



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e. g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.



THE UNIVERSITY *of* EDINBURGH

Novel extremophilic metalloproteases for consumer product application

Benjamin Michael Astles

Thesis submitted for the degree of Doctor of Philosophy

The University of Edinburgh

2023

Abstract

Proteases are one of the largest classes of commercial enzymes worldwide, with applications in a range of industries, including detergents, tanning and dairy. This project aims to develop novel metalloproteases for use in the detergents industry.

Metagenomic libraries were screened for likely protease candidates, predicting biochemical characteristics based on their primary sequence. Candidate DNA was then synthesised, codon-optimised for expression in *Bacillus subtilis*. A protease deficient *B. subtilis* strain was engineered for this purpose through CRISPR-MAD7. This included a system for overexpression of the foldase *prsA*, that has been shown to increase secretion of heterologous proteins. Proteases were expressed in this *B. subtilis* chassis downstream of a signal peptide library to optimise secretion. Five protease candidates were successfully expressed and secreted, and of these, four had optimal activities in the desired ranges of pH and temperature.

Of these candidates, one was purified and further characterised. This protease was optimally active at 40 °C, pH 9.0 and was stable across a pH range of 5.0 – 10.0. It was found to be composed of an N-terminal propeptide, a catalytic domain and a C-terminal domain. The C-terminal domain, although inactive by itself, proved to be crucial for achieving full protease activity. This protease displayed a non-specific enzymatic profile, showing particular effectiveness in tackling blood and gravy stains, indicating its potential for use in detergent applications. Furthermore, the enzyme exhibited remarkable resilience in the presence of sequestrants, suggesting its suitability for various detergent formulations. In addition, comparative analyses against the commercial proteases Neutrase® and Carnival Ecity™ 16 L highlighted the protease's potential advantages in terms of salt tolerance and its activity in hard water. These encouraging initial findings warrant further investigation into the practical applications of this protease in the detergent industry.

Lay Summary

In this study, we embarked on a mission to make your laundry routine greener and more energy efficient. You might not know it, but the detergents you use to clean your clothes contain special proteins called proteases. These proteins are like tiny cleaning agents that break down protein-based stains, such as blood stains. However, most proteases in detergents work best in hot water, which means more energy is needed to wash your clothes. Our goal was to find proteases that could do their job just as effectively in cold water.

Proteases have tonnes of uses in Nature. One of these is to break down proteins in the environment on which microbes can feed, and these proteases generally work well in detergents. There are over a million proteases that we know about, though. Using advanced computer programs, we scoured through this vast sea of proteases to identify potential candidates. These candidates were then put through rigorous testing in a laboratory to see which ones would be good to use in detergents. We managed to isolate one of these proteases and tested its ability to break down stains.

We found that this protease is incredibly versatile. It can handle a wide range of stains, including those notorious troublemakers like blood and gravy stains. Another great feature is that it is not bothered by the minerals and stuff that can make your water hard. This is a common problem for other proteases used in detergents, but not for this one.

So, what does this mean for you? Well, in the future, you might find this protease in your laundry detergent. This simple change can make your laundry routine more environmentally friendly and save you money on your energy bills. With this protease, you can wash your clothes in cold water, reduce your carbon footprint, and still have your clothes come out sparkling clean. It's like a win-win for you and the planet. So, keep an eye out for this because it may revolutionise the way you do laundry!

Declaration

I declare that this thesis is an original report of my research, has been written by me and has not been submitted for any previous degree. The experimental work is almost entirely my own work; the collaborative contributions have been indicated clearly and acknowledged. I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification.

Benjamin Michael Astles

01 October 2023

Impact of COVID-19

The work contained in this thesis was in part performed during the COVID-19 pandemic, which significantly affected the amount of time available in laboratory facilities.

Acknowledgements

I truly cannot believe that this is the end of my PhD, which has been such a fundamental part of my adult life. I have been in Prof. Susan Rosser's lab for around five years now, and it is scary to think that I will be moving on.

There have been so many people that have supported me throughout this project. Of course, I would primarily like to thank Susan for this amazing opportunity. Without her I may not have been able to stay in Edinburgh, the city I love and call home, so I am incredibly fortunate that I could pursue my research here.

I could not have asked for a better experience. Everybody in the lab has been so supportive during my time here, and I have always felt at home. I have made so many connections, and I truly hope that we will maintain these friendships.

Firstly, I'd like to thank Marcus for his constant support and advice (although he does have his off days!), as well as his assistance in brewing. Just remember that you could never do it without me! In all seriousness though, I have always thought that you are one of the most capable people that I know, and if I were to leave you with all of my (very expensive) brewing equipment, I'm sure you would make the most of it. Just make sure to save some for me! Of course, you are now moving on to a new chapter in your life as well, and I wish you and Fiona all the best.

To all of the post-docs in the lab, past and present, I would like to give my heartfelt appreciation. Christine, thanks for your unwavering support and guidance. You have always made me feel more talented than I am. I'd like to thank DJ for introducing me to the lab and demonstrating the importance of colony PCR!

I'd also like to acknowledge the other PhD students in the lab. Jess has been particularly supportive in recent weeks while I've been pulling my hair out trying to get this written! I really appreciate their kind words, especially considering that they had their own thesis to worry about. James has always been a calm, level-headed presence in my life, and I am so grateful for his counsel and companionship. I would like to thank Alex for his upbeat attitude! He definitely thinks he's the funniest person in the lab, and while I may disagree with him there, he does have his moments. Special thanks are due to Ben Baker, for being a constant reminder that I could

always be better. Finally, to all of the others in my cohort, thanks for all the good times, I just wish I could remember them!

I would of course like to thank my partner Hannah for her encouragement and reassurance over these years. I'm sure without her I would have more results, but more grey hairs! While I'm reticent to move onto the next chapter of my life, I'm so glad that we get to do it together.

I am also grateful to my family and other friends who have supported me in various ways, offering encouragement, understanding, and consolation when needed most. I would first and foremost like to thank Marek, who's genius and intellect inspire me every day. His graceful skill reminds me to be careful with lab equipment, his attention to detail keeps my scientific mind sharp. He is an ideal to which I can only aspire, unlike Patrick who is mediocre. Thank you.

I would like to thank Tjalling Siersma and Leendert Hamoen for graciously sharing the mini*Bacillus* strains mentioned in Chapter 3.

Finally, I'd like to thank Unilever for letting me take on this project. It's a lot to trust a PhD student with, and I hope that my performance was satisfactory! It was great to get to stay in Liverpool during my placement, and I thoroughly enjoyed getting to know everybody. I wish all of you the best in your future endeavours, and I hope that we'll have the chance to meet again someday!

Contents

Abstract	2
Lay Summary	3
Declaration	4
Impact of COVID-19	5
Acknowledgements.....	6
Contents	8
List of Figures	14
List of Figures	14
List of Tables.....	16
List of Equations	17
List of Abbreviations.....	18
Chapter 1 Introduction.....	20
1.1 Context and motivation	20
1.2 Detergent formulations	21
1.2.1 Surfactants	22
1.2.2 Builders	23
1.2.3 Other components.....	24
1.2.4 Enzymes	25
1.3 Protease structure and function	26
1.3.1 Protease classification	26
1.3.2 Metalloprotease clans	29
1.3.3 Inhibition by propeptide domains.....	44
1.3.4 Activation of metalloproteases.....	45
1.3.5 Further roles of propeptide domains	46
1.3.6 Metalloprotease mechanism of action	47

1.4 Bioinformatic approaches of predicting enzyme activity	48
1.4.1 The ideal candidate protease	48
1.4.2 Identification of candidate proteases	49
1.4.3 Predicting biochemical characteristics	50
1.4.4 Machine learning based methods of predicting protein characteristics	50
1.4.5 Bioinformatic tools for database mining	51
1.4.6 Prediction of protein 3D structure	53
1.5 <i>Bacillus subtilis</i> as a host for protease production	54
1.5.1 <i>B. subtilis</i> in industry	54
1.5.2 Native proteases of <i>B. subtilis</i>	55
1.5.3 Genetic competence of <i>B. subtilis</i>	55
1.5.4 Minimal <i>B. subtilis</i> strains	56
1.5.5 Protein secretion in <i>B. subtilis</i>	57
1.5.6 Improving secretion of heterologous proteins	58
1.5.7 CRISPR-Cas systems in <i>B. subtilis</i>	59
1.6 Aims	61
Chapter 2 – Materials and methods	62
2.1 Strains and plasmids	62
2.2 Growth and maintenance of bacterial cultures	62
2.2.1 General growth and storage of bacterial cultures	62
2.2.2 Visualisation of <i>B. subtilis</i> cells by microscopy	63
2.2.3 Plate reader growth curves	63
2.2.4 BioLector analysis	64
2.3 Bacterial transformations	64
2.3.1 Preparation and transformation of chemically competent <i>E. coli</i> cells	64
2.3.2 Preparation and transformation of competent <i>B. subtilis</i>	65
2.4 Chemicals	65

2.5 DNA manipulation	66
2.5.1 Purification of plasmid DNA.....	66
2.5.2 Isolation of genomic DNA from <i>B. subtilis</i>	66
2.5.3 DNA quantification.....	67
2.5.4 Oligonucleotides.....	67
2.5.5 Polymerase chain reaction (PCR)	69
2.5.6 Agarose gel electrophoresis	72
2.5.7 Purification of DNA following agarose gel electrophoresis	72
2.5.8 DNA assembly methods.....	72
2.6 DNA Sequencing	74
2.7 Gene editing in <i>B. subtilis</i>	74
2.7.1 Homologous recombination.....	74
2.7.2 CRISPR-MAD7 mediated gene editing	74
2.8 Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis (SDS-PAGE).....	76
2.8.1 Protease sample denaturation by sulphuric acid.....	76
2.8.2 Protease sample concentration by trichloroacetic acid (TCA) precipitation	76
2.8.3 Separation of protein samples by SDS-PAGE.....	77
2.8.4 Visualisation of protein by staining with InstantBlue™	77
2.8.5 Visualisation of protein by western blot	77
2.8.6 Separation of protease samples by gelatin zymography.....	78
2.9 Peptide sequencing	78
2.9.1 Identification of propeptide domains.....	78
2.9.2 Protease cleavage site identification	79
2.10 Protease activity detection.....	79
2.10.1 Qualitative detection of protease activity on skimmed milk agar	79
2.10.2 Quantification of protease activity by the azocasein assay	79

2.10.3 Quantification of protease activity by the Protazyme AK assay	79
2.11 Determination of protein stability	80
2.11.1 Determination of pH stability by azocasein assay	80
2.11.2 Determination of NaCl stability by azocasein assay	80
2.11.3 Determination of melting temperature by nano differential scanning fluorimetry (nanoDSF)	80
2.12 Protease purification techniques	81
2.12.1 Ammonium sulphate precipitation	81
2.12.2 Production and purification of protease by immobilised metal ion chromatography (IMAC)	81
2.12.3 Purification of protease by SEC	82
2.12.4 Quantification of protease concentration	82
2.13 Stain removal by tergotometer	82
2.14 <i>In silico</i> data analysis	84
Chapter 3 – <i>Bacillus subtilis</i> strain engineering	86
3.1 Introduction	86
3.2 Material & Methods	88
3.2.1 Strains and plasmids	88
3.2.2 Plasmid construction	88
3.2.3 Strain construction	89
3.3 Results	90
3.3.1 Deletion of native <i>B. subtilis</i> proteases	90
3.3.2 Improving secretion of heterologous proteases	93
3.4 Discussion	95
3.4.1 Extracellular protease deletions had no effect on transformation efficiency, growth rate or morphology	95
3.4.2 <i>prsA</i> overexpression increases secretion of heterologous protease	95
3.5 Conclusion	96

Chapter 4 – Protease identification and expression	97
4.1 Introduction	97
4.2 Material & Methods	98
4.2.1 Plasmids and strains	98
4.2.2 Dataset construction	101
4.2.3 Plasmid construction	102
4.3 Results.....	103
4.3.1 Identification of metalloprotease candidates	103
4.3.2 Protease expression	107
4.3.3 Signal peptide sequences	107
4.3.4 Qualitative evaluation of signal peptides	111
4.3.5 Quantitative evaluation of signal peptides	112
4.3.6 Effect of incubation temperature and growth medium on expression of proteases	114
4.3.7 Inducible expression of protease candidates	115
4.3.8 Expression of protease candidates at 18 °C over time.....	116
4.3.9 Protease candidate activity at temperature and pH.....	117
4.4 Discussion	118
4.4.1 Prediction accuracy of bioinformatic models	118
4.4.2 Efficiency of signal peptide is protein-dependent	119
4.5 Conclusion.....	120
Chapter 5 – Protease purification and characterisation.....	121
5.1 Introduction	121
5.2 Material & Methods	123
5.2.1 Plasmids and strains	123
5.2.2 Plasmid construction	123
5.3 Results.....	124

5.3.1 Ammonium sulphate precipitation	124
5.3.2 Purification of Pro_10	125
5.3.3 Pro_10 salt stability	127
5.3.4 Pro_10 sequence and predicted structure	128
5.3.5 Pro_10 cleavage site.....	130
5.3.6 Pro_10 C-terminal deletion retains protease activity	132
5.3.7 Alignments of Pro_10 and PrtA/B	133
5.3.8 Effect of temperature and pH on activity and stability of Pro_10	135
5.3.9 Effect of NaCl and water hardness on Pro_10 activity and stability	136
5.3.10 Effect of sequestrants on Pro_10 activity	138
5.3.11 Effect of protease inhibitors on Pro_10 activity.....	140
5.3.12 Mutations in the HEXXH motif are catalytically inactive	142
5.3.13 Removal of proteinaceous stains by Pro_10	142
5.4 Discussion	144
5.4.1 Pro_10 C-terminus is required for full protease activity	144
5.4.2 Pro_10 is active in a range of industrially relevant conditions	145
5.4.3 Pro_10 is resistant to a range of chelating agents	145
5.5 Conclusion.....	146
Chapter 6 – Conclusions and future work	147
Appendix A	155
Appendix B	157
Appendix C	158
Appendix D	160
Appendix E	161
Bibliography	165

List of Figures

List of Figures

Figure 1.1: Estimated growth of enzyme business from 1875 until 2025 and historic events, classified by industry.....	22
Figure 1.2: Distribution of sequences in the MEROPS database by protease class	28
Figure 1.3: Systems of protease classification.	29
Figure 1.4: Richardson diagrams of proteases from clan MA.....	30
Figure 1.5: Richardson diagram of protease from clan MC.....	31
Figure 1.6: Richardson diagram of protease from clan MD.....	32
Figure 1.7: Richardson diagrams of protease from clan ME.....	33
Figure 1.8: Richardson diagram of protease from clan MF.....	34
Figure 1.9: Richardson diagram of protease from clan MG.....	35
Figure 1.10: Richardson diagram of protease from clan MH.....	36
Figure 1.11: Richardson diagram of protease from clan MJ.....	37
Figure 1.12: Richardson diagram of protease from clan MM.....	38
Figure 1.13: Richardson diagrams of protease from clan MN.....	39
Figure 1.14: Richardson diagram of protease from clan MO.....	39
Figure 1.15: Richardson diagram of protease from clan MP.....	40
Figure 1.16: Richardson diagram of protease from clan MQ.....	41
Figure 1.17: Richardson diagram of protease from clan MT.....	42
Figure 1.18: Richardson diagram of protease from clan MU.....	42
Figure 1.19: Illustration of cysteine switch mechanism of inhibition and activation in the case of matrix metalloproteinases (MMPs).....	44
Figure 1.20: Possible mechanisms of proteolysis by <i>Bacillus thermoproteolyticus</i> thermolysin.....	47
Figure 1.21: Illustration of protein secretion in <i>B. subtilis</i> through the Sec pathway.	57
Figure 3.1: Transformation efficiency and growth rate of <i>B. subtilis</i> strains.....	91
Figure 3.2: Images of <i>B. subtilis</i> cells.....	92
Figure 3.3: Comparison of protease activity with and without <i>prsA</i> overexpression .	94
Figure 4.1: Workflow for SP identification.....	103
Figure 4.2: Candidate metalloproteases from nr and env_nr databases.....	105
Figure 4.3: Candidate metalloproteases from SWISSPROT database.....	106

Figure 4.4: Candidate metalloproteases from assembly database.....	106
Figure 4.5: Qualitative evaluation of SPs.	111
Figure 4.6: Quantitative evaluation of SPs.	113
Figure 4.7: Comparison of protease expression at different growth temperatures.	114
Figure 4.8: Comparison of protease expression in different growth media.....	115
Figure 4.9: Induction of protease expression by IPTG.	115
Figure 4.10: Optical density and protease activity in supernatant while growing at 18 °C.	117
Figure 4.11: Protease candidate activity at temperature and pH.	118
Figure 5.1: Ammonium sulphate precipitation of protease candidates.	124
Figure 5.2: Chromatographs for ÄKTA Purification of Pro_10 and associated electrophoresis gels	126
Figure 5.3: T _m values (°C) of Pro_10 for Durham salt screen.	127
Figure 5.4: Amino acid sequence and structure of Pro_10.....	129
Figure 5.5: Specificity matrix for Pro_10 unique cleavage sites	131
Figure 5.6: Chromatograph for ÄKTA Purification of Pro_10 ΔC-term, electrophoresis gels and comparison of protease activity with Pro_10.....	133
Figure 5.7: Pairwise structural alignments of Alphafold predicted structures and sequence alignment of Pro_10 and PrtA and PrtB from <i>M. xanthus</i>	134
Figure 5.8: Activity and stability of Pro_10 at a range of temperature and pH.....	135
Figure 5.9: Protease activity and stability across a range of NaCl concentrations and water hardness levels.....	137
Figure 5.10: Protease activity following incubation with a range of sequestrants... ..	139
Figure 5.11: Protease inhibition by a range of protease inhibitors.....	141
Figure 5.12: Protease activity of Pro_10 HEXXH active site variants.....	142
Figure 5.13: Percentage cleaning of proteinaceous stains by proteases	143

List of Tables

Table 1.1: Qualitative advantages and disadvantages of select sequestrants.....	23
Table 2.1: Stock and working concentrations of antibiotics used in this study.	63
Table 2.2: Parameters used in calculation of maximum specific growth rate (μ_{max})	64
Table 2.3: Reaction mixture components for phosphorylation and annealing of ssDNA oligonucleotides	68
Table 2.4: Reaction conditions for phosphorylation and annealing of ssDNA oligonucleotides.....	68
Table 2.5: Standard PCR reaction using Q5® High-Fidelity DNA polymerase.	69
Table 2.6: Standard thermocycling reaction conditions for use with Q5® High-Fidelity DNA polymerase.	69
Table 2.7: Standard cPCR reaction using OneTaq® Quick-Load® 2x Master Mix with Standard Buffer.	70
Table 2.8: Standard cPCR thermo-cycling reaction conditions for use with OneTaq® Quick-Load® 2x Master Mix with Standard Buffer.....	70
Table 2.9: Standard OE-PCR reaction using Q5® High-Fidelity DNA polymerase....	71
Table 2.10: Standard OE-PCR thermo-cycling reaction conditions for use with Q5® High-Fidelity DNA polymerase.	71
Table 2.11: Conditions of T4 DNA Ligase reaction.	73
Table 2.12: CRISPR-MAD7 plasmids used throughout this study.	75
Table 3.1: Strains and plasmids used Chapter 3.	88
Table 4.1: Strains and plasmids used in Chapter 4	98
Table 4.2: Proteases used as query sequences for BLAST and HMMER searches and relevant characteristics.....	104
Table 4.3: Comparison of signal peptides used to secrete heterologous Pro_10... ..	108
Table 4.4: Comparison of signal peptides used to secrete heterologous Pro_12... ..	108
Table 4.5: Comparison of signal peptides used to secrete heterologous Pro_20... ..	109
Table 4.6: Comparison of signal peptides used to secrete heterologous Pro_23... ..	110
Table 4.7: Comparison of signal peptides used to secrete heterologous Pro_26... ..	110
Table 5.1: Average values of French hardness, calcium: magnesium ratio and wash temperature for various countries supplied by Unilever.....	122
Table 5.2: Strains and plasmids used in Chapter 5.	123

List of Equations

Equation 2.1: Method of least squares	64
Equation 2.2: Calculation to determine transformation efficiency.	65
Equation 2.3: Beer-Lambert Law	67
Equation 2.4: CIE 2000 colour difference formula.	83
Equation 2.5: Formula for calculating SRI values.....	84
Equation 2.6: Formula for calculation of percentage cleaning.....	84
Equation 2.7: Formula to blank percentage cleaning by no-enzyme control.	84

List of Abbreviations

°C – degrees Celsius	EDTA – ethylenediaminetetraacetic acid
Å – angstrom	EFSA – European Food Safety Authority
A _x – absorbance at x nm	ESI-MS – electrospray ionization mass spectrometry
ADAM – a disintegrin and metalloproteinase	FDA – Food and Drug Administration
ADAMTS – a disintegrin and metalloproteinase with thrombospondin motifs	g – gram (can be combined with metric SI prefixes)
ADP – adenosine diphosphate	g – gravity
Amp – ampicillin	gDNA – genomic DNA
ANN – artificial neural network	GLDA – glutamic acid N,N-diacetic acid
Å r.m.s.d. ₉₅ – Cα root-mean-square deviation at 95% residue coverage	GRAS – generally regarded as safe
ATP – adenosine triphosphate	gRNA – guide RNA
BLAST – basic local alignment search tool	HDR – homology directed repair
BMP – bone morphogenic protein	HEDP – 1-hydroxyethylidene-1,1-diphosphonic acid
bp – base pairs (can be combined with metric SI prefixes)	HMM – hidden Markov model
Cm – chloramphenicol	HR – homologous recombination
CMC – carboxymethyl cellulose	IMAC – immobilised metal ion chromatography
cPCR – colony PCR	IPTG – isopropyl β-D-1-thiogalactopyranoside
CRISPR – clustered regularly interspersed palindromic repeats	Kan – kanamycin
Cas – CRISPR associated	L – litre (can be combined with metric SI prefixes)
CASP – Critical Assessment of Structure Prediction	LB – lysogeny broth
CFU – colony forming unit	LC-MS/MS – liquid chromatography with tandem mass spectrometry
CIE – Commission Internationale d’Eclairage or International Commission on Light	M – moles per litre (can be combined with metric SI prefixes)
Cpf – CRISPR from <i>Prevotella</i> and <i>Francisella</i>	MGDA – methylglycinediacetic acid
crRNA – CRISPR RNA	min – minute
dATP – deoxyadenosine triphosphate	MMP – matrix metalloproteinase
dCTP – deoxycytidine triphosphate	MSA – multiple sequence alignment
dDNA – donor DNA	NEB – New England Biolabs
DDT – dithiothreitol	NHEJ – non-homologous end joining
dGTP – deoxyguanosine triphosphate	Ni-NTA – nickel-nitrilotriacetic acid
DMPP – diethylenetriamine penta(methylene phosphonic acid)	NTA – nitrilotriacetic acid
DNA – deoxyribonucleic acid	nt – nucleotide
DSB – double-strand break	OD ₆₀₀ – optical density at 600 nm
DSF – differential scanning fluorimetry	OE-PCR – overlap extension PCR
dNTPs – Deoxyribonucleotide mix of bases dATP, dTTP, dGTP and dCTP	PAGE – polyacrylamide gel electrophoresis
dTTP – deoxythymidine triphosphate	PAM – protospacer adjacent motif
E – expectation (value)	PCM – phase contrast microscopy

PCR – polymerase chain reaction
PDB – protein data bank
PEG – polyethylene glycol
Pi – inorganic phosphate
pLDDT – predicted local distance difference test
PMSF – phenylmethylsulfonyl fluoride
QPS – qualified presumption of safety
RNA – ribonucleic acid
rpm – revolutions per minute
SDS – sodium dodecyl sulphate
SEC – size exclusion chromatography
SEM – standard error of the mean
s - second
sgRNA – single guide RNA
SP – signal peptide
Spec – spectinomycin
SRP – signal recognition particle
ssDNA – single-stranded DNA

SVM – support vector machine
T – score assigned to alignments during BLAST searching
TAE – tris-acetate-EDTA
Tat – twin arginine translocation
TCA – trichloroacetic acid
TLCK – tosyl-L-lysyl-chloromethane hydrochloride
T_m – melting temperature
TPCK – tosyl phenylalanyl chloromethyl ketone
UV – ultra violet
V – volts (can be combined with metric SI prefixes)
v/v – volume per volume
WT – wild-type
w/v – weight per volume

Chapter 1 Introduction

1.1 Context and motivation

Proteases are one of the largest classes of commercial enzymes worldwide, with applications in a broad range of industries. In detergents, they are used for removal of proteinaceous stains, such as blood, grass, milk, and egg stains (Maurer, 2004). In the food industry proteases are used for flavour-enhancement of fish, meat, and dairy products (Gupta et al., 2002). For the textiles industry, proteases are used in soaking and dehairing of wool fibres (Khandelwal et al., 2015; Tian et al., 2019). They have also found uses in medical applications, such as aiding in wound healing and the removal of dead tissue (Mekkes et al., 1998). Out of these, their use in detergents is greatest with regards to market volume (Maurer, 2004).

The use of proteases in detergents began in 1913 with the addition of pancreatic extract to detergent formulations by Röhms & Haas (Hede, 2020). However, trypsin is not stable in detergent formulations. As such, proteases did not become widespread in detergents until the 1960s, when subtilisins were first isolated from *Bacillus* species (Maurer, 2004). Subtilisins are a family of serine proteases with a wide range of functions, and members of this family have been identified in practically all known organisms (Siezen & Leunissen, 1997). Subtilisins have dominated the detergent protease market, with most commercially available proteases originating from this family. There is, however, evidence that metalloproteases may be more effective at removal of certain stains, particularly bloodstains (Zhang et al., 2019). The breakdown of bloodstains by subtilisins has been shown to cause redeposition, that is, the yellowing of fabric that is washed alongside bloodstained fabric. Metalloproteases have been found to be particularly effective at removal of bloodstains and the avoidance of redeposition (Unilever, 2023a).

The majority of commercially available detergent proteases are thermostable alkaline proteases that have optimal activities at temperatures between 50 °C and 70 °C, as thermostable enzymes are generally more active and stable overall (Gupta et al., 2002). Due to the current energy crisis, it is imperative that we adopt processes, both industrially and domestically, that require lower energy inputs. There is, therefore, a

need to discover novel proteases that are effective at lower washing temperatures. In addition, the use of novel enzymes with improved washing performance has the potential to reduce the carbon footprint of detergent formulations by replacing petrochemical-derived ingredients, as described in Unilever's Clean Future strategy (Unilever, 2023b).

1.2 Detergent formulations

Soaps produced from sodium or potassium-neutralised fatty acids have been in use since at least 2,200 BCE. These formed the basis of laundry products up until the Second World War, when the shortage of animal fats in Germany necessitated the production of synthetic alternatives (Bajpai & Tyagi, 2007). Initially, these synthetic detergents were based on sodium diphosphate. However, the widespread release of phosphates into the water course led to eutrophication, where algal blooms deplete oxygen in water systems, causing the death of aquatic organisms. The use of phosphates in detergents has largely been phased out in the USA and Europe, with a range of alternatives being utilised.

The composition of current detergent formulations will be described in the following sections. They are generally composed of surfactants, builders, anti-redeposition agents, fragrances, bleaching agents, and enzymes. Detergents are designed to clean two types of soils: organic and inorganic. Inorganic soils may be easily removed through the action of surfactants and builders. On the other hand, organic soils like sebum or blood, mainly comprised of organic compounds such as proteins and fats, often exhibit low solubility in water and strong adhesion to textile surfaces. This adhesion makes it challenging to thoroughly remove this type of soiling using surfactants. The use of enzymes in detergents can aid in cleaning through hydrolysis of organic stains (Zhang et al., 2014), which has led to the detergents market forming the bulk of the global enzyme industry (Figure 1.1).

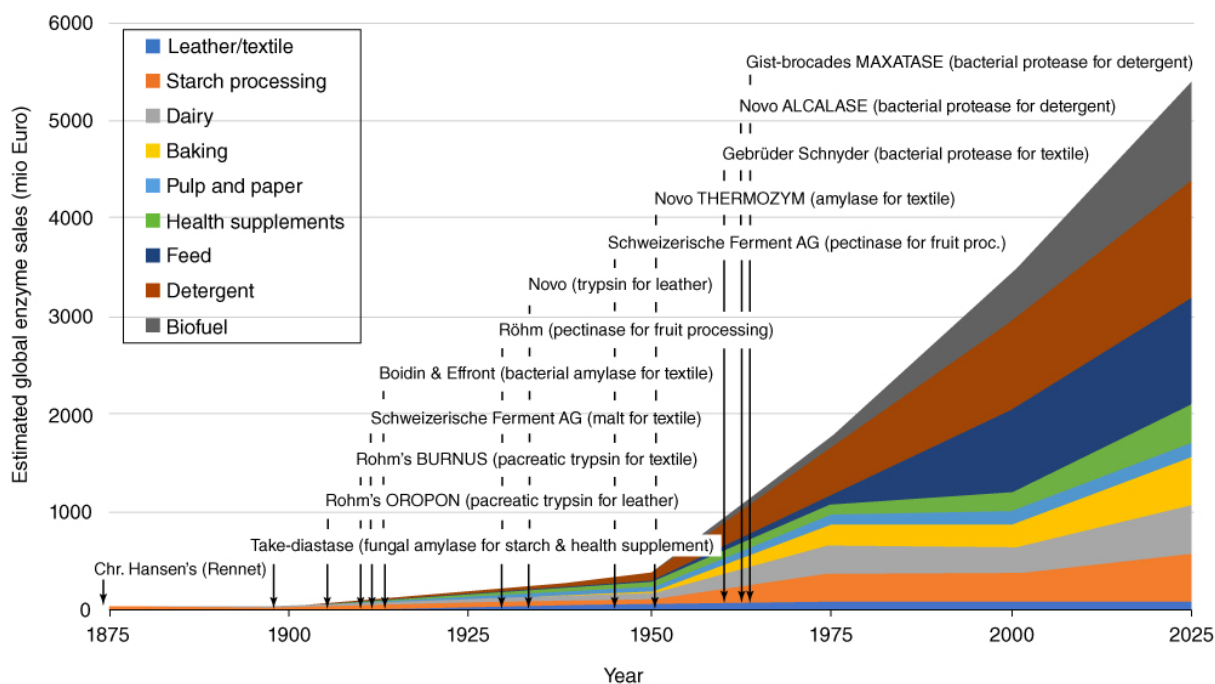


Figure 1.1: Estimated growth of enzyme business from 1875 until 2025 and historic events, classified by industry (Vogel & May, 2019).

1.2.1 Surfactants

Surfactants are amphiphilic compounds responsible for the majority of stain removal by detergents (Cheng et al., 2020), comprising around 15-40% of the total detergent formulation (Scheibel, 2004). They may be classified as anionic, cationic, and non-ionic, depending on the nature of the polar head group. Generally, anionic surfactants are efficient in soil removal but are susceptible to the influence of multivalent ions found in hard water. This is due to sequestration by divalent cations such as calcium and magnesium ions, causing precipitation.

On the other hand, non-ionic surfactants exhibit excellent solubility and are largely unaffected by hard water, yet they tend to be less effective than anionic surfactants in soil removal. Cationic surfactants are typically employed as fabric softeners, as they reduce friction between fabric fibres and skin (Yu et al., 2008).

Combinations of these surfactants have proven to be effective at stain removal, however, 75-80% of these are derived from petrochemical sources and are generally toxic and not biodegradable (Inès et al., 2023; Rebello et al., 2014). There have been recent efforts to switch to more renewable surfactants, particularly those of microbial origin. These generally anionic surfactants include emulsane produced by *Acinetobacter calcoaceticus*, surfactin of *Bacillus subtilis*, sophorolipids of *Candida*

bombicola and rhamnolipids produced by *Pseudomonas aeruginosa*, which have shown promise in consumer applications (Jadhav et al., 2019; Khaje Bafghi & Fazaelipour, 2012; Unilever, 2022).

1.2.2 Builders

Builders enhance the cleaning capabilities of surfactants. They may be sequestering (sequestrants, or chelants) or precipitating builders.

Sequestrants are organic chemicals that can bind to metal ions to create stable and typically soluble complexes, preventing metal ions from reacting with other species. Sequestrants possess two or more electron donor atoms that form coordinate bonds with a single metal ion, resulting in the formation of one or more rings containing the metal ion. Organic groups that establish strong bonds with metal ions include carboxylic acids, amines, alcohols, phosphonic acids, and phosphates (Anastas et al., 2014).

Sequestrants are added to detergent formulations to counteract the effects of water hardness through chelation of calcium and magnesium ions, preventing buildup of limescale. They also chelate transition metal ions that are present in lower concentrations but can reduce effectiveness of bleaching agents and contribute to stains (Anastas et al., 2014). Finally, sequestrants aid in dispersal of soils, preventing stain redeposition.

Initially, phosphate-based sequestrants (e.g. polyphosphonates) formed the basis of sequestering builders, however these are due to be replaced due to their action in eutrophication (Rott et al., 2018). Examples of polyphosphonates are 1-hydroxyethylidene-1,1-diphosphonic acid (HEDP or Dequest 2010) and diethylenetriamine penta(methylene phosphonic acid) (DMPP or Dequest 2066), both of which are still used in detergent formulations and are derived from petrochemical sources (Asemave, 2018).

Alternative chelating agents have been explored, initially the aminocarboxylate ethylenediaminetetraacetic acid (EDTA) although its use is now controlled due to a lack of biodegradability, with levels up to 12 mg/L being found in European bodies of water (Koodynska, 2011; Oviedo & Rodríguez, 2003). Carboxylates such as citrate may be utilised, although they are weaker sequestrants. Nitrilotriacetic acid (NTA) is a biodegradable aminocarboxylate that could be used in detergents, however it has

been found to have carcinogenic effects (Nesslany et al., 2008). NTA variants have been explored as alternative sequestrants, including MGDA (methylglycinediacetic acid) and GLDA (glutamic acid N,N-diacetic acid). These are also biodegradable and have similar metal binding strengths to NTA, but do not exhibit carcinogenic activity (Anastas et al., 2014). Each of these sequestrants have different binding strengths and preferences for metal ions. Advantages and disadvantages of select sequestrants can be seen in Table 1.1.

Table 1.1: Qualitative advantages and disadvantages of select sequestrants (Anastas et al., 2014).

Sequestrant	Advantage	Disadvantage
EDTA	Strong binding of all metals	Less selective Poorly biodegradable
DMPP	Strong binding of all metals	Less selective Very poorly biodegradable
HEDP	Ca ²⁺ /Mg ²⁺ chelation	Lower binding of transition metals
NTA	Ca ²⁺ /Mg ²⁺ chelation Biodegradable	Lower binding of transition metals Carcinogen
MGDA, GLDA, citrate	Ca ²⁺ /Mg ²⁺ chelation Readily biodegradable	Weaker for transition metals

Conversely, precipitating builders such as sodium carbonate or sodium silicate remove water hardness ions through a non-reversible interaction to form an insoluble precipitant. Precipitating builders are particularly effective against calcium ions (Bajpai & Tyagi, 2007).

1.2.3 Other components

Other adjuncts included in detergent formulations are anti-redeposition agents such as carboxymethyl cellulose (CMC), which prevents lifted stains from redepositing onto cleaned fabrics. Corrosion inhibitors are added to slow corrosion of washing machine components. Solvents such as alcohols are included to aid in solubility of detergent compounds and lower the melting point to prevent freezing during storage. Bleaching agents may be added, including oxygen bleaches such as hydrogen peroxide, to aid in stain removal from white cloth. Fragrance compounds are another

additive, used to cover the smell of detergent compounds and impart a fresh scent to washed clothes (Bajpai & Tyagi, 2007).

1.2.4 Enzymes

Enzymes used in detergents include amylases, lipases, cellulases and proteases. Proteases and amylases form the greatest market share of the global detergent enzyme industry at 60% and 25%, respectively (Zhang et al., 2014). These provide a variety of functions as adjuncts to surfactants and have the advantage of being renewable and biodegradable. An important feature of all of these detergent enzymes is stability in the presence of other components of detergent formulations, including surfactants, bleaching agents and sequestrants.

The α -amylases of *Bacillus* species are commonly added to detergents. These hydrolases act on α -1,4-glycosidic bonds of starch and related molecules, and are mostly metalloenzymes that require calcium ions for activity and stability (Pandey et al., 2000). In detergent processes, they break down starchy stains such as residues of potato, gravy, chocolate, and custard (Souza & Magalhães, 2010).

Lipases facilitate the breakdown of triacylglycerides into diacylglycerides, monoacylglycerides, free fatty acids, and glycerol at the boundary between oil and water (Nerurkar et al., 2013). They are included in detergent formulations for removal of fatty stains such as butter, oil, sauces, and stains from human sebum. Such stains are normally difficult to remove at low temperatures with conventional detergents, but the addition of lipases can help with removal in cold conditions (Hasan et al., 2010).

Cellulases used in detergents are primarily produced by *Bacillus* species. As hydrolases that cleave the glucosidic linkages of cellulose and related cello-oligosaccharides, they are produced in nature to degrade plant cell walls (Doi & Kosugi, 2004). Cellulases alter the composition of cellulose fibers in cotton and cotton blends. When incorporated into a detergent, this modification leads to enhanced colour vibrancy, increased softness, and improved soil removal (Hasan et al., 2010; Niyonzima, 2019).

As proteases hydrolyse peptide bonds in proteins, they are used for degradation of proteinaceous stains, such as blood, gravy, grass, and chocolate stains. One

consideration in the use of proteases in detergents is the cross-compatibility with other detergent enzymes, as they will degrade, for example, lipases during storage. This requires a careful balance of which enzymes are included in formulation to maximise stain removal and prevent unwanted proteolysis.

Proteases were the first enzymes to be included in detergent formulations, starting with the addition of trypsin in 1913 (Hede, 2020), and have dominated the global enzyme market (Figure 1.1). The protease industry has since expanded, and has mostly focused on the subtilisins, a family of serine proteases, proceeding with the launch of Subtilisin Carlsberg, marketed as Aldolase® (Novozymes, Denmark) in 1963 and Maxatase® in 1965 by Gist-Brocades (now DSM, Netherlands). These were both produced by *Bacillus licheniformis* (Beck & Strobel, 1988), and the majority of detergent proteases since have been subtilisins from *Bacillus* species. This includes the more recent Savinase® (Novozymes, Denmark) from *Bacillus lentus* (Mikkelsen et al., 2012) and the Savinase® variant Carnival Ecity™ 16 L (Novozymes, Denmark), which is the current gold standard protease (Mole et al., 2020).

1.3 Protease structure and function

1.3.1 Protease classification

Proteases, or peptidases, are a class of hydrolase that cleave peptide bonds in proteins. They are ubiquitous in Nature, accounting for at least 2% of any known genome (Carnevale et al., 2006), with over 1.1 million known and predicted protease sequences (Rawlings et al., 2010). The possible roles of proteases are incredibly varied. For example, in *Homo sapiens*, proteases facilitate myriad functions, including degradation of proteins by the proteasome (Kim et al., 2023; Tanaka, 2009), and associated de-ubiquitination to remove ubiquitin tags that target proteins to the proteasome (Lee et al., 2010; Lee et al., 2011). The matrix metalloproteinases (MMPs) are associated with the extracellular matrix and are involved in angiogenesis, promoting bone growth and development, and the degradation of collagen, among others (Cabral-Pacheco et al., 2020). Similarly, all mammals have 19 distinct ADAMTS enzymes (A Disintegrin and Metalloproteinase with Thrombospondin motifs), which are involved in development (Kelwick et al., 2015). In

plants, proteases are involved in photomorphogenesis, embryogenesis, circadian rhythms, flower and trichome development, and hormone signalling (Vierstra, 2003).

There are many possible functions in microorganisms as well. They may play a role in pathogenesis through degradation of host proteins, such as collagen (Fricke et al., 2001). In *Bacillus subtilis* there are several proteases involved in secretion of proteins. Signal peptidases such as SppA cleave signal peptides (SPs) from secreted proteins, while the quality control proteases HtrA/HtrB and WprA degrade misfolded proteins in the secretory pathway. They are also involved in feeding through degradation of proteins in the extracellular space (Huddy & Coyne, 2015; Wu & Chen, 2011).

Proteases are classified by the peptidase database, MEROPS (Rawlings et al., 2010), according to their catalytic type: aspartic, cysteine, glutamic, threonine, asparagine, serine, or metallo, based on the functional group present at their active site (Figure 1.2). Metalloproteases encompass a wide and diverse array of sequences, with functions relating to cell growth, death, signalling and secretion, among others. They are termed metalloproteases as they use a metal ion to coordinate a water molecule that performs a nucleophilic attack on the substrate peptide bond (Figure 1.5) (Klein et al., 2018). This is most commonly a zinc ion, although cobalt (Fabry et al., 2006), iron (Alvarado et al., 2009), nickel (Mazzei et al., 2020) and manganese (Graham et al., 2005) have also been observed in the catalytic site, and it is also possible to substitute the zinc ion with other metal ions, such as cobalt, to restore activity (Kuzuya & Inouye, 2001). Structural calcium-binding domains are also common in metalloproteases, as calcium can have a large role in their stabilisation. For example, the metalloprotease thermolysin binds four calcium ions and mutation of calcium binding sites significantly reduced thermostability (Veltman et al., 1996, 1998).

The MEROPS database (Rawlings et al., 2010) currently lists over 390,000 metalloproteases with wide variation in structure, metal binding and catalytic mechanism, split into 16 clans and 73 families based on conserved residues and mechanisms of action. These may then be further divided, in some cases, into subclans and subfamilies (Figure 1.3).

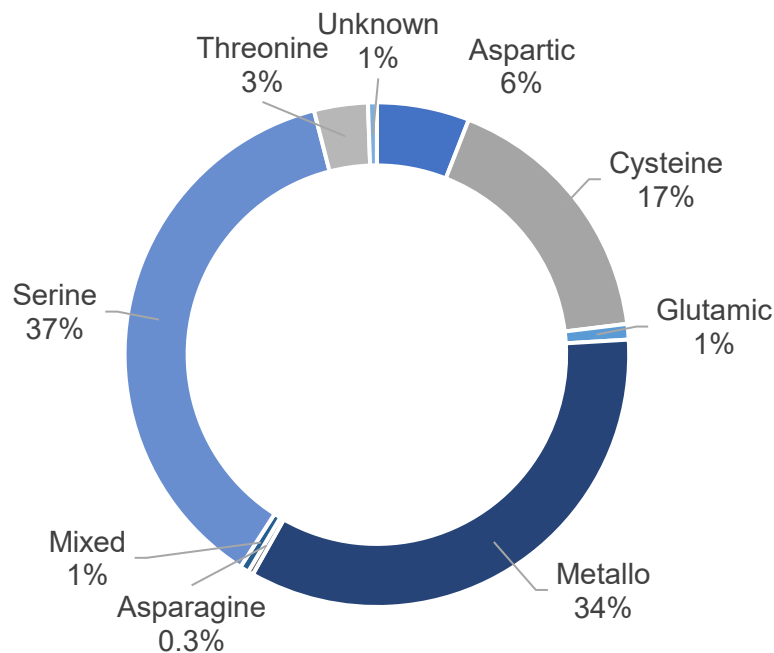


Figure 1.2: Distribution of sequences in the MEROPS database by protease class (Rawlings et al., 2010).

Metalloproteases can also be grouped based on the number of catalytic metal ions. Over half are monometallic (binding one catalytic metal ion) although 44.5% of metalloproteases are bimetallic (Cheggour et al., 2000; Kim et al., 2013; Rawlings et al., 2010). All metalloproteases that use ions other than zinc are bimetallic, although there are also bimetallic zinc metalloproteases (Rawlings & Barrett, 2013). The distribution of mono- and bimetallic proteases is split across clans (Figure 1.3).

Proteases may be further classified as exopeptidases, which act on the terminal amino acids of a substrate protein, or endopeptidases, which cleave within the substrate (Figure 1.3). Exopeptidases may then be split into aminopeptidases and carboxypeptidases, which cleave at the N- or C-terminus of the substrate, respectively. There are also di- and tripeptidases, which target di- and tripeptides. The primary role of both metalloexo- and metalloendopeptidases is to facilitate the hydrolysis of peptide bonds by catalysing the incorporation of a water molecule. As a result of this, the outcome is the formation of a protonated amine and an ionized carboxylate group. It's worth noting that the energy needed for this process may exhibit minor variations depending on the specific type of peptide bond, owing to slight differences in the pKa values of the resulting amines produced by amino-, carboxy-, and endopeptidases (Auld, 2013).

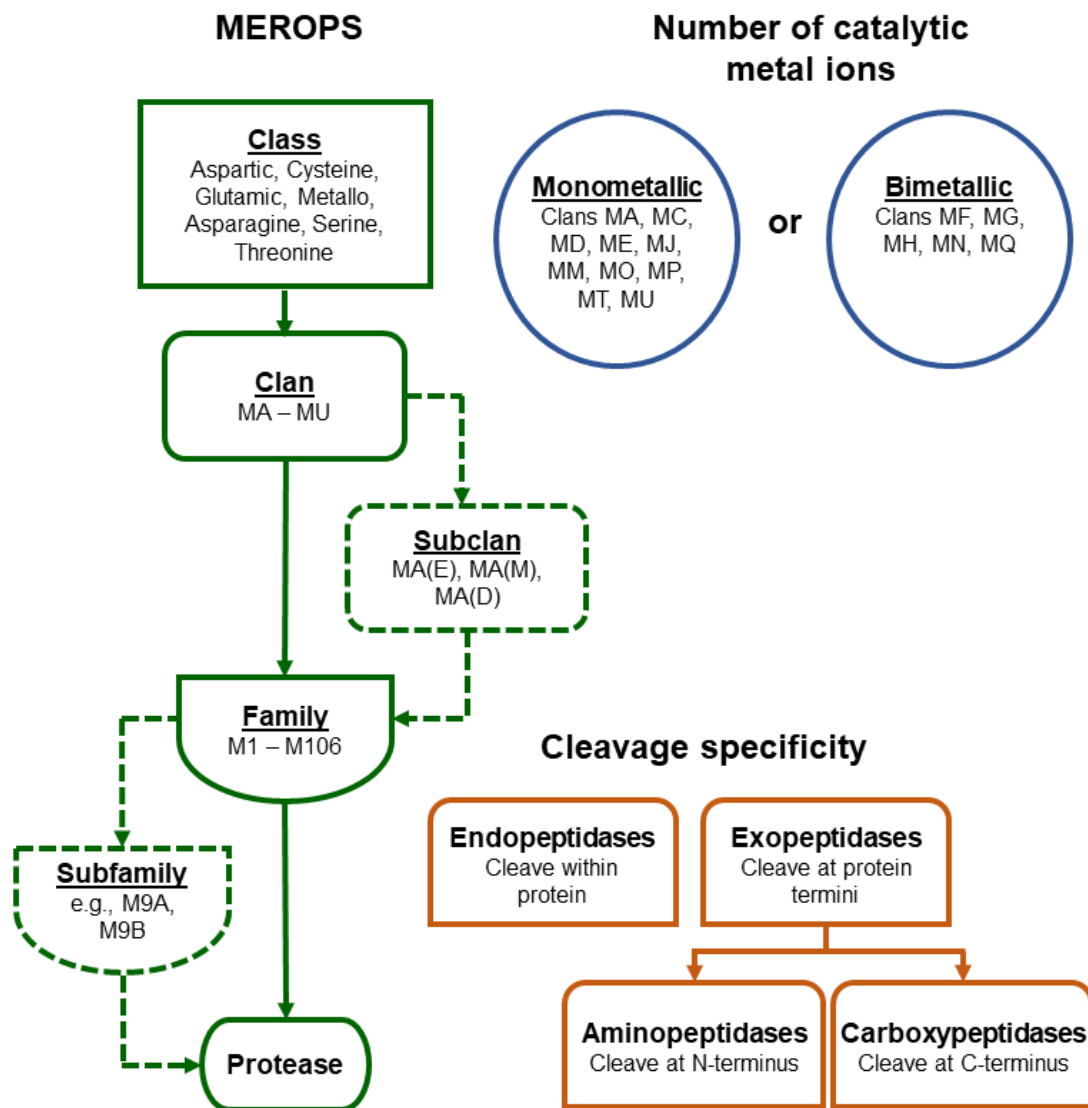


Figure 1.3: Systems of protease classification.

1.3.2 Metalloprotease clans

The 16 clans of metalloproteases are classified by different structures, mechanisms of action and conserved motifs (where the zinc-coordinating ligands are highlighted in bold).

Clan MA (the zincin clan) is the most widespread and best characterised clan of metalloproteases, comprising 49 of the 73 known families, and is characterised by the **HEXXH** active site motif (in which X denotes any amino acid). The HEXXH motif may be further defined as: Xaa-Xbb-Xcc-**His**-Glu-Xbb-Xbb-**His**-Xbb-Xdd, where Xaa is hydrophobic or threonine, Xbb is uncharged, Xcc is any amino acid except proline, and Xdd is hydrophobic. This is known as the Jongeneel consensus (Jongeneel et al., 1989). The two histidine residues of the HEXXH active site are the first two zinc-

coordinating ligands. Clan MA may be further divided into the gluzincin, metzincin and aspzincin subclans, by the third zinc-coordinating ligand (glutamate, histidine/aspartate, and aspartate, respectively). The fourth zinc ligand is a water molecule, which is used to perform a nucleophilic attack on the protease substrate. The HEXXH active site is found on an α -helix, as is the third zinc ligand, such that a turn is necessary between the helices to bring these ligands together (Figure 1.4). The gluzincins – MA(E) – and metzincins – MA(M) – are the most studied of this clan, and each shows a different mechanism of action. Clan MA(E) contains both endo- and exoproteases, while clan MA(M) is composed entirely of endoproteases (Rawlings & Barrett, 2013). Clan MA(M) also has an extended active site motif: **HEXXHXXGXX(H/D)**, and may be characterised by a strictly-conserved “Met-turn”, in which a methionine forms a hydrophobic cushion for the active site zinc ion (Gomis-Rüth, 2003).

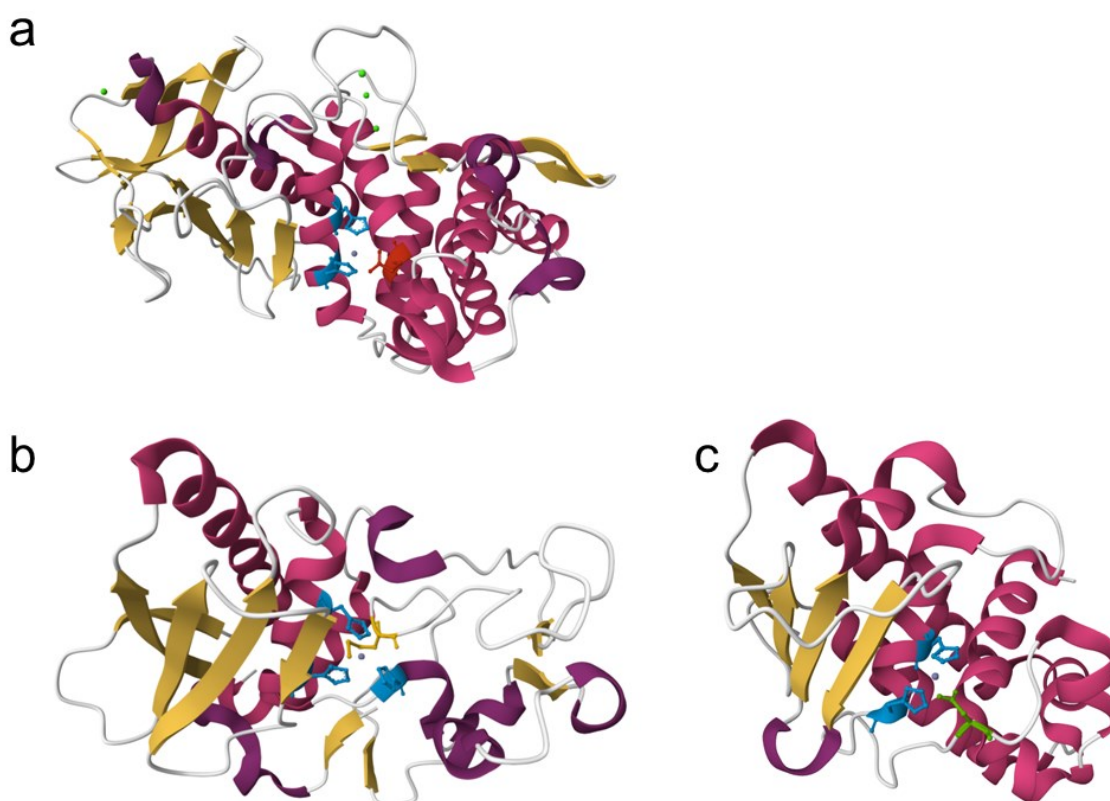


Figure 1.4: Richardson diagrams of proteases from clan MA coloured according to secondary structure, with zinc ligands shown in ball and stick representation (histidine residues coloured in blue, glutamate residues coloured in red, methionine residues coloured in yellow and aspartate residues coloured in green, zinc ion shown in grey, calcium ions shown in green). a) Structure of the gluzincin thermolysin from *Bacillus thermoproteolyticus* (PDB ID: 1FJQ) (English, Groom & Hubbard, 2001). b) Structure of the metzincin astacin

from *Astacus astacus* (PDB ID: 1AST) (Bode et al., 1992). c) Structure of the aspzincin deuterolysin from *Aspergillus oryzae* (PDB ID: 1EB6) (McAuley et al., 2001).

Clan MC contains carboxypeptidases of family M14 that contain a zinc ion coordinated by two histidine residues, a glutamate residue, and a water molecule. The active site is composed of a **His-Xaa-Xaa-Glu** motif, with the third zinc-coordinating histidine 103 to 143 residues C-terminal to the motif (Figure 1.5). The active sites contain a fold, although this is not similar to that of the well-characterised proteases thermolysin, from the MA(E) clan, or astacin, from the MA(M) clan. It does show some similarity to the folds of the bimetallic proteases leucyl aminopeptidase and *Vibrio* aminopeptidase, although the zinc ligands are not conserved between these and as such, they belong to different clans (Rawlings & Barrett, 2013).

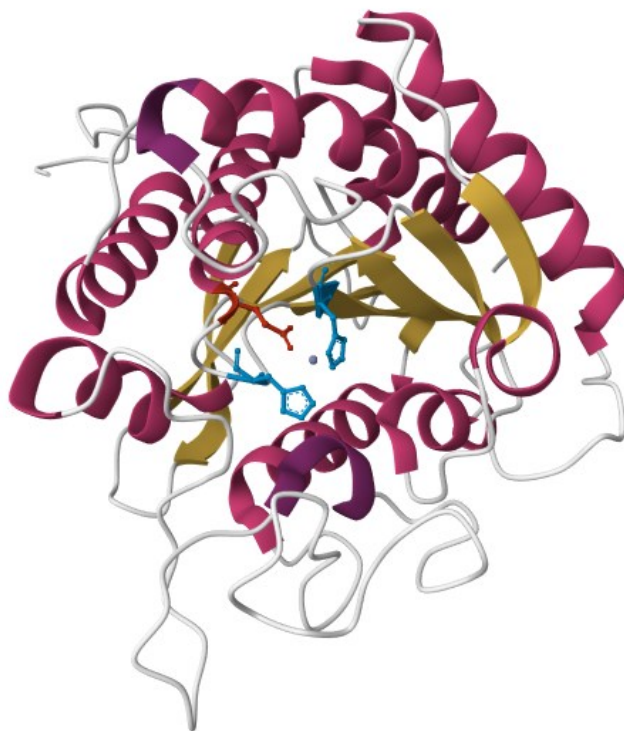


Figure 1.5: Richardson diagram of *Homo sapiens* carboxypeptidase A1 (PDB ID: 2V77) (Pallarès et al., 2008) from clan MC coloured according to secondary structure, with zinc ligands shown in ball and stick representation (histidine residues coloured in blue, glutamate residue coloured in red, zinc ion shown in grey).

Clan MD is composed of the carboxypeptidases in family M15 and the endopeptidases of family M74, both of which have functions related to bacterial cell wall synthesis and lysis. The four zinc-coordinating ligands are two histidine

residues, an aspartate residue, and a water molecule. The active site residues are as follows: **His**-(X)₆-**Asp**, with the second zinc-coordinating histidine being found 34–54 residues C-terminal to this motif, in the sequence Asp/Glu-X-X-**His** (Figure 1.6). The aspartate or glutamate in this second motif has been shown to be necessary for catalysis by site-directed mutagenesis (McCafferty et al., 1997).

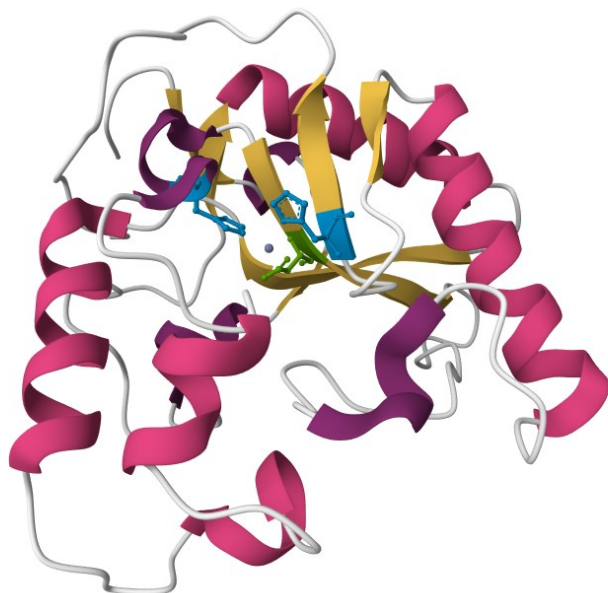


Figure 1.6: Richardson diagram of *Enterococcus faecium* VanX (PDB ID: 1R44) (Bussiere et al., 1998) from clan MD coloured according to secondary structure, with zinc ligands shown in ball and stick representation (histidine residues coloured in blue, aspartate residue coloured in green, zinc ion shown in grey).

The proteases of clan ME contain a **HXXEH** motif, the reverse of clan MA. The two histidine residues in this motif coordinate the zinc ion, along with a glutamate residue C-terminal to the motif (Figure 1.7), although the glutamate within the HXXEH motif has been shown to have a function in catalysis (Becker & Roth, 1992).

