9.
29 Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem*. 1976;**72**:248–54.
30 Strohalm M, Hassman M, Kosata B, Kodícek M. mMass data miner: an open source alternative for mass spectrometric data analysis. *Rapid Commun Mass Spectrom*. 2008;**22**:905–8.
31 Miller AJ, Roman B, Norstrom E. A method for easily customizable gradient gel electrophoresis. *Anal Biochem*. 2016;**509**:12–4.
32 Martinez R, Jakob F, Tu R, Siegert P, Maurer K-H, Schwaneberg U. Increasing activity and thermal resistance of *Bacillus gibsonii* alkaline protease (BgAP) by directed evolution. *Biotechnol Bioeng*. 2013;**110**: 711–20.
33 Siezen RJ, Renckens B, Boekhorst J. Evolution of prokaryotic subtilases: genome-wide analysis reveals novel subfamilies with different catalytic residues. *Proteins*. 2007;**67**:681–94.
34 The UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res*. 2021;**49**:D480–9.
35 Yamagata Y, Isshiki K, Ichishima E. Subtilisin Sendai from alkalophilic *Bacillus* sp.: molecular and enzymatic properties of the enzyme and molecular cloning and characterization of the gene, aprS. *Enzyme Microb Technol*. 1995;**17**:653–63.
36 Betzel C, Klupsch S, Papendorf G, Hastrup S, Branner S, Wilson KS. Crystal structure of the alkaline proteinase Savinase from *Bacillus lentus* at 1.4 a resolution. *J Mol Biol*. 1992;**223**:427–45.
37 Teplyakov AV, van der Laan JM, Lammers AA, Kelders H, Kalk KH, Misset O, et al. Protein engineering of the high-alkaline serine protease PB92 from *Bacillus alcalophilus*: functional and structural consequences of mutation at the S4 substrate binding pocket. *Protein Eng*. 1992;**5**:413–20.
38 Matsubara H, Kasper CB, Brown DM, Smith EL. Subtilisin BPN'. *J Biol Chem*. 1965;**240**:1125–30.
39 Smith EL, Markland FS, Kasper CB, DeLange RJ, Landon M, Evans WH. The complete amino acid sequence of two types of subtilisin, BPN' and Carlsberg. *J Biol Chem*. 1966;**241**:5974–6.
40 Yamagata Y, Maeda H, Nakajima T, Ichishima E. The molecular surface of proteolytic enzymes has an important role in stability of the enzymatic activity in extraordinary environments. *Eur J Biochem*. 2002;**269**:4577–85.
41 Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*. 2008;**9**:40.
42 Goddette DW, Paech C, Yang SS, Mielenz JR, Bystroff C, Wilke ME, et al. The crystal structure of the *Bacillus lentus* alkaline protease, subtilisin BL, at 1.4 a resolution. *J Mol Biol*. 1992;**228**: 580–95.
43 Li Z, Roccatano D, Lorenz M, Martinez R, Schwaneberg U. Insights on activity and stability of subtilisin E towards guanidinium chloride and sodium dodecyl sulfate. *J Biotechnol*. 2014;**169**:87–94.
44 Maeda H, Mizutani O, Yamagata Y, Ichishima E, Nakajima T. Alkaline-resistance model of subtilisin ALP I, a novel alkaline subtilisin. *J Biochem*. 2001;**129**:675–82.
45 Maurer K-H. Detergent proteases. *Curr Opin Biotechnol*. 2004;**15**:330–4.
46 Thebti W, Riahi Y, Belhadj O. Purification and characterization of a new thermostable, Haloalkaline, solvent stable, and detergent compatible serine protease from *Geobacillus toebii* strain LBT 77. *Biomed Res Int*. 2016;**2016**:9178962.

47 Ibrahim ASS, Al-Salamah AA, El-Badawi YB, El-Tayeb MA, Antranikian G. Detergent-, solvent- and salt-compatible thermoactive alkaline serine protease from halotolerant alkaliphilic *Bacillus sp*. NPST-AK15: purification and characterization. *Extremophiles*. 2015;**19**:961–71.
48 Toyokawa Y, Takahara H, Reungsang A, Fukuta M, Hachimine Y, Tachibana S, et al. Purification and characterization of a halotolerant serine proteinase from thermotolerant *Bacillus licheniformis* RKK-04 isolated from Thai fish sauce. *Appl Microbiol Biotechnol*. 2010;**86**:1867–75.
49 Kazan D, Denizci AA, Oner MNK, Erarslan A. Purification and characterization of a serine alkaline protease from *Bacillus clausii* GMBAE 42. *J Ind Microbiol Biotechnol*. 2005;**32**:335–44.
50 Saeki K, Hitomi J, Okuda M, Hatada Y, Kageyama Y, Takaiwa M, et al. A novel species of alkaliphilic *Bacillus* that produces an oxidatively stable alkaline serine protease. *Extremophiles*. 2002;**6**:65–72.
51 Phrommao E, Yongsawatdigul J, Rodtong S, Yamabhai M. A novel subtilase with NaCl-activated and oxidant-stable activity from *Virgibacillus* sp. SK37. *BMC Biotechnol*. 2011;**11**:65.
52 Tekin A, Uzuner U, Sezen K. Homology modeling and heterologous expression of highly alkaline subtilisin-like serine protease from *Bacillus halodurans* C-125. *Biotechnol Lett*. 2020;**43**:479–94.
53 Ferreira L, Ramos M, Dordick J, Gil M. Influence of different silica derivatives in the immobilization and stabilization of a *Bacillus licheniformis* protease (subtilisin Carlsberg). *J Mol Catal B Enzym*. 2003;**21**:189–99.
54 Tanaka T, Matsuzawa H, Ohta T. Stability of thermostable enzyme, Aqualysin I; a subtilisin-type serine protease from Thermus aquaticus YT-1. *Biosci Biotechnol Biochem*. 1998;**62**:1806–8.
55 Back JF, Oakenfull D, Smith MB. Increased thermal stability of proteins in the presence of sugars and polyols. *Biochemistry*. 1979;**18**:5191–6.
56 Joo H-S, Koo Y-M, Choi J-W, Chang C-S. Stabilization method of an alkaline protease from inactivation by heat, SDS and hydrogen peroxide. *Enzyme Microb Technol*. 2005;**36**:766–72.
57 Masui A, Fujiwara N, Yamamoto K, Takagi M, Imanaka T. Rational design for stabilization and optimum pH shift of serine protease AprN. *J Ferment Bioeng*. 1998;**85**:30–6.
58 Gurunathan R, Huang B, Ponnusamy VK, Hwang J-S, Dahms H-U. Novel recombinant keratin degrading subtilisin like serine alkaline protease from *Bacillus cereus* isolated from marine hydrothermal vent crabs. *Sci Rep*. 2021;**11**:12007.
59 Kobayashi T, Lu J, Li Z, Hung VS, Kurata A, Hatada Y, et al. Extremely high alkaline protease from a deep-subsurface bacterium, *Alkaliphilus transvaalensis*. *Appl Microbiol Biotechnol*. 2007;**75**:71–80.
60 Bhatt HB, Singh SP. Cloning, expression, and structural elucidation of a biotechnologically potential alkaline serine protease from a newly isolated haloalkaliphilic *Bacillus lehensis* JO-26. *Front Microbiol*. 2020;**11**:941.
61 Joshi S, Satyanarayana T. Characteristics and applications of a recombinant alkaline serine protease from a novel bacterium *Bacillus lehensis*. *Bioresour Technol*. 2013;**131**:76–85.
62 Sellami-Kamoun A, Haddar A, Ali NE-H, Ghorbel-Frikha B, Kanoun S, Nasri M. Stability of thermostable alkaline protease from *Bacillus licheniformis* RP1 in commercial solid laundry detergent formulations. *Microbiol Res*. 2008;**163**:299–306.
63 Joo H-S, Kumar CG, Park G-C, Paik SR, Chang C-S. Oxidant and SDS-stable alkaline protease from *Bacillus clausii* I-52: production and some properties. *J Appl Microbiol*. 2003;**95**:267–72.
64 Shaw AK, Pal SK. Activity of subtilisin Carlsberg in macromolecular crowding. *J Photochem Photobiol B*. 2007;**86**:199–206.
65 Stauffer CE, Etson D. The effect on subtilisin activity of oxidizing a methionine residue. *J Biol Chem*. 1969;**244**:5333–8.
66 Bott R, Ultsch M, Kossiakoff A, Graycar T, Katz B, Power S. The three-dimensional structure of *Bacillus amyloliquefaciens* subtilisin at 1.8 a and an analysis of the structural consequences of peroxide inactivation. *J Biol Chem*. 1988;**263**:7895–906.
67 Bryan P, Pantoliano MW, Quill SG, Hsiao HY, Poulos T. Site-directed mutagenesis and the role of the oxyanion hole in subtilisin. *Proc Natl Acad Sci USA*. 1986;**83**:3743–5.
68 Estell DA, Graycar TP, Wells JA. Engineering an enzyme by site-directed mutagenesis to be resistant to chemical oxidation. *J Biol Chem*. 1985;**260**:6518–21.
69 Grøn H, Bech LM, Branner S, Breddam K. A highly active and oxidation-resistant subtilisin-like enzyme produced by a combination of site-directed mutagenesis and chemical modification. *Eur J Biochem*. 1990;**194**:897–901.
70 Yang Y, Jiang L, Zhu L, Wu Y, Yang S. Thermal stable and oxidation-resistant variant of subtilisin E. *J Biotechnol*. 2000;**81**:113–8.
71 Vojcic L, Despotovic D, Maurer K-H, Zacharias M, Bocola M, Martinez R, et al. Reengineering of subtilisin Carlsberg for oxidative resistance. *Biol Chem*. 2013;**394**:79–87.
72 Nonaka T, Fujihashi M, Kita A, Saeki K, Ito S, Horikoshi K, et al. The crystal structure of an oxidatively stable subtilisin-like alkaline serine protease, KP-43, with a C-terminal beta-barrel domain. *J Biol Chem*. 2004;**279**:47344–51.

73 Haddar A, Agrebi R, Bougatef A, Hmidet N, Sellami-Kamoun A, Nasri M. Two detergent stable alkaline serine-proteases from *Bacillus mojavensis* A21: purification, characterization and potential application as a laundry detergent additive. *Bioresour Technol*. 2009;**100**:3366–73.
74 Ding Y, Yang Y, Ren Y, Xia J, Liu F, Li Y, et al. Extracellular production, characterization, and engineering of a polyextremotolerant subtilisin-like protease from feather-degrading *Thermoactinomyces vulgaris* strain CDF. *Front Microbiol*. 2020;**11**:605771.
75 Rekik H, Zaraî Jaouadi N, Gargouri F, Bejar W, Frikha F, Jmal N, et al. Production, purification and biochemical characterization of a novel detergent-stable serine alkaline protease from *Bacillus safensis* strain RH12. *Int J Biol Macromol*. 2019;**121**:1227–39.
76 North MJ. Comparative biochemistry of the proteinases of eucaryotic microorganisms. *Microbiol Rev*. 1982;**46**:308–40.
77 Siezen RJ, de Vos WM, Leunissen JA, Dijkstra BW. Homology modelling and protein engineering strategy of subtilases, the family of subtilisin-like serine proteinases. *Protein Eng*. 1991;**4**:719–37.
78 Dodia MS, Rawal CM, Bhimani HG, Joshi RH, Khare SK, Singh SP. Purification and stability characteristics of an alkaline serine protease from a newly isolated haloalkaliphilic bacterium sp. AH-6. *J Ind Microbiol Biotechnol*. 2008;**35**:121–31.
79 Takenaka S, Yoshinami J, Kuntiya A, Techapun C, Leksawasdi N, Seesuriyachan P, et al. Characterization and mutation analysis of a halotolerant serine protease from a new isolate of *Bacillus subtilis*. *Biotechnol Lett*. 2018;**40**:189–96.
80 Farooq S, Nazir R, Ganai SA, Ganai BA. Isolation and characterization of a new cold-active protease from psychrotrophic bacteria of Western Himalayan glacial soil. *Sci Rep*. 2021;**11**:12768.
81 Vidyasagar M, Prakash S, Mahajan V, Shouche YS, Sreeramulu K. Purification and characterization of an extreme halothermophilic protease from a halophilic bacterium *Chromohalobacter sp*. TVSP101. *Braz J Microbiol*. 2009;**40**:12–9.
82 Raval VH, Pillai S, Rawal CM, Singh SP. Biochemical and structural characterization of a detergent-stable serine alkaline protease from seawater haloalkaliphilic bacteria. *Process Biochem*. 2014;**49**:955–62.
83 Setyorini E, Takenaka S, Murakami S, Aoki K. Purification and characterization of two novel halotolerant extracellular proteases from *Bacillus subtilis* strain FP-133. *Biosci Biotechnol Biochem*. 2006;**70**:433–40.
84 Mokashe N, Chaudhari B, Patil U. Operative utility of salt-stable proteases of halophilic and halotolerant bacteria in the biotechnology sector. *Int J Biol Macromol*. 2018;**117**:493–522.
85 Rhodes ME, Fitz-Gibbon ST, Oren A, House CH. Amino acid signatures of salinity on an environmental scale with a focus on the Dead Sea. *Environ Microbiol*. 2010;**12**:2613–23.
86 Takenaka S, Takada A, Kimura Y, Watanabe M, Kuntiya A. Improvement of the halotolerance of a *Bacillus* serine protease by protein surface engineering. *J Basic Microbiol*. 2022;**62**:174–84.
87 Georgieva DN, Stoeva S, Voelter W, Genov N, Betzel C. Differences in the specificities of the highly alkalophilic proteinases Savinase and Esperase imposed by changes in the rigidity and geometry of the substrate binding sites. *Arch Biochem Biophys*. 2001;**387**:197–201.
88 Bieth J, Spiess B, Wermuth CG. The synthesis and analytical use of a highly sensitive and convenient substrate of elastase. *Biochem Med*. 1974;**11**:350–7.
89 Schechter I, Berger A. On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun*. 1967;**27**:157–62.
90 Perona JJ, Craik CS. Structural basis of substrate specificity in the serine proteases. *Protein Sci*. 1995;**4**:337–60.
91 Georgieva D, Genov N, Voelter W, Betzel C. Catalytic efficiencies of alkaline proteinases from microorganisms. *Z Naturforsch C J Biosci*. 2006;**61**:445–52.
92 Tsai YC, Yamasaki M, Tamura G. Substrate specificity of a new alkaline elastase from an alkalophilic *Bacillus*. *Biochem Int*. 1984;**8**:283–8.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.
Table S1. Preculture medium.
Table S2. Fermentation medium.
Table S3. Trace element solution.
Fig. S1. MALDI-TOF mass spectra of SPAO.
Fig. S2. Effect of pH on the stability of purified SPAO, BPN', Savinase, and subtilisin Carlsberg.

FEBS Open Bio **12** (2022) 1729–1746 © 2022 The Authors. *FEBS Open Bio* published by John Wiley & Sons Ltd on behalf of Federation of European Biochemical Societies.

2.3 New robust subtilisins from halotolerant and halophilic *Bacillaceae*

Fabian Falkenberg, Leonie Voß, Michael Bott, Johannes Bongaerts**,** Petra Siegert (2023) Applied Microbiology and Biotechnology, DOI: 10.1007/s00253-023-12553-w

Authors contributions:
FF, JB, and PS conceived and designed the experiments.
FF conducted the experiments except those of LV, collected and analysed the data.
LV cloned the gene for SPMI and did pre-experiments for SPMI.
FF wrote the original draft.
FF, JB, MB, and PS revised the manuscript.

All figures and tables are from this study.

Overall contribution FF: 95 %

Applied Microbiology and Biotechnology
https://doi.org/10.1007/s00253-023-12553-w

BIOTECHNOLOGICALLY RELEVANT ENZYMES AND PROTEINS

New robust subtilisins from halotolerant and halophilic *Bacillaceae*

Fabian Falkenberg[1] · Leonie Voß[1] · Michael Bott[2] · Johannes Bongaerts[1] · Petra Siegert[1]

Received: 21 March 2023 / Revised: 17 April 2023 / Accepted: 19 April 2023

Abstract
The aim of the present study was the characterisation of three true subtilisins and one phylogenetically intermediate subtilisin from halotolerant and halophilic microorganisms. Considering the currently growing enzyme market for efficient and novel biocatalysts, data mining is a promising source for novel, as yet uncharacterised enzymes, especially from halophilic or halotolerant *Bacillaceae*, which offer great potential to meet industrial needs. Both halophilic bacteria *Pontibacillus marinus* DSM 16465^T and *Alkalibacillus haloalkaliphilus* DSM 5271^T and both halotolerant bacteria *Metabacillus indicus* DSM 16189 and *Litchfieldia alkalitelluris* DSM 16976^T served as a source for the four new subtilisins SPPM, SPAH, SPMI and SPLA. The protease genes were cloned and expressed in *Bacillus subtilis* DB104. Purification to apparent homogeneity was achieved by ethanol precipitation, desalting and ion-exchange chromatography. Enzyme activity could be observed between pH 5.0–12.0 with an optimum for SPPM, SPMI and SPLA around pH 9.0 and for SPAH at pH 10.0. The optimal temperature for SPMI and SPLA was 70 °C and for SPPM and SPAH 55 °C and 50 °C, respectively. All proteases showed high stability towards 5% (w/v) SDS and were active even at NaCl concentrations of 5 M. The four proteases demonstrate potential for future biotechnological applications.

Key points
- *Halophilic and halotolerant Bacillaceae are a valuable source of new subtilisins.*
- *Four new subtilisins were biochemically characterised in detail.*
- *The four proteases show potential for future biotechnological applications.*

Keywords Halotolerant protease · Subtilases · Subtilisin · *Bacillaceae* · Biotechnological application

Introduction

Microorganisms that can survive in environments with extreme temperatures, pH and salinity produce enzymes called extremozymes (Ferrer et al. 2007). These microorganisms and their enzymes developed molecular mechanisms of adaptation to extreme physico-chemical conditions (Tehei and Zaccai 2005). Of particular interest are proteases, which are one of the most important enzymes used commercially, with subtilisins or alkaline proteases from microbial sources having the largest market share (Gupta et al. 2002; Naveed et al. 2021). Subtilisins are a group of subtilases classified as S8 in the MEROPS database, one of the largest families of serine peptidases (Rawlings et al. 2018). Subtilisins are further subdivided, among others, into true subtilisins, high-alkaline subtilisins, intracellular subtilisins and phylogenetically intermediate subtilisins (PIS) (Siezen and Leunissen 1997; Saeki et al. 2003; Falkenberg et al. 2022a). Especially the genus *Bacillus* proved to be a valuable source of alkaline proteases such as BPN′, subtilisin Carlsberg and Savinase, which are mainly used as detergent enzymes due to their good performance and high stability towards extreme temperatures, pH values, organic solvents, detergents and oxidising agents (Kalisz 1988; Contesini et al. 2017). In addition, subtilisins are used in leather and food processing, sewage purification and as a cosmetic ingredient (Kalisz 1988; Solanki et al. 2021; Azrin et al. 2022).

Enzymes obtained from extremophilic microorganisms are not per se extremozymes in terms of their properties (Ferrer et al. 2007). In environments of extreme pH or salinity, for example, the intracellular enzymes are exposed

✉ Petra Siegert
siegert@fh-aachen.de

[1] Institute of Nano- and Biotechnologies, Aachen University of Applied Sciences, 52428 Jülich, Germany

[2] Institute of Bio- and Geosciences, IBG-1: Biotechnology, Forschungszentrum Jülich, 52425 Jülich, Germany

Published online: 09 May 2023

to conditions more typical of non-extremophiles, as the microorganisms outlast such environments by intracellularly excluding or compensating for such an environment (Ferrer et al. 2007). However, extracellular subtilisins of microbial background are mainly involved in nutrient supply and are therefore directly exposed to the environmental conditions (Kalisz 1988). Therefore, extracellular enzymes isolated from microorganisms found in environments with extreme pH, temperature and especially salinity offer huge potential to meet the needs of industry, as shown by the growing number of newly characterised subtilisins with polyextremophilic properties (Salwan and Sharma 2019; Alberto Cira-Chávez et al. 2019; Coker 2016; Falkenberg et al. 2022b; Mokashe et al. 2018). Besides the labour-intensive search for microorganisms harbouring new enzymes in such extreme environments, genome sequencing and automatic annotation offer an alternative way to search for new protease genes for industrial purposes. Sequence data of uncharacterised proteins are becoming more prevalent due to the growing number of genome sequencing projects (Rawlings 2013). Recently, we reported on a data-mining-based search for new uncharacterised subtilisins from the *Bacillaceae* family (Falkenberg et al. 2022a). Within a phylogenetic tree, these sequences were categorised to the subgroups true subtilisins, PIS and high-alkaline subtilisins. We reported about SPAO from *Alkalihalobacillus okhensis* Kh10-101^T, which has a high stability towards H_2O_2 and NaCl concentrations of up to 5.0 M and belongs to the subgroup of high-alkaline subtilisins (Falkenberg et al. 2022b). Here, we selected three sequences from the phylogenetic tree of true subtilisins and one sequence from the phylogenetically intermediate subtilisins obtained from halotolerant or halophilic bacteria for biochemical characterisation. The true subtilisins (WP_051255158.1, WP_029565418.1, WP_078544469.1) were identified in *Pontibacillus marinus*, *Metabacillus indicus* and *Litchfieldia alkalitelluris* and the PIS WP_146817050.1 in *Alkalibacillus haloalkaliphilus*.

P. marinus DSM 16465^T is a moderately halophilic bacterium isolated by Lim et al. (2005) from a saline in Korea. The strain is Gram-positive, aerobic and endospore-forming. It grew optimally on media containing 2–5% NaCl (w/v), but did not grow without NaCl or with more than 10% (w/v) NaCl. The optimum growth was observed at pH 7.0–7.5 at 30 °C. *M. indicus* DSM 16189 is a Gram-variable, endospore-forming and halotolerant bacterium isolated by Yoon et al. (2005) from jeotgal, a traditional fermented dish from Korea that contains seafood. The strain was first classified as *Bacillus cibi* in 2005 and then reclassified into *Bacillus indicus* and later into *Metabacillus indicus* (Yoon et al. 2005; Stropko et al. 2014; Patel and Gupta 2020). The strain grew optimally at 37 °C, pH 6.5–7.5 and in the presence of 0–1% (w/v) NaCl, but did not grow with more than 12% (w/v) NaCl (Yoon et al. 2005). The colonies are characteristically orange/yellow pigmented due to the production of carotenoids (Le Duc et al. 2006). *L. alkalitelluris* DSM 16976^T is an alkaliphilic bacterium isolated from sandy soil in Korea (Lee et al. 2008). The strain is Gram-positive, endospore-forming and grew optimally at 30 °C and pH 9.0–9.5 (Lee et al. 2008). The strain was reclassified from *Bacillus alkalitelluris* to *Litchfieldia alkalitelluris* (Gupta et al. 2020). The optimal NaCl concentration for growth is 0–1% (w/v), while growth occurs until 4% (w/v) NaCl (Lee et al. 2008). *A. haloalkaliphilus* DSM 5271^T was isolated by Weisser and Trüper (1985) from a saline lake of the Wadi Natrun in Egypt. It is a moderate halophilic, Gram-positive, alkaliphilic and spore-forming bacterium. The strain was first classified as *Bacillus haloalkaliphilus* and in 2005 reclassified into *Alkalibacillus haloalkaliphilus* (Fritze 1996; Jeon et al. 2005). It grows at salt concentrations between 1 and 20% (w/v) NaCl with an optimum of 5% (w/v), while cells form a flocculated and slimy sediment without growth in the absence of NaCl (Weisser and Trüper 1985). Optimal growth can be observed at pH 8.5–10.0 and 15–45 °C (Weisser and Trüper 1985; Fritze 1996).

The genes of these four extracellular subtilisins were cloned, overexpressed in *B. subtilis* DB104 and purified. This is the first report on the biochemical characterisation of the recombinant subtilisin proteases of *P. marinus* (SPPM), *M. indicus* (SPMI), *L. alkalitelluris* (SPLA) and *A. haloalkaliphilus* (SPAH).

Material and methods

Bioinformatic analysis

The sequence similarity between the four proteases and different well-known characterised subtilisins was investigated within a multiple sequence alignment (MSA) using the peptidase unit sequences from various *Bacillus* strains. The four protein sequences were blasted by using the blastp suite of NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi) (Sayers et al. 2021). The signal peptide and propeptide sequences were excluded before alignment and phylogenetic tree construction was performed via *Phylogeny.fr* (http://www.phylogeny.fr/index.cgi) using the "One-Click" option (Dereeper et al. 2008). The signal peptides were identified by the SignalP6.0 software (https://services.healthtech.dtu.dk/service.php?SignalP-6.0) (Teufel et al. 2022). Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) was used for MSA before analysis with ESPript 3.0 (Robert and Gouet 2014; Sievers et al. 2011). ESPript 3.0 was applied using %strict option (percentage of strictly conserved residues per column) for the colouring scheme (https://espript.ibcp.fr/ESPript/ESPript/). The phylogenetic tree was visualised with the iTOL software (https://itol.embl.de/) (Letunic and

Bork 2016). Structure predictions were performed through the I-TASSER server, including ligand binding prediction with COACH and COFACTOR (https://zhanggroup.org/I-TASSER/) using the amino acid sequence of the peptidase unit of the four proteases (Roy et al. 2012; Yang et al. 2013; Yang et al. 2015). The homology models were displayed with the Mol* Viewer (https://www.rcsb.org/3d-view) (Sehnal et al. 2021). For the determination of the surface-exposed residues and the calculation of the electrostatic potential with the Swiss-PdbViewer (http://www.expasy.org/spdbv/), standard settings using the Poisson-Boltzmann equation were used (Guex and Peitsch 1997). The molecular mass and the theoretical pI of the peptidase unit were determined with the Expasy system (https://web.expasy.org/compute_pi/) (Wilkins et al. 1999).

Strains and growth conditions

Bacterial strains were bought from the DSMZ—German Collection of Microorganisms and Cell Cultures GmbH and cultivated according to their recommendations: *Pontibacillus marinus* DSM 16465^T (Lim et al. 2005) and *Metabacillus indicus* DSM 16189 (Yoon et al. 2005; Stropko et al. 2014; Patel and Gupta 2020) in medium 514 + 10 mg/L $MnSO_4$ at 30 °C; *Litchfieldia alkalitelluris* DSM 16976^T (Lee et al. 2008; Gupta et al. 2020) in medium 830, pH 9.0 at 30 °C; *Alkalibacillus haloalkaliphilus* DSM 5271^T (Weisser and Trüper 1985; Fritze 1996; Jeon et al. 2005) in medium 31 with 5% NaCl, pH 9.7 at 30 °C. For the preparation of genomic DNA from an overnight culture, the InnuSPEED Bacteria/Fungi DNA Kit (Analytik Jena™, Jena, Germany) was used. For cloning and protein production, *Bacillus subtilis* DB104 was used as previously described (Kawamura and Doi 1984; Falkenberg et al. 2022b).

Plasmid construction and cloning

For recombinant protease production with *B. subtilis* DB104, pFF-RED, a pBC16-based expression plasmid (Acct. No. U32369.1) was used (Bernhard et al. 1978), as described previously (Falkenberg et al. 2022b). The genomic DNA was used to amplify the DNA sequences encoding the protease genes (including signal peptide, propeptide and the peptidase unit) in a PCR using the Phusion® Hot Start II High-Fidelity polymerase (Thermo Fisher Scientific GmbH, Karlsruhe, Germany) according to the manufacturer's recommendations. The following NCBI reference sequences were used to design primers for the *aprE* genes (extracellular alkaline protease) (Sayers et al. 2021; Clark et al. 2016): NZ_AVPF01000099.1 for *Pontibacillus marinus* DSM 16465^T encoding the protein WP_051255158.1; due to an assumed misannotation causing a partial lack of the signal peptide, in this case the annotated ORF was extended by eight codons (24 bp) at the 5′-end, leading to a TTG start codon; NZ_JNVC02000001.1 for *Metabacillus indicus* DSM 16189 encoding the protein WP_029565418.1; NZ_KV917374.1 for *Litchfieldia alkalitelluris* DSM 16976^T encoding the protein WP_078544469.1; NZ_BJYA01000014.1 for *Alkalibacillus haloalkaliphilus* DSM 5271^T encoding the protein WP_146817050.1. The oligonucleotides listed in Table S1 were obtained from Eurofins Genomics GmbH (Ebersberg, Germany) and used for PCR amplification and the introduction of two BbsI restriction sites and appropriate overhangs to the cloning site of pFF-RED. The PCR products were cloned into the BbsI sites of pFF-RED via Golden Gate cloning (Engler et al. 2008). The resulting plasmids were amplified with rolling-circle mechanism using Illustra TempliPhi 100 amplification kit (Cytiva, Marlborough, USA) before being used to transform naturally competent cells of *B. subtilis* DB104 as described elsewhere (Vojcic et al. 2012). The control of successful cloning and transformation was performed as previously described (Falkenberg et al. 2022b). The production of the protease was confirmed by proteolytic activity assays using N-succinyl-Ala-Ala-Pro-Phe-p-nitroanilide (suc-AAPF-pNA) and azocasein as substrate and by sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE).

Recombinant protease production and purification

Production of the proteases by *Bacillus subtilis* DB104 was carried out on a 1-L scale using the DASGIP® parallel reactor system (DASGIP, Jülich, Germany) as described previously (Falkenberg et al. 2022b). The suc-AAPF-pNA assay, azocasein assay and SDS-PAGE were used to confirm the protease production. The protease purification was performed in a three-step protocol as described previously (Falkenberg et al. 2022b). For the proteases an anion exchanger (25 ml Q-Sepharose FF, GE Healthcare, IL, USA) and a pH of 7.0 for running (10 mM HEPES-NaOH buffer) and elution buffer (10 mM HEPES-NaOH, 1 M NaCl) were used.

Enzyme activity assay

The hydrolytic activity of the proteases was determined using the tetrapeptide substrate suc-AAPF-pNA (BACHEM, Bubendorf, Switzerland) at 30 °C in 100 mM Tris-HCl buffer, pH 8.6, containing 0.1% (w/v) Brij®35 as described previously (DelMar et al. 1979; Falkenberg et al. 2022b). In addition, protease activities were determined using azocasein (Sigma-Aldrich, Schnelldorf, Germany) as substrate according to Brock et al. at 37 °C in 100 mM Tris-HCl buffer, pH 8.6 as described before (Brock et al. 1982; Falkenberg et al. 2022b). The substrate specificity of the proteases for

different synthetic 4-nitroanilide substrates was determined as previously described (Falkenberg et al. 2022b).

Analytical methods, protein measurement and electrophoresis

Protein concentrations were determined according to Bradford (1976) using Roti®Nanoquant (Carl Roth, Germany) and bovine serum albumin fraction V (Carl Roth, Germany) as a standard and measuring the absorbance ratio 590 nm/450 nm. Matrix-assisted laser desorption ionisation-time of flight mass spectrometry (MALDI-TOF-MS) was performed with the Axima confidence instrument (Shimadzu Europe, Duisburg, Germany) as previously described (Falkenberg et al. 2022b). SDS-PAGE analysis was performed as described by Miller et al. (2016) using an 8–20% (v/v) resolving gel and a 6% (v/v) stacking gel. The sample preparation and electrophoresis were performed as described before (Falkenberg et al. 2022b). For isoelectric focusing–polyacrylamide gel electrophoresis (IEF-PAGE), purified proteases were re-buffered in 10 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)-NaOH pH 7.0 using centrifugal spin columns (VWR, Radnor, USA) with a molecular mass cut-off of 3 kDa. The SERVALYT™ PRECOTES™ 3-10 gel (SERVA, Heidelberg, Germany) was used according to the manufacturer's recommendations.

Effect of SDS, hydrogen peroxide and PMSF on enzyme activity and stability

SPFA was incubated with H_2O_2 (1 and 5% (v/v)) and SDS (1 and 5% (w/v)) in 10 mM HEPES-NaOH, pH 8.0 for 1 h at 10 °C. The influence of the protease inhibitor phenylmethylsulfonyl fluoride (PMSF) was investigated by incubating the proteases in 10 mM HEPES-NaOH with 1 mM PMSF, pH 8.0, for 30 min on ice. Residual activity was measured in the standard suc-AAPF-pNA activity assay and residual activity of the proteases incubated in buffer with no additives was defined as 100%. The stored samples were used undiluted in the activity assay, so that 0.1 and 0.5% of H_2O_2 and SDS remained during the measurement.

Effect of NaCl, $CaCl_2$ and EDTA on enzyme activity and stability

The effect of NaCl on proteolytic activity was measured using the suc-AAPF-pNA assay under standard conditions with the addition of NaCl (0–5 M) in the reaction buffer as described before (Falkenberg et al. 2022b). Proteases were incubated in 10 mM HEPES-NaOH pH 8.0 with NaCl (0–5 M) at 20 °C for 2 h to investigate the effects of NaCl on enzyme stability. The activities measured before incubation were defined as 100%. The effect of ethylenediaminetetraacetic acid (EDTA) and $CaCl_2$ was examined as described before (Falkenberg et al. 2022b).

Effects of temperature and pH on enzyme activity and stability

Within a thermal shift assay, the thermal protein unfolding and the melting points of the proteases were determined by using the fluorescent dye SYPRO™ Orange (Thermo Fisher Scientific GmbH, Karlsruhe, Germany) as described previously (Falkenberg et al. 2022b). The temperature optimum was assayed with the suc-AAPF-pNA assay between 20 and 90 °C in 5 °C steps and the temperature stability was determined by measuring the residual activity after incubating the proteases at 20 and 50 °C for 3 h in 10 mM HEPES-NaOH, pH 8.0. The pH optimum of the proteases was determined in 0.1 M Tris-maleate buffer (pH 5.0–7.0), 0.1 M Tris-HCl (pH 7.0–9.0) and 0.1 M glycine-NaOH (pH 9.0–12.5) at 30 °C using the suc-AAPF-pNA assay. Residual activities of the proteases were measured with the suc-AAPF-pNA assay after incubating the proteases in said buffers for 24 h at 4 °C.

Results

Cloning and expression of the *aprE* genes in *B. subtilis* DB104

Four uncharacterised proteases were selected from our previous report about a data-mining-based search for new subtilisins from *Bacillaceae* (Falkenberg et al. 2022a). The coding sequences of *aprE_P. marinus* for the protease SPPM, *aprE_M. indicus* for SPMI, *aprE_A. haloalkaliphilus* for SPAH and *aprE_L. alkalitelluris* for SPLA were amplified as described in methods and fragments of 1172 bp, 1152 bp, 1151 bp and 1148 bp were obtained, respectively. The PCR products containing the sequence for the signal peptide, the propeptide and the peptidase unit of the proteases were cloned into the vector pFF-RED and transferred into *B. subtilis* DB104. Successful transformation was checked with arising clearing zones on LB agar plates complemented with 2.5% (w/v) skim milk around the colonies and was analysed by plasmid preparation and restriction analysis. The DNA sequences of the cloned genes were determined by Sanger sequencing and verified that they were identical to the genomic nucleotide sequences of the four *aprE* genes.

Homology modelling and bioinformatic analysis

The *aprE* genes from *M. indicus* DSM 16189 and *A. haloalkaliphilus* DSM 5271^T comprise 1128 bp each and encode proteins of 375 amino acids. The *aprE* gene from *L. alkalitelluris* DSM 16976^T comprises 1131 bp encoding a protein

of 376 amino acids. The *aprE* gene from *P. marinus* DSM 16465^T comprises 1152 bp encoding a protein of 383 amino acids. In this case, we corrected the automatic annotation by extension of the ORF by eight codons at the 5′-end, leading to a TTG start codon. The signal peptide prediction revealed the presence of a Sec signal peptide for all four selected proteases with a probability above 97% (Fig. 1). The propeptides were identified by multiple sequence alignment and are marked in Fig. 1. The MSA shows that SPAH has a double insertion within a loop between position 42 and 43 in contrast to BPN′ (numbering refers to the mature part of BPN′). In comparison to the other investigated proteases, SPAH has an insertion between position 159 and 160. At the C-terminus, SPPM displays an extension of ten amino acids in comparison to BPN′ and nine amino acids in comparison to SPAH. Without these nine amino acids, the theoretical mass of SPPM is 27.90 kDa, which is congruent to the MALDI-TOF MS analysis as shown later. For the phylogenetic analysis and the homology modelling, the SPPM mature part without its probable C-terminal extension was used. The in silico analysis of mature proteins revealed a molecular mass of 27.48 kDa and a pI of 5.5 for SPMI, 27.47 kDa and a pI of 5.1 for SPLA, 28.6 kDa and a pI of 4.3 for SPAH and 27.90 kDa and a pI of 4.2 for SPPM without the C-terminal extension. The catalytic triad consists of Asp^{32}, His^{64} and Ser^{221} in SPPM, SPMI and SPLA and Asp^{32}, His^{66} and Ser^{224} in SPAH (numbers based on the mature protease sequences).

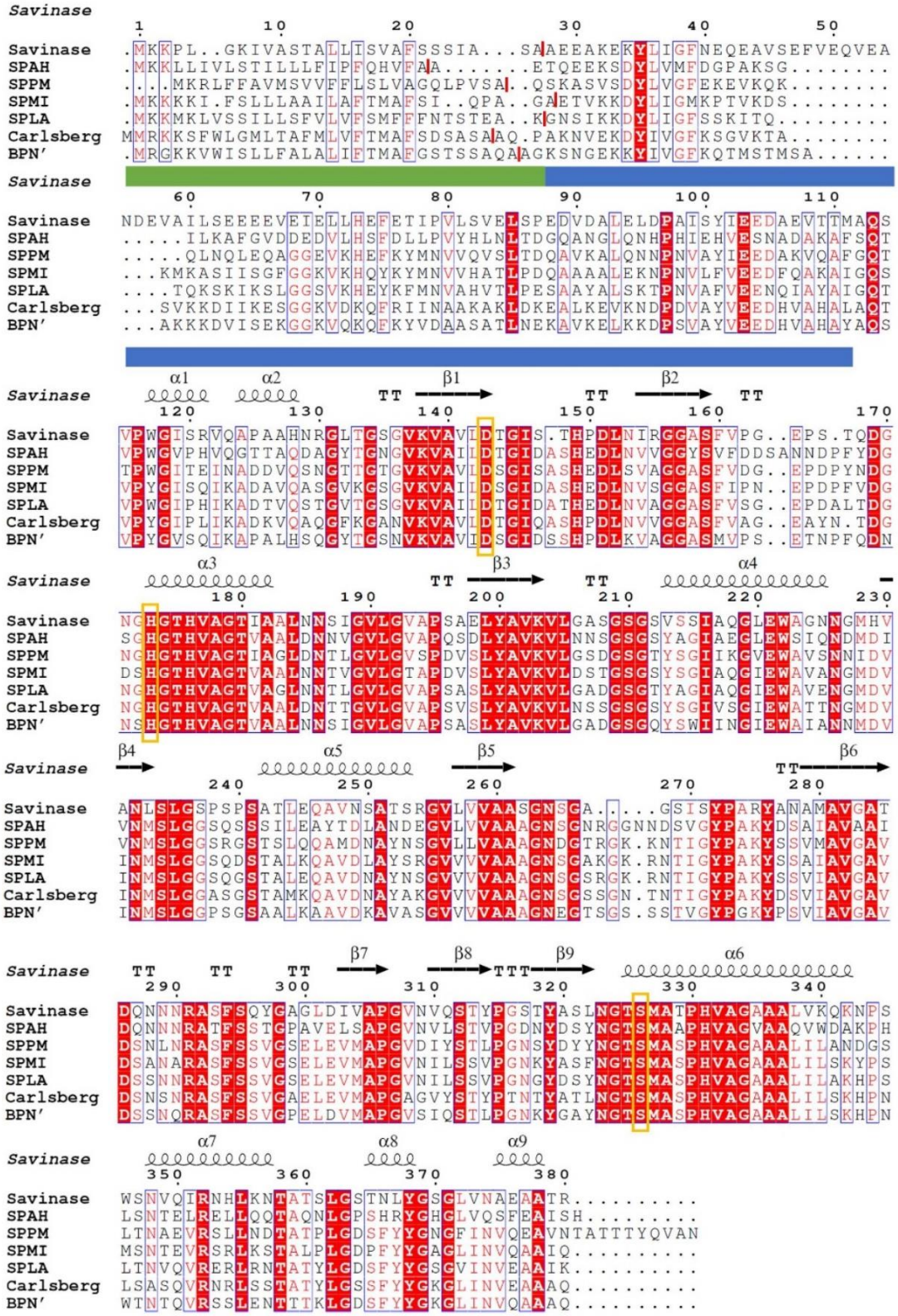
The amino acid sequences of the peptidase unit of the four proteases were aligned and compared in a phylogenetic tree with well-characterised proteases of the three subtilisin families (true, phylogenetically intermediate, high-alkaline) retrieved from the UniProt (The UniProt Consortium 2021) and MEROPS database (Rawlings et al. 2018) as well as our previously characterised high-alkaline subtilisin from *Alkalihalobacillus okhensis* K10-101^T (Falkenberg et al. 2022b) (Fig. 2). There, SPAH is clearly a member of the PIS subgroup with a sequence identity of 73.4% to LD1 from *Bacillus* sp. KSM-LD1 (Saeki et al. 2003), 57.7% to ALTP from *Alkaliphilus transvaalensis* (Kobayashi et al. 2007), and a more distant relationship to the well-characterised true subtilisins BPN′ (53.1%) (Matsubara et al. 1965) and the high-alkaline subtilisin Savinase from *Lederbergia lenta* (formerly *Bacillus lentus*) (55.4%) (Betzel et al. 1992). SPPM, SPLA and SPMI cluster together within the subgroup of true subtilisins. The sequence identity between SPPM and SPLA is 76.4% and between SPPM and SPMI 69.8%, while the sequence identity between SPLA and SPMI is 78.6%. The highest sequence identity to the well-characterised subtilisin Carlsberg (Smith et al. 1966) was displayed by SPLA with 74.8%. A more detailed phylogenetic comparison with all subtilisin sequences from *Bacillaceae* was reported before (Falkenberg et al. 2022a).

A homology modelling of the four mature proteases is shown in Fig. S1. The C-score of the models of SPPM and SPMI is 1.52 and of SPLA and SPAH 1.51 and 1.40, respectively. The C-score ranges from −5 to 2 and higher values indicate higher confidence of the model (Zhang 2008). A high TM-score (template modelling) is indicated for BPN′ (PDB: 1S01) with 0.997 for SPPM, 0.995 for SPAH, 0.977 for SPMI and 0.966 for SPLA, where a TM-score of 1 suggests a perfect match between two structures (Zhang and Skolnick 2004). As mentioned above, SPPM contains a presumably nine amino acid long C-terminal extension, and when the 3D structure is calculated with this extension, this extension projects away from the core molecule (data not shown). By using the Swiss-PdbViewer, the 3D structures were used to calculate the electrostatic potential at pH 7.0 as shown in Fig. 3. All four proteases are mainly negatively charged around the active site, while on the back side SPPM and SPAH are completely negatively charged. SPMI shows a more neutral to positive charge, while SPLA is more negatively charged but shows some neutral to positive charged areas at the back side. The in silico analysis of the homology models for Ca^{2+}-binding sites suggested that all four proteases harbour two potential Ca^{2+}-binding sites. The side chains Gln^2 and Asp^{41} and several side chains of the loop-forming residues 75–81 including Leu^{75}, Asn^{77}, Leu^{79} (SPPM and SPLA), Val^{79} (SPMI and SPAH) and Val^{81} (Savinase numbering) are involved in the first Ca^{2+}-binding site. The side chains Ala^{169}, Tyr^{171}, Val^{174} (SPPM and SPLA) and Ala^{174} (SPMI and SPAH) are involved in the second Ca^{2+}-binding site.

Recombinant protease production and purification

Culture supernatants produced from recombinant *B. subtilis* DB104 in a 1-L fermentation showed activity with suc-AAPF-pNA as substrate of 63 U/ml for SPPM, 69 U/ml for SPMI, 73 U/ml for SPLA and 18 U/ml for SPAH. The four supernatants were used for a three-step purification process as described previously (Falkenberg et al. 2022b). The successful purification of the four proteases to apparent homogeneity was confirmed via SDS-PAGE (Fig. 4). The proteases SPPM and SPAH migrate at approximately 35 kDa in contrast to the theoretical molecular mass of 27.9 and 28.6 kDa, respectively. SPMI and SPLA migrate also higher than the theoretical molecular mass of 27.5 kDa at approximately 30 kDa. The MALDI-TOF MS analysis revealed a molecular mass of 27.49 kDa for SPMI, 27.48 kDa for SPLA, 28.60 kDa for SPAH and 27.97 kDa for SPPM (Fig. S2).

Purified SPPM and SPMI had specific activities of 208 and 160 U/mg for the suc-AAPF-pNA substrate and 1371 and 1085 U/mg for azocasein, respectively. Purified SPLA and SPAH had a specific activity of 233 and 314 U/mg for the AAPF substrate and 1036 and 2719 U/mg for azocasein.


Savinase
SPAH
SPPM
SPMI
SPLA
Carlsberg
BPN′
α1
α2
β1
β2
α3
β3
α4
β4
α5
β5
β6
β7
β8
β9
α6
α7
α8
α9

◂**Fig. 1** SPPM, SPMI, SPLA, SPAH, Savinase (WP_094423791.1), subtilisin Carlsberg (WP_020450819.1) and BPN′ (WP_013351733.1) within a multiple sequence alignment. The alignment was calculated using Clustal Omega and analysed using ESPript 3.0 and Savinase (PDB: 1C9J) as a template. Solid blue and green bars indicate the propeptide and signal peptide sequence of Savinase. Individual signal peptide cleavage sites are marked with a red bar. Helices are marked with squiggles, β-strands with arrows, and turns with TT letters. The residues of the catalytic triad are marked by orange boxes (Asp^{143}, His^{173}, Ser^{326}; Savinase numbering)

The analysis of the isoelectric point of the purified and rebuffered proteases showed a pI for SPPM of approx. 4.3, which is near the predicted pI of 4.2 (Fig. S5). For SPMI, a pI of approx. 5.5 was measured, which corresponds to the theoretical value. Furthermore, SPLA showed a pI of approx. 5.0, close to the theoretical pI of 5.1, and SPAH had a pI of approx. 4.9, which deviates from the theoretical pI of 4.3. All four proteases have an acidic pI and an AB ratio above 1.0 with a high number of Asp residues (Table S2).

Effects of temperature and pH on enzyme activity and stability

The influence of temperature on enzyme activity was investigated in a temperature range from 20 to 90 °C at a pH of 8.6, as described in the methods section (Fig. 5). The activity of SPLA and SPMI gradually increased from 20 °C to the optimum of 70 °C and decreased to 48 and 47% residual activity at 90 °C, respectively. The lowest temperature optimum of 50 °C showed SPAH, and activity measurement was only possible up to 75 °C with residual activity of 14%. The temperature profile is comparable to that of SPPM with an optimum at 55 °C and 17% residual activity at 80 °C.

Enzyme stability towards different temperatures was investigated either by incubation at 20 and 50 °C or by monitoring thermal protein unfolding in a thermal shift assay. Figure 6 shows the remaining protease activity during an incubation at 20 and 50 °C for 4 h. The activity of SPPM, SPLA and SPAH remained quite stable with a residual activity of over 75% after 4 h, while SPMI lost 85% of its activity in this period. The loss of activity was more distinct during incubation at 50 °C. While SPPM and SPMI lost all their remaining activity after 4 h at 50 °C, SPAH retained 10% and SPLA 52% of their activity. However, a comparison of the temperature stability of proteases is difficult due to the possible autoproteolysis during incubation. In order to monitor thermal protein unfolding instead of autoproteolysis, the proteases were tested in a thermal shift assay. Proteases were inhibited with phenylmethylsulfonyl fluoride (PMSF) and denaturation curves were recorded (Fig. S4). SPMI and SPLA revealed melting points (T_m) of 62.5 and 61.5 °C, respectively. It was not possible to obtain a melting curve for SPPM and SPAH.

At a pH range of 5.0–12.0 showed all four proteases a comparable pH profile until pH 11.0 (Fig. 7). For SPPM and SPMI, the highest activity was observed at pH 9.0–9.5. The relative activities at pH 5.0 and 12.0 were 1% and 63% for SPPM and 4% and 62% for SPMI. SPLA and SPAH showed pH optima at pH 9.0 and 10.0, respectively. The relative activities at pH 5.0 and 12.0 were 7% and 91% for SPLA and 5% and 37% for SPAH. The stability test of the proteases at different pH values showed that all four proteases retained a residual activity of at least 65% at pH 5.0, 94% around the pH optimum and 83% at pH 12.0 (Fig. S3).

Effect of SDS and H_2O_2 on enzyme activity

The activity of the four proteases after 1-h incubation with 1 and 5% SDS (w/v) at 10 °C showed that they possess a high stability towards SDS (Table 1). All proteases exhibited an enhanced activity after incubation with 1% SDS of up to 178% (SPLA). After incubation with 5% SDS, only SPPM revealed a reduced residual activity of 76%, while SPLA with 97% residual activity showed almost no decrease. SPMI displayed the highest stability and had a residual increased activity of 165% after incubation with 5% SDS.

The influence of 1 and 5% (v/v) H_2O_2 is shown in Table 1. After 1 h of treatment with 1% H_2O_2, all proteases showed a high residual activity of 78 to 92%. Treatment with 5% H_2O_2 reduced the remaining activity to 31–40%. PMSF is a classical inhibitor for serine proteases (North 1982), and the incubation of the four serine proteases SPPM, SPMI, SPLA and SPAH with 1 mM PMSF led to a complete inhibition (Table 1).

The purified proteases were incubated with 1 and 5% (v/v) H_2O_2, 1 and 5% (w/v) SDS and 1 mM PMSF in 10 mM HEPES-NaOH pH 7.0 for 1 h at 10 °C. Remaining activity of the proteases incubated in buffer with no additions was defined as 100%. During the measurement, 0.1 and 0.5% of H_2O_2 and SDS remained, respectively. All experiments were performed at least in triplicates and data are shown as mean values ± SD

Effect of NaCl, $CaCl_2$ and EDTA on enzyme activity and stability

In the suc-AAPF-pNA activity assay with different NaCl concentrations (0–5 M), the protease SPPM showed the highest activity without NaCl and the activity gradually decreased with higher NaCl concentrations to 51% at 5 M NaCl, comparable to the proteases SPMI and SPLA with 44 and 41% residual activity (see Fig. 8a). The protease SPAH revealed the highest activity at 0 M NaCl, which dropped to 60% at 1 M NaCl, but then displayed no further loss at higher NaCl concentrations. In addition, the stability of the proteases was investigated in the presence of salt by incubation with NaCl (0–5 M) in 10

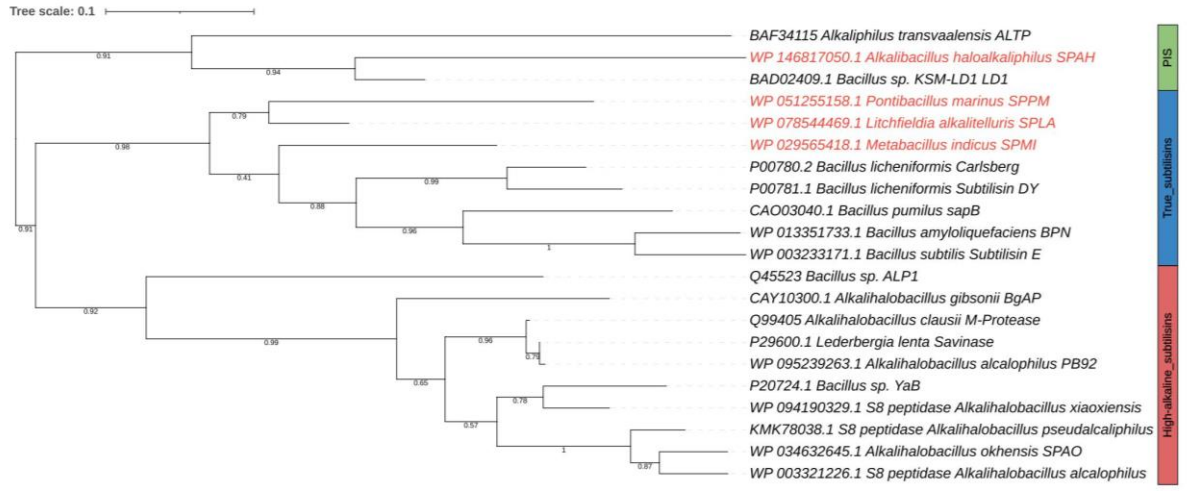


Fig. 2 Phylogenetic tree analysis of SPPM, SPMI, SPLA and SPAH with various characterised subtilisins from different species of the family *Bacillaceae*. The Phylogeny.fr server was used for maximum likelihood phylogenetic analysis of the peptidase unit. Branch support is indicated with numbers obtained by approximate likelihood ratio test (SH-like aLRT)

mM HEPES-NaOH, pH 8.0 for 2 h at 20 °C (Fig. 8b). SPPM, SPMI and SPLA were stable with and without NaCl, while SPAH lost activity with increasing NaCl concentration.

The effect of Ca^{2+} on the activity of the proteases was studied as subtilisins are calcium-dependent and the binding of Ca^{2+} is essential for enzyme activity and stability (Siezen et al. 1991). The in silico studies revealed two calcium binding sites for the four proteases, as mentioned above. When incubated with or without EDTA, SPPM and SPAH showed almost no difference in activity (data not shown). Recovery of activity was not possible for SPLA and SPMI after the addition of $CaCl_2$.

Proteolytic activity on synthetic peptides

The specificity of the four proteases towards ten synthetic peptide 4-nitroanilide substrates with three or four amino acids was analysed (Table 2). With the exception of suc-AAA-pNA, which is an elastase substrate, all are typical subtilisin substrates (Georgieva et al. 2001; Bieth et al. 1974). The proteases showed a very low specificity for suc-YVAD-pNA, suc-TVAA-pNA, suc-AAA-pNA and suc-AAVA-pNA. Highest activity was displayed for suc-ALPF-pNA and suc-AGPP-pNA.

Discussion

Subtilisins are extremely versatile serine peptidases from the subtilase family and due to their properties such as thermostability, broad pH range and broad specificity, of particular interest for industrial applications (Azrin et al. 2022). Especially subtilisins from halophilic or halotolerant *Bacillaceae* have a high potential for meeting industrial needs (Salwan and Sharma 2019; Alberto Cira-Chávez et al. 2019; Coker 2016). Therefore, in this study, we characterise four subtilisins from halophilic and halotolerant *Bacillaceae* found in our previous publication through a data mining search (Falkenberg et al. 2022a).

The coding sequences of *aprE_P. marinus* for the protease SPPM, *aprE_M. indicus* for SPMI, *aprE_A. haloalkaliphilus* for SPAH and *aprE_L. alkalitelluris* for SPLA were amplified. Most subtilases consist of a signal peptide for translocation, a propeptide for maturation, a protease domain and sometimes additional domains (Siezen et al. 2007). However, the annotated gene of SPPM would lack eight amino acids of the signal peptide. In this case, we corrected the automatic annotation by extension of the ORF by eight codons at the 5′-end, leading to a TTG start codon, which is not uncommon for *Bacillus* sp. (Rocha et al. 1999).

When analysing the amino acid sequence of the four proteases in a multiple sequence alignment, shows that SPAH has a double insertion within a loop between position 42 and 43, and an insertion between position 159 and 160 in contrast to BPN′. These insertions are typical of phylogenetically intermediate subtilisins, as our earlier study shows (Falkenberg et al. 2022a). Additionally, around position 160 high-alkaline subtilisins have a four-amino acid deletion in common (Falkenberg et al. 2022a), which cannot be observed for SPPM, SPMI, SPLA and SPAH. Position 160 is within a loop associated with the P1 binding site, which is the first position N-terminal to the cleavage site and

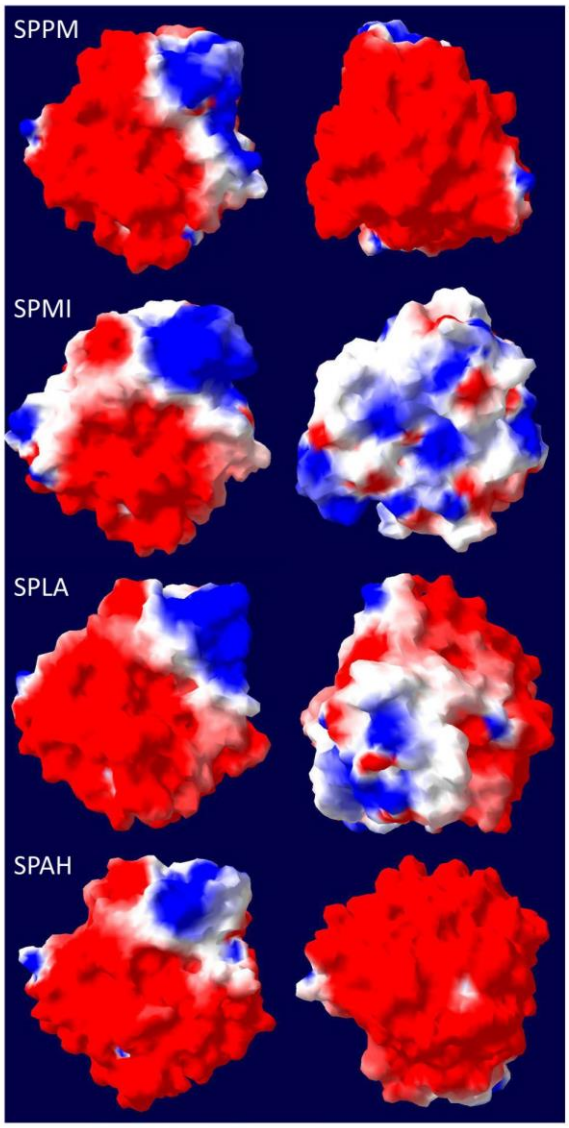


Fig. 3 Homology models of SPPM, SPMI, SPLA and SPAH with its calculated protein surface electrostatic potential. Front view of the active site (left); back side of the active site (right). The red (negative) and blue (positive) areas show the electrostatic potential at pH 7.0 calculated with the Swiss-PdbViewer

therefore may be involved in P1 preference and steric conformation (Wells et al. 1987; Betzel et al. 1992; Yamagata et al. 1995; Schechter and Berger 1967). In addition, shorter loops can increase enzyme stability (Gavrilov et al. 2015).

Furthermore, the subtilisins of *Bacillus* species usually contain one strong and one weak Ca^{2+}-binding site (Siezen and Leunissen 1997). The strong Ca^{2+}-binding site, which is conserved in diverse subtilases, requires the side chains of residues Gln^{2} and Asp^{41} and several side chains of the loop-forming residues 75–81, which are also found for the four proteases (Siezen and Leunissen 1997; Betzel et al. 1992). The occupancy of the weak site is dependent of the $CaCl_2$ concentration in the solution and the side chains Ala^{169}, Tyr^{171}, Val^{174} (SPPM and SPLA) and Ala^{174} (SPMI and SPAH) are involved (Savinase numbering) (Siezen et al. 1991; Betzel et al. 1992). When the proteases were incubated with EDTA, SPPM and SPAH showed nearly no difference in activity between incubation with and without EDTA. SPLA and SPMI, however, lost activity. Since Ca^{2+} has a stabilising effect on the protease, this could be due to a general loss of activity during the incubation with EDTA (Siezen et al. 1991), which was also observed by others (Dodia et al. 2008; Thaz and Jayaraman 2014; Vidyasagar et al. 2009). A protein-engineered version of BPN′ led to a calcium-independent protease that is fully active, but has a lower thermal stability in the absence of stabilising mutations (Almog et al. 1998).

SPPM, SPMI, SPLA and SPAH migrate higher on the SDS-PAGE than would be expected on the basis of the theoretical masses. A migration behaviour of proteins during SDS-PAGE deviating from the expected molecular mass is not uncommon and may occur due to partial refolding or altered detergent binding (Matagne et al. 1991; Rath et al. 2009). The MALDI-TOF-MS analysis thus confirmed the theoretical masses. SPPM displays an extension of ten amino acids in comparison to BPN′ and nine amino acids in comparison to SPAH at the C-terminus when comparing the amino acid sequences. Without these nine amino acids, the theoretical mass of SPPM is 27.90 kDa. A mass of 27.97 kDa was determined by MALDI-TOF MS, close to the value of 27.90 kDa predicted for the variant lacking the C-terminal nine amino acids. Therefore, the C-terminus seems to be processed in an intermolecular process by other SPPM molecules, as this extension projects away from the core molecule, supporting the assumption that it is prone to proteolytic processing. However, it could also be that only 8 amino acids are cleaved off, because with the additional Ala, the calculated molecular weight would agree with the experimental one.

The four proteases have a temperature optimum of 50–70 °C, which is in the range of other subtilisins such as subtilisin Carlsberg, BPN′ and Savinase, which have their optimum at 65 °C, 55 °C and 60 °C, respectively (Falkenberg et al. 2022b). Comparing the optimal temperatures with the optimal growth temperature of the bacterial origins, differences of up to 20 °C can be observed. *P. marinus* and *L. alkalitelluris*, the native hosts of SPPM and SPLA, grow between 15 and 40 °C (Lim et al. 2005; Lee et al. 2008). *M. indicus* and *A. haloalkaliphilus*, the native hosts of SPMI and SPAH, grow at 4–50 °C and 15–45 °C, respectively (Yoon et al. 2005; Fritze 1996). This observation is consistent with the study of Engqvist (2018) who reported that proteins from

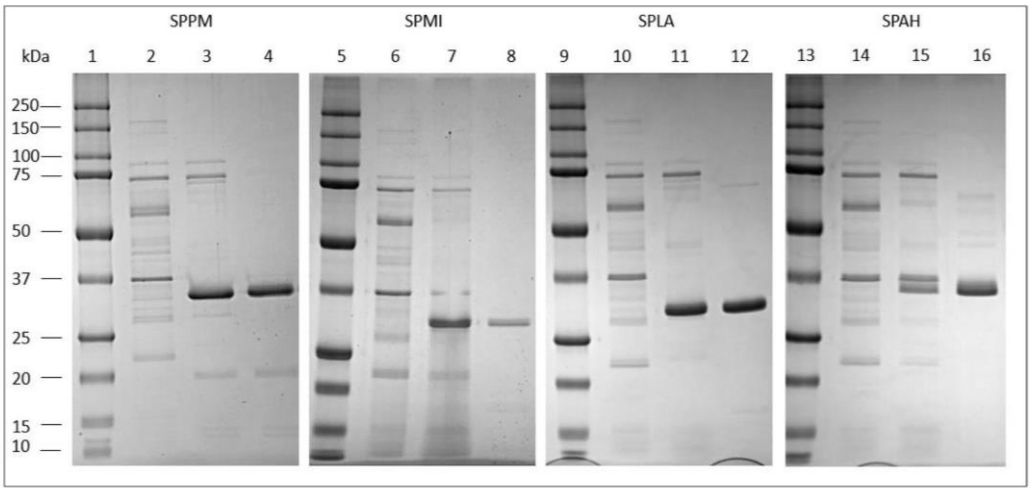


Fig. 4 SDS-PAGE analysis of recombinant SPPM, SPMI, SPLA and SPAH. An 8–20% SDS polyacrylamide gel was used for electrophoresis. Bio-Rad Precision Plus Dual Color length marker (1, 5, 9, 13). *B. subtilis* DB104-pFF-RED culture supernatant as negative control (2, 6, 10, 14); *B. subtilis* DB104 culture supernatant producing SPPM (3), after purification (4); SPMI (7, 8), SPLA (11, 12), SPAH (15, 16) before and after purification

mesophiles (15–50 °C) tend to be catalytically active at higher temperatures than expected based on the growth temperature. When analysing the melting point of the four proteases, it was not possible to obtain a melting point for SPPM and SPAH. According to Boivin et al., this could be due to protein precipitation, aggregation, some intrinsically disordered regions with complicated folding landscape or a high hydrophobic background masking the melting transition (Boivin et al. 2013). The high melting point of SPMI and SPLA compared to the value obtained for the subtilisin SPAO from *Alkalihalobacillus okhensis* (T_m of 53.0 °C) correlates with the higher temperature optimum (Fig. 5) (Falkenberg et al. 2022b).

