



# University of Reading

## Altering the thermal lability properties of industrial proteases

PhD thesis

Mr. Gediminas Baltulionis

Schools of Food Sciences and Biological Sciences

Supervisors: Drs. D. Charalampopoulos and K.A. Watson

in collaboration with Biocatalysts Ltd

(a BBSRC Case studentship)

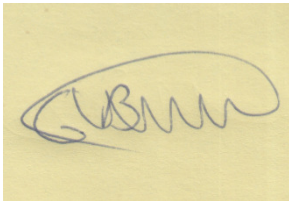
September 2017

*I dedicate this work to my Mum, Dad,  
brothers, sister, wife and my child*

## **Declaration**

I, Gediminas Baltulionis, confirm that this thesis is based upon my own work and that the use of all materials from other sources have been properly and fully acknowledged. This thesis has not been presented previously, in part or in whole, for any other degree.

**Gediminas Baltulionis**

A handwritten signature in blue ink on a light yellow background. The signature is stylized, starting with a large 'G' and 'B' followed by several loops and a final flourish.

**Reading, September 2017**

## **Acknowledgements**

I would like to express my sincere gratitude to supervisors Professor Dimitris Charalampopoulos and Dr Kimberly Ann Watson for providing an opportunity to work on this interesting and challenging project, for the whole academic, research and emotional support and guidance throughout these years. Their excellent mentoring allowed to excel in this project and develop into an independent thinker and problem solver.

I would like to also thank all the experienced scientists from Biocatalysts for their immense contribution to this challenging project. In particular, my big gratitude goes to Dr Mark Blight, Dr Aelig Robin, Dr Maeve O'Neil, Dr Andrew Ellis, Dr Rhian White, Dr Denise Gallagher, Dr Sheryl Philip from Biocatalysts organisation for their invaluable support, insightful ideas and research directions.

This work was delivered in collaboration with several other research institutions and laboratories. Special thanks go to the research scientists, Dr Raymond Owens, Dr Louise Bird, Dr Joanne Nettleship, Anil Verma, from the Oxford Protein Production Facility group. I wish to acknowledge the assistance provided by Dr Isabel De Moraes and Ms Maria Rosa from MPL laboratory at Diamond Light Source synchrotron. Also, I would like to thank all the beamline scientists at the macromolecular crystallography beamlines I-02, I03, I04 and I04-1 and Dr Edward Lowe for providing sufficient amount of synchrotron time for each data collection.

I would like to express additional gratitude to academic and technical staff at Reading University for various contributions at different stages of this project: Professor Simon Andrews, Dr David Leake, Professor Angela Clerk, Dr Stephen Fuller, Mrs Julia Janes, Mrs Young Wha Lee. I wish to thank Dr Afroditi Chatzifragkou and Dr Sameer Khalil Ghawifor their help in the last part of my experimental work.

My special appreciation goes to former PhD students Dr Francis Atanu, Dr Katerina Lazidou, Dr Oliver Ellingham, Ms Amy Elizabeth Danson and Ms Dannielle Kydd-Sinclair for the delivered support and friendship at different stages of this journey.

Finally, I wish to thank my dear wife, Julija and my lovely little child, Benas for their belief, patience, lasting love and emotional support in these years.

**Gediminas Baltulionis,**

**Reading, September 2017**

# Table of contents

<b>CHAPTER 1 INTRODUCTION.....</b>	<b>1</b>
1.1 Extremophiles.....	2
1.2 Protein synthesis and folding in psychrophiles.....	3
1.3 Biocatalysis at low temperatures .....	4
1.4 Stability of extremophilic proteins.....	7
1.5 Structural characteristics of psychrophilic enzymes .....	10
1.5.1 Activity-flexibility-thermolability relationships .....	12
1.5.2 Activity-stability trade-off and active site flexibility in psychrophilic enzymes .....	13
1.5.3 The folding funnel model .....	14
1.6 Cold-adapted thermolabile proteases.....	16
1.6.1 Engineering low-temperature activity and thermolability in hydrolases.....	17
1.6.2 Identified strategies to achieve thermolability and cold-adaptivity.....	19
1.7 Industrial applications of cold-adapted enzymes .....	20
1.7.1 Industrial application of cold-adapted proteases .....	21
1.8 Production of protein hydrolysates .....	23
1.8.1 Protein enzymatic hydrolysis in functional food development .....	23
1.8.2 Enzymatic hydrolysis of whey protein .....	23
1.9 Importance of this project.....	26
1.10 Target enzymes in this project.....	27
1.10.1 Subtilisin Carlsberg (sC).....	27
1.10.2 Leucine Aminopeptidase (LapA).....	27
1.11 Aims and objectives .....	29
<b>CHAPTER 2 MATERIALS AND METHODS.....</b>	<b>30</b>
2.1 Sequence and structure bioinformatics.....	31
2.1.1 Subtilisin Carlsberg (sC) from Bacillus licheniformis .....	31
2.1.2 Leucine aminopeptidases from Aspergillus oryzae.....	31

<b>2.2 DNA manipulation .....</b>	<b>32</b>
2.2.1 Polymerase chain reaction.....	32
2.2.2 Plasmid DNA purification.....	32
2.2.3 DNA extraction from agarose gel.....	32
2.2.4 DNA restriction digest.....	33
2.2.5 In-Fusion molecular cloning.....	33
2.2.6 Quickchange site-directed mutagenesis protocol .....	34
<b>2.3 Electrophoresis .....</b>	<b>34</b>
2.3.1 Agarose gel electrophoresis.....	34
2.3.2 SDS – PAGE gel electrophoresis .....	35
2.3.3 Western Blot Analysis.....	35
<b>2.4 Quantification methods .....</b>	<b>36</b>
2.4.1 Quantification of recombinant DNA .....	36
2.4.2 Quantification of recombinant LapA.....	36
<b>2.5 Microbiology methods .....</b>	<b>36</b>
2.5.1 Bacterial transformation .....	36
2.5.2 Glycerol stocks .....	37
2.5.3 Preparation of competent cells .....	37
<b>2.6 Recombinant Protein Expression in E.coli .....</b>	<b>37</b>
2.6.1 pOPIN expression vectors .....	37
2.6.2 Bacterial expression.....	37
2.6.3 Small-scale expression .....	38
2.6.4 Disruption of bacterial cells.....	39
2.6.5 Periplasmic extraction .....	39
<b>2.7 Protein purification.....</b>	<b>39</b>
2.7.1 Immobilised metal affinity chromatography (IMAC).....	39
2.7.2 Purification and refolding of refLapA protein from E.coli inclusion bodies .....	40
2.7.3 Native LapA (flavLapA) purification from Flavourzyme.....	41
2.7.4 Protein concentration.....	41
<b>2.8 Protein crystallisation .....</b>	<b>41</b>
2.8.1 High throughput crystallisation screening.....	41
2.8.2 24-well plate crystallisation screening .....	42
2.8.1 Data Collection, Structure Determination & Refinement .....	44
<b>2.9 Protein thermal and spectroscopic analysis.....</b>	<b>44</b>

2.9.1	Differential Scanning Fluorimetry (DSF) measurements.....	44
2.9.2	Differential Scanning Calorimetry (DSC) measurements .....	45
2.9.3	Circular Dichroism (CD) spectroscopy of LapA proteins.....	45
<b>2.10</b>	<b>LapA enzymatic assays.....</b>	<b>45</b>
2.10.1	Lap enzymatic assay – 3 mL format .....	45
2.10.2	Lap enzymatic assay – 300µL microplate assay .....	46
2.10.3	LapA thermolability assay .....	46
2.10.4	LapA enzyme kinetic assay.....	47
<b>2.11</b>	<b>Industrial pasteurisation for LapA protein inactivation.....</b>	<b>47</b>
<b>2.12</b>	<b>Other analyses .....</b>	<b>47</b>
2.12.1	Deglycosylation with PNGase F .....	47
2.12.2	DNA sequencing .....	48
2.12.3	Mass Spectrometry.....	48
<b>2.13</b>	<b>Recombinant cloning and expression of LapA in eukaryotic host – Pichia pastoris... 48</b>	
2.13.1	Strains, plasmids and culture conditions .....	48
2.13.2	Construction of recombinant expression vectors and transformations .....	48
2.13.3	Over-expression of recombinant LapA in P.pastoris .....	49
2.13.4	Purification of recombinant LapA from P.pastoris supernatant.....	50
2.13.5	Heterologous protein expression in Pichia pastoris - experimental workflow.....	53

## **CHAPTER 3 FLAVOURZYME – ISOLATION OF NATIVE ASPERGILLUS ORYZAE LEUCINE AMINOPEPTIDASE A .....55**

<b>3.1</b>	<b>Introduction.....</b>	<b>56</b>
<b>3.2</b>	<b>Isolation of native Aspergillus aminopeptidases from Flavourzyme.....</b>	<b>57</b>
3.2.1	Isolation of flavLapA .....	59
3.2.2	Isolation of Lap2.....	60
3.2.3	Initial challenges in flavLapA and Lap2 crystallisation trials.....	62
3.2.4	Crystallisation of flavLapA .....	65
3.2.5	Data collection and refinement of flavLapA crystal structure .....	66
3.2.6	Evidence of primary thermal unfolding states in flavLapA .....	67
<b>3.3</b>	<b>Isolation of NP1 and AP1 proteases from Flavourzyme.....</b>	<b>71</b>
3.3.1	Improved isolation protocol of AP1 and NP1 proteases .....	72

## **CHAPTER 4 OPTIMISATION OF HETEROLOGOUS EXPRESSION OF TARGET PROTEINS IN PROKARYOTIC AND EUKARYOTIC HOSTS ..... 75**

<b>4.1</b>	<b>Escherichia coli as host for recombinant protein expression.....</b>	<b>76</b>
4.1.1	Selected endoprotease – Subtilisin Carlsberg (sC).....	76
4.1.2	Selected exoprotease – leucine aminopeptidase A (LapA) .....	86
<b>4.2</b>	<b>Concluding remarks from target protein expression in E.coli.....</b>	<b>97</b>
<b>4.3</b>	<b>Pichia pastoris as host for recombinant protein expression.....</b>	<b>98</b>
4.3.1	Initial attempts of wild-type LapA over-expression.....	98
4.3.2	Ab initio development of wtproLapA and wtΔproLapA P. pastoris expression strains	107
4.3.3	Downstream processing and analysis of expressed LapA proteins .....	115
4.3.4	Optimization of purification protocol of activated wtLapA t.....	123
4.3.5	Purification of wtproLapA precursor from P. pastoris supernatant for crystallization .	125
4.3.6	Considerations of LapA protein purification from P. pastoris supernatant.....	129
4.3.7	Exploring the cleavage of propeptide in wtproLapA proenzyme.....	133

## **CHAPTER 5 STRUCTURAL STUDIES OF RECOMBINANT WILD – TYPE LAPA: MATURATION, INHIBITION AND SUBSTRATE BINDING..... 138**

<b>5.1</b>	<b>Crystal structures of LapA precursor and mature enzyme: role of propeptide in inhibition and as intramolecular chaperone.....</b>	<b>139</b>
5.1.1	Introduction .....	139
5.1.2	Materials and Methods .....	142
5.1.3	Results .....	147
<b>5.2</b>	<b>Crystal structure of LapA with natural substrate – Leucine (wtLapA-Leu) .....</b>	<b>175</b>
<b>5.3</b>	<b>Discussion.....</b>	<b>177</b>

## **CHAPTER 6 STRUCTURAL AND FUNCTIONAL STUDIES OF LAPA MUTANTS: CHARACTERISATION OF THERMOLABILITY ..... 182**

<b>6.1</b>	<b>LapA thermolabile mutants M1 – M6: 1st round mutagenesis .....</b>	<b>183</b>
6.1.1	Rationale and structural bioinformatics.....	183
6.1.2	Recombinant cloning and expression of M1 – M6 mutants .....	187

6.1.3	Purification and AP1 – mediated enzymatic activation of M1 – M6 mutant precursors	191
<b>6.2</b>	<b>LapA thermolabile mutants M7 – M13: 2nd round mutagenesis.....</b>	<b>194</b>
6.2.1	Rationale and structural bioinformatics.....	194
6.2.2	Recombinant cloning and expression of M7 – M13 mutants.....	198
6.2.3	M7 – M13 recombinant expression in YNB synthetic media .....	200
6.2.4	Purification and enzymatic activation of M7 – M13 mutant precursors .....	202
<b>6.3</b>	<b>Crystallisation of LapA thermolabile mutant structures .....</b>	<b>204</b>
<b>6.4</b>	<b>Analysis and comparison of wild-type LapA and thermolabile mutant structures .....</b>	<b>206</b>
6.4.1	Comparison of LapA precursor (wtproLapA) and M7, M8, M12 structures .....	206
<b>6.5</b>	<b>Functional studies of LapA .....</b>	<b>214</b>
6.5.1	High-temperature inactivation assays for M1 – M6 thermolabile mutants .....	214
6.5.2	High-temperature inactivation assays for M7 – M13 thermolability mutants.....	222
6.5.3	Comparison of the most thermolabile mutants by additional parameters .....	234
6.5.4	Stability of recombinant LapA proteins at the room temperature (RT) .....	237
6.5.5	Exploring low-temperature activity of recombinant LapA proteins .....	239
6.5.6	Industrial pasteurisation for LapA protein inactivation.....	241
<b>6.6</b>	<b>Kinetic characterisation of wild-type LapA and mutant enzymes .....</b>	<b>244</b>
6.6.1	Determination of kinetic parameters of LapA hydrolytic reaction.....	244
<b>6.7</b>	<b>Thermodynamic characterisation of wild-type LapA and mutant enzymes .....</b>	<b>249</b>
6.7.1	LapA characterisation by differential scanning fluorimetry (DSF).....	249
6.7.2	LapA characterisation by differential scanning calorimetry (DSC).....	253
 <b>CHAPTER 7 EXPLORATION OF THE ALTERNATIVE ROUTES TO RECOMBINANT LAPA SECRETION IN P.PASTORIS .....</b>		<b>256</b>
<b>7.1</b>	<b>Introduction .....</b>	<b>257</b>
<b>7.2</b>	<b>LapA trans – acting propeptide .....</b>	<b>257</b>
<b>7.3</b>	<b>Kex2/Ste13 – mediated cleavage of LapA .....</b>	<b>260</b>
7.3.1	Purification of <sub>EA</sub> LapA protein .....	264
<b>7.4</b>	<b>MAT<math>\alpha</math>-mediated secretion of mature LapA .....</b>	<b>265</b>
7.4.1	Background rationale.....	265

7.4.2	MAT $\alpha$ -LapA cloning and expression in <i>P. pastoris</i> .....	267
7.5	Considerations for pilot scale production of LapA .....	271
<b>CHAPTER 8 AMINOPEPTIDASE FROM GLACIOZYMA</b>		
<b>ANTARCTICA (GAP): A PSYCHROPHILIC HOMOLOGUE OF LAP</b>		
.....		<b>274</b>
8.1	Introduction and preliminary bioinformatics study .....	275
8.2	Recombinant cloning and expression of GAP in <i>P.pastoris</i> .....	276
8.3	Conclusions and future possibilities .....	280
<b>CHAPTER 9 GENERAL DISCUSSION AND FUTURE DIRECTIONS</b>		
.....		<b>282</b>
9.1	Industrial relevance of cold-adapted proteases: the driving force .....	283
9.2	Project highlights: from inception to completion.....	285
9.3	Isolation of native peptidases from Flavourzyme.....	290
9.4	Exploration of suitable routes to recombinant production of active proteases.....	292
9.4.1	Secretion and post – translational activation of recombinant LapA precursor.....	294
9.5	LapA proenzyme structure: elucidation of the mechanism of prodomain-mediated inhibition and its role as intramolecular chaperone .....	296
9.6	Achieving thermolability .....	297
9.7	Future structure-to-function evolution of LapA thermolability and cold adaptivity ....	302
9.8	Cold adaptation of a psychrophilic homologue – GAP aminopeptidase.....	303
9.9	Concluding remarks.....	304
<b>APPENDICES.....</b>		<b>306</b>
<b>LITERATURE REFERENCES .....</b>		<b>319</b>

## List of abbreviations

<b>aa</b>	Amino acid
<b>FPLC</b>	Fast protein liquid chromatography
<b>IPTG</b>	Isopropyl- $\beta$ -D-thiogalactopyranoside
<b>MW</b>	Molecular weight
<b>NCBI</b>	National center for biotechnology information
<b>OD</b>	Optical density
<b>LPNA</b>	Leucine <i>p</i> -nitroanilide (LapA substrate)
<b>SDS-PAGE</b>	Sodium dodecyl sulfate polyacrilamide gel electrophoresis
<b>PCR</b>	Polymerase chain reaction
<b>Zeo<sup>R</sup></b>	Zeocin resistance marker
<b>CCP4</b>	Collaborative Computational Project Number 4
<b>PEG</b>	Polyethylene glycol
<b>OPPF</b>	Oxford protein production facility
<b>DLS</b>	Diamond Light Source
<b>LapA</b>	Leucine aminopeptidase from <i>Aspergillus oryzae</i>
<b><i>flav</i>LapA</b>	Native LapA purified from Flavourzyme
<b><i>wt</i>LapA</b>	Recombinant wild – type mature active LapA
<b><u><i>wtpro</i></u>LapA</b>	Inactive recombinant LapA precursor
<b><sub>EA</sub>LapA</b>	Recombinant LapA from Kex2 cleavage construct
<b><i>Ma</i>-LapA</b>	Recombinant LapA from MAT $\alpha$ cleavage construct
<b>GAP</b>	Leucine aminopeptidase from <i>Glaciozyma antarctica</i>
<b>AOX1</b>	Alcohol oxidase 1
<b>P<sub>AOX</sub></b>	Alcohol oxidase 1 promotor
<b>His-tag</b>	Histidine tag
<b>MOPS</b>	3-morpholinopropane-1-sulfonic acid

<b>IMAC</b>	Immobilized-Metal Affinity Chromatography
<b>AEX</b>	Anion exchange chromatography
<b>CAX</b>	Cation exchange chromatography
<b>SEC</b>	Size exclusion chromatography
<b>BLAST</b>	Basic local alignment search tool
<b>HTP</b>	High throughput
<b>Ni-NTA</b>	Nickel-Nitrilotriacetic acid
<b>Tris</b>	2-Amino-2-(hydroxymethyl)-1,3-propanediol
<b>RMSD</b>	Root mean square deviation
<b>FPLC</b>	Fast protein liquid chromatography
<b>PISA</b>	Protein interfaces, surfaces and assemblies
<b>PDB</b>	Protein data bank
<b>IntFOLD</b>	Integrated protein structure and function prediction server
<b>MeOH</b>	methanol
<b>EtOH</b>	ethanol
<b>NaCl</b>	Sodium chloride
<b>BMGY</b>	buffered minimal glycerol-complex medium
<b>BMMY</b>	buffered minimal methanol-complex medium
<b>DMSO</b>	dimethyl sulfoxide
<b>ER</b>	endoplasmic reticulum
<b>MAT<math>\alpha</math></b>	$\alpha$ -mating factor secretion signal
<b>YNB</b>	yeast nitrogen base
<b>YPD</b>	yeast peptone dextrose
<b>His6</b>	hexahistidine
<b>pI</b>	Isoelectric point
<b>NaOH</b>	Sodium Hydroxide

<b>UV</b>	Ultraviolet
<b>w/v</b>	weight/volume
<b>AmSO<sub>4</sub></b>	Ammonium sulphate
<b>WPI</b>	Whey protein isolate
<b>WPC</b>	Whey protein concentrate
<b>WPH</b>	Whey protein hydrolysate

## List of Figures

<b>Figure 1.1</b> Optimisation of activity by decreasing substrate affinity in cold-adapted enzymes. ....	5
<b>Figure 1.2</b> Temperature dependence on the activity of psychrophilic (blue) and mesophilic (black) protein homologues.. ....	6
<b>Figure 1.3</b> Thermal unfolding of extremophilic proteins.....	8
<b>Figure 1.4</b> Permeability of homologues of $\alpha$ -amylases at room temperature, as estimated by fluorescence quenching.....	9
<b>Figure 1.5</b> Structural superimposition of psychrophilic (blue) and mesophilic (red) homologous $\alpha$ -amylases.....	11
<b>Figure 1.6</b> Folding funnel model of extremophilic proteins. ....	15
<b>Figure 1.7</b> Structure of Subtilisin Carlsberg (pdb:3UNX).....	17
<b>Figure 1.8</b> Hydrolysis of whey protein by the simultaneous action of endoprotease and exopeptidase enzymes.....	24
<b>Figure 1.9</b> Crystal structure of leucine aminopeptidase from <i>Vibrio proteolytica</i> (AAP) (pdb: 1AMP). ....	28
<b>Figure 2.1</b> Schematic of recombinant LapA purification workflow. ....	52
<b>Figure 3.1</b> Separation of Flavourzyme proteins by size using a gel filtration column (HiLoad 16/60 200pg, GE Healthcare). ....	57
<b>Figure 3.2</b> Analysis of Flavourzyme purification and 3-step isolation of <i>flav</i> LapA.....	58
<b>Figure 3.3</b> UV trace (left) and SDS-PAGE gel (right) of AEX fractions from Fraction II purification.....	61
<b>Figure 3.4</b> Crystals of the TAKA alpha-amylase from <i>Aspergillus oryzae</i> (Flavourzyme).. ....	62
<b>Figure 3.5</b> Analysis of LapA and Lap2 deglycosylation.....	64
<b>Figure 3.6</b> Needle – shaped crystals of <i>flav</i> LapA. ....	65

<b>Figure 3.7</b> Crystal structure of <i>flavLapA</i> at 2.71 Å resolution. Four protein monomers (A – D) were found in the asymmetric unit. ....	66
<b>Figure 3.8</b> Superimposition of different asymmetric unit monomers of <i>flavLapA</i> .....	68
<b>Figure 3.9</b> Superimposition of Monomer B ( <i>flavLapA</i> ) and <i>wtLapA</i> structures. ....	69
<b>Figure 3.10</b> SDS-PAGE of AEX purification of proteins in Flavourzyme <i>fraction IV</i> . ....	71
<b>Figure 3.11</b> Purification of Flavourzyme neutral proteases by CAX.....	73
<b>Figure 3.12</b> AEX purification analysis of 50mM NaCl CAX Flavourzyme fraction. ....	74
<b>Figure 4.1</b> The amino acid sequence for sC in the pOPINP vector.. ....	81
<b>Figure 4.2</b> Subtilisin Carlsberg expression vector constructed, employing InFusion cloning as described in section 2.2.5. ....	82
<b>Figure 4.3</b> Recombinant sC small scale expression analysis by SDS-PAGE. ....	83
<b>Figure 4.4</b> Recombinant sC small scale expression analysis by Western Blot.. ....	84
<b>Figure 4.5</b> Western Blot analysis of recombinant expression and purification of the sC-pOPINE construct. ....	85
<b>Figure 4.6</b> SDS-PAGE gel of LapA and Lap2 mutant expression trial. ....	88
<b>Figure 4.7</b> Assessing expression of soluble recombinant LapA (20.3_10).....	90
<b>Figure 4.8</b> Purification of recombinant 20.3_10 LapA by IMAC (A) and IEX (B). ....	91
<b>Figure 4.9</b> Purification of recombinant LapA from inclusion bodies by IMAC (A) and SEC (B). ....	94
<b>Figure 4.10</b> Recombinant pro-LapA expression analysis. ....	95
<b>Figure 4.11</b> Western Blot analysis of recombinant expression and purification of the pro-LapA-pOPINE construct.....	96
<b>Figure 4.12</b> 1% DNA agarose gel of SacI-HF linearized pD-npro-LapA plasmid DNA. ....	100
<b>Figure 4.13</b> 1% DNA agarose gel analysis of PCR amplicons of pD-npro-LapA transformants.....	100
<b>Figure 4.14</b> SDS-page analysis of <i>wtproLapA</i> supernatants, following 120h induction..	102

<b>Figure 4.15</b> SDS-PAGE analysis of <i>wtproLapA</i> supernatants and subsequent AEX purification.....	103
<b>Figure 4.16</b> 1% DNA agarose gel of PCR amplicons of pD- $\Delta$ pro-LapA <i>E.coli</i> clones..	104
<b>Figure 4.17</b> 1.5% DNA agarose gel analysis of <i>HinfI</i> -digested pD-npro-LapA-His6 plasmid DNA clones.....	105
<b>Figure 4.18</b> 1% DNA agarose gel of <i>SacI</i> -HF linearized pD- $\Delta$ pro-LapA and pD-npro-LapA-His6 plasmid DNA.....	106
<b>Figure 4.19</b> 1% DNA agarose gel of PCR amplicons of pD-npro-LapA <i>P.pastoris</i> transformants.....	108
<b>Figure 4.20</b> 1% DNA agarose gel of PCR amplicons of pD- $\Delta$ pro-LapA <i>P. pastoris</i> transformants.....	108
<b>Figure 4.21</b> SDS-PAGE of <i>wtproLapA</i> and <i>wt<math>\Delta</math>proLapA</i> supernatants, following 120h MeOH induction. ....	111
<b>Figure 4.22</b> SDS-PAGE analysis of <i>wtLapA</i> Clones 1 – 10 supernatant, following 120h MeOH induction. ....	114
<b>Figure 4.23</b> Enzymatic activation of <i>wtproLapA</i> protein. ....	115
<b>Figure 4.24</b> Enzymatic activation of <i>wtproLapA</i> protein. ....	116
<b>Figure 4.25</b> AEX purification and deglycosylation of activated <i>wt<math>\Delta</math>proLapA</i> and <i>wtLapA</i> proteins.....	118
<b>Figure 4.26</b> N – terminal sequencing results for the different activated forms of <i>wtproLapA</i> protein. ....	120
<b>Figure 4.27</b> Pairwise sequence alignment of expected (top) and observed <i>wtLapA</i> (bottom) sequences. ....	122
<b>Figure 4.28</b> <i>wtLapA</i> Clone 12 purification and analysis.....	125
<b>Figure 4.29</b> <i>wtproLapA</i> Clone 16 SEC purification and SDS-PAGE. ....	126
<b>Figure 4.30</b> <i>wtproLapA</i> Clone 5 purification and SDS-PAGE analysis. ....	128

<b>Figure 4.31</b> Wild-type LapA purification on SEC column following AmSO <sub>4</sub> precipitation..	130
<b>Figure 4.32</b> Standard peak elution pattern for LapA protein from 5 mL Q HP AEX column	131
<b>Figure 4.33</b> Typical elution pattern of LapA protein from MonoQ AEX column.....	132
<b>Figure 4.34</b> SDS-PAGE analysis of putative LapA autocatalytic maturation. ....	133
<b>Figure 4.35</b> SDS-PAGE analysis of <i>wtproLapA</i> Clone 16 enzymatic treatment with proteases.....	135
<b>Figure 4.36</b> Analysis of enzymatic processing of different LapA zymogen proteins.....	136
<b>Figure 5.1.</b> Multiple, full open reading frame alignment of homologous leucine aminopeptidases.....	150
<b>Figure 5.2</b> Purification of Lap proteins.....	152
<b>Figure 5.3</b> CD spectra and deduced folded state of LapA proteins. ....	154
<b>Figure 5.4</b> Crystal structure of LapA catalytic domain ( <i>wtLapA</i> ) determined to 1.97Å resolution ( <b>A</b> ) and a detailed view of the active site ( <b>B</b> ). ....	156
<b>Figure 5.5</b> Crystal structure of LapA precursor ( <i>wtproLapA</i> ) determined to 1.61 Å resolution.....	158
<b>Figure 5.6</b> Contacts between the catalytic (left) and prodomain (right). ....	161
<b>Figure 5.7</b> Superimposition of LapA precursor (cyan), mature LapA (yellow) and AAP-LLL (PDBid: 2iq6) (wheat) crystal structures. ....	163
<b>Figure 5.8</b> Differences in the secondary structure of LapA proenzyme and mature proteins. Purple boxes indicate the differences in SSE.....	164
<b>Figure 5.9</b> Comparison of residue B-factor distribution in LapA proenzyme and activated protein. ....	165
<b>Figure 5.10</b> Superimposition of <i>wtproLapA</i> (cyan) and mature <i>wtLapA</i> (yellow) Ca traces. ....	166

<b>Figure 5.11</b> Detailed analysis of the Leu substrate binding modes in <i>wtproLapA</i> (cyan) and mature <i>wtLapA</i> (yellow) structures.....	169
<b>Figure 5.12</b> Mechanism of propeptide inhibition in LapA precursor ( <i>wtproLapA</i> ). .....	171
<b>Figure 5.13</b> Multiple hydrophobic contacts of prodomain Leu49' (orange). .....	174
<b>Figure 5.14</b> Comparison of the substrate binding pockets in <i>wtLapA</i> -Leu and AAP-LLL crystal structures. ....	175
<b>Figure 6.1</b> Analysis of the secretion profiles of the M1 – M6 LapA thermolability mutants. ....	188
<b>Figure 6.2</b> Recombinant expression data from large – scale trial of wild-type and M1-M6 LapA mutants.....	190
<b>Figure 6.3</b> SDS-PAGE analysis of the purification of M5 LapA mutant. ....	193
<b>Figure 6.4</b> Proposed LapA thermolability mutations mapped onto the <i>wtLapA</i> crystal structure.....	196
<b>Figure 6.5</b> SDS-PAGE analysis of the secretion profiles of M7 – M13 LapA mutants. ....	199
<b>Figure 6.6</b> SDS-PAGE analysis of LapA protein supernatants, following 96h induction with MeOH. ....	201
<b>Figure 6.7</b> SDS-PAGE analysis for purification and enzymatic activation of LapA mutants. ....	203
<b>Figure 6.8</b> Crystal images of LapA thermolability mutants.....	205
<b>Figure 6.9</b> Superimposition of LapA wild-type and mutant proenzyme structures.....	207
<b>Figure 6.10</b> Detailed structural comparison of LapA wild-type and thermolabile mutant structures. ....	211
<b>Figure 6.11</b> Comparison of overall B factor representation for <i>wtproLapA</i> and M12 structures. ....	212
<b>Figure 6.12</b> Temperature inactivation profiles of LapA variants.....	218
<b>Figure 6.13</b> Comparison of remaining percentage activity of LapA proteins at 60 °C, within the first 60 minutes of incubation. ....	220

<b>Figure 6.14</b> A single exponential rate model fitted to LapA protein inactivation data at 60 °C by non-linear regression. ....	220
<b>Figure 6.15</b> Comparison of remaining percentage activity of LapA proteins, at 50 °C, within the first 90 minutes of incubation. ....	223
<b>Figure 6.16</b> Comparison of remaining % activity of LapA proteins at 60 °C within first 90 minutes of incubation.....	225
<b>Figure 6.17</b> A single exponential rate model fitted to LapA protein inactivation data, at 60 °C by non-linear regression. ....	226
<b>Figure 6.18</b> Plot of residuals linked to LapA inactivation data fit in <b>Figure 6.17</b> .....	229
<b>Figure 6.19</b> Overlay of observed versus predicted remaining activity values relative to the LapA inactivation data fit in <b>Figure 6.17</b> . ....	229
<b>Figure 6.20</b> Comparison of determined LapA specific activity values, following 60 min incubation at increasing temperatures.....	230
<b>Figure 6.21</b> Comparison of remaining % activity of LapA proteins at 60 °C, within the first 90 minutes of incubation.....	232
<b>Figure 6.22</b> Visual representation of the determined specific activity differences of LapA proteins.....	235
<b>Figure 6.23</b> Comparison of <i>wt</i> LapA and thermolability mutants by $T_{60/30}$ estimate.....	236
<b>Figure 6.24</b> Specific activity values for LapA mutants over a lower temperature range..	240
<b>Figure 6.25</b> Reaction velocity plots versus LPNA substrate concentration for LapA enzymes.....	248
<b>Figure 6.26</b> Reaction velocity plots versus LPNA substrate concentration for LapA enzymes.....	248
<b>Figure 6.27</b> Plot showing the derivative of the fluorescence signal versus temperature. ‘ ..... .....	249
<b>Figure 6.28</b> Plot of the derivative of the fluorescence signal versus temperature. ....	251
<b>Figure 6.29</b> Plot of the derivative of the fluorescence signal versus temperature. ....	252

<b>Figure 6.30</b> Analysis of LapA proteins by DSC. ....	254
<b>Figure 7.1</b> <i>P. pastoris</i> cassettes for LapA <b>A.</b> pro- and <b>B.</b> mature domain expression in <i>trans</i> . .....	258
<b>Figure 7.2</b> Experimental plan to generation of LapA propeptide ( <b>pLapA</b> ) and mature domain ( <b>mLapA</b> ) <i>P.pastoris</i> GS115 expression strains. ....	259
<b>Figure 7.3</b> GS115 agar plates generated in the transformation experiment. ....	260
<b>Figure 7.4</b> Engineered LapA expression cassettes for integration into <i>P. pastoris</i> X33 strain. .....	262
<b>Figure 7.5</b> Recombinant protein secretion analysis of LapA Kex2/Ste13 constructs (1 – 4). .....	263
<b>Figure 7.6</b> Structural comparison of MAT $\alpha$ prodomain ( <b>A</b> ) and LapA prodomain ( <b>B</b> ). ..	266
<b>Figure 7.7</b> Model of MAT $\alpha$ propeptide obtained in the study by Lin-Cereghino et al., 2013. .....	267
<b>Figure 7.8</b> Engineered LapA expression cassettes for integration into <i>P. pastoris</i> X33 strain. .....	268
<b>Figure 7.9</b> Recombinant protein secretion analysis of M $\alpha$ -AVT-LapA and M $\alpha$ -AVT-LapA constructs. ....	269
<b>Figure 7.10</b> Secretion levels of different LapA thermolability mutants. ....	272
<b>Figure 8.1</b> Pairwise alignment of LapA and GAP full open reading frame sequences.....	275
<b>Figure 8.2</b> Structural superimposition of the crystal structure of <i>wt</i> LapA (yellow) and the homology model of GAP (blue). ....	276
<b>Figure 8.3</b> 1% DNA agarose gel analysis of PCR amplicons of pD-npro-GAP transformants. .....	278
<b>Figure 8.4</b> Expression analysis of secreted putative GAP protein. ....	279
<b>Figure 9.1</b> Experimental synopsis and highlights of the LapA thermolability project. ....	289
<b>Figure 10.1</b> MS identification of native LapA enzyme in Flavourzyme .....	307
<b>Figure 10.2</b> MS identification of native Lap2 enzyme in Flavourzyme .....	308

<b>Figure 10.3</b> MS identification of native TAKA $\alpha$ -amylase enzyme in Flavourzyme .....	309
<b>Figure 10.4</b> N-terminal sequencing result of $_{EA}$ LapA protein.....	309
<b>Figure 10.5</b> Multiple sequence alignment of LapA with homologous psychrophilic and thermophilic sequences.....	311
<b>Figure 10.6</b> Superimposition of LapA crystal structure and homologous psychrophilic proteins.....	312
<b>Figure 10.7</b> Homology model of LapA from <i>Aspergillus oryzae</i> .. .....	312
<b>Figure 10.8</b> LapA precursor topology diagram.....	318

## List of Tables

<b>Table 2.1</b> InFusion cloning assembly in 10 µL reaction. ....	34
<b>Table 2.2</b> Components of LB broth (left) and M9 minimal media (right) used for bacterial growth .....	38
<b>Table 2.3</b> Buffers used in IMAC for HiTrap™ 5mL columns. ....	40
<b>Table 2.4</b> Experimental pipeline for development of <i>P.pastoris</i> wild-type LapA strains and initial expression trial.....	53
<b>Table 2.5</b> Experimental pipeline for development of <i>P.pastoris</i> mutant strains, LapA protein expression and purification. ....	54
<b>Table 3.1</b> LAP specific activities measured in SEC-isolated Flavourzyme fractions. ....	59
<b>Table 3.2</b> Effect of glycosylation on the specific activity (U/mg) of LAP enzymes. ....	65
<b>Table 4.1</b> The selected site-directed mutations for sC based on an extensive literature survey and multiple sequence alignment, as described in the text. ....	80
<b>Table 4.2</b> The list of LapA (sequence denoted 20.3) and Lap2 (sequence denoted 20.1) mutants predicted to show thermolability, according to predictive work of J. Mullins (Cardiff). ....	87
<b>Table 4.3</b> Open reading frame amino acid sequences of LapA expression constructs synthesised in pD912 <i>P. pastoris</i> expression vector.....	99
<b>Table 4.4</b> Cell density measured in the primary culture of different <i>wtpro</i> LapA clones..	101
<b>Table 4.5</b> Cell density measured in the secondary culture of different <i>wtpro</i> LapA clones. ....	102
<b>Table 4.6</b> Cell density measured in the primary culture of different <i>wtpro</i> LapA clones..	109
<b>Table 4.7</b> Cell density measured in the primary culture of different <i>wtΔpro</i> LapA clones. ....	109
<b>Table 4.8</b> Cell density measured in the secondary culture of different <i>wtpro</i> LapA clones. ....	110
<b>Table 4.9</b> LapA specific activities measured in the crude <i>P. pastoris</i> supernatants. ....	112

<b>Table 4.10</b> LapA precursor activation data. ....	117
<b>Table 5.1</b> Specific activities of purified LapA proteins. ....	153
<b>Table 5.2</b> Secondary structure prediction via CAPITO, based on the CD data analysis. .	155
<b>Table 5.3</b> Summary of the protein surface statistics calculated between mature and prodomains in the crystal structure of LapA precursor.....	159
<b>Table 5.4</b> Contacts between the catalytic and prodomain. ....	161
<b>Table 5.5</b> Hydrogen bond and non-bonded distances between polar groups of <i>wt</i> LapA and leucine ligand, in the complexed state. ....	177
<b>Table 6.1</b> Library of M1 – M6 LapA thermolability mutants. ....	186
<b>Table 6.2</b> Primers used to generate M1 – M6 point mutations ....	187
<b>Table 6.3</b> LapA mutant clones selected for subsequent large-scale expression.....	189
<b>Table 6.4</b> Specific activity values of LapA proteins before and after enzymatic treatment with AP1 protease. ....	192
<b>Table 6.5</b> CASTp pocket analysis of LapA and GAP proteins. ....	194
<b>Table 6.6</b> Library of M7 – M13 LapA proposed thermolability mutants. ....	197
<b>Table 6.7</b> Primers used to generate M7 – M13 point mutations. ....	198
<b>Table 6.8</b> LapA mutant clones selected for subsequent large-scale expression.....	200
<b>Table 6.9</b> Cell density measured in the secondary culture of LapA supernatants.....	201
<b>Table 6.10</b> TM-align results from the superimposition of <i>wtpro</i> LapA and mutant structures for the catalytic domain residues. ....	208
<b>Table 6.11</b> Temperature factor (B) analysis of LapA crystal structures. ....	213
<b>Table 6.12.</b> Specific activity values observed for LapA variants in the 1 <sup>st</sup> round of mutagenesis.....	217
<b>Table 6.13</b> Remaining percentage activities of LapA proteins, following incubation at 60 °C. ....	219

<b>Table 6.14</b> Kinetic parameters for thermal inactivation of wild-type and mutant LapA enzymes.....	221
<b>Table 6.15</b> Remaining % activities of LapA proteins following incubation at 50 °C.....	223
<b>Table 6.16</b> Kinetic parameters for thermal inactivation of wild-type and mutant LapA enzymes.....	224
<b>Table 6.17</b> Remaining % activities of LapA proteins, following incubation at 60 °C. Data are normalised to 155 U/mg (100%) activity value. ....	225
<b>Table 6.18</b> Kinetic parameters for thermal inactivation of wild-type and mutant LapA enzymes.....	226
<b>Table 6.19</b> Statistical evaluation of exponential one-phase decay model fitted to the inactivation data of select LapA thermolability variants. ....	228
<b>Table 6.20</b> Kinetic parameters for thermal inactivation of LapA enzymes. ....	233
<b>Table 6.21</b> Determined absolute specific activity values of LapA proteins.....	234
<b>Table 6.22</b> Remaining percentage activities of LapA proteins following incubation at 60 °C. ....	236
<b>Table 6.23</b> Remaining percentage activities of WT and mutant LapA enzymes following prolonged incubation at room temperature. ....	238
<b>Table 6.24</b> Room temperature inactivation profiles for <i>wt</i> LapA and mutants.....	239
<b>Table 6.25</b> Measured specific activity values for WT LapA and mutants at lower temperature range.....	240
<b>Table 6.26</b> Remaining percentage activities of <i>flav</i> LapA in post-pasteurisation samples. ....	241
<b>Table 6.27</b> Inactivation profiles of <i>wt</i> LapA, M7 and M8 proteins by pasteurisation. ....	242
<b>Table 6.28</b> Comparison of kinetic constants between recombinant LapA and native LapA enzymes.....	245
<b>Table 6.29</b> LapA enzyme steady-state kinetic parameters calculated from non-linear regression fits to reaction data.. ....	247

<b>Table 6.30</b> Differences in $T_m$ values for wild-type LapA and thermolability mutants..	250
<b>Table 6.31</b> Differences in $T_m$ values of wild-type LapA and thermolability mutants.....	251
<b>Table 6.32</b> Differences in $T_m$ values of wild-type LapA and thermolability mutants.....	253
<b>Table 6.33</b> Thermodynamic parameters for LapA variants, as analysed by DSC. ....	255
<b>Table 7.1</b> Primers used in Quickchange™ cloning PCR reaction.....	262
<b>Table 7.2</b> Primers used in Quickchange™ deletion mutagenesis reaction. ....	268
<b>Table 7.3</b> Cell density measured in the secondary culture of different clones of M $\alpha$ -AVT-LapA and M $\alpha$ -AVT-LapA strains.....	268
<b>Table 8.1</b> Open reading frame sequences of LapA expression cassettes synthesized in pD902 and pD912 <i>P. pastoris</i> expression vectors. ....	277
<b>Table 8.2</b> Cell density measured in the secondary culture of different GAP clones.....	279
<b>Table 9.1</b> Experimental timeline for the 4-year PhD project .....	288
<b>Table 10.1</b> Data collection and refinement statistics of <i>flav</i> LapA. Statistics for the highest-resolution shell are shown in parentheses. ....	313
<b>Table 10.2</b> Data collection and refinement statistics of <i>wt</i> LapA. Statistics for the highest-resolution shell are shown in parentheses. ....	314
<b>Table 10.3</b> Data collection and refinement statistics of <i>wtpro</i> LapA. Statistics for the highest-resolution shell are shown in parentheses. ....	315
<b>Table 10.4</b> Data collection and refinement statistics of LapA M7 mutant. Statistics for the highest-resolution shell are shown in parentheses. ....	316
<b>Table 10.5</b> Data collection and refinement statistics of LapA M8 mutant. Statistics for the highest-resolution shell are shown in parentheses. ....	317
<b>Table 10.6</b> Resolution and accession codes for LapA structures deposited in the public PDB database.....	318

## Abstract

Food processing at low temperatures has many advantages as it minimises undesirable chemical reactions as well as lowers the risk of microbial contamination, which is higher when processes are carried out at higher temperatures. Cold adapted enzymes have higher catalytic efficiencies than their analogous mesophilic enzymes, allowing the use of lower concentrations, which subsequently can be inactivated by relatively low heat processing and therefore reduce heat consumption. Consequently, cold adapted enzymes are becoming increasingly important and can help considerably in developing clean label foods and sustainable processes.