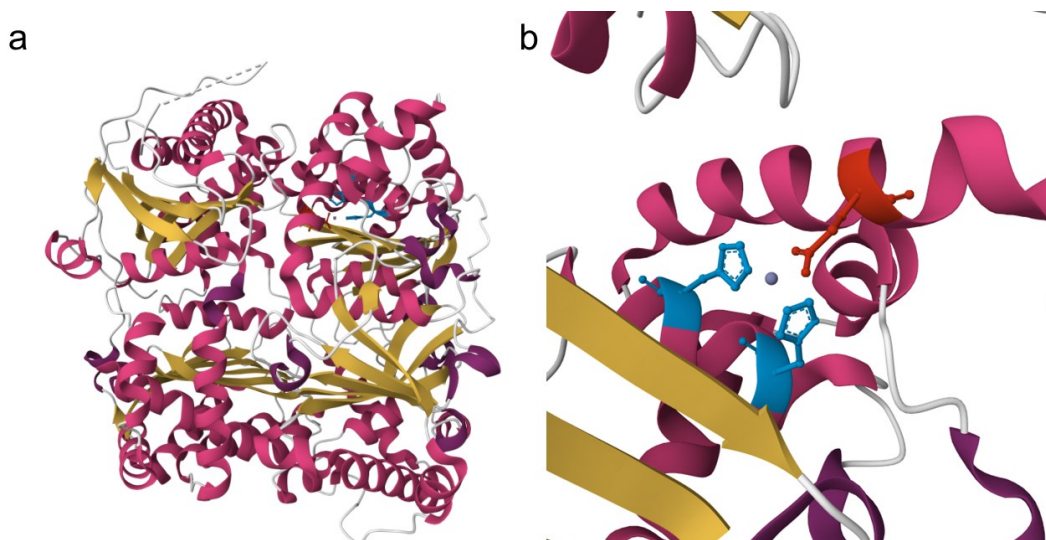


Figure 1.7 Richardson diagrams of *Homo sapiens* insulin-degrading enzyme (PDB ID: 2JG4) (Im et al., 2007) from clan ME coloured according to secondary structure, with zinc ligands shown in ball and stick representation (histidine residues coloured in blue, aspartate residue coloured in green, zinc ion shown in grey). a) full protein. b) Close up of the active site.

Clan MF is the first of the clans to be comprised of only bimetallic metalloproteases, containing cocatalytic metal ions. For all metalloproteases that use cocatalytic metal ions, five residues, instead of six, act as ligands, and one residue ligates both metal ions. With regards to cocatalytic metal ions, there is less preference for histidine as a ligand, with the ratio of aspartate and glutamate to histidine being three to one in these sites, compared to a ratio of two histidine residues to one glutamate or aspartate for metalloproteases as a whole (Auld, 2013). While monometallic proteases can be exopeptidases or endopeptidases, all known bimetallic metalloproteases are exopeptidases. HmrA from *Staphylococcus aureus* was found by Botelho et al. (2011) to be an endopeptidase, however this has since been shown to be a carboxypeptidase (Jamdar et al., 2015). Clan MF is no exception, as it contains only family M17, the leucyl aminopeptidases (Figure 1.8), which are mostly aminopeptidases. The zinc ligands in leucyl aminopeptidase are three aspartate residues, a lysine residue, and a glutamate residue, with no histidine residues seen as zinc ligands or catalytic residues, which is unusual for metalloproteases (Rawlings & Barrett, 2013). The motifs in which zinc ligands occur in this clan are as follows: (Val/Ile)-Gly-**Lys**-(Gly/Ser)-Xaa-(Thr/Ile/Val)-Xbb-**Asp**-(Ser/Thr/Ala)-Gly-Gly, in which Xaa is aliphatic and hydrophobic and Xbb is mostly aromatic and hydrophobic; Met-Xaa-Xbb-**Asp**-(Met/Lys)-Xbb-Gly-(Ala/Ser/Gly)-(Ala-Gly), where Xaa charged and

Xbb is uncharged, and Asn-Thr-**Asp**-Ala-**Glu**-Gly-Arg-Leu. The final motif is well conserved in family M17. Interestingly, while this clan is widespread in bacteria and eukaryotes, with most genomes containing one homologue, plants have two homologues, one encoded in the nucleus and the other in the chloroplast (Rawlings & Barrett, 2013).



Figure 1.8: Richardson diagram of *Bos taurus* leucyl aminopeptidase (PDB ID: 1BPN) (Kim & Lipscomb, 1993) from clan MF coloured according to secondary structure, with zinc ligands shown in ball and stick representation (aspartate residues coloured in green, glutamate residue coloured in red, lysine residue coloured in pink, zinc ions shown in grey).

The proteases of clan MG all contain a “pita-bread” fold, where the active site may be found between two domains that are structurally similar and thought to be derived from an ancient gene duplication (Figure 1.9) (Bazan et al., 199). All proteases from this clan are also exopeptidases that require cocatalytic metal ions, although in this case cobalt or manganese are used almost exclusively. One notable exception is the methionyl aminopeptidase from *Pyrococcus furiosus*, a hyperthermophilic archaeon, which binds iron. This is thought to be an adaptation to high temperature activity (Meng et al., 2002). There is only one associated family in clan MG (M24) containing aminopeptidases or dipeptidases, that is, peptidases that cleave the bond between dipeptides to release individual amino acids (Myara et al., 1994). Most proteases of this family are cytosolic and do not require proteolytic activation. Family M24 is one of only two ubiquitous peptidase families, being found in all genomes sequenced to

date (Rawlings & Barrett, 2010). The family may be split into two subfamilies: M24A which contains the methyl aminopeptidases and M24B which contains X-Pro aminopeptidases and dipeptidases, each with different active site motifs. X-Pro aminopeptidases cleave any amino acid that is bound to a proline residue, including in those of di- and tripeptides (Yaron & Mlynar, 1968). The metal binding ligands are the same for both for both subfamilies, however, being two aspartate residues, a histidine and two glutamates (Rawlings et al., 2010).

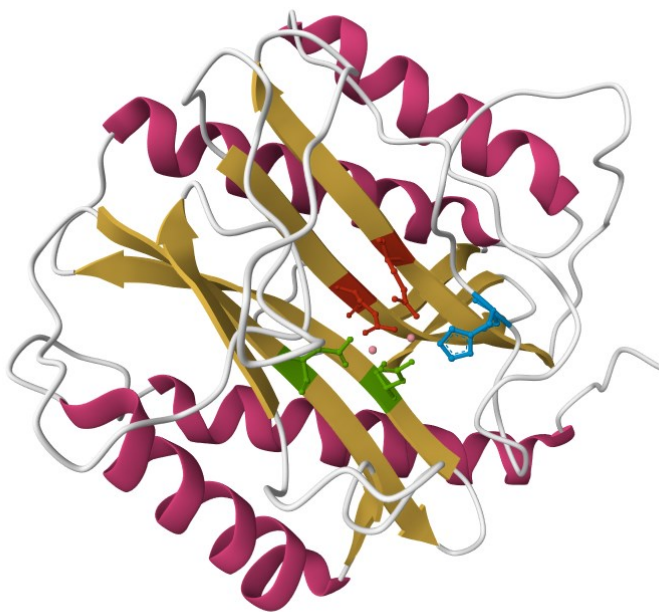


Figure 1.9: Richardson diagram of *Escherichia coli* methionyl aminopeptidase (PDB ID: 1MAT) (Roderick & Matthews, 1993) from clan MG coloured according to secondary structure, with zinc ligands shown in ball and stick representation (aspartate residues coloured in green, glutamate residues coloured in red, histidine residue coloured in blue, cobalt ions shown in grey).

Clan MH contains a diverse range of proteases from families M18, M20 and M42, all of which are cocatalytic zinc-dependent exopeptidases. Families M20, M28 and M42 display similar protein folds (Figure 1.10), although folding pathways differ for some members of this clan (Murzin, 1998). M18 is included in this clan due to conservation of probable active site residues that match other families in clan MH. The zinc ligands in clan MH are varied, and as such there is no consensus motif.

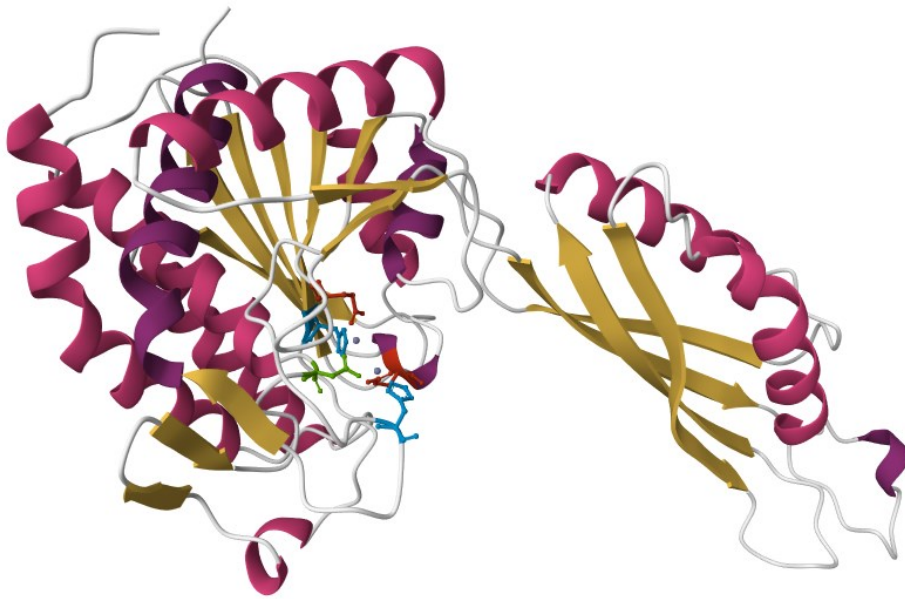


Figure 1.10: Richardson diagram of *Pseudomonas* sp. RS-16 carboxypeptidase G2 (PDB ID: 1CG2) (Rowse et al., 1997) from clan MH coloured according to secondary structure, with zinc ligands shown in ball and stick representation (aspartate residue coloured in green, glutamate residues coloured in red, histidine residues coloured in blue, zinc ions shown in grey).

Metalloprotease clan MJ consists of two families of bimetallic dipeptidases, all of which contain two cocatalytic zinc ions. They may be defined by what they are not, in that they are not subject to restrictions common to other dipeptidases. This includes membrane dipeptidase (Figure 1.11a), which can cleave a variety of dipeptides regardless of whether the C-terminal residue is in the D or L conformation (Hooper, 2013). Non-peptidase enzymes are also found in this clan due to similarities in structure and catalytic mechanism, such as *Klebsiella* urease (Figure 1.11b).

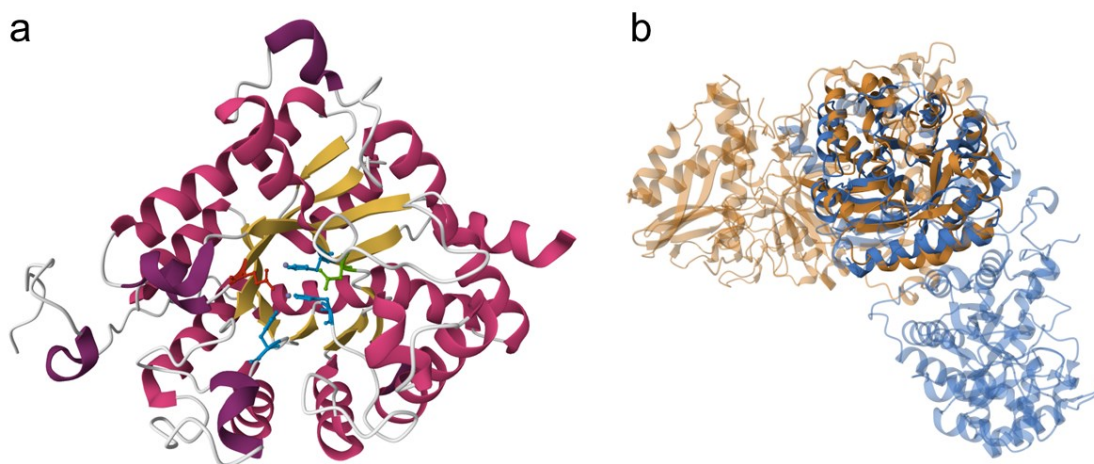


Figure 1.11: a) Richardson diagram of *Homo sapiens* membrane dipeptidase (PDB ID: 1ITQ) (Nitanai et al., 2002) from clan MJ coloured according to secondary structure, with zinc ligands shown in ball and stick representation (aspartate residue coloured in green, glutamate residue coloured in red, histidine residues coloured in blue, zinc ions shown in grey). b) Pairwise structural alignment of 3D structures of *Homo sapiens* membrane dipeptidase (PDB ID: 1ITQ, shown in blue) and *Klebsiella aerogenes* urease subunit A (PDB ID: 1A5K, shown in orange) (Pearson et al., 1998). Aligned residues are shown in solid colour, unaligned residues are faded.

Clan MM contains membrane-bound proteases with the active site located in the membrane. They are involved in regulation of gene expression through hydrolysis of transcription factors. While they all contain the **HEXXH** motif characteristic of clan MA, they are distinguished by a unique protein fold (Figure 1.12), whereby water is allowed access to the zinc ion following a conformational change where two helices move to expose a channel opening to the cytoplasm (Feng et al., 2007). The zinc ion in members of this clan is ligated by the two histidine residues in the HEXXH motif and an aspartate up to 300 residues C-terminal to this.

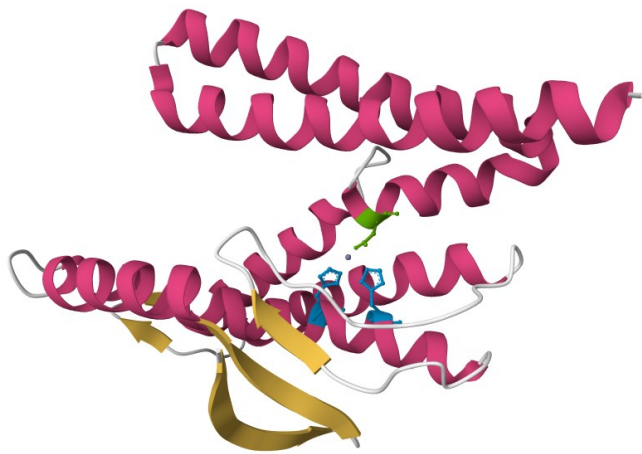


Figure 1.12: Richardson diagram of *Methanocaldococcus jannaschii* Site-2 protease (PDB ID: 3B4R) (Feng et al., 2007) from clan MM coloured according to secondary structure, with zinc ligands shown in ball and stick representation (aspartate residue coloured in green, histidine residues coloured in blue, zinc ions shown in grey).

Clan MN contains only one protease: the bimetallic aminopeptidase DppA from *Bacillus subtilis*, belonging to family M55, which cleaves the sequences d-Ala-d-Ala and d-Ala-Gly-Gly (Cheggour et al., 2000). This is an example of a self-compartmentalising protease, such as the threonine protease proteasomes in family T1 and Clp peptidases in the serine protease family S14 (Lupas et al., 1997). DppA forms a homodecamer composed of two homopentamers that form a ring, stacked into a barrel (Figure 1.13), that is stabilised without disulphide bonds. The central cavity of each ring contains the five active sites arranged in a zig-zag pattern, such that the active sites are shielded from solvent (Remaut et al., 2001). The metal ion ligands of DppA are Asp8, Glu10 and His60 for the first zinc ion, and Asp8, His104 and Glu133 for the second ion. The active site motif Ser-Xaa-**Asp**-Xaa-**Glu**-(Gly/Val) is conserved within the family M55.

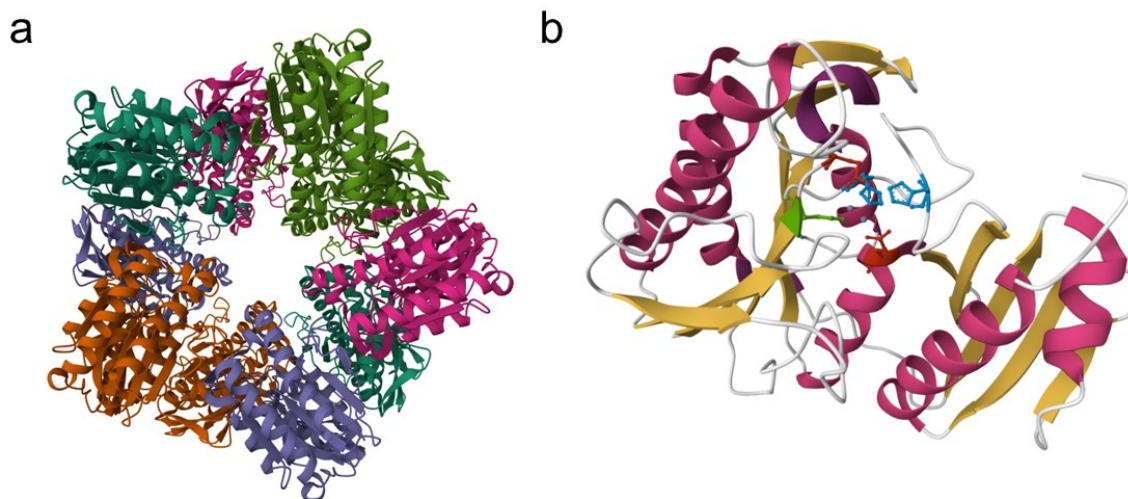


Figure 1.13: Richardson diagrams of *Bacillus subtilis* DppA (PDB ID: 1HI9) (Remaut et al., 2001) from clan MN coloured according to secondary structure. a) Homodecameric complex. b) Single subunit, with zinc ligands shown in ball and stick representation (aspartate residue coloured in green, glutamate residues coloured in red, histidine residues coloured in blue, zinc ions shown in grey).

Clan MO is composed of one family: the monometallic zinc proteases of family M23 (Figure 1.14). These are utilised by certain bacteria to break down cell walls of other bacteria (Ahmed et al., 2003). Members of this family have zinc ligands in the motifs $H(X)_3-6D$ and $H/E-X-H$, where the zinc ligands are the histidine and aspartate of the first motif and the second histidine of the second motif, with the first histidine or glutamate in the second motif likely being a catalytic residue (Odintsov et al., 2004).

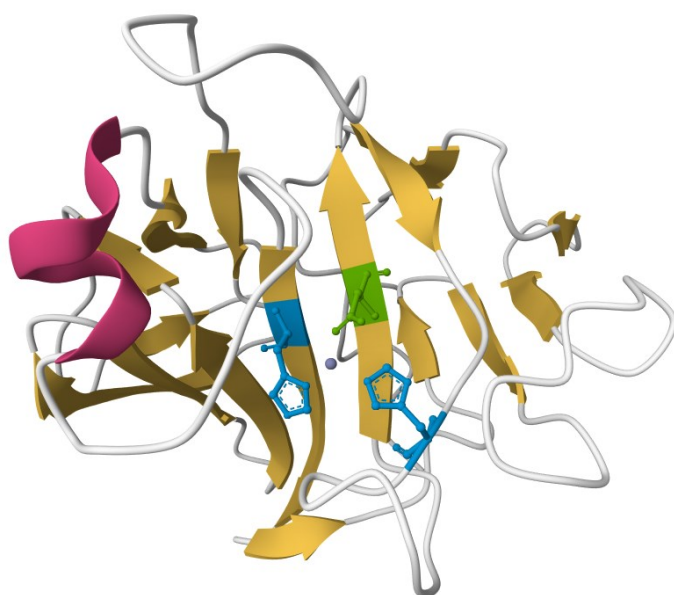


Figure 1.14: Richardson diagram of *Pseudomonas aeruginosa* LasA (PDB ID: 3IT5) (Spencer et al., 2010) from clan MO coloured according to secondary structure, with zinc

ligands shown in ball and stick representation (aspartate residue coloured in green, histidine residues coloured in blue, zinc ion shown in grey).

The clan MP contains zinc monometallic proteases of family M67, the only metalloproteases involved in deubiquitination, a role most commonly attributed to cysteine proteases. One of the best characterised examples from this family is RPN11 (Figure 1.15), which is one of three deubiquitinases associated with the mammalian proteasome, of which the other two are cysteine proteases (Lee et al., 2010; Verma et al., 2002; Yao et al., 2006). The proteasome's function is to degrade and recycle unneeded or damaged proteins, which are first fused with a ubiquitin tag. The two associated cysteine proteases both function to regulate ubiquitination by trimming ubiquitin from protein substrates to prevent their degradation (Lee et al., 2010; Lee et al., 2011). RPN11, on the other hand, cleaves the ubiquitin chain from substrates prior to their degradation so that ubiquitin tags can be recycled, and the unwanted protein substrate can be degraded. Without the function of RPN11, protein substrates would be sterically hindered from entering the proteasome, preventing degradation. RPN11 is therefore essential to proteasome function (Worden et al., 2017). In addition, as RPN11 is a metalloprotease, it is more resistant to oxidative stress than cysteine proteases involved in deubiquitination, as the catalytic cysteine can be oxidated to form sulfenic acid (Devarie-Baez et al., 2015; Mevissen & Komander, 2017). The zinc ligands of family M67 are two histidine residues and an aspartate residue within the conserved **HSHP(X)₉D** motif (Rawlings et al., 2010).

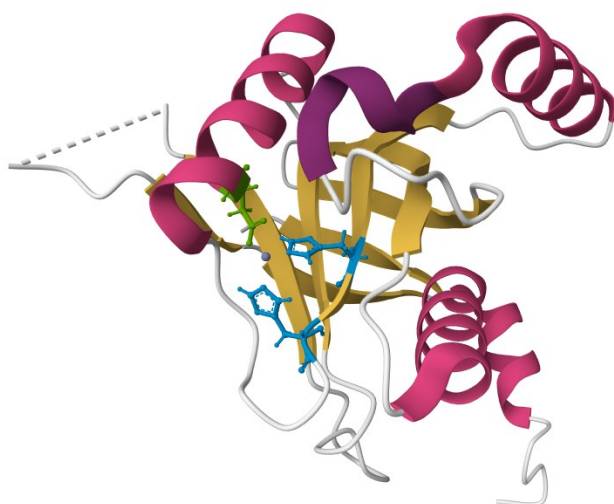


Figure 1.15: Richardson diagram of *Saccharomyces cerevisiae* S288C Rpn11 (PDB ID: 4O8X) (Worden, Padovani & Martin, 2014) from clan MP coloured according to secondary structure, with zinc ligands shown in ball and stick representation (aspartate residue coloured in green, histidine residues coloured in blue, zinc ion shown in grey).

Clan MQ contains only one family of proteases: M29. These are all aminopeptidases that contain cocatalytic zinc or cobalt ions (Kuo et al., 2003; Rawlings & Barrett, 2013; Stoll et al., 1976), that are ligated by, in order, two glutamate residues, two histidine residues, and an aspartate residue (Figure 1.16).

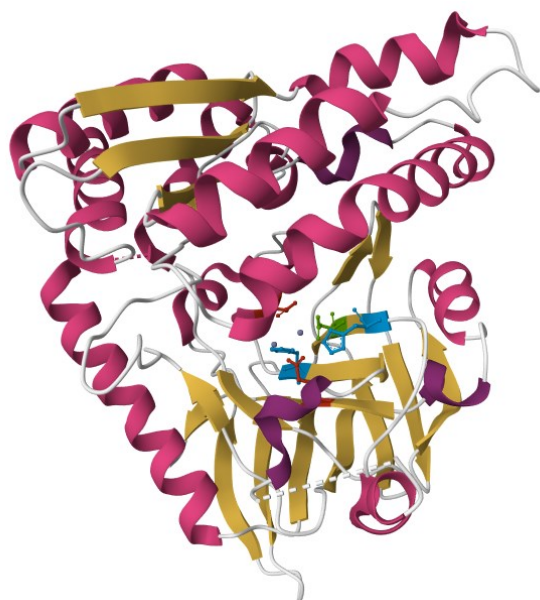


Figure 1.16: Richardson diagram of *Thermus thermophilus* AmpT (PDB ID: 2AYI) (Odintsov et al., 2005) from clan MQ coloured according to secondary structure, with zinc ligands shown in ball and stick representation (aspartate residue coloured in green, glutamate residues coloured in red, histidine residues coloured in blue, zinc ions shown in grey).

The imelysins of family M75 are found in clan MS, classified as peptidases due to the HxxE motif, however these have since been shown to have no peptidase activity (Xu et al., 2011).

The protease clan MT is formed of a single protease: MlrC from *Sphingomonas* sp. ACM-3962 (Figure 1.17). This M81 family protease degrades the hepatotoxin microcystin LR, the most common form of microcystin, which are heptapeptide toxins produced by cyanobacteria of the genera *Microcystis*, *Anabaena*, *Nostoc*, and *Oscillatoria* (Bourne et al., 1996; Carmichael et al., 2009).

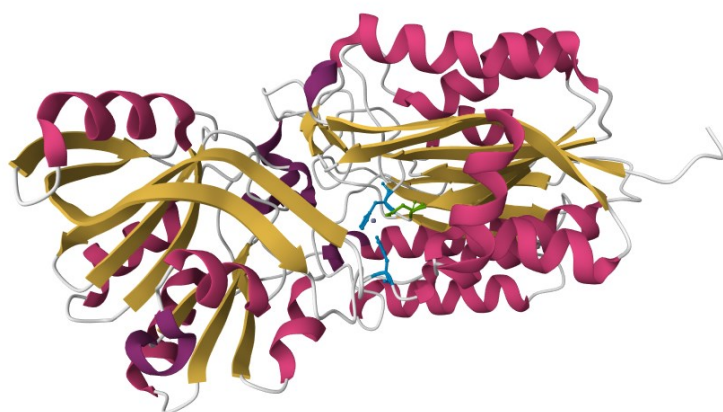


Figure 1.17: Richardson diagram of *Chelativorans* sp. BNC1 Putative metallopeptidase (PDB ID: 3IUU) from clan MT coloured according to secondary structure, with zinc ligands shown in ball and stick representation (aspartate residue coloured in green, histidine residues coloured in blue, zinc ion shown in grey).

Clan MU is composed of family M103, the representative example of which is TldD peptidase from *Escherichia coli* (Figure 1.18), which is involved in the degradation of CcdA of the CcdB/A toxin/antitoxin system (Hu et al., 2012).

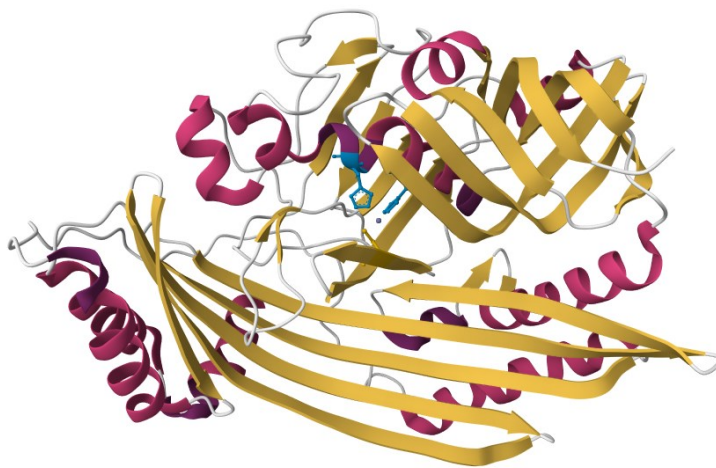


Figure 1.18: Richardson diagram of *Escherichia coli* str. K-12 substr. MG1655 TldD (PDB ID: 5NJ5) (Ghilarov et al., 2017) from clan MU coloured according to secondary structure, with zinc ligands shown in ball and stick representation (histidine residues coloured in blue, cysteine coloured in yellow, zinc ion shown in grey).

Finally, there is family M73, which is not assigned to any known metalloprotease clan. This includes camelysin from *Bacillus cereus* and spore coat associated protein N from *Bacillus subtilis*, both of which are monometallic zinc endopeptidases.

Camelysin may have implications in disease, as *B. cereus* may cause secondary infections in hospitals and is a causative agent of food poisoning. Camelysin possesses the ability to break down collagen, a component of connective tissue, and activate plasminogen. This activation has the potential to aid in the infiltration of fibrin clots and the extracellular matrix during bacterial invasion (Fricke et al., 2001). The active site residues and metal binding ligands of family M73 are unknown, although there are no conserved histidine residues, which is unusual among metalloproteases. The only other family that does not use histidine to coordinate the metal ion is family M63 (Rawlings et al., 2010).

Another key feature to characterise proteases is the presence or absence of a propeptide domain, which functions in regulating protease activity. While this is not strictly conserved across clans, the prevalence of propeptide domains does follow trends within clans. For example, propeptide domains are commonly found in the proteases of clan MA (Demidyuk et al., 2008).

1.3.3 Inhibition by propeptide domains

Propeptide domains are often observed in metalloproteases. In the thermolysin family of metalloproteases this may be either around 50 or around 200 residues in length (Demidyuk et al., 2008), however they can be greater than 200 (Boon et al., 2020). This propeptide inhibits catalytic activity, thereby preventing uncontrolled proteolysis, as well as having an impact on protein folding and export. The propeptide may then be cleaved, which may be autocatalytic or mediated by other proteases, or otherwise modified (such as through substrate binding) to allow protease function.

One well-characterised method of metalloprotease inhibition by propeptide domains is the so-called “cysteine switch”. This was first observed in the matrix metalloproteinases, or MMPs (Springman et al., 1990; Vallee & Auld, 1990; Van Wart & Birkedal-Hansen, 1990), a group of metalloproteases found in the M10 family of the metzincin subclan. The cysteine switch uses a cysteine residue within the propeptide sequence to act as the fourth ligand of the catalytic zinc ion in place of the water molecule. Dissociation of this cysteine residue leads to ligation of a water molecule and the catalytic zinc, activating the protease (Figure 1.19). *Bacteroides fragilis* fragilysin employs a similar mechanism known as the aspartate switch, in which an aspartate residue (Asp194) coordinates the zinc ion in place of a water molecule while the propeptide remains associated (Goulas et al., 2011). Examples of the cysteine switch can be found in other members of the metzincin subclan, and indeed, it has been found that all metzincins either possess a cysteine switch or a catalytic Tyr149 residue, except for the notable exception of *Bacteroides fragilis* fragilysin. The catalytic Tyr149 was first observed in the astacins of family M12 (Gomis-Rüth et al., 1993; Miyatake et al., 1995), and may be found C-terminal of the Met-turn motif. This tyrosine residues acts as a weak fourth protein ligand for the catalytic zinc ion. Mutation of *Astacus astacus* astacin Tyr149 to alanine resulted in total loss of protease activity, while mutation to phenylalanine retained 2.5% activity (Yiallourous et al., 2000), which indicates that Tyr149 is involved in transition state stabilisation during hydrolysis of the protein substrate.

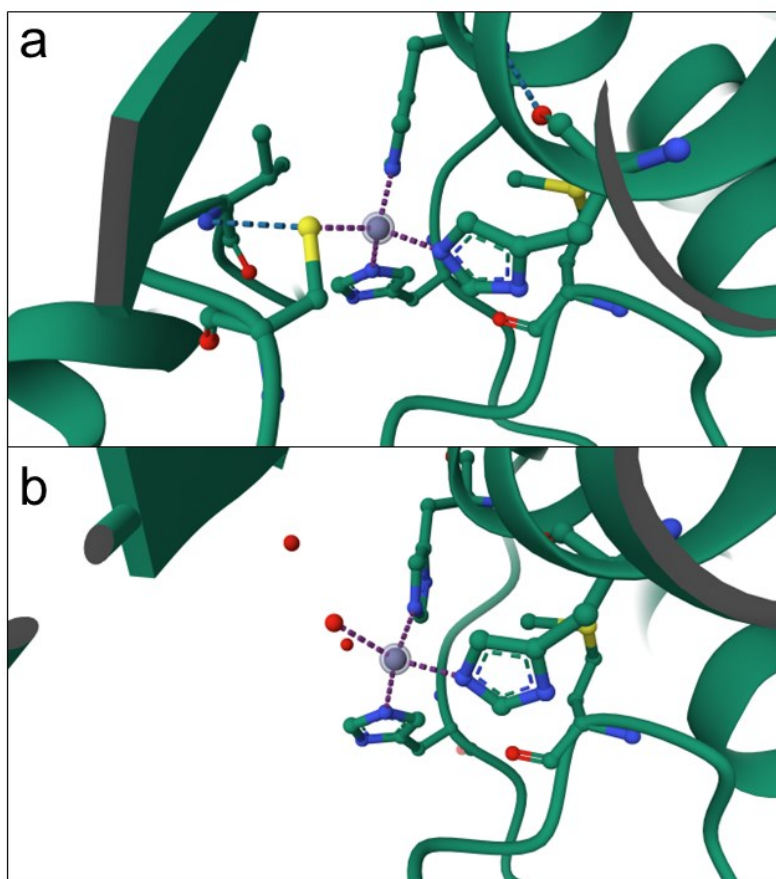


Figure 1.19: Illustration of cysteine switch mechanism of inhibition and activation in the case of *Homo sapiens* matrix metalloproteinase 1 (MMP1). Metal coordination is shown by purple dashed lines, hydrogen bonds are shown by blue dashed lines, water molecules are shown as red dots. a) proMMP1 (PDB ID: 1SU3) (Jozic et al., 2005), in which the active site zinc is coordinated by three histidine residues and the cysteine-switch cysteine. b) Active MMP1 (PDB ID: 2CLT) (Iyer et al., 2006), in which the zinc-coordinating cysteine has been displaced and the fourth zinc ligand is a water molecule.

1.3.4 Activation of metalloproteases

Metalloproteases may be activated in various ways, most commonly through cleavage of the propeptide. This may be autocatalytic, as in the case of thermolysin (Marie-Claire et al., 1998), or mediated by other proteases, but chemical addition or posttranslational modification can also affect activation. Some proteases may even be activated without removal of the propeptide.

The activation of metalloproteases through action of separate proteases is often a complex one, with an intricate web of interactions (Fortelny et al., 2014). For example, human MMP-9, a secreted metalloprotease containing a cysteine switch, may be activated through cleavage of the propeptide domain by other metalloproteases including MMP-3 and meprin, as well as the serine proteases

neutrophil elastase and tissue kallikrein (Desrivieres et al., 1993; Geurts et al., 2012; Jackson et al., 2010; Ogata et al., 1992). Intracellular MMPs, on the other hand, may be activated in the Golgi by preprotein convertases such as furin (Desrivieres et al., 1993).

Metalloprotease activation may also proceed via trimming of the propeptide domain by other proteases followed by autocatalytic processing. A notable example is the ATP-dependent *m*-AAA protease Apg3l2, which is first processed by mitochondrial processing peptidase to produce an intermediate form that is processed autocatalytically (Koppen et al., 2009).

Posttranslational modifications can also affect metalloprotease activation, as is the case with human MMP-9, which has a propeptide that may be glycosylated at Asn38. Glycosylation of this propeptide then inhibits processing of proMMP9, providing another layer of control for MMP-9 activation.

Protease activation can proceed through chemical means, as with MMPs containing a cysteine switch. The zinc-coordinating cysteine in these proteases may be oxidized or nitrosylated following exposure to reactive oxygen species or reactive nitrogen species, respectively, thereby activating the protease (Schulz, 2009).

In some cases, proteases may be activated without the removal of their propeptide domain. MMP-9, for example, shows activity upon binding with gelatin, despite the presence of its propeptide, suggesting that substrate binding causes displacement of the propeptide from the active site (Bannikov et al., 2002). Similarly, activity can be seen for ADAMTS13 in the absence of propeptide cleavage (Majerus et al., 2003).

1.3.5 Further roles of propeptide domains

Beyond the fine-tuning of protease activity, propeptide domains have been shown to have several functions in physiology. They can be required for correct folding of the protease, which has been observed for ADAM10, which is involved in cleavage of the Notch receptor in mammals (Anders et al., 2001; van Tetering et al., 2009). They may also have roles in protease trafficking, as with ADAMTS9, which requires glycosylation of the propeptide for secretion (Koo et al., 2007). Finally, propeptide domains may be implicated in interaction with other proteins. The propeptide of BMP-1, for example, binds to BMP-2 and BMP-4 (Jasuja et al., 2007).

1.3.6 Metalloprotease mechanism of action

Whether or not the protease has a propeptide domain or is produced directly in its active form, it will function in the hydrolysis of peptide bonds. The mechanism of action of metalloproteases is difficult to elucidate, compounded by difficulty in cocrystallisation due to rapid hydrolysis of the substrate. Like all proteases, they most likely act preferentially on beta strands (Tyndall et al., 2005).

The best characterised family of metalloproteases is M4, the thermolysin-like proteases. Thermolysin was one of the first metalloproteases to be sequenced and to have its tertiary structure solved (Matthews et al., 1972; Titani et al., 1972). The thermolysins are all gluzincins and are mostly endoproteases. The prototypical example of this family is thermolysin from *Bacillus thermoproteolyticus*. Thermolysin has been extensively characterised, with 190 crystal structures submitted to the PDB to date, and the mechanism of action is relatively well understood. First, the thermolysin recognises the substrate, preferentially close to an aromatic residue. It then undergoes a hinge-bending motion of the N- and C-terminal domains with respect to the active site pocket (Carvalho et al., 2015; Holland et al., 1992). Beyond this there are two proposed methods of action. In the first, the glutamate residue of the HEXXH motif (in this case HELTH) acts as a general base, polarising the zinc-bound water molecule. The polarised water molecule then performs a nucleophilic attack on the substrate peptide bond, forming an intermediate with the carbonyl carbon which is stabilised by a protonated histidine from the active site (Figure 1.20a). The second proposed mechanism of action has a histidine in the active site acting as a general base instead of the glutamate (Figure 1.20b) (Adekoya & Sylte, 2009; van den Burg & Eijssink, 2013). The thermolysin model of action is not universal amongst metalloproteases, or even amongst the zincin clan, for example, in metzincins the zinc ligand is coordinated by a histidine or aspartate residue (Bode et al., 1993; Rawlings & Barrett, 2013) and aspzincins coordinate the zinc ligand in a pyramidal state, compared to the tetrahedral state common of gluzincins and metzincins (Hori et al., 2001). Despite these, and other, differences, the thermolysin mechanism of action is a good example of metalloprotease function, specifically how the water molecule is coordinated for nucleophilic attack, which is a feature common to all metalloproteases.

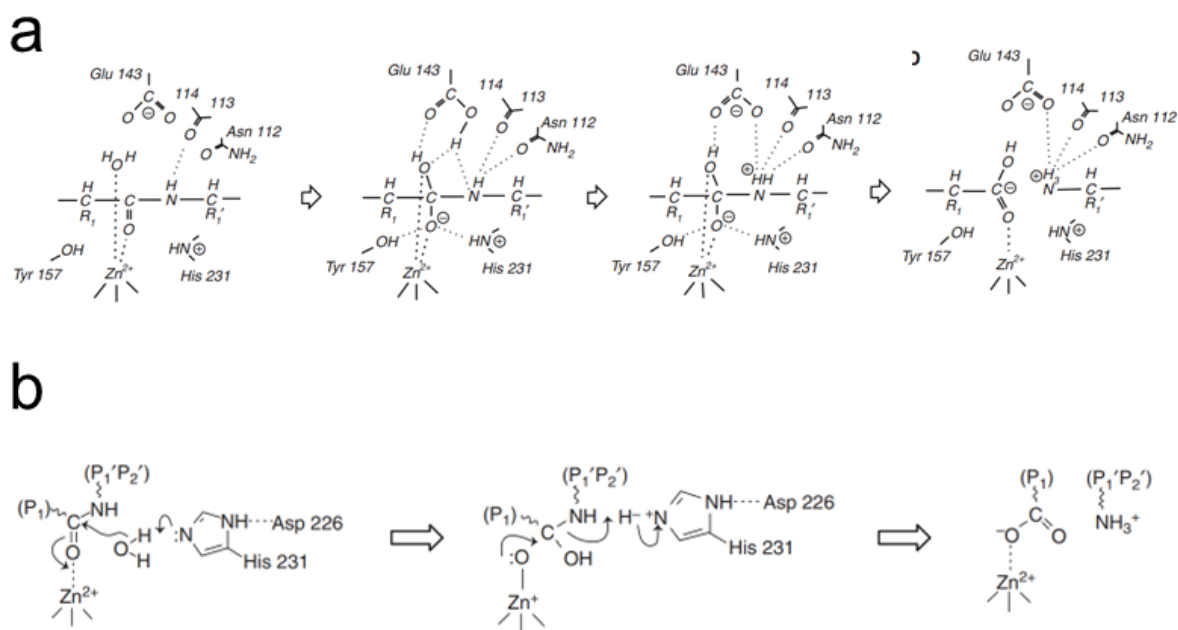


Figure 1.20: Two possible mechanisms of proteolysis by *Bacillus thermoproteolyticus* thermolysin a) The first proposed method of thermolysin action, in which the glutamate residue of the HELTH active site acts as a general base, polarising the catalytic water molecule which then performs a nucleophilic attack on the substrate peptide bond. b) The second proposed mechanism of thermolysin action, in which an active-site histidine polarises the catalytic water molecule through its action as a general base. (Taken from Adekoya & Sylte, 2009)

1.4 Bioinformatic approaches of predicting enzyme activity

1.4.1 The ideal candidate protease

The ideal protease candidate will exhibit enhanced activity at cold temperatures, increased halophilicity and greater acid tolerance, relative to industry standards.

The majority of detergent proteases currently in use have optimal activities at high temperature, due to claims of increased overall activity (Mansfeld & Ulbrich-Hofmann, 2007; Queiroga et al., 2012). There has been a recent trend, however, towards the implementation of cold-active enzymes within industrial processes, which would reduce energy costs and environmental impact. In relation to detergent applications, a reduced washing temperature will also lower damage to garments (Kuddus & Ramteke, 2012). These cold-active enzymes are often isolated from psychrophilic, psychrotolerant or mesophilic microorganisms, with temperature optima between 0-40 °C (Białkowska et al., 2016).

In addition to cold-activity, halophilicity – increased activity or stability at high salt concentrations – has been identified as a target for novel proteases. Halophilicity can improve activity in saline and coastal water sources, where NaCl concentrations can reach 2% (w/v), thereby improving detergent efficiency in these areas (H. Zhang et al., 2018). Secondary to this, activity in a range of water hardness levels is desirable, to ensure broad applicability in different water sources. This can be expressed as a ratio of calcium to magnesium ions (Evans et al., 2011). Furthermore, high concentrations of sodium chloride are used in the process of granulation, which is often undergone when the protease will be added to detergent powders (Jaouadi et al., 2012).

Finally, higher activity at the slightly-acidic to neutral pH range is desirable as the industrial partner may be shifting towards lower pH's in their detergent formulations, to reduce the need for petrochemical-derived preservatives. Additionally, there should be significant activity in the presence of sequestrants, such as EDTA, as these are often included in formulation for improved stain removal by chelating calcium and magnesium ions to effectively reduce water hardness (Niyonzima & More, 2015) (Section 1.2.2). However, this may pose a significant challenge as these are potent inhibitors of metalloproteases due to chelation of stabilising calcium ions and the catalytic metal ion(s) (Fontana, 1988).

1.4.2 Identification of candidate proteases

Microorganisms display a vast potential for novel catalyst identification, although identification of specific proteases for study from over 390,000 documented candidates (in the MEROPS database alone) is a significant challenge. Structure, function, and substrate specificity vary within protease clans and even within families and subfamilies.

Traditional screening for proteases involves identification and culturing of protease-secreting organisms, an approach that has been utilised extensively for discovery of a vast range of proteases (Chen et al., 2011; Reddy et al., 2017; Riffel et al., 2003; Wang et al., 2008). While this method has shown many successes, it is unwieldy; taking a long time to optimise the culturing conditions, and inherently excludes the vast majority of possible proteases due to the uncultivability of between 99% and 99.9% of prokaryotes (Berini et al., 2017). Metagenomics seeks to access this

“microbial dark matter” through direct sequencing of DNA from an environmental sample, bypassing any need for culturing.

Bioinformatic analysis of these metagenomic sequences may then lead to identification of new enzymes through detection of conserved residues and motifs. However, the prediction of enzyme characteristics, e.g. thermostability, is not an easy task, as the source organism is not always a reliable indicator. For example, both hAQP10 and CutA1 are thermostable proteins, yet they are derived from a mesophilic source – *Homo sapiens* (Bagautdinov et al., 2015; Oberg et al., 2011). Similarly, malate dehydrogenase from *Salinibacter ruber*, an extreme halophile, behaves like a non-halophilic enzyme (Madern & Zaccai, 2004).

1.4.3 Predicting biochemical characteristics

While it is not yet possible to definitively determine protein stability without experimental characterisation, some generalisations have been made based on sequence analysis. One such example is the ratio of glutamate and lysine to glutamine and histidine, which was derived by Farias et al. (2003) to distinguish between hyperthermophilic, thermophilic and mesophilic proteins. Similarly, several trends in cold adaptation were posited by Adekoya et al. (2006), including a lower ratio of arginine to arginine plus lysine, lower percentages of aliphatic residues and more asparagine residues – which are non-catalytically deaminated, reducing thermostability. These differences are posited to increase overall flexibility of the protein, thereby decreasing thermostability. However, there have been many such rules suggested and when taken together they are often found to be confused and contradictory (Chakravorty et al., 2017).

1.4.4 Machine learning based methods of predicting protein characteristics

Greater clarity may be obtained through the use of machine learning approaches, which have been employed to elucidate and refine the rules governing protein stability. Various methods have been explored for this purpose, including genetic algorithms, artificial neural networks (ANNs), support vector machines (SVMs), rotational forest and random forest.

Genetic algorithms can be viewed as computational analogues to biological evolution, iteratively refining results over successive generations to optimise the prediction of biological traits (Weber, 1998). ANNs, on the other hand, are inspired by

neurological processes, and they are particularly effective when there is limited prior knowledge of the model, however, they can be prone to over-fitting (Tu, 1996). SVMs are specialized classifiers that find the best separating hyperplanes to distinguish different biological patterns, making them effective in situations where there is significant understanding of the model, and they are less susceptible to over-fitting than ANNs (Awad & Khanna, 2015). Rotational and random forests are based on decision trees, with both methods utilizing ensembles of trees to improve predictive performance. However, they differ in their approach to constructing these ensembles. In a rotational forest, each decision tree in the ensemble is trained on a different subset of features selected through a rotation process, enhancing the model's robustness and adaptability, particularly in high-dimensional or noisy datasets (Rodriguez et al., 2006). In contrast, a random forest builds an ensemble by training each tree on a randomly selected sample of the data, with feature selection done randomly at each split. This randomness reduces overfitting and promotes diversity among the trees (Breiman, 2001).

All of these approaches have been successful in prediction of protein stability. Nath & Subbiah (2014) implemented a rotation forest approach to predict psychrophilicity of proteins, leading to the generation of rules which may be easily understood from a human perspective. Support vector machine (Lin & Chen, 2011; Zhang & Fang, 2006), genetic algorithms (Wang & Li, 2014) and neural networks (Gromiha & Suresh, 2007) have also been utilised to train models to predict thermostability with recorded accuracies often above ninety percent. Similar algorithms have been developed to predict halophilicity (Ebrahimie et al., 2011; Nath, 2016) and acidophilicity (Lin et al., 2013). Some studies observed differences in prediction accuracy between models (Nath & Subbiah, 2014), and others found that the choice of model made little impact (Gromiha & Suresh, 2007).

1.4.5 Bioinformatic tools for database mining

Prior to prediction of activity, a dataset of protein sequences must be generated. Generally, this begins with local alignment searching to find homologous sequences to enzymes of known function.

Local alignment searching has been dominated by the basic local alignment search tool (BLAST) since its development by Altschul et al. in 1990 (Altschul, 1997;

Camacho et al., 2009). BLAST gained popularity due to its ease of use and speed, which is achieved through a heuristic method which begins by searching for short matches between sequences (seeding) followed by expansion of these alignments. The size of initial matches is defined as the word size, which has a default length of 5 for protein sequences, and 28 for nucleotide sequences. Each local alignment is given a score of T based on the distance between sequences. For protein sequences this is computed according to a substitution matrix, the default of which is BLOSUM62 (Pearson, 2013). This ensures that similar residues (e.g. aspartate and glutamate) are given a higher score than dissimilar residues (e.g. aspartate and phenylalanine). Following initial seeding, alignments are extended in both directions and the T score is adjusted. Once T drops below a given threshold, alignment extension stops, and if the alignment has a score above the threshold, it will be included in the final BLAST output. This method ensures speed at the cost of accuracy, as it cannot be guaranteed that all possible alignments will be detected, particularly if the word size or T -thresholds are high.

Another method of local alignment is HMMER (Eddy, 2011), which is more effective than BLAST when searching for remote homologues from different organisms (Goussarov et al., 2022). It uses profile hidden Markov models (HMMs), which are probabilistic models that capture the statistical characteristics of biological sequences. HMMER begins by creating a profile HMM from a multiple sequence alignment (MSA) or a single sequence. Probabilities of observing different residues or gap states at each position within the alignment are estimated. Transition probabilities between states in the model are also determined, which reflects the likelihood of moving from one position to another during sequence alignment. HMMER then conducts a systematic search of a target sequence database to find potential homologs. This exploration is executed via the Forward algorithm, which computes the cumulative probability of generating each target sequence while traversing the model. The alignment scores obtained from the Forward algorithm are subsequently employed to rank the sequences within the database, offering a measure of their resemblance to the query model, and sequences with computed expectation (E)-values above a given threshold are included in the final result. HMMER uses ensemble algorithms that take into account all possible alignments, and also considers the positioning of residues, as opposed to BLAST which

searches for optimal high-scoring alignments. This ensures that HMMER has a greater accuracy when searching for homologous sequences with low sequence identity.

Another useful tool for database mining is CD-HIT, which may be used to reduce redundancy in a database through removal of similar sequences (Li & Godzik, 2006). It begins by calculating similarities between sequences, grouping highly similar ones into clusters based on user-defined thresholds. Within each cluster, a representative sequence is selected. This process results in a reduced database with representative sequences, eliminating redundancy while preserving dataset diversity. CD-HIT's efficiency and clustering algorithms make it significantly faster than BLAST for simplifying sequence datasets.

It is often helpful to determine whether a protein is intracellular, secreted or membrane bound. For example, extracellular or “feeding” proteases are more likely to be effective detergent proteases (Maurer, 2004). There are a variety of programs designed for this purpose, with SignalP being the most utilised due to its user-friendly web-based interface and high level of performance (Brockmeier et al., 2006; Choo et al., 2009; Klee & Ellis, 2005; Menne et al., 2000; Nielsen, 2017; Zhang et al., 2009). SignalP 4.0 (Nielsen, 2017) employs a multi-step process, commencing with sequence analysis to identify potential signal peptides (SPs) based on characteristic motifs and hydrophobic regions. It then extracts features from the protein sequence, such as amino acid composition and physicochemical properties. Machine learning, in the form of ANNs and HMMs, is applied to classify sequences into SP or non-SP categories, based on these features. Moreover, SignalP 4.0 then provides predictions for the precise cleavage site of the SP, so that the processed peptide sequence can be determined.

1.4.6 Prediction of protein 3D structure

Having identified candidate enzyme sequences, it is often useful to model their 3D structure. As these enzymes are likely to have minimal or no experimental characterisation, their structures must be predicted purely based on their primary sequence. There have been many attempts to predict the 3D structure of a protein based on its sequence alone. These fall into two categories: the physical interaction programme and the evolutionary programme.

The physical interaction programme is based on simulating the physical forces and interactions between atoms or amino acids to determine the most energetically favorable protein conformation, although this has proven to be too computationally demanding (Brini et al., 2020; Jumper et al., 2021; Sippl, 1990). Conversely, the evolutionary programme seeks to predict structure based on conservation and co-evolutionary patterns observed in multiple sequence alignments and homology to known structures (Jones et al., 2012; Jumper et al., 2021; Marks et al., 2011; Roy et al., 2010; Weight et al., 2009). The most successful of these methods to date has been AlphaFold2 (Jumper et al., 2021), which uses ANNs to predict structure. This relies on the evolutionary programme but also considers physical constraints and energy potentials as part of its model training and refinement process.

The Critical Assessment of Structure Prediction (CASP) assessment is the benchmark for 3D structure prediction. This uses solved structures that have not been deposited into the protein data bank (PDB) to ensure that all predictions are unbiased. In CASP14, AlphaFold2 outcompeted all other models for structure prediction, with a median backbone accuracy of 0.96 Å r.m.s.d.₉₅ where the next best performer showed an accuracy of 2.8 Å r.m.s.d.₉₅ (Jumper et al., 2021), a significant result considering that the width of a carbon atom is 1.4 Å.

1.5 *Bacillus subtilis* as a host for protease production

1.5.1 *B. subtilis* in industry

In this work, *B. subtilis* was chosen as a host for expression of candidate proteases. *B. subtilis* has a long history as a host for industrial enzyme production. *Bacillus* species have demonstrated robust secretory capacities, with reported secretion levels up to 25 g/L (Degering et al., 2010; Schallmey et al., 2011). Enzymes produced in *Bacillus* species account for 60% of the global enzyme market (Degering et al., 2010), which was valued at \$5.8b in 2021 and is predicted to reach \$10.2b by 2031 (Manjrekar S et al., 2021). Due to a lack of endotoxin production, it has the status of “generally regarded as safe” (GRAS) by the US Food and Drug Administration (FDA), and “qualified presumption of safety” (QPS) by the European Food Safety Authority (EFSA). Additionally, it has a well-understood pathway for natural genetic competence and has a host of available tools for genetic

manipulation (Hohmann et al., 2017; Schallmey et al., 2011). As such, *B. subtilis* is currently used for production of many commercial enzymes, including amylases, amylomaltase, cellulases, xylanases, and proteases (Hohmann et al., 2017; Maurer, 2004).

In addition to its use as a host for enzyme production, *B. subtilis* is also used in industry for the manufacture of high value chemicals, including riboflavin, folic acid, and nucleosides (Sauer et al., 1998), that are used in the food industry as flavour-enhancers. Production of adenosine in *B. subtilis*, for example, has been achieved at levels upwards of 14 g/L (Chen et al., 2013; Li et al., 2019).

1.5.2 Native proteases of *B. subtilis*

According to the MEROPS database, the *B. subtilis* lab strain 168 produces 177 native proteases, including putative and predicted proteases (Rawlings et al., 2010). Three of these are known to be associated with the cell wall (HtrA, HtrB and WprA), and seven of these are extracellular (AprE, Bpr, Epr, Mpr, NprB, NprE and Vpr). This affects background protease activity, decreasing the sensitivity of protease activity measurements (Tu et al., 2011), and has been shown to decrease yield of heterologous proteins (Westers et al., 2008; Wu et al., 2002; Wu et al., 1993). A number of protease deficient strains have been constructed (Pohl et al., 2013; Wu et al., 1991; Zhang et al., 2016). These have been previously used for heterologous expression of proteases (Liu et al., 2019; Tian et al., 2019), and it was found by Tu et al. (2011) that deletion of all seven extracellular proteases (and perhaps the quality control protease *wprA*) is necessary to effectively reduce background activity.

1.5.3 Genetic competence of *B. subtilis*

One of the attractive features of *B. subtilis* for use in industry is the ease by which it may be genetically manipulated through induction of natural competence, first investigated by Spizizen (1958).

When *B. subtilis* cells enter the stationary phase, typically triggered by nutrient scarcity and high cell density, they undergo differentiation into various subpopulations. Among these subpopulations, some acquire motility as observed by Nishihara & Freese (1975), while others initiate the formation of biofilms (Vlamakis et al., 2008), secrete degradative enzymes and antibiotics (González-Pastor et al., 2003), or ultimately proceed to sporulation (Piggot & Hilbert, 2004; Rudner & Losick,

2001). Additionally, a small subpopulation differentiates into competent cells capable of taking up extracellular DNA (Dubnau, 1991; Dubnau & Provvedi, 2000).

The transcriptional master regulator of sporulation is Spo0A, which governs the process of differentiation. The level of phosphorylated Spo0A determines the physiological state, which is controlled through an intricate process that involves the kinases KinA, B, C, D and E as well as a phosphotransferase (Spo0B) and Spo0F – a phosphate acceptor (Burbulys et al., 1991). Phosphorylated Spo0A in turn acts as a transcriptional repressor of *abrB* (Bird et al., 1993; Strauch et al., 1990), which is a repressor of *comK* expression (Hahn et al., 1995). ComK is the master regulator of competency, therefore an increase in phosphorylated Spo0A will derepress *comK* expression and lead to initiation of competency. This, along with several other means of regulation (including CodY, Rok and the two-component system DegS-DegU), ensures tight control of *comK* expression, such that only 5-10% of cells in a culture grown to competence will express *comK* (Berka et al., 2002). When ComK reaches a sufficient level, it will bind to the *comK* promoter as a dimer of dimers and activate its own expression (Maamar & Dubnau, 2005; Smits et al., 2005). ComK then initiates expression of over 100 genes involved in DNA recombination, repair, binding, and uptake (Berka et al., 2002).

The method developed by Spizizen (1958) induces genetic competency by starvation, which provided the template for most current methods (Anagnostopoulos & Spizizen, 1961; Bennallack et al., 2014; Kunst & Rapoport, 1995). Inducible expression of the master regulator *comK* and *comS* (which prevents ComK degradation) has also been used for induction of competency, exhibiting transformation efficiencies up to 10^7 CFU/ μ g of DNA (Duitman et al., 2007; Hahn et al., 1996; Liu et al., 1996; Nijland et al., 2010; Rahmer et al., 2015; Zhang & Zhang, 2011).

1.5.4 Minimal *B. subtilis* strains

Genome reduction has been explored in several bacterial species, including *E. coli*, *Streptomyces avermitilis*, *Pseudomonas putida* and *B. subtilis*, and has been shown to improve yields of high-value products through the removal of non-essential genes (Aguilar et al., 2020; Juhas et al., 2014; Komatsu et al., 2010; Leprince et al., 2012; Trinh et al., 2008). In *B. subtilis*, 253 genes have been identified that are essential for

growth on rich media at 37 °C (Kobayashi et al., 2003), and successive attempts at deletion of non-essential genes have generated strains with total genome deletions of 7.7%, 25% and 36% (Ara et al., 2007; Reuß et al., 2017; Westers, 2003). The last of these, strain PG10 developed by Reuß et al. (2017), required the inclusion of a *comK/comS* cassette expressed under the mannitol-inducible *mtlP* promoter for overexpression of components in the natural competency pathway (Rahmer et al., 2015) (Section 1.5.3). This was added to ensure sufficient transformation efficiency for these deletions.