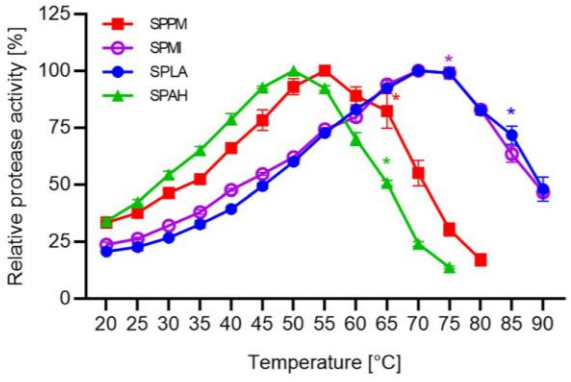


Fig. 5 Activity of purified SPPM, SPMI, SPLA and SPAH at different temperatures. Protease activity at temperatures between 20 and 90 °C was measured with the suc-AAPF-pNA assay. Maximum activity for each protease was defined as 100%: 263 U/mg for SPPM (red squares), 540 U/mg for SPMI (violet open circles), 1092 U/mg for SPLA (blue circles) and 371 U/mg for SPAH (green triangles). *The enzyme was not stable for the intended 5 min. The experiments were performed in triplicates and data are plotted as means ± SD

Although the proteases do not belong to the high-alkaline proteases phylogenetically, they showed a high activity at alkaline pH until pH 12.0, which shows the potential for various industrial applications (Tekin et al. 2021; Phrommao et al. 2011; Gurunathan et al. 2021). In particular, SPLA is characterised by a high residual activity of 91% at pH 12.0, which is even higher than the residual activity of 53% of the high-alkaline subtilisin Savinase (Falkenberg et al. 2022b). The two other true subtilisins SPPM and SPMI with a residual activity of over 60% also stand out, as other proteases within this subgroup such as BPN′, subtilisin Carlsberg and endopeptidase Q show lower activity with 6, 19 and about 8% relative residual activity at pH 12.0, respectively (Falkenberg et al. 2022b; Han and Damodaran 1998). In contrast to SPAH with a residual activity of 37%, the phylogenetically intermediate subtilisin ALTP from *Alkaliphilus transvaalensis* showed its pH optimum at a pH above 12.6 (Kobayashi et al. 2007). The adaptation to higher alkaline conditions by high-alkaline subtilisins is indicated by an altered surface charge with an increased pI value and, in particular, by an increased amount of Arg and a reduced amount of Lys residues (Masui et al. 1998). Interestingly this correlation does not hold with the four investigated proteases, as they all have an acidic pI and, compared to the two high-alkaline subtilisins SPAO and Savinase, a decreased number of Arg residues and an increased number of Asp

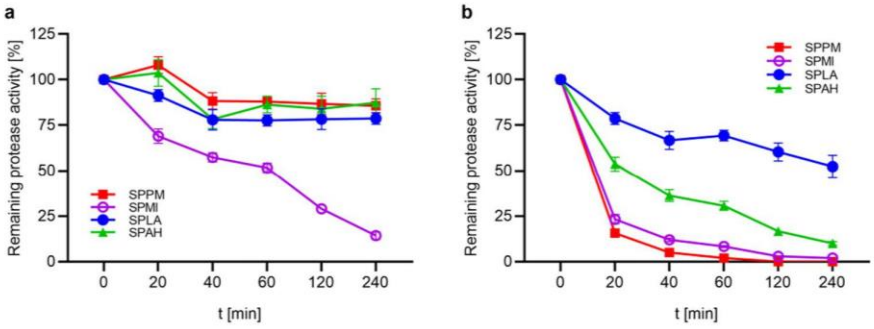


Fig. 6 Temperature stability of purified SPPM, SPMI, SPLA and SPAH. Stability was studied at 20 °C (**a**) and 50 °C (**b**) in 10 mM HEPES-NaOH buffer, pH 8.0. The suc-AAPF-pNA assay was used to measure the activity at 30 °C. The activity at 0 min was defined as 100%: 121 U/mg for SPPM (red squares), 151 U/mg for SPMI (violet open circles), 212 U/mg for SPLA (blue closed circles) and 197 U/mg for SPAH (green triangles). The experiment was performed in triplicates and data are plotted as means ± SD

residues (Falkenberg et al. 2022b). An increased number of charged amino acids on the protein surface leads to better ionic interactions, thus maintaining stability and solubility (Panja et al. 2020). The stability test of the proteases at different pH values showed that all four proteases retained a residual activity of at least 65% at pH 5.0, 94% around the pH optimum and 83% at pH 12.0 (Fig. S3). This is comparable to other subtilisins as previously discussed (Falkenberg et al. 2022b).

When incubated with 1 and 5% SDS (w/v), SPPM, SPMI, SPLA and SPAH all show high stability towards SDS. As reported in our previous study, the three known proteases subtilisin Carlsberg, BPN′ and Savinase also showed higher activities after incubation with 1 and 5% (w/v) SDS than without (Falkenberg et al. 2022b). Instead of supporting protein unfolding, SDS can help to achieve a favourable protein conformation in some subtilisins, as has been reported for other subtilisins as well (Falkenberg et al. 2022b; Bhatt and Singh 2020; Joshi and Satyanarayana 2013; Thebti et al. 2016). For a salt-tolerant and thermostable protease from *B. subtilis*, no loss of activity was observed even at an SDS concentration of 10% (w/v) (Kembhavi et al. 1993). Rekik et al. reported for an alkaline serine protease from *Bacillus safensis* RH12 a reduction of the residual activity to 90 and 60% after incubation with 1 and 5% SDS (Rekik et al. 2019). However, for SPAO from *A. okhensis*, incubation with SDS led to a complete loss of activity, which is quite unusual for highly-alkaline subtilisins (Falkenberg et al. 2022b).

When analysing stability against H_2O_2, the four new proteases are highly stable against oxidation, contrary to Savinase, subtilisins Carlsberg and BPN′, which lost up to 92% of their activity (Savinase) under these conditions (Falkenberg et al. 2022b). However, our previously reported high-alkaline subtilisin SPAO showed an even higher resistance to H_2O_2 with an increased activity of 108% at 1% H_2O_2 and a remaining activity of 58% after incubation with 5% H_2O_2 (Falkenberg et al. 2022b). The sensitivity to H_2O_2 is probably due to a conserved methionine near the catalytic serine, which is oxidised to its sulphoxide. The sulphoxide oxygen is directed towards the oxyanion hole and destabilises the tetrahedral intermediate formed with the carbonyl group of the substrate (Stauffer and Etson 1969; Bott et al. 1988; Bryan et al. 1986). The comparison of the obtained data with

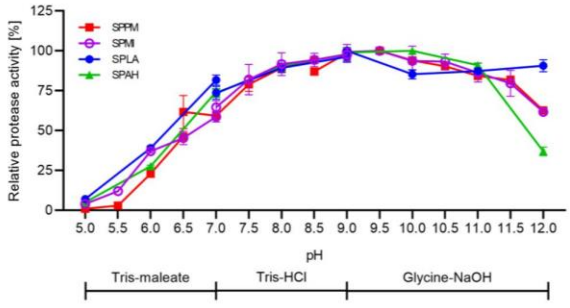


Fig. 7 SPPM, SPMI, SPLA and SPAH activity at different pH. The test was performed with the suc-AAPF-pNA assay in the pH range of 5.0–12.0 at 30 °C. Maximum activity for each protease was defined as 100%: 98 U/mg for SPPM (red squares), 104 U/mg for SPMI (violet open circles), 201 U/mg or SPLA (blue closed circles) and 176 U/mg for SPAH (green triangles). The experiments were performed in triplicates and data are plotted as mean values ± SD

Table 1 Influence of SDS, H_2O_2 and PMSF on enzyme activity

Protease	Residual protease activity (%)				
	1% SDS	5% SDS	1% H_2O_2	5% H_2O_2	1 mM PMSF
SPPM	121 ± 4	76 ± 2	92 ± 4	41 ± 0	0 ± 0
SPMI	178 ± 4	165 ± 1	81 ± 1	31 ± 2	0 ± 0
SPLA	108 ± 4	97 ± 3	78 ± 2	40 ± 1	0 ± 0
SPAH	120 ± 7	106 ± 2	89 ± 1	40 ± 2	0 ± 0

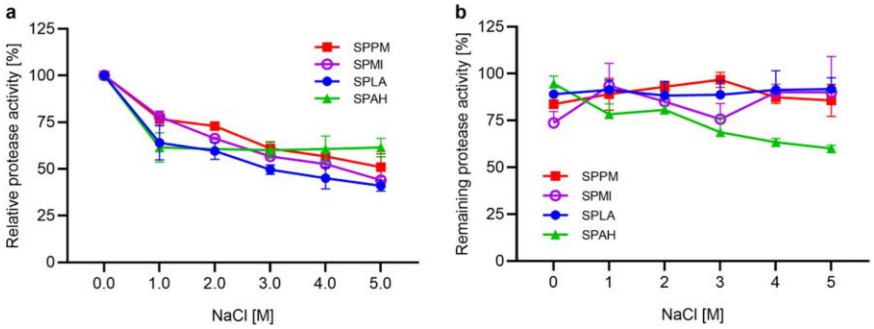


Fig. 8 The effect of NaCl on the activity (**a**) and stability (**b**) of the purified proteases SPPM, SPMI, SPLA and SPAH. Experiment was performed with the suc-AAPF-pNA assay at 30 °C with different NaCl concentrations (0–5 M). Maximum activity for each protease was given as 100%: 551 U/mg for SPPM, 601 U/mg for SPMI, 337 U/mg for SPLA and 207 U/mg for SPAH. The proteases were incubated in 10 mM HEPES-NaOH buffer, pH 8.0, with the indicated NaCl concentrations to test the stability. Residual activity was assayed after 2-h incubation at 20 °C with the suc-AAPF-pNA assay in standard buffer at pH 8.6. The initial activity at each NaCl concentration was considered as 100% activity. The experiments were performed in triplicates and data are given as mean values ± SD

literature data is difficult due to the different experimental conditions. Subtilisin LD1 from *Bacillus* sp. KSM-LD1 maintained 40% of its activity after incubation with 3.4% (v/v) H_2O_2 at 30 °C for 30 min, and a protease from *Bacillus patagoniensis* was not affected by H_2O_2 (10% v/v) after incubation for 30 min at 25 °C (Olivera et al. 2006; Saeki et al. 2003).

With regard to tolerance and activity under saline conditions, it is interesting to look at the bacterial origin. *A. haloalkaliphilus*, the native host of SPAH, grows at salt concentrations between 1 and 20% with an optimum of 5%, while *P. marinus*, the origin of SPPM, grows between 1 and 9% NaCl with an optimum between 2 and 5% (Lim et al. 2005; Weisser and Trüper 1985). SPMI and SPLA derive from halotolerant bacteria, with *L. alkalitelluris* growing between 0 and 4% NaCl and *M. indicus* between 0 and 12% NaCl. Both species show an optimum between 0 and 1% NaCl (Lee et al. 2008; Yoon et al. 2005). Interestingly, the proteases are still active at NaCl concentrations at which the bacterial strains no longer grow. The highest activity was found for all proteases without NaCl, while for the two halophilic strains at least 1% NaCl is required for growth. The PIS enzyme SPAH maintained its activity after an initial loss of activity, whereas the three true subtilisins SPMI, SPPM and SPLA constantly lost activity. The activity of the previously investigated high-alkaline subtilisins is even induced by high salt concentrations (Falkenberg et al. 2022b). In contrast to the true subtilisins studied here, BPN′, which belongs to the same subgroup, showed higher activity with increasing NaCl levels (Falkenberg et al. 2022b). This is probably because BPN′ has five of the seven amino acid positions identified that are beneficial for salt adaptation

Table 2 Substrate specificities of the indicated proteases against 10 synthetic substrates (suc-XXXX-pNA)

Protease	Relative activity (%)									
	FAAF	AAA	AAVA	ALPF	AGPF	AAPF	TVAA	YVAD	AGPP	AAPL
SPPM	25	1	3	140	97	100 (190 U/mg)	1	2	127	78
SPMI	44	1	3	139	88	100 (223 U/mg)	1	1	146	68
SPLA	26	1	3	130	100	100 (93 U/mg)	1	2	173	73
SPAH	32	1	3	140	102	100 (159 U/mg)	1	1	157	67
SPAO[a]	753	12	21	127	107	100 (42 U/mg)	3	3	216	34
Subtilisin Carlsberg[a]	57	0	2	60	90	100 (570 U/mg)	1	1	147	104
Savinase[a]	605	8	22	117	96	100 (180 U/mg)	5	5	144	12
BPN′[a]	96	0	6	106	96	100 (181 U/mg)	0	0	61	67

Kinetic experiments were performed with 17 mM substrate at 30 °C in 0.1 M Tris-HCl buffer pH 8.6 containing 0.1% (w/v) Brij®35 for 5 min. The experiments were performed in triplicates and the standard deviation was <5%. Enzyme activity against AAPF was defined as 100% relative activity

[a]Falkenberg et al. (2022b)

(Takenaka et al. 2022). The finding that SPAH lost more activity at higher NaCl concentrations could be related to the constant activity at high salt concentrations and thus a higher autoproteolytic activity, which was also observed for subtilisin Carlsberg (Falkenberg et al. 2022b). No loss of activity at up to 5 M NaCl was also reported for a salt-tolerant and thermostable protease from *B. subtilis* (Kembhavi et al. 1993). Although the mechanism of halotolerance is not yet fully understood, a charged surface of the protein leads to increased hydration of the enzyme surface, which provides protection against aggregation at high salt concentrations (Mokashe et al. 2018; Takenaka et al. 2018). In addition to predicting adaptation to higher pH values by looking at the charge on the protein surface, this can also be used to predict salt tolerance. This is indicated by an acidic isoelectric point and the ratio of glutamate, aspartate to lysine, histidine and arginine (AB ratio) (Rhodes et al. 2010; Mokashe et al. 2018). All four proteases have an acidic pI and an AB ratio above 1.0 with a high number of Asp residues (Table S2). Whereas the surface charge of high-alkaline subtilisins is predominantly positive, it is predominantly negative for the four proteases of this study (Fig. 3). As mentioned above, salt adaptation is increased by a high number of negative or positive charges on the surface of the enzyme (Takenaka et al. 2022). The differences in the AB ratio of all residues seems to have no influence on the salt tolerance, as the AB ratio of SPPM (1.9) and SPAH (1.7) is quite high, but much lower for SPMI (1.0) and SPLA (1.1) (Fig. 3, Table S2).

Most subtilisins have broad substrate specificity and mainly have a nutritional role by supplying peptides and amino acids for cell growth (Siezen and Leunissen 1997). Variations in substrate specificity occur due to modulations of residues in the substrate-binding region, especially whose side chains interact with substrate residues P1 and P4, which dominate substrate preference in subtilisins (Siezen and Leunissen 1997; Grøn et al. 1992). The use of ten synthetic peptide 4-nitroanilide substrates with three or four amino acids makes it possible to compare the preference with others. As we have previously reported, the high-alkaline subtilisin SPAO and Savinase revealed the highest activity for suc-FAAF-pNA, while BPN′ preferred suc-ALPF-pNA and the subtilisin Carlsberg suc-AGPP-pNA, which is comparable to the preferences of the proteases from this work (Falkenberg et al. 2022b). Georgieva et al. showed for several subtilisins and proteinase K that they show lower activity when alanine, glutamate, lysine or valine are in position P1 nomenclature of Schechter and Berger (1967) (Georgieva et al. 2006). Based on the substrate specificity, SPPM, SPMI, SPLA and SPAH can be regarded as typical subtilisins.

In summary, we describe the production, purification and biochemical characterisation of the four extracellular subtilisin proteases SPPM, SPMI, SPLA and SPAH. The sequences were obtained from a data-mining search for new subtilisins from *Bacillaceae*, as this family has proven to be a valuable source of alkaline proteases with industrial applications. The genes were isolated from the two halophilic bacteria *P. marinus* DSM 16465^T and *A. haloalkaliphilus* DSM 5271^T and the two halotolerant species *M. indicus* DSM 16189 and *L. alkalitelluris* DSM 16976^T. The proteases showed high halotolerance up to 5 M NaCl and activity within a broad pH spectrum of pH 5.0–12.0 with an optimum between pH 9.0–10.0. The optimum temperature was found to be 50 and 55 °C for SPAH and SPPM and 70 °C for SPMI and SPLA. In addition, a high stability towards 5% (w/v) SDS and a good stability towards 5% (v/v) H_2O_2 were observed. With their biochemical properties, the four proteases show the potential for future biotechnological applications and that bacteria of halotolerant or halophilic origin are a promising source for novel enzymes.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00253-023-12553-w.

Author contributions FF, JB and PS conceived and designed the experiments. FF collected and analysed the data. LV cloned the gene for SPMI and did pre-experiments for SPMI. FF wrote the original draft; FF, JB, MB and PS revised the manuscript. All authors contributed to the final manuscript. All authors read and approved the manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. This work was supported by the FH Aachen University of Applied Sciences within the framework of the program for the promotion of young scientists.

Data availability The original contributions presented in this study are included in the article/Supplementary material; further inquiries can be directed to the corresponding author.

Declarations

Ethical approval This article does not contain any studies involving human or animal participants conducted by any of the authors.

Conflict of interest The authors declare no competing interests.

References

Alberto Cira-Chávez L, Guevara-Luna J, Yadira Soto-Padilla M, Román-Ponce B, Soledad Vásquez-Murrieta M, Isabel Estrada-Alvarado M (2019) Kinetics of halophilic enzymes. In: Fernandez C (ed) Rajendran L. Kinetics of enzymatic synthesis, IntechOpen, pp 3–26

Almog O, Gallagher T, Tordova M, Hoskins J, Bryan P, Gilliland GL (1998) Crystal structure of calcium-independent subtilisin BPN' with restored thermal stability folded without the prodomain. Proteins 31:21–32. https://doi.org/10.1002/(SICI)1097-0134(19980401)31:1<21:AID-PROT3>3.0.CO;2-K

Azrin NAM, Ali MSM, Rahman RNZRA, Oslan SN, Noor NDM (2022) Versatility of subtilisin: a review on structure, characteristics, and applications. Biotechnol Appl Biochem 69:2599–2616. https://doi.org/10.1002/bab.2309

Bernhard K, Schrempf H, Goebel W (1978) Bacteriocin and antibiotic resistance plasmids in *Bacillus cereus* and *Bacillus subtilis*. J Bacteriol 133:897–903. https://doi.org/10.1128/jb.133.2.897-903.1978

Betzel C, Klupsch S, Papendorf G, Hastrup S, Branner S, Wilson KS (1992) Crystal structure of the alkaline proteinase Savinase from *Bacillus lentus* at 1.4 A resolution. J Mol Biol 223:427–445. https://doi.org/10.1016/0022-2836(92)90662-4

Bhatt HB, Singh SP (2020) Cloning, expression, and structural elucidation of a biotechnologically potential alkaline serine protease from a newly isolated haloalkaliphilic *Bacillus lehensis* JO-26. Front Microbiol 11:941. https://doi.org/10.3389/fmicb.2020.00941

Bieth J, Spiess B, Wermuth CG (1974) The synthesis and analytical use of a highly sensitive and convenient substrate of elastase. Biochem Med 11:350–357. https://doi.org/10.1016/0006-2944(74)90134-3

Boivin S, Kozak S, Meijers R (2013) Optimization of protein purification and characterization using Thermofluor screens. Protein Expr Purif 91:192–206. https://doi.org/10.1016/j.pep.2013.08.002

Bott R, Ultsch M, Kossiakoff A, Graycar T, Katz B, Power S (1988) The three-dimensional structure of *Bacillus amyloliquefaciens* subtilisin at 1.8 A and an analysis of the structural consequences of peroxide inactivation. J Biol Chem 263:7895–7906. https://doi.org/10.1016/S0021-9258(18)68582-5

Bradford MM (1976) 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. https://doi.org/10.1006/abio.1976.9999

Brock FM, Forsberg CW, Buchanan-Smith JG (1982) Proteolytic activity of rumen microorganisms and effects of proteinase inhibitors. Appl Environ Microbiol 44:561–569

Bryan P, Pantoliano MW, Quill SG, Hsiao HY, Poulos T (1986) Site-directed mutagenesis and the role of the oxyanion hole in subtilisin. Proc Natl Acad Sci USA 83:3743–3745. https://doi.org/10.1073/pnas.83.11.3743

Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2016) GenBank. Nucleic Acids Res 44:D67–D72. https://doi.org/10.1093/nar/gkv1276

Coker JA (2016) Extremophiles and biotechnology: current uses and prospects. F1000Res 5:396. https://doi.org/10.12688/f1000research.7432.1

Contesini FJ, Melo RR, Sato HH (2017) An overview of *Bacillus* proteases: from production to application. Crit Rev Biotechnol 38:321–334. https://doi.org/10.1080/07388551.2017.1354354

DelMar EG, Largman C, Brodrick JW, Geokas MC (1979) A sensitive new substrate for chymotrypsin. Anal Biochem 99:316–320. https://doi.org/10.1016/S0003-2697(79)80013-5

Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard J-F, Guindon S, Lefort V, Lescot M, Claverie J-M, Gascuel O (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res 36:W465–W469. https://doi.org/10.1093/nar/gkn180

Dodia MS, Rawal CM, Bhimani HG, Joshi RH, Khare SK, Singh SP (2008) Purification and stability characteristics of an alkaline serine protease from a newly isolated Haloalkaliphilic bacterium sp. AH-6. J Ind Microbiol Biotechnol 35:121–131. https://doi.org/10.1007/s10295-007-0273-x

Engler C, Kandzia R, Marillonnet S (2008) A one pot, one step, precision cloning method with high throughput capability. PLoS One 3:e3647. https://doi.org/10.1371/journal.pone.0003647

Engqvist MKM (2018) Correlating enzyme annotations with a large set of microbial growth temperatures reveals metabolic adaptations to growth at diverse temperatures. BMC Microbiol 18:177. https://doi.org/10.1186/s12866-018-1320-7

Falkenberg F, Bott M, Bongaerts J, Siegert P (2022a) Phylogenetic survey of the subtilase family and a data-mining-based search for new subtilisins from *Bacillaceae*. Manuscript submitted for publication. Front Microbiol 13:1017978. https://doi.org/10.3389/fmicb.2022.1017978

Falkenberg F, Rahba J, Fischer D, Bott M, Bongaerts J, Siegert P (2022b) Biochemical characterization of a novel oxidatively stable, halotolerant, and high-alkaline subtilisin from *Alkalihalobacillus okhensis* Kh10-101^{T}. FEBS Open Bio 12:1729–1746. https://doi.org/10.1002/2211-5463.13457

Ferrer M, Golyshina O, Beloqui A, Golyshin PN (2007) Mining enzymes from extreme environments. Curr Opin Microbiol 10:207–214. https://doi.org/10.1016/j.mib.2007.05.004

Fritze D (1996) *Bacillus haloalkaliphilus* sp. nov. Int J Syst Evol Microbiol 46:98–101. https://doi.org/10.1099/00207713-46-1-98

Gavrilov Y, Dagan S, Levy Y (2015) Shortening a loop can increase protein native state entropy. Proteins 83:2137–2146. https://doi.org/10.1002/prot.24926

Georgieva D, Genov N, Voelter W, Betzel C (2006) Catalytic efficiencies of alkaline proteinases from microorganisms. Z Naturforsch C J Biosci 61:445–452. https://doi.org/10.1515/znc-2006-5-623

Georgieva DN, Stoeva S, Voelter W, Genov N, Betzel C (2001) Differences in the specificities of the highly alkalophilic proteinases Savinase and Esperase imposed by changes in the rigidity and geometry of the substrate binding sites. Arch Biochem Biophys 387:197–201. https://doi.org/10.1006/abbi.2000.2249

Grøn H, Meldal M, Breddam K (1992) Extensive comparison of the substrate preferences of two subtilisins as determined with peptide substrates which are based on the principle of intramolecular quenching. Biochem 31:6011–6018. https://doi.org/10.1021/bi00141a008

Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18:2714–2723. https://doi.org/10.1002/elps.1150181505

Gupta R, Beg QK, Lorenz P (2002) Bacterial alkaline proteases: molecular approaches and industrial applications. Appl Microbiol Biotechnol 59:15–32. https://doi.org/10.1007/s00253-002-0975-y

Gupta RS, Patel S, Saini N, Chen S (2020) Robust demarcation of 17 distinct *Bacillus* species clades, proposed as novel *Bacillaceae* genera, by phylogenomics and comparative genomic analyses: description of *Robertmurraya kyonggiensis* sp. nov. and proposal for an emended genus *Bacillus* limiting it only to the members of the Subtilis and Cereus clades of species. Int J Syst Evol Microbiol 70:5753–5798. https://doi.org/10.1099/ijsem.0.004475

Gurunathan R, Huang B, Ponnusamy VK, Hwang J-S, Dahms H-U (2021) Novel recombinant keratin degrading subtilisin like serine alkaline protease from *Bacillus cereus* isolated from marine hydrothermal vent crabs. Sci Rep 11:12007. https://doi.org/10.1038/s41598-021-90375-4

Han X-Q, Damodaran S (1998) Purification and characterization of protease Q: a detergent- and urea-stable serine endopeptidase from

Bacillus pumilus. J Agric Food Chem 46:3596–3603. https://doi.org/10.1021/jf980361m

Jeon CO, Lim J-M, Lee J-M, Xu L-H, Jiang C-L, Kim C-J (2005) Reclassification of *Bacillus haloalkaliphilus* Fritze 1996 as *Alkalibacillus haloalkaliphilus* gen. nov., comb. nov. and the description of *Alkalibacillus salilacus* sp. nov., a novel halophilic bacterium isolated from a salt lake in China. Int J Syst Evol Microbiol 55:1891–1896. https://doi.org/10.1099/ijs.0.63456-0

Joshi S, Satyanarayana T (2013) Characteristics and applications of a recombinant alkaline serine protease from a novel bacterium *Bacillus lehensis*. Bioresour Technol 131:76–85. https://doi.org/10.1016/j.biortech.2012.12.124

Kalisz HM (1988) Microbial proteinases. Adv Biochem Eng Biotechnol 36:1–65. https://doi.org/10.1007/bfb0047944

Kawamura F, Doi RH (1984) Construction of a *Bacillus subtilis* double mutant deficient in extracellular alkaline and neutral proteases. J Bacteriol 160:442–444

Kembhavi AA, Kulkarni A, Pant A (1993) Salt-tolerant and thermostable alkaline protease from *Bacillus subtilis* NCIM no. 64. Appl Biochem Biotechnol 38:83–92. https://doi.org/10.1007/BF02916414

Kobayashi T, Lu J, Li Z, Hung VS, Kurata A, Hatada Y, Takai K, Ito S, Horikoshi K (2007) Extremely high alkaline protease from a deep-subsurface bacterium, *Alkaliphilus transvaalensis*. Appl Microbiol Biotechnol 75:71–80. https://doi.org/10.1007/s00253-006-0800-0

Le Duc H, Fraser PD, Tam NKM, Cutting SM (2006) Carotenoids present in halotolerant *Bacillus* spore formers. FEMS Microbiol Lett 255:215–224. https://doi.org/10.1111/j.1574-6968.2005.00091.x

Lee J-C, Lee GS, Park D-J, Kim C-J (2008) *Bacillus alkalitelluris* sp. nov., an alkaliphilic bacterium isolated from sandy soil. Int J Syst Evol Microbiol 58:2629–2634. https://doi.org/10.1099/ijs.0.65733-0

Letunic I, Bork P (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res 44:W242–W245. https://doi.org/10.1093/nar/gkw290

Lim J-M, Jeon CO, Park D-J, Kim H-R, Yoon B-J, Kim C-J (2005) *Pontibacillus marinus* sp. nov., a moderately halophilic bacterium from a solar saltern, and emended description of the genus *Pontibacillus*. Int J Syst Evol Microbiol 55:1027–1031. https://doi.org/10.1099/ijs.0.63489-0

Masui A, Fujiwara N, Yamamoto K, Takagi M, Imanaka T (1998) Rational design for stabilization and optimum pH shift of serine protease AprN. J Ferment Bioeng 85:30–36. https://doi.org/10.1016/S0922-338X(97)80349-2

Matagne A, Joris B, Frère JM (1991) Anomalous behaviour of a protein during SDS/PAGE corrected by chemical modification of carboxylic groups. Biochem J 280(Pt 2):553–556. https://doi.org/10.1042/bj2800553

Matsubara H, Kasper CB, Brown DM, Smith EL (1965) Subtilisin BPN′. J Biol Chem 240:1125–1130. https://doi.org/10.1016/S0021-9258(18)97548-4

Miller AJ, Roman B, Norstrom E (2016) A method for easily customizable gradient gel electrophoresis. Anal Biochem 509:12–14. https://doi.org/10.1016/j.ab.2016.07.003

Mokashe N, Chaudhari B, Patil U (2018) Operative utility of salt-stable proteases of halophilic and halotolerant bacteria in the biotechnology sector. Int J Biol Macromol 117:493–522. https://doi.org/10.1016/j.ijbiomac.2018.05.217

Naveed M, Nadeem F, Mehmood T, Bilal M, Anwar Z, Amjad F (2021) Protease—a versatile and ecofriendly biocatalyst with multi-industrial applications: an updated review. Catal Lett 151:307–323. https://doi.org/10.1007/s10562-020-03316-7

North MJ (1982) Comparative biochemistry of the proteinases of eucaryotic microorganisms. Microbiol Rev 46:308–340. https://doi.org/10.1128/MMBR.46.3.308-340.1982

Olivera N, Sequeiros C, Siñeriz F, Breccia JD (2006) Characterisation of alkaline proteases from a novel alkali-tolerant bacterium *Bacillus patagoniensis*. World J Microbiol Biotechnol 22:737–743. https://doi.org/10.1007/s11274-005-9099-8

Panja AS, Maiti S, Bandyopadhyay B (2020) Protein stability governed by its structural plasticity is inferred by physicochemical factors and salt bridges. Sci Rep 10:1822. https://doi.org/10.1038/s41598-020-58825-7

Patel S, Gupta RS (2020) A phylogenomic and comparative genomic framework for resolving the polyphyly of the genus *Bacillus*: proposal for six new genera of *Bacillus* species, *Peribacillus* gen. nov., *Cytobacillus* gen. nov., *Mesobacillus* gen. nov., *Neobacillus* gen. nov., *Metabacillus* gen. nov. and *Alkalihalobacillus* gen. nov. Int J Syst Evol Microbiol 70:406–438. https://doi.org/10.1099/ijsem.0.003775

Phrommao E, Yongsawatdigul J, Rodtong S, Yamabhai M (2011) A novel subtilase with NaCl-activated and oxidant-stable activity from *Virgibacillus* sp. SK37. BMC Biotechnol 11:65. https://doi.org/10.1186/1472-6750-11-65

Rath A, Glibowicka M, Nadeau VG, Chen G, Deber CM (2009) Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. Proc Natl Acad Sci USA 106:1760–1765. https://doi.org/10.1073/pnas.0813167106

Rawlings ND (2013) Identification and prioritization of novel uncharacterized peptidases for biochemical characterization. J Biol Databases 2013:bat022. https://doi.org/10.1093/database/bat022

Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A, Finn RD (2018) The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. Nucleic Acids Res 46:D624–D632. https://doi.org/10.1093/nar/gkx1134

Rekik H, Zaraî Jaouadi N, Gargouri F, Bejar W, Frikha F, Jmal N, Bejar S, Jaouadi B (2019) Production, purification and biochemical characterization of a novel detergent-stable serine alkaline protease from *Bacillus safensis* strain RH12. Int J Biol Macromol 121:1227–1239. https://doi.org/10.1016/j.ijbiomac.2018.10.139

Rhodes ME, Fitz-Gibbon ST, Oren A, House CH (2010) Amino acid signatures of salinity on an environmental scale with a focus on the Dead Sea. Environ Microbiol 12:2613–2623. https://doi.org/10.1111/j.1462-2920.2010.02232.x

Robert X, Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res 42:W320–W324. https://doi.org/10.1093/nar/gku316

Rocha EP, Danchin A, Viari A (1999) Translation in *Bacillus subtilis*: roles and trends of initiation and termination, insights from a genome analysis. Nucleic Acids Res 27:3567–3576. https://doi.org/10.1093/nar/27.17.3567

Roy A, Yang J, Zhang Y (2012) COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. Nucleic Acids Res 40:W471–W477. https://doi.org/10.1093/nar/gks372

Saeki K, Magallones MV, Takimura Y, Hatada Y, Kobayashi T, Kawai S, Ito S (2003) Nucleotide and deduced amino acid sequences of a new subtilisin from an alkaliphilic *Bacillus* isolate. Curr Microbiol 47:337–340. https://doi.org/10.1007/s00284-002-4018-9

Salwan R, Sharma V (2019) Trends in extracellular serine proteases of bacteria as detergent bioadditive: alternate and environmental friendly tool for detergent industry. Arch Microbiol 201:863–877. https://doi.org/10.1007/s00203-019-01662-8

Sayers EW, Beck J, Bolton EE, Bourexis D, Brister JR, Canese K, Comeau DC, Funk K, Kim S, Klimke W, Marchler-Bauer A, Landrum M, Lathrop S, Lu Z, Madden TL, O'Leary N, Phan L, Rangwala SH, Schneider VA et al (2021) Database resources of

the National Center for Biotechnology Information. Nucleic Acids Res 49:D10–D17. https://doi.org/10.1093/nar/gkaa892

Schechter I, Berger A (1967) On the size of the active site in proteases. I Papain Biochem Biophys Res Commun 27:157–162. https://doi.org/10.1016/s0006-291x(67)80055-x

Sehnal D, Bittrich S, Deshpande M, Svobodová R, Berka K, Bazgier V, Velankar S, Burley SK, Koča J, Rose AS (2021) Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. Nucleic Acids Res 49:W431–W437. https://doi.org/10.1093/nar/gkab314

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539. https://doi.org/10.1038/msb.2011.75

Siezen RJ, de Vos WM, Leunissen JA, Dijkstra BW (1991) Homology modelling and protein engineering strategy of subtilases, the family of subtilisin-like serine proteinases. Protein Eng 4:719–737. https://doi.org/10.1093/protein/4.7.719

Siezen RJ, Leunissen JA (1997) Subtilases: the superfamily of subtilisin-like serine proteases. Protein Sci 6:501–523. https://doi.org/10.1002/pro.5560060301

Siezen RJ, Renckens B, Boekhorst J (2007) Evolution of prokaryotic subtilases: genome-wide analysis reveals novel subfamilies with different catalytic residues. Proteins 67:681–694. https://doi.org/10.1002/prot.21290

Smith EL, Markland FS, Kasper CB, DeLange RJ, Landon M, Evans WH (1966) The complete amino acid sequence of two types of subtilisin, BPN' and Carlsberg. J Biol Chem 241:5974–5976. https://doi.org/10.1016/S0021-9258(18)96365-9

Solanki P, Putatunda C, Kumar A, Bhatia R, Walia A (2021) Microbial proteases: ubiquitous enzymes with innumerable uses. 3 Biotech 11:428. https://doi.org/10.1007/s13205-021-02928-z

Stauffer CE, Etson D (1969) The effect on subtilisin activity of oxidising a methionine residue. J Biol Chem 244:5333–5338. https://doi.org/10.1016/S0021-9258(18)63664-6

Stropko SJ, Pipes SE, Newman JD (2014) Genome-based reclassification of *Bacillus cibi* as a later heterotypic synonym of *Bacillus indicus* and emended description of *Bacillus indicus*. Int J Syst Evol Microbiol 64:3804–3809. https://doi.org/10.1099/ijs.0.068205-0

Takenaka S, Takada A, Kimura Y, Watanabe M, Kuntiya A (2022) Improvement of the halotolerance of a *Bacillus* serine protease by protein surface engineering. J Basic Microbiol 62:174–184. https://doi.org/10.1002/jobm.202100335

Takenaka S, Yoshinami J, Kuntiya A, Techapun C, Leksawasdi N, Seesuriyachan P, Chaiyaso T, Watanabe M, Tanaka K, Yoshida K-I (2018) Characterization and mutation analysis of a halotolerant serine protease from a new isolate of *Bacillus subtilis*. Biotechnol Lett 40:189–196. https://doi.org/10.1007/s10529-017-2459-2

Tehei M, Zaccai G (2005) Adaptation to extreme environments: macromolecular dynamics in complex systems. Biochim Biophys Acta 1724:404–410. https://doi.org/10.1016/j.bbagen.2005.05.007

Tekin A, Uzuner U, Sezen K (2021) Homology modeling and heterologous expression of highly alkaline subtilisin-like serine protease from *Bacillus halodurans* C-125. Biotechnol Lett 43:479–494. https://doi.org/10.1007/s10529-020-03025-6

Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, Winther O, Brunak S, Heijne G von, Nielsen H (2022) SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol, 40:1023–1025. https://doi.org/10.1038/s41587-021-01156-3

Thaz CJ, Jayaraman G (2014) Stability and detergent compatibility of a predominantly β-sheet serine protease from halotolerant *B. aquimaris* VITP4 strain. Appl Biochem Biotechnol 172:687–700. https://doi.org/10.1007/s12010-013-0524-4

The UniProt Consortium (2021) UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res 49:D480–D489. https://doi.org/10.1093/nar/gkaa1100

Thebti W, Riahi Y, Belhadj O (2016) Purification and characterization of a new thermostable, haloalkaline, solvent stable, and detergent compatible serine protease from *Geobacillus toebii* strain LBT 77. Biomed Res Int 2016:9178962. https://doi.org/10.1155/2016/9178962

Vidyasagar M, Prakash S, Mahajan V, Shouche YS, Sreeramulu K (2009) Purification and characterization of an extreme halothermophilic protease from a halophilic bacterium *Chromohalobacter sp.* TVSP101. Braz J Microbiol 40:12–19. https://doi.org/10.1590/S1517-83822009000100002

Vojcic L, Despotovic D, Martinez R, Maurer K-H, Schwaneberg U (2012) An efficient transformation method for *Bacillus subtilis* DB104. Appl Microbiol Biotechnol 94:487–493. https://doi.org/10.1007/s00253-012-3987-2

Weisser J, Trüper HG (1985) Osmoregulation in a new haloalkaliphilic *Bacillus* from the Wadi Natrun (Egypt). Syst Appl Microbiol 6:7–11. https://doi.org/10.1016/S0723-2020(85)80003-5

Wells JA, Powers DB, Bott RR, Graycar TP, Estell DA (1987) Designing substrate specificity by protein engineering of electrostatic interactions. Proc Natl Acad Sci USA 84:1219–1223. https://doi.org/10.1073/pnas.84.5.1219

Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, Hochstrasser DF (1999) Protein identification and analysis tools in the ExPASy server. Methods Mol Biol 112:531–552. https://doi.org/10.1385/1-59259-584-7:531

Yamagata Y, Isshiki K, Ichishima E (1995) Subtilisin Sendai from alkalophilic *Bacillus* sp.: molecular and enzymatic properties of the enzyme and molecular cloning and characterization of the gene, aprS. Enzyme Microb Technol 17:653–663. https://doi.org/10.1016/0141-0229(94)00109-5

Yang J, Roy A, Zhang Y (2013) Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. Bioinformatics 29:2588–2595. https://doi.org/10.1093/bioinformatics/btt447

Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y (2015) The I-TASSER Suite: protein structure and function prediction. Nat Methods 12:7–8. https://doi.org/10.1038/nmeth.3213

Yoon J-H, Lee C-H, Oh T-K (2005) *Bacillus cibi* sp. nov., isolated from jeotgal, a traditional Korean fermented seafood. Int J Syst Evol Microbiol 55:733–736. https://doi.org/10.1099/ijs.0.63208-0

Zhang Y (2008) I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9:40. https://doi.org/10.1186/1471-2105-9-40

Zhang Y, Skolnick J (2004) Scoring function for automated assessment of protein structure template quality. Proteins 57:702–710. https://doi.org/10.1002/prot.20264

2.4 Biochemical characterisation of a novel broad pH spectrum subtilisin from *Fictibacillus arsenicus* DSM 15822^T

Falkenberg F, Kohn S, Bott M, Bongaerts J, Siegert P (2023). Biochemical characterisation of a novel broad pH spectrum subtilisin from *Fictibacillus arsenicus* DSM 15822^T. FEBS Open Bio, DOI: 10.1002/2211-5463.13701

Authors contributions:

FF, JB, and PS conceived and designed the experiments.

FF conducted the experiments except those of SK, collected and analysed the data.

SK purified SPFA and did the MALDI-TOF MS experiment.

FF wrote the original draft.

FF, JB, MB, and PS revised the manuscript.

All figures and tables are from this study.

Overall contribution FF: 90 %

FEBS openbio

FEBS PRESS science publishing by scientists

RESEARCH ARTICLE

Biochemical characterisation of a novel broad pH spectrum subtilisin from *Fictibacillus arsenicus* DSM 15822^T

Fabian Falkenberg[1], Sophie Kohn[1], Michael Bott[2], Johannes Bongaerts[1] and Petra Siegert[1]

1 Institute of Nano- and Biotechnologies, Aachen University of Applied Sciences, Jülich, Germany
2 Institute of Bio- and Geosciences, IBG-1: Biotechnology, Forschungszentrum Jülich, Germany

Keywords
Bacillaceae; biotechnological application; broad pH spectrum; subtilases; subtilisin

Correspondence
P. Siegert, Institute of Nano- and Biotechnologies, Aachen University of Applied Sciences, Jülich 52428, Germany
Tel: +49 241 6009 53124
E-mail: siegert@fh-aachen.de
website: https://www.fh-aachen.de/forschung/institute/inb/

(Received 25 April 2023, revised 23 August 2023, accepted 29 August 2023)

doi:10.1002/2211-5463.13701

Subtilisins from microbial sources, especially from the *Bacillaceae* family, are of particular interest for biotechnological applications and serve the currently growing enzyme market as efficient and novel biocatalysts. Biotechnological applications include use in detergents, cosmetics, leather processing, wastewater treatment and pharmaceuticals. To identify a possible candidate for the enzyme market, here we cloned the gene of the subtilisin SPFA from *Fictibacillus arsenicus* DSM 15822^T (obtained through a data mining-based search) and expressed it in *Bacillus subtilis* DB104. After production and purification, the protease showed a molecular mass of 27.57 kDa and a pI of 5.8. SPFA displayed hydrolytic activity at a temperature optimum of 80 °C and a very broad pH optimum between 8.5 and 11.5, with high activity up to pH 12.5. SPFA displayed no NaCl dependence but a high NaCl tolerance, with decreasing activity up to concentrations of 5 M NaCl. The stability enhanced with increasing NaCl concentration. Based on its substrate preference for 10 synthetic peptide 4-nitroanilide substrates with three or four amino acids and its phylogenetic classification, SPFA can be assigned to the subgroup of true subtilisins. Moreover, SPFA exhibited high tolerance to 5% (w/v) SDS and 5% H_2O_2 (v/v). The biochemical properties of SPFA, especially its tolerance of remarkably high pH, SDS and H_2O_2, suggest it has potential for biotechnological applications.

Hydrolytic enzymes produced by microorganisms play a crucial role in a variety of industrial applications due to their commercial value. Here, the largest market share is accounted for by subtilisins or alkaline proteases from microbial sources, particularly from the *Bacillaceae* family [1,2]. Subtilisins are a group of subtilases classified in the MEROPS database as S8, one of the largest families of serine peptidases [3]. Subtilisins can be further structured into true subtilisins, phylogenetically intermediate subtilisins (PIS), high-alkaline subtilisins and intracellular subtilisins, among others [4–6]. In Bacilli and related bacteria, intracellular subtilisins form the main component of the degradome, while extracellular subtilisins primarily contribute to nutritional

Abbreviations
aa, amino acid; *aprE*, extracellular alkaline protease gene; bp, base pairs; EDTA, ethylenediaminetetraacetic acid; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; IEF, isoelectric focussing; LB, lysogeny broth; LM, length marker; MALDI-TOF-MS, matrix-assisted laser desorption ionisation—time-of-flight mass spectrometry; MSA, multiple sequence alignment; PAGE, polyacrylamide gel electrophoresis; PCR, polymerase chain reaction; PDB, protein data bank; pI, isoelectric point; PIS, phylogenetically intermediate subtilisins; PMSF, phenylmethylsulfonyl fluoride; pNA, para-nitroanilide; SDS, sodium dodecyl sulphate; SPFA, subtilisin protease *F. arsenicus*; suc, N-succinyl; T_m, melting temperature.

provisioning due to their broad substrate specificity [4,7,8]. In particular, the genus *Bacillus* has been recognised as a valuable source of alkaline proteases such as subtilisin Carlsberg, Savinase, BPN' and PB92, which are mainly used as active ingredients in detergents due to their properties, such as good performance against proteinaceous stains, high stability towards components of the detergent matrix and high operating temperatures [9,10]. In addition, they are used in industrial applications such as food processing, wastewater treatment and cosmetics instead of aggressive chemicals [9,11,12]. Because of their industrial importance, subtilisins have been intensively studied for their biological function in order to understand the mechanism of catalysis and the structure–function relationship [13]. An extracellular subtilisin typically has a molecular weight of about 27 kDa, consists of a signal peptide, propeptide and a peptidase domain and is secreted as a precursor of about 39 kDa [14–16].

Nature holds with its almost unlimited microbial biodiversity great potential for enriching the repertoire of known and new enzymes [17]. However, the classical approach of isolation and cultivation of microorganisms from a specific environment with the extraction of genomic DNA or the purification of proteins and gene identification is a time and labour-intensive procedure [18]. The increasing number of genomic or metagenomic sequencing projects, which provide an exponentially growing amount of data on uncharacterised proteins due to automatic annotation, offers an alternative source for finding new protease genes for industrial applications [19]. We recently analysed this unused information and conducted a search for new uncharacterised subtilisins from the *Bacillaceae* family [6]. These sequences were phylogenetically categorised into the subgroups true subtilisins, PIS and high-alkaline subtilisins. Here, we selected a sequence from the phylogenetic tree of true subtilisins obtained from *Fictibacillus arsenicus* DSM 15822^T for biochemical characterisation, as the sequence originates from a yet unexplored sequence section. *F. arsenicus* is a Gram-positive, endospore-forming and arsenic-resistant bacterium, which was isolated from a concretion of arsenic ore obtained from a borehole in India by Shivaji *et al.* [20]. The strain grows optimally at a temperature of 30 °C and a pH of 7.0, while also tolerating up to 1.0% NaCl. The bacterium was initially classified as *Bacillus arsenicus* in 2005 and then reclassified as *Fictibacillus arsenicus* [21].

Research into more robust subtilisins with polyextremotolerant properties is not only scientifically crucial to better understand the mechanism of enzyme adaptation but is also of huge practical importance for the development of proteolytic biocatalysts with enhanced versatility in relation to different extreme conditions that are often encountered in industrial applications. Here, we report on the gene for subtilisin WP_077360649 from *Fictibacillus arsenicus* DSM 15822^T, which was cloned, overexpressed in *B. subtilis* DB104 and purified by ion-exchange chromatography. This is the first report on the biochemical characterisation of a recombinant subtilisin protease from *F. arsenicus* (SPFA) and for the first time from the genus *Fictibacillus*.

Material and methods

Bacterial strains and growth conditions

Fictibacillus arsenicus DSM 15822^T was purchased from the German Collection of Microorganisms and Cell Cultures GmbH (DSMZ) and cultivated in medium 220 with 10 mg·L^{-1} $MnSO_4$ at a temperature of 30 °C as recommended by the DSMZ. The InnuSPEED Bacteria/Fungi DNA Kit (Analytik Jena™, Jena, Germany) was used for genomic DNA preparation from an overnight culture. *Bacillus subtilis* DB104 was used for cloning and protein production and cultivated in LB medium (Luria/Miller; Carl Roth, Karlsruhe, Germany) at 37 °C [22].

Plasmid construction/cloning and bioinformatic analysis

For recombinant protease production with *B. subtilis* DB104, pFF-RED, a pBC16-based expression plasmid (Acct. No. U32369.1) [23], was used as previously described [24]. The DNA sequence encoding the gene for SPFA, including the signal peptide, the propeptide and the peptidase domain, was amplified from the genomic DNA in a polymerase chain reaction (PCR) using the Phusion® Hot Start II high-fidelity polymerase (Thermo Fisher Scientific GmbH, Karlsruhe, Germany) according to the manufacturer's recommendations. The NCBI reference sequence NZ_MQMF01000001.1 was used to design primers for the *aprE* gene (extracellular alkaline protease) of WP_077360649 [25,26]. The oligonucleotides for PCR were obtained from Eurofins Genomics GmbH (Ebersberg, Germany; Table S1). Cloning and transformation were performed as previously described [27]. Bioinformatic analysis of the amino acid sequence, homology modelling and multiple sequence alignment (MSA) as previously reported [27].

Recombinant protease production and purification

A 1-L scale DASGIP® parallel reactor system (DASGIP, Jülich, Germany) was used for the production of SPFA by

Bacillus subtilis DB104 as previously described [24]. Successful production was confirmed by SDS/PAGE (sodium dodecyl sulphate–polyacrylamide gel electrophoresis) and by a proteolytic activity assay using suc-AAPF-pNA (*N*-succinyl-Ala-Ala-Pro-Phe-p-nitroanilide) and azocasein as substrate. Protease purification was performed as previously detailed [24]. For SPFA, an anion exchanger (25 mL Q-Sepharose FF, GE Healthcare, Chicago, IL, USA) and a pH of 8.0 for running (10 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)-NaOH buffer) and elution buffer (10 mM HEPES-NaOH, 1 M NaCl) was used.

Enzyme activity assay

The determination of hydrolytic activity was performed by using the tetrapeptide substrate suc-AAPF-pNA (BACHEM, Bubendorf, Switzerland) at 30 °C in 100 mM Tris–HCl buffer, pH 8.6, containing 0.1% (w/v) Brij®35 as described elsewhere [24,28]. Azocasein (Sigma-Aldrich, Schnellendorf, Germany) was used as a more complex substrate at 37 °C in 100 mM Tris–HCl buffer, pH 8.6, as previously described [24,29]. In addition, various synthetic 4-nitroanilide substrates were used to determine the substrate specificity of SPFA as previously reported [24].

Protein electrophoresis, measurement and analytical methods

Protein concentration was quantified using Roti®Nanoquant (Carl Roth) with bovine serum albumin fraction V (Carl Roth) as reference according to the manufacturer's recommendations. The molecular mass of the purified SPFA was analysed by SDS/PAGE and matrix-assisted laser desorption ionisation time-of-flight mass spectrometry (MALDI-TOF-MS), as previously described [24]. For the determination of the isoelectric point (pI), purified SPFA was rebuffered in 10 mM HEPES-NaOH pH 8.0 using centrifugal spin columns (VWR, Radnor, PA, USA) with a molecular mass cut-off of 3 kDa. Isoelectric focussing–polyacrylamide gel electrophoresis (IEF-PAGE) was performed using an IEF SERVALYT™ PRECOTES™ 3–10 gel (SERVA, Heidelberg, Germany) according to the manufacturer's recommendations.

Effects of pH and temperature

The pH optimum was determined in 0.1 M Tris-maleate buffer (pH 5.0–7.0), 0.1 M Tris–HCl (pH 7.0–9.0) and 0.1 M glycine-NaOH (pH 9.0–12.5) at 30 °C using the suc-AAPF-pNA assay. The stability of SPFA at different pH levels was assayed by preincubating SPFA in said buffers for 24 h at 4 °C and residual activities were measured under standard reaction conditions for the suc-AAPF-pNA assay. For the determination of the melting point, the fluorescent dye SYPRO™ Orange (Thermo Fisher Scientific GmbH) was used in the thermal shift assay as previously described [24]. Temperature stability was investigated by measuring the residual activity after incubation of SPFA in 10 mM HEPES-NaOH, pH 8.0, at 20 and 50 °C for 3 h, as previously detailed [24].

Effects of different additives

To test the impact of additives, 1 and 5% (w/v) SDS and 1 and 5% (v/v) H_2O_2 were added to the enzyme solution and incubated for 1 h at 10 °C in 10 mM HEPES-NaOH, pH 8.0. The inhibition of SPFA with PMSF was investigated by incubating SPFA in 10 mM HEPES-NaOH with 1 mM PMSF, pH 8.0, for 30 min on ice. Standard suc-AAPF-pNA activity assay was used to determine the residual activity. To investigate the influence of NaCl (0–5 M) on SPFA, hydrolytic activity was measured using the above-described suc-AAPF-pNA assay with the addition of NaCl in the reaction buffer as previously reported [24]. In addition, stability at different NaCl concentrations (0–5 M) was investigated by incubating SPFA in 10 mM HEPES-NaOH, pH 8.0, at 20 °C for 2 h. The effect of ethylenediaminetetraacetic acid (EDTA) was investigated by incubating the protease for 12 h at 4 °C with 20 mM EDTA as previously described [24]. The influence of different metal ions was investigated by incubating the protease with 1 mM each of $MgSO_4$, $ZnCl_2$, $MnCl_2$, $CaCl_2$, $CoCl_2$, $NiSO_4$, $FeSO_4$ and $CuSO_4$ in 10 mM HEPES for 1 h at room temperature. Standard suc-AAPF-pNA activity assay was used to determine the residual activity.

Results

Cloning and expression of *aprE_F. arsenicus* in *B. subtilis* DB104

The uncharacterised protease SPFA was chosen from our previous report on a data mining approach screen for new subtilisins from *Bacillaceae*, as it displays a more distant sequence homology to known subtilisins [6]. The coding sequence of aprE_*F. arsenicus* for the protease SPFA, including the signal peptide, the propeptide and the peptidase domain, was amplified, and a fragment of 1147 bp was obtained. The PCR product was cloned into the vector pFF-RED. The transformation of *B. subtilis* DB014 was verified by the emerging clearing zones on LB agar plates with 2.5% (w/v) skim milk around the colonies and analysed by plasmid preparation and restriction analysis. Sequence data obtained by Sanger sequencing confirmed that it was identical to the genomic nucleotide sequence of the *aprE_F. arsenicus* gene available in GenBank.

Bioinformatic analysis

The *aprE* gene from *Fictibacillus arsenicus* DSM 15822^T comprises 1140 bp encoding a protein of 380 amino acids (aa). The signal peptide prediction revealed the presence of a Sec signal peptide with a probability above 99%. Using MSAs, the propeptide was identified and is labelled in Fig. 1. An S8 peptidase domain containing 275 amino acids, a 32-aa signal peptide and a 73-aa propeptide were identified. The *in silico* analysis of mature SPFA showed a molecular mass of 27.57 kDa and a pI of 5.8. The catalytic triad consists of Asp^{32}, His^{64} and Ser^{221} (numbers

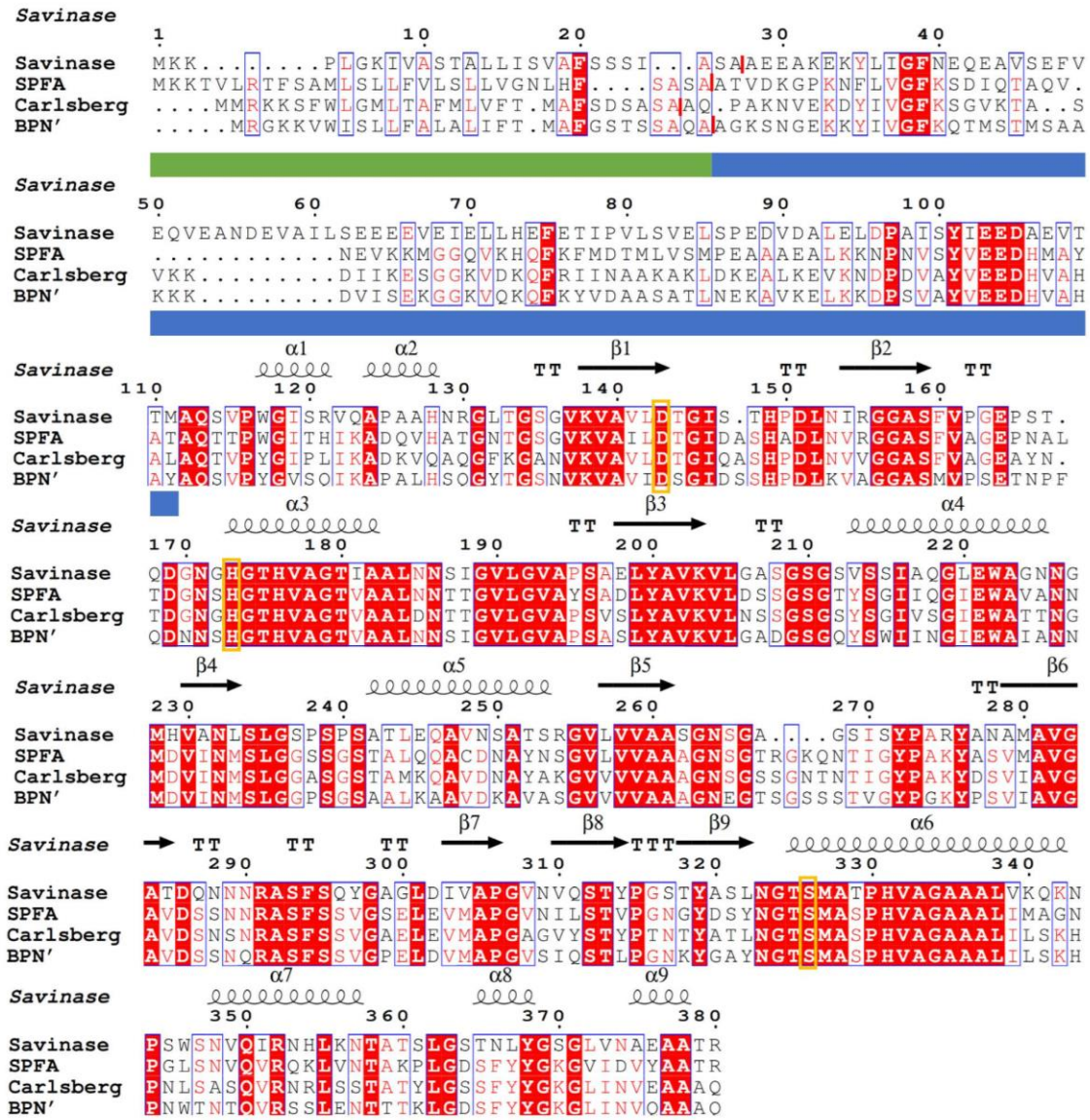


Fig. 1. Multiple sequence alignment (MSA) of SPFA with Savinase (WP_094423791.1), subtilisin Carlsberg (WP_020450819.1) and BPN' (WP_013351733.1). CLUSTAL OMEGA was used for the alignment [59]. ESPript 3.0 with Savinase (PDB: 1C9J) as a template was used to analyse the MSA [60]. Signal peptide sequence (green bars); propeptide (blue bars) of SPFA. Red bars indicate individual signal peptide cleavage sites. Secondary structure elements: helices with squiggles, β-strands with arrows and turns with TT letters. The catalytic triad (Asp^{143}, His^{173}, Ser^{326}; Savinase numbering) is marked with orange boxes.

based on the mature protease sequence). The analysis of our previously reported phylogenetic tree with well-characterised proteases of the three subtilisin families (true, high-alkaline, phylogenetically intermediate) retrieved from MEROPS [3] and the UniProt database [30], as well as data mined sequences, demonstrates that SPFA is clearly a member of the true subtilisin subgroup [6].

A homology modelling-based structural analysis with the I-TASSER server was performed with the peptidase unit of SPFA, yielding a model with a C-score of 1.55 (Fig. S1). The C-score ranges from −5 to 2 and higher values indicate higher confidence of the model [31]. A high template modelling (TM) score is given for BPN' (PDB: 1S01) of 0.997, where a TM value of 1 specifies a perfect correlation for two structures [32]. The metal-binding prediction suggested two Ca^{2+}-binding sites involving the side chains Gln^2 and Asp^{41} and several side chains of the loop-forming residues 75–81 for the first Ca^{2+}-binding site. The side chains of Ala^{169}, Tyr^{171} and Val^{174} are involved in the second Ca^{2+}-binding site. In addition, the homology model was used to calculate the electrostatic potential at pH 7.0, as shown in Fig. 2. SPFA is mainly negatively charged around the active site, while neutral, positive and negative charges are balanced on the backside.

Protease production and purification

Production of recombinant SPFA by *B. subtilis* DB104 resulted in an activity of 32 U·mL^{-1} for suc-AAPF-pNA substrate and 418 U·mL^{-1} for azocasein in the supernatant. The purification to apparent homogeneity of SPFA was confirmed by SDS/PAGE (Fig. 3). The protease migrates at approximately 27 kDa, which is congruent with the theoretical molecular mass of 27.57 kDa. Furthermore, the molecular mass was confirmed to be 27.57 kDa by MALDI-TOF MS analysis (Fig. S2). The purified SPFA had a specific activity of 195 U·mg^{-1} for the suc-AAPF-pNA substrate and 539 U·mg^{-1} for azocasein, respectively. The recovery yield was 5.92% with a purification fold of 2.98. Isoelectric point analysis of the purified and rebuffered protease revealed a pI of about 5.8, consistent with the predicted pI and an AB ratio of 0.9 with a high number of Asp residues (Fig. S3, Table S2).

Effects of pH and temperature

The effects of pH on the activity of SPFA within the suc-AAPF-pNA assay were investigated at a pH range of 5.0–12.5 at 30 °C (Fig. 4). The protease showed a broad pH optimum between pH 8.5 and 11.5, while a relative activity of more than 75% was observed between pH 7.0 and 12.5. The relative activity at pH 5.0 and 12.0 was 8 and 76%, respectively. The stability test revealed a residual activity of at least 72% between pH 6.0 and 12.0, while SPFA was unstable at pH 5.0 (Fig. 4).

The enzyme activity was investigated at pH 8.6 between 20 and 90 °C (Fig. 5A). The temperature optimum was reached at 80 °C and decreased to 79% residual activity at 90 °C. The temperature stability was investigated by incubating SPFA at 20 and 50 °C for 4 h (Fig. 5B). The activity of SPFA decreased to 25% after 4 h at 20 °C and to 3% at 50 °C. Possible autoproteolytic cleavage complicates the comparison of protease temperature stability. Hence, to monitor thermal protein unfolding rather than autoproteolysis, a thermal shift assay was conducted on SPFA. The protease was inhibited with phenylmethylsulfonyl fluoride (PMSF) and a denaturation curve with a melting point (T_m) of 62.5 °C was obtained (Fig. S4).

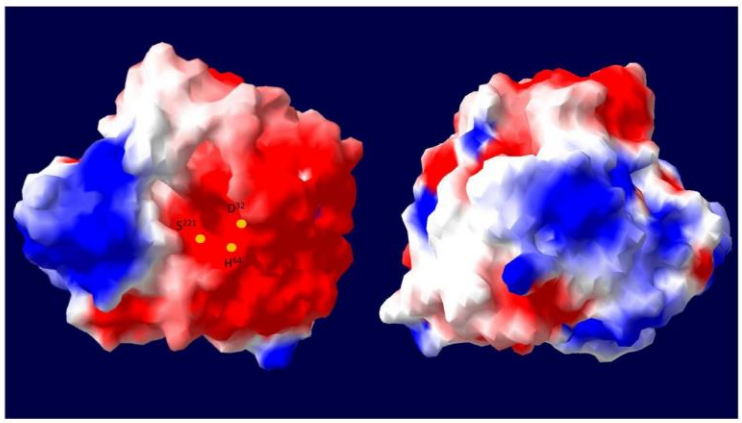


Fig. 2. Structural model of SPFA with its calculated surface electrostatic potential. (Left) top view of the active site; (right) rear view to the active site. Swiss-PdbViewer was used to calculate the electrostatic potential at pH 7.0 and negative charges (red), positive (blue) and neutral (white) are shown.