Whey protein is incorporated in a variety of foods and beverages due to its proven beneficial effects on bone health, improved cognitive performance and even prevention of cancer. These whey protein hydrolysates (WPH) are produced by enzymatic digestion followed by heat inactivation of active enzymes. There is high market demand for cold-adapted and thermolabile exopeptidases that could be inactivated at  $<50^{\circ}\text{C}$ , following WPH production. Low temperature inactivation of active peptidases is essential to overcome the high-temperature-induced gelling of the resulting WPH peptides and successive loss of commercial product. Additionally, exopeptidases shape the resultant flavour profile of the WPH product, by eliminating terminal hydrophobic amino acids responsible for bitterness effect.

The aim of this work was to understand the fundamental structure-function relationships underpinning the thermal properties of a protease enzyme. The main focus was toward engineering the thermolability of leucine aminopeptidase A (LapA), a glycoprotein originating from *Aspergillus oryzae*. To this end, a recombinant expression platform, utilizing *P. pastoris* X33 yeast strain for secretory expression of LapA into culture supernatant, was created. The thermolability engineering strategy, developed in the course of this work, was successfully tested using LapA. By reducing the number of rigidifying residues and/or intramolecular interactions, site-directed mutants were found that showed a 50% decrease in half-life at  $60^{\circ}\text{C}$ , compared to wild-type protein. These thermolabile mutants displayed comparable specific activity and kinetic parameters to the wild-type enzyme and maintained their structural and functional integrity.

In addition, three dimensional structures of the wild-type LapA precursor and mature protein were solved during this work, which are the first of their kind. These structures have provided a detailed understanding of the potential role for the LapA pro-domain in premature inactivation and its role as an intra-molecular chaperone in LapA proteogenesis.

Furthermore, high resolution structures for six LapA thermolability mutants were determined, which helps provide a deeper structural understanding of highly compromised thermostability for these mutants.

This work embodies a novel structure-to-function approach to unconventional engineering of thermal lability of industrially relevant peptidases and lays essential foundations for further work in this area.

# Chapter 1 INTRODUCTION

---

## 1.1 Extremophiles

Environmental conditions on earth's biosphere present different challenges to microorganisms. The temperatures encountered, over a variety of biotopes, spans from -80 °C in Antarctic regions up to >100 °C in volcanic hot springs. A certain genus of bacteria and archaea have evolved to thrive in such harsh conditions. These are referred to as extremophiles. As such, extreme temperature, salinity, pressure, and pH are absolutely tolerated by respective extremophilic microorganisms. Very specific adaptations, at the molecular level, ensures the survival of such microorganisms under various environmental stress conditions (Demirjian et al., 2001).

Adaptations of enzymes play a crucial role in withstanding thermal extremes. Optimization of protein activity and stability is the key element in maintaining appropriate cellular functions at very low or high temperatures (Cavicchioli et al., 2002). With regard to temperature, some enzymes have been isolated from organisms viable between 80-100 °C. *Thermus aquaticus* –a famous source organism of the first thermostable DNA polymerase - was first found in several of the Yellowstone National Park hot springs (Brock and Edwards, 1970). Thermo- and hyperthermophiles are adapted to thrive in the range of 60-80 °C and >80 °C, respectively (Vieille et al., 1996). In response to the high thermal motion, thermophilic enzymes have evolved exceptional stability and intrinsic structural rigidity, leading to inherent low activity at low-to-moderate temperatures (Merz et al., 2000).

At the other extremity of the biological temperature scale, there are organisms living at temperatures ranging from -20 °C to 20 °C, with optimal growth temperatures  $\leq 15$  °C (Deming, 2002, Morita, 1975). These are collectively called psychrophiles. There is also a sub-group referred to as psychrotolerant organisms, able to grow at low temperatures, but with optimal growth in the range of 20-35 °C (Bowman and McCuaig, 2003).

The bulk of the earth's biosphere is cold ( $< 5$  °C) and inhabited by psychrophiles (Siddiqui et al., 2013). Such mean temperatures are linked to the fact that 70% of the surface of our planet is covered by oceans where, due to its physical properties, the bulk of water maintains a temperature of 4 °C. In addition, permafrost soils occupy another 20% of earth's terrestrial surface (Feller, 2010). A variety of microorganism species have been discovered, at seemingly inhabitable biotopes, such as porous rocks in Arctic dry valleys (-60°C), brine veins between polar sea ice crystals (-20 °C) and super-coiled cloud droplets (Vaitilingom et al., 2010, Feller, 2010). If one takes into account eukaryotic species that find niches in cold environments, then psychrophiles can be concluded to be the most abundant

extremophiles on earth. Interestingly, psychrophilic bacteria exhibit the ability to maintain reaction rates, comparable to their mesophilic counterparts at higher temperatures, and it is worth noting that doubling times at 4 °C for certain cold-adapted strains have been demonstrated to compare with those of *E. coli* at 37 °C (Pomeroy and Wiebe, 2001, Feller and Gerday, 2003).

Despite the significant variety of cold-adapted microorganisms in the biosphere, their survival strategies, their functioning at the cellular and even molecular level, as yet, remains poorly understood, compared to their mesophilic and thermophilic counterparts. In recent decades, it has been discovered that, primarily, cold adaptation is associated with enzymes that function effectively at low temperatures (Smalas et al., 2000, Georlette et al., 2004). However, other adaptive traits have also been identified, such as increased membrane fluidity (Russell, 1997) and expression of cold-shock and antifreeze proteins (Michel et al., 1997, Muryoi et al., 2004).

## 1.2 Protein synthesis and folding in psychrophiles

For a long time, protein synthesis and folding at low temperatures were considered as requiring specific adaptations and possessing constraints in these temperature-sensitive processes (Araki, 1991, Hebraud and Potier, 1999). Only recently have these challenges been projected to the level of proteomics and transcriptomics (Casanueva et al., 2010, Piette et al., 2011), which have yielded contrasting patterns in cold adaptation (Watanabe et al., 2012, Rodrigues et al., 2008). However, some general trends have emerged:

**Translation.** It appears that protein synthesis may be a rate-limiting step in cold-adapted bacteria e.g. the Antarctic bacterium *P. haloplanktis* was shown to devote 30% of its upregulated proteins at 4 °C to protein synthesis (Piette et al., 2010).

**Protein folding.** The over-expression and trigger of chaperonins (e.g. TF – ribosome bound trigger factor) was observed in cold-adapted microorganisms (Qiu et al., 2006, Kawamoto et al., 2007).

**Proline isomerization.** This is an important aspect of polypeptide folding in psychrophilic proteins. Since proline isomerization is a rate-limiting step for folding of most proteins (Baldwin, 2008), at low temperatures this process should become even slower. Thus, cold-adapted polypeptides exhibit reduced proline content (Feller and Gerday, 2003). The observed over-expression of prolyl isomerases in psychrophilic bacteria further supports the

evidence for the requirement to accelerate protein folding in psychrophiles (Zheng et al., 2007, Goodchild et al., 2004).

### 1.3 Biocatalysis at low temperatures

Extreme temperatures present different challenges to proteins: a low-temperature environment increases viscosity and, due to insufficient kinetic energy, enzymatic activation barriers are difficult to overcome, which remarkably reduces reaction rates (Siddiqui et al., 2013). The latter can be decreased 30- to 80-fold when the reaction temperature is lowered from 37 °C to 0 °C (Ermak and McCammon, 1978).

The catalytic constant  $k_{cat}$  defines the turnover rate at which the maximum number of substrate molecules are converted to product, per active site per unit of time. The temperature dependence of chemical reactions is commonly described by the Arrhenius equation:

$$k_{cat} = Ae^{-Ea/RT}$$

where  $k_{cat}$  is the catalytic rate constant,  $A$  is the pre-exponential factor;  $Ea$  is the activation energy (which is also written as  $\Delta G^\#$ ),  $R$  is the gas constant (8.314 kJ mol<sup>-1</sup> K<sup>-1</sup>) and  $T$  is the absolute temperature. Consequently, even a small reduction in temperature will cause an exponential decrease in the reaction rate.

One of the main characteristic features of psychrophilic enzymes is the shift of the apparent optimal  $T$  with a concomitant decrease in stability. Cold-adapted enzymes compensate for a shortage of kinetic reaction energy by increasing the catalytic rate, which can be up to 10-fold higher compared to heat-stable orthologs (Siddiqui and Cavicchioli, 2006). An increase in catalytic rates arises from the decrease in the activation energy barrier, in catalysed reactions (Figure 1.1).

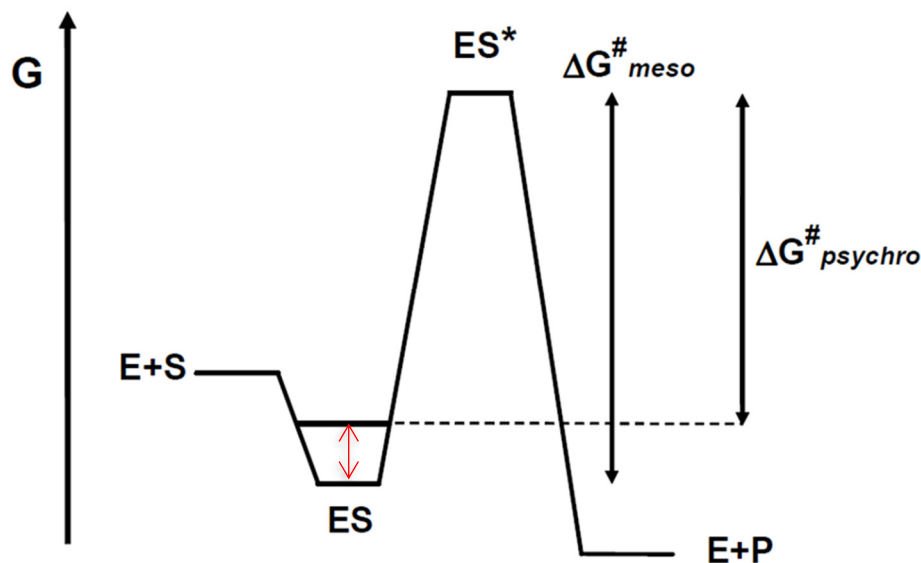
The following calculation for activation energy ( $\Delta G^\#$ ) stems from the well-known Gibbs free energy equation:

$$\Delta G^\# = \Delta H^\# - T\Delta S^\#,$$

where  $\Delta G^\#$  is the change in activation energy,  $\Delta H^\#$  is the change in activation enthalpy,  $\Delta S^\#$  is the change in activation entropy, and  $T$  is absolute temperature. This equation suggests that in order to decrease the activation energy one needs either to make the activation enthalpy more negative or the activation entropy more positive. Cold-adapted enzymes

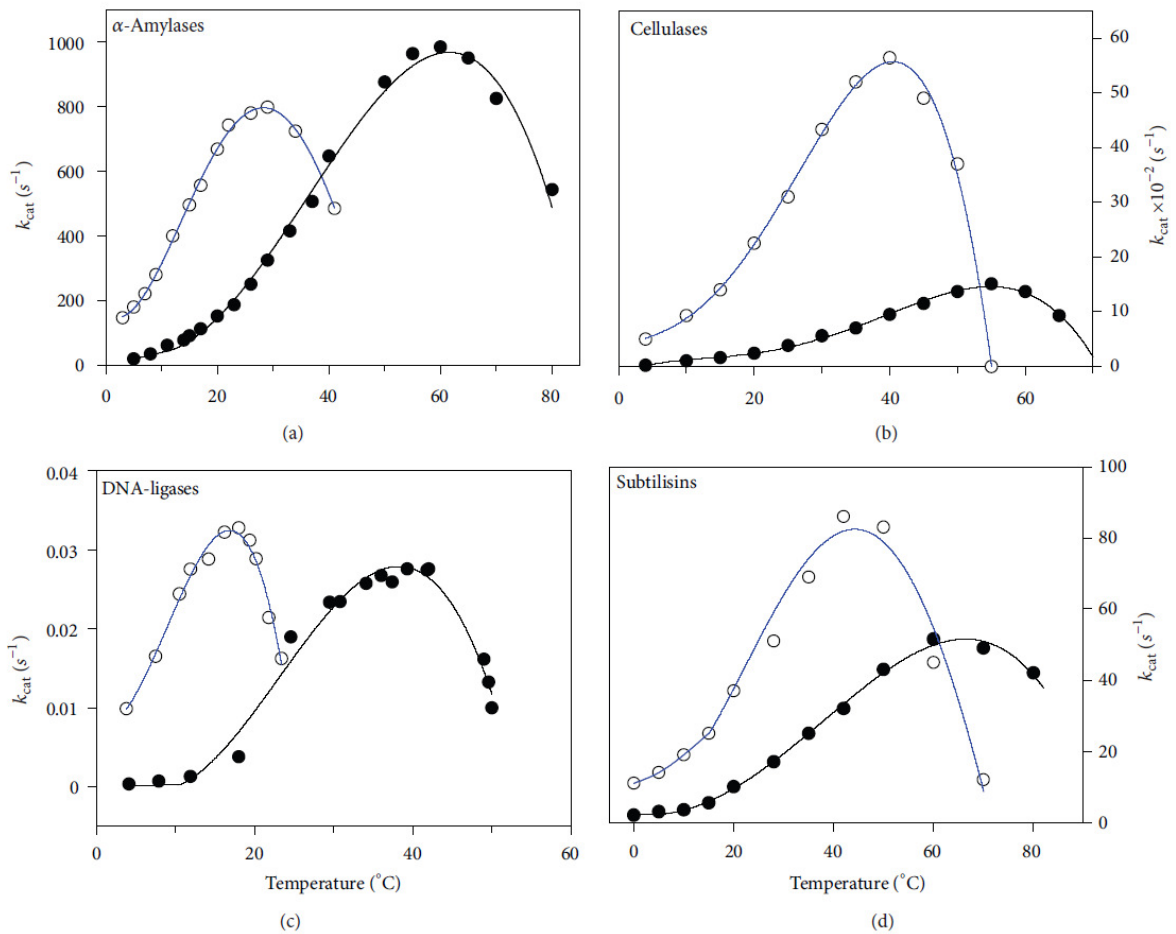
proceed with low enthalpy change (this is, in fact, primary kinetic adaptation in psychrophilic enzymes), which is structurally achieved by a requirement to disrupt less enthalpy-driven interactions within the active site during the activation process (Low et al., 1973, Feller and Gerday, 2003). These interactions are required to be broken during the formation of the transition state from the ground state substrate.

Psychrophilic enzymes achieve higher  $k_{cat}$  rates by decreasing the activation free energy ( $\Delta G^\ddagger$ ) barrier between the ground state (substrate) and the transition state ( $TS^\ddagger$ ) - enzyme-substrate complex ( $ES^\ddagger$ ) (**Figure 1.1**). Also, the figure represents the evolutionary adaptive drift of psychrophilic enzymes towards lowering affinity for the substrate (higher  $K_m$ ), which contributes to the increase in reaction rates. Transition state theory implies that when the enzyme meets the substrate, the enzyme-substrate complex (ES) falls into the energy pit. The  $ES^\ddagger$  activated transition state has to be achieved before the reaction can complete yielding enzyme and product. As can be seen in **Figure 1.1**, cold-active enzymes encounter a less deep energy pit due to their lower affinity for the substrate. Thus the magnitude of the energy barrier is reduced, yielding higher catalytic activity.



**Figure 1.1** Optimisation of activity by decreasing substrate affinity in cold-adapted enzymes. The energy pit is indicated by a red arrow. Weak substrate binding (dashed line) decreases the activation energy barrier for cold-adapted enzymes ( $\Delta G^\ddagger_{psychro}$ ) compared to mesophilic counterparts ( $\Delta G^\ddagger_{meso}$ ). Adapted from (Feller, 2013).

Indeed, cold-adapted enzymes are often characterised by increased  $K_m$  values i.e. reduced affinity for the substrate. This phenomenon is associated with increased disorder of the enzyme-substrate ground state complex, which is a consequence of overall flexibility of the active site and the whole molecular edifice (Fields, 2001). Similarly, there is a trend toward decreased affinity for various ions, cofactors, inhibitors and reduced substrate specificity in cold-adapted enzymes (Davail et al., 1994, D'Amico et al., 2006, Okubo et al., 1999, Alvarez et al., 1998, Smalas et al., 2000). The evolutionary pressure of decreasing substrate affinity ( $K_m$ ) in order to maximise the reaction rates ( $k_{cat}$ ) has been well illustrated by the examples of lactate dehydrogenase from Antarctic fish (Fields and Somero, 1998) and cold-adapted  $\alpha$ -amylase (D'Amico et al., 2001). The decrease in the magnitude of  $\Delta G^\ddagger$  is also related to the decrease in  $\Delta H^\ddagger$  (contribution from intra-molecular interactions) leading to the increased flexibility of the structural elements forming the active site. This adaptation permits cold-active proteins to outperform warm-adapted homologues at low temperatures.



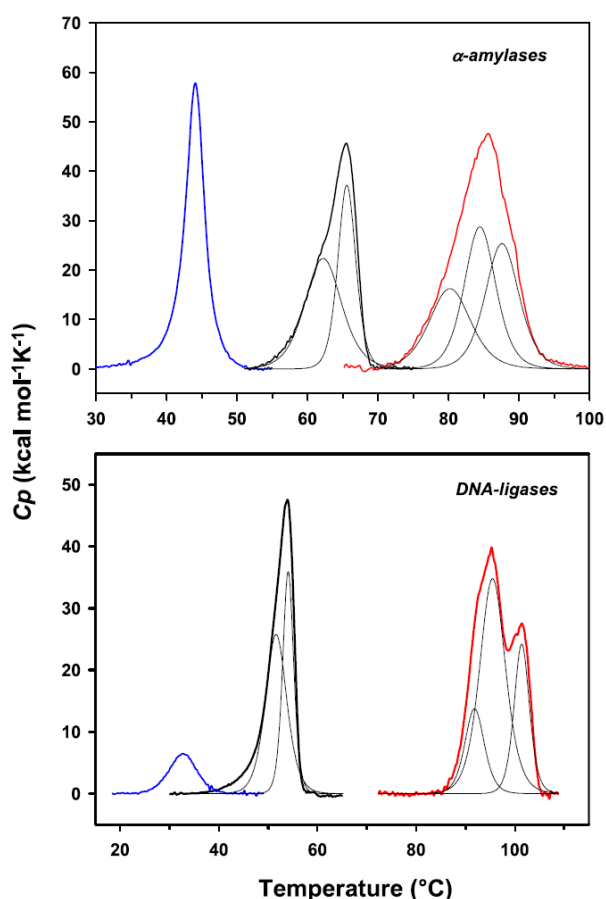
**Figure 1.2** Temperature dependence on the activity of psychrophilic (blue) and mesophilic (black) protein homologues. Adapted from (Feller, 2013).

**Figure 1.2** illustrates how different functional enzyme classes adopt the same strategy in compensating for slow reaction rates at low temperatures by i) a significant increase in specific activity and ii) shift of  $T_{\text{opt}}$  to lower temperature range and resulting heat lability. Although the catalytic rate of psychrophilic  $\alpha$ -amylase is lower than its mesophilic homologue, thus cold adaptation for this enzyme can be considered as incomplete, suggesting potential to further increase the activity at low temperatures.

#### **1.4 Stability of extremophilic proteins**

Protein stability determinants reside within the folded conformation and network of molecular forces and interactions. The driving force for protein folding and the main determinant of protein stability is the hydrophobic effect. Once other residues are brought into contact, weaker molecular interactions (H-bonds, van der Waals) further modulate stability. Salt bridges, disulfide bonds, and aromatic interactions also contribute to increased rigidity. Structural factors such as the occurrence of glycine and proline residues are also involved, whereby the former allows large dihedral rotations along the alpha carbon, and the latter restricts dihedral angles of its own and preceding residues (Piette et al., 2011).

To understand the origin of temperature adaptation and find ways to alter the thermal properties of enzymes, it is useful to compare the thermodynamic behaviour and temperature responses of protein orthologs along the temperature scale.

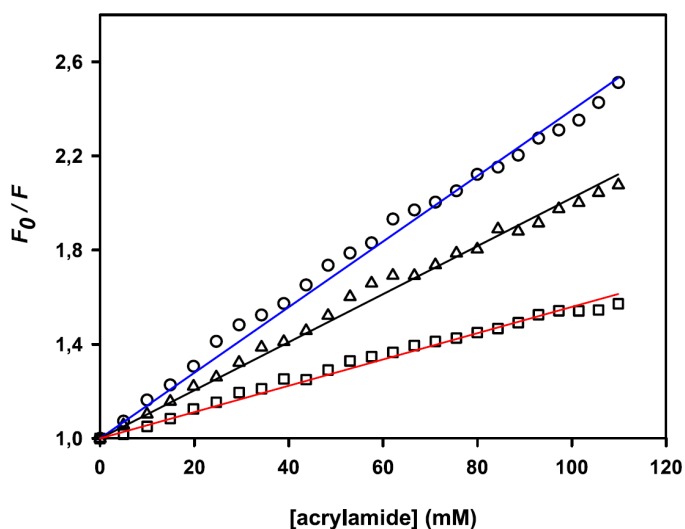


**Figure 1.3** Thermal unfolding of extremophilic proteins. Thermograms of  $\alpha$ -amylases and DNA ligases as recorded by DSC. Psychrophilic (blue), mesophilic (black), and hyperthermophilic (red) protein homologues. The stability domains are indicated by thin black lines under the curves of meso- and thermophilic proteins. Adapted from (Georlette et al., 2003, D'Amico et al., 2001)

Microcalorimetric studies indicate the effects of thermally induced unfolding of proteins adapted to distinct temperatures. In **Figure 1.3**, the area under the curve reflects the amount of heat absorbed during unfolding. It is evaluated by calorimetric enthalpy -  $\Delta H_{cal}$ , which tends to decrease going from thermophilic-to-psychrophilic proteins. The transition for psychrophilic proteins is sharp and symmetric as opposed to its meso- and thermophilic counterparts. This relates to the low stability of the psychrophilic protein fold and indicates the cooperative mode of unfolding, whereby fewer intermolecular interactions results in lower stability and causes pronounced unfolding. Whereas, thermophilic proteins undergo stepwise denaturation and resist unfolding with various intermediate unfolded states arising (Georlette et al., 2003, D'Amico et al., 2001).

The flexibility of protein in solution is best described as fluctuations of larger populations of conformers. These conformational transitions are very complex, involving backbone and side chain motions, including but not limited to bond rotations, stretching and torsional motions (Feller, 2010). Thus, determination of these vast parameters (comprising flexibility) can be further complicated. However, the permeability of protein structure can serve as a great flexibility indicator as was shown by fluorescence quenching experiments with extremophilic proteins (D'Amico et al., 2003, Collins et al., 2003). The reported permeability

study (**Figure 1.4**) investigated three structurally homologous  $\alpha$ -amylases ( $\alpha$ -amylase from *Pseudoalteromonas haloplanktis* (AHA) is the best characterized psychrophilic enzyme isolated from an Antarctic bacterium). The quencher – acrylamide – was used to measure the accessibility of tryptophan across the protein molecule and, more specifically, its ability to penetrate the structure is a good indication of protein permeability. The nature of the flexible protein backbone of the psychrophilic homologue allows the quencher molecule to penetrate with a higher propensity, while the more compact meso- and thermophilic  $\alpha$ -amylases both display a lower extent of permeability and thus flexibility.



**Figure 1.4** Permeability of homologues of  $\alpha$ -amylases at room temperature, as estimated by fluorescence quenching. Psychrophilic protein (blue), mesophilic (black), thermophilic (red). The larger intrinsic fluorescence attenuation ( $F_0/F$ ) and the steep slope for the psychrophilic protein indicate that the acrylamide can easily permeate the structure. Adapted from (Feller, 2013).

The apparent differences in structural flexibility across thermally adapted proteins are further supported by neutron scattering experiments (Tehei et al., 2004). Since neutron wavelengths and energies match protein motion amplitudes and frequencies, neutron spectroscopy allows quantification of macromolecular dynamics, including thermal motion. Using this technique, the decrease in intramolecular rigidity and enhanced dynamics were again confirmed in psychrophilic proteins (Tehei et al., 2004).

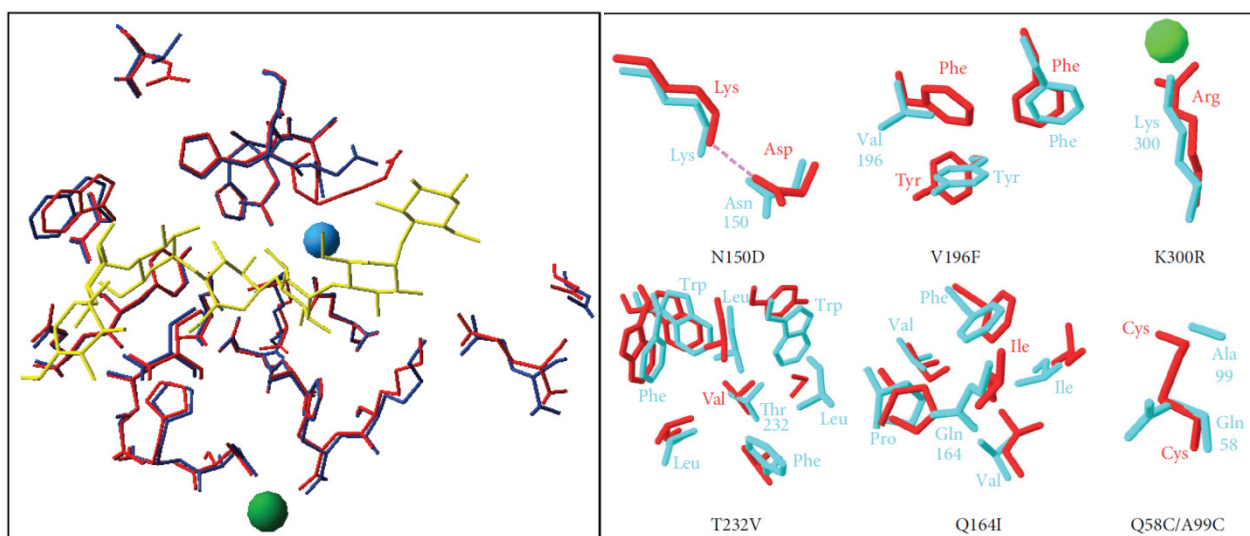
## **1.5 Structural characteristics of psychrophilic enzymes**

In the course of evolution, psychrophilic enzymes developed a spectrum of adaptive properties in order to maintain their intramolecular flexibility. Individual studies and exhaustive comparisons with mesophilic and thermophilic homologues revealed a number of structural adaptations of psychrophilic proteins. For example, local mobility is provided by clustering glycine residues (Kulakova et al., 2004, Mavromatis et al., 2002); the flexibility within the secondary structure elements is achieved by reducing the number of proline residues in the loops joining these elements (Sakaguchi et al., 2007). Prolines are highlighted, particularly as an increased number of proline residues decreases the conformational entropy of the unfolded state (more than the folded state) and thus reduces the force of entropy-driven unfolding. Proline residues restrict their own dihedral angles and that of the preceding residue and, therefore, reduce the conformational freedom at this location of the polypeptide chain (Arnorsdottir et al., 2007). There are several factors that influence thermostability of psychrophilic enzymes. For example, psychrophilic enzymes exhibit a reduced number of aromatic interactions, H-bonds and salt bridges (seen in a lower content of arginine residues, which can form multiple electrostatic interactions) compared to mesophilic proteins (Smalas et al., 1994, Xie et al., 2009, Siddiqui et al., 2006). Furthermore, since the hydrophobic effect is mainly responsible for compactness of structure within the protein core, psychrophilic proteins soften this effect by reducing the size and relative hydrophobicity of nonpolar residue clusters (Saelensminde et al., 2009). Cold-adapted proteins incorporate larger cavities in their interior with more water molecules embedded. There are several other factors such as; stabilising metal cofactors, such as calcium, are bound by affinity several orders of magnitude lower compared to mesophilic enzymes (Davail et al., 1994, Feller et al., 1994); disulphide bridges, stabilising tertiary structure, are sometimes dismissed in psychrophilic proteins (Siddiqui et al., 2005b, D'Amico et al., 2002b); in certain cold-adapted proteins an increase in the number of surface exposed nonpolar residues is observed, which is entropy-driven destabilising factor (Aghajari et al., 1998, Russell et al., 1998, Georlette et al., 2003); more acidic pI values are observed for cold-adapted proteins, which is apparently linked to their more extensive network of surface exposed negatively charged residues providing improved interactions with solvent and resulting in essential flexibility at close-to-zero temperatures (Feller et al., 1999); and finally, improved dynamics of the external shell of psychrophilic proteins appears linked to their lower content of arginine-mediated surface ion pairs, where the number of electrostatic interactions clearly rises going from psychrophilic to hyperthermophilic proteins (Vetriani et al., 1998, Yip et al., 1995). The thermostabilizing effect of arginine

guanidinium groups was demonstrated with psychrophilic AHA protein, whereby lower lysine-to-arginine ratio was attributed to cold-adaptation of this enzyme (Siddiqui et al., 2006).

All of these structural attributes are not found in every psychrophilic enzyme but, rather it appears that individual enzymes adopt their own combination of structural adaptations to function at low temperatures and there is no universal solution to achieve the localized or overall flexibility within a molecular scaffold. Importantly, this also implies that there are multiple routes for any protein system to achieve cold adaptation.

**Figure 1.5** illustrates the weakened intramolecular interaction network within a psychrophilic homologue of  $\alpha$ -amylase, which contributes to flexibility and dynamics of the protein backbone that is required for cold adaptivity.



**Figure 1.5** Structural superimposition of psychrophilic (blue) and mesophilic (red) homologous  $\alpha$ -amylases. Left: overlay of the active site residues; substrate analog, chloride and calcium ions are indicated in yellow, green and blue, respectively; the 24 residues forming direct or water-mediated interactions superimpose almost identically, suggesting that determinants of cold adaptation are lying outside the active site center. Right: the mutation N150D disrupts a salt bridge with the corresponding Lys side chain in the cold-adapted protein. The mutation V196F disrupts a triple face-to-edge aromatic interaction. Mutant K300R results in a monodentate coordination of the chloride ion, instead of a bidentate coordination by Arg in the mesophilic protein, as demonstrated by the crystal structure of the single mutant (Aghajari et al., 2002). Mutations T232V and Q164I decrease

the apolarity of the hydrophobic core clusters in the psychrophilic enzyme and the double mutation Q58C/A99C eliminates a disulfide bond. Adapted from (Cipolla et al., 2011).

Also, thermolability and low temperature activity can be achieved via very subtle cooperative action of long loop insertion and stabilization of an ordered water molecule chain in the vicinity of the active, as demonstrated for the psychrophilic  $\beta$  – glucosidase (BglU) enzyme (Miao et al., 2016). A single H299 residue (stabilizing ordered waters) was attributed to active site heat lability of this protein. Also, in the following mutational study of BglU, lower contents of Pro, Arg and Glu were shown to contribute to the determined psychrophilic properties of this enzyme (Miao et al., 2017).

Some very unique cold adaptation determinants also have been found in eukaryotic proteins. A seven amino acid stretch was shown to confer cold adaptivity and required flexibility of the cytosolic C-terminal region of the transport protein in Antarctic icefish (*Chionodraco hamatus*). Moreover, upon translocation of this region (VDMSRKS) to a mesophilic ortholog protein, the recipient gained cold adaptation (Rizzello et al., 2013).

### **1.5.1 Activity-flexibility-thermolability relationships**

The catalytic activity of enzymes generally involves ‘breathing’ movements, upon accommodation of the substrate associated with conformational changes in the protein structure. The catalytic efficiency, at a given temperature, is governed by the ease of these movements. In fact, the enzyme needs to establish a proper balance between rigidity, which allows the enzyme to maintain the required conformation to keep the substrate in place and flexibility, which enables the enzyme to carry out its function (Jaenicke, 1991a, Jaenicke, 1991b). A low-temperature environment presents higher viscosity and low kinetic energy of the molecules. Thus psychrophilic proteins tend to increase plasticity and reduce the number of interactions within their molecular edifice. The acquired flexibility along with lowered affinity for the substrate determine the increased catalytic rates of cold-adapted proteins (D'Amico et al., 2001). The concomitant heat lability of psychrophilic proteins is a result of attained local and global flexibility within the molecular scaffold. In principle, this would imply that destabilisation of a heat stable protein molecule, via disruption/weakening of the intramolecular interactions, would allow one to engineer a thermolabile enzyme.

Numerous directed evolution experiments have revealed that when mutant libraries are created, to screen for improved activity at low temperatures (no other evolutionary constraints applied), the best mutant candidates display the canonical properties of psychrophilic enzymes (D'Amico et al., 2002a). Random mutations improving both stability and activity are rare as shown by examining hundreds of mutants (Cherry et al., 1999, Giver et al., 1998). Consequently, it seems that from an evolutionary point of view the most frequent and accessible event to improved catalytic activity at low temperatures would be the loss of thermal stability.

The positive correlation between cold-activity and heat-lability was also challenged by concluding that there was a lack of selective pressure for stable proteins such that instability of these proteins is a result of genetic drift in natural evolution (Wintrobe and Arnold, 2000). However, this situation most likely is more complicated, as nature created multi-domain cold-adapted enzymes that contain a heat-labile catalytic domain and a mesophilic-like (more stable) non-catalytic domain (Claverie et al., 2003, Lonhienne et al., 2001). Thus, it is unlikely that genetic drift is the dominant factor for different domains in the same protein.

In contrast, heat-stable enzymes from thermophilic organisms are known to be less active at low temperatures and acquire larger values of activation entropy and enthalpy (D'Amico et al., 2003). Therefore, an inverse correlation between substrate turnover at low temperatures and molecular rigidity of thermophilic proteins is apparent. This signifies the stability – activity trade – off evolved in hot-adapted enzymes.

### **1.5.2 Activity-stability trade-off and active site flexibility in psychrophilic enzymes**

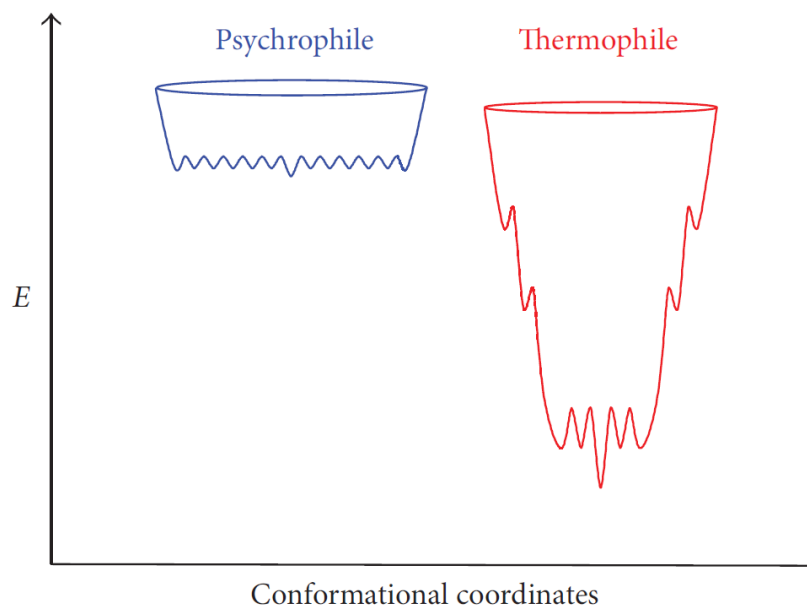
Previous observations can be summarised in a single activity-stability trade-off model. Psychrophilic proteins show high intrinsic activity at low temperatures at the expense of reduced stability (hence thermolability). While the increased stability and structural rigidity in thermophilic proteins results in diminished catalytic rates at similar temperatures. Thermolabile and cold-active  $\alpha$  – amylase (AHA) was demonstrated to display localised active site flexibility, which is attained via unfavourable active site geometry that also promotes higher activity (Siddiqui, 2017). A flexible active site bearing unfavourable interactions permits more effective binding of the transition state (TS<sup>#</sup>). Competitive inhibitors mimicking the transition state were shown to significantly improve the thermostability ( $T_m$ : from 44 to 61 °C) of highly active AHA, thereby binding to residues via favourable interactions. The thermophilic homologue from *B.amyloliquefaciens* showed

no improvement in  $T_m$  upon inhibitor binding, because the active site pocket is already optimized for higher stability at the expense of activity (Siddiqui et al., 2005a). Therefore, the thermal lability the active sites of cold-adapted enzymes stems from highly flexible geometry as a result of fewer enthalpic interactions. The pronounced active site heat lability can be exploited in engineering mesophilic enzymes towards inactivation at lower temperatures.

### 1.5.3 The folding funnel model

Integrated kinetic, structural and biophysical data from enzymes adapted to extreme temperatures has been summarized by a model based on folding funnels (D'Amico et al., 2003). Activity-stability relationships and energy landscapes of psychrophilic and thermophilic enzymes can be described using these models (**Figure 1.6**). In general, the model represents folding-unfolding events in psychrophilic and mesophilic proteins. The height of each funnel (free energy of folding) corresponds to the conformational stability (which is lower in psychrophilic enzymes). The upper edge of the funnel is occupied by the unfolded state (random coil) conformations. This edge in psychrophilic proteins is slightly longer due to the larger distribution of unfolded states. Upon the folding event of protein, the level of free energy decreases along with the size of the conformational ensemble (represented as the funnel narrowing). The bumpiness of the funnel edges for thermophilic proteins corresponds to intermediate folding-unfolding states, resembling local energy minima. In contrast, the conformational states of psychrophilic protein appear to unfold in a cooperative fashion, representing no unfolding intermediates (also see **Figure 1.3**). This is linked to a lower number of stabilizing interactions and the absence of stable domains. As a result, the edges of the funnel are smooth for psychrophilic proteins. The bottom of the funnel corresponds to the stability of the native state ensemble, and it also displays the significant difference between the two extremophilic enzymes. The bottom of the rigid thermophilic enzyme can be characterized as a single global minimum or as having only a few minima with relatively high energy barriers between them (Kumar et al., 2000, Tsai et al., 1999). The rugged bottom observed for unstable psychrophilic enzymes represents a large population of interconverting conformers with low energy barriers. Thus, the rigidity of the native state is directly correlated to the energy barrier height. This is in agreement with the results of the fluorescence quenching (**Figure 1.4**) and neutron scattering experiments discussed earlier (Tehei et al., 2004). In this context, the activity-stability relationship in these extremozymes depends on the properties of the base of the folding funnel. Indeed, it has been argued that upon substrate binding to the association-competent

sub-population, the equilibrium between all conformers is shifted towards this sub-population, leading to the active conformational ensemble (Kumar et al., 2000, Tsai et al., 1999). In the case of the rugged bottom of the folding funnel of psychrophilic enzymes, this equilibrium shift only requires a modest free energy change and a low enthalpy change for inter-conversion of the different conformations of the numerous conformational isomers existing in the wide conformer ensemble. The low energy barriers across a spectrum of micro-states would result in an overall decreased activation energy and a high  $k_{cat}$ . Moreover, the increased flexibility would cause the cold-adapted enzyme to spend more time in conformations that are not optimal for substrate binding and result in a high  $K_m$  (Somero, 2004).



**Figure 1.6** Folding funnel model of extremophilic proteins. These schematic energy landscapes represent the plot of the free energy of folding ( $E$ ) versus conformational diversity. The height of the funnels relates to conformational stabilities, the top of the funnels – unfolded states in random coil conformations, bottom – native and catalytically active conformations. The ruggedness of the bottom corresponds to energy barriers for interconversion indicative of the structural fluctuations of the native state. Adapted from (D'Amico et al., 2003).

## 1.6 Cold-adapted thermolabile proteases

There are only a few cold-adapted proteases that have been structurally and functionally characterised. With respect to known exopeptidases, the well-described aminopeptidase from *Colwellia psychrerythraea* (ColAP) displays common features of psychrophilic and thermolabile enzymes (Bauvois et al., 2008, Huston et al., 2008). The comparison with its mesophilic homologue, human leukotriene A4 hydrolase, revealed that  $T_m$  values were at least 15 °C lower and  $k_{cat}$  were 8-fold higher for ColAP, yielding lower thermal stability and higher specific activity of the psychrophilic enzyme. Structural permeability studies by fluorescence quenching also identified ColAP as exhibiting more permeable compared to its mesophilic homologue. A lower number of proline residues, ions pairs, and buried hydrophobic residues were identified as the main structural determinants of weak conformational stability and thermolability for ColAP (Huston et al., 2008).

By contrast, known cold-adapted endoproteases are extremely well represented in the literature. One of the first insights into the determinants of cold adaptation came from studies of subtilisins S39 and S41 secreted by psychrophilic *Bacillus* (Narinx et al., 1997, Davail et al., 1994). Mutagenesis of subtilisin S39 and, in comparison to Subtilisin Carlsberg, showed three orders of magnitude lower affinity for calcium and a higher number of Asp residues in loops for the psychrophilic homologue (Narinx et al., 1997). Importantly, a single mutation T85D significantly increased the affinity for calcium and dramatically improved thermostability for S39, which was comparable to that of Subtilisin Carlsberg. This study highlights the importance of bound calcium ions for the thermal stability of subtilases. Similarly, subtilisin S41 was shown to be able to induce a more flexible and heat labile conformation due to four extended surface loops, which create a very hydrophilic surface (11 extra Asp residues) and a lower affinity for calcium, compared to the mesophilic Subtilisin Carlsberg (Davail et al., 1994). Another psychrophilic subtilase, termed P6, was isolated from Antarctic bacteria in which Ala284 was identified as an important hot spot in the cold adaptation of mesophilic subtilases during laboratory evolution (Acevedo et al., 2013).

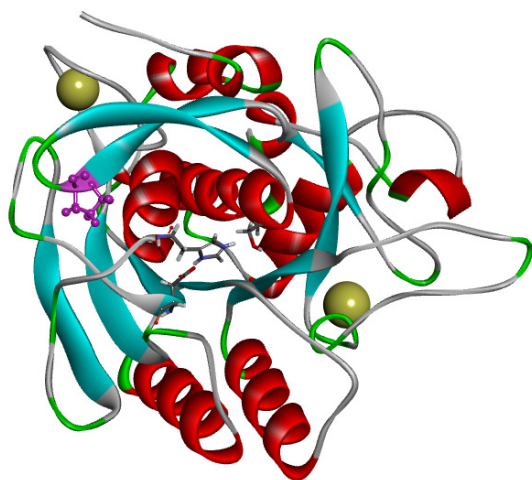
Some of the psychrophilic subtilases have evolved very diverse structural composition, particularly, eukaryotic homologous family members. The MCP-01 protease secreted by the deep-sea cold-adapted bacterium *Pseudoalteromonas sp.* SM9913 was reported to have a multi-domain architecture in contrast to the homologous S8 family members that are comprised of a single domain polypeptide. This protease, termed deseasin, displayed a catalytic region, a linker, a P-protein domain and a PKD domain, with only the catalytic

domain bearing a moderate degree of homology to the S8 family subtilases (Chen et al., 2007).

### 1.6.1 Engineering low-temperature activity and thermolability in hydrolases

Highly flexible residues have a low number of contacts with other amino acids and produce a local perturbation inside the complex network of non-covalent interactions. They can trigger protein unfolding due to their large thermal fluctuations (Yu and Huang, 2014). Therefore, most of the below described cases of protein engineering aim to introduce a certain degree of local flexibility and counteract on rigid ‘hot spots’.