1.5.5 Protein secretion in *B. subtilis*

Protein secretion in *B. subtilis* can proceed through the Sec pathway, the Tat pathway, ABC transporters or the pseudopilin export pathway (Fu et al., 2007). Most proteins are exported through the Sec pathway (Figure 1.21), in a process beginning with targeting of the unfolded protein to the cytoplasmic membrane via a signal peptide (SP) that may be found at the N-terminus of the protein. This SP is composed of a positively charged N-terminal region, a hydrophobic region and a polar C-terminal region, the recognition site for the signal peptidase (Brockmeier et al., 2006). Peptides are held in an unfolded state following translation by the signal recognition particle (SRP) and FtsY chaperones (Zanen et al., 2006), and are then channelled through the membrane by the Sec apparatus. This consists of the membrane channel proteins SecY, SecE, SecG (SecYEG) and SecDF, and the SecA ATPase motor (Fu et al., 2007). Following translocation through the SecYEG channel, the SPs of extracellular proteins are cleaved at the signal peptidase recognition site by signal peptidases, such as SppA, and the SP is degraded, preventing blockage of Sec pathway components (Bolhuis, Matzen, et al., 1999). The protein is then folded in the extracellular space, facilitated by the chaperone PrsA, and sometimes by the thiol-disulphide oxidoreductases BdbB, BdbC and BdbD (Kouwen & van Dijk, 2009; Quesada-Ganuza et al., 2019). Finally, the quality control proteases HtrA, HtrB and WprA function to prevent blockages in the secretion system by degradation of misfolded proteins (Tjalsma et al., 2004).

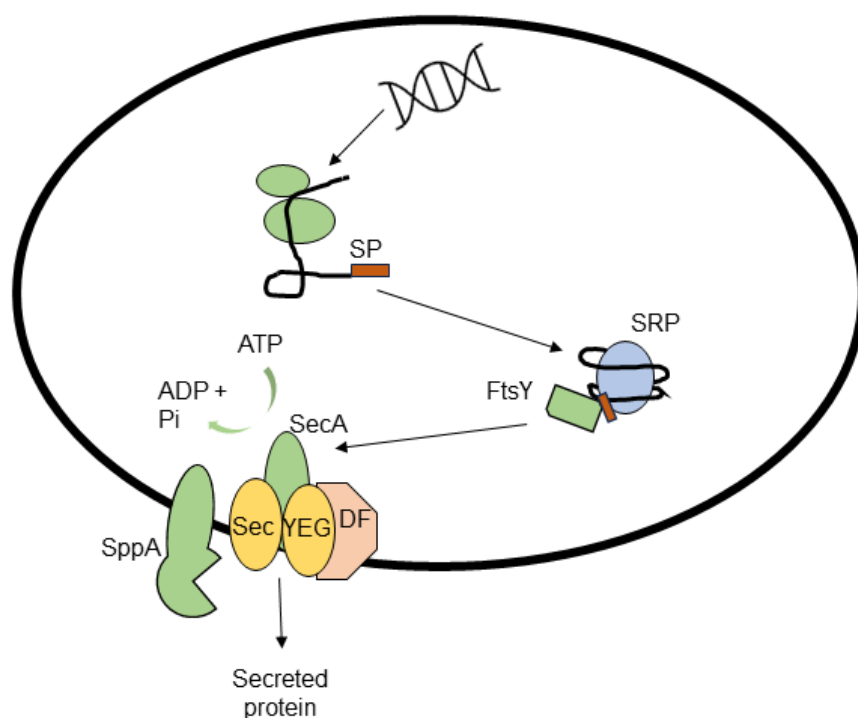


Figure 1.21: Illustration of protein secretion in *B. subtilis* through the Sec pathway (Adapted from Fu et al., 2007).

1.5.6 Improving secretion of heterologous proteins

While high yields of protein can be achieved for expression of homologous proteins in *B. subtilis* (Schallmey et al., 2011), secretion of heterologous proteins can lead to bottlenecks due to blockages (Li et al., 2004). This may be due to variations in folding efficiencies among proteins that result in the obstruction of the Sec pathway components. Consequently, this obstruction can lead to heightened degradation of secreted proteins, whether they are correctly or incorrectly folded, by the quality control proteases HtrA/HtrB as indicated by Brockmeier et al. (2006). There have been many attempts to tackle this, including optimisation of SPs and the secretion pathway.

The choice of SP has been found to have a huge impact on the level of protein secretion through systematic screening of large libraries of SPs (Brockmeier et al., 2006; Degering et al., 2010). The level of secretion is not only organism dependent (Hemmerich et al., 2016), it is protein dependent. After SP library screening Brockmeier et al. (2006) found that in *B. subtilis* the most efficient SP for secretion of a serine protease (cutinase) was inefficient at secretion of an esterase (EstCL1), and vice versa. This is believed to be related to a wide range of factors, including the SP

translation rate, the availability of chaperones and the protease cleavage sites present in the mature protein.

Another notable aspect is the influence of the propeptide region, that is, the N-terminal portion of the mature protein (Choo & Ranganathan, 2008). Le Loir et al. (1998) found that the use of a synthetic propeptide region in the lactic acid bacterium *Lactococcus lactis* significantly improved protein secretion, although the exact mechanism of this has not yet been fully elucidated. Similarly, many studies have been conducted on the effect of the *Staphylococcus hyicus* lipase prepro-region on heterologous expression in *B. subtilis* (Bolhuis, Venema, et al., 1999; Kouwen et al., 2010; Meens et al., 1997), finding that use of the pro-region in conjunction with the pre-region (the SP) can increase secretion into the medium compared with the SP on its own.

There have also been efforts to improve secretion through optimisation of elements of the Sec pathway, as these can get blocked when overexpressing heterologous proteins (Chen et al., 2015). This includes the overexpression of *ftsY* and SRP components (Quax et al., 2002), overexpression of *prsA*, *groE* and *dnaK* (Wu et al., 1998) and modification of SecA (Kakeshtia et al., 2010). A combinatorial study that investigated all Sec pathway components found that overexpression of *prsA* and a partial *dnaK* operon had the greatest impact on improving secretion of heterologous α -amylases (Chen et al., 2015). Expression of heterologous secretory components has also been shown to improve secretion, such as expression of *Pseudomonas glumae* chaperones (Bos et al., 1998) and *secB* from *E. coli* (Diao et al., 2012).

Other efforts to improve protein secretion involve the deletion of the surfactin *srfAC*, which has been implicated in foam production during culturing. This has led to increase in yields of pullulanase from *Bacillus naganoensis*, when secreted in *B. subtilis* (Liu et al., 2018b). Optimisation of promoters and ribosome binding sites has also been explored to improve protease expression (Li et al., 2022; Liu et al., 2018a; Mao et al., 2021).

1.5.7 CRISPR-Cas systems in *B. subtilis*

In recent years, clustered regularly interspaced short palindromic repeats (CRISPR)-associated (Cas) systems have witnessed widespread adoption, revolutionizing genome editing techniques (Jinek et al., 2012). This technology has led to

substantial progress in genome manipulation, particularly in microorganisms of industrial significance like *B. subtilis* (Altenbuchner, 2016; Burby & Simmons, 2017; Price et al., 2019; Westbrook et al., 2016). The most broadly used CRISPR system is the type II CRISPR-Cas9 from *Streptococcus pyogenes*, which has had broad applications across prokaryotes and eukaryotes (Altenbuchner, 2016; Burby & Simmons, 2017; DiCarlo et al., 2013; Jakociūnas et al., 2016; Jakutyte-Giraitiene & Gasiunas, 2016; Jinek et al., 2012; Lu et al., 2019; Mertens et al., 2019; Peng et al., 2017; Peters et al., 2016; Price et al., 2019, 2020; Westbrook et al., 2016; Zhang et al., 2016; Zhu et al., 2017).

Cas9, guided by either a two-component crRNA–tracrRNA complex or a single guide RNA complex for targeting of a specific DNA sequence, induces a blunt DNA double-strand break (DSB) upstream of the protospacer adjacent motif (PAM) (Jinek et al., 2012), which must then be repaired for host survival. The inefficiency or absence of a nonhomologous end-joining (NHEJ) system in most bacterial species predominantly leads to DSB repair by the homology-directed repair (HDR) pathway (Adli, 2018; Altenbuchner, 2016; Burby & Simmons, 2017; Jinek et al., 2012; Shuman & Glickman, 2007; Westbrook et al., 2016). Repair of the DSB via the HDR pathway then permits the introduction of donor DNA (dDNA) through homologous recombination, facilitating the introduction of specific mutations.

While there are reservations about using the commonly employed CRISPR-Cas9 technology for commercial purposes, these concerns might be circumvented by opting for the royalty-free CRISPR nuclease MAD7 (Inscripta, USA). MAD7 (or ErCas12a) is an alternative CRISPR nuclease that has shown success in genome engineering of *B. subtilis* (Price et al., 2020). As a Cpf1 (or Cas12a) nuclease, MAD7 differs from Cas9 in that it has a thymine-rich PAM, forms a staggered DSB and uses only one guide RNA molecule to form a complex with the nuclease for DNA targeting (Zetsche et al., 2015). These features make MAD7 ideal for editing of DNA with low GC content.

1.6 Aims

This study aims to identify novel metalloproteases with characteristics that are desirable for use in the detergents industry. These will be chosen from within the well-characterised zincin family of metalloproteases, as this clan contains a large proportion of endoproteases. Protease sequences will be analysed with models for thermostability (Lin & Chen, 2011), psychrophilicity (Nath & Subbiah, 2014), halophilicity (Nath, 2016) and acidophilicity (Lin et al., 2013). Redundant sequences will be removed using CD-HIT (Li & Godzik, 2006), and sequences will be screened for the presence of an SP by SignalP (Nielsen et al., 1997).

The resultant subset of proteases will be codon optimised and synthesised, as in (Bayer et al., 2009), and expressed in a *B. subtilis* strain engineered for improved protease secretion. The proteases will be fused with an SP library and protease secretion will be measured to find the optimal SP for each candidate.

Finally, protease candidates will be purified, and their suitability as detergent proteases will be assessed by assaying activity in a range of industrially relevant conditions.

Chapter 2 – Materials and methods

This chapter details the materials and experimental methods used throughout this study.

2.1 Strains and plasmids

All strains and plasmids used within this study are given at the beginning of the relevant chapter. All cloning of plasmids was performed in *Escherichia coli* Top10 (Invitrogen (*F*-, *mcrA*, $\Delta(mrr-hsdRMS-mcrBC)$, $\phi80lacZ\Delta M15$, $\Delta lacX74$, *recA1*, *araD139*, $\Delta(araleu)$, 7697, *galU*, *galK*, *rpsL*, (Str^R), *endA1*, *nupG*)).

2.2 Growth and maintenance of bacterial cultures

2.2.1 General growth and storage of bacterial cultures

Lysogeny broth (LB) medium was used for all general-purpose *E. coli* and *Bacillus subtilis* growth cultures. Growth media components are detailed in Appendix A. Growth was monitored by measuring optical density at 600 nm (OD₆₀₀) in a spectrophotometer. Unless otherwise stated, growth cultures were performed in non-baffled 50 mL Erlenmeyer flasks at 37 °C with orbital shaking at 250 rpm. When appropriate, isopropyl β -D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM to induce protein expression. When appropriate, antibiotics were added to the relevant concentration as described in Table 2.1. For optimal protein secretion, SR or SRX medium was used for growth of *B. subtilis* cultures.

Unless otherwise stated, to ensure comparability between strains and replicates, growth was synchronised through the use of a pre-culture to grow diluted overnight cultures to exponential growth phase, in the same conditions as the assay culture. Generally, overnight cultures were diluted to OD₆₀₀ 0.02 for subsequent pre-culture growth. For protein expression, cultures were grown to mid-exponential phase (OD₆₀₀ 0.6-0.8) at 37 °C, 200rpm prior to incubation at 18 °C, 200 rpm. For long term storage of bacterial cultures, overnight LB cultures were supplemented to 10% (v/v) glycerol. These were stored at -80 °C in cryovials.

Table 2.1: Stock and working concentrations of antibiotics used in this study.

Antibiotic	Solvent	Stock concentration (mg/mL)	Working concentration (µg/mL)
Ampicillin (Amp)	70% (v/v) ethanol	100	100
Chloramphenicol (Cam)	100% (v/v) ethanol	40	10
Kanamycin (Kan)	H ₂ O	50	10
Spectinomycin (Spec)	H ₂ O	100	100

2.2.2 Visualisation of *B. subtilis* cells by microscopy

B. subtilis strains were cultured in LB medium overnight at 37 °C, 200 rpm. Cultures were then used to inoculate 10 mL LB medium to OD₆₀₀ 0.01 and incubated at 37 °C, 200 rpm until cultures reached mid-exponential phase (OD₆₀₀ 0.6-0.8). These were then visualised by phase-contrast microscopy (PCM) using the Leica DMI8 inverted microscope (Leica Microsystems, Germany).

2.2.3 Plate reader growth curves

LB medium overnight cultures (10 mL), supplemented with the appropriate antibiotics, were grown at 37 °C, 200 rpm, and used to inoculate 10 mL LB medium pre-cultures to OD₆₀₀ 0.01. These were subsequently grown to mid-exponential phase (OD₆₀₀ 0.6-0.8) at 37 °C, 200 rpm. At this point the cell states have been normalised to improve the reproducibility of experiments, and the pre-cultures were used to inoculate the pre-warmed assay cultures. OD₆₀₀ measurement was carried out in 96-well flat bottom plates (Greiner Bio-One, Austria) using the FLUOstar Omega Microplate Reader (BMG LABTECH, Germany). 200 µL cultures were incubated at 37 °C, 700 rpm. OD₆₀₀ was measured every 90 s and maximum specific growth rate ($\mu_{\max}$) values were calculated by fitting to a parametric growth model using the method of least squares, as described in Equation 2.1, where y_i is the dependent variable, β_0 , is the intercept of the model, X_j corresponds to the j th explanatory variable of the model ($j= 1$ to p), with parameters shown in Table 2.2. Statistical analysis was performed using the growthrates package (v0.8.4; Petzoldt, 2022) of the R Statistical Software (v4.1.1; R Core Team, 2021).

Equation 2.1: Method of least squares

$$y_i = \beta_0 + \sum_{j=1..p} \beta_j X_{ij}$$

Table 2.2: Parameters used in calculation of maximum specific growth rate ($\mu_{\max}$)

Parameter	y0	$\mu_{\max}$	K
Start (p)	0.01	0.2	2
Lower	1e ⁻⁶	0	0
Upper	0.05	5	5

2.2.4 BioLector analysis

The BioLector Pro microfermentation system (m2p-labs, Germany) was used to ensure sufficient aeration in small-scale fermentation studies. Prior to BioLector Pro analysis, single colonies were used to inoculate 500 μ L SRX medium cultures, supplemented with the appropriate antibiotics, and grown overnight in 96-well deep-well plates. The pre-cultures were diluted to OD₆₀₀ 0.1 in 1 mL pre-warmed SRX medium in 48-well FlowerPlate®. These plates were sealed with Breathe-EASIER micro-titre plate seals (Thistle Scientific, UK) and incubated in the BioLector Pro at 37 °C, 900 rpm, 85% humidity. Biomass (excitation: 620 nm, gain: 4) was monitored with measurements taken every 15 minutes. Following growth to mid-exponential phase, cultures were incubated at 18 °C, 900 rpm.

2.3 Bacterial transformations

2.3.1 Preparation and transformation of chemically competent *E. coli* cells

E. coli cells were streaked out on LB agar plate and grown overnight at 37° C. Single colonies were then inoculated in 10 mL LB medium and grown overnight at 37° C, 200 rpm. The overnight bacterial culture was inoculated at 1:400 dilution in pre-warmed LB media, supplemented with 20 mM MgSO₄, and grown at 37 °C, 200 rpm until reaching OD₆₀₀ ~0.48. The bacterial culture was then incubated on ice for 10 min and pelleted by centrifugation at 3752 x *g* for 5 min at 4 °C in an Avanti J-26XP high performance centrifuge (Beckman Coulter, UK). The supernatant was discarded, and the pellet was resuspended in 40 mL of ice-cold TFB1 buffer (30 mM CH₃COOK, 100 mM RbCl, 10 mM CaCl₂, 50 mM MnCl₂, 15% (v/v) glycerol, pH 5.8) per 100 mL cell culture. Following incubation on ice for 5 min, cells were centrifuged at 1351 x *g* for 10 min at 4 °C. The supernatant was discarded, and the pellet was

resuspended in 4 mL of ice-cold TFB2 buffer (10 mM 3-(N-morpholino)propanesulfonic acid, 75 mM CaCl₂, 10 mM RbCl, 15% (v/v) glycerol, pH 6.5) per 100 mL cell culture. 50 µL aliquots of cells were transferred into 1.5 mL tubes and stored at -80 °C after snap freezing in a dry ice/ethanol bath.

To transform the chemically competent *E. coli* cells, an aliquot of competent cells was thawed on ice. DNA was added to the cells and the cells were subjected to heatshock at 42 °C for 90 s, then incubated on ice for 1 min. 400 µL SOC media (Appendix A) was added and the cells were incubated for 1 h at 37 °C, 200 rpm. Cells were then plated on LB agar selection plates and incubated overnight at 37 °C.

2.3.2 Preparation and transformation of competent *B. subtilis*

The natural competence pathway in *B. subtilis* based on the maltose-inducible expression of ComK and ComS was used to prepare competent *B. subtilis*. The method described by Rahmer et al. (2015) was used with some modifications.

Single colonies were used to inoculate 2 mL LB in a 5 mL tube, which was then incubated overnight at 37 °C, 200 rpm. 5 mL LB was inoculated with 500 µL overnight culture and incubated at 37 °C, 200 rpm for 90 min. 5 mL LBM (Appendix A) was added and cultures were incubated at 37 °C, 200 rpm for 90 min. 1 mL of this culture was centrifuged at 2000 x g and the supernatant was removed. The pellet was resuspended in 500 µL LB with 500 ng DNA and incubated at 37 °C, 200rpm. Cells were then plated on LB or PA agar (Appendix A) selection plates.

Transformation efficiency (TE) was determined by counting colonies on transformation plates and applying Equation 2.1, using the count of colonies on a selection plate (Co), the mass of DNA added (m_{DNA}, in µg) and the total dilution factor of DNA prior to plating (DF).

Equation 2.2: Calculation to determine transformation efficiency.

$$TE = \frac{Co}{m_{DNA}} \times DF$$

2.4 Chemicals

Unless otherwise stated, all chemicals and reagents were obtained from Merck (UK).

2.5 DNA manipulation

General molecular biology techniques for DNA manipulation were performed as described in (Sambrook & Russell, 2001). Further detail is provided in the following sections.

2.5.1 Purification of plasmid DNA

Plasmid DNA was purified from *E. coli* overnight cultures using the E.Z.N.A.[®] Plasmid DNA Mini Kit I (VWR International, UK) following the manufacturer's instructions, with some modifications. This utilises a column-based system that contains a silica membrane to which plasmid DNA can be bound, washed, and eluted. 2 mL of bacterial culture was spun down and resuspended in 200 μ L Resuspension Buffer. For *B. subtilis* cultures, the pellet was resuspended in 200 μ L Resuspension Buffer with 1 mg/mL lysozyme and incubated at 37 °C for 10 minutes prior to addition of Lysis Buffer. 250 μ L Lysis Buffer was then added and mixed well by inversion before incubating for 1 min. 350 μ L Neutralisation Buffer was added and mixed well by inversion. This lysate was clarified by centrifugation at 18,000 x *g* for 10 min and the supernatant was added to a spin column. The plasmid DNA was bound to the silica by centrifugation at 18,000 x *g* for 1 minute. The column was washed with 500 μ L Wash Buffer 1 by centrifugation at 18,000 x *g* for 1 minute, before being washed again with 500 μ L of Wash Buffer 2 by centrifugation at 18,000 x *g* for 1 minute. Plasmid DNA was eluted with molecular biology grade H₂O.

2.5.2 Isolation of genomic DNA from *B. subtilis*

To purify genomic DNA from *B. subtilis*, a single colony was cultured in LB medium containing the appropriate antibiotic (where applicable) at 37 °C, 200 rpm. 1 mL culture was harvested at 18,000 x *g* for 1 min at room temperature. The supernatant was removed and the pellet was resuspended in 100 μ L 0.85% (w/v) sodium chloride, supplemented with 5 μ L 10 mg/mL RNase ONE (Promega, UK) and 0.8 μ L of 125 mg/mL Lysozyme (Thermo Fisher Scientific, UK). This was then incubated at 37 °C for 5 min. 300 μ L Cell Lysate Solution (25 mM ethylenediaminetetraacetic acid (EDTA), 2% (w/v) sodium dodecyl sulphate (SDS)) was added and mixed before addition of 168 μ L Protein Precipitation Solution (10 M ammonium acetate). This was mixed by vortex and incubated on ice for 5 min. Cellular debris was removed by centrifugation at 18,000 x *g* for 10 minutes and the supernatant was transferred to a fresh 1.5 mL tube. 600 μ L 100% (v/v) isopropanol was added and mixed well. DNA

was isolated by centrifugation at 18,000 x *g*. The supernatant was removed, and the DNA was washed with 200 µL 70% (v/v) ethanol. Following centrifugation at 18,000 x *g* for 2 min, the supernatant was discarded. Any remaining ethanol was removed by incubation at 55 °C for 10 min. The isolated genomic DNA (gDNA) was resuspended in 100 µL molecular biology grade H₂O and stored at 4 °C.

2.5.3 DNA quantification

DNA concentration was quantified using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific, UK). This was calculated by measuring the absorbance of 1 µL of the sample at 260 nm. The spectrophotometer was blanked using molecular biology grade H₂O. The Beer-Lambert law (Equation 2.1) was used to determine the DNA concentration using the absorbance at 260 nm, the molar extinction co-efficient (ϵ) of double stranded DNA ($0.020 (\mu\text{g/mL})^{-1} \text{ cm}^{-1}$), and the known pathlength (*l*) of 1 mm (auto-ranging to 0.2 mm based on concentration).

Equation 2.3: Beer-Lambert Law

$$A_{280} = \epsilon Cl$$

2.5.4 Oligonucleotides

All single strand DNA (ssDNA) oligonucleotides used in this study were obtained from Integrated DNA Technologies (IDT). All oligonucleotides are described in Appendix B. Oligonucleotides used for single guide RNA (sgRNA) protospacer cloning were designed as described in Section 2.7.2.1. Oligonucleotides used for PCR were designed using the New England Biolabs (NEB, UK) T_m Calculator (NEB, UK) to calculate primer annealing temperatures. Lyophilised oligonucleotides were resuspended in molecular biology grade H₂O. 100 µM stocks were stored at -20 °C, and 10 µM working solutions were stored 4 °C.

2.5.4.1 Phosphorylation and annealing of oligonucleotides

For ligation into pBAC0158, ssDNA oligonucleotide pairs were designed with corresponding overhangs. Oligos were 5' phosphorylated, using T4 Polynucleotide Kinase (NEB, UK), and annealed in a one-pot reaction. The reaction mixture and thermo-cycler protocol utilised for this procedure are described in Table 2.3 & Table 2.4.

Table 2.3: Reaction mixture components for phosphorylation and annealing of ssDNA oligonucleotides

Reaction Mixture		
Component	Volume (μL)	Final Concentration
100 μM oligonucleotide 1	5	10 μM
100 μM oligonucleotide 2	5	10 μM
10X T4 polynucleotide kinase buffer	5	1X
10 mM ATP	5	1 mM
100 mM DTT	2.5	5 mM
50% (w/v) PEG8000	5	5% (w/v)
10 units/μL T4 polynucleotide kinase	1	0.2 units/μL
H ₂ O	21.5	

Table 2.4: Reaction conditions for phosphorylation and annealing of ssDNA oligonucleotides

Reaction Conditions	
Lid temperature: 105 °C	
Temperature (°C)	Time (minutes)
37	30
65	20
90	5
85	1
80	1
75	1
70	1
65	1
60	0.5
55	0.5
50	0.5
45	0.5
40	0.5
35	0.5
30	0.5
25	0.5
20	3

2.5.5 Polymerase chain reaction (PCR)

DNA amplification was performed *in vitro* using the polymerase chain reaction (PCR).

2.5.5.1 High fidelity PCR

DNA to be used in plasmid cloning or genomic integration was prepared using Q5[®] High-Fidelity DNA polymerase (NEB) as per the manufacturer's instructions (Table 2.5). The thermo-cycling conditions for a routine PCR reaction using Q5[®] High-Fidelity DNA polymerase is described in Table 2.6.

Table 2.5: Standard PCR reaction using Q5[®] High-Fidelity DNA polymerase.

Component	Volume (μL)	Final concentration
5x Q5 HF buffer	10	1x
10 μM forward primer	2.5	0.5 μM
10 μM reverse primer	2.5	0.5 μM
10 mM dNTPs	1	200 μM
Template DNA	Variable	< 5 ng
2 units/μL Q5 [®] High-Fidelity DNA polymerase	0.5	20 units/mL
H ₂ O	Up to 50 μL	

Table 2.6: Standard thermocycling reaction conditions for use with Q5[®] High-Fidelity DNA polymerase.

Step	Temperature (°C)	Time	Cycles
Initial denaturation	98	30 s	1
Denaturation	98	10 s	35
Primer annealing	45-72	30 s	
Extension	72	30 s per kbp	
Final extension	72	2 min	1

2.5.5.2 Colony PCR

To identify colonies in which cloning of a DNA fragment into a vector had been successful, or to find colonies that had effectively had a region of DNA deleted or inserted, colony PCR was used. A colony was picked and suspended in 10 μL LB

medium with the relevant antibiotic and used as the PCR template. This colony PCR (cPCR) procedure was performed using OneTaq® Quick-Load® 2x Master Mix with Standard Buffer (NEB, UK). Table 2.7 details the composition of the reaction mix, and Table 2.8 the thermo-cycling reaction conditions when using OneTaq® Quick-Load® 2x Master Mix with Standard Buffer.

Table 2.7: Standard cPCR reaction using OneTaq® Quick-Load® 2x Master Mix with Standard Buffer.

Component	Volume (µL)	Final concentration
OneTaq® Quick-Load® 2x Master Mix with Standard Buffer	10	1x
10 µM forward primer	1.5	0.75 µM
10 µM reverse primer	1.5	0.75 µM
Colony suspension template DNA	Variable	
H ₂ O	Up to 20 µL	

Table 2.8: Standard cPCR thermo-cycling reaction conditions for use with OneTaq® Quick-Load® 2x Master Mix with Standard Buffer.

Step	Temperature (°C)	Time	Cycles
Initial denaturation	94	5 min	1
Denaturation	94	30 s	30
Primer annealing	45-68	30 s	
Extension	68	1 min per kbp	
Final extension	68	5 min	1

2.5.5.3 Overlap extension PCR

DNA fragments, prepared through PCR, were assembled together in a two-step process known as overlap extension PCR (OE-PCR). Fragments shared 20 bp with one hundred percent homology at the extreme end where two fragments were to be fused. The fragments are then added together at the same molar ratio, and the regions of homology act as priming regions in the first PCR phase (Tables 2.9 & 2.10). Oligonucleotide primers are added during a pause in the thermo-cycling

program, hybridising to the 5' end of the first fragment and the 3' end of the second fragment. These primers amplify the full construct during the second PCR phase. The products of OE-PCR were then isolated for purification by agarose gel electrophoresis (Section 2.5.6).

Table 2.9: Standard OE-PCR reaction using Q5® High-Fidelity DNA polymerase.

	Component	Volume (µL)	Final concentration
Phase 1	5x Q5 HF buffer	10	1x
	Template PCR 1	Variable	2.2 ng/µL
	Template PCR 2	Variable	
	10 mM dNTPs	1	200 µM
	2 units/µL Q5® High-Fidelity DNA polymerase	0.5	20 units/mL
	H ₂ O	Up to 45 µL	
Phase 2	10 µM forward primer	2.5	0.5 µM
	10 µM reverse primer	2.5	0.5 µM

Table 2.10: Standard OE-PCR thermo-cycling reaction conditions for use with Q5® High-Fidelity DNA polymerase.

	Step	Temperature (°C)	Time	Cycles
Phase 1	Initial denaturation	98	30 s	1
	Denaturation	98	10 s	15
	Primer annealing	45-72	30 s	
	Extension	72	30 s per kbp	
Pause to add oligonucleotides				
Phase 2	Denaturation	98	10 s	20
	Primer annealing	45-72	30 s	
	Extension	72	30 s per kbp	
	Final extension	72	5 min	1

2.5.6 Agarose gel electrophoresis

Agarose gel electrophoresis was used to analyse DNA. 0.8 g UltraPure agarose (Invitrogen, UK) was suspended in 100 mL Tris-Acetate-EDTA (TAE) buffer (40 mM Tris, 20 mM Acetic acid, 1 mM EDTA) and dissolved by heating in a microwave. 10 μ L Thiazole Orange (1-Methyl-4-[(3-methyl-2(3*H*)-benzothiazolylidene)methyl]quinolinium *p*-tosylate) was added to facilitate DNA visualisation. The gel was cast and, once set, the DNA containing solution (with addition of gel loading dye, purple (NEB, UK) to 1x concentration) was added. The loaded gel then underwent electrophoresis at 100 V for 30 minutes in TAE buffer. DNA bound with Thiazole Orange was visualised with the Safe Imager™ 2.0 Blue Light Transilluminator (Thermo Fisher Scientific, UK) system (excitation: 470 nm, emission: 530 nm). DNA band sizes were determined by comparison with 1 kb DNA Ladder (NEB, UK).

2.5.7 Purification of DNA following agarose gel electrophoresis

Following separation by agarose gel electrophoresis, DNA was excised using a razor blade and subsequently purified using the Zymoclean™ Gel DNA Recovery Kit (Cambridge Bioscience, UK), following the manufacturer's instructions. This kit uses columns containing a silica DNA binding matrix to which DNA binds, is washed and then eluted. The gel slice was added to a 2 mL tube and 3 volumes Agarose Dissolving Buffer was added (1 mg agarose gel equates to 1 μ L buffer). This mixture was incubated for 15 min at 55 °C, until complete dissolution of the gel slice, prior to addition to a spin column. The DNA was then bound to the column membrane by centrifugation at 18,000 x *g* for 1 min. The DNA was washed twice with 200 μ L DNA Wash Buffer, with centrifugation steps of 18,000 x *g* for 1 min after addition of buffer. DNA was then eluted with molecular biology grade H₂O into a 1.5 mL tube.

2.5.8 DNA assembly methods

DNA was assembled via traditional restriction/ligation and Gibson Assembly (Gibson et al., 2009).

2.5.8.1 Restriction endonuclease-mediated digestion of DNA

Restriction endonucleases can cleave DNA fragments either in a blunt or staggered fashion, resulting in the exposure of either blunt or "sticky ends". Type IIP restriction endonucleases cleave DNA within their target region, leaving a consistent region for facilitating the cloning of DNA fragments. Conversely, Type IIS restriction

endonucleases cleave DNA outside their recognition site, leading to potentially different "sticky ends" among DNA fragments cut by the same type IIS restriction endonuclease. To join DNA fragments together, ligases are used at compatible sites. For ligases to effectively combine DNA fragments cut in this manner, compatible "sticky ends" are necessary. On the other hand, bluntly cleaved DNA can be ligated to another bluntly cleaved DNA fragment, provided there is a free 5' phosphate group available to facilitate the connection of the DNA backbone (Loenen et al., 2014). DNA digestion with restriction endonucleases was carried out according to the manufacturer's recommendations, and unless specified otherwise, all restriction nucleases were sourced from NEB.

2.5.8.2 Dephosphorylation of DNA

Vector DNA for use in restriction/ligation was 5' dephosphorylated to improve efficiency of DNA insertion by prevention of vector self-ligation. This was carried out using Antarctic Phosphatase (NEB, UK), following manufacturer's instructions. DNA was then separated by agarose gel electrophoresis (Section 2.5.6) and purified (Section 2.5.7).

2.5.8.3 Ligation of restriction-digested DNA

T4 DNA Ligase (NEB, UK) was used to ligate DNA fragments. A total DNA mass of 100 ng was used in each reaction, with a molar ratio of 3:1 insert:vector (Table 2.11). Reactions were incubated at room temperature for 15 min prior to heat inactivation at 65 °C for 10 min and subsequent transformation of *E. coli* cells (Section 2.3.1).

Table 2.11: Conditions of T4 DNA Ligase reaction.

Component	Volume (μL)	Final concentration
Vector DNA	Variable	Variable
Insert DNA	Variable	3x mM vector DNA
10x T4 DNA Ligase Buffer	2	1x
400 units/μL T4 DNA Ligase	1	20 units/μL
H ₂ O	Up to 20 μL	

2.5.8.4 Gibson DNA Assembly

Gibson Assembly (Gibson et al., 2009) was used to assemble of multiple DNA fragments with overlapping regions in a one-pot isothermal reaction. Typically, 20

base pairs of homology were designed between adjacent fragments. For all assembly reactions, the Gibson Assembly® Master Mix (NEB, UK), was employed. The reaction master mix was diluted in an equal volume of the DNA fragments intended for assembly. The insert DNA was combined with the vector in a 3:1 molar ratio. Subsequently, all reactions were incubated at 50 °C for 15 min before being utilised for the transformation of *E. coli* cells (Section 2.3.1).

2.6 DNA Sequencing

Sanger sequencing of DNA was primarily performed by Azenta (UK).

2.7 Gene editing in *B. subtilis*

2.7.1 Homologous recombination

Gene deletion was achieved in *B. subtilis* through transformation with DNA containing an antibiotic resistance cassette and homology arms upstream and downstream of the target site. Transformants were plated onto LB agar containing the relevant antibiotic and gene deletion was confirmed through cPCR followed by sequencing of the PCR-amplified section of genomic DNA.

2.7.2 CRISPR-MAD7 mediated gene editing

The methods described in Price et al. (2020) were followed for CRISPR-MAD7 mediated gene editing in *B. subtilis*.

2.7.2.1 sgRNA protospacer design

The online tool, CHOPCHOP, has been adapted for use with *B. subtilis* and used for protospacer design (Labun et al., 2016; Montague et al., 2014). CHOPCHOP is applicable for AsCpf1 protospacer identification, and due to the similarities between the PAM sites for AsCpf1 and MAD7, CHOPCHOP can also identify appropriate sites for MAD7 based experiments. For identification of protospacer sequences, the target DNA sequence is uploaded and candidate protospacers are scored based on GC percentage, predicted efficiency, detection of off-target effects and level of self-complementarity to identify stem loops in the eventual RNA sequence. However, as this software was designed for AsCpf1, the scoring system may not be relevant for MAD7 until this is further characterised. Where possible, a PAM site was then selected at a similar distance from the 5'- and 3'-ends of the target area.

2.7.2.2 Cloning of protospacer DNA into MAD7 plasmids

The MAD7 parental plasmid was digested with Sapl (Section 2.5.8.1) and dephosphorylated (Section 2.5.8.2). Oligonucleotides were designed containing the protospacer region and 3 bp on the 5' and 3' to generate single stranded overhangs compatible with Sapl sticky ends. These oligonucleotides were annealed and phosphorylated (Section 2.5.4.1) prior to ligation into the Sapl digested parental plasmid. Successful insertion was confirmed by cPCR (Section 2.5.5.2) followed by Sanger sequencing. Final plasmids for CRISPR-MAD7 are shown in Table 2.11.

Table 2.11: CRISPR-MAD7 plasmids used throughout this study.

Plasmid	Target gene	PAM		Oligonucleotides used for cloning protospacer DNA
		Sequence (5'-3')	Distance from start codon	
pBAC0158\amyE2	<i>amyE</i>	TTTC	1333 bp downstream	crMAD7_Bs_amyE_F2/ crMAD7_Bs_amyE_R2
pBAC0158\mpr1	<i>mpr</i>	TTTC	155 bp downstream	oMAP0847/0848
pBAC0158\spec1	<i>aad9</i>	TTTC	386 bp downstream	crMAD7_Bs_aad9_F1/ crMAD7_Bs_aad9_R1

2.7.2.3 Donor DNA (dDNA) preparation

Linear dDNA was prepared by OE-PCR (Section 2.5.5.3), with modifications of interest introduced into the primers used for the first stage PCR products, and 1 kbp regions of homology designed to allow homologous recombination into the target region.

2.7.2.4 Co-transformation of *B. subtilis*

B. subtilis was induced to a state of natural competency (Section 2.3.2) and transformed with 200 ng of the editing plasmid along with 1 µg dDNA and plated onto LB agar containing the relevant antibiotic.

2.7.2.5 Curing of editing plasmid

Plasmid loss of the CRISPR-MAD7 editing plasmids was achieved in the edited strain by culturing in LB medium supplemented with 1 mM IPTG overnight at 37 °C,

200 rpm. This introduces the pressure of CRISPR-MAD7 production in the absence of antibiotic selection, leading to decreased fitness in cells harbouring the editing plasmid. Plasmid loss was then confirmed by cross-plating on LB agar plates with and without supplementation of the relevant antibiotic. Following incubation at 37 °C, colonies that had lost the editing plasmid could be distinguished by growth on LB without antibiotic and lack of growth on LB with antibiotic. This was then confirmed through culturing in LB medium with and without the antibiotic at 37 °C, 200 rpm.

2.8 Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Protease samples were visualised through SDS-PAGE and protease activity was visualised by zymography.

2.8.1 Protease sample denaturation by sulphuric acid

Protein samples are routinely boiled prior to electrophoresis in SDS-PAGE, however, in protease-containing samples it has been found that this boiling step can lead to autolysis of the protease. This affects accuracy of results and results in a large number of bands due to autolysis products. As an alternative to boiling, denaturation with sulphuric acid has been shown to reduce the level of autolysis prior to SDS-PAGE (Fujita et al., 2019). The method developed by Fujita et al. (2019) was adapted for use in this study.

To 19 µL protease-containing sample, 1 µL 10 M sulphuric acid was added and incubated at room temperature for 10 min. This is followed by addition of 10 µL 5x Laemmli buffer (4% (w/v) SDS, 250 mM Tris-HCl, 890 µM bromophenol blue, 500 mM dithiothreitol, 30% (v/v) glycerol, pH 6.8) and incubation at room temperature for a further 10 min. To this, 20 µL 2 M Tris base was added to increase the pH to within suitable range for separation by SDS-PAGE.

2.8.2 Protease sample concentration by trichloroacetic acid (TCA) precipitation

Another method of reducing autolysis in protease-containing samples, which simultaneously concentrates the protein sample, is precipitation with TCA. To 1 mL protease-containing sample, 100 µL 100% (w/v) TCA was added and the sample was mixed by vortexing. The protein was then precipitated by incubation on ice for 30 min, and precipitated protein was pelleted by centrifugation at 10,000 x *g* at 4 °C

for 10 min. The supernatant was removed and the pellet resuspended in 50 μ L 2x Laemmli buffer.

2.8.3 Separation of protein samples by SDS-PAGE

Protein samples prepared by denaturation with sulphuric acid or TCA were separated based on their molecular weight using the Mini-PROTEAN® TGX™ Precast Protein Gel system (Bio-Rad, UK), with 1 mm gel thickness and 4-20% polyacrylamide concentration. Proteins were separated by electrophoresis at 120 V for 1 hour in TGS running buffer (24.8 mM Tris, 192 mM glycine, 0.1% (w/v) SDS).

2.8.4 Visualisation of protein by staining with InstantBlue™

Protein samples separated by the Mini-PROTEAN® TGX™ Precast Protein Gel system were visualised using the colloidal Coomassie stain, InstantBlue™ (Abcam, UK). This was directly applied to the gel and incubated at room temperature with gentle agitation for 15-30 min. Following incubation, the InstantBlue™ protein stain was removed, the gel was washed with deionised H₂O and de-stained in deionised H₂O overnight at room temperature with gentle agitation.

2.8.5 Visualisation of protein by western blot

Immunolabelled proteins were detected by western blot. Following SDS-PAGE, gels were incubated in semi-dry transfer buffer (48 mM Tris, 39 mM Glycine, 20% (v/v) methanol, 0.025% (w/v) SDS, pH 8.3) for 20 min at 4 °C. Proteins were transferred from gel onto a nitrocellulose membrane by semi-dry transfer using a Trans-Blot® SD Semi-Dry Transfer Cell with a voltage of 10 V for 25 min. Membranes were then incubated in blocking buffer (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na₂HPO₄, 0.1% (v/v) Tween-20, 5% (w/v) skimmed milk powder) for 45 minutes with gentle agitation. Membranes were incubated with blocking buffer and Monoclonal Anti-polyHistidine antibody (Merck, UK), diluted 1/10,000 at 4 °C for 18 h. Nitrocellulose membranes were washed three times with 1xPBS/Tween-20 (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na₂HPO₄, 0.1% (v/v) Tween-20) at room temperature for 10 min with gentle agitation, then incubated with blocking buffer and IRDye® 800CW Donkey anti-Mouse IgG Secondary Antibody (LI-COR Biosciences, UK) at 1/10,000 dilution at room temperature for 1 h. Membranes were washed once with 1xPBS/Tween-20 at room temperature for 10 min with gentle agitation, then twice in 1xPBS (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na₂HPO₄) at room temperature for 10 min with gentle agitation.

Membranes were then imaged in the Odyssey[®] CLx Imaging System (LI-COR Biosciences, UK) at a wavelength of 800 nm.

2.8.6 Separation of protease samples by gelatin zymography

To separate a protein sample and visualise protease activity, gelatin zymography was employed. This method uses a polyacrylamide gel containing gelatin to separate a protein sample by SDS-PAGE. The gel is then washed in buffer containing the nonionic surfactant Triton X-100 to remove SDS and the gel is incubated to facilitate proteolysis of the gelatin substrate by any proteases with gelatinase activity within the sample. The gel is stained to visualise the remaining gelatin such that protease activity may be observed by zones of clearing. In this study, the method developed by Abcam (Abcam, UK) was followed (Abcam, 2021).

Zymogram gels were prepared with 7 mL separating gel (375 mM Tris-HCl, 12.5% (w/v) acrylamide, 1 mg/mL gelatin, 0.1% (w/v) SDS, 0.1% (w/v) ammonium persulphate (APS), 0.125% (v/v) tetramethylethylenediamine (TEMED), pH 8.8) and 2 mL stacking gel (125 mM Tris-HCl, 4% (w/v) acrylamide, 0.1% (w/v) SDS, 0.1% (w/v) APS, 0.2% (v/v) TEMED, pH 6.8). 5x non-reducing sample buffer (4% (w/v) SDS, 20% (v/v) glycerol, 150 μ M bromophenol blue, 125 mM Tris-HCl, pH 6.8) was added to the protein samples to 1x concentration and separated by electrophoresis at 120 V for 1 hour in TGS running buffer. Gels were then washed twice with washing buffer (50 mM Tris-HCl, 2.5% (v/v) Triton X-100, 5 mM CaCl₂, 1 μ M ZnCl₂, pH 7.5) for 30 min at room temperature with gentle agitation, rinsed with incubation buffer (50 mM Tris-HCl, 1% (v/v) Triton X-100, 5 mM CaCl₂, 1 μ M ZnCl₂, pH 7.5) and incubated overnight at 37 °C in incubation buffer. Gels were then stained with staining solution (40% (v/v) methanol, 10% (v/v) acetic acid, 0.5% (w/v) Coomassie Brilliant blue G-250) for 30-60 min, rinsed with deionised H₂O and destained with destaining solution (40% (v/v) methanol, 10% (v/v) acetic acid) until bands were visible.

2.9 Peptide sequencing

2.9.1 Identification of propeptide domains

Bands in SDS-PAGE gels were analysed by LC-MS/MS to determine the peptide sequence by the Mass Spectrometry and Proteomics facility at the University of St. Andrews.

2.9.2 Protease cleavage site identification

To determine protease cleavage site specificity, aldolase from rabbit muscle at 5 mg/mL was digested with the protease sample at 0.1 mg/mL for 1 hour at 37 °C in Tris-HCl, pH 9.0. Digests were analysed by electrospray ionisation by the Mass Spectrometry and Proteomics facility at the University of St. Andrews, to determine the sequence of cleavage products.

2.10 Protease activity detection

2.10.1 Qualitative detection of protease activity on skimmed milk agar

Protease activity was visualised qualitatively by plating protease-secreting *B. subtilis* colonies on PA agar (Appendix A), as in Zaghloul et al. (2011). Following transformation of *B. subtilis* (Section 2.3.2), transformants were plated onto PA agar supplemented with 1% (w/v) Oxoid™ skimmed milk powder (Thermo Fisher Scientific, UK) and protease secretion and activity was observed by zones of clearing around colonies.

2.10.2 Quantification of protease activity by the azocasein assay

Azocasein is a protease substrate composed of soluble casein labelled with azo dye, such that proteolysis releases the dye and protease activity can be quantified by absorbance at 420 nm. Protease activity was detected by the azocasein assay following the method developed by Daroit et al. (2009), modified for use in micro-titre plates.

The buffer systems used were as follows: 100 mM sodium acetate (pH 5.0), MES (pH 6.0), Tris-HCl (pH 7-9), sodium bicarbonate (pH 10.0). Azocasein was dissolved in buffer of choice to 1% (w/v). 20 µL buffer of choice, 60 µL 1% (w/v) azocasein and 20 µL protease sample were incubated at assay temperature for 1 hour. The reaction was then terminated by addition of 120 µL 10% (w/v) TCA. Azocasein was precipitated on ice for 10 min, and precipitated protein was separated by centrifugation at 3,500 x g for 10 min at 4 °C. 160 µL supernatant was then added to 40 µL 1.8 N NaOH in well of a micro-titre plate and A₄₂₀ was measured using the FLUOstar Omega Microplate Reader (BMG LABTECH, Germany).

2.10.3 Quantification of protease activity by the Protazyme AK assay

The Protazyme AK assay (Megazyme, Ireland) is more representative of stain samples, as it uses insoluble casein. The protease substrate is produced as tablets

of insoluble casein labelled with azurine dye that can be detected following proteolysis by absorbance at 590 nm. In this study this assay was followed according to the manufacturer's instructions, with modifications for use in micro-titre plates. Each tablet was suspended in 1 mL buffer with stirring, and 100 μ L of this suspended substrate was added to 100 μ L buffer (or additive of choice) and 20 μ L protease sample at 100 μ g/mL (pre-incubated with additive of choice). This mixture was incubated at 40 °C, 900 rpm for 30 min, then incubated at 4 °C for 10 min. The insoluble substrate was separated by centrifugation at 2,500 x *g* for 20 min at 4 °C. 100 μ L supernatant was then transferred to a micro-titre plate and A_{590} was measured using the Varioskan® Flash Microplate Reader (Thermo Fisher Scientific, UK).

2.11 Determination of protein stability

2.11.1 Determination of pH stability by azocasein assay

Protease pH stability was assayed by incubation in a buffer range of pH 5-10 for 1-3 hours. The buffer was then exchanged to a suitable assay pH and protease activity was determined by the azocasein assay (Section 2.10.2).

2.11.2 Determination of NaCl stability by azocasein assay

NaCl stability was assayed by incubation of the protease in an NaCl range of 0-20% (w/v) for 144 h. The buffer was then exchanged to remove NaCl and protease activity was determined by the azocasein assay (Section 2.10.2).

2.11.3 Determination of melting temperature by nano differential scanning fluorimetry (nanoDSF)

NanoDSF is a technique used to determine the temperature at which a protein unfolds (Bruce et al., 2019). This relies on the bathochromic shift of tryptophan residues within non-polar regions of a protein molecule such that when they are buried within the molecule, they exhibit a fluorescence emission maximum of 330 nm. Upon unfolding of the protein, tryptophan residues are exposed to solvent and the emission maximum shifts to 350 nm (Burstein et al., 1973). During nanoDSF a temperature ramp is applied to the protein sample and fluorescence is measured. The ratio of fluorescence intensities at 350 and 330 nm can then be used to determine the temperature at which 50% of protein molecules are unfolded (the melting temperature, or T_m). In this study, the Prometheus NT.Plex (Nanotemper,

Germany) was used according to manufacturer's instructions. Samples were incubated at 0.1 mg/mL in the relevant conditions for 30 min at room temperature prior to measurement.

2.12 Protease purification techniques

2.12.1 Ammonium sulphate precipitation

The percentage of ammonium sulphate saturation at which protease candidates precipitate was calculated by addition of ammonium sulphate as in England & Seifter (1990). Precipitates were then resuspended in appropriate buffer and protease activity was determined by the azocasein assay (Section 2.10.2).

2.12.2 Production and purification of protease by immobilised metal ion chromatography (IMAC)

The protease candidates Pro_10 and Pro_10 Δ C-term were purified from supernatant samples by nickel-nitrilotriacetic acid (Ni-NTA) affinity chromatography, a form of IMAC that utilises the interaction between the His-tag and immobilised nickel ions on a resin to selectively capture the target protein. *B. subtilis* strain BA04 harbouring the relevant plasmid was cultured overnight from a single colony in LB medium containing chloramphenicol at 37 °C, 200 rpm. Cultures were used to inoculate 1 L LB medium containing chloramphenicol to OD₆₀₀ 0.01. These were grown to mid-exponential phase (OD₆₀₀ 0.6-0.8) at 37 °C, 200 rpm, then chilled at 4 °C for 15 min. IPTG was added to 1 mM final concentration and cultures were incubated at 18 °C for 18 h. Cell mass was then separated by centrifugation at 5,000 x *g* for 15 min and the supernatant was isolated. Protein was precipitated by addition of ammonium sulphate to the relevant saturation and precipitated protein was separated by centrifugation at 25,000 x *g* for 30 min. The supernatant was discarded, and precipitates were resuspended in 200 mL binding buffer (20 mM sodium phosphate, 1 M NaCl, pH 7). To clarify the samples through removal of precipitated protein, the samples were then centrifuged at 25,000 x *g* for 30 min and the supernatant was passed through a MF-Millipore® Membrane Filter, 0.45 µm, followed by a MF-Millipore® Membrane Filter, 0.22 µm. Samples were then loaded onto a 5 mL HisTrap™ High Performance column (Merck, UK) and eluted on a linear gradient with elution buffer (1M imidazole, 20 mM sodium phosphate, 1 M NaCl, pH 7.0). The absorbance at 280 nm was observed. Throughout the purification process, 5 mL fractions were collected using the ÄKTA Start protein purification system

(Merck, UK). Successful isolation of the protease was confirmed by SDS-PAGE (Section 2.8.3). IMAC purified samples were then concentrated using Amicon® Ultra-15 (3 kDa molecular weight cut-off) spin columns, and either buffer exchanged 1/10,000 with storage buffer (20 mM sodium phosphate, 1.5 M NaCl, pH 7) or further purified by size exclusion chromatography (SEC) (Section 2.12.3).

2.12.3 Purification of protease by SEC

SEC, or gel-filtration, employs isocratic elution to separate proteins based on their respective sizes, with larger proteins being eluted from the column prior to smaller proteins. Concentrated samples following IMAC purification (Section 2.12.2) were loaded onto a Superdex 75 pg 120 mL 16/600 size exclusion column (Merck, UK) and eluted in storage buffer (20 mM sodium phosphate, 1.5 M NaCl, pH 7). The absorbance at 280 nm was recorded, and 1 mL fractions were collected using the ÄKTA Pure protein purification system (Merck, UK). Protease activity of each fraction was measured using the azocasein assay (Section 2.11.1). Successful isolation of the protease was then confirmed by SDS-PAGE (Section 2.8.3) and pure, active fractions were pooled and concentrated.

2.12.4 Quantification of protease concentration

Protein concentration was quantified using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific, UK). This was calculated by measuring the absorbance of 1 µL of the sample at 280 nm. The spectrophotometer was blanked using the relevant buffer. The Beer-Lambert law (Equation 2.1) was used to determine the protein concentration using the absorbance at 280 nm, the molar extinction coefficient (ϵ) of the protein – predicted using the ProtParam tool (Gasteiger et al., 2005) – and the known pathlength (l) of 1 mm (auto-ranging to 0.2 mm based on concentration).

2.13 Stain removal by tergotometer

To determine the effectiveness of protease samples at removal of stains from cloth, a tergotometer was utilised, which mimics a washing machine through application of a constant temperature in a water bath that contains cylindric vessels in which washing materials are mixed and in which a paddle rotates at 100 rpm to mimic the agitation in a top-loading instrument.

A white A4 knitted cotton stain sheet containing ten 2x2 cm protease-sensitive stains (aged for four days) was used (Warwick Equest, UK). The protease enzymes were diluted in 50 mM Tris-HCl buffer, pH 9.0 after which 500 µL of the solution was applied onto the stain. A control sheet where 50 mM Tris-HCl buffer pH 9.0 was applied without addition of protease was prepared in parallel. The sheets were left at room temperature overnight before laundering in a tergotometer, 1 L wash in 50 mM Tris-HCl buffer, pH 9.0 at 40°C. The stain sheets were dried at room temperature overnight before scanning with the ArtixScan F2 Microtek scanner (Microtek, Taiwan) to generate L^* , a^* and b^* colour coordinates as defined by the CIE (Commission Internationale d'Eclairage or International Commission on Light), where L^* represents darkness to lightness, a^* represents greenness to redness, and b^* represents blueness to yellowness. Percentage cleaning was then calculated as follows:

DeltaE* values between unstained fabric and the unwashed and washed stains were calculated using the CIE 2000 colour difference formula (Equation 2.3) (Luo et al., 2001), using the ColorNameR package (v0.1.0; Sánchez Beeckman, 2021) of the R Statistical Software (v4.1.1; R Core Team, 2021).

Equation 2.4: CIE 2000 colour difference formula.

$$\Delta E^* = \sqrt{\left(\frac{\Delta L'}{k_L S_L}\right)^2 + \left(\frac{\Delta C'}{k_C S_C}\right)^2 + \left(\frac{\Delta H'}{k_H S_H}\right)^2 + R_T \frac{\Delta C'}{k_C S_C} \frac{\Delta H'}{k_H S_H}}$$

where

$$\begin{aligned} L' &= L^* & a' &= a^*(1 + G) & G &= 0.5 \left(1 - \sqrt{\frac{\bar{C}^* ab^7}{\bar{C}^* ab^7 + 25^7}} \right) \\ b' &= b^* & C' &= \sqrt{(a')^2 + (b')^2} & h' &= \tan^{-1} \left(\frac{b'}{a'} \right) \end{aligned}$$

with weighting coefficients defined as:

$$S_L = 1 + \frac{0.015(\bar{L}' - 50)^2}{\sqrt{20 + (\bar{L}' - 50)^2}} \quad S_C = 1 + 0.045\bar{C}' \quad S_H = 1 + 0.015\bar{C}'T$$

$$T = 1 - 0.17\cos(\bar{h}' - 30) + 0.24\cos(2\bar{h}') + 0.32\cos(3\bar{h}' + 6) - 0.20\cos(4\bar{h}' - 63)$$

and rotation factor (RT) determined as below (hue angle and Θ are in degrees):

$$R_T = -\sin(2\Delta\theta)R_C \quad \Delta\theta = 30\exp\left(-\left(\frac{\bar{h}'-275}{25}\right)^2\right) \quad R_C = 2\sqrt{\frac{\bar{c}'^7}{\bar{c}'^7+25^7}}$$

The bar over the values indicates the average value of the two colors.

The parametric coefficients k_L , k_C , and k_H can vary according to the actual measurement conditions. k_C and k_H were set to 1 under the standard conditions specified for CIE 2000, and k_L was set to 1.5 as is standard for textile applications (Luo, 2002).

Stain removal index (SRI) values were then calculated using Equation 2.4, where ΔE^* represents the colour difference determined by the CIE 2000 colour difference formula (Equation 2.3). This represents the difference in colour between the stained fabric and standard white.

Equation 2.5: Formula for calculating SRI values.

$$SRI = 100 - \Delta E^*$$

Percentage cleaning was calculated using Equation 2.5.

Equation 2.6: Formula for calculation of percentage cleaning.

$$PC = \frac{SRI_{washed} - SRI_{unwashed}}{(100 - SRI_{unwashed})} * 100$$

Values were blanked by the negative (no enzyme) control using Equation 2.6.

Equation 2.7: Formula to blank percentage cleaning by no-enzyme control.

$$PC_{blanked} = PC - PC_{ctrl}$$

2.14 *In silico* data analysis

All data calculations were carried out using the R Statistical Software (v4.1.1; R Core Team, 2021), and represented in graphs using the ggplot2 package (Ginestet, 2011) or Microsoft Excel (Microsoft, USA).

In silico DNA manipulation was performed using Benchling (Benchling, USA). Alignment of sequence data with reference sequences was performed using Benchling (Benchling, USA).

Phylogenetic trees were constructed using the phylogeny.fr online tool (Dereeper et al., 2008).

Multiple sequence alignment was performed using ClustalW (Thompson, Higgins & Gibson, 1994) using the genome.jp online tool, with a word size of 1, window size of 5, gap penalty of 3, top diagonals of 5 and percentage scoring method.

3D protein structures were modelled using Alphafold 2 (Jumper et al., 2021).

Pairwise structural alignments using jFATCAT (Li et al., 2020) were performed using the RSCB PDB Pairwise Structure Alignment tool (Berman et al., 2000). 3D protein structures were visualised using Mol* (Sehnal et al., 2021).

Chapter 3 – *Bacillus subtilis* strain engineering

3.1 Introduction

Expression of heterologous proteases has been achieved in many different microbial chassis, including *Escherichia coli* (Lin et al., 2009; Toplak et al., 2013), *Bacillus subtilis* (Tian et al., 2019; Zaghloul et al., 2011), *Pichia pastoris* (Fernández et al., 2013; Olędzka et al., 2003) and *Yarrowia lipolytica* (Ni et al., 2009; Yu et al., 2013). We decided to focus on expression in *B. subtilis*, as it has a robust secretory capability, with reported secretion levels up to 25 g/L in *Bacillus* species, a well understood mechanism of natural genetic competence, and has a host of available tools for genetic manipulation (Hohmann et al., 2017; Schallmey et al., 2011). In addition to this, it has a long history as a host for expression of industrial proteases, with most current commercial proteases being expressed in *B. subtilis* or other *Bacillus* species (Maurer, 2004).

B. subtilis produces at least fourteen native proteases, seven of which are secreted (Pohl et al., 2013). For expression of heterologous proteases, a strain deficient in native proteases would be desired, as this would both reduce background protease activity and would prevent unwanted proteolysis of the protease of interest. In addition, the ideal strain would have a high transformation efficiency to facilitate high-throughput screening for directed evolution of the target protease and would have a high growth rate to minimise culturing time. Finally, it has been seen that some *B. subtilis* strains can form extended filaments of unseparated cells (Reuß et al., 2017), which could interfere with certain methods of directed evolution that involve encapsulation of single cells, such as in-vitro compartmentalisation (Miller et al., 2006). The ideal strain would therefore have a normal cell morphology to ensure compatibility with these techniques.