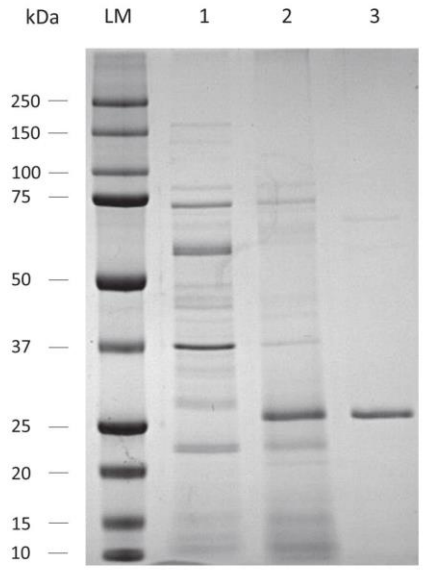


Fig. 3. SDS/PAGE analysis of recombinant SPFA. Electrophoresis was performed using an 8–20% SDS polyacrylamide gel. Bio-Rad Precision Plus Dual Color length marker (LM); culture supernatant of *B. subtilis* DB104 carrying pFF-RED (1); culture supernatant of *B. subtilis* DB104 carrying pFF producing SPFA (2), after purification (3).

Effects of SDS, H_2O_2 and metal ions

The activity of SPFA after 1-h incubation with 1% and 5% SDS (w/v) at 10 °C revealed high stability towards SDS, with increased activity after incubation with 1% and 5% SDS of 182% and 169%, respectively. SPFA showed a residual activity of 81% and 52% after 1 h of treatment with 1% and 5% H_2O_2 (v/v), respectively. Incubation of SPFA with 1 mM PMSF, a classical inhibitor for serine proteases [33], resulted in a complete inhibition. The effect of metal ions on protease activity after 1-h incubation demonstrated residual activities of Mg^{2+} (94% ± 1%), Zn^{2+} (83% ± 1%), Mn^{2+} (95% ± 2%), Ca^{2+} (101% ± 0%), Co^{2+} (93% ± 1%), Ni^{2+} (90% ± 1%), Fe^{2+} (96% ± 1%) and Cu^{2+} (89% ± 1%). The incubation with EDTA showed no decrease in enzyme activity at 20 mM concentration.

Effects of NaCl and proteolytic activity on synthetic peptides

Examination of the effect of different NaCl concentrations (0–5 M) in the suc-AAPF-pNA assay showed that SPFA reached maximum activity without NaCl and the activity declined gradually with higher NaCl levels to 36% at 5 M NaCl (Fig. 6). In contrast, the lowest stability was observed without NaCl with a residual activity of 43%, while the stability at 1–5 M NaCl was above 75%.

Against the synthetic peptide 4-nitroanilide substrates, SPFA exhibited very low specificity for suc-TVAA-pNA, suc-YVAD-pNA, suc-AAA-pNA and suc-AAVA-pNA. The highest activity was shown for suc-ALPF-pNA and suc-AGPP-pNA (Table 1).

Discussion

Serine peptidases, especially the group of subtilisins from the subtilase family, are extremely versatile and

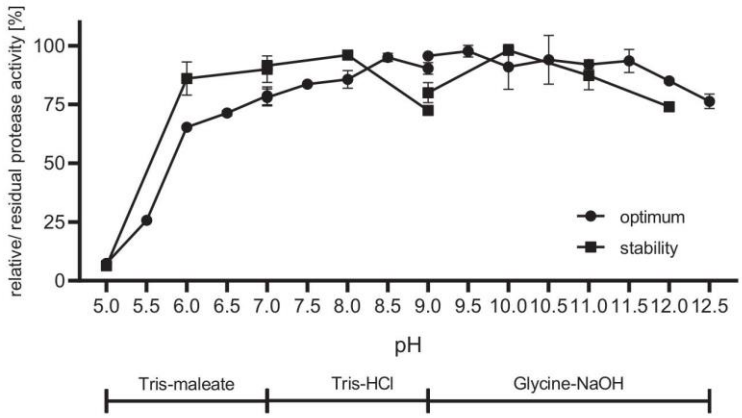


Fig. 4. Influence of pH on the activity and stability of purified SPFA. Enzyme activity was determined using the suc-AAPF-pNA assay at 30 °C in a pH range of 5.0–12.5 (closed circles). The average maximum activity was considered as 100%: 64 U·mg^{-1}. The effect of pH on the stability of purified SPFA (squares). The residual activity was measured with the standard suc-AAPF-pNA assay after incubation for 24 h at 4 °C in Tris-maleate buffer (pH 5–7), in Tris–HCl (pH 7–9) and in glycine-NaOH (pH 9–12). The activity at 0 h was considered as 100%; highest residual activity: 106 U·mg^{-1}. Experiments were performed in triplicate, and data are presented as means ± SD.

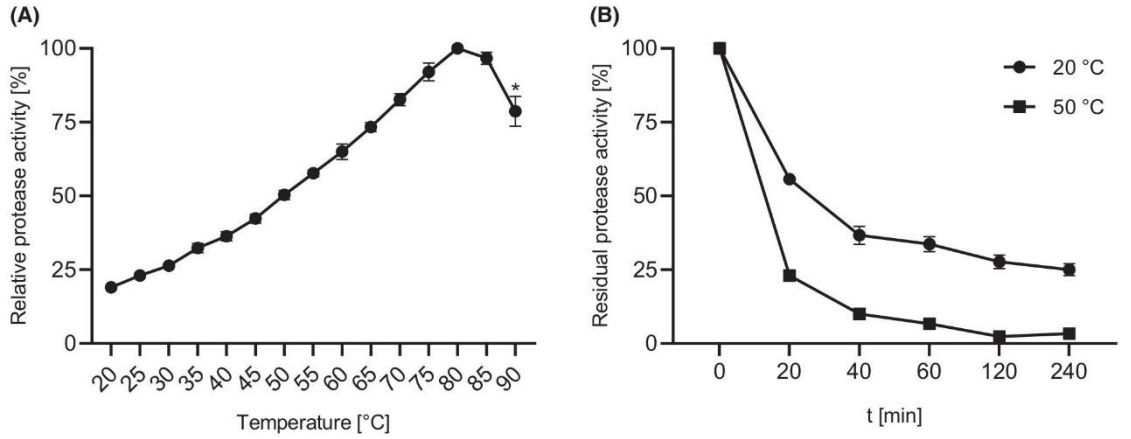


Fig. 5. Effect of temperature on the activity (A) and stability (B) of purified SPFA. Enzyme activity was analysed at temperatures between 20 and 90 °C using the suc-AAPF-assay. The maximum activity was defined as 100%: 272 U·mg^{-1}. * Enzyme stability did not last the intended 5 min. Stability was examined at 20 and 50 °C in 10 mM HEPES-NaOH buffer, pH 8.0. Residual activity was measured using the suc-AAPF pNA assay at 30 °C. The activity at 0 min was set at 100%: 85 U·mg^{-1}. Experiments were performed in triplicate, and data are plotted as mean values ± SD.

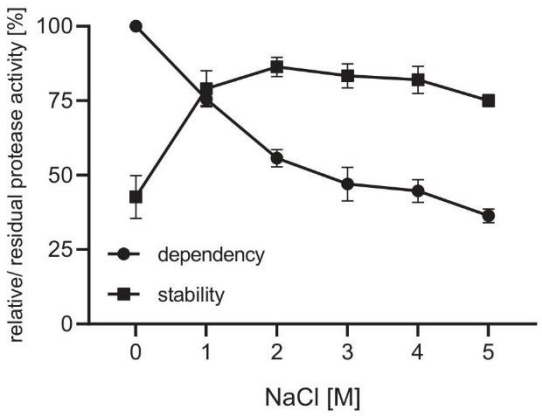


Fig. 6. Activity and stability of purified SPFA at different NaCl concentrations. Activity was measured with the suc-AAPF-pNA assay in standard buffer (pH 8.6) at 30 °C with different NaCl concentrations (0–5 M). Maximum activity was defined as 100%: 104 U·mg^{-1}. Stability was tested in 10 mM HEPES-NaOH buffer, pH 8.0, with NaCl (0–5 M). The residual activity was measured with the suc-AAPF-pNA assay in standard buffer at pH 8.6 after incubation for 2 h at 20 °C. The activity before incubation for each NaCl concentration was defined as 100%. The experiments were performed in triplicate, and data are displayed as means ± SD.

of specific interest for biotechnological applications due to their properties such as broad pH range, high specificity and thermostability [12]. The *Bacillaceae* family has been recognised as a valuable source of subtilisins with high potential for industrial applications [2]. In this study, we therefore characterised the subtilisin SPFA from *Fictibacillus arsenicus* DSM 15822^T, which was found through a data mining-based search, as we previously published [6].

The comparison of the mature SPFA amino acid sequence to the well-characterised true subtilisins Carlsberg [34], BPN' [35] and subtilisin DY [36] reveals a sequence identity of 61.8%, 61.3% and 71.9%, respectively. When comparing SPFA amino acid sequence to our previously reported true subtilisins from halophilic and halotolerant background a sequence identity of 64.2% (SPMI), 65.4% (SPPM) and 71.7% (SPLA) was reached [27]. A more distant relationship exists to the high-alkaline subtilisin Savinase from *Lederbergia lenta* (formerly *Bacillus lentus*; 52.2%) [37] and our previously reported high-alkaline subtilisin SPAO from *Halalkalibacter okhensis* K10-101^T (47.8%) [24], as well as the PIS SPAH (48.1%) from *Alkalibacillus haloalkaliphilus* DSM 5271^T [27]. The comparison of the amino acid sequences in the MSA confirms that SPFA belongs to the group of true subtilisins. Here, compared with BPN', SPFA did not exhibit an amino acid deletion around position 160 that high-alkaline subtilisins have in common [6]. Additionally, insertions compared with BPN' that are common for PIS cannot be observed [6]. Furthermore, SPFA displayed a molecular mass of 27.57 kDa, which is typical for subtilisins found from *Bacillaceae* used in the detergent industry [38].

The classification of SPFA as a member of the true subtilisin subgroup is especially remarkable because

Table 1. Substrate specificities of SPFA against 10 synthetic substrates (suc-XXXX-pNA). Experiments were performed for 5 min at 30 °C with 17 mM substrate in 0.1 M Tris–HCl-Puffer pH 8.6 with 0.1% (w/v) Brij®35. The experiment was executed in triplicates with a standard deviation of <5%. The enzyme activity against AAPF was defined as 100%.

	Relative activity [%]									
Protease	FAAF	AAA	AAVA	ALPF	AGPF	AAPF	TVAA	YVAD	AGPP	AAPL
SPFA	39	1	3	151	98	100 (117 U·mg^{-1})	3	1	181	81
Subtilisin Carlsberg[a]	57	0	2	60	90	100 (570 U·mg^{-1})	1	1	147	104
Savinase[a]	605	8	22	117	96	100 (180 U·mg^{-1})	5	5	144	12
BPN'[a]	96	0	6	106	96	100 (181 U·mg^{-1})	0	0	61	67

[a]Falkenberg *et al.* [24].

high activity at alkaline pH of 12.0 is usually subjected to members of the high-alkaline subgroup [12]. Another example is the phylogenetically intermediate subtilisin ALTP from *Alkaliphilus transvaalensis*, which showed increasing activity up to the optimum of pH 12.0 [39]. SPFA is active and stable over a broad pH range (6.0–12.5) with a broad pH optimum. This broad pH optimum is quite unusual, while a broad pH stability is common to subtilases [40–42]. Tekin *et al.* [43] reported on the high-alkaline subtilisin aprM from *Halalkalibacterium halodurans* C-125 that also displays a high activity (> 70%) between pH 7.0 and 12.0 but with a clear optimum at pH 12.0. We also recently reported on three true subtilisins and one PIS derived from halotolerant and halophilic *Bacillaceae* that exhibited good activity even at a pH of 12.0, probably due to the salt adaptation through a charged molecular surface that facilitates the adaptation to high pH [27]. This charged molecular surface can also be observed in SPFA. However, especially on the back of the active site, positive and negative charges are almost evenly distributed. The fact that SPFA is still active at high NaCl concentrations is likely due to the charged amino acids on the protein surface, which form a hydrate shell around the protein and thus maintain its solubility. Here, negatively charged amino acids are the most favourable, followed by positively charged and charge-neutral amino acids [44]. The increasing stability at higher NaCl concentrations is most likely due to the reduced activity and thus lower autoproteolysis.

Furthermore, the high temperature optimum of 80 °C is particularly noteworthy since *F. arsenicus*, the microbial origin of SPFA, is only able to grow below 40 °C [20]. Higher catalytic activity than the optimal growth temperature is common to enzymes originating from mesophiles (15–50 °C) [45]. Other subtilisins such as BPN', Savinase and subtilisin Carlsberg displayed lower optimal temperatures (between 55 and 65 °C) than SPFA [24]. Similar to SPFA, the subtilase Aqualysin I from thermophilic *Thermus aquaticus* YT 1 also showed a high temperature optimum of 75–90 °C [46]. The phylogenetically intermediate subtilisin ALTP from *Alkaliphilus transvaalensis* exhibited a temperature optimum of 70 °C [39].

The stability towards SDS and H_2O_2 is interesting for industrial applications such as detergents. Here, SPFA showed increased activity when incubated with SDS, which has also been observed in other subtilisins [47–49]. In contrast, the two subtilisins from *Bacillus mojavensis* A21 or the subtilisin from *Bacillus pumilus* BO1 lost their activity when incubated with 1% (w/v) SDS [50,51]. In terms of stability against oxidation with H_2O_2, SPFA showed good stability and retained more than half of its activity when incubated with 5% (v/v) H_2O_2. The two *B. mojavensis* A21 subtilisins also showed comparable stability to SPFA, while subtilisin Carlsberg, BPN' and Savinase lost more than two thirds of their activity [24,50]. Even higher stability was observed for the subtilisin from *Bacillus safensis* RH12 or the alkaline protease from *Bacillus patagoniensis* at higher hydrogen peroxide concentrations [42,52]. The sensitivity of proteases to oxidants is probably due to the oxidation of a conserved methionine near the catalytic site leading to inactivation [53]. However, it was also found that this oxidation can lead to an altered substrate spectrum instead of inactivation [54]. Metal ions like Mg^{2+}, Mn^{2+}, Co^{2+} and Fe^{2+} at a concentration of 1 mM after incubating SPFA for 1 h caused only a slight decrease in protease activity of up to 7%. On the contrary, Zn^{2+}, Ni^{2+}, Cu^{2+} reduced protease activity by up to 17%, while Ca^{2+} had no influence. However, the incubation with EDTA showed no decrease in activity. After incubation with Fe^{2+}, Cu^{2+}, Zn^{2+}, Ca^{2+} and Mn^{2+}, a subtilisin from *Bacillus pumilus* BO1 even displayed a slight increase of up to 13% in activity, while Co^{2+} reduced the residual activity to 81% and EDTA to 78% [51].

A higher decrease in activity after incubation with Ni^{2+}, Cu^{2+} was observed for a subtilase from a *Bacillus cereus* strain, while the decrease after Co^{2+} incubation was comparable to that of SPFA [55]. The proteases of the S8 family are known to be calcium-dependent and contain usually two binding sites [4,56]. The observation that the addition of calcium had no influence on SPFA could be due to the fact that the two calcium binding sites are already sufficiently filled. However, even the addition of EDTA causes no decrease in activity. No effect of EDTA after incubating the serine protease from *B. clausii* GMBAE 42 was also reported by Kazan *et al.* [57] and by Raval *et al.* [58]. Kembhavi *et al.* demonstrated that EDTA for a protease from *Bacillus subtilis* only has a destabilising effect at higher temperatures [41].

The substrate spectrum of SPFA towards the different synthetic substrates exhibits a preference comparable to that of other subtilisins from the true subtilisin and PIS subgroup [24,27]. SPFA demonstrated comparable specificity to the industrially relevant subtilisins Carlsberg and BPN' which belong to the true subtilisins subgroup. Hence, the substrate specificity towards the synthetic substrate is less similar to Savinase, which belongs to the high-alkaline subgroup.

In conclusion, this is the first report of cloning, production, purification and biochemical characterisation of the true subtilisin SPFA from *Fictibacillus arsenicus* DSM 15822^T. With its substrate preference towards 10 synthetic peptide-4-nitroanilide substrates with three or four amino acids and its phylogenetic classification, SPFA can be assigned to the group of true subtilisins. Furthermore, SPFA displayed a temperature optimum of 80 °C and a very broad pH optimum between 8.5 and 11.5 with high activity (> 75%) in an extremely wide range between pH 6.0 and 12.5. SPFA is still active at NaCl concentrations up to 5 M and very stable against 5% (w/v) SDS and stable against 5% (v/v) H_2O_2. Due to its unique biochemical properties, SPFA has the potential for use in biotechnological applications.

Acknowledgement

This work was supported by the FH Aachen University of Applied Sciences within the framework of the programme for the promotion of young scientists. Open Access funding enabled and organized by Projekt DEAL.

Conflict of interest

The authors declare no conflict of interest.

Data accessibility

The original contributions presented in this study are included in the article/Supplementary material, and further inquiries can be directed to the corresponding author.

Author contributions

FF, JB and PS conceived and designed the experiments. FF conducted the experiments, collected and analysed the data. SK purified SPFA and did the MALDI-TOF MS experiment. FF wrote the original draft; FF, JB, MB and PS revised the manuscript. All authors contributed to the final manuscript. All authors read and approved the manuscript.

References

1 Gupta R, Beg QK and Lorenz P (2002) Bacterial alkaline proteases: molecular approaches and industrial applications. *Appl Microbiol Biotechnol* **59**, 15–32.
2 Naveed M, Nadeem F, Mehmood T, Bilal M, Anwar Z and Amjad F (2021) Protease—a versatile and ecofriendly biocatalyst with multi-industrial applications: an updated review. *Catal Lett* **151**, 307–323.
3 Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A and Finn RD (2018) The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res* **46**, D624–D632.
4 Siezen RJ and Leunissen JA (1997) Subtilases: the superfamily of subtilisin-like serine proteases. *Protein Sci* **6**, 501–523.
5 Saeki K, Magallones MV, Takimura Y, Hatada Y, Kobayashi T, Kawai S and Ito S (2003) Nucleotide and deduced amino acid sequences of a new subtilisin from an alkaliphilic *Bacillus* isolate. *Curr Microbiol* **47**, 337–340.
6 Falkenberg F, Bott M, Bongaerts J and Siegert P (2022) Phylogenetic survey of the subtilase family and a data-mining-based search for new subtilisins from *Bacillaceae*. *Front Microbiol* **13**, 1017978.
7 Burnett TJ, Shankweiler GW and Hageman JH (1986) Activation of intracellular serine proteinase in *Bacillus subtilis* cells during sporulation. *J Bacteriol* **165**, 139–145.
8 Gamble M, Künze G, Dodson EJ, Wilson KS and Jones DD (2011) Regulation of an intracellular subtilisin protease activity by a short propeptide sequence through an original combined dual mechanism. *Proc Natl Acad Sci USA* **108**, 3536–3541.
9 Kalisz HM (1988) Microbial proteinases. *Adv Biochem Eng Biotechnol* **36**, 1–65.

10 Contesini FJ, Melo RR and Sato HH (2017) An overview of *bacillus* proteases: from production to application. *Crit Rev Biotechnol* **38**, 321–334.
11 Solanki P, Putatunda C, Kumar A, Bhatia R and Walia A (2021) Microbial proteases: ubiquitous enzymes with innumerable uses. *3 Biotech* **11**, 428.
12 Azrin NAM, Ali MSM, Rahman RNZRA, Oslan SN and Noor NDM (2022) Versatility of subtilisin: a review on structure, characteristics, and applications. *Biotechnol Appl Biochem* **69**, 2599–2616.
13 Gupta R, Beg QK, Khan S and Chauhan B (2002) An overview on fermentation, downstream processing and properties of microbial alkaline proteases. *Appl Microbiol Biotechnol* **60**, 381–395.
14 Power SD, Adams RM and Wells JA (1986) Secretion and autoproteolytic maturation of subtilisin. *Proc Natl Acad Sci USA* **83**, 3096–3100.
15 Markland FS and Smith EL (1971) 16 subtilisins: primary structure, chemical and physical properties. In *Hydrolysis: Peptide Bonds*, Vol. 3 (Boyer PD, ed.), pp. 561–608. Elsevier, Amsterdam, The Netherlands.
16 Tjalsma H, Bolhuis A, Jongbloed JD, Bron S and van Dijl JM (2000) Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome. *Microbiol Mol Biol Rev* **64**, 515–547.
17 Aharoni A (2009) Mining for new enzymes. *Microb Biotechnol* **2**, 128–129.
18 Luo X-J, Yu H-L and Xu J-H (2012) Genomic data mining: an efficient way to find new and better enzymes. *Enzyme Eng* **01**, 104.
19 Rawlings ND (2013) Identification and prioritization of novel uncharacterized peptidases for biochemical characterization. *Database (Oxford)* **2013**, bat022.
20 Shivaji S, Suresh K, Chaturvedi P, Dube S and Sengupta S (2005) *Bacillus arsenicus* sp. nov., an arsenic-resistant bacterium isolated from a siderite concretion in West Bengal, India. *Int J Syst Evol Microbiol* **55**, 1123–1127.
21 Glaeser SP, Dott W, Busse H-J and Kämpfer P (2013) *Fictibacillus phosphorivorans* gen. Nov., sp. nov. and proposal to reclassify *Bacillus arsenicus*, *Bacillus barbaricus*, *bacillus macauensis*, *Bacillus nanhaiensis*, *Bacillus rigui*, *Bacillus solisalsi* and *Bacillus gelatini* in the genus *Fictibacillus*. *Int J Syst Evol Microbiol* **63**, 2934–2944.
22 Kawamura F and Doi RH (1984) Construction of a *Bacillus subtilis* double mutant deficient in extracellular alkaline and neutral proteases. *J Bacteriol* **160**, 442–444.
23 Bernhard K, Schrempf H and Goebel W (1978) Bacteriocin and antibiotic resistance plasmids in *Bacillus cereus* and *Bacillus subtilis*. *J Bacteriol* **133**, 897–903.
24 Falkenberg F, Rahba J, Fischer D, Bott M, Bongaerts J and Siegert P (2022) Biochemical characterization of a novel oxidatively stable, halotolerant, and high-alkaline subtilisin from *Alkalihalobacillus okhensis* Kh10-101^T. *FEBS Open Bio* **12**, 1729–1746.
25 Sayers EW, Beck J, Bolton EE, Bourexis D, Brister JR, Canese K, Comeau DC, Funk K, Kim S, Klimke W *et al.* (2021) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **49**, D10–D17.
26 Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J and Sayers EW (2016) GenBank. *Nucleic Acids Res* **44**, D67–D72.
27 Falkenberg F, Voß L, Bott M, Bongaerts J and Siegert P (2023) New robust subtilisins from halotolerant and halophilic *Bacillaceae*. *Appl Microbiol Biotechnol* **107**, 3939–3954.
28 DelMar EG, Largman C, Brodrick JW and Geokas MC (1979) A sensitive new substrate for chymotrypsin. *Anal Biochem* **99**, 316–320.
29 Brock FM, Forsberg CW and Buchanan-Smith JG (1982) Proteolytic activity of rumen microorganisms and effects of proteinase inhibitors. *Appl Environ Microbiol* **44**, 561–569.
30 The UniProt Consortium (2021) UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* **49**, D480–D489.
31 Zhang Y (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40.
32 Zhang Y and Skolnick J (2004) Scoring function for automated assessment of protein structure template quality. *Proteins* **57**, 702–710.
33 North MJ (1982) Comparative biochemistry of the proteinases of eucaryotic microorganisms. *Microbiol Rev* **46**, 308–340.
34 Smith EL, Markland FS, Kasper CB, DeLange RJ, Landon M and Evans WH (1966) The complete amino acid sequence of two types of subtilisin, BPN' and Carlsberg. *J Biol Chem* **241**, 5974–5976.
35 Matsubara H, Kasper CB, Brown DM and Smith EL (1965) Subtilisin BPN. *J Biol Chem* **240**, 1125–1130.
36 Peters K, Pauli D, Hache H, Boteva RN, Genov NC and Fittkau S (1989) Subtilisin DY—kinetic characterization and comparison with related proteinases. *Curr Microbiol* **18**, 171–177.
37 Betzel C, Klupsch S, Papendorf G, Hastrup S, Branner S and Wilson KS (1992) Crystal structure of the alkaline proteinase Savinase from *Bacillus lentus* at 1.4 a resolution. *J Mol Biol* **223**, 427–445.
38 Maurer K-H (2004) Detergent proteases. *Curr Opin Biotechnol* **15**, 330–334.
39 Kobayashi T, Lu J, Li Z, Hung VS, Kurata A, Hatada Y, Takai K, Ito S and Horikoshi K (2007) Extremely high alkaline protease from a deep-subsurface bacterium, *Alkaliphilus transvaalensis*. *Appl Microbiol Biotechnol* **75**, 71–80.
40 Kobayashi T, Hakamada Y, Adachi S, Hitomi J, Yoshimatsu T, Koike K, Kawai S and Ito S (1995)

Purification and properties of an alkaline protease from alkalophilic *Bacillus sp*. KSM-K16. *Appl Microbiol Biotechnol* **43**, 473–481.

41 Kembhavi AA, Kulkarni A and Pant A (1993) Salt-tolerant and thermostable alkaline protease from *Bacillus subtilis* NCIM no. 64. *Appl Biochem Biotechnol* **38**, 83–92.

42 Rekik H, Zaraî Jaouadi N, Gargouri F, Bejar W, Frikha F, Jmal N, Bejar S and Jaouadi B (2019) Production, purification and biochemical characterization of a novel detergent-stable serine alkaline protease from *Bacillus safensis* strain RH12. *Int J Biol Macromol* **121**, 1227–1239.

43 Tekin A, Uzuner U and Sezen K (2021) Homology modeling and heterologous expression of highly alkaline subtilisin-like serine protease from *Bacillus halodurans* C-125. *Biotechnol Lett* **43**, 479–494.

44 Qiao B, Jiménez-Ángeles F, Nguyen TD, Olvera L and de Cruz M (2019) Water follows polar and nonpolar protein surface domains. *Proc Natl Acad Sci USA* **116**, 19274–19281.

45 Engqvist MKM (2018) Correlating enzyme annotations with a large set of microbial growth temperatures reveals metabolic adaptations to growth at diverse temperatures. *BMC Microbiol* **18**, 177.

46 Tanaka T, Matsuzawa H and Ohta T (1998) Substrate specificity of Aqualysin I, a bacterial thermophilic alkaline serine protease from *Thermus aquaticus* YT-1: comparison with proteinase K, subtilisin BPN' and subtilisin Carlsberg. *Biosci Biotechnol Biochem* **62**, 2161–2165.

47 Joshi S and Satyanarayana T (2013) Characteristics and applications of a recombinant alkaline serine protease from a novel bacterium *Bacillus lehensis*. *Bioresour Technol* **131**, 76–85.

48 Thebti W, Riahi Y and Belhadj O (2016) Purification and characterization of a new thermostable, haloalkaline, solvent stable, and detergent compatible serine protease from *Geobacillus toebii* strain LBT 77. *Biomed Res Int* **2016**, 9178962.

49 Bhatt HB and Singh SP (2020) Cloning, expression, and structural elucidation of a biotechnologically potential alkaline serine protease from a newly isolated haloalkaliphilic *Bacillus lehensis* JO-26. *Front Microbiol* **11**, 941.

50 Haddar A, Agrebi R, Bougatef A, Hmidet N, Sellami-Kamoun A and Nasri M (2009) Two detergent stable alkaline serine-proteases from *Bacillus mojavensis* A21: purification, characterization and potential application as a laundry detergent additive. *Bioresour Technol* **100**, 3366–3373.

51 Farooq S, Nazir R, Ganai SA and Ganai BA (2021) Isolation and characterization of a new cold-active protease from psychrotrophic bacteria of Western Himalayan glacial soil. *Sci Rep* **11**, 12768.

52 Olivera N, Sequeiros C, Siñeriz F and Breccia JD (2006) Characterisation of alkaline proteases from a novel alkali-tolerant bacterium *Bacillus patagoniensis*. *World J Microbiol Biotechnol* **22**, 737–743.

53 Stauffer CE and Etson D (1969) The effect on subtilisin activity of oxidising a methionine residue. *J Biol Chem* **244**, 5333–5338.

54 Nonaka T, Fujihashi M, Kita A, Saeki K, Ito S, Horikoshi K and Miki K (2004) The crystal structure of an oxidatively stable subtilisin-like alkaline serine protease, KP-43, with a C-terminal beta-barrel domain. *J Biol Chem* **279**, 47344–47351.

55 Gurunathan R, Huang B, Ponnusamy VK, Hwang J-S and Dahms H-U (2021) Novel recombinant keratin degrading subtilisin like serine alkaline protease from *Bacillus cereus* isolated from marine hydrothermal vent crabs. *Sci Rep* **11**, 12007.

56 Siezen RJ, de Vos WM, Leunissen JA and Dijkstra BW (1991) Homology modelling and protein engineering strategy of subtilases, the family of subtilisin-like serine proteinases. *Protein Eng* **4**, 719–737.

57 Kazan D, Denizci AA, Oner MNK and Erarslan A (2005) Purification and characterization of a serine alkaline protease from *Bacillus clausii* GMBAE 42. *J Ind Microbiol Biotechnol* **32**, 335–344.

58 Raval VH, Pillai S, Rawal CM and Singh SP (2014) Biochemical and structural characterization of a detergent-stable serine alkaline protease from seawater haloalkaliphilic bacteria. *Process Biochem* **49**, 955–962.

59 Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 2011;**7**:539.

60 Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res*. 2014;**42**:W320–4.

Supporting information

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

Fig. S1. Homology model of the mature SPFA obtained using I-TASSER software. *In silico* metal-binding analysis predicted the existence of two Ca2 + −binding sites (yellow balls). The catalytic residues Asp32, His64 and Ser221 are shown in red.

Fig. S2. MALDI-TOF mass spectra of SPFA. The labels on the peaks indicate the measured average molecular mass. The peaks correspond from right to left M/z up to M/5z.

Fig. S3. Determination of the pI of the purified proteases. Isoelectric focussing was performed with a SERVALYT™ PRECOTES™ wide range pH 3–10 precast gel according

to the manufacturer's recommendations. Lane M, SERVA IEF marker 3–10; lanes 1 purified SPFA rebuffered in 10 mM HEPES-NaOH pH 8.0.

Fig. S4. Normalised melting curve of purified SPFA. The melting point (Tm) at which 50% of the protein is unfolded (−) was determined using SYPRO® Orange as a fluorescent probe (Ex/Em = 470/550 nm) (5 x SYPRO® Orange, 10 mM HEPES-NaOH pH 8.0, 3 mM PMSF). The experiment was performed in triplicates and data are plotted as mean values ± SD.

Table S1. Oligonucleotides for amplification of the gene for SPFA by PCR using genomic DNA of *F. arsenicus* as template.

Table S2. pI value and AB ratio calculation.

2.5 *Metabacillus indicus* DSM 16189 as promising host for recombinant protease production

Fabian Falkenberg, Vincent Vonderbank, Jessica Herten, Nihal Yasemin Kus, Michael Bott, Petra Siegert, Johannes Bongaerts

Manuscript

Authors contributions:

FF, JB, and PS conceived and designed the experiments.
FF conducted the experiments except those of VV, JH, and NK, collected and analysed the data.
VV did the cloning and transformation experiments.
VV, JH and NK did pre-experiments and media optimisation.
FF wrote the original draft.
FF, PS, MB, and JB revised the manuscript.

All figures and tables are from this study.

Overall contribution FF: 75 %

Title Page

Title:

Metabacillus indicus DSM 16189 as promising host for recombinant protease production

Author's names:

Fabian Falkenberg[1], Vincent Vonderbank[1], Jessica Herten[1], Nihal Yasemin Kus[1], Michael Bott[2], Petra Siegert[1], Johannes Bongaerts[1]

Addresses:

1 Institute of Nano- and Biotechnologies, Aachen University of Applied Sciences, 52428, Jülich, Germany.
2 Institute of Bio- and Geosciences, IBG-1: Biotechnology, Forschungszentrum Jülich, 52425 Jülich, Germany.

ORCID:

Fabian Falkenberg: 0000-0003-2374-0418

Michael Bott: 0000-0002-4701-8254

Petra Siegert: 0000-0001-7172-3060

Corresponding author:

J. Bongaerts, Institute of Nano- and Biotechnologies, Aachen University of Applied Sciences, 52428, Jülich, Germany
Tel.: +49 241 6009 53941
E-mail: bongaerts@fh-aachen.de
https://www.fh-aachen.de/forschung/inb

Running title:

Protease production with *Metabacillus indicus* DSM 16189

Acknowledgment:

This work was supported by the FH Aachen University of Applied Sciences within the framework of the program for the promotion of young scientists.

Keywords:

***Metabacillus indicus*, protease production, Sec dependent protein secretion, subtilisin, *Bacillaceae*, Biotechnological application**

Abstract

Production of enzymes such as proteases at a high level is of interest and relevance for both research and industry. The selection of microbial host and expression system depends on the origin and characteristics of the protein to be produced, as well as the ease of fermentation on cheap substrates followed by simple downstream processing. Of particular importance are Gram-positive *Bacillus* species such as *Bacillus licheniformis*, *Bacillus subtilis* and *Bacillus amyloliquefaciens* as industrial production organisms for technical enzymes or chemicals. However, often the yield is not adequate and an expansion of the host spectrum is necessary. For this purpose, the halotolerant *Metabacillus indicus* DSM 16189, with its natural production of subtilisins and its growth on cheap media without detectable sporulation, is a promising candidate. Genetic accessibility could be shown with the transformation by triparental conjugation with the expression vector, which carries the gene for the homologous subtilisin SPMI. The yield of protease production was compared with that of non-recombinant *M. indicus* and *B. subtilis* DB104, which carries the same plasmid. The recombinant SPMI produced by *M. indicus* in a 1-L fermentation over 48 hours had an average activity of 170 U/mL (AAPF) and 1426 U/mL (azocasein), while the non-recombinant strain reached 4 U/mL (AAPF) and 121 U/mL (azocasein), respectively. The successful transformation and high yields demonstrate that *M. indicus* DSM 16189 has the potential as new production organism and could achieve competitive protein yields with further process and strain development.

Introduction

In recent years, the biotechnological production of enzymes and chemicals such as insecticides, vitamins, sugars or hyaluronic acid has gained in importance [1–3]. Technical enzymes like proteases, lipases, cellulases and α-amylases are produced by microorganisms [1]. In particular, proteases are important with the highest share of the global enzymes market and an expected compound annual growth rate of 3.6 % in market value from 2019 to 2026 [4, 5]. The versatility and biochemical properties of proteases are in demand in various industries, such as food, feed, pharmaceuticals, textiles and the largest sector, detergents [4, 6]. The use of proteases in these sectors can increase production efficiency and environmental sustainability as fewer harsh chemicals are used [6]. Increasing amounts of biotechnological production are accompanied by a search for easy-to-use, cost-effective production strains. For protease production, strains such as *Bacillus subtilis*, *Bacillus licheniformis* or *Bacillus amyloliquefaciens* are established microorganisms that have been further genetically optimised to improve the production of various proteins and their yield at concentrations greater than 20 g/L [7]. This is due to the fact that *Bacillus* species naturally produce hydrolytic enzymes, as they cannot absorb macromolecules and must first break them down. Especially in their natural habitats, available sources of plant nutrients such as protein, starch, pectin and cellulose must first be hydrolysed [7]. Compared to other species, *Bacillus* species exhibit non-toxicity, ease of gene modification, short fermentation cycles, superior protein secretion abilities, and robustness in industrial fermentations [8, 3].

In addition to the industrially established *Bacillus* species, other strains from the *Bacillaceae* family were also evaluated for recombinant protein production, e.g. *Bacillus methanolicus* [9], *Bacillus pumilus* [10], *Brevibacillus choshinensis* [11] or *Geobacillus thermodenitrificans* [12]. However, the production of proteases is not limited to strains from the *Bacillaceae* family, but also to other bacteria or yeasts such as *Pichia pastoris* [13, 14].

The ability to secrete target proteins in high concentrations and thus simplify further processing makes Bacteria from the *Bacillaceae* family ideal candidates for industrial production [3]. However, problems can arise in the production of heterologous proteins [15, 16]. The reasons for this lie in the target

protein itself and in the secretion system, which includes a number of quality control points to prevent potentially lethal blockages [7]. In order to have an alternative source for the production of target proteins, especially proteases, the search for new production strains is necessary. Here, proteases with halotolerant or halophilic background provide immense potential for meeting industrial demands such as high tolerance to pH, temperature and osmotic stress [17–20]. We recently reported about the extracellular subtilisin protease SPMI from *Metabacillus indicus* DSM 16189, which showed high activity at pH 12.0, stability towards 5 % (w/v) SDS and was active at NaCl concentrations of 5 M [21]. The objective of this study is therefore to investigate whether the halotolerant strain *Metabacillus indicus* DSM 16189 can be used for the recombinant production of proteases. Specifically, we started the investigation by the recombinant production of its homologous subtilisin SPMI.

M. indicus DSM 16189 was isolated from seafood in Korea by Yoon et al. and originally identified as *Bacillus cibi* [22]. A genome-based study led to the reclassification to *Bacillus indicus* and later to *Metabacillus indicus* [23, 24]. The strain is halotolerant and can grow in environments with up to 12 % (w/v) NaCl, with an optimum between 0 and 1 %. Gram variability and possible spore formation is also observed with growth between 4 and 50 °C with an optimum at 37 °C [22]. This is the first report of the transformation of *M. indicus* DSM 16189 by triparental conjugation and the recombinant production of its homologous subtilisin.

Material and Methods

Bacterial strains and Plasmid construction/cloning

Metabacillus indicus DSM 16189 was purchased from the German Collection of Microorganisms and Cell Cultures GmbH (DSMZ) and cultivated in marine broth with 10 mg/L $MnSO_4$ (Carl Roth, Karlsruhe, Germany) at 30 °C. *Bacillus subtilis* DB104 [25] was cultivated in LB medium [26] (Carl Roth, Karlsruhe, Germany) or agar plates at 37 °C. *E. coli* DH5α competent cells were used for cloning and as donor strain for the triparental conjugation experiments (Thermo Fisher Scientific GmbH, Karlsruhe, Germany); *E. coli* HB101 pRK2013 was used for triparental conjugation, both were cultivated in LB medium [26] (Carl Roth, Karlsruhe, Germany) or on agar plates at 37 °C. Strains containing a plasmid were all cultivated with 50 µg/mL kanamycin sulphate (Carl Roth, Karlsruhe, Germany).

Cloning of the gene encoding the protease SPMI (WP_029565418.1) into the pBACOV plasmid [27] was performed after previous PCR amplification using the primers (5'-TAAACGCGTATGAAAAAGAAAAAGATTTTCAGTC-3' and 5'-ATAGAATTCTTATTGAATGGCGGCCTG-3') with the genomic DNA of *M. indicus* as a template. The genomic DNA was obtained by using the InnuSPEED Bacteria/Fungi DNA Kit (Analytik JenaTM, Jena, Germany). The PCR was performed with the Phusion® Hot Start II High-Fidelity polymerase (Thermo Fisher Scientific GmbH, Karlsruhe, Germany) according to the manufacturer's recommendations. The PCR products were analysed on an agarose gel and PCR products were purified using the GeneJet PCR purification kit (Thermo Fisher Scientific GmbH, Karlsruhe, Germany). Purified PCR product and pBACOV were treated with *EcoRI* and *MluI* FastDigest™ restriction enzymes according to manufacturer's recommendations, with following agarose gel purification using GeneJET Gel extraction kit (Thermo Fisher Scientific GmbH, Karlsruhe, Germany). Transformation of *E. coli* DH5α with pVV01 was performed by heat shock according to manufacturer's recommendation (Thermo Fisher Scientific GmbH, Karlsruhe, Germany). Transformation of *B. subtilis* DB104 with pVV01 was performed by natural competence as described previously [21]. The transformation of *M. indicus* was performed by triparental conjugation according to Heinze et al. (2018). Selection after conjugation was performed using 50 µg/mL kanamycin sulphate (Carl Roth, Karlsruhe, Germany), 5 % (w/v) NaCl, 2.5 % (w/v) skim milk for the LB agar plates and by visual identification of the orange colonies and clearing zones. The correct cloning and transformation of *E. coli* DH5α pVV01, *M. indicus* pVV01 and *B. subtilis* DB104 pVV01 was confirmed by double

restriction digestion and Sanger sequencing (Eurofins Genomics, Ebersberg, Germany) by analysing the isolated plasmid DNA obtained with the GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific GmbH, Karlsruhe, Germany).

Recombinant protease production

The production of SPMI with *B. subtilis* DB104 in a 1-L fed-batch fermentation was performed as previously described [28]. For recombinant protein production and cultivation of the wild type of *M. indicus*, a day culture was performed in 10 mL marine broth containing 10 mg/L $MnSO_4$ (Carl Roth, Karlsruhe, Germany) with a freshly plated clone from an LB agar plate. At all cultivation steps 50 µg/mL kanamycin sulphate (Carl Roth, Karlsruhe, Germany) was used for *M. indicus* pVV01 and *B. subtilis* DB104. Cultivation was carried out for 7 hours at 180 rpm and 30 °C. With this culture, a 50 mL overnight culture was inoculated with the preculture medium (Table S1, Table S3) to an optical density at 600 nm (OD_{600}) of 0.2 and cultured at 180 rpm and 30 °C. After cultivation overnight, the bioreactors were inoculated to an OD_{600} of 0.2. Fermentation was carried out for 48 h in fermentation medium with a glucose feed (Table S2, Table S3) as for *B. subtilis* DB104. The glucose feed started for recombinant and wild type *M. indicus* 18 h and 20 h after inoculation, respectively.

Analytical methods

The fermentation supernatant was analysed after centrifugation for 20 min at 12.000 g on a SDS-Gel as described previously [28]. The protein concentration was determined using Roti® Nanoquant (Carl Roth, Karlsruhe, Germany) with bovine serum albumin fraction V (Carl Roth, Karlsruhe, Germany) as a standard according to manufacturer's recommendations based on the method of Bradford [29]. The analysis of the hydrolytic activity of the protease was determined using the substrate suc-AAPF-pNA and azocasein as described before [28].

The genome of *M. indicus* was analysed by searching for S8 family serine peptidases in the protein list of the reference genome (NZ_JNVC00000000.2). The sequences were aligned to the MEROPS S8 holotype dataset (https://www.ebi.ac.uk/merops/cgi-bin/famsum?family=S8) using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) [30, 31].

Results and discussion

Cloning of the *aprE* gene from *M. indicus*

The analysis of the protein list of *M. indicus* reference genome shows that the strain harbours 11 open reading frames that encodes for S8 family serine peptidases. The alignment and the classification according to our previous publication [32] show that four proteases belong to the subtilisin group, two of which are intracellular subtilisins, one extremophilic subtilisin and one true subtilisin which are secreted (Table 1). The true subtilisin (SPMI) was used in this report as a model for cloning and transformation of *M. indicus*. The production with *B. subtilis* DB104, the purification and the biochemical characterisation was described in our previous publication [21]. The other subtilase genes encode a bacillopeptidase F (bpF subtilase), a thermitase, a tripeptidyl peptidase subtilase (TPPS) and four pyrolysins.

Table 1 Subtilases of the *M. indicus* genome

Protein sequence	Group	Subgroup
WP_029565418.1 (SPMI)	Subtilisins	True subtilisins
WP_029278844.1	Subtilisins	Intracellular subtilisins
WP_029565391.1	Subtilisins	Intracellular subtilisins
WP_035207592.1	Subtilisins	Extremophilic subtilisins
WP_029565567.1	bpF subtilases	-
WP_051827326.1	Thermitases	-
WP_029566183.1	TPPS subtilases	-
WP_029566552.1	Pyrolysins	High-mass subtilases
KEZ49076.1	Pyrolysins	High-mass subtilases
WP_035207616.1	Pyrolysins	SAM-P45 pyrolysins
WP_051827410.1	Pyrolysins	Oxidatively stable proteases

The gene of the protease SPMI was successfully amplified, including the signal peptide, the propeptide and the mature peptidase, and cloned into the pBACOV plasmid, resulting in the plasmid pVV01. The pBACOV plasmid is a shuttle vector for conjugation transfer and contains an aprE protease promoter from *B. subtilis* for expression of the target gene, an ampicillin and kanamycin resistance gene and the origin of replication for *E. coli* and *Bacillus* [27]. Successful transformation of *B. subtilis* DB0104 and *M. indicus* DSM 16189 was verified by restriction enzyme analysis and Sanger sequencing after plasmid preparation. The production of SPMI by both strains was confirmed by the appearance of clearing zones around the colonies on skim milk/LB agar plates and enzyme activity assays using suc-AAPF-pNA and azocasein as substrates. The method of transformation of *M. indicus* was the conjugative transfer [27]. No evidence of natural competence could be found experimentally, by using the protocol for *B. subtilis* [33] (data not shown). Although the conjugative transfer requires an additional cloning step via *E. coli* compared to the natural competence that can be used to transform *B. subtilis* [33], the method proved to be simple and efficient and enabled recombinant gene expression in several other *Bacillus* species for the first time [27]. Endogenous restriction modification systems can hinder successful transformation and pre-methylation of the plasmid is required, as in strains such as *B. amyloliquefaciens*, *B. cereus* [34] and *B. halodurans* [35]. Within this work, this constraint could not be observed for *M. indicus*.

Protease production

The production of SPMI on a 1-L scale using the DASGIP® bioreactors by *M. indicus* pVV01 compared to *B. subtilis* DB104 pVV01 and the wild type *M. indicus* strain after 48 hours shows that the recombinant *M. indicus* is able to produce high amounts of SPMI. This is expressed by about twice the protease activity for the suc-AAPF-pNA substrate and a 1.5-fold higher protein content (Table 2), which can also be seen in the SDS-PAGE analysis (Figure 1). However, when looking at the cell mass achieved in relation to the protease activities and protein content, *B. subtilis* DB104 is still more efficient in production. The growth of the two strains is difficult to compare because they grow in different media, as cell growth of *M. indicus* is insufficient in the medium of *B. subtilis* and *vice versa*. The total amount of nitrogen is about 3 % higher than for *B. subtilis,* while the amount of carbon is comparable. Other parameters such as glucose concentration, pH and temperature may influence growth and protease production [36]. Comparing the fermentation results of recombinant *M. indicus* with the wild type strain, it was found that the wild type strain achieved about 30 % higher optical density compared to the recombinant *M. indicus*, but had only about 2 % activity in suc-AAPF pNA assay, 9 % in azocasein assay and 17 % of protein content, respectively. This shows that *M. indicus* is well suited for recombinant protein production. The lower cell density of the recombinant strain might be due to the additional metabolic energy required to maintain the plasmid and the higher protease production. With optimised production processes in industrial *Bacillus* strains, protein concentrations of more than 20 g/L are possible [7]. In other strains of the *Bacillaceae* family such as *Brevibacillus choshinensis*, the production of an α-amylase with a yield of 3.7 g/L was possible [11]. Furthermore, after applying strain and process optimisation strategies, *B. pumilus* had a 14 % higher protease titre than the industrially optimised strain *B. licheniformis* [10]. Recombinant production of the protease from *B. halodurans* CM1 also resulted in significantly higher protease production compared to the wild type [35].

Table 2 Analysis of the fermentation culture regarding their OD_{600}, and the supernatants regarding protease activity and protein concentration

Strain/ Plasmid	OD_{600}	U/mL AAPF	U/mL Azocasein	Protein conc. [mg/mL]
***M. indicus* wt**	91.8	3.4	101	0.2
***M. indicus* wt**	113.4	5.1	139	0.3
***M. indicus* wt**	99.4	4.2	124	0.2
***M. indicus* pVV01**	76.2	170	1359	1.5
***M. indicus* pVV01**	81.2	164	1386	1.3
***M. indicus* pVV01**	77.1	175	1533	1.4
***B. subtilis* DB104 pVV01**	30.2	90	928	1.1
***B. subtilis* DB104 pVV01**	31.1	75	767	0.8

The optical density (OD_{600}) was determined after 48 h of fermentation. The protease activities and the protein concentration were analysed with the supernatant using the suc-AAPF-pNA assay, azocasein assay and Roti®Nanoquant, respectively. Each entry shows the results of an independent fermentation.

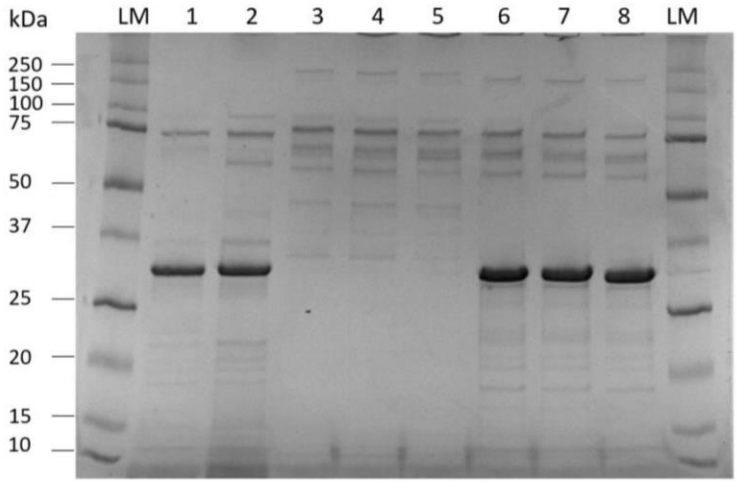


Figure 1 SDS-PAGE analysis of the fermentation supernatants. Samples were electrophoresed with an 8 - 20 % SDS polyacrylamide gel. Bio-Rad Precision Plus Dual Color length marker (LM); *B. subtilis* pVV01 (1, 2); *M. indicus* wt (3 - 5); *M. indicus* pVV01 (6 - 8). Each line shows the protein content of independent fermentations.

Improvement of protease production with *M. indicus* can be achieved by strain engineering and process development. A starting point for engineering could be the elimination of carotenoid production by deletion of the *crtM* gene, one of the genes for carotenoid biosynthesis [37–39], in order to reduce the metabolic burden of the strain. However, the orange/yellow colour of the colonies helps to quickly identify the strain and possible contamination. Furthermore, the endogenous α-amylase genes encoding the proteins WP_029283087.1 and WP_029283839.1 could be deleted which was observed to increase the protein yield in *Bacillus licheniformis* [40]. Although the background proteolytic activity of *M. indicus* is low, deletion of the protease gene could increase productivity, as has been observed in *B. amyloliquefaciens* or *B. subtilis* [41, 42].

To increase protease yield, a signal peptide screen in *B. subtilis* and *B. licheniformis* using homologous and heterologous signal peptides helped [43]. However, the optimal signal peptide for a certain protein and production host is not predictable and must be tested individually for each new protein [44, 43]. Since spores can cause difficulties in sterility, endospore formation can be eliminated, such as in *B. pumilus* [10]. However, we were unable to induce sporulation in *M. indicus* (data not shown), although sporulation has been reported [22]. As mentioned above, the pBACOV plasmid harbours an *aprE* protease promotor from *B. subtilis* to express the target gene [27]. Replacing the promoter with the homologous subtilisin promoter or other promoters that have been shown to be effective in the production of proteases, such as the p43 promoter that has enhanced expression in *B. licheniformis*, could improve the expression level [10, 45].

In conclusion, *B. subtilis* DB104 and *Metabacillus indicus* DSM 16189 were successfully transformed with plasmid pVV01, which contains the gene for the protease SPMI, by natural competence and triparental conjugation, respectively. The recombinant SPMI produced by *M. indicus* in a 48-hour fed-batch fermentation achieved higher activity levels than the wild type strain. This result is the first report of successful genetic modification of the halotolerant *M. indicus* making this bacterium a promising candidate for recombinant protease production. The investigation of producing heterologous proteases and improvements to competitive protein yields seem achievable by additional strain development and process optimisation.

Author Contributions

FF, JB, and PS conceived and designed the experiments. FF collected and analysed the data. VV did the cloning and transformation experiments. VV, JH and NK did pre-experiments and media optimisation. FF wrote the original draft; FF, PS, MB, and JB revised the manuscript. All authors contributed to the final manuscript. All authors read and approved the manuscript.

Funding

This work was supported by the FH Aachen University of Applied Sciences within the framework of the program for the promotion of young scientists.

Data availability

The original contributions presented in this study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Declarations:

Conflict of interest

The authors declare no conflict of interest.

Ethical approval

This article does not contain any studies involving human or animal participants conducted by any of the authors.

References

1 Cai D, Rao Y, Zhan Y, Wang Q & Chen S (2019) Engineering *Bacillus* for efficient production of heterologous protein: current progress, challenge and prospect. *J Appl Microbiol* **126,** 1632–1642.

2 Su Y, Liu C, Fang H & Zhang D (2020) *Bacillus subtilis*: a universal cell factory for industry, agriculture, biomaterials and medicine. *Microb Cell Fact* **19,** 173.

3 Schallmey M, Singh A & Ward OP (2004) Developments in the use of *Bacillus* species for industrial production. *Can J Microbiol* **50,** 1–17.

4 Ward OP Proteases. In *Comprehensive Biotechnology,* 571–582 (2011). Elsevier.

5 Research Reports World (2021) global alkaline proteases market insights and forecast to 2027. https://www.researchreportsworld.com/global-alkaline-proteases-market-18086190.

6 Azrin NAM, Ali MSM, Rahman RNZRA, Oslan SN & Noor NDM (2022) Versatility of subtilisin: A review on structure, characteristics, and applications. *Biotechnol Appl Biochem* **69,** 2599–2616.

7 Pohl S & Harwood CR (2010) Heterologous protein secretion by *Bacillus* species. *Advances in Applied Microbiology,* 1–25.

8 Cui W, Han L, Suo F, Liu Z, Zhou L & Zhou Z (2018) Exploitation of *Bacillus subtilis* as a robust workhorse for production of heterologous proteins and beyond. *World J Microbiol Biotechnol* **34,** 145.

9 Irla M, Drejer EB, Brautaset T & Hakvåg S (2020) Establishment of a functional system for recombinant production of secreted proteins at 50 °C in the thermophilic *Bacillus methanolicus*. *Microb Cell Fact* **19,** 151.

10 Küppers T, Steffen V, Hellmuth H, O'Connell T, Bongaerts J, Maurer K-H & Wiechert W (2014) Developing a new production host from a blueprint: *Bacillus pumilus* as an industrial enzyme producer. *Microb Cell Fact* **13,** 46.

11 Mizukami M, Hanagata H & Miyauchi A (2010) *Brevibacillus* expression system: host-vector system for efficient production of secretory proteins. *Curr Pharm Biotechnol* **11,** 251–258.

12 Daas MJA, Nijsse B, van de Weijer AHP, Groenendaal BWAJ, Janssen F, van der Oost J & van Kranenburg R (2018) Engineering *Geobacillus thermodenitrificans* to introduce cellulolytic activity; expression of native and heterologous cellulase genes. *BMC Biotechnol* **18,** 42.

13 Liu B, Zhang J, Gu L, Du G, Chen J & Liao X (2014) Comparative analysis of bacterial expression systems for keratinase production. *Appl Biochem Biotechnol* **173,** 1222–1235.

14 Radha S & Gunasekaran P (2009) Purification and characterization of keratinase from recombinant *Pichia* and *Bacillus* strains. *Protein Expr Purif* **64,** 24–31.

15 Li W, Zhou X & Lu P (2004) Bottlenecks in the expression and secretion of heterologous proteins in *Bacillus subtilis*. *Res Microbiol* **155,** 605–610.

16 Anné J, Economou A & Bernaerts K (2017) Protein secretion in gram-positive bacteria: from multiple pathways to biotechnology. *Curr Top Microbiol Immunol* **404,** 267–308.

17 Mechri S, Bouacem K, Jabeur F, Mohamed S, Addou NA, Dab A, Bouraoui A, Bouanane-Darenfed A, Bejar S, Hacène H, Baciou L, Lederer F & Jaouadi B (2019) Purification and biochemical characterization of a novel thermostable and halotolerant subtilisin SAPN, a serine protease from *Melghiribacillus*

thermohalophilus Nari2AT for chitin extraction from crab and shrimp shell by-products. *Extremophiles* **23,** 529–547.

18 Joo H-S, Kumar CG, Park G-C, Paik SR & Chang C-S (2003) Oxidant and SDS-stable alkaline protease from *Bacillus clausii* I-52: production and some properties. *J Appl Microbiol* **95,** 267–272.

19 Gupta A, Roy I, Patel RK, Singh SP, Khare SK & Gupta MN (2005) One-step purification and characterization of an alkaline protease from haloalkaliphilic Bacillus sp. *J Chromatogr A* **1075,** 103–108.

20 Mokashe N, Chaudhari B & Patil U (2018) Operative utility of salt-stable proteases of halophilic and halotolerant bacteria in the biotechnology sector. *Int J Biol Macromol* **117,** 493–522.

21 Falkenberg F, Voß L, Bott M, Bongaerts J & Siegert P (2023) New robust subtilisins from halotolerant and halophilic Bacillaceae. *Appl Microbiol Biotechnol* **107,** 3939–3954.

22 Yoon J-H, Lee C-H & Oh T-K (2005) *Bacillus cibi* sp. nov., isolated from jeotgal, a traditional Korean fermented seafood. *Int J Syst Evol Microbiol* **55,** 733–736.

23 Stropko SJ, Pipes SE & Newman JD (2014) Genome-based reclassification of *Bacillus cibi* as a later heterotypic synonym of *Bacillus indicus* and emended description of *Bacillus indicus*. *Int J Syst Evol Microbiol* **64,** 3804–3809.

24 Patel S & Gupta RS (2020) A phylogenomic and comparative genomic framework for resolving the polyphyly of the genus *Bacillus*: Proposal for six new genera of *Bacillus* species, *Peribacillus* gen. nov., *Cytobacillus* gen. nov., *Mesobacillus* gen. nov., *Neobacillus* gen. nov., *Metabacillus* gen. nov. and *Alkalihalobacillus* gen. nov. *Int J Syst Evol Microbiol* **70,** 406–438.

25 Kawamura F & Doi RH (1984) Construction of a *Bacillus subtilis* double mutant deficient in extracellular alkaline and neutral proteases. *J Bacteriol* **160,** 442–444.

26 Bertani G (1951) Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J Bacteriol* **62,** 293–300.

27 Heinze S, Kornberger P, Grätz C, Schwarz WH, Zverlov VV & Liebl W (2018) Transmating: conjugative transfer of a new broad host range expression vector to various *Bacillus* species using a single protocol. *BMC Microbiol* **18,** 56.

28 Falkenberg F, Rahba J, Fischer D, Bott M, Bongaerts J & Siegert P (2022) Biochemical characterization of a novel oxidatively stable, halotolerant, and high-alkaline subtilisin from *Alkalihalobacillus okhensis* Kh10-101^{T}. *FEBS Open Bio* **12,** 1729–1746.

29 Bradford MM (1976) 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.

30 Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD & Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7,** 539.

31 Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A & Finn RD (2018) The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res* **46,** D624-D632.

32 Falkenberg F, Bott M, Bongaerts J & Siegert P (2022) Phylogenetic survey of the subtilase family and a data-mining-based search for new subtilisins from *Bacillaceae*. *Front Microbiol* **13,** 1017978.

33 Vojcic L, Despotovic D, Martinez R, Maurer K-H & Schwaneberg U (2012) An efficient transformation method for *Bacillus subtilis* DB104. *Appl Microbiol Biotechnol* **94,** 487–493.

34 Zhang G, Wang W, Deng A, Sun Z, Zhang Y, Liang Y, Che Y & Wen T (2012) A mimicking-of-DNA-methylation-patterns pipeline for overcoming the restriction barrier of bacteria. *PLoS Genet* **8,** e1002987.

35 Priyanka VA, Ulfah M, Mulyawati L, Abinawanto & Helianti I (2022) Conjugational tansformation of wild Type *Bacillus halodurans* CM1 by methylated recombinant plasmid harbouring a gene encoding for alkaline protease and the protease activity assay of the transformant. *HAYATI J Biosci* **29,** 799–805.

36 Rozanov AS, Shekhovtsov SV, Bogacheva NV, Pershina EG, Ryapolova AV, Bytyak DS & S EP (2021) Production of subtilisin proteases in bacteria and yeast. *Vavilov J. Genet. Breed* **25,** 125–134.

37 Le Duc H, Fraser PD, Tam NKM & Cutting SM (2006) Carotenoids present in halotolerant *Bacillus* spore formers. *FEMS Microbiol Lett* **255,** 215–224.

38 Khaneja R, Perez-Fons L, Fakhry S, Baccigalupi L, Steiger S, To E, Sandmann G, Dong TC, Ricca E, Fraser PD & Cutting SM (2010) Carotenoids found in *Bacillus*. *J Appl Microbiol* **108,** 1889–1902.

39 Köcher S, Breitenbach J, Müller V & Sandmann G (2009) Structure, function and biosynthesis of carotenoids in the moderately halophilic bacterium *Halobacillus halophilus*. *Arch Microbiol* **191,** 95–104.

40 Waldeck J, Meyer-Rammes H, Wieland S, Feesche J, Maurer K-H & Meinhardt F (2007) Targeted deletion of genes encoding extracellular enzymes in *Bacillus licheniformis* and the impact on the secretion capability. *J Biotechnol* **130,** 124–132.

41 Chen W, Li L, Ye C, Zhao Z, Huang K, Zou D & Wei X (2022) Efficient production of extracellular alkaline protease in *Bacillus amyloliquefaciens* by host strain construction. *LWT* **163,** 113620.

42 Wu XC, Lee W, Tran L & Wong SL (1991) Engineering a *Bacillus subtilis* expression-secretion system with a strain deficient in six extracellular proteases. *J Bacteriol* **173,** 4952–4958.

43 Degering C (2011) Optimization of heterologous protein secretion in *Bacillus* with non-homologous and artificial signal peptides.

44 Freudl R (2018) Signal peptides for recombinant protein secretion in bacterial expression systems. *Microb Cell Fact* **17,** 52.

45 Drejer EB, Hakvåg S, Irla M & Brautaset T (2018) Genetic tools and techniques for recombinant expression in thermophilic *Bacillaceae*. *Microorganisms* **6**.

3 Discussion

Subtilisins are extremely versatile serine proteases from the subtilase family. Due to their favourable properties such as thermostability, broad pH range and broad substrate specificity, subtilisins are of particular interest for various industries producing detergents, cosmetics, or pharmaceuticals or being involved in leather processing or wastewater treatment [54]. As a result, there is scientific and commercial interest in finding new subtilisins and, moreover, understanding the relationship between amino acid sequence and biochemical properties can only be improved if more of these enzymes are comprehensively studied.