In a recent study, Subtilisin Carlsberg (sC) was engineered to have increased thermolability (12-fold) and specific activity (1.5-fold) via a single P210G substitution in the  $\beta 8$ - $\beta 9$  turn region (**Figure 1.7**) (Fuchita et al., 2012). This region has been shown to display a higher abundance of proline residues in mesophilic and thermophilic subtilases. Interestingly, a 148 amino acid insertion in the  $\beta 8$ - $\beta 9$  region is found in the subtilisin-like serine endoprotease from *Shewanella* sp. AS-11, which may introduce the required flexibility between the secondary structure elements in the cold-adapted enzyme (Dong et al., 2005).



**Figure 1.7** Structure of Subtilisin Carlsberg (pdb:3UNX). Active site residues are shown as sticks (grey), Pro210 in  $\beta 8$ - $\beta 9$  turn is shown in magenta, and calcium coordination sites are depicted as yellow spheres.

Mesophilic subtilisin Savinase was engineered towards low-temperature activity and thermolability by inserting a highly flexible eight amino acid fragment from psychrophilic subtilisin TA39 (Tindbaek et al., 2004). This H5 hybrid displayed an increase in low-temperature activity, a decrease in thermostability and, improved flexibility in both the substrate binding region and globally, as shown by molecular dynamics and DSC (Tindbaek et al., 2004).

Subtilisin BPN<sup>+</sup> was engineered extensively to confer cold adaptation by *in vitro* random mutagenesis. The single V84I mutation was shown to slightly increase the activity of mutant

subtilisin BPN<sup>ˆ</sup> in the 5-10 °C temperature range, but had no effect on the thermolability compared to the wild - typeenzyme (Kano et al., 1997). In a similar study, taking advantage of a directed evolution approach, cold adaptation of subtilisin BPN<sup>ˆ</sup> was achieved through V72I, A92T, and G131D mutations, and the catalytic efficiency of the resultant mutant was found to be 100% higher at 10 °C (Taguchi et al., 1998b). Interestingly, the V72I mutation on its own had a significant effect on thermolability of subtilisin BPN<sup>ˆ</sup>. However, the triple mutant (V72I+A92T+G131D) recovered the thermal stability. In addition, randomisation and saturation mutagenesis at position G131 identified a series of subtilisin BPN<sup>ˆ</sup> mutants with improved proteolytic activities at 10 °C and conferred thermolability to varying degrees, revealing potential hot spots for cold adaptation in subtilisins (Taguchi et al., 2000).

A saturation mutagenesis approach on identified rigid ‘hot spots’ in protein structure was also implemented for engineering proteins with increased thermolability. The lowest B-factor residues (indicative of high rigidity) are selected for iterative saturation mutagenesis strategy, whereby the randomisation at every identified ‘hot spot’ produces large libraries of mutants requiring extensive screening. This approach has lead to destabilisation and significantly increased thermolability of lipase from *Pseudomonas aeruginosa* (Reetz et al., 2009). The best mutant displayed a 20 °C decrease in inactivation temperature without affecting its catalytic performance.

One of the other factors that have an impact on enzyme low-temperature activity is uncompetitive substrate inhibition. The industrial protease Savinase (*Bacillus lentus*) was used as a model to show that by chemically modifying the enzyme it was possible to increase its low-temperature productivity by sterically blocking the allosteric site, thereby reducing uncompetitive substrate inhibition (Siddiqui et al., 2009).

### **1.6.2 Identified strategies to achieve thermolability and cold-adaptivity**

Based on the literature review in previous sections, a number of strategies were identified in order to address the problem in this project. The desired thermolability with potentially added enhanced cold-activity properties in target enzymes can be mainly achieved in the following ways.

#### ***Increasing overall conformational flexibility***

This can be achieved by reducing the number of intra-molecular interactions (electrostatic, covalent and hydrophobic), improving interactions with solvent (introducing more negatively charged surface residues, solvent-exposed apolar groups), reducing the number of surface salt bridges.

#### ***Surface loop engineering – entropic approach***

Multiple studies indicate the importance of surface residues (particularly in loops) in defining the thermal properties of an enzyme. Restricted rotation around C $\alpha$  – N bond in pyrrolidine ring of proline renders this residue to have the lowest conformational entropy (especially in unfolded state). In contrast, due to the missing  $\beta$  – carbon glycine shows the highest conformational entropy. To further destabilise the temperature – denatured state we can exploit these entropic residue effects.

Proline residues were shown both to enhance the thermostability (when introduced in surface loops of psychrophilic proteins) and improve the thermolability (when eliminated from surface loops in mesophilic proteins). Therefore, a special emphasis will be given to proline residues in our mutagenesis strategy to confer local disorder. Several reports indicated that the unfolding process might be initiated from the disordered surface loop (Deng et al., 2014, Anbar et al., 2012).

#### ***Enhancing thermal lability of the active site***

Since the main aim of this study is to inactivate the enzyme at low temperatures, mutations may be targeted to introduce local flexibility in the active site pocket, thereby distorting the active site geometry at moderate temperatures even before the melting of overall protein structure.

### ***Rigid cluster ('hot spot') engineering***

Intrinsic overall structural rigidity is a defining property of protein thermostabilisation. Therefore, identifying the rigid residues ('hot spots') defined by low B-factors in a crystal structure of the target protein may help to direct rational substitutions at these sites which in turn would confer localised or overall flexibility to protein structure.

## **1.7 Industrial applications of cold-adapted enzymes**

Despite the enormous potential to reveal the diversity of biomolecules and bioprocesses occurring in psychrophilic organisms, which could have a significant impact in industrial applications, our understanding of the ecology, physiology, and enzymology of cold-adapted organisms still remains limited (Huston, 2008). The discovered and characterised psychrophilic microorganisms constitute only <1% of the total successful attempts (to isolate these) since the traditional isolation and cultivation laboratory methods are usually detrimental to these extremophiles (Giovannoni et al., 1990). Consequently, the existing diversity of biocatalysts from psychrophiles is under-represented. However, recent advances in metagenomics and high throughput sequencing allows researchers to bypass the cultivation step in order to isolate novel genes encoding unique biocatalysts (Huston, 2008).

The biotechnological potential of these cold-adapted biocatalysts originates from their structural and thermodynamic properties: the inherent flexibility and low substrate affinity allow them to display higher  $k_{cat}$  values in low-to-moderate temperatures, high thermolability at elevated temperatures and functionality in organic solvents (Cavicchioli et al., 2002, Margesin and Feller, 2010). Current industrial processes, utilising high-temperature adapted enzymes, can benefit from substantial energy savings if more productive (at low temperatures) cold-adapted biocatalysts are employed. Thus, it was not too long before they found application in a variety of industries from household detergents, to molecular biology and baking (Cavicchioli et al., 2011). Industrial processes, sensitive to the occurrence of secondary reactions, heat-sensitive substrates or increased risk of microbial contamination, all benefit from reduced temperature biocatalysis (Jeon et al., 2009). In food and feed industries, it is imperative to avoid harming or changing the nutritional value and flavour of the heat-sensitive substrates and products. Furthermore, selective inactivation at moderate temperatures by heat-labile enzymes permits controllable inhibition of sequential reactions at intermediate steps in the food industry (Cavicchioli et al., 2011).

### **1.7.1 Industrial application of cold-adapted proteases**

The market value of global enzyme sales was estimated to reach \$6 billion by 2011, increasing by 7% annually (McCoy, 2001). Proteases represent one of three largest groups of industrial enzymes and account for about 60-65% of global sales of enzymes, where alkaline proteases constitute 25% (Rao et al., 1998). The vast majority of commercial proteases (two-thirds) come from microbial sources, although the enzymes from plants and animals also are produced (Kumar and Takagi, 1999). Most of the neutral and alkaline commercial proteases originate from *Bacillus sp.* These proteases typically are active in the pH range 5-8, display broad substrate specificity, and are not very thermo-tolerant (Rao et al., 1998).

Only in the last decade, the full potential of cold-adapted proteases (regarding catalytic capabilities at low temperatures and thermolability) has been recognised by a number of industries (Kuddus and Ramteke, 2012). Some industrial uses of proteases in the industry are discussed briefly below.

#### ***Detergents***

Enzymes have been added to laundry detergents for over 50 years to remove proteinaceous stains. The very first protease incorporated in washing formulations was Alcalase (Subtilisin Carlsberg) back in 1960.

The apparent benefits of cold washing are decreased energy consumption, protected colours and prolonged life of the fabric. Due to higher or equivalent efficiency of cold-adapted proteases (as opposed to warm-adapted analogues), the processing times become shorter which contributes to extra energy savings. The lack of stability within other detergents and chemicals in washing formulations limited the utilisation of cold-adapted enzymes in the final product (Kuddus and Ramteke, 2012). However, several suitable proteases were recently identified: cold active serine protease –CP70 (*Flavobacterium balustinum*) – has a temperature optimum 20 °C lower compared to the typical detergent protease – Savinase. It is also tolerant to a range of pH 6.5 – 10 and is resistant to surface-active/bleaching agents. The psychrophilic alkaline protease from *Stenotrophomonas maltophilia* also appeared as a potential candidate for improving future washing formulations due to its excellent stability and proteinaceous stain removal capabilities (Kuddus and Ramteke, 2009, Kuddus and Ramteke, 2011).

## ***Food***

Food companies are required to maintain low food processing temperatures to minimise undesirable chemical reactions and to reduce the chances of microbial contamination, both of which are more pronounced at elevated temperatures. In this regard, cold-adapted proteases find important use in a variety of food processing applications; accelerated maturation of cheese, treatment of beer (Margesin and Schinner, 1994), tenderization of meat and taste improvement of refrigerated meat (He et al., 2004).

## ***Molecular biology***

Heat-labile proteases could be used as inactivation agents to other catalytic enzymes once the desired reaction is complete, as demonstrated with hydrolysis of *Taq* polymerase mediated by thermolabile A9 protease (Moran et al., 2001). This permits to avoid PCR product purification step, which is time-efficient and cost-effective in recombinant cloning protocols.

## ***Textiles***

Cold-active enzymes have found numerous niches in the textile industry. These are used for fabric finishing, degumming threads of raw silk to remove sericin (a proteinaceous agent covering silk fibers) and modifying the surface of wool (Najafi et al., 2005, Kuddus and Ramteke, 2012).

## ***Bioremediation***

The advantages of high catalytic efficiency at low temperatures of psychrophilic proteases can serve a purpose in wastewater treatment, bioremediation of polluted soils and protein contaminate cold environments (Kuddus and Ramteke, 2012).

## ***Therapeutics***

ColdZyme<sup>®</sup> (Enzymatica AB, Lund, Sweden), a mouth spray, uses a cold-adapted protease from cod. It was shown to reduce the incidence of common cold and to suppress the infectivity of human rhinovirus 16 and herpes simplex virus 1 *in vitro*.

## **1.8 Production of protein hydrolysates**

### **1.8.1 Protein enzymatic hydrolysis in functional food development**

Enzymatic hydrolysis of a vast array of protein isolates are known to be beneficial and significantly increase their functional properties and nutritional value. For instance, enzymatic hydrolysis forms a basis in the production of fermented soy products, which in turn display improved flavor and absorptivity properties (Sun, 2011). Due to reduced antigenicity of the resultant protein hydrolysates, they find use in peptide-based hypoallergenic infant formulas, hospital weight control, enteric and other therapeutic diets (Cordle, 1994, Adlernissen, 1977). Food-derived bioactive peptides (BAPs) have a potential to exert many beneficial, health-enhancing effects *in vivo* (Rutherford-Markwick, 2012). These naturally inaccessible oligomeric amino acid combinations are released by the hydrolytic action of proteases both *in vivo* and *in vitro*. Many biological functions have been ascribed to these food-derived peptides, such as antioxidant, antihypertensive (ACE inhibition), immunoregulatory properties (Najafian and Babji, 2014, Power et al., 2013).

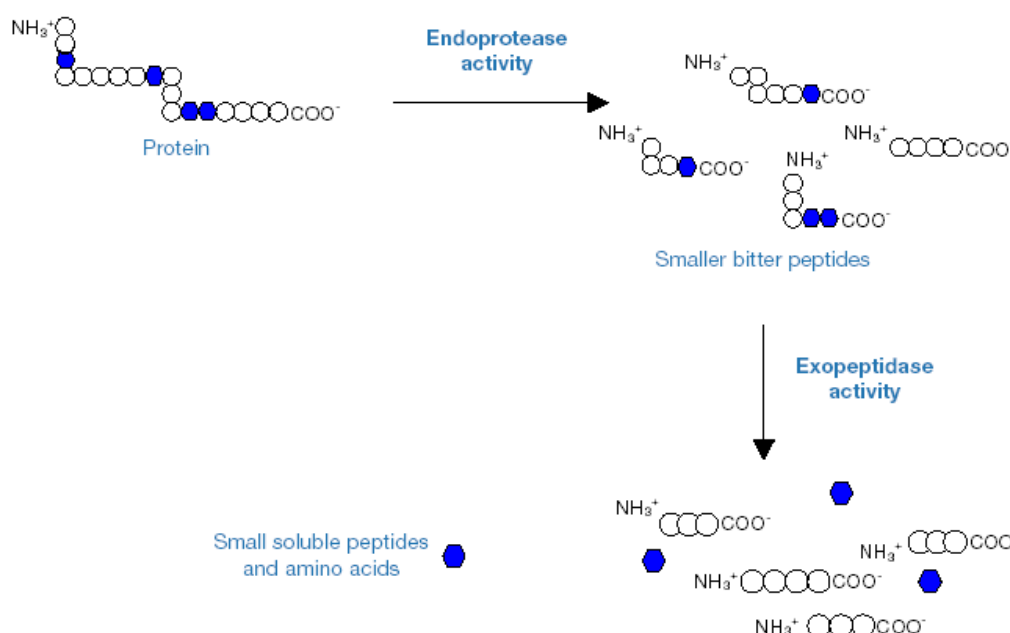
Controlling degree of hydrolysis (DH) becomes the key in an effort to reduce the bitterness in the resulting hydrolysates, which directly affects customer acceptance. Enzymatic debittering protocols utilizing exo- and endopeptidases can both improve the DH and bitter taste (FitzGerald and O'Cuinn, 2006). Therefore there is an ever increasing demand in the functional food industry for effective peptidase blends to produce bioactive, hypoallergenic, non-bitter tasting functional small molecular peptides.

### **1.8.2 Enzymatic hydrolysis of whey protein**

Whey protein hydrolysates (WPH) due to their nutritional value have been incorporated into functional foods for several years. The WPH are used in infant formulae, sports nutrition products, clinical foods, instant beverages, weight management products and much more. Whey proteins have a high protein quality score and contain a relatively high proportion of branched-chain amino acids – Val, Leu, Ile (BCAA, ~26%) (Bos et al., 2000). In this respect, the significant amount of leucine present in whey protein can be extremely beneficial, as leucine plays a key role in protein metabolism and the translation initiation pathway of muscle protein synthesis (Anthony et al., 2001).

Enzymatic hydrolysis is usually the method of choice to produce WPH, as it is conducted under mild conditions thus retaining the nutritional quality. The resultant WPH display a high degree of hydrolysis and a variety of short peptides, which are less allergenic due to the destruction of allergenic epitopes (Siemensma et al., 1993). In the process of WPH

hydrolysis, a low content of free amino acids is preferred due to the absorbance differences in the intestine; a product with more di- and tripeptides will be absorbed more efficiently (Boza et al., 2000). To achieve such a digestion profile, endoprotease (e.g. Alcalase) is usually employed for prolonged hydrolysis of whey protein, ensuring the absence of free amino acids in the final product. One of the cornerstones in commercial production of hydrolysates is developing an enzyme mixture that produces the desired ratio of amino acids, di-, tri- and oligopeptides (Lahl and Braun, 1994). Production of hydrolysates, by the action of endopeptidases and exopeptidases, is considered the most effective way to obtain protein hydrolysates with defined characteristics (Clemente, 2000) (**Figure 1.8**).



**Figure 1.8** Hydrolysis of whey protein by the simultaneous action of endoprotease and exopeptidase enzymes. The bitter peptides containing C- or N-terminal hydrophobic amino acids are degraded by the specific action of exopeptidases. A hydrophobic amino acid is represented by blue hexagon.

### ***Bitterness of WPH***

In their natural state, proteins do not contribute chemically to the formation of flavour in foods. However, protein hydrolysates have an array of flavours. The industry faces a real challenge to try and reduce the bitterness of whey protein hydrolysates, which reduces customer acceptance. Bitterness is caused by peptides (>10 kDa) containing terminal hydrophobic amino acids such as Pro, Leu, Phe, etc. (Saha and Hayashi, 2001). The bitterness of protein hydrolysates is particularly undesirable in beverages.

Enzymatic de-bittering serves as the main tool to further process the WPH for flavour development. For example, soybean protein is incubated with Alcalase (Subtilisin Carlsberg) and the resultant bitter product is further hydrolyzed by Flavourzyme, yielding an improved taste profile with the reduced bitterness of the soybean polypeptide mixture (Ma et al., 2013). Flavourzyme is a fungal extract from *Aspergillus oryzae* containing two exopeptidases, namely LapA and Lap2, either of which can remove the N-terminal hydrophobic amino acids contributing to the pronounced bitter flavor of protein hydrolysates. On the other hand, the essential and bioactive amino acids are usually those containing hydrophobic and/or aromatic side chains. Therefore the segregation of these amino acids from small functional peptides may compromise their bioactive properties (Carreira et al., 2003).

### ***Inactivation of proteases***

From a practical point of view, when the desired characteristics of the hydrolysate have been achieved, the active endoprotease or/and exopeptidase need to be deactivated, preventing further hydrolysis (Conesa and FitzGerald, 2013). This is usually achieved by heat treatment causing thermal denaturation of previously active peptidases. Yet, upon heating various undesirable chemical (Maillard reaction products) and physical (peptide denaturation) effects have been observed. Hence, the biological, functional and nutritional properties of hydrolysates are affected by heat processing, and energy input must be minimised (Nicoli et al., 1997, Lijun et al., 2012). Therefore, this phase of the enzymatic WPH production is of significant importance to the food industry (Xue et al., 2012).

Mesophilic proteins utilised in current formulations of commercial peptidase enzyme blends (e.g. Alcalase, Flavourzyme) require high temperature (>85 – 90 °C) treatment to achieve complete inactivation. To date, there are no reported alternative industrial formulations containing either cold – adapted or thermolabile peptidases for whey protein hydrolysis market.

### ***Gelation of WPH***

Upon heating (>70 °C) the mixture of WPI/WPC and proteolytic peptidases, the forming bioactive oligopeptides (WPH) along with partly hydrolysed globular proteins are susceptible to gelation. The mechanistic principle of the high temperature induced complex gelation process is not fully understood. Nonetheless, several models have been proposed, whereby formation of reactive -SH groups upon denaturation and aggregation were deemed to promote non-specific disulphide covalent linkages. Concomitantly, the aggregation, is

supported by hydrophobically interacting interpeptide networks (Oldfield et al., 1998). During the heating process, an unfolded form of  $\beta$  – Lactoglobulin 2 ( $\beta$  – Lg 2), abundant in whey protein, aggregates with  $\alpha$  – Lactalbumin and  $\kappa$  – casein on the surface of the casein micelle (via the -SH group). The latter gelation effects were observed following WPI treatment with Alcalase and were attributed to hydrophobic interactions of  $\beta$  – Lg derived 2 – 6 kDa peptides (Otte et al., 1997). These insoluble aggregates can also clog and damage heat exchangers, pasteurisation equipment vessels and form sediments in heated WPH mixture, spoiling the end product.

In conclusion, there is a high demand in Industry for thermolabile endo- and exopeptidases that can be inactivated by mild heat treatment such that the beneficial properties and quality of the vegetable and whey protein hydrolysates are preserved.

## **1.9 Importance of this project**

The industrial partner – Biocatalysts Ltd – is offering novel enzyme solutions for food and healthcare industries. Biocatalysts identified a commercial need to engineer proteases that can be inactivated  $<50$  °C. Development of cold-active and thermolabile proteases, through protein engineering existing mesophilic variants (of, for example, Subtilisin Carlsberg, and Leucine aminopeptidase A), would allow the industry to benefit from improved properties of these enzymes. Advantages would include: increased catalytic rates at low and moderate temperatures (detergent industry, energy savings), thermolability (complete inactivation of the functional enzyme at moderate temperatures) and elimination of gelation of WPH at elevated temperatures.

The need for combined endo- and exopeptidase activities for efficient hydrolysis of whey protein and flavour development influenced the selection of target enzymes in this project. Additionally, the knowledge gained in this project will contribute to fundamental understanding of protein engineering to affect the thermodynamic behaviour of protein in an unconventional way, specifically, increasing the thermolability and flexibility of the polypeptide chain to destabilise intramolecular interactions.

## **1.10 Target enzymes in this project**

Combining the interests from an industrial partner and our academic interest in fundamental structure-function approaches, it was agreed that one endoprotease and one aminopeptidase used in industry, will be selected for mutagenesis experiments in an effort to engineer more thermolabile enzymes.

### **1.10.1 Subtilisin Carlsberg (sC)**

Subtilisin Carlsberg, a serine endoprotease with 274aa, is an extracellular enzyme produced by *Bacillus licheniformis* and belonging to the superfamily of S8\_S53 proteases (BLAST). It is classified as a member of the superfamily of Subtilases: subtilisin-like serine proteases (Barrett and Rawlings, 1995, Rawlings and Barrett, 1994). Subtilisin Carlsberg is a single polypeptide protein displaying a characteristic alpha/beta subtilisin-like fold with a 7-stranded parallel beta sheet (SCOP). The active site is composed of the classic D32/H64/S221 catalytic triad, where serine acts as a nucleophile, aspartate as an electrophile and histidine as a base. Two  $\text{Ca}^{2+}$  coordination sites are present in the active form, exhibiting different affinities for  $\text{Ca}^{2+}$  ions. Subtilisin Carlsberg is non-specific alkaline endoprotease, containing a binding surface channel that can accommodate at least six amino acid residues (Siezen, 1996). The temperature optimum for activity of wild-type sC is 60 °C in alkaline pH 7-10. Thermal deactivation is achieved at 80-85 °C in 5 to 10 minutes (data provided by Biocatalysts Ltd).

### **1.10.2 Leucine Aminopeptidase (LapA)**

The other commercial enzyme that has been selected is LapA, which is an exopeptidase that catalyses the removal of an N-terminal amino acid from a polypeptide chain and, as the name implies, displays a preference for leucine. LapA originates from a eukaryotic fungus *Aspergillus oryzae*, and is present in the commercial *A.oryzae* peptidase extract – Flavourzyme. LapA was reported to display an optimal temperature at 60 °C and optimal pH of 8.5, while the enzyme was inactivated at temperatures >70 °C (Matsushita-Morita et al., 2011). LapA was predicted to have two glycosylation sites with the N-linked attachment of sugars. As extracellular LapA originates from a eukaryotic fungus, N-linked glycosylation probably plays an important role in the structure and/or function of this enzyme (Imperiali and O'Connor, 1999). The presence of disulphide bonds in LapA was also implicated to be important for the activity (Matsushita-Morita et al., 2011).

There is no crystal structure currently available for LapA. However, the closest structural homologue – leucine aminopeptidase from *Vibrio proteolytica* (AAP) shares 34% sequence identity with LapA (BLAST) (**Figure 1.9**).



**Figure 1.9** Crystal structure of leucine aminopeptidase from *Vibrio proteolytica* (AAP) (pdb: 1AMP). AAP is represented as a homodimer. The essential cofactors –Zn ions- are shown as grey spheres.

The recombinant expression of this natively extracellular exopeptidase might result in the formation of inclusion bodies, as described in the protocols established for AAP (Hartley et al., 2009). Also, the N-terminal propeptide was shown to be a prerequisite for appropriately folded and functional AAP (Bzymek et al., 2004). Therefore, it is very likely that these same principals would apply to LapA such that the propeptide sequence should precede the mature sequence in any recombinant expression constructs of LapA.

### **1.11 Aims and objectives**

The aim of the present study is to engineer industrially relevant proteases to obtain thermolability and cold adaptivity. The two proteins of interest (to Biocatalysts) are LapA a leucine aminopeptidase from *Aspergillus oryzae* and Subtilisin Carlsberg sC from *Bacillus licheniformis*. Improved catalytic activities of these cold-adapted proteases at low and moderate temperatures would allow their replacement of thermostable equivalents in industrial whey protein digestion and would allow the company to use less active enzymes. Such cold-adapted proteases could offer a substantial energy savings during the production process. The added thermolabile character of these proteases would facilitate rapid and effective inactivation of proteolysis (at intermediate steps) and would prevent heat-sensitive protein hydrolysates from gelation.

#### ***Specific objectives are to:***

- Identify structurally viable mutations, using the known crystal structure of Subtilisin Carlsberg, which may lead to improved thermolability.
- Clone and express potential thermolabile mutants of Subtilisin Carlsberg and test for activity at low temperature.
- Determine the three-dimensional structure of LapA using X-ray crystallography.
- Identify structurally viable mutations, using the new crystal structure of LapA, which may lead to improved thermolability.
- Implement the production of the best thermolabile mutants on a pilot scale at Biocatalysts, focusing on yield, stability and optimising specific activity.

## **Chapter 2 MATERIALS AND METHODS**

---

## **2.1 Sequence and structure bioinformatics**

### **2.1.1 Subtilisin Carlsberg (sC) from *Bacillus licheniformis***

The sequences of pre-pro and mature sC were obtained from UNIPROT and Protein Data Bank (PDB), respectively. The accession codes were P00780 for the former and 1R0R for the latter.

Pairwise and multiple sequence alignments were performed using the TCOFFEE (Notredame et al., 2000) sequence alignment package, making use of the MUSCLE algorithm (Edgar, 2004). The BLAST server from the National Centre of Biotechnology Information (NCBI) (Altschul et al., 1997) was employed in order to determine homologous proteins.

The lowest B-factor values, as determined in the X-ray crystal structure, for the selected model (PDBid 3UNX) were analysed using B-FITTER (Reetz and Carballeira, 2007).

For visualizing and rendering multiple sequence alignments ESPript and ENDscript interfaces were used (Robert and Gouet, 2014).

For mutational analysis and annotating three-dimensional models, Accelrys Studio Visualiser (Accelrys Software Inc., San Diego, CA, 2009) and PyMOL (The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC) were used.

### **2.1.2 Leucine aminopeptidases from *Aspergillus oryzae***

LapA was modeled using the online Integrated Protein Structure and Function Prediction Server (IntFOLD) server (Roche et al., 2011). Alternative models were also produced using PHYRE2 (Kelley and Sternberg, 2009) and RaptorX (Kallberg et al., 2012) servers. Tertiary structures of the predicted LapA thermolability mutants were modelled using SWISS-MODEL server (Biasini et al., 2014).

LapA and GAP pocket and cavity analysis was performed on CASTp server (Liang et al., 1998). Rigidity analysis was conducted on KINARI-Web server (Fox et al., 2011).

Pairwise structural alignment was conducted using the Dali Lite pairwise comparison server (Holm and Rosenstrom, 2010). Structural comparisons and superimpositions were also generated using the TM-align server (Zhang and Skolnick, 2005)

For the prediction of LapA melting temperature ( $T_m$ ) the Predictor server (by P.C. Lyu Lab., Institute of Bioinformatics and Structural Biology, National Tsing-Hua University, Hsin-Chu, Taiwan, <http://tm.life.nthu.edu.tw>) was used.

## **2.2 DNA manipulation**

### **2.2.1 Polymerase chain reaction**

PCR was performed in accordance with standard protocols used with Premixed 2X solution of *Taq* DNA Polymerase (Promega, UK) and CloneID™ 1X Colony PCR Master Mix (Lucigen, US). The PCR steps used the following conditions: *Denaturation* – initial denaturation step included heating the sample to 98°C for 30s; *Denaturation* for another 5s at 98°C, *Primer annealing* – 60-65 °C for 30 sec., *Elongation* – at 72 °C for 1 min/kb. A final elongation step, for 3 minutes at 72 °C, was allowed to ensure that all ssDNA was extended. Finally, all samples were incubated at 4 °C.

### **2.2.2 Plasmid DNA purification**

Plasmid DNA purifications from bacterial biomass were performed using the QIAprep Spin Miniprep Kit (QIAGEN), PureYield™ Plasmid Miniprep System (Promega, UK) and PureYield™ Plasmid Midiprep System (Promega, UK) according to the manufacturer's instructions. The cells were harvested from a 3 - 5 ml overnight culture for purification using Miniprep system, and from a 100 mL culture for purification using Midiprep system.

### **2.2.3 DNA extraction from agarose gel**

QIAquick Gel Extraction Kit from QIAGEN was used according to the manufacturer's instructions to purify DNA samples from 1% agarose gels, as follows: 1.5 mL Eppendorf tube containing a slice of agarose gel and 3 volumes of buffer QG (0.5M ammonium acetate, 10 mM magnesium acetate, 1 mM EDTA, 0.1% SDS, pH 8.0) were added followed by incubation at 50 °C water bath for 10 min. The dissolved sample was transferred to QIAquick spin column and centrifuged at  $12.000 \times g$  for 1 min to allow the DNA to bind to the silica matrix. A wash step (0.75 mL of buffer containing ethanol was added to the spin column followed by incubation of 5 min.) was performed. DNA was eluted into a 1.5 mL Eppendorf tube, using 50 µL of 50 °C ddH<sub>2</sub>O following the pre-incubation for 4 min for increased DNA yield.

#### **2.2.4 DNA restriction digest**

Other DNA endonucleases used in this project (*SacI*-HF, *HinfI*) were obtained from NEB (UK) and digestion protocols were performed according to manufacturer's instructions.

#### **2.2.5 In-Fusion molecular cloning**

Recombinant LapA and sC expression constructs were obtained utilising InFusion cloning kit. The In-Fusion™ HD Cloning Kit (Clontech, USA) was composed of CloneAmp HiFi PCR Premix, 5X In-Fusion HD Enzyme Premix, pUC19 Control Vector, linearized (50 ng/μl), 2kb Control Insert (40ng//μl), PCR Clean-up kit, Stellar® competent cells.

Gel-purified PCR products (genes to be cloned) and gel-purified digested vectors were combined with In-Fusion™ enzyme, followed by incubation for 15 min. at 50 °C. 2.5 μL of the reaction mixture was transformed into Stellar® competent cells.

##### **2.2.5.1 pOPINE/pOPINP vector restriction digest and clean-up**

pOPINE and pOPINP vector DNA were digested with *NcoI*-*PmeI* and *KpnI*-*PmeI* endonuclease pairs, respectively. In a 200 μL PCR tube, 50 μL total reaction mix was assembled. 1 unit of enzyme was used to digest 1000 ng of DNA using optimal buffering conditions with CutSmart™ buffer 1X (50 mM Potassium Acetate, 20 mM Tris-acetate, 10 mM Magnesium Acetate, 100 μg/ml BSA, pH 7.9). The reaction mixture was incubated for 1h at 37 °C and the reaction products were analysed on an agarose gel.

The PCR generated amplicons of appropriate gene insert, and linearised pOPINE/P vectors were purified using NucleoSpin Gel and PCR Clean-up Kit (Clontech, UK) according to manufacturer's instructions.

##### **2.2.5.2 InFusion reaction**

Gene inserts with PCR-generated 15bp homologous overlaps were combined with linearized vectors in InFusion reaction. Typical quantities are represented in Table 2.1.

	<b>pOPINE-sC construction</b>	<b>pOPINP-sC construction</b>	<b>Negative control</b>	<b>Positive control</b>
<b>Gene PCR product</b>	1 $\mu$ L (80ng)	0.8 $\mu$ L (80ng)	-----	2 $\mu$ L of control insert
<b>pOPINE/P linear</b>	4 $\mu$ L (80ng)	5 $\mu$ L (80ng)	4 $\mu$ L	1 $\mu$ L of pUC19 vector
<b>5XInFusion Mix</b>	2 $\mu$ L	2 $\mu$ L	2 $\mu$ L	2 $\mu$ L
<b>ddH<sub>2</sub>O</b>	3 $\mu$ L	2.2 $\mu$ L	4 $\mu$ L	5 $\mu$ L

**Table 2.1** InFusion cloning assembly in 10  $\mu$ L reaction.

The assembled reactions were allowed to proceed for 15 min. at 50 °C.

1  $\mu$ L of completed reaction mixture was transformed to 50  $\mu$ L of chemically competent Stellar<sup>TM</sup> (Clontech, UK) cells.

100  $\mu$ L of regenerated cells were plated on agar supplemented with ampicillin (100mg/mL).

### 2.2.6 Quickchange site-directed mutagenesis protocol

LapA thermolability mutants were generated using Quickchange II Lightning (Agilent) site-directed mutagenesis kit according to manufacturer's instructions. In brief, mutagenic primers were designed using an online tool: <http://www.genomics.agilent.com/primerDesignProgram.jsp>. Site-directed mutations were introduced by performing mutagenic PCR as a first step, utilising wild-type LapA expression cassette as a template. Amplification products were treated with a DpnI enzyme to digest the parent DNA, followed by transformation to XL1-Blue Supercompetent cells. Several *E.coli* transformants from each mutant were propagated. The extracted plasmid DNA was verified to contain the desired mutation by Sanger sequencing. 100 mL of *E.coli* cultures were cultivated to obtain the required biomass for medium scale plasmid DNA preparation and subsequent linearisation with *SacI* endonuclease before transforming into competent *P.pastoris* X33 strain.

## 2.3 Electrophoresis

### 2.3.1 Agarose gel electrophoresis

1% w/v agarose gels were prepared by dissolving the respective amount of agarose in 1X TAE buffer (40mM Tris, 20mM acetic acid, 1mM EDTA) buffer, supplemented with

SYBR® Safe DNA gel stain (Life Technologies, UK). Samples were prepared using 1X Purple Dye (NEB, UK) (2.5% Ficoll®-400, 10mM EDTA, 3.3mM Tris-HCl, 0.08% SDS, 0.02% Dye 1, 0.001% Dye 2, pH 8.0). DNA bands were separated using horizontal gel tank Sub-gel midi Submarine (Fisher, UK) by running the gel for 40 min under 120V. Bands were visualised under UV light using a benchtop UV gel doc system (UVP, Cambridge, UK).

### **2.3.2 SDS – PAGE gel electrophoresis**

Two gel running systems have been employed in this work; either precast gels (NuPAGE® Novex® 4-12% Bis-Tris Gels, 1.0 mm, 15 well) together with the Invitrogen NuPage® Novex® Gel System or Tricine-glycine custom-poured gels together with Bio-Rad Mini-PROTEAN® Tetra cell system.

All protein samples were prepared by taking 15 µL of protein sample and mixing with 5 µL of NuPAGE® SDS Sample Loading Buffer (Life Technologies, UK). Before loading, the mixture was incubated in a TechneDri-block® DB-2A at 95°C for 3 minutes. The electrophoresis was allowed to proceed for 50 min. under 200V. Staining of the gels was performed using InstantBlue® (Generon, UK) quick stain, incubating for 15 min.

### **2.3.3 Western Blot Analysis**

Western blotting was used to investigate the expression of recombinant Subtilisin Carlsberg and LapA proteases, as each construct tested possessed a hexahistidine tag at the N or C-terminus, which was used for cross-reactivity with 6xHis specific antibody. Proteins separated by SDS-PAGE were blotted onto a nitrocellulose membrane using a semi-dry method, as follows: thin filter papers were immersed in Transfer buffer (25 mM Tris, 192 mM glycine, 10% methanol) for 2 min. The filters were placed on the Blotter. The membrane was also kept in Transfer buffer (TB) for 5 min and placed on the top of filter papers. Analysed SDS-gel was kept in TB for 5 min and placed on the top of the membrane. Another three TB pre-soaked filter papers were placed on the top. Some TB was pipetted on the top to keep the system wet; the sandwich was rolled over with the pen to remove trapped air bubbles. The Blotter was left to run for 1h and 30 min at 15V (120mA). Blocking solution (TBST (Tris Buffered Saline with Tween® 20: 20 mM Tris pH 7.5, 150 mM NaCl, 0.1% Tween 20) + 5% skimmed milk) was prepared. The membrane was inspected for sufficient transfer of proteins from the SDS-gel. The membrane was incubated in Blocking solution with gentle rocking for 1h at ambient temperatures or overnight at 4 °C. 5-10 mL of Blocking solution was saved, and Histidine-specific antibody was added according to the ratio 1:500, or 10 µL in 5 mL. The membrane was washed 3 times with TBST, changing to fresh buffer

each time. The solution from step 11 was poured over the TBST-washed membrane and left gently rocking for 1h at RT. The membrane was then washed 3 times with TBST in the same way as in step 12 and repetitively placed on cling film to get rid of any residual buffer. Chemiluminescent Clarity™ Western ECL Substrate (Bio-Rad) was applied with a pipette, followed by 5 min incubation time at RT and another layer of cling film was placed over the membrane for longer term storage. The resulting bands were visualized using an ImageQuant LAS 4000 (GE Healthcare).

## **2.4 Quantification methods**

### **2.4.1 Quantification of recombinant DNA**

A NanoDrop® 1000 (Thermo Scientific) spectrophotometer was used to measure the concentration of linear and plasmid DNA fragments at 260 nm. Purity was estimated by the ratio 260/280 nm.

### **2.4.2 Quantification of recombinant LapA**

LapA proteins were quantified either by Bradford assay (Bio-Rad, USA), using BSA as a standard or spectrophotometrically, employing a NanoDrop® ND-1000 (Thermo Scientific) using the LapA absorbance extinction coefficient ( $23,505 \text{ M}^{-1}\text{cm}^{-1}$ ). The concentration of recombinant LapA in crude *P.pastoris* supernatant could only be determined by Bradford assay due to impurities interfering with spectrophotometric measurements.

## **2.5 Microbiology methods**

### **2.5.1 Bacterial transformation**

Unless otherwise stated, the transformation protocol was performed as follows: Competent cells were taken out of the -80 °C freezer, briefly centrifuged and kept on ice for 5 min to thaw. 5-10 ng of plasmid DNA was mixed with 50-100 µL of competent cells. The mixture was kept on ice for 30 min. Heat shock was carried out for 30 s at 42 °C, and then the mixture was placed back on the ice for 2 min. SOC outgrowth medium (1X SOC Outgrowth Medium: 2% Vegetable Peptone, 0.5% Yeast Extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl<sub>2</sub>, 10 mM MgSO<sub>4</sub>, 20 mM Glucose) was added to the mixture and then placed in a 37 °C shaker incubator for 1 h at 220 rpm. The bacterial solution was transferred onto an agar plate

containing selection marker and spread evenly across the plate and incubated at 37 °C overnight.

### **2.5.2 Glycerol stocks**

500 µL glycerol stocks were prepared by mixing 375 µL cell culture ( $OD_{600} = 0.4-0.6$ ) and 125 µL of sterilized 32% glycerol, combined in a sterile cryogenic or 1.5 mL Eppendorf tube, followed by freezing at -80 °C.

### **2.5.3 Preparation of competent cells**

LB was inoculated with a single colony of bacteria and cultured overnight at 37 °C. The culture was further grown the next day until  $OD_{600} = 0.4$ , following inoculation using the overnight culture diluted 100-fold. Cell culture was placed on ice for 20 min, followed by centrifugation at  $3,000 \times g$  for 10 min. at 4 °C. Cells were resuspended in ice-cold 0.1M  $CaCl_2$  followed by incubation for 30 min. and subsequent centrifugation at  $3,000 \times g$  for 10 min. at 4 °C. The supernatant was discarded and the pellet resuspended in ice-cold 0.1M  $CaCl_2$  supplemented with 15% v/v glycerol. 50 µL aliquots have been made and frozen in liquid nitrogen for storage at -80 °C.

## **2.6 Recombinant Protein Expression in *E.coli***

### **2.6.1 pOPIN expression vectors**

Recombinant protein expression was conducted, taking advantage of the T7 expression system utilized in the pOPIN family of vectors (OPPF, UK) and pET-21a vector (Novagen). These vectors provide a 6X His-tag either at the N-terminus (pET-21a and pOPINF vectors) or the C-terminus (pOPINE vector), as well as providing a pelB signal peptide (pOPINP vector). The recombinant tag can be removed by enterokinase or Rhinovirus 3C protease. The antibiotic resistance marker in these vectors is ampicillin.

### **2.6.2 Bacterial expression**

BL21 (DE3) *E. coli* expression strain and its derivatives, Lemo21 (NEB), B834 (NEB), Rosetta (NEB), SoluBI21 (AMSBio) were used. Both Luria-Bertani (LB) Broth and M9 Minimal Media were used to grow the bacterial cultures (Table 2.2).

Components (LB)	Components (M9) in 1L
1.0% <u>Tryptone</u>	Na <sub>2</sub> HPO <sub>4</sub> (anhydrous) 6g, KH <sub>2</sub> PO <sub>4</sub> 3g, NaCl 0.5 g, NH <sub>4</sub> Cl 1g
0.5% Yeast Extract	100mM CaCl <sub>2</sub> 1mL
1.0% Sodium Chloride (NaCl)	1M MgSO <sub>4</sub> 1mL
pH 7.0	Glycerol 0.3%

**Table 2.2** Components of LB broth (left) and M9 minimal media (right) used for bacterial growth

### 2.6.3 Small-scale expression

All preliminary recombinant LapA over-expression trials were carried out using 10 mL cultures. A single BL21 (DE3) colony harbouring recombinant LapA expression construct was transferred to 2 mL LB broth and incubated at 37 °C shaker (220 rpm) overnight. 200 µL of the overnight inoculum was transferred to 10 mL of fresh LB media. The optical density (OD) at 600 nm was regularly monitored to attain a value between 0.5 and 1.0 absorbance units, after which IPTG (1 mM final) was added to induce protein expression. The pre-induction sample was taken for further analysis on SDS – PAGE. The induced cultures were then incubated for 16-20h at either 37 °C or 20 °C. Following the harvest at *16,000 x g*, cells were lysed with the BugBuster® Master Mix (Novagen) and soluble versus insoluble protein fractions were analysed on SDS-PAGE.

Minimal media expression trials were performed as follows: A colony of the SoluBL21 was inoculated into 1-2 ml of M9 minimal media with ampicillin (100 µg/mL final). The inoculum was left to grow overnight at room temperature in a shaker/ incubator at 200 rpm. 300 µL of the overnight culture was inoculated into 5mL of fresh M9. Cells were left to grow at room temperature until an OD<sub>600</sub> of at least 0.4, followed by addition of IPTG to a final concentration of 1 mM. Cells were incubated overnight at room temperature, in a shaker/incubator at 200 rpm. Post-induction sample was taken, and the cells were pelleted by centrifugation at *16,000 x g*.

#### **2.6.4 Disruption of bacterial cells**

Mechanical (ultrasonication) and chemical cell disruption were used to disrupt the cells. A Sonifier® and an S-450D digital ultrasonic processor (Branson) equipped with a double step 1/8" (3.2mm) ultrasonic probe immersed into the cell solution, were used for all experiments. For chemical cell disruption, BugBuster® Master Mix (Novagen) was chosen.