Improved yields of high-value products have been achieved through large-scale genome deletion of non-essential genes (Aguilar et al., 2020; Juhas et al., 2014; Komatsu et al., 2010; Leprince et al., 2012; Trinh et al., 2008). One such example is the *B. subtilis* strains developed by Reuß et al. (2017). Many of these fulfil some or all of the criteria listed above, namely IIG-Bs27-24, IIG-Bs27-31 and PG10, with total genome deletions of 17%, 22% and 36%, respectively. Full genome sequences of

these strains can be found in Appendix C. IIG-Bs27-24 has a maltose-inducible overexpression system for *comK/comS* – two genes involved in regulation of natural competency – that has been shown to improve transformation efficiency (Rahmer et al., 2015). However, this strain expresses two extracellular proteases: *aprE* and *mpr*. IIG-Bs27-31 also contains the *comK/comS* overexpression cassette and is deficient in all native extracellular proteases. PG10 is similarly deficient in all extracellular proteases and contains the maltose-inducible *comK/comS* cassette, but due to the extensive genome deletions it forms extended filaments.

In addition to deletion of native proteases, secretion of heterologous proteins by *B. subtilis* has been improved through overexpression of molecular chaperones. For example, overexpression of the foldase *prsA*, a Sec pathway component that is responsible for folding of secreted proteins, by Chen et al. (2015) improved secretion of two α -amylases (AmyL and AmyS) 3.2- and 5.5-fold, respectively. An ideal protease-production strain would therefore include such a system to improve protease secretion. In this chapter, a protease candidate (Pro_20, identified in Chapter 4) was expressed in *B. subtilis* to compare protease activity.

In recent years, there has been a growing adoption of clustered regularly interspaced short palindromic repeats (CRISPR)-associated (Cas) systems. This technology has significantly improved genome editing techniques, especially in industrially relevant microorganisms like *B. subtilis* (Jinek et al., 2012). Although there are some concerns regarding the application of the widely used CRISPR-Cas9 for commercial applications due to the intellectual property landscape, this may be avoided through use of the royalty-free CRISPR nuclease MAD7 (Inscripta, USA). This has been applied in *B. subtilis* for precise genome modifications (Price et al., 2020).

In this chapter, we have tested several minimal *B. subtilis* strains to investigate their suitability as hosts for protease production. We then engineered a strain deficient in all native extracellular proteases in order to reduce background protease activity and prevent native degradation of the heterologously expressed proteases. We then investigated a method of improving secretion of heterologous proteases through overexpression of the Sec pathway component *prsA*. For these modifications, we employed CRISPR-MAD7 for precise, marker-free genome editing.

3.2 Material & Methods

3.2.1 Strains and plasmids

The strains and plasmids used in this chapter can be seen in Table 3.1.

Table 3.1: Strains and plasmids used in this chapter.

Strain/plasmid	Genotype/description	Reference
Strains		
BAC0114	<i>B. subtilis</i> 168 $\Delta aprE::aad9$	Laboratory stock
IIG-Bs27-24	<i>B. subtilis</i> 168 with 17% total genome deletion and <i>PmtIA-comK-comS</i> cassette	Reuß <i>et al.</i> , 2017
IIG-Bs27-31	<i>B. subtilis</i> 168 with 22% total genome deletion	Reuß <i>et al.</i> , 2017
PG10	<i>B. subtilis</i> 168 with 36% total genome deletion	Reuß <i>et al.</i> , 2017
BA01	IIG-Bs27-24 $\Delta aprE::aad9$	This chapter
BA02	BA01 Δmpr	This chapter
BA03	BA02 $\Delta aad9$	This chapter
BA04	BA03 $\Delta amyE::xylR-pxylA-prsA$	This chapter
Plasmids		
pBAC0158	pHT01 with SapI sites removed, rrnB T1 T2; crRNA MAD7 handle; protospacer cloning site; Pveg; Pgrac; MAD7	Price <i>et al.</i> , 2020
pBAC0158\amyE2	pBAC0158 with gRNA targeting 1333 bp downstream of the start codon within <i>B. subtilis</i> 168 amyE	This chapter
pBAC0158\mpr1	pBAC0158 with gRNA targeting 134 bp downstream of the start codon within <i>B. subtilis</i> 168 mpr	This chapter
pBAC0158\spec1	pBAC0158 with gRNA targeting 386 bp downstream of the start codon within <i>aad9</i>	This chapter
pBE-S	<i>E. coli</i> / <i>B. subtilis</i> shuttle vector carrying PaprE, KNTase, bla.	Takara Bio
Pro_20_S7_pBES	Pro_20 in pBES downstream of WalM SP	Chapter 4

3.2.2 Plasmid construction

Plasmids used for MAD7-mediated editing were assembled from the SapI-digested parental plasmid, pBAC0158, using 5'-phosphorylated and annealed oligonucleotide duplexes as described in Section 2.5.4.1.

3.2.3 Strain construction

Strain BA01 was constructed through transformation of strain IIG-Bs27-24 following the method described in Section 2.3.2 with a PCR product containing the spectinomycin resistance gene, *aad9*, flanked by 1 kbp homology arms upstream and downstream of *aprE*. This PCR product was obtained through PCR of genomic DNA of strain BAC0114 using oligonucleotides oMAP0152 and oMAP0217. Deletion of *aprE* was confirmed through PCR of the $\Delta aprE::aad9$ locus from purified BA01 genomic DNA, using oligonucleotides *aad9_F1*, *aad9_R1*, oMAP0152 and oMAP0155, which bind outside of the homology arm regions, and subsequent sequencing of the PCR product.

Strains BA02 and BA03 were constructed by deletion of *mpr* and *aad9*, respectively, through CRISPR-MAD7 mediated editing following the approach described in (Price et al., 2020). Donor DNA (dDNA) was constructed through OE-PCR, as described in Section 2.5.5.3, amplified from genomic DNA of the parental strain using oligonucleotide sets oMAP0843/oMAP0844/oMAP0845/oMAP0846 and oMAP0152/dDNA_Bs_aad9_R1/ dDNA_Bs_aad9_F2/oMAP0155, to be used alongside pBAC0158/*mpr1*/CRISP and pBAC0158/*spec1*/CRISP, respectively. Deletion of *mpr* and *aad9* was confirmed through PCR of each locus using oligonucleotides oMAP0851/oMAP0849 and QC_Bs_aad9_F1/ oMAP0217 followed by sequencing with oligonucleotides oMAP0843/oMAP0846 and oMAP0155/oMAP0220. CRISPR-MAD7/sgRNA plasmids were removed from the edited strains as described in Section 2.7.2.5.

Strain BA04 was constructed by substitution of *amyE* with a *xylR-pxyIA-prsA* cassette by CRISPR-MAD7 mediated editing following the method described in Section 2.7.2. Two sets of dDNA were constructed through OE-PCR as described in Section 2.5.5.3, amplified from genomic DNA of strain BA03 using oligonucleotide sets *amyE_dDNA_F1/amyE_dDNA_R1/xylA_dDNA_F/xylA_dDNA_R* and *prsA_dDNA_F/prsA_dDNA_R/amyE_dDNA_F2/amyE_dDNA_R2*, followed by a second round of OE-PCR to fuse these two constructs using oligonucleotides *amyE_dDNA_F1/amyE_dDNA_R2*. Strain BA03 was transformed with this dDNA alongside pBAC0158\amyE2. Integration was confirmed through PCR of the *amyE* locus using oligonucleotides *amyE_QC_F/amyE_QC_R*, which bind outside of the homology arm region, followed by sequencing with oligonucleotides

prsA_seq_F1/prsA_seq_F2/prsA_seq_R1/prsA_seq_R2. The CRISPR-MAD7/sgRNA plasmids were cured from the edited strains as described in Section 2.7.2.5.

3.3 Results

3.3.1 Deletion of native *B. subtilis* proteases

The transformation efficiencies and specific growth rates of the minimal *B. subtilis* strains developed by Reuß et al. (2017), namely IIG-Bs27-24, IIG-Bs27-31 and PG10, were compared (Sections 2.3.2 & 2.2.3). It was found that IIG-Bs27-24 had a significantly higher transformation efficiency than both IIG-Bs27-31 and PG10, and a significantly higher growth rate than PG10 (Figure 3.1a & Figure 3.1b). It should be noted that while growth was slowed for PG10, the carrying capacity of the culture remained the same (Figure 3.1c). As IIG-Bs27-24 expresses two native extracellular proteases (*mpr* and *aprE*), these were knocked out to reduce background protease activity and improve yield of protease candidates (Westers et al., 2008; Wu et al., 2002; Wu et al., 1993). The serine protease *aprE* was deleted through introduction of *aad9* – a spectinomycin resistance marker – to yield strain BA01. The serine protease *mpr* was deleted by marker-less editing with CRISPR-MAD7, to give strain BA02. The *aad9* resistance gene was then deleted through CRISPR-MAD7 mediated editing to produce BA03, which was found to have significantly higher transformation efficiency than both IIG-Bs27-31 and PG10 and a similar growth rate to all strains except PG10 (Figure 3.1a & Figure 3.1b).

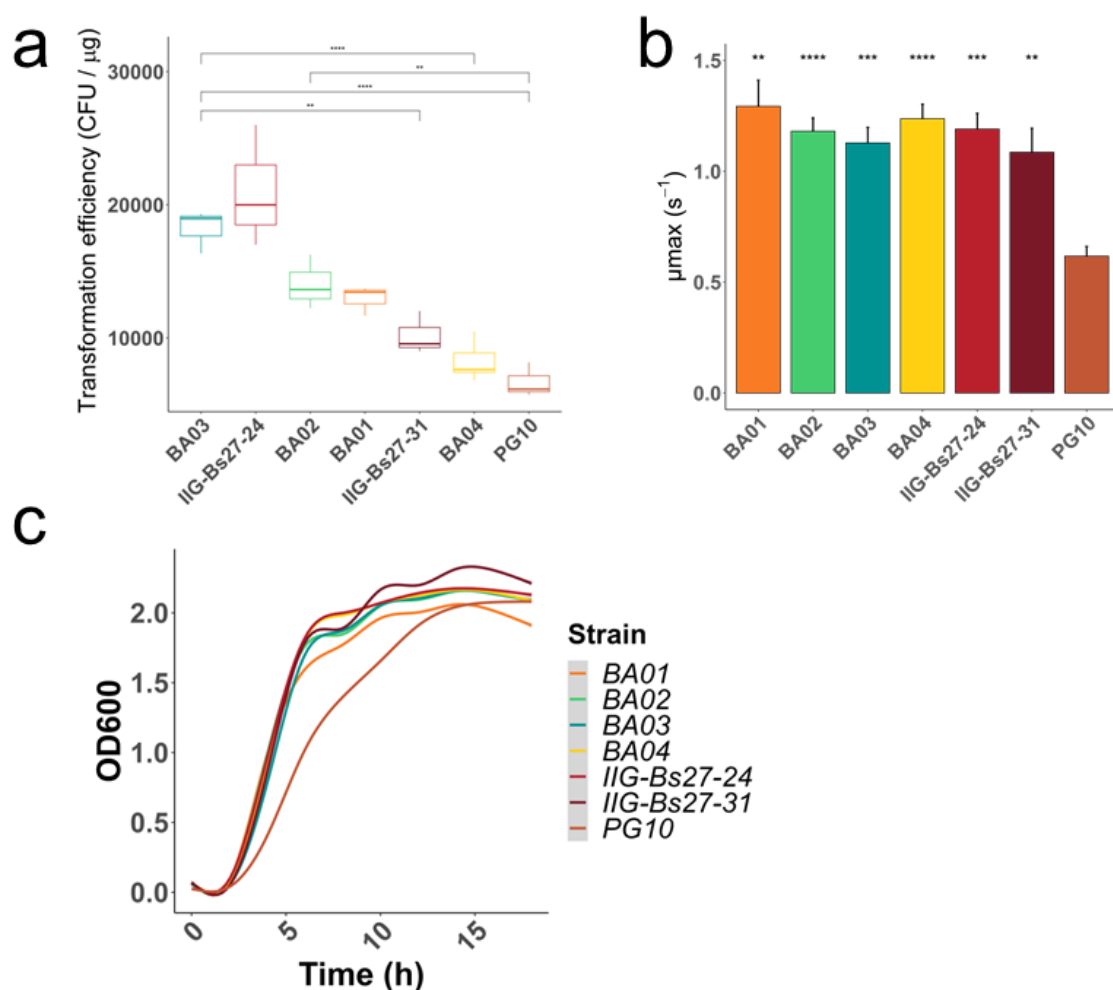


Figure 3.1: a) Transformation efficiency of *B. subtilis* strains (one-way ANOVA, $p = 0.05$, $n = 3$). b) Specific growth rates of *B. subtilis* strains (independent t-test in reference to PG10, p-values: $* \leq 0.05$, $** \leq 0.01$, $*** \leq 0.001$, $**** \leq 0.0001$, $n = 6$, error = $\pm$ SEM). c) Optical density of cultures of *B. subtilis* strains during growth for 18 h at 37 °C, 700 rpm.

B. subtilis strains were imaged at mid-log phase by phase contrast microscopy (Section 2.2.2) and their cell morphologies were observed. All strains constructed in this chapter exhibit the same morphology as the parental strain, IIG-Bs27-24, as does strain IIG-Bs27-31 (Figure 3.2). The morphology of PG10 appeared as long, unseparated filaments, as seen in (Reuß et al., 2017), which would be incompatible with methods of directed evolution that involve encapsulation of single cells.

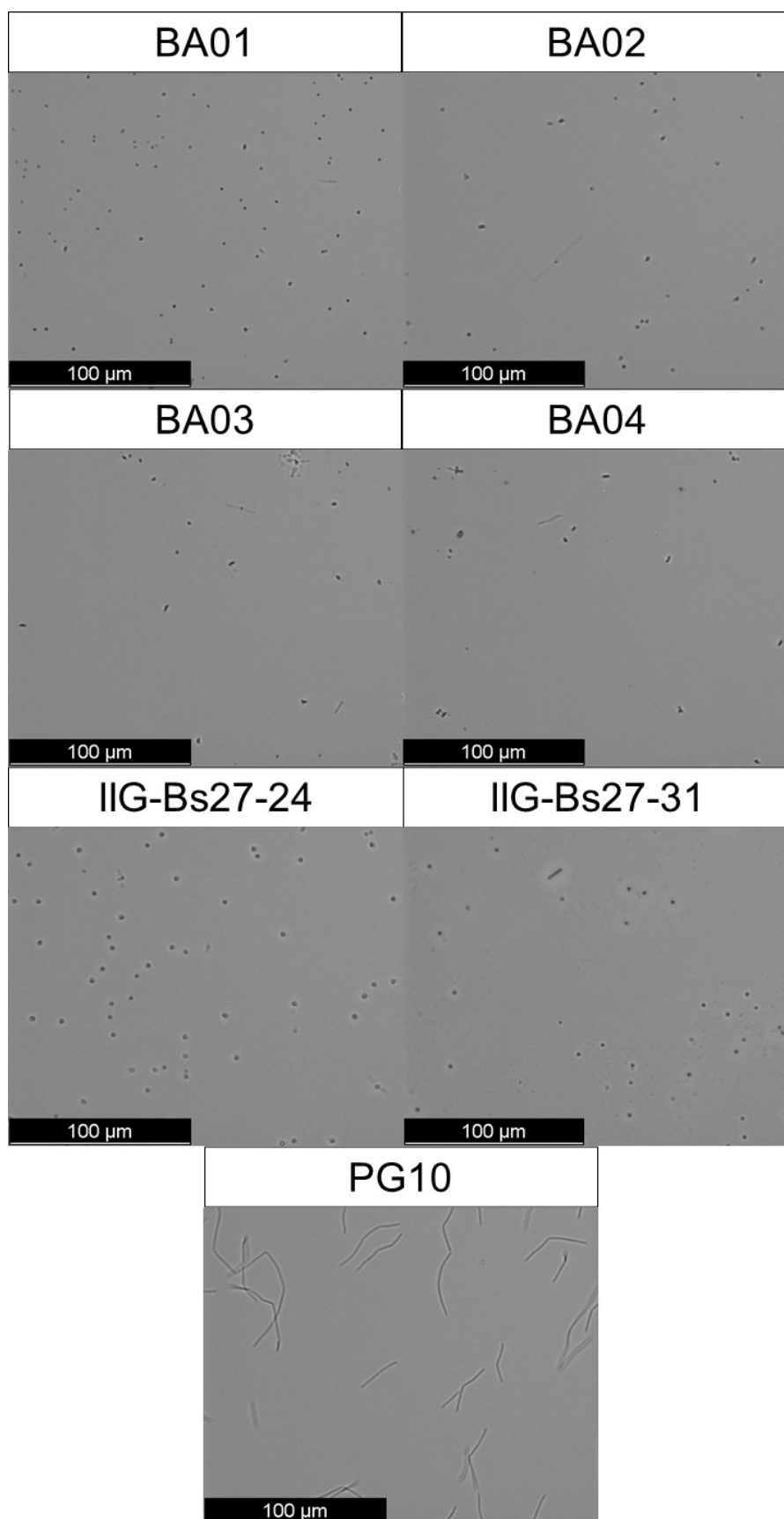


Figure 3.2: Phase contrast images of *B. subtilis* cells at mid-log growth phase.

3.3.2 Improving secretion of heterologous proteases

To improve protease secretion, the foldase *prsA* was overexpressed (Chen et al., 2015). Strain BA04 was constructed by integration of a xylose-inducible *prsA* overexpression cassette that was assembled through OE-PCR (Section 2.5.5.3) at the *amyE* locus through CRISPR-MAD7. This has the added benefit of aiding in downstream processing by reducing contaminants in the supernatant through deletion of the α -amylase *amyE*. The effect of xylose induction on protease secretion was investigated by expression of a protease candidate (Pro_20) (Section 4.3.2) in strains BA03 and BA04 in SR medium with (SRX) and without (SR) xylose (Figure 3.3). Strain BA04 had higher optical density than BA03 in both cases, with an increase of 33% and 19% with and without the inclusion of 4% (w/v) xylose, respectively. Although there were lower protease levels in the supernatant for BA04 than BA03 when grown in SR, in both cases xylose addition increased protease activity – 1.57-fold for BA03 and 11.6-fold for BA04 – such that BA04 grown with xylose had the highest level of protease activity overall.

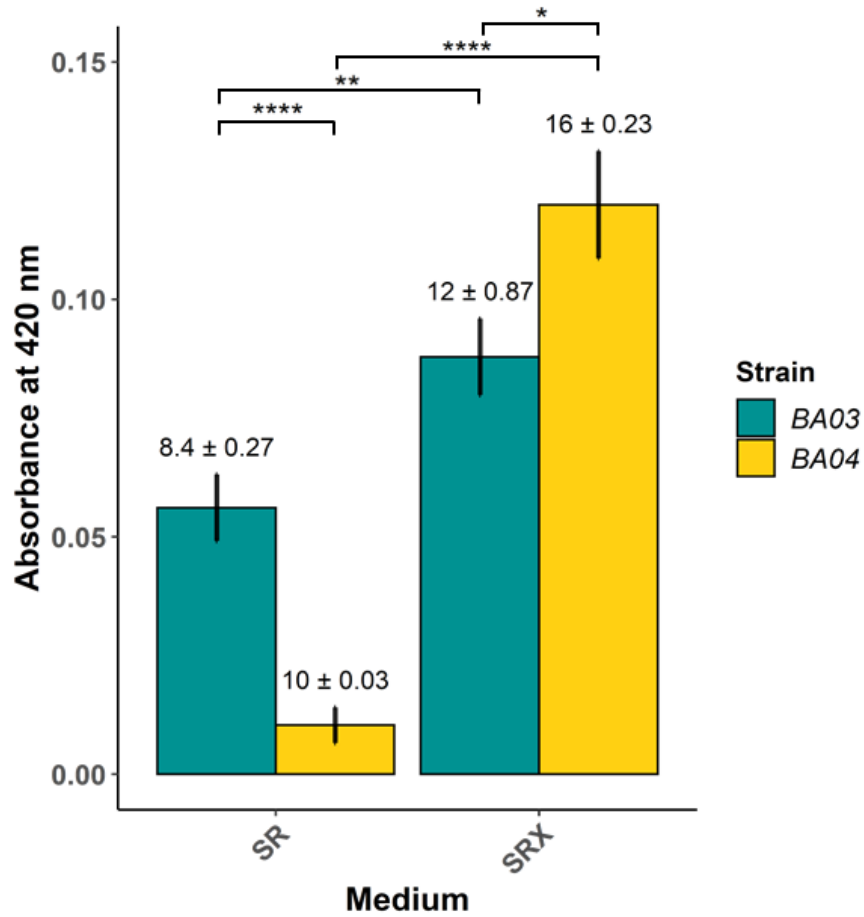


Figure 3.3: Protease activity of supernatant of strains BA03 and BA04 expressing Pro_20, normalised to OD₆₀₀, following 48 h growth at 18 °C when grown in SR and SRX medium, as measured by azocasein assay (independent t-test, p-values: * ≤ 0.05, ** ≤ 0.01, *** ≤ 0.001, **** ≤ 0.0001, n = 9, error = +/- SEM). Culture OD₆₀₀ values labelled above bars (n = 3, error = +/- SEM).

3.4 Discussion

3.4.1 Extracellular protease deletions had no effect on transformation efficiency, growth rate or morphology

It was found that the protease-deficient strains BA01 and BA02 exhibit lower transformation efficiency than the ancestral strain, IIG-Bs27-24. However, deletion of the spectinomycin resistance cassette, *aad9*, used for deletion of *aprE*, restored the transformation efficiency to the level seen in IIG-Bs27-24 (Figure 3.1a). It was also seen that the specific growth rate was unaffected, with all strains constructed in this chapter having growth rates equivalent to IIG-Bs27-24 and IIG-Bs27-31, and the only strain with a significantly lower growth rate being PG10 (Figure 3.1b), albeit with a similar final optical density, demonstrating that the carrying capacity for this culture remained unchanged (Figure 3.1c). Finally, the cell morphology was not impacted for any strain produced. All strains showed a similar morphology, except PG10 which forms extended filaments as seen in (Reuß et al., 2017). It may be concluded, therefore, that deletion of the native extracellular proteases does not have a significant effect on transformation efficiency, growth rate or cell morphology.

3.4.2 *prsA* overexpression increases secretion of heterologous protease

Insertion of the xylose-inducible *prsA* cassette and simultaneous deletion of *amyE* appeared to decrease the secretion of the protease candidate, Pro_20, in the absence of xylose (Figure 3.3). The increase in protease activity in BA03 in the absence of xylose could be explained by the rupturing of cells due to overexpression of the protease and subsequent blockage of Sec pathway components, releasing the candidate protease and intracellular proteases into the medium. BA04, on the other hand, would have a basal level of *prsA* overexpression due to leakiness of the PxylA promoter, which has been found to have a low dynamic range (Radeck et al., 2013). This would prevent the blockage of Sec pathway components in the absence of xylose, avoiding the rupture of cells.

With the addition of xylose, an increase in protease activity was observed for both strains, showing that this alternative carbon source can improve protease production. There was significantly higher protease activity in the supernatant of strain BA04 expressing Pro_20, showing that the induction of *prsA* expression increased protease secretion. However, the introduction of this *prsA* overexpression cassette did lead to a reduction in transformation efficiency (Figure 3.1a)

3.5 Conclusion

We have developed several *B. subtilis* strains in this chapter as chassis for protease production. Through homologous recombination and CRISPR-MAD7 directed editing, we produced a strain deficient in all native extracellular proteases, and this was shown to have no effect on growth rate, transformation efficiency or cell morphology. We also generated a strain with improved secretory capacity through overexpression of the molecular chaperone *prfA*. This strain had a reduced transformation efficiency when compared with the parental strain, and so can be viewed as an expression strain. Further modifications could be introduced to increase protease secretion, such as deletion of the *srfAC* to reduce foam formation as in Liu et al. (2018), which was still present in the parental strain IIG-Bs27-24 (Appendix C).

Chapter 4 – Protease identification and expression

4.1 Introduction

There is a great deal of diversity found within metalloproteases with over 390,000 sequences currently deposited in the MEROPS database, comprising 15 clans and 106 families (MEROPS - the Peptidase Database, 2017). With this amount of variation, it is a significant challenge to identify novel proteases that possess a given set of traits. There are, however, a range of tools available for the prediction of enzyme biochemical characteristics. This includes machine-learning based algorithms that purport to be able to predict activity based solely on the enzyme's primary sequence. In this chapter, several of these algorithms were tested on datasets of metalloprotease sequences and those predicted to be active in industrially relevant conditions were expressed in *B. subtilis*.

The specific characteristics searched for in this chapter were as follows: mesophilicity or psychrophilicity to facilitate the lowering of washing temperatures in laundry; activity in slightly acidic to neutral pH to ensure laundry formulation stability and activity at a standard wash pH of 7; halophilicity to ensure activity in a broad range of water sources.

B. subtilis was chosen as a host for protease expression as it has a robust secretory capability and is easily manipulated, with a wealth of tools for genetic manipulation (Section 1.5) (Hohmann et al., 2017; Schallmeyer et al., 2011). It has been shown that protein secretion through the Sec pathway is often protein-dependent, with variation in secretion levels for different proteins and signal peptides (Brockmeier et al., 2006). There are many possible reasons for this, as protein secretion is governed by an intricate process of events. For example, the fusion of the SP to a target protein may result in different mRNA transcript stabilities, affecting the amount of the final protein. There will also be differences in folding efficiencies between proteins. This can lead to blockage of the Sec pathway components, which in turn can lead to increased degradation of secreted proteins (both correctly and incorrectly folded) by the quality control proteases HtrA/HtrB (Section 1.5.5) (Brockmeier et al., 2006). For optimal protein secretion, then, there must be a balance between efficiency of targeting by the SP and the folding efficiency of the protein. One strategy for optimisation of protein secretion is to create a combinatorial library of the target protein fused with a

range of signal peptides. In this study, protease candidates were expressed in a commercially available vector (*B. subtilis* Secretory Protein Expression System, Takara Bio, Japan) containing a library of all 173 *B. subtilis* Sec signal peptides. The efficacy of these signal peptides for secretion of each candidate was then explored. The effect of various culturing conditions was also investigated to optimise protease secretion.

4.2 Material & Methods

4.2.1 Plasmids and strains

The strains and plasmids used in this chapter can be seen in Table 4.1.

Table 4.1: Strains and plasmids used in this chapter

Strain/plasmid	Genotype/description	Reference
Strains		
BA04	BA03 $\Delta amyE::prsA$	Chapter 3
Plasmids		
pBE-S	<i>E. coli/B. subtilis</i> shuttle vector carrying PaprE, KNTase, bla.	Takara Bio
pBE-S_SPLib	<i>E. coli/B. subtilis</i> shuttle vector carrying PaprE, SP library, KNTase, bla.	Takara Bio
Pro_1_pBES_SPLib	WP_014202324.1 in pBES downstream of SP library	This chapter
Pro_2_pBES_SPLib	WP_012239334.1 in pBES downstream of SP library	This chapter
Pro_3_pBES_SPLib	WP_041509222.1 in pBES downstream of SP library	This chapter
Pro_4_pBES_SPLib	WP_012800440.1 in pBES downstream of SP library	This chapter
Pro_5_pBES_SPLib	WP_008858092.1 in pBES downstream of SP library	This chapter
Pro_6_pBES_SPLib	PLX79621.1 in pBES downstream of SP library	This chapter
Pro_7_pBES_SPLib	WP_090493785.1 in pBES downstream of SP library	This chapter
Pro_8_pBES_SPLib	WP_030462359.1 in pBES downstream of SP library	This chapter
Pro_9_pBES_SPLib	WP_012326503.1 in pBES downstream of SP library	This chapter
Pro_10_pBES_SPLib	WP_006970331.1 in pBES downstream of SP library	This chapter
Pro_11_pBES_SPLib	A0A0C5CJR8 in pBES downstream of SP library	This chapter
Pro_12_pBES_SPLib	P0C0Q3 in pBES downstream of SP library	This chapter

Pro_13_pBES_SPLib	Q9LCJ5 in pBES downstream of SP library	This chapter
Pro_14_pBES_SPLib	Q9RG23 in pBES downstream of SP library	This chapter
Pro_15_pBES_SPLib	P55111 in pBES downstream of SP library	This chapter
Pro_16_pBES_SPLib	O05485 in pBES downstream of SP library	This chapter
Pro_17_pBES_SPLib	Q8D4Y9 in pBES downstream of SP library	This chapter
Pro_18_pBES_SPLib	CAL42347 in pBES downstream of SP library	This chapter
Pro_19_pBES_SPLib	CAL42348 in pBES downstream of SP library	This chapter
Pro_20_pBES_SPLib	WP_163531214.1 in pBES downstream of SP library	This chapter
Pro_21_pBES_SPLib	WP_075037004.1 in pBES downstream of SP library	This chapter
Pro_22_pBES_SPLib	NZ_LT800499.1 in pBES downstream of SP library	This chapter
Pro_23_pBES_SPLib	WP_102400416.1 in pBES downstream of SP library	This chapter
Pro_24_pBES_SPLib	WP_082233925.1 in pBES downstream of SP library	This chapter
Pro_25_pBES_SPLib	WP_180963949.1 in pBES downstream of SP library	This chapter
Pro_26_pBES_SPLib	WP_135326494.1 in pBES downstream of SP library	This chapter
Eu_1_pBES_SPLib	CAA7262796.1 in pBES downstream of SP library	This chapter
Eu_2_pBES_SPLib	KFG79808.1 in pBES downstream of SP library	This chapter
Eu_3_pBES_SPLib	MER0247143 in pBES downstream of SP library	This chapter
Eu_4_pBES_SPLib	MER0468173 in pBES downstream of SP library	This chapter
Eu_5_pBES_SPLib	KFY30966.1 in pBES downstream of SP library	This chapter
Eu_6_pBES_SPLib	KID94564.1 in pBES downstream of SP library	This chapter
Eu_7_pBES_SPLib	XP_008602715.1 in pBES downstream of SP library	This chapter
Eu_8_pBES_SPLib	SMY28112.1 in pBES downstream of SP library	This chapter
Eu_9_pBES_SPLib	XP_003029310.1 in pBES downstream of SP library	This chapter
Eu_10_pBES_SPLib	XP_016758957.1 in pBES downstream of SP library	This chapter
Eu_11_pBES_SPLib	A6P7M6 in pBES downstream of SP library	This chapter
Eu_12_pBES_SPLib	A6P7M6 in pBES downstream of SP library	This chapter
Pro_10_S1_pBES	WP_006970331.1 in pBES downstream of MltG SP	This chapter
Pro_10_S2_pBES	WP_006970331.1 in pBES downstream of SpoIIQ SP	This chapter
Pro_10_S3_pBES	WP_006970331.1 in pBES downstream of WP_004399085.1 SP	This chapter
Pro_10_S4_pBES	WP_006970331.1 in pBES downstream of AbnA SP	This chapter

Pro_10_S6_pBES	WP_006970331.1 in pBES downstream of WalM SP	This chapter
Pro_12_S1_pBES	P0C0Q3 in pBES downstream of YbbE SP	This chapter
Pro_12_S2_pBES	P0C0Q3 in pBES downstream of AprE SP	This chapter
Pro_12_S3_pBES	P0C0Q3 in pBES downstream of DacF SP	This chapter
Pro_12_S4_pBES	P0C0Q3 in pBES downstream of PhrK SP	This chapter
Pro_12_S5_pBES	P0C0Q3 in pBES downstream of YjdB SP	This chapter
Pro_12_S6_pBES	P0C0Q3 in pBES downstream of YxiT SP	This chapter
Pro_12_S7_pBES	P0C0Q3 in pBES downstream of YkoJ SP	This chapter
Pro_12_S8_pBES	P0C0Q3 in pBES downstream of AmyQ SP	This chapter
Pro_20_S1_pBES	WP_163531214.1 in pBES downstream of AbnA SP	This chapter
Pro_20_S3_pBES	WP_163531214.1 in pBES downstream of chitosanase SP	This chapter
Pro_20_S4_pBES	WP_163531214.1 in pBES downstream of BslB SP	This chapter
Pro_20_S5_pBES	WP_163531214.1 in pBES downstream of AprE SP	This chapter
Pro_20_S7_pBES	WP_163531214.1 in pBES downstream of WalM SP	This chapter
Pro_20_S8_pBES	WP_163531214.1 in pBES downstream of YkwD SP	This chapter
Pro_23_S1_pBES	WP_102400416.1 in pBES downstream of pectate lyase SP	This chapter
Pro_23_S2_pBES	WP_102400416.1 in pBES downstream of BSP4_32340 SP	This chapter
Pro_23_S3_pBES	WP_102400416.1 in pBES downstream of TatAC SP	This chapter
Pro_23_S4_pBES	WP_102400416.1 in pBES downstream of levansucrase SP	This chapter
Pro_23_S5_pBES	WP_102400416.1 in pBES downstream of WalM SP	This chapter
Pro_23_S6_pBES	WP_102400416.1 in pBES downstream of WP_004399085.1 SP	This chapter
Pro_23_S7_pBES	WP_102400416.1 in pBES downstream of AbnA SP	This chapter
Pro_23_S8_pBES	WP_102400416.1 in pBES downstream of PTU29680.1 SP	This chapter
Pro_26_S1_pBES	WP_135326494.1 in pBES downstream of WprA SP	This chapter
Pro_26_S2_pBES	WP_135326494.1 in pBES downstream of chitosanase SP	This chapter
Pro_26_S3_pBES	WP_135326494.1 in pBES downstream of CwIO SP	This chapter
Pro_26_S4_pBES	WP_135326494.1 in pBES downstream of AprE SP	This chapter
Pro_26_S5_pBES	WP_135326494.1 in pBES downstream of WP_003231385.1 SP	This chapter

Pro_26_S6_pBES	WP_135326494.1 in pBES downstream of XynC SP	This chapter
Pro_26_S7_pBES	WP_135326494.1 in pBES downstream of YggA SP	This chapter
Pro_26_S8_pBES	WP_135326494.1 in pBES downstream of WP_121591378.1 SP	This chapter
pBAC0001	E. coli/B. subtilis shuttle vector carrying Pgrac and lacI, bla, cat	Price et al., 2019
Pro_10_S3_M28	WP_006970331.1 in pBAC0001 downstream of WP_004399085.1 SP	This chapter
Pro_12_S7_M28	P0C0Q3 in pBAC0001 downstream of YkoJ SP	This chapter
Pro_20_S7_M28	WP_163531214.1 in pBAC0001 downstream of WalM SP	This chapter

4.2.2 Dataset construction

Unless otherwise stated, all computations were performed in R x64 3.6.3 on a desktop running Windows 10 with a 3.2 GHz hex-core processor and 16 GB of RAM.

For protein databases, BLAST searches were performed using the BLOSUM62 matrix on NCBI's BLASTP server and the top 10,000 hits were taken for further analysis. Iterative HMMER (JACKHMMER) searches were performed locally with 3 iterations, inclusion threshold of 0.001 for per-sequence and per-domain E-values and reporting threshold E-values of 10. For DNA databases, nhmmer searches were performed locally with 3 iterations, inclusion threshold E-value of 0.01 and reporting threshold E-value of 10. These datasets were pre-processed to remove sequences without a start codon or the HEXXH motif, with ambiguous residues ("B", "J", "O", "U", "X" & "Z"), those described as "putative", "predicted", "hypothetical" or "fragment" and those less than 250 residues or greater than 750 residues. This size range was chosen due to the presence of propeptide domains in metalloproteases from clan MA, which are commonly between 50 and 200 residues in length. A cutoff of 250 residues would then result in a minimum size of the active protease between 50 and 200 residues. Sequences annotated as housekeeping proteases were excluded.

Psychrophilicity was determined following the category I rules generated by Nath & Subbiah (2014) and predicted non-psychrophilic proteins were excluded.

Redundant sequences from the resulting dataset were removed with CD-HIT (W. Li & Godzik, 2006) run locally with an identity cut-off of 0.4 and a word size of 2, followed by BLAST with an upper cut-off of 25% identity to ensure non-redundancy.

Acidophilicity was determined using the Acalpred server (*AcalPred*, 2013) and sequences below 60% confidence were removed. Mesophilicity was confirmed using the Thermopred server (*ThermoPred*, 2011) and sequences below 99% confidence were removed. Halophilicity was determined by Dr Abhigyan Nath using the model developed in Nath (2016) and non-halophilic sequences were removed.

Signal peptides were predicted using the SignalP 4.1 web server (Nielsen, 2017). Sequences without a predicted signal sequence of eukaryotic or Gram-positive origin were removed. The resulting dataset was then filtered to those of fungal or unicellular origin as these are generally extracellular in nature (Gupta et al., 2002).

4.2.3 Plasmid construction

pBE-S_SPLib was constructed via Gibson assembly, as described in Section 2.5.8.4, by insertion of a signal peptide library (Takara Bio, Japan) into the MluI and EagI digested pBE-S vector.

Protease candidate DNA was synthesised by Integrated DNA Technologies, with BsaI restriction sites included in the 5' and 3' ends compatible with EagI and XbaI restriction site overhangs. Polyhistidine-tags containing six histidine residues were included on the C-terminus of each construct to aid in purification and detection by western blot. SP library plasmids were constructed through ligation of the BsaI-HFv2 digested candidate with pBE-S_SPLib digested with EagI and XbaI. This library was transformed into *E. coli* strain Top10 (Section 2.3.1), colonies were pooled, and the plasmid DNA was extracted (Section 2.5.1). This generated libraries for each protease candidate that were then transformed into *B. subtilis* strain BA04 (Section 2.3.2). The eight highest-secreting colonies identified for each candidate following transformation on skimmed milk agar were sequenced and duplicates were discarded, yielding plasmids with distinct SPs. This process is summarised in Figure 4.1.

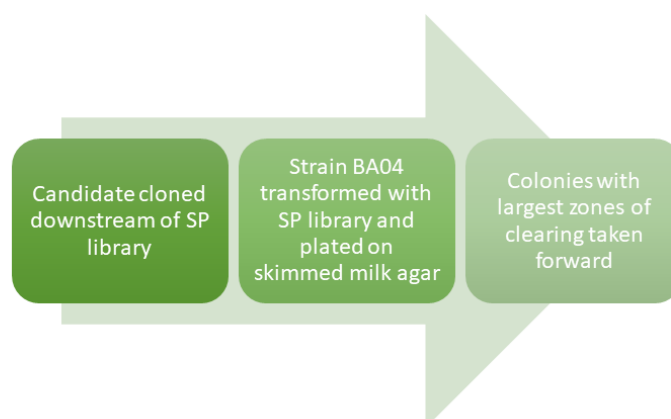


Figure 4.1: Process workflow used to identify functional SPs for each protease candidate. Pro_10_S3_M28, Pro_12_S7_M28 and Pro_20_S7_M28 were constructed by PCR of each candidate from the respective pBE-S plasmid, with Bsal overhangs added compatible with BamHI and XbaI. PCR products were then digested with Bsal-HFv2 and ligated into the BamHI and XbaI digested pBAC0001.

4.3 Results

4.3.1 Identification of metalloprotease candidates

Four metalloproteases from three subclans of the zincin family were chosen as inputs for BLAST and HMMER searches. These were selected as they possess some of the traits described in Section 4.1. The relevant characteristics of these four proteases are shown in Table 4.2 (Secades, Alvarez & Guijarro, 2001; 2003; Adekoya et al., 2006, Austin & Austin, 2016).

Table 4.2: Proteases used as query sequences for BLAST and HMMER searches and relevant characteristics. Subclan is given as defined by the MEROPS (Rawlings et al., 2010) database.

Source organism	Protease	Subclan	Optimal temperature (°C)	Optimal pH	Size (kDa)
<i>Flavobacterium psychrophilum</i>	Fpp1	Metzincin	25-30	6.5	55
<i>Flavobacterium psychrophilum</i>	Fpp2	Metzincin	24	6-7	62
<i>Aeromonas salmonicida</i> subsp. <i>achromogenes</i>	AsaP1	Aspzincin	Psychrophilic	-	37
<i>Antarctic bacterium str. 643</i>	VAB	Gluzincin	Psychrophilic	-	78

A dataset of homologous sequences was then generated using the non-redundant protein sequences (NCBI-nr) and metagenomic proteins (NCBI-env_nr) databases. BLAST searches retrieved 10,000 results for Fpp1, Fpp2 and VAB, and 2,269 for AsaP1. Running three iterations of JACKHMMER searches against the MEROPS database identified 24,234 remote homologues for Fpp1; 20,374 for Fpp2; 638 for AsaP1; and 7,742 for VAB.

After pre-processing and prediction of psychrophilicity according to the rules defined in Nath & Subbiah (2014), this dataset was reduced to 17,193 sequences. Removal of redundant sequences by CD-HIT further refined this to 1,700 sequences. Machine-learning algorithms were then applied to predict thermostability, halophilicity and acidophilicity (Section 1.4.4), and those without signal peptides as predicted by SignalP 4.1 were removed in order to enrich the dataset for extracellular enzymes. The taxonomies of this refined dataset were then analysed and those of metazoan or viridiplantae origin were removed as proteases from unicellular hosts are more likely to be extracellular (Gupta et al., 2002) and therefore more effective detergent proteases. The resultant dataset consisted of twenty non-redundant candidate proteases predicted to be psychrophilic, halophilic, acidophilic, and secreted. Ten of these are from prokaryotic sources and ten are from eukaryotic sources, from a range of taxonomic classes (Appendix D) and with a high degree of sequence variation (Figure 4.2).

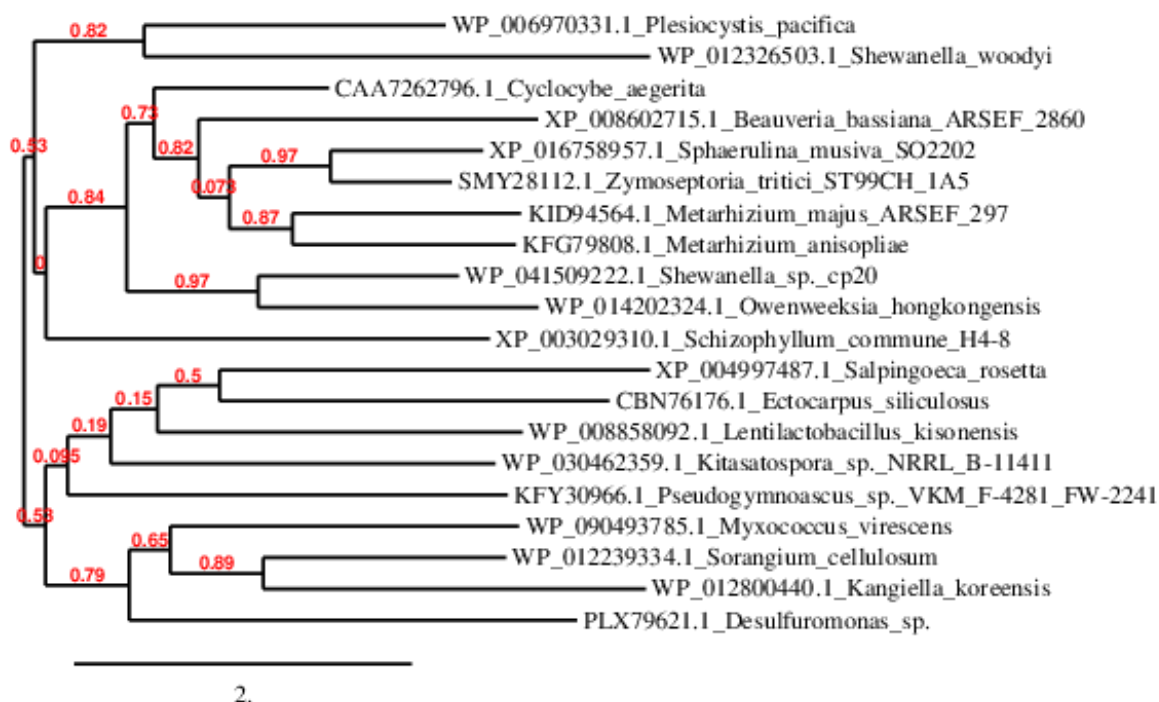


Figure 4.2: Phylogenetic tree showing candidate metalloproteases taken from BLAST and HMMER searches of nr and env_nr databases and refined to non-redundant sequences predicted to be acidophilic, psychrophilic, halophilic, and secreted.

Another dataset of protease candidates was generated from the SWISSPROT database. Following three iterations of JACKHMMER searches, 642 proteases were identified with homology to the query sequences in Table 4.2. After pre-processing this was reduced to 247 sequences. Following prediction of acidophilicity, psychrophilicity, presence of a signal peptide and removal of redundant sequences this was narrowed down to 10 candidates. As Swissprot is a much smaller database, the degree of variation was much reduced when compared with the previously identified candidates (Figure 4.3). Due to the reduced size of this dataset, halophilicity was not predicted.

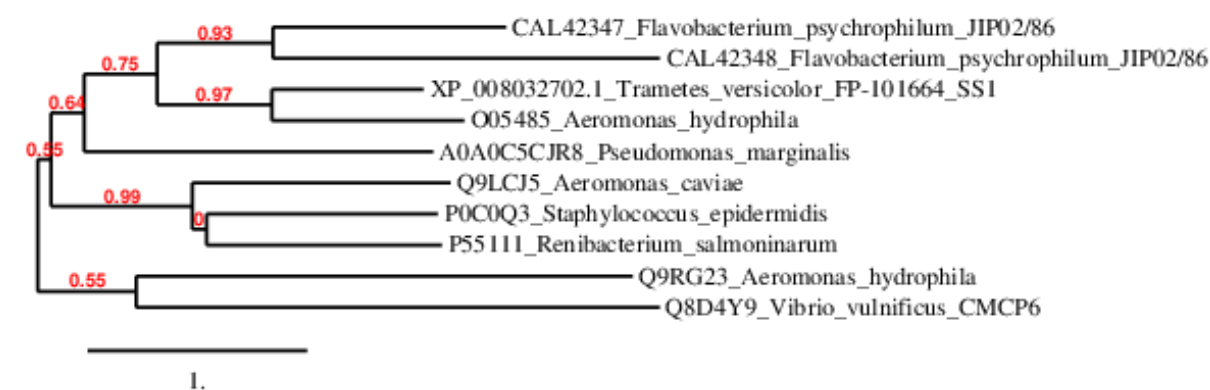


Figure 4.3: Phylogenetic tree showing candidate metalloproteases taken from HMMER searches of SWISSPROT databases and refined to non-redundant sequences predicted to be acidophilic, psychrophilic, and secreted.

As *B. subtilis* belongs to the phylum *Bacillota*, a third set of candidates was identified from the same phylum to increase the likelihood of protease expression. To further enrich this dataset for the desired characteristics, a database was created with all the *Bacillota* sequences from the NCBI assembly database containing “psychro”, “acido” or “halo” in the species name, resulting in a database of 1,893,574 sequences. Two iterations of nhmmer using the queries in Table 4.2 against this database resulted in 55,820 hits, which were refined down to 95 proteases following pre-processing to remove ambiguous residues, duplicates, and sequences without a HEXXH motif. Following prediction of optimal activities, presence of a signal peptide and after removal of redundant sequences, this resulted in 7 candidate proteases with a lower level of sequence variation than both previous sets of candidates (Figure 4.4). Halophilicity was not predicted for this dataset due to its much-reduced size.

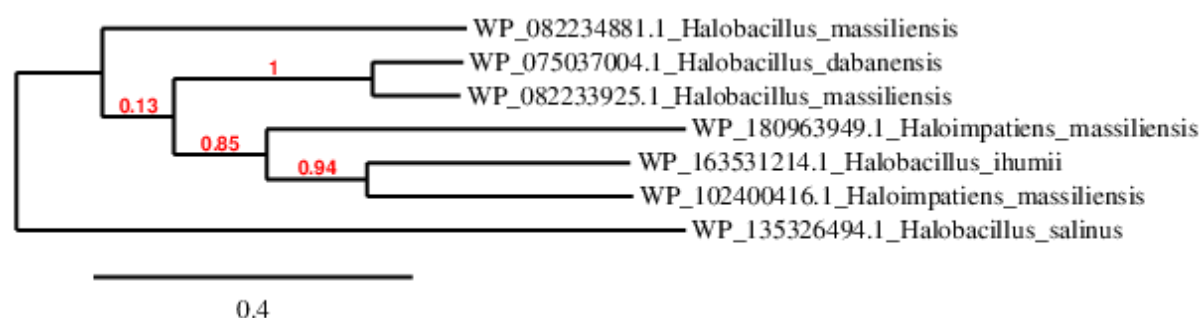


Figure 4.4: Phylogenetic tree showing candidate metalloproteases taken from HMMER searches of all assembly databases with “halo”, “acid” or “psychro” and refined to non-redundant sequences predicted to be acidophilic, psychrophilic, and secreted.

4.3.2 Protease expression

Protease candidate DNA was synthesised and integrated downstream of a pBE-S signal peptide library containing 173 Sec signal peptides from *B. subtilis*. This library was transformed into *B. subtilis* strain BA04 and plated on 1% skimmed milk agar. Following incubation at 30 °C for 48 hours, protease secretion and activity was determined by the presence of zones of clearing around colonies.

Out of the 20 candidates identified in Figure 4.2, one was found to be active and secreted, WP_006970331.1 from *P. pacifica*, hereafter designated Pro_10. Of the candidates in Figure 4.3, one was shown to be active and secreted – P0C0Q3 from *S. epidermidis* (Pro_12). From Figure 4.4, three proteases were shown to be active and secreted – WP_163531214.1 from *H. ihumii* (Pro_20), WP_102400416.1 from *H. massiliensis* (Pro_23) and WP_135326494.1 from *H. salinus* (Pro_26).

4.3.3 Signal peptide sequences

Following growth of transformants on skimmed milk agar, the 8 colonies with the largest zones of clearing from colonies of equal size for each candidate were cultured, plasmid DNA was isolated, and the signal peptide region of the plasmids were sequenced. The sequences of signal peptides which effectively secrete each candidate are shown in Tables 4.3 - 4.7, along with net charge and hydrophobicity. Net charge and hydrophobicity were calculated as in Brockmeier et al. (2006) to investigate whether this influences secretion of each protease candidate.

Table 4.3: Comparison of signal peptides used to secrete heterologous Pro_10.

SP name	SP sequence	Charge N region	Hydrophobicity (%)
MltG	MYINQQKKSFFNKKR IILSSIVVLFLIIGGAFL	5	60.61
SpolIQ	MREEEKTSQVKKLQ QFFRKRWWFPAIYLV SAAVILTAVL	5	52.5
WP_004399085.1	MFKKLLLATSLTFSL SLVLPLDGHAKA	2	64.29

AbnA	MKKKKTWKRFLHFSS AALAAGLIFTSAAPAE A	5	59.38
WalM	MLKKVILAAFILVGSTL GAFSFSSDASA	1	67.86
Average		3.6	60.92

Table 4.4: Comparison of signal peptides used to secrete heterologous Pro_12.

SP name	SP sequence	Charge N region	Hydrophobicity (%)
YbbE	MKTKTLFIFSAILTSLIFA PNETFAAA	1	62.96
AprE	MRSKKLWISLLFALTLLIF TMAFSNMSVQAAA	3	64.52
DacF	MKRLSTLLIGIMLLTC APSAFAAA	2	72
PhrK	MKKLVLCVSILAVILSG VAAA	2	76.19
YjdB	MNFKKTVVSALSISALA LSVSGVASAAA	2	64.29
YxiT	MKWNNMLKAAGIAVLL FSVFAYAAPSLKAVQA AA	3	73.53
YkoJ	MVGLLAGCLAAGGFS YNAFAAA	0	81.82
AmyQ	MFAKRFKTSLLPLFAG FLLLFHLVLAGPAAASA AA	3	80
Average		2.0	71.91

Table 4.5: Comparison of signal peptides used to secrete heterologous Pro_20.

SP name	SP sequence	Charge N region	Hydrophobicity (%)
AbnA	MKKKKTWKRFLHFSS AALAAGLIFTSAAPAEA AA	5	61.76
Chitosanase	MKISMQKADFWKKAAL SLLVFTMFFTLMMSET VFAAA	2	64.86
BsIB	MLKRTSFVSSLFISSAV LLSILLPSGQAHAAA	2	62.50
AprE	MRSKKLWISLLFALTILF TMAFSNMSVQAAA	3	64.52
WalM	MLKKVILAAFILVGSTL GAFSFSSDASAAA	1	70
YkwD	MKKAFILSAAAAGLFT FGGVQQASAAA	2	75
Average		2.5	66.44

Table 4.6: Comparison of signal peptides used to secrete heterologous Pro_23.

SP name	SP sequence	Charge N region	Hydrophobicity (%)
Pectate lyase	MKKVMLATALFLGLTP AGANAAA	2	78.26
BSP4_32340	MAKPLSKGGILVKKVL IAGAVGTAVLFGTLSS GIPGLPAADAAA	3	77.27
TatAC	MELSFTKILVILFVGFL VFGPDKLPALGAA	0	80
Levansucrase	MNIKKFAKQATVLTFT TALLAGGATQAFAAA	3	64.52
WalM	MLKKVILAAFILVGSTL GAFSFSSDASAAA	1	70

WP_004399085.1	MFKKLLLSALTFSL SLVLPLDGHAKAAA	2	66.67
Average		2.4	71.36

Table 4.7: Comparison of signal peptides used to secrete heterologous Pro_26.

SP name	SP sequence	Charge N region	Hydrophobicity (%)
WprA	MKRRKFSSVVATVLIF ALIFSLFSPGTKAAAA A	5	66.67
Chitosanase	MKISMQKADFWKKA AISLLVFTMFFTLMMMS ETVFAAA	2	64.86
CwlO	MRKSLITLGLASVIGT SSFLIPFTSKTASAAA	3	59.38
AprE	MRSKKLWISLLFALTLI FTMAFSNMSVQAAA	3	64.52
WP_003231385.1	MKLKTKASIKFGICVG LLCLSITGFTPFFNST HAEAAA	3	57.89
XynC	MIPRIKKTICVLLVCFT MLSVMLGPGATEVLA AA	2	70.59
YqgA	MKQGKFVFLILLML TLVVAPKGKAEAAA	3	73.33
WP_121591378.1	MKTKTLFIFSAILTSLIF APNETFAAA	1	62.96
Average		2.75	65.03

4.3.4 Qualitative evaluation of signal peptides

The secretion efficiency of each signal peptide for secretion of Pro_10 was qualitatively assessed by comparing the sizes of zones of clearing around single colonies on skimmed milk agar plates (Figure 4.5). Zones of clearing were found to vary significantly, and so quantitative methods of assaying SP efficiency were explored for Pro_10 and all other protease candidates.

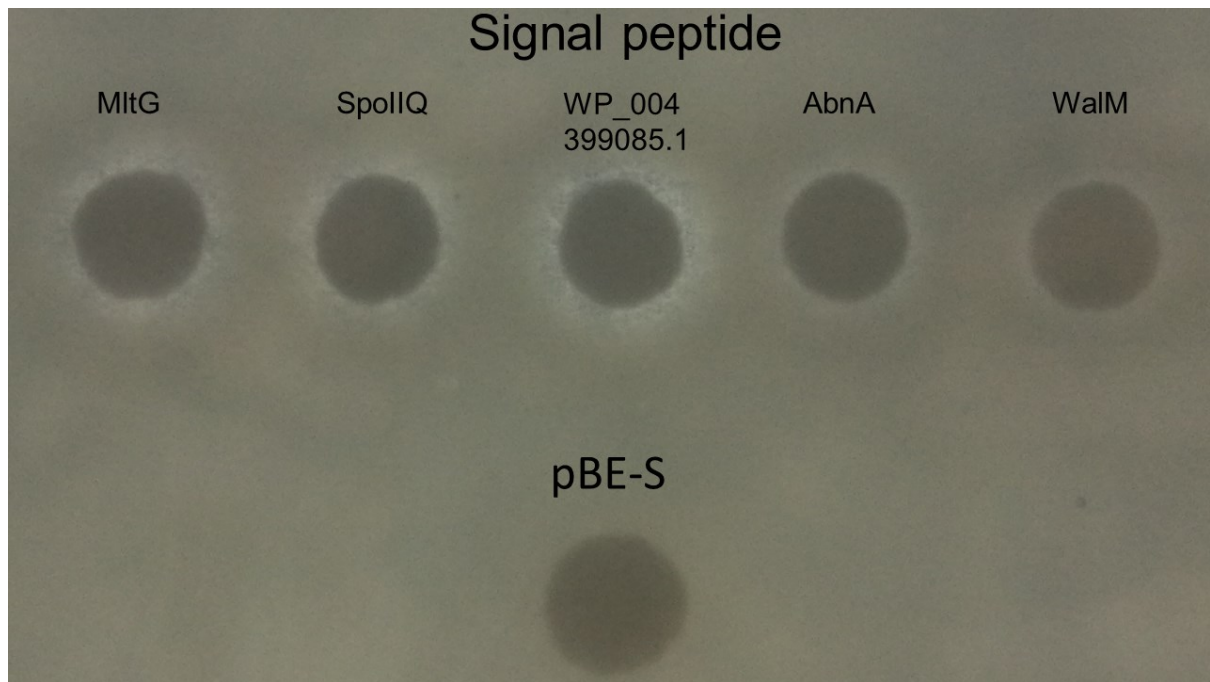


Figure 4.5: Skimmed milk agar plate showing zones of clearing surrounding *B. subtilis* colonies secreting heterologous Pro_10 from a range of signal peptides, and empty pBE-S vector.

4.3.5 Quantitative evaluation of signal peptides

To quantify the relative secretion level of protease candidates by each signal peptide, protease secreting strains were cultured using the Biolector microbioreactor system, as this ensures appropriate aeration in small-scale fermentation studies. Following incubation in SRX medium, protease activity in culture supernatants was measured for each signal peptide using the azocasein assay (Section 2.10.2) (Figure 4.6). The signal peptide had a large impact on protease secretion for all proteases, with the exception of Pro_10 which showed similar activity in the supernatant for all SPs tested. A large variation in activity could be seen for all other candidates, with fold changes of 5.3, 7.0, 2.2 and 50-fold for Pro_12, Pro_20 and Pro_23 and Pro_26, respectively, from the SPs with the lowest to the highest secretion level. However, protease activity of Pro_26 downstream of either the chitosanase or XynC SPs was not significantly higher than the empty vector control. The WalM signal peptide was found to be effective for secretion of three out of the five protease candidates, showing the highest levels of activity for Pro_20 and Pro_23, and significant activity for Pro_10. The combination of SP and protease that displayed the highest protease secretion for each candidate was then taken forward for all future experiments.