One subject of this thesis is an updated overview of the classic subtilase family (D-H-S family), as with the increasing number of proteases characterised, the classification reported by Siezen and Leunissen in 1997 was no longer up to date [24]. The current overview was applied to the categorisation of newly found subtilisin sequences derived from a data mining search in public databases. In particular, subtilisins from halophilic or halotolerant *Bacillaceae* have the greatest potential for meeting industrial needs [72, 107, 108]. Selected subtilisin genes were isolated from their wild-type origin and cloned into a expression plasmid. Transformation of *B. subtilis* DB104 and selection of positive clones showing clearing zones around colonies on skim-milk agar plates was followed by a first productivity assay in shake flasks. Productive clones were subjected to production in a 1-L bioreactor and the subtilisins were subsequently purified chromatographically (Figure 5). In addition, the wild-type strains were investigated for the first time for their genetic accessibility and their potential for the production of proteases. Here, *Metabacillus indicus* DSM 16189 proved to be a promising candidate. In the following, the enzymes used in this work were classified with regard to their biochemical properties in comparison to each other and to industrially relevant subtilisins. In addition, the subtilisins were evaluated for use in industrial applications, with a special emphasis on their use in detergents.

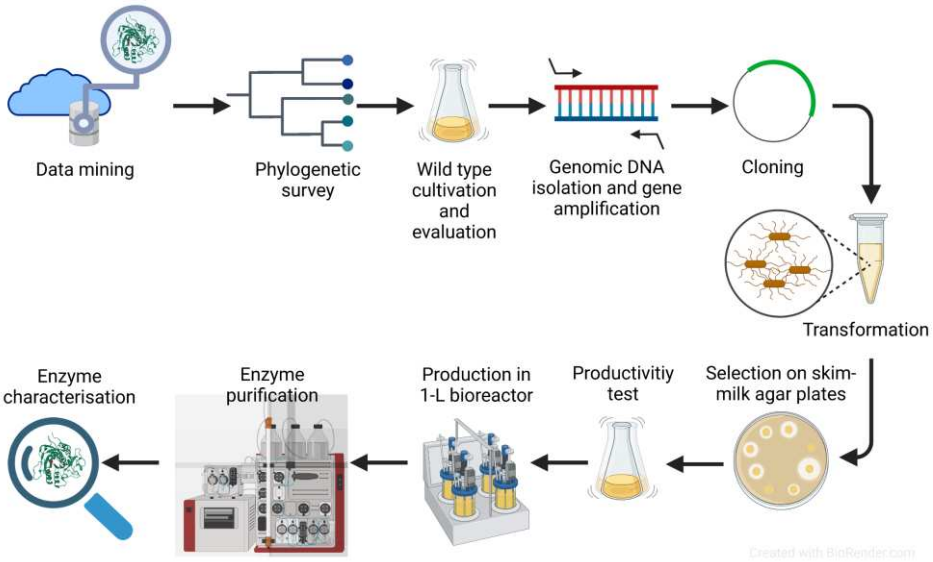


Figure 5 Schematic workflow of this thesis.

3.1 Phylogenetic survey of the subtilase family

Understanding the phylogeny and relationships between proteases within the subtilase family has proven to be difficult. As outlined in 1.1.3, the subtilase family is the third largest family of serine proteases, with regard to the number of sequences as well as the characterised proteases [9]. It is a very diverse family with low sequence identity, often restricted to regions around the catalytic triad [33]. When searching for new subtilisin sequences and placing these sequences in the context of the subtilase family, it is necessary to have a comprehensive phylogenetic overview of this family. In addition, it is worth noting that in the literature the terms subtilisins, subtilisin-like proteases and subtilases are sometimes used misleadingly, leading to confusion. The phylogenetic overview addresses this issue. The MEROPS S8 holotype dataset with the mature protease sequences and a selection of biochemically characterised proteases from the PDB were used for the analysis. The advantage of using the MEROPS S8 holotype dataset was that other structural domains, if present, can be excluded. Signal peptides and propeptides were excluded as these parts usually have low sequential similarity and are therefore not suitable for phylogenetic comparisons [163, 164]. In addition, C-terminal domains, if present, were excluded because not all subtilases contain such domains, which would affect the multiple sequence alignment (MSA).

When constructing a phylogenetic tree, several things need to be considered, as the quality and reliability of a phylogenetic tree is highly dependent on the accuracy of the underlying MSA [94]. It also depends on the tools used, as the creation of a phylogenetic context of a molecular sequence usually involves several steps, including the identification of homologous sequences, the MSA, the phylogenetic reconstruction and the graphical representation of the tree [165]. There are several algorithms for MSA that use different strategies: MAFFT [166], Muscle [167], Clustal Omega [168], or T-Coffee [169], to name a few. Most of these alignment programs can correctly identify highly conserved regions,

but a large number of insertions, deletions or additional domains can lead to a less reliable MSA and thus errors in the tree calculation [170]. Simultaneous use of different alignment methods to increase tree accuracy leads to only minor improvements [170]. Therefore, the identification and removal of unreliable parts of the alignment is performed by various automatic filtering methods (TrimAI [171], Gblocks [172], Noisy [173], BMGE [174]), which often lead to the deletion of positions that comprise a gap [175]. However, gaps may contain important phylogenetic information [170]. Furthermore, even a single incomplete or incorrect sequence in a dataset can lead to a large loss of information if a strict filtering method was used. Tan *et al.* 2015 compared several automatic filtering algorithms and concluded that they often reduce tree accuracy, while a less stringent filtering algorithm has a minimal effect on the accuracy of the phylogenetic tree, but it results in a compromise in terms of the computational time saved when constructing the tree [175]. Therefore, in this work, trimAI was used to curate the alignment, which uses a less stringent algorithm.

There are also various programs with different algorithms for creating a phylogenetic tree (IQTree [176], PhylML [177], FastME [178] FastTree [179], etc.). The Maximum Likelihood (ML) tree construction method was used, which is widely considered the most accurate molecular phylogenetic approach along with the Bayesian method [180, 165]. Statistically, the reliability of the constructed tree was assessed by using a bootstrap-based bias correction method, in which branch support is calculated by repeating the tree construction [181]. Nevertheless, a detailed comparison of a generated tree with the literature is difficult because the different alignment parameters, curation methods and different tree generation methods lead to different phylogenetic trees. The selected methods led to the workflow shown in the following Figure 6.

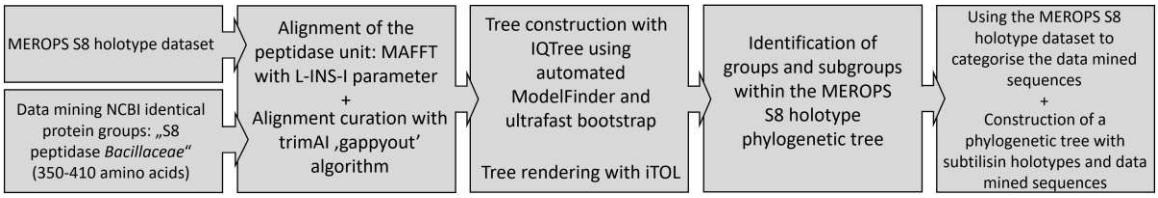


Figure 6 Workflow of data mining and phylogenetic analysis [182].

The constructed phylogenetic tree in Figure 2 (2.1) shows the phylogenetic overview of the subfamilies S8A and S8B with the classic catalytic triad (D-H-S). The S8B subfamily was chosen as the outgroup to create and restructure the tree, as the sequences are outside the ingroup but closely related to it [183]. As outlined in 1.1.3, Siezen and Leunissen divided the subtilases into six groups in 1997 [24], and the same division was observed in the newly constructed phylogenetic tree. Since the first classification, many new proteases have been described, so that the overview by Siezen and Leunissen is no longer up-to-date. Some of the newly described proteases have already been identified as new

groups or subgroups, but have never been fully phylogenetically placed in the context of the other subtilases. To account for the new diversity of sequences resulting from their different positions in the phylogenetic tree, their biological functions, biochemical properties, structural similarity and the taxa- and species-specific clusters formed, the classification of subtilase groups was revised and more precisely defined subgroups were proposed. The analysis resulted in the finding of eight previously uncharacterised groups and 13 subgroups within the S8 family (2.1). In particular, the group of pyrolysins harbours nine new subgroups, as it is a heterogeneous group of enzymes of different origins with low sequence conservation [24]. Furthermore, eight new groups, namely dentilisins, subAB subtilases, CDF subtilases, bpF subtilases, TPPS subtilases, apicomplexa subtilases, sporangins and perkinsins, originate from different sources and fulfil different biological functions, as discussed in 2.1. Interesting in the context of this thesis is the previously undescribed group of dentilisins, which lies between the thermitases and the subtilisins. Dentilisins and thermitases are co-located in the clade with subtilisins, which underlines their similarity.

Analysis of the sequences within the subtilisin group revealed four new subgroups, namely EPR subtilisins, PopC subtilisins, extremophilic subtilisins and Archaea subtilisins. The majority of the holotype sequences within the subtilisins originate from Bacilli, whereas aerolysin (S08.105) [57] and Tk-subtilisin (*Thermococcus kodakarensis* subtilisin) derives from Archaea (S08.129) [184], PopC, (S08.143) from Myxococcia [185], and ALTP (*Alkaliphilus transvaalensis* protease) (S08.028) [56] from Clostridia, respectively. The created phylogenetic tree of the subtilase family gives an overview of the pyhlogenetic relationship of the protease sequences. However, when the subgroups are further analysed, additional subdivision into classes is most likely, as it is described for plant pyrolysins [44, 186, 187] or fungal pyrolysins [188–190]. Nevertheless, it is now possible to place new sequences, whether from classical screening approaches or from data mining searches, in the context of the subtilases.

3.2 The search for new subtilisins

The search for new subtilisins is important as outlined in 1.1.4, as subtilisins find wide industrial application in various products and processes, such as detergents, cosmetics, pharmaceuticals, leather processing, and waste water treatment [54]. With the market volume of alkaline proteases such as subtilisins expected to reach 2.9 billion USD by the end of 2026, with the largest market share in detergents [54]. The subtilisins that are used in detergents are mainly isolated from species such as *Bacillus licheniformis* with subtilisin Carlsberg [68], *Lederbergia lenta* (formerly *Bacillus lentus*) with Savinase [69] and *B. amyloliquefaciens* with BPN' [23]. Many enzyme engineering studies have been conducted to improve these known proteases for industrial use and mutations in more than half of the amino acids of the known subtilisins have been reported [79, 80, 82, 72]. However, the existing stock of enzymes is not yet sufficient to meet the needs of industry.

Therefore, there is still a need to discover and develop new and better enzymes [91]. With its almost unlimited microbial biodiversity, nature offers great potential for enriching the repertoire of subtilisins [92].

As outlined in 1.1.5, there are several methods to exploit this potential. Besides the time-consuming classical approach of isolating and cultivating microorganisms secreting subtilisins or creating metagenomic libraries, the exponentially growing online database contains previously unused sequence information [91, 93, 94]. Only a small proportion of the publicly available sequences have been subjected to experimental characterisation [191]. Therefore, the genomic data search for new enzymes is becoming an established routine [91]. As pointed out in 1.1.5, several bacteria within the genus *Bacillus* have been classified into distinct clades, and therefore new subtilisin sequences were searched from the entire family of *Bacillaceae*. However, when searching for S8 peptidases from *Bacillaceae*, the number of available sequences is still overwhelming with 18.881 sequence entries compared to 247.897 without the family refinement at the time of the survey. Therefore, another search parameter was set. Since the subtilisins of the *Bacillaceae* have a very similar length of about 380 amino acids, including the signal peptide and the propeptide [61–64], the peptide sequence length filter was set to 350 - 410 amino acids. Major differences in length are in the surface loops or at the termini [192]. In addition, the analysis in 2.1 showed that sequences typical of AprX, lantibiotic peptidases, kexins, oxidatively stable proteases (OSP) and high-mass subtilases (HMS) were larger than the set filter and were therefore excluded. The AprX group could be of interest for industrial applications, as Phrommao *et al.* noted that the subtilase from *Virgibacillus* sp. SK37 was relatively stable up to 5 % (v/v) H_2O_2 and active under alkaline conditions [193]. However, this subtilase is not secreted and therefore not predestined for easy production.

The set parameters still yielded 1424 sequences, which were divided into 375 clusters with an identity threshold of 85 %, of which one representative sequence was used for further analysis. As mentioned above, only secreted subtilisins are of interest, reducing the dataset from 375 to 240 sequences. Using the sequences from the phylogenetic survey, it was easy to identify the sequences that did not fall within the sequence space of subtilisins, leaving a set of 120 sequences. Most of the discarded sequences were identified as thermitases and dentilisins. Although thermitases and dentilisins are phylogenetically closely related to subtilisins, as shown in Figure 2 (2.1), they were excluded because of the focus on subtilisins. Using the sequences forming the group of subtilisins derived from the phylogenetic overview of the subtilases, it was possible to construct a phylogenetic tree with the sequences obtained (2.1). Interestingly, the resulting phylogenetic tree showed that all remaining 120 sequences are found in the three subgroups PIS, high-alkaline subtilisins and true subtilisins with a sequence identity of 55 – 58 % to each other. None of the remaining sequences belong to the newly defined

subgroups of EPR subtilisins, PopC subtilisins, extremophilic subtilisins and Archaea subtilisins, as these subgroups are either not formed by *Bacillaceae* or the sequences fall outside the defined length parameters. The subgroups have a sequence identity of less than 45% with each other and with the three subgroups PIS, high-alkaline subtilisins and true subtilisins.

Most of the sequences found in the data mining search are phylogenetically classified within the subgroup of true subtilisins near well-characterised representatives, as shown in the bar chart with cluster sizes up to 127 sequences in Figure 2 (2.1). This means that the majority of the publicly available sequences have a high sequence identity with the known representatives, which could be due to a certain bias with which the bacteria were isolated, analysed and deposited in the database. This assumption is supported by the fact that the three largest clusters mainly harbour sequences from *B. pumilus*, *B. subtilis* and *B. safensis*. Interestingly, Salwan *et al.*, who reviewed the scientific literature from 2000 to 2019 dealing with proteases, especially in detergents, showed that 62 % of proteases originate from mesophilic bacteria [72]. This tendency can also be observed in the sequences found in this work, as most of them originate from mesophilic *Bacillaceae*. The phylogenetic tree helped to classify the sequences found in comparison to the characterised subtilisins from the MEROPS database. The advantage of using phylogenetics in the discovery of new subtilisins is that the phylogenetic relationship to already known enzymes can be quickly visually estimated and conclusions can possibly be drawn about their biochemical properties. The disadvantage, however, is that, as explained in 3.1, the quality of the phylogenetic tree and its informative value is strongly influenced by the quality of the underlying sequence alignment [94].

A comprehensive examination of all the insertions and deletions among the three subgroups of PIS, high-alkaline and true subtilisins revealed that each subgroup showed conserved mutations. In particular, for each subgroup, specific mutations are found in the loop of amino acids 50-59, a known region of subtilisin structure that exhibits high variability [70]. It is interesting that high-alkaline subtilisins show a deletion of four amino acids around position 160 [69], while all PIS have an insertion between positions 159 and 160. This region is localised in a loop that is involved in the substrate binding [194]. Furthermore, a shorter loop can increase the stability of an enzyme, which is confirmed by comparison with some highly thermostable proteases of the proteinase K and thermitases group, where this loop is also shortened [195, 196]. It is also worth mentioning that all insertions or deletions are located on the surface of the protease. This could be due to the fact that the overall structure within the subtilisins is highly conserved [197].

The number of 120 remaining sequences within the phylogenetic tree analysis was still too large to analyse all of the corresponding enzymes biochemically and therefore a look

at the microbial origin was important. As described in 1.1.6, enzymes derived from bacteria that can survive in harsh conditions are of particular interest. Therefore, special emphasis in the selection was placed on bacteria that are either halotolerant or halophilic, as such species proved to harbour enzymes capable of withstanding various extreme conditions relevant to the industry [72, 111]. In detail, six sequences belonging to the true subtilisin subgroup were selected from *Pontibacillus marinus* DSM 16465^T, *Metabacillus indicus* DSM 16189, *Litchfieldia alkalitelluris* DSM 16976^T, *Fictibacillus arsenicus* DSM 15822^T, *Alcalicoccus saliphilus* DSM 15402^T and *Halobacillus faecis* DSM 21559^T. In addition, sequences of the phylogenetically intermediate subgroup from *Alkalibacillus haloalkaliphilus* DSM 5271^T, *Alkalicoccus daliensis* DSM 26767^T and *Piscibacillus halophilus* DSM 21633^T were selected. From the subgroup of high-alkaline subtilisins, sequences from *Halalkalibacter okhensis* Kh10-101^T (former *Alkalihalobacillus okhensis*) and *Halalkalibacterium ligniniphilum* DSM 26145^T were chosen.

P. marinus, *A. saliphilus*, *A. haloalkaliphilus, P. halophilus* and *H. okhensis* are moderately to extremely halophilic [121, 198–201]. *M. indicus*, *L. alkalitelluris, A. daliensis, H. ligniniphilum* and *H. faecis* are moderately to extremely halotolerant [159, 202–205], while *F. arsenicus* tolerates only low concentrations of NaCl (1 % w/v) [206]. However, as pointed out in 1.1.6, not all enzymes from extremophilic bacterial origin are extremozymes [101]. Therefore, an in-depth bioinformatic analysis, the cloning of the subtilisin genes, the recombinant expression and production by *B. subtilis* DB104, and the purification and biochemical characterisation is discussed in the following chapters. The well-known subtilisins Savinase, subtilisin Carlsberg and BPN' were used for comparison.

3.3 Recombinant protease production and purification

For the biochemical characterisation of the proteases, it is necessary to produce the enzymes in sufficient quantities. For this purpose, the proteases were produced recombinantly with *B. subtilis* DB104. Representatives of the genus *Bacillus* have been used for the industrial production of enzymes and offer several advantages compared to *E. coli*. As there is no outer membrane, protein secretion is facilitated, thereby simplifying separation from cell biomass, which in turn reduces process costs and facilitates downstream processing [143]. Furthermore, cost-efficient production is possible through the use of agro-industrial residues such as sugar cane molasses and corn steep liquor as well as various types of bran and cake [144]. As outlined in section 1.2, *Bacillus* species naturally secrete various hydrolytic enzymes such as proteases and amylases and are therefore the ideal production system for the recombinant production of proteases, in particular by deleting the genes responsible for endogenous protease production [147]. However, due to the natural secretion of proteases, the mutant strain *B. subtilis* DB104 was used, a double mutant deficient in extracellular alkaline and neutral proteases, with a residual protease activity of 2 - 4 % compared to the wild-type [207]. For the cloning of the protease genes it is necessary to analyse its gene and the encoded amino acid sequence, which will be further outlined in 3.4.

The successful transformation and production of the protease in *B. subtilis* DB104 was checked by the appearance of clearing zones around the colonies on LB/skim-milk agar plates. This selection method is a decisive advantage, as in addition to successful transformation, the secretion and functionality of the recombinant protease can be checked at the same time. Correct cloning was then further confirmed by Sanger sequencing. Before producing the proteases in a 1-L fermentation, the activities of the proteases were evaluated within a shake flask production experiment. Here, the proteases from *A. daliensis* and one protease from *P. halophilus* were sorted out due to very low activity (data not shown). Furthermore, the proteases of *H. ligniniphilum, P. halophilus, A. saliphilus* and *H. faecis* were not produced sufficiently after fermentation and therefore showed only low activity in the supernatant of the fermentation. Thus, only the six proteases from *P. marinus*, *A. haloalkaliphilus*, *H. okhensis*, *M. indicus*, *L. alkalitelluris* and *F. arsenicus* were selected for further evaluation. As mentioned in 1.2, secreted protein concentrations up to 20 g/l are possible in highly optimised industrial fermentation processes with *Bacillus* strains [147]. In the unoptimised fermentation of the six proteases, only 0.4 - 0.6 g/l extracellular total protein could be achieved. However, this can be due to the fact that heterologous proteins were secreted [139]. It is known that the yield of heterologous proteins is often lower than that of native proteins or that they are not secreted at all [155]. The observation that not all cloned proteases were produced in sufficient quantities is therefore not unexpected. Furthermore, the fermentation process is not optimised and efficient protease production also depends on nutritional factors such as the fermentation medium and physicochemical parameters

such as initial pH, agitation, incubation time, temperature and the protease activity itself [144].

Since the proteases are secreted, the supernatant was used for a three-step purification process. Purification is a critical step, as proteases tend to autoproteolytic cleavage under unfavourable conditions [208]. However, it was possible to purify all proteases in almost pure form, as confirmed by SDS-PAGE. While the observed molecular masses of the proteases SPAO and SPFA matched their calculated values, SPPM, SPAH, and SPMI exhibited a molecular mass exceeding their theoretical values. Although it is known that most reduced proteins bind SDS in a constant weight ratio under suitable conditions [209], an altered migration behavior deviating from the expected molecular mass of the proteins during SDS-PAGE is not unusual and is referred to as "gel shifting" [210]. This may be due to partial refolding or altered detergent binding [211, 210]. Therefore, the molecular mass of the purified proteases was determined by MALDI-TOF MS analysis. The values obtained for SPMI, SPLA, SPAH, SPAO and SPFA were in agreement with the predicted values (Table 3). MALDI-TOF MS revealed a mass of 27.97 kDa for SPPM, a result in proximity to the predicted value of 27.90 kDa for the variant lacking the nine C-terminal amino acids, as discussed in section 3.4.

The isoelectric point plays a role in understanding the relationship between surface charge and biochemical properties, as will be explained later, and was also important for enzyme purification as ion exchange chromatography was used. The experimentally determined pI values of the six proteases almost agree with the calculated pI values (Table 1). Here, the difference between the high-alkaline subtilisin SPAO and the others was observed, as SPAO with a pI of 9.8 is in line with other high-alkaline subtilisins that have a high pI in common [212]. An acidic pI for proteins from halophilic or halotolerant sources is known, as shown by SPPM, SPMI, SPLA and SPAH [131]. SPFA, although from a microbial source that can tolerate only low concentrations of salt (1 % NaCl), also exhibited an acidic pI [206]. Other true subtilisins, such as subtilisin Carlsberg and BPN', have a neutral to slightly basic pI [213].

The activities after purification of the six proteases were measured with the two substrates suc-AAPF-pNA and azocasein (Table 2), as described in 2.2. Higher activities were obtained for azocasein, as it is a more complex substrate with more potential hydrolysis sites. Comparing the activities with literature data is of limited value as different substrates or experimental conditions were used. Under the same assay conditions, subtilisin E with 486 U/mg (azocasein) and 25 U/mg (AAPF) was obtained after production with *B. subtilis* and purification by Li *et al.* [214].

Table 3 Molecular mass, pI and activities after purification

Protease	Calc. mol. mass	Exp. mol. mass	Calc. pI	Exp. pI	U/mg (AAPF)	U/mg (Azocasein)
SPPM	27.90	27.97	4.2	~4.3	208	1371
SPMI	27.48	27.49	5.5	~5.5	160	1085
SPLA	27.47	27.48	5.1	~5.0	233	314
SPAH	28.60	28.60	4.3	~4.9	314	2719
SPAO	27.14	27.12	9.6	~9.8	139	528
SPFA	27.57	27.57	5.8	~5.8	195	539

3.3.1 Protease production with *M. indicus* DSM 16189

Due to the success of *B. subtilis* as a host for the production of recombinant proteins, other *Bacillaceae* were also investigated for this application [215–218]. In this thesis the wild-type strains from which the new subtilisin sequences were obtained were investigated for the first time for their cultivability, genetic accessibility, and their potential for the production of proteases (data not shown). Among the strains studied, *Metabacillus indicus* DSM 16189 emerged as a promising candidate (2.5). *M. indicus* has some beneficial properties, such as the production of carotenoids that turn the colonies orange, which helps to quickly detect possible contaminations. The ability to withstand salt concentrations of up to 12 % (w/v) [159], can reduce the need for antibiotics. In addition, saline waste products such as molasses could be directly used as substrate [219]. *M. indicus* was isolated from a traditional food, so its products could gain GRAS status. [220]. In addition, the closely related strain *M. indicus* HU36 is used as a probiotic [221].
In order to test its genetic accessibility, methods such as electroporation [222], natural competence [223] and protoplast transformation [224] were evaluated (data not shown). However, the only successful method was the triparental conjugation according to Heinze *et al.* [162]. The method involves an additional cloning step in *E. coli* DH5α compared to direct transformation with the other methods. *E. coli* DH5α harbouring the target plasmid serves as the donor strain for the transformation of *M. indicus* by conjugation together with the helper strain *E. coli* HB101 harbouring the plasmid pRK2013, which contains a broad host-spectrum transfer system that promotes conjugation transfer [162]. The advantage of halotolerance of *M. indicus* was demonstrated when, after conjugation, a selection method was lacking, so that a higher NaCl concentration helped to inhibit the growth of donor and helper strains (2.5). Cultivation of recombinant *M. indicus* for 48 hours showed that the strain is able to produce its homologous protease recombinantly in sufficient quantities, as shown by SDS gel and activity measurements with an average of 169.7 U/ml in the suc-AAPF-pNA assay and 1426 U/ml in the azocasein assay. While the wild-type strain showed only low protease production with an average of 4.2 U/ml activity in the suc-AAPF-pNA assay and 121 U/ml in the azocasein assay. A comparison with the

fermentation experiment of *B. subtilis* DB104 harbouring the same expression plasmid showed that *M. indicus* produces about twice the amount of the protease when comparing the volumetric enzyme activity in the suc-AAPF-pNA assay. However, the final cell density of *M. indicus* is about 2.5 times higher than that of *B. subtilis* DB104. Therefore, protease production is more efficient in *B. subtilis* than in *M. indicus*. The difference in cell growth can be explained by the use of different media, as growth of *M. indicus* is insufficient in the medium of *B. subtilis* and *vice versa*. The total amount of nitrogen in the medium of *M. indicus* is about 3 % higher than for *B. subtilis,* while the amount of carbon is comparable. The production of the proteases in *Bacillus* species is regulated by the *aprE* promoter that is activated at the beginning of the stationary phase alongside with the sporulation [225] The transition towards the stationary phase can be caused by carbon and nitrogen limitations [144]. Furthermore, the agitation parameter that maintains an adequate amount of dissolved oxygen might not be optimal for *B. subtilis.*

An improvement in production can be achieved by a combination of methods with the optimisation of process parameters and by genetic modification of the production strain. A combination of both led to remarkable results for the strain *B. amyloliquefaciens* K11, which produced its own neutral protease recombinantly more efficiently than the wild type and an industrial *B. subtilis* strain [226]. By applying the process and strain development of the industrial production strain *B. licheniformis* to *B. pumilus*, the protease titre was increased by 14 % [227]. Zhang *et al.* recently showed that the introduction of multiple translation initiation sites increases protein production in *Bacillus licheniformis* [228]. Production can be further improved by testing different promoters, e.g. the native subtilisin promoter of *M. indicus* or other promoters that have been shown to be effective [227, 215], because the plasmid used contains a subtilisin promoter from *B. subtilis* [162]. However, an increase in transcription and translation might not be helpful if secretion is insufficient. Therefore, optimising secretion by screening signal peptides can improve production yield, although the signal peptide must be assessed individually for each protein [153, 157]. Further improvement can be achieved by relieving the secretion system by reducing the number of secreted proteins through genome minimisation [229]. The knock out of spore formation can also reduce the metabolic burden and sterilisation problems. However, it was not possible to induce sporulation in *M. indicus* although it was originally reported [159]. Hence, *M. indicus* proved to be suitable for the recombinant production of its homologous protease, and it remains to be seen whether this potential can be exploited with heterologous proteases, especially those that cannot be produced with *B. subtilis* DB104.

3.4 Bioinformatic analysis and homology modelling

For the cloning of the proteases, it is necessary to analyse their gene and the encoded amino acid sequence. Moreover, the use of bioinformatic analysis not only aids in the interpretation of results from biochemical characterisation, but also has the potential to reveal possible connections between protein structure, amino acid composition, and biochemical properties. As described in 1.1.3, most subtilases have a multi-domain structure with a signal peptide for translocation, a propeptide for maturation, a peptidase domain and often one or more additional domains [33]. The genes of the proteases SPPM, SPMI, SPLA, SPAH, SPAO and SPFA comprise 1128 bp to 1152 bp with 375 to 383 amino acids. In the case of SPPM, the automatic annotation was corrected by an extension of the open reading frame by eight codons at the 5' end, as the gene would be lacking eight amino acids of the signal peptide. This extension leads to a TTG start codon, which is not unusual for *Bacillus* sp. [230]. Signal peptide prediction revealed that for all six proteases there is a probability of more than 96 % of a Sec signal peptide with a length of 23 to 32 amino acids. In general, signal peptides do not have sequence similarities, but feature a typical composition of three parts, with a positively charged N-terminus, a central hydrophobic core and a polar C-terminal domain containing the recognition site of the signal peptidase [149]. This structure can also be observed in the signal peptides of the six proteases and subtilisin Carlsberg, BPN' and Savinase. Additionally, the signal peptides have less than 50 % sequence identity to each other [231, 213].

Identification of the propeptide by multiple sequence alignment revealed sizes from 71 to 86 amino acids, and the sequence identity of the propeptide between the six proteases, subtilisin Carlsberg, BPN' and Savinase is less than 59 %. As outlined in section 1.1.3, the propeptide is present in almost all subtilases and acts as an intramolecular chaperone during maturation, supporting the folding of the catalytic domain and acting as an inhibitor of the subtilase [50]. Without its propeptide, the peptidase domain alone is unable to fold into its native structure [47]. Interestingly, Shinde *et al.* showed that subtilisin E folds differently when a mutant propeptide is used, such that differences in secondary structures, thermostability, and substrate specificity were observed. The propeptide facilitates the folding of the subtilisin by acting as a chaperon [232, 164]. Comparison of the C-terminus of the proteases showed that SPPM has an extension of ten amino acids compared to BPN' and nine amino acids compared to SPAH. Without the nine amino acid extension, the theoretical mass of SPPM is 27.90 kDa, which is consistent with the MALDI-TOF MS analysis (3.3). It appears that the C-terminus must be processed in some way by other SPPM molecules or another protease.

Calculation of the theoretical isoelectric point revealed an acidic pI of 4.2 - 5.8 for SPPM, SPMI, SPLA, SPAH and SPFA and an alkaline pI of 9.6 for SPAO. As described in 1.1.7, an acidic pI is typical for halophilic or halotolerant proteins and an alkaline pI is typical for high-alkaline subtilisins. The MSA showed that the six proteases have deletions and

insertions typical of their subgroup (see 3.2). SPAH has the double insertion within a loop between position 42 and 43 common to phylogenetically intermediate subtilisins studied in this work (2.1). Unlike the others, SPAO shares the four-amino acid deletion around position 160, which is located within a loop near the substrate binding site. Sequences of the PIS subgroup have an additional amino acid in this loop compared to the high-alkaline and true subtilisins, which is also observed in SPAH. Since shorter loops can increase stability, the insertion could have a negative effect for SPAH in terms of stability [195].

As can be seen from the phylogenetic tree, SPAH belongs to the subgroup of phylogenetically intermediate subtilisins. SPPM, SPMI, SPLA and SPFA belong to the subgroup of true subtilisins and SPAO to the high-alkaline subtilisins. This also becomes clear when comparing the sequence identities, as shown in Table 4. SPPM, SPMI, SPLA and SPFA have a sequence identity with the other true subtilisins BPN' and Carlsberg of 65.09 to 74.82 %, while SPAO, in contrast, has 82.16 % sequence identity with Savinase. SPAH shows a high identity with the true subtilisin SPLA at 61.45 %. However, SPAH is clearly a member of the PIS subgroup, as it has the PIS-typical mutations and shows a sequence identity of 73.4 % with the PIS LD1 from *Bacillus* sp. KSM-LD1 [58], and cluster within the PIS subgroup in the phylogenetic tree (2.1). As mentioned in 1.1, the MEROPS database assigns unique identifiers to an enzyme of a biochemically characterised protein species. For each protein species, there is a holotype that represents the first biochemically characterised peptidase [10]. Therefore, SPPM, SPMI, SPLA, SPAH and SPFA most likely form new holotypes. However, SPAO can be assigned to subtilisin sendai (S08.098) from an alkaliphilic *Bacillus* sp. G-825-6 with a sequence identity of 91.45 % [194]. A new holotype identifier would be formed if the substrate specificity of SPAO differs from subtilisin sendai [10]. However, the available data are difficult to compare due to the use of different substrates and experimental conditions [194].

Table 4 Percent identity matrix of the mature protease sequences examined in the biochemical study

	Savinase	SPAO	SPAH	BPN'	Carlsberg	SPPM	SPMI	SPFA	SPLA
Savinase	100.00	82.16	55.39	60.97	61.94	55.39	60.97	62.83	63.20
SPAO	82.16	100.00	53.53	56.13	58.21	55.02	54.65	58.74	57.99
SPAH	55.39	53.53	100.00	53.09	56.20	55.80	59.27	58.55	61.45
BPN'	60.97	56.13	53.09	100.00	69.71	65.09	70.18	69.09	70.18
Carlsberg	61.94	58.21	56.20	69.71	100.00	67.88	72.63	72.63	74.82
SPPM	55.39	55.02	55.80	65.09	67.88	100.00	69.82	74.18	76.36
SPMI	60.97	54.65	59.27	70.18	72.63	69.82	100.00	73.82	78.55
SPFA	62.83	58.74	58.55	69.09	72.63	74.18	73.82	100.00	82.55
SPLA	63.20	57.99	61.45	70.18	74.82	76.36	78.55	82.55	100.00

Red: high-alkaline subtilisins; green: PIS; blue: true subtilisins. Percent identity was calculated with Clustal Omega.

To understand the structure-related biochemical properties, a 3D model of the proteases is helpful. Therefore, a structure prediction was performed with the mature proteases using the I-TASSER server [233]. Since the overall structure of subtilisins is conserved and several crystal structures are available as templates, homology modelling is possible and sufficient [197]. The quality of the models can be assessed with the C-score, which ranges from -5 to 2, with higher values indicating a higher reliability of the model [233]. The C-score for the models of SPPM, SPMI, SPLA, SPAH, SPAO and SPFA is above 1.4, so that a high quality can be assumed. When analysing the 3D structure of SPPM with the nine amino acid long C-terminal extension mentioned above, this extension projects away from the core molecule (data not shown), supporting the assumption that it is susceptible to proteolytic processing. With the 3D structures, it is possible to analyse the electrostatic potential of the molecular surface, which contributes to the interpretation of the adaptation to high pH and NaCl concentrations, as discussed later. Furthermore, it is possible to analyse the homology models for metal binding sites. As expected, all six proteases contain two Ca^{2+}-binding sites, as commonly observed for subtilisins from *Bacillus* species, since the binding of Ca^{2+} is essential for the stability and activity of the enzyme [24, 64]. The two binding sites differ in the strength of Ca^{2+}-binding [24]. The first site binds strongly to Ca^{2+} ($K_d \approx 1 \times 10^{-8}$ M) and is highly conserved in subtilisins, requiring the side chains of residues Gln^{2} and Asp^{41} as well as several side chains of loop-forming residues 75-81 (Savinase numbering) [24, 69, 234]. This site is not formed before autoprocessing of the propeptide, because the structural reorganisation of the N-terminal region during autoprocessing is necessary for the formation of this site [235]. The second binding site is a weak binding site ($K_d \approx 32$ mM) and is occupied by Na^+ or K^+ ions at low calcium ion concentration in solution, involving the side chain of Ala^{169}, Ala^{174} and Tyr^{171} (Savinase numbering) [64, 69, 234]. Other subtilases may have the Ca^{2+}-binding sites at different positions in the structure than the subtilisins, as a comparison with proteinase K and thermitase shows [192]. In addition, seven Ca^{2+}-binding sites are known for the

Tk-subtilisin from *Thermococcus kodakarensis* and three for the subtilase Ak1 from *Bacillus* sp. Ak.1 [236, 237].

3.5 Comparable biochemical characterisation of the new subtilisins

Besides the research of proteases for industrial use, the biochemical characterisation of the different subtilisins is necessary to understand the structure-function relationship of the different backbones and their biochemical properties [24, 144]. As mentioned in 1.1.7, the various subgroups of subtilisins differ in their biochemical characteristics. The differences between high-alkaline subtilisins, PIS, and true subtilisins are mainly that the high-alkaline subtilisins have a high activity in the alkaline pH range. In the following, the six subtilisins are evaluated comparatively with regard to their suitability for industrial use and in the context of the literature.

3.5.1 Effects of temperature and pH on enzyme activity and stability

The optimum temperature of the investigated subtilisins ranged between 50 - 80 °C (2.2, 2.3, 2.4). The highest temperature optimum was displayed by SPFA at 80 °C, followed by SPMI and SPLA at 70 °C. The lowest optimum temperature was found for SPAH at 50 °C, while SPPM and SPAO showed the highest catalytic activity at 55 °C (Table 5). The commercial proteases subtilisin Carlsberg, BPN' and Savinase showed a temperature optimum of 65 °C, 55 °C and 60 °C, respectively. The environmental factors of the habitat in which the organisms live shape their physiological and evolutionary adaptation, and as already mentioned in 1.1.7, temperature is an environmental factor that overcomes physical barriers [101]. Consequently, unlike extreme pH or salinity, organisms are unable to counteract higher temperatures by relying on steep concentration gradients across biological membranes [125]. Extracellular enzymes are directly exposed to the environmental conditions of the bacterial habitat, so that conclusions can be drawn about their biochemical properties. Here, all proteases showed a temperature optimum above the maximum growth temperature of the species from which they were derived. *P. marinus* and *L. alkalitelluris* grow between 15 - 40°C and *H. okhensis* between 25 - 40 °C [198, 202, 238]. *M. indicus* and *A. haloalkaliphilus* grow at 4 - 50 °C and 15 - 45 °C, respectively [159, 239]. *F. arsenicus* grows between 20 and 40 °C [206]. The bacterial origins of commercial proteases also grow in the mesophilic range (15 - 50 °C) [98–100]. This observation is in line with the study of Engqvist, who compared the optimal enzyme temperatures with the optimal growth temperatures of all available enzyme data from the BRENDA database and the culture collection centre websites [125]. He reported that proteins from mesophiles (15 - 50 °C) tend to be catalytically active at higher temperatures than would be expected based on the growth temperature, and that proteins from thermophiles (50 - 80 °C) or hyperthermophiles (>80 °C) tend to have catalytic optima that are below the growth temperature. This trend may be due to the fact that random mutations in thermostable proteins are more likely to reduce stability

than a random mutation in a moderately stable protein [240]. In thermophiles and hyperthermophiles, extrinsic factors such as compatible solutes, enhanced effect of chaperones, molecular crowding, higher protein turnover rates and other mechanisms might play a role in addition to adaptations in protein sequence and folding [125].

Table 5 pH, temperature optima and melting points of the proteases

Microbial origin	Protease name	Subtilisin subgroup	Relative activity at pH 12.0 [%]	Temperature optimum [°C]	T_m [°C]	Bacterial growth optimum [°C]
P. marinus	SPPM	True	62	55	-	30
M. indicus	SPMI	True	62	70	62.5	30
L. alkalitelluris	SPLA	True	91	70	61.5	30
F. arsenicus	SPFA	True	87	80	62.5	30
A. haloalkaliphilus	SPAH	PIS	37	50	-	30
H. okhensis	SPAO	High-alkaline	60	55	53.0	37
B. licheniformis	Carlsberg	True	19	65	64.0	37
L. lenta	Savinase	High-alkaline	53	60	-	26
B. amyloliquefaciens	BPN‘	True	6	55	58.5	30

As outlined in 1.1.7, the sequence adaptation of thermophilic proteins to stability is attributed to an increased number of charged residues, disulphide bonds and a hydrophobic core [125]. This leads to a more compact conformation with less global flexibility [124]. Since temperature is the most important external factor for protein dynamics and catalytic efficiency, an increase in temperature leads to an increase in the kinetic energy of the atoms, which in turn can lead to increased mobility and instability of the proteins [124]. After reaching the temperature optimum, the enzyme begins to thermally denature, which can also be observed in all proteases examined, as the catalytic activity decreases after reaching the temperature optimum (2.2, 2.3, 2.4). Thermal denaturation or inactivation can be observed for SPMI, SPAO and SPFA already at 5 °C above the optimum, for SPPM at 10 °C and for SPLA and SPAH at 15 °C above the optimum, as the enzymes were not stable for the intended 5 min of the suc-AAPF-pNA assay. However, the investigation of the thermal resistance of a peptidase is difficult because autoproteolysis always plays a key role [136]. As already mentioned, thermophilic proteins have a more compact conformation, which also protects against autoproteolysis [137]. However, as the unfolding of certain regions of the protease begins, these regions act as substrates for the still native molecules [137, 208]. Hence, the irreversibility of thermal inactivation of proteases is mainly due to autoproteolysis [241]. Therefore, in order to monitor thermal protein unfolding rather than autoproteolysis, the melting points of the proteases were determined (2.2, 2.3, 2.4). In the thermal shift assay, the proteases were irreversibly inhibited by phenylmethanesulfonylfluoride (PMSF), so no proteolytic activity is expected [242]. SPMI and SPFA both had melting points of 62.5 °C, while SPLA and SPAO had their melting points at 61.5 °C and 53.0 °C, respectively. The two commercial reference proteases BPN' and subtilisin Carlsberg had melting points of

58.5 °C and 64.0 °C, respectively (Table 5). The measured melting points are consistent with the observed optimal temperature, as the proteases with the higher melting points also had a higher optimal temperature. As mentioned earlier, thermophilic enzymes require greater structural rigidity and a more compact conformation. As the temperature increases, the entropy of the solvent increases as the hydrogen bonds between the solvent molecules are weakened, forcing hydrophobic groups inside the protein, which stabilises the protein to a certain extent until the denaturation temperature is reached [126]. The high temperature destroys the stability of the protein structure, which could lead to the exposure of internal hydrophobic groups detected in the thermal shift assay. Unfortunately, it was not possible to determine the melting point for SPPM and SPAH with this assay. The reason could be protein precipitation, aggregation, a complicated folding landscape or a high hydrophobic background masking the melting transition [243]. The above mentioned autoproteolysis can be observed when testing thermal stability at 20 and 50 °C, as the temperatures chosen are below the temperature optima and melting points, but a decrease in activity can be observed for the time course of four hours. The activity of SPPM, SPLA and SPAH remained relatively stable at 20 °C with a residual activity of over 75% after four hours, while SPFA lost 75%, SPMI 85% and SPAO all their activity during this period. Subtilisin Carlsberg and Savinase also lost all their activity, while BPN' retained about one third of its activity. At 50 °C incubation temperature, the loss of activity was even more pronounced and only SPAH retained 10 % and SPLA 52 % of its activity, while the others lost almost all their activity after only one hour, indicating a stronger susceptibility to autoproteolysis [231]. Since the three commercially available enzymes also show a sharp reduction in activity after four hours of incubation, it can be assumed that these enzymes are highly active in the buffer used, which leads to autoproteolysis. This is a major problem in aqueous environments and the reduction of autoproteolysis in liquid detergents is overcome with boron compounds in combination with polyols [16, 244]. The supernatants of the fermentative production of the six proteases were stored with 10 % (v/v) 1,2-propanediol at 4 °C, with no loss of activity after one year, highlighting the stabilising effect of such polyols (data not shown).

In industrial applications in detergents, changes in washing preference towards low temperatures (15 - 25 °C) to reduce energy consumption lead to the search for psychrophilic proteases [16]. However, the proteases described in this work showed lower catalytic efficiency at 20 °C than at their temperature optimum. Cold-adapted subtilisins were found in nature, but their performance in terms of stability prevented their commercial use [79]. Therefore, for example, in the case of a mesophilic subtilisin from *Alkalihalobacillus gibsonii* (formerly *Bacillus gibsonii*), enzyme engineering campaigns simultaneously improved low-temperature activity and heat resistance [79]. In addition to the optimum temperature, temperature stability is also important for biotechnological applications, as the enzymes are stored for a long time in detergents, for example [16]. To achieve this, various enzyme engineering strategies were used, showing that sometimes

only specific mutations were required to increase the stability of a protease [82]. Stability can be increased by introducing mutations that reduce either the unfolding or the rate of autoproteolysis [127]. For example, the introduction of six point mutations in BPN' resulted in a 300-fold reduction in the rate constant for thermal inactivation [245]. Other possibilities are the introduction of disulfide bridges or the deletion of known autoproteolysis sites [246, 247]. It is also possible to replace the amino acids Asn and Gln, which are sensitive for deamidation [80].

In terms of pH optimum, SPAH showed a high pH optimum at 10.0, while SPPM, SPMI, SPLA and the reference proteases were in the range of 9.0 - 9.5 (2.2, 2.3). The protease SPFA showed a very broad pH spectrum with an optimum between 8.5 and 11.5 (2.4). When comparing the activity at pH 12.0, the difference between true and high-alkaline subtilisins becomes clear, as SPAO and Savinase as high-alkaline subtilisins showed a relative catalytic activity of more than 50 %, while the true subtilisins Carlsberg and BPN' showed less than 20 % (Table 5). Other high-alkaline subtilisins showed an even higher pH optimum such as BgAP from *Alkalihalobacillus gibsonii* at pH 11.0 [79] or YaB from *Bacillus* sp. YaB at pH 11.75 [248]. However, the difference between the subgroups is no longer as clear when comparing SPPM, SPMI, SPFA and SPLA, as they all show high relative activity at pH 12.0 (2.3, 2.4). SPLA retains even 91 % relative activity, which is higher than that of the high-alkaline subtilisins measured in this work. Here SPFA clearly stands out with its very broad pH spectrum with over 65 % relative activity at pH 6.0 to pH 12.5 (2.4). This behaviour has not yet been observed in the literature for a subtilisin, as most have a clear optimum [56, 249–251]. The phylogenetically intermediate subtilisin SPAH had below 50 % relative activity at pH 12.0, while the subtilisin ALTP from *Alkaliphilus transvaalensis*, which belongs to the same subgroup, had its pH optimum even above pH 12.6 [56]. Unfortunately, there is only one other PIS described with LD1 from *Bacillus* sp. KSM-LD1, with limited biochemical data [252], so it is difficult to make a general statement about this subgroup. Other potential subtilisins published in recent years cannot be phylogenetically classified because they were either isolated directly from their bacterial origin and only limited or no sequence data are available.

As outlined in 1.1.7, enzymes active at high pH must adapt to high pH as it affects the formation of salt bridges and hydrogen bonds that influence the three-dimensional shape of the protein. A high-alkaline protease adapts to higher pH values by an altered surface charge, which results in a higher pI value with more Arg and fewer Lys, Asp and Glu residues [129]. With the increased Arg content, the Arg residues contribute to stabilisation with an increase in hydrogen bonding and ion pairs, as the pK_a of the guanidino group of Arg (ca. 13.8) is higher than that of the ε-amino group of Lys (ca. 10.5) and can form ion pairs with amino acid residues more easily under alkaline conditions [130, 129]. In addition, as mentioned above, a deletion in the loop around position 160 can be observed in all high-alkaline subtilisins and a shorter loop increases the stability of the enzyme

[195]. These adaptations can also be observed for SPAO and Savinase as a typical representative of the high-alkaline subtilisins (2.2). However, SPPM, SPMI, SPLA, SPAH and SPFA all have an acidic pI and therefore a greater number of acidic amino acids, resulting in a mainly negatively charged protein surface (2.3, 2.4). Hence, a generally increased number of charged amino acids on the protein surface leads to better ionic interactions, thus maintaining stability and solubility [253]. Changes in the charge profile of a protein can therefore be used to influence enzyme activity, pH optimum and stability [80]. The working ranges up to a pH of 12.0 of all six proteases investigated indicates their great potential for various industrial applications, as is the case with other alkaline proteases previously reported [193, 254, 255]. The test of stability at different pH values for 24 h at 4 °C showed that SPPM, SPMI, SPLA and SPAH retain at least 65 % activity at a pH of 5.0, 92 % at their pH optimum (9.0 – 10.0) and 83 % at a pH of 12.0 (2.3). SPFA showed comparable stability, except at pH 5.0 (2.4). On the other hand, SPAO showed a loss of activity of more than 50 % around its pH optimum at pH 9.0, while the commercial proteases BPN', Savinase and subtilisin Carlsberg showed almost no residual protease activity under the test conditions, which could be due to higher autoproteolysis, as mentioned above. In comparison, ALTP from *Alkaliphilus transvaalensis* was stable between pH 5 and 11 after a short incubation of 10 minutes at 50°C [56]. The stability of the high-alkaline subtilisin sendai at pH 12.0 and 30°C after four hours was comparable, as the remaining activity was more than 90 %, whereas the true subtilisin NAT lost more than 80 % of its activity after only 60 min under the same conditions [194].

3.5.2 Effect of NaCl on enzyme activity and stability

Tolerance and stability to NaCl play a role in industries requiring low to high NaCl concentrations, including various applications in peptide synthesis, detergent formulation and fermented foods [111]. Therefore, the study of tolerance to high salinity is important. The proteases SPPM, SPMI, SPLA, SPAH and SPFA showed the highest activity without the addition of NaCl, and the activity gradually decreased to 36 % residual activity at 5 M NaCl (2.3, 2.4). The protease SPAH showed no further reduction after decreasing to 60 % remaining activity at 1 M NaCl. The activity of SPAO was observed to increase with increasing concentrations of NaCl, reaching its maximum at 4 M NaCl (2.2). A comparable increase in activity was also observed for Savinase. While the activity of subtilisin Carlsberg decreased constantly, it was comparable to that of SPPM, SPMI, SPLA, SPAH and SPFA. The activity of BPN' was relatively constantly high and only dropped to 84 % residual activity at 5 M NaCl (2.2).

In addition to the activity at different NaCl concentrations, the stability of the proteases towards high NaCl concentrations is also important. SPPM, SPMI and SPLA were stable with and without NaCl, while SPAH lost activity with increasing NaCl concentration after incubation at different NaCl concentrations for 2 h and measuring the residual activity (2.3). Subtilisin Carlsberg, SPAO and SPFA are more stable at higher NaCl concentrations,

while BPN' and Savinase are stable with and without NaCl (2.2, 2.4). However, SPAO showed stability above 50 % only with 1 M NaCl, while it lost stability with increasing NaCl concentration. Therefore, NaCl is beneficial for the stability and activity of SPAO until the increased activity leads to increased autoproteolysis, which is also true for SPAH. For subtilisin Carlsberg, increasing NaCl concentrations are beneficial as the activity decreases to protect against autoproteolysis.

A look at the bacterial background of the proteases shows that SPPM, SPAH and SPAO originate from a halophilic background [198, 199, 121]. Here *A. haloalkaliphilus*, the microbial background of SPAH, proves to be the most tolerant, growing with salt concentrations up to 20 % and an optimum at 5 % [199]. The other proteases SPMI, SPLA, subtilisin Carlsberg, Savinase and BNP' derive from halotolerant backgrounds [159, 202, 98, 256–258], while *F. arsenicus*, the bacterial origin of SPFA, tolerates only small amounts of NaCl (1 % w/v) [206]. SPPM, SPAH and SPAO showed halotolerant behaviour despite their halophilic origin, as they also function in the absence of NaCl. However, for SPAO, the activity increased at higher NaCl concentrations. Interestingly, all proteases examined are still active at NaCl concentrations at which the bacterial strains no longer grow. Therefore, a prediction of the behaviour towards salt based on the growth characteristics of the bacterial origin is limited, but a more or less pronounced halotolerance can be expected if the extracellular protease originates from a halotolerant or halophilic background [111, 135, 259, 260].

As outlined in 1.1.7, high salt concentrations affect the structure and function of proteins if they are not adapted to them, as the local water structure around the protein is affected and the propensity for interaction between intermolecular hydrogen bonds decreases, leading to protein aggregation and precipitation [132]. On a molecular basis, most halophilic enzymes contain a relatively high proportion of acidic amino acids (Asp and Glu) and therefore predictions about the adaptation of a protein to high salt concentrations can be made with regard to the proportion of acidic amino acids such as Glu and Asp compared to the basic amino acids Lys, His, and Arg (referred to as the AB ratio) [134, 111]. However, this rule can only be used partially to predict the salt adaptation of halotolerant proteases, as it only partially fits to the proteases studied, which is related to the fact that the mechanisms of halotolerance have not been fully elucidated yet [135, 134]. Rather, the general surface charge plays a role here. As explained in 1.1.7, adaptation to a high salt content is based on a high proportion of negative or positive charges on the protein surface [131]. The charged amino acids play a crucial role in forming a hydrate shell around the protein that maintains its solubility [133]. According to Qiao *et al.* the negatively charged amino acids are the most beneficial ones, followed by the positively charged and finally the charge-neutral amino acids [133]. However, SPAO with a predominantly positive surface charge was even more active in the presence of a high salt concentration than SPPM, SPMI, SPAH and SPLA with a predominantly negative

surface charge. All investigated subtilisins are even active in the absence of NaCl, although SPPM, SPAH and SPAO originate from halophilic background. That halophilic enzymes are active even in the absence of salt has also been reported for other enzymes, with the explanation that the protein may be structurally rigid enough to remain folded in the absence of NaCl [132].

According to the isoelectric point and the number of charged amino acids, BPN' should be more salt sensitive. However, in contrast to subtilisin Carlsberg, BPN' has five of eight acidic and polar amino acids located on the protein surface that are conserved within halophilic enzymes according to Takenaka *et al.* [131, 135]. As already mentioned, charged amino acids stabilise water and/or ion binding, protein refolding is facilitated and aggregation is prevented, making halotolerant proteases tolerant to high pH values. Therefore, halotolerant proteases may be interesting for industrially relevant applications [251]. Hence, it is interesting to note that not only representatives of the subgroup of high-alkaline subtilisins, but also true subtilisins with halotolerant or halophilic properties are highly active at pH 12.0, such as SPPM, SPMI, SPLA and SPFA.

3.5.3 Effect of SDS, H_2O_2 and Ca^{2+} on enzyme activity and stability

With regard to stability towards 1 and 5 % SDS (w/v) at 10 °C, all proteases, with the exception of SPAO, were found to be highly stable [231, 213]. As shown in Table 6, SPAO lost all its activity with 1 % SDS. SPAO is thus very sensitive to SDS, while the other proteases were even more active with than without SDS. Only SPPM and SPLA had slightly reduced activity with 76 % and 97 %, respectively, at 5 % SDS. Usually, the anionic character of SDS causes it to interact with hydrophobic amino acid residues of proteins, leading to protein unfolding and loss of activity. However, the increasing activity of some proteases could be explained by the fact that SDS does not unfold the protein, but instead helps it to achieve a favourable protein conformation, which increases activity [261]. This behaviour has also been described in the literature for a salt-tolerant and thermostable serine protease from *B. subtilis*, which does not lose activity even at an SDS concentration of 10 % (w/v) [262]. An alkaline serine protease from haloalkaliphilic *Shouchella oshimensis* (formerly *Bacillus lehensis*) JO-26 showed an increased activity of 275 % after incubation with 1 % SDS for 30 min [251]. The subtilisin BLAP from another *Shouchella oshimensis* strain lost no activity at 1 % SDS, but increased to 160 % activity after incubation with 2 % SDS [261]. For an alkaline protease from *Shouchella clausii* I-52, a slight decrease in activity to 73 % was observed after incubation with 5 % SDS [109]. A complete loss of activity is therefore rather unusual for a high-alkaline subtilisin. Another example is the wild-type variant of subtilisin E, which lost more than 80 % of its activity at 4 % SDS [214], or the intracellular subtilase AprX-SK37 from *Virgibacillus* sp. SK 37, which revealed a complete loss of activity after 30 min with 0.5 % SDS [193]. Within the scope of various enzyme engineering approaches, it was possible to increase the tolerance of subtilisin E to chaotropic compounds [214]. Interestingly, in this study it was found that

the active site and the surrounding substrate binding pocket were the key factors in enhancing chaotolerance [214].

The influence of hydrogen peroxide on the stability of the enzymes showed that all proteases are very resistant to 1 % (v/v) H_2O_2 and had a residual activity of 78 to 108 % (Table 6). Treatment with 5 % (v/v) H_2O_2 reduced the activity to 31 - 58 %, whereas subtilisin Carlsberg, Savinase and BPN' lost up to 92 % of their activity at 5 % H_2O_2. SPAO showed the highest resistance with increased activity at 1 % H_2O_2 and 58 % residual activity at 5 % H_2O_2. A higher increase in activity than SPAO by 14 % and 16 % at 1 % and 5 % H_2O_2 was observed for a protease from *S. clausii* I-52 [109]. A protease from the alkaliphilic *Bacillus* sp. NPST-AK15 showed an increase in activity of 2 % at 1 % H_2O_2 and a loss of 6 % at 5 % H_2O_2 [105]. From other reports on the stability of alkaline proteases to oxidants, subtilisin LD1 from *Bacillus* sp. KSM-LD1 retained 40 % of its activity after incubation with 3.4 % H_2O_2 at 30 °C for 30 min [58]. Almost unaffected by 10 % H_2O_2 was an alkaline protease from *Shouchella patagoniensis* (formerly *Bacillus patagoniensis*) after incubation at 25 °C for 30 min [263]. A higher loss of activity was reported for the subtilase from *Thermoactinomyces vulgaris* strain CDF, which lost 90 % of its activity after 1 hour at 40 °C and 5 % H_2O_2 [76]. The subtilase KP-43 from the group of oxidatively stable proteases (OSP) lost their hydrolysis ability already after 30 min incubation with 3 % H_2O_2 [264]. However, comparing results to data found in the literature can be challenging because the conditions under which the activity studies were performed, such as the temperature, buffer, pH, substrate concentration, and H_2O_2 concentration, often vary. The sensitivity against oxidants such as H_2O_2 is especially relevant for the use in detergents, as hydrogen peroxide is often added to remove stains by bleaching [16].

The sensitivity to H_2O_2 is likely due to the oxidation of a conserved methionine residue near the catalytic serine residue, transforming it into a sulfoxide and resulting in the inactivation of the enzyme [265]. The sulfoxide oxygen destabilises the tetrahedral intermediate formed with the carbonyl group of the bound peptide [266, 32]. Enzyme engineering approaches in which methionine residue 222 (BPN' numbering) was replaced by non-oxidizable amino acids (i.e. Ser, Ala, and Leu) showed increased resistance to oxidants [267–269, 83]. Interestingly, examination of the digestion pattern of β-casein of the subtilase KP-43 from *Bacillus* sp. KSM-KP43, revealed that it differs when the methionine residue 222 (BPN' numbering) is oxidised [264].

Table 6 Influence of H_2O_2 and SDS on enzyme activity (2.2, 2.3, 2.4)

	Residual protease activity [%]			
Protease	1 % SDS	5 % SDS	1 % H_2O_2	5 % H_2O_2
SPPM	121±4	76±2	92±4	41±0
SPMI	178±4	165±1	81±1	31±2
SPLA	108±4	97±3	78±2	40±1
SPAH	120±7	106±2	89±1	40±2
SPAO	0±0	0±0	108±4	58±3
SPFA	182±8	169±4	81±4	52±3
Sub. Carlsberg	325±6	189±6	71±9	27±2
Savinase	138±4	113±6	64±3	8±0
BPN’	205±3	164±7	81±5	11±0

The purified proteases were incubated with 1 and 5 % (v/v) H_2O_2; 1 and 5 % (w/v) SDS in 10 mM HEPES-NaOH pH 7.0 for 1 h at 10 °C. Activity was measured with the suc-AAPF-pNA assay under standard conditions. Remaining activity of the proteases incubated in buffer with no additions was defined as 100 %. All experiments were performed at least in triplicates and data are shown as mean values ± SD.

As mentioned in 3.4, SPPM, SPMI, SPLA, SPAH, SPAO and SPFA contain two Ca^{2+} binding sites. To investigate the effect of Ca^{2+} on the activity and stability of the proteases, they were incubated with EDTA. SPPM, SPAH and SPFA showed almost no difference in activity between incubation with or without EDTA, which was also observed for BPN' (data not shown). No effect on activity was also observed for a serine protease from *S. clausii* GMBAE 42 [270]. For SPLA, SPMI, SPAO and Savinase it was not possible to completely restore the activity after addition of $CaCl_2$ (2.2, 2.3). An overall loss of protease activity could explain the lack of recovery of the protease activity, as Ca^{2+} has a stabilising effect that reduces thermal denaturation and autolysis due to the reduction in molecular flexibility [64], which was also observed by others [259, 271, 272]. The binding of Ca^{2+}, especially within a loop, stabilises it against unfolding and thus reduces access to the primary binding sites for autoproteolysis [273]. Interestingly, a protein-engineered version of BPN' resulted in a calcium-independent protease that is fully active but has lower thermal stability in the absence of stabilising mutations [274].

3.5.4 Substrate spectrum

Most subtilisins have broad substrate specificity and mainly play a role in nutrition by providing peptides and amino acids for cell growth [24]. Variations in substrate specificity occur due to different residues in the substrate-binding region. In particular their side chains interact with the P1 and P4 substrate residues that dominate substrate preference in subtilisins [24, 275]. Exceptions to broad substrate specificity occur when residue 166 (BPN’ numbering) at the bottom of the S1 binding site is an Asp, leading to a preference for cleavage after P1-Arg residues, which has also been observed in other subtilases such

as the members of proteinase K, the lantibiotic leader peptidases and in all members of the kexin group [24]. This is not the case with the proteases investigated, including the reference proteases. To compare the substrate specificity of the six proteases with each other and with Savinase, subtilisin Carlsberg and BPN', ten synthetic peptide-4-nitroanilide substrates with three or four amino acids were used. The selected substrates are all typical subtilisin substrates, with the exception of suc-AAA-pNA, which is an elastase substrate [276, 277]. The experiment showed that the six proteases are typical subtilisins (2.2, 2.3, 2.4). Furthermore, SPAO is a typical high-alkaline subtilisin like Savinase, as it prefers the substrate suc-FAAF-pNA. In addition, as discussed in 3.3, the proteases showed higher activity for the substrate azocasein, as it is a more complex substrate with more potential hydrolysis sites. A comparison with the data available in the literature on the hydrolysis of substrates is difficult, because the experimental conditions are different. Georgieva *et al.* showed comparable results for Savinase with highest specificity for suc-FAAF-pNA compared to Esperase with a similar preference [276]. A high specificity towards suc-AAPF-pNA by an alkaline serine protease from *S. clausii* GMBAE 42 was discovered by Kazan et al [270]. For the elastase-specific substrate Suc-AAA-pNA, for which the proteases studied in this work showed low activity, the alkaline elastase YaB from *Bacillus* strain YaB showed high activity [278]. It was also reported that several subtilisins and proteinase K have higher activity when Ala, Glu, Lys or Val are in position P1 [279]. In subtilisins, the S1 and S4 binding sites are large and hydrophobic, which explains the preference for aromatic or large non-polar P1 and P4 substrate residues [24]. This can also be observed for the investigated proteases. However, SPAO and Savinase are able to hydrolyse the substrate when alanine is at the P1 and P2 positions, but with higher efficiency when one of the positions is alanine and the other one an amino acid with a larger hydrophobic group. That SPAO and Savinase show this different substrate specificity could be due to the four amino acid deletion around residue 160, which is in the S1 recognition loop, compared to the true subtilisins and the PIS [192].

4 Conclusion and outlook

In this work, phylogenetic analysis of the MEROPS holotype dataset of the S8 family showed that a large number of different subtilases formed new groups and subgroups. Based on their position in the phylogenetic tree, their biochemical properties or their origin, eight new groups were identified within the S8A subfamily, in addition to the known groups such as proteinase K, pyrolysins, kexins, subtilisins, thermitases and lantibiotic peptidases first described by Siezen and Leunissen in 1997 [24]. Furthermore, 13 new subgroups were identified. In particular, the group of pyrolysins harbours nine and the group of subtilisins four new subgroups. Considering the increasing data on characterised sequences, the finding of new groups is very likely and also a further subdivision into classes, as described for plant pyrolysins [44, 186, 187], or fungal pyrolysins [188–190]. The phylogenetic analysis was necessary to place newly found sequences from a data mining search into the context of the subtilase family. Thus, about 1400 sequences were evaluated in the search for promising subtilisin sequences from *Bacillaceae*. Using the updated phylogenetic context and further bioinformatic analyses, a set of 120 sequences remained that could be analysed.