##### *Ultrasonication:*

The instrument was set to 40% amplitude, 5-second pulse, for 2 minutes only once. For small scale preparations (<1 mL) a 3.2 mm diameter probe or 40% amplitude, 5-second pulse, for 5 minutes, 3 times was used. For large scale preparations (> 30 mL), a 6 mm diameter, tapered 1/4" (6.4mm) probe was used.

##### *BugBuster® treatment:*

5 mL of BugBuster® was used for approximately 1g of wet cell paste, followed by 10-20 min incubation on a rotating mixer at RT.

Once the crude cell lysate was obtained, 50 µL sample was taken. The cell lysate was centrifuged for 10 min. at  $16,000 \times g$  and the clear supernatant was sampled. The two samples were retained for analysis on SDS-PAGE.

#### **2.6.5 Periplasmic extraction**

The recombinant expression of Subtilisin Carlsberg (sC) carried the pelB signal peptide, which directed the protein to the periplasm where it was subsequently extracted using osmotic shock treatment.

In this method, the *E. coli* cells carrying the sC protein were treated with an ice-cold osmotic shock solution (20-25% w/v sucrose, 5 mM EDTA, 20 mM Tris, pH 8.0). The resulting suspension, the 'sucrose' fraction, was centrifuged at  $8,000 \times g$ . The pellet was re-suspended in 5-10 mM MgCl<sub>2</sub> to complete the osmotic shock. The suspension was centrifuged at  $8,000 \times g$ . The resulting supernatant was the 'periplasmic' fraction.

### **2.7 Protein purification**

#### **2.7.1 Immobilised metal affinity chromatography (IMAC)**

All recombinant proteins (sC, LapA) were tagged with a 6x-Histidine tag to facilitate purification using a HiTrap<sup>™</sup> affinity column.

HiTrap<sup>TM</sup> 5mL pre-packed affinity columns (GE Healthcare) were used in conjunction with a ÄKTA FPLC system to isolate the desired proteins. The buffers used are provided in Table 2.3.

Binding buffer (A)	Elution buffer (B)
50mM Tris pH=8.0	50mM Tris pH=8.0
500mM NaCl	500mM NaCl
30mM imidazole	500mM Imidazole

**Table 2.3** Buffers used in IMAC for HiTrap<sup>TM</sup> 5mL columns.

Purification followed the following steps unless otherwise stated. The NiNTA column was charged using 0.1 NiSO<sub>4</sub> and washed with 10 column volumes (CV) of Milli-Q water to remove residual Ni<sup>2+</sup> ions. The column was then pre-equilibration with 5-10 CV of Buffer A. Cell lysate was loaded, following filtration through a Millipore 0.2 µm sterile filter. The unbound and non-specifically bound proteins were washed with 3-5 CV Buffer A. The bound recombinant protein was eluted with 5 CV of 100% Buffer B, or a 10 CV linear gradient was applied. Aliquots representing whole cell lysate, column flow-through (unbound protein), wash stages and final elution were taken, and the contents were visualised using SDS-PAGE.

### **2.7.2 Purification and refolding of *refLapA* protein from *E.coli* inclusion bodies**

Pellet with inclusion bodies were re-solubilised in 8M Urea buffer (50mM Tricine pH = 8.0, 500 mM NaCl, 8M Urea) gently agitating for 1 hour. The non-dissolved debris was separated by centrifugation for 20 min at 16,000 *x g*. IMAC: the supernatant was loaded on HisTrap<sup>TM</sup> FF 5 ml (GE Healthcare, UK) nickel affinity column pre-equilibrated with wash buffer (50 mM Tricine pH = 8.0, 500 mM NaCl, 8M Urea) and re-solubilised *refLapA* eluted in 50 mM Tricine pH = 8.0, 500 mM NaCl, 8M Urea and 200 mM Imidazole. *refLapA* was allowed to slowly refold by reducing the concentration of Urea in the buffer: IMAC-eluted fraction was dialysed against refolding buffer (10 mM Tricine pH = 8.0, 1 mM ZnCl<sub>2</sub>) for 20h at 4 °C. The *refLapA* was further purified on HiLoad Superdex<sup>TM</sup> 16/60 200pg (GE Healthcare, UK) gel filtration column pre-equilibrated with 20 mM Tricine pH = 8.0, 1 mM ZnCl<sub>2</sub>. Collected 2 ml fractions with most specific activity were pooled and concentrated on Vivaspin 15R – 10,000 Dalton Centrifugal Filter Units (Sartorius, UK) containing HydroSart® membrane.

### 2.7.3 Native LapA (*flav*LapA) purification from Flavourzyme

Flavourzyme 500MG commercial *Aspergillus oryzae* peptidase preparation was a gift from Biocatalysts Ltd. (UK). 2 ml of 2 g/L (total protein content) Flavourzyme solution was 0.22  $\mu$ M filtered and loaded on HiLoad Superdex<sup>TM</sup> 16/60 200pg (GE Healthcare, UK) gel filtration column equilibrated in 20 mM Tris pH = 8.0, 500 mM NaCl. This way a mixture of 8 peptidases present in Flavourzyme (Merz et al., 2015) could be crudely separated based on their molecular weight. Fraction containing most of the native LapA (*flav*LapA) from the latter step was loaded on HiPrep<sup>TM</sup> 26/10 53mL Desalting column and buffer exchanged into 20 mM Tris pH = 8.0 prior to loading on HiTrap<sup>TM</sup> Q HP anion exchange column pre-equilibrated with 20 mM Tris pH = 8.0. Column was washed with 10 CV of equilibration buffer or until no trace of protein appeared in the eluate. *flav*LapA was eluted with increasing gradient (0-1M) of NaCl (20 mM Tris pH = 8.0, 1M NaCl). To remove residual brown pigment associated with the protein, *flav*LapA from previous anion exchange step was loaded on HiTrap<sup>TM</sup> Phenyl HP 5ml (GE Healthcare, UK) hydrophobic interaction (HIC) column pre-equilibrated with high salt buffer (20mM Tris pH = 8.0, 1.5M Ammonium sulphate) and eluted with low salt buffer (20mM Tris pH = 8.0). This step was required repeating several times to remove traces of pigment prior to crystallisation. Final two purification steps of native LapA involved: 1) HIC-purified fraction was heat-treated to 67 °C for 20 min in the presence of 500mM NaCl to remove any residual contaminating *A.oryzae* proteins ( $\alpha$ -amylase and neutral protease I). Protein precipitate (contaminants) was separated from supernatant (*flav*LapA) by centrifugation at 16,000  $\times$  g for 10 min, 2) the successive supernatant was loaded on HiLoad Superdex<sup>TM</sup> 16/60 200pg (GE Healthcare, UK) gel filtration column equilibrated in 20mM HEPES pH = 8.0. Eluted homogenous *flav*LapA from Flavourzyme was concentrated and stored at -20 °C for further analysis.

### 2.7.4 Protein concentration

All proteins were concentrated using Amicon Ultra -15 Centrifugal Filter Units – 10,000 NMWL (Merck Millipore), Amicon Ultra-0.5 mL Centrifugal Filter Units – 10,000 NMWL (Merck Millipore) or Vivaspin 15R – 10,000 Dalton Centrifugal Filter Units (Sartorius).

## 2.8 Protein crystallisation

### 2.8.1 High throughput crystallisation screening

The OPPF has commercial blocks of crystallisation screens available (containing 96 different conditions for each 96-well plate used). Crystallisation was achieved by the vapour

diffusion sitting drop method (Chayen, 1998a) in 96-well plates. Native and wild-type LapA crystallisation screens were performed using:

PACT premier HT-96 (Molecular dimensions)

JCSG-plus<sup>TM</sup> Screen (Molecular dimensions)

INDEX HT (Hampton Research)

Morpheus® HT-96 (Molecular dimensions)

PEG/Ion HT (Hampton Research)

WIZARD (Molecular dimensions)

The resultant protein crystals were mounted in microscopic loops and cryo-protected with mother liquor supplemented with 20-25% glycerol (v/v) or ethylene glycol.

### **2.8.2 24-well plate crystallisation screening**

Several 24-well crystallization plates were set up in order to optimize conditions identified in the 96-well plate screen to obtain diffraction quality crystals. Crystallization was achieved by vapour diffusion hanging drop method (Chayen, 1998b). In brief, 1 µl of purified protein solution and 1 µl of crystallisation solution were combined on a clean 22mm square coverslip. The cover slip was placed over top of the reservoir chamber containing the screening reagents, with the crystallisation drop facing the reservoir solution. Vacuum grease was applied around the perimeter of the edge of each reservoir chamber providing a sealant for the coverslip over the well. The crystallisation trays were incubated at RT and monitored under a light microscope at regular intervals (typically, after 1 day, then 3 days, 1 week, weekly then monthly) for the appearance of crystals. The stock solutions used were 4M Ammonium Phosphate, 1M Sodium Acetate, 1M Citrate, 2M ZnCl<sub>2</sub>, 2M KSCN, 5M NaCl, 2M HEPES. All reagents were purchased either from Molecular Dimensions or Sigma-Aldrich, and the solutions were made using Milli-Q pure water.

#### **2.8.2.1 Crystallisation of native *flav*LapA**

*flav*LapA from Flavourzyme was crystallised at 6 mg/ml in 20mM HEPES, pH = 8.0. Using robotics at OPPF, *flav*LapA crystallization trials were set up in 200 nl (100 nl protein plus 100 nl reservoir solution) drops using JCSG-plus<sup>TM</sup> HT-96 screen (Molecular dimensions, UK). Crystals of *flav*LapA appeared in 3-4 days under F2 condition (0.1M Citrate, pH = 5.0, 3.2 M Ammonium sulphate) at 21 °C.

### 2.8.2.2 Crystallisation of recombinant LapA wild-type proteins

Crystallisation screening was carried out using HTP robotics at Oxford Protein Production facility (OPPF, UK) (96 well commercial sparse matrix screens, sitting-drop vapour diffusion method) and crystallisation optimisation was carried out in the home laboratory using 24 well, hanging-drop vapour diffusion.

A pre-crystallisation test (PCT) for each purified protein was performed prior to crystallisation screening. The PCT Testing Kit (Hampton Research Ltd) was used according to the following protocol: 4 reservoirs are used in the Greiner crystallization plate to pipette approximately 100  $\mu$ L of screening solutions A1, A2, B1 and B2. Half of the microliter of reservoir solution and protein solution are combined in one drop. Wells used are sealed with clear sealing foil. Drops were monitored under the microscope after 15 min. If A1 and/or A2 wells contained a light granular precipitate, then protein was deemed close to the right concentration. If A1 and/or A2 wells contained heavy precipitate, then most likely the protein was too concentrated and should be diluted to 50-70% of existing concentration. If the drops remained clear, then the protein concentration is too dilute and further concentration is needed to increase the concentration at least two-fold.

*wtproLapA*, *wtLapA* recombinant proteins were crystallised at 13 mg/ml and 11 mg/ml concentrations, respectively. Both proteins were buffer exchanged into 20mM Tricine pH = 8.0, 1mM ZnCl<sub>2</sub> prior to crystallisation. *wtproLapA* crystals with dimensions 0.5  $\times$  0.04  $\times$  0.005 mm formed under F10 condition (0.1 M Bis-Tris, pH = 5.5, 0.2 M NaCl, 25% PEG 3,350) in the Index HT™ screen (Hampton Research, USA) in 3 days at 21 °C. *wtLapA* crystals were produced in 24-well plate format at room temperature under previously identified crystallisation conditions for *flavLapA* from Flavourzyme (0.1M Citrate, pH = 5.0, 3.2 M Ammonium sulphate).

### 2.8.2.3 Crystallisation of LapA thermolability mutant proteins

Crystallisation of LapA mutants were conducted under previously determined Index HT™ F10 (0.1 M Bis-Tris, pH = 5.5, 0.2 M NaCl, 25% PEG 3,350) condition. Crystallisation drops were set-up in 24-well Grainer plate by hanging-drop vapour diffusion method. LapA mutant protein solutions were prepared at very high 30 – 50 mg/ml concentration. Crystals were visible after 24 h incubation at 21 °C.

### **2.8.1 Data Collection, Structure Determination & Refinement**

Prior to data collection, crystals were cryo-protected by quick transfer through a mixture of mother liquor and 20% ethylene glycol (with the exception of *wtp*roLapA, which had 25% PEG in the mother liquor) and flash cooled to 100K in liquid nitrogen.

Diffraction data were collected at Diamond Light Source (UK), on beamlines I02 and I03. The diffraction data were recorded, by fine slicing (0.1° degree oscillation), on a Pilatus 6M detector. Indexing and integration were completed automatically by *XDS* (Kabsch, 2010) and *xia2* (Winter et al., 2013).

All the LapA structures were solved by maximum likelihood molecular replacement with PHASER (McCoy et al., 2005), using the coordinates of an established LapA homology model; LapA homology model was generated using the IntFold2 server (Roche et al., 2011), which identified aminopeptidase from *Vibrio sp.* (34% sequence identity) as a closely related structural homologue (PDBid: 1RTQ), as the template. Resulting LapA models were manually built, fitting to the resulting electron density maps, using COOT (Emsley and Cowtan, 2004). All structures were refined using PHENIX (Adams et al., 2010). MolProbity4 was used for structure validation. Structural images were prepared using PyMOL.

## **2.9 Protein thermal and spectroscopic analysis**

### **2.9.1 Differential Scanning Fluorimetry (DSF) measurements**

In brief, this assay measures increase in fluoresce signal of Sypro-Orange dye, which binds to increasingly exposed hydrophobic patches of unfolded protein populations arising from increased temperature of the system. The obtained fluoresce curve midpoint value  $T_m$  corresponds to the temperature of half-unfolded protein population. 50  $\mu$ L samples for DSF measurements were prepared in the following order: 2  $\mu$ L of concentrated 5,000X Sypro-Orange (Invitrogen, Carlsbad, US) dye was diluted with 38  $\mu$ L of sample buffer (20 mM Tricine, pH=8.0, 1 mM  $ZnCl_2$ ). 1  $\mu$ L of at least 5 – 10 mg/ml LapA protein was mixed with 42  $\mu$ L of sample buffer and 7  $\mu$ L of previously diluted dye, avoiding the formation of bubbles. The prepared samples were transferred to 96 – well plate and measurements were recorded on Mx3000P qPCR System (Agilent Technologies, CA, US) within 25 – 90 °C interval at a scanning rate of 1 °C/min. Melting temperature ( $T_m$ ) was transition midpoint

value between the start point and the maximum point which was calculated using the manufacturer's software.

### 2.9.2 Differential Scanning Calorimetry (DSC) measurements

LapA protein samples were analysed using NanoDSC 6300 instrument (TA instruments, New Castle, DE) equipped with 300  $\mu$ L platinum cell. Protein concentrations were in the range of 0.5 – 1.5 mg/ml. Heat flow and endothermic peaks were monitored at a scanning rate of 1  $^{\circ}$ C/min. Protein endotherm peaks were adjusted to their respective concentration and molecular weight and converted to molar heat capacity. Analysis with NanoAnalyze v3.7 software package yielded protein melting point ( $T_m$ ) and thermodynamic parameters ( $\Delta H$ ,  $\Delta S$ ).

### 2.9.3 Circular Dichroism (CD) spectroscopy of LapA proteins

The circular dichroism spectra for all resultant proteins were recorded in the range of 360 to 190 nm at 20  $^{\circ}$ C using a Chirascan series 800 spectrophotometer. Measurements were taken using a 1 nm step size in a 0.1 mm path length quartz cell. CD data was averaged from independent 5 scans for each sample. Samples of *refLapA*, *flavLapA*, *wtLapA*, *wtproLapA* and *wt $\Delta$ proLapA* proteins were analysed at 36  $\mu$ M, 30  $\mu$ M, 21  $\mu$ M, 25  $\mu$ M and 40  $\mu$ M concentrations, respectively, buffered in 20 mM HEPES, pH = 8.0.

## 2.10 LapA enzymatic assays

### 2.10.1 Lap enzymatic assay – 3 mL format

LapA specific activity was determined spectrophotometrically using a synthetic substrate (Leu-*p*NA) and monitoring the formation of the *p*-nitroanilide moiety (*p*NA) as absorbance at 405nm ( $\epsilon = 9.9 \text{ mmol}^{-1}\text{cm}^{-1}$ ) in a cuvette with 1 cm path length. Peptidase assays were performed on CE1020 1000 Series spectrophotometer (Cecil Instruments).

The reagents used for the assay were as follows:

1. 0.05 M Phosphate Buffer, pH 7.2 (at 37 $^{\circ}$ C) – stable for 3 months at room temperature.
2. 0.024M L-Leucine-*p*-nitroanilide -stable for 3 months when frozen

120.6 mg of L-Leucine-*p*-nitroanilide (Sigma, UK) was dissolved in approximately 15 mL of methanol and made up to a final volume of 20 mL with phosphate buffer.

Calculation of specific activity:

$$U_{g^{-1}} = \frac{A_{405nm}^{min} \times V_t \times \text{dilution} \times 1000}{[E] \times V_s \times E}, \text{ where}$$

$V_t = 1\text{ml}$

$V_s = 0.333\text{ ml}$

1000 = convert from mg to g in enzyme concentration

$[E]$  = enzyme concentration in mg/ml

$E$  = extinction coefficient of L-leucine and p- nitroaniline (9.9 mmol/cm)

Unit definition:

One unit will hydrolyze 1.0  $\mu\text{mole}$  of Leu-pNA to L-leucine and p-nitroaniline per minute at

pH 7.2, at 37 °C, and 24 mM substrate concentration.

### **2.10.2 Lap enzymatic assay – 300 $\mu\text{L}$ microplate assay**

All the reaction conditions and specific activity calculations were equivalent to 3 mL assay format, except the reaction volume was reduced to 300 $\mu\text{L}$ . This was necessary to adopt LapA assay to measurements in a plate format. Greiner CELLSTAR® 96 well plates were used along with SpectraMax Plus 384 Microplate Reader (Molecular Devices, CA, USA) to measure the increase in absorbance at 405 nm. Rainin electronic multichannel pipettes were utilized to dispense 10  $\mu\text{L}$  of enzyme followed by programmed 3 mixing cycles to ensure even distribution of reactants. Absorbance values were recorded for 60 seconds on SoftMax Pro Software (Molecular Devices, CA, USA).

### **2.10.3 LapA thermolability assay**

LapA enzymes were diluted to a concentration, which would give a change in absorption between 0.1 and 1.0 for the most accurate measurement. These diluted aliquots were incubated at 5 different temperatures: 37, 50, 60, 70 and 80 °C. The samples were withdrawn every 15 minutes and cooled either on ice for 3 minutes (3 mL cuvette assay) or 4 °C PCR thermo-block for 5 minutes (300  $\mu\text{L}$  microplate assay) prior to measuring residual specific activity. The cooling step was required to normalize reaction kinetics of the sampled enzymes from different temperatures.

#### **2.10.4 LapA enzyme kinetic assay**

Enzyme reaction kinetic data was obtained by measuring LapA reaction velocity at increasing concentrations (0.008 – 0.8 mM) of LPNA at 37 °C. The protocol was analogous to LapA activity plate assay, except the LPNA substrate varied and was not in excess. Formation of the liberated *p*-nitroanilide product was monitored for 60 seconds from initiation. The increase in absorbance was recorded every 5 seconds. Reaction rates were calculated from slopes within first 40 seconds, where the reaction response was linear. Reaction velocity (mM/sec) versus initial concentration of LPNA in the reaction was analysed by non-linear regression using Graph-Prism software, and best-fit values for  $V_{\max}$  and  $K_m$  were obtained. The concentration of LapA active sites ( $E_a$ ) was equivalent to a molar concentration of LapA added to the reaction (mM). Catalytic constant -  $k_{\text{cat}}$  ( $\text{s}^{-1}$ ) – was defined as  $V_{\max}/E_a$  ratio. Catalytic efficiency was calculated as  $k_{\text{cat}}/K_m$  ratio.

#### **2.11 Industrial pasteurisation for LapA protein inactivation**

Industrial small-scale pasteuriser Armfield FT74XTS (HTST/UHT) was utilised to perform industrial pasteurisation trials with Flavourzyme and recombinantly expressed LapA proteins. The instrument was capable of 10L/h sample feeding rate. LapA protein samples were diluted to the concentration required to give a change of absorbance between 0.1 – 0.5 in standard 3 mL LapA cuvette assay. The instrument was set to either 15 or 30 second holding time for a range (65 – 80°C) of different temperatures. Samples were collected following 15 – 30 second pasteurisation at the predetermined temperatures. Residual specific activity was measured by a standard LapA assay following cooling the enzyme sample to the room temperature.

#### **2.12 Other analyses**

##### **2.12.1 Deglycosylation with PNGase F**

The glycoproteins were processed using peptide -N-Glycosidase F, also known as PNGase F, an amidase that cleaves between the innermost GlcNAc and asparagine residues of high mannose, hybrid, and complex oligosaccharides from N-linked glycoproteins.

##### *Denaturing Reaction Conditions:*

20 µg of glycoprotein, 1 µl of 1X Glycoprotein Denaturing Buffer (0.5% SDS, 40 mM DTT) and H<sub>2</sub>O were combined to make a 10 µl total reaction volume. Glycoprotein was denatured by heating reaction at 95°C for 5 minutes. Reaction chilled on ice and centrifuged for 10

seconds. Total reaction volume made to 20 µl by adding 2 µl 1X GlycoBuffer 2 (50 mM Sodium Phosphate, pH 7.5), 2 µl 10% NP40 and 6 µl H<sub>2</sub>O. 1 µl of PNGase F added followed by gentle mixing. Reaction incubated at 37°C for 1 hour. Reaction analysed by SDS-PAGE.

#### *Non-Denaturing Reaction Conditions:*

20 µg of glycoprotein, 2 µl of 10X GlycoBuffer 2 (50 mM Sodium Phosphate, pH 7.5) and H<sub>2</sub>O were combined to make a 20 µl total reaction volume. 2 µl of PNGase F added followed by gentle mixing. Reaction incubated at 37°C for 24 hours. Reaction analysed by SDS-PAGE.

### **2.12.2 DNA sequencing**

All constructs were verified by Sanger sequencing before expression trials at the Source Bioscience Facility, Oxford.

### **2.12.3 Mass Spectrometry**

All MS analyses were carried out at the University of St. Andrews BBSRC Mass Spectrometry and Proteomics Facility. Protein identification was performed by conventional tryptic digest of the shipped SDS gel band, followed by MALDI MS-MS analysis of the resulting peptides.

## **2.13 Recombinant cloning and expression of LapA in eukaryotic host – *Pichia pastoris***

### **2.13.1 Strains, plasmids and culture conditions**

*P.pastoris* X33 strain was obtained from Biogrammatix, Inc (Carlsbad, CA, USA). Secretory expression vector - pJexpress 912 – was provided by DNA 2.0 (USA).

*E.coli* BL21 (DE3) (*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHI* Δ*EcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*) expression strain and DH5α (*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*) cloning strain were purchased from New England Biolabs (NEB,UK).

Primers used in this study were obtained from Sigma (UK). Zeocin<sup>TM</sup> was purchased from Invitrogen (USA). All other reagents were purchased from Sigma (UK) and were of analytical grade.

### **2.13.2 Construction of recombinant expression vectors and transformations**

Recombinant LapA open reading frames (Genbank accession: XP\_001825745.1) were codon-optimised and cloned into the pJ902 vector, by DNA 2.0, using the *EcoRI* and *NotI* restriction sites. In this way, an expression construct, harbouring the LapA native signal peptide and propeptide sequences (pJ-npro-LapA), was obtained. A second expression construct, containing a deletion of the LapA propeptide (pJ-Δpro-LapA) was produced, utilising the QuickChange™ II site-directed mutagenesis kit from Stratagene (USA). Deletion of the propeptide sequence was achieved by performing PCR with designed primers 1; 5'-GGACGCTATCTGGATAAGCTAATGCGCTTGCC-3' and 2; 5'-GGCAAGCGCATTAGCTTATCCAGATAGCGTCC-3', using the pJ-npro-LapA construct as a template.

Following *SacI* linearization and clean-up, recombinant plasmids were transformed into an X33 expression strain by electroporation (1500 V single pulse in a 2 mm electroporation cuvette). Immediately following electroporation, the cells were resuspended in 0.5ml 1M sorbitol and 0.5ml YPD (1% yeast extract Y, 2% soy peptone P, 2% dextrose D) and incubated for 2h at 30°C under gentle agitation. The revived cells were then plated onto YPD agar (1% yeast extract, 2% soy peptone, 2% dextrose, 2% agar), containing either 100 µg/ml or 2000 µg/ml Zeocin™ and incubated at 30 °C 2-4 days. The resulting *P.pastoris* colonies were verified, by PCR, for appropriate genomic integration using the method previously described (Looke et al., 2011). In each case, 10 transformants were screened, picking the largest, single colonies. Each single colony was patch-plated onto a separate YPD agar plate and residual cells were re-suspended in 100µL lysis solution (200 mM lithium acetate, 1% SDS), followed by incubation at 70 °C for 15 min. Following 96% ethanol precipitation, the DNA pellet was re-suspended in 100 µL sterile water and 1µL (in each case) was used in each subsequent genomic PCR. Clone ID (Lucigen, USA), PCR pre-mix and a pair of AOX primers (AOX1-F; 5'-GACTGGTTCCAATTGACAAGC-3' and AOX1-R; 5'-GCAAATGGCATTCTGACATCC-3') were used, under the following conditions: 1 cycle of 98 °C for 2 min, 30 cycles at 98 °C for 30 s, 56 °C for 30 s, 72 °C for 1 min, finishing with 72 °C for 5 min. The PCR products of predicted size were visualised on a 1% agarose gel.

### **2.13.3 Over-expression of recombinant LapA in *P.pastoris***

Recombinant proteins were over-expressed in a *P. pastoris* X-33 strain (Mut<sup>+</sup>), under control of the *P<sub>AOX1</sub>* methanol inducible promoter. Shake flask expression trials were conducted in BMGY (1% yeast extract, 2% soy peptone, 1% glycerol, 1.34% YNB\*, 4E-5% biotin,

100mM potassium phosphate, pH 6.0) and BMMY (1% glycerol replaced by 0.5% methanol). Genomic PCR-verified *P.pastoris* clones were inoculated into 50 ml BMGY medium in 250 ml baffled conical flasks and incubated in a shaker (250 rpm) for 16h at 28 °C. Once the OD<sub>600</sub> reached 2-8, an equivalent amount of cells (corresponding to OD<sub>600</sub> of 1) were harvested and re-suspended in 100 ml of BMMY and shaken (250 rpm) in 500 ml conical baffled flasks at 28 °C for 4-5 days. Every 24h, the cultures were supplemented with 0.5% final concentration of methanol to maintain induction and OD<sub>600</sub> cell densities were recorded. After 120h of maintaining recombinant protein expression, the cell cultures were harvested by centrifugation at 6000 x g for 15 min. The supernatant was sampled and recombinant LapA protein detected on SDS-PAGE (NuPAGE® Novex® 4-12% Bis-Tris gels).

\*YNB – yeast nitrogen base, with no amino acids. Rich growth medium containing nitrogen sources, vitamins, trace elements, salts.

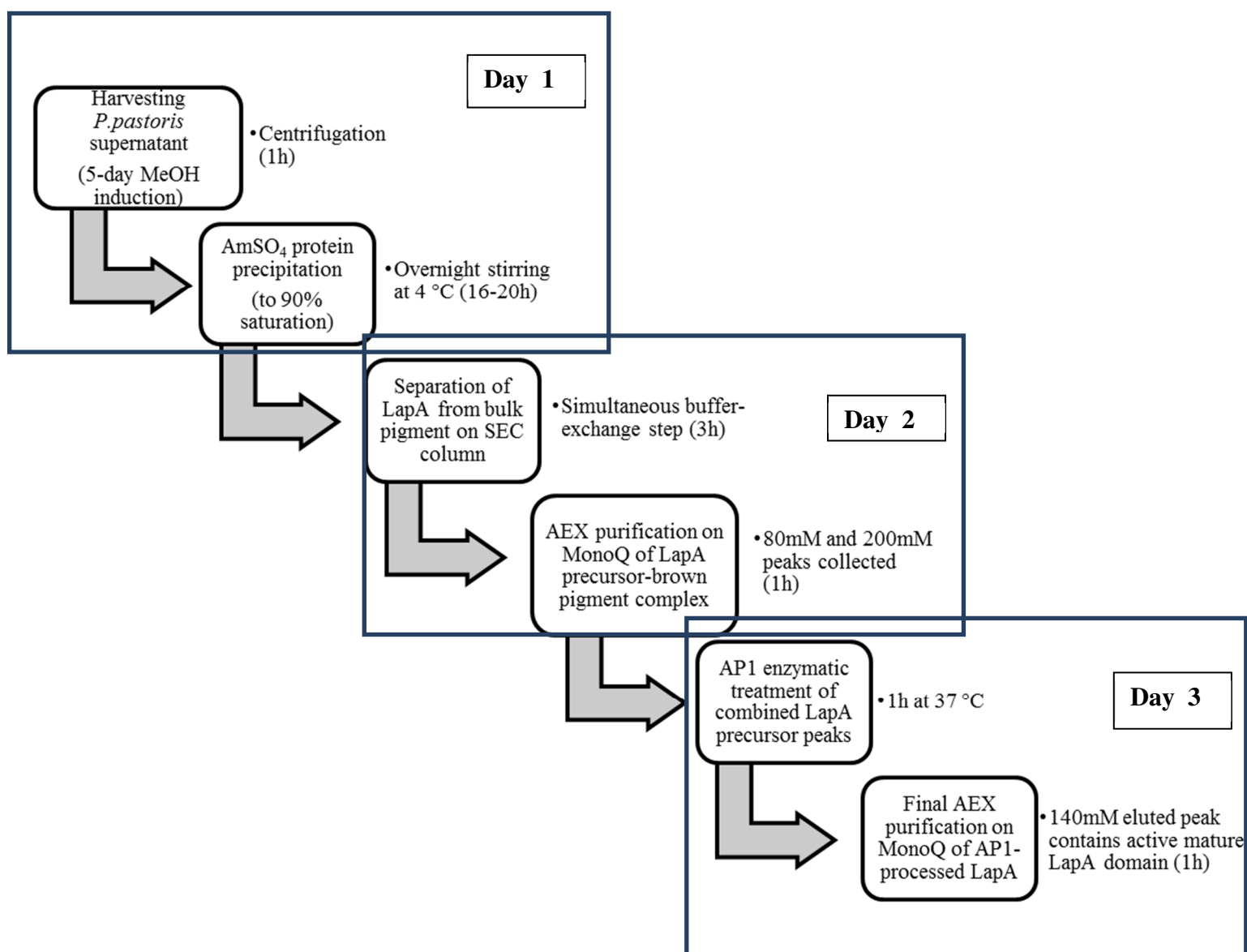
*P.pastoris* micro-fermentation trials were conducted using Biolector (M2P Labs, UK) micro-fermenter with controlled levels of humidity, oxygen and agitation. A 48 well Biolector FlowerPlate (M2P Labs, UK) plate was inoculated with a selection of recombinant colonies from agar plates into 1ml of BMGY medium and incubated at 30°C, 85% relative humidity with 1100 rpm agitation on a 2 mm orbital axis for 16 h (overnight). A 1:100 dilution in BMGY of 20 µl of biomass from several wells was measured for OD<sub>600</sub> to assist in converting Biolector AU readings for biomass into OD<sub>600</sub> values. Typically, growth was between 25 to 35 OD<sub>600</sub>. 1 ml cultures were recovered in sterile Eppendorf tubes, centrifuged for 1 min at 13,000 rpm and resuspended in 500 µl of BMMY. A new Flowerplate was inoculated with 1ml BMMY and a volume of each culture resuspension providing a final OD<sub>600</sub> of approximately 1.0. Cultures were then grown as for the BMGY inocula with additions of 50% Methanol to a final concentration of 0.5% (10 µl per 1 ml of culture) every 24 h. *P.pastoris* supernatants were harvested after 120 h and analysed on SDS-PAGE.

#### **2.13.4 Purification of recombinant LapA from *P.pastoris* supernatant**

Over-expressed *wtproLapA*, *wtLapA*, *wtΔproLapA* and thermolability mutant proteins were purified from *P.pastoris* supernatant employing 2 different protocols: 1) crude *P.pastoris* supernatant, containing recombinant LapA, was dialysed against 20 mM Tricine pH = 8.0, 1 mM ZnCl<sub>2</sub> for 16h at 4 °C, then filtered through a 0.22 µm steri-filter prior to loading onto a 5 mL HiTrap™ Q HP anion exchange column, pre-equilibrated with 20 mM Tricine pH = 8.0, 1 mM ZnCl<sub>2</sub>. The column was washed (10 column volumes) with equilibration buffer

until no trace of protein appeared in the eluate. Recombinant LapA was eluted using an increasing gradient (0 - 1M) of NaCl (20 mM Tricine pH = 8.0, 1 mM ZnCl<sub>2</sub>, 1M NaCl). Active fractions were pooled and concentrated, using 10 kDa MW cut-off ultrafiltration membranes. In each case, the respective protein concentrate was loaded onto a HiLoad Superdex<sup>TM</sup> 16/60 200pg (GE Healthcare, UK) gel filtration column, pre-equilibrated with 20 mM Tricine pH = 8.0, 1 mM ZnCl<sub>2</sub>. Homogeneous, recombinant LapA fractions were concentrated and stored at -20 °C for further analysis; or 2) Alternatively, crude *P.pastoris* supernatant, containing the respective recombinant LapA constructs, was saturated to 90% ammonium sulphate by adding hard salt, followed by stirring for 1h at ambient temperature. LapA precipitant was pelleted by centrifugation at 16,000 g for 20 minutes. The collected pellet was re-suspended in 20mM Tricine pH = 8.0, 1 mM ZnCl<sub>2</sub>, followed by dialysis against the same buffer overnight at 4 °C. At this point the purification procedure was equivalent to 1), starting at the anion exchange chromatography step. Recombinant protein purified, using protocols 1) and 2), was analysed on SDS-PAGE.

The purification protocol depicted in **Figure 2.1** was found to be most optimal to obtain protein suitable for crytallisation and subsequent thermolability assays.



**Figure 2.1** Schematic of recombinant LapA purification workflow. From crude *P.pastoris* supernatant to homogenous activated recombinant LapA protein: time required – 3 days.

#### 2.13.4.1 Activation of purified LapA proenzymes

Activation of purified LapA proenzymes was achieved by incubating the target protein with AP1 protease fractions purified from Flavourzyme. The hydrolysis of propeptide domain was allowed to proceed for 1 – 2 hours at 37 °C. The resultant mixture was subsequently loaded on MonoQ AEX column to isolate active mature LapA domain for downstream analysis and thermolability assays.

#### 2.13.5 Heterologous protein expression in *Pichia pastoris* - experimental workflow

Experimental workflow of generating wild-type and mutant recombinant LapA *P.pastoris* strains is represented in **Table 2.4** and **Table 2.5**, respectively.

Step order	Step description	Duration
1	Gene synthesis and codon optimisation	14 days
2	Linearized DNA transformation to <i>P.pastoris</i> X33 strain	1 day
3	Growth of transformants	3 days
4	Colony PCR clone verification	1 day
5	250 - 500mL shake-flask expression trial	6 days
6	Glycerol stock preparation for positive <i>P.pastoris</i> mutant strains	1 day
	<b>Total</b>	26 days

**Table 2.4** Experimental pipeline for development of *P.pastoris* wild-type LapA strains and initial expression trial.

Step order	Step description	Duration
1	Quickchange PCR and <i>E.coli</i> transformation	1 day
2	Colony PCR transformant verification	1 day
3	Plasmid DNA purification and sequencing	1 day
4	Large scale plasmid DNA preparation and linearization	1 day

5	Mutant DNA transformation to <i>P.pastoris</i> X33	1 day
6	Mutant colony growth	3 days
7	Colony PCR clone verification	1 day
8	Small scale Biolector expression trial	5 days
9	Glycerol stock preparation for positive <i>P.pastoris</i> mutant strains	1 day
10	Large scale shake flask trial	7 days
11	Chromatography purification steps	4 days
12	Enzymatic activation of LapA mutant precursors	1 day
13	Final purification step to isolate active mature LapA variants	3 days
	<b>Total</b>	30 days

**Table 2.5** Experimental pipeline for development of *P.pastoris* mutant strains, LapA protein expression and purification. Duration column represents number of days equivalent to preparing 5 different LapA mutant proteins.

**Table 2.5** describes sequential experimental workflow of cloning, expression and purification of LapA thermolability mutants. Recombinant expression of LapA proteins in eukaryotic host – *Pichia pastoris* – presented considerable challenges regarding the number of experimental steps, duration and complexity, which in turn increases probability of failure. To obtain 5 homogenous pure LapA mutant proteins for preliminary analysis at least 30 days was required, provided every experiment in this sequence was successful. In fact, most of the times number of days would increase to at least 50 if travelling times to Biocatalysts and weekends were taken into account. Every experimental step was carried out in Reading, except steps 5 – 8 were implemented at Biocatalysts due to requirement of specialist equipment: electroporator for transformation of *P.pastoris* cells and Biolector for micro-fermentation run of multiple *P.pastoris* transformants.

## **Chapter 3 FLAVOURZYME – ISOLATION OF NATIVE *ASPERGILLUS ORYZAE* LEUCINE AMINOPEPTIDASE A**

---

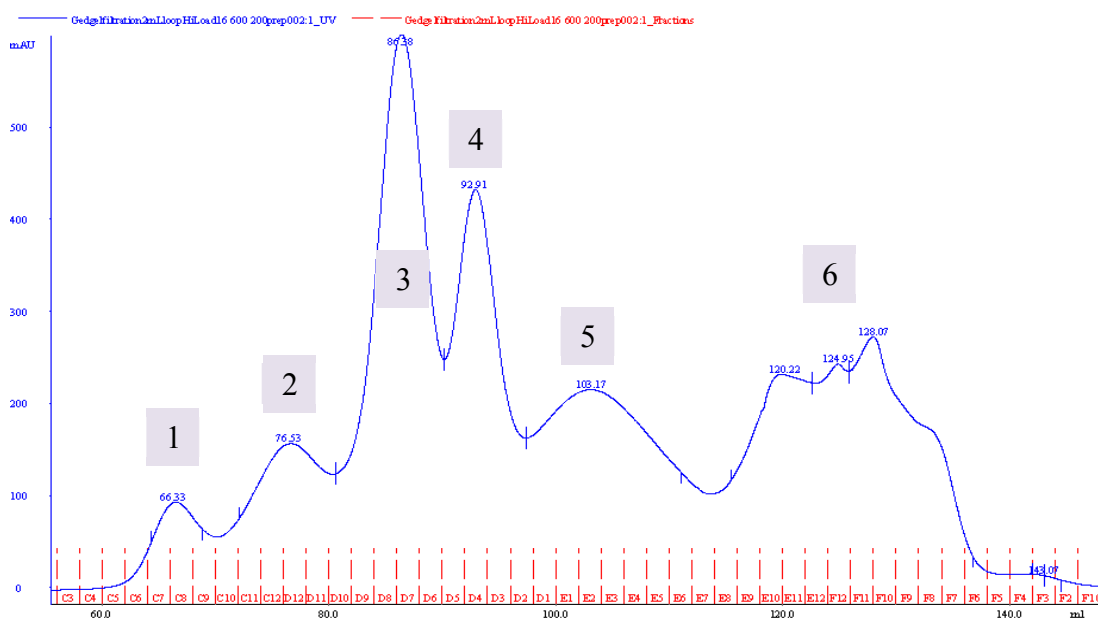
### 3.1 Introduction

Flavourzyme is a commercial fungal extract obtained from *Aspergillus oryzae* fermentation broth. The commercial preparation is known to display endoprotease and exopeptidase activities and contains at least five known proteases (Blinkovsky et al., 2000). Flavourzyme is utilised across a number of food hydrolysis applications, mainly for flavour development of resultant bitter whey protein hydrolysates (WPH). Enzymatic de-bittering serves as the main tool to further process WPH for flavour development. For example, soybean protein is incubated with Alcalase (*Subtilisin Carlsberg*) and the resultant bitter product is further hydrolysed by Flavourzyme, yielding an improved taste profile with reduced bitterness of the soybean polypeptide mixture (Ma et al., 2013). Flavourzyme fungal extract contains two exopeptidases, namely, leucine aminopeptidase A (LapA) and leucine aminopeptidase 2 (Lap2), either of which can remove the N-terminal hydrophobic amino acid residues that contribute to the pronounced bitter flavour of protein hydrolysates. Comprehensive structural and functional characterization of Flavourzyme proteases would be of benefit to determine the contribution of each protein to the process of enzymatic whey hydrolysis.

At the start of the project, attempts were made to purify aminopeptidases directly from Flavourzyme, as this is a readily available, abundant commercial protein mixture (provided by Biocatalysts Ltd, UK). Thus, Flavourzyme provides a potential source of the enzymes of interest, from which, upon successful purification, could provide native LapA and Lap2 proteins for structure determination and functional characterisation. In order to rationally design potential point mutations, which could render LapA thermolabile, it is desirable to obtain a three-dimensional structure. This would then allow direct incorporation of mutations that could result in altered thermal properties of this commercial aminopeptidase. Moreover, this structure-based mutagenesis approach would circumvent any inaccuracies arising, as a result of utilising an homology model, and would reduce speculative mutations that could lead to misfolded versions of the recombinant protein. In other words, having a three-dimensional structure of the target protein, obtained by X-ray crystallography, reduces the extent of mutant screening effort and allows a more accurate prediction of the implications of the rationally selected mutations. This has been a priority objective from the beginning of the project. Lap2, a non-specific aminopeptidase, is also produced by *Aspergillus* and, therefore, already is present in the commercially available Flavourzyme. It was decided to purify this aminopeptidase, along with LapA, for crystallisation studies and structural determination, which would provide insight into its mode of action and provide a potential new route for thermal adaptation for these leucine aminopeptidases.