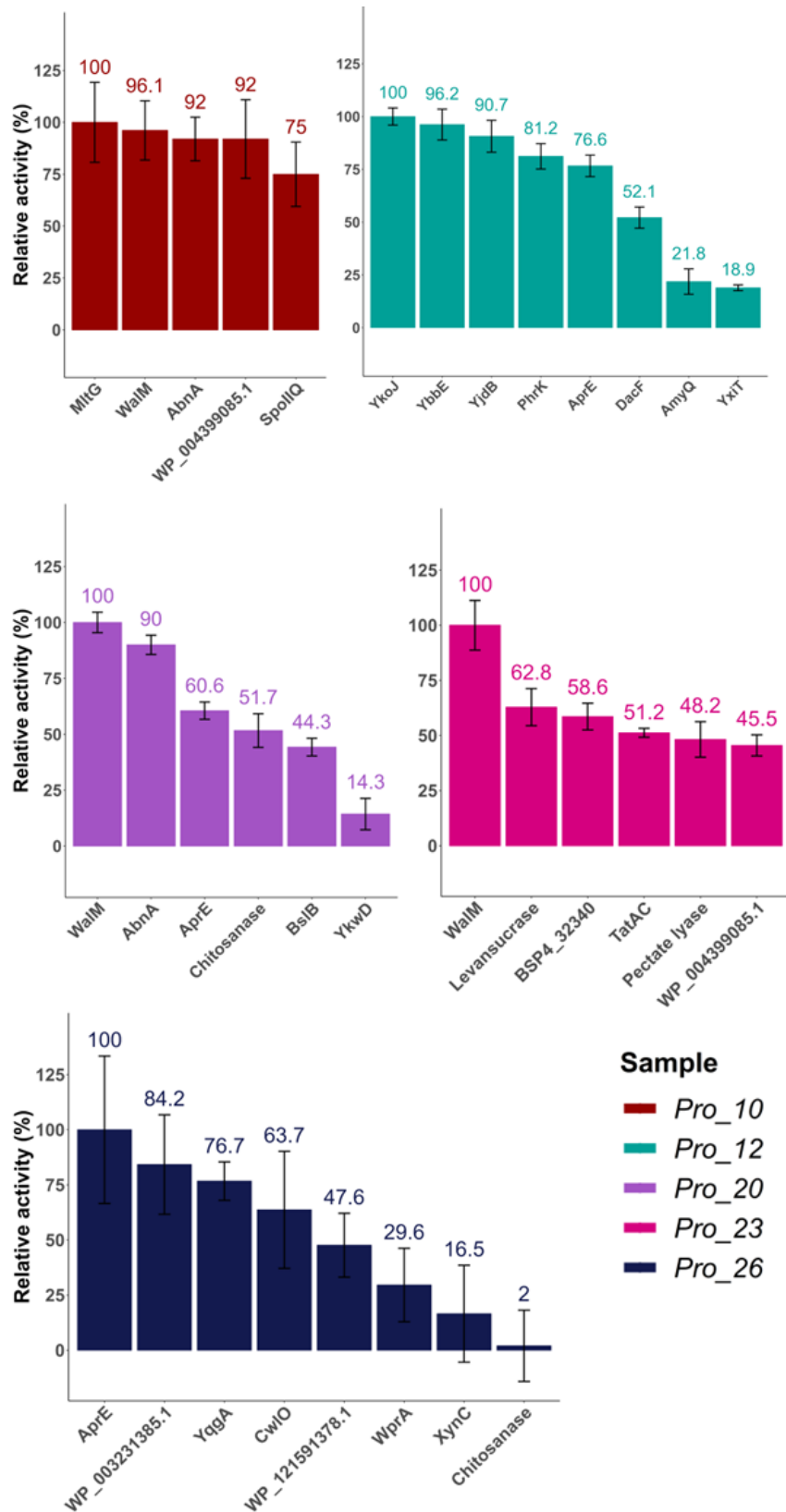


Figure 4.6: Comparison of signal peptides used for secretion of protease candidates expressed heterologously in *B. subtilis*, as measured by azocasein assay of culture supernatants. Activity is given relative to the maximal activity for each candidate. Relative activity values are shown above bars. Values are blanked by the empty vector control (n = 6, error = +/- SEM).

4.3.6 Effect of incubation temperature and growth medium on expression of proteases

The effect of incubation temperature and growth medium was investigated by western blot of the protease candidates Pro_10, Pro_12 and Pro_20, as Pro_23 and Pro_26 could not be detected by western blot (Section 2.8.5). Following growth to mid-log phase (OD 0.8) at 37 °C, candidates were cultured at 18 °C and 37 °C for 18 hours and expression was compared by western blot of culture supernatants (Figure 4.7). In all cases, there were higher levels of secretion following growth at 18 °C, with Pro_12 and Pro_20 being undetectable after growth at 37 °C.



Figure 4.7: Western blot of culture supernatants following growth at 18 °C and 37 °C. Supernatant samples were normalised by OD prior to concentration by TCA precipitation. Lanes represent: M – Blue Prestained Protein Standard, Broad Range (11-250 kDa) (New England Biolabs); 1 – Pro_10 (18 °C); 2 – Pro_12 (18 °C); 3 – Pro_20 (18 °C); 4 – Pro_10 (37 °C); 5 – Pro_12 (37 °C); 6 – Pro_20 (37 °C).

Various growth media were tested for expression of protease candidates. Cultures were incubated at 18 °C for 48 hours, following growth to mid-log phase in LB, TB, SR and SRX media (J. Chen et al., 2015). Protein levels in culture supernatants were visualised by western blot (Figure 4.8a). It was seen that growth in SRX medium was optimal for expression of all candidates. Bands in the western blot were of equivalent or higher intensity compared with LB and SR. Total protein in the supernatant (Section 2.12.4) was over three times that of LB (Figure 4.8b), and as the samples were normalised by total protein, the mass of each specific protease in the supernatant was much greater in SRX than LB.

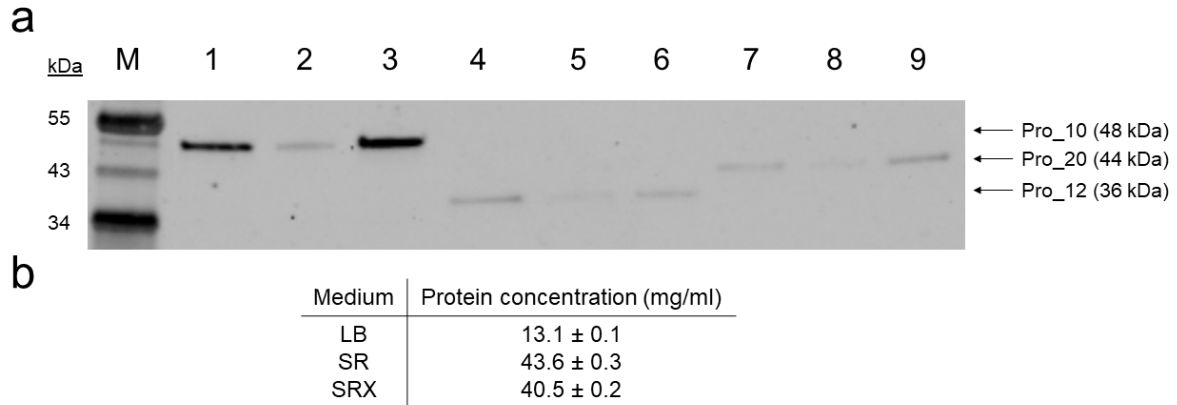


Figure 4.8: a) Western blot of culture supernatants following growth in various media. Normalised by total protein of supernatant. Lanes represent: M – Blue Prestained Protein Standard, Broad Range (11-250 kDa) (New England Biolabs); 1 – Pro_10 (LB); 2 – Pro_10 (SR); 3 – Pro_10 (SRX); 4 – Pro_12 (LB); 5 – Pro_12 (SR); 6 – Pro_12 (SRX); 7 – Pro_20 (LB); 8 – Pro_20 (SR); 9 – Pro_20 (SRX). b) Protein concentration of supernatant samples from cultures grown for 48 h at 18 °C, as measured by A_{280} .

4.3.7 Inducible expression of protease candidates

Protease candidates were integrated into pBAC0001 (Price et al., 2019), a *B. subtilis* shuttle vector containing the Pgrac promoter, which is inducible via addition of IPTG. Secretion of each candidate was then compared by western blot of culture supernatants following growth in either LB or SRX medium, with or without the addition of IPTG. It was found that Pro_10 expression could be induced by IPTG, with the highest expression seen when grown in LB (Figure 4.9). All other candidates could not be detected by western blot following IPTG induction.

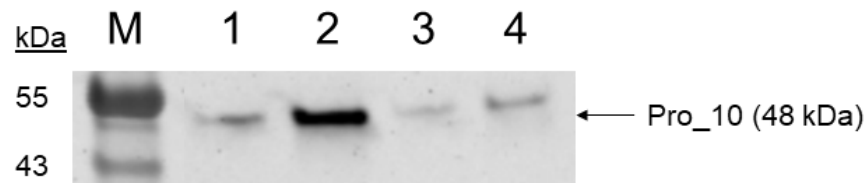


Figure 4.9: Western blot of Pro_10 culture supernatants following growth in LB and SRX medium with or without addition of 1mM IPTG. Supernatant samples were normalised by OD prior to western blotting. Lanes represent: M – Blue Prestained Protein Standard, Broad Range (11-250 kDa) (New England Biolabs); 1 – LB; 2 – LB, 1 mM IPTG; 3 – SRX; 4 – SRX, 1 mM IPTG.

4.3.8 Expression of protease candidates at 18 °C over time

Candidate proteases were expressed downstream of the *aprE* promoter in *B. subtilis* strain BA04 and cultured in SRX medium at 18 °C for a period of 240 h. Protease activity in the culture supernatant and culture OD was measured every 24 h. All protease producing strains followed the same growth profile, reaching stationary phase after culturing for 200 h (Figure 4.10a), although timing of protease expression was protein-dependent (Figure 4.10b & 4.10c), despite the fact that the *aprE* promoter is normally only active during stationary phase (Barbieri et al., 2016). Pro_10 and Pro_23 had maximal activities at 240 h and 168 h, respectively, followed by a plateau in activity. Pro_12, Pro_20 and Pro_26 had maximal activities at 144 h, 120 h and 96 h, respectively, followed by a decline in activity. Similarly, when normalising by optical density, Pro_10 and Pro_23 had maximal activities at 240 h and 144 h, respectively, with a plateau in activity. Interestingly, when normalised by optical density, Pro_12, Pro_20 and Pro_26 all had maximal activities at 120 h, followed by a decline in activity.

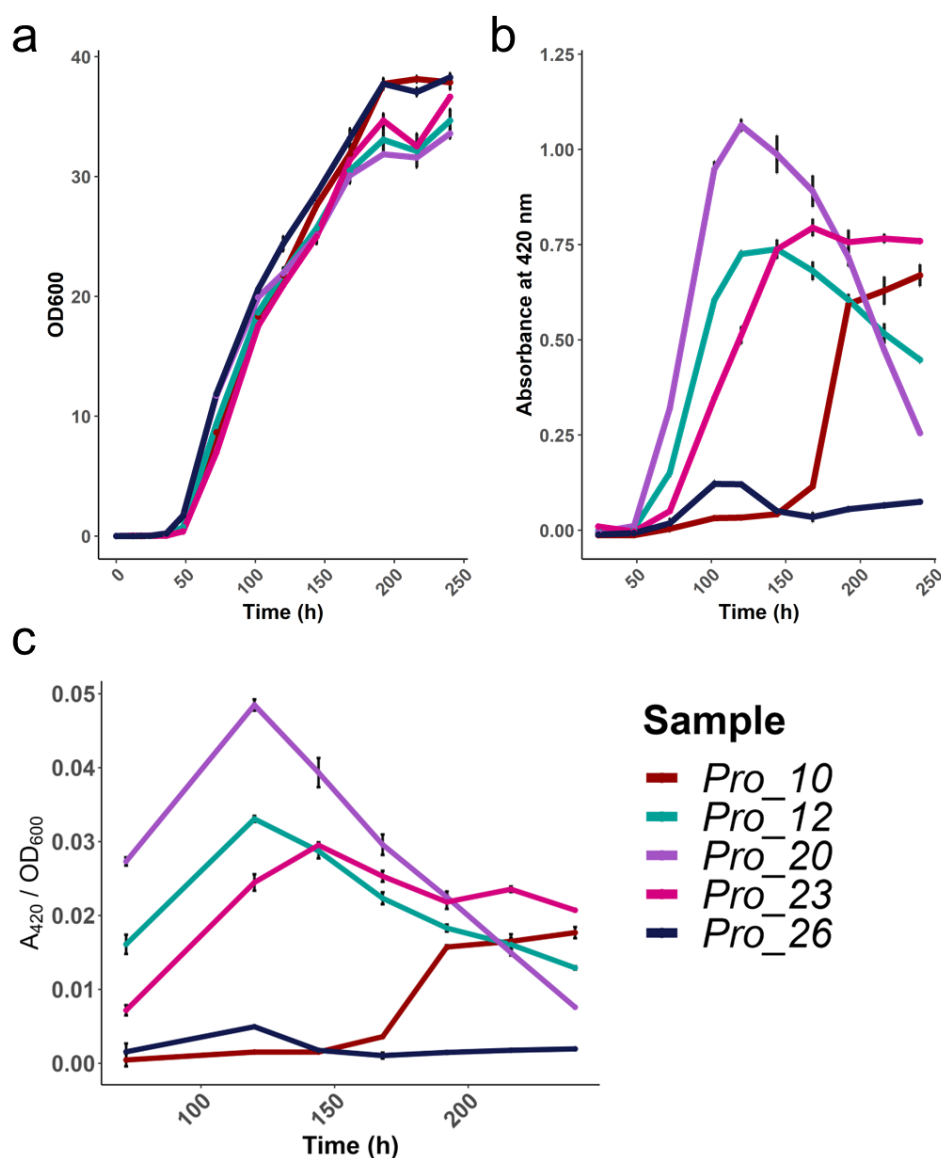


Figure 4.10: a) Optical density (OD₆₀₀) of protease candidate producing cultures during growth at 18 °C. b) Protease activity of culture supernatants as measured by azocasein assay at pH 7.0 (n = 3, error = +/- SEM). c) Protease activity of culture supernatants normalised to optical density of culture (n = 3, error = +/- SEM).

4.3.9 Protease candidate activity at temperature and pH

Protease activity of culture supernatants was assayed at a temperature range between 10 °C and 60 °C (Figure 4.11a). It was found that all proteases had optimal activities in the mesophilic range (20-45 °C): Pro_26 had optimal activity at 30 °C, Pro_10, Pro_12 and Pro_20 had optimal activities at 35 °C and Pro_23 had optimal activity at 40 °C. Protease activity was then assayed at a pH range of 5-10 (Figure 4.11b). Pro_12, Pro_20 and Pro_26 had optimal activities at pH 6.0, Pro_23 had optimal activity at pH 7.0, and Pro_10 had optimal activity at pH 9.0.

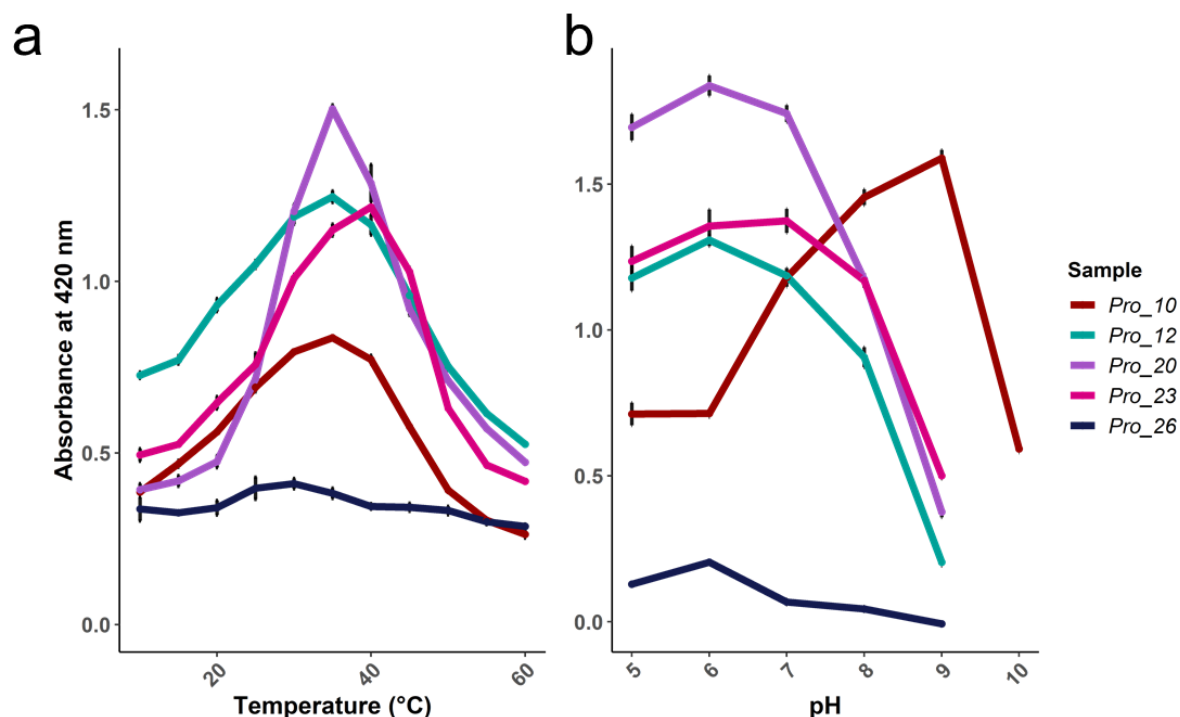


Figure 4.11: a) Azocasein assay of culture supernatants (pH 7.0) at a range of temperatures. (n = 3, error = $\pm$ SEM). Error bars show standard error of the mean. b) Azocasein assay of culture supernatants at a range of pH (n = 3, error = $\pm$ SEM).

4.4 Discussion

4.4.1 Prediction accuracy of bioinformatic models

Due to the incredible diversity of protein sequences found in Nature, it is a significant challenge to identify novel enzymes that exhibit a desired characteristic. In this chapter, we aimed to address this issue through the use of bioinformatic approaches to predict enzyme characteristics based solely on the primary sequence of said enzyme. One challenge to this approach was the scarcity of published methods of predicting either psychrophilicity or halophilicity. Thermostability has long been a target for novel enzyme discovery, as thermostable enzymes are often more active and stable overall. As such, there have been many attempts to predict thermostability, including that of Lin & Chen (2011), in which Thermopred was developed. With that in mind, we utilised two models for predicting psychrophilicity: the rotation forest-based model developed by Nath & Subbiah (2014); and the SVM-based Thermopred (*ThermoPred*, 2011) to classify sequences as not thermophilic.

Our results demonstrate that prediction of psychrophilicity remains a challenge. Of the five proteases tested here, all had optimal activity in the mesophilic range (Figure 4.11a), suggesting that prediction of psychrophilicity by category I rules generated by Nath & Subbiah (2014) was sub-optimal, however, prediction of mesophilicity by Thermopred (*ThermoPred*, 2011) was accurate. This follows the results found in the literature, where maximum accuracy of psychrophilicity prediction by Nath & Subbiah (2014) was found to be 70.5%, while prediction of mesophilicity by Lin & Chen (2011) was 92.7%. Prediction of acidophilicity had a high degree of accuracy also, with four out of five proteases having high levels of activity in the slightly acidic range, that is, pH 5.0 – 7.0 (Figure 4.11b). This may be compared with a recorded accuracy of 96.3% (Lin et al., 2013). This disparity may demonstrate the disconnect between experimental and bioinformatic results, although further experimentation with a higher sample size is required to conclusively state that there is such a difference in prediction accuracies.

4.4.2 Efficiency of signal peptide is protein-dependent

To optimise secretion of protease candidates, a range of SPs were tested for each through use of a SP library containing all 173 Sec signal peptides found in *B. subtilis* (Takara Bio, Japan). It was shown that the efficiency of secretion was largely dependent on the protease being secreted, with each having a different suite of SPs. A large variation in length, charge and hydrophobicity was seen for each SP. This did not appear to be correlated with secretion efficiency, with the exception of Pro_12 which exhibited higher secretion levels for SPs with lower charge in the N-region (Tables 4.3 - 4.7; Figure 4.6). There were several SPs that were effective for multiple candidates, namely those of AbnA, AprE, WalM. These SPs may be particularly efficient at secretion of metalloproteases, or it could be that they are overrepresented in the signal peptide library. To ensure that this is not the case, the SP library and protease candidate library constructs should be sequenced to ensure that there is complete library coverage.

4.5 Conclusion

In this chapter, we have identified five protease candidates that are expressed and secreted in *B. subtilis*, four of which were previously uncharacterised. Candidates were expressed downstream of an SP library, and it was seen that the optimal SP for protease secretion was protein dependent. Optimal activities were predicted using several machine-learning based algorithms and it was found experimentally that prediction of mesophilicity was accurate in all five cases, and prediction of acidophilicity was accurate for four out of five candidates. These proteases exhibit activities in the desired ranges, and are promising candidates to take forward for purification and wash studies.

Chapter 5 – Protease purification and characterisation

5.1 Introduction

Of the five candidates identified in Chapter 4, Pro_10 was selected for further characterisation as it was expressed and secreted to a high level and could therefore be easily purified. In addition, purification of the other candidates by IMAC was unsuccessful due to insufficient binding to the column. Purification of the remaining candidates was attempted by anion exchange followed by size exclusion chromatography; however this is still ongoing.

Pro_10 was annotated as an M57 family metalloprotease, falling under the metzincin, MA(M), subclan. The metzincins are characterised by the HEXXHXXGXXH motif (where X denotes any amino acid), in which the three histidine residues constitute the zinc ligands, and the glutamate acts as a general base. In all metzincins, a “met-turn” can be observed, in which a methionine forms a hydrophobic cushion for the active site zinc ion (Gomis-Rüth, 2003). There are many characterised metzincins, but only one previously characterised M57 family metalloprotease – PrtB from *Myxococcus xanthus*. This protease is not active on its own but requires the action of a serine protease for its function (Quillet et al., 1997). With that in mind, the mechanism of Pro_10 activity and domain architecture were investigated.

There are many characteristics that are of interest with regards to protease activity in detergent applications. The ideal protease should: be non-specific to ensure that it is active against a range of stains. It should show activity at pH 7.0, as this is the standard pH during washing, and should be stable across a broad pH range to ensure formulation stability. It should also be active in a range of water sources, that is, a range of French hardness levels up to 40 °FH and NaCl concentrations up to 2% w/v. The average washing conditions for various countries can be seen in Table 5.1.

Table 5.1: Average values of French hardness, calcium: magnesium ratio and wash temperature for various countries supplied by Unilever (nr = not recorded). (Unilever, 2023a).

Country	Water hardness (°FH)	Ca ²⁺ :Mg ²⁺	Wash temperature (°C)
Argentina	12	nr	30
Australia	6	nr	30
Brazil	4	nr	25
Chile	40	nr	30
China	25	3	25
Colombia	2	nr	30
Ecuador	17	nr	25
EU	25	nr	35
France	25	nr	35
India	20	nr	27
Indonesia	30	2	28
KSA	30	nr	40
Malaysia	8	3	30
Morocco	30	nr	40
Philippines	6	2	28
South Africa	8	1.4	20
Thailand	12	3	28
Turkey	20	nr	40
UK	25	nr	35
Vietnam	18	3	25

In addition, resistance to effects of sequestrants commonly used in detergent formulations is desired, as these are commonly added to improve lifting of stains and as viscosity builders. This is an area of particular importance for metalloproteases, as these sequestrants can chelate zinc and calcium ions that are required for activity and stability. In this chapter we have characterised the M57 family metalloprotease Pro_10 and have tested its activity in a range of industrially relevant conditions.

5.2 Material & Methods

5.2.1 Plasmids and strains

The strains and plasmids used in this chapter can be seen in Table 5.1.

Table 5.2: Strains and plasmids used in this chapter.

Strain/plasmid	Genotype/description	Reference
Strains		
BA04	BA03 $\Delta amyE::prsA$	Chapter 3
Plasmids		
Pro_10_S3_pBES	Pro_10 in pBES downstream of WP_004399085.1 SP	Chapter 4
Pro_12_S7_pBES	Pro_12 in pBES downstream of YkoJ SP	Chapter 4
Pro_20_S7_pBES	Pro_20 in pBES downstream of WalM SP	Chapter 4
Pro_23_S5_pBES	Pro_23 in pBES downstream of WalM SP	Chapter 4
Pro_26_S4_pBES	Pro_26 in pBES downstream of AprE SP	Chapter 4
Pro_10_S3_M28	Pro_10 in pMAP028 downstream of WP_004399085.1 SP	Chapter 4
pBAC0001	<i>E. coli/B. subtilis</i> shuttle vector carrying Pgrac and lacI, bla, cat	Price et al., 2019
Pro_10_noC_S3_M28	Pro_10 Δ C-term in pMAP028 downstream of WP_004399085.1 SP	This chapter
pBE-S	<i>E. coli/B. subtilis</i> shuttle vector carrying PaprE, KNTase, bla.	Takara Bio
Pro_10_S3_pBES_H173A	WP_006970331.1 with H173A substitution in pBES downstream of WP_004399085.1 SP	This chapter
Pro_10_S3_pBES_E174A	WP_006970331.1 with E174A substitution in pBES downstream of WP_004399085.1 SP	This chapter
Pro_10_S3_pBES_H177A	WP_006970331.1 with H177A substitution in pBES downstream of WP_004399085.1 SP	This chapter

5.2.2 Plasmid construction

Pro_10_noC_S3_M28 was constructed using PCR to amplify a portion of the Pro_10 sequence lacking the C-terminal domain, using Pro_10_S3_pBES as a template.

PCR primers were designed to introduce a C-terminal 6xHis tag and BamHI and XbaI restriction sites. The PCR product was then digested with BamHI-HF and XbaI and ligated into the pBAC0001 vector digested with the same restriction enzymes.

The HEXXH substitution construct (Pro_10_S3_pBES_H173A, Pro_10_S3_pBES_E174A and Pro_10_S3_pBES_H177A) DNA was synthesised by Integrated DNA Technologies, with MluI and XbaI restriction sites included in the 5' and 3' ends. Plasmids were constructed through ligation of the MluI and XbaI

digested HEXXH mutation DNA with pBE-S digested with the same restriction enzymes.

5.3 Results

5.3.1 Ammonium sulphate precipitation

As a first step in purification, an ammonium sulphate cut was performed on supernatant samples to determine the ammonium sulphate saturation at which the protease precipitates. While Pro_10 is the focus of this chapter, all candidates were included here as a basis for future attempts at purification. The optimal ammonium sulphate concentration for precipitation is shown in Figure 5.1.

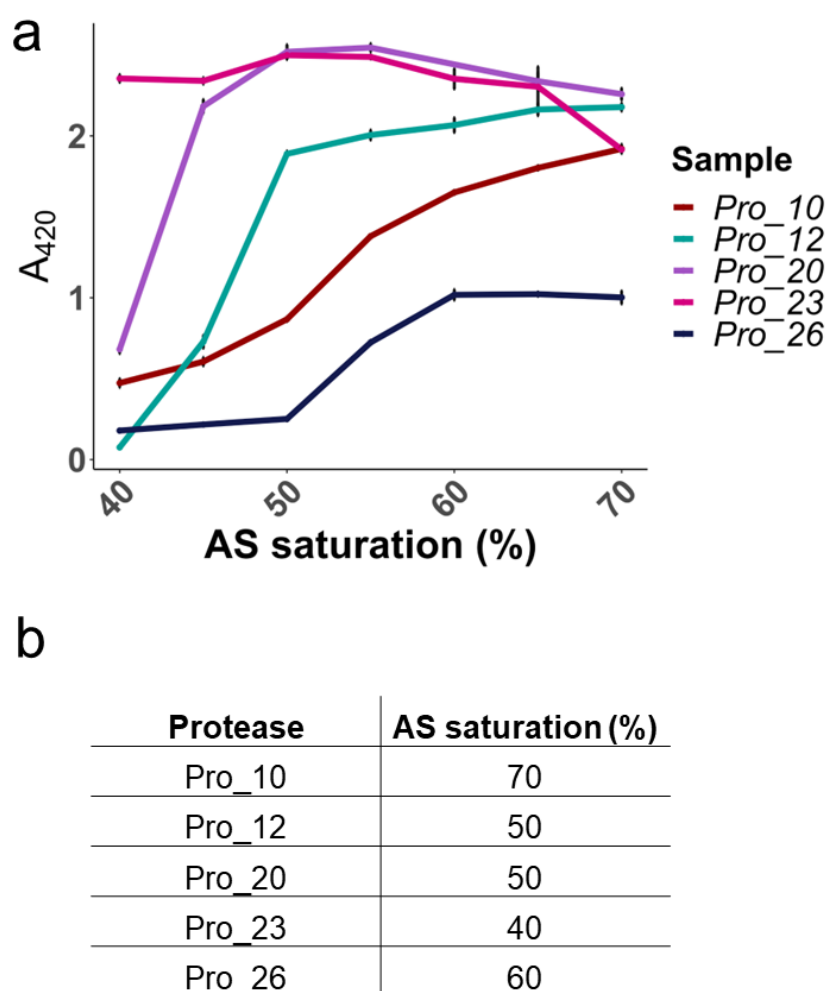


Figure 5.1: a) Protease activity of resuspended precipitate following precipitation at ammonium sulphate saturation range of 40 – 70%, as measured by azocasein assay ($n = 3$, error = $\pm$ SEM). b) Ammonium sulphate saturation at which each protease candidate precipitates.

5.3.2 Purification of Pro_10

Nickel-nitrilotriacetic acid (Ni-NTA) affinity chromatography was used to purify Pro_10 from culture supernatant. Ni-NTA affinity chromatography is a method for purifying proteins with a histidine (His)-tag. It utilises the interaction between the His-tag and immobilised nickel ions on a resin to selectively capture the target protein while removing contaminants. Samples were first concentrated by ammonium sulphate precipitation then loaded onto the column and eluted with imidazole. The absorbance at 280 nm was observed, as shown in Figure 5.2a. Throughout the purification process, 5 mL fractions were collected using the ÄKTA Start protein purification system from Cytiva.

Following Ni-NTA affinity chromatography, fractions were pooled, concentrated and further purified by size exclusion chromatography (SEC). SEC employs isocratic elution to separate proteins based on their respective sizes, with larger proteins being eluted from the column prior to smaller proteins. The absorbance at 280 nm was recorded (Figure 5.2b), and 1 mL fractions were collected using the Cytiva ÄKTA Pure protein purification system. Protease activity of each fraction was measured using the azocasein assay (Figure 5.2b). After each purification step, the active fractions were analysed using SDS-PAGE to confirm successful isolation of the protein (Figure 5.2d & 5.2e) and activity of each band was visualised by zymography (Figure 5.2c). Western blots were performed however no bands could be detected due to cleavage of the C-terminal immunolabel. It can be seen from Figures 5.2b & Figure 5.2e that the purity of the sample increased following SEC. The banding pattern in Figure 5.2c shows that there are two species with protease activity, likely corresponding to the mature protease and the catalytic domain. However, this does not align with the protein standard due to the differences in folding, as the samples here are not reduced.

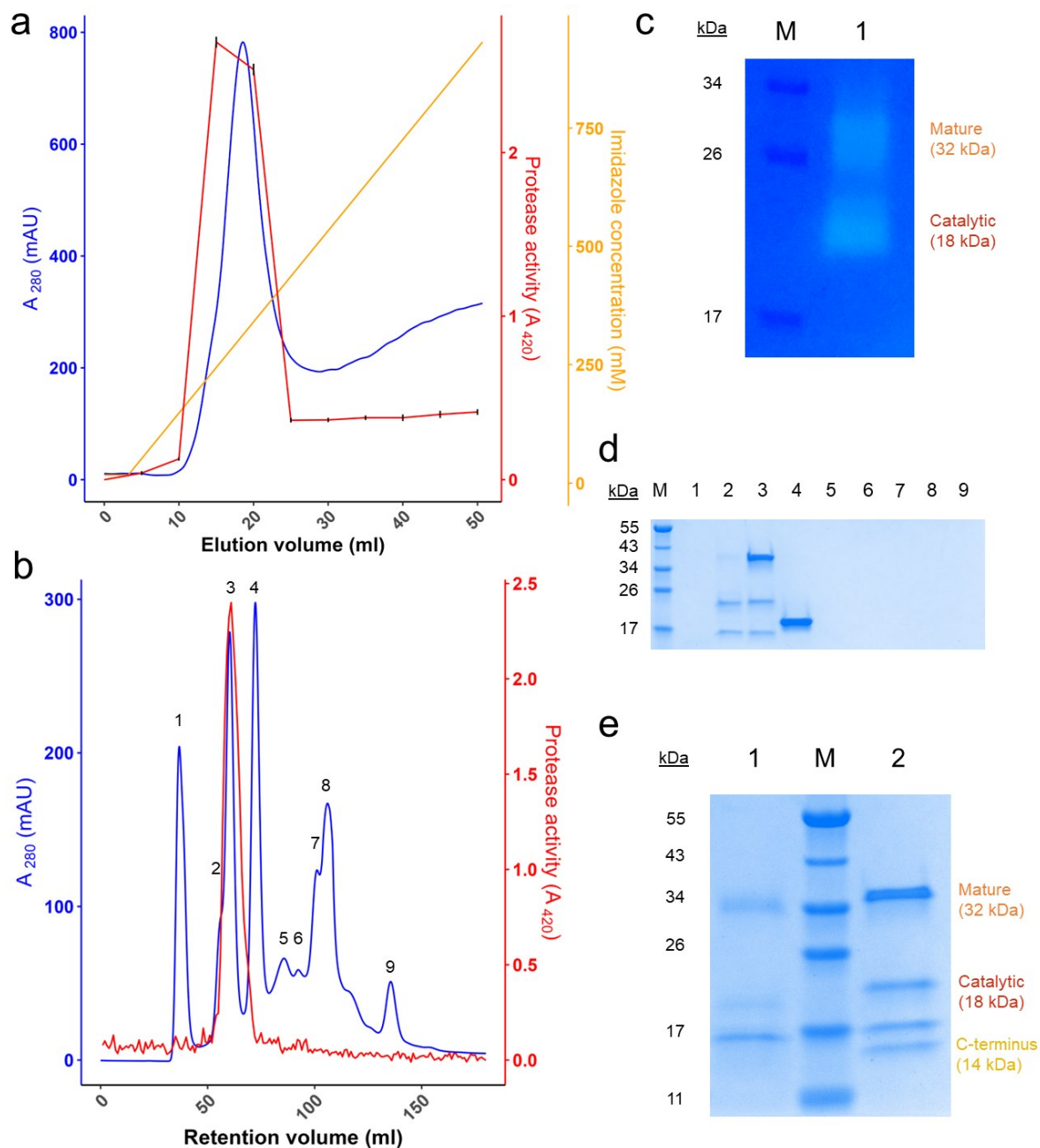


Figure 5.2: Chromatographs for ÄKTA Purification of Pro_10 (Blue line = UV280 nm) and corresponding PAGE gels following purification. a) Chromatograph for Ni-NTA chromatography (Red line = protease activity as measured by azocasein assay at pH 9.0, 40 °C, n = 3, error = +/- SEM. Orange line = imidazole concentration). b) Chromatograph for size exclusion chromatography, with first three peaks numbered (Red line = protease activity as measured by azocasein assay at pH 9.0, 40 °C). c) Gelatin zymogram following Ni-NTA chromatography. Lanes represent: M – Blue Prestained Protein Standard, Broad Range (11-250 kDa) (New England Biolabs); 1 – pooled active fractions following Ni-NTA (3 µg total protein). d) SDS-PAGE gel of peak fractions following size exclusion chromatography; M – Blue Prestained Protein Standard, Broad Range (11-250 kDa) (New England Biolabs); 1-9 – peaks 1-9 as labelled in Figure 5.2(b). e) SDS-PAGE gel following purification showing domains identified by LC-MS/MS. Lanes represent: 1 – pooled active fractions following Ni-NTA (3 µg total protein); M – Blue Prestained Protein Standard, Broad Range (11-250 kDa) (New England Biolabs); 2 – pooled fractions from peak 2 of SEC chromatograph (3 µg total protein).

5.3.3 Pro_10 salt stability

As a first step in crystallography experiments, melting temperature (T_m) of Pro_10 was assessed by nanoDSF (Section 2.11.3) in a range of concentration for various salts using the Durham salt screen (Molecular Dimensions, UK). This was performed by Dr Marleen Renders at the Unilever Research & Development Port Sunlight Laboratory. Addition of salts increased melting temperature in general, with a mean increase of 9.8 °C across all conditions (Figure 5.3). The largest increase can be seen for well E4, containing 0.4 M rubidium chloride, with an increase in T_m of 38.5 °C. The largest decrease in T_m corresponds to 4 mM europium (III) chloride hexahydrate (well G12), at 2.7 °C.

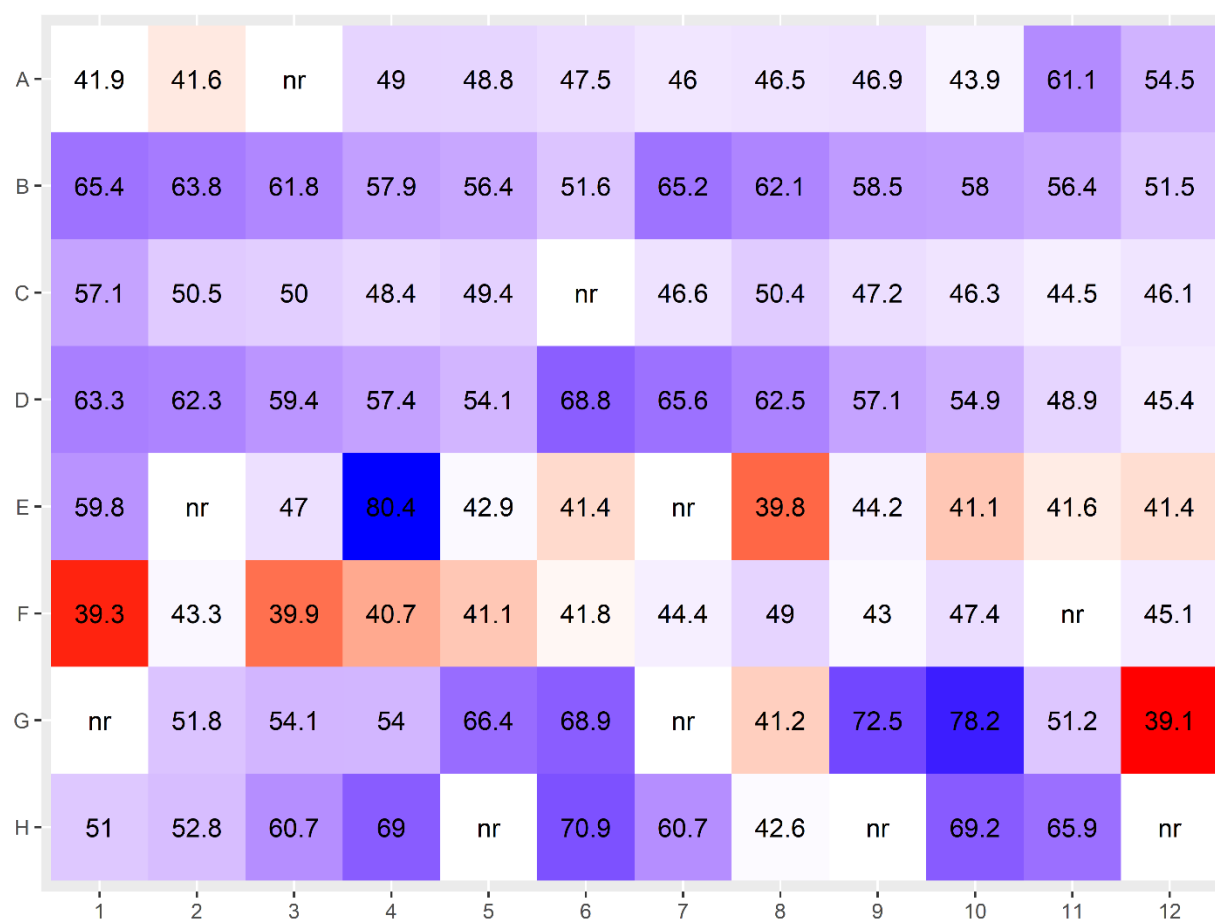


Figure 5.3: T_m values (°C) of Pro_10 for Durham salt screen calculated from ratio of fluorescence intensities at 350 nm vs 330 nm for label-free nanoDSF. Blue values indicate an increase and red values a decrease in T_m with reference to the control (ultrapure water). Well contents can be seen in Appendix E.

5.3.4 Pro_10 sequence and predicted structure

Protein bands from SDS-PAGE gels following Pro_10 purification were analysed by liquid chromatography with tandem mass spectrometry (LC-MS/MS) (Section 2.9.1) to determine N- and C-terminal domain sequences. The sites at which these domains are cleaved were confirmed by matching the sequences with the specificity matrix shown in Figure 5.5. The full protease sequence with annotations can be seen in Figure 5.4a. Fractions from peak 3 of the chromatograph shown in Figure 5.2b were also analysed by LC-MS/MS and found to contain only the C-terminal domain. The 3D structures of the pre-protease and mature protease (without the N-terminal domain, which is cleaved during maturation as described in Section 1.3.4) were predicted using AlphaFold 2 (Jumper et al., 2021) and visualised with Mol* viewer (Sehna et al., 2021). AlphaFold generates predicted local distance difference test (pLDDT) scores to assess the confidence of the model on a per-residue basis. Values over 90 indicate high confidence, and lower than 50 indicate low confidence. In this case, prediction accuracy was high, with overall pLDDT values of 85.51 and 92.83 for the pre-protease and mature protease, respectively (Figure 5.4b & Figure 5.4c). The catalytic domain of the pre-protease has a predicted structure characteristic of the metzincin upper subdomain, with a five-stranded beta-sheet and two prominent alpha-helices. However, in the mature protease this domain is predicted to have only a three-stranded beta sheet and three alpha-helices. In both cases though, the characteristic “met-turn” can be seen, with M223 situated in close proximity to the consensus HEXXHXXGXXH motif (Figure 5.4d & Figure 5.4e).

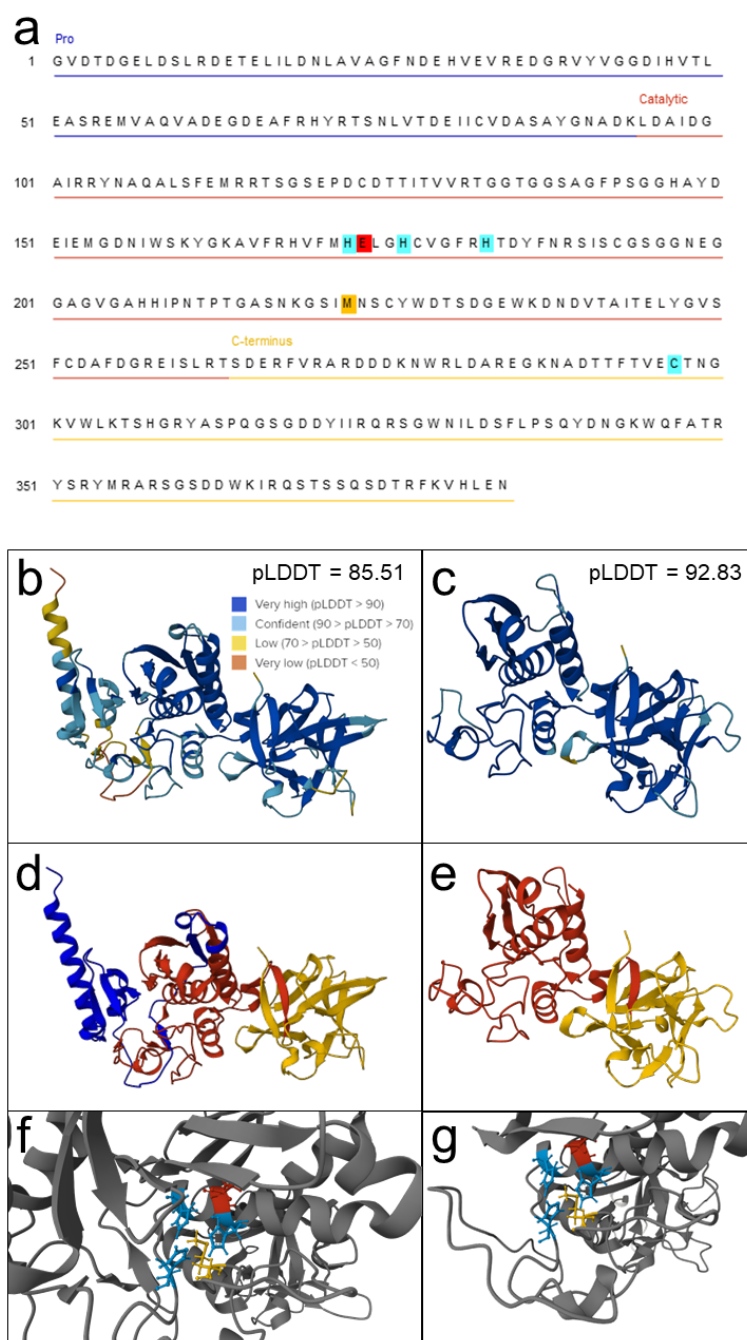


Figure 5.4: Amino acid sequence and structure of Pro_10 (blue = N-terminal propeptide, red = catalytic domain, yellow = C-terminal domain). a) Single letter amino acid sequence of Pro_10, with N-terminal propeptide (Pro), mature and C-terminal sequences annotated. Zinc binding sites predicted with high confidence using ZincExplorer (Chen et al., 2013) are highlighted in cyan, catalytic glutamate is highlighted in red, “met-turn” methionine is highlighted in yellow. b) 3D structure of Pro_10 pre-protease as predicted by AlphaFold, coloured by pLDDT, with average pLDDT shown. c) 3D structure of mature Pro_10 as predicted by AlphaFold, coloured by pLDDT, with average pLDDT shown. d) 3D structure of Pro_10 as predicted by AlphaFold, coloured by domain. e) 3D structure of mature Pro_10 as predicted by AlphaFold, coloured by domain. f) Active site of Pro_10 pre-protease as predicted by AlphaFold (cyan = zinc co-ordinating histidine, red = catalytic glutamate, yellow = “met-turn” methionine). g) Active site of Pro_10 mature protease as predicted by AlphaFold (cyan = zinc co-ordinating histidine, red = catalytic glutamate, yellow = “met-turn” methionine).

5.3.5 Pro_10 cleavage site

Detergent proteases should ideally be non-specific in their preferred cleavage site to ensure that they are active against a broad range of stains. To determine cleavage site specificity, aldolase from rabbit muscle was digested with Pro_10, as this is a large protein standard (160 kDa) that provides many possible cleavage sites. The Mass Spectrometry and Proteomics facility at the University of St. Andrews analysed the digests using electrospray ionization (ESI-MS) (Section 2.9.2) to determine the sequence of the cleavage products. The unique cleavage sites within the protein were then used to generate a specificity matrix (Figure 5.5a). This shows the probability of each amino acid being present in the P4 – P4' positions, where P1 to P4 represent the first four residues in the N-terminal direction of the cleavage site, and P1' to P4' represent those in the C-terminal direction. The data reveals Pro_10 to be non-specific in which residues are required for cleavage. There does however seem to be a preference against bulky hydrophobic residues such as tryptophan and phenylalanine, and sulphur-containing residues. A similar cleavage preference can be seen for Subtilisin Carlsberg of *Bacillus licheniformis* (Figure 5.5b), a standard detergent protease. This shows a broad range of possible cleavage sites, with a preference against tryptophan and sulphur containing residues, among others. On the other hand, *Bacillus subtilis* NprE, a well-characterised metalloprotease, is much more specific, with a strong preference for glycine, leucine and alanine in the P1, P1' and P2' positions, respectively (Figure 5.5c).

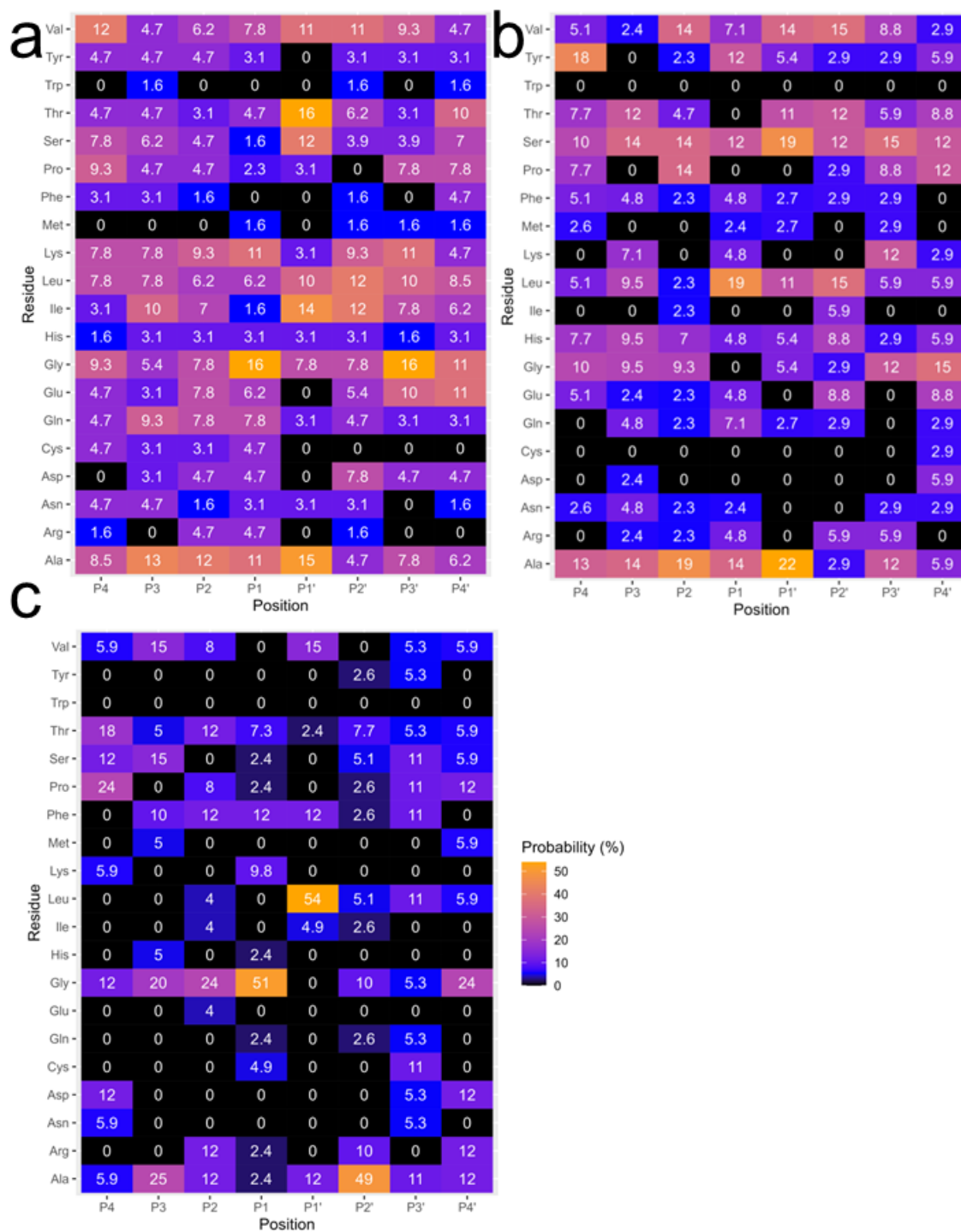


Figure 5.5: a) Specificity matrix for Pro_10 unique cleavage sites based on digestion of aldolase from rabbit muscle. b) Specificity matrix for *Bacillus licheniformis* Subtilisin Carlsberg c) Specificity matrix for *Bacillus subtilis* NprE (MEROPS, 2023).

5.3.6 Pro_10 C-terminal deletion retains protease activity

The overall yield of Pro_10 may be affected by the cleavage of the C-terminal His-tag prior to IMAC purification (Section 2.12.2). In an attempt to prevent cleavage of the purification tag, Pro_10 variants lacking either the N- or C-termini identified in Figure 5.4a were then constructed, with the concurrent addition of an N- or C-terminal His-tag. *B. subtilis* strain BA04 was transformed with these constructs. It was observed by plating onto skimmed milk agar that the N-terminal deletion was not secreted and active, however, the C-terminal deletion (Pro_10 Δ C-term) retained protease activity.

This C-terminal deletion variant was purified by IMAC following the method described in (Section 2.12.2), with the corresponding chromatograph shown in Figure 5.6a. The variant was visualised by SDS-PAGE and western blot (Figure 5.6b, Figure 5.6c), in contrast with the full length protease which could not be detected by western blot following purification. Protease activity of Pro_10 Δ C-term was compared with full length Pro_10 (Figure 5.6d), finding that the C-terminal deletion retained 31% of the activity of the full-length protease, as measured by the azocasein assay (Section 2.10.2).

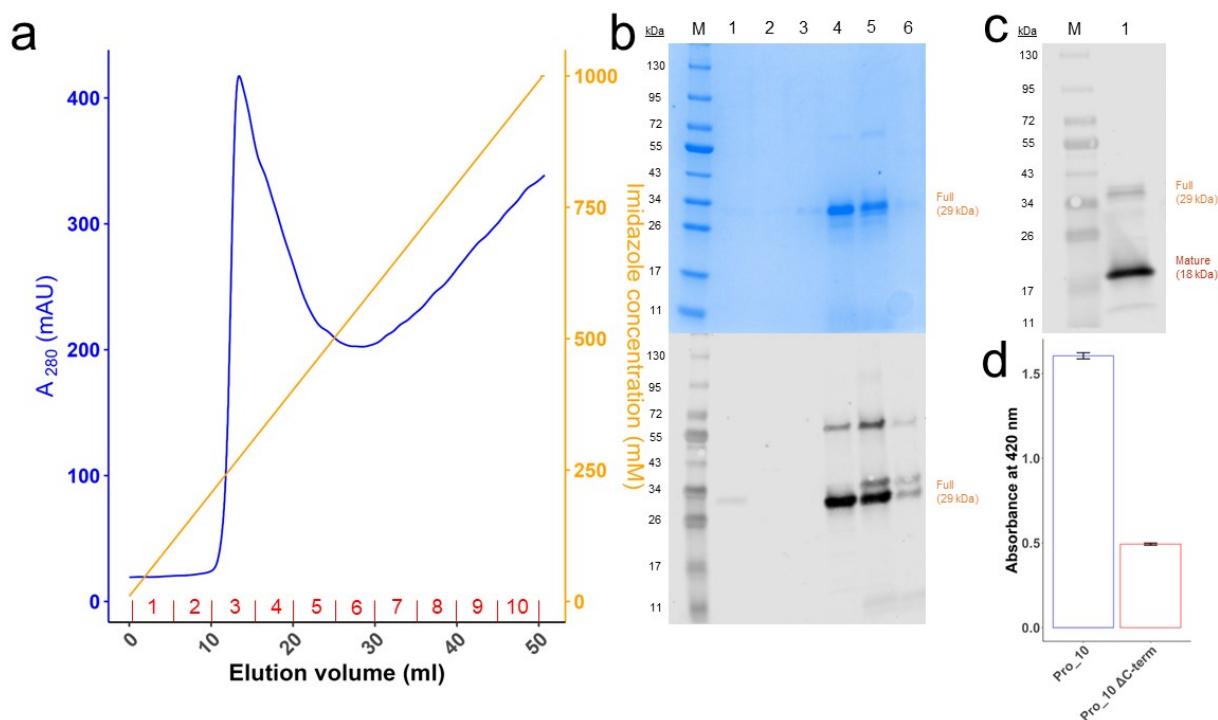


Figure 5.6: a) Chromatograph for ÄKTA Purification of Pro_10 Δ C-term (Blue line = UV280 nm, Orange line = imidazole concentration, Red lines and numbers = collection fractions). b) SDS-PAGE gel and western blot immediately following Ni-NTA chromatography. Lanes represent: M – Blue Prestained Protein Standard, Broad Range (11-250 kDa) (New England Biolabs); 1 – culture supernatant; 2 – Flowthrough; 3 – Wash; 4 – eluent fraction 3; 5 – eluent fraction 4; 6 – eluent fraction 5. c) SDS-PAGE gel of pooled eluents 7 days following Ni-NTA chromatography. Lanes represent: M – Blue Prestained Protein Standard, Broad Range; 1 – pooled fractions (3 μ g total protein). d) Protease activity of Ni-NTA purified Pro_10 and Pro_10 Δ C-term as measured by azocasein assay at pH 9.0 (n = 3, error = $\pm$ SEM).

5.3.7 Alignments of Pro_10 and PrtA/B

To compare the structure of Pro_10 with the only currently characterised M57 metalloprotease, the 3D structures of PrtB and PrtA from *M. xanthus* were predicted by AlphaFold (pLDDT values were 85.7 and 88.09, respectively). These structures were compared with that of Pro_10 by pairwise structural alignment (PSA) using jFATCAT (Li et al., 2020). Although Pro_10 and PrtB share only 50% identity, there are 236 in the 264 residues of the Pro_10 N-terminal and catalytic domains that share structural homology to those in PrtB (Figure 5.7a & Figure 5.7c), and the active site residues are aligned (Figure 5.7e). It can also be seen that of the 120 amino acids of the Pro_10 C-terminal domain, there are 83 structurally equivalent residues with PrtA (Figure 5.7b & Figure 5.7d), despite only having 19% sequence identity to each other. There is also sequence alignment with the Pro_10 C-terminal domain at two out of three PrtA active site residues (Figure 5.7f).