In particular, extremophilic organisms such as thermophiles, psychrophiles and halophiles or halotolerants offer enormous potential for meeting industrial requirements [72]. Since enzymes with a halophilic or halotolerant background have been shown to be suitable for industrial applications, the focus was placed on proteases of this microbial background. The subtilisins derived from *Pontibacillus marinus* DSM 16465^T (SPPM), *Metabacillus indicus* DSM 16189 (SPMI), *Litchfieldia alkalitelluris* DSM 16976^T (SPLA), *Fictibacillus arsenicus* DSM 15822^T (SPFA), *Alkalibacillus haloalkaliphilus* DSM 5271^T (SPAH), and *Halalkalibacter okhensis* Kh10-101^T (SPAO) were recombinantly produced with *Bacillus subtilis* DB104, then purified and biochemically characterised for the first time. The studied subtilisins belong to different subgroups: SPPM, SPMI, SPLA and SPFA are members of the true subtilisins, SPAH is a member of the phylogenetically intermediate subtilisins and SPAO belongs to the high-alkaline subtilisins. Here, it was confirmed that subtilisins with halotolerant and halophilic backgrounds have improved biochemical properties compared to subtilisins already used industrially. The proteases exhibited pH optima between 9.0 and 10.0 and temperature optima between 50 and 80 °C. SPFA particularly stood out as it showed a broad pH optimum from 8.5 to 11.5 and a generally high activity between pH 6.0 and 12.0 with a temperature optimum of 80 °C. All proteases were active at NaCl concentrations up to 5 M, stable against hydrogen peroxide and, except for SPAO, stable against SDS. The enzymes showed remarkable properties, especially considering that they have not been engineered, and proved to be interesting for industrial applications, especially for detergents, as shown by the two patents applications filed based on this work. The biochemical data of the halotolerant subtilisins and their structure-function analysis with emphasis on surface charge analysis form the

starting point for further enzyme engineering studies. Despite the selection and cloning of individual sequences from the phylogenetic tree, a high-throughput library could accelerate the search for biotechnologically relevant proteases through gene synthesis and subsequent functional screening after transformation, with the potential disadvantage of losing sequences that cannot be produced and or secreted in the selected host. In addition, an investigation of the other groups and subgroups of the subtilase family would be interesting for biotechnological applications.

Problems with the production of divers subtilisins by *B. subtilis* DB104 were to be expected. Thus, out of twelve proteins, only six were produced in sufficient quantity. For this reason, the wild-type strains from which the new subtilisins derived were examined for their cultivability and genetic accessibility. While *B. subtilis* is a known host for the production of proteases, *Metabacillus indicus* DSM 16189 proved to be a suitable host for the recombinant production of its homologous subtilisin for the first time in this work. However, it remains to be seen whether this potential can also be exploited for heterologous proteases, especially those that we cannot produce with *B. subtilis* DB104. After further improvement through strain development and process optimisation, *M. indicus* could be suitable competitive protein yields for recombinant proteases.

5 References

1 Rawlings ND, Fraser RM & Barrett AJ An introduction to peptidases and the MEROPS database. In *Industrial enzymes*: *Structure, function, and applications.* (Polaina J & MacCabe AP, eds), 161–179 (2007). Dordrecht. Springer.

2 Webb EC (1992) Enzyme nomenclature: Recommendations of the Nomenclature Commitee of the International Union of Biochemistry and Molecular Biology on the nomenclature and classification of enzymes. Academic Press, London.

3 Hartley BS (1960) Proteolytic enzymes. *Annu Rev Biochem* **29,** 45–72.

4 Kondo MY, Okamoto DN, Santos JAN, Juliano MA, Oda K, Pillai B, James MNG, Juliano L & Gouvea IE (2010) Studies on the catalytic mechanism of a glutamic peptidase. *J Biol Chem* **285,** 21437–21445.

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

6 Rawlings ND (2020) Twenty-five years of nomenclature and classification of proteolytic enzymes. *Biochim Biophys Acta Proteins Proteom* **1868,** 140345.

7 Mentlein R (2004) Cell-surface peptidases. *Int Rev Cytol* **235,** 165–213.

8 Barrett AJ & Rawlings ND (1995) Families and clans of serine peptidases. *Arch Biochem Biophys* **318,** 247–250.

9 Rawlings ND, Waller M, Barrett AJ & Bateman A (2014) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* **42,** D503-9.

10 Rawlings ND & Bateman A (2021) How to use the MEROPS database and website to help understand peptidase specificity. *Protein Sci* **30,** 83–92.

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

12 Southan C (2001) A genomic perspective on human proteases as drug targets. *Drug Discov Today* **6,** 681–688.

13 Robbins BH (1930) A proteolytic enzyme in Ficin, the antheleminitic principle of Leche de Higueron. *J Biol Chem* **87,** 251–257.

14 Sawjalow W (1905) Zur Frage nach der Identität von Pepsin und Chymosin. *Hoppe-Seyler´s Zeitschrift für physiologische Chemie* **46,** 307–331.

15 Isabela Avila-Rodríguez M, Meléndez-Martínez D, Licona-Cassani C, Manuel Aguilar-Yañez J, Benavides J & Lorena Sánchez M (2020) Practical context of enzymatic treatment for wound healing: A secreted protease approach (Review). *Biomed Rep* **13,** 3–14.

16 Maurer K-H (2004) Detergent proteases. *Curr Opin Biotechnol* **15,** 330–334.

17 Di Cera E (2009) Serine proteases. *IUBMB Life* **61,** 510–515.

18 Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A & Finn RD (2018) The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res* **46,** D624-D632.

19 Page MJ & Di Cera E (2008) Serine peptidases: classification, structure and function. *Cell Mol Life Sci* **65,** 1220–1236.

20 Hedin SG (1905) Observations on the action of trypsin. *J Physiol* **32,** 468–485.

21 Le Bonniec BF Thrombin. In *Handbook of proteolytic enzymes.* (Rawlings ND & Salvesen G, eds), 2915–2932 (2013). Amsterdam. Elsevier Academic Press.

22 Gráf L, Szilágyi L & Venekei I Chymotrypsin. In *Handbook of proteolytic enzymes.* (Rawlings ND & Salvesen G, eds), 2626–2633 (2013). Amsterdam. Elsevier Academic Press.

23 Matsubara H, Kasper CB, Brown DM & Smith EL (1965) Subtilisin BPN'. *J Biol Chem* **240,** 1125–1130.

24 Siezen RJ & Leunissen JA (1997) Subtilases: the superfamily of subtilisin-like serine proteases. *Protein Sci* **6,** 501–523.

25 LeMosy EK, Hong CC & Hashimoto C (1999) Signal transduction by a protease cascade. *Trends Cell Biol* **9,** 102–107.

26 Sim RB & Laich A (2000) Serine proteases of the complement system. *Biochem Soc Trans* **28,** 545–550.

27 Barros C, Crosby JA & Moreno RD (1996) Early steps of sperm-egg interactions during mammalian fertilization. *Cell Biol Int* **20,** 33–39.

28 Bender ML & Kezdy J (1965) Mechanism of action of proteolytic enzymes. *Annu Rev Biochem* **34,** 49–76.

29 Polgár L (2005) The catalytic triad of serine peptidases. *Cell Mol Life Sci* **62,** 2161–2172.

30 Hedstrom L (2002) Serine protease mechanism and specificity. *Chem Rev* **102,** 4501–4524.

31 Berg JM, Tymoczko JL, Gatto GJ & Stryer L (2018) Stryer Biochemie. Springer, Berlin.

32 Bryan P, Pantoliano MW, Quill SG, Hsiao HY & Poulos T (1986) Site-directed mutagenesis and the role of the oxyanion hole in subtilisin. *Proc Natl Acad Sci USA* **83,** 3743–3745.

33 Siezen RJ, Renckens B & Boekhorst J (2007) Evolution of prokaryotic subtilases: genome-wide analysis reveals novel subfamilies with different catalytic residues. *Proteins* **67,** 681–694.

34 Nardini M & Dijkstra BW (1999) α/β Hydrolase fold enzymes: the family keeps growing. *Curr Opin Struct Biol* **9,** 732–737.

35 Fuller RS Kexin. In *Handbook of proteolytic enzymes.* (Rawlings ND & Salvesen G, eds), 3270–3277 (2013). Amsterdam. Elsevier Academic Press.

36 Declercq J & Creemers JW Furin. In *Handbook of proteolytic enzymes.* (Rawlings ND & Salvesen G, eds), 3281–3285 (2013). Amsterdam. Elsevier Academic Press.

37 Tunlid A & Jansson S (1991) Proteases and their involvement in the infection and immobilization of nematodes by the nematophagous fungus *Arthrobotrys oligospora*. *Appl Environ Microbiol* **57,** 2868–2872.

38 Li J, Yu L, Yang J, Dong L, Tian B, Yu Z, Liang L, Zhang Y, Wang X & Zhang K (2010) New insights into the evolution of subtilisin-like serine protease genes in *Pezizomycotina*. *BMC Evol Biol* **10,** 68.

39 Leger R, Charnley AK & Cooper RM (1987) Characterization of cuticle-degrading proteases produced by the entomopathogen *Metarhizium anisopliae*. *Arch Biochem Biophys* **253,** 221–232.

40 Sahl HG, Jack RW & Bierbaum G (1995) Biosynthesis and biological activities of lantibiotics with unique post-translational modifications. *Eur J Biochem* **230,** 827–853.

41 Coutte L, Antoine R, Drobecq H, Locht C & Jacob-Dubuisson F (2001) Subtilisin-like autotransporter serves as maturation protease in a bacterial secretion pathway. *EMBO J* **20,** 5040–5048.

42 Masayama A, Hamasaki K, Urakami K, Shimamoto S, Kato S, Makino S, Yoshimura T, Moriyama M & Moriyama R (2006) Expression of germination-related enzymes, CspA, CspB, CspC, SleC, and SleM, of *Clostridium perfringens* S40 in the mother cell compartment of sporulating cells. *Genes Genet Syst* **81,** 227–234.

43 Renn SC, Tomkinson B & Taghert PH (1998) Characterization and cloning of tripeptidyl peptidase II from the fruit fly, *Drosophila melanogaster*. *J Biol Chem* **273,** 19173–19182.

44 Schaller A Plant Subtilisins. In *Handbook of proteolytic enzymes.* (Rawlings ND & Salvesen G, eds), 3247–3254 (2013). Amsterdam. Elsevier Academic Press.

45 Garcia-Sanchez A, Cerrato R, Larrasa J, Ambrose NC, Parra A, Alonso JM, Hermoso-de-Mendoza M, Rey JM, Mine MO, Carnegie PR, Ellis TM, Masters AM, Pemberton AD & Hermoso-de-Mendoza J (2004) Characterisation of an extracellular serine protease gene (nasp gene) from *Dermatophilus congolensis*. *FEMS Microbiol Lett* **231,** 53–57.

46 Dave JA, van Gey Pittius NC, Beyers AD, Ehlers MRW & Brown GD (2002) Mycosin-1, a subtilisin-like serine protease of *Mycobacterium tuberculosis*, is cell wall-associated and expressed during infection of macrophages. *BMC Microbiol* **2,** 30.

47 Eder J, Rheinnecker M & Fersht AR (1993) Folding of subtilisin BPN': role of the pro-sequence. *J Mol Biol* **233,** 293–304.

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

49 Kobayashi H, Yoshida T, Miyakawa T, Tashiro M, Okamoto K, Yamanaka H, Tanokura M & Tsuge H (2015) Structural basis for action of the external chaperone for a propeptide-deficient serine protease from *Aeromonas sobria*. *J Biol Chem* **290,** 11130–11143.

50 Subbian E, Yabuta Y & Shinde UP (2005) Folding pathway mediated by an intramolecular chaperone: intrinsically unstructured propeptide modulates stochastic activation of subtilisin. *J Mol Biol* **347,** 367–383.

51 Jain SC, Shinde U, Li Y, Inouye M & Berman HM (1998) The crystal structure of an autoprocessed Ser221Cys-subtilisin E-propeptide complex at 2.0 A resolution. *J Mol Biol* **284,** 137–144.

52 Li Y, Hu Z, Jordan F & Inouye M (1995) Functional analysis of the propeptide of subtilisin E as an intramolecular chaperone for protein folding. Refolding and inhibitory abilities of propeptide mutants. *J Biol Chem* **270,** 25127–25132.

53 Yabuta Y, Takagi H, Inouye M & Shinde U (2001) Folding pathway mediated by an intramolecular chaperone: propeptide release modulates activation precision of pro-subtilisin. *J Biol Chem* **276,** 44427–44434.

54 Azrin NAM, Ali MSM, Rahman RNZRA, Oslan SN & Noor NDM (2022) Versatility of subtilisin: A review on structure, characteristics, and applications. *Biotechnol Appl Biochem* **69,** 2599–2616.

55 Graycar TP, Bott RR, Power SD & Estell DA Subtilisins. In *Handbook of proteolytic enzymes.* (Rawlings ND & Salvesen G, eds), 3148–3155 (2013). Amsterdam. Elsevier Academic Press.

56 Kobayashi T, Lu J, Li Z, Hung VS, Kurata A, Hatada Y, Takai K, Ito S & Horikoshi K (2007) Extremely high alkaline protease from a deep-subsurface bacterium, *Alkaliphilus transvaalensis*. *Appl Microbiol Biotechnol* **75,** 71–80.

57 Völkl P, Markiewicz P, Stetter KO & Miller JH (1994) The sequence of a subtilisin-type protease (aerolysin) from the hyperthermophilic archaeum *Pyrobaculum aerophilum* reveals sites important to thermostability. *Protein Sci* **3,** 1329–1340.

58 Saeki K, Magallones MV, Takimura Y, Hatada Y, Kobayashi T, Kawai S & Ito S (2003) Nucleotide and deduced amino acid sequences of a new subtilisin from an alkaliphilic *Bacillus* isolate. *Curr Microbiol* **47,** 337–340.

59 Burnett TJ, Shankweiler GW & Hageman JH (1986) Activation of intracellular serine proteinase in *Bacillus subtilis* cells during sporulation. *J Bacteriol* **165,** 139–145.

60 Gamble M, Künze G, Dodson EJ, Wilson KS & Jones DD (2011) Regulation of an intracellular subtilisin protease activity by a short propeptide sequence through an original combined dual mechanism. *Proc Natl Acad Sci USA* **108,** 3536–3541.

61 Power SD, Adams RM & Wells JA (1986) Secretion and autoproteolytic maturation of subtilisin. *Proc Natl Acad Sci USA* **83,** 3096–3100.

62 Markland FS & Smith EL 16 subtilisins: primary structure, chemical and physical properties. In *Hydrolysis: Peptide Bonds* **3,** 561–608 (1971). Elsevier.

63 Tjalsma H, Bolhuis A, Jongbloed JD, Bron S & van Dijl JM (2000) Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome. *Microbiol Mol Biol Rev* **64,** 515–547.

64 Siezen RJ, Vos WM de, Leunissen JA & Dijkstra BW (1991) Homology modelling and protein engineering strategy of subtilases, the family of subtilisin-like serine proteinases. *Protein Eng* **4,** 719–737.

65 Bryan P, Alexander P, Strausberg S, Schwarz F, Lan W, Gilliland G & Gallagher DT (1992) Energetics of folding subtilisin BPN'. *Biochem* **31,** 4937–4945.

66 Research Reports World (2021) global alkaline proteases market insights and forecast to 2027. https://www.researchreportsworld.com/global-alkaline-proteases-market-18086190.

67 Ward OP Proteases. In *Comprehensive Biotechnology,* 571–582 (2011). Elsevier.

68 Smith EL, Markland FS, Kasper CB, DeLange RJ, Landon M & Evans WH (1966) The complete amino acid sequence of two types of subtilisin, BPN' and Carlsberg. *J Biol Chem* **241,** 5974–5976.

69 Betzel C, Klupsch S, Papendorf G, Hastrup S, Branner S & Wilson KS (1992) Crystal structure of the alkaline proteinase Savinase from *Bacillus lentus* at 1.4 A resolution. *J Mol Biol* **223,** 427–445.

70 van der Laan JM, Teplyakov AV, Kelders H, Kalk KH, Misset O, Mulleners LJ & Dijkstra BW (1992) Crystal structure of the high-alkaline serine protease PB92 from *Bacillus alcalophilus*. *Protein Eng* **5,** 405–411.

71 Sahm H, Antranikian G, Stahmann K-P & Takors R (2013) Industrielle Mikrobiologie. Springer.

72 Salwan R & Sharma V (2019) Trends in extracellular serine proteases of bacteria as detergent bioadditive: alternate and environmental friendly tool for detergent industry. *Arch Microbiol* **201,** 863–877.

73 Mechri S, Bouacem K, Amziane M, Dab A, Nateche F & Jaouadi B (2019) Identification of a new serine alkaline peptidase from the moderately halophilic *Virgibacillus natechei* sp. nov., strain FarDT and its application as bioadditive for peptide synthesis and laundry detergent formulations. *Biomed Res Int* **2019,** 6470897.

74 Mechri S, Bouacem K, Zaraî Jaouadi N, Rekik H, Ben Elhoul M, Omrane Benmrad M, Hacene H, Bejar S, Bouanane-Darenfed A & Jaouadi B (2019) Identification of a novel

protease from the thermophilic *Anoxybacillus kamchatkensis* M1V and its application as laundry detergent additive. *Extremophiles* **23,** 687–706.

75 Farooq S, Nazir R, Ganai SA & Ganai BA (2021) Isolation and characterization of a new cold-active protease from psychrotrophic bacteria of Western Himalayan glacial soil. *Sci Rep* **11,** 12768.

76 Ding Y, Yang Y, Ren Y, Xia J, Liu F, Li Y, Tang X-F & Tang B (2020) Extracellular production, characterization, and engineering of a polyextremotolerant subtilisin-like protease from feather-degrading *Thermoactinomyces vulgaris* strain CDF. *Front Microbiol* **11,** 605771.

77 Tindbaek N, Svendsen A, Oestergaard PR & Draborg H (2004) Engineering a substrate-specific cold-adapted subtilisin. *Protein Eng Des Sel* **17,** 149–156.

78 Fujinami S & Fujisawa M (2010) Industrial applications of alkaliphiles and their enzymes-past, present and future. *Environ Technol* **31,** 845–856.

79 Martinez R, Jakob F, Tu R, Siegert P, Maurer K-H & Schwaneberg U (2013) Increasing activity and thermal resistance of *Bacillus gibsonii* alkaline protease (BgAP) by directed evolution. *Biotechnol Bioeng* **110,** 711–720.

80 Jakob F, Martinez R, Mandawe J, Hellmuth H, Siegert P, Maurer K-H & Schwaneberg U (2013) Surface charge engineering of a *Bacillus gibsonii* subtilisin protease. *Appl Microbiol Biotechnol* **97,** 6793–6802.

81 Li Z, Roccatano D, Lorenz M & Schwaneberg U (2012) Directed evolution of subtilisin E into a highly active and guanidinium chloride- and sodium dodecylsulfate-tolerant protease. *Chembiochem* **13,** 691–699.

82 Bryan PN (2000) Protein engineering of subtilisin. *BBA Protein Struct Mol Enzym* **1543,** 203–222.

83 Vojcic L, Despotovic D, Maurer K-H, Zacharias M, Bocola M, Martinez R & Schwaneberg U (2013) Reengineering of subtilisin Carlsberg for oxidative resistance. *Biol Chem* **394,** 79–87.

84 Kaori Kurata, Masaya Itoh, Masahiro Matsumiya, Akira Dobashi, Atsushi Nakamura, Tomohiko Kondo, Katsuhiro Osajima, Yasuharu Itagaki (2018) Reduction of potential IgE reactivity in fish meat of japanese anchovy by enzymatic degradation of parvalbumin and collagen. *J Cook Sci Jpn,* 205–216.

85 Yu J, Mikiashvili N, Bonku R & Smith IN (2021) Allergenicity, antioxidant activity and ACE-inhibitory activity of protease hydrolyzed peanut flour. *Food Chem* **360,** 129992.

86 Li H, Yang J, Qin A, Yang F, Liu D, Li H & Yu J (2022) Milk protein hydrolysates obtained with immobilized alcalase and neutrase on magnetite nanoparticles: Characterization and antigenicity study. *J Food Sci* **87,** 3107–3116.

87 Chen Y, Toth EA, Ruan B, Choi EJ, Simmerman R, Chen Y, He Y, Wang R, Godoy-Ruiz R, King H, Custer G, Travis Gallagher D, Rozak DA, Solomon M, Muro S, Weber DJ, Orban J, Fuerst TR & Bryan PN (2021) Engineering subtilisin proteases that specifically degrade active RAS. *Commun Biol* **4,** 299.

88 Yatagai C, Maruyama M, Kawahara T & Sumi H (2008) Nattokinase-promoted tissue plasminogen activator release from human cells. *Pathophysiol Haemost Thromb* **36,** 227–232.

89 Bacheva AV, Plieva FM, Lysogorskaya EN, Filippova IY & Lozinsky VI (2001) Peptide synthesis in organic media with subtilisin 72 immobilized on poly(vinyl alcohol)-cryogel carrier. *Bioorganic Med Chem Lett* **11,** 1005–1008.

90 Mechri S, Bouacem K, Jabeur F, Mohamed S, Addou NA, Dab A, Bouraoui A, Bouanane-Darenfed A, Bejar S, Hacène H, Baciou L, Lederer F & Jaouadi B (2019) Purification and biochemical characterization of a novel thermostable and halotolerant subtilisin SAPN, a serine protease from *Melghiribacillus thermohalophilus* Nari2AT for chitin extraction from crab and shrimp shell by-products. *Extremophiles* **23,** 529–547.

91 Luo X-J, Yu H-L & Xu J-H (2012) Genomic data mining: an efficient way to find new and better enzymes. *Enzyme Engg* **01,** 104.

92 Aharoni A (2009) Mining for new enzymes. *Microb Biotechnol* **2,** 128–129.

93 Rawlings ND (2013) Identification and prioritization of novel uncharacterized peptidases for biochemical characterization. *J Biol Databases* **2013,** bat022.

94 Robinson SL, Piel J & Sunagawa S (2021) A roadmap for metagenomic enzyme discovery. *Nat Prod Rep* **38,** 1994–2023.

95 Sayers EW, Beck J, Bolton EE, Bourexis D, Brister JR, Canese K, Comeau DC, Funk K, Kim S, Klimke W, Marchler-Bauer A, Landrum M, Lathrop S, Lu Z, Madden TL, O'Leary N, Phan L, Rangwala SH, Schneider VA, Skripchenko Y, Wang J, Ye J, Trawick BW, Pruitt KD & Sherry ST (2021) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **49,** D10-D17.

96 Patel S & Gupta RS (2020) A phylogenomic and comparative genomic framework for resolving the polyphyly of the genus *Bacillus*: Proposal for six new genera of *Bacillus* species, *Peribacillus* gen. nov., *Cytobacillus* gen. nov., *Mesobacillus* gen. nov., *Neobacillus* gen. nov., *Metabacillus* gen. nov. and *Alkalihalobacillus* gen. nov. *Int J Syst Evol Microbiol* **70,** 406–438.

97 Gupta RS, Patel S, Saini N & Chen S (2020) Robust demarcation of 17 distinct *Bacillus* species clades, proposed as novel *Bacillaceae* genera, by phylogenomics and comparative genomic analyses: description of *Robertmurraya kyonggiensis* sp. nov. and proposal for an emended genus *Bacillus* limiting it only to the members of the Subtilis and Cereus clades of species. *Int J Syst Evol Microbiol* **70,** 5753–5798.

98 Priest FG, Goodfellow M, Shute LA & Berkeley RCW (1987) *Bacillus amyloliquefaciens* sp. nov., nom. rev. *Int J Syst Evol Microbiol* **37,** 69–71.

99 Curators of the CCUG (1972) CCUG 1816^{T} - *Bacillus lentus*.

100 Curators of the CCUG (1978) CCUG 7422^{T} - *Bacillus licheniformis*.

101 Ferrer M, Golyshina O, Beloqui A & Golyshin PN (2007) Mining enzymes from extreme environments. *Curr Opin Microbiol* **10,** 207–214.

102 Antranikian G, Vorgias CE & Bertoldo C (2005) Extreme environments as a resource for microorganisms and novel biocatalysts. *Adv Biochem Eng Biotechnol* **96,** 219–262.

103 Narinx E, Baise E & Gerday C (1997) Subtilisin from psychrophilic antarctic bacteria: characterization and site-directed mutagenesis of residues possibly involved in the adaptation to cold. *Protein Eng* **10,** 1271–1279.

104 Ulucay O, Gormez A & Ozic C (2022) Identification, characterization and hydrolase producing performance of thermophilic bacteria: geothermal hot springs in the Eastern and Southeastern Anatolia Regions of Turkey. *Antonie Van Leeuwenhoek* **115,** 253–270.

105 Ibrahim ASS, Al-Salamah AA, El-Badawi YB, El-Tayeb MA & Antranikian G (2015) Detergent-, solvent- and salt-compatible thermoactive alkaline serine protease from halotolerant alkaliphilic *Bacillus sp.* NPST-AK15: purification and characterization. *Extremophiles* **19,** 961–971.

106 Tehei M & Zaccai G (2005) Adaptation to extreme environments: macromolecular dynamics in complex systems. *Biochim Biophys Acta* **1724,** 404–410.

107 Alberto Cira-Chávez L, Guevara-Luna J, Yadira Soto-Padilla M, Román-Ponce B, Soledad Vásquez-Murrieta M & Isabel Estrada-Alvarado M Kinetics of halophilic enzymes. In *Kinetics of Enzymatic Synthesis.* (Rajendran L & Fernandez C, eds) **1,** 3–26 (2019). IntechOpen.

108 Coker JA (2016) Extremophiles and biotechnology: current uses and prospects. *F1000Res* **5,** 396.

109 Joo H-S, Kumar CG, Park G-C, Paik SR & Chang C-S (2003) Oxidant and SDS-stable alkaline protease from *Bacillus clausii* I-52: production and some properties. *J Appl Microbiol* **95,** 267–272.

110 Gupta A, Roy I, Patel RK, Singh SP, Khare SK & Gupta MN (2005) One-step purification and characterization of an alkaline protease from haloalkaliphilic *Bacillus* sp. *J Chromatogr A* **1075,** 103–108.

111 Mokashe N, Chaudhari B & Patil U (2018) Operative utility of salt-stable proteases of halophilic and halotolerant bacteria in the biotechnology sector. *Int J Biol Macromol* **117,** 493–522.

112 Oren A (2008) Microbial life at high salt concentrations: phylogenetic and metabolic diversity. *Saline Syst* **4,** 2.

113 Antón J Halophile. In *Encyclopedia of Astrobiology.* (Gargaud M, Amils R, Quintanilla JC, Cleaves HJ, Irvine WM, Pinti DL & Viso M, eds), 725–727 (2011). Berlin. Springer.

114 Kushner D (1978) Microbial life in extreme environments. Academic Press, London, New York.

115 Oren A (2002) Diversity of halophilic microorganisms: Environments, phylogeny, physiology, and applications. *J Ind Microbiol Biotech* **28,** 56–63.

116 Ruginescu R, Gomoiu I, Popescu O, Cojoc R, Neagu S, Lucaci I, Batrinescu-Moteau C & Enache M (2020) Bioprospecting for novel halophilic and halotolerant sources of hydrolytic enzymes in backish, alkaline and hypersaline lakes of Romania. *Microorganisms* **8**.

117 Ventosa A, Nieto JJ & Oren A (1998) Biology of moderately halophilic aerobic bacteria. *Microbiol Mol Biol Rev* **62,** 504–544.

118 Gunde-Cimerman N, Plemenitaš A & Oren A (2018) Strategies of adaptation of microorganisms of the three domains of life to high salt concentrations. *FEMS Microbiol Rev* **42,** 353–375.

119 Wood JM (2011) Bacterial osmoregulation: a paradigm for the study of cellular homeostasis. *Annu Rev Microbiol* **65,** 215–238.

120 Kirsch F, Klähn S & Hagemann M (2019) Salt-regulated accumulation of the compatible solutes sucrose and glucosylglycerol in Cyanobacteria and its biotechnological potential. *Front Microbiol* **10,** 2139.

121 Krishna PS, Raghunathan S & Prakash JSS (2021) Comparative genome analysis of *Alkalihalobacillus okhensis* Kh10-101^{T} reveals insights into adaptive mechanisms for halo-alkali tolerance. *3 Biotech* **11,** 392.

122 Aono R, Ito M & Machida T (1999) Contribution of the cell wall component teichuronopeptide to pH homeostasis and alkaliphily in the alkaliphile *Bacillus lentus* C-125. *J Bacteriol* **181,** 6600–6606.

123 Krulwich TA, Hicks DB & Ito M (2009) Cation/proton antiporter complements of bacteria: why so large and diverse? *Mol Microbiol* **74,** 257–260.

124 Sang P, Liu S-Q & Yang L-Q (2020) New insight into mechanisms of protein adaptation to high temperatures: A comparative molecular dynamics simulation study of hermophilic and mesophilic subtilisin-like serine proteases. *Int J Mol Sci* **21**.

125 Engqvist MKM (2018) Correlating enzyme annotations with a large set of microbial growth temperatures reveals metabolic adaptations to growth at diverse temperatures. *BMC Microbiol* **18,** 177.

126 Du X, Sang P, Xia Y-L, Li Y, Liang J, Ai S-M, Ji X-L, Fu Y-X & Liu S-Q (2017) Comparative thermal unfolding study of psychrophilic and mesophilic subtilisin-like serine proteases by molecular dynamics simulations. *J Biomol Struct Dyn* **35,** 1500–1517.

127 Klibanov AM Stabilization of enzymes against thermal inactivation. In *Advances in Applied Microbiology Volume 29* **29,** 1–28 (1983). Elsevier.

128 Parvizpour S, Hussin N, Shamsir MS & Razmara J (2021) Psychrophilic enzymes: structural adaptation, pharmaceutical and industrial applications. *Appl Microbiol Biotechnol* **105,** 899–907.

129 Shirai T, Suzuki A, Yamane T, Ashida T, Kobayashi T, Hitomi J & Ito S (1997) High-resolution crystal structure of M-protease: phylogeny aided analysis of the high-alkaline adaptation mechanism. *Protein Eng* **10,** 627–634.

130 Fitch CA, Platzer G, Okon M, Garcia-Moreno BE & McIntosh LP (2015) Arginine: Its pKa value revisited. *Protein Sci* **24,** 752–761.

131 Takenaka S, Takada A, Kimura Y, Watanabe M & Kuntiya A (2022) Improvement of the halotolerance of a *Bacillus* serine protease by protein surface engineering. *J Basic Microbiol* **62,** 174–184.

132 Sinha R & Khare SK (2014) Protective role of salt in catalysis and maintaining structure of halophilic proteins against denaturation. *Front Microbiol* **5,** 165.

133 Qiao B, Jiménez-Ángeles F, Nguyen TD & La Olvera de Cruz M (2019) Water follows polar and nonpolar protein surface domains. *Proc Natl Acad Sci USA* **116,** 19274–19281.

134 Rhodes ME, Fitz-Gibbon ST, Oren A & House CH (2010) Amino acid signatures of salinity on an environmental scale with a focus on the Dead Sea. *Environ Microbiol* **12,** 2613–2623.

135 Takenaka S, Yoshinami J, Kuntiya A, Techapun C, Leksawasdi N, Seesuriyachan P, Chaiyaso T, Watanabe M, Tanaka K & Yoshida K-I (2018) Characterization and mutation analysis of a halotolerant serine protease from a new isolate of *Bacillus subtilis*. *Biotechnol Lett* **40,** 189–196.

136 van den Burg B, Eijsink VG, Stulp BK & Venema G (1990) Identification of autodigestion target sites in *Bacillus subtilis* neutral proteinase. *Biochem J* **272,** 93–97.

137 Stawiski EW, Baucom AE, Lohr SC & Gregoret LM (2000) Predicting protein function from structure: unique structural features of proteases. *Proc Natl Acad Sci USA* **97,** 3954–3958.

138 Tatsumi H, Ikegaya K, Murakami S, Kawabe H, Nakano E & Motai H (1994) Elucidation of the thermal stability of the neutral proteinase II from *Aspergillus oryzae*. *BBA Protein Struct Mol Enzym* **1208,** 179–185.

139 Freudl R (2018) Signal peptides for recombinant protein secretion in bacterial expression systems. *Microb Cell Fact* **17,** 52.

140 Terpe K (2006) Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol* **72,** 211–222.

141 Rozanov AS, Shekhovtsov SV, Bogacheva NV, Pershina EG, Ryapolova AV, Bytyak DS & S EP (2021) Production of subtilisin proteases in bacteria and yeast. *Vavilov J. Genet. Breed* **25,** 125–134.

142 Schallmey M, Singh A & Ward OP (2004) Developments in the use of *Bacillus* species for industrial production. *Can J Microbiol* **50,** 1–17.

143 Su Y, Liu C, Fang H & Zhang D (2020) *Bacillus subtilis*: a universal cell factory for industry, agriculture, biomaterials and medicine. *Microb Cell Fact* **19,** 173.

144 Contesini FJ, Melo RRd & Sato HH (2017) An overview of *Bacillus* proteases: from production to application. *Crit Rev Biotechnol* **38,** 321–334.

145 Gupta R, Beg QK, Khan S & Chauhan B (2002) An overview on fermentation, downstream processing and properties of microbial alkaline proteases. *Appl Microbiol Biotechnol* **60,** 381–395.

146 Patel AR, Mokashe NU, Chaudhari DS, Jadhav AG & Patil UK (2019) Production optimisation and characterisation of extracellular protease secreted by newly isolated *Bacillus subtilis* AU-2 strain obtained from *Tribolium castaneum* gut. *Biocatal Agric Biotechnol* **19,** 101122.

147 Pohl S & Harwood CR (2010) Heterologous protein secretion by *Bacillus* species. *Adv Appl Microbiol,* 1–25.

148 Tsirigotaki A, Geyter J de, Šoštaric N, Economou A & Karamanou S (2017) Protein export through the bacterial Sec pathway. *Nat Rev Microbiol* **15,** 21–36.

149 Rusch SL & Kendall DA (2007) Interactions that drive Sec-dependent bacterial protein transport. *Biochem* **46,** 9665–9673.

150 Denks K, Vogt A, Sachelaru I, Petriman N-A, Kudva R & Koch H-G (2014) The Sec translocon mediated protein transport in prokaryotes and eukaryotes. *Mol Membr Biol* **31,** 58–84.

151 Zhang K, Su L & Wu J (2019 // 2020) Recent advances in recombinant protein production by Bacillus subtilis // Recent Advances in Recombinant Protein Production by Bacillus subtilis. *Annu Rev Food Sci Technol* **11,** 295–318.

152 Tufvesson P, Lima-Ramos J, Nordblad M & Woodley JM (2011) Guidelines and cost analysis for catalyst production in biocatalytic processes. *Org. Process Res. Dev.* **15,** 266–274.

153 Degering C, Eggert T, Puls M, Bongaerts J, Evers S, Maurer K-H & Jaeger K-E (2010) Optimization of protease secretion in *Bacillus subtilis* and *Bacillus licheniformis* by screening of homologous and heterologous signal peptides. *Appl Environ Microbiol* **76,** 6370–6376.

154 Kabisch J, Pratzka I, Meyer H, Albrecht D, Lalk M, Ehrenreich A & Schweder T (2013) Metabolic engineering of *Bacillus subtilis* for growth on overflow metabolites. *Microb Cell Fact* **12,** 72.

155 Li W, Zhou X & Lu P (2004) Bottlenecks in the expression and secretion of heterologous proteins in *Bacillus subtilis*. *Res Microbiol* **155,** 605–610.

156 Anné J, Vrancken K, van Mellaert L, van Impe J & Bernaerts K (2014) Protein secretion biotechnology in Gram-positive bacteria with special emphasis on *Streptomyces lividans*. *Biochim Biophys Acta* **1843,** 1750–1761.

157 Freudl R (2017) Beyond amino acids: Use of the *Corynebacterium glutamicum* cell factory for the secretion of heterologous proteins. *J Biotechnol* **258,** 101–109.

158 Morello E, Bermúdez-Humarán LG, Llull D, Solé V, Miraglio N, Langella P & Poquet I (2008) *Lactococcus lactis*, an efficient cell factory for recombinant protein production and secretion. *J Mol Microbiol Biotechnol* **14,** 48–58.

159 Yoon J-H, Lee C-H & Oh T-K (2005) *Bacillus cibi* sp. nov., isolated from jeotgal, a traditional Korean fermented seafood. *Int J Syst Evol Microbiol* **55,** 733–736.

160 Stropko SJ, Pipes SE & Newman JD (2014) Genome-based reclassification of *Bacillus cibi* as a later heterotypic synonym of *Bacillus indicus* and emended description of *Bacillus indicus*. *Int J Syst Evol Microbiol* **64,** 3804–3809.

161 Le Duc H, Fraser PD, Tam NKM & Cutting SM (2006) Carotenoids present in halotolerant *Bacillus* spore formers. *FEMS Microbiol Lett* **255,** 215–224.

162 Heinze S, Kornberger P, Grätz C, Schwarz WH, Zverlov VV & Liebl W (2018) Transmating: conjugative transfer of a new broad host range expression vector to various *Bacillus* species using a single protocol. *BMC Microbiol* **18,** 56.

163 Heijne G von (1990) The signal peptide. *J Membr Biol* **115,** 195–201.

164 Shinde U & Inouye M (2000) Intramolecular chaperones: polypeptide extensions that modulate protein folding. *Semin Cell Dev Biol* **11,** 35–44.

165 Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard J-F, Guindon S, Lefort V, Lescot M, Claverie J-M & Gascuel O (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* **36,** W465-9.

166 Katoh K, Rozewicki J & Yamada KD (2019) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinformatics* **20,** 1160–1166.

167 Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32,** 1792–1797.

168 Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD & Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7,** 539.

169 Notredame C, Higgins DG & Heringa J (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302,** 205–217.

170 Dessimoz C & Gil M (2010) Phylogenetic assessment of alignments reveals neglected tree signal in gaps. *Genome Biol* **11,** R37.

171 Sánchez R, Serra F, Tárraga J, Medina I, Carbonell J, Pulido L, María A de, Capella-Gutíerrez S, Huerta-Cepas J, Gabaldón T, Dopazo J & Dopazo H (2011) Phylemon 2.0: a suite of web-tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing. *Nucleic Acids Res* **39,** W470-4.

172 Talavera G & Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* **56,** 564–577.

173 Dress AWM, Flamm C, Fritzsch G, Grünewald S, Kruspe M, Prohaska SJ & Stadler PF (2008) Noisy: identification of problematic columns in multiple sequence alignments. *Algorithms Mol Biol* **3,** 7.

174 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 Evol Biol* **10,** 210.

175 Tan G, Muffato M, Ledergerber C, Herrero J, Goldman N, Gil M & Dessimoz C (2015) Current methods for automated filtering of multiple sequence alignments frequently worsen single-gene phylogenetic inference. *Syst Biol* **64,** 778–791.

176 Trifinopoulos J, Nguyen L-T, Haeseler A von & Minh BQ (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res* **44,** W232-5.

177 Guindon S, Lethiec F, Duroux P & Gascuel O (2005) PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res* **33,** W557-9.

178 Lefort V, Desper R & Gascuel O (2015) FastME 2.0: A Comprehensive, Accurate, and Fast Distance-Based Phylogeny Inference Program. *Mol Biol Evol* **32,** 2798–2800.

179 Price MN, Dehal PS & Arkin AP (2009) FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* **26,** 1641–1650.

180 Kuhner MK & Felsenstein J (1994) A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Mol Biol Evol* **11,** 459–468.

181 Hoang DT, Chernomor O, Haeseler A von, Minh BQ & Le Vinh S (2018) UFBoot2: Improving the ultrafast bootstrap approximation. *Mol Biol Evol* **35,** 518–522.

182 Falkenberg F, Bott M, Bongaerts J & Siegert P (2022) Phylogenetic survey of the subtilase family and a data-mining-based search for new subtilisins from *Bacillaceae*. *Front Microbiol* **13,** 1017978.

183 Sanderson MJ & Shaffer HB (2002) Troubleshooting molecular phylogenetic analyses. *Annu. Rev. Ecol. Syst.* **33,** 49–72.

184 Kannan Y, Koga Y, Inoue Y, Haruki M, Takagi M, Imanaka T, Morikawa M & Kanaya S (2001) Active subtilisin-like protease from a hyperthermophilic archaeon in a form with a putative prosequence. *Appl Environ Microbiol* **67,** 2445–2452.

185 Rolbetzki A, Ammon M, Jakovljevic V, Konovalova A & Søgaard-Andersen L (2008) Regulated secretion of a protease activates intercellular signaling during fruiting body formation in *M. xanthus*. *Dev Cell* **15,** 627–634.

186 Taylor A & Qiu Y-L (2017) Evolutionary history of subtilases in land plants and their involvement in symbiotic interactions. *Mol Plant Microbe Interact* **30,** 489–501.

187 Xu Y, Wang S, Li L, Sahu SK, Petersen M, Liu X, Melkonian M, Zhang G & Liu H (2019) Molecular evidence for origin, diversification and ancient gene duplication of plant subtilases (SBTs). *Sci Rep* **9,** 12485.

188 Hu G & Leger RJS (2004) A phylogenomic approach to reconstructing the diversification of serine proteases in fungi. *J Evol Biol* **17,** 1204–1214.

189 Muszewska A, Taylor JW, Szczesny P & Grynberg M (2011) Independent subtilases expansions in fungi associated with animals. *Mol Biol Evol* **28,** 3395–3404.

190 Li J, Gu F, Wu R, Yang J & Zhang K-Q (2017) Phylogenomic evolutionary surveys of subtilase superfamily genes in fungi. *Sci Rep* **7,** 45456.

191 Hon J, Borko S, Stourac J, Prokop Z, Zendulka J, Bednar D, Martinek T & Damborsky J (2020) EnzymeMiner: automated mining of soluble enzymes with diverse structures, catalytic properties and stabilities. *Nucleic Acids Res* **48,** W104-W109.

192 Betzel C, Klupsch S, Branner S & Wilson KS (1996) Crystal structures of the alkaline proteases savinase and esperase from *Bacillus lentus*. *Adv Exp Med Biol* **379,** 49–61.

193 Phrommao E, Yongsawatdigul J, Rodtong S & Yamabhai M (2011) A novel subtilase with NaCl-activated and oxidant-stable activity from *Virgibacillus* sp. SK37. *BMC Biotechnol* **11,** 65.

194 Yamagata Y, Isshiki K & Ichishima E (1995) Subtilisin Sendai from alkalophilic *Bacillus* sp.: Molecular and enzymatic properties of the enzyme and molecular cloning and characterization of the gene, aprS. *Enzyme Microb Technol* **17,** 653–663.

195 Gavrilov Y, Dagan S & Levy Y (2015) Shortening a loop can increase protein native state entropy. *Proteins* **83,** 2137–2146.

196 Schauperl M, Fuchs JE, Waldner BJ, Huber RG, Kramer C & Liedl KR (2015) Characterizing protease specificity: How many substrates do we need? *PLoS ONE* **10,** e0142658.

197 Goddette DW, Paech C, Yang SS, Mielenz JR, Bystroff C, Wilke ME & Fletterick RJ (1992) The crystal structure of the *Bacillus lentus* alkaline protease, subtilisin BL, at 1.4 A resolution. *J Mol Biol* **228,** 580–595.

198 Lim J-M, Jeon CO, Park D-J, Kim H-R, Yoon B-J & Kim C-J (2005) *Pontibacillus marinus* sp. nov., a moderately halophilic bacterium from a solar saltern, and emended description of the genus *Pontibacillus*. *Int J Syst Evol Microbiol* **55,** 1027–1031.

199 Weisser J & Trüper HG (1985) Osmoregulation in a New Haloalkaliphilic *Bacillus* from the Wadi Natrun (Egypt). *Systematic and Applied Microbiology* **6,** 7–11.

200 Romano I, Lama L, Nicolaus B, Gambacorta A & Giordano A (2005) *Bacillus saliphilus* sp. nov., isolated from a mineral pool in Campania, Italy. *Int J Syst Evol Microbiol* **55,** 159–163.

201 Amoozegar MA, Sánchez-Porro C, Rohban R, Hajighasemi M & Ventosa A (2009) *Piscibacillus halophilus* sp. nov., a moderately halophilic bacterium from a hypersaline Iranian lake. *Int J Syst Evol Microbiol* **59,** 3095–3099.

202 Lee J-C, Lee GS, Park D-J & Kim C-J (2008) *Bacillus alkalitelluris* sp. nov., an alkaliphilic bacterium isolated from sandy soil. *Int J Syst Evol Microbiol* **58,** 2629–2634.

203 An S-Y, Kanoh K, Kasai H, Goto K & Yokota A (2007) *Halobacillus faecis* sp. nov., a spore-forming bacterium isolated from a mangrove area on Ishigaki Island, Japan. *Int J Syst Evol Microbiol* **57,** 2476–2479.

204 Zhai L, Liao T, Xue Y & Ma Y (2012) *Bacillus daliensis* sp. nov., an alkaliphilic, Gram-positive bacterium isolated from a soda lake. *Int J Syst Evol Microbiol* **62,** 949–953.

205 Zhu D, Tanabe S-H, Xie C, Honda D, Sun J & Ai L (2014) *Bacillus ligniniphilus* sp. nov., an alkaliphilic and halotolerant bacterium isolated from sediments of the South China Sea. *Int J Syst Evol Microbiol* **64,** 1712–1717.

206 Shivaji S, Suresh K, Chaturvedi P, Dube S & Sengupta S (2005) *Bacillus arsenicus* sp. nov., an arsenic-resistant bacterium isolated from a siderite concretion in West Bengal, India. *Int J Syst Evol Microbiol* **55,** 1123–1127.

207 Kawamura F & Doi RH (1984) Construction of a *Bacillus subtilis* double mutant deficient in extracellular alkaline and neutral proteases. *J Bacteriol* **160,** 442–444.

208 Dürrschmidt P, Mansfeld J & Ulbrich-Hofmann R (2005) An engineered disulfide bridge mimics the effect of calcium to protect neutral protease against local unfolding. *FEBS J* **272,** 1523–1534.

209 Reynolds JA & Tanford C (1970) The Gross Conformation of Protein-Sodium Dodecyl Sulfate Complexes. *J Biol Chem* **245,** 5161–5165.

210 Rath A, Glibowicka M, Nadeau VG, Chen G & Deber CM (2009) Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proc Natl Acad Sci USA* **106,** 1760–1765.

211 Matagne A, Joris B & Frère JM (1991) Anomalous behaviour of a protein during SDS/PAGE corrected by chemical modification of carboxylic groups. *Biochem J* **280 (Pt 2),** 553–556.

212 Maeda H, Mizutani O, Yamagata Y, Ichishima E & Nakajima T (2001) Alkaline-resistance model of subtilisin ALP I, a novel alkaline subtilisin. *J Biochem* **129,** 675–682.

213 Falkenberg F, Rahba J, Fischer D, Bott M, Bongaerts J & Siegert P (2022) Biochemical characterization of a novel oxidatively stable, halotolerant, and high-alkaline subtilisin from *Alkalihalobacillus okhensis* Kh10-101^{T}. *FEBS Open Bio* **12,** 1729–1746.

214 Li Z, Roccatano D, Lorenz M, Martinez R & Schwaneberg U (2014) Insights on activity and stability of subtilisin E towards guanidinium chloride and sodium dodecylsulfate. *J Biotechnol* **169,** 87–94.

215 Drejer EB, Hakvåg S, Irla M & Brautaset T (2018) Genetic tools and techniques for recombinant expression in thermophilic *Bacillaceae*. *Microorganisms* **6**.

216 Irla M, Drejer EB, Brautaset T & Hakvåg S (2020) Establishment of a functional system for recombinant production of secreted proteins at 50 °C in the thermophilic *Bacillus methanolicus*. *Microb Cell Fact* **19,** 151.

217 Mizukami M, Hanagata H & Miyauchi A (2010) *Brevibacillus* expression system: host-vector system for efficient production of secretory proteins. *Curr Pharm Biotechnol* **11,** 251–258.

218 Daas MJA, Nijsse B, van de Weijer AHP, Groenendaal BWAJ, Janssen F, van der Oost J & van Kranenburg R (2018) Engineering *Geobacillus thermodenitrificans* to introduce cellulolytic activity; expression of native and heterologous cellulase genes. *BMC Biotechnol* **18,** 42.

219 Scoma A, Coma M, Kerckhof F-M, Boon N & Rabaey K (2017) Efficient molasses fermentation under high salinity by inocula of marine and terrestrial origin. *Biotechnol Biofuels* **10,** 23.

220 Sewalt V, Shanahan D, Gregg L, La Marta J & Carrillo R (2016) The generally recognized as safe (GRAS) process for industrial microbial enzymes. *Ind Biotechnol* **12,** 295–302.

221 Marzorati M, van den Abbeele P, Bubeck S, Bayne T, Krishnan K & Young A (2021) Treatment with a spore-based probiotic containing five strains of *Bacillus* induced changes in the metabolic activity and community composition of the gut microbiota in a SHIME® model of the human gastrointestinal system. *Food Res Int* **149,** 110676.

222 Xue G-P, Johnson JS & Dalrymple BP (1999) High osmolarity improves the electro-transformation efficiency of the gram-positive bacteria *Bacillus subtilis* and *Bacillus licheniformis*. *J Microbiol Methods* **34,** 183–191.

223 Vojcic L, Despotovic D, Martinez R, Maurer K-H & Schwaneberg U (2012) An efficient transformation method for *Bacillus subtilis* DB104. *Appl Microbiol Biotechnol* **94,** 487–493.

224 Lévi-Meyrueis C, Fodor K & Schaeffer P (1980) Polyethyleneglycol-induced transformation of *Bacillus subtilis* protoplasts by bacterial chromosomal DNA. *Mol Gen Genet* **179,** 589–594.

225 Jan J, Valle F, Bolivar F & Merino E (2000) Characterization of the 5' subtilisin (aprE) regulatory region from *Bacillus subtilis*. *FEMS Microbiol Lett* **183,** 9–14.

226 Wang H, Yang L, Ping Y, Bai Y, Luo H, Huang H & Yao B (2016) Engineering of a *Bacillus amyloliquefaciens* strain with high neutral protease producing capacity and optimization of its fermentation conditions. *PLoS ONE* **11,** e0146373.

227 Küppers T, Steffen V, Hellmuth H, O'Connell T, Bongaerts J, Maurer K-H & Wiechert W (2014) Developing a new production host from a blueprint: *Bacillus pumilus* as an industrial enzyme producer. *Microb Cell Fact* **13,** 46.

228 Zhang M, Song J, Xiao J, Jin J, Nomura CT, Chen S & Wang Q (2022) Engineered multiple translation initiation sites: a novel tool to enhance protein production in *Bacillus licheniformis* and other industrially relevant bacteria. *Nucleic Acids Res* **50,** 11979–11990.

229 Aguilar Suárez R, Stülke J & van Dijl JM (2019) Less is more: Toward a genome-reduced *Bacillus* cell factory for "difficult proteins". *ACS Synth Biol* **8,** 99–108.

230 Rocha EP, Danchin A & Viari A (1999) Translation in *Bacillus subtilis*: roles and trends of initiation and termination, insights from a genome analysis. *Nucleic Acids Res* **27,** 3567–3576.

231 Falkenberg F, Voß L, Bott M, Bongaerts J & Siegert P (2023) New robust subtilisins from halotolerant and halophilic *Bacillaceae* // New robust subtilisins from halotolerant and halophilic Bacillaceae. *Appl Microbiol Biotechnol* **107,** 3939–3954.

232 Shinde UP, Liu JJ & Inouye M (1997) Protein memory through altered folding mediated by intramolecular chaperones. *Nature* **389,** 520–522.

233 Zhang Y (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9,** 40.

234 Kidd RD, Yennawar HP, Sears P, Wong C-H & Farber GK (1996) A weak calcium binding site in subtilisin BPN‘ has a dramatic effect on protein stability. *J Am Chem Soc* **118,** 1645–1650.

235 Uehara R, Angkawidjaja C, Koga Y & Kanaya S (2013) Formation of the high-affinity calcium binding site in pro-subtilisin E with the insertion sequence IS1 of pro-Tk-subtilisin. *Biochem* **52,** 9080–9088.

236 Tanaka S-I, Matsumura H, Koga Y, Takano K & Kanaya S (2007) Four new crystal structures of Tk-subtilisin in unautoprocessed, autoprocessed and mature forms: insight into structural changes during maturation. *J Mol Biol* **372,** 1055–1069.

237 Smith CA, Toogood HS, Baker HM, Daniel RM & Baker EN (1999) Calcium-mediated thermostability in the subtilisin superfamily: the crystal structure of *Bacillus* Ak.1 protease at 1.8 A resolution. *J Mol Biol* **294,** 1027–1040.

238 Nowlan B, Dodia MS, Singh SP & Patel BKC (2006) *Bacillus okhensis* sp. nov., a halotolerant and alkalitolerant bacterium from an Indian saltpan. *Int J Syst Evol Microbiol* **56,** 1073–1077.

239 Fritze D (1996) *Bacillus haloalkaliphilus* sp. nov. *Int J Syst Evol Microbiol* **46,** 98–101.

240 Zeldovich KB, Chen P & Shakhnovich EI (2007) Protein stability imposes limits on organism complexity and speed of molecular evolution. *Proc Natl Acad Sci USA* **104,** 16152–16157.

241 Dürrschmidt P (2003) Entfaltung und Autoproteolyse der neutralen Protease aus *Bacillus stearothermophilus* und einer Disulfid-modifizierten Variante, Halle.

242 Chu NM, Chao Y & Bi RC (1995) The 2 A crystal structure of subtilisin E with PMSF inhibitor. *Protein Eng* **8,** 211–215.

243 Boivin S, Kozak S & Meijers R (2013) Optimization of protein purification and characterization using Thermofluor screens. *Protein Expr Purif* **91,** 192–206.

244 Stoner MR, Dale DA, Gualfetti PJ, Becker T, Manning MC, Carpenter JF & Randolph TW (2004) Protease autolysis in heavy-duty liquid detergent formulations: effects of thermodynamic stabilizers and protease inhibitors. *Enzyme Microb Technol* **34,** 114–125.

245 Pantoliano MW, Whitlow M, Wood JF, Dodd SW, Hardman KD, Rollence ML & Bryan PN (1989) Large increases in general stability for subtilisin BPN' through incremental changes in the free energy of unfolding. *Biochem* **28,** 7205–7213.

246 Takagi H, Yamamoto M, Ohtsu I & Nakamori S (1998) Random mutagenesis into the conserved Gly154 of subtilisin E: isolation and characterization of the revertant enzymes. *Protein Eng* **11,** 1205–1210.

247 Bae KH, Jang JS, Park KS, Lee SH & Byun SM (1995) Improvement of thermal stability of subtilisin J by changing the primary autolysis site. *Biochem Biophys Res Commun* **207,** 20–24.

248 Tsai YC, Yamasaki M, Yamamoto-Suzuki Y & Tamura G (1983) A new alkaline elastase of an alkalophilic *Bacillus* . *Biochem Int* **7,** 577–583.

249 Kobayashi T, Hakamada Y, Adachi S, Hitomi J, Yoshimatsu T, Koike K, Kawai S & Ito S (1995) Purification and properties of an alkaline protease from alkalophilic *Bacillus sp.* KSM-K16. *Appl Microbiol Biotechnol* **43,** 473–481.

250 Takami H, Akiba T & Horikoshi K (1990) Characterization of an alkaline protease from *Bacillus* sp. no. AH-101. *Appl Microbiol Biotechnol* **33,** 519–523.

251 Bhatt HB & Singh SP (2020) Cloning, expression, and structural elucidation of a biotechnologically potential alkaline serine protease from a newly isolated haloalkaliphilic *Bacillus lehensis* JO-26. *Front Microbiol* **11,** 941.

252 Takimura Y, Saito K, Okuda M, Kageyama Y, Saeki K, Ozaki K, Ito S & Kobayashi T (2007) Alkaliphilic *Bacillus sp*. strain KSM-LD1 contains a record number of subtilisin-like serine proteases genes. *Appl Microbiol Biotechnol* **76,** 395–405.

253 Panja AS, Maiti S & Bandyopadhyay B (2020) Protein stability governed by its structural plasticity is inferred by physicochemical factors and salt bridges. *Sci Rep* **10,** 1822.

254 Tekin A, Uzuner U & Sezen K (2021) Homology modeling and heterologous expression of highly alkaline subtilisin-like serine protease from *Bacillus halodurans* C-125. *Biotechnol Lett* **43,** 479–494.

255 Gurunathan R, Huang B, Ponnusamy VK, Hwang J-S & Dahms H-U (2021) Novel recombinant keratin degrading subtilisin like serine alkaline protease from *Bacillus cereus* isolated from marine hydrothermal vent crabs. *Sci Rep* **11,** 12007.

256 Welker NE & Campbell LL (1967) Unrelatedness of *Bacillus amyloliquefaciens* and *Bacillus subtilis*. *J Bacteriol* **94,** 1124–1130.

257 Gordon RE, Hyde JL & Moore JA (1977) *Bacillus firmus-Bacillus lentus*: a series or one species? *Int J Syst Evol Microbiol* **27,** 256–262.

258 Schroeter R, Hoffmann T, Voigt B, Meyer H, Bleisteiner M, Muntel J, Jürgen B, Albrecht D, Becher D, Lalk M, Evers S, Bongaerts J, Maurer K-H, Putzer H, Hecker M, Schweder T & Bremer E (2013) Stress responses of the industrial workhorse *Bacillus licheniformis* to osmotic challenges. *PLoS ONE* **8,** e80956.

259 Thaz CJ & Jayaraman G (2014) Stability and detergent compatibility of a predominantly β-sheet serine protease from halotolerant *B. aquimaris* VITP4 strain. *Appl Biochem Biotechnol* **172,** 687–700.

260 Daoud L, Jlidi M, Hmani H, Hadj Brahim A, El Arbi M & Ben Ali M (2017) Characterization of thermo-solvent stable protease from *Halobacillus sp.* CJ4 isolated from Chott Eldjerid hypersaline lake in Tunisia. *J Basic Microbiol* **57,** 104–113.

261 Joshi S & Satyanarayana T (2013) Characteristics and applications of a recombinant alkaline serine protease from a novel bacterium *Bacillus lehensis*. *Bioresour Technol* **131,** 76–85.

262 Kembhavi AA, Kulkarni A & Pant A (1993) Salt-tolerant and thermostable alkaline protease from *Bacillus subtilis* NCIM no. 64. *Appl Biochem Biotechnol* **38,** 83–92.

263 Olivera N, Sequeiros C, Siñeriz F & Breccia JD (2006) Characterisation of alkaline proteases from a novel alkali-tolerant bacterium *Bacillus patagoniensis*. *World J Microbiol Biotechnol* **22,** 737–743.

264 Nonaka T, Fujihashi M, Kita A, Saeki K, Ito S, Horikoshi K & Miki K (2004) The crystal structure of an oxidatively stable subtilisin-like alkaline serine protease, KP-43, with a C-terminal beta-barrel domain. *J Biol Chem* **279,** 47344–47351.

265 Stauffer CE & Etson D (1969) The effect on subtilisin activity of oxidising a methionine residue. *J Biol Chem* **244,** 5333–5338.

266 Bott R, Ultsch M, Kossiakoff A, Graycar T, Katz B & Power S (1988) The three-dimensional structure of *Bacillus amyloliquefaciens* subtilisin at 1.8 A and an analysis of the structural consequences of peroxide inactivation. *J Biol Chem* **263,** 7895–7906.

267 Estell DA, Graycar TP & Wells JA (1985) Engineering an enzyme by site-directed mutagenesis to be resistant to chemical oxidation. *J Biol Chem* **260,** 6518–6521.

268 Grøn H, Bech LM, Branner S & Breddam K (1990) A highly active and oxidation-resistant subtilisin-like enzyme produced by a combination of site-directed mutagenesis and chemical modification. *Eur J Biochem* **194,** 897–901.

269 Yang Y, Jiang L, Zhu L, Wu Y & Yang S (2000) Thermal stable and oxidation-resistant variant of subtilisin E. *J Biotechnol* **81,** 113–118.

270 Kazan D, Denizci AA, Oner MNK & Erarslan A (2005) Purification and characterization of a serine alkaline protease from *Bacillus clausii* GMBAE 42. *J Ind Microbiol Biotechnol* **32,** 335–344.

271 Dodia MS, Rawal CM, Bhimani HG, Joshi RH, Khare SK & Singh SP (2008) Purification and stability characteristics of an alkaline serine protease from a newly isolated Haloalkaliphilic bacterium sp. AH-6. *J Ind Microbiol Biotechnol* **35,** 121–131.

272 Vidyasagar M, Prakash S, Mahajan V, Shouche YS & Sreeramulu K (2009) Purification and characterization of an extreme halothermophilic protease from a halophilic bacterium *Chromohalobacter sp.* TVSP101. *Braz J Microbiol* **40,** 12–19.

273 Abraham LD & Breuil C (1995) Factors affecting autolysis of a subtilisin-like serine proteinase secreted by *Ophiostoma piceae* and identification of the cleavage site. *Biochim Biophys Acta Gen Subj* **1245,** 76–84.

274 Almog O, Gallagher T, Tordova M, Hoskins J, Bryan P & Gilliland GL (1998) Crystal structure of calcium-independent subtilisin BPN' with restored thermal stability folded without the prodomain. *Proteins* **31,** 21–32.

275 Grøn H, Meldal M & Breddam K (1992) Extensive comparison of the substrate preferences of two subtilisins as determined with peptide substrates which are based on the principle of intramolecular quenching. *Biochem* **31,** 6011–6018.

276 Georgieva DN, Stoeva S, Voelter W, Genov N & Betzel C (2001) Differences in the specificities of the highly alkalophilic proteinases Savinase and Esperase imposed by changes in the rigidity and geometry of the substrate binding sites. *Arch Biochem Biophys* **387,** 197–201.

277 Bieth J, Spiess B & Wermuth CG (1974) The synthesis and analytical use of a highly sensitive and convenient substrate of elastase. *Biochem Med* **11,** 350–357.

278 Tsai YC, Yamasaki M & Tamura G (1984) Substrate specificity of a new alkaline elastase from an alkalophilic *Bacillus*. *Biochem Int* **8,** 283–288.

279 Georgieva D, Genov N, Voelter W & Betzel C (2006) Catalytic efficiencies of alkaline proteinases from microorganisms. *Z Naturforsch C J Biosci* **61,** 445–452.