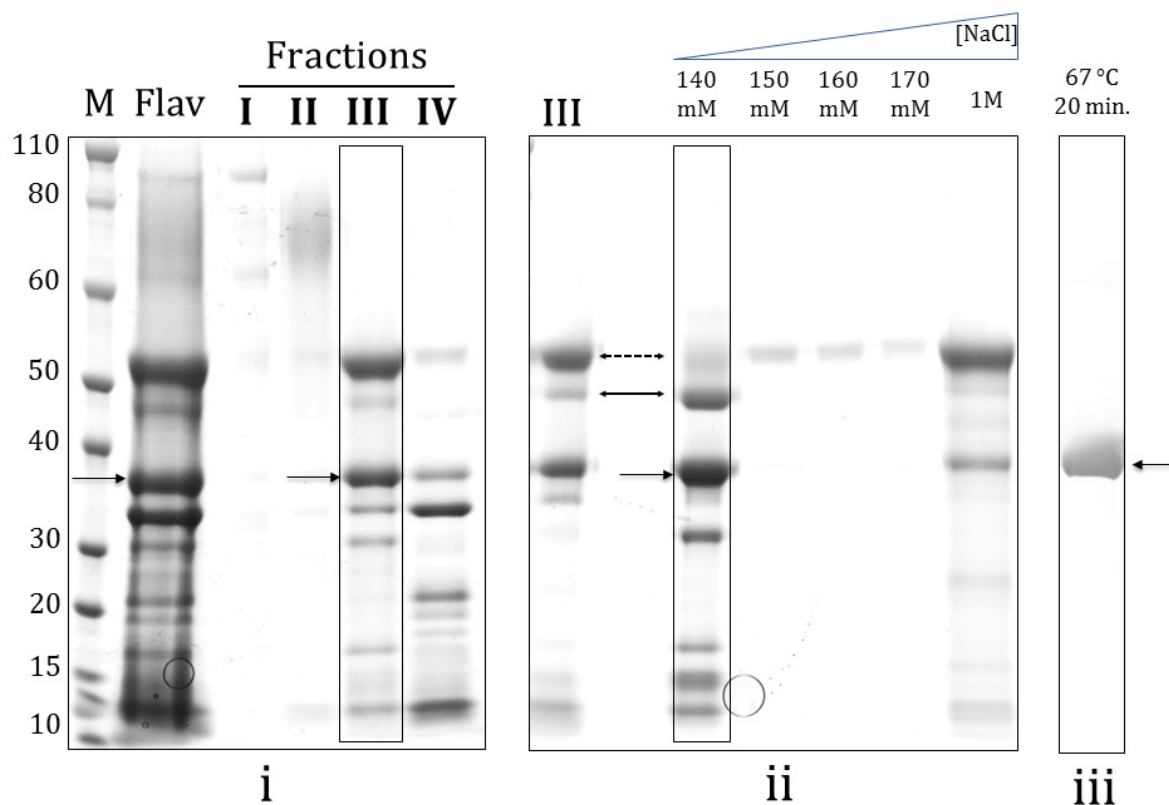
### 3.2 Isolation of native *Aspergillus* aminopeptidases from Flavourzyme

Partial separation of Flavourzyme proteases was achieved using size exclusion chromatography (SEC), even after a single purification step, as indicated in **Figure 3.1**. As none of the proteases isolated from the commercial preparation are recombinantly produced, any subsequent purification relies solely on the native properties of the given enzymes, such hydrodynamic volume (in the case of SEC purification).



**Figure 3.1** Separation of Flavourzyme proteins by size using a gel filtration column (HiLoad 16/60 200pg, GE Healthcare). Fractions 1-5 contained protein (see **Figure 3.2**), and fraction 6 contained unidentified small molecular weight compounds.

**Figure 3.1** shows the chromatogram obtained from an initial separation of Flavourzyme, following a single run on a size exclusion column (SEC), loading 3 mg of crude Flavourzyme extract and elution of 2 mL fractions from the SEC column. This purification step permitted a partial separation of an otherwise complex mixture of varying molecular weight Flavourzyme proteins. The identified Flavourzyme peaks were analysed for protein content, as illustrated in **Figure 3.2** (i).



**Figure 3.2** Analysis of Flavourzyme purification and 3-step isolation of *flavLapA*. Purification protocol for *flavLapA* (indicated by single arrow): i) separation Flavourzyme fractions on gel filtration column, Flav – Flavourzyme proteins prior to separation, I-IV- protein elution in four distinct peaks based on molecular weight, Fraction II contains Lap2 (smeary band) protein, Fraction III (rectangle) – further purified using anion exchange (AEX) chromatography; ii) and further separation of Fraction III proteins (amylase – dashed double arrow; neutral protease I – solid double arrow) by AEX chromatography, using a linear gradient elution with increasing concentration of NaCl; iii) completion of *flavLapA* purification by adjusting 140 mM fraction (rectangle) to 500 mM NaCl and heating to 67 °C for 20 min., followed by final purification using SEC.

Following a successful initial separation of Flavourzyme proteins, it was not clear which fraction contain the target aminopeptidases (*flavLapA* and Lap2). The desired LapA and Lap2 enzymes were identified in the complicated mixture of proteins, using the LAP specific activity assay (see section 2.10.1) and mass spectrometry (section 2.12.3). MS analysis of 3 gel bands, clearly revealed the identity of LapA, Lap2 and TAKA alpha-amylase (**Figures 10.1 - 10.3**; Appendix I). LAP activity assay (**Table 3.1**), of each of the SEC fractions, showed Fraction II and Fraction III possessed the highest activity, and coincidentally these were expected to contain Lap2 and LapA, respectively. Fraction IV showed moderate activity, and Fraction V showed negligible activity (not shown on the gel).

Purified Flavourzyme fraction	Specific activity (U/mg)
Fraction II	8.0 ± 0.5
Fraction III	20.0 ± 0.2
Fraction IV	5.0 ± 0.3
Fraction V	0.7 ± 0.1

**Table 3.1** LAP specific activities measured in SEC-isolated Flavourzyme fractions.

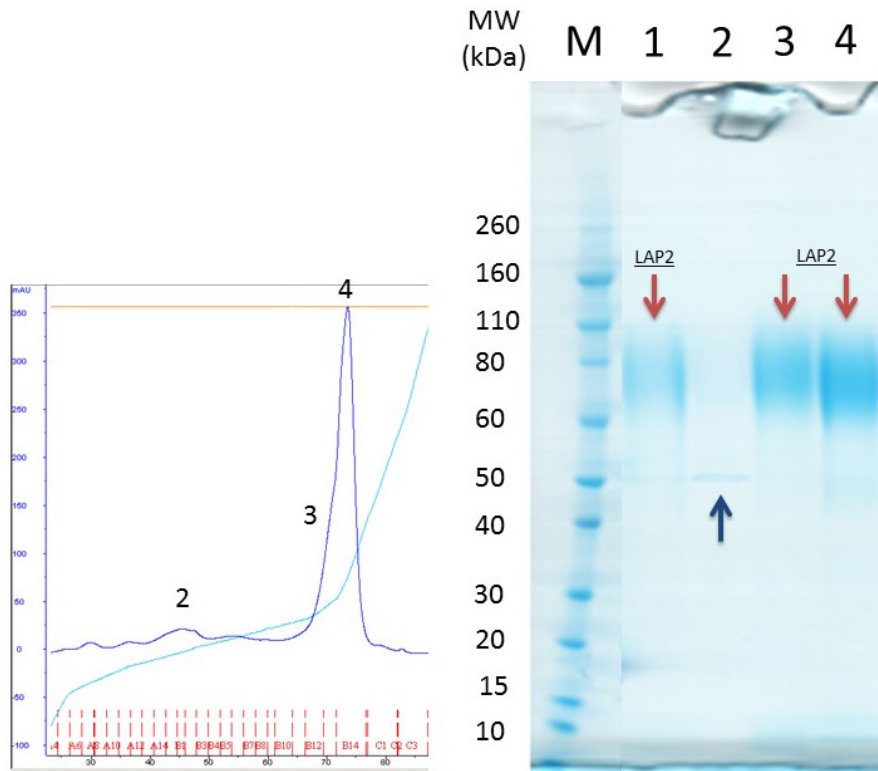
### 3.2.1 Isolation of *flavLapA*

Recently, an automated protocol, involving 9 steps, for purification of all Flavourzyme proteins (including *flavLapA*) was reported (Merz et al., 2015). In this study, we took a quite distinct approach to purification of native *Aspergillus oryzae* leucine aminopeptidase A, involving only 3 steps to isolate *flavLapA* from Flavourzyme commercial enzyme blend (**Figure 3.2**, i-ii-iii). A comprehensive description of *flavLapA* purification protocol is provided in section 2.7.3. In brief, the successive Fraction III from the initial SEC separation step was firstly loaded on AEX column. A subsequent AEX step was required to separate *flavLapA* from  $\alpha$ -amylase and an unidentified brown pigment that is known to associate with the protein (Schalk et al., 1992). The latter we found very challenging and repetitive cycles using AEX chromatography were necessary to eliminate traces of this amber pigment, which was mandatory for crystallisation trials. Following repetitive AEX cycles, the eluent obtained was pigment-free (at 140 mM NaCl) and almost amylase-free, whilst at 1M NaCl the fraction contained most of the amylase and pigment (appeared dark brown) (**Figure 3.2**, ii). This mysterious brown pigment ( $M_r < 2\text{kDa}$ ) was previously reported to associate with other Lap peptidases, and was briefly characterised by EPR to be spectroscopically active and contain Fe(III) ligand (Schalk et al., 1992, Hartley et al., 2009, Aphale and Strohl, 1993, Bennett and Holz, 1997).

Heat treatment was required as a penultimate step in the purification of *flav*LapA to efficiently separate it from contaminating neutral protease I and other lower MW proteins present in Flavourzyme (**Figure 3.2**, iii). Characteristic high thermal stability of native LapA allowed incubation at 67-70 °C for prolonged times, without precipitating the enzyme and/or losing activity. However, it is worth noting, that precipitation of other contaminating proteins in 20 minutes could only be achieved in the presence of 500mM NaCl or equivalent ionic strength of ammonium sulphate (results not shown). Similarly, intrinsic thermostability was utilised in the purification of an homologous aminopeptidase from *Aeromonas proteolytica* (AAP) (Prescott and Wilkes, 1976). As a result, homogenous, pigment-free *flav*LapA was obtained, following heat treatment at 67 °C for 20 minutes in the presence of 500 mM NaCl (**Figure 3.2**, iii). Purification of *flav*LapA by temperature denaturation proved particularly useful in the successful isolation of *flav*LapA. The discovery of this simple LapA purification method facilitated the ability to obtain higher yields of pure enzyme.

### 3.2.2 Isolation of Lap2

**Figure 3.2** shows that Lap2 is the predominant enzyme in Fraction II and it appears to be homogenous. Thus, only one further purification step was attempted in an effort to obtain pure Lap2. In this instance, standard AEX chromatography protocol was applied for the successful separation of Lap2 from contaminant protein, as shown in **Figure 3.3**.



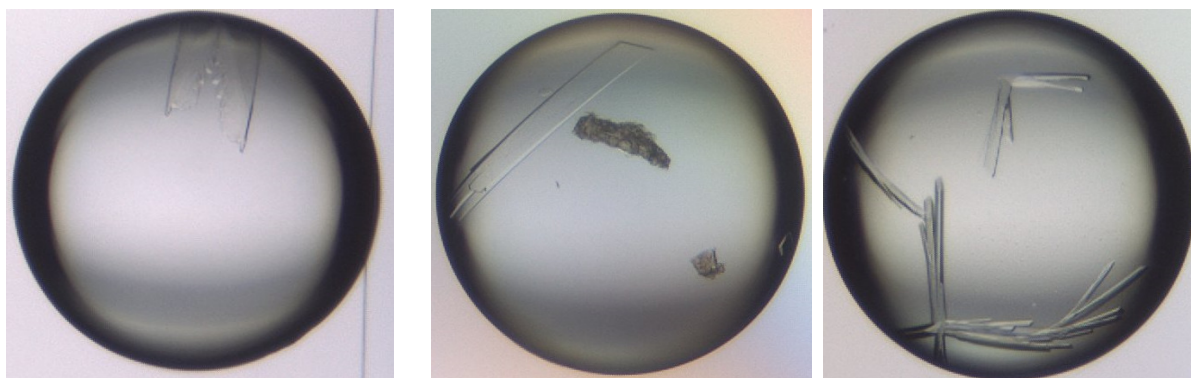
**Figure 3.3** UV trace (left) and SDS-PAGE gel (right) of AEX fractions from Fraction II purification. Gel analysis: 1 – Fraction II from SEC; 2 – 80 mM NaCl AEX peak; 3, 4 – 200 – 250 mM NaCl AEX elution range; Lap2 is indicated by red arrows. The smearing of Lap2 protein band indicates hyperglycosylation. Major contaminant protein is indicated by the dark blue arrow.

It is easy to observe that the Lap2 protein 'smears' on the SDS gel (**Figure 3.3**). This is characteristic of the N-linked glycosylation of this eukaryotic protein (Blinkovsky et al., 2000). These added sugars significantly contribute to the total MW of Lap2, and as a result, the migration of Lap2 on SDS-PAGE is observed at around 70 kDa. The predicted MW of Lap2, using ProtParam, is 52 kDa whereby the protein was reported to migrate around 56 kDa following deglycosylation of the N-linked sugars (Blinkovsky et al., 2000). The smearing on the gel is a result of the heterogeneity of the glycosylated protein in the solution i.e. different degrees of glycosylation results in a variety of similar molecular weights for the same protein.

### 3.2.3 Initial challenges in *flavLapA* and Lap2 crystallisation trials

#### 3.2.3.1 Crystals of TAKA alpha-amylase

Before the purification protocol for *flavLapA* was optimised, Fraction III was used to perform initial crystallisation screens in 96 – well plates, using high throughput robotics at Oxford Protein Purification Facility (OPPF-UK). It was anticipated that *flavLapA* crystals could still form, following the identification of a favourable condition using multiple commercial random screening matrices. On the other hand, Taka alpha-amylase was a rather abundant protein in Fraction III, therefore, not surprisingly, crystals of this protein preferentially formed overnight under JCSG F5 condition (Figure 3.4). Unfortunately, no crystals were obtained for *flavLapA*, even after two months, suggesting a more homogenous *flavLapA* preparation was required.



**Figure 3.4** Crystals of the TAKA alpha-amylase from *Aspergillus oryzae* (Flavourzyme). These crystals appeared in 24h under JCSG F5 condition (0.2 M Magnesium chloride hexahydrate, 0.1 M Tris pH = 8.5, 50 % v/v Ethylene glycol).

The identity of TAKA alpha-amylase crystals were verified by X-ray crystallography and molecular replacement structure determination. We obtained a model at 1.69 Å resolution, which is a minor improvement to the highest resolution model (2.1 Å) available in PDB database at the time.

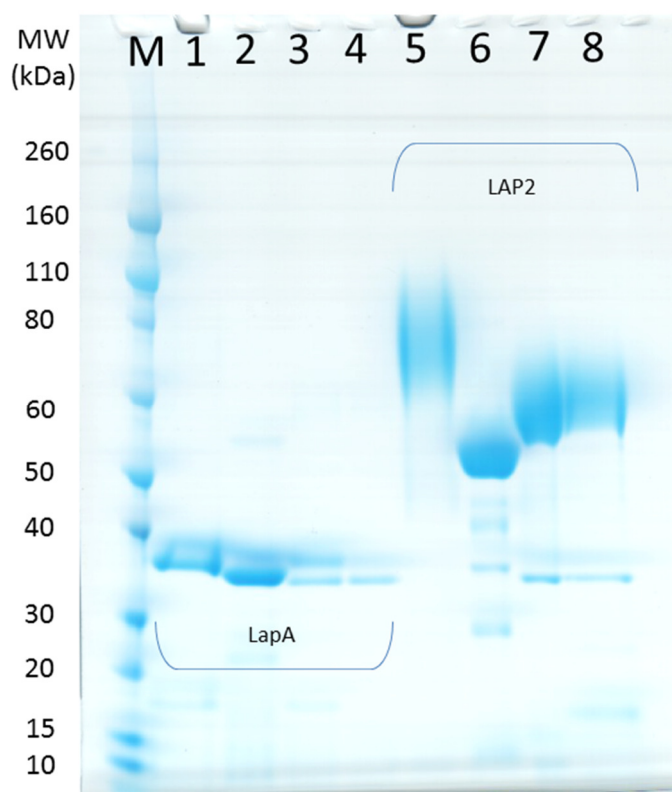
Twenty crystallisation screens were set up in order to obtain diffraction quality crystals of LapA and Lap2, as isolated from Flavourzyme. In all trials, either Taka alpha-amylase or salt crystals were observed. It may be that the inhibition of crystal formation of the desired proteins is a result of the multiple N-glycosylation sites for these particular proteases (Matsushita-Morita et al., 2011, Blinkovsky et al., 2000). Based on the SDS-PAGE and MS

analysis of LapA and Lap2, the contribution to the molecular weight by these sugars is quite significant, i.e. an additional 7 kDa for the former and an additional 18 kDa for the latter. Thus, the effect of such heavy chains of sugars on crystal packing is unpredictable. By contrast, Taka alpha-amylase crystallises easily overnight from the first preparation of enzyme sample (**Figure 3.4**). This raises the question of the actual contribution of the glycosylation sites to the crystallisation process, as a single linked GlcNAc site for this amylase has been reported previously (Vujicic-Zagar and Dijkstra, 2006). However, this Taka alpha-amylase has only one glycosylation site as opposed to multiple sites in LapA (2 sites) and Lap2 (6 sites), which would have a more pronounced effect on crystallisation, due to the inherent increased heterogeneity. Enzymatic deglycosylation can be used to overcome these problems (see the following section 3.2.3.2).

### **3.2.3.2 Deglycosylation of Flavourzyme-isolated native Lap enzymes**

The glycoproteins *flav*LapA (further LapA) and Lap2 were subjected to PNGase F treatment in order to investigate the extent of glycosylation. LapA and Lap2 can potentially possess 2 and 6 glycosylation sites, respectively. The removal of flexible N-linked sugars could facilitate the crystallisation of these proteins, since uncontrolled glycosylation can introduce heterogeneity into the sample preparation, which will interfere with crystallisation.

20 µg of LapA and Lap2 were treated with PNGase and half of the reaction mixture was analysed on SDS-PAGE (**Figure 3.5**). The effect of PNGase was investigated under denaturing and native reaction conditions.



**Figure 3.5** Analysis of LapA and Lap2 deglycosylation. Left to right: 1 – Flavourzyme-purified LapA untreated, 2 – denatured LapA treated with PNGase, 3 – native LapA treated with PNGase after 4h, 4 - native LapA treated with PNGase after 24h, 5 - Flavourzyme-purified Lap2 untreated, 6 - denatured Lap2 treated with PNGase, 7 - native Lap2 treated with PNGase after 4h, 8 - native Lap2 treated with PNGase after 24h.

As demonstrated in **Figure 3.5**, the treatment with PNGase was 100% effective on denatured and native LapA (after 24h). However, the equivalent degree of deglycosylation for Lap2 could be achieved for the denatured protein only. The inaccessibility of some glycosylation sites (for PNGase enzyme) could explain why deglycosylation was incomplete in the case of native Lap2. This can be further verified, as smearing of both the 4 h and 24 h samples for Lap2 was still present, as observed by SDS-PAGE. However, some deglycosylation has occurred, as seen by the downshift in the molecular weight compared to untreated Lap2. This experiment verified the presence of fewer glycosylation sites on the LapA enzyme. Therefore, these sugars can potentially be removed prior crystallisation of *flav*LapA.

To study the effect of deglycosylation on function, Lap activity assays were performed with fully deglycosylated LapA and partially deglycosylated Lap2. The activity of LapA was almost completely abolished – with only 1.5% activity recovered compared to the native

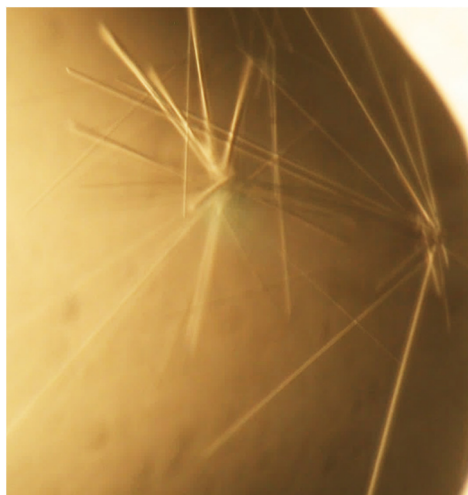
glycosylated LapA. Lap2 showed a 2-fold reduction in activity compared to the native enzyme (**Table 3.2**). Coincidentally, the findings from this experiment are in agreement with elimination of the N-glycosylation site in recombinant LapA (M5 mutant) by site-directed mutagenesis (sections 6.1.3 and 6.5.1.1), whereby the removal of N-linked sugar was detrimental to enzyme activity.

Lap protein	Glycosylated form (U/mg)	Deglycosylated form (U/mg)
<i>flavLapA</i>	94.6 ± 2.2	1.42 ± 0.5
Lap2	21 ± 2.0	10.40 ± 0.4

**Table 3.2** Effect of glycosylation on the specific activity (U/mg) of LAP enzymes.

### 3.2.4 Crystallisation of *flavLapA*

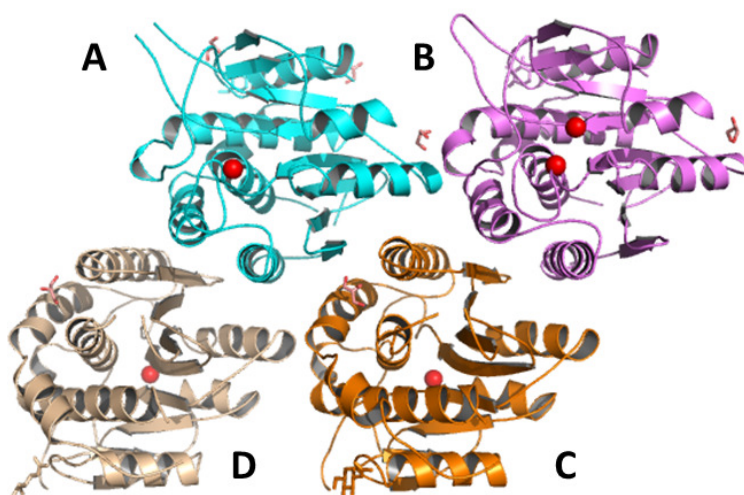
Once the purification of *flavLapA* from Fraction III was optimised, by introduction of a heating step, the resulting homogenous LapA population readily crystallised in 3 – 4 days (**Figure 3.6**). *flavLapA* from Flavourzyme crystallised at 6 mg/ml in 20mM HEPES, pH = 8.0 under JCSG F2 condition (0.1M Citrate, pH = 5.0, 3.2 M Ammonium sulphate) at 21 °C.



**Figure 3.6** Needle – shaped crystals of *flavLapA*.

### 3.2.5 Data collection and refinement of *flavLapA* crystal structure

The crystal structure of *flavLapA* was determined to 2.71 Å resolution (see **Table 10.1**, Appendix III). The crystal packing showed spacegroup C2 with four LapA monomers in the asymmetric unit (**Figure 3.7**). At this resolution, a single N-acetylglucosamine (NAG) moiety was found covalently linked to the ND nitrogen of asparagine 87 (N87), showing N-linked glycosylation. Further evidence of N-glycans were not definitive.



**Figure 3.7** Crystal structure of *flavLapA* at 2.71 Å resolution. Four protein monomers (A – D) were found in the asymmetric unit. Catalytic Zn ions are depicted as red spheres.

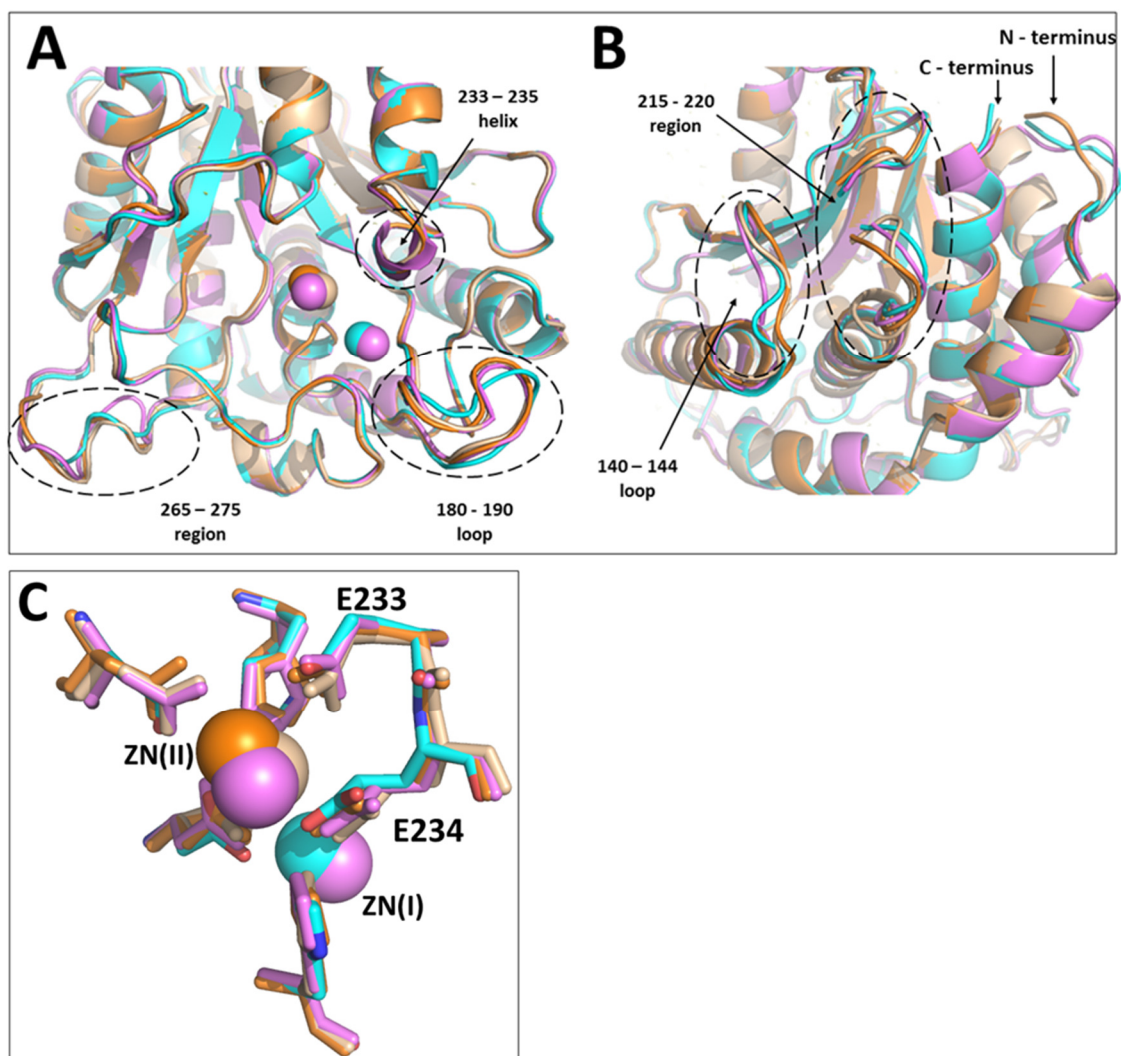
By contrast, the observed crystal packing for the mature recombinant LapA (*wtLapA*) structure was quite distinct (see section 5.1). The *wtLapA* structure showed P6<sub>3</sub>22 spacegroup and only 1 molecule in the asymmetric unit. The observed differences in the crystal packing can be attributed to the different N-linked carbohydrate structures present in *Aspergillus* versus *Pichia*, which appears to affect the spatial packing and crystal contacts. The additional heat-induced purification step could result in structural alterations in *flavLapA* and thus play a role in the deglycosylation process. Therefore, it was concluded that the observed four molecules in the asymmetric unit of *flavLapA* is likely a crystallographic artefact and does not represent the biologically active molecule.

### 3.2.6 Evidence of primary thermal unfolding states in *flavLapA*

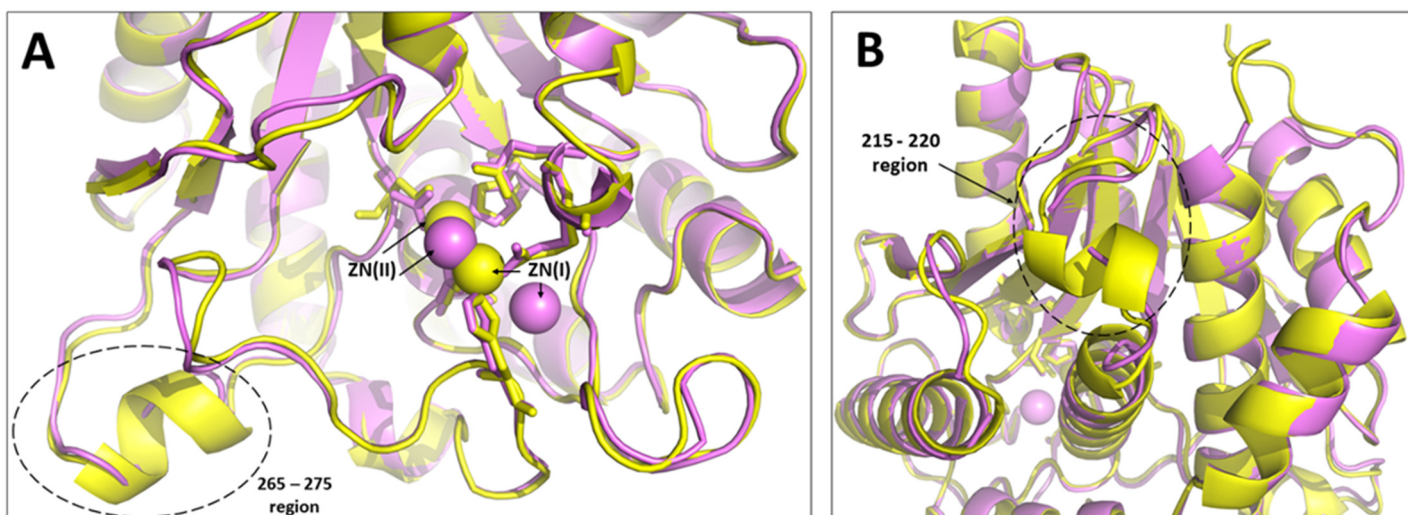
The implemented heating step (67 °C, 20 minutes) in the purification protocol of *flavLapA* had a noticeable effect on protein fold integrity.

Quite remarkably, each of the monomers in the asymmetric unit exhibit unique structural differences. Analysis and superimposition of C $\alpha$  traces of each monomer (compared to each other) indicates significant surface loop movements (**Figure 3.8**). The most affected regions can be found in residues 140 - 144, 180 – 190, 215 – 220 and 265 – 275. Monomer B shows the most secondary structure content with two ZN ions coordinated at the active site pocket, as expected. A helix at 233 – 235 appears only in Monomer B. This structural element holds a catalytically conserved glutamate E233 motif in place at the active site. Thus, the observed movement of the E234 sidechain was the most significant in Monomers A, C and D with 1.5 – 1.7 Å shift with respect to each other (**Figure 3.8C**). Altered positions of the sidechains of other active site residues were also apparent.

Interestingly, only a single ZN ion was found in the active sites of Monomers A, C and D. Also, the corresponding ZN atom in each monomer was not directly superimposable. Taken together, these observations seem to suggest that we have trapped different states of thermal unfoldedness in each of the four LapA monomers.



**Figure 3.8** Superimposition of different asymmetric unit monomers of *flavLapA*. Monomer A – cyan, Monomer B – violet, Monomer C – orange, Monomer D – wheat; **A** Active site view: ZN ions are indicated as spheres; **B** 180° rotation view along y-axis; **C** Superimposition of active site residue sidechains and catalytic ZN atoms.



**Figure 3.9** Superimposition of Monomer B (*flavLapA*) and *wtLapA* structures. Monomer B of *flavLapA* – violet, *wtLapA* – yellow. **A** Active site view: ZN ions are indicated as spheres, catalytic residues are shown as sticks; **B** 180° rotation view along y-axis.

**Figure 3.9** illustrates the major structural changes that occur in *flavLapA* protein compared to the crystal structure of recombinant LapA (*wtLapA*) (reported in section 5.1). The latter corresponds to appropriately folded, fully active LapA protein. As a result of thermal unfolding, two surface helices appear to be unravelled in *flavLapA*. The 215 – 220 region appears to have become so mobile that the K215-A218 residue stretch could not be traced in the electron density map and therefore could not be modeled (**Figure 3.8, B**; **Figure 3.9, B**). Additionally, the position of the active site residues and catalytic ZN atoms (I and II) are significantly altered. In particular, in *flavLapA*, ZN(I) atom appears to be completely expelled from the tetrahedral coordination geometry, resulting in a 4.2 Å movement. At the same time, the position of ZN(II) atom is shifted by 2.6 Å with respect to the same atoms in *wtLapA* protein (**Figure 3.9, A**).

The observed distorted active site geometry, in the *flavLapA* structure, was also reflected in the reduced specific activity towards Leu-*p*-nitroanilide substrate. Surprisingly, it still retained a substantial level of activity and showed  $94.6 \pm 2.2$  U/mg, which is approximately 67% of that observed for the *wtLapA* ( $140.15 \pm 4.8$  U/mg). Overall structural comparison, using the TM-Align server, showed an RMS difference of 0.59 over 291 residues, between *flavLapA* Monomer B and *wtLapA* coordinates, confirming significant structural differences between the two crystal structures.

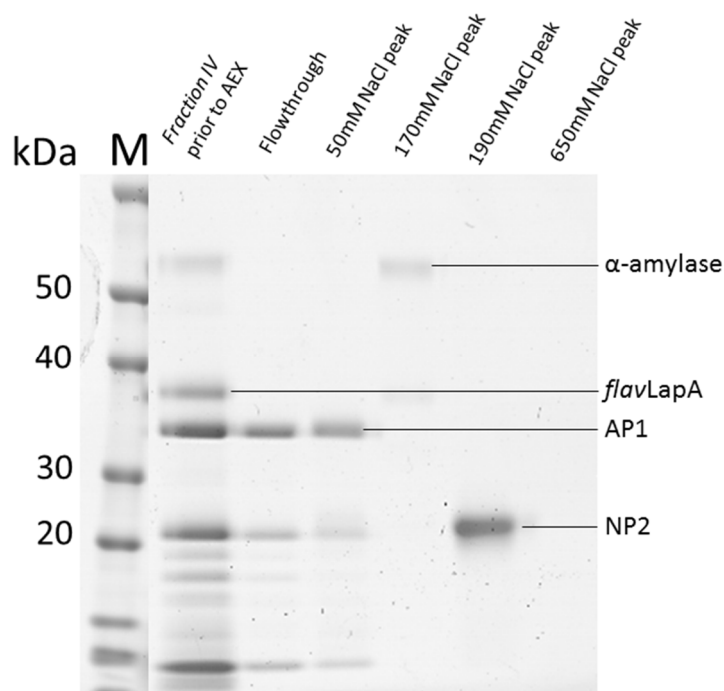
The *flav*LapA crystal structure has offered a deeper understanding of the thermal unfolding process of LapA. It seems apparent that unfolding has initiated at the surface loop regions and then propagated to the secondary structure elements (unravelling of 265 – 275 and 215 – 220 helical elements). Coincidentally, in our thermolability mutagenesis strategy, we have also targeted the 267 – 272 helix, specifically with D271T (M8) mutation. It is likely that the success of this destabilising mutation could be attributed to the fortuitous selection of this region. Consequently, we have managed to mirror the thermal unfolding process required to inactivate the LapA protein.

Analysis of *flav*LapA structure permits a more intuitive design of future mutagenesis libraries. Following the same rationale, future efforts toward LapA thermolability engineering could involve selecting mutations across a number of surface loops, including but not limited to 215 – 220, 140 – 144 and 180 – 190.

### 3.3 Isolation of NP1 and AP1 proteases from Flavourzyme

Since LapA aminopeptidase is natively expressed in the form of precursor protein and is not a self-activating enzyme, it was speculated that LapA zymogen could be processed by one of the non-specific proteases present in Flavourzyme: neutral protease I (NPI) and/or alkaline protease I (API). These proteases in the Flavourzyme commercial enzyme mixture have been isolated previously by Merz et al. (Merz et al., 2015).

Isolated *Fraction IV* from Flavourzyme contains at least two non-specific proteases (AP1 and NP2) (**Figure 3.2**, i) that potentially are capable of effectively cleaving LapA propeptide, as shown previously (4.3.7.1). Thus, a 20% w/v Flavourzyme solution was separated into individual fractions by SEC (as described previously, section 3.2). Next, *Fraction IV* was loaded onto a 5ml Q HP AEX column, pre-equilibrated with 20mM Tricine pH 8.0, 1mM ZnCl<sub>2</sub> and 1mM CaCl<sub>2</sub>. After 100mL wash, to remove unbound proteins, elution using a 1% stepwise gradient was initiated, using an equivalent buffer containing 1M NaCl. The main fractions eluted at 50, 170, 190 and 650 mM NaCl concentration increments in the elution buffer (**Figure 3.10**).



**Figure 3.10** SDS-PAGE of AEX purification of proteins in Flavourzyme *fraction IV*. Eluted fractions of increasing NaCl concentration are indicated. NP2 – neutral protease 2; AP1 – alkaline protease 1; *flavLapA* – native LapA.

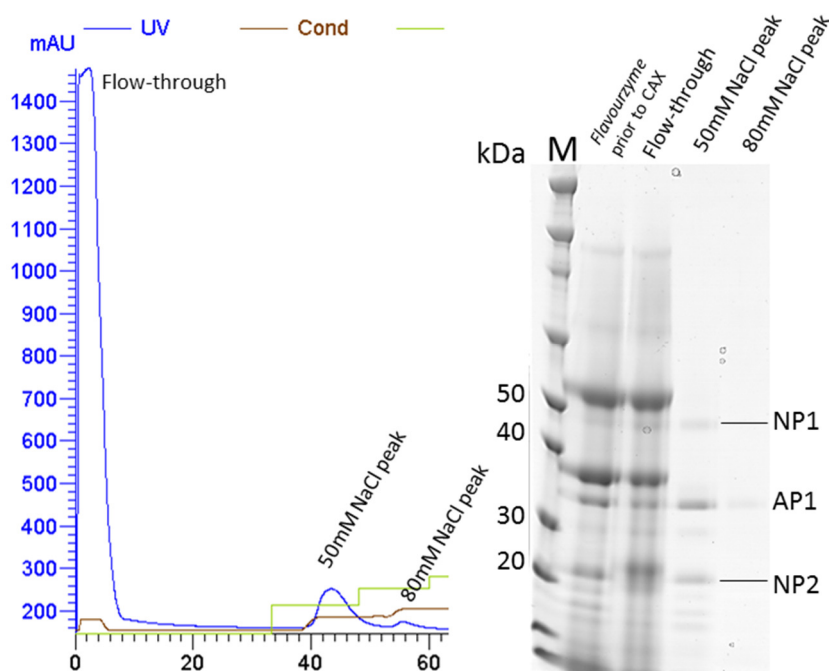
In this AEX purification experiment, AP1 protease did not strongly bind to the positively charged resin and continuously eluted in the flow-through and in the 50 mM NaCl fraction. Also, some NP2 protease and protein degradation products co-eluted in these fractions (**Figure 3.10**). Taka  $\alpha$ -amylase and *flav*LapA eluted at 170 mM NaCl. NP2 separated at 190 mM NaCl and was the most homogenous protein fraction obtained.

It was not possible to separate AP1 and NP2 at 50 mM NaCl. However, there was no AP1 present at 190 mM NaCl, which purely consisted of NP2, suggesting it has a considerably higher net negative charge and therefore bound to the AEX resin with increased affinity. This distinct separation of AP1 and NP2 proteins into discrete AEX fractions allowed investigation into which of the two non-specific proteases naturally activates LapA precursor.

### **3.3.1 Improved isolation protocol of AP1 and NP1 proteases**

The current AP1 purification protocol consists of two purification steps: SEC separation of Flavourzyme proteins into 5 distinct pools (*fractions 1 – 5*), followed by separation of proteins in *Fraction IV* using AEX resin. In the protocol suggested by Merz et al. (Merz et al., 2015), it was possible to isolate another neutral protease (NP1) from Flavourzyme, along with AP1, in a single cation exchange chromatography (CAX) run. NP1 is another potential activator of LapA zymogen in Flavourzyme. Therefore, it was important to test whether NP1 and AP1 are both equally important in LapA processing or whether they work synergistically to produce the appropriate N-terminal processing of LapA.

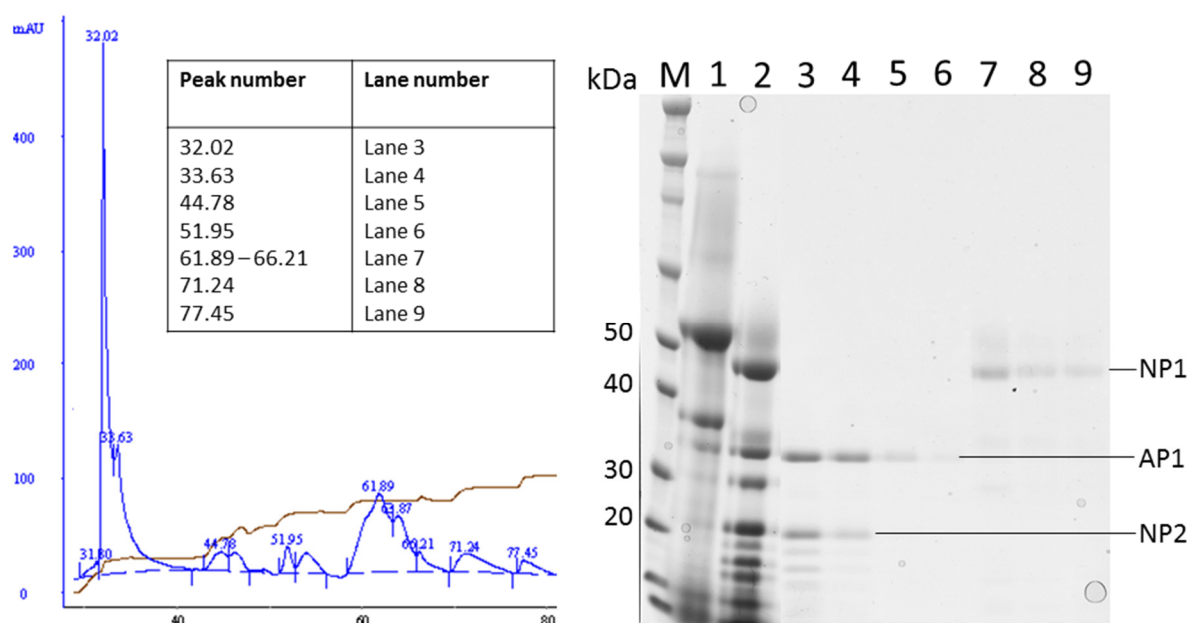
A 20% w/v Flavourzyme solution (corresponding to approximately 12 mg of total protein) was prepared in cation exchange chromatography binding buffer (10mM Sodium acetate pH 4.85, 0.1 ZnCl<sub>2</sub>, 0.1mM CaCl<sub>2</sub>) and loaded onto a 5mL SP HP CAX column. Fractions were eluted with equivalent buffer containing 1M NaCl.



**Figure 3.11** Purification of Flavourzyme neutral proteases by CAX.

Based on SDS-PAGE of the eluted CAX fractions, binding to the purification resin was not very efficient as there were still a large proportion of NP1, AP1 and NP2 proteins visible in the flow through (**Figure 3.11**). However, all the non-specific proteases present in Flavourzyme eluted in the 50 mM NaCl fraction. Also, AP1 protease showed a marginally increased affinity to the negatively charged CAX resin and eluted in the 80 mM NaCl fraction. Thus, it became possible to isolate AP1 in a single purification step, as opposed to the alternative lengthy protocol described above. Obtaining the NP1 protease in the same pool of enzymes (50 mM NaCl fraction) provided an opportunity to further separate it from AP1 protease and subsequently characterise it in the processing of the LapA zymogen.

Furthermore, the 50mM NaCl CAX peak was applied to a MonoQ AEX column (Figure 3.12) in a further effort to improve the separation of the target proteins.