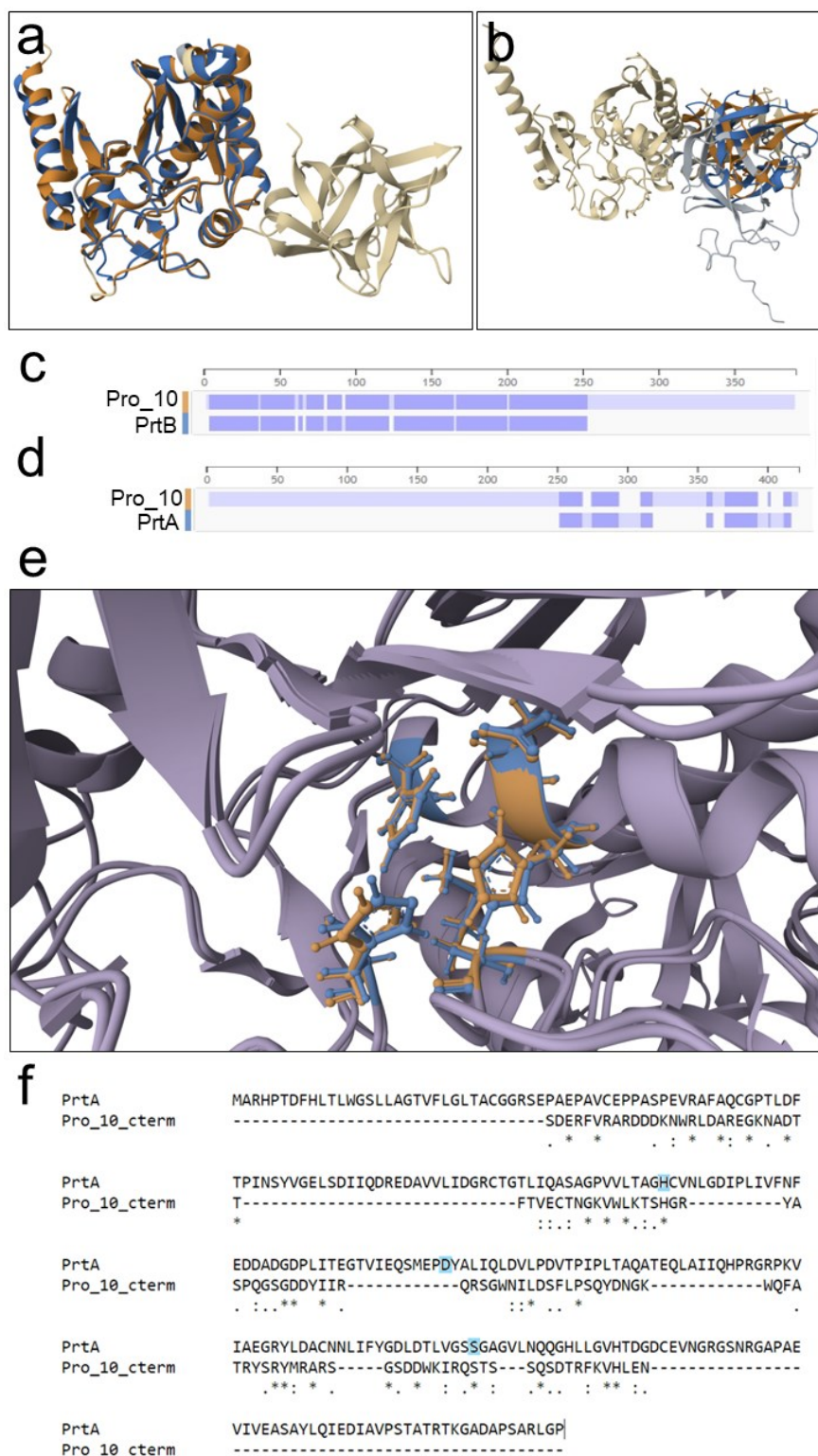


Figure 5.7: Pairwise structural alignments of Alphafold predicted structures of Pro_10 and PrtA and PrtB from *M. xanthus* (orange = Pro_10 residues aligned with PrtA/B, blue = PrtA/B residues aligned with Pro_10, grey = inequivalent residues within the aligned region, cream = unaligned residues outside the aligned region). a) 3D structure alignment of Pro_10 and PrtB. b) 3D structure alignment of Pro_10 and PrtA. c) 1D sequence alignment of Pro_10 and PrtB. d) 1D sequence alignment of Pro_10 and PrtA. e) 3D structure alignment of Pro_10 and PrtB active site residues. f) Clustal 2.1 multiple sequence alignment of PrtA and Pro_10 c-terminus.

5.3.8 Effect of temperature and pH on activity and stability of Pro_10

The effects of temperature and pH on Pro_10 activity and stability were investigated using the azocasein assay. The optimal temperature of Pro_10 was found to be 40 °C, with relative activities of >50% between 25 °C and 50 °C, and 48% activity at 20 °C (Figure 5.8a). Optimal pH was found to be 9.0, and >50% activity was observed between pH 7.0 and 9.0 (Figure 5.8b), which corresponds with the values found in Figure 4.11b. The protease was found to be stable across a pH range of 5.0 – 10.0, with melting temperatures calculated by label-free nanoDSF above 50 °C (Figure 5.8c) and residual activities all exceeding 86% following incubation for 3 hours (Figure 5.8d).

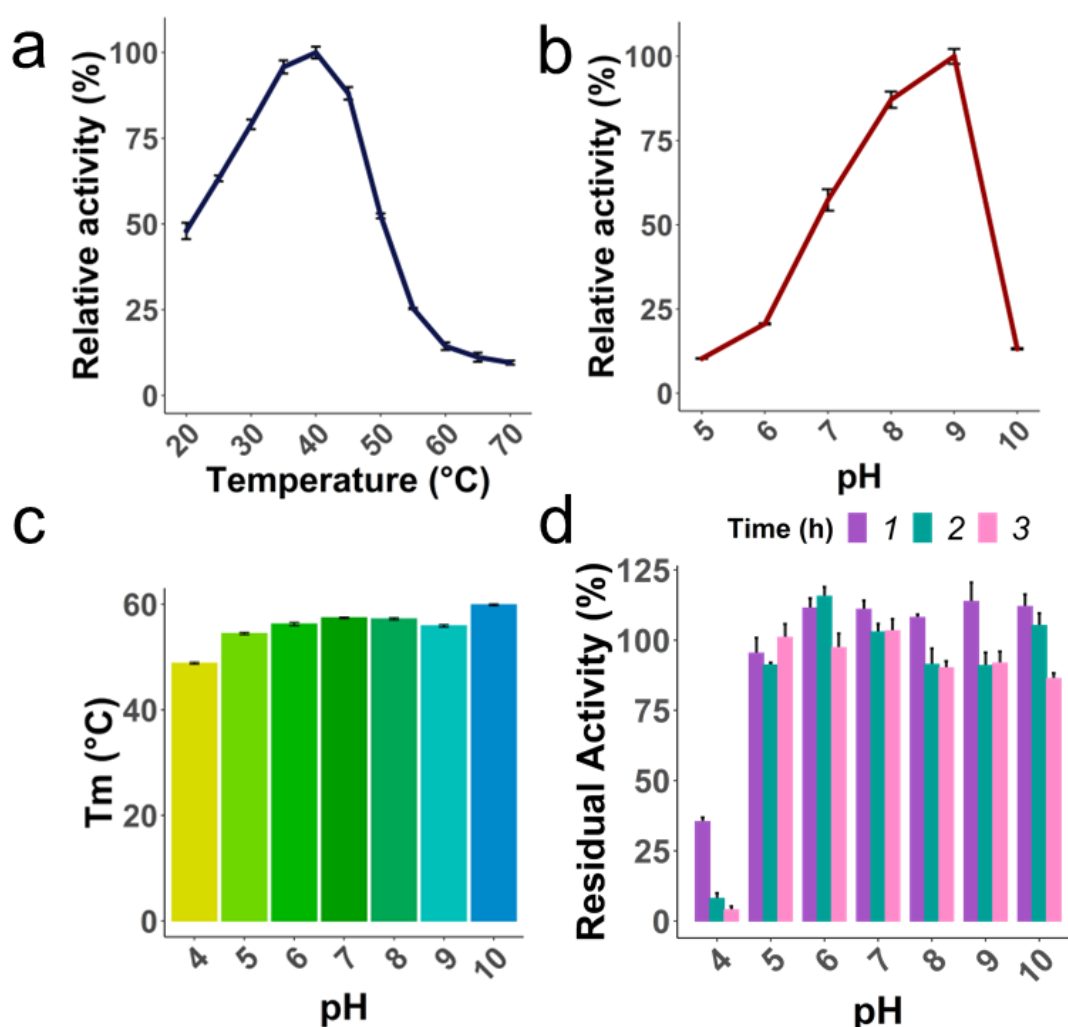


Figure 5.8: Activity and stability of Pro_10 at a range of temperature and pH (n = 3, error = +/- SEM). a) Protease activity of Pro_10 as measured by azocasein assay at pH 9.0 at a range of temperatures. b) Protease activity of Pro_10 as measured by azocasein assay at 40 °C at a range of pH. c) Average T_m values of Pro_10 at a range of pH calculated from ratio of fluorescence intensities at 350 nm vs 330 nm for label-free nanoDSF d) Residual activity of Pro_10 following incubation at a range of pH for 1, 2 and 3 h at 20 °C, as measured by azocasein assay at pH 9.0, 40 °C.

5.3.9 Effect of NaCl and water hardness on Pro_10 activity and stability

Protease activity was measured in a range of NaCl and water hardness levels and compared with two commercially available proteases. Carnival Ecity™ 16 L (Mole et al., 2020) is a serine protease of the subtilisin family developed by Novozymes (Denmark) and is a current gold-standard detergent protease. Neutrase® is a commercially available metalloprotease from *Bacillus amyloliquefaciens* (Novozymes, Denmark).

It was found that Pro_10 retained >50% activity in NaCl concentrations up to 1.5% w/v (0.26 M), however, activity dropped steeply from this point, with a total loss of activity at 7.5% w/v (1.3 M) (Figure 5.9a). This is an improvement upon Neutrase®, another metalloprotease, which retained >50% activity up to 1% NaCl, although Carnival Ecity™ 16 L outperformed both, with >50% activity up to 4% NaCl. Pro_10 was shown to have high activity at a range of water hardness levels, within the range found in Table 5.1. Activity increased up to 24 °FH, with a maximum of 152% activity relative to the 0 °FH control. From this point, activity decreased to a relative activity of 115% at 40 °FH (Figure 5.9b). This is a marked improvement over both Carnival Ecity™ 16 L and Neutrase®, both of which displayed decreased activity with increasing water hardness. Carnival Ecity™ 16 L retained >50% activity up to 36 °FH and Neutrase® retained >50% activity up to 32 °FH.

It was found that Pro_10 stability increased with increasing NaCl concentration, with melting temperatures increasing significantly as NaCl concentration is increased – from 40.2 °C at 0% NaCl up to 61.8 °C at 20% w/v (3.4 M) NaCl (Figure 5.9c). A similar result was seen following incubation with NaCl, with protease activity remaining stable when incubating with NaCl concentrations up to 20% w/v over 144 h. Over the same period in the absence of NaCl, protease activity dropped down to 6.2% residual activity (Figure 5.9d).

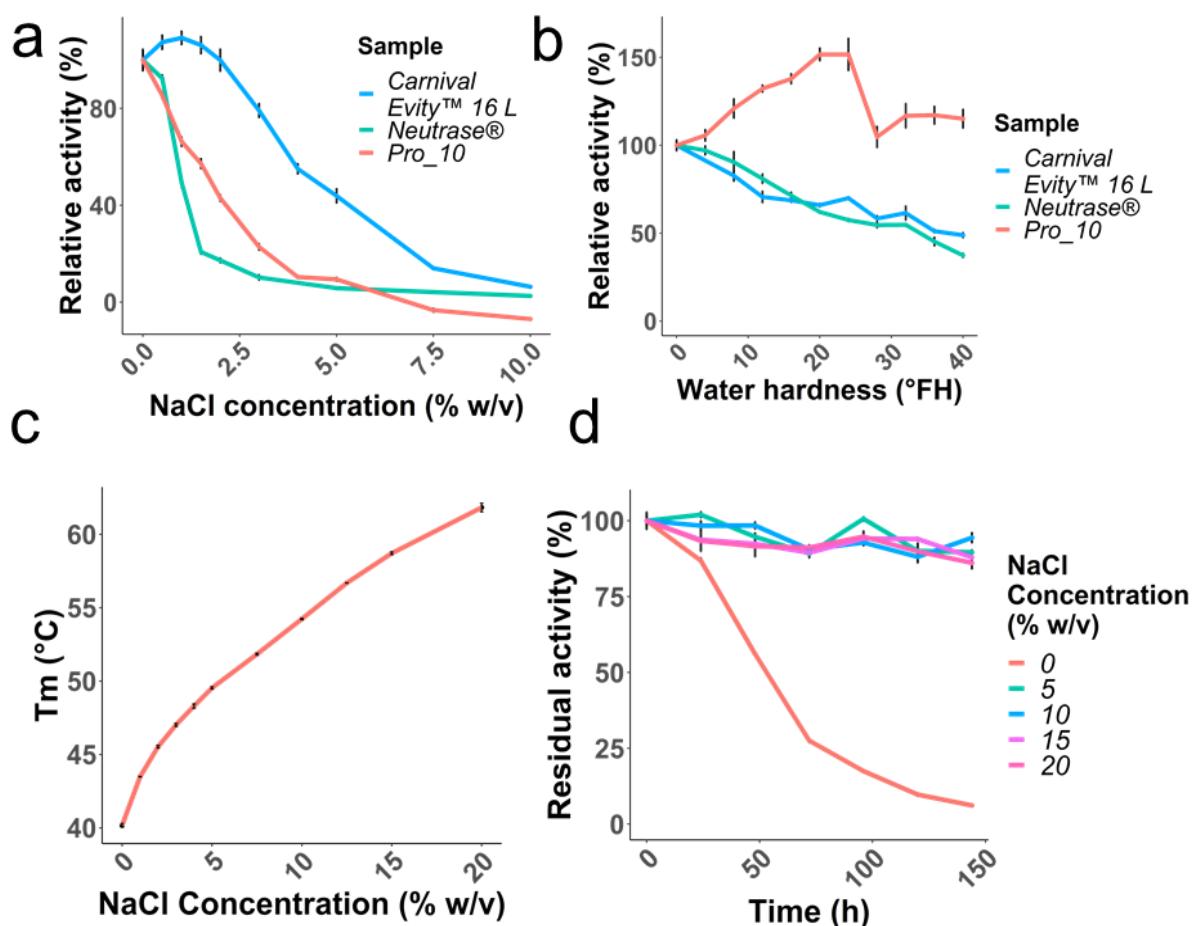


Figure 5.9: a) Protease activity across a range of NaCl concentrations, as measured by protazyme assay at pH 9.0, 40 °C (n = 3, error = +/- SEM). b) Protease activity across a range of water hardness (2:1 Ca²⁺:Mg²⁺), as measured by protazyme assay at pH 9.0, 40 °C (n = 3, error = +/- SEM). c) Average T_m values of Pro_10 at a range of NaCl concentrations calculated from ratio of fluorescence intensities at 350 nm vs 330 nm for label-free nanoDSF (n = 2, error = +/- SEM). d) Residual activity of Pro_10 following incubation in a range of NaCl concentrations at 20 °C, as measured by azocasein assay at pH 9.0, 37 °C (n = 4, error = +/- SEM).

5.3.10 Effect of sequestrants on Pro_10 activity

Sequestrants are commonly added to detergent formulations to counteract the effects of hard water through chelation of metal ions. This makes them particularly effective at inhibition of metalloproteases, however, due to chelation of the catalytic metal ions. The activity of Pro_10 and two commercially available proteases was investigated following incubation with a range of sequestrants at standard working concentrations.

Activity of Pro_10 was not significantly decreased with the addition of sequestrants, relative to the no sequestrant control (Figure 5.10a). Addition of citrate or MGDA did not affect activity at any concentration tested. Dequest 2010 (HEDP), Dequest 2066 (DTPMP) and GLDA addition increased activity proportional to concentration of sequestrant, with a 149%, 47% and 34% increase in activity, respectively, at a sequestrant concentration of 68 ppm.

Minor decreases in activity were observed for Carnival Evity™ 16 L with the addition of the highest working concentrations of citrate, Dequest 2010, Dequest 2066, GLDA and MGDA, with decreases of 17%, 18%, 21% and 12%, respectively (Figure 5.10b). An increase in protease activity was observed for Dequest 2010, with an increase of 54% at a concentration of 68 ppm.

Protease activity of Neutrase® was not significantly impacted by the addition of citrate at any concentration tested here (Figure 5.10c). A slight increase in activity was observed with addition of Dequest 2010 (a 30% increase at 68 ppm). For all other sequestrants, activity was significantly decreased, with activities of 19%, -1%, and 4% in the presence of Dequest 2066, GLDA and MGDA, respectively.

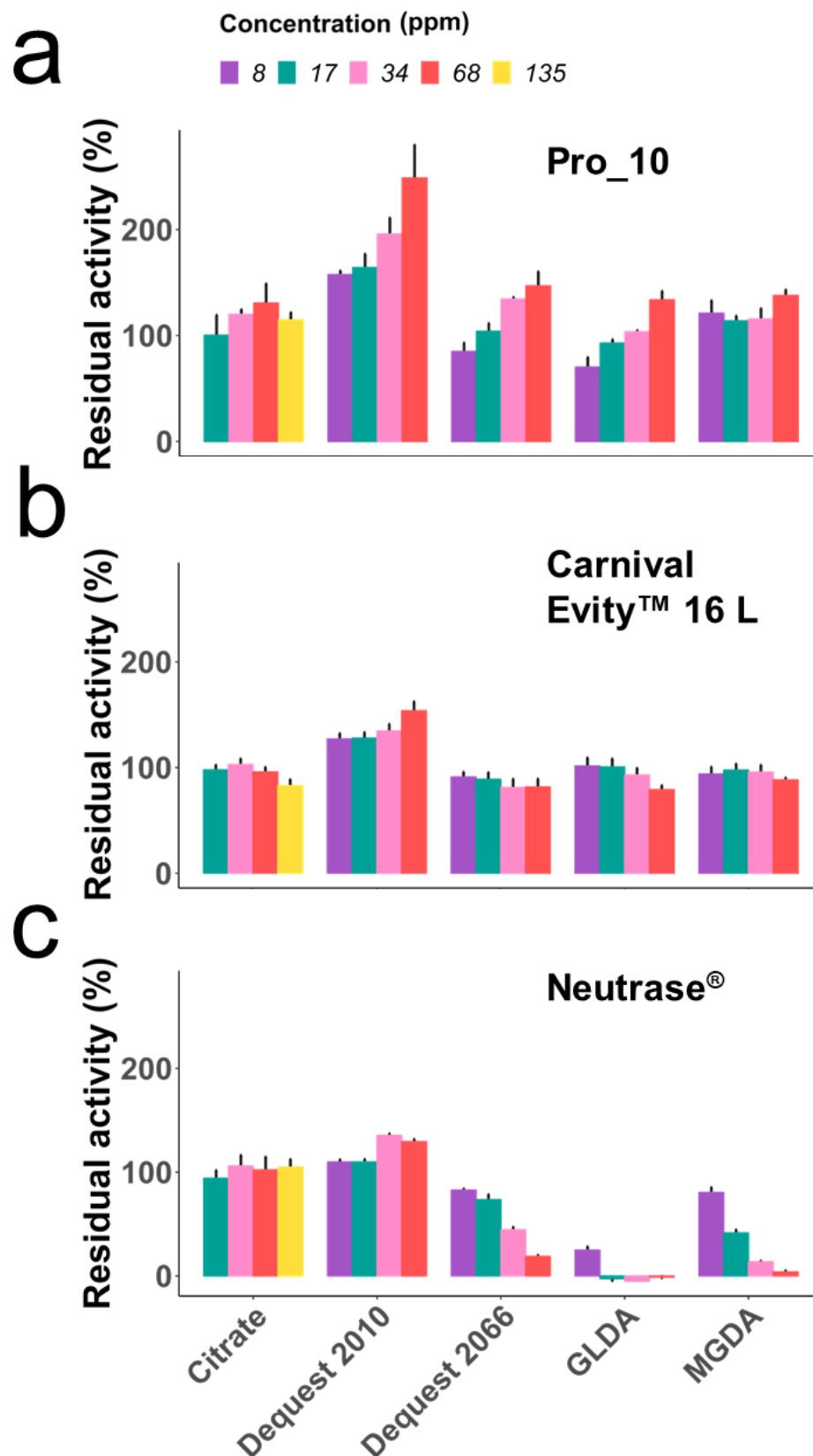


Figure 5.10: Protease activity following incubation with a range of sequestrants as measured by protazyme assay at 40 °C, pH 9.0 (n = 3, error = +/- SEM) for a) Pro_10 b) Carnival Evity™ 16 L c) Neutrase®.

5.3.11 Effect of protease inhibitors on Pro_10 activity

Protease activity was tested in the presence of a range of protease inhibitors at standard working concentrations, that target a variety of protease classes. This was blanked by a no-enzyme control for each inhibitor. Protease activity was not decreased to below 70% by any inhibitor except EDTA at 100 mM, which decreased activity to 58% (Figure 5.11a). Moderate decreases in activity were observed for pepstatin A, PMSF and TPCK, however, this is likely due to the concurrent addition of methanol required to solubilise these inhibitors, at final concentrations of 0.5, 2.5 and 5% v/v for the lowest to highest concentrations tested here. Protease activity increased at EDTA concentrations up to 10 mM, with increases of 16, 41 and 46% for 1, 5 and 10 mM EDTA, respectively, even though EDTA is a standard metalloprotease inhibitor.

Carnival Evity™ 16L was inhibited by EDTA, most likely due to chelation of calcium ions which serine proteases require for stability, with a moderate decrease in activity as EDTA concentration is increased to 10 mM, and a decrease in activity to 25% at 100 mM EDTA. Activity was not significantly decreased upon addition of E-64 or pepstatin A. As would be expected for a serine protease, addition of PMSF greatly reduced activity, with activity decreasing to 32% at 10 mM.

Addition of EDTA significantly decreased the activity of Neutrase®, which is unsurprising considering that it is a metalloprotease, with no activity observed upon addition of 1 mM EDTA and above. Neutrase activity decreased with PMSF addition, but this is likely due to the simultaneous addition of methanol.

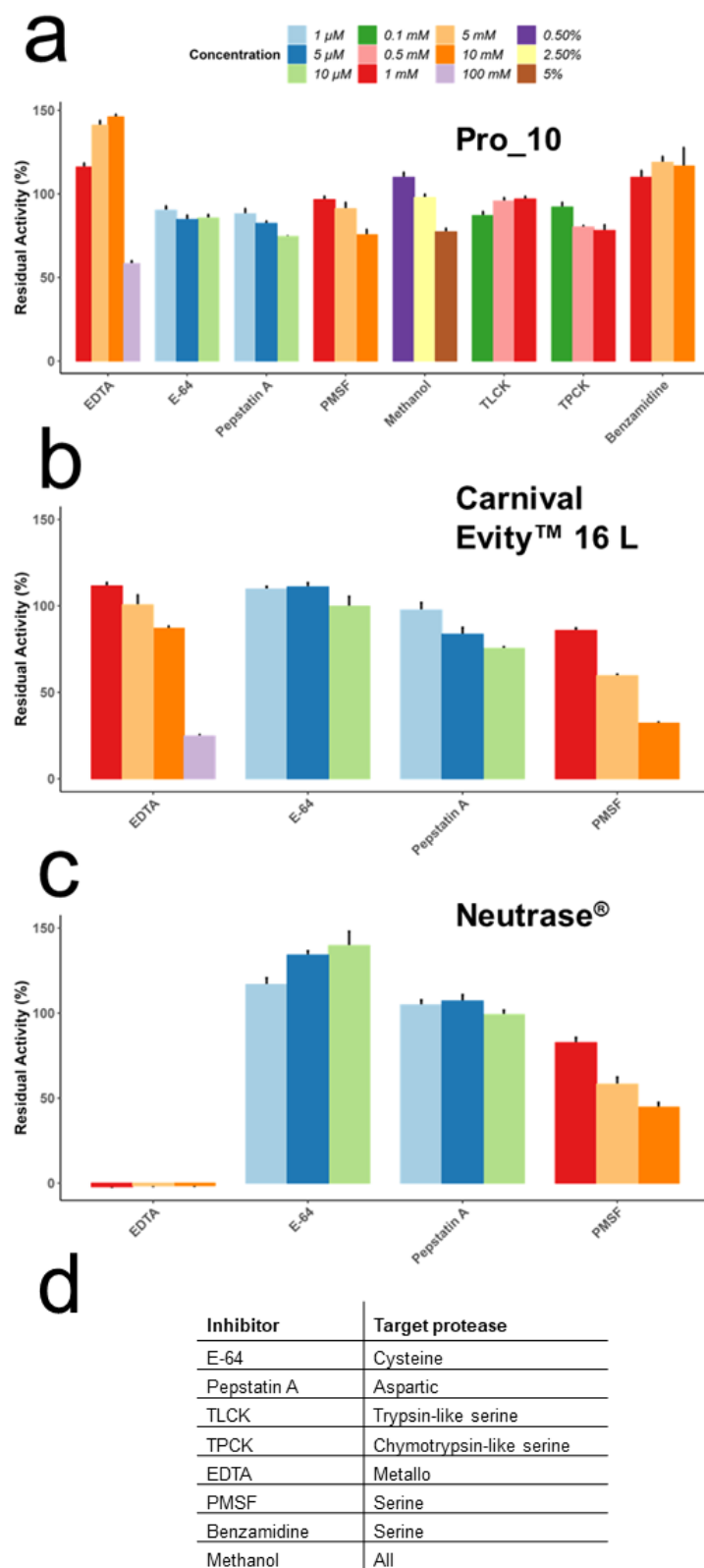


Figure 5.11: Protease activity of proteases following incubation with a range of protease inhibitors as measured by protazyme assay at 40 °C, pH 9.0 (n = 3, error = +/- SEM), and standard targets of these inhibitors. a) Pro_10. b) Carnival Ecity™ 16L. c) Neutrase®. d) Class of protease that each inhibitor is active against.

5.3.12 Mutations in the HEXXH motif are catalytically inactive

To confirm that Pro_10 is a metalloprotease, constructs were designed with mutations in the HEXXH active site motif that is characteristic of the zincin clan of metalloproteases. *B. subtilis* strain BA04 was transformed with these constructs. Following growth on skimmed milk agar, protease activity was assessed by zones of clearing around the transformants. No clearing was seen for any of the transformants, confirming that these residues are required for protease activity (Figure 5.12).

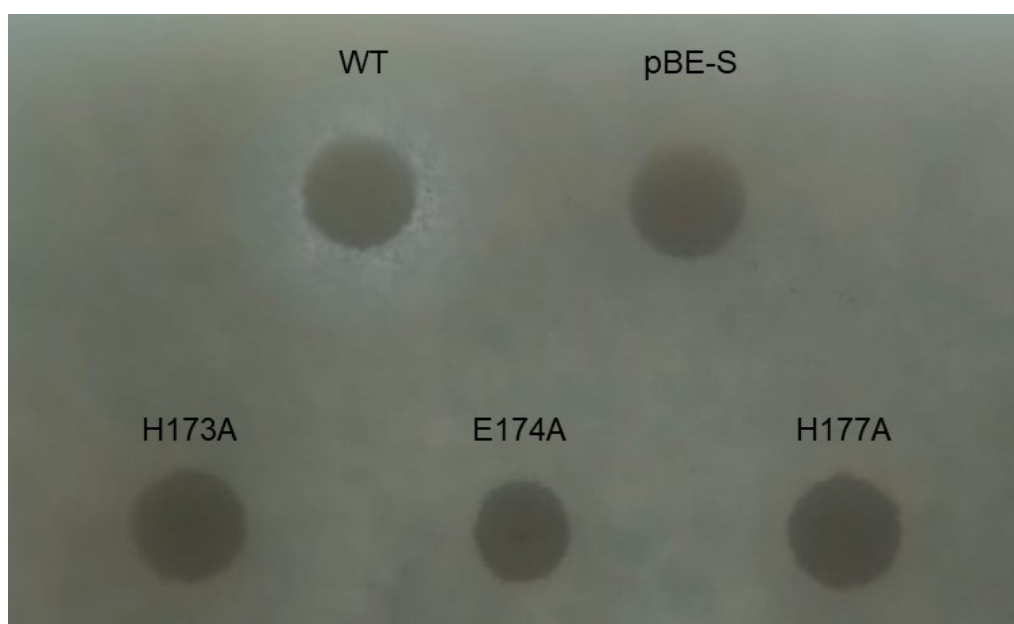


Figure 5.12: Skimmed milk agar plate showing zones of clearing surrounding *B. subtilis* colonies secreting Pro_10 HEXXH active site substitutions (H173A, E174A, H177A), wild-type Pro_10 (WT) and empty pBE-S vector (pBE-S).

5.3.13 Removal of proteinaceous stains by Pro_10

The efficacy of Pro_10 on removal of proteinaceous stains was compared with that of the current gold standard protease, Carnival Ecity™ 16 L, following the method described in Section 2.13. This was carried out by Dr Marleen Renders at the Unilever Research & Development Port Sunlight Laboratory. Initial study of Pro_10 detergent formulation stability indicated that it was not active in standard formulation, so this will have to be engineered in the future. As such, cleaning assays here were performed in Tris-HCl buffer. Protease samples were normalised by activity, as determined by the azocasein assay (Section 2.10.2), where one unit of activity was defined as an increase in A_{420} of 1 (Figure 5.13a). It was found that Pro_10 was more effective than Carnival Ecity™ 16 L at removal of blood and gravy stains

(Figure 5.13b). Comparable levels of stain removal were seen for brown sauce, droste cocoa and Frijj chocolate milk stains, and Carnival Evity™ 16 L was more effective at removal of chocolate ice cream, chocolate milk and scrubbed grass stains in the conditions tested.

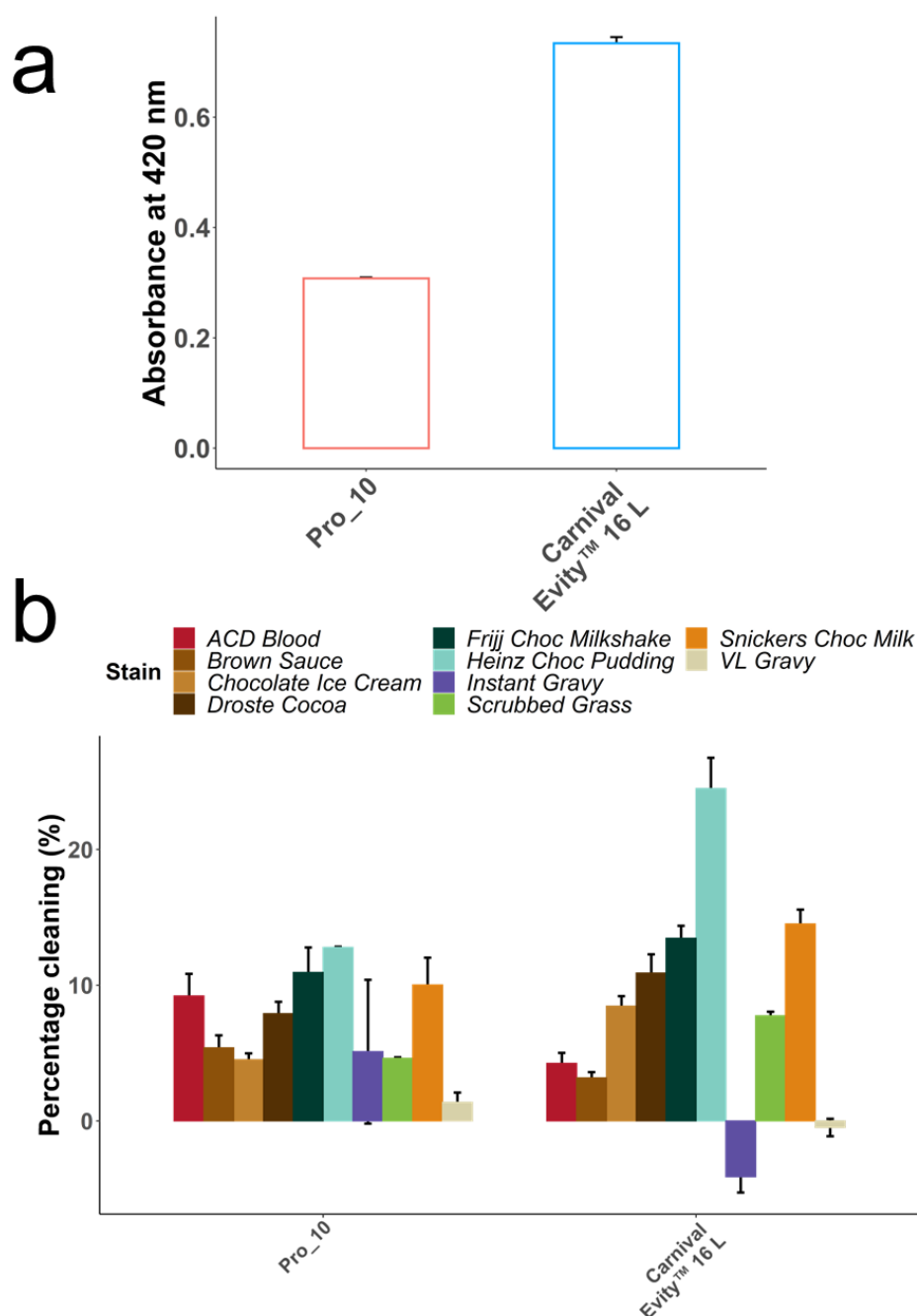


Figure 5.13: a) Protease activity of Pro_10 and Carnival Evity™ 16 L at 100 µg/mL as measured by azocasein assay (n = 3, error = +/- SEM). b) Percentage cleaning of proteinaceous stains after washing in tergotometer following pre-treatment with Pro_10 and Carnival Evity™ 16 L at 3 units/mL (n = 2, error = +/- SEM). Blanked by no enzyme control.

5.4 Discussion

5.4.1 Pro_10 C-terminus is required for full protease activity

It was found that Pro_10 is composed of an N-terminal propeptide that is cleaved during maturation, a catalytic core domain and a C-terminal domain that may be cleaved off – as seen in peak 3 of Figure 5.2b. However, this C-terminus must be present for full function of the protease, as deletion of this domain reduces protease activity by 69% (Figure 5.6d). The association of the catalytic domain and C-terminus may be disrupted by SDS-PAGE, suggesting that the C-terminus is only loosely bound (Figure 5.2d). It may be seen from the predicted structure in Figure 5.4b and Figure 5.4c that the C-terminal domain is not tightly associated with the catalytic domain, providing further evidence for this claim. This may be compared with the only other characterised protease in family M57 (at time of writing): PrtB from *Myxococcus xanthus*. PrtB was found to only exhibit protease activity in conjunction with PrtA, a trypsin-like serine protease (Quillet et al., 1997). While these proteases do not share significant sequence identity (50% between Pro_10 and PrtB), they appear to have a high degree of structural homology in the N-terminal and catalytic domains (Figure 5.7). Interestingly, the C-terminal domain shares a significant level of structural elements with PrtA, with 69% of its amino acids showing equivalent residues in PrtA, despite having only 19% sequence identity. This suggests that the C-terminal domain has a similar function to PrtA of *M. xanthus*, with regards to activation of full protease activity. While the C-terminal domain has no azocaseinase activity, it may still have protease activity that can activate Pro_10. This is supported by the alignment of the Pro_10 C-terminus with two of three catalytic residues in PrtA. On the other hand, this effect may be due to binding of the C-terminus to the catalytic domain and either increasing stability or activity.

This finding is, however, dependent on predicted 3D structures. While the pLDDT values implied significant confidence in these models, it would be more reliable to produce the 3D structure through experimental means such as X-ray crystallography. To this end, a Durham screen was carried out to find stabilising conditions for the protease, with the results shown in Figure 5.3, which will aid in future crystallography experiments.

5.4.2 Pro_10 is active in a range of industrially relevant conditions

The activity of Pro_10 was tested in conditions applicable for use in the detergent industry. It was found that its optimal temperature was 40 °C, with 48% activity at 20 °C (Figure 5.8). This is applicable to a broad range of washing temperatures, from 20 °C in South Africa to 40 °C in Morocco, for example (Table 5.1). The protease was stable between pH 5.0 and 10.0, which would be applicable to a wide range of detergent formulations and had over 50% activity at pH 7.0, the standard pH during washing. It retained over 50% activity up to 1.5% w/v NaCl, with 43% activity at 2%, which is sufficient for most washing conditions (Figure 5.9a). The protease also showed increased activity at French hardness levels up to 40 °FH: the maximum average water hardness seen in countries supplied by Unilever (Table 5.1), while two commercially available proteases showed decreased activity in these conditions (Figure 5.9b). Finally, protease activity was not significantly lowered by common sequestrants used in detergent formulations (Figure 5.10), and in most cases increased at the highest sequestrant concentrations tested. In addition, Pro_10 was found to be more effective at removal of blood and gravy stains than the current industry standard, Carnival Eivity™ 16 L (Figure 5.13b).

5.4.3 Pro_10 is resistant to a range of chelating agents

It was found that Carnival Eivity™ 16 L was active in a range of sequestrants, with an increase in activity upon addition of Dequest 2010. Neutrase®, on the other hand, was significantly inhibited by addition of several sequestrants. This would be expected, as Neutrase® is a metalloprotease and Carnival Eivity™ 16 L a serine protease, and metalloproteases are sensitive to addition of chelating agents due to chelation of catalytic and structural metal ions. Pro_10 activity was not significantly affected by various sequestrants (Figure 5.10), with activity increasing with addition of Dequest 2010, Dequest 2066 and GLDA. This is surprising, as there is no apparent correlation with binding strength. GLDA has one of the lowest binding strengths of the sequestrants tested, while Dequest 2066 has the highest, with a higher logK for Ca²⁺ and Zn²⁺ than EDTA (Anastas et al., 2014). It was also seen that activity increased with the addition of EDTA up to concentrations of 10 mM (Table 5.4). It is unclear why activity increased, although one possibility is that nickel ions were carried over from IMAC purification that inhibited protease activity. Addition of

sequestrants would then reverse this inhibition through chelation of nickel. Regardless, it was found that protease activity was only significantly affected at 100 mM EDTA, over 100 times the standard working concentration. This suggests that the catalytic zinc ion is buried deeply within the protease, shielding it from chelation.

5.5 Conclusion

In this chapter, we have characterised a protease from the relatively uncharacterised M57 family of metalloproteases. It was found to contain an N-terminal propeptide and a C-terminal domain that is required for full protease function. This protease was shown to share structural homology with the only other characterised M57 protease, despite limited sequence identity. Activity and stability of Pro_10 was then tested, and it was found to be active and stable in a range of industrially relevant conditions. It was also shown to be highly resistant to the effects of chelating agents such as EDTA. Finally, it was found to be active against a range of proteinaceous stains, and more effective at removal of blood and gravy stains than the industry standard, Carnival Ecity™ 16 L.

Chapter 6 – Conclusions and future work

It is essential that we adopt processes with lower energy consumption, due to the ongoing energy crisis. Historically, the search for new proteases for use in detergents has focused on thermostability as this is generally correlated with increased activity overall (Gupta et al., 2002). However, discovery of novel proteases with activity at lower temperatures will facilitate reduced washing temperatures, thereby decreasing energy costs for consumers and the global carbon footprint of laundry processes. In addition, metalloproteases have been found to be more effective than the more conventional serine proteases at removal of certain stains, bloodstains in particular (Zhang et al., 2019).

The search for novel proteases for use in the detergents industry generally involves cultivation of protease producing organisms (Chen et al., 2011; Reddy et al., 2017; Riffel et al., 2003; Wang et al., 2008). However, this is often time consuming and costly, and inherently excludes up to 99.9% of prokaryotic organisms due to issues with cultivation (Berini et al., 2017). An alternative may be found through the use of metagenomic databases, composed of sequences obtained through direct sequencing of environmental samples. This leads to the identification of a vast wealth of hypothetical protein sequences that may be used in industry. However, the difficulty lies in refining these datasets as there are, for example, over 390,000 known and predicted metalloprotease sequences (Rawlings et al., 2010). This study aimed to provide a targeted approach for protease discovery through prediction of protease characteristics based solely on amino acid sequence. To do so, datasets were constructed using BLAST and HMMER searches of a range of databases, with known metalloproteases as query sequences (Altschul et al., 1990; Eddy, 2011). These were then refined by predicted activities using several machine-learning based algorithms, based on rotation forest and SVM, to classify proteins as mesophilic, psychrophilic, halophilic, and acidophilic (Lin et al., 2013; Lin & Chen, 2011; Nath, 2016; Nath & Subbiah, 2014). These algorithms have recorded prediction accuracies greater than 70%. Datasets were then further refined by presence of signal peptides (SPs) to suggest that they are secreted, and redundant sequences were removed. The resultant dataset of protease candidates was then taken forward for expression and purification.

This approach requires a robust system for heterologous expression of proteases to ensure the greatest likelihood of sufficient expression. *B. subtilis* was chosen as a host for this study due to its robust secretory capabilities and known capacity for protease expression (Degering et al., 2010; Maurer, 2004; Schallmey et al., 2011). There are many *B. subtilis* strains that have been adapted for use in industry for improved protein secretion. Approaches generally involve deletion of native proteases to reduce degradation of secreted proteins, as well as upregulation and modification of secretory pathway genes, and deletion of genes implicated in sporulation and formation of foam, as this can affect protein recovery (Chen et al., 2015; Kakeshtia et al., 2010; Pohl et al., 2013; Quax et al., 2002; Westers et al., 2008; Wu et al., 1998, 2002; Wu et al., 1993). In this study, a range of *B. subtilis* strains were constructed through the use of CRISPR-MAD7 for precise, targeted genome modifications (Price et al., 2020). Minimal *B. subtilis* strains were chosen as parental strains, as genome reduction has been shown to improve yields of proteins that are difficult to express (Aguilar et al., 2020; Reuß et al., 2017).

Initially, all native extracellular proteases were deleted to both reduce background activity and improve yield. This was found to have no effect on specific growth rate, cell morphology or transformation efficiency. In an effort to increase protein secretion, the Sec pathway component *prsA* was overexpressed through integration of a xylose-inducible system. This foldase is involved in folding of secreted proteins following translocation through Sec pathway components into the extracellular space. Upregulation of *prsA* can increase protein yield by preventing blockages of these Sec pathway components (Chen et al., 2015). In addition, the α -amylase *amyE* was deleted during integration of the *prsA* overexpression cassette to lower protein levels in the culture supernatant, which could impact purification. It was found that the integration of this cassette reduced protease secretion in the absence of xylose (the inducing molecule), however overexpression of *prsA* improved secretion compared to the parental strain upon the addition of xylose. An increase in protease level in the supernatant of 11.6-fold was observed upon addition of xylose. While this method did demonstrate a marked improvement in secretion of a heterologous protease, these strains could be developed further through the knockout of further genes. For example, deletion of *srfAC* has been shown to reduce foam formation and improve yields of secreted proteins (Liu et al., 2018b). Furthermore, optimisation

of promoters and ribosome binding sites could be explored to improve protease expression (Li et al., 2022; Liu et al., 2018a; Mao et al., 2021).

Following the generation of strains for improved protease secretion, expression of the candidate proteases identified through bioinformatic approaches was investigated. As the efficacy of SPs for protein secretion is often highly protein-dependent, and *in silico* prediction of SP efficiency is not currently feasible (Fu et al., 2018), candidates were expressed using a combinatorial library of SPs (Brockmeier et al., 2006; Degering et al., 2010). Protease secretion and activity was then confirmed through the observation of zones of clearing following transformation on skimmed milk agar. Of the 38 protease candidates identified, 5 were found to be secreted and active. One of these was from the Gram-negative bacterium *Plesiocystis pacifica*, while the remaining four were all from the Gram-positive phylum *Bacillota*, the same phylum as the host, *B. subtilis*. This demonstrates the importance of the choice of host for the expression of heterologous proteins.

We then demonstrated the importance of screening an SP library for optimisation of protein secretion. SPs that were effective at secretion of each protease candidate were compared, and it was found that for most candidates the SP impacted the secretion of active protease. The efficiency of secretion was found to heavily rely on the type of protease being secreted, each having a distinct set of SPs. There was significant variability in the length, charge, and hydrophobicity of these SPs. Interestingly, this variability did not seem to have a consistent correlation with secretion efficiency, except for Pro_12, which displayed higher secretion levels for SPs with lower charge in the N-region. Some SPs were effective for multiple candidates, notably AbnA, AprE, and WalM, suggesting that these SPs might be particularly efficient at facilitating the secretion of metalloproteases, or it's possible that they are more prevalent in the signal peptide library. Further study is required to ensure that this library had complete coverage. A decreased library coverage could be due to random variation, or incompatibility of specific SPs with transformation into either *E. coli* or *B. subtilis*. For example, Brockmeier et al. (2006) found that of the 173 Sec SPs tested for expression of cutinase, 25 could not be transformed into *B. subtilis* as they had a lethal effect. Similarly, depending on choice of promoter, certain SP constructs may be targeted to the *E. coli* periplasm, affecting growth during library construction (Li et al., 2022).

The activity of candidate proteases that were secreted and active was then assayed at temperature and pH, to determine the accuracy of the predictions of psychrophilicity, mesophilicity and acidophilicity. Our findings highlight the ongoing challenge in accurately predicting psychrophilicity. Among the five proteases examined in this study, all exhibited their optimal activity within the mesophilic temperature range. This suggests that the predictions of psychrophilicity based on the category I rules established by Nath & Subbiah in 2014 were less than optimal. However, the predictions of mesophilicity using Thermopred (Lin & Chen, 2011) were found to be accurate. These results align with previous literature findings, where the highest accuracy achieved in psychrophilicity prediction by Nath & Subbiah (2014) was 70.5%, while Lin & Chen (2011) achieved an accuracy of 92.7% in predicting mesophilicity.

Furthermore, our analysis indicated a high level of accuracy in predicting acidophilicity, as four out of five proteases displayed substantial activity within the slightly acidic pH range of 5 to 7. This level of accuracy is less than that reported by Lin et al. (2013), of 96.3%. This highlights the discrepancy between bioinformatic and experimental results. However, it is worth noting that these results are from a sample size of 5 proteases. To conclusively establish such a difference in prediction accuracies, further experimentation with a larger sample size is warranted.

Of these candidate proteases, one was successfully purified through IMAC and SEC, designated Pro_10. This is the second protease in family M57 to be characterised. The only other characterised member of this family is *prtB* from *Myxococcus xanthus*, which was expressed in *E. coli* alongside *prtA*, a polycistronic serine protease of family S1 (Quillet et al., 1997). It was found that PrtB only exhibited protease activity when co-expressed with PrtA. This was hypothesised to be due to the activation of PrtB by proteolytic cleavage mediated by PrtA. A similar mechanism can be seen in the activation of MMP-9 by serine and metalloproteases (Desrivieres et al., 1993; Geurts et al., 2012; Jackson et al., 2010; Ogata et al., 1992). In this study, we found that mature Pro_10 is composed of an N-terminal propeptide, a catalytic domain and a C-terminal domain. This C-terminal domain does not exhibit azocaseinase activity on its own and has no significant homology with any known protease. However, the C-terminal domain is required for full protease activity of Pro_10, as truncation leads to a 69% reduction in activity.

The 3D structures of Pro_10, PrtA and PrtB were modelled using AlphaFold2 and their structures were compared by pairwise structural alignment. We found that while they do not share a great deal of sequence identity (50%), they do have a high degree of structural alignment, with 89.4% of Pro_10 residues in the catalytic and N-terminal domains having a structural equivalent in PrtB. Interestingly, the C-terminal domain of Pro_10 shared significant structural homology with PrtA, with 69% of residues having a structural equivalent, despite very limited sequence identity (19%), and no innate protease activity. This suggests that PrtA and the Pro_10 C-terminus may perform similar functions with regards to activation of protease activity of PrtB and Pro_10, respectively. The Pro_10 C-terminal sequence aligns to two out of three catalytic residues of PrtA, suggesting that it may have serine protease activity. This may be sufficient to fully activate Pro_10. Another possibility is that neither of these proteases are activated by proteolytic cleavage, but may instead be due to binding of the respective peptides. To investigate whether this is indeed the underlying mechanism of activation, it should be determined whether PrtA binds to PrtB through pulldown assays or SEC.

The mechanism of action of other proteases from family M57 should also be investigated, to determine whether they have a similar mechanism of action. The Pro_10 C-terminal domain shares alignment with the C-terminal domain from only one other M57 protease: a predicted metalloprotease from a *Deltaproteobacteria* bacterium (Genbank: MCH9686466, 28.7% identity). Characterisation of this protease would be a good starting point for further investigation, and could help elucidate whether the M57 C-terminal domain has protease activity of its own.

It should be noted that these structural analyses are based on predicted 3D structures, and while AlphaFold2 has proven to have a high accuracy, predictions are based on known structures. None of these proteins have significant homology to any protein with a solved crystal structure, with the closest known structure in the PDB having identities of 24%, 37% and 37% to PrtA, PrtB and Pro_10, respectively (corresponding to *Staphylococcus aureus* exfoliative toxin D2, *Homo sapiens* polycomb protein SUZ12, and *Homo sapiens* Fascin, respectively). Therefore, the accuracy of the models produced by AlphaFold2 may be in question, despite the high pLDDT values reported here. To reliably confirm that there is structural

homology between these proteins, their 3D structures should be determined experimentally.

The applicability of Pro_10 for use in detergents was then determined. It was found that it is non-specific, which is important for use in detergents as this implies it will be active against a range of stains, although it does have a bias against bulky hydrophobic residues and sulphur-containing residues. Its activity was assayed in a range of temperatures and pH values, and it was found to be most active at 40 °C, pH 9, although retained approximately 50% activity at 20 °C and at pH 7. Future work should focus on directed evolution to optimise this protease for use at lower temperatures and pH values, as well as to improve stability in detergent formulations. This would facilitate washing at lower temperatures and at neutral pH, as found in standard wash conditions. The protease did, however, appear stable across a pH range of 5-10, ensuring stability in a range of possible pH values for detergent formulation. To confirm applicability of this protease for use in the detergents industry, studies should be undertaken to test pH stability over a longer timeframe, as well as to determine stability in various detergent formulations.

The activity of Pro_10 at a range of NaCl concentrations and water hardness levels was compared with that of two commercially available proteases: the metalloprotease Neutrase® and the gold standard serine protease Carnival Ecity™ 16 L. It was found that Pro_10 was more active than Neutrase® at increasing NaCl concentrations, however Carnival Ecity™ 16 L had significantly higher activity than the other two proteases with increasing NaCl concentration. Future studies could be undertaken to improve the halophilicity of Pro_10, increasing applicability in saline and coastal water sources. Pro_10 was much more tolerant to increasing water hardness than either Neutrase® or Carnival Ecity™ 16 L. Both of the commercial proteases showed decreasing activity levels as water hardness increased, while Pro_10 activity actually increased with water hardness.

We then compared protease activity with the addition of several sequestrants that are commonly added to detergent formulations. Carnival Ecity™ 16 L was shown to not be significantly impacted by addition of sequestrants, which would make sense as it is a serine protease and therefore does not rely on metal ions for activity, only stability. Marginal decreases in activity were observed at some of the highest

concentrations tested, although activity did increase with addition of Dequest 2010. The activity of Neutrase® was severely impacted by three of the five sequestrants tested, which would be expected as it is a metalloprotease and requires zinc for activity. Surprisingly, the activity of Pro_10 was not significantly decreased by addition of sequestrants, and in most cases activity increased at the highest concentration tested, despite the fact that this is a metalloprotease. The binding strength of the sequestrant did not seem to correlate with the observed increase in protease activity either (Anastas et al., 2014), and activity even increased with addition of EDTA, a standard metalloprotease inhibitor. The activity was only significantly decreased upon addition of 100 mM EDTA, which is over 100 times the standard working concentration when used as a metalloprotease inhibitor. This suggests that the catalytic metal ion is buried deeply within a hydrophobic pocket of the enzyme, such that it is inaccessible to chelating agents.

Another possibility is that Pro_10 is not actually a metalloprotease. To test this, activity was measured following addition of a range of protease inhibitors that are used to target various classes of protease. However, Pro_10 was not significantly inhibited by any of these. We then mutated the predicted HEXXH active site residues, which includes two of the metal binding histidine residues and the catalytically active glutamate. It was found from a lack of clearing on skimmed milk agar plates that these mutants did not retain protease activity, suggesting that Pro_10 is in fact a metalloprotease. However, it could also be that the variants were not secreted, or they retained activity but it was not sufficient to produce visible zones of clearing. A more robust method would be to purify these variants and compare activity *in vitro*. Metalloproteomic approaches should also be investigated to confirm metal ion binding to Pro_10, for example by furnace atomic absorption spectrometry (FAAS), inductively coupled plasma mass spectrometry (ICP-MS) or X-ray absorption spectroscopy (Marshall et al., 2004; Rigueira et al., 2016; Shi et al., 2011).

Finally, stain removal of a range of proteinaceous stains was compared between Pro_10 and the gold standard detergent protease, Carnival Evity™ 16 L. We found that Pro_10 was active against all stains tested and had greater stain removal of blood stains than Carnival Evity™ 16 L, and marginally better removal of gravy stains. This was tested in a standard buffer as formulation stability has not been

optimised, although this is a promising initial result. Any future work should focus on testing stain removal in formulation, following optimisation of formulation stability, as this is more representative of a real-world application.

In summary, this study highlights the significance of addressing energy efficiency in detergent processes and underscores the potential advantages of discovering novel proteases adapted to lower temperatures, offering more energy-efficient laundry processes and cost savings for consumers while reducing environmental impacts. We streamlined the protease discovery process by combining bioinformatic approaches, host optimisation, and signal peptide library screening. Notably, our successful purification and characterisation of Pro_10, a member of family M57, revealed a multifaceted enzyme structure with distinct domains. The C-terminal domain, while inactive on its own, proved essential for full protease activity. Pro_10 exhibited a non-specific enzymatic profile, particularly effective against blood and gravy stains, demonstrating its potential for detergent applications. This is particularly promising as blood is a problematic stain that is difficult to remove. Additionally, the enzyme displayed remarkable tolerance to sequestrants, suggesting its suitability in various detergent formulations. Lastly, comparative analyses with commercial proteases Neutrase® and Carnival Evely™ 16 L revealed the potential advantages of Pro_10 in terms of halophilicity and tolerance of water hardness. This promising initial result invites further exploration of Pro_10's utility in real-world detergent applications.

Appendix A

Growth media used in this study.

LB

Component	Concentration
Tryptone	10 g/L
NaCl	10 g/L
Yeast extract	5 g/L
pH	7.2 – 7.5

LBM

Component	Concentration
Tryptone	10 g/L
NaCl	10 g/L
Yeast extract	5 g/L
Mannitol	0.5% (v/v)
MgCl ₂	5 mM
pH	7.2 – 7.5

LB agar

Component	Concentration
Tryptone	10 g/L
NaCl	10 g/L
Yeast extract	5 g/L
Agar	15 g/L
pH	7.2 – 7.5

SOC

Component	Concentration
Tryptone	20 g/L
NaCl	0.5 g/L
Yeast extract	5 g/L
pH	7

SR

Component	Concentration
Tryptone	15 g/L
K ₂ HPO ₄	3 g/L
Yeast extract	25 g/L
pH	7.2

SRX

Component	Concentration
Tryptone	15 g/L
K ₂ HPO ₄	3 g/L
Yeast extract	25 g/L
Xylose	40 g/L
pH	7.2

PA agar

Component	Concentration
Peptone	10 g/L
NaCl	20 g/L
Yeast extract	5 g/L
Agar	15 g/L
Skimmed milk powder	10 g/L
pH	7.2

Appendix B

Crucial oligonucleotides used in this study.

Designation	Sequence
oMAP0152	AAGAAAACGTTGGATAGAGC
oMAP0217	TACAACATTGGTGACGCTGCC
aad9_F1	TAGATAATGGCTCTTCTCACATCA
aad9_R1	GATTGATGTGAGAAGAGCCATTAT
oMAP0155	TTGACAGAGAACAGAGAAGC
oMAP0843	CAAGCTGACGATGGATCGGG
oMAP0844	TTTGTCACTCTCCCTCCTTTAGTG
oMAP0845	TAAAGGAGGGAGATGACAAAGGGAACAAGGGTGACGAACG
oMAP0846	AGATGTGCCGTCTCAAGTCC
dDNA_Bs_aad9_R1	CGTTCAGGCAGTACGCCTCTTTT
dDNA_Bs_aad9_F2	AGAGGCGTACTGCCTGAACGTAGTAAAAAGAAGCAGGTTTCCTC C
oMAP0849	AGGAAGGATTTACGTTATCC
oMAP0851	AAATTCACGATATGGGACCC
QC_Bs_aad9_F1	CACCATCCTTCGGCAAATCC
oMAP0220	ATCGGCATTTTTTTCTGCCTGC
amyE_dDNA_F1	TTGTATTTATTCCAATCCTGGC
amyE_dDNA_R1	TGGACATGATTACAATGTAATCTTGACACTCCTTATTTGATTTTTT G
xylA_dDNA_F	TCAAATAAGGAGTGTCAAGATTACATTGTAATCATGTCCAGAAAA TG
xylA_dDNA_R	GCTATTGCGATTTTCTTCATGTGATTTCCCCCTTAAAAATAAATTC
prsA_dDNA_F	ATTTTAAAGGGGAAATCACATGAAGAAAATCGCAATAGCAG
prsA_dDNA_R	ATGGTTTCTTTCGGTAAGTCTTATTTAGAATTGCTTGAAGATGAA G
amyE_dDNA_F2	CTTCAAGCAATTCTAAATAAGACTTACCGAAAGAAACCATC
amyE_dDNA_R2	ATATGTGCTGACAGTTAATATGC
amyE_QC_F	CTTCATTTTCTCATCTCTTCAGG
amyE_QC_R	TCAGTGATACCTGCGATC
prsA_seq_F1	CGGCGTTCTGTTTCTGCTTC
prsA_seq_F2	CGATCCCTATTCCTGTGCTG
prsA_seq_R1	TTTCTTGCACTAGCTGTGTCAG
prsA_seq_R2	CAACACTTCACAACTTTTGC

Appendix C

The genotypes of two minimal *Bacillus subtilis* strains used in this study are shown below.