6 Supplementary

6.1 Supplementary: Phylogenetic survey of the subtilase family and a data-mining-based search for new subtilisins from *Bacillaceae*

Table S1 Clustered data mining S8 *Bacillaceae* sequences

Number #	*Bacterial origin*
S08.105_aerolysin	*Pyrobaculum aerophilum*
S08.129_Tk-sub.	*Thermococcus kodakarensis*
S08.113_sfericase	*Lysinibacillus sphaericus*
S08.016_WF146_pep.	*Bacillus* sp. WF146
S08.140_sub._S41	*Bacillus subtilis* TA41
P29139_INT72	*Paenibacillus polymyxa*
S08.030_IspA_pep.	*Bacillus subtilis*
Q45621 Isp-Q	*Bacillus* sp. NKS-21
P74937_TIAP	*Thermoactinomyces sp.*
S08.143_PopC_pep.	*Myxococcus xanthus*
08.126_EPR_pep.	*Bacillus subtilis*
S08.028_ALTP	*Alkaliphilus transvaalensis*
WP 146817052.1	*Alkalibacillus haloalkaliphilus*
WP 168009413.1	*Alkalicoccus luteus*
WP 171051829.1	*Alteribacter natronophilus*
WP 026691136.1	*Alteribacter aurantiacus*
WP 096188791.1	*Bacillus* sp. FJAT 44876
WP 202080138.1	*Bacillus* sp. YIM B00319
WP 091776386.1	*Piscibacillus halophilus*
WP 134338579.1	*Filobacillus milosensis*
WP 101332746.1	*Halalkalibacillus sediminis*
WP 167261846.1	*Alkalibacillus almallahensis*
WP 027963976.1	*Halalkalibacillus halophilus*
WP 144089130.1	*Allobacillus salarius*
WP 124221886.1	*Aquisalibacillus elongatus*
WP 091776380.1	*Piscibacillus halophilus*
WP 134339482.1	*Filobacillus milosensis*
WP 146817048.1	*Alkalibacillus haloalkaliphilus*
WP 188208160.1	*Alkalibacillus aidingensis*
WP 027965007.1	*Halalkalibacillus halophilus*
WP 101331250.1	*Halalkalibacillus sediminis*
WP 107583584.1	*Alkalicoccus saliphilus*
WP 168007760.1	*Alkalicoccus luteus*
WP 090843404.1	*Alkalicoccus daliensis*
WP 168006597.1	*Alkalicoccus luteus*
WP 105960433.1	*Alkalicoccus urumqiensis*
WP 147804655.1	*Alkalicoccus halolimnae*
WP 146817050.1	*Alkalihalobacillus haloalkaliphilus*

WP 091776383.1	*Piscibacillus halophilus*
WP 134339480.1	*Filobacillus milosensis*
WP 138811387.1	*Alteribacter natronophilus*
WP 230895209.1	*Salipaludibacillus* sp. CUR1
WP 110612024.1	*Salipaludibacillus keqinensis*
WP 122897711.1	*Bacillus* sp. KQ 3
WP 047973355.1	*Bacillus* sp. LL01
WP 075683870.1	*Alkalihalobacillus pseudofirmus*
WP 100374144.1	*Bacillus* sp. FJAT 45037
WP 202078324.1	*Bacillus* sp. YIM B00319
WP 226516443.1	*Bacillus shivajii*
WP 216831833.1	*Alkalihalobacterium elongatum*
WP 129080804.1	*Anaerobacillus alkaliphilus*
WP 035661169.1	*Halalkalibacter akibai*
WP 100334247.1	*Bacillus alkalisoli*
WP 084380659.1	*Sutcliffiella cohnii*
S08.133_sub._LD-1	*Bacillus* sp. KSM-LD1
WP 017729072.1	*Halalkalibacterium ligniniphilus*
WP 122896828.1	*Bacillus* sp. KQ 3
WP 047973137.1	*Bacillus* sp. LL01
WP 022628745.1	*Alkalihalophilus marmarensis*
S08.045_sub._ALP_1	*Bacillus* sp. nks-21
WP 210595747.1	*Bacillus* sp. YZJH907 2
S08.046_sub._aprM	*Halalkalibacterium halodurans*
WP 100374143.1	*Bacillus* sp. FJAT 45037
WP 035666680.1	*Alkalihalobacillus akibai*
WP 078596166.1	*Evansella clarkii*
WP 216831504.1	*Alkalihalobacterium elongatum*
WP 129077943.1	*Anaerobacillus alkaliphilus*
WP 203088820.1	*Shouchella gibsonii*
WP 143849870.1	*Bacillus* sp. P16 2019
WP 059104808.1	*Shouchella shacheensis*
WP 090775603.1	*Alkalihalobacillus lonarensis*
WP 090774843.1	*Alkalihalobacillus lonarensis*
KMK76635.1 pept	*Alkalihalobacillus pseudalcaliphilus*
WP 034632645.1	*Alkalihalobacillus okhensis*
S08.098_sub._sendai	*Bacillus* sp. G-825-6
S08.010_M-pep.	*Shouchella clausii*
S08.003_Savinase	*Lederbergia lentus*
S08.038_PB92	*Alkalihalobacillus alcalophilus*
WP 060704798.1	*Shouchella miscanthi*
S08.157_sub._YaB	*Bacillus* sp. YAB
WP 203087429.1	*Shouchella gibsonii*
WP 143850013.1	*Bacillus* sp. P16 2019
WP 059105057.1	*Shouchella shacheensis*

EZH65969.1 pept	*Bacillaceae bacterium JMAK1*
WP 078393865.1	*Shouchella patagoniensis*
WP 035392836.1	*Bacillus* sp. JCM19047
WP 163537364.1	*Gracilibacillus sp. YIM 98692*
WP 164853199.1	*Peribacillus asahii*
WP 099092793.1	*Bacillus weihaiensis*
WP 018922084.1	*Salsuginibacillus kocurii*
WP 106589713.1	*Salsuginibacillus halophilus*
WP 022794977.1	*Marinococcus halotolerans*
WP 107586282.1	*Alkalicoccus saliphilus*
WP 078597775.1	*Evansella clarkii*
WP 100832725.1	*Bacillus* sp. FJAT 45348
WP 090774498.1	*Alkalihalobacillus lonarensis*
WP 110520788.1	*Bacillus lacisalsi*
WP 096186536.1	*Bacillus* sp. FJAT 44876
WP 026691049.1	*Alteribacter aurantiacus*
WP 122900894.1	*Bacillus* sp. KQ 3
WP 231417544.1	*Pontibacillus* sp. HN14
WP 051255158.1	*Pontibacillus marinus*
WP 206945444.1	*Halobacillus* sp. GSS1
WP 173918387.1	*Halobacillus* sp. Marseille-Q1614
WP 188377243.1	*Halobacillus andaensis*
WP 070119644.1	*Bacillus marinisedimentorum*
WP_077360649.1	*Fictibacillus arsenicus*
WP 153236691.1	*Fictibacillus phosphorivorans*
WP 181472841.1	*Halobacillus locisalis*
WP 136946078.1	*Pseudalkalibacillus hwajinpoensis*
WP 224844257.1	*Bacillus timonensis*
WP 094921089.1	*Lottiidibacillus patelloidae*
WP 193538138.1	*Cytobacillus luteolus*
WP 090849877.1	*Litchfieldia salsus*
WP 078544469.1	*Litchfieldia alkalitelluris*
WP 224844255.1	*Bacillus timonensis*
WP 230500606.1	*Bacillus* sp. JJ 125
WP 088017821.1	*Sutcliffiella horikoshii*
WP 224838688.1	*Bacillus tianshenii*
WP 066412694.1	*Sutcliffiella cohnii*
WP 100334303.1	*Bacillus alkalisoli*
AST90329.1 pept	*Sutcliffiella cohnii*
WP 096155439.1	*Bacillus alkalisoli*
WP 230500539.1	*Bacillus* sp. JJ 125
WP 152444042.1	*Bacillus* sp. THAF10
WP 078381234.1	*Sutcliffiella halmapala*
WP 060666810.1	*Bacillus* sp. CHD6a
TDL80277.1 pept	*Brevibacterium frigoritolerans*
WP 029565418.1	*Metabacillus indicus*

WP 053604255.1	*Bacillus gobiensis*
AAS86761.1_kera	*Bacillus licheniformis*
S08.001_sub._Carlsberg	*Bacillus licheniformis*
NPC92104.1 S8 f	*Bacillus* sp. WMMC1349
NUJ19608.1 S8 f	*Bacillus glycinifermentans*
S08.037_sub._DY	*Bacillus subtilis*
WP 199800957.1	*Bacillus pumilus*
S08.005_endopep._Q	*Bacillus pumilus*
CAO03040.1 sapB	*Bacillus pumilus*
WP 003327717.1	*Bacillus atrophaeus*
S08.034_sub._BPN'	*Bacillus amyloliquefaciens*
WP 039073463.1	*Bacillus vallismortis*
S08.044_sub._NAT	*Bacillus subtilis*
S08.036_sub._E	*Bacillus subtilis*
S08.002_mesentericopep.	*Bacillus pumilus*
S08.042_amylosacchariticus	*Bacillus subtilis*
S08.035_sub._J	*Geobacillus stearothermophilus*

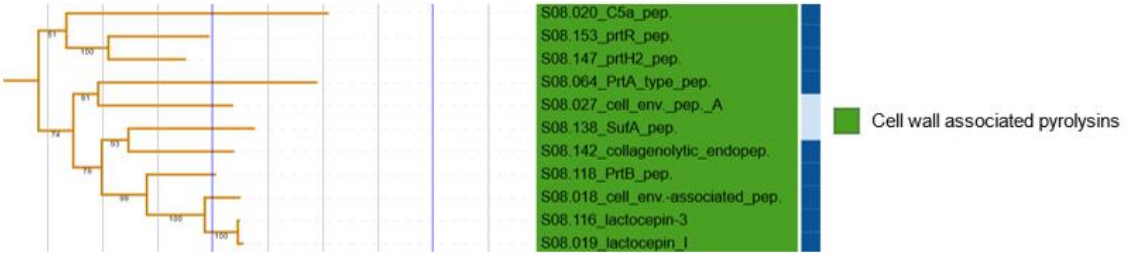


Figure S1 Tree of cell wall associated subgroup (pyrolysin) of MEROPS holotype dataset.

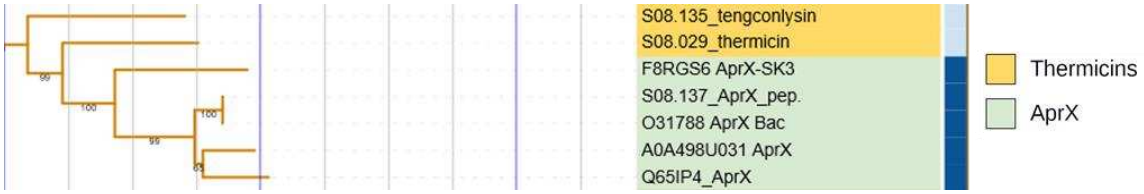


Figure S2 Tree of thermicin and AprX subgroup (pyrolysin) of MEROPS holotype dataset.

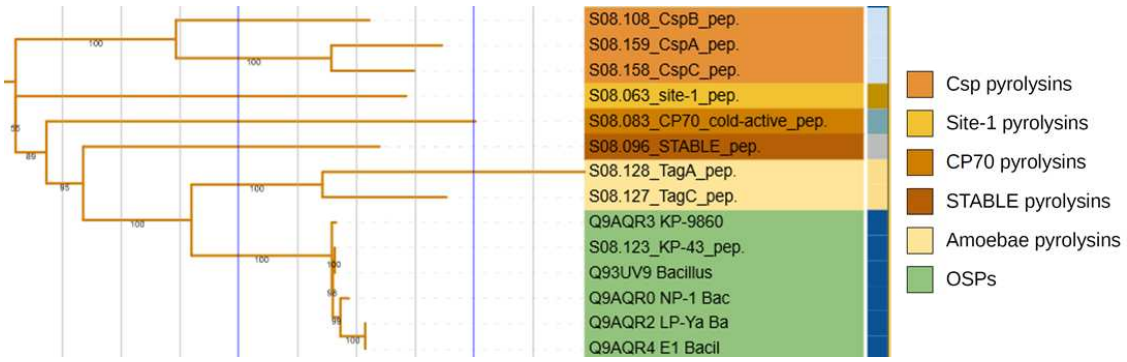


Figure S3 Tree of Csp, Site-1, CP70, STABLE, Amoebae and OSP (pyrolysin) of MEROPS holotype dataset.

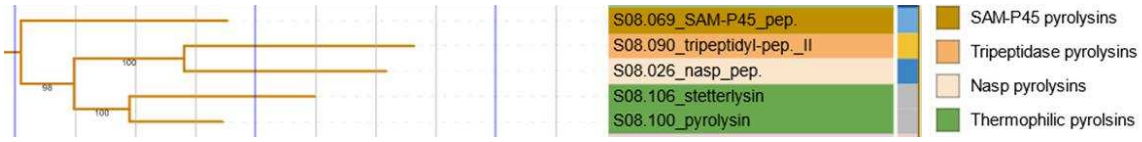


Figure S4 Tree of SAM-P45, TPPII, Nasp and pyrolysin subgroup (pyrolysin) of MEROPS holotype dataset.

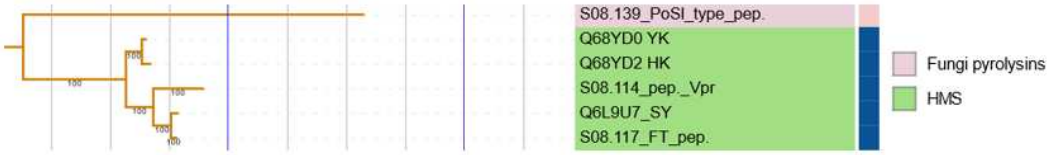


Figure S5 Tree of fungal pyrolysins and HMS subgroup (pyrolysin) of MEROPS holotype dataset.

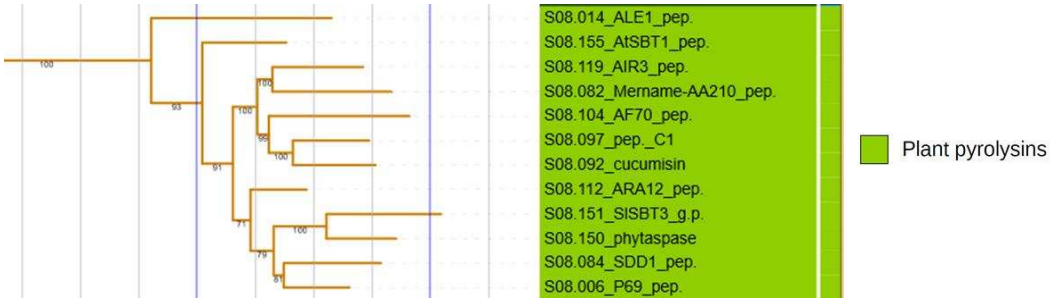


Figure S6 Tree of plant subgroup (pyrolysin) of MEROPS holotype dataset.

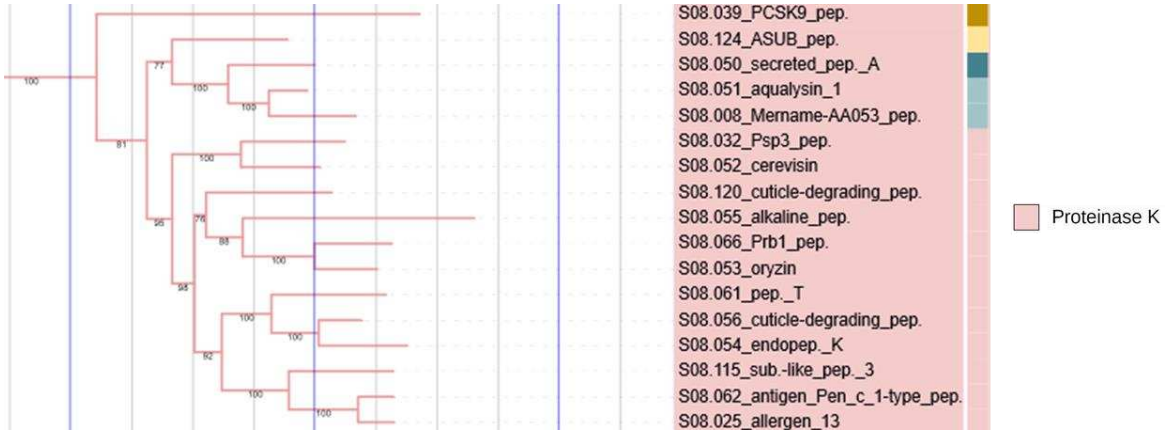


Figure S7 Tree of proteinase K group of MEROPS holotype dataset.

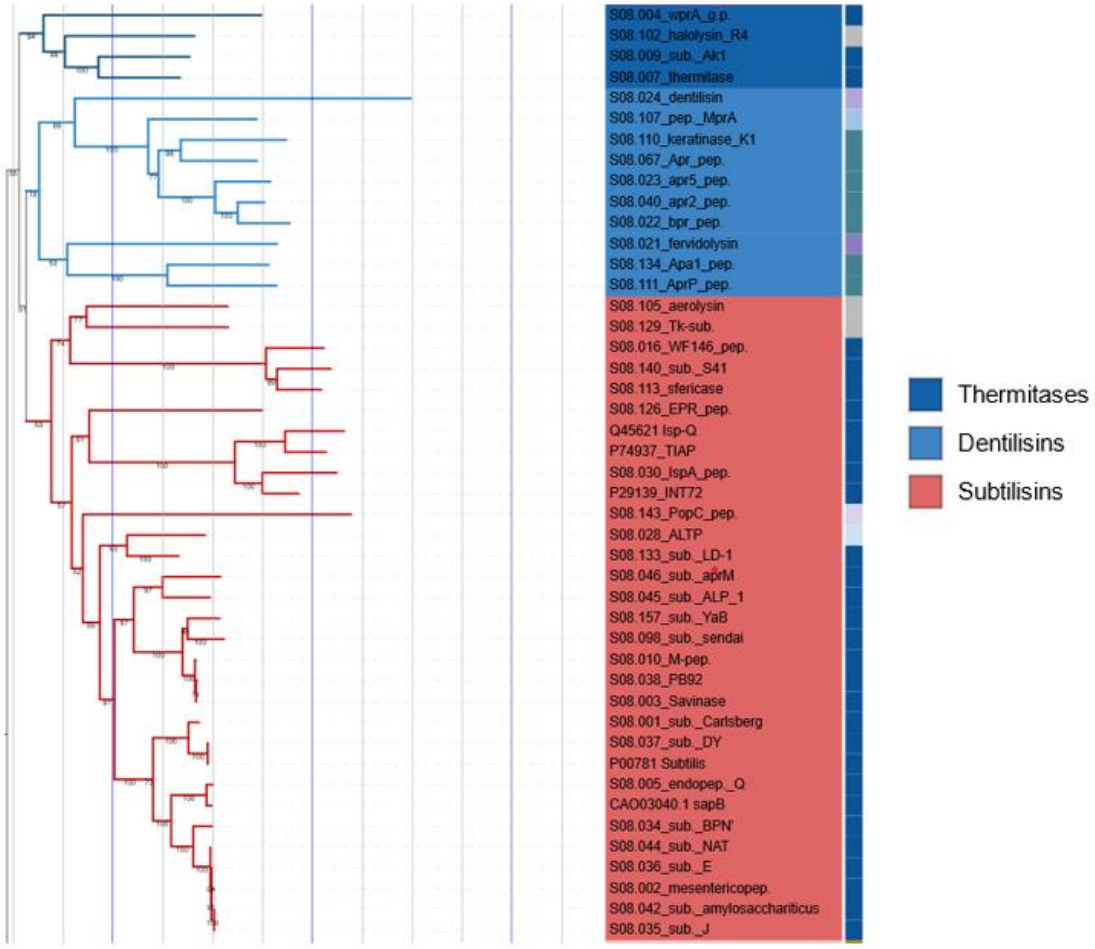


Figure S8 Tree of subtilisin, dentilisin and thermitase group of MEROPS holotype dataset.

Figure S9 Tree of bpF, CDF, SubAB, mycosins and transamidating groups of MEROPS holotype dataset.

Figure S10 Tree of TPPS and autotransporter group of MEROPS holotype dataset.

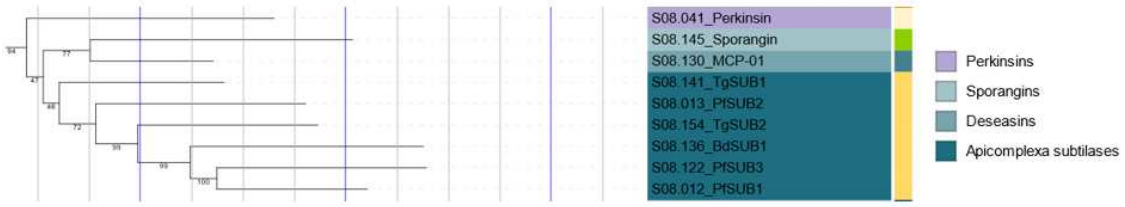


Figure S11 Tree of perkinsin, sporangin, deseasin and apicomplexa group of MEROPS holotype dataset.

Figure S12 Tree of lantibiotic peptidase group of MEROPS holotype dataset.

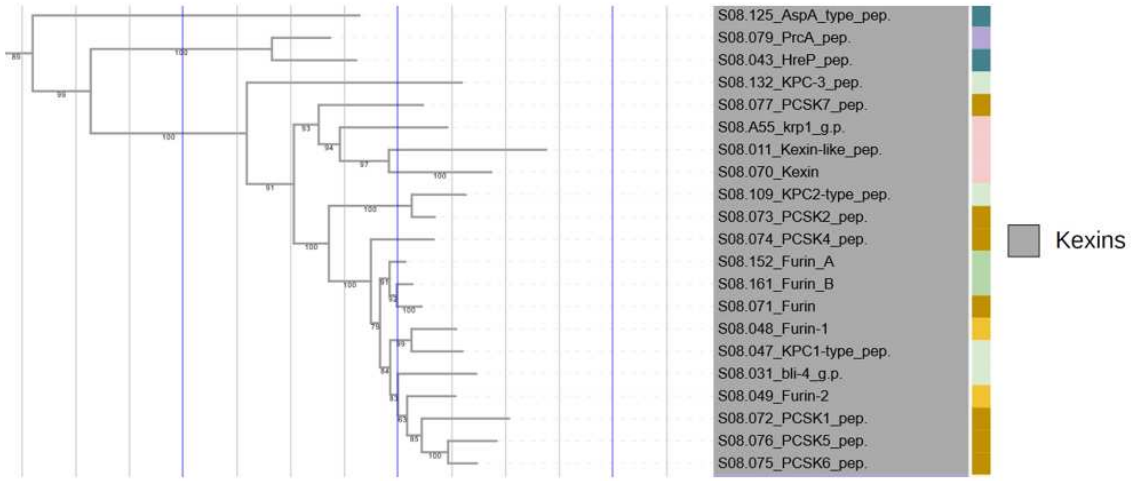


Figure S13 Tree of kexin group of MEROPS holotype dataset.

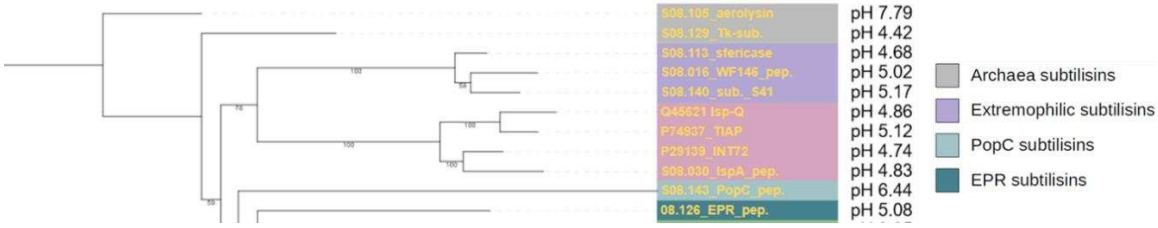


Figure S14 Tree of several subgroups of subtilisin group of MEROPS holotype dataset.

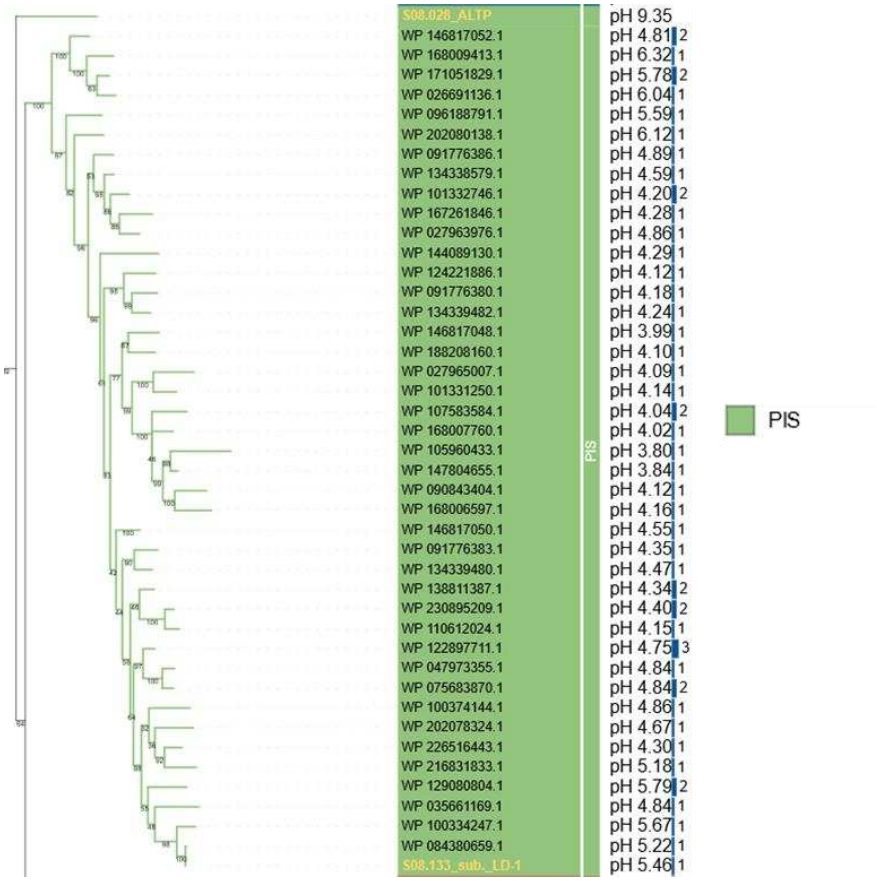


Figure S15 Tree of PIS subgroup of subtilisin group.

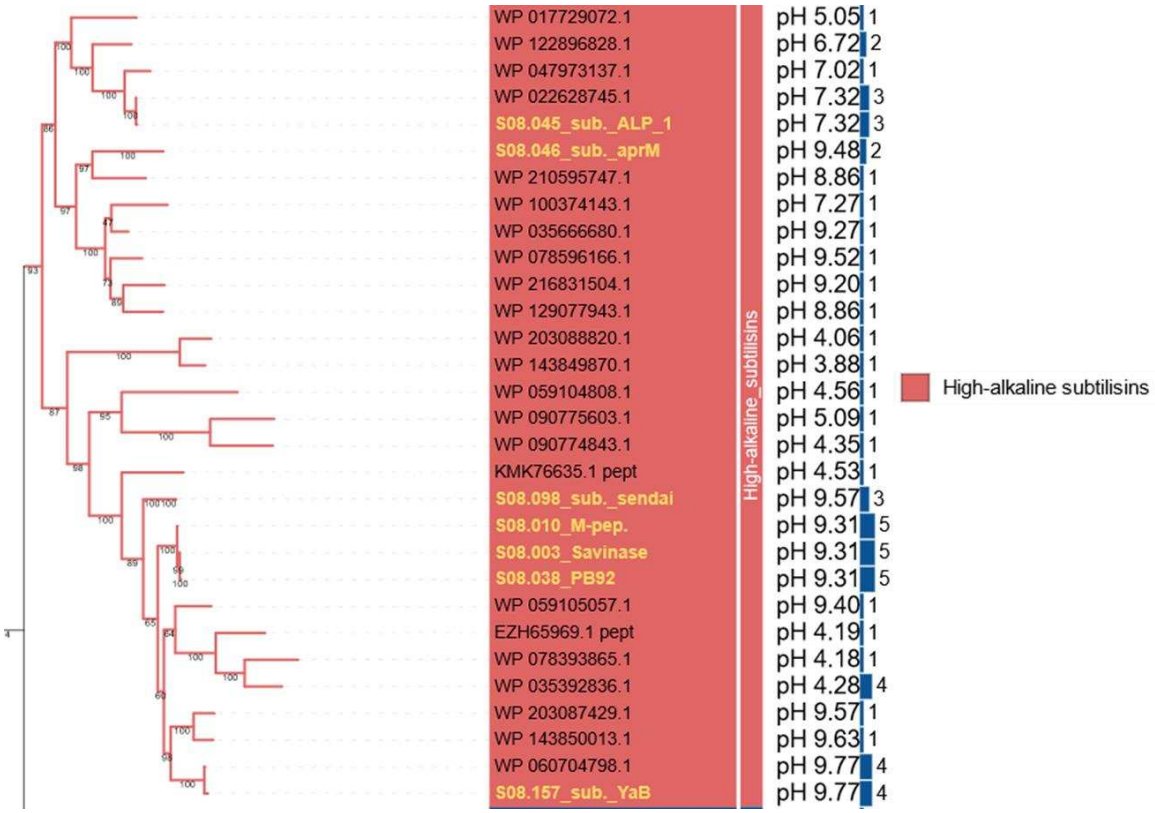


Figure S16 Tree of high-alkaline subtilisins subgroup of subtilisin group.

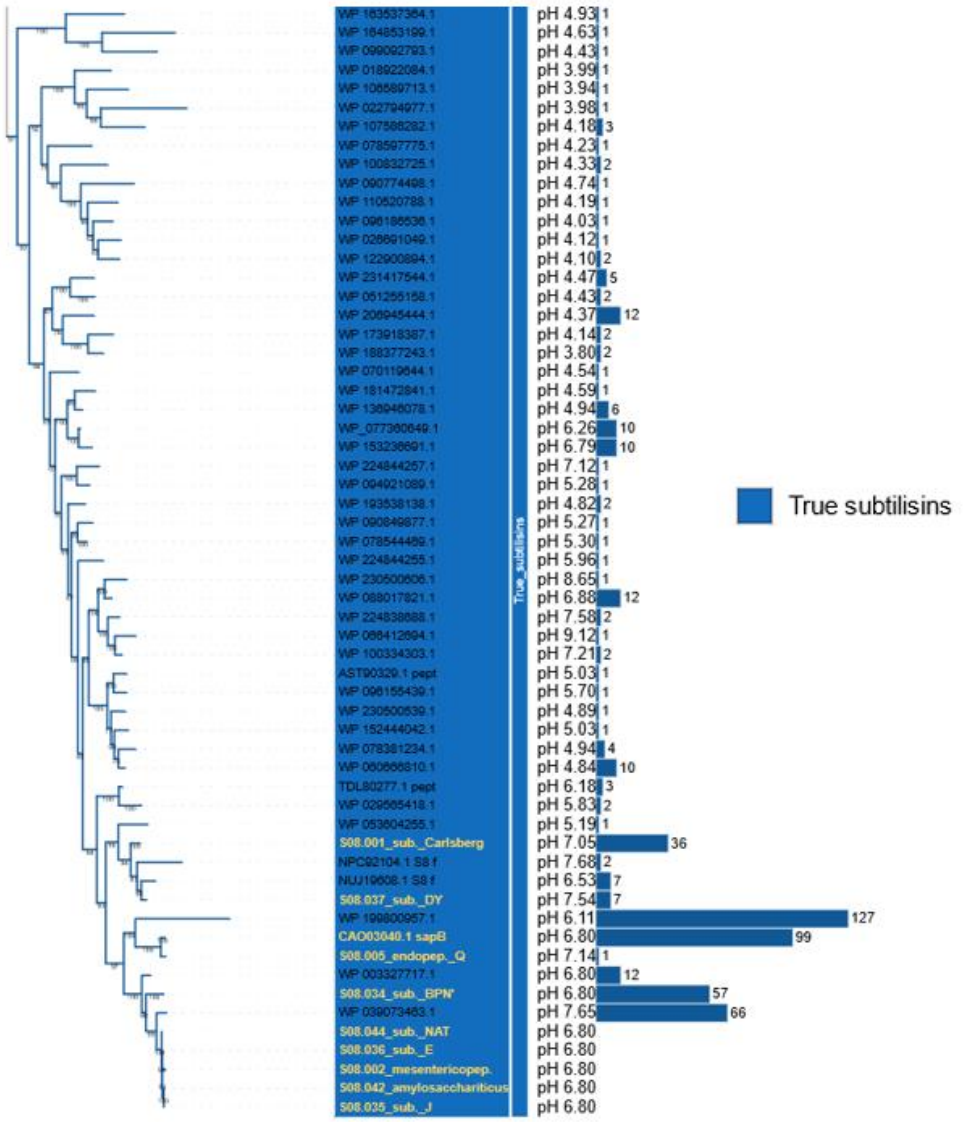
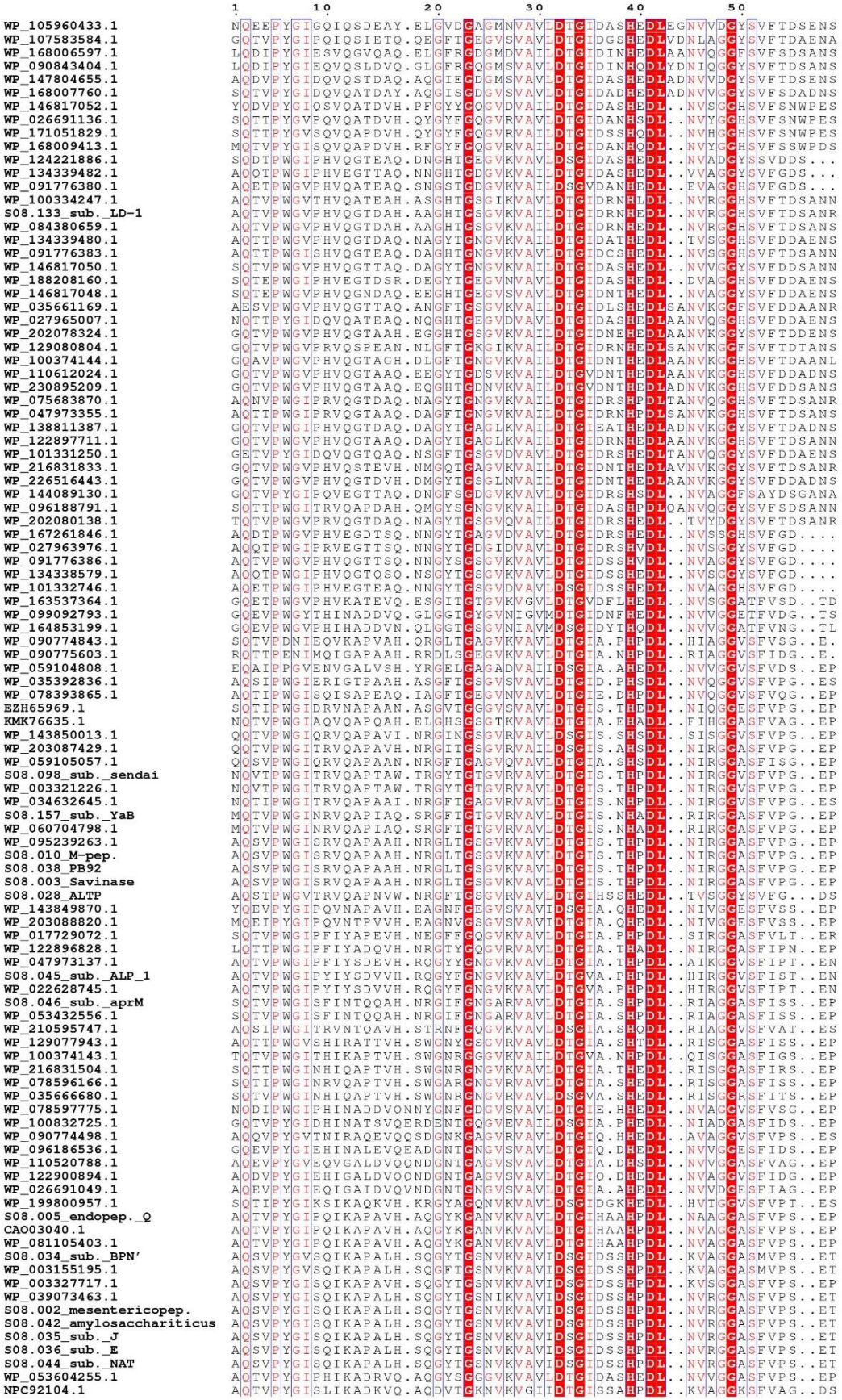
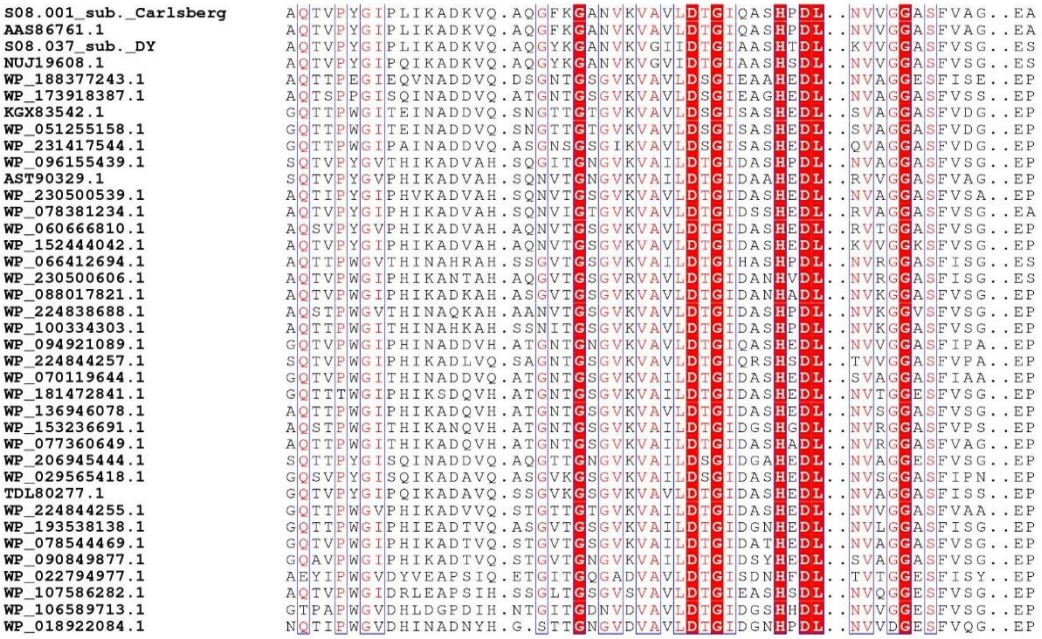
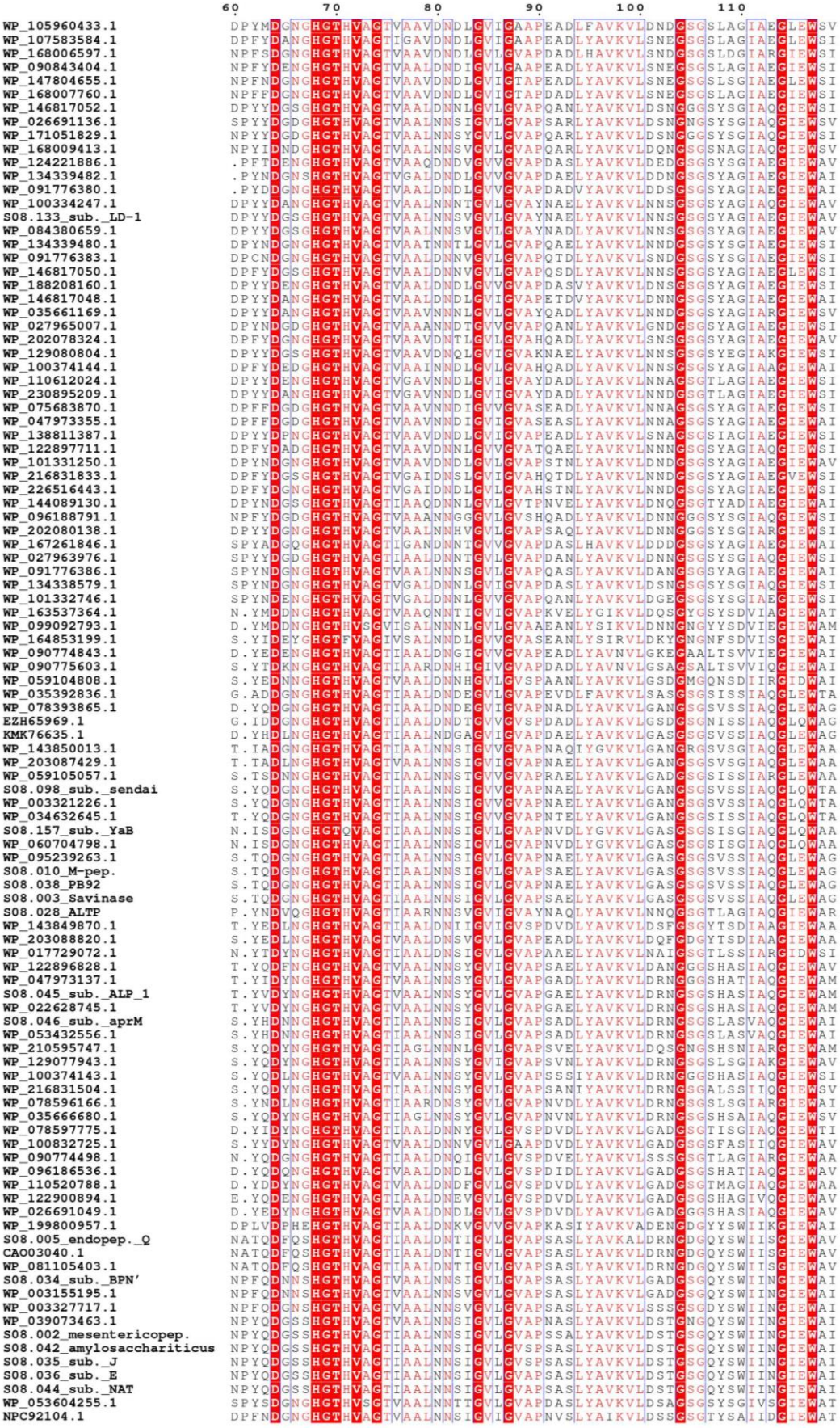
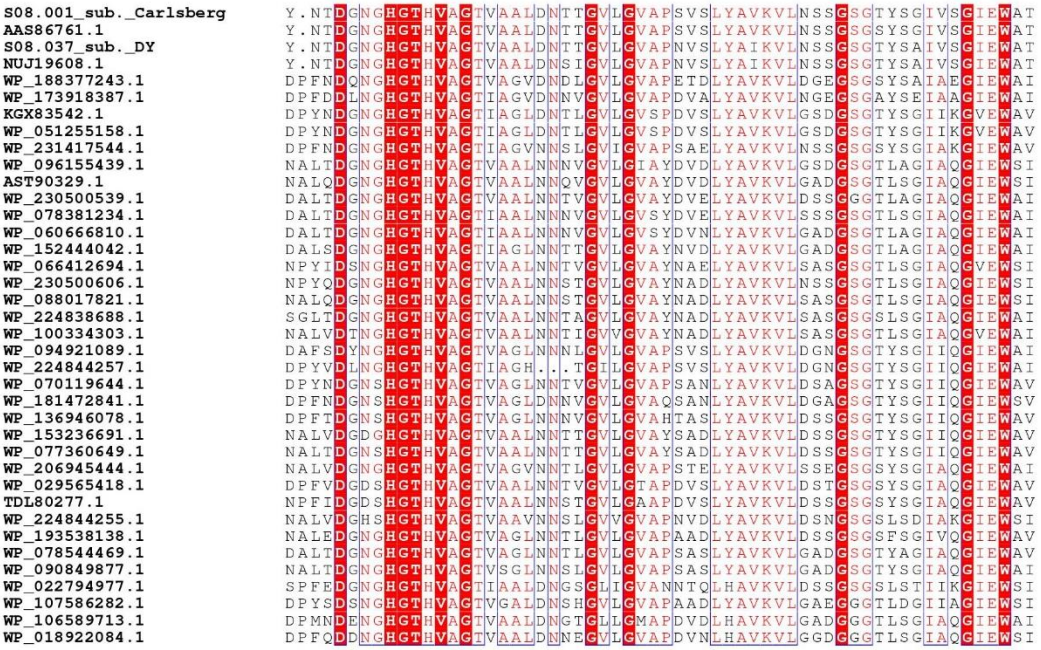


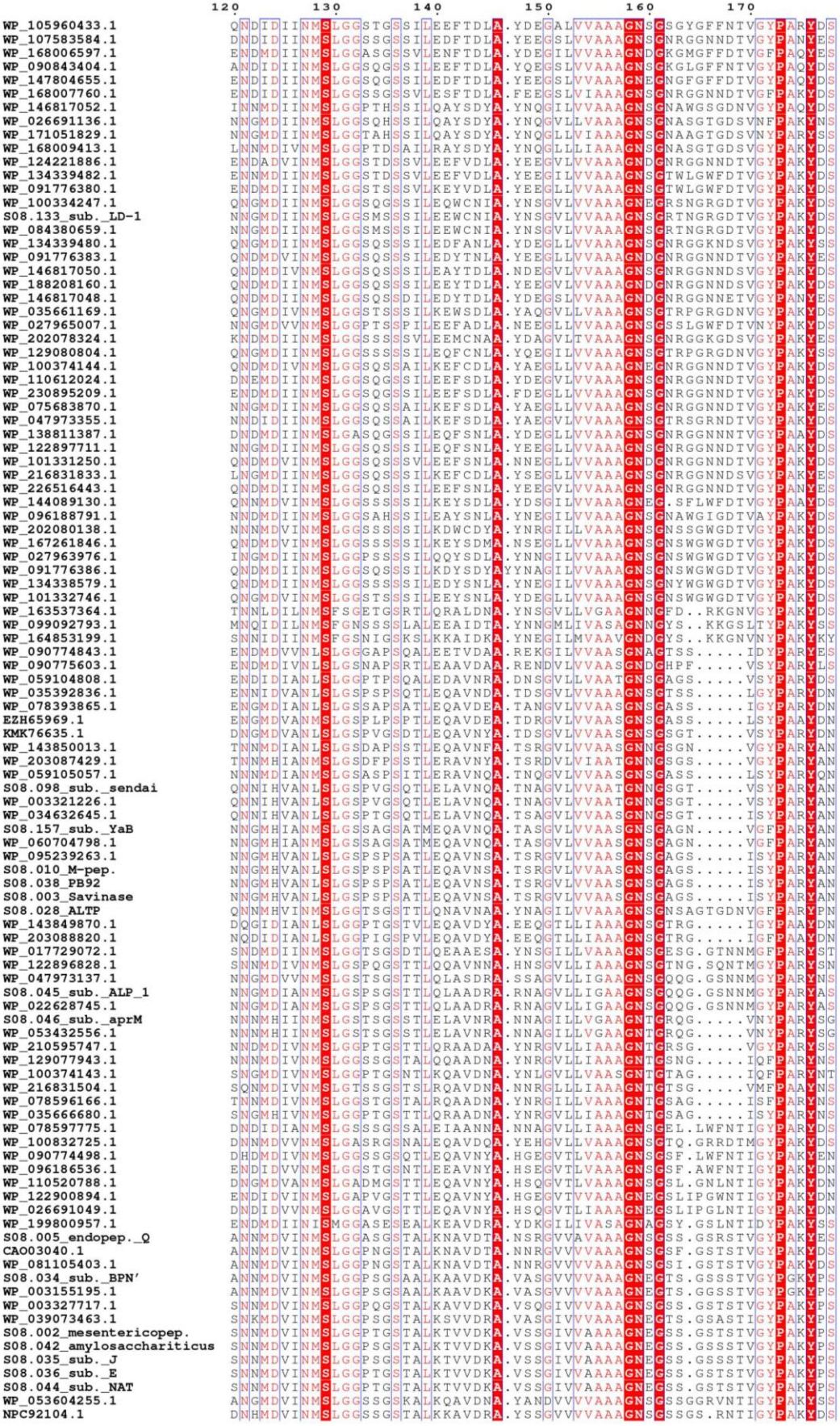
Figure S17 Tree of true subtilisins subgroup of subtilisin group.

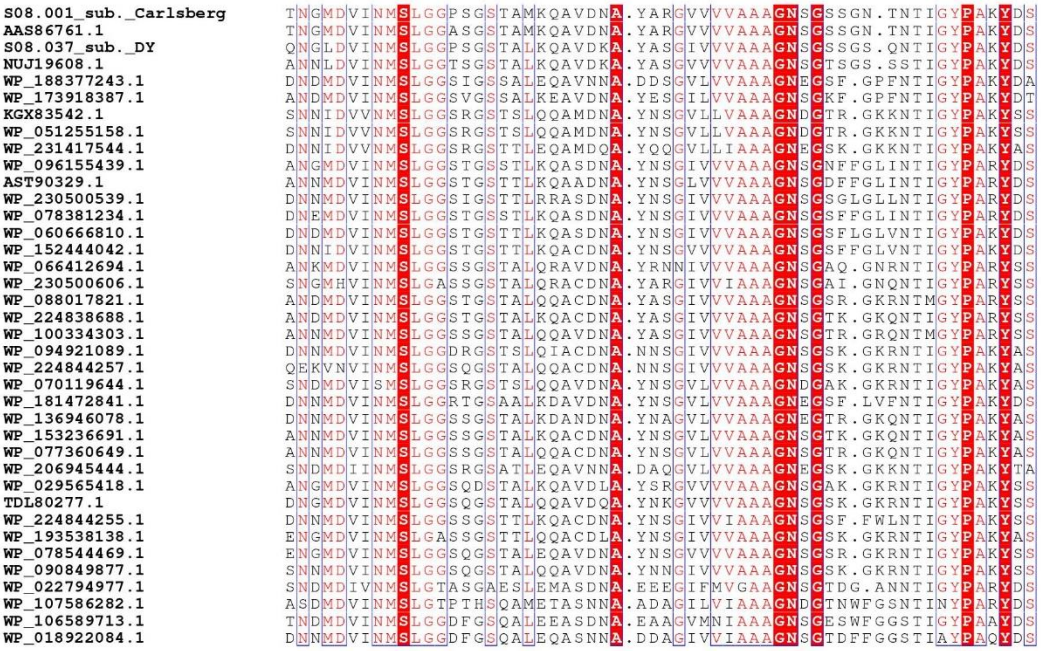

1 10 20 30 40 50
WP_105960433.1 NQEEPYGIGQIQSDEAY.ELGVDGAGMNVAVLDTGIDASHEDLEGNVVDGYSVFTDSENS
WP_107583584.1 GQTVPYGIPQIQSIETQ.QEGFTGEGVSVAVLDTGIDGSHEDLVDNLAGGYSVFTDAENA
WP_168006597.1 LQDIPYGIESVQGVQAQ.ELGFRGDGMDVAILDTGIDINHEDLAANVQGGFSVFSDSANS
WP_090843404.1 LQDIPYGIEQVQSLDVQ.GLGFRGQGMSVAVLDTGIDINHQDLYDNIQGGYSVFTDSANS
WP_147804655.1 AQDVPYGIDQVQSTDAQ.AQGIEGDGMSVAVLDTGIDASHEDLADNVVDGYSVFTDSENS
WP_168007760.1 SQTVPYGIDQVQATDAY.AQGISGDGVSVAVLDTGIDADHEDLADNVVGGYSVFSDADNS
WP_146817052.1 YQDVPYGIQSVQATDVH.PFGYYGQGVDVAILDTGIDASHEDL..NVSGGHSVFSNWPES
WP_026691136.1 SQTTPYGVPQVQATDVH.QYGYFGQGVRVAVLDTGIDANHSDL..NVYGGHSVFSNWPES
WP_171051829.1 SQTTPYGVSQVQAPDVH.QYGYFGQGVRVAILDTGIDSSHQDL..NVHGGHSVFSNWPES
WP_168009413.1 MQTVPYGISQVQAPDVH.RFGYFGQGVRVAILDTGIDANHQDL..NVYGGHSVFSSWPDS
WP_124221886.1 SQDTPWGIPHVQGTEAQ.DNGHTGEGVKVAVLDSGIDASHEDL..NVADGYSSVDDS...
WP_134339482.1 AQQTPWGIPHVEGTTAQ.NNGHTGDGVKVAILDTGIDASHEDL..VVAGGYSVFGDS...
WP_091776380.1 AQETPWGVPHVQATEAQ.SNGSTGDGVKVAILDSGVDANHEDL..EVAGGHSVFGDS...
WP_100334247.1 AQTVPWGVTHVQATEAH.ASGHTGSGIKVAVLDTGIDRNHLDL..NVRGGHSVFTDSANN
S08.133_sub._LD-1 AQTVPWGVPHVQGTDAH.AAGHTGSGVKVAILDTGIDRNHEDL..NVRGGHSVFTDSANR
WP_084380659.1 AQTVPWGVPHVQGTDAH.AAGHTGSGVKVAILDTGIDRNHEDL..NVRGGHSVFTDSANS
WP_134339480.1 AQTTPWGVPHVQGTDAQ.NAGYTGNGVKVAILDTGIDATHEDL..TVSGGHSVFDDAENS
WP_091776383.1 AQTTPWGISHVQGTEAQ.DAGHTGNGVKVAVLDTGIDCSHEDL..NVSGGHSVFTDSANN
WP_146817050.1 SQTVPWGVPHVQGTTAQ.DAGYTGNGVKVAILDTGIDASHEDL..NVVGGYSVFDDSANN
WP_188208160.1 SQTEPWGIPHVEGTDSR.DEGYTGEGVNVAVLDTGIDASHEDL..DVAGGHSVFDDAENS
WP_146817048.1 SQTEPWGVPHVQGNDAQ.EEGHTGEGVSVAVLDTGIDNTHEDL..NVAGGYSVFDDAENS
WP_035661169.1 AESVPWGVPHVQGTTAQ.ANGFTGSGVKVAILDTGIDLSHEDLSANVKGGFSVFDDAANR
WP_027965007.1 NQTTPYGIDQVQATEAQ.NQGHTGEGVDVAVLDTGIDASHEDLAANVQGGHSVFDDAENS
WP_202078324.1 GQTVPWGVPHVQGTAAH.EGGHTGSGVKVAILDTGIDNEHEDLAANVKGGYSVFTDDENS
WP_129080804.1 GQTVPWGVPRVQSPEAN.NLGFTGKGIKVAILDTGIDRNHEDLSANVKGGFSVFTDTANS
WP_100374144.1 GQAVPWGVPHVQGTAGH.DLGFTGNGVKVAILDTGIDNTHEDLAANVKGGHSVFTDAANL
WP_110612024.1 GQTVPWGVPHVQGTAAQ.EEGYTGDSVKVAILDTGVDNTHEDLAANVKGGHSVFDDADNS
WP_230895209.1 GQTVPWGVPHVQGTAAQ.EQGHTGDNVKVAILDTGVDNTHEDLADNVKGGHSVFDDSANS
WP_075683870.1 AQNVPWGIPRVQGTDAQ.NAGYTGNGVKVAILDTGIDRSHPDLTANVQGGHSVFTDSANR
WP_047973355.1 AQTTPWGIPRVQGTAAQ.DAGFTGNGVKVAILDTGIDRNHPDLSANVKGGHSVFTDSANS
WP_138811387.1 GQTVPWGIPHVQGTDAQ.DAGYTGAGLKVAILDTGIEATHEDLADNVKGGYSVFTDADNS
WP_122897711.1 GQTVPWGVPHVQGTAAQ.DAGYTGAGLKVAILDTGIDRNHEDLAANVKGGHSVFTDSANN
WP_101331250.1 GETVPYGIDQVVGTQAQ.NSGFTGSGVDVAVLDTGIDRSHEDLTANVQGGYSVFDDAENS
WP_216831833.1 GQTVPWGVPHVQSTEVH.NMGQTGAGVKVAILDTGIDNTHEDLAVNVKGGYSVFTDSANR
WP_226516443.1 GQTVPWGVPHVQGTDVH.DMGYTGSGLNVAILDTGIDNTHEDLAANVKGGYSVFDDADNS
WP_144089130.1 GQTVPYGIPQVEGTTAQ.DNGFSGDGVKVAVLDTGIDRSHSDL..NVAGGFSAYDSGANA
WP_096188791.1 SQTTPWGITRVQAPDAH.QMGYSGNGVKVAILDTGIDASHPDLQANVQGGYSVFSDSANN
WP_202080138.1 TQTVPWGVPRVQGTDAQ.NAGYTGSGVQVAILDTGIDRSHEDL..TVYDGYSVFTDSANR
WP_167261846.1 AQDTPWGVPHVEGDTSQ.NNGYTGAGVDVAVLDTGIDRSHEDL..NVSSGHSVFGD....
WP_027963976.1 AQQTPWGIPRVEGTTSQ.NNGYTGDGIDVAVLDTGIDRSHVDL..NVSGGYSVFGD....
WP_091776386.1 AQTVPWGVPHVQGTTSQ.NNGYSGSGVKVAVLDTGIDSSHVDL..NVAGGYSVFGD....
WP_134338579.1 AQQTPWGIPHVQGTQSQ.NSGYTGSGVKVAVLDTGIDSSHEDL..NVSGGYSVFGD....
WP_101332746.1 AQETPWGIPHVEGTEAQ.NNGYTGSGVDVAVLDSGIDSSHEDL..NVAGGHSVFGD....
WP_163537364.1 GQETPWGVPHVKATEVQ.ESGITGTGVKVGVLDTGVDFLHEDL..NVSGGATFVSD..TD
WP_099092793.1 GQEVPWGYTHINADDVQ.GLGGTGYGVNIGVMDTGIDNFHEDL..NVVGGETFVDG..TS
WP_164853199.1 GQEVPWGVPHIHADDVN.QLGGTGSGVNIAVMDSGIDYTHQDL..NVVGGATFVNG..TL
WP_090774843.1 SQTVPDNIEQVKAPVAH.QRGLTGAGVKVAVLDTGIA.PHPDL..HIAGGVSFVSG..E.
WP_090775603.1 RQTTPENIMQIGAPAAH.RRDLSGEGVKVAVLDTGIA.NHPDL..RIAGGVSFIDG..E.
WP_059104808.1 EQAIPPGVENVGALVSH.YRGELGAGADVAIIDSGIA.AHEDL..NVVGGVSFVDS..EP
WP_035392836.1 AQSIPWGIERIGTPAAH.ASGFTGSGVSVAVLDTGID.PHSDL..NVQGGVSFVPG..ES
WP_078393865.1 AQTIPWGISQISAPEAQ.IAGFTGEGVNVAVLDTGIE.DHPDL..NVQGGVSFVQG..EP
EZH65969.1 SQTIPWGIDRVNAPAAN.ASGVTGGGVSVAVLDTGIS.THEDL..NIQGGESFVPG..EP
KMK76635.1 NQTVPWGIAQVQAPQAH.ELGHSGSGTKVAVLDTGIA.EHADL..FIHGGASFVAG..EP
WP_143850013.1 QQTVPWGIQRVQAPAVI.NRGINGSGVRVAVLDSGIS.SHSDL..SISGGVSFVPG..EP
WP_203087429.1 QQTVPWGITRVQAPAVH.NRGITGSGVRVAILDSGIS.AHSDL..NIRGGASFVPG..EP
WP_059105057.1 QQSVPWGIQRVQAPAAN.NRGFTGAGVQVAVLDTGIA.SHSDL..NIRGGASFISG..EP
S08.098_sub._sendai NQVTPWGITRVQAPTAW.TRGYTGTGVRVAVLDTGIS.THPDL..NIRGGVSFVPG..EP
WP_003321226.1 NQVTPWGITRVQAPTAW.TRGYTGTGVRVAVLDTGIS.THPDL..NIRGGVSFVPG..EP
WP_034632645.1 NQTIPWGITRVQAPAAI.NRGFTGAGVRVAVLDTGIS.NHPDL..NIRGGVSFVPG..ES
S08.157_sub._YaB MQTVPWGINRVQAPIAQ.SRGFTGTGVRVAVLDTGIS.NHADL..RIRGGASFVPG..EP
WP_060704798.1 MQTVPWGINRVQAPIAQ.SRGFTGTGVRVAVLDTGIS.NHADL..RIRGGASFVPG..EP
WP_095239263.1 AQSVPWGISRVQAPAAH.NRGLTGSGVKVAVLDTGIS.THPDL..NIRGGASFVPG..EP
S08.010_M-pep. AQSVPWGISRVQAPAAH.NRGLTGSGVKVAVLDTGIS.THPDL..NIRGGASFVPG..EP
S08.038_PB92 AQSVPWGISRVQAPAAH.NRGLTGSGVKVAVLDTGIS.THPDL..NIRGGASFVPG..EP
S08.003_Savinase AQSVPWGISRVQAPAAH.NRGLTGSGVKVAVLDTGIS.THPDL..NIRGGASFVPG..EP
S08.028_ALTP AQSTPWGVTRVQAPNVW.NRGFTGSGVRVAVLDTGIHSSHEDL..TVSGGYSVFG...DS
WP_143849870.1 YQEVPYGIPQVNAPAVH.EAGNFGEGVSVAVIDSGIA.QHEDL..NIVGGESFVSS..EP
WP_203088820.1 MQEIPYGIPQVNTPVVH.EAGNVGSGVSVAVIDTGIA.QHEDL..NIVGGESFVSS..EP
WP_017729072.1 SQTVPWGIPFIYAPEVH.NEGFFGQGVKVAVLDTGIA.PHPDL..SIRGGASFVLT..ER
WP_122896828.1 LQTTPWGIPFIYADQVH.NRGTYGQGVRVAVLDTGIA.THADL..NIRGGASFIPN..EP
WP_047973137.1 AQTVPWGIPFIYSDEVH.RQGYYGNGVKVAVLDTGIA.THPDL..AIKGGVSFIPT..EN
S08.045_sub._ALP_1 AQTVPWGIPYIYSDVVH.RQGYFGNGVKVAVLDTGVA.PHPDL..HIRGGVSFIST..EN
WP_022628745.1 AQTVPWGIPYIYSDVVH.RQGYFGNGVKVAVLDTGVA.PHPDL..HIRGGVSFIPT..EN
S08.046_sub._aprM SQTVPWGISFINTQQAH.NRGIFGNGARVAVLDTGIA.SHPDL..RIAGGASFISS..EP
WP_053432556.1 SQTVPWGISFINTQQAH.NRGIFGNGARVAVLDTGIA.SHPDL..RIAGGASFISS..EP
WP_210595747.1 AQSIPWGITRVNTQAVH.STRNFGQGVKVAVLDSGIA.SHQDL..RIAGGVSFVAT..ES
WP_129077943.1 AQTTPWGVSHIRATTVH.SWGNYGSGVRVAVLDTGIA.SHTDL..RISGGASFISS..EP
WP_100374143.1 TQTVPWGITHIKAPTVH.SWGNRGGGVKVAILDTGVA.NHPDL..QISGGASFIGS..EP
WP_216831504.1 SQTTPWGINHIQAPTVH.SWGNRGNGVKVAVLDTGIA.THEDL..RISGGASFIRS..EP
WP_078596166.1 SQTIPWGINRVQAPTVH.SWGARGNGVRVAVLDTGIA.SHEDL..RISGGASFISS..EP
WP_035666680.1 SQTVPWGINHIQAPTVH.SWGNRGNGVRVAVLDSGVA.SHEDL..RISGGRSFITS..EP
WP_078597775.1 NQDIPWGIPHINADDVQNNYGNFGDGVSVAVLDTGIE.HHEDL..NVAGGVSFVSG..EP
WP_100832725.1 GQTVPYGIDHINATSVQERDENTGQGVEVAILDSGIA.PHEDL..NIADGASFIDS..EP
WP_090774498.1 AQQVPYGVTNIRAQEVQQSDGNKGAGVRVAVLDTGIQ.HHEDL..AVAGGVSFVPS..ES
WP_096186536.1 GQEVPYGIEHINALEVQEADGNTGNGVSVAVLDTGIQ.DHEDL..NVVGGESFIDS..EP
WP_110520788.1 AQEVPYGVEQVGALDVQQNDGNTGAGVSVAVLDTGIA.DHSDL..NVVGGASFVAG..EP
WP_122900894.1 AQDVPYGIEHIGALDVQQNDGNTGAGVSVAVLDTGIQ.DHEDL..NVAGGVSFIDG..EP
WP_026691049.1 AQEVPYGIEQIGAIDVQVNDGNTGNGVSVAVLDTGIA.AHEDL..NVVDGESFIAG..EP
WP_199800957.1 SQTIPYGIKSIKAQKVH.KRGYAGQNVKVAVLDSGIDGKHEDL..HVTGGVSFVPT..ES
S08.005_endopep._Q AQTVPYGIPQIKAPAVH.AQGYKGANVKVAVLDTGIHAAHPDL..NAAGGASFVPS..EP
CAO03040.1 AQTVPYGIPQIKAPAVH.AQGYKGANVKVAVLDTGIHAAHPDL..NVAGGASFVPS..EP
WP_081105403.1 AQTVPYGIPQIKAPAVH.AQGYKGANVKVAVLDTGIHAAHPDL..NVAGGASFVPS..EP
S08.034_sub._BPN' AQSVPYGVSQIKAPALH.SQGYTGSNVKVAVIDSGIDSSHPDL..KVAGGASMVPS..ET
WP_003155195.1 AQSVPYGVSQIKAPALH.SQGFTGSNVKVAVIDSGIDSSHPDL..KVAGGASMVPS..ET
WP_003327717.1 AQSVPYGISQIKAPAVH.SQGYTGSNVKVAVIDSGIDSSHPDL..KVSGGASFVPS..EP
WP_039073463.1 AQSVPYGISQIKAPALH.SQGYTGSNIKVAVIDSGIDSSHPDL..NVRGGASFVPS..ET
S08.002_mesentericopep. AQSVPYGISQIKAPALH.SQGYTGSNVKVAVIDSGIDSSHPDL..NVRGGASFVPS..ET
S08.042_amylosachariticus AQSVPYGISQIKAPALH.SQGYTGSNVKVAVIDSGIDSSHPDL..NVRGGASFVPS..ET
S08.035_sub._J AQSVPYGISQIKAPALH.SQGYTGSNVKVAVIDSGIDSSHPDL..NVRGGASFVPS..ET
S08.036_sub._E AQSVPYGISQIKAPALH.SQGYTGSNVKVAVIDSGIDSSHPDL..NVRGGASFVPS..ET
S08.044_sub._NAT AQSVPYGISQIKAPALH.SQGYTGSNVKVAVIDSGIDSSHPDL..NVRGGASFVPS..ET
WP_053604255.1 AQTVPYGIPHIKADRVQ.AQGYSGGGVKVAVLDTGIDASHEDL..NVVGGASFVPS..EP
NPC92104.1 AQTVPYGISLIKADKVQ.AQDVTGKNVKVGIIDTGISSAHPDL..KVAGGKSFVAG..DS