**Figure 3.12** AEX purification analysis of 50mM NaCl CAX Flavourzyme fraction. NaCl elution peaks are shown as UV absorption chromatogram (left). SDS-page analysis of identified peaks (right): 1 – CAX flow through; 2 – 50mM NaCl CAX peak applied to AEX MonoQ column; 3, 4 – 50mM NaCl AEX peak; 5 – 100mM NaCl AEX peak; 6 – 120mM NaCl AEX peak; 7 – 140mM NaCl AEX peak; 8 – 160mM NaCl AEX peak; 9 – 180mM NaCl AEX peak.

AP1 and NP1 Flavourzyme proteases were successfully separated using a MonoQ column, due to their slightly distinct binding affinity to the positively charged AEX resin. AP1 eluted at 50 – 120 mM NaCl, along with some NP2 protein, whereas the more strongly bound NP1 eluted at 140 – 180mM NaCl. This rapid, two – step purification protocol enabled efficient isolation of the AP1 and NP1 non-specific proteases from Flavourzyme by combining complementary CAX – AEX purification techniques. Isolation of both NP1 and AP1 proteases from Flavourzyme enabled the trials of LapA propeptide cleavage described in sections 4.3.7.1 and 4.3.7.2.

**Chapter 4 OPTIMISATION OF HETEROLOGOUS  
EXPRESSION OF TARGET PROTEINS IN PROKARYOTIC  
AND EUKARYOTIC HOSTS**

---

#### **4.1 *Escherichia coli* as host for recombinant protein expression**

At the onset of this thesis work, our industrial partner (Biocatalysts Ltd) had identified a number of commercially utilised exo- and endopeptidase enzymes that could be subjected to protein engineering in an effort to manipulate and exploit thermal properties, particularly thermolability. The selected peptidases (Subtilisin Carlsberg – sC and leucine aminopeptidase – LapA) originated from prokaryotic (sC) and eukaryotic (LapA) hosts. Due to the reduced complexity of an *E.coli* expression system and available expertise in our laboratory, this prokaryotic host was selected to conduct initial cloning and over-expression trials for the selected target proteins. A high-throughput cloning and expression platform also was available at the OPPF-UK research complex at Harwell, which could be used to screen for multiple expression constructs in *E.coli*, if deemed necessary.

##### **4.1.1 Selected endoprotease – Subtilisin Carlsberg (sC)**

Typically, subtilisin-like serine proteases are secreted into the extracellular environment upon expression in its natural host. And, in fact, *Bacillus*-derived sC endopeptidase is an extracellularly secrete enzyme. Therefore, a suitable *E.coli* expression strategy was first established to increase the success of obtaining soluble recombinant protein, since recombinant expression of such excreted proteins in the *E. coli* cytoplasm often results in the formation of insoluble protein aggregates or cell toxicity. As a result, laborious protein refolding techniques often need to be employed to recover the protein of interest from inclusion bodies (Parry, 2002). Despite these efforts, several successful cases of cytosolic expression of subtilisin-like serine proteases have been reported (Ghasemi et al., 2012, Taguchi et al., 1998a).

To overcome any potential solubility problems, directing expression of sC in *E.coli* to the periplasm must be attempted, which could provide a less reducing environment with a lower abundance of naturally produced host proteins. The well-established and effortless sucrose extraction of periplasmic proteins was another added benefit of such a recombinant protein expression strategy (Neu and Heppel, 1965) and anticipated to provide a ready source of protein. It was expected that periplasmic expression in *E.coli*, of recombinant sC, would likely result in expression of the different forms of sC (mature and immature sC) as previously described (Sroga and Dordick, 2002). *E.coli* signal peptidases should effectively remove the signal sequence from the polypeptide chain, however, it is known that the propeptide induced automaturation of sC is a complex and slow process yielding a mixture of mature and pro-subtilisins (Shinde and Inouye, 1996, Yabuta et al., 2001). Nevertheless, directing expression of the recombinant protein to the periplasmic space in *E. coli* appears

to be a viable and robust method of obtaining soluble serine proteases (Koma et al., 2007, Wu et al., 2004). Furthermore, it is known that the mature part of the sC gene requires co-expression with a propeptide, which acts as an intramolecular chaperone (Shinde and Inouye, 1996, Ikemura et al., 1987). This propeptide has been shown to be obligatory for appropriate folding of subtilisin. Thus, the propeptide should precede the mature sequence in recombinant constructs of Subtilisin Carlsberg. Therefore, it was decided to over-express sC as zymogen both intracellularly and directed to the periplasmic space in *E.coli*.

#### **4.1.1.1 Bioinformatic analysis of Subtilisin Carlsberg (sC)**

An extensive literature survey was conducted in order to identify determinants of cold-adaptation and heat lability in the superfamily of subtilisin-like serine proteases (see **Table 4.1**). Such a comprehensive search enabled a list of potential site-directed mutations to be compiled, together with a rationalisation for each mutation. Taken together with a multiple sequence alignment (MSA) (section 2.1.1) of psychrophilic and thermophilic homologues of sC, numerous potential mutagenesis targets have been summarised in **Table 4.1**.

The approach outlined herein utilises rationally designed mutations, based on accumulated knowledge from the literature that has been shown to destabilise the structure of mesophilic enzymes and to confer more flexibility to loops, break existing salt bridges, disrupt hydrophobic packing within the core of the protein, reduce the content of secondary structure (extend loop regions) and change the coordination pattern of any bound ions, in this case  $\text{Ca}^{2+}$ . In addition, for the known structures (namely sC), crystallographic B-factor values also can be utilised as a guide for each of the target residues to be mutated, since a crystallographic B factor provides additional information regarding the mobility/flexibility of any given atom and/or residue (Reetz et al., 2009). In this case, the individual B-factors were assessed, using the on-line programme B-FITTER (Reetz and Carballeira, 2007), in the sC PDB coordinate file (PDBid: 3UNX).

Target site	Target residue	B-factor (Å <sup>2</sup> )	Rationale
Pro	Pro 210	8.58	Could be mutated to Gly or Ala. Reported in the literature to increase thermolability of sC (Fuchita et al., 2012)
	Pro 5	8.09	Located at the N-terminus, rigid (according to B factor). Increasing flexibility in C and/or N termini has been shown to have a destabilising effect on a protein (Riise et al., 2007)
	Pro 9	11.17	Located further along the N-terminus, moderately rigid according to B factor, contained within a helical motif. Mutating Pro9 could help unravel the helix, which is predicted to increase flexibility (in general, reduction of secondary structures is desirable). This Pro is not conserved, Lys is found in homologous psychrophilic proteases.
	Pro 40	11.52	His9 (B factor 7.38), Pro40 (B factor 11.52), Asp41 (B factor 7.43) known as 'HPD' region. It may be that Pro40 is restricting the conformational flexibility of nearby residues (low B-factor values are observed in this region). Asp41 is coordinating a Ca <sup>2+</sup> ion. These 3 residues are a potential site for randomisation (saturation mutagenesis).
	Pro 85	8.99	Pro85 is well conserved within the superfamily. However, it is found in a loop, which could alter the conformation of the loop and subsequent Ca <sup>2+</sup> coordination.
	Pro 167	9.14	Located in a loop. Strictly conserved.
	Pro 200	7.75	Part of a very conserved stretch of residues.
	Pro 225	5.26	The most rigid residue in the structure. Close to Ser221 – active site residue. Potential to introduce a bulky residue that would not disrupt the helix completely but would introduce some conformational flexibility. Likely not relevant in the first round of mutations but could become important for later consideration, when

			a shift of lower optimal temperature is desired. It also could introduce flexibility in the region of the active site, which could make the active site more readily accessible.
Arg	Arg 247	12.01	Disrupt E197-R247 salt bridge to weaken thermostability.
	Arg 249	15.3	This residue has one contact with Q275, which might be responsible for stabilisation of the C-terminus.
Hydrophobic residues	Met 124	8.74	M124T mutation would change the hydrophobicity pattern and present a more solvent exposed pocket, as Thr does not form favourable hydrophobic contacts, by comparison, and has a smaller side chain. Strictly conserved, away from the active site with respect to sequence, but close in proximity with respect to structure. Could be considered later, if there is a need to increase flexibility in the vicinity of the active site to increase accessibility.
	Met 222	9.87	Strictly conserved, pointing towards the active site, preceding active site residue Ser221. Both M222 and M124 could be mutated to Thr, as it has been reported to be favourable in psychrophilic proteins.
	Val 84	7.01	V84I. This mutation was shown to increase low-temperature activity in subtilisin BPN' compared to wild-type enzyme (Kano et al., 1997).
	Val 72	6.99	V72I mutation was shown to have a large effect on thermostability of subtilisin BPN' (Taguchi et al., 1998b).
	Leucines		Substitution of Leu to Ala/Val was reported to be favourable in psychrophilic enzymes. There are a relatively high number of Leu residues present in sC, therefore, several could be mutated.

Residues on the protein surface	G131	25.77	G131M/G131W substitutions were shown to confer increased low-temperature activity and thermolability for subtilisin BPN'; sC Carlsberg & BPN' share 70% identity (Taguchi et al., 2000). Also, G131D and D197N were reported to cold adapt subtilisin BPN (Tange et al., 1994).
	Ser 98	12.56	S98D, S99D/ S105D, G106D. An increased number of negatively charged amino acids were observed in homologous cold-adapted proteases of sC, particularly, a high percentage of these were exposed on the surface of loops (Kulakova et al., 1999). A negative surface potential is a common feature of cold-adapted enzymes, as many of these are characterized as acidic proteins.
	Ser 99	15.27	
	Ser 105	14.45	
	Gly 106	11.08	
Alternati ve targets	S260-F261- Y262 motif		Destabilisation of the N and/or C termini. The C-terminus could be unravelled to make it more flexible. Flexible C and N termini were reported to lead to thermolability (Riise et al., 2007). Unravelling of the S260-F261-Y262 helical motif could infer thermolability.
	Glu-Asp mutation		Asp has a less favourable conformational entropy for stability than Glu and, Glu to Asp substitutions were shown to decrease the melting temperature of hyperthermophilic enzymes (Lee et al., 2004, Bauvois et al., 2008).

**Table 4.1** The selected site-directed mutations for sC based on an extensive literature survey and multiple sequence alignment, as described in the text.

#### 4.1.1.2 Optimisation of recombinant cloning and expression of Subtilisin Carlsberg

In order to direct the recombinantly expressed sC protein to either the *E.coli* cytoplasm or periplasm, specifically tailored pOPIN vectors (kindly donated by OPPF-UK) were utilised. The recombinant sC gene was designed to be inserted into both pOPIN $\mathbf{P}$  and pOPIN $\mathbf{E}$  expression vectors. These vectors present similar features, except the pOPIN $\mathbf{P}$  (periplasmic expression) vector, also contains the *pelB* leader sequence, which is absent in the pOPIN $\mathbf{E}$  vector (cytoplasmic expression). Also, pOPIN $\mathbf{F}$  was later tried to test the effect of an additional N-terminal 6xHis-tag.

The full open reading frame, containing the sC gene in the pOPIN $\mathbf{P}$  vector, is shown in **Figure 4.1**. Cleavage of the *pelB* leader sequence was predicted, using the SignalP 4.1 server, as being between positions 22 and 23, namely AMA-AQ was identified (Petersen et al., 2011).

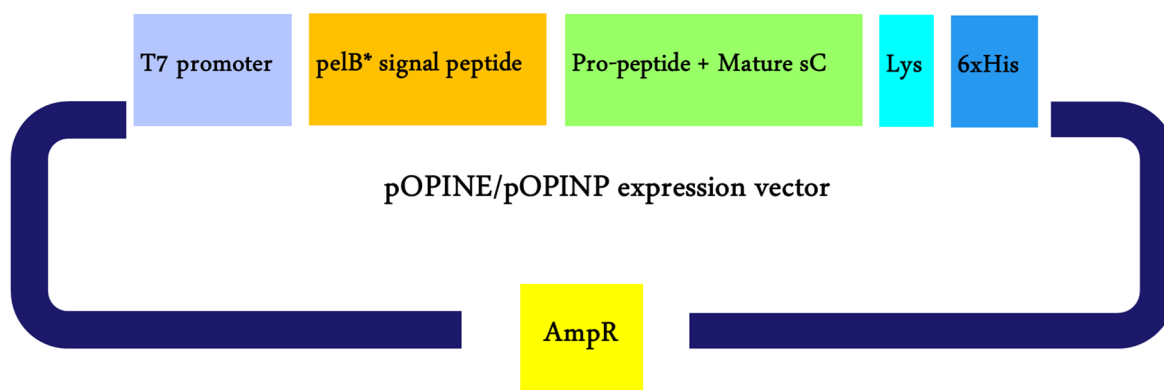
```

MKYLLPTAAAGLLLLAAQPAMAAQPAKNVEKDYIVGFKSGVK TASVKKDI I KESGGKVDKQFRI INAAKAKLD
KEALKEVKNDPDVAYVEEDHVAHALAQTVPYGIPLIKADKVQAQGFKGANVKVAVLDTGIQASHPDLNVVGGA
SFVAGEAYNTDGNHGHGTHVAGTVAALDNTTGVLGVAPSVSLYAVKVLNSSGSGSYSGIVSGIEWATTNGMDVI
NMSLGGASGSTAMKQAVDNAYARGVVVVAAGNSGNSGSTNTIGYPAKYDSVIAVGAVDSNSNRASFSSVGAE
LEVMAPGAGVYSTYPTNTYATLNGTSMASPHVAGAAALILSKHPNLSASQVRNRLSSTATYLGSSFYYGKGLI
NVEAAQQLHHHHHH

```

**Figure 4.1** The amino acid sequence for sC in the pOPIN $\mathbf{P}$  vector. Colour coding corresponds to *pelB* + propeptide + mature polypeptide + L + 6xHis. *PelB* is the signal peptide sequence that directs the expressed recombinant protein from the cytosol to the periplasm in *E. coli*, which in turn increases the chance of obtaining more soluble protein due to a less reducing environment and isolation from the vast majority of other *E.coli* proteins.

A more visual representation of the components of the expression system used is provided in **Figure 4.2**.



**Figure 4.2** Subtilisin Carlsberg expression vector constructed, employing InFusion cloning as described in section 2.2.5. The expression is driven by the T7 promoter; the *pelB* signal peptide directs recombinant protein expression to the periplasm and the propeptide sequence is required for correct folding and activation of Subtilisin Carlsberg. Lys – lysine is recognised by enterokinase for the effective hexahistidine tag cleavage.

\*there is no *pelB* signal peptide present in the pOPINE vector.

The expression strains selected were the conventionally used BL21 (DE3) and Lemo21 (a derivative of the latter). Lemo21 allows titration of the expression levels by addition of L-Rhamnose. This was important due to potential inclusion body formation, as the strong T7 promoter-induced high rate of expression could interfere with the required maturation and correct folding of Subtilisin Carlsberg. The additional capabilities of the Lemo21 expression strain eliminate the need to design further constructs with a more tunable promoter (such as pBAD) in order to control the expression levels of recombinant protein (Guzman et al., 1995).

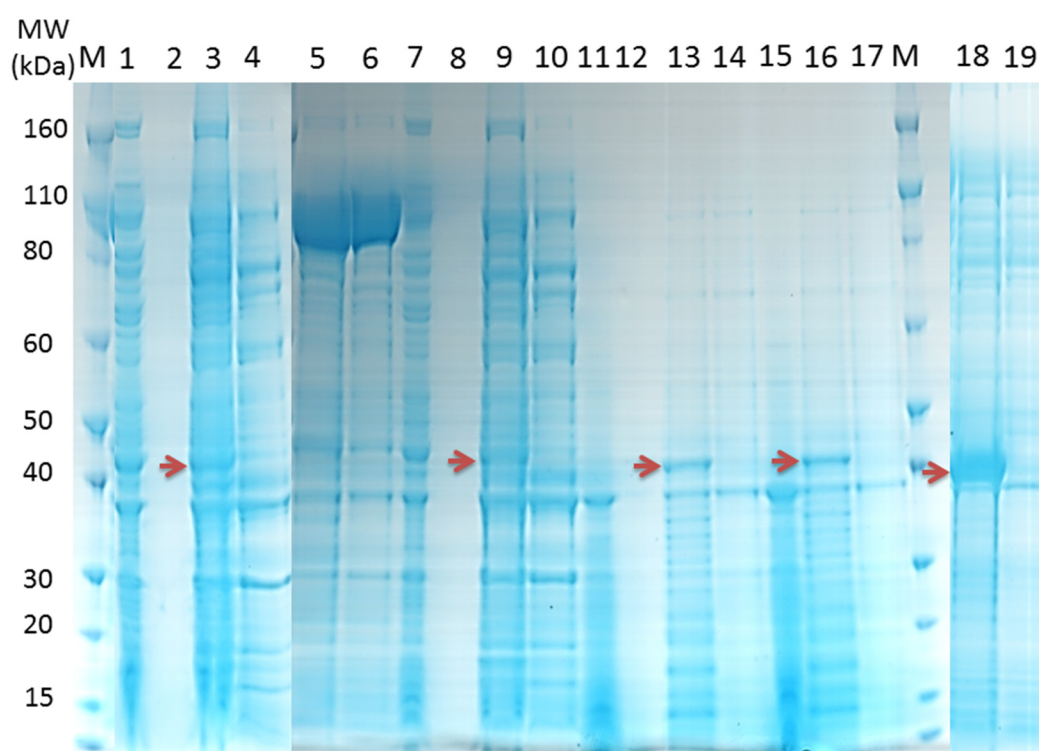
#### 4.1.1.3 Challenging over-expression of sC in *E.coli*

Following the successful cloning into target expression vectors, no recombinant sC was detected by SDS-PAGE and Western Blot in the first three expression trials (data not shown). To help resolve the lack of sC recombinant protein expression, a time course expression experiment, i.e. sample the expression at several continuous time points as the expression proceeds, was carried out. Samples at 2h, 4h, and 16h growth were taken. This was necessary in order to capture the expression of sC at an early stage, in the event that the protein was toxic to the *E.coli* cells and/or autolysis of sC had occurred. Upon ultrasonication processing, the cell resuspension buffer was supplemented with 1mM PMSF, to prevent autolytic degradation of sC. However, the outcome was similar – still, no recombinant sC was

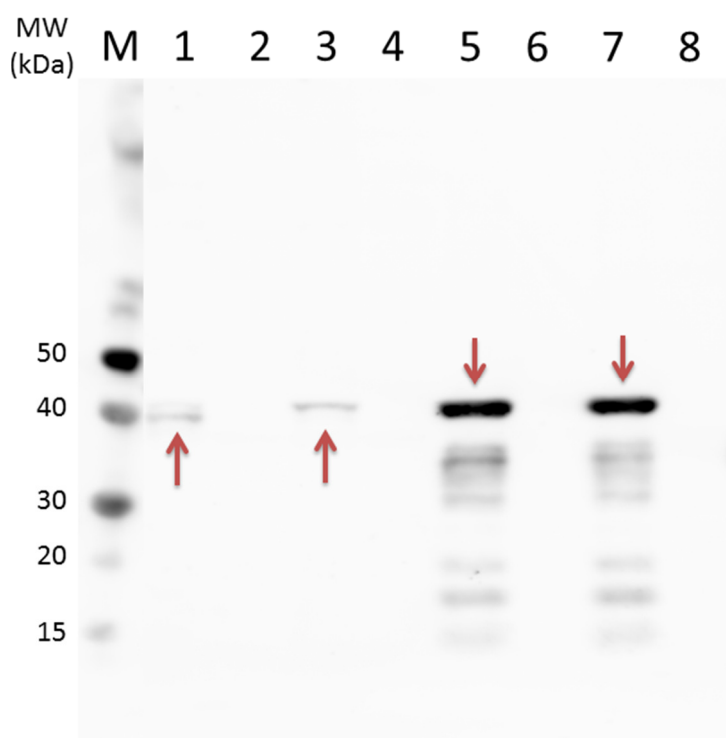
detected. As the initial expression attempts did not yield any recombinant protein, it was decided to re-clone the sC constructs into pOPINP, pOPINE and pOPINF expression vectors utilising In-Fusion cloning as described in section 2.2.5 and re-evaluate expression in *E.coli*.

Following construct verification by Sanger sequencing, sC-pOPINP, sC-pOPINE and sC-pOPINF constructs were obtained, and a small scale expression trial was carried out in both B834 (DE3) and Lemo21 (DE3) expression strains. The cultures were incubated at 37 °C post-induction with IPTG, as described in section 2.6.3.

The over-expressed protein was analysed by SDS-PAGE (**Figure 4.3**) and Western Blot (**Figure 4.4**).



**Figure 4.3** Recombinant sC small scale expression analysis by SDS-PAGE. sC-pOPINP construct: 1,2,3,4 – pre-induction, culture supernatant, lysate, soluble fractions, respectively, in B834 cells; 5,6 – positive control of empty pOPINP; 7,8,9,10 - pre-induction, culture supernatant, lysate, soluble fractions, respectively, in Lemo21 cells; sC-pOPINF construct: 11,12,13,14 - pre-induction, culture supernatant, lysate, soluble fractions, respectively, in B834 cells; 15,16,17 - pre-induction, lysate, soluble fractions, respectively, in Lemo21 cells; sC-pOPINE construct: 18,19 – lysate, soluble fractions, respectively, in B834 cells; Red arrows indicate the location (based on expected molecular weight 38 kDa) of recombinant sC proenzyme.

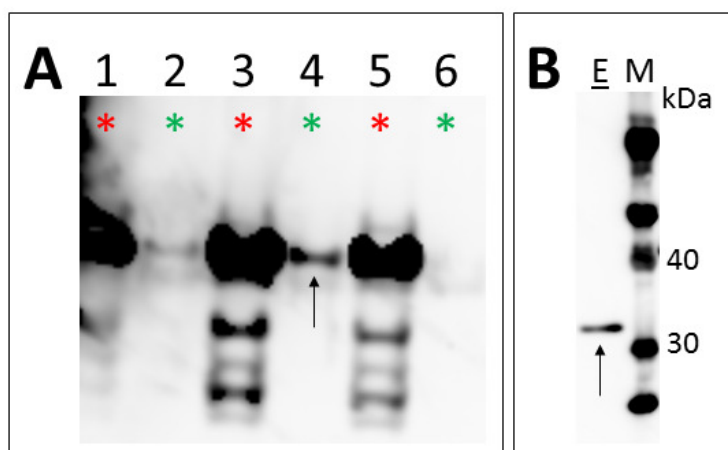


**Figure 4.4** Recombinant sC small scale expression analysis by Western Blot. sC-pOPINP construct: 1,2 - lysate, soluble fractions, respectively, in B834 cells; 3,4 - lysate, soluble fractions, respectively, in Lemo21 cells; sC-pOPINF construct: 5,6 - lysate, soluble fractions, respectively, in B834 cells; 7,8 - lysate, soluble fractions, respectively, in Lemo21 cells; Red arrows indicate the location of recombinant sC proenzyme (expected molecular weight 38 kDa).

SDS-PAGE and Western Blot analysis indicated successful over-expression of sC-pOPINP, E,F constructs as inclusion bodies in the insoluble fractions (**Figure 4.3** and **Figure 4.4**), possibly due to the high post-induction temperature and inappropriate time allowed for polypeptide folding. The predicted migration pattern of these protein constructs was near 38 kDa, which was confirmed by the SDS-PAGE gels. The recombinant sC over-expression levels were quite low for all the constructs tested. This may be linked to the *E.coli* toxicity caused by recombinant Subtilisin Carlsberg and its subsequent degradation during the expression process. In particular, the periplasm-extracted sC (**Figure 4.4** lanes 1 and 3) appeared to be in extremely low quantity.

The next obvious experimental step was to reduce the post-induction temperature in an effort to slow down protein expression and provide a longer time for the folding process. Also, the cells were only allowed to produce sC for 5 hours, in an effort to reduce the extent of

autocatalytic degradation and improve the yield. sC-pOPINE construct was selected as the best expressor, as identified in the previous round (**Figure 4.3**, lane 18). In this next round of sC expression, the temperature was dropped to 28 °C, post-induction.



**Figure 4.5** Western Blot analysis of recombinant expression and purification of the sC-pOPINE construct. **A.** insoluble and soluble fractions are indicated by red and green stars, respectively; 1,2 – sC expressed in Lemo21 cells; 3,4 - sC expressed in B834 cells, harvested after 5 h; 5,6 - sC expressed in B834 cells, harvested after 16 h; **B.** IMAC His-Trap purification: E – imidazole elution (100% buffer B). Soluble sC population is indicated by arrows.

**Figure 4.5** demonstrates that reduction of post-induction temperature had a substantial effect on soluble expression of sC. Although a small fraction of recombinant sC appeared as a soluble protein in B834 cells (Fig. 4.5 lane 4), the vast majority stayed in inclusion bodies. Also, the prolonged expression time resulted in sC degradation (Fig. 4.5 lane 6).

A small quantity (only traceable on Western Blot) of active sC could be recovered, following an IMAC His-Trap purification (lane E) via the engineered C-terminal hexahistidine tag, suggesting a high rate of autocatalytic degradation. As a result, only a trace amount of pure sC was obtained from 1.6 litres of *E.coli* B834 culture.

#### **4.1.2 Selected exoprotease – leucine aminopeptidase A (LapA)**

*Aspergillus* – derived LapA exopeptidase is an extracellularly secreted enzyme. LapA originates from a eukaryotic organism and is known to contain N-linked glycosylation and intrinsic disulphide linkage, which is predicted to be essential for activity (Matsushita-Morita et al., 2011). In addition, as the N-terminal propeptide must precede the mature sequence in all successive expression constructs, suggests a requirement of enzymatic activation to obtain a functional LapA protein. Therefore, the recombinant *E.coli* cloning and expression strategy for LapA included directing the protein to the periplasmic space to eliminate the disulphide-unfavourable reducing environment of the *E.coli* cytoplasm. LapA was produced, as a precursor, both intracellularly and in the periplasmic space (section 4.1.2.1). At this stage, post-translational N-glycosylation was considered non-essential for the activity of LapA.

##### **4.1.2.1 Expression of recombinant Lap thermolability mutants**

At the onset of this work, a number of LapA and Lap2 expression constructs were obtained from Biocatalysts Ltd (the project industrial partner). The constructs obtained were designed *in silico* by Dr. Jonathan Mullins (Cardiff University), as part of a separate project with Biocatalysts Ltd, prior to our involvement in the project. However, none of these constructs had been tested for over-expression (**Table 4.2**). At the same time, the superiority of the SoluBL21 expression strain, to improve the recombinant protein solubility, was planned to be evaluated. Previous attempts (by Biocatalysts Ltd) showed that, for the constructs tested, the Lap mutants expressed as inclusion bodies (see section 3.6). Therefore, new attempts focused on the use of the SoluBL21 expression strain, in an effort to obtain soluble expression of the Lap mutants, as required by Biocatalysts Ltd.

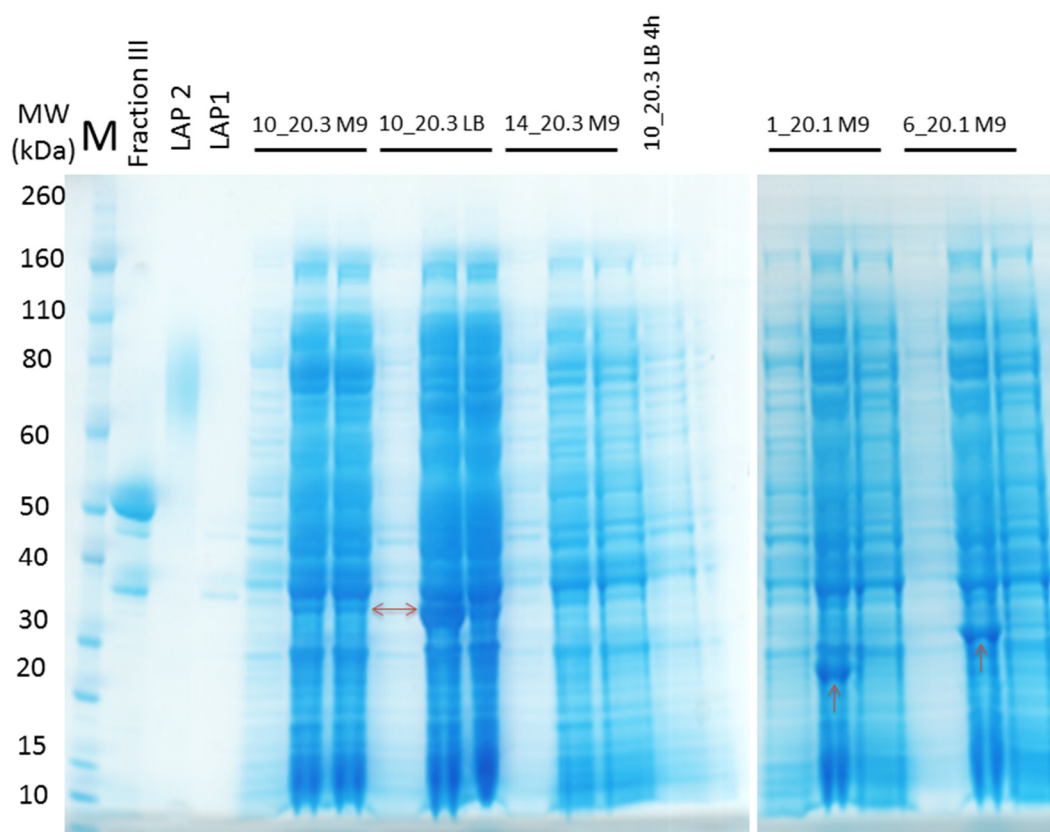
Sample	Plasmid	Plasmid size / kbp	Protein	Protein molecular weight / kDa
1-20.1_133-361	pET21a_1-20.1	6.118	<i>Aspergillus oryzae</i> RIB40 Leucine Aminopeptidase 20.1 MUTANT 1	25.572
2-20.1_133-361_stripped_all	pET21a_2-20.1	6.037	<i>Aspergillus oryzae</i> RIB40 Leucine Aminopeptidase 20.1 MUTANT 2	22.923
3-20.1_133-361_minus257-261	pET21a_3-20.1	6.103	<i>Aspergillus oryzae</i> RIB40 Leucine Aminopeptidase 20.1 MUTANT 3	25.116
4-20.1_133-361minus309-323	pET21a_4-20.1	6.073	<i>Aspergillus oryzae</i> RIB40 Leucine Aminopeptidase 20.1 MUTANT 4	24.072
5-20.1_133-361plus20.3_1-74	pET21a_5-20.1	6.340	<i>Aspergillus oryzae</i> RIB40 Leucine Aminopeptidase 20.1 MUTANT 5	34.039
6-20.1_133-361plus1-30_20.3	pET21a_6-20.1	6.208	<i>Aspergillus oryzae</i> RIB40 Leucine Aminopeptidase 20.1 MUTANT 6	29.087
7-20.1_133-361minus309-323plus1-74_20.3	pET21a_7-20.1	6.295	<i>Aspergillus oryzae</i> RIB40 Leucine Aminopeptidase 20.1 MUTANT 7	32.538
8-20.1_133-361minus309-323plus19-74_20.3	pET21a_8-20.1	6.241	<i>Aspergillus oryzae</i> RIB40 Leucine Aminopeptidase 20.1 MUTANT 8	30.471
9-20.1_133-361plus20.3_1-30_220-221	pET21a_9-20.1	6.214	<i>Aspergillus oryzae</i> RIB40 Leucine Aminopeptidase 20.1 MUTANT 9	29.321
10-20.3_1-298	pET21a_10-20.1	6.325	<i>Aspergillus oryzae</i> RIB40 Leucine Aminopeptidase 20.3 MUTANT10	33.428
11-20.3_minus1-74	pET21a11-20.1	6.103	<i>Aspergillus oryzae</i> RIB40 Leucine Aminopeptidase 20.3 MUTANT11	24.962
12-20.3 Removed: 1-74 102-104 106-109 133-138 190-192 194-199	pET21a_12-20.1	6.037	<i>Aspergillus oryzae</i> RIB40 Leucine Aminopeptidase 20.3 MUTANT12	22.676
13-20.3_1-298minus1-74minus220-221	pET21a_13-20.1	6.097	<i>Aspergillus oryzae</i> RIB40 Leucine Aminopeptidase 20.3 MUTANT13	24.727
14-20.3_1-298plus20.1_257-261	pET21a_14-20.1	6.340	<i>Aspergillus oryzae</i> RIB40 Leucine Aminopeptidase 20.3 MUTANT14	33.885
15-20.3_1-298plus20.1_247-253, 257-261	pET21a_15-20.1	6.361	<i>Aspergillus oryzae</i> RIB40 Leucine Aminopeptidase 20.3 MUTANT15	34.576
16-20.3_1-298minus1-74_220-221plus20.1_247-253_257-261	pET21a_16-20.1	6.133	<i>Aspergillus oryzae</i> RIB40 Leucine Aminopeptidase 20.3 MUTANT16	25.876

**Table 4.2** The list of LapA (sequence denoted 20.3) and Lap2 (sequence denoted 20.1) mutants predicted to show thermolability, according to predictive work of J. Mullins (Cardiff). Adapted from a Biocatalysts Ltd internal report.

All LapA and Lap2 mutants (**Table 4.2**) were engineered (*in silico*) towards achieving thermolability, using a novel thermodynamic approach developed by Dr J. Mullins. The sequence alterations mainly included deletions, insertions and truncations of N- and C-terminal parts of Lap proteins. Only the mutant numbered 10\_20.3 (LapA) was the unmodified mature version of the wild - type(WT) enzyme.

An initial expression trial was conducted with 10\_20.3, 14\_20.3, 1\_20.1, 6\_20.1 constructs. The SoluBL21 *E.coli* cell line and M9 minimal media were utilised, and the protein over-expression was carried out, according to sections 2.6.2 – 2.6.4. Induced cells were grown at RT for 16 h. Also, construct 10\_20.3 (LapA WT) was expressed in SoluBL21 combined with LB Broth. Induced cells were grown at 18 °C for 16 h.

Figure 4.6 shows that recombinant expression was successfully obtained for 10\_20.3 (LapA WT), 1\_20.1 (Lap2), 6\_20.1 (Lap2). A fraction of 10\_20.3 protein was expressed as soluble protein, whereas both 20.1 mutants were expressed as inclusion bodies (observed in the cell lysate but not in the soluble supernatant). Constructs 1\_20.1 and 6\_20.1 match the expected molecular weights of 25.5 kDa and 29 kDa, respectively (**Table 4.2**). No obvious recombinant expression was detected for the 14\_20.3 (LapA) mutant.



**Figure 4.6** SDS-PAGE gel of LapA and Lap2 mutant expression trial. Left to right: M – protein ladder, Fraction III from SEC of Flavourzyme, Lap2 – purified from Flavourzyme, LapA – purified from Flavourzyme, 10\_20.3 M9 – pre-induction+lysate+soluble supernatant samples (expressed in M9 minimal media), 10\_20.3 LB – pre-induction+lysate+soluble supernatant samples (expressed in LB Broth), 14\_20.3 M9 – pre-induction+lysate+soluble supernatant samples (expressed in M9 minimal media), 10\_20.3 LB 4h – sample after 4h of induction, 1\_20.1 M9 – pre-induction+lysate+soluble supernatant samples (expressed in M9

minimal media), 6\_20.1 M9 - pre-induction+lysate+soluble supernatant samples (expressed in M9 minimal media). Recombinant proteins are indicated by red arrows.

Additional Lap proteins from the mutant library (**Table 4.2**) were tested for expression in the SoluBL21 strain. However, none of the constructs tested (as illustrated in **Figure 4.6**) produced soluble protein in the intracellular *E.coli* expression trials (results not shown). This was verified by both SDS-PAGE and Western Blot analysis. Only wild - typeLapA (10\_20.3) appeared in the soluble fraction using the SoluBL21 strain (confirmed by Western Blot).

The insolubility of LapA and Lap2 thermal mutants was most likely due to misfolding. Upon receipt of the actual engineered sequences, a thorough investigation of the effects of these mutations on an homology model of Lap (**Figure 10.7**, Appendix II) used prior to our subsequent structure determination, suggested that the engineered alterations for the majority of these mutants were extensive, often involving complete removal of essential, core regions of secondary structure that would very likely affect proper folding of the polypeptide. Thus, it was difficult to establish whether SoluBL21 strain was advantageous to the general solubility of the recombinant protein. However, the fact that only the wild - typesequences showed soluble expression, further suggests the detrimental effect of the mutations on the protein integrity.

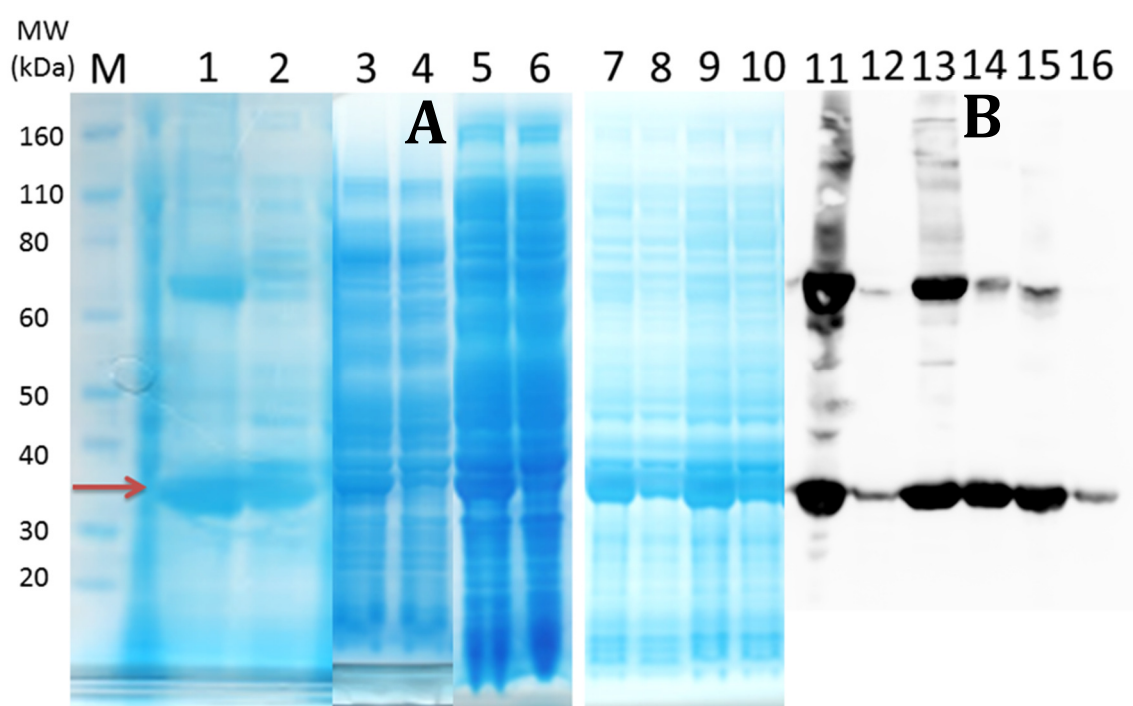
SoluBL21 cell lysate of 20.3\_10 (LapA WT) was tested for activity. Only about 10% activity ( $9.1 \pm 0.2$  U/mg) was observed compared to the activity of native Flavourzyme-purified LapA ( $94.6 \pm 2.2$  U/mg). As this was crude cell lysate, some non-specific processing of Leucine-p-nitroanilide (substrate) might have occurred. Although 20.3\_10 recombinant protein was active, it may have suffered from the lack of available  $\text{Zn}^{2+}$  ions required for enhanced activity. And, even though 20.3\_10 (LapA WT) was present in the soluble supernatant of the SoluBL21 cell lysate, the weak activity might stem from the protein being inappropriately or only partially folded.

#### 4.1.2.2 Characterisation of recombinant mature LapA (20.3\_10)

Since 20.3\_10 LapA construct provided access to soluble mature protein in *E.coli* lysate, it was decided to explore its properties further.

##### 4.1.2.2.1 Expression levels of recombinant mature LapA (20.3\_10)

In order to estimate the expression levels of soluble recombinant LapA in different *E.coli* expression strains, 10\_20.3 plasmid DNA was transformed into BL21 (DE3), Lemo21 (DE3) and SoluBL21 expression strains. Additionally, the effect of culturing BL21 cells in M9 minimal media versus LB media was investigated. The expression levels were verified by SDS-PAGE and Western Blot analysis (**Figure 4.7**).



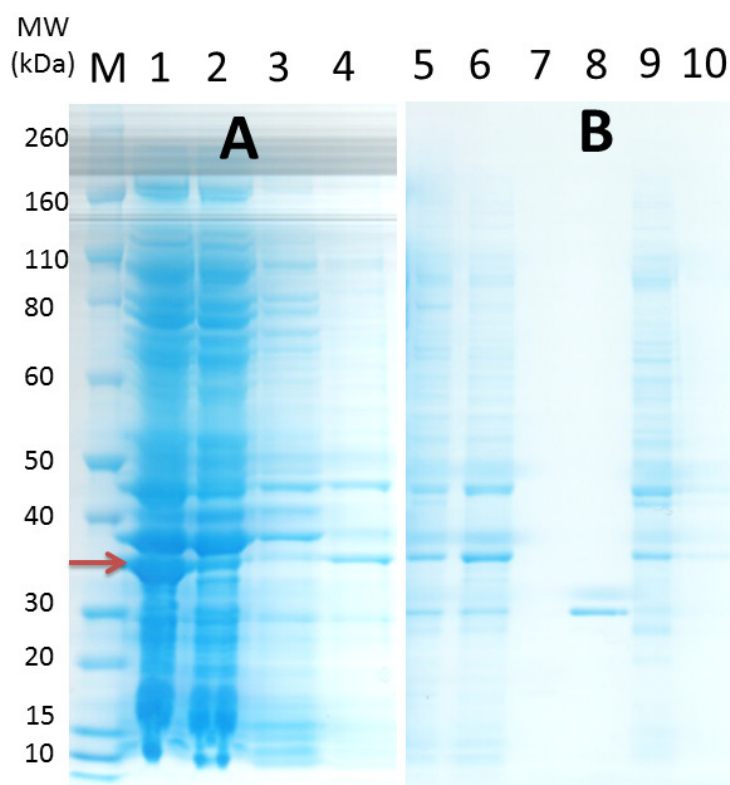
**Figure 4.7** Assessing expression of soluble recombinant LapA (20.3\_10). **A.** SDS-PAGE 1, 2 – insoluble and soluble fractions, respectively, BL21 (DE3); 3, 4 - insoluble and soluble fractions, respectively, Lemo21 (DE3); 5,6 - insoluble and soluble fractions, respectively, SoluBL21; 7,8 - insoluble and soluble fractions, respectively, BL21 (DE3) in M9 media; 9,10 - insoluble and soluble fractions, respectively, BL21 (DE3) in LB media; **B.** WB 11, 12 – insoluble and soluble fractions, respectively, in WB of SoluBL21; 13,14 - insoluble and soluble fractions, respectively, in WB of BL21 (DE3); 15,16 - insoluble and soluble fractions, respectively, in WB of Lemo21; red arrow indicates the location of recombinant LapA enzyme.

BL21 (DE3) produced the most soluble LapA enzyme (**Figure 4.7 A**, lane 2). The levels of expression from Lemo21 (DE3) and SoluBL21 appear comparable (**Figure 4.7 A**, lanes 4 and 6). This was further verified by Western Blot, where most of the soluble LapA was obtained in BL21 (DE3) cells (**Figure 4.7 B**, lane 14). No increase in the expression of soluble LapA protein was observed in BL21 (DE3) cells when M9 versus LB expression was compared (**Figure 4.7 A**, lanes 8 and 9). Therefore, BL21 (DE3) expression strain, combined with growth in LB medium, was selected for further over-expression of recombinant LapA.