IIIG-Bs-27-24

B. subtilis 168 Δ phy, Δ ytN, Δ ytM, Δ ytL, Δ ytK, Δ ytJ, Δ ytI, Δ ytH, Δ ytG, Δ ytF, Δ ytE, Δ ytD, Δ ytC, Δ ytB, Δ sspC, Δ ysX, Δ ysW, Δ ysV, Δ yojW, Δ ysU, Δ ysT, Δ ysS, Δ ysR, Δ nrdF, Δ ysQ, Δ nrdEB, Δ nrdI, Δ ysL, Δ ysK, Δ ysJ, Δ ysl, Δ ysH, Δ ysG, Δ ysF, Δ ysE, Δ ysD, Δ ysC, Δ ysB, Δ ysA, Δ yorZ, Δ yorY, Δ yorX, Δ yorW, Δ yorV, Δ mtbP, Δ yorT, Δ yorS, Δ yorR, Δ yorQ, Δ yorP, Δ yorO, Δ yorN, Δ yorM, Δ yorL, Δ yorK, Δ yorJ, Δ yorI, Δ yorH, Δ yorG, Δ yorF, Δ yorE, Δ yorD, Δ yorC, Δ yorB, Δ yorA, Δ yoqZ, Δ yoqY, Δ yoqX, Δ yoqW, Δ ligB, Δ yoqU, Δ yoqT, Δ yoqS, Δ yoqR, Δ yoqP, Δ yoqO, Δ yoqN, Δ yoqM, Δ yoqL, Δ yoqK, Δ yoqJ, Δ yoqI, Δ yoqH, Δ yoqG, Δ yoqF, Δ yoqE, Δ yoqD, Δ yoqC, Δ yoqB, Δ yoqA, Δ yopZ, Δ yopY, Δ yopX, Δ yopW, Δ yopV, Δ yopU, Δ yopT, Δ yopS, Δ yopR, Δ yopQ, Δ yopP, Δ yopO, Δ yopN, Δ yopM, Δ yopL, Δ yopK, Δ yopJ, Δ yopI, Δ yopH, Δ yopG, Δ yopF, Δ yopE, Δ yoyH, Δ yoyI, Δ yopD, Δ yopC, Δ yopB, Δ yopA, Δ yonX, Δ yonV, Δ yonU, Δ yoyJ, Δ yonT, Δ yonS, Δ yonR, Δ yonP, Δ yonO, Δ yonN, Δ yonK, Δ yonJ, Δ yonI, Δ yonH, Δ yonG, Δ yonF, Δ yonE, Δ yonD, Δ yonC, Δ yonB, Δ yonA, Δ yomZ, Δ yomY, Δ yomX, Δ yomW, Δ yomV, Δ yomU, Δ yoyA, Δ yomT, Δ yomS, Δ yomR, Δ yomQ, Δ yomP, Δ yomO, Δ yomN, Δ yomM, Δ yozP, Δ yomL, Δ yoyB, Δ yomK, Δ yomJ, Δ yomI, Δ yomH, Δ yomG, Δ yomF, Δ yomE, Δ yomD, Δ blyA, Δ bhlA, Δ bhlB, Δ bdbB, Δ yolJ, Δ bdbA, Δ sunT, Δ sunA, Δ sunI, Δ uvrX, Δ yolD, Δ yolC, Δ yolB, Δ yolA, Δ yokL, Δ yokK, Δ yokJ, Δ yokI, Δ yokH, Δ yokG, Δ yokF, Δ yokE, Δ yokD, Δ yokC, Δ yokB, Δ yokAarsB, Δ yqcK, Δ arsR, Δ yqcl, Δ rapE, Δ phrE, Δ yqzI, Δ yqcG, Δ yqcF, Δ yqxJ, Δ yqxI, Δ cwlA, Δ yqxH, Δ yqxG, Δ yqcE, Δ yqcD, Δ yqcC, Δ yqcB, Δ yqcA, Δ yqbT, Δ yqbS, Δ yqbR, Δ yqbQ, Δ yqbP, Δ yqbO, Δ yqbN, Δ yqbM, Δ yqbK, Δ yqzN, Δ yqbJ, Δ yqbl, Δ yqbH, Δ yqbG, Δ yqbF, Δ yqbE, Δ yqbD, Δ yqbC, Δ yqbB, Δ yqbA, Δ yqaT, Δ yqaS, Δ yqaR, Δ yqaQ, Δ yqaP, Δ yqaO, Δ yqaN, Δ yqzO, Δ yqaM, Δ yqaL, Δ yqaK, Δ yqaJ, Δ yqal, Δ yqaH, Δ yqaG, Δ yqdA, Δ yqaF, Δ yqaE, Δ yqaD, Δ yqaCyjoAB, Δ rapA, Δ phrA, Δ yjpA, Δ xlyB, Δ yjqABC, Δ xkdA, Δ xre, Δ yjzJ, Δ xkdB, Δ xkdC, Δ yzkK, Δ ykdd, Δ xtrA, Δ xpf, Δ xtmA, Δ xtmB, Δ xkdE, Δ xkdF, Δ xkdG, Δ yzkL, Δ ykdh, Δ ykdl, Δ ykdd, Δ yzkM, Δ xkdK, Δ ykdm, Δ ykdn, Δ xkzB, Δ ykdo, Δ ykdp, Δ ykdq, Δ ykdr, Δ ykds, Δ ykdt, Δ ykdu, Δ xkzA, Δ xkdV, Δ xkdW, Δ xkdX, Δ xepA, Δ xhlAB, Δ xlyAalkA, Δ adaA, Δ adaB, Δ ndhF, Δ ybcC, Δ ybcF, Δ ybcH, Δ ybcI, Δ ybzH, Δ ybcL, Δ ybcM, Δ skfA, Δ skfB, Δ skfC, Δ skfE, Δ skfF, Δ skfG, Δ skfHpkasA, Δ B, Δ C, Δ D, Δ E, Δ acpK, Δ pksG, Δ pksH, Δ pksI, Δ pksJ, Δ pksK, Δ pksL, Δ pksM, Δ pksR, Δ pksSDproF3, Δ ydiM, Δ ydzuV, Δ ydiO, Δ ydiP, Δ ydzuW, Δ ydiR, Δ ydiS, Δ ydiJ, Δ ydiB, Δ ydiCmanP, Δ manAynxB, Δ ynzFG, Δ ynaB, Δ ynaC, Δ ynaD, Δ ynaE, Δ ynaF, Δ ynaG, Δ ynzI, Δ ynalyozV, Δ yobD, Δ yozH, Δ yozI, Δ yobE, Δ yobF, Δ yozW, Δ yozX, Δ yozJ, Δ yozY, Δ yozZ, Δ rapK, Δ phrK, Δ yobH, Δ yozK, Δ yozL, Δ yozM, Δ yobl, Δ yoyA, Δ yobJ, Δ yobK, Δ yobL, Δ yobM, Δ yobN, Δ yobOCmRsboA, Δ sboX, Δ albA, Δ albB, Δ albC, Δ albD, Δ albE, Δ albF, Δ albGppsA, Δ ppsB, Δ ppsC, Δ ppsD, Δ ppsEbacA, Δ bacB, Δ bacC, Δ bacD, Δ bacE, Δ bacF, Δ ywfHyrkK, Δ yrkJ, Δ yrkl, Δ yrkH, Δ yrkF, Δ yrkE, Δ yrkD, Δ yrzM, Δ yrzN, Δ yrkC, Δ yrkB, Δ bltR, Δ blt, Δ bltD, Δ yrkA, Δ yrzO, Δ yrdR, Δ yrdQ, Δ yrdP, Δ czcD, Δ yrdN, Δ gltR, Δ yrdK, Δ brnQ, Δ azlB, Δ azlC, Δ azlD, Δ yrdF, Δ cypA, Δ yrdD, Δ yrdC, Δ yrdB, Δ yrdA, Δ aadK, Δ yrpB, Δ yrpC, Δ yrpD, Δ yrpEypA, Δ ytpB, Δ ytoAsdpA, Δ sdpB, Δ sdpC, Δ sdpI, Δ sdpRbpr, Δ spolIGA, Δ sigE, Δ sigGydgG, Δ ydgH, Δ ydgI, Δ ydgJ, Δ ydgK, Δ ydhB, Δ ydhC, Δ ydhD, Δ ydhE, Δ ydhF, Δ phoB, Δ ydhG, Δ ydhH, Δ ydhI, Δ ydhJ, Δ ydhK, Δ pbuE, Δ gmuB, Δ gmuA, Δ gmuC, Δ gmuD, Δ gmuR, Δ gmuE, Δ gmuF, Δ gmuG, Δ ydhUspolIAA, Δ spolIAB, Δ sigF, Δ spoVAA, Δ spoVAB, Δ spoVAC, Δ spoVAD, Δ spoVAEB, Δ spoVAEA, Δ spoVAFepsO, Δ epsN, Δ epsM, Δ epsL, Δ epsK, Δ epsJ, Δ epsI, Δ epsH, Δ epsG, Δ epsF, Δ epsE, Δ epsD, Δ epsC, Δ epsB, Δ epsA, Δ slrRyqxM, Δ sipW, Δ tasAglcP, Δ ntdC, Δ ntdB, Δ ntdA, Δ yhjMlytD, Δ yvyI, Δ gerBA, Δ gerBB, Δ gerBC, Δ ywtG, Δ ywtF, Δ ywtEyqeF, Δ cwlH, Δ yqeD, Δ yqeC, Δ yqeB, Δ spolVCB, Δ spolVCA, Δ arsC, Δ arsB, Δ yqcK, Δ arsR, Δ yqcl, Δ rapE, Δ phrE, Δ yqzI, Δ yqcG, Δ yqcF, Δ yqxJ, Δ yqxI, Δ cwlA, Δ yqxH, Δ yqxG, Δ yqcE, Δ yqcD, Δ yqcC, Δ yqcB, Δ yqcA, Δ yqbT, Δ yqbS, Δ yqbR, Δ yqbQ, Δ yqbP,

ΔyqbO, ΔyqbN, ΔyqdB, ΔyqbM, ΔyqbK, ΔyqzN, ΔyqbJ, Δyqbl, ΔyqbH, ΔyqbG, ΔyqbF, ΔyqbE, ΔyqbD, ΔyqbC, ΔyqbB, ΔyqbA, ΔyqaT, ΔyqaS, ΔyqaR, ΔyqaQ, ΔyqaP, ΔyqaO, ΔyqzO, ΔyqaM, ΔyqaL, ΔyqaK, ΔyqaJ, Δyqal, ΔyqaH, ΔyqaG, ΔyqdA, ΔyqaF, ΔyqaE, ΔyqaD, ΔyqaC, ΔyqaB, ΔspolIIC, ΔyrkSfliE, ΔfliF, ΔfliG, ΔfliH, ΔfliI, ΔfliJ, ΔylxH, ΔfliK, ΔflgD, ΔflgE, ΔylzI, ΔfliL, ΔfliM, ΔfliY, ΔcheY, ΔfliZ, ΔfliP, ΔfliQ, ΔfliR, ΔflhB, ΔflhA, ΔflhF, ΔylxH, ΔcheB, ΔcheA, ΔcheW, ΔcheC, ΔcheDspolVBgerPF, ΔgerPE, ΔgerPD, ΔgerPC, ΔgerPB, ΔgerPA, ΔyisI, ΔyisJ, ΔyisK, ΔyisL, ΔwprAyyvzG, ΔfliT, ΔfliS, ΔfliD, ΔyvyC, Δhag, ΔcsrA, ΔfliW, ΔyviE, ΔflgL, ΔfliKcotO, ΔcotZ, ΔcotY, ΔcotX, ΔcotW, ΔcotV, ΔyjcA, ΔyjkK, ΔyjcZ, ΔspoVIFyclG, ΔgerKA, ΔgerKC, ΔgerKB, ΔyclH, ΔyclI, ΔyclJ, ΔyclK, ΔrapC, ΔphrC, ΔyczM, ΔyczN, ΔyclM, ΔyclN, ΔyclO, ΔyclP, ΔyclQ, ΔycnB, ΔycnC, ΔycnD, ΔycnE, ΔyczGyybP, ΔyybO, ΔyyzI, ΔyyzJ, ΔyyzK, ΔyyzL, ΔyybN, ΔyybM, ΔyybL, ΔyybK, ΔyybJ, ΔyybI, ΔyybH, ΔyybG, ΔyybF, ΔyybE, ΔyybD, ΔyybC, ΔyybB, ΔyybA, ΔyyaT, ΔyyaS, ΔyyaR, ΔyyaQ, ΔyyaP, ΔtetB, ΔtetL, ΔyyaO, ΔyyaN, ΔyyaM, ΔyyaL, ΔyyaK, ΔyyaJytM, ΔyitO, ΔyitP, ΔyizB, ΔyitQ, ΔyitR, ΔnprB, ΔyitSvpryxxB, ΔyxeR, ΔyxeQ, ΔyxeP, ΔyxeO, ΔyxeN, ΔyxeM, ΔyxeL, ΔyxeK, ΔyxeJ, ΔyxeI, ΔyxeH, ΔyxeG, ΔyxeF, ΔyxeE, ΔyxeD, ΔyxeC, ΔyxeB, ΔyxeA, ΔyxdM, ΔyxdL, ΔyxdK, ΔyxdJ, ΔiolJ, ΔiolI, ΔiolH, ΔiolG, ΔiolF, ΔiolE, ΔiolD, ΔiolC, ΔiolB, ΔmmsA, ΔiolR, ΔiolS, ΔyxcE, ΔyxcDyktA, ΔyktB, ΔyktI, ΔsuhB, ΔyktC, ΔyktD, ΔnprEywB, ΔywbA, ΔeprycA::PmtIA-comK-comS

IIG-Bs27-31

IIG-Bs27-24 ΔyydD, ΔyydC, ΔyydB, ΔyydA, ΔyyzF, ΔyycS, ΔyycR, ΔyycG, ΔyycQ, ΔyycP, ΔyycO, ΔyycN, ΔrapG, ΔphrGmpr, ΔybfJkinD, ΔmhqR, ΔmotA, ΔmotBymzD, ΔymzC, ΔpksA, ΔpksB, ΔpksC, ΔpksD, ΔpksE, ΔacpK, ΔpksF, ΔpksG, ΔpksH, ΔpksI, ΔpksJ, ΔpksL, ΔpksM, ΔpksN, ΔpksR, ΔpksS, ΔymzB, ΔymaE, ΔaprX, ΔymzE, ΔymaC, ΔymaD, ΔebrB, ΔebrA, ΔymaGyulE, ΔyulD, ΔyulC, ΔyulB, ΔyuxG, ΔtlpB, ΔmcpA, ΔtlpA, ΔmcpB, Δtgi, ΔyuzH, ΔyuzU, ΔyuzT, ΔyuzS, ΔyuzP, ΔyuzI, ΔmstX, ΔyuzO, ΔyuzN, ΔyuzMyefB, ΔyefC, ΔyeeA, ΔyeeB, ΔyeeC, ΔyeeD, ΔyezA, ΔyezG, ΔyeeF, ΔyeeG, ΔrapH, ΔphrH, Δyeel, ΔyeeK, ΔyezE, ΔyesE, ΔyesF, ΔcotJA, ΔcotJB, ΔcotJC, ΔyesJ, ΔyesK, ΔyesL, ΔyesM, ΔyesN, ΔyesO, ΔyesP, ΔyesQ, ΔyesR, ΔyesS, ΔrhgT, ΔyesU, ΔyesV, ΔyesW, ΔyesX, ΔyesY, ΔyesZ, ΔyetA, ΔlplA, ΔlplB, ΔlplC, ΔlplD, ΔyetF, ΔyetG, ΔyetH, ΔyetI, ΔyezD, ΔyetJ, ΔyetK, ΔyetL, ΔyetM, ΔyetN, ΔyetOyhfM, ΔyhfN, ΔaprE, ΔyhfO, ΔyhfP, ΔyhfQ

Appendix D

The phylogenies and accession numbers of candidate proteases are shown below.

Domain	Class	Species	Accession number
Prokaryotic	Actinomycetia	<i>Kitasatospora</i> sp. NRRL B-11411	WP_030462359.1
		<i>Renibacterium salmoninarum</i>	P55111
	Bacilli	<i>Halobacillus dabanensis</i>	WP_075037004.1
		<i>Halobacillus ihumii</i>	WP_163531214.1
		<i>Halobacillus massiliensis</i>	WP_082233925.1
			WP_082234881.1
		<i>Halobacillus salinus</i>	WP_135326494.1
		<i>Lactobacillus kisonensis</i>	WP_008858092.1
		<i>Staphylococcus epidermidis</i>	P0C0Q3
	Clostridia	<i>Haloimpatiens massiliensis</i>	WP_102400416.1
			WP_180963949.1
	Deltaproteobacteria	<i>Desulfuromonas</i> sp.	PLX79621.1
		<i>Myxococcus virescens</i>	WP_090493785.1
		<i>Plesiocystis pacifica</i>	WP_006970331.1
		<i>Sorangium cellulosum</i>	WP_012239334.1
	Flavobacteriia	<i>Flavobacterium psychrophilum</i> JIP02/86	CAL42347
		<i>Flavobacterium psychrophilum</i> JIP02/87	CAL42348
		<i>Owenweeksia hongkongensis</i>	WP_014202324.1
	Gammaproteobacteria	<i>Aeromonas caviae</i>	Q9LCJ5
		<i>Aeromonas hydrophila</i>	O05485
			Q9RG23
		<i>Kangiella koreensis</i>	WP_012800440.1
		<i>Pseudomonas marginalis</i>	A0A0C5CJR8
		<i>Shewanella</i> sp. cp20	WP_041509222.1
		<i>Shewanella woodyi</i>	WP_012326503.1
		<i>Vibrio vulnificus</i> CMCP6	Q8D4Y9
Eukaryotic	Agaricomycetes	<i>Agrocybe aegerita</i>	CAA7262796.1
		<i>Schizophyllum commune</i> H4-8	XP_003029310.1
		<i>Trametes versicolor</i> FP-101664 SS1	XP_008032702.1
	Choanoflagellatea	<i>Salpingoeca</i> sp. ATCC 50818	XP_004997487.1
	Dothideomycetes	<i>Pseudogymnoascus pannorum</i>	KFY30966.1
		<i>Sphaerulina musiva</i> SO2202	XP_016758957.1
		<i>Zymoseptoria tritici</i> ST99CH_1A5	SMY28112.1
	Phaeophyceae	<i>Ectocarpus siliculosus</i>	CBN76176.1
	Sordariomycetes	<i>Cordyceps bassiana</i>	XP_008602715.1
		<i>Metarhizium anisopliae</i>	KFG79808.1
		<i>Metarhizium brittlebankisoides</i>	KID94564.1

Appendix E

The Durham salt screen (Molecular Diagnostics, UK) conditions used in **Error! Reference source not found.** are listed below.

Well	Concentration (M)	Reagent1
A1		Water
A2		Water
A3	8	Urea
A4	6	Guanidine hydrochloride
A5	2	Guanidine hydrochloride
A6	1.6	Guanidine hydrochloride
A7	1.2	Guanidine hydrochloride
A8	0.8	Guanidine hydrochloride
A9	0.4	Guanidine hydrochloride
A10	0.01	Guanidine hydrochloride
A11	1	Sodium citrate
A12	0.4	Sodium citrate
B1	3	Sodium malonate
B2	2	Sodium malonate
B3	1.6	Sodium malonate
B4	1.2	Sodium malonate
B5	0.8	Sodium malonate
B6	0.4	Sodium malonate
B7	3	Ammonium sulfate
B8	2	Ammonium sulfate
B9	1.6	Ammonium sulfate
B10	1.2	Ammonium sulfate
B11	0.8	Ammonium sulfate
B12	0.4	Ammonium sulfate
C1	3	Sodium chloride
C2	2	Sodium chloride
C3	1.6	Sodium chloride
C4	1.2	Sodium chloride
C5	0.8	Sodium chloride

C6	0.4	Sodium chloride
C7	3	Ammonium chloride
C8	2	Ammonium chloride
C9	1.6	Ammonium chloride
C10	1.2	Ammonium chloride
C11	0.8	Ammonium chloride
C12	0.4	Ammonium chloride
D1	2	Magnesium sulfate
D2	1.6	Magnesium sulfate
D3	1.2	Magnesium sulfate
D4	0.8	Magnesium sulfate
D5	0.4	Magnesium sulfate
D6	2	Sodium sulfate
D7	1.6	Sodium sulfate
D8	1.2	Sodium sulfate
D9	0.8	Sodium sulfate
D10	0.4	Sodium sulfate
D11	1	Potassium chloride
D12	0.4	Potassium chloride
E1	1	Lithium chloride
E2	0.4	Lithium chloride
E3	1	Rubidium chloride
E4	0.4	Rubidium chloride
E5	1	Cesium chloride
E6	0.4	Cesium chloride
E7	0.8	Sodium fluoride
E8	0.2	Sodium fluoride
E9	3	Sodium bromide
E10	0.8	Sodium bromide
E11	0.2	Sodium bromide
E12	0.8	Sodium iodide
F1	0.2	Sodium iodide

F2	0.8	Magnesium chloride
F3	0.01	Magnesium chloride
F4	0.01	Calcium chloride
F5	0.01	Strontium chloride
F6	0.002	Zinc chloride
F7	0.0002	Zinc chloride
F8	0.002	Nickel chloride
F9	0.0002	Nickel chloride
F10	0.01	Manganese chloride
F11	0.001	Manganese chloride
F12	0.002	Cobalt chloride
G1	0.0002	Cobalt chloride
G2	0.002	Copper(II) sulfate
G3	0.0002	Copper(II) sulfate
G4	0.002	Cadmium sulfate
G5	0.01	EDTA pH 7.5
G6	0.01	EGTA pH 7.5
G7	0.004	I3C
G8	0.004	Lanthanum (III) nitrate hexahydrate
G9	0.004	Praseodymium (III) chloride hydrate
G10	0.004	Neodymium (III) chloride hexahydrate
G11	0.004	Samarium (III) chloride hexahydrate
G12	0.004	Europium (III) chloride hexahydrate
H1	0.004	Gadolinium (III) chloride hexahydrate
H2	0.004	Dysprosium (III) chloride hexahydrate
H3	0.004	Holmium (III) chloride hexahydrate
H4	0.004	Ytterbium (III) chloride hexahydrate

H5	0.004	Lutetium (III) chloride hexahydrate
H6	0.01	Sodium phosphate dibasic
H7	0.01	Sodium orthovanadate
H8	0.01	Sodium tungstate
H9	0.01	Sodium molybdate
H10	0.01	DTT
H11	0.01	TCEP pH 7.0
H12	0.01	β-Mercaptoethanol

Bibliography

- Abcam. (2021). *Gelatin zymography protocol*. <https://www.abcam.com/protocols/gelatin-zymography-protocol>
- AcalPred. (2013). <http://lin-group.cn/server/AcalPred>
- Adekoya, O. A., Helland, R., Willassen, N. P., & Sylte, I. (2006). Comparative sequence and structure analysis reveal features of cold adaptation of an enzyme in the thermolysin family. *Proteins: Structure, Function and Genetics*, 62(2), 435–449. <https://doi.org/10.1002/prot.20773>
- Adekoya, O. A., & Sylte, I. (2009). The thermolysin family (M4) of enzymes: Therapeutic and biotechnological potential. In *Chemical Biology and Drug Design* (Vol. 73, Issue 1, pp. 7–16). John Wiley & Sons, Ltd. <https://doi.org/10.1111/j.1747-0285.2008.00757.x>
- Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. *Nature Communications*, 9(1), 1911. <https://doi.org/10.1038/s41467-018-04252-2>
- Aguilar, R., Suárez, S., Rg Stü, J., & Maarten Van Dijn, J. (2020). Less Is More: Toward a Genome-Reduced *Bacillus* Cell Factory for “Difficult Proteins.” *ACS Synthetic Biology*, 13, 27. <https://doi.org/10.1021/acssynbio.8b00342>
- Ahmed, K., Chohnan, S., Ohashi, H., Hirata, T., Masaki, T., & Sakiyama, F. (2003). Purification, bacteriolytic activity, and specificity of β -lytic protease from *Lysobacter* sp. IB-9374. *Journal of Bioscience and Bioengineering*, 95(1), 27–34. [https://doi.org/10.1016/S1389-1723\(03\)80144-5](https://doi.org/10.1016/S1389-1723(03)80144-5)
- Altenbuchner, J. (2016). Editing of the *Bacillus subtilis* Genome by the CRISPR-Cas9 System. *Applied and Environmental Microbiology*, 82(17), 5421–5427. <https://doi.org/10.1128/AEM.01453-16>
- Altschul, S. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25(17), 3389–3402. <https://doi.org/10.1093/nar/25.17.3389>
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic Local Alignment Search Tool. *J. Mol. Biol* (Vol. 215).
- Alvarado, J. J., Nemkal, A., Sauder, J. M., Russell, M., Akiyoshi, D. E., Shi, W., Almo, S. C., & Weiss, L. M. (2009). Structure of a microsporidian methionine aminopeptidase type 2 complexed with fumagillin and TNP-470. *Molecular and Biochemical Parasitology*, 168(2), 158–167. <https://doi.org/10.1016/J.MOLBIOPARA.2009.07.008>

- Anagnostopoulos, C., & Spizizen, J. (1961). Requirements For Transformation in *Bacillus subtilis*. *Journal of Bacteriology*, 81(5), 741–746. <https://doi.org/10.1128/jb.81.5.741-746.1961>
- Anastas, P. T., Boethling, R., & Voutchkova-Kostal, A. (2014). Green Processes, Volume 9: Designing Safer Chemicals. Wiley. <https://books.google.co.uk/books?id=oGFuAwAAQBAJ>
- Anders, A., Gilbert, S., Garten, W., Postina, R., & Fahrenholz, F. (2001). Regulation of the α -secretase ADAM10 by its prodomain and proprotein convertases. *The FASEB Journal*, 15(10), 1837–1839. <https://doi.org/10.1096/fj.01-0007fje>
- Ara, K., Ozaki, K., Nakamura, K., Yamane, K., Sekiguchi, J., & Ogasawara, N. (2007). *Bacillus* minimum genome factory: effective utilization of microbial genome information. *Biotechnology and Applied Biochemistry*, 46(3), 169. <https://doi.org/10.1042/BA20060111>
- Asemave, K. (2018). Greener Chelators for Recovery of Metals and Other Applications. *Organic & Medicinal Chemistry International Journal*, 6(4). <https://doi.org/10.19080/OMCIJ.2018.06.555694>
- Auld, D. S. (2013). Catalytic Mechanisms for Metallopeptidases. *Handbook of Proteolytic Enzymes* (pp. 370–396). Elsevier. <https://doi.org/10.1016/B978-0-12-382219-2.00078-8>
- Austin, B., & Austin, D. A. (2016). *Aeromonadaceae* Representative (*Aeromonas salmonicida*) BT - Bacterial Fish Pathogens: Disease of Farmed and Wild Fish (B. Austin & D. A. Austin, Eds.; pp. 215–321). Springer International Publishing. https://doi.org/10.1007/978-3-319-32674-0_5
- Awad, M., & Khanna, R. (2015). Efficient Learning Machines. Apress. <https://doi.org/10.1007/978-1-4302-5990-9>
- Bagautdinov, B., Matsuura, Y., Yamamoto, H., Sawano, M., Ogasahara, K., Takehira, M., Kunishima, N., Katoh, E., & Yutani, K. (2015). Thermodynamic analysis of unusually thermostable CutA1 protein from human brain and its protease susceptibility. *Journal of Biochemistry*, 157(3), 169–176. <https://doi.org/10.1093/jb/mvu062>
- Bajpai, D., & Tyagi, V. K. (2007). Laundry Detergents: An Overview. *Journal of Oleo Science*, 56(7), 327–340. <https://doi.org/10.5650/jos.56.327>
- Bannikov, G. A., Karelina, T. V., Collier, I. E., Marmer, B. L., & Goldberg, G. I. (2002). Substrate Binding of Gelatinase B Induces Its Enzymatic Activity in the Presence of Intact Propeptide. *Journal of Biological Chemistry*, 277(18), 16022–16027. <https://doi.org/10.1074/jbc.M110931200>

- Barbieri, G., Albertini, A. M., Ferrari, E., Sonenshein, A. L., & Belitsky, B. R. (2016). Interplay of CodY and ScoC in the regulation of major extracellular protease genes of *Bacillus subtilis*. *Journal of Bacteriology*, 198(6), 907–920. <https://doi.org/10.1128/JB.00894-15>
- Bayer, T. S., Widmaier, D. M., Temme, K., Mirsky, E. A., Santi, D. V., & Voigt, C. A. (2009). Synthesis of Methyl Halides from Biomass Using Engineered Microbes. *Journal of the American Chemical Society*, 131(18), 6508–6515. <https://doi.org/10.1021/ja809461u>
- Bazan, J. F., Weaver, L. H., Roderick, S. L., Huber, R., & Matthews, B. W. (1994). Sequence and structure comparison suggest that methionine aminopeptidase, prolidase, aminopeptidase P, and creatinase share a common fold. *Proceedings of the National Academy of Sciences*, 91(7), 2473–2477. <https://doi.org/10.1073/pnas.91.7.2473>
- Beck, C. M., & Strobel, R. J. J. (1988). Low-temperature active alkaline protease from *Paecilomyces marquandii* and its preparation (Patent WO1988003948A1). WIPO.
- Becker, A. B., & Roth, R. A. (1992). An unusual active site identified in a family of zinc metalloendopeptidases. *Proceedings of the National Academy of Sciences*, 89(9), 3835–3839. <https://doi.org/10.1073/PNAS.89.9.3835>
- Bennallack, P. R., Burt, S. R., Heder, M. J., Robison, R. A., & Griffiths, J. S. (2014). Characterization of a Novel Plasmid-Borne Thiopeptide Gene Cluster in *Staphylococcus epidermidis* Strain 115. *Journal of Bacteriology*, 196(24), 4344–4350. <https://doi.org/10.1128/JB.02243-14>
- Berini, F., Casciello, C., Marcone, G. L., & Marinelli, F. (2017). Metagenomics: Novel enzymes from non-culturable microbes. In *FEMS Microbiology Letters* (Vol. 364, Issue 21). Oxford University Press. <https://doi.org/10.1093/femsle/fnx211>
- Berka, R. M., Hahn, J., Albano, M., Draskovic, I., Persuh, M., Cui, X., Sloma, A., Widner, W., & Dubnau, D. (2002). Microarray analysis of the *Bacillus subtilis* K-state: genome-wide expression changes dependent on ComK. *Molecular Microbiology*, 43(5), 1331–1345. <https://doi.org/10.1046/j.1365-2958.2002.02833.x>
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E. The Protein Data Bank (2000) *Nucleic Acids Research* 28: 235-242 <https://doi.org/10.1093/nar/28.1.235>. <https://www.rcsb.org/alignment>
- Białkowska, A., Gromek, E., Florczak, T., Krysiak, J., Szulczewska, K., & Turkiewicz, M. (2016). Extremophilic Proteases: Developments of Their Special Functions, Potential Resources and Biotechnological Applications (pp. 399–444). *Springer*. https://doi.org/10.1007/978-3-319-13521-2_14

- Bird, T. H., Grimsley, J. K., Hoch, J. A., & Spiegelman, G. B. (1993). Phosphorylation of Spo0A activates its stimulation of in vitro transcription from the *Bacillus subtilis* *spoIIG* operon. *Molecular Microbiology*, 9(4), 741–749. <https://doi.org/10.1111/j.1365-2958.1993.tb01734.x>
- Bode, W., Gomis-Rüth, F. X., Huber, R., Zwillig, R., & Stöcker, W. (1992). Structure of astacin and implications for activation of astacins and zinc-ligation of collagenases. *Nature* 1992 358:6382, 358(6382), 164–167. <https://doi.org/10.1038/358164a0>
- Bode, W., Gomis-Rüth, F. X., & Stöckler, W. (1993). Astacins, serralytins, snake venom and matrix metalloproteinases exhibit identical zinc-binding environments (HEXXHXXGXXH and Met-turn) and topologies and should be grouped into a common family, the “metzincins.” *FEBS Letters*, 331(1–2), 134–140. [https://doi.org/10.1016/0014-5793\(93\)80312-I](https://doi.org/10.1016/0014-5793(93)80312-I)
- Bolhuis, A., Matzen, A., Hyyryläinen, H.-L., Kontinen, V. P., Meima, R., Chapuis, J., Venema, G., Bron, S., Freudl, R., & van Dijk, J. M. (1999). Signal Peptide Peptidase- and ClpP-like Proteins of *Bacillus subtilis* Required for Efficient Translocation and Processing of Secretory Proteins. *Journal of Biological Chemistry*, 274(35), 24585–24592. <https://doi.org/10.1074/jbc.274.35.24585>
- Bolhuis, A., Venema, G., Quax, W. J., Bron, S., & Maarten Van Dijk, J. (1999). Functional Analysis of Paralogous Thiol-disulfide Oxidoreductases in *Bacillus subtilis*. *Journal of Biological Chemistry* <http://www.jbc.org/>
- Boon, L., Ugarte-Berzal, E., Vandooren, J., & Opdenakker, G. (2020). Protease propeptide structures, mechanisms of activation, and functions. *Critical Reviews in Biochemistry and Molecular Biology*, 55(2), 111–165. <https://doi.org/10.1080/10409238.2020.1742090>
- Bos, J. W., Frenken, L. G., Verrips, C. T., & Visser, C. (1998). Production of active *Pseudomonas glumae* lipase in homologous or heterologous hosts (Patent US6410262B1).
- Botelho, T. O., Guevara, T., Marrero, A., Arêde, P., Fluxa, V. S., Reymond, J. L., Oliveira, D. C., & Xavier Gomis-Rüth, F. (2011). Structural and functional analyses reveal that *Staphylococcus aureus* antibiotic resistance factor *hmra* is a zinc-dependent endopeptidase. *Journal of Biological Chemistry*, 286(29), 25697–25709. <https://doi.org/10.1074/jbc.M111.247437>
- Bourne, D. G., Jones, G. J., Blakeley, R. L., Jones, A., Negri, A. P., & Riddles, P. (1996). Enzymatic pathway for the bacterial degradation of the cyanobacterial cyclic peptide toxin microcystin LR. *Applied and Environmental Microbiology*, 62(11), 4086–4094. <https://doi.org/10.1128/aem.62.11.4086-4094.1996>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>

- Brini, E., Simmerling, C., & Dill, K. (2020). Protein storytelling through physics. *Science*, 370(6520). <https://doi.org/10.1126/science.aaz3041>
- Brockmeier, U., Caspers, M., Freudl, R., Jockwer, A., Noll, T., & Eggert, T. (2006). Systematic Screening of All Signal Peptides from *Bacillus subtilis*: A Powerful Strategy in Optimizing Heterologous Protein Secretion in Gram-positive Bacteria. *Journal of Molecular Biology*, 362(3), 393–402. <https://doi.org/10.1016/j.jmb.2006.07.034>
- Bruce, D., Cardew, E., Freitag-Pohl, S., & Pohl, E. (2019). How to Stabilize Protein: Stability Screens for Thermal Shift Assays and Nano Differential Scanning Fluorimetry in the Virus-X Project. *Journal of Visualized Experiments*, 144. <https://doi.org/10.3791/58666>
- Burbulys, D., Trach, K. A., & Hoch, J. A. (1991). Initiation of sporulation in *B. subtilis* is controlled by a multicomponent phosphorelay. *Cell*, 64(3), 545–552. [https://doi.org/10.1016/0092-8674\(91\)90238-T](https://doi.org/10.1016/0092-8674(91)90238-T)
- Burby, P. E., & Simmons, L. A. (2017). MutS2 Promotes Homologous Recombination in *Bacillus subtilis*. *Journal of Bacteriology*, 199(2). <https://doi.org/10.1128/JB.00682-16>
- Burstein, E. A., Vedenkina, N. S., & Ivkova, M. N. (1973). Fluorescence and the Location of Tryptophan Residues in Protein Molecules. *Photochemistry and Photobiology*, 18(4), 263–279. <https://doi.org/10.1111/j.1751-1097.1973.tb06422.x>
- Bussiere, D. E., Pratt, S. D., Katz, L., Severin, J. M., Holzman, T., & Park, C. H. (1998). The Structure of VanX Reveals a Novel Amino-Dipeptidase Involved in Mediating Transposon-Based Vancomycin Resistance. *Molecular Cell*, 2(1), 75–84. [https://doi.org/10.1016/S1097-2765\(00\)80115-X](https://doi.org/10.1016/S1097-2765(00)80115-X)
- Cabral-Pacheco, G. A., Garza-Veloz, I., Castruita-De la Rosa, C., Ramirez-Acuña, J. M., Perez-Romero, B. A., Guerrero-Rodriguez, J. F., Martinez-Avila, N., & Martinez-Fierro, M. L. (2020). The Roles of Matrix Metalloproteinases and Their Inhibitors in Human Diseases. *International Journal of Molecular Sciences*, 21(24), 9739. <https://doi.org/10.3390/ijms21249739>
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10(1), 421. <https://doi.org/10.1186/1471-2105-10-421>
- Carmichael, W. W., Jones, C. L. A., Mahmood, N. A., & Theiss, W. C. (2009). Algal toxins and water-based diseases. *Critical reviews in environmental control* 15(3), 275–313. <https://doi.org/10.1080/10643388509381734>

- Carnevale, V., Raugei, S., Micheletti, C., & Carloni, P. (2006). Convergent dynamics in the protease enzymatic superfamily. *Journal of the American Chemical Society*, 128(30), 9766–9772. <https://doi.org/10.1021/ja060896t>
- Carvalho, H. F., Roque, A. C. A., Iranzo, O., & Branco, R. J. F. (2015). Comparison of the Internal Dynamics of Metalloproteases Provides New Insights on Their Function and Evolution. *Plos One* 10(9). <https://doi.org/10.1371/journal.pone.0138118>
- Chakravorty, D., Khan, M. F., & Patra, S. (2017). Multifactorial level of extremostability of proteins: can they be exploited for protein engineering? *Extremophiles*, 21(3), 419–444. <https://doi.org/10.1007/s00792-016-0908-9>
- Cheggour, A., Fanuel, L., Duez, C., Joris, B., Bouillenne, F., Devreese, B., Van Driessche, G., Van Beeumen, J., Frère, J. M., & Goffin, C. (2000). The *dppA* gene of *Bacillus subtilis* encodes a new D-aminopeptidase. *Molecular Microbiology*, 38(3), 504–513. <https://doi.org/10.1046/J.1365-2958.2000.02117.X>
- Chen, J., Fu, G., Gai, Y., Zheng, P., Zhang, D., & Wen, J. (2015). Combinatorial Sec pathway analysis for improved heterologous protein secretion in *Bacillus subtilis*: Identification of bottlenecks by systematic gene overexpression. *Microbial Cell Factories*, 14(1), 1–15. <https://doi.org/10.1186/S12934-015-0282-9/FIGURES/6>
- Chen, K. N., Huang, J. C., Chung, C. I., Kuo, W. Y., & Chen, M. J. (2011). Identification and characterization of H10 enzymes isolated from *Bacillus cereus* H10 with keratinolytic and proteolytic activities. *World Journal of Microbiology and Biotechnology*, 27(2), 349–358. <https://doi.org/10.1007/s11274-010-0465-9>
- Chen, X., Zhang, C., Cheng, J., Shi, X., Li, L., Zhang, Z., Bai, J., Chen, Y., Li, S., & Ying, H. (2013). Enhancement of adenosine production by *Bacillus subtilis* CGMCC 4484 through metabolic flux analysis and simplified feeding strategies. *Bioprocess and Biosystems Engineering*, 36(12), 1851–1859. <https://doi.org/10.1007/s00449-013-0959-6>
- Chen, Z., Wang, Y., Zhai, Y. F., Song, J., & Zhang, Z. (2013). ZincExplorer: an accurate hybrid method to improve the prediction of zinc-binding sites from protein sequences. *Molecular BioSystems*, 9(9), 2213–2222. <https://doi.org/10.1039/C3MB70100J>
- Cheng, K. C., Khoo, Z. S., Lo, N. W., Tan, W. J., & Chemmangattuvalappil, N. G. (2020). Design and performance optimisation of detergent product containing binary mixture of anionic-nonionic surfactants. *Heliyon*, 6(5), e03861. <https://doi.org/10.1016/j.heliyon.2020.e03861>

- Choo, K. H., & Ranganathan, S. (2008). Flanking signal and mature peptide residues influence signal peptide cleavage. *BMC Bioinformatics* 9(12). <https://doi.org/10.1186/1471-2105-9-S12-S15>
- Choo, K. H., Tan, T. W., & Ranganathan, S. (2009). A comprehensive assessment of N-terminal signal peptides prediction methods. *BMC Bioinformatics*, 10(S15), S2. <https://doi.org/10.1186/1471-2105-10-S15-S2>
- Daroit, D. J., Corrêa, A. P. F., & Brandelli, A. (2009). Keratinolytic potential of a novel *Bacillus* sp. P45 isolated from the Amazon basin fish *Piaractus mesopotamicus*. *International Biodeterioration and Biodegradation*, 63(3), 358–363. <https://doi.org/10.1016/j.ibiod.2008.11.008>
- 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. *Applied and Environmental Microbiology*, 76(19), 6370–6376. <https://doi.org/10.1128/AEM.01146-10>
- Demidyuk, I. V., Gasanov, E. V., Safina, D. R., & Kostrov, S. V. (2008). Structural organization of precursors of thermolysin-like proteinases. *Protein Journal*, 27(6), 343–354. <https://doi.org/10.1007/s10930-008-9143-2>
- 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 Research*, 36, 465–469. <https://doi.org/10.1093/nar/gkn180>
- Desrivieres, S., Lu, H., Peyri, N., Soria, C., Legrand, Y., & Ménashi, S. (1993). Activation of the 92 kDa type IV collagenase by tissue kallikrein. *Journal of Cellular Physiology*, 157(3), 587–593. <https://doi.org/10.1002/jcp.1041570319>
- Devarie-Baez, N. O., Lopez, E. I. S., & Furdui, C. M. (2015). Biological chemistry and functionality of protein sulfenic acids and related thiol modifications. *Free Radic Res* 50(2), 172–194. <https://doi.org/10.3109/10715762.2015.1090571>
- Diao, L., Dong, Q., Xu, Z., Yang, S., Zhou, J., & Freudl, R. (2012). Functional Implementation of the Posttranslational SecB-SecA Protein-Targeting Pathway in *Bacillus subtilis*. *Applied and Environmental Microbiology*, 78(3), 651–659. <https://doi.org/10.1128/AEM.07209-11>
- DiCarlo, J. E., Norville, J. E., Mali, P., Rios, X., Aach, J., & Church, G. M. (2013). Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Research*, 41(7), 4336–4343. <https://doi.org/10.1093/nar/gkt135>

- Doi, R. H., & Kosugi, A. (2004). Cellulosomes: plant-cell-wall-degrading enzyme complexes. *Nature Reviews Microbiology*, 2(7), 541–551. <https://doi.org/10.1038/nrmicro925>
- Dubnau, D. (1991). Genetic competence in *Bacillus subtilis*. *Microbiological Reviews*, 55(3), 395–424. <https://doi.org/10.1128/MMBR.55.3.395-424.1991>
- Dubnau, D., & Provvedi, R. (2000). Internalizing DNA. *Research in Microbiology*, 151(6), 475–480. [https://doi.org/10.1016/S0923-2508\(00\)00166-2](https://doi.org/10.1016/S0923-2508(00)00166-2)
- Duitman, E. H., Wyczawski, D., Boven, L. G., Venema, G., Kuipers, O. P., & Hamoen, L. W. (2007). Novel Methods for Genetic Transformation of Natural *Bacillus subtilis* Isolates Used to Study the Regulation of the Mycosubtilin and Surfactin Synthetases. *Applied and Environmental Microbiology*, 73(11), 3490–3496. <https://doi.org/10.1128/AEM.02751-06>
- Ebrahimie, E., Ebrahimi, M., Sarvestani, N. R., & Ebrahimi, M. (2011). Protein attributes contribute to halo-stability, bioinformatics approach. *Saline Systems*, 7(1), 1. <https://doi.org/10.1186/1746-1448-7-1>
- Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Computational Biology*, 7(10), e1002195. <https://doi.org/10.1371/journal.pcbi.1002195>
- Englard, S., & Seifter, S. (1990). Precipitation techniques. *Methods in Enzymology* (182) Pg. 285-300 [https://doi.org/10.1016/0076-6879\(90\)82024-V](https://doi.org/10.1016/0076-6879(90)82024-V)
- English, A. C., Groom, C. R., & Hubbard, R. E. (2001). Experimental and computational mapping of the binding surface of a crystalline protein. *Protein Engineering*, 14(1), 47–59.
- Evans, A. O., Marsh, J. M., & Wickett, R. R. (2011). The structural implications of water hardness metal uptake by human hair. *International Journal of Cosmetic Science*, 33(5), 477–482. <https://doi.org/10.1111/j.1468-2494.2011.00659.x>
- Fabry, S., Ruigrok, R. W. H., Ebel, C., ¶3, A. R., & Franzetti, B. (2006). An Archaeal Peptidase Assembles into Two Different Quaternary Structures: A tetrahedron and a giant octahedron. *Journal of Biological Chemistry* 281(47). <https://doi.org/10.1074/jbc.M604417200>
- Farias, S. T., Christina Bonato, M. M., Farias, S., & Bonato, M. (2003). Preferred amino acids and thermostability. In *Genetics and Molecular Research* (Vol. 2, Issue 4). www.funpecrp.com.br
- Feng, L., Yan, H., Wu, Z., Yan, N., Wang, Z., Jeffrey, P. D., & Shi, Y. (2007). Structure of a Site-2 Protease Family Intramembrane Metalloprotease. *Science*, 318(5856), 1608–1612. <https://doi.org/10.1126/science.1150755>

- Fernández, D., Russi, S., Vendrell, J., Monod, M., & Pallarès, I. (2013). A functional and structural study of the major metalloprotease secreted by the pathogenic fungus *Aspergillus fumigatus*. *Acta Crystallographica. Section D, Biological Crystallography*, 69(Pt 10), 1946–1957.
<https://doi.org/10.1107/S0907444913017642>
- Fontana, A. (1988). Structure and stability of thermophilic enzymes Studies on thermolysin. In *Biophysical Chemistry* (Vol. 29).
- Fortelny, N., Cox, J. H., Kappelhoff, R., Starr, A. E., Lange, P. F., Pavlidis, P., & Overall, C. M. (2014). Network Analyses Reveal Pervasive Functional Regulation Between Proteases in the Human Protease Web. *PLOS Biology*, 12(5), e1001869.
<https://doi.org/10.1371/JOURNAL.PBIO.1001869>
- Fricke, B., Dröbber, K., Willhardt, I., Schierhorn, A., Menge, S., & Rücknagel, P. (2001). The cell envelope-bound metalloprotease (camelysin) from *Bacillus cereus* is a possible pathogenic factor. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1537(2), 132–146.
[https://doi.org/10.1016/S0925-4439\(01\)00066-7](https://doi.org/10.1016/S0925-4439(01)00066-7)
- Fu, G., Liu, J., Li, J., Zhu, B., & Zhang, D. (2018). Systematic Screening of Optimal Signal Peptides for Secretory Production of Heterologous Proteins in *Bacillus subtilis*. *Journal of Agricultural and Food Chemistry*, 66(50), 13141–13151.
<https://doi.org/10.1021/acs.jafc.8b04183>
- Fu, L. L., Zi Rong Xu, Wei Fen Li, Jiang Bing Shuai, Ping Lu, & Chun Xia Hu. (2007). Protein secretion pathways in *Bacillus subtilis*: Implication for optimization of heterologous protein secretion. *Biotechnology Advances*, 25(1), 1–12.
<https://doi.org/10.1016/j.biotechadv.2006.08.002>
- Fujita, T., Shiota, K., Yoshikawa, J., Ogawa, S., & Aoyagi, H. (2019). Simple method for analyzing the purity of protease-containing samples by acid-treatment SDS-PAGE. *J Biosci Bioeng* 128(5). <https://doi.org/10.1016/j.jbiosc.2019.04.014>
- Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.; Protein Identification and Analysis Tools on the Expasy Server; (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press (2005). pp. 571-607
- Geurts, N., Becker-Pauly, C., Martens, E., Proost, P., Van den Steen, P. E., Stöcker, W., & Opdenakker, G. (2012). Meprins process matrix metalloproteinase-9 (MMP-9)/gelatinase B and enhance the activation kinetics by MMP-3. *FEBS Letters*, 586(24), 4264–4269.
<https://doi.org/10.1016/j.febslet.2012.10.033>

- Ghilarov, D., Serebryakova, M., Stevenson, C. E. M., Hearnshaw, S. J., Volkov, D. S., Maxwell, A., Lawson, D. M., & Severinov, K. (2017). The Origins of Specificity in the Microcin-Processing Protease TldD/E. *Structure*, 25(10), 1549-1561.e5. <https://doi.org/10.1016/J.STR.2017.08.006>
- Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, 6(5), 343–345. <https://doi.org/10.1038/nmeth.1318>
- Ginestet, C. (2011). ggplot2: Elegant Graphics for Data Analysis. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 174(1), 245–246. https://doi.org/10.1111/j.1467-985X.2010.00676_9.x
- Gomis-Rüth, F. X. (2003). Structural aspects of the metzincin clan of metalloendopeptidases. *Applied Biochemistry and Biotechnology - Part B Molecular Biotechnology*, 24(2), 157–202. <https://doi.org/10.1385/MB:24:2:157>
- Gomis-Rüth, F. X., Stöcker, W., Huber, R., Zwilling, R., & Bode, W. (1993). Refined 1·8 Å X-ray Crystal Structure of Astacin, a Zinc-endopeptidase from the Crayfish *Astacus astacus* L.: Structure Determination, Refinement, Molecular Structure and Comparison with Thermolysin. *Journal of Molecular Biology*, 229(4), 945–968. <https://doi.org/10.1006/JMBI.1993.1098>
- González-Pastor, J. E., Hobbs, E. C., & Losick, R. (2003). Cannibalism by Sporulating Bacteria. *Science*, 301(5632), 510–513. <https://doi.org/10.1126/science.1086462>
- Goulas, T., Arolas, J. L., & Gomis-Rüth, F. X. (2011). Structure, function and latency regulation of a bacterial enterotoxin potentially derived from a mammalian adamalysin/ADAM xenolog. *Proceedings of the National Academy of Sciences*, 108(5), 1856–1861. <https://doi.org/10.1073/pnas.1012173108>
- Goussarov, G., Mysara, M., Vandamme, P., & Van Houdt, R. (2022). Introduction to the principles and methods underlying the recovery of metagenome-assembled genomes from metagenomic data. *Microbiology Open*, 11(3). <https://doi.org/10.1002/mbo3.1298>
- Graham, S. C., Bond, C. S., Freeman, H. C., & Guss, J. M. (2005). Structural and functional implications of metal ion selection in aminopeptidase P, a metalloprotease with a dinuclear metal center. *Biochemistry*, 44(42), 13820–13836. <https://doi.org/10.1021/bi0512849>
- Gromiha, M. M., & Suresh, M. X. (2007). Discrimination of mesophilic and thermophilic proteins using machine learning algorithms. *Proteins: Structure, Function, and Bioinformatics*, 70(4), 1274–1279. <https://doi.org/10.1002/prot.21616>

- Gupta, R., Beg, Q. K., Khan, S., & Chauhan, B. (2002). An overview on fermentation, downstream processing and properties of microbial alkaline proteases. In *Applied Microbiology and Biotechnology* (Vol. 60, Issue 4, pp. 381–395). Springer Verlag. <https://doi.org/10.1007/s00253-002-1142-1>
- Hahn, J., Luttinger, A., & Dubnau, D. (1996). Regulatory inputs for the synthesis of ComK, the competence transcription factor of *Bacillus subtilis*. *Molecular Microbiology*, 21(4), 763–775. <https://doi.org/10.1046/j.1365-2958.1996.371407.x>
- Hahn, J., Roggiani, M., & Dubnau, D. (1995). The major role of Spo0A in genetic competence is to downregulate *abrB*, an essential competence gene. *Journal of Bacteriology*, 177(12), 3601–3605. <https://doi.org/10.1128/jb.177.12.3601-3605.1995>
- Hasan, F., Shah, A. A., Javed, S., & Hameed, A. (2010). Enzymes used in detergents: Lipases. *African Journal of Biotechnology*, 9(31), 4836–4844.
- Hede, P. D. (2020, May 7). *A beginner's guide to enzymes in detergents*. Novozymes - Household Care. [https://biosolutions.novozymes.com/en/dish/insights/article/beginners-guide-enzymes-detergents#:~:text=Enzymes%20in%20\(heavy%2Dduty\),attached%20to%20microfibrils%20are%20removed.](https://biosolutions.novozymes.com/en/dish/insights/article/beginners-guide-enzymes-detergents#:~:text=Enzymes%20in%20(heavy%2Dduty),attached%20to%20microfibrils%20are%20removed.)
- Hemmerich, J., Rohe, P., Kleine, B., Jurischka, S., Wiechert, W., Freudl, R., & Oldiges, M. (2016). Use of a Sec signal peptide library from *Bacillus subtilis* for the optimization of cutinase secretion in *Corynebacterium glutamicum*. *Microbial Cell Factories*, 15, 208. <https://doi.org/10.1186/s12934-016-0604-6>
- Hohmann, H., van Dijk, J. M., Krishnappa, L., & Prágai, Z. (2017). Host Organisms: *Bacillus subtilis*. In *Industrial Biotechnology* (pp. 221–297). Wiley. <https://doi.org/10.1002/9783527807796.ch7>
- Holland, D. R., Tronrud, D. E., Matthews, B. W., Pley, H. W., Flaherty, K. M., McKay, D. B., Stark, W., & Jansonius, J. N. (1992). Structural Comparison Suggests That Thermolysin and Related Neutral Proteases Undergo Hinge-Bending Motion During Catalysis. *Biochemistry*, 31(46), 11310–11316. <https://doi.org/10.1021/bi00161a008>
- Hooper, N. M. (2013). Membrane dipeptidase. In *Handbook of Proteolytic Enzymes* (pp. 1670–1673). Elsevier. <https://doi.org/10.1016/B978-0-12-382219-2.00377-X>

- Hori, T., Kumasaka, T., Yamamoto, M., Nonaka, T., Tanaka, N., Hashimoto, Y., Ueki, T., & Takio, K. (2001). Structure of a new “aspzincin” metalloendopeptidase from *Grifola frondosa*: Implications for the catalytic mechanism and substrate specificity based on several different crystal forms. *Acta Crystallographica Section D: Biological Crystallography*, 57(3), 361–368. <https://doi.org/10.1107/S09074444900019740>
- Hu, Y., Peng, N., Han, W., Mei, Y., Chen, Z., Feng, X., Liang, Y. X., & She, Q. (2012). An archaeal protein evolutionarily conserved in prokaryotes is a zinc-dependent metalloprotease. *Bioscience Reports*, 32(6), 609–618. <https://doi.org/10.1042/BSR20120074>
- Huddy, R. J., & Coyne, V. E. (2015). Characterisation of the role of an alkaline protease from *Vibrio midae* SY9 in enhancing the growth rate of cultured abalone fed a probiotic-supplemented feed. *Aquaculture*, 448, 128–134. <https://doi.org/10.1016/j.aquaculture.2015.05.048>
- Im, H., Manolopoulou, M., Malito, E., Shen, Y., Zhao, J., Neant-Fery, M., Sun, C. Y., Meredith, S. C., Sisodia, S. S., Leissring, M. A., & Tang, W. J. (2007). Structure of Substrate-free Human Insulin-degrading Enzyme (IDE) and Biophysical Analysis of ATP-induced Conformational Switch of IDE. *Journal of Biological Chemistry*, 282(35), 25453–25463. <https://doi.org/10.1074/JBC.M701590200>
- Inès, M., Mouna, B., Marwa, E., & Dhouha, G. (2023). Biosurfactants as Emerging Substitutes of Their Synthetic Counterpart in Detergent Formula: Efficiency and Environmental Friendly. *Journal of Polymers and the Environment*, 31(7), 2779–2791. <https://doi.org/10.1007/s10924-023-02778-1>
- Iyer, S., Visse, R., Nagase, H., & Acharya, K. R. (2006). Crystal Structure of an Active Form of Human MMP-1. *Journal of Molecular Biology*, 362(1), 78–88. <https://doi.org/10.1016/J.JMB.2006.06.079>
- Jackson, P. L., Xu, X., Wilson, L., Weathington, N. M., Clancy, J. P., Blalock, J. E., & Gaggar, A. (2010). Human neutrophil elastase-mediated cleavage sites of MMP-9 and TIMP-1: Implications to cystic fibrosis proteolytic dysfunction. *Molecular Medicine*, 16(5–6), 159–166. <https://doi.org/10.2119/molmed.2009.00109>
- Jadhav, J. V., Anbu, P., Yadav, S., Pratap, A. P., & Kale, S. B. (2019). Sunflower Acid Oil-Based Production of Rhamnolipid Using *Pseudomonas aeruginosa* and Its Application in Liquid Detergents. *Journal of Surfactants and Detergents*, 22(3), 463–476. <https://doi.org/10.1002/jsde.12255>

- Jakočiūnas, T., Jensen, M. K., & Keasling, J. D. (2016). CRISPR/Cas9 advances engineering of microbial cell factories. *Metabolic Engineering*, 34, 44–59.
<https://doi.org/10.1016/j.ymben.2015.12.003>
- Jakutyte-Giraitiene, L., & Gasiunas, G. (2016). Design of a CRISPR-Cas system to increase resistance of *Bacillus subtilis* to bacteriophage SPP1. *Journal of Industrial Microbiology and Biotechnology*, 43(8), 1183–1188. <https://doi.org/10.1007/s10295-016-1784-0>
- Jamdar, S. N., Are, V. N., Navamani, M., Kumar, S., Nagar, V., & Makde, R. D. (2015). The members of M20D peptidase subfamily from *Burkholderia cepacia*, *Deinococcus radiodurans* and *Staphylococcus aureus* (HmrA) are carboxydipeptidases, primarily specific for Met-X dipeptides. *Arch Biochem Biophys* 587(18-30)
<https://doi.org/10.1016/j.abb.2015.10.003>
- Jaouadi, N. Z., Jaouadi, B., Aghajari, N., & Bejar, S. (2012). The overexpression of the SAPB of *Bacillus pumilus* CBS and mutated sapB-L31I/T33S/N99Y alkaline proteases in *Bacillus subtilis* DB430: New attractive properties for the mutant enzyme. *Bioresource Technology*, 105, 142–151. <https://doi.org/https://doi.org/10.1016/j.biortech.2011.11.115>
- Jasuja, R., Ge, G., Voss, N. G., Lyman-Gingerich, J., Branam, A. M., Pelegri, F. J., & Greenspan, D. S. (2007). Bone Morphogenetic Protein 1 Prodomain Specifically Binds and Regulates Signaling by Bone Morphogenetic Proteins 2 and 4. *Journal of Biological Chemistry*, 282(12), 9053–9062. <https://doi.org/10.1074/jbc.M610929200>
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, 337(6096), 816–821. <https://doi.org/10.1126/science.1225829>
- Jones, D. T., Buchan, D. W. A., Cozzetto, D., & Pontil, M. (2012). PSICOV: precise structural contact prediction using sparse inverse covariance estimation on large multiple sequence alignments. *Bioinformatics*, 28(2), 184–190. <https://doi.org/10.1093/bioinformatics/btr638>
- Jongeneel, C. V., Bouvier, J., & Bairoch, A. (1989). A unique signature identifies a family of zinc-dependent metallopeptidases. *FEBS Letters*, 242(2), 211–214. [https://doi.org/10.1016/0014-5793\(89\)80471-5](https://doi.org/10.1016/0014-5793(89)80471-5)
- Jozic, D., Bourenkov, G., Lim, N. H., Visse, R., Nagase, H., Bode, W., & Maskos, K. (2005). X-ray Structure of Human proMMP-1: NEW INSIGHTS INTO PROCOLLAGENASE ACTIVATION AND COLLAGEN BINDING. *Journal of Biological Chemistry*, 280(10), 9578–9585.
<https://doi.org/10.1074/JBC.M411084200>