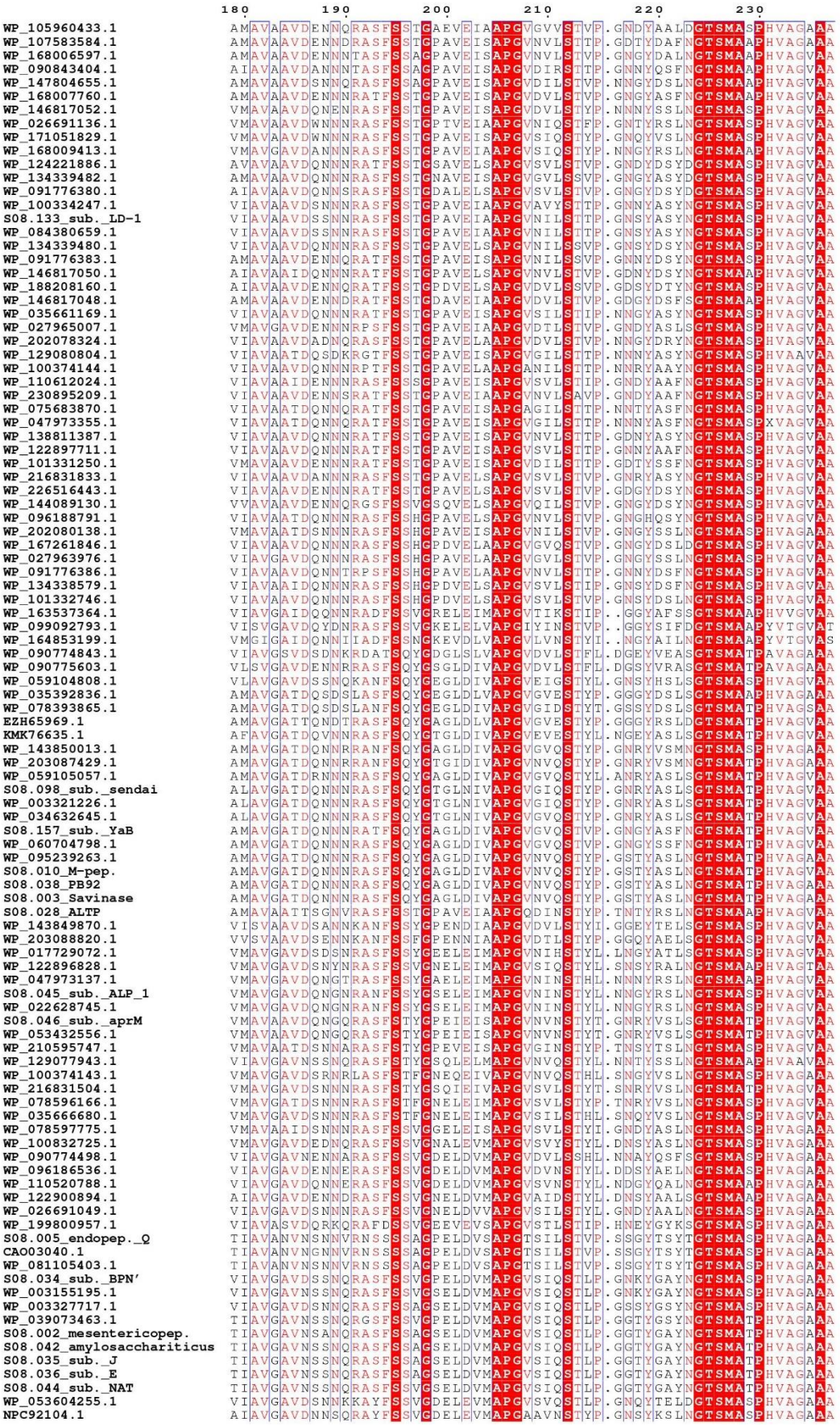

S08.001_sub._Carlsberg AQTVPYGIPLIKADKVQ.AQGFKGANVKVAVLDTGIQASHPDL..NVVGGASFVAG..EA
AAS86761.1 AQTVPYGIPLIKADKVQ.AQGFKGANVKVAVLDTGIQASHPDL..NVVGGASFVAG..EA
S08.037_sub._DY AQTVPYGIPLIKADKVQ.AQGYKGANVKVGIIDTGIAASHTDL..KVVGGASFVSG..ES
NUJ19608.1 AQTVPYGIPQIKADKVQ.AQGYKGANVKVGVIDTGIAASHSDL..NVVGGASFVSG..ES
WP_188377243.1 AQTTPEGIEQVNADDVQ.DSGNTGSGVKVAVLDSGIEAAHEDL..NVAGGESFISE..EP
WP_173918387.1 AQTSPPGISQINADDVQ.ATGNTGSGVKVAVLDSGIEAGHEDL..NVAGGASFVSS..EP
KGX83542.1 GQTTPWGITEINADDVQ.SNGTTGTGVKVAVLDSGISASHEDL..SVAGGASFVDG..EP
WP_051255158.1 GQTTPWGITEINADDVQ.SNGTTGTGVKVAVLDSGISASHEDL..SVAGGASFVDG..EP
WP_231417544.1 GQTTPWGIPAINADDVQ.ASGNSGSGIKVAVLDSGISASHEDL..QVAGGASFVDG..EP
WP_096155439.1 SQTVPYGVTHIKADVAH.SQGITGNGVKVAILDTGIDASHPDL..NVAGGASFVSG..EP
AST90329.1 SQTVPYGVPHIKADVAH.SQNVTGNGVKVAILDTGIDAAHEDL..RVVGGASFVAG..EP
WP_230500539.1 AQTIPYGIPHVKADVAH.SQNVTGSGVKVAVLDTGIDASHEDL..NVAGGASFVSA..EP
WP_078381234.1 AQTVPYGIPHIKADVAH.SQNVIGTGVKVAVLDTGIDSSHEDL..RVAGGASFVSG..EA
WP_060666810.1 AQSVPYGVPHIKADVAH.AQNVTGSGVKVAVLDTGIDASHEDL..RVTGGASFVSG..EP
WP_152444042.1 AQTVPYGIPHIKADVAH.AQNVTGSGVKVAVLDTGIDASHEDL..KVVGGKSFVSG..EP
WP_066412694.1 AQTTPWGVTHINAHRAH.SSGVTGSGVKVAILDTGIHASHPDL..NVRGGASFISG..ES
WP_230500606.1 AQTVPWGIPHIKANTAH.AQGVTGSGVRVAVLDTGIDANHVDL..NVRGGASFISG..ES
WP_088017821.1 AQTVPWGIPHIKADKAH.ASGVTGSGVKVAVLDTGIDANHADL..NVKGGASFVSG..EP
WP_224838688.1 AQSTPWGVTHINAQKAH.AANVTGSGVKVAVLDTGIDASHPDL..NVKGGVSFVSG..EP
WP_100334303.1 AQTTPWGITHINAHKAH.SSNITGSGVKVAVLDTGIDASHPDL..NVKGGASFVSG..EP
WP_094921089.1 GQTVPWGIPHINADDVH.ATGNTGNGVKVAVLDTGIQASHEDL..NVVGGASFIPA..EP
WP_224844257.1 SQTVPWGIPHIKADLVQ.SAGNTGSGVKVAVLDTGIQRSHSDL..TVVGGASFVPA..EP
WP_070119644.1 GQTVPWGITHINADDVQ.ATGNTGSGVKVAILDTGIDASHEDL..SVAGGASFIAA..EP
WP_181472841.1 GQTTTWGIPHIKSDQVH.ATGNTGSGVKVAILDTGIDASHEDL..NVTGGESFVSG..EP
WP_136946078.1 AQTTPWGIPHIKADQVH.ATGNTGSGVKVAVLDTGIDASHEDL..NVSGGASFVSG..EP
WP_153236691.1 AQSTPWGITHIKANQVH.ATGNTGSGVKVAILDTGIDGSHGDL..NVRGGASFVPS..EP
WP_077360649.1 AQTTPWGITHIKADQVH.ATGNTGSGVKVAILDTGIDASHADL..NVRGGASFVAG..EP
WP_206945444.1 SQTTPYGISQINADDVQ.AQGTTGNGVKVAILDSGIDGAHEDL..NVAGGESFVSG..EP
WP_029565418.1 GQSVPYGISQIKADAVQ.ASGVKGSGVKVAILDSGIDASHEDL..NVSGGASFIPN..EP
TDL80277.1 GQTVPYGIPQIKADAVQ.SSGVKGSGVKVAVLDTGIDASHEDL..NVAGGASFISS..EP
WP_224844255.1 GQTTPWGVPHIKADVVQ.STGTTGTGVKVAILDTGIDASHEDL..NVVGGASFVAA..EP
WP_193538138.1 GQTTPWGIPHIEADTVQ.ASGVTGSGVKVAILDTGIDGNHEDL..NVLGGASFISG..EP
WP_078544469.1 GQTVPWGIPHIKADTVQ.STGVTGSGVKVAILDTGIDATHEDL..NVAGGASFVSG..EP
WP_090849877.1 GQAVPWGIPHIKADTVQ.STGVTGNGVKVAILDTGIDSYHEDL..SVAGGASFVSG..EP
WP_022794977.1 AEYIPWGVDYVEAPSIQ.ETGITGQGADVAVLDTGISDNHFDL..TVTGGESFISY..EP
WP_107586282.1 AQTVPWGIDRLEAPSIH.SSGLTGSGVSVAVLDTGIEASHSDL..NVQGGESFVSG..EP
WP_106589713.1 GTPAPWGVDHLDGPDIH.NTGITGDNVDVAVLDTGIDGSHHDL..NVVGGESFVSG..EP
WP_018922084.1 NQTIPWGVDHINADNYH.G.STTGNGVDVAVLDTGIDGNHPDL..NVVDGESFVQG..EP

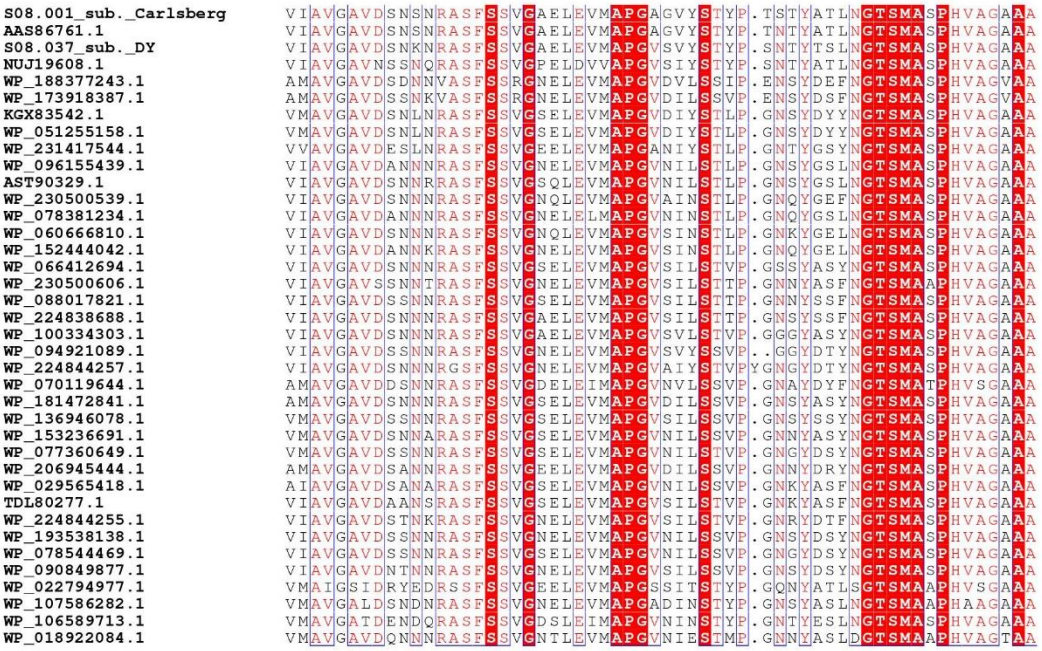

60
70
80
90
100
110
WP_105960433.1 DPYMDGNGHGTHVAGTVAAVDNDLGVIGAAPEADLFAVKVLDNDGSGSLAGIAEGLEWSV
WP_107583584.1 DPFYDANGHGTHVAGTIGAVDNDLGVIGAAPEADLYAVKVLSNEGSGSLAGIAEGLEWSI
WP_168006597.1 NPFSDGNGHGTHVAGTVAAVDNDLGVLGVAPDADLHAVKVLSNDGSGSLDGIARGIEWSI
WP_090843404.1 NPFYDENGHGTHVAGTVAALDNDIGVLGAAPEADLYAVKVLSNEGSGSLAGIAEGIEWSI
WP_147804655.1 NPFNDGNGHGTHVAGTVAAVDNDLGVIGTAPEADLYAVKVLSNEGSGSLSGIAEGLEWSI
WP_168007760.1 NPFFDGNGHGTHVAGTVAAVDNDLGVIGTAPDADLYAVKVLSNEGSGSLAGIAEGIEWSI
WP_146817052.1 DPYYDGSGHGTHVAGTVAALDNNLGVLGVAPQANLYAVKVLDSNGGGSYSGIAQGIEWSI
WP_026691136.1 SPYYDGDGHGTHVAGTVAALNNSIGVLGVAPSARLYAVKVLDSNGNGSYSGIAQGIEWSV
WP_171051829.1 NPYYDGDGHGTHVAGTVAALNNSYGVLGVAPQARLYAVKVLDSNGGGSYSGIAQGIEWSI
WP_168009413.1 NPYIDNDGHGTHVAGTVAALNNSVGVLGVAPQARLYAVKVLDQNGSGSNAGIAQGIEWSV
WP_124221886.1 .PFTDENGHGTHVAGTVAAQDNDVGVVGVAPDASLYAVKVLDEDGSGSYSGIAEGIEWAV
WP_134339482.1 .PYNDGNSHGTHVAGTVGALDNDLGVVGVAPDAELYAVKVLDDNGSGSYAGIAEGIEWAI
WP_091776380.1 .PYDDGNGHGTHVAGTVAALDNDLGVVGVAPDADVYAVKVLDDSGSGSYAGIAEGIEWAI
WP_100334247.1 DPYYDANGHGTHVAGTVAALNNNTGVLGVAYNAELYAVKVLNNSGSGSYAGIAQGIEWAV
S08.133_sub._LD-1 DPYYDGSGHGTHVAGTVAALNNSVGVLGVAYNAELYAVKVLNNSGSGSYAGIAEGIEWAV
WP_084380659.1 DPYYDGSGHGTHVAGTVAALNNSVGVLGVAYNADLYAVKVLNNSGSGSYAGIAEGIEWAV
WP_134339480.1 DPYNDGNGHGTHVAGTVAATNNTLGVLGVAPQAELYAVKVLNNDGSGSYSGIAEGIEWSI
WP_091776383.1 DPCNDGNGHGTHVAGTVAALDNNVGVLGVAPQTDLYAVKVLSNDGSGSYSGIAEGIEWSI
WP_146817050.1 DPFYDGSGHGTHVAGTVAALDNNVGVLGVAPQSDLYAVKVLNNSGSGSYAGIAEGLEWSI
WP_188208160.1 DPYYDENGHGTHVAGTVAALDNDLGVVGVAPDASVYAVKVLDNSGSGSYAGIAEGIEWSI
WP_146817048.1 DPYYDANGHGTHVAGTVAALDNDLGVIGVAPETDVYAVKVLDNNGSGSYAGIAEGIEWAI
WP_035661169.1 DPYYDANGHGTHVAGTVAAVNNNLGVLGVAYQADLYAVKVLNNDGSGSYAGIARGIEWSV
WP_027965007.1 DPYNDGDGHGTHVAGTVAAANNDTGVVGVAPQANLYAVKVLGNDGSGSYAGIAEGIEWSI
WP_202078324.1 DPFYDGNGHGTHVAGTVAAVDNTLGVLGVAHQADLYAVKVLSNSGSGSYAGIAEGIEWAV
WP_129080804.1 DPYYDGSGHGTHVAGTVAAVDNQLGVIGVAKNAELYAVKVLNNSGSGSYEGIAKGIEWSI
WP_100374144.1 DPFYDEDGHGTHVAGTVAAINNDLGVLGVAHQADLYAVKVLNNSGSGSYAGIAAGIEWAI
WP_110612024.1 DPYYDENGHGTHVAGTVGAVNNDLGVIGVAYDADLYAVKVLNNAGSGTLAGIAEGIEWSI
WP_230895209.1 DPYYDANGHGTHVAGTVGAVNNDLGVIGVAYDADLYAVKVLNNDGSGSTLAGIAEGIEWSI
WP_075683870.1 DPFFDGDGHGTHVAGTVAAVNNDIGVVGVASEADLYAVKVLNNAGSGSYAGIAEGIEWSI
WP_047973355.1 DPFFDGDGHGTHVAGTVAAVNNDIGVIGVASEASLYAVKVLNNAGSGSYAGIAEGIEWAI
WP_138811387.1 DPYYDPNGHGTHVAGTVAAVDNDLGVIGVAPEADLYAVKVLSNAGSGSIAGIAEGIEWSI
WP_122897711.1 DPFYDADGHGTHVAGTVAAVDNNLGVVGVATQAELYAVKVLNNNGSGSYAGIAQGIEWSI
WP_101331250.1 DPYNDGNGHGTHVAGTVAAVDNDLGVLGVAPSTNLYAVKVLDNDGSGSYAGIAEGIEWAV
WP_216831833.1 DPFYDGSGHGTHVAGTVGAIDNSLGVIGVAHQTDLYAVKVLNNDGSGSYAGIAEGVEWSI
WP_226516443.1 DPFYDGNGHGTHVAGTVGAIDNDLGVLGVAHSTNLYAVKVLNNDGSGSYAGIAEGIEWSI
WP_144089130.1 DPYNDGSGHGTHVAGTIAAQDNNLGVLGVTPNVELYAVKVLDNQGSGSYTYADIAEGIEWAI
WP_096188791.1 NPFYDGDGHGTHVAGTVAAANNGGGVLGVSHQADLYAVKVLDNNGGGSYSGIAQGIEWSI
WP_202080138.1 DPYYDGNGHGTHVAGTVAALNNHVGVLGVAPSAQLYAVKVLDNNGGGSYSGIARGIEWSI
WP_167261846.1 SPYADGQGHGTHVAGTIGANDNNTGVVGVAPDASLHAVKVLDDDGSGSYAGIAEGIEWAI
WP_027963976.1 SPYYDGDGHGTHVAGTIAALDNNTGVLGVAPDANLYAVKVLDNNGSGSYSGIAQGIEWSI
WP_091776386.1 SPYNDGNGHGTHVAGTVAALNNNSGVLGVAPQASLYAVKVLDANGSGSYAGIAEGIEWAI
WP_134338579.1 SPYNDGNGHGTHVAGTVGALDNNLGVIGVAPDASLYAVKVLDSNGSGSYSGIAQGIEWSI
WP_101332746.1 SPYNDENGHGTHVAGTVGALDNNLGVVGVAPQANLYAVKVLDGEGSGSYSGIAEGIEWSI
WP_163537364.1 N.YMDDNGHGTHVAGTVAAQNNTIGVIGVAPKVELYGIKVLDQSGYGSYSDVIAGIEWAT
WP_099092793.1 D.YMDDNGHGTHVSGVISALNNNLGVLGVAAEANLYSIKVLDNNGNGYYSDVIEGIEWAM
WP_164853199.1 S.YIDEYGHGTFVAGIVSALNNDLGVVGVASEANLYSIRVLDKYGNGNFSDVISGIEWAI
WP_090774843.1 D.YEDENGHGTHVAGTIAALDNGIGVVGVAPEADLYAVNVLGKEGAALTSVVIEGIEWAI
WP_090775603.1 S.YTDKNGHGTHVAGTIAARDNHIGIVGVAPDADLYAVNVLGSAGSALTSVVIQGIEWAI
WP_059104808.1 S.YEDNNGHGTHVAGTVAALDNNHGVLGVSPAANLYAVKVLGSDGMGQNSDIIRGIDWAI
WP_035392836.1 G.ADDGNGHGTHVAGTIAALDNDEGVLGVAPEVDLFAVKVLSASGSGSISSIAQGLEWTA
WP_078393865.1 D.YQDGNGHGTHVAGTIAALDNDEGVIGVAPNADLYAVKVLGANGSGSVSSIAQGLEWAG
EZH65969.1 G.IDDGNGHGTHVAGTIAALDNDTGVVGVSPDADLYAVKVLGSDGSGNISSIAQGLQWAG
KMK76635.1 D.YHDLNGHGTHVAGTIAALNDGAGVIGVAPDAELYAVKVLGASGSGSVSSIAQGLEWAG
WP_143850013.1 T.IADGNGHGTHVAGTIAALNNSIGVVGVAPNAQIYGVKVLGANGRGSVSGIAQGLEWAA
WP_203087429.1 T.TADLNGHGTHVAGTVAALNNSIGVIGVAPNAELYAVKVLGANGSGSVSGIAQGLEWAA
WP_059105057.1 S.TSDNNGHGTHVAGTIAALNNSTGVVGVAPRAELYAVKVLGADGSGSISSIARGLEWAA
S08.098_sub._sendai S.YQDGNGHGTHVAGTIAALNNSIGVVGVAPNAELYAVKVLGANGSGSVSSIAQGLQWTA
WP_003321226.1 S.YQDGNGHGTHVAGTIAALNNSIGVVGVAPNAELYAVKVLGANGSGSVSSIAQGLQWTA
WP_034632645.1 T.YQDGNGHGTHVAGTIAALNNSIGVVGVAPNTELYAVKVLGANGSGSISSIAQGLQWTA
S08.157_sub._YaB N.ISDGNGHGTQVAGTIAALNNSIGVLGVAPNVDLYGVKVLGASGSGSISGIAQGLQWAA
WP_060704798.1 N.ISDGNGHGTHVAGTIAALNNSIGVLGVAPNVDLYGVKVLGASGSGSISGIAQGLQWAA
WP_095239263.1 S.TQDGNGHGTHVAGTIAALNNSIGVLGVAPNAELYAVKVLGASGSGSVSSIAQGLEWAG
S08.010_M-pep. S.TQDGNGHGTHVAGTIAALNNSIGVLGVAPSAELYAVKVLGASGSGSVSSIAQGLEWAG
S08.038_PB92 S.TQDGNGHGTHVAGTIAALNNSIGVLGVAPNAELYAVKVLGASGSGSVSSIAQGLEWAG
S08.003_Savinase S.TQDGNGHGTHVAGTIAALNNSIGVLGVAPSAELYAVKVLGASGSGSVSSIAQGLEWAG
S08.028_ALTP P.YNDVQGHGTHVAGTIAARNNSVGVIGVAYNAQLYAVKVLNNQGSGTLAGIAQGIEWAR
WP_143849870.1 T.YEDLNGHGTHVAGTIAALDNIIGVIGVSPDVDLYAVKVLDSFGSGYTSDIAAGIEWAA
WP_203088820.1 S.YEDLNGHGTHVAGTVAALDNSVGVLGVAPEADLYAVKVLDQFGDGYTSDIAAGIEWAA
WP_017729072.1 N.YTDYNGHGTHVAGTIAALDNSIGVLGVAPAAELYAVKVLNAIGSGTLSSIARGIDWSI
WP_122896828.1 T.YQDFNGHGTHVAGTVAALNNSYGVIGVAPSAELYAVKVLDANGGGSHASIAQGIEWAV
WP_047973137.1 T.YIDYNGHGTHVAGTVAALNNSYGVLGVAPGADLYAVKVLDRNGGGSHATIAQGIDWAM
S08.045_sub._ALP_1 T.YVDYNGHGTHVAGTVAALNNSYGVLGVAPGAELYAVKVLDRNGGGSHASIAQGIEWAM
WP_022628745.1 T.YVDYNGHGTHVAGTVAALNNSYGVLGVAPGAELYAVKVLDRNGGGSHASIAQGIEWAM
S08.046_sub._aprM S.YHDNNGHGTHVAGTIAALNNSIGVLGVAPSADLYAVKVLDRNGGGSLASVAQGIEWAI
WP_053432556.1 S.YHDNNGHGTHVAGTIAALNNSIGVLGVAPSADLYAVKVLDRNGGGSLASVAQGIEWAI
WP_210595747.1 S.YQDYNGHGTHVAGTIAGLNNNLGVLGVAPSVELYAVKVLDQSGNGSHSNIARGIDWAM
WP_129077943.1 S.YQDYNGHGTHVAGTIAALNNSYGVIGVAPSVNLYAVKVLDRNGGGSLSGIAKGIEWAV
WP_100374143.1 S.YQDLNGHGTHVAGTVAALNNSYGVLGVAPSSSIYAVKVLDRNGGGSHASIAQGIEWSI
WP_216831504.1 S.YQDYNGHGTHVAGTIAALNNSYGVLGVAPSANIYAVKVLDRNGSGALSSIIQGIEWSV
WP_078596166.1 S.YNDLNGHGTHVAGTIAARDNSYGVLGVAPNVDLYAVKVLDRNGGGSLSGIARGIEWAI
WP_035666680.1 S.YQDYNGHGTHVAGTIAGLNNSYGVLGVAPNVNLYAVKVLDRNGGGSHSAIAQGIEWSV
WP_078597775.1 D.YIDYNGHGTHVAGTIAALDNNYGVLGVSPDVDLYAVKVLGADGSGTISGIAQGGIEWTI
WP_100832725.1 S.YYDYNGHGTHVAGTVAALDNNVGVLGAAPDVDLYAVKVLGADGSGSFASIIQGIEWAV
WP_090774498.1 N.YQDGNGHGTHVAGTIAALDNQIGVLGVSPDVELYAVKVLSSSGSGTLAGIARGIEWAA
WP_096186536.1 D.YQDQNGHGTHVAGTVAALDNDLGVLGVSPDIDLYAVKVLGADGSGSHATIAQGIEWAV
WP_110520788.1 D.YDDYNGHGTHVAGTVAALDNDFGVLGVSPDVDLYAVKVLGADGSGTMAGIAQGIEWAA
WP_122900894.1 E.YQDENGHGTHVAGTIAALDNEVGVLGVSPDVDLYAVKVLGADGSGSHAGIVQGIEWAV
WP_026691049.1 D.YEDYNGHGTHVAGTVAALDNDLGVLGVSPDVDLYAVKVLGADGGGSHASIAQGIEWAV
WP_199800957.1 DPLVDPHEHGTHVAGTIAALDNKVGVVGVAPKASIYAVKVADENGDGYYSWIIKGIEWAI
S08.005_endopep._Q NATQDFQSHGTHVAGTIAALDNTIGVLGVAPSASLYAVKALDRNGDGQYSWIISGIEWAV
CAO03040.1 NATQDFQSHGTHVAGTIAALDNTIGVLGVAPSASLYAVKVLDRNGDGQYSWIISGIEWAV
WP_081105403.1 NATQDFQSHGTHVAGTIAALDNTIGVLGVAPSASLYAVKVLDRNGDGQYSWIISGIEWAV
S08.034_sub._BPN' NPFQDNNSHGTHVAGTVAALNNSIGVLGVAPSASLYAVKVLGADGSGQYSWIINGIEWAI
WP_003155195.1 NPFQDNNSHGTHVAGTVAALNNSVGVLGVAPSASLYAVKVLGADGSGQYSWIINGIEWAI
WP_003327717.1 NPFQDGNSHGTHVAGTVAALNNSVGVLGVAPSASLYAVKVLSSSGSGDYSWIINGIEWAI
WP_039073463.1 NPYQDGSSHGTHVAGTVAALNNSIGVLGVAPNASLYAVKVLDSTGNGQYSWIINGIEWAI
S08.002_mesentericopep. NPYQDGSSHGTHVAGTIAALNNSIGVLGVAPSSALYAVKVLDSTGSGQYSWIINGIEWAI
S08.042_amylosacchariticus NPYQDGSSHGTHVAGTIAALNNSIGVLGVSPSASLYAVKVLDSTGSGQYSWIINGIEWAI
S08.035_sub._J NPYQDGSSHGTHVAGTIAALNNSIGVLGVSPSASLYAVKVLDSTGSGQYSWIINGIEWAI
S08.036_sub._E NPYQDGSSHGTHVAGTIAALNNSIGVLGVAPSASLYAVKVLDSTGSGQYSWIINGIEWAI
S08.044_sub._NAT NPYQDGSSHGTHVAGTIAALNNSIGVLGVAPSASLYAVKVLDSTGSGQYSWIINGIEWAI
WP_053604255.1 SPYSDGNGHGTHVSGTVAALNNTTGVLGVAPDASLYAVKVLDSAGSGSYSGIVSGIEWAT
NPC92104.1 DPFNDGNGHGTHVAGTVAALNNSIGVIGVAPNVSLYAIKVLDSSGSGTYSAIIDGIEWAT

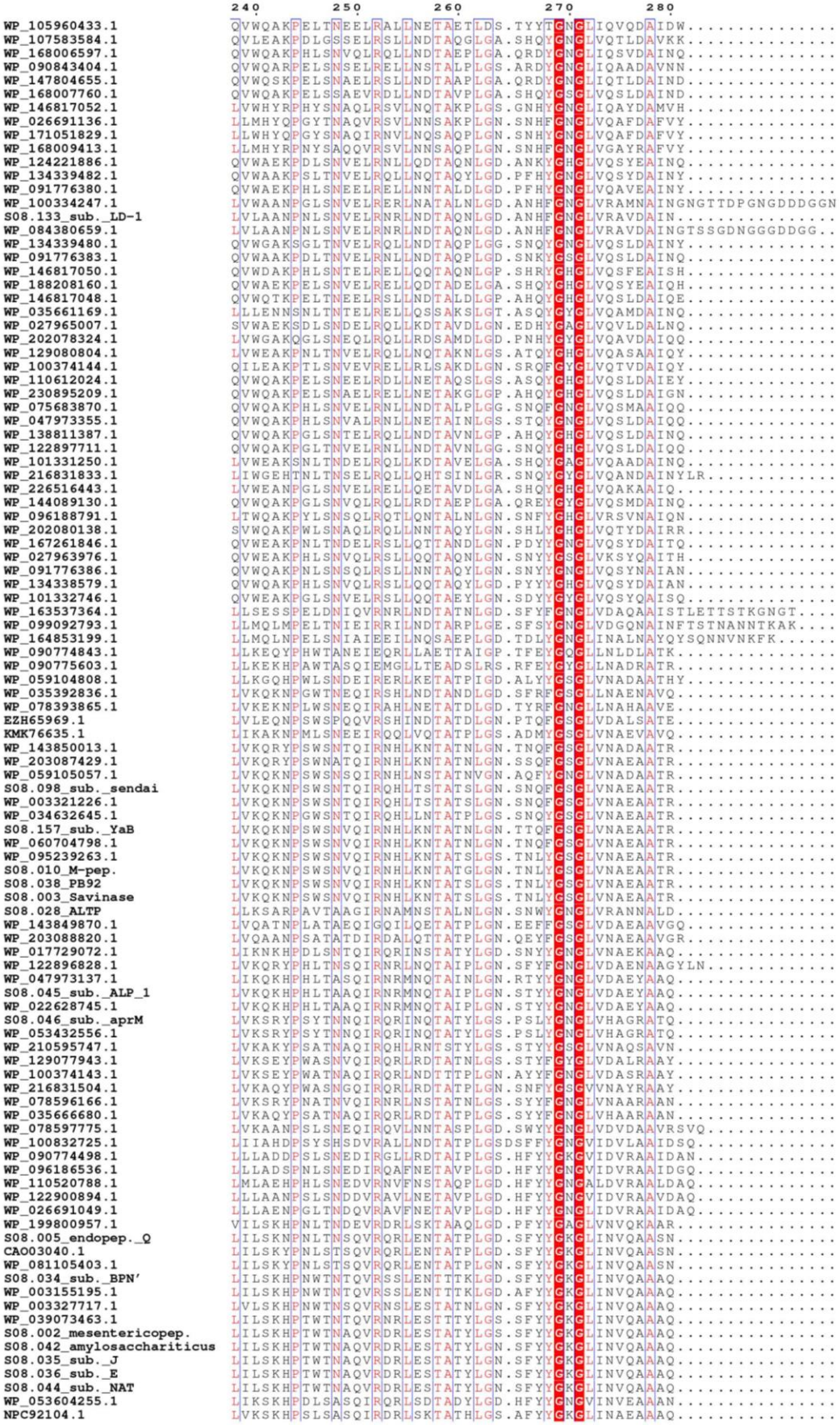

S08.001_sub._Carlsberg Y.NTDGNGHGTHVAGTVAALDNTTGVLGVAPSVSLYAVKVLNSSGSGTYSGIVSGIEWAT
AAS86761.1 Y.NTDGNGHGTHVAGTVAALDNTTGVLGVAPSVSLYAVKVLNSSGSGSYSGIVSGIEWAT
S08.037_sub._DY Y.NTDGNGHGTHVAGTVAALDNTTGVLGVAPNVSLYAIKVLNSSGSGTYSAIVSGIEWAT
NUJ19608.1 Y.NTDGNGHGTHVAGTVAALDNSIGVLGVAPNVSLYAIKVLNSSGSGTYSAIVSGIEWAT
WP_188377243.1 DPFNDQNGHGTHVAGTVAGVDNDLGVLGVAPETDLYAVKVLDGEGSGSYSAIAEGIEWAI
WP_173918387.1 DPFDDLNGHGTHVAGTIAGVDNNVGVLGVAPDVALYAVKVLNGEGSGAYSEIAAGIEWAI
KGX83542.1 DPYNDGNGHGTHVAGTIAGLDNTLGVLGVSPDVSLYAVKVLGSDGSGTYSGIIKGVEWAV
WP_051255158.1 DPYNDGNGHGTHVAGTIAGLDNTLGVLGVSPDVSLYAVKVLGSDGSGTYSGIIKGVEWAV
WP_231417544.1 DPFNDGNGHGTHVAGTIAGVNNSLGVIGVAPSAELYAVKVLNSSGSGSYSGIAKGIEWAV
WP_096155439.1 NALTDGNGHGTHVAGTVAALNNNVGVLGIAYDVDLYAVKVLGSDGSGTLAGIAQGIEWSI
AST90329.1 NALQDGNGHGTHVAGTVAALNNQVGVLGVAYDVDLYAVKVLGADGSGTLSGIAQGIEWSI
WP_230500539.1 DALTDGNGHGTHVAGTVAALNNTVGVLGVAYDVELYAVKVLDSSGGGTLAGIAQGIEWAI
WP_078381234.1 DALTDGNGHGTHVAGTIAALNNNVGVLGVSYDVELYAVKVLSSSGSGTLSGIAQGIEWAI
WP_060666810.1 DALTDGNGHGTHVAGTIAALNNNVGVLGVSYDVNLYAVKVLGADGSGTLAGIAQGIEWAI
WP_152444042.1 DALSDGNGHGTHVAGTIAGLNNTTGVLGVAYNVDLYAVKVLGADGSGTLAGIAQGIEWAI
WP_066412694.1 NPYIDSNGHGTHVAGTVAALNNTVGVLGVAYNAELYAVKVLSASGSGTLSGIAQGVEWSI
WP_230500606.1 NPYQDGNGHGTHVAGTVAALNNSTGVLGVAYNADLYAVKVLNSSGSGTLSGIAQGIEWSI
WP_088017821.1 NALQDGNGHGTHVAGTVAALNNSTGVLGVAYNADLYAVKVLSASGSGTLSGIAQGIEWSI
WP_224838688.1 SGLTDGNGHGTHVAGTVAALNNTAGVLGVAYNADLYAVKVLSASGSGSLSGIAQGIEWAI
WP_100334303.1 NALVDTNGHGTHVAGTVAALNNTIGVVGVAYNADLYAVKVLSASGSGTLSGIAQGVEWAI
WP_094921089.1 DAFSDYNGHGTHVAGTVAGLNNNLGVLGVAPSVSLYAVKVLDGNGSGTYSGIIQGIEWAI
WP_224844257.1 DPYVDLNGHGTHVAGTIAGH...TGILGVAPSVSLYAVKVLDGNGSGTYSGIIQGIEWAI
WP_070119644.1 DPYNDGNSHGTHVAGTVAGLNNTVGVLGVAPSANLYAVKVLDSAGSGTYSGIIQGIEWAV
WP_181472841.1 DPFNDGNSHGTHVAGTVAGLDNNVGVLGVAQSANLYAVKVLDGAGSGTYSGIIQGIEWSV
WP_136946078.1 DPFTDGNSHGTHVAGTVAGLNNNVGVLGVAHTASLYAVKVLDSSGSGTYSGIIQGIEWAV
WP_153236691.1 NALVDGDGHGTHVAGTVAALNNTTGVLGVAYSADLYAVKVLDSSGSGTYSGIIQGIEWAV
WP_077360649.1 NALTDGNSHGTHVAGTVAALNNTTGVLGVAYSADLYAVKVLDSSGSGTYSGIIQGIEWAV
WP_206945444.1 NALVDGNGHGTHVAGTVAGVNNTLGVLGVAPSTELYAVKVLSSEGSGSYSGIAQGIEWAI
WP_029565418.1 DPFVDGDSHGTHVAGTVAALNNTVGVLGTAPDVSLYAVKVLDSTGSGSYSGIAQGIEWAV
TDL80277.1 NPFIDGDSHGTHVAGTVAALNNSTGVLGAAPDVSLYAVKVLDSSGSGSYSGIAQGIEWAV
WP_224844255.1 NALVDGHSHGTHVAGTVAAVNNSLGVVGVAPNVDLYAVKVLDSNGSGSLSDIAKGIEWSI
WP_193538138.1 NALEDGNGHGTHVAGTVAGLNNTLGVLGVAPAADLYAVKVLDSSGSGSFSGIVQGIEWAV
WP_078544469.1 DALTDGNGHGTHVAGTVAGLNNTLGVLGVAPSASLYAVKVLGADGSGTYAGIAQGIEWAV
WP_090849877.1 NALTDGNGHGTHVAGTVSGLNNSLGVLGVAPSASLYAVKVLGADGSGTYSGIAQGIEWAI
WP_022794977.1 SPFEDGNGHGTHVAGTIAALDNGSGLIGVANNTQLHAVKVLDSSGSGSLSTIIKGIEWSI
WP_107586282.1 DPYSDSNGHGTHVAGTVGALDNSHGVLGVAPAADLYAVKVLGAEGGGTLDGIIAGIEWSI
WP_106589713.1 DPMNDENGHGTHVAGTVAALDNGTGLLGMAPDVDLHAVKVLGADGGGTLSGIAQGIEWAI
WP_018922084.1 DPFQDDNGHGTHVAGTVAALDNNEGVLGVAPDVNLHAVKVLGGDGGGTLSGIAQGIEWSI

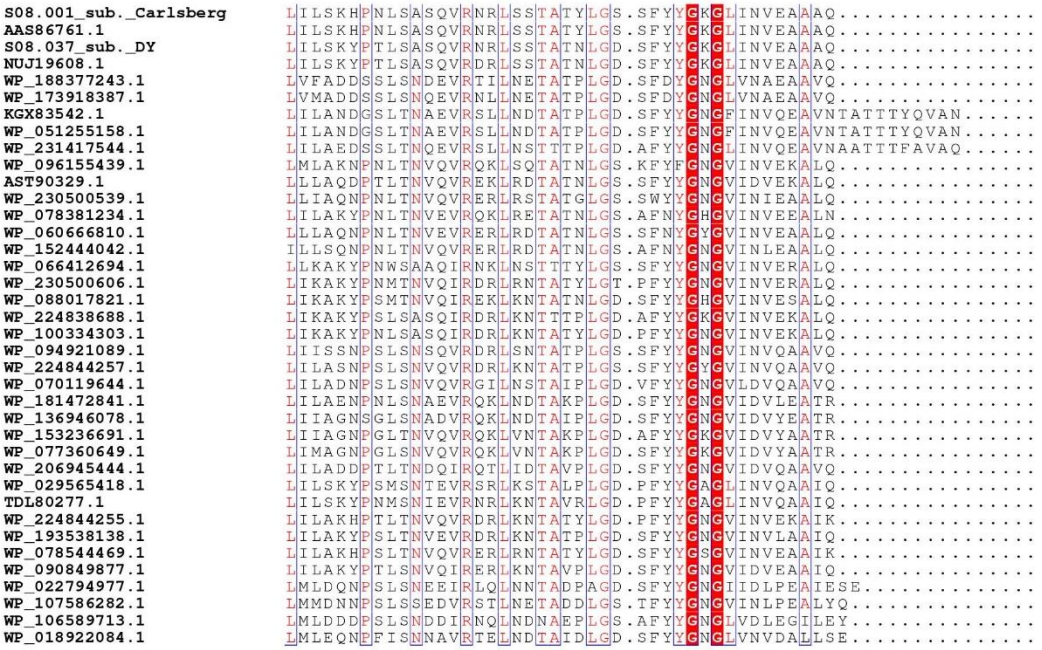

120
130
140
150
160
170
WP_105960433.1
WP_107583584.1
WP_168006597.1
WP_090843404.1
WP_147804655.1
WP_168007760.1
WP_146817052.1
WP_026691136.1
WP_171051829.1
WP_168009413.1
WP_124221886.1
WP_134339482.1
WP_091776380.1
WP_100334247.1
S08.133_sub._LD-1
WP_084380659.1
WP_134339480.1
WP_091776383.1
WP_146817050.1
WP_188208160.1
WP_146817048.1
WP_035661169.1
WP_027965007.1
WP_202078324.1
WP_129080804.1
WP_100374144.1
WP_110612024.1
WP_230895209.1
WP_075683870.1
WP_047973355.1
WP_138811387.1
WP_122897711.1
WP_101331250.1
WP_216831833.1
WP_226516443.1
WP_144089130.1
WP_096188791.1
WP_202080138.1
WP_167261846.1
WP_027963976.1
WP_091776386.1
WP_134338579.1
WP_101332746.1
WP_163537364.1
WP_099092793.1
WP_164853199.1
WP_090774843.1
WP_090775603.1
WP_059104808.1
WP_035392836.1
WP_078393865.1
EZH65969.1
KMK76635.1
WP_143850013.1
WP_203087429.1
WP_059105057.1
S08.098_sub._sendai
WP_003321226.1
WP_034632645.1
S08.157_sub._YaB
WP_060704798.1
WP_095239263.1
S08.010_M-pep.
S08.038_PB92
S08.003_Savinase
S08.028_ALTP
WP_143849870.1
WP_203088820.1
WP_017729072.1
WP_122896828.1
WP_047973137.1
S08.045_sub._ALP_1
WP_022628745.1
S08.046_sub._aprM
WP_053432556.1
WP_210595747.1
WP_129077943.1
WP_100374143.1
WP_216831504.1
WP_078596166.1
WP_035666680.1
WP_078597775.1
WP_100832725.1
WP_090774498.1
WP_096186536.1
WP_110520788.1
WP_122900894.1
WP_026691049.1
WP_199800957.1
S08.005_endopep._Q
CAO03040.1
WP_081105403.1
S08.034_sub._BPN'
WP_003155195.1
WP_003327717.1
WP_039073463.1
S08.002_mesentericopep.
S08.042_amylosaccharitic
S08.035_sub._J
S08.036_sub._E
S08.044_sub._NAT
WP_053604255.1
NPC92104.1


S08.001_sub._Carlsberg TNGMDVINMSLGGPSGSTAMKQAVDNA.YARGVVVVAAAGNSGSSGN.TNTIGYPAKYDS
AAS86761.1 TNGMDVINMSLGGASGSTAMKQAVDNA.YARGVVVVAAAGNSGSSGN.TNTIGYPAKYDS
S08.037_sub._DY QNGLDVINMSLGGPSGSTALKQAVDKA.YASGIVVVAAAGNSGSSGS.QNTIGYPAKYDS
NUJ19608.1 ANNLDVINMSLGGTSGSTALKQAVDKA.YASGVVVVAAAGNSGTSGS.SSTIGYPAKYDS
WP_188377243.1 DNDMDVINMSLGGSIGSSALEQAVNNA.DDSGVLVVAAAGNEGSF.GPFNTIGYPAKYDA
WP_173918387.1 ANDMDVINMSLGGSVGSSALKEAVDNA.YESGILVVAAAGNSGKF.GPFNTIGYPAKYDT
KGX83542.1 SNNIDVVNMSLGGSRGSTSLQQAMDNA.YNSGVLLVAAAGNDGTR.GKKNTIGYPAKYSS
WP_051255158.1 SNNIDVVNMSLGGSRGSTSLQQAMDNA.YNSGVLLVAAAGNDGTR.GKKNTIGYPAKYSS
WP_231417544.1 DNNIDVVNMSLGGSRGSTTLEQAMDQA.YQQGVLLIAAAGNEGSK.GKKNTIGYPAKYAS
WP_096155439.1 ANGMDVINMSLGGSTGSSTLKQASDNA.YNSGIVVVAAAGNSGNFFGLINTIGYPARYDS
AST90329.1 ANNMDVINMSLGGSTGSTTLKQAADNA.YNSGLVVVAAAGNSGDFFGLINTIGYPARYDS
WP_230500539.1 DNNMDVINMSLGGSIGSTTLRRASDNA.YNSGIVVVAAAGNSGSGLGLLNTIGYPARYDS
WP_078381234.1 DNEMDVINMSLGGSTGSSTLKQASDNA.YNSGIVVVAAAGNSGSFFGLINTIGYPARYDS
WP_060666810.1 DNDMDVINMSLGGSTGSTTLKQASDNA.YNSGIVVVAAAGNSGSFLGLVNTIGYPAKYDS
WP_152444042.1 DNNIDVINMSLGGSTGSTTLKQACDNA.YNSGVVVVAAAGNSGSFFGLVNTIGYPAKYDS
WP_066412694.1 ANKMDVINMSLGGSSGSTALQRAVDNA.YRNNIVVVAAAGNSGAQ.GNRNTIGYPARYSS
WP_230500606.1 SNGMHVINMSLGASSGSTALQRACDNA.YARGIVVIAAAGNSGAI.GNQNTIGYPARYSS
WP_088017821.1 ANDMDVINMSLGGSTGSTALQQACDNA.YASGIVVVAAAGNSGSR.GKRNTMGYPARYSS
WP_224838688.1 ANDMDVINMSLGGSTGSTALKQACDNA.YASGIVVVAAAGNSGTK.GKQNTIGYPARYSS
WP_100334303.1 ANNMDVINMSLGGSSGSTALQQAVDNA.YASGIVVVAAAGNSGTR.GRQNTMGYPARYSS
WP_094921089.1 DNNMDVINMSLGGDRGSTSLQIACDNA.NNSGIVVVAAAGNSGSK.GKRNTIGYPAKYAS
WP_224844257.1 QEKVNVINMSLGGSQGSTALQQACDNA.NNSGIVVVAAAGNSGSK.GKRNTIGYPAKYAS
WP_070119644.1 SNDMDVISMSLGGSRGSTSLQQAVDNA.YNSGVLVVAAAGNDGAK.GKRNTIGYPAKYAS
WP_181472841.1 DNNMDVINMSLGGRTGSAALKDAVDNA.YNSGVLVVAAAGNEGSF.LVFNTIGYPAKYDS
WP_136946078.1 DNNMDVINMSLGGSSGSTALKDANDNA.YNAGVLVVAAAGNEGTR.GKQNTIGYPAKYAS
WP_153236691.1 ANNMDVINMSLGGSSGSTALKQACDNA.YNSGVLVVAAAGNSGTK.GKQNTIGYPAKYAS
WP_077360649.1 ANNMDVINMSLGGSSGSTALQQACDNA.YNSGVLVVAAAGNSGTR.GKQNTIGYPAKYAS
WP_206945444.1 SNDMDIINMSLGGSRGSATLEQAVNNA.DAQGVLVVAAAGNEGSK.GKKNTIGYPAKYTA
WP_029565418.1 ANGMDVINMSLGGSQDSTALKQAVDLA.YSRGVVVVAAAGNSGAK.GKRNTIGYPAKYSS
TDL80277.1 DNGMDVINMSLGGSQGSTALQQAVDQA.YNKGVVVVAAAGNSGSK.GKRNTIGYPAKYSS
WP_224844255.1 DNNMDVINMSLGGSSGSTTLKQACDNA.YNSGIVVIAAAGNSGSF.FWLNTIGYPAKYSS
WP_193538138.1 ENGMDVINMSLGASSGSTTLQQACDLA.YNSGIVVVAAAGNSGSK.GKRNTIGYPAKYAS
WP_078544469.1 ENGMDVINMSLGGSQGSTALEQAVDNA.YNSGVVVVAAAGNSGSR.GKRNTIGYPAKYSS
WP_090849877.1 SNNMDVINMSLGGSQGSTALQQAVDNA.YNNGIVVVAAAGNSGSK.GKRNTIGYPAKYSS
WP_022794977.1 SNDMDIVNMSLGTASGAESLEMASDNA.EEEGIFMVGAAGNSGTDG.ANNTIGYPARYDS
WP_107586282.1 ASDMDVINMSLGTPTHSQAMETASNNA.ADAGILVIAAAGNDGTNWFGSNTINYPARYDS
WP_106589713.1 TNDMDVINMSLGGDFGSQALEEASDNA.EAAGVMNIAAAGNSGESWFGGSTIGYPAAYDS
WP_018922084.1 DNNMDVINMSLGGDFGSQALEQASNNA.DDAGIVVIAAAGNSGTDFFGGSTIAYPAQYDS


180 190 200 210 220 230
WP_105960433.1 AMAVAAVDENNQRASFSSTGAEVEIAAPGVGVVSTVP.GNDYAALDGTSMASPHVAGAAA
WP_107583584.1 AMAVAAVDENNDRASFSSTGPAVEISAPGVNVLSTTP.GDTYDAFNGTSMAAPHVAGAAA
WP_168006597.1 AMAVAAVDENNNTASFSSAGPAVEISAPGVNVLSTVP.GNGYDALNGTSMAAPHVAGVAA
WP_090843404.1 AIAVAAVDANNNTASFSSAGPAVEISAPGVDIRSTTP.GNNYQSFNGTSMAAPHVAGVAA
WP_147804655.1 AMAVAAVDSNNQRASFSSAGPAVEISAPGVDILSTVP.NNGYDSLNGTSMAAPHVAGVAA
WP_168007760.1 AMAVAAVDENNNRATFSSTGPAVEISAPGVDVLSTVP.GNGYASFNGTSMAAPHVAGVAA
WP_146817052.1 VMAVAAVDQNENRASFSSTGPAVEISAPGVDVLSTYP.NNNYSSLNGTSMASPHVAGVAA
WP_026691136.1 VMAVAAVDWNNNRASFSSTGPTVEIAAPGVNIQSTFP.GNTYRSLNGTSMASPHVAGVAA
WP_171051829.1 VMAVAAVDWNNNRASFSSTGPAVEISAPGVSIQSTYP.GNQYVSLNGTSMASPHVAGVAA
WP_168009413.1 VMAVGAVDANNNRASFSSTGPAVEIAAPGVSIQSTYP.NNGYRSLNGTSMAAPHVAGVAA
WP_124221886.1 AVAVAAVDQNNNRATFSSTGSAVELSAPGVSVLSTVP.GNDYDSYDGTSMASPHVAGVAA
WP_134339482.1 AMAVAAVDQNNNRASFSSTGNAVEISAPGVGVLSSVP.GNGYASYDGTSMASPHVAGVAA
WP_091776380.1 AIAVAAVDQNNSRASFSSTGDALELSAPGVSVLSTVP.GNGYDSYDGTSMASPHVAGVAA
WP_100334247.1 VIAVAAVDSNNNRASFSSTGPAVEIAAPGVAVYSTTP.GNNYASYNGTSMASPHVAGVAA
S08.133_sub._LD-1 VIAVAAVDSSNNRASFSSTGPAVEIAAPGVNILSTTP.GNSYASYNGTSMASPHVAGVAA
WP_084380659.1 VIAVAAVDSSNNRASFSSTGPAVEIAAPGVNILSTTP.GNSYASYNGTSMASPHVAGVAA
WP_134339480.1 VIAVAAVDQNNNRASFSSTGPAVELSAPGVNILSSVP.GNSYDSYNGTSMASPHVAGVAA
WP_091776383.1 AMAVAAVDENNQRATFSSTGPAVEISAPGVNILSSVP.GNNYDSYNGTSMASPHVAGVAA
WP_146817050.1 AIAVAAIDQNNNRATFSSTGPAVELSAPGVNVLSTVP.GDNYDSYNGTSMAAPHVAGVAA
WP_188208160.1 AIAVAAVDENNQRATFSSTGPDVELSAPGVDVLSSVP.GDSYDTYNGTSMASPHVAGVAA
WP_146817048.1 AMAVAAVDENNDRATFSSTGDAVEIAAPGVDVLSTVP.GDGYDSFSGTSMAAPHVAGVAA
WP_035661169.1 VIAVAAVDQNNNRATFSSTGPAVEISAPGVSILSTIP.NNGYASYNGTSMASPHVAGVAA
WP_027965007.1 VMAVGAVDENNNRPSFSSTGPAVEIAAPGVDTLSTVP.GNDYASLSGTSMASPHVAGVAA
WP_202078324.1 VIAVAAVDADNQRASFSSTGPAVELAAPGVDVLSTVP.NNGYDRYNGTSMASPHVAGVAA
WP_129080804.1 VIAVAATDQSDKRGTFSSTGPAVEISAPGVGILSTTP.NNNYASYNGTSMASPHVAAVAA
WP_100374144.1 VIAVAAVDQNNNRPTFSSTGPAVELAAPGANILSTTP.NNRYAAYNGTSMASPHVAGVAA
WP_110612024.1 VIAVAAIDENNNRASFSSSGPAVEISAPGVSVLSTIP.GNDYAAFNGTSMASPHVAGVAA
WP_230895209.1 VIAVAAVDENNSRATFSSTGPAVEIAAPGVNVLSAVP.GNDYAAFNGTSMASPHVAGVAA
WP_075683870.1 VIAVAATDQNNQRATFSSTGPAVEISAPGAGILSTTP.NNTYASFNGTSMASPHVAGVAA
WP_047973355.1 VIAVAATDQNNQRATFSSTGPAVEISAPGVGILSTTP.NNNYASFNGTSMASPXVAGVAA
WP_138811387.1 VIAVAAVDQNNNRATFSSTGPAVEISAPGVNVLSTTP.GDNYASYNGTSMASPHVAGVAA
WP_122897711.1 VIAVAAVDQNNNRATFSSTGPAVEISAPGVSVLSTTP.GNNYAAFNGTSMASPHVAGVAA
WP_101331250.1 VMAVAAVDENNNRATFSSTGPAVEISAPGVDILSTTP.GDTYSSFNGTSMASPHVAGVAA
WP_216831833.1 VIAVAAVDANNNRASFSSTGPAVELSAPGVSVLSTVP.GNRYASYNGTSMASPHVAGVAA
WP_226516443.1 VIAVAAVDENNNRATFSSTGPAVELSAPGVSVLSTVP.GDGYDSYNGTSMASPHVAGVAA
WP_144089130.1 VVAVAAVDENNQRGSFSSVGSQVELSAPGVQILSTVP.GNGYDSYNGTSMASPHVAGVAA
WP_096188791.1 VIAVAATDQNNNRASFSSHGPAVELSAPGVNVLSTVP.GNGHQSYNGTSMASPHVAGVAA
WP_202080138.1 VMAVAATDSNNNRASFSSHGPAVEISAPGVNILSTVP.GNGYSSLNGTSMASPHVAGAAA
WP_167261846.1 VIAVAAVDQNNNRASFSSHGPDVELAAPGVGVQSTVP.GNGYDSLDGTSMASPHVAGVAA
WP_027963976.1 VIAVGAVDQNNNRASFSSHGPAVELAAPGVGVLSTVP.GNGYSSLNGTSMASPHVAGVAA
WP_091776386.1 VIAVAAVDQNNTRPSFSSHGPAVELAAPGVNVLSTTP.GNNYDSFNGTSMASPHVAGVAA
WP_134338579.1 VIAVAAIDQNNNRASFSSHGPDVELSAPGVSVLSTIP.GNSYDSFNGTSMASPHVAGVAA
WP_101332746.1 VIAVAAVDQNNNRASFSSHGPDVELSAPGVSVLSTVP.GNSYDSLNGTSMASPHVAGVAA
WP_163537364.1 VIAVGAIDQQNNRADFSSVGRELEIMAPGVTIKSTIP..GGYAFSSGTSMAAPHVVGVAA
WP_099092793.1 VISVGAVDQYDNRASFSSVGKELELVAPGIYINSTVP..GGYSIFDGTSMAAPYVTGVAT
WP_164853199.1 VMGIGAIDQNNIIADFSSNGKEVDLVAPGVLVNSTYI..NGYAILNGTSMAAPYVTGVAS
WP_090774843.1 VIAVGSVDSDNKRDATSQYGDGLSLVAPGVDVLSTFL.DGEYVEASGTSMATPAVAGAAA
WP_090775603.1 VLSVGAVDENNRRASFSQYGSGLDIVAPGVDVLSTFL.DGSYVRASGTSMATPAVAGAAA
WP_059104808.1 VLAVGAVDSSNQKANFSQYGEGLDIVAPGVEIGSTYL.GNSYHSLSGTSMASPHVAGVAA
WP_035392836.1 AMAVGATDQSDSLASFSQYGEGLDLVAPGVGVESTYP.GGGYDSLSGTSMAAPHVAGAAA
WP_078393865.1 AMAVGATDQSDSLANFSQYGEGLDIVAPGVGIDSTYT.GSSYDSLSGTSMATPHVAGSAA
EZH65969.1 AMAVGATTQNDTRASFSQYGAGLDLVAPGVGVESTYP.GGGYRSLDGTSMATPHVAGVAA
KMK76635.1 AFAVGATDQVNNRASFSQYGTGLDIVAPGVEVESTYL.NGEYASLSGTSMATPHVAGVAA
WP_143850013.1 AMAVGATDQNNRRANFSQYGAGLDIVAPGVGVQSTYP.GNRYVSMNGTSMASPHVAGVAA
WP_203087429.1 AMAVGATDQNNRRANFSQYGTGIDIVAPGVNVQSTYP.GNRYVSMNGTSMATPHVAGAAA
WP_059105057.1 AMAVGATDRNNNRASFSQYGAGLDIVAPGVGVQSTYL.ANRYASLSGTSMATPHVAGVAA
S08.098_sub._sendai ALAVGATDQNNNRASFSQYGTGLNIVAPGVGIQSTYP.GNRYASLSGTSMATPHVAGVAA
WP_003321226.1 ALAVGATDQNNNRASFSQYGTGLNIVAPGVGIQSTYP.GNRYASLSGTSMATPHVAGVAA
WP_034632645.1 ALAVGATDQNNNRASFSQYGTGLNIVAPGVGVQSTYP.GNRYASLSGTSMATPHVAGVAA
S08.157_sub._YaB AMAVGATDQNNNRATFSQYGAGLDIVAPGVGVQSTVP.GNGYASFNGTSMATPHVAGVAA
WP_060704798.1 AMAVGATDQNNNRASFSQYGAGLDIVAPGVGVQSTVP.GNGYSSFNGTSMATPHVAGVAA
WP_095239263.1 AMAVGATDQNNNRASFSQYGAGLDIVAPGVNVQSTYP.GSTYASLNGTSMATPHVAGAAA
S08.010_M-pep. AMAVGATDQNNNRASFSQYGAGLDIVAPGVNVQSTYP.GSTYASLNGTSMATPHVAGVAA
S08.038_PB92 AMAVGATDQNNNRASFSQYGAGLDIVAPGVNVQSTYP.GSTYASLNGTSMATPHVAGAAA
S08.003_Savinase AMAVGATDQNNNRASFSQYGAGLDIVAPGVNVQSTYP.GSTYASLNGTSMATPHVAGAAA
S08.028_ALTP AMAVAATTSGNVRASFSSTGPAVEIAAPGQDINSTYP.TNTYRSLNGTSMAAPHVAGVAA
WP_143849870.1 VISVAAVDSANNKANFSSYGPENDIAAPGVDVLSTYI.GGEYTELSGTSMASPHVAGVAA
WP_203088820.1 VVSVAAVDSENNKANFSSFGPENNIAAPGVDTLSTYP.GGQYAELSGTSMASPHVAGVAA
WP_017729072.1 VMAVGAVDSDSNRASFSSYGEELEIMAPGVNIHSTYL.LNGYATLSGTSMASPHVAGVAA
WP_122896828.1 VMAVGAVDSNYNRASFSSVGNELEIMAPGVSIQSTYL.SNSYRALNGTSMAAPHVAGTAA
WP_047973137.1 VMAVGAVDQNGTRASFSSYGAELEIMAPGVNINSTHL.NNNYRSLNGTSMAAPHVAGVAA
S08.045_sub._ALP_1 VMAVGAVDQNGNRANFSSYGSELEIMAPGVNINSTYL.NNGYRSLNGTSMASPHVAGVAA
WP_022628745.1 VMAVGAVDQNGNRANFSSYGSELEIMAPGVNINSTYL.NNGYRSLNGTSMASPHVAGVAA
S08.046_sub._aprM VMAVAAVDQNGQRASFSTYGPEIEISAPGVNVNSTYT.GNRYVSLSGTSMATPHVAGVAA
WP_053432556.1 VMAVAAVDQNGQRASFSTYGPEIEISAPGVNVNSTYT.GNRYVSLSGTSMATPHVAGVAA
WP_210595747.1 VMAVAATDSNNARASFSTYGPEVEISAPGVGINSTYP.TNSYTSLNGTSMASPHVAGVAA
WP_129077943.1 VIAVGAVNSSNQRASFSTYGSQLELMAPGVNVQSTYL.NNTYSSLNGTSMAAPHVAAVAA
WP_100374143.1 VMAVGAVDSRNRLASFSTFGNEQEIVAPGVNVQSTHL.SNGYVSLNGTSMASPHVAGAAA
WP_216831504.1 VMAVGAVDSNNNRASFSTYGSQIEIVAPGVSVLSTYT.SNRYVSLNGTSMATPHVAGVAA
WP_078596166.1 VMAVGATDSNNNRASFSTFGNELEIMAPGVSVLSTYP.TNRYVSLNGTSMASPHVAGVAA
WP_035666680.1 VMAVGAVDSNNNRASFSTFGNELEIMAPGVSILSTHL.SNQYVSLNGTSMASPHVAGVAA
WP_078597775.1 VMAVAAIDSNNNRASFSSVGGELEISAPGVSVLSTYI.GNDYASLNGTSMASPHVAGAAA
WP_100832725.1 VMAVGAVDEDNQRASFSSVGNALEVMAPGVSVYSTYL.DNSYASLNGTSMASPHVAGAAA
WP_090774498.1 VIAVGAVNENNARASFSSVGDELDVMAPGVDVLSSHL.NNAYQSFSGTSMASPHVAGAAA
WP_096186536.1 VIAVGAVDENNERASFSSVGDELDVMAPGVDVNSTYL.DDSYAELNGTSMASPHVAGAAA
WP_110520788.1 VIAVGAVDQNNERASFSSVGDELDVMAPGVSVNSTYL.NDGYQALNGTSMAAPHVAGAAA
WP_122900894.1 AIAVGAVDENNDRASFSSVGNELDVMAPGVAIDSTYL.DNSYAALSGTSMAAPHVAGAAA
WP_026691049.1 VIAVGAVDSNNNRASFSSVGNELDVVAPGVSILSTYL.GNDYAALNGTSMASPHVAGAAA
WP_199800957.1 VIAVASVDQRKQRAFDSSVGEEVEVSAPGVSTLSTIP.HNEYGYKSGTSMASPHVAGAAA
S08.005_endopep._Q TIAVANVNSNNVRNSSSSAGPELDVSAPGTSILSTVP.SSGYTSYTGTSMASPHVAGAAA
CAO03040.1 TIAVANVNGNNVRNSSSSAGPELDVSAPGTSILSTVP.SSGYTSYTGTSMASPHVAGAAA
WP_081105403.1 TIAVANVNSNNVRNSSSSAGPELDVSAPGTSILSTVP.SSGYTSYTGTSMASPHVAGAAA
S08.034_sub._BPN' VIAVGAVDSSNQRASFSSVGPELDVMAPGVSIQSTLP.GNKYGAYNGTSMASPHVAGAAA
WP_003155195.1 VIAVGAVNSSNQRASFSSVGSELDVMAPGVSIQSTLP.GNKYGAYNGTSMASPHVAGAAA
WP_003327717.1 VIAVGAVDSNNQRASFSSAGSELDVMAPGVSIQSTLP.GSSYGSYNGTSMASPHVAGAAA
WP_039073463.1 TIAVGAVNSSNQRGSFSSVGPELDVMAPGVSIQSTLP.GGTYGSYNGTSMATPHVAGAAA
S08.002_mesentericopep. TIAVGAVNSSANQRASFSSAGSELDVMAPGVSIQSTLP.GGTYGAYNGTSMATPHVAGAAA
S08.042_amylosachariticus TIAVGAVNSSNQRASFSSAGSELDVMAPGVSIQSTLP.GGTYGAYNGTSMATPHVAGAAA
S08.035_sub._J TIAVGAVNSSNQRASFSSAGSELDVMAPGVSIQSTLP.GGTYGAYNGTSMATPHVAGAAA
S08.036_sub._E TIAVGAVNSSNQRASFSSAGSELDVMAPGVSIQSTLP.GGTYGAYNGTSMATPHVAGAAA
S08.044_sub._NAT TIAVGAVNSSNQRASFSSVGSELDVMAPGVSIQSTLP.GGTYGAYNGTSMATPHVAGAAA
WP_053604255.1 VIAVGAVDSNNKKAYFSSVGDELEVMAPGVSVQSTLP.GNQYTELDGTSMASPHVAGAAA
NPC92104.1 AIAVGAVDNNSQRAYFSSVGDELEVMAPGAAVNSTYP.GNSYKSLNGTSMASPHVAGAAA