It is important to note that 20.3\_10 protein was prone to formation of higher oligomeric states (~70 kDa band, **Figure 4.7 B**, lanes 11 - 15) in both insoluble and soluble (lower extent) fractions, as indicated by Western-Blot analysis.

#### 4.1.2.2.2 Purification and characterisation of soluble LapA (20.3\_10)

Soluble recombinant LapA, obtained from *E.coli* BL21 (DE3) lysate, was further purified and analysed. 20.3\_10 LapA was purified using a 2-step protocol, whereby IMAC purification was followed by anion exchange chromatography, as described in section 2.7.1.



**Figure 4.8** Purification of recombinant 20.3\_10 LapA by IMAC (**A**) and IEX (**B**). 1 – soluble supernatant of BL21 (DE3) containing LapA (red arrow), 2 – flow through from IMAC, 3 – unbound protein wash from IMAC, 4 – eluted fraction containing LapA from IMAC, 5-

dialysed LapA fraction following IMAC, 6 – IMAC-eluted LapA fraction loaded on IEX column, 7 – IEX flow through, 8, 9, 10 – peaks collected from IEX. The red arrow indicates the location of recombinant mature LapA enzyme (expected molecular weight 33.3 kDa).

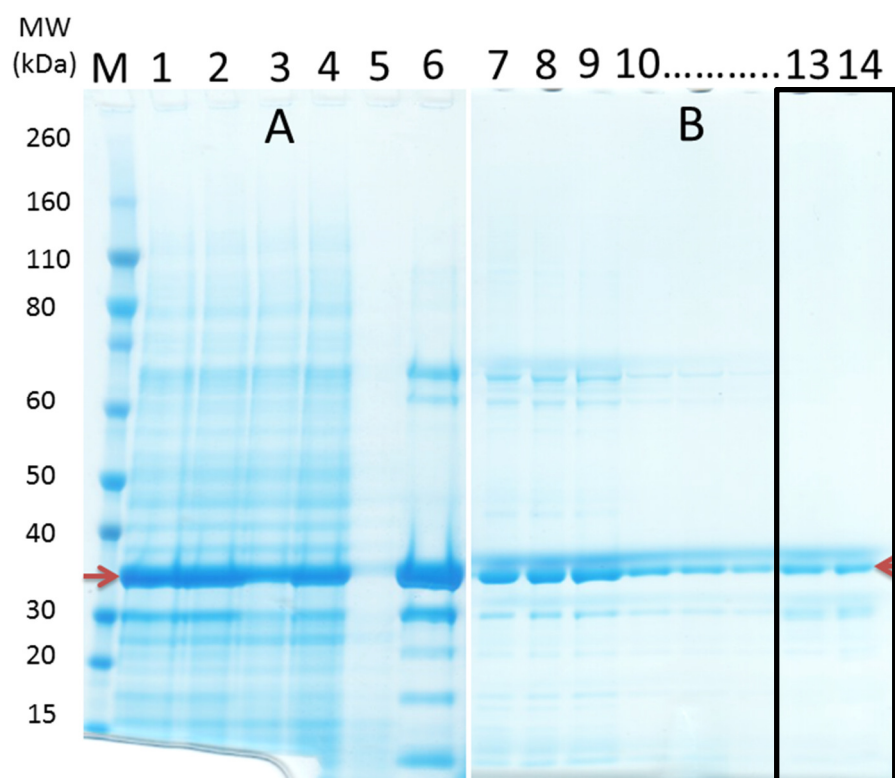
The IMAC-eluted fraction, containing recombinant LapA (**Figure 4.8A**, lane 4) had several contaminant proteins present. Thus, IEX chromatography was performed to further purify recombinant LapA, following overnight dialysis in 50 mM Bis-Tris Propane pH 8.0, 1mM ZnCl<sub>2</sub>. Dialysis was carried out to remove 0.5 M sodium chloride and 0.2 M imidazole, resulting from IMAC elution. The peaks in the IEX purification step were eluted at 110 mM, 400-800 mM, 850-1000 mM of NaCl for lane 8, 9 and 10, respectively (**Figure 4.8**). Most of the recombinant LapA is contained within the fraction corresponding to lane 9. Elution of the LapA fraction at a high NaCl concentration implies a very tight interaction with the anion exchange resin. Thus, although IMAC purification provided separation of LapA from crude cell lysate, not much further separation was achieved in the subsequent IEX step, likely due to the very highly charged nature of LapA or possibly due to misfolding (which potentially exposes more hydrophobic and/or charged residues for enhanced interaction with other proteins).

Lap activity assay was performed on the IMAC- and IEX-purified LapA fractions (see section 2.10.1), which exhibited only about 10% ( $9.1 \pm 0.2$  U/mg) of the activity obtained in the Flavourzyme-purified native LapA fraction ( $94.6 \pm 2.2$  U/mg). Since the 20.3\_10 construct possesses a hexahistidine tag anchored to its N-terminus, followed by the mature domain of Lap, it is likely that the inability to acquire a full catalytically competent enzyme stems from the lack of the ‘chaperone-like’ propeptide domain.

#### **4.1.2.3 Refolding studies of recombinant mature LapA (20.3\_10)**

Since the observed activity of the recombinant LapA, obtained from soluble BL21 (DE3) over-expression, was particularly low, it was decided to re-solubilise the inclusion bodies using 8 M urea. Inclusion bodies, containing insoluble recombinant LapA, were re-solubilised in 200 mL buffer containing 50 mM Tricine pH 8.0, 500 mM NaCl and 8 M Urea, followed by stirring for 24 h at 4 °C. The solubilised supernatant was loaded onto an IMAC column and the purification proceeded according to section 2.7.1, except 8 M urea was included in the purification buffers. The IMAC-eluted fraction, containing LapA in 8 M urea (**Figure 4.9**, lane 6), was dialysed against 10 mM Tricine pH 8.0, 1 mM ZnCl<sub>2</sub>.

Subsequently, 8 mg of dialysed and re-folded LapA preparation was subjected to SEC purification, and the active fractions were pooled together (**Figure 4.9 B**, lanes 13 and 14).



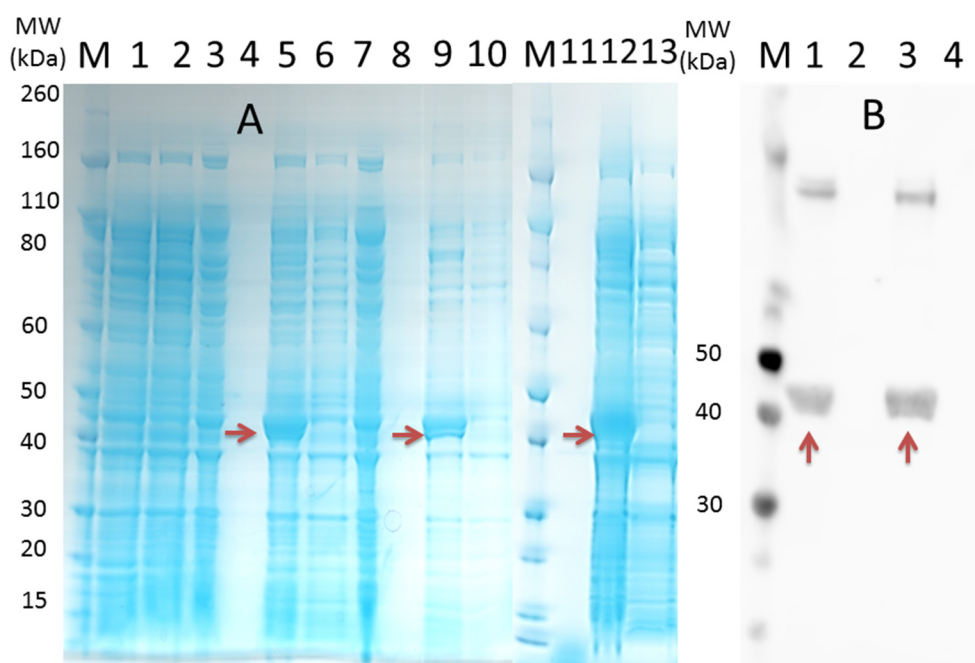
**Figure 4.9** Purification of recombinant LapA from inclusion bodies by IMAC (**A**) and SEC (**B**). 1 – sample of re-solubilised pellet, 2,3,4 – flow through steps, 5 – unbound wash, 6 – eluted fraction; 7,8,9,10-12 – inactive SEC fractions, 13, 14 – active SEC fractions. Red arrows indicate the location of recombinant LapA enzyme. Black rectangle indicates pooled active LapA fractions.

Binding of recombinant LapA to the IMAC column was not particularly efficient, as seen by substantial protein in the flow-through from the column (**Figure 4.9 A**, lanes 2 - 4). The imidazole-eluted fraction still contained several contaminant proteins. The bulk of the protein eluted from the size exclusion column was inactive, possibly due to mis-folding and/or non-specific aggregation. However, a correctly folded subset of the recombinant LapA eluted in separate fractions (**Figure 4.9 B**, lanes 13-14) and was combined for subsequent crystallisation trials, which were not successful (data not presented).

The final pool of the refolded recombinant LapA (*refLapA*) showed a specific activity of  $59.5 \pm 1.1$  U/mg, which was slightly lower compared to the purified *flavLapA* ( $94.6 \pm 2.2$  U/mg).

#### 4.1.2.4 Cloning and expression of recombinant LapA proenzyme (pro-LapA)

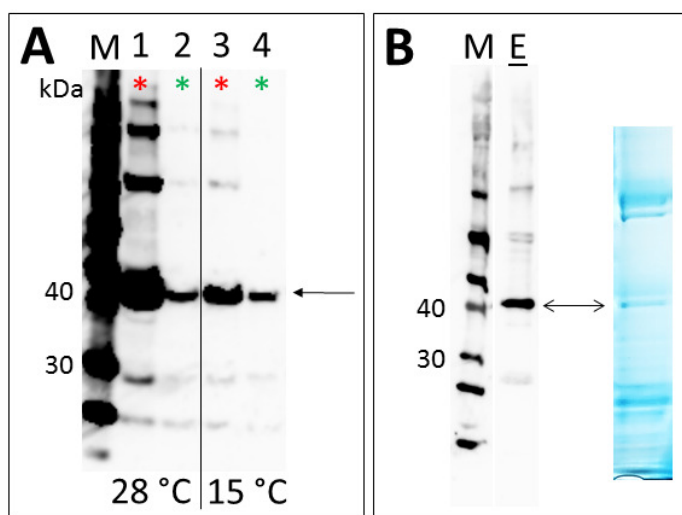
The previous work indicated potential problems with the folding of mature LapA polypeptide in *E.coli* (sections 4.1.2.2.2 and 4.1.2.3). Recombinant mature sequence of LapA was cloned, along with its native propeptide sequence, as it had been reported previously that the propeptide sequence was necessary for proper folding in the case of Lap from *Vibrio sp.* (Bzymek et al., 2004) - the closest structural homologue to LapA. Thus, pro-LapA was cloned into the pOPINP expression vector, utilising In-Fusion cloning, as described in section 2.2.5. Following construct verification by Sanger sequencing (section 2.12.2), a pro-LapA-pOPINP construct was obtained. A small scale expression trial for this new construct was carried out in B834 (DE3), Lemo21 (DE3) and Rosetta (DE3) expression strains. The cultures were incubated at 37 °C, post-induction with IPTG, as described in section 2.6.3. After 16 h, expression was analysed by SDS-PAGE and Western Blot (**Figure 4.10**).



**Figure 4.10** Recombinant pro-LapA expression analysis. **A** (SDS-PAGE): 1,2 – negative control of empty pOPINP vector: lysate (1), soluble (2); 3,4,5,6 – pre-induction, culture supernatant, lysate, soluble fractions, respectively, in B834 cells; 7,8,9,10 - pre-induction, culture supernatant, lysate, soluble fractions, respectively, in Lemo21 cells; 11,12,13 - culture supernatant, lysate, soluble fractions, respectively, in Rosetta cells; **B** (Western Blot): 1, 2 - lysate, soluble fractions, respectively, in B834 cells; 3,4 - lysate, soluble fractions, respectively, in Lemo21 cells; Red arrows indicate the location of recombinant pro-LapA enzyme.

SDS-PAGE and Western Blot analysis indicated expression of the pro-LapA-pOPINP construct in inclusion bodies only, which could be due to the high post-induction temperature and insufficient time allowed for proper/complete polypeptide folding. The predicted migration pattern of this protein construct was at 40 kDa, which was observed. Western Blot analysis also revealed that recombinant LapA might form a dimer (high MW band above 50 kDa), which could not be separated under reducing and denaturing SDS-PAGE conditions.

Recombinant LapA production in the periplasm was not advantageous, with regard to its solubility. Therefore, it was decided to attempt cytoplasmic over-expression and improve protein solubility by controlling the rate of folding at lower temperatures. For this purpose, a pro-LapA-pOPINE construct was made, harbouring a C-terminal hexahistidine tag for downstream purification, with both 28 and 15 °C post-induction temperatures tried. The decreased post-induction temperature had a positive effect on pro-LapA solubility. **Figure 4.11 A** shows that more LapA proenzyme was produced at 28 °C. However, no further gain in soluble pro-LapA population was noticeable when the temperature was reduced to 15 °C. LapA proenzyme bound the IMAC affinity resin, however, the elution profile indicated either the presence of co-eluting contaminant proteins or other oligomeric states of pro-LapA recombinant protein (**Figure 4.11**). The latter is more apparent from the SDS-PAGE analysis of IMAC-eluted fraction, where additional bands are present above and below pro-LapA (**Figure 4.11 B**).



**Figure 4.11** Western Blot analysis of recombinant expression and purification of the pro-LapA-pOPINE construct. **A** insoluble and soluble fractions are indicated by red and green stars, respectively; 1,2 – expression in B834 cells at 28 °C; 3,4 - expression in B834 cells at 15 °C; **B** IMAC purification: E – imidazole elution (100% buffer B); an equivalent E loaded on SDS-PAGE is shown on the far right; Soluble pro-LapA population is indicated by arrows.

From the initial expression trials in *E.coli*, it was clear that pro-LapA was prone to forming non-specific aggregates and formation of inclusion bodies could be slightly alleviated by reducing the post-induction temperature. However, only a trace amount of inactive LapA precursor could be recovered, following purification, and the contaminants constituted a higher proportion in the eluate.

## **4.2 Concluding remarks from target protein expression in *E.coli***

Recombinant expression of Subtilisin Carlsberg in *E.coli* proved to be very challenging, mainly due to insolubility and autocatalytic degradation of the target protein sequence tried. Directing protein transport to the periplasmic space did not circumvent these challenges. Only traces of purified sC could be detected on Western Blot, obtained from a relatively large amount of biomass (1.6 L). This would hinder subsequent *in vitro* studies, including thermolability mutant expression and crystallisation, of Subtilisin Carlsberg.

Similar challenges were experienced with recombinant over-expression of leucine aminopeptidase A in *E.coli*. Although the refolded population of mature LapA (*refLapA*) from *E.coli* inclusion bodies was moderately active, it was highly likely that propeptide sequence was required to obtain a fully functional enzyme. However, the bulk of recombinant LapA, produced with its propeptide domain, appeared in inclusion bodies, even following optimisation of the expression conditions. Additionally, the intracellularly obtained soluble fraction of pro-LapA could not be readily isolated from other *E.coli* lysate proteins, was susceptible to aggregation, and was shown to be inactive.

Coincident with this work, preliminary recombinant LapA expression trials in *Pichia pastoris* were showing promise (worked conducted by research scientists at Biocatalysts Ltd). Therefore, in conclusion to these two pieces of work, it was agreed that food grade recombinant expression host *Pichia pastoris* could provide a more viable route for over-production of the target enzyme sequences. This approach could offer additional advantages, as discussed below.

### 4.3 *Pichia pastoris* as host for recombinant protein expression

LapA enzymes originate from the eukaryotic fungus *Aspergillus oryzae*. Post-translational modifications, such as N-linked glycosylation, were predicted to be present in native LapA produced by *A. oryzae* and subsequently could be important either for folding and/or function of the enzyme, providing a further driver for the choice of *Pichia pastoris* as a suitable expression host.

#### 4.3.1 Initial attempts of wild-type LapA over-expression

For this work, two expression constructs were produced; one harbouring a native signal secretion sequence plus the propeptide sequence preceding the LapA N-terminal mature domain, the other harbouring a MAT $\alpha$  *S. cerevisiae* secretion sequence plus the propeptide sequence preceding the LapA N-terminal mature domain. The preliminary work (by Biocatalysts Ltd) showed that both of these constructs yielded comparable levels of extracellular secreted recombinant LapA precursor. However, due to the presence of propeptide, the expressed protein required subsequent enzymatic activation, which was not a cost-effective process for Biocatalyst Ltd. Nevertheless, this work provided the rationale to continue recombinant LapA expression in *P. pastoris* and to explore a variety of other expression constructs, in an effort to obtain soluble, functional mature protein.

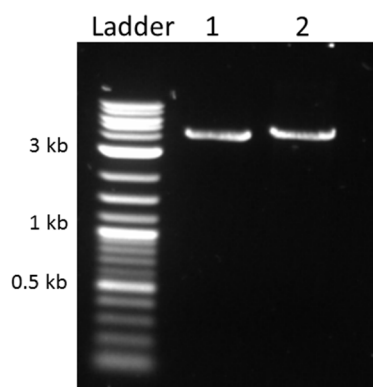
LapA (codon optimised and synthesised gene by DNA 2.0, California, USA) containing its N-terminal native signal sequence and propeptide was placed under the control of a P<sub>AOX</sub> methanol-inducible promoter in a pD912 expression vector. In this way, an expression construct harbouring the LapA native signal peptide and propeptide sequences (pD-npro-LapA) was obtained. Other variants of LapA also were produced (pD- $\Delta$ pro-LapA and pD-npro-LapA-His6, **Table 4.3**), as detailed in the following sections. The rationale behind making the last two constructs was to determine whether LapA could be expressed and properly folded in the absence of propeptide in *P.pastoris* ( $\Delta$ pro construct) and whether subsequent expression could be facilitated by co-linking a hexahistidine tag to the C-terminus of extracellular expressed LapA.

Expressi on construct	Protein construct	Open reading frame sequence
pD-npro- LapA	<i>wtproLapA</i>	MRFLPC <sup>red</sup> IATLA <sup>red</sup> ATASALA <sup>red</sup> IGD <sup>blue</sup> HVRSDDQYVLELAPGQTKV VTEAEK <sup>blue</sup> WALRAEGKRFFDITERASSLELASNKKQKLAVTY PDSVQHNETVQNLIKSLDKKNFETVLQPFSEFHNRYKSDN GKKSS <sup>black</sup> EWLQGGIQEIISASGAKGVTVEPFKHSFPQSSLIAKIP GKSDKTIVLGAHQDSINLDSPEGRAPGADDDGSGVV <sup>black</sup> TILE AFRVLLTDEKVAAGEAPNTVEFHFYAGEEGLLGSQDIFEQ YSQKSRDVKAMLQQDMTGYTKGTTDAGKPESIGIITDNVD ENLTKFLKVIVDAYCTIPTVDSKCGYGCS <sup>black</sup> DHASATKYGY AAFAFESAFGDDSPYIHSADDTIETVNF <sup>black</sup> DHVLQHGR <sup>black</sup> LT <sup>black</sup> LG <sup>black</sup> F AYELAFADSL
pD-npro- LapA- His6	<i>wtproLapA<sup>his</sup></i>	MRFLPC <sup>red</sup> IATLA <sup>red</sup> ATASALA <sup>red</sup> IGD <sup>blue</sup> HVRSDDQYVLELAPGQTKV VTEAEK <sup>blue</sup> WALRAEGKRFFDITERASSLELASNKKQKLAVTY PDSVQHNETVQNLIKSLDKKNFETVLQPFSEFHNRYKSDN GKKSS <sup>black</sup> EWLQGGIQEIISASGAKGVTVEPFKHSFPQSSLIAKIP GKSDKTIVLGAHQDSINLDSPEGRAPGADDDGSGVV <sup>black</sup> TILE AFRVLLTDEKVAAGEAPNTVEFHFYAGEEGLLGSQDIFEQ YSQKSRDVKAMLQQDMTGYTKGTTDAGKPESIGIITDNVD ENLTKFLKVIVDAYCTIPTVDSKCGYGCS <sup>black</sup> DHASATKYGY AAFAFESAFGDDSPYIHSADDTIETVNF <sup>black</sup> DHVLQHGR <sup>black</sup> LT <sup>black</sup> LG <sup>black</sup> F AYELAFADSLK <sup>purple</sup> HHHHHH
pD-Δpro- LapA	<i>wtΔproLapA</i>	MRFLPC <sup>red</sup> IATLA <sup>red</sup> ATASALA <sup>red</sup> YPDSVQHNETVQNLIKSLDKKNF ETVLQPFSEFHNRYKSDNGKKSS <sup>black</sup> EWLQGGIQEIISASGAK GVTVEPFKHSFPQSSLIAKIPGKSDKTIVLGAHQDSINLDS PEGRAPGADDDGSGVV <sup>black</sup> TILEAFRVLLTDEKVAAGEAPNTVE FHFYAGEEGLLGSQDIFEQYSQKSRDVKAMLQQDMTGY TKGTTDAGKPESIGIITDNVDENLTKFLKVIVDAYCTIPTVD SKCGYGCS <sup>black</sup> DHASATKYGYPAFAFESAFGDDSPYIHSADD TIETVNF <sup>black</sup> DHVLQHGR <sup>black</sup> LT <sup>black</sup> LG <sup>black</sup> FAYELAFADSL

**Table 4.3** Open reading frame amino acid sequences of LapA expression constructs synthesised in pD912 *P. pastoris* expression vector. Native LapA secretion signal, propeptide, mature sequence and His6 tag are depicted in red, blue, black and purple, respectively.

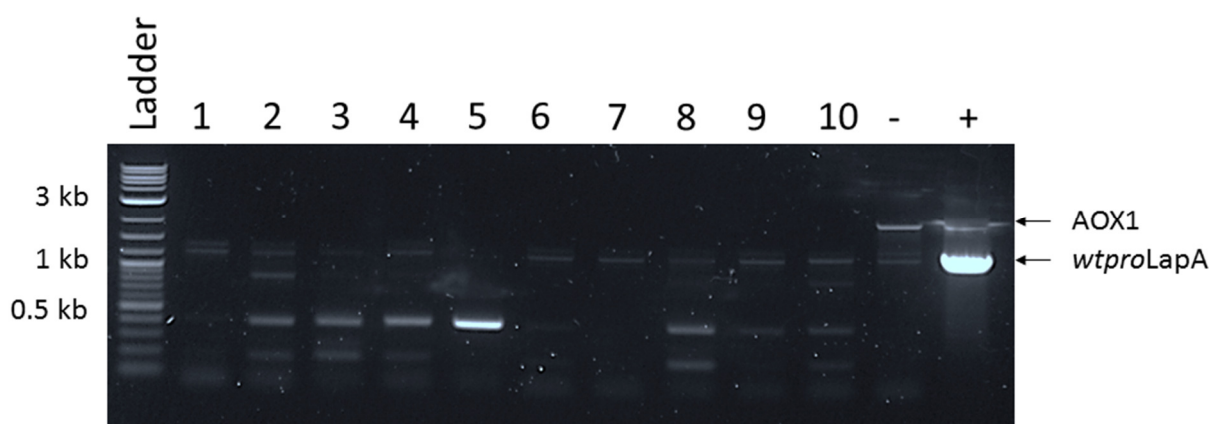
#### 4.3.1.1 Generating a *P. pastoris* expression strain for *wtproLapA*

20 µg of pD-npro-LapA plasmid DNA was linearized using SacI-HF endonuclease, as described in section 2.13.1. Approximately 7 µg of linearized DNA was recovered for transformation. The quality of linearized DNA was evaluated on 1% agarose gel (**Figure 4.12**).



**Figure 4.12** 1% DNA agarose gel of SacI-HF linearized pD-npro-LapA plasmid DNA. Ladder - 1 kb DNA Ladder (NEB, UK), Lane 1 - Linear empty pD912 vector, Lane 2 – linear pD-npro-LapA.

**Figure 4.12** shows a single, narrow DNA band for the linearized pD-npro-LapA plasmid. If SacI-HF digestion was inefficient, typically, the resulting DNA bands would become smeared, indicating supercoiled plasmid DNA. The pD-npro-LapA transformation protocol did not include screening for ‘Bonus’ transformants. Electroschocked X33 cells were dispersed on agar plates containing 100 µg/mL Zeocin. Approximately 70 colonies were observed, following a 3-day incubation at 28°C. Five large and five small colonies were picked for colony PCR to account for any phenotypic difference among *P.pastoris* transformants. PCR reactions were performed, according to section 2.13.1, using AOX1 primers: 5’ – GACTGGTTCCAATTGACAAGC – 3’ and 5’ – GCAAATGGCATTCTGACATCC – 3’.



**Figure 4.13** 1% DNA agarose gel analysis of PCR amplicons of pD-npro-LapA transformants. Ladder - 1 kb DNA Ladder (NEB, UK), Lanes 1 – 10 represent different

transformants. Positive (+, PCR using pD-npro-LapA as a template) and negative (-, PCR using empty X33 strain) controls are indicated. AOX1 (2.2 kb) and *wtproLapA* (1.2 kb) amplified loci are indicated by arrows.

It is clear that colony PCR produced many non-specific bands (below that indicated for *wtproLapA*; Fig 4.13). Transformants 6 and 7 gave single profiles of only a weak signal for the putative amplicon of *wtproLapA* (expected size 1267 bp). The negative control also has a band of equivalent size to the *wtproLapA* amplicon. Therefore, a false positive signal with other transformants clearly was possible. Nevertheless, clones 6, 7 and 9 were selected for further expression trials. Clones 6 and 7 exhibited the most clean amplification profiles. Clone 9 was randomly selected.

#### 4.3.1.2 Shake flask expression trial of *wtproLapA*

Positive transformants were subjected to a shake flask expression trial, according to section 2.13.3. In brief, 250 mL secondary cultures were initiated in 2L baffled flasks to ensure appropriate aeration. **Table 4.4** shows the cell density measured in cultures after 23h incubation and the volume of primary culture that was centrifuged to inoculate a secondary culture, equivalent to  $OD_{600} = 1$ . Importantly, the following morning primary cultures were found shaking at a temperature just under 36 °C, due to a fault with the incubator. Temperatures above 30 °C may induce cell death of *P. pastoris*. However, the experiment was continued, since it was not clear how the incubator failure to maintain the correct temperature (during the course of the experiment) could affect protein expression.

Flask	$OD_{600}$ 23h	Volume (mL) of culture added to BMMY (250mL)
Clone 6	20	12.5
Clone 7	27	9.26
Clone 9	29	8.62

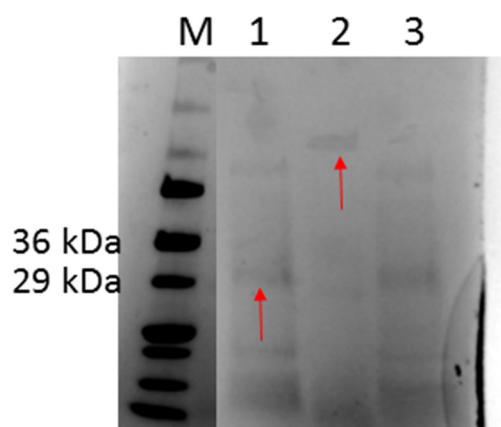
**Table 4.4** Cell density measured in the primary culture of different *wtproLapA* clones.

**Table 4.5** shows the cell densities observed in the secondary cultures of different *wtproLapA* clones within a 5-day continuous MeOH induction. Likely, due to previous overheating of the primary cultures, it was observed that growth in the secondary cultures was not steadily increasing (Clone 9) or the OD's were fluctuating (Clone 7).

Flask	OD <sub>600</sub> 24h	OD <sub>600</sub> 48h	OD <sub>600</sub> 72h	OD <sub>600</sub> 96h	OD <sub>600</sub> 120h
Clone 6	22	24	19	26	33
Clone 7	25	25	23	23	32
Clone 9	17	17	17	22	26

**Table 4.5** Cell density measured in the secondary culture of different *wtproLapA* clones.

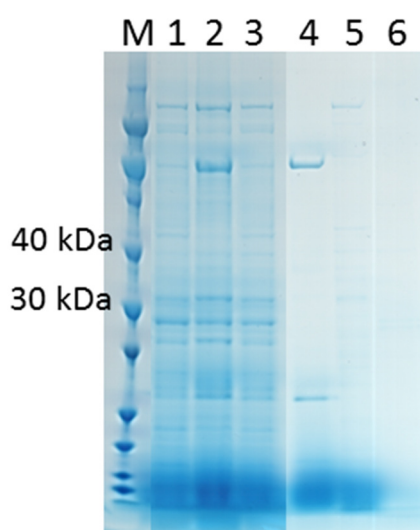
After 120h, *wtproLapA* supernatants were sampled and analysed by SDS-PAGE (**Figure 4.14**). However, no firm conclusions could be drawn regarding the expression of recombinant *wtproLapA*. Multiple protein bands of varying molecular weights were observed for clones 6, 7 and 9 (**Figure 4.14**). The predicted molecular weight of *wtproLapA* is 39.3 kDa, excluding glycosylation. Therefore, the observed band for each of clone 6 and 9, close to the 29 kDa molecular weight marker, was unlikely to represent a *wtproLapA* protein. No LapA specific activity was detected from clone 6, 7 or 9 supernatants, suggesting absence of the expressed target protein. Additionally, the presence of other lower and higher molecular weight bands could be an indication of ER-associated degradation (ERAD) or initiation of the unfolded protein response (UPR) in *P. pastoris* (Vanz et al., 2014). This could be a consequence of the over-heating in the primary culture step.



**Figure 4.14** SDS-page analysis of *wtproLapA* supernatants, following 120h induction. Lane 1 – Clone 6; Lane 2 – Clone 7; Lane 3 – Clone 9; red arrows indicate the lower and higher molecular weight proteins observed. Image was taken utilising gel stain-free technology by Bio-Rad.

#### 4.3.1.3 Purification of *wtproLapA* Clone 7

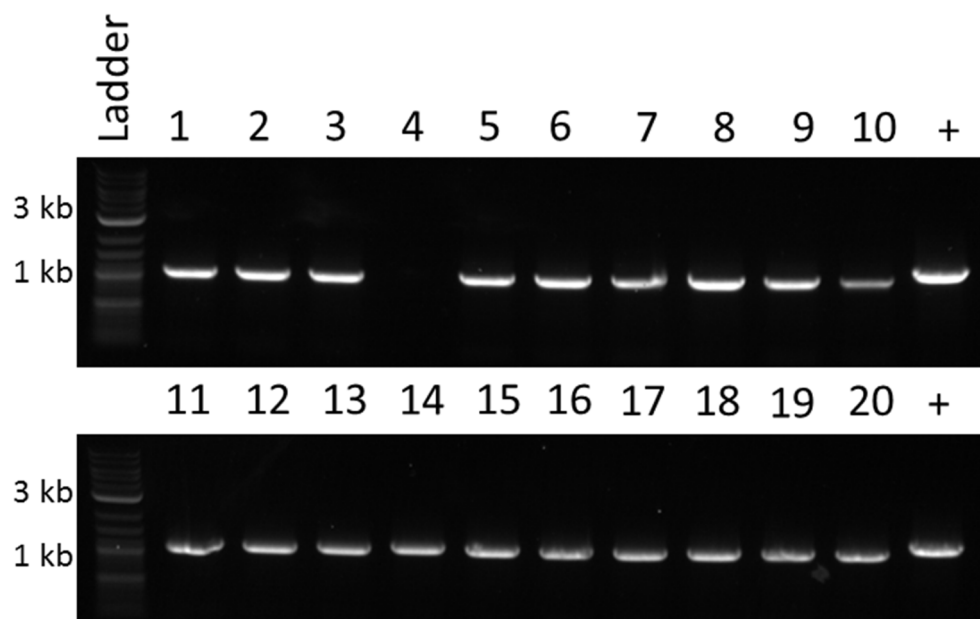
Despite the problems experienced, AEX purification was attempted, using clone 7 supernatant, in order to ensure no inactive LapA precursor or mature protein was expressed in this trial. Eluted fractions were subjected to LapA assay, but none of these displayed activity. Only accumulation of low molecular weight polypeptides was noticeable (at the bottom of the gel, Figure 4.15) suggesting severe protein degradation products. This provided further evidence to suggest that the ERAD response in *P. pastoris* had been triggered during cultivation of the primary culture.



**Figure 4.15** SDS-PAGE analysis of *wtproLapA* supernatants and subsequent AEX purification. Lanes 1, 2, and 3 correspond to sampled supernatants of clones 6, 7 and 9, respectively. Lanes 4, 5 and 6 correspond to eluted 0.1, 0.2 and 1M NaCl AEX fractions of clone 7 supernatant only.

#### 4.3.1.4 Generating a *P. pastoris* expression strain of *wtΔproLapA*

A second expression construct, containing a deletion of the LapA propeptide (pD-Δpro-LapA) was produced, utilising the QuickChange™ II site-directed mutagenesis kit (see section 2.13.1 for details). Deletion of the propeptide sequence was achieved by performing PCR with the following designed primers: 5'-GGACGCTATCTGGATAAGCTAATGCGCTTGCC-3' and 5'-GGCAAGCGCATTAGCTTATCCAGATAGCGTCC-3', using the pD-npro-LapA construct as a template. Following the Quickchange™ PCR reaction and subsequent transformation into *E.coli*, successive clones were screened by colony PCR, using the AOX1 primer set. Twenty single colonies (numbered 1-20) were randomly picked and patch plated on LB agar, containing 100 µg/mL Zeocin. With the exception of number 4, all of the pD-Δpro-LapA *E.coli* transformants carried the pD912 mutated expression plasmid (**Figure 4.16**).



**Figure 4.16** 1% DNA agarose gel of PCR amplicons of pD- $\Delta$ pro-LapA *E.coli* clones. Positive (+), PCR using pD- $\Delta$ pro-LapA as a template, control is indicated. The expected size of  $\Delta$ pro-LapA amplicon is 1084 bp.

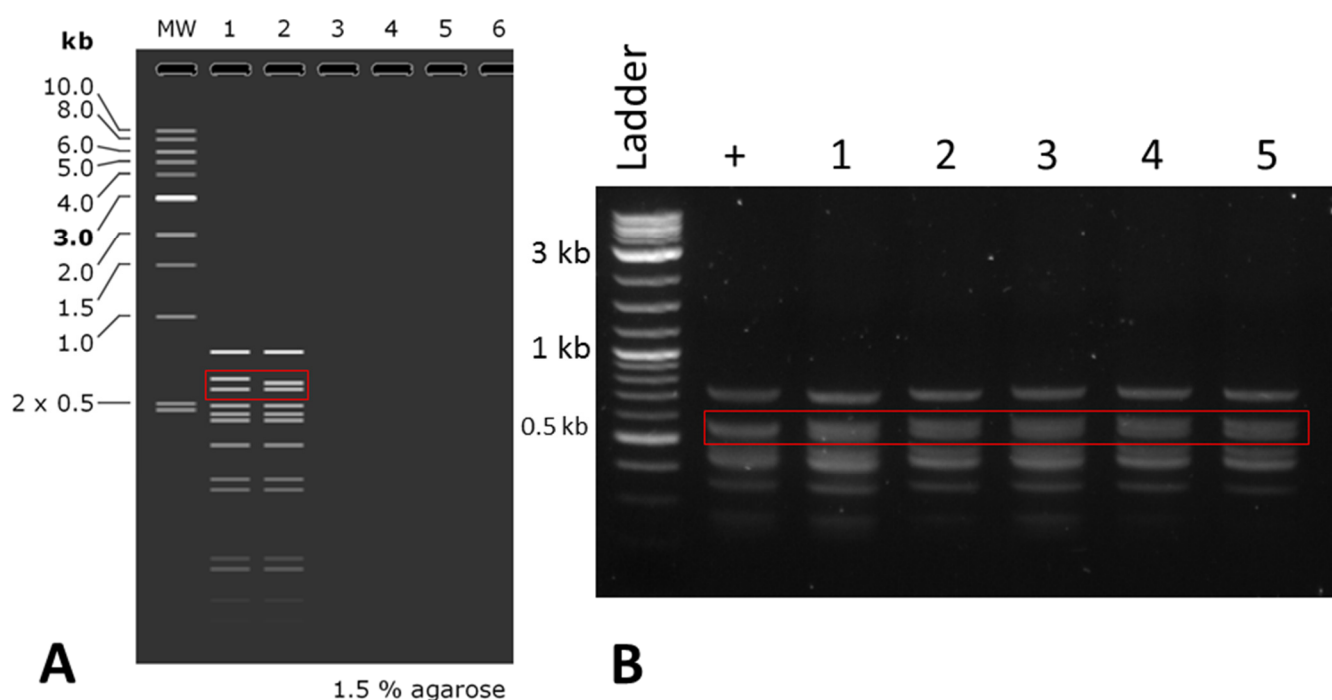
Transformant number 12 was randomly selected for large scale plasmid DNA preparation (see section 2.2.2 for details). Following MidiPrep plasmid DNA purification, 354  $\mu$ g of pD- $\Delta$ pro-LapA vector was obtained and 20  $\mu$ g was subsequently linearized with SacI-HF. As demonstrated in **Figure 4.18**, a successful DNA linearization quality check was required before transformation. Around 5  $\mu$ g of linearized DNA was transformed into a *P. pastoris* X33 strain, screening for multiple genome recombination events ('Bonus' transformants) on agar supplemented with 2 mg/mL Zeocin.

#### 4.3.1.5 Generating a *P. pastoris* expression strain of *wtproLapA**His*

A third expression construct, harbouring a C-terminal hexahistidine (His6) tag also was cloned using a QuickChange<sup>TM</sup> II site-directed mutagenesis kit and pD-npro-LapA plasmid as a template. The addition of a His6-tag in the pD-npro-LapA-His6 expression construct was implemented by performing PCR with the following designed primers: 5'-TTGCTGACTCTCTGAAACATCACCATCACCATCACTAAGGGGCGGCCGC-3' and 5'-GCGGCCGCCCCTTAGTGATGGTGGTGGTGGTGGTTCAGAGAGTCAGCAA-3'. A carboxypeptidase A recombinant His-tag cleavage site (Lys) also was engineered upstream of the histidine repeats.

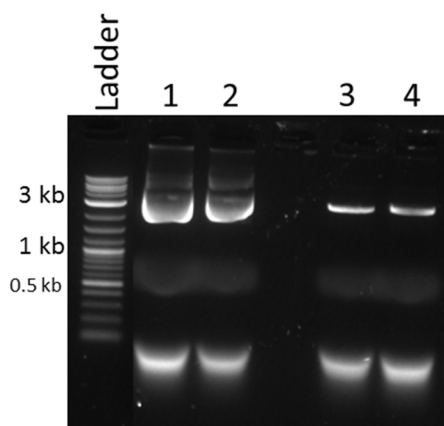
Following a Quickchange™ PCR reaction and subsequent transformation into *E.coli*, plasmid DNA of successive clones was extracted and screening was performed using HinfI digestion (see section 2.2.4 for details). PCR-based colony screening was not feasible, due to a modest difference between pD-npro-LapA and pD-npro-LapA-His6 amplicons. HinfI digestion of plasmid DNA should have revealed a small band difference of 628 bp and 604 bp between the template above and the mutant DNA, respectively. 300 ng each of control and 5 mutant clone plasmid DNA were treated with HinfI and loaded on a 1.5% agarose gel for analysis (**Figure 4.17**).

**Figure 4.17** shows that all five mutant clones displayed a smaller (604 bp) DNA band and, therefore, had the His6-tag successfully cloned. Clone number 1 was selected for large-scale plasmid DNA preparation, as described in section 2.2.2.



**Figure 4.17** 1.5% DNA agarose gel analysis of HinfI-digested pD-npro-LapA-His6 plasmid DNA clones. **A.** Simulation of the expected digestion pattern: lane 1 – band pattern in original vector, positive control (pD-npro-LapA), lane 2 – band pattern in the mutated pD-npro-LapA-His6 vector; **B.** Actual HinfI digestion patterns observed for positive (+) control and transformants 1-5. Red boxes indicate the region where band shift occurred.

Following MidiPrep plasmid DNA purification, 168.2 µg of the pD-npro-LapA-His6 vector was obtained and 20 µg was subsequently linearized with SacI-HF. Following verification of linearisation (**Figure 4.18**), around 6 µg of linearized DNA was transformed into a *P. pastoris* X33 strain, screening for multiple genome recombination events ('Bonus' transformants) on agar supplemented with 2 mg/mL Zeocin.



**Figure 4.18** 1% DNA agarose gel of SacI-HF linearized pD-Δpro-LapA and pD-npro-LapA-His6 plasmid DNA. Lanes 1 and 3 correspond to undigested and digested pD-Δpro-LapA, respectively; lanes 2 and 4 correspond to undigested and digested pD-npro-LapA-His6, respectively. 1.5 µg of DNA was loaded in each sample lane.

**Figure 4.18** shows single, narrow DNA bands for pD-Δpro-LapA and pD-npro-LapA-His6 linearized expression plasmid (lanes 3 and 4). As stated previously, if the SacI-HF digestion was inefficient, the DNA bands would be smeared, indicating super-coiled plasmid DNA. In this case, smearing of untreated plasmid DNA is clearly seen in lanes 1 and 2.

#### 4.3.1.6 Optimising *P. pastoris* colony PCR

Previous colony PCR for *wtproLapA* transformants was conducted in the laboratory at Biocatalysts Ltd. Following formation of *P. pastoris* colonies, strains pD-Δpro-LapA and pD-npro-LapA-His6 were brought back to Reading to resume with colony PCR clone identification, which appeared to be difficult to replicate. Due to unidentified reasons, colony PCR was not producing any visible DNA bands on the agarose gel, even when using the previously established protocol (section 2.13.1). It was assumed that the failure of the PCR reaction could arise from problems in either the DNA extraction step or differences in

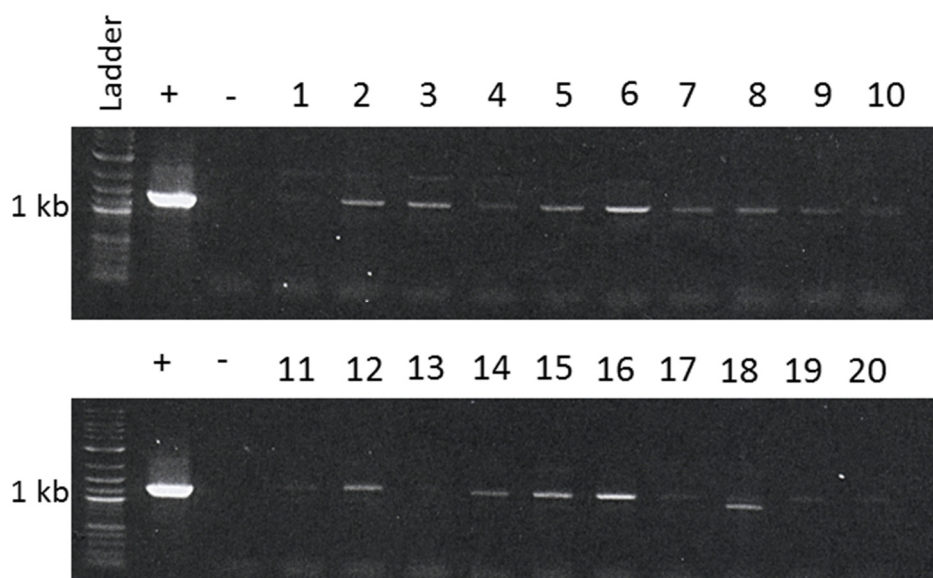
heating block temperatures of different PCR machines. Therefore, for the purposes of these and further experiments, requiring colony PCR for *wtproLapA* transformants, these were carried out on site at Biocatalysts Ltd (as described below), then transferred to Reading.

#### **4.3.2 *Ab initio* development of *wtproLapA* and *wtΔproLapA* *P. pastoris* expression strains**

Due to complications in the previous rounds of expressing recombinant LapA precursor and the challenges with colony PCR clone identification at Reading, it was decided to re-generate the *P. pastoris* expression strains for *wtproLapA* and *wtΔproLapA* proteins.

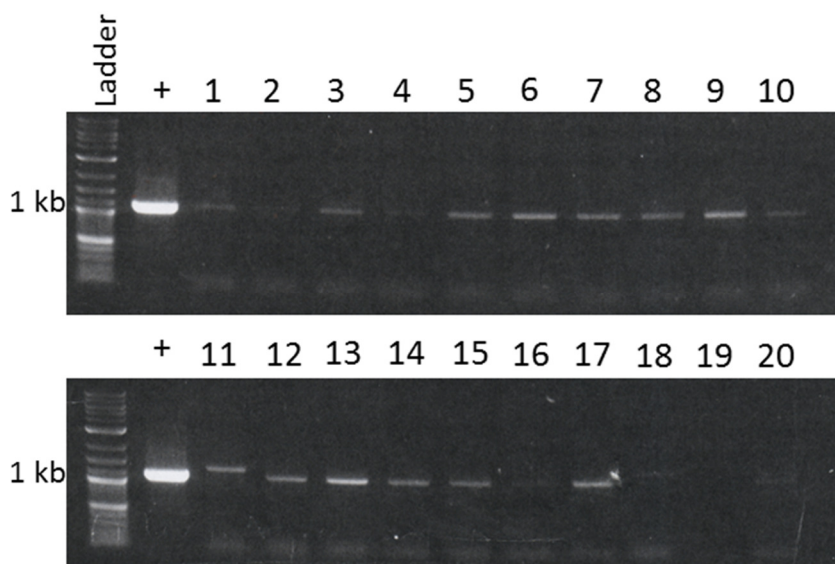
However, this time all the protocol steps outlined in **Table 2.4** were carried out at Biocatalysts Ltd. Screening for multiple genome recombination events ('Bonus' transformants), on agar supplemented with 2 mg/mL Zeocin, yielded around 30 single colonies for both *wtproLapA* and *wtΔproLapA* transformants. Ten large and ten small colonies were picked for colony PCR, from both strains, in the event that there was a phenotypic difference among *P. pastoris* transformants. PCR reactions were set up, according to section 2.13.1, using the following AOX1 primers: 5' – GACTGGTTCCAATTGACAAGC – 3' and 5' – GCAAATGGCATTCTGACATCC – 3'.

**Figure 4.19** shows that the colony PCR of pD-npro-LapA clones produced clean amplification signals just above the 1 kb reference mark (expected amplicon size 1267 bp). Clones 2, 3, 5, 6, 12, 15 and 16 were selected for shake flask expression trials. Since all randomly selected colonies were growing on extremely high concentration Zeocin plates, clones that produced no amplification product were likely to contain the recombinant LapA gene. However, the absence of a PCR signal was likely due to impurities in the reaction.



**Figure 4.19** 1% DNA agarose gel of PCR amplicons of pD-npro-LapA *P.pastoris* transformants. Positive (+, PCR using pD-npro-LapA as a template) and negative (-, PCR using empty X33 strain) controls are indicated.

Similarly, the colony PCR of pD- $\Delta$ pro-LapA clones produced clean amplification signals just above 1 kb reference mark (expected amplicon size 1084 bp) (**Figure 4.20**). Clones 5, 6, 7, 8, 9, 12 and 13 were selected for shake flask expression trial.



**Figure 4.20** 1% DNA agarose gel of PCR amplicons of pD- $\Delta$ pro-LapA *P. pastoris* transformants. Positive (+) control PCR using pD- $\Delta$ pro-LapA as a template is indicated.

#### 4.3.2.1 Shake flask expression trial of *wtproLapA* and *wtΔproLapA*

Selected transformants were subjected to shake flask expression trials, according to section 2.13.3. 250 mL secondary cultures were initiated in 2L baffled flasks to ensure appropriate aeration. **Table 4.6** and **Table 4.7** show the cell density measured in cultures after 23h incubation and the volume of primary cultures that were centrifuged to inoculate the secondary cultures, equivalent to OD<sub>600</sub>=1.

<i>wtproLapA</i>	OD <sub>600</sub> 23h	Volume (mL) of culture added to BMMY (100mL)
Clone 2	47	2.13
Clone 3	45	2.2
Clone 5	46	2.2
Clone 6	52	1.9
Clone 12	47	2.13
Clone 15	43	2.33
Clone 16	46	2.2

**Table 4.6** Cell density measured in the primary culture of different *wtproLapA* clones.

<i>wtΔproLapA</i>	OD <sub>600</sub> 23h	Volume (mL) of culture to BMMY (100mL)
Clone 5	47	2.2
Clone 6	43	2.2
Clone 7	40	2.5
Clone 8	32	3.2
Clone 9	48	2.1
Clone 12	41	2.5
Clone 13	45	2.2

**Table 4.7** Cell density measured in the primary culture of different *wtΔproLapA* clones.

**Table 4.8** shows the cell densities observed in the secondary cultures of different *wtproLapA* clones within a 5-day continuous MeOH induction. After each addition of MeOH, a steady increase in cell density was noticeable across all *wtproLapA* clones tested. Before harvesting supernatants at 120h, OD<sub>600</sub> reached an expected range between 35 – 50, indicating healthy *P. pastoris* cell growth.

<i>wtproLapA</i>	OD <sub>600</sub> 24h	OD <sub>600</sub> 48h	OD <sub>600</sub> 72h	OD <sub>600</sub> 96h	OD <sub>600</sub> 120h
Clone 2	12	15	21	33	39
Clone 3	14	21	25	35	46
Clone 5	11	17	20	34	37
Clone 6	12	15	22	30	35
Clone 12	15	20	22	36	46
Clone 15	13	18	23	41	49
Clone 16	10	16	20	28	33

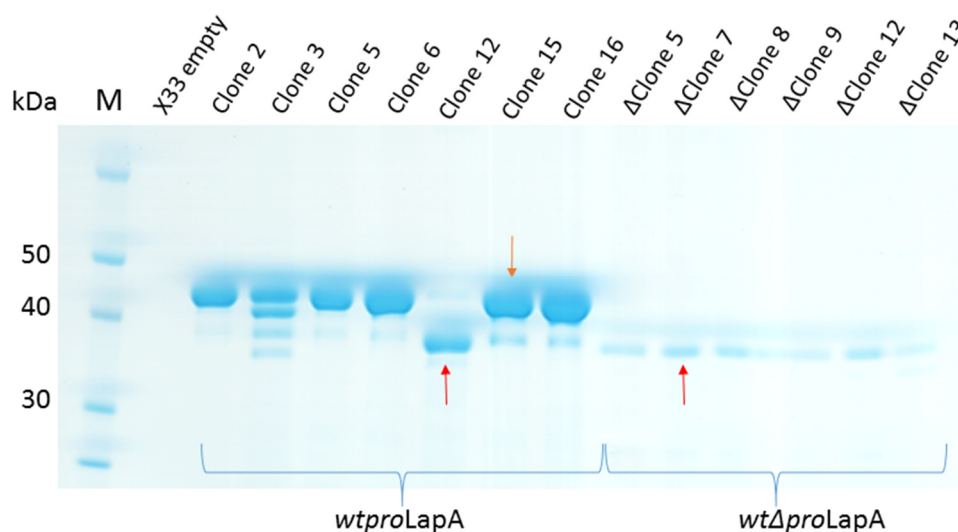
**Table 4.8** Cell density measured in the secondary culture of different *wtproLapA* clones.

**Table 4.9** shows cell densities observed in secondary cultures of different *wtΔproLapA* clones within a 5-day continuous MeOH induction. Similar healthy *P. pastoris* cell growth was observed in all *wtΔproLapA* clones tested.

<i>wtΔproLapA</i>	OD <sub>600</sub> 24h	OD <sub>600</sub> 48h	OD <sub>600</sub> 72h	OD <sub>600</sub> 96h	OD <sub>600</sub> 120h
Clone 5	13	16	21	31	35
Clone 6	15	19	27	37	40
Clone 7	16	19	25	31	36
Clone 8	16	18	25	32	37
Clone 9	15	18	26	32	37
Clone 12	17	21	34	40	42
Clone 13	12	19	25	38	45

**Table 4.9** Cell density measured in the secondary culture of different *wtΔproLapA* clones.

After 120h, *wtproLapA* and *wtΔproLapA* supernatants were sampled and analysed on SDS-PAGE (**Figure 4.21**).



**Figure 4.21** SDS-PAGE of *wtproLapA* and *wtΔproLapA* supernatants, following 120h MeOH induction. Lanes corresponding to different clone numbers are labelled. Expression control – *P. pastoris* X33 empty strain is also shown. Red and orange arrows indicate mature (~37 kDa) and precursor (~44 kDa) forms of LapA protein, respectively. 10  $\mu$ L of each supernatant was loaded onto the gel.

In general, the recombinant protein secreted by *P. pastoris* in the supernatant was very pure, with almost no contaminant proteins. The clonal variation was apparent in the supernatant profiles of the *wtproLapA* strain (**Figure 4.21**, Clones 3, 12). Most of the *wtproLapA* proteins were expressed as precursors, except Clone 12, where the bulk of recombinant protein appeared in the active form. Some pre-processed protein also was noticeable in other *wtproLapA* supernatants (

**Figure 4.21**, Clones 15 and 16). Conversely, no obvious clonal variation was present in the expression profiles of the *wtΔproLapA* strain. Since the *wtΔproLapA* expression construct contained a deletion of the propeptide domain, the recombinant protein was secreted in the mature form, across all clones. Based on the SDS-PAGE analysis, it was evident that the *wtΔproLapA* strain exhibited significantly lower secretion levels in *P. pastoris* supernatant (**Figure 4.21**). The secretion levels of the recombinant protein were also estimated by Bradford assay in each supernatant: clones of the *wtproLapA* strain showed protein content

in the range of 0.2 – 0.4 mg/ml, whereby equivalent values for *wtΔproLapA* clones varied between 0.05 - 0.07 mg/ml. The 4-fold reduction in secretion levels of *wtΔproLapA* highlight the likely requirement of the prodomain in LapA expression.

Specific activity levels in each crude supernatant were measured using a standard LapA assay, as described in section 2.10.1, and provided in **Table 4.9**.

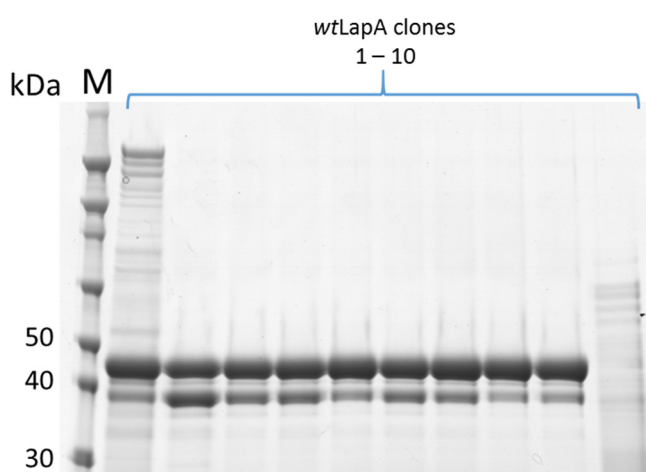
<i>wtproLapA</i>	Specific activity (U/mg)	<i>wtΔproLapA</i>	Specific activity (U/mg)
Clone 2	NA	Clone 5	NA
Clone 3	0.03	Clone 6	NA
Clone 5	NA	Clone 7	15.2
Clone 6	0.05	Clone 8	NA
Clone 12	2.1	Clone 9	0.42
Clone 15	0.03	Clone 12	0.3
Clone 16	0.8	Clone 13	NA

**Table 4.9** LapA specific activities measured in the crude *P. pastoris* supernatants. NA – no activity. Mostly active clones are highlighted in grey.

Since most of the *wtproLapA* clones were secreted as precursors, none or very low levels of activity were detected in the respective supernatants, suggesting that the LapA propeptide acted as an intra-molecular inhibitor. Only Clone 12 of *wtproLapA* showed significant levels of specific activity, owing to the activated form of the enzyme present in this supernatant. Interestingly, only Clone 7 of *wtΔproLapA* showed a substantial increase in activity. However, the resultant mature *wtΔproLapA* recombinant proteins were very sensitive to changes in temperature. Several repetitive cycles of incubation on ice lead to complete enzyme inactivation. A single freeze-thaw cycle with Clone 7 showed the same effect. No significant change in activity was observed for *wtproLapA* Clone 12 when the equivalent stability tests were performed. These observations might suggest a lack of stable protein fold for the *wtΔproLapA* secreted enzymes.

#### 4.3.2.2 Replicating expression trial for *wtLapA* Clone 12 deviant

To further investigate the observed completely mature secretion of *wtLapA* (an apparent outlier), it was decided to replicate the expression trial in *P. pastoris* from prepared glycerol stocks. Some *wtLapA* glycerol stock was streaked on YPD agar supplemented with 2000  $\mu$ g to ensure the growth of ‘Bonus’ transformants with multiple recombinant gene insertions. 10 individual clones were picked to inoculate primary cultures to account for potential clonal variation. Standard shake flask expression protocol was implemented (section 2.13.3). Following 120h, supernatants were sampled and analysed on SDS-PAGE (**Figure 4.22**).



**Figure 4.22** SDS-PAGE analysis of *wtLapA* Clones 1 – 10 supernatant, following 120h MeOH induction. Lanes correspond to clone numbers from 1 to 10.

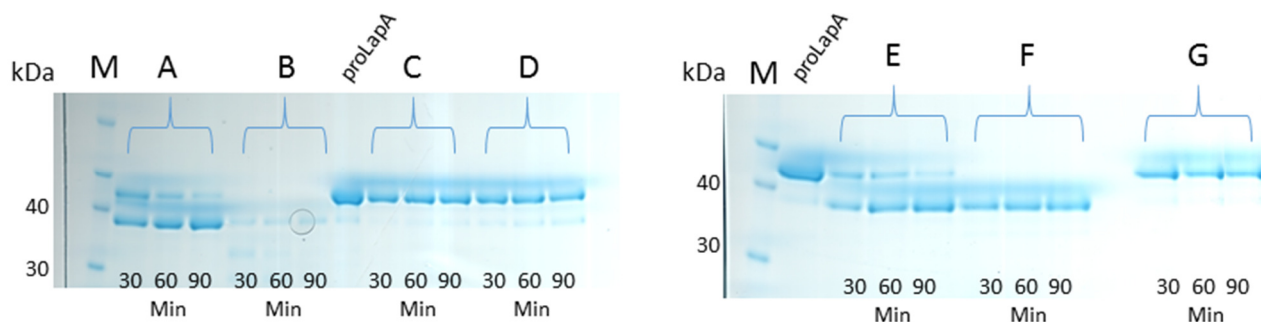
In this expression trial, of 10 individual *wtLapA* clones, the completely mature population could not be found in any of the supernatants analysed (**Figure 4.22**). Obvious clonal variation was present, where Clone 1 secreted additional high molecular weight proteins, while Clone 10 failed to secrete any significant amount of recombinant enzyme. Clone 2 displayed the highest proportion of mature LapA enzyme after 120h. It was concluded, that due to explicit clonal variation in recombinant LapA expression, formerly observed mature secretion *wtLapA* cannot be reliably replicated and, therefore, was accidental. Consequently, any successive recombinantly expressed LapA precursors would require enzymatic activation by exogenous proteases.

### 4.3.3 Downstream processing and analysis of expressed LapA proteins

#### 4.3.3.1 Enzymatic activation of *wtproLapA* precursor protein

To enable verification of secreted LapA specific activity, the LapA precursor required activation by enzymatic treatment with either commercially available non-specific proteases or one of the native LapA activators present in Flavourzyme. Previously isolated Flavourzyme fractions (section 3.2) were used to treat recombinant LapA precursor, in the crude supernatant. An underlying aim was to identify a protease in the commercial enzyme blend, which would naturally activate LapA precursor protein.

To activate the enzyme, 10 µg of Proteinase K and isolated Flavourzyme fractions were combined with 100 µg of target *wtproLapA* precursor (Clone2). The increase in LapA specific activity was monitored over the course of 90 minutes at 37 °C, also withdrawing samples, for subsequent SDS-PAGE analysis, every 30 minutes (**Figure 4.23**).



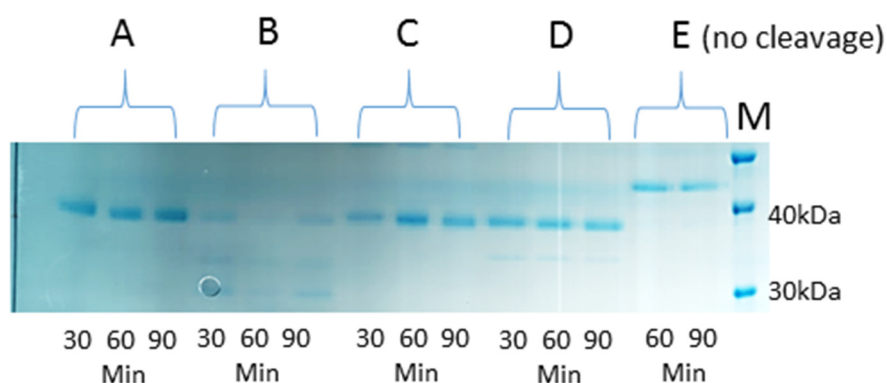
**Figure 4.23** Enzymatic activation of *wtproLapA* protein. A – addition of Flavourzyme; B – addition of Proteinase K; C – addition of Flavourzyme *fraction I*; D - addition of Flavourzyme *fraction II*; E – addition of Flavourzyme *fraction III*; F - addition of Flavourzyme *fraction IV*; G - addition of Flavourzyme *fraction IV*; proLapA – untreated LapA precursor in crude supernatant.

Enzymatic cleavage of LapA propeptide was most evident in samples A, B, and F (**Figure 4.23**). A band shift of around 7 kDa (predicted by ProtParam tool, ExPASy) is clearly observed. This experiment established that LapA activator most likely was found in isolated *fraction IV*, which contains AP1 and NP2 proteases, also as described previously (Merz et al., 2015). The action of Proteinase K caused severe proteolysis of recombinant LapA enzyme (sample B), whereby propeptide cleavage by non-specific AP1 and NP2 proteases left mature protein intact (sample F).

Apart from sample F, no significant increase in specific activity was noticeable following 90-minute treatment of LapA precursor (data not shown). Also, it was noticed that baseline LapA activity arising from native LapA proteases (Lap2 present in *fraction II*, *flavLapA* present in *fraction III*) was dramatically reduced, when Flavourzyme fractions were combined with *P. pastoris* crude supernatant (NOTE: as these were crude supernatants, activity data are not presented herein, see **Table 4.10** for specific activity from corresponding purified proteins). This suggested that purification of LapA precursor was necessary to re-evaluate enzymatic activation, since impurities in *P. pastoris* compromised the activity of both recombinant activated LapA and native Lap enzymes in Flavourzyme.

#### 4.3.3.2 Purification and further activation of *wtproLapA* precursor protein

LapA precursor was purified by AEX chromatography (see section 2.13.4 for details) from Clone 2 *wtproLapA* crude supernatant. Enzymatic activation was replicated using crude Flavourzyme solution, Proteinase K, and purified Flavourzyme fractions (III, IV, and V), as before. This time removal of LapA propeptide was immediately visible even after 30 minutes incubation at 37 °C (**Figure 4.24**). Proteinase K still degraded most of the recombinant LapA precursor (Fig. 4.24, sample B). Efficient propeptide cleavage with crude Flavourzyme and *fraction III* suggest only trace amounts of AP1 or NP2 were required for activation. Isolated *fraction III* contains 100-fold less proteases.



**Figure 4.24** Enzymatic activation of *wtproLapA* protein. A – addition of Flavourzyme; B – addition of Proteinase K; C – addition of Flavourzyme *fraction III*; D - addition of Flavourzyme *fraction IV*; E – addition of Flavourzyme *fraction V*.

**Table 4.10** illustrates that purified LapA precursor possessed low activity of 1.5 U/mg compared to Clone 2 crude supernatant with no activity, suggesting the presence of a small

activated protein population in the non-treated sample, which exhibits activity upon removal of bulk contaminants. In fact, this can be seen in

**Figure 4.23** as a very weak band below inactive LapA precursor. The final observed activities following enzymatic treatment corresponded to 6, 7-fold activation of treated LapA enzyme in relation to inactive precursor (**Table 4.10**, last column).

	Samples	Absorbance rate			Background absorbance after 90 min	Specific activity after 90min*	Increase in activity (-fold)
		30 min	60 min	90 min			
	Untreated LapA precursor	0.135	0.167	0.149	0.149	1.5 ± 0.1 U/mg	-
<b>A</b>	Flavourzyme	0.72	0.87	0.95	0.05	9.1 ± 0.2 U/mg	6.4
<b>B</b>	Proteinase K	0.78	0.85	0.95	No activity	9.6 ± 0.5 U/mg	6.4
<b>C</b>	Fraction III	1.28	1.14	1.32	0.22	11.1 ± 0.3 U/mg	7.4
<b>D</b>	Fraction IV	0.95	0.91	0.85	0.07	7.9 ± 0.4 U/mg	5.3
<b>E</b>	Fraction V	0.208	0.22	0.305	No activity	3.1 ± 0.2 U/mg	1.58

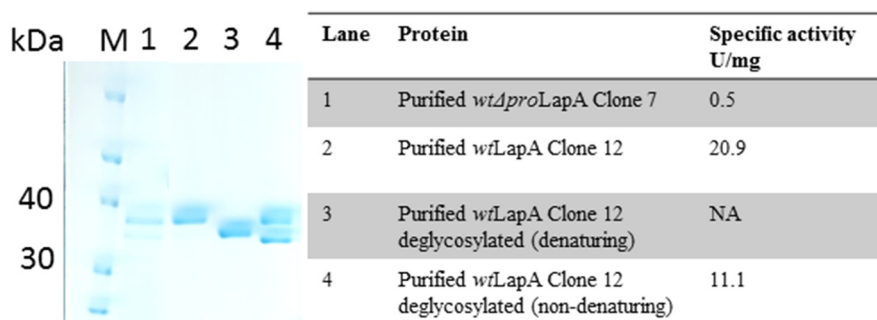
\*Background - adjusted

**Table 4.10** LapA precursor activation data. Background absorbance signal from each fraction was subtracted from the final measured rate after 90 minutes and specific activity was calculated from the residual absorbance rate.

#### 4.3.3.3 Purification and deglycosylation of activated LapA proteins

Chromatographic AEX purification on 5 mL Q HP column was performed using the endogenously activated Clone 12 deviant (*wtLapA*) protein and maturely expressed Clone 7 *wtAproLapA* protein (**Figure 4.21**). Mostly active protein eluted in the range of 140 - 200 mM NaCl for both proteins, thus, higher salt elution fractions were discarded.

Following AEX purification, *wtLapA* protein was treated with PNGase to test if recombinant LapA enzyme expressed in *P. pastoris* was glycosylated and to estimate the effect of N-linked glycan removal on the activity (**Figure 4.25**).



**Figure 4.25** AEX purification and deglycosylation of activated *wtΔproLapA* and *wtLapA* proteins. SDS-PAGE analysis (left); Corresponding lanes with the respective measured specific activities (right); NA – no activity.

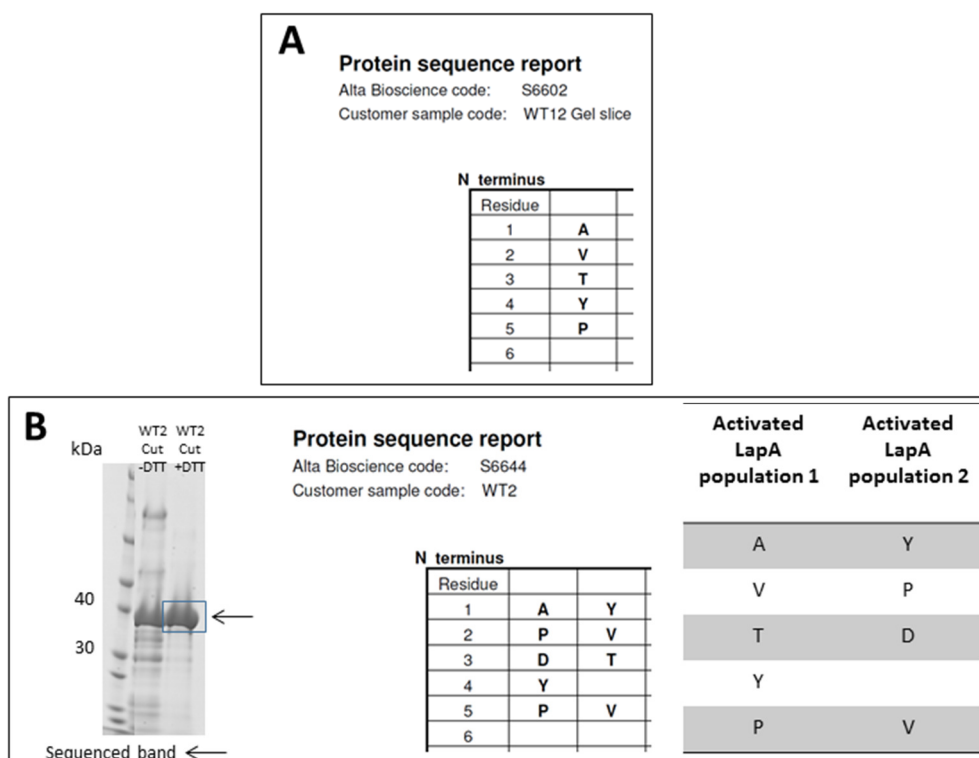
In general, the observed specific activities for recombinant purified *wtΔproLapA* and *wtLapA* proteins were quite low ( $0 - 20.9 \pm 0.5$  U/mg) compared to the previously reported specific activity for native LapA of  $142.0 \pm 0.2$  U/mg (Matsushita-Morita et al., 2011). Also, purified native *flavLapA* (from Flavourzyme) and refolded (*refLapA*), expressed in *E.coli*, displayed activities of  $94.6 \pm 2.2$  and  $59.5 \pm 1.1$  U/mg, respectively. A negligible amount of activity from *wtΔproLapA* enzyme was likely due to misfolded and/or partially folded enzyme. Furthermore, purified *wtLapA* still had a considerable amount of brown pigment bound. Therefore, it was speculated that this could cause partial enzyme inhibition. Similar pigment – induced inhibition of activity was already observed in the purification process of *flavLapA* (section 3.2). Additional AEX purification did not eliminate all of the associated pigment from *wtLapA*. It was decided to explore alternative ways of purifying activated *wtLapA* from *P. pastoris* supernatant (section 4.3.4).

Recombinant *wtLapA* protein appeared to be N-glycosylated. The estimated size of N-glycan could be in the range of 3 – 5 kDa, based on the band shift observed on SDS-PAGE. This is potentially distributed across two predicted asparagine glycosites (N87 and N288). The effect of N-linked glycan removal on *wtLapA* specific activity was quite detrimental; 50% loss in activity was observed following 24h incubation with PNGase. It is important to note that deglycosylation effect was only partial, under non-denaturing conditions, likely due to limited glycosite accessibility in fully folded *wtLapA* protein (**Figure 4.25**, lane 4). Further investigation was required to understand the importance of N-linked glycosylation for specific activity and for potential thermostability of recombinant LapA.

#### 4.3.3.4 N – terminal sequencing of activated LapA proteins

To gain a better mechanistic understanding of LapA precursor processing in *P. pastoris*, N – terminal sequencing was performed on different enzymatically processed LapA variants. Purified activated *wtLapA* Clone 12 seemed to possess a fully processed N – terminus. Additionally, purified *wtproLapA* Clone 2 was subjected to treatment with a small amount of Flavourzyme. The anticipated N-terminus, resulting from the action of AP1 and NP2 proteases, should have been equivalent to mature native LapA from Flavourzyme (*flavLapA*).

Unexpectedly, the processed N – terminus of *wtLapA* Clone 12 protein displayed an 'AVT' extension (Figure 4.26) compared to the original extremity reported by Morita et. al (Matsushita-Morita et al., 2011). This was later confirmed by the structural solution of *wtLapA* Clone 12 enzyme. Even more interesting, the N – terminus presented by *wtproLapA* Clone 2 protein, following enzymatic processing by Flavourzyme, showed two distinct populations of protein were present in a single sequenced protein band. The actual N-terminus could only be resolved by having prior knowledge of the exact protein sequence (**Figure 4.26**, B(right)). It is possible that two different enzymes, with diverse specificity, are responsible for the appropriate processing of LapA propeptide in *Aspergillus oryzae*. Alternatively, insufficient incubation time with Flavourzyme yielded not fully processed LapA (population 1).



**Figure 4.26** N – terminal sequencing results for the different activated forms of *wtproLapA* protein. **A.** First five residues at the N – terminus of *wtLapA* Clone 12; **B.** SDS – PAGE of Flavourzyme - activated *wtproLapA* Clone 2 protein (left). Original N – terminal amino acid sequence (middle). Reordered amino acids using knowledge of actual polypeptide sequence (right). The sequencing was performed by Alta Biosciences, UK.

Clone 12 was an outlier in the expression trial of LapA precursor protein (**Figure 4.21**), as it was present as mature enzyme. This suggests that it was most likely processed by a co-secreted exogenous proteolytic enzyme from *P. pastoris*. Interestingly, the specificity of this unidentified enzyme could be similar to the one present in Flavourzyme, since the N-terminus of Flavourzyme – activated Clone 2 was identical to that of the Clone 12 outlier, possessing an additional 3 – residue 'AVT' extension to the original LapA N – terminus (**Figure 4.26, B**). The observed N – terminus, in a fully refined structural solution of *flavLapA*, showed tyrosine (Y) in position 1.

#### 4.3.3.5 Genomic sequencing of *wtLapA* Clone 12 deviant

It was important to verify that the explicitly different secretion profile of Clone 12 was not a result of genetic mutation within the AOX locus, in which the *wtLapA* expression cassette was integrated. For that purpose *P. pastoris* colony PCR was performed to amplify the target integrated *wtLapA* portion of the expression cassette, followed by sequencing of the resultant amplicon and matching against the expected sequence (**Figure 4.27**). The primers used in the amplification: 5' – GAGGCAGAGAAATGGGCACT – 3' and 5' – TCATCTGCTGCGAGATAGGC – 3'. See section 2.2.1 for further details of the method.



**Figure 4.27** Pairwise sequence alignment of expected (top) and observed *wtLapA* (bottom) sequences. Translated propeptide sequence is overlaid with the DNA sequence: LapA propeptide and mature domain boundary is highlighted in purple and indicated by outward facing arrows.

The Sanger-sequenced *wtLapA* amplicon (section 2.12.2) was found to match the expected sequence, containing LapA propeptide. Hence, no deletion was seen in the propeptide region of integrated pD-npro-LapA expression cassette in *P. pastoris* AOX locus. Therefore, as speculated, maturation of *wtLapA* Clone 12 must be a post-translational event, most likely mediated by co-secreted endogenous protease activity.

#### 4.3.4 Optimization of purification protocol of activated *wtLapA* from *P. pastoris* supernatant

Since previous attempts of AEX chromatography alone did not fully separate activated LapA from tightly bound brown pigment (section 4.3.3.3), alternative methods for isolation of purified protein were required. Two mysterious brown pigments ( $M_r < 2\text{kDa}$ ) were previously reported to associate with other Lap peptidases and were briefly characterized by EPR to be spectroscopically active and contain Fe(III) ligand (Schalk et al., 1992, Hartley et al., 2009, Aphale and Strohl, 1993, Bennett and Holz, 1997).

Ammonium sulfate precipitation seemed an attractive way of cleaning up recombinant protein from the bulk of *P. pastoris* supernatant. Also, it served as a quick concentration method of the target protein from a large volume of supernatant. Based on previous knowledge regarding the properties of *P. pastoris* growth media and partial characterization of brown pigment, selective gel filtration seemed appropriate to separate the target protein from low molecular weight peptides and pigments present in the growth media.

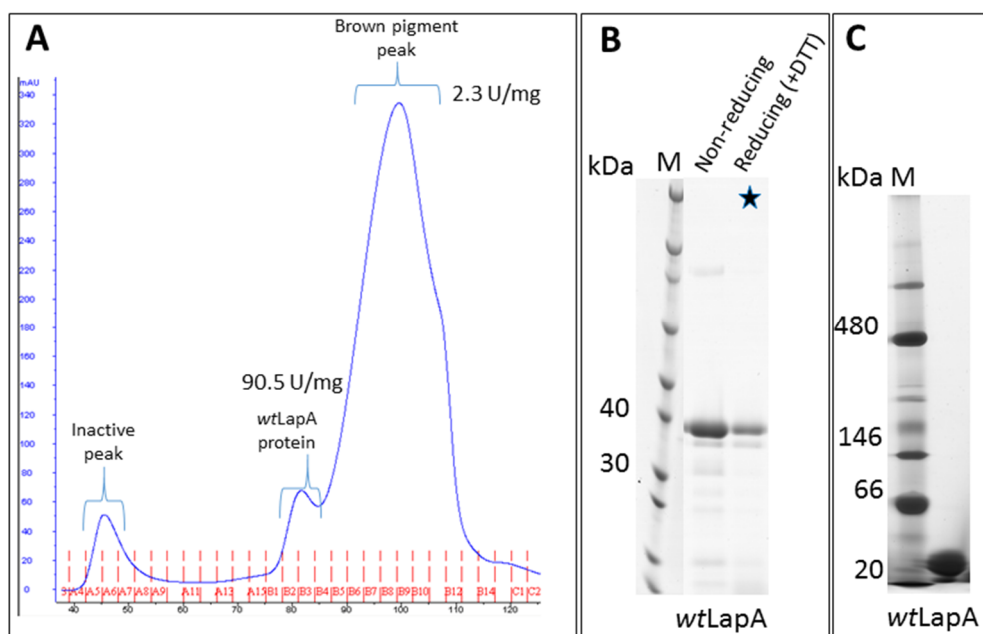
Clone 12 *wtLapA* supernatant was saturated to 90% with  $\text{AmSO}_4$  and left to equilibrate at 4 °C for 24h. The resultant protein pellet was harvested by high speed ( $30,000 \times g$ ) centrifugation and subsequently re-suspended in gel filtration buffer. 0.2  $\mu\text{m}$  filtered protein suspension was loaded onto a SEC column (as described in section 2.13.4).

**Figure 4.28** shows three distinct peaks. The nature of the first peak could not be established, since it exhibited no LapA activity and no observable band on SDS-PAGE (not shown). A broad peak from 90 mL to 110 mL showed extremely high UV absorption, however, it too showed no particular band on the SDS-PAGE (not shown). Once eluted, this fraction appeared intense brown in color and, therefore, possibly contained most of the pigments present in the *P. pastoris* supernatant. It showed a tiny amount of specific activity ( $2.3 \pm 0.5$  U/mg). However, this was possibly due to residual carry over from an overlapping active *wtLapA* peak at 80 mL. The characteristic high absorption of these associating ‘brown pigments’ at 280 nm hindered LapA protein quantification using purification chromatogram and spectrophotometric instruments, superseding the absorption arising from aromatic amino acids in the protein chain.

The eluted *wtLapA* peak at around 80 mL was 4.3-fold more active than AEX-alone formerly purified batch, with  $90.5 \pm 0.3$  U/mg specific activity after SEC, compared to  $20.9 \pm 0.5$

U/mg after AEX-alone. This was a significant improvement in specific activity. However, the activity level still was comparable to purified *flav*LapA ( $94.6 \pm 2.2$  U/mg). By concentrating the 80 mL *wt*LapA fraction, using an ultracentrifugation device, a light brown protein layer was visible on the membrane, suggesting some fraction of pigment was still bound to *wt*LapA, thus likely preventing the release of full activity potential.

The most highly active *wt*LapA peak from SEC could be conveniently loaded directly onto an AEX Q HP column, without the need of an intermediate buffer exchange step. The peak eluted at 140 mM NaCl, during this purification step, contained no traces of brown pigment, which appeared to remain strongly bound to the AEX resin and could only be removed completely by performing a 2 M NaCl wash. This final pigment-free *wt*LapA fraction displayed maximal specific activity of  $140.15 \pm 4.8$  U/mg, which was higher to that observed for *flav*LapA. Also, complete elimination of brown pigments, in the final AEX purification step, ensured an homogeneous *wt*LapA protein population for subsequent crystallization trials (**Figure 4.28**, B). Interestingly, a higher molecular weight association was noticeable in the absence of reducing agent in protein sample buffer, probably mediated by covalent disulphide bonds between SDS-unfolded *wt*LapA proteins (non-reducing lane, **Figure 4.28**, B). Also, the absence of possible *wt*LapA higher oligomeric states was confirmed by Native-PAGE analysis (**Figure 4.28**, C), which further verified *wt*LapA sample homogeneity, a crucial feature for obtaining single crystals for structural studies.



**Figure 4.28** *wtLapA* Clone 12 purification and analysis. **A.** Elution profile from Superdex 75 SEC matrix. Peaks and their corresponding specific activities are indicated; **B.** SDS-PAGE analysis of AEX purified, pigment-free *wtLapA* protein. The star indicates a fraction used in crystallization trial; **C.** Native-PAGE analysis of AEX purified, pigment-free *wtLapA* protein.

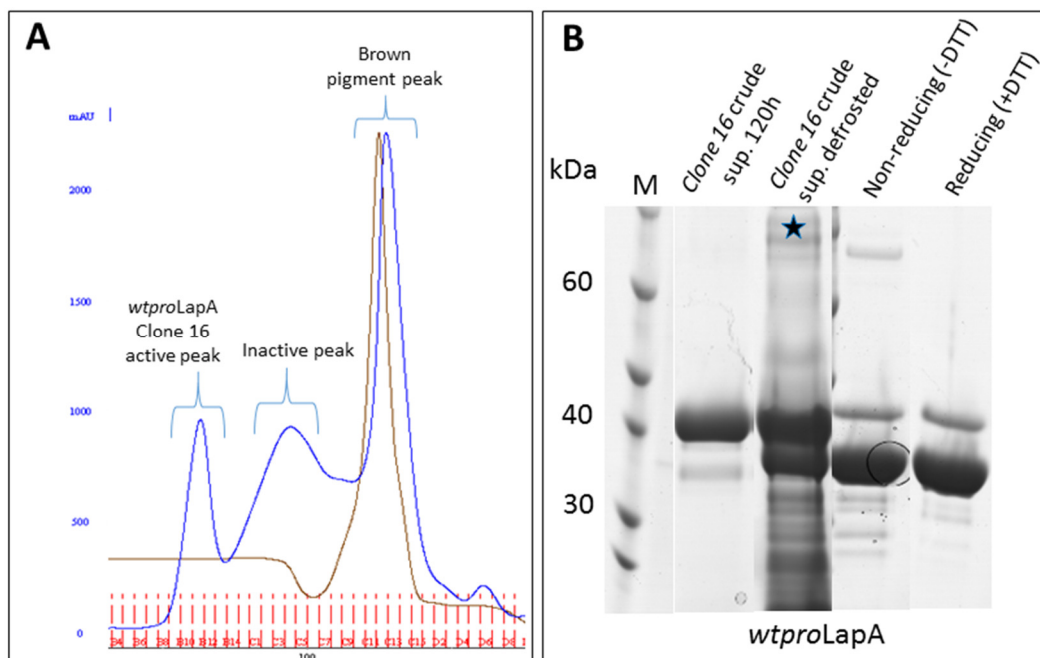
#### 4.3.5 Purification of *wtproLapA* precursor from *P. pastoris* supernatant for crystallization

Similar to the anticipated crystal structure of activated mature *wtLapA*, it was decided to attempt the structural solution of the LapA precursor enzyme, which would help elucidate the structural and possibly functional role of the propeptide domain. Since recombinant LapA was secreted in the form of proenzyme (*wtproLapA*), unprocessed enzyme population was readily available for purification. An equivalent purification procedure to that of the mature *wtLapA* protein was assumed and followed (section 4.3.4).

##### 4.3.5.1 Purification of *wtproLapA* Clone 16

Clone 16 *wtproLapA* supernatant was saturated to 90% with  $\text{AmSO}_4$  and left to equilibrate at 4 °C for 24h. The resultant protein pellet was harvested by high speed (30,000 x *g*) centrifugation and subsequently re-suspended in gel filtration buffer, as previously described (see section 2.13.4). 0.2  $\mu\text{m}$  filtered protein suspension was loaded on SEC column. **Figure 4.29** shows three distinct peaks eluted from the SEC column. The last two broad peaks

showed no LapA activity and eluted as very pigmented fractions. Interestingly, the observed dynamics of the conductivity curve followed the pattern of the last UV absorption peak. This might be a result of eluting  $\text{AmSO}_4$  from the LapA sample loaded or, more likely, to the brown pigment  $\text{Fe(III)}$  moieties giving rise to a striking increase in conductivity.



**Figure 4.29** *wtproLapA* Clone 16 SEC purification and SDS-PAGE. **A.** Elution profile from Superdex 75 SEC matrix; trace in brown corresponds to changes in conductivity; **B.** SDS-PAGE analysis of SEC purification. Lane 2 – crude supernatant of Clone 16 LapA precursor after 120 h harvesting. Lane 3 (starred) – the same supernatant, following a single freeze-thaw cycle. Lanes 4 and 5 - eluted active 80 mL peak.

The most active peak eluted at around 80 mL, similarly to the elution pattern of mature *wtLapA* (see **Figure 4.28**). Therefore, the additional 7 kDa propeptide domain did not significantly affect the hydrodynamic volume of the *wtproLapA* protein and, consequently, its retention time on the gel filtration matrix. The specific activity of the 80 mL peak was surprisingly high ( $75 \pm 0.3$  U/mg), since this should have been an inactive precursor. Subsequent analysis on SDS-PAGE revealed induced maturation of *wtproLapA* Clone 16, which can be noticed from a clear molecular weight band shift (**Figure 4.29**, last three fractions). While the observed high specific activity in the previous SEC fraction (corresponding to mature protein) could easily be explained, the phenomenon of this *wtproLapA* autoprocessing on the gel filtration column could not be explained. It appears