- Juhas, M., Reuß, D. R., Zhu, B., & Commichau, F. M. (2014). *Bacillus subtilis* and *Escherichia coli* essential genes and minimal cell factories after one decade of genome engineering. *Microbiology*, 160(11), 2341–2351. <https://doi.org/10.1099/mic.0.079376-0>
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 2021 596:7873, 596(7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>
- Kakeshtia, H., Kageyama, Y., Ara, K., Ozaki, K., & Nakamura, K. (2010). Enhanced Extracellular Production of Heterologous Proteins in *Bacillus subtilis* by Deleting the C-terminal Region of the SecA Secretory Machinery. *Molecular Biotechnology*, 46(3), 250–257. <https://doi.org/10.1007/s12033-010-9295-0>
- Kelwick, R., Desanlis, I., Wheeler, G. N., & Edwards, D. R. (2015). The ADAMTS (A Disintegrin and Metalloproteinase with Thrombospondin motifs) family. *Genome Biology*, 16(1), 113. <https://doi.org/10.1186/s13059-015-0676-3>
- Khaje Bafghi, M., & Fazelipour, M. H. (2012). Application of Rhamnolipid in the Formulation of a Detergent. *Journal of Surfactants and Detergents*, 15(6), 679–684. <https://doi.org/10.1007/s11743-012-1386-4>
- Khandelwal, H. B., More, S. V., Kalal, K. M., & Laxman, R. S. (2015). Eco-friendly enzymatic dehairing of skins and hides by *C. brevidianus* protease. *Clean Technologies and Environmental Policy*, 17(2), 393–405. <https://doi.org/10.1007/s10098-014-0791-y>
- Kim, H., & Lipscomb, W. N. (1993). Differentiation and identification of the two catalytic metal binding sites in bovine lens leucine aminopeptidase by x-ray crystallography. *Proceedings of the National Academy of Sciences*, 90(11), 5006–5010. <https://doi.org/10.1073/PNAS.90.11.5006>
- Kim, J. K., Natarajan, S., Park, H., Huynh, K. H., Lee, S. H., Kim, J. G., Ahn, Y. J., & Kang, L. W. (2013). Crystal structure of XoLAP, a leucine aminopeptidase, from *Xanthomonas oryzae* pv. *oryzae*. *Journal of Microbiology*, 51(5), 627–632. <https://doi.org/10.1007/S12275-013-3234-2>/METRICS
- Kim, Y., Kim, E.-K., Chey, Y., Song, M.-J., & Jang, H. H. (2023). Targeted Protein Degradation: Principles and Applications of the Proteasome. *Cells*, 12(14), 1846. <https://doi.org/10.3390/cells12141846>

- Klee, E. W., & Ellis, L. B. (2005). Evaluating eukaryotic secreted protein prediction. *BMC Bioinformatics*, 6(1), 256. <https://doi.org/10.1186/1471-2105-6-256>
- Klein, T., Eckhard, U., Dufour, A., Solis, N., & Overall, C. M. (2018). Proteolytic Cleavage - Mechanisms, Function, and “omic” Approaches for a Near-Ubiquitous Posttranslational Modification. In *Chemical Reviews* (Vol. 118, Issue 3, pp. 1137–1168). American Chemical Society. <https://doi.org/10.1021/acs.chemrev.7b00120>
- Kobayashi, K., Ehrlich, S. D., Albertini, A., Amati, G., Andersen, K. K., Arnaud, M., Asai, K., Ashikaga, S., Aymerich, S., Bessieres, P., Boland, F., Brignell, S. C., Bron, S., Bunai, K., Chapuis, J., Christiansen, L. C., Danchin, A., Débarbouillé, M., Dervyn, E., Ogasawara, N. (2003). Essential *Bacillus subtilis* genes. *Proceedings of the National Academy of Sciences*, 100(8), 4678–4683. <https://doi.org/10.1073/pnas.0730515100>
- Komatsu, M., Uchiyama, T., Ōmura, S., Cane, D. E., & Ikeda, H. (2010). Genome-minimized *Streptomyces* host for the heterologous expression of secondary metabolism. *Proceedings of the National Academy of Sciences*, 107(6), 2646–2651. <https://doi.org/10.1073/pnas.0914833107>
- Koo, B.-H., Longpré, J.-M., Somerville, R. P. T., Alexander, J. P., Leduc, R., & Apte, S. S. (2007). Regulation of ADAMTS9 Secretion and Enzymatic Activity by Its Propeptide. *Journal of Biological Chemistry*, 282(22), 16146–16154. <https://doi.org/10.1074/jbc.M610161200>
- Koodynska, D. (2011). Chelating Agents of a New Generation as an Alternative to Conventional Chelators for Heavy Metal Ions Removal from Different Waste Waters. In *Expanding Issues in Desalination*. InTech. <https://doi.org/10.5772/21180>
- Koppen, M., Bonn, F., Ehse, S., & Langer, T. (2009). Autocatalytic Processing of m-AAA Protease Subunits in Mitochondria. *Molecular Biology of the Cell*, 20(19), 4216–4224. <https://doi.org/10.1091/mbc.e09-03-0218>
- Kouwen, T. R. H. M., Nielsen, A. K., Denham, E. L., Dubois, J.-Y. F., Dorenbos, R., Rasmussen, M. D., Quax, W. J., Freudl, R., & Maarten Van Dijk, J. (2010). Contributions of the Pre- and Pro-Regions of a *Staphylococcus hyicus* Lipase to Secretion of a Heterologous Protein by *Bacillus subtilis*. *Applied and Environmental Microbiology*, 76(3), 659–669. <https://doi.org/10.1128/AEM.01671-09>
- Kuddus, M., & Ramteke, P. W. (2012). Recent developments in production and biotechnological applications of cold-active microbial proteases. In *Critical Reviews in Microbiology* (Vol. 38, Issue 4, pp. 330–338). <https://doi.org/10.3109/1040841X.2012.678477>

- Kunst, F., & Rapoport, G. (1995). Salt stress is an environmental signal affecting degradative enzyme synthesis in *Bacillus subtilis*. *Journal of Bacteriology*, 177(9), 2403–2407. <https://doi.org/10.1128/jb.177.9.2403-2407.1995>
- Kuo, L. Y., Hwang, G. Y., Lai, Y. J., Yang, S. L., & Lin, L. L. (2003). Overexpression, Purification, and Characterization of the Recombinant Leucine Aminopeptidase II of *Bacillus stearothermophilus*. *Current Microbiology* 2003 47:1, 47(1), 0040–0045. <https://doi.org/10.1007/S00284-002-3950-Z>
- Kuzuya, K., & Inouye, K. (2001). Effects of Cobalt-Substitution of the Active Zinc Ion in Thermolysin on its Activity and Active-Site Microenvironment. *J. Biochem*, 130(18), 783–788. <https://academic.oup.com/jb/article/130/6/783/798471>
- Labun, K., Montague, T. G., Gagnon, J. A., Thyme, S. B., & Valen, E. (2016). CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Research*, 44(W1), W272–W276. <https://doi.org/10.1093/nar/gkw398>
- Le Loir, Y., Gruss, A., Ehrlich, S. D., & Langella, A. P. (1998). A Nine-Residue Synthetic Propeptide Enhances Secretion Efficiency of Heterologous Proteins in *Lactococcus lactis*. In *Journal of Bacteriology* (Vol. 180, Issue 7). <http://jb.asm.org/>
- Lee, B. H., Lee, M. J., Park, S., Oh, D. C., Elsasser, S., Chen, P. C., Gartner, C., Dimova, N., Hanna, J., Gygi, S. P., Wilson, S. M., King, R. W., & Finley, D. (2010). Enhancement of proteasome activity by a small-molecule inhibitor of USP14. *Nature* 2010 467:7312, 467(7312), 179–184. <https://doi.org/10.1038/NATURE09299>
- Lee, M. J., Lee, B.-H., Hanna, J., King, R. W., & Finley, D. (2011). Trimming of Ubiquitin Chains by Proteasome-associated Deubiquitinating Enzymes. *Molecular & Cellular Proteomics* 10(5). <https://doi.org/10.1074/mcp.R110.003871>
- Leprince, A., de Lorenzo, V., Völler, P., van Passel, M. W. J., & Martins dos Santos, V. A. P. (2012). Random and cyclical deletion of large DNA segments in the genome of *Pseudomonas putida*. *Environmental Microbiology*, 14(6), 1444–1453. <https://doi.org/10.1111/j.1462-2920.2012.02730.x>
- Li, B., Yan, Z.-Y., Liu, X.-N., Zhou, J., Wu, X.-Y., Wei, P., Jia, H.-H., & Yong, X.-Y. (2019). Increased fermentative adenosine production by gene-targeted *Bacillus subtilis* mutation. *Journal of Biotechnology*, 298, 1–4. <https://doi.org/10.1016/j.jbiotec.2019.04.007>
- Li, W., & Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13), 1658–1659. <https://doi.org/10.1093/bioinformatics/btl158>

- Li, W., Zhou, X., & Lu, P. (2004). Bottlenecks in the expression and secretion of heterologous proteins in *Bacillus subtilis*. *Research in Microbiology*, 155(8), 605–610.
<https://doi.org/10.1016/j.resmic.2004.05.002>
- Li, Y., Wu, Y., Liu, Y., Li, J., Du, G., Lv, X., & Liu, L. (2022). A genetic toolkit for efficient production of secretory protein in *Bacillus subtilis*. *Bioresource Technology*, 363, 127885.
<https://doi.org/10.1016/j.biortech.2022.127885>
- Li, Z., Jaroszewski, L., Iyer, M., Sedova, M., & Godzik, A. (2020). FATCAT 2.0: towards a better understanding of the structural diversity of proteins. *Nucleic Acids Research*, 48(W1), W60–W64. <https://doi.org/10.1093/NAR/GKAA443>
- Lin, H., & Chen, W. (2011). Prediction of thermophilic proteins using feature selection technique. *Journal of Microbiological Methods*, 84(1), 67–70.
<https://doi.org/https://doi.org/10.1016/j.mimet.2010.10.013>
- Lin, H., Chen, W., & Ding, H. (2013). AcalPred: A Sequence-Based Tool for Discriminating between Acidic and Alkaline Enzymes. *PLoS One* 8(10).
<https://doi.org/10.1371/journal.pone.0075726>
- Lin, H. H., Yin, L. J., & Jiang, S. T. (2009). Cloning, expression, and purification of *Pseudomonas aeruginosa* keratinase in *Escherichia coli* AD494(DE3)pLysS expression system. *Journal of Agricultural and Food Chemistry*, 57(9), 3506–3511. <https://doi.org/10.1021/jf803752j>
- Liu, D., Mao, Z., Guo, J., Wei, L., Ma, H., Tang, Y., Chen, T., Wang, Z., & Zhao, X. (2018a). Construction, Model-Based Analysis, and Characterization of a Promoter Library for Fine-Tuned Gene Expression in *Bacillus subtilis*. *ACS Synthetic Biology*, 7(7), 1785–1797.
<https://doi.org/10.1021/acssynbio.8b00115>
- Liu, L., Nakano, M. M., Lee, O. H., & Zuber, P. (1996). Plasmid-amplified *comS* enhances genetic competence and suppresses *sinR* in *Bacillus subtilis*. *Journal of Bacteriology*, 178(17), 5144–5152. <https://doi.org/10.1128/jb.178.17.5144-5152.1996>
- Liu, X., Wang, H., Wang, B., & Pan, L. (2018b). Efficient production of extracellular pullulanase in *Bacillus subtilis* ATCC6051 using the host strain construction and promoter optimization expression system. *Microbial Cell Factories*, 17(1), 163. <https://doi.org/10.1186/s12934-018-1011-y>
- Liu, Z., Zheng, W., Ge, C., Cui, W., Zhou, L., & Zhou, Z. (2019). High-level extracellular production of recombinant nattokinase in *Bacillus subtilis* WB800 by multiple tandem promoters. *BMC Microbiology*, 19(89). <https://doi.org/10.1186/s12866-019-1461-3>

- Loenen, W. A. M., Dryden, D. T. F., Raleigh, E. A., Wilson, G. G., & Murray, N. E. (2014). Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic Acids Research*, 42(1), 3–19. <https://doi.org/10.1093/nar/gkt990>
- Lu, Z., Yang, S., Yuan, X., Shi, Y., Ouyang, L., Jiang, S., Yi, L., & Zhang, G. (2019). CRISPR-assisted multi-dimensional regulation for fine-tuning gene expression in *Bacillus subtilis*. *Nucleic Acids Research*, 47(7), e40–e40. <https://doi.org/10.1093/nar/gkz072>
- Luo, M. R. (2002). Development of colour-difference formulae. *Review of Progress in Coloration and Related Topics*, 32, 28–39. <https://doi.org/10.1111/j.1478-4408.2002.tb00248.x>
- Luo, M. R., Cui, G., & Rigg, B. (2001). The development of the CIE 2000 colour-difference formula: CIEDE2000. *Color Research and Application*, 26(5), 340–350. <https://doi.org/10.1002/col.1049>
- Lupas, A., Flanagan, J. M., Tamura, T., & Baumeister, W. (1997). Self-compartmentalizing proteases. *Trends in Biochemical Sciences*, 22(10), 399–404. [https://doi.org/10.1016/S0968-0004\(97\)01117-1](https://doi.org/10.1016/S0968-0004(97)01117-1)
- Maamar, H., & Dubnau, D. (2005). Bistability in the *Bacillus subtilis* K-state (competence) system requires a positive feedback loop. *Molecular Microbiology*, 56(3), 615–624. <https://doi.org/10.1111/j.1365-2958.2005.04592.x>
- Madern, D., & Zaccai, G. (2004). Molecular adaptation: the malate dehydrogenase from the extreme halophilic bacterium *Salinibacter ruber* behaves like a non-halophilic protein. *Biochimie*, 86(4), 295–303. <https://doi.org/https://doi.org/10.1016/j.biochi.2004.04.004>
- Majerus, E. M., Zheng, X., Tuley, E. A., & Sadler, J. E. (2003). Cleavage of the ADAMTS13 Propeptide Is Not Required for Protease Activity. *Journal of Biological Chemistry*, 278(47), 46643–46648. <https://doi.org/10.1074/jbc.M309872200>
- Manjrekar S, Wadekar T, & Sumant O. (2021). *Enzymes market, allied market research*. <https://www.alliedmarketresearch.com/enzymes-market>
- Mansfeld, J., & Ulbrich-Hofmann, R. (2007). The stability of engineered thermostable neutral proteases from *Bacillus stearothermophilus* in organic solvents and detergents. *Biotechnology and Bioengineering*, 97(4), 672–679. <https://doi.org/10.1002/bit.21292>
- Mao, X., Huang, Z., Sun, G., Zhang, H., Lu, W., Liu, Y., Lv, X., Du, G., Li, J., & Liu, L. (2021). High level production of diacetylchitobiose deacetylase by refactoring genetic elements and cellular metabolism. *Bioresource Technology*, 341, 125836. <https://doi.org/10.1016/j.biortech.2021.125836>

- Marie-Claire, C., Roques, B. P., & Beaumont, A. (1998). Intramolecular Processing of Prothermolysin. *Journal of Biological Chemistry* 10(6).
<https://doi.org/10.1074/jbc.273.10.5697>
- Marks, D. S., Colwell, L. J., Sheridan, R., Hopf, T. A., Pagnani, A., Zecchina, R., & Sander, C. (2011). Protein 3D Structure Computed from Evolutionary Sequence Variation. *PLoS ONE*, 6(12), e28766. <https://doi.org/10.1371/journal.pone.0028766>
- Marshall, P. S., Leavens, B., Heudi, O., & Ramirez-Molina, C. (2004). Liquid chromatography coupled with inductively coupled plasma mass spectrometry in the pharmaceutical industry: selected examples. *Journal of Chromatography A*, 1056(1–2), 3–12.
<https://doi.org/10.1016/j.chroma.2004.07.073>
- Matthews, B. W., Jansonius, J. N., Colman, P. M., Schoenborn, B. P., & Dupourque, D. (1972). Three-dimensional Structure of Thermolysin. *Nature New Biology*, 238(80), 37–41.
<https://doi.org/10.1038/newbio238037a0>
- Maurer, K. H. (2004). Detergent proteases. In *Current Opinion in Biotechnology* (Vol. 15, Issue 4, pp. 330–334). <https://doi.org/10.1016/j.copbio.2004.06.005>
- Mazzei, L., Musiani, F., & Ciurli, S. (2020). The structure-based reaction mechanism of urease, a nickel dependent enzyme: tale of a long debate. *Journal of Biological Inorganic Chemistry* (Vol. 25, Issue 6, pp. 829–845). <https://doi.org/10.1007/s00775-020-01808-w>
- McAuley, K. E., Jia-Xing, Y., Dodson, E. J., Lehmbeck, J., Østergaard, P. R., & Wilson, K. S. (2001). A quick solution: ab initio structure determination of a 19 kDa metalloproteinase using ACORN. *Urn:Issn:0907-4449*, 57(11), 1571–1578.
<https://doi.org/10.1107/S090744490101335X>
- McCafferty, D. G., Lessard, I. A. D., & Walsh, C. T. (1997). Mutational analysis of potential zinc-binding residues in the active site of the enterococcal D-Ala-D-Ala dipeptidase VanX. *Biochemistry*, 36(34), 10498–10505. <https://doi.org/10.1021/bi970543u>
- Meens, J., Herbort, M., Klein, M., & Freudl, R. (1997). Use of the Pre-Pro Part of *Staphylococcus hyicus* Lipase as a Carrier for Secretion of *Escherichia coli* Outer Membrane Protein A (OmpA) Prevents Proteolytic Degradation of OmpA by Cell-Associated Protease(s) in Two Different Gram-Positive Bacteria. In *Applied and Environmental Microbiology* (Vol. 63, Issue 7). <http://aem.asm.org/>

- Mekkes, J. R., Le Poole, I. C., Das, P. K., Bos, J. D., & Westerhof, W. (1998). Efficient debridement of necrotic wounds using proteolytic enzymes derived from Antarctic krill: A double-blind, placebo-controlled study in a standardized animal wound model. *Wound Repair and Regeneration*, 6(1), 50–57. <https://doi.org/10.1046/J.1524-475X.1998.60108.X>
- Meng, L., Ruebush, S., D'souza, V. M., Copik, A. J., Tsunasawa, S., & Holz, R. C. (2002). Overexpression and divalent metal binding properties of the methionyl aminopeptidase from *Pyrococcus furiosus*. *Biochemistry*, 41(23), 7199–7208. <https://doi.org/10.1021/BI020138P/ASSET/IMAGES/LARGE/BI020138PF00007.JPEG>
- Menne, K. M. L., Hermjakob, H., & Apweiler, R. (2000). A comparison of signal sequence prediction methods using a test set of signal peptides. *Bioinformatics*, 16(8), 741–742. <https://doi.org/10.1093/bioinformatics/16.8.741>
- MEROPS - the Peptidase Database*. (n.d.). Retrieved September 15, 2023, from <https://www.ebi.ac.uk/merops/>
- Mertens, S., Gallone, B., Steensels, J., Herrera-Malaver, B., Cortebeek, J., Nolmans, R., Saels, V., Vyas, V. K., & Verstrepen, K. J. (2019). Reducing phenolic off-flavors through CRISPR-based gene editing of the FDC1 gene in *Saccharomyces cerevisiae* x *Saccharomyces eubayanus* hybrid lager beer yeasts. *PLoS ONE*, 14(1). <https://doi.org/10.1371/journal.pone.0209124>
- Mevissen, T. E. T., & Komander, D. (2017). Mechanisms of Deubiquitinase Specificity and Regulation. *Annu Rev Biochem* 86:159-192. <https://doi.org/10.1146/annurev-biochem>
- Mikkelsen, M., Ryom, N. M., Ladefoged, C., & Friis-Jensen, S. (2012). Detergent Compositions and the Use of Enzyme Combinations Therein (Patent 8329632). U.S. Patent and Trademark Office.
- Miller, O. J., Bernath, K., Agresti, J. J., Amitai, G., Kelly, B. T., Mastrobattista, E., Taly, V., Magdassi, S., Tawfik, D. S., & Griffiths, A. D. (2006). Directed evolution by in vitro compartmentalization. *Nature Methods*, 3(7), 561–570. <https://doi.org/10.1038/nmeth897>
- Miyatake, H., Hata, Y., Fujii, T., Hamada, K., Morihara, K., & Katsube, Y. (1995). Crystal Structure of the Unliganded Alkaline Protease from *Pseudomonas aeruginosa* IFO3080 and Its Conformational Changes on Ligand Binding. *The Journal of Biochemistry*, 118(3), 474–479. <https://doi.org/10.1093/oxfordjournals.jbchem.a124932>
- Mole, V. C., Priimov, G. U., Sanderson, A. R., & Wright, M. P. (2020). Liquid Laundry Detergent Composition (Patent US Patent 10731111). U.S. Patent and Trademark Office.

- Montague, T. G., Cruz, J. M., Gagnon, J. A., Church, G. M., & Valen, E. (2014). CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Research*, 42(W1), W401–W407. <https://doi.org/10.1093/nar/gku410>
- Murzin, A. G. (1998). How far divergent evolution goes in proteins. *Current Opinion in Structural Biology*, 8(3), 380–387. [https://doi.org/10.1016/S0959-440X\(98\)80073-0](https://doi.org/10.1016/S0959-440X(98)80073-0)
- Myara, I., Cosson, C., Moatti, N., & Lemonnier, A. (1994). Human Kidney Prolidase-Purification, Preincubation Properties and Immunological Reactivity. *J. Biochem* (Vol. 26, Issue 2).
- Nath, A. (2016). Insights into the sequence parameters for halophilic adaptation. *Amino Acids*, 48(3), 751–762. <https://doi.org/10.1007/s00726-015-2123-x>
- Nath, A., & Subbiah, K. (2014). Inferring biological basis about psychrophilicity by interpreting the rules generated from the correctly classified input instances by a classifier. *Computational Biology and Chemistry*, 53, 198–203. <https://doi.org/https://doi.org/10.1016/j.compbiolchem.2014.10.002>
- Nerurkar, M., Joshi, M., Pariti, S., & Adivarekar, R. (2013). Application of Lipase from Marine Bacteria *Bacillus sonorensis* as an Additive in Detergent Formulation. *Journal of Surfactants and Detergents*, 16(3), 435–443. <https://doi.org/10.1007/s11743-012-1434-0>
- Nesslany, F., Simar-Meintières, S., Watzinger, M., Talahari, I., & Marzin, D. (2008). Characterization of the genotoxicity of nitrilotriacetic acid. *Environmental and Molecular Mutagenesis*, 49(6), 439–452. <https://doi.org/10.1002/em.20403>
- Ni, X., Yue, L., Chi, Z., Li, J., Wang, X., & Madzak, C. (2009). Alkaline protease gene cloning from the marine yeast *Aureobasidium pullulans* HN2-3 and the protease surface display on *Yarrowia lipolytica* for bioactive peptide production. *Marine Biotechnology*, 11(1), 81–89. <https://doi.org/10.1007/s10126-008-9122-9>
- Nielsen, H. (2017). Predicting secretory proteins with signalP. In *Methods in Molecular Biology* (Vol. 1611, pp. 59–73). Humana Press Inc. https://doi.org/10.1007/978-1-4939-7015-5_6
- Nielsen, H., Engelbrecht, J., Brunak, S., & von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering, Design and Selection*, 10(1), 1–6. <https://doi.org/10.1093/protein/10.1.1>
- Nijland, R., Burgess, J. G., Errington, J., & Veening, J.-W. (2010). Transformation of Environmental *Bacillus subtilis* Isolates by Transiently Inducing Genetic Competence. *PLoS ONE*, 5(3), e9724. <https://doi.org/10.1371/journal.pone.0009724>

- Nishihara, T., & Freese, E. (1975). Motility of *Bacillus subtilis* during growth and sporulation. *Journal of Bacteriology*, 123(1), 366–371. <https://doi.org/10.1128/jb.123.1.366-371.1975>
- Nitanai, Y., Satow, Y., Adachi, H., & Tsujimoto, M. (2002). Crystal Structure of Human Renal Dipeptidase Involved in β -Lactam Hydrolysis. *Journal of Molecular Biology*, 321(2), 177–184. [https://doi.org/10.1016/S0022-2836\(02\)00632-0](https://doi.org/10.1016/S0022-2836(02)00632-0)
- Niyonzima, F., & More, S. (2015). Detergent-Compatible Proteases: Microbial Production, Properties, and Stain Removal Analysis. *Preparative Biochemistry & Biotechnology*, 45(3), 233–258. <https://doi.org/10.1080/10826068.2014.907183>
- Niyonzima, F. N. (2019). Detergent-compatible bacterial cellulases. *Journal of Basic Microbiology*, 59(2), 134–147. <https://doi.org/10.1002/jobm.201800436>
- Oberg, F., Sjöhamn, J., Fischer, G., Moberg, A., Pedersen, A., Neutze, R., & Hedfalk, K. (2011). Glycosylation Increases the Thermostability of Human Aquaporin 10 Protein. *Journal of Biological Chemistry*, 286(36), 31915–31923. <https://doi.org/10.1074/jbc.M111.242677>
- Odintsov, S. G., Sabala, I., Marcyjaniak, M., & Bochtler, M. (2004). Latent LytM at 1.3Å Resolution. *Journal of Molecular Biology*, 335(3), 775–785. <https://doi.org/10.1016/j.jmb.2003.11.009>
- Odintsov, S. G., Sabala, I., Bourenkov, G., Rybin, V., & Bochtler, M. (2005). Substrate Access to the Active Sites in Aminopeptidase T, a Representative of a New Metallopeptidase Clan. *Journal of Molecular Biology*, 354(2), 403–412. <https://doi.org/10.1016/J.JMB.2005.09.042>
- Ogata, Y., Enghild, J. J., & Nagase, H. (1992). Matrix metalloproteinase 3 (stromelysin) activates the precursor for the human matrix metalloproteinase 9. *Journal of Biological Chemistry*, 267(6), 3581–3584. [https://doi.org/10.1016/S0021-9258\(19\)50563-4](https://doi.org/10.1016/S0021-9258(19)50563-4)
- Olędzka, G., Dąbrowski, S., & Kur, J. (2003). High-level expression, secretion, and purification of the thermostable aqualysin I from *Thermus aquaticus* YT-1 in *Pichia pastoris*. *Protein Expression and Purification*, 29(2), 223–229. [https://doi.org/10.1016/S1046-5928\(03\)00060-3](https://doi.org/10.1016/S1046-5928(03)00060-3)
- Oviedo, C., & Rodríguez, J. (2003). EDTA: the chelating agent under environmental scrutiny. *Química Nova*, 26(6), 901–905. <https://doi.org/10.1590/S0100-40422003000600020>

- Pallarès, I., Fernández, D., Comellas-Bigler, M., Fernández-Recio, J., Ventura, S., Avilés, F. X., Bode, W., & Vendrell, J. (2008). Direct interaction between a human digestive protease and the mucoadhesive poly(acrylic acid). *Acta Crystallographica Section D: Biological Crystallography*, 64(7), 784–791.
<https://doi.org/10.1107/S09074444908013474/HV5106SUP1.PDF>
- Pandey, A., Nigam, P., Soccol, C. R., Soccol, V. T., Singh, D., & Mohan, R. (2000). Advances in microbial amylases. *Biotechnology and Applied Biochemistry*, 31(2), 135.
<https://doi.org/10.1042/BA19990073>
- Pearson, M. A., Schaller, R. A., Michel, L. O., Karplus, P. A., & Hausinger, R. P. (1998). Chemical rescue of *Klebsiella aerogenes* urease variants lacking the carbamylated-lysine nickel ligand. *Biochemistry*, 37(17), 6214–6220.
<https://doi.org/10.1021/BI980021U/ASSET/IMAGES/LARGE/BI980021UF00005.JPEG>
- Pearson, W. R. (2013). Selecting the Right Similarity-Scoring Matrix. *Current Protocols in Bioinformatics*, 43(1). <https://doi.org/10.1002/0471250953.bi0305s43>
- Peng, F., Wang, X., Sun, Y., Dong, G., Yang, Y., Liu, X., & Bai, Z. (2017). Efficient gene editing in *Corynebacterium glutamicum* using the CRISPR/Cas9 system. *Microbial Cell Factories*, 16(1), 201. <https://doi.org/10.1186/s12934-017-0814-6>
- Peters, J. M., Colavin, A., Shi, H., Czarny, T. L., Larson, M. H., Wong, S., Hawkins, J. S., Lu, C. H. S., Koo, B.-M., Marta, E., Shiver, A. L., Whitehead, E. H., Weissman, J. S., Brown, E. D., Qi, L. S., Huang, K. C., & Gross, C. A. (2016). A Comprehensive, CRISPR-based Functional Analysis of Essential Genes in Bacteria. *Cell*, 165(6), 1493–1506.
<https://doi.org/10.1016/j.cell.2016.05.003>
- Petzoldt, T. (2022). growthrates: Estimate Growth Rates from Experimental Data.
<https://CRAN.R-project.org/package=growthrates>
- Piggot, P. J., & Hilbert, D. W. (2004). Sporulation of *Bacillus subtilis*. *Current Opinion in Microbiology*, 7(6), 579–586. <https://doi.org/10.1016/j.mib.2004.10.001>
- Pohl, S., Bhavsar, G., Hulme, J., Bloor, A. E., Misirli, G., Leckenby, M. W., Radford, D. S., Smith, W., Wipat, A., Williamson, E. D., Harwood, C. R., & Cranenburgh, R. M. (2013). Proteomic analysis of *Bacillus subtilis* strains engineered for improved production of heterologous proteins. *Proteomics*, 13(22), 3298–3308. <https://doi.org/10.1002/pmic.201300183>
- Price, M. A., Cruz, R., Baxter, S., Escalettes, F., & Rosser, S. J. (2019). CRISPR-Cas9 In Situ engineering of subtilisin E in *Bacillus subtilis*. *PLoS ONE*, 14(1).
<https://doi.org/10.1371/journal.pone.0210121>

- Price, M. A., Cruz, R., Bryson, J., Escalettes, F., & Rosser, S. J. (2020). Expanding and understanding the CRISPR toolbox for *Bacillus subtilis* with MAD7 and dMAD7. *Biotechnology and Bioengineering*, 117(6), 1805–1816. <https://doi.org/10.1002/BIT.27312>
- Quax, W. J., Kerkman, R., & Broekhuizen, C. P. (2002). Secretion factors for gram-positive microorganisms genes encoding them and methods of using it (Patent US6410262B1).
- Queiroga, A. C., Pintado, M. E., & Malcata, F. X. (2012). Search for novel proteolytic enzymes aimed at textile and agro-industrial applications: An overview of current and novel approaches. *Biocatalysis and Biotransformation*, 30(1), 154–169. <https://doi.org/10.3109/10242422.2012.650856>
- Quillet, L., Bensmail, L., Barray, S., & Guespin-Michel, J. (1997). Cloning and sequencing of two genes, *prtA* and *prtB*, from *Myxococcus xanthus*, encoding PrtA and PrtB proteases, both of which are required for the protease activity. *Gene*, 198(1–2), 135–140. [https://doi.org/10.1016/S0378-1119\(97\)00303-X](https://doi.org/10.1016/S0378-1119(97)00303-X)
- R Core Team. (2021). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*. <https://www.R-project.org/>
- Radeck, J., Kraft, K., Bartels, J., Cikovic, T., Dürr, F., Emenegger, J., Kelterborn, S., Sauer, C., Fritz, G., Gebhard, S., & Mascher, T. (2013). The *Bacillus* BioBrick Box: generation and evaluation of essential genetic building blocks for standardized work with *Bacillus subtilis*. In *Journal of Biological Engineering* (Vol. 7). <http://www.jbioleng.org/content/7/1/29>
- Rahmer, R., Heravi, K. M., & Altenbuchner, J. (2015). Construction of a super-competent *Bacillus subtilis* 168 using the PmtIA-comKS inducible cassette. *Frontiers in Microbiology*, 6(DEC). <https://doi.org/10.3389/fmicb.2015.01431>
- Rawlings, N. D., & Barrett, A. J. (2013). Metallopeptidases and Their Clans. In *Handbook of Proteolytic Enzymes* (pp. 325–370). Elsevier. <https://doi.org/10.1016/B978-0-12-382219-2.00077-6>
- Rawlings, N. D., Barrett, A. J., & Bateman, A. (2010). MEROPS: the peptidase database. *Nucleic Acids Research*, 38(suppl_1), D227–D233. <https://doi.org/10.1093/nar/gkp971>
- Rebello, S., Asok, A. K., Mundayoor, S., & Jisha, M. S. (2014). Surfactants: toxicity, remediation and green surfactants. *Environmental Chemistry Letters*, 12(2), 275–287. <https://doi.org/10.1007/s10311-014-0466-2>

- Reddy, M. R., Sathi Reddy, K., Ranjita Chouhan, Y., Bee, H., & Reddy, G. (2017). Effective feather degradation and keratinase production by *Bacillus pumilus* GRK for its application as bio-detergent additive. *Bioresource Technology*, 243, 254–263. <https://doi.org/10.1016/j.biortech.2017.06.067>
- Remaut, H., Bompard-Gilles, C., Goffin, C., Frère, J. M., & Van Beeumen, J. (2001). Structure of the *Bacillus subtilis* d-aminopeptidase DppA reveals a novel self-compartmentalizing protease. *Nature Structural Biology* 2001 8:8, 8(8), 674–678. <https://doi.org/10.1038/90380>
- Reuß, D. R., Altenbuchner, J., Mäder, U., Rath, H., Ischebeck, T., Sappa, P. K., Thürmer, A., Guérin, C., Nicolas, P., Steil, L., Zhu, B., Feussner, I., Klumpp, S., Daniel, R., Commichau, F. M., Völker, U., & Stülke, J. (2017). Large-scale reduction of the *Bacillus subtilis* genome: consequences for the transcriptional network, resource allocation, and metabolism. *Genome Res* 27(2). <https://doi.org/10.1101/gr.215293.116>
- Riffel, A., Lucas, F., Heeb, P., & Brandelli, A. (2003). Characterization of a new keratinolytic bacterium that completely degrades native feather keratin. *Archives of Microbiology*, 179(4), 258–265. <https://doi.org/10.1007/s00203-003-0525-8>
- Rigueira, L. M. B., Lana, D. A. P. D., dos Santos, D. M., Pimenta, A. M., Augusti, R., & Costa, L. M. (2016). Identification of metal-binding to proteins in seed samples using RF-HPLC-UV, GFAAS and MALDI-TOF-MS. *Food Chemistry*, 211, 910–915. <https://doi.org/10.1016/j.foodchem.2016.05.061>
- Roderick, S. L., & Matthews, B. W. (1993). Structure of the Cobalt-Dependent Methionine Aminopeptidase from *Escherichia coli*: A New Type of Proteolytic Enzyme. *Biochemistry*, 32(15), 3907–3912. https://doi.org/10.1021/BI00066A009/ASSET/BI00066A009.FP.PNG_V03
- Rodriguez, J. J., Kuncheva, L. I., & Alonso, C. J. (2006). Rotation Forest: A New Classifier Ensemble Method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(10), 1619–1630. <https://doi.org/10.1109/TPAMI.2006.211>
- Rott, E., Steinmetz, H., & Metzger, J. W. (2018). Organophosphonates: A review on environmental relevance, biodegradability and removal in wastewater treatment plants. *Science of The Total Environment*, 615, 1176–1191. <https://doi.org/10.1016/j.scitotenv.2017.09.223>
- Rowell, S., Pauptit, R. A., Tucker, A. D., Melton, R. G., Blow, D. M., & Brick, P. (1997). Crystal structure of carboxypeptidase G2, a bacterial enzyme with applications in cancer therapy. *Structure*, 5(3), 337–347. [https://doi.org/10.1016/S0969-2126\(97\)00191-3](https://doi.org/10.1016/S0969-2126(97)00191-3)

- Roy, A., Kucukural, A., & Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*, 5(4), 725–738.
<https://doi.org/10.1038/nprot.2010.5>
- Rudner, D. Z., & Losick, R. (2001). Morphological Coupling in Development. *Developmental Cell*, 1(6), 733–742. [https://doi.org/10.1016/S1534-5807\(01\)00094-6](https://doi.org/10.1016/S1534-5807(01)00094-6)
- Sambrook, J., & Russell, D. W. (2001). *Molecular Cloning: A Laboratory Manual* (3rd ed., Vol. 1). Cold Spring Harbor Laboratory Press.
- Sánchez Beeckman, M. (2021). ColorNameR. <https://CRAN.R-project.org/package=ColorNameR>
- Sauer, U., Cameron, D. C., & Bailey, J. E. (1998). Metabolic capacity of *Bacillus subtilis* for the production of purine nucleosides, riboflavin, and folic acid. *Biotechnology and Bioengineering*, 59(2), 227–238. [https://doi.org/10.1002/\(SICI\)1097-0290\(19980720\)59:2<227::AID-BIT10>3.3.CO;2-F](https://doi.org/10.1002/(SICI)1097-0290(19980720)59:2<227::AID-BIT10>3.3.CO;2-F)
- Schallmey, M., Singh, A., & Ward, O. P. (2011). Developments in the use of *Bacillus* species for industrial production. <https://doi.org/10.1139/W03-076>, 50(1), 1–17.
<https://doi.org/10.1139/W03-076>
- Scheibel, J. J. (2004). The evolution of anionic surfactant technology to meet the requirements of the laundry detergent industry. *Journal of Surfactants and Detergents*, 7(4), 319–328.
<https://doi.org/10.1007/s11743-004-0317-7>
- Schulz, R. (2009). Activation of MMP-2 as a key event in oxidative stress injury to the heart. *Frontiers in Bioscience, Volume*(14), 699. <https://doi.org/10.2741/3274>
- Secades, P., Alvarez, B., & Guijarro, J. A. (2001). Purification and Characterization of a Psychrophilic, Calcium-Induced, Growth-Phase-Dependent Metalloprotease from the Fish Pathogen *Flavobacterium psychrophilum*. *Applied and Environmental Microbiology*, 67(6), 2436–2444. <https://doi.org/10.1128/AEM.67.6.2436-2444.2001>
- Secades, P., Alvarez, B., & Guijarro, J. A. (2003). Purification and properties of a new psychrophilic metalloprotease (Fpp2) in the fish pathogen *Flavobacterium psychrophilum*. *FEMS Microbiology Letters*, 226(2), 273–279. [https://doi.org/10.1016/S0378-1097\(03\)00599-8](https://doi.org/10.1016/S0378-1097(03)00599-8)
- Sehna, D., Bittrich, S., Deshpande, M., Svobodová, R., Berka, K., Bazgier, V., Velankar, S., Burley, S. K., Koča, J., & Rose, A. S. (2021). Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Research*, 49(W1), W431–W437. <https://doi.org/10.1093/NAR/GKAB314>

- Shi, W., Punta, M., Bohon, J., Sauder, J. M., D'Mello, R., Sullivan, M., Toomey, J., Abel, D., Lippi, M., Passerini, A., Frasconi, P., Burley, S. K., Rost, B., & Chance, M. R. (2011). Characterization of metalloproteins by high-throughput X-ray absorption spectroscopy. *Genome Research*, 21(6), 898–907. <https://doi.org/10.1101/gr.115097.110>
- Shuman, S., & Glickman, M. S. (2007). Bacterial DNA repair by non-homologous end joining. *Nature Reviews Microbiology*, 5(11), 852–861. <https://doi.org/10.1038/nrmicro1768>
- Siezen, R. J., & Leunissen, J. A. M. (1997). Subtilases: The superfamily of subtilisin-like serine proteases. *Protein Science*, 6(3), 501–523. <https://doi.org/10.1002/PRO.5560060301>
- Sippl, M. J. (1990). Calculation of conformational ensembles from potentials of mean force. *Journal of Molecular Biology*, 213(4), 859–883. [https://doi.org/10.1016/S0022-2836\(05\)80269-4](https://doi.org/10.1016/S0022-2836(05)80269-4)
- Smits, W. K., Eschevins, C. C., Susanna, K. A., Bron, S., Kuipers, O. P., & Hamoen, L. W. (2005). Stripping *Bacillus*: ComK auto-stimulation is responsible for the bistable response in competence development. *Molecular Microbiology*, 56(3), 604–614. <https://doi.org/10.1111/J.1365-2958.2005.04488.X>
- Souza, P. M. de, & Magalhães, P. de O. e. (2010). Application of microbial α -amylase in industry - A review. *Brazilian Journal of Microbiology*, 41(4), 850–861. <https://doi.org/10.1590/S1517-83822010000400004>
- Spencer, J., Murphy, L. M., Connors, R., Sessions, R. B., & Gamblin, S. J. (2010). Crystal Structure of the LasA Virulence Factor from *Pseudomonas aeruginosa*: Substrate Specificity and Mechanism of M23 Metallopeptidases. *Journal of Molecular Biology*, 396(4), 908–923. <https://doi.org/10.1016/J.JMB.2009.12.021>
- Spizizen, J. (1958). Transformation Of Biochemically Deficient Strains of *Bacillus subtilis* By Deoxyribonucleate. *Proceedings of the National Academy of Sciences*, 44(10), 1072–1078. <https://doi.org/10.1073/pnas.44.10.1072>
- Springman, E. B., Angleton, E. L., Birkedal-Hansen, H., & Van Wart, H. E. (1990). Multiple modes of activation of latent human fibroblast collagenase: evidence for the role of a Cys73 active-site zinc complex in latency and a “cysteine switch” mechanism for activation. *Proceedings of the National Academy of Sciences*, 87(1), 364–368. <https://doi.org/10.1073/PNAS.87.1.364>
- Stoll, E., Weder, H.-G., & Zuber, H. (1976). Aminopeptidase II from *Bacillus stearothermophilus*. *Biochimica et Biophysica Acta (BBA) - Enzymology*, 438(1), 212–220. [https://doi.org/10.1016/0005-2744\(76\)90237-0](https://doi.org/10.1016/0005-2744(76)90237-0)

- Strauch, M., Webb, V., Spiegelman, G., & Hoch, J. A. (1990). The SpoOA protein of *Bacillus subtilis* is a repressor of the abrB gene. *Proceedings of the National Academy of Sciences*, 87(5), 1801–1805. <https://doi.org/10.1073/pnas.87.5.1801>
- Tanaka, K. (2009). The proteasome: Overview of structure and functions. *Proceedings of the Japan Academy, Series B*, 85(1), 12–36. <https://doi.org/10.2183/pjab.85.12>
- Thompson, J.D., Higgins, D.G., Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994 Nov 11;22(22):4673-80. <https://doi.org/10.1093/nar/22.22.4673>
- ThermoPred. (2011). <http://lin-group.cn/server/ThermoPredv1>
- Tian, J., Xu, Z., Long, X., Tian, Y., & Shi, B. (2019). High-expression keratinase by *Bacillus subtilis* SCK6 for enzymatic dehairing of goatskins. *International Journal of Biological Macromolecules*, 135, 119–126. <https://doi.org/10.1016/j.ijbiomac.2019.05.131>
- Titani, K., Hermodson, M. A., Ericsson, L. H., Walsh, K. A., & Neurath, H. (1972). Amino-acid Sequence of Thermolysin. *Nature New Biology*, 238(80), 35–37. <https://doi.org/10.1038/newbio238035a0>
- Tjalsma, H., Antelmann, H., Jongbloed, J. D. H., Braun, P. G., Darmon, E., Dorenbos, R., Dubois, J.-Y. F., Westers, H., Zanen, G., Quax, W. J., Kuipers, O. P., Bron, S., Hecker, M., & van Dijk, J. M. (2004). Proteomics of Protein Secretion by *Bacillus subtilis*: Separating the “Secrets” of the Secretome. *Microbiology and Molecular Biology Reviews*, 68(2), 207–233. <https://doi.org/10.1128/MMBR.68.2.207-233.2004>
- Toplak, A., Wu, B., Fusetti, F., Quaedflieg, P. J. L. M., & Janssena, D. B. (2013). Proteolysin, a novel highly thermostable and cosolvent-compatible protease from the thermophilic bacterium *Coprophthermobacter proteolyticus*. *Applied and Environmental Microbiology*, 79(18), 5625–5632. <https://doi.org/10.1128/AEM.01479-13>
- Trinh, C. T., Unrean, P., & Srienc, F. (2008). Minimal *Escherichia coli* Cell for the Most Efficient Production of Ethanol from Hexoses and Pentoses. *Applied and Environmental Microbiology*, 74(12), 3634–3643. <https://doi.org/10.1128/AEM.02708-07>
- Tu, R., Martinez, R., Prodanovic, R., Klein, M., & Schwaneber, U. (2011). A Flow Cytometry–Based Screening System for Directed Evolution of Proteases. *Journal of Biomolecular Screening*, 16(3), 285–294. <https://doi.org/10.1177/1087057110396361>

- Tu, J. V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of Clinical Epidemiology*, 49(11), 1225–1231. [https://doi.org/10.1016/S0895-4356\(96\)00002-9](https://doi.org/10.1016/S0895-4356(96)00002-9)
- Tyndall, J. D. A., Nall, T., & Fairlie, D. P. (2005). Proteases Universally Recognize Beta Strands in Their Active Sites. *Chemical Reviews*, 105(3), 973–1000. <https://doi.org/10.1021/cr040669e>
- Unilever. (2022, January 18). Building a clean green foam-production machine. <https://www.unilever.com/news/news-search/2022/building-a-clean-green-foamproduction-machine/>
- Unilever. (2023a). Personal correspondence.
- Unilever. (2023b, June 12). Unilever and Arzeda use AI to develop performance-boosting enzymes. <https://www.unilever.com/news/news-search/2023/unilever-and-arzeda-use-ai-to-develop-performanceboosting-enzymes/#:~:text=Unilever%20and%20Arzeda%20use%20AI%20to%20develop%20perform,ance%2Dboosting%20enzymes,-Published%3A%2012%20June&text=The%20new%20enzymes%2C%20used%20to,to%20deliver%20effective%20cleaning%20performance.>
- Vallee, B. L., & Auld, D. S. (1990). Zinc Coordination, Function, and Structure of Zinc Enzymes and Other Proteins. *Biochemistry*, 29(24), 5647–5659. <https://doi.org/10.1021/bi00476a001>
- van den Burg, B., & Eijssink, V. (2013). Thermolysin and Related *Bacillus* Metalloproteases. In *Handbook of Proteolytic Enzymes* (Vol. 1, pp. 540–553). Elsevier Ltd. <https://doi.org/10.1016/B978-0-12-382219-2.00111-3>
- van Tetering, G., van Diest, P., Verlaan, I., van der Wall, E., Kopan, R., & Vooijs, M. (2009). Metalloprotease ADAM10 Is Required for Notch1 Site 2 Cleavage. *Journal of Biological Chemistry*, 284(45), 31018–31027. <https://doi.org/10.1074/jbc.M109.006775>
- Van Wart, H. E., & Birkedal-Hansen, H. (1990). The cysteine switch: A principle of regulation of metalloproteinase activity with potential applicability to the entire matrix metalloproteinase gene family (collagenase/gelatinase/stromelysin/zinc enzyme). In *Proc. Natl. Acad. Sci. USA* (Vol. 87). <https://www.pnas.org>
- Veltman, O. R., Vriend, G., Berendsen, H. J. C., Van Den Burg, B., Venema, G., & Eijssink, V. G. H. (1998). A single calcium binding site is crucial for the calcium-dependent thermal stability of thermolysin-like proteases. *Biochemistry*, 37(15), 5312–5319. <https://doi.org/10.1021/bi9725879>

- Veltman, O. R., Vriend, G., Middelhoven, P. J., van den Burg, B., Venema, G., & Eijssink, V. G. H. (1996). Analysis of structural determinants of the stability of thermolysin like proteases by molecular modelling and site-directed mutagenesis. *Protein Engineering, Design and Selection*, 9(12), 1181–1189. <https://doi.org/10.1093/protein/9.12.1181>
- Verma, R., Aravind, L., Oania, R., McDonald, W. H., Yates, J. R., Koonin, E. V., & Deshaies, R. J. (2002). Role of Rpn11 metalloprotease in deubiquitination and degradation by the 26S proteasome. *Science*, 298(5593), 611–615. <https://doi.org/10.1126/science.1075898>
- Vierstra, R. (2003). The ubiquitin/26S proteasome pathway, the complex last chapter in the life of many plant proteins. *Trends in Plant Science*, 8(3), 135–142. [https://doi.org/10.1016/S1360-1385\(03\)00014-1](https://doi.org/10.1016/S1360-1385(03)00014-1)
- Vlamakis, H., Aguilar, C., Losick, R., & Kolter, R. (2008). Control of cell fate by the formation of an architecturally complex bacterial community. *Genes & Development*, 22(7), 945–953. <https://doi.org/10.1101/gad.1645008>
- Vogel, A., & May, O. (Eds.). (2019). Industrial Enzyme Applications. *Wiley*. <https://doi.org/10.1002/9783527813780>
- Wang, L. Q., & Li, C. F. (2014). Optimal subset selection of primary sequence features using the genetic algorithm for thermophilic proteins identification. *Biotechnology Letters*, 36(10), 1963–1969. <https://doi.org/10.1007/s10529-014-1577-3>
- Wang, S. L., Hsu, W. T., Liang, T. W., Yen, Y. H., & Wang, C. L. (2008). Purification and characterization of three novel keratinolytic metalloproteases produced by *Chryseobacterium indologenes* TKU014 in a shrimp shell powder medium. *Bioresource Technology*, 99(13), 5679–5686. <https://doi.org/10.1016/j.biortech.2007.10.024>
- Weber, L. (1998). Applications of genetic algorithms in molecular diversity. *Current Opinion in Chemical Biology*, 2(3), 381–385. [https://doi.org/10.1016/S1367-5931\(98\)80013-6](https://doi.org/10.1016/S1367-5931(98)80013-6)
- Weigt, M., White, R. A., Szurmant, H., Hoch, J. A., & Hwa, T. (2009). Identification of direct residue contacts in protein–protein interaction by message passing. *Proceedings of the National Academy of Sciences*, 106(1), 67–72. <https://doi.org/10.1073/pnas.0805923106>
- Westbrook, A. W., Moo-Young, M., & Chou, C. P. (2016). Development of a CRISPR-Cas9 Tool Kit for Comprehensive Engineering of *Bacillus subtilis*. *Applied and Environmental Microbiology*, 82(16), 4876–4895. <https://doi.org/10.1128/AEM.01159-16>

Westers, H. (2003). Genome Engineering Reveals Large Dispensable Regions in *Bacillus subtilis*. *Molecular Biology and Evolution*, 20(12), 2076–2090.
<https://doi.org/10.1093/molbev/msg219>

- Westers, L., Westers, H., Zanen, G., Antelmann, H., Hecker, M., Noone, D., Devine, K. M., van Dijk, J. M., & Quax, W. J. (2008). Genetic or chemical protease inhibition causes significant changes in the *Bacillus subtilis* exoproteome. *PROTEOMICS*, 8(13), 2704–2713.
<https://doi.org/10.1002/pmic.200800009>
- Worden, E. J., Padovani, C., & Martin, A. (2014). Structure of the Rpn11–Rpn8 dimer reveals mechanisms of substrate deubiquitination during proteasomal degradation. *Nature Structural & Molecular Biology* 2014 21:3, 21(3), 220–227. <https://doi.org/10.1038/nsmb.2771>
- Worden, E. J., Dong, K. C., & Martin, A. (2017). An AAA Motor-Driven Mechanical Switch in Rpn11 Controls Deubiquitination at the 26S Proteasome. *Molecular Cell*, 67(5), 799–811.e8. <https://doi.org/10.1016/J.MOLCEL.2017.07.023>
- Wu, J. W., & Chen, X. L. (2011). Extracellular metalloproteases from bacteria. *Applied Microbiology and Biotechnology* (Vol. 92, Issue 2, pp. 253–262). Springer.
<https://doi.org/10.1007/s00253-011-3532-8>
- Wu, S.-C., Ye, R., Wu, X.-C., Ng, S.-C., & Wong, S.-L. (1998). Enhanced Secretory Production of a Single-Chain Antibody Fragment from *Bacillus subtilis* by Coproduction of Molecular Chaperones. *Journal of Bacteriology*, 180(11), 2830–2835.
<https://doi.org/10.1128/JB.180.11.2830-2835.1998>
- Wu, S.-C., Yeung, J. C., Duan, Y., Ye, R., Szarka, S. J., Habibi, H. R., & Wong, S.-L. (2002). Functional Production and Characterization of a Fibrin-Specific Single-Chain Antibody Fragment from *Bacillus subtilis*: Effects of Molecular Chaperones and a Wall-Bound Protease on Antibody Fragment Production. *Applied and Environmental Microbiology*, 68(7), 3261–3269. <https://doi.org/10.1128/AEM.68.7.3261-3269.2002>
- Wu, X., Lee, W., Tran, L., & Wong, S.-L. (1991). Engineering a *Bacillus subtilis* Expression-Secretion System with a Strain Deficient in Six Extracellular Proteases. In *Journal of Bacteriology* (Vol. 173, Issue 16).
- Wu, X.-C., Ng, S.-C., Near, R. I., & Wong, S.-L. (1993). Efficient Production of a Functional Single-Chain Antidigoxin Antibody via an Engineered *Bacillus subtilis* Expression-Secretion System. *Nature Biotechnology*, 11(1), 71–76. <https://doi.org/10.1038/nbt0193-71>
- Xu, Q., Rawlings, N. D., Farr, C. L., Chiu, H.-J., & Grant, J. C. (2011). Structural and Sequence Analysis of Imelysin-Like Proteins Implicated in Bacterial Iron Uptake. *PLoS ONE*, 6(7), 21875. <https://doi.org/10.1371/journal.pone.0021875>

- Yao, T., Song, L., Xu, W., DeMartino, G. N., Florens, L., Swanson, S. K., Washburn, M. P., Conaway, R. C., Conaway, J. W., & Cohen, R. E. (2006). Proteasome recruitment and activation of the Uch37 deubiquitinating enzyme by Adrm1. *Nature Cell Biology* 2006 8:9, 8(9), 994–1002. <https://doi.org/10.1038/NCB1460>
- Yaron, A., & Mlynar, D. (1968). Aminopeptidase-P. *Biochemical and Biophysical Research Communications*, 32(4), 658–663. [https://doi.org/10.1016/0006-291X\(68\)90289-1](https://doi.org/10.1016/0006-291X(68)90289-1)
- Yiallourous, I., Große Berkhoff, E., & Stöcker, W. (2000). The roles of Glu93 and Tyr149 in astacin-like zinc peptidases. *FEBS Letters*, 484(3), 224–228. [https://doi.org/10.1016/S0014-5793\(00\)02163-3](https://doi.org/10.1016/S0014-5793(00)02163-3)
- Yu, X. J., Chi, Z., Wang, F., Li, J., Chi, Z. M., & Madzak, C. (2013). Expression of the acid protease gene from *Saccharomycopsis fibuligera* in the marine-derived *Yarrowia lipolytica* for both milk clotting and single cell protein production. *Applied Biochemistry and Biotechnology*, 169(7), 1993–2003. <https://doi.org/10.1007/s12010-013-0118-1>
- YU, Y., ZHAO, J., & Bayly, A. E. (2008). Development of Surfactants and Builders in Detergent Formulations. *Chinese Journal of Chemical Engineering*, 16(4), 517–527. [https://doi.org/10.1016/S1004-9541\(08\)60115-9](https://doi.org/10.1016/S1004-9541(08)60115-9)
- Zaghloul, T. I., Embaby, A. M., & Elmahdy, A. R. (2011). Biodegradation of chicken feathers waste directed by *Bacillus subtilis* recombinant cells: Scaling up in a laboratory scale fermentor. *Bioresource Technology*, 102(3), 2387–2393. <https://doi.org/10.1016/j.biortech.2010.10.106>
- Zanen, G., Antelmann, H., Meima, R., Jongbloed, J. D. H., Kolkman, M., Hecker, M., van Dijl, J. M., & Quax, W. J. (2006). Proteomic dissection of potential signal recognition particle dependence in protein secretion by *Bacillus subtilis*. *PROTEOMICS*, 6(12), 3636–3648. <https://doi.org/10.1002/pmic.200500560>
- Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., Volz, S. E., Joung, J., van der Oost, J., Regev, A., Koonin, E. V., & Zhang, F. (2015). Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell*, 163(3), 759–771. <https://doi.org/10.1016/j.cell.2015.09.038>
- Zhang, G., & Fang, B. (2006). Support Vector Machine for Discrimination of Thermophilic and Mesophilic Proteins Based on Amino Acid Composition. *Protein & Peptide Letters*, 13(10), 965–970. <https://doi.org/10.2174/092986606778777560>

- Zhang, H., Li, H., Lang, D. A., Xu, H., & Zhu, H. (2018). Purification, characterization and potential applications of a novel halotolerant metalloprotease from marine bacterium *Vibrio* sp. LA-05. *Journal of Chemical Technology and Biotechnology*, 93(12), 3627–3637.
<https://doi.org/10.1002/jctb.5744>
- Zhang, H., Li, H., Liu, H., Lang, D. A., Xu, H., & Zhu, H. (2019). The application of a halotolerant metalloprotease from marine bacterium *Vibrio* sp. LA-05 in liquid detergent formulations. *International Biodeterioration and Biodegradation*, 142, 18–25.
<https://doi.org/10.1016/j.ibiod.2019.04.006>
- Zhang, J., Zhang, Y., Li, W., Li, X., & Lian, X. (2014). Optimizing Detergent Formulation with Enzymes. *Journal of Surfactants and Detergents*, 17(6), 1059–1067.
<https://doi.org/10.1007/s11743-014-1626-x>
- Zhang, K., Duan, X., & Wu, J. (2016). Multigene disruption in undomesticated *Bacillus subtilis* ATCC 6051a using the CRISPR/Cas9 system. *Scientific Reports*, 6(1), 27943.
<https://doi.org/10.1038/srep27943>
- Zhang, X., Li, Y., & Li, Y. (2009). Evaluating signal peptide prediction methods for Gram-positive bacteria. *Biologia*, 64(4), 655–659. <https://doi.org/10.2478/s11756-009-0118-3>
- Zhang, X. Z., & Zhang, Y. H. P. (2011). Simple, fast and high-efficiency transformation system for directed evolution of cellulase in *Bacillus subtilis*. *Microbial Biotechnology*, 4(1), 98–105.
<https://doi.org/10.1111/j.1751-7915.2010.00230.x>
- Zhu, X., Zhao, D., Qiu, H., Fan, F., Man, S., Bi, C., & Zhang, X. (2017). The CRISPR/Cas9-facilitated multiplex pathway optimization (CFPO) technique and its application to improve the *Escherichia coli* xylose utilization pathway. *Metabolic Engineering*, 43, 37–45.
<https://doi.org/10.1016/j.ymben.2017.08.003>