S08.001_sub._Carlsberg VIAVGAVDSNSNRASFSSVGAELEVMAPGAGVYSTYP.TSTYATLNGTSMASPHVAGAAA
AAS86761.1 VIAVGAVDSNSNRASFSSVGAELEVMAPGAGVYSTYP.TNTYATLNGTSMASPHVAGAAA
S08.037_sub._DY VIAVGAVDSNKNRASFSSVGAELEVMAPGVSVYSTYP.SNTYTSLNGTSMASPHVAGAAA
NUJ19608.1 VIAVGAVNSSNQRASFSSVGPELDVVAPGVSIYSTYP.SNTYATLNGTSMASPHVAGAAA
WP_188377243.1 AMAVGAVDSDNNVASFSSRGNELEVMAPGVDVLSSIP.ENSYDEFNGTSMASPHVAGVAA
WP_173918387.1 AMAVGAVDSSNKVASFSSRGNELEVMAPGVDILSSVP.ENSYDSFNGTSMASPHVAGVAA
KGX83542.1 VMAVGAVDSNLNRASFSSVGSELEVMAPGVDIYSTLP.GNSYDYYNGTSMASPHVAGAAA
WP_051255158.1 VMAVGAVDSNLNRASFSSVGSELEVMAPGVDIYSTLP.GNSYDYYNGTSMASPHVAGAAA
WP_231417544.1 VVAVGAVDESLNRASFSSVGEELEVMAPGANIYSTLP.GNTYGSYNGTSMASPHVAGAAA
WP_096155439.1 VIAVGAVDANNNRASFSSVGNELEVMAPGVNILSTLP.GNSYGSLNGTSMASPHVAGAAA
AST90329.1 VIAVGAVDSNNRRASFSSVGSQLEVMAPGVNILSTLP.GNSYGSLNGTSMASPHVAGAAA
WP_230500539.1 VIAVGAVDSNNNRASFSSVGNQLEVMAPGVAINSTLP.GNQYGEFNGTSMASPHVAGAAA
WP_078381234.1 VIAVGAVDANNNRASFSSVGNELELMAPGVNINSTLP.GNQYGSLNGTSMASPHVAGAAA
WP_060666810.1 VIAVGAVDSNNNRASFSSVGNQLEVMAPGVSINSTLP.GNKYGELNGTSMASPHVAGAAA
WP_152444042.1 VIAVGAVDANNKRASFSSVGNELEVMAPGVSINSTLP.GNQYGELNGTSMASPHVAGAAA
WP_066412694.1 VIAVGAVDSNNNRASFSSVGSELEVMAPGVSILSTVP.GSSYASYNGTSMASPHVAGAAA
WP_230500606.1 VIAVGAVSSNNTRASFSSVGNELEVMAPGVSILSTTP.GNNYASFNGTSMAAPHVAGAAA
WP_088017821.1 VIAVGAVDSSNNRASFSSVGSELEVMAPGVSILSTTP.GNNYSSFNGTSMASPHVAGAAA
WP_224838688.1 VIAVGAVDSNNNRASFSSVGAELEVMAPGVSILSTTP.GNSYSSFNGTSMASPHVAGAAA
WP_100334303.1 VIAVGAVDSNNNRASFSSVGAELEVMAPGVSVLSTVP.GGGYASYNGTSMASPHVAGAAA
WP_094921089.1 VIAVGAVDSSNNRASFSSVGNELEVMAPGVSVYSSVP..GGYDTYNGTSMASPHVAGAAA
WP_224844257.1 VIAVGAVDSNNNRGSFSSVGNELEVMAPGVAIYSTVPYGNGYDTYNGTSMASPHVAGAAA
WP_070119644.1 AMAVGAVDDSNNRASFSSVGDELEIMAPGVNVLSSVP.GNAYDYFNGTSMATPHVSGAAA
WP_181472841.1 AMAVGAVDSNNNRASFSSVGSELEVMAPGVDILSSVP.GNSYASYNGTSMASPHVAGAAA
WP_136946078.1 VMAVGAVDSSNNRASFSSVGSELEVMAPGVSILSSVP.GNSYSSYNGTSMASPHVAGAAA
WP_153236691.1 VMAVGAVDSNNARASFSSVGSELEVMAPGVNILSSVP.GNNYASYNGTSMASPHVAGAAA
WP_077360649.1 VMAVGAVDSSNNRASFSSVGSELEVMAPGVNILSTVP.GNGYDSYNGTSMASPHVAGAAA
WP_206945444.1 AMAVGAVDSANNRASFSSVGEELEVMAPGVDILSSVP.GNNYDRYNGTSMASPHVAGAAA
WP_029565418.1 AIAVGAVDSANARASFSSVGSELEVMAPGVNILSSVP.GNKYASFNGTSMASPHVAGAAA
TDL80277.1 VIAVGAVDAANSRASFSSVGSELEVMAPGVSILSTVP.GNKYASFNGTSMASPHVAGAAA
WP_224844255.1 VIAVGAVDSTNKRASFSSVGNELEVMAPGVSILSTVP.GNRYDTFNGTSMASPHVAGAAA
WP_193538138.1 VIAVGAVDSSNNRASFSSVGNELEVMAPGVNILSSVP.GNSYDSYNGTSMASPHVAGAAA
WP_078544469.1 VIAVGAVDSSNNRASFSSVGSELEVMAPGVNILSSVP.GNGYDSYNGTSMASPHVAGAAA
WP_090849877.1 VIAVGAVDNTNNRASFSSVGNELEVMAPGVSILSSVP.GNSYDSYNGTSMASPHVAGAAA
WP_022794977.1 VMAIGSIDRYEDRSSFSSVGEELEVMAPGSSITSTYP.GQNYATLSGTSMAAPHVSGAAA
WP_107586282.1 VMAVGALDSNDNRASFSSVGNELEVMAPGADINSTYP.GNSYASLNGTSMAAPHAAGAAA
WP_106589713.1 VMAVGATDENDQRASFSSVGDSLEIMAPGVNINSTYP.GNTYESLNGTSMASPHVAGAAA
WP_018922084.1 VMAVGAVDQNNNRASFSSVGNTLEVMAPGVNIESTMP.GNNYASLDGTSMAAPHVAGTAA


240
250
260
270
280
WP_105960433.1 QVWQAKPELTNEELRALLNETAETLDS.TYYTGNGLIQVQDAIDW.....
WP_107583584.1 QVLEAKPDLGSSELRSLLNDTAQGLGA.SHQYGNGLVQTLDAVKK.....
WP_168006597.1 QVWQAKPHLSNVQLRQLLNDTAEPLGA.QRDYGNGLIQSVDAINQ.....
WP_090843404.1 QVWQARPELSNSELRELLNSTALPLGS.ARDYGNGLIQAADAVNN.....
WP_147804655.1 QVWQSKPELSNAELRSLLNDTAAPLGA.QRDYGNGLIQTLDAIND.....
WP_168007760.1 QVWQAKPELSSAEVRDLLNDTAVPLGA.SHQYGSGLVQSLDAIND.....
WP_146817052.1 LVWHYRPHYSNAQLRSVLNQTAKPLGS.GNHYGNGLIQAYDAMVH.....
WP_026691136.1 LLMHYQPGYTNAQVRSVLNNSAKPLGN.SNHFGNGLVQAFDAFVY.....
WP_171051829.1 LLWHYQPGYSNAQIRNVLNQSAQPLGN.SNHYGNGLVQAFDAFVY.....
WP_168009413.1 LLMHYRPNYSAQQVRSVLNNSAQPLGN.SNHFGNGLVGAYRAFVY.....
WP_124221886.1 QVWAEKPDLSNVELRNLLQDTAQNLGD.ANKYGHGLVQSYEAINQ.....
WP_134339482.1 QVWAAKPSLTNVELRQLLNQTAQYLGD.PFHYGNGLVQSYDAINY.....
WP_091776380.1 QVWAEKPHLSNEELRELLNNTALDLGD.PFHYGNGLVQAVEAINY.....
WP_100334247.1 LVWAANPGLSNVELRERLNATALNLGD.ANHFGNGLVRAMNAINGNGTTDPGNGDDDGGN
S08.133_sub._LD-1 LVLAANPNLSNVELRNRLNDTAQNLGD.ANHFGNGLVRAVDAIN.....
WP_084380659.1 LVLAANPNLSNVELRNRLNDTAQNLGD.ANHFGNGLVRAVDAINGTSSGDNGGGDDGG..
WP_134339480.1 QVWGAKSGLTNVELRQLLNDTAQPLGG.SNQYGNGLVQSLDAINY.....
WP_091776383.1 QVWAAKPDLTNVELRNLLNDTAQPLGD.SNKYGSGLVQSLDAINQ.....
WP_146817050.1 QVWDAKPHLSNTELRELLQQTAQNLGP.SHRYGHGLVQSFEAISH.....
WP_188208160.1 QVWAEKPELSNVELRELLQDTADELGA.SHQYGHGLVQSYEAIQH.....
WP_146817048.1 QVWQTKPELTNEELRSLLNDTALDLGP.AHQYGHGLVQSLDAIQE.....
WP_035661169.1 LLLENNSNLTNTELRELLQSSAKSLGT.ASQYGYGLVQAMDAINQ.....
WP_027965007.1 SVWAEKSDLSNDELRQLLKDTAVDLGN.EDHYGAGLVQVLDALNQ.....
WP_202078324.1 LVWGAKQGLSNEQLRQLLRDSAMDLGD.PNHYGYGLVQAVDAIQQ.....
WP_129080804.1 LVWEAKPNLTNVELRQLLNQTAKNLGS.ATQYGHGLVQASAAIQY.....
WP_100374144.1 QILEAKPTLSNVEVRELLRLSAKDLGN.SRQFGYGLVQTVDAIQY.....
WP_110612024.1 QVWQAKPELSNEELRDLLNETAQSLGS.ASQYGHGLVQSLDAIEY.....
WP_230895209.1 QVWQAKPELSNAELRELLNETAKGLGP.AHQYGHGLVQSLDAIGN.....
WP_075683870.1 QVWQAKPHLSNVELRNLLNDTALPLGG.SNQFGNGLVQSMAAIQQ.....
WP_047973355.1 QVWQAKPHLSNVALRNLLNETAINLGS.STQYGNGLVQSLDAIQQ.....
WP_138811387.1 QVWQAKPGLSNTELRQLLNDTAVNLGP.AHQYGHGLVQSLDAINQ.....
WP_122897711.1 QVWQAKPGLTNVELRNLLNDTAVNLGG.SNQYGHGLVQSLDAIQQ.....
WP_101331250.1 LVWEAKSNLTNDELRQLLKDTAVELGA.SHQYGAGLVQAADAINQ.....
WP_216831833.1 LIWGEHTNLTNSELRQLLQHTSINLGR.SNQYGYGLVQANDAINYLR.....
WP_226516443.1 LVWEANPGLSNVELRELLQETAVDLGA.SHQYGHGLVQAKAAIQ.....
WP_144089130.1 QVWQAKPGLSNVQLRQLLRDTAEPLGA.QREYGYGLVQSMDAINQ.....
WP_096188791.1 LTWQAKPYLSNSQLRQTLQNTALNLGN.SNFYGHGLVRSVNAIQN.....
WP_202080138.1 SVWQAKPWLSNAQLRQLLNNTAQYLGN.SHLYGHGLVQTYDAIRR.....
WP_167261846.1 QVWEAKPNLTNDELRSLLQTTANDLGN.PDYYGNGLVQSYDAITQ.....
WP_027963976.1 QVWEAKPHLSNVQLRSLLQQTAQNLGN.SNYYGSGLVKSYQAITH.....
WP_091776386.1 QVWQAKPNLSNSQLRSLLNNTAQYLGN.SNYYGNGLVQSYNAIAN.....
WP_134338579.1 QVWQAKPHLSNVQLRSLLQQTAQYLGD.PYYYGHGLVQSYDAIAN.....
WP_101332746.1 QVWEAKPGLSNVELRSLLQQTAEYLGN.SDYYGYGLVQSYQAISQ.....
WP_163537364.1 LLSESSPELDNIQVRNRLNDTATNLGD.SFYFGNGLVDAQAAISTLETTSTKGNGT....
WP_099092793.1 LLMQLMPELTNIEIRRILNDTARPLGE.SFSYGNGLVDGQNAINFTSTNANNTKAK....
WP_164853199.1 LLMQLNPELSNIAIEEILNQSAEPLGD.TDLYGNGLINALNAYQYSQNNVNKFK.....
WP_090774843.1 LLKEQYPHWTANEIEQRLLAETTAIGP.TFEYGQGLLNLDLATK.....
WP_090775603.1 LLKEKHPAWTASQIEMGLLTEADSLRS.RFEYGYGLLNADRATR.....
WP_059104808.1 LLKGQHPWLSNDEIRERLKETATPIGD.ALYYGSGLVNADAATHY.....
WP_035392836.1 LVKQKNPGWTNEQIRSHLNDTANDLGD.SFRFGSGLLNAENAVQ.....
WP_078393865.1 LVKEKNPLWSNEQIRAHLNETATDLGD.TYRFGNGLLNAHAAVE.....
EZH65969.1 LVLEQNPSWSPQQVRSHINDTATDLGN.PTQFGSGLVDALSATE.....
KMK76635.1 LIKAKNPMLSNEEIRQQLVQTATPLGS.ADMYGSGLVNAEVAVQ.....
WP_143850013.1 LVKQRYPSWSNTQIRNHLKNTATNLGN.TNQFGSGLVNADAATR.....
WP_203087429.1 LVKQRYPSWNATQIRNHLKNTATNLGN.SSQFGSGLVNAEAATR.....
WP_059105057.1 LVKQKNPSWSNSQIRNHLNSTATNVGN.AQFYGNGLVNADAATR.....
S08.098_sub._sendai LVKQKNPSWSNTQIRQHLTSTATSLGN.SNQFGSGLVNAEAATR.....
WP_003321226.1 LVKQKNPSWSNTQIRQHLTSTATSLGN.SNQFGSGLVNAEAATR.....
WP_034632645.1 LVKQKNPGWSNTQIRQHLLNTATPLGS.SNQYGSGLVNAEAATR.....
S08.157_sub._YaB LVKQKNPSWSNVQIRNHLKNTATNLGN.TTQFGSGLVNAEAATR.....
WP_060704798.1 LVKQKNPSWSNVQIRNHLKNTATNLGN.TNQFGSGLVNAEAATR.....
WP_095239263.1 LVKQKNPSWSNVQIRNHLKNTATSLGS.TNLYGSGLVNAEAATR.....
S08.010_M-pep. LVKQKNPSWSNVQIRNHLKNTATGLGN.TNLYGSGLVNAEAATR.....
S08.038_PB92 LVKQKNPSWSNVQIRNHLKNTATSLGS.TNLYGSGLVNAEAATR.....
S08.003_Savinase LVKQKNPSWSNVQIRNHLKNTATSLGS.TNLYGSGLVNAEAATR.....
S08.028_ALTP LLKSARPAVTAAGIRNAMNSTALNLGN.SNWYGNGLVRANNALD.....
WP_143849870.1 LVQATNPLATAEQIGQILQETATPLGN.EEFFGSGLVDAEAAVGQ.....
WP_203088820.1 LVQAANPSATATDIRDALQTTATPLGN.QEYFGSGLVNAEAAVGR.....
WP_017729072.1 LIKNKHPDLSNTQIRQRINSTATYLGD.SNYYGNGLVNAEKAAQ.....
WP_122896828.1 LVKQRYPHLTNSQIRNRLNQTAIPLGN.SFYFGNGLVDAENAAGYLN.....
WP_047973137.1 LIKQKHPHLTASQIRNRMNQTAINLGN.RTYYGNGLVDAEYAAQ.....
S08.045_sub._ALP_1 LVKQKHPHLTAAQIRNRMNQTAIPLGN.STYYGNGLVDAEYAAQ.....
WP_022628745.1 LVKQKHPHLTAAQIRNRMNQTAIPLGN.STYYGNGLVDAEYAAQ.....
S08.046_sub._aprM LVKSRYPSYTNNQIRQRINQTATYLGS.PSLYGNGLVHAGRATQ.....
WP_053432556.1 LVKSRYPSYTNNQIRQRINQTATYLGS.PSLYGNGLVHAGRATQ.....
WP_210595747.1 LVKAKYPSATNAQIRQHLRNTSTYLGS.STYYGSGLVNAQSAVN.....
WP_129077943.1 LVKSEYPWASNVQIRQRLRDTATNLGS.STYFGYGLVDALRAAY.....
WP_100374143.1 LVKSEYPWATNAQIRQRLNDTTTPLGN.AYYFGNGLVDASRAAY.....
WP_216831504.1 LVKAQYPWASNGQIRQRLRDTATPLGN.SNFYGSGVVNAYRAAY.....
WP_078596166.1 LVKSRYPNATNVQIRNRLNSTATNLGS.SYYFGNGLVNAARAAN.....
WP_035666680.1 LVKAQYPSATNAQIRQRLRDTATPLGS.SYYFGNGLVHAARAAN.....
WP_078597775.1 LVKAANPSLSNEQIRQVLNNTASPLGD.SWYYGNGLVDVDAAVRSVQ.....
WP_100832725.1 LIIAHDPSYSHSDVRALLNDTATPLGSDSFFYGNGVIDVLAAIDSQ.....
WP_090774498.1 LLLADDPSLSNEDIRGLLRDTAIPLGS.HFYYGKGVIDVRAAIDAN.....
WP_096186536.1 LLLADSPNLSNEDIRQAFNETAVPLGD.HFYYGNGVIDVRAAIDGQ.....
WP_110520788.1 LMLAEHPHLSNEDVRNVFNSTAQPLGD.HFYYGNGALDVRAALDAQ.....
WP_122900894.1 LLLAANPSLSNDDVRAVLNETAVPLGD.HFYYGNGVIDVRAAVDAQ.....
WP_026691049.1 LLLAENPGLTNDQVRAVFNETAVPLGD.HFYYGNGLIDVRAAIDAQ.....
WP_199800957.1 VILSKHPNLTNDEVRDRLSKTAAQLGD.PFYYGAGLVNVQKAAR.....
S08.005_endopep._Q LILSKNPNLTNSQVRQRLENTATPLGD.SFYYGKGLINVQAASN.....
CAO03040.1 LILSKYPNLSTSQVRQRLENTATPLGD.SFYYGKGLINVQAASN.....
WP_081105403.1 LILSKYPNLSTSQVRQRLENTATPLGN.SFYYGKGLINVQAASN.....
S08.034_sub._BPN' LILSKHPNWTNTQVRSSLENTTTKLGD.SFYYGKGLINVQAAAQ.....
WP_003155195.1 LILSKHPNWTNTQVRSSLENTTTKLGD.AFYYGKGLINVQAAAQ.....
WP_003327717.1 LVLSKHPNWTNSQVRNSLESTATNLGN.SFYYGKGLINVQAAAQ.....
WP_039073463.1 LILSKHPTWTNTQVRNRLESTTTYLGS.SFYYGKGLINVQAAAQ.....
S08.002_mesentericopep. LILSKHPTWTNAQVRDRLESTATYLGS.SFYYGKGLINVQAAAQ.....
S08.042_amylosachariticus LILSKHPTWTNAQVRDRLESTATYLGN.SFYYGKGLINVQAAAQ.....
S08.035_sub._J LILSKHPTWTNAQVRDRLESTATYLGN.SFYYGKGLINVQAAAQ.....
S08.036_sub._E LILSKHPTWTNAQVRDRLESTATYLGN.SFYYGKGLINVQAAAQ.....
S08.044_sub._NAT LILSKHPTWTNAQVRDRLESTATYLGN.SFYYGKGLINVQAAAQ.....
WP_053604255.1 LIKSKHPDLSASQIRQRLSDTADYLGD.HFYYGNGVINVEAAAN.....
NPC92104.1 LVKSKHPSLSASQIRDRLSKTATHLGS.AFYYGKGLINAEAAAQ.....


S08.001_sub._Carlsberg LILSKHPNLSASQVRNRLSSTATYLGS.SFYYGKGLINVEAAAQ.....
AAS86761.1 LILSKHPNLSASQVRNRLSSTATYLGS.SFYYGKGLINVEAAAQ.....
S08.037_sub._DY LILSKYPTLSASQVRNRLSSTATNLGD.SFYYGKGLINVEAAAQ.....
NUJ19608.1 LILSKYPTLSASQVRDRLSSTATNLGD.SFYYGKGLINVEAAAQ.....
WP_188377243.1 LVFADDSSLSNDEVRTILNETATPLGD.SFDYGNGLVNAEAAVQ.....
WP_173918387.1 LVMADDSSLSNQEVRNLLNETATPLGD.SFDYGNGLVNAEAAVQ.....
KGX83542.1 LILANDGSLTNAEVRSLLNDTATPLGD.SFYYGNGFINVQEAVNTATTTYQVAN.....
WP_051255158.1 LILANDGSLTNAEVRSLLNDTATPLGD.SFYYGNGFINVQEAVNTATTTYQVAN.....
WP_231417544.1 LILAEDSSLTNQEVRSLLNSTTTPLGD.AFYYGNGLINVQEAVNAATTTFAVAQ.....
WP_096155439.1 LMLAKNPNLTNVQVRQKLSQTATNLGS.KFYFGNGVINVEKALQ.....
AST90329.1 LLLAQDPTLTNVQVREKLRDTATNLGS.SFYYGNGVIDVEKALQ.....
WP_230500539.1 LLIAQNPNLTNVQVRERLRSTATGLGS.SWYYGNGVINIEAALQ.....
WP_078381234.1 LILAKYPNLTNVEVRQKLRETATNLGS.AFNYGHGVINVEEALN.....
WP_060666810.1 LLLAQNPNLTNVEVRERLRDTATNLGS.SFNYGYGVINVEAALQ.....
WP_152444042.1 ILLSQNPNLTNVQVRERLRDTATNLGS.AFNYGNGVINLEAALQ.....
WP_066412694.1 LLKAKYPNWSAAQIRNKLNSTTTYLGS.SFYYGNGVINVERALQ.....
WP_230500606.1 LIKAKYPNMTNVQIRDRLRNTATYLGT.PFYYGNGVINVERALQ.....
WP_088017821.1 LIKAKYPSMTNVQIREKLKNTATNLGD.SFYYGHGVINVESALQ.....
WP_224838688.1 LIKAKYPSLSASQIRDRLKNTTTPLGD.AFYYGKGVINVEKALQ.....
WP_100334303.1 LIKAKYPNLSASQIRDRLKNTATYLGD.PFYYGNGVINVEKALQ.....
WP_094921089.1 LIISSNPSLSNSQVRDRLSNTATPLGS.SFYYGNGVINVQAAVQ.....
WP_224844257.1 LILASNPSLSNVQVRDRLKNTATPLGS.SFYYGYGVINVQAAVQ.....
WP_070119644.1 LILADNPSLSNVQVRGILNSTAIPLGD.VFYYGNGVLDVQAAVQ.....
WP_181472841.1 LILAENPNLSNAEVRQKLNDTAKPLGD.SFYYGNGVIDVLEATR.....
WP_136946078.1 LIIAGNSGLSNADVRQKLNDTAIPLGD.SFYYGNGVIDVYEATR.....
WP_153236691.1 LIIAGNPGLTNVQVRQKLVNTAKPLGD.AFYYGKGVIDVYAATR.....
WP_077360649.1 LIMAGNPGLSNVQVRQKLVNTAKPLGD.SFYYGKGVIDVYAATR.....
WP_206945444.1 LILADDPTLTNDQIRQTLIDTAVPLGD.SFYYGNGVIDVQAAVQ.....
WP_029565418.1 LILSKYPSMSNTEVRSRLKSTALPLGD.PFYYGAGLINVQAAIQ.....
TDL80277.1 LILSKYPNMSNIEVRNRLKNTAVRLGD.PFYYGAGLINVQAAIQ.....
WP_224844255.1 LILAKHPTLTNVQVRDRLKNTATYLGD.PFYYGNGVINVEKAIK.....
WP_193538138.1 LILAKYPSLTNVEVRDRLKNTATPLGD.SFYYGNGVINVLAAIQ.....
WP_078544469.1 LILAKHPSLTNVQVRERLRNTATYLGD.SFYYGSGVINVEAAIK.....
WP_090849877.1 LILAKYPTLSNVQIRERLKNTAVPLGD.SFYYGNGVIDVEAAIQ.....
WP_022794977.1 LMLDQNPSLSNEEIRLQLNNTADPAGD.SFYYGNGIIDLPEAIESE.....
WP_107586282.1 LMMDNNPSLSSEDVRSTLNETADDLGS.TFYYGNGVINLPEALYQ.....
WP_106589713.1 LMLDDDPSLSNDDIRNQLNDNAEPLGS.AFYYGNGLVDLEGILEY.....
WP_018922084.1 LMLEQNPFISNNAVRTELNDTAIDLGD.SFYYGNGLVNVDALLSE.....

```
WP_105960433.1              ................
WP_107583584.1              ................
WP_168006597.1              ................
WP_090843404.1              ................
WP_147804655.1              ................
WP_168007760.1              ................
WP_146817052.1              ................
WP_026691136.1              ................
WP_171051829.1              ................
WP_168009413.1              ................
WP_124221886.1              ................
WP_134339482.1              ................
WP_091776380.1              ................
WP_100334247.1              PGPGKPGNGKGNGRN
S08.133_sub._LD-1           ................
WP_084380659.1              SEPTKPGNGKGNGRN
WP_134339480.1              ................
WP_091776383.1              ................
WP_146817050.1              ................
WP_188208160.1              ................
WP_146817048.1              ................
WP_035661169.1              ................
WP_027965007.1              ................
WP_202078324.1              ................
WP_129080804.1              ................
WP_100374144.1              ................
WP_110612024.1              ................
WP_230895209.1              ................
WP_075683870.1              ................
WP_047973355.1              ................
WP_138811387.1              ................
WP_122897711.1              ................
WP_101331250.1              ................
WP_216831833.1              ................
WP_226516443.1              ................
WP_144089130.1              ................
WP_096188791.1              ................
WP_202080138.1              ................
WP_167261846.1              ................
WP_027963976.1              ................
WP_091776386.1              ................
WP_134338579.1              ................
WP_101332746.1              ................
WP_163537364.1              KKPKK...........
WP_099092793.1              K...............
WP_164853199.1              ................
WP_090774843.1              ................
WP_090775603.1              ................
WP_059104808.1              ................
WP_035392836.1              ................
WP_078393865.1              ................
EZH65969.1                  ................
KMK76635.1                  ................
WP_143850013.1              ................
WP_203087429.1              ................
WP_059105057.1              ................
S08.098_sub._sendai         ................
WP_003321226.1              ................
WP_034632645.1              ................
S08.157_sub._YaB            ................
WP_060704798.1              ................
WP_095239263.1              ................
S08.010_M-pep.              ................
S08.038_PB92                ................
S08.003_Savinase            ................
S08.028_ALTP                ................
WP_143849870.1              ................
WP_203088820.1              ................
WP_017729072.1              ................
WP_122896828.1              ................
WP_047973137.1              ................
S08.045_sub._ALP_1          ................
WP_022628745.1              ................
S08.046_sub._aprM           ................
WP_053432556.1              ................
WP_210595747.1              ................
WP_129077943.1              ................
WP_100374143.1              ................
WP_216831504.1              ................
WP_078596166.1              ................
WP_035666680.1              ................
WP_078597775.1              ................
WP_100832725.1              ................
WP_090774498.1              ................
WP_096186536.1              ................
WP_110520788.1              ................
WP_122900894.1              ................
WP_026691049.1              ................
WP_199800957.1              ................
S08.005_endopep._Q          ................
CAO03040.1                  ................
WP_081105403.1              ................
S08.034_sub._BPN'           ................
WP_003155195.1              ................
WP_003327717.1              ................
WP_039073463.1              ................
S08.002_mesentericopep.     ................
S08.042_amylosachariticus   ................
S08.035_sub._J              ................
S08.036_sub._E              ................
S08.044_sub._NAT            ................
WP_053604255.1              ................
NPC92104.1                  ................
```

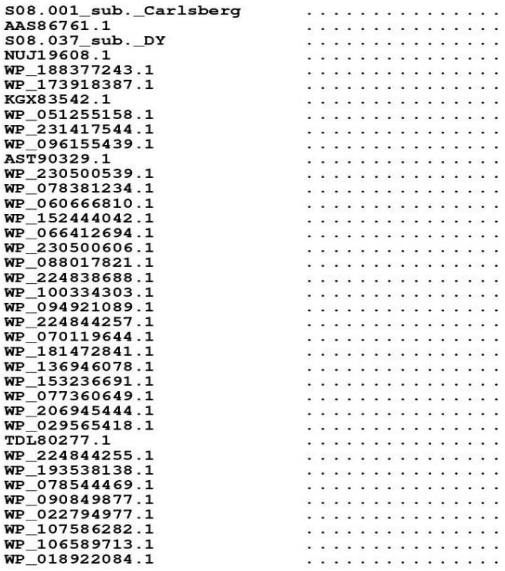


Figure S18 MSA of mature subtilisin sequences of the S8 *Bacillaceae* data mining sequences.

6.2 Supplementary: Biochemical characterisation of a novel oxidatively stable, halotolerant, and high-alkaline subtilisin from *Alkalihalobacillus okhensis* Kh10-101^T

Table S1 Preculture medium

Component	Concentration
$MgSO_4$	0.7 $g \cdot L^{-1}$
KH_2PO_4	7.7 $g \cdot L^{-1}$
$(NH_4)_2SO_4$	2.8 $g \cdot L^{-1}$
$FeSO_4 \times 7H_2O$	0.09 $g \cdot L^{-1}$
$MnSO_4 \times H_2O$	0.09 $g \cdot L^{-1}$
Peptone from soy	40 $g \cdot L^{-1}$
Yeast extract	5 $g \cdot L^{-1}$
Tetracycline	20 $\mu g \cdot mL^{-1}$
$CaCl_2$	1 $g \cdot L^{-1}$
Glucose	20 $g \cdot L^{-1}$
Trace elements	see Table S3

Table S2 Fermentation medium

Component	Concentration
$MgSO_4$	0.7 $g \cdot L^{-1}$
KH_2PO_4	7.7 $g \cdot L^{-1}$
$(NH_4)_2SO_4$	2.8 $g \cdot L^{-1}$
$FeSO_4 \times 7H_2O$	0.09 $g \cdot L^{-1}$
$MnSO_4 \times H_2O$	0.09 $g \cdot L^{-1}$
Peptone from soy	40 $g \cdot L^{-1}$
Yeast extract	5 $g \cdot L^{-1}$
Tetracycline	20 $\mu g \cdot mL^{-1}$
$CaCl_2$	1 $g \cdot L^{-1}$
Glucose	8 $g \cdot L^{-1}$ start glucose + glucose feed
Trace elements	see Table S3
Polypropylene glycol 2000 (PPG)	2 mL + PPG feed

Table S3 Trace element solution

Chemical	Concentration
Citric acid∗H_2O	40 g·L^{-1}
$MnSO_4$∗H_2O	4 g·L^{-1}
$ZnSO_4$∗$7H_2O$	5 g·L^{-1}
$CuSO_4$∗$5H_2O$	4 g·L^{-1}
$FeSO_4$∗$7H_2O$	10.67 g·L^{-1}
$NiSO_4$∗$6H_2O$	0.25 g·L^{-1}
$CoCl_2$∗$6H_2O$	0.324 g·L^{-1}
H_3BO_3	0.06 g·L^{-1}
Na_2MoO_4∗$2H_2O$	0.655 g·L^{-1}

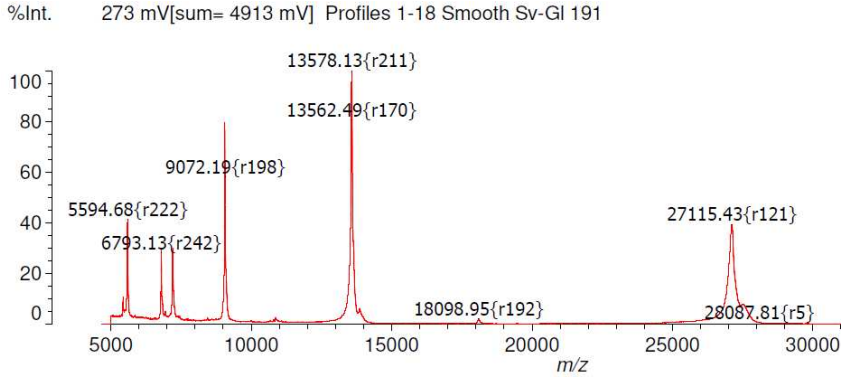


Figure S1 MALDI-TOF mass spectra of SPAO. The labels on the peaks indicate the measured average molecular mass. The peaks correspond from right to left M/z up to M/5z.

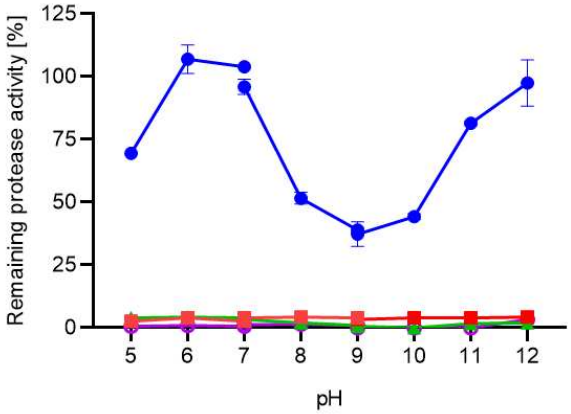


Figure S2 The effect of pH on the stability of purified SPAO, BPN', Savinase, and subtilisin Carlsberg. The activity was measured with the suc-AAPF-pNA assay in standard buffer at pH 8.6 after incubation for 4 h at 4°C in Tris-maleate buffer (pH 5 – 7), in Tris-HCl (pH 7 – 9), and in glycine-NaOH (pH 9 – 12). The activity at 0 h was considered as 100 % activity; highest residual activities: SPAO (closed circles; 288 U/mg), BPN' (squares; 17 U/mg), Savinase (triangles; 15 U/mg) and subtilisin Carlsberg (open circles; 23 U/mg). The experiments were performed in triplicates and data are plotted as mean values ± SD.

6.3 Supplementary: New robust subtilisins from halotolerant and halophilic *Bacillaceae*

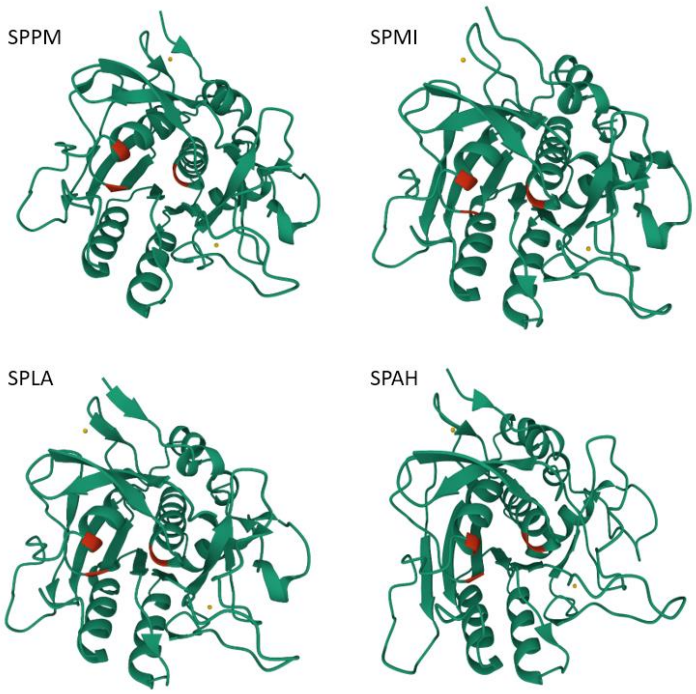


Figure S1 Homology models of the mature forms of SPPM, SPMI, SPLA and SPAH obtained using I-TASSER software. *In silico* metal-binding analysis predicted the existence of two Ca^{2+}-binding sites (yellow balls). The catalytic residues Asp^{32}, His^{64}, and Ser^{221} for SPPM, SPMI, SPLA and Asp^{32}, His^{66}, and Ser^{224} for SPAH are shown in red.

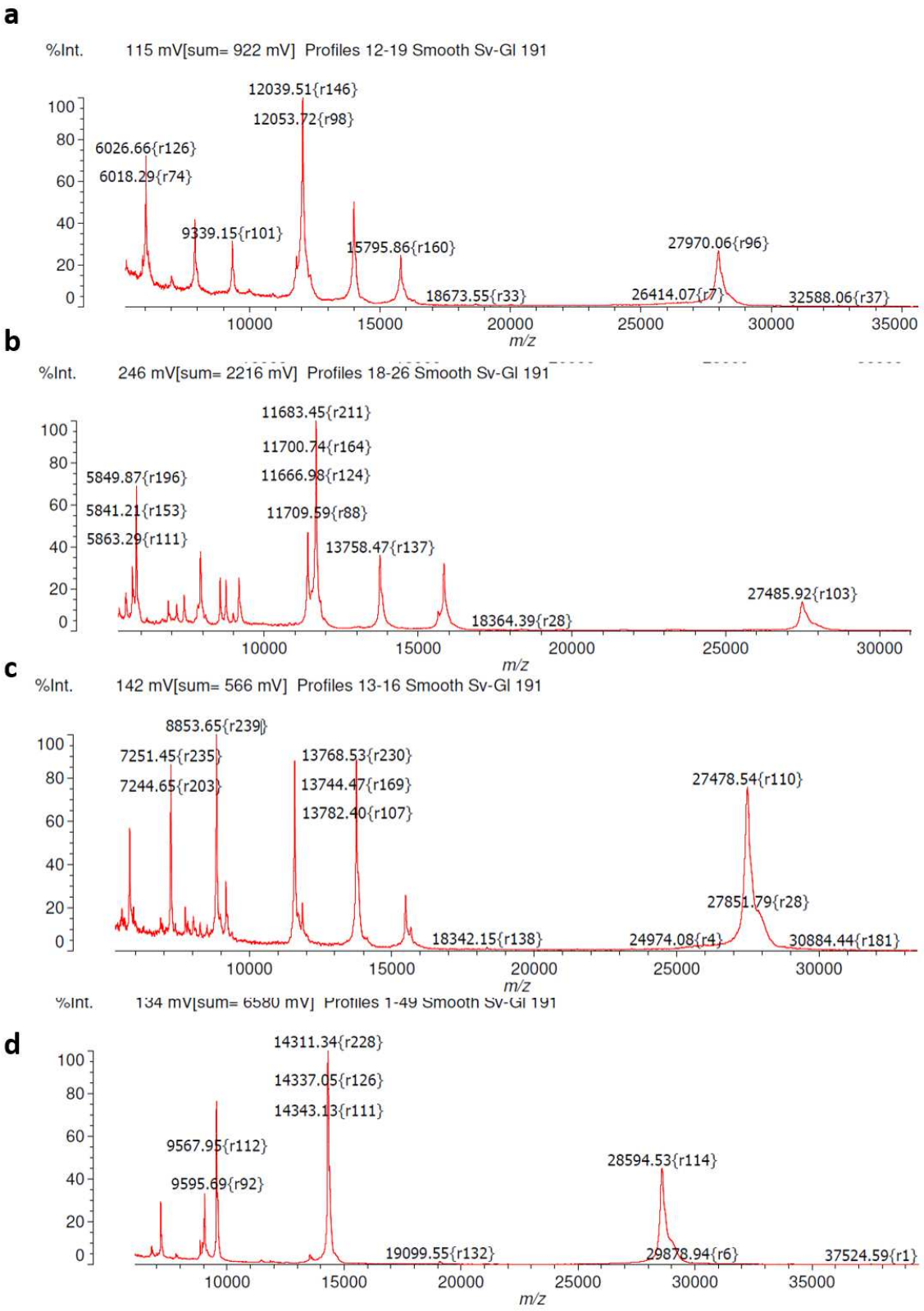


Figure S2 MALDI-TOF mass spectra of SPPM (a), SPMI (b), SPLA (c) and SPAH (d). The labels on the peaks indicate the measured average molecular mass. The peaks correspond from right to left M/z up to M/5z.

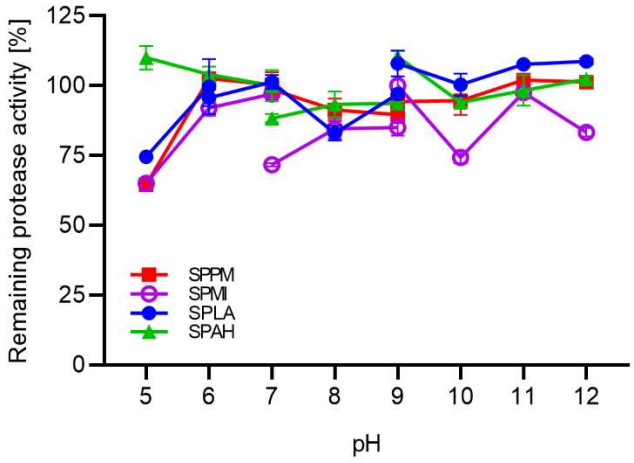


Figure S3 The effect of pH on the stability of purified SPPM, SPMI, SPLA and SPAH. The activity was measured with the suc-AAPF-pNA assay in standard buffer at pH 8.6 after incubation for 24 h at 4 °C in Tris-maleate buffer (pH 5 – 7), in Tris-HCl (pH 7 – 9), and in glycine-NaOH (pH 9 – 12). The activity at 0 h was considered as 100 %; highest residual activities: 115 U/mg for SPPM (red squares), 165 U/mg for SPMI (violet open circles), 219 U/mg for SPLA (blue closed circles), and 221 U/mg for SPAH (green triangles). The experiments were performed in triplicates and data are plotted as mean values ± SD.

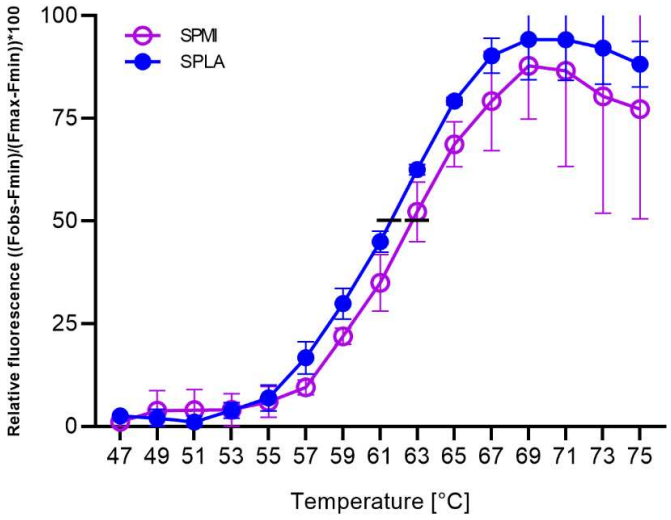


Figure S4 Melting curves of purified SPMI and SPLA. The effect of temperature on the stability of the enzyme using SYPRO® Orange as a fluorescence probe, based on the changes in fluorescence emission intensity (Ex/Em = 470/550 nm) (5 x SYPRO® Orange, 10 mM HEPES-NaOH pH 8.0, 3 mM PMSF), is shown as normalized denaturation curves of the thermal shift assay for the proteases SPMI (violet open circles) and SPLA (blue closed circles). The inflection point corresponds to the melting temperature (T_m), at which 50 % of the protein is unfolded (-). The experiment was performed in triplicates and data are plotted as mean values ± SD.

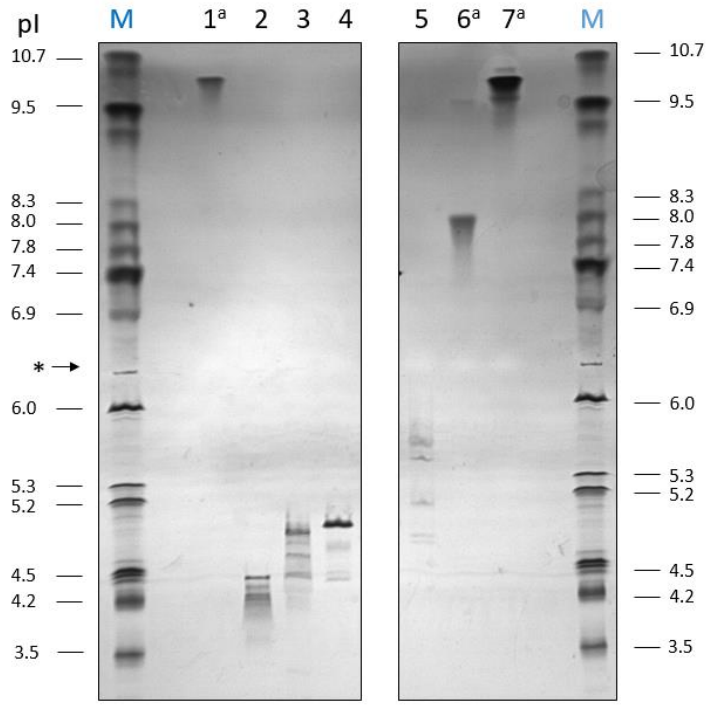


Figure S5 Determination of the pI of the purified proteases. Isoelectric focusing was performed with a SERVALYT™ PRECOTES™ wide range pH 3-10 precast gel according to manufacturer recommendations. Lane M, SERVA IEF marker 3-10; lanes 1-7, purified proteases rebuffered in 10 mM HEPES-NaOH pH 7.0. *Sample application point. SPAO (1); SPPM (2); SPAH (3); SPLA (4); SPMI (5); subtilisin Carlsberg (6); Savinase (7). [a] (Falkenberg *et al.* 2022b).

Table S1 Oligonucleotides for amplification of the indicated subtilisin genes by PCR using genomic DNA of *P. marinus*, *M. indicus*, *L. alkalitelluris*, and *A. haloalkaliphilus* as template

Name	Sequence 5'->3'
aprE_P.marinus_fw	AGAAGACTTAATGAAAAGATTATTTTTTGCAGTGATG
aprE_P.marinus_rv	AGAAGACAAGTTAATTCGCTACTTGATAAGTAGTAGT
aprE_M.indicus_fw	AAAGAAGACGGAATGAAAAAGAAAAAGATTTTCAGTCT GCTTC
aprE_M.indicus_rv	AAAGAAGACCCGTTATTGAATGGCGGCCTGGAC
aprE_L.alkalitelluris_fw	AGAAGACTTAATGAAGAAAATGAAATTAGTAAGTAGTAT TTTGC
aprE_L.alkalitelluris_rv	AGAAGACTTGTTACTTTATAGCAGCTTCTACATTAATGAC
aprE_A.haloalkaliphilus_fw	AGAAGACATAATGAAAAAGCTTCTTATTGTTTTAAGTACC
aprE_A.haloalkaliphilus_rv	AGAAGACATGTTAGTGAGAGATTGCTTCAAAAGAC

Table S2 pI values and number of amino acids in the proteases

Protease	experimental pI	Number of residues					AB ratio[a]
		Arg	Asp	Glu	His	Lys	
SPPM	4.3	4	19	8	4	6	1.9
SPMI	5.5	5	14	6	4	11	1.0
SPLA	5.0	6	12	9	6	7	1.1
SPAH	4.9	4	20	9	9	4	1.7
SPAO[b]	9.8	9	4	4	6	3	0.4

[a] The AB ratio [AB = (Glu + Asp)/(Lys + His + Arg)] was calculated as described in (Rhodes *et al.* 2010).

[b] SPAO from *Alkalihalobacillus okhensis* Kh10-101^T (Falkenberg *et al.* 2022b)

6.4 Supplementary: Biochemical characterisation of a novel broad pH spectrum subtilisin from *Fictibacillus arsenicus* DSM 15822^{T}

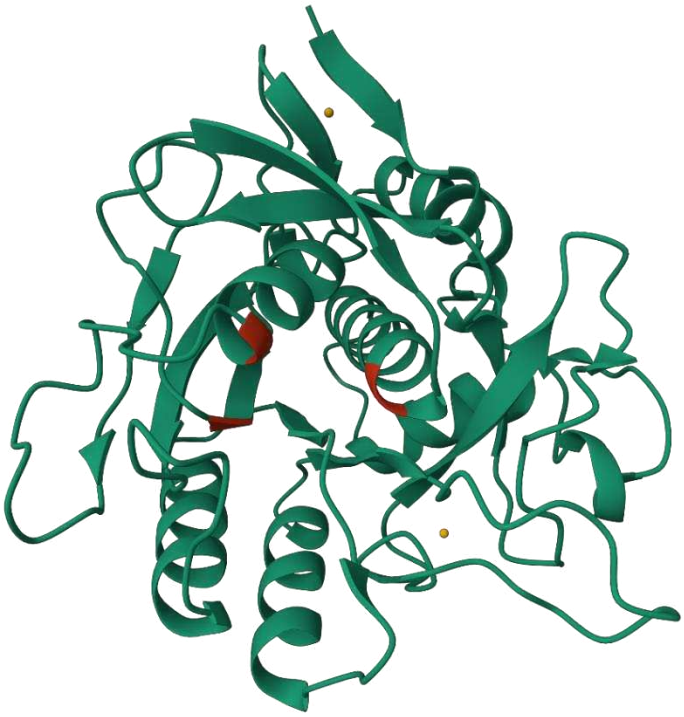


Figure S1 Homology model of the mature SPFA obtained using I-TASSER software. *In silico* metal-binding analysis predicted the existence of two Ca^{2+}-binding sites (yellow balls). The catalytic residues Asp^{32}, His^{64}, and Ser^{221} are shown in red.

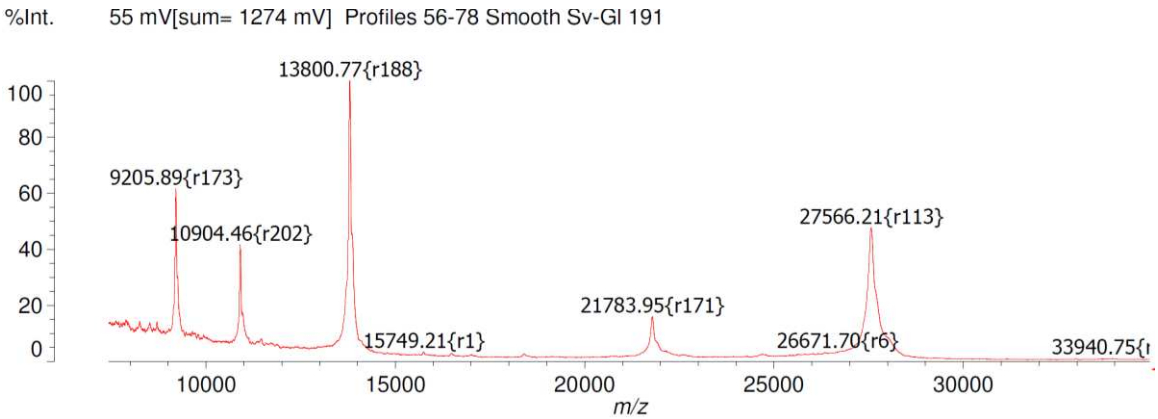


Figure S2 MALDI-TOF mass spectra of SPFA. The labels on the peaks indicate the measured average molecular mass. The peaks correspond from right to left M/z up to M/5z.

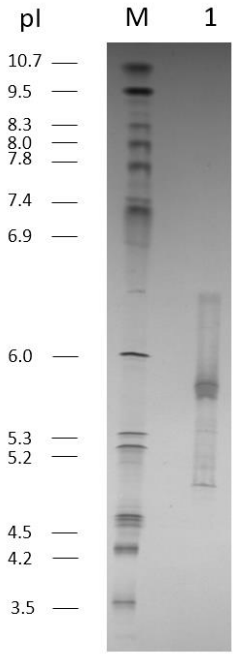


Figure S3 Determination of the pI of the purified proteases. Isoelectric focusing was performed with a SERVALYT™ PRECOTES™ wide range pH 3-10 precast gel according to manufacturer recommendations. Lane M, SERVA IEF marker 3-10; lanes 1 purified SPFA rebuffered in 10 mM HEPES-NaOH pH 8.0.

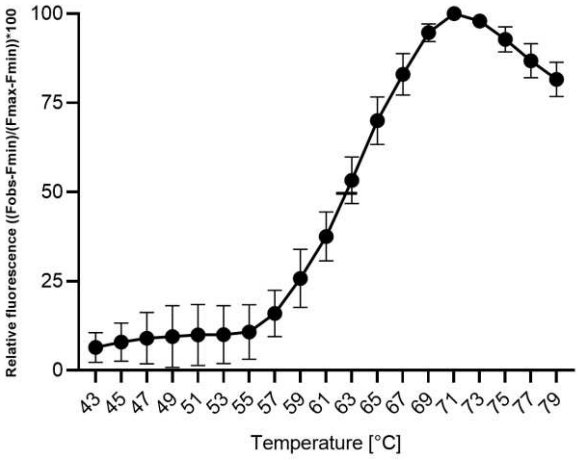


Figure S4 Normalised melting curve of purified SPFA. The melting point (T_m) at which 50 % of the protein is unfolded (-) was determined using SYPRO® Orange as a fluorescent probe (Ex/Em = 470/550 nm) (5 x SYPRO® Orange, 10 mM HEPES-NaOH pH 8.0, 3 mM PMSF). The experiment was performed in triplicates and data are plotted as mean values ± SD.

Table S1 Oligonucleotides for amplification of the gene for SPFA by PCR using genomic DNA of *F. arsenicus* as template

Name	Sequence 5'->3'
aprE_F.arsenicus_fw	AGAAGACGAAATGAAAAAAACTGTATTACGCACG
aprE_F.arsencius_rv	AGAAGACATGTTATCTTGTCGCTGCGTAAAC

Table S2 pI value and AB ratio calculation

Protease	experimental pI	Number of residues					AB ratio[a]
		Arg	Asp	Glu	His	Lys	
SPFA	~5.8	5	13	4	6	8	0.9

[a] The AB ratio [AB = (Glu + Asp)/(Lys + His + Arg)] was calculated as described in (Rhodes *et al.* 2010).

6.5 Supplementary: *Metabacillus indicus* DSM 16189 as promising host for recombinant protease production

Table S1 Preculture medium

Component	Concentration
Marine broth (Carl Roth, Karlsruhe, Germany)	40.1 $g \cdot L^{-1}$
Tryptone/Peptone from Casein	35 $g \cdot L^{-1}$
Yeast extract	4 $g \cdot L^{-1}$
$MnSO_4$ x H_2O	0.01 $g \cdot L^{-1}$
(Kanamycin)	50 $\mu g \cdot mL^{-1}$
Glucose	10 $g \cdot L^{-1}$
Trace elements	1:1000 of the stock solution (Table S3)
pH	7.0

Table S2 Fermentation medium

Component	Concentration
$MnSO_4$ x H_2O	0.01 $g \cdot L^{-1}$
Tryptone/Peptone from Casein	35 $g \cdot L^{-1}$
Yeast extract	4 $g \cdot L^{-1}$
(Kanamycin)	50 $\mu g \cdot mL^{-1}$
Glucose	8 $g \cdot L^{-1}$ start glucose + glucose feed
Trace elements	1:1000 of the stock solution (Table S3)
Polypropylene glycol 2000 (PPG)	2 mL + PPG feed
pH	7.0

Table S3 Trace element stock solution

Chemical	Concentration
Citric acid$*H_2O$	40 $g \cdot L^{-1}$
$MnSO_4*H_2O$	4 $g \cdot L^{-1}$
$ZnSO_4*7H_2O$	5 $g \cdot L^{-1}$
$CuSO_4*5H_2O$	4 $g \cdot L^{-1}$
$FeSO_4*7H_2O$	10.67 $g \cdot L^{-1}$
$NiSO_4*6H_2O$	0.25 $g \cdot L^{-1}$
$CoCl_2*6H_2O$	0.324 $g \cdot L^{-1}$
H_3BO_3	0.06 $g \cdot L^{-1}$
$Na_2MoO_4*2H_2O$	0.655 $g \cdot L^{-1}$

Danksagung

Gerne möchte ich die Gelegenheit nutzen, um allen meinen herzlichen Dank auszusprechen, ohne deren Unterstützung und Beiträge die Erstellung dieser Arbeit nicht möglich gewesen wäre.

Ein besonderer Dank gebührt Prof. Dr. Petra Siegert und Prof. Dr. Johannes Bongaerts für die Überlassung des überaus interessanten wissenschaftlichen Themas. Ihre fachliche Kompetenz, Ratschläge und Leidenschaft waren eine wertvolle Unterstützung im Verlauf des Promotionsprojektes. Ich bin Ihnen aufrichtig dankbar für Ihre Zeit, Geduld und Ihr Vertrauen in meine Fähigkeiten. Besonders bedanken möchte ich mich auch bei Prof. Dr. Michael Bott für die Übernahme der Betreuung, das Interesse am Fortgang meines Projekts, die zahlreichen Diskussionen und die Unterstützung.

Ein großer Dank geht außerdem an alle Mitglieder des Fachbereichs Chemie und Biotechnologie der Fachhochschule Aachen für ihre Hilfe im Verlauf dieser Arbeit. Besonders danke ich Fabian Hüllen und Gerrit Haeger für die unzähligen Diskussionen, Kaffeepausen und die großartige Zeit im Labor. Eure Unterstützung hat entscheidend zum Erfolg dieser Arbeit beigetragen. Ebenso möchte ich mich bei Sabine Jossek, Christoph Horst, Brigitte Lehan und Melanie Gellissen für die tatkräftige Unterstützung bei allen wissenschaftlichen und organisatorischen Fragen bedanken.

Ich möchte mich bei allen Studenten bedanken, die mich durch ihre Praktika und Abschlussarbeiten unterstützt haben. Ein besonderer Dank gilt Sophie, Vincent, Jessica, Nihal, Tobias und Jonas.

Danken möchte ich auch unseren Kooperationspartnern der Firma Henkel AG & Co. KGaA. Ihre Zusammenarbeit und Engagement haben meine Forschung bereichert und erweitert. Durch unsere gemeinsamen Bemühungen konnten wichtige Erkenntnisse gewonnen und neue Perspektiven erschlossen werden.

Ein herzliches Dankeschön geht an meine Familie und meine Freundin Rebecca, die mir stets den Rücken gestärkt haben. Ihr Glaube an meine Fähigkeiten und bedingungslose Unterstützung haben mir die nötige Motivation gegeben, meine Ziele zu verfolgen und Hindernisse zu überwinden. Dieser Erfolg gehört auch euch.

Ich möchte allen meine aufrichtige Dankbarkeit für ihre Mitwirkung, Unterstützung und Inspiration während meiner gesamten Doktorarbeit zum Ausdruck bringen. Ohne euch wäre dieses Ergebnis nicht möglich gewesen. Ich bin dankbar für die wertvollen Erfahrungen, die ich während dieser Zeit sammeln durfte.

Erklärung

Ich versichere an Eides Statt, dass die vorgelegte Dissertation von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist. Die Dissertation wurde in der vorgelegten oder in ähnlicher Form noch bei keiner anderen Institution eingereicht. Ich habe bisher keine erfolglosen Promotionsversuche unternommen.

_____

Neuss, den 06.06.2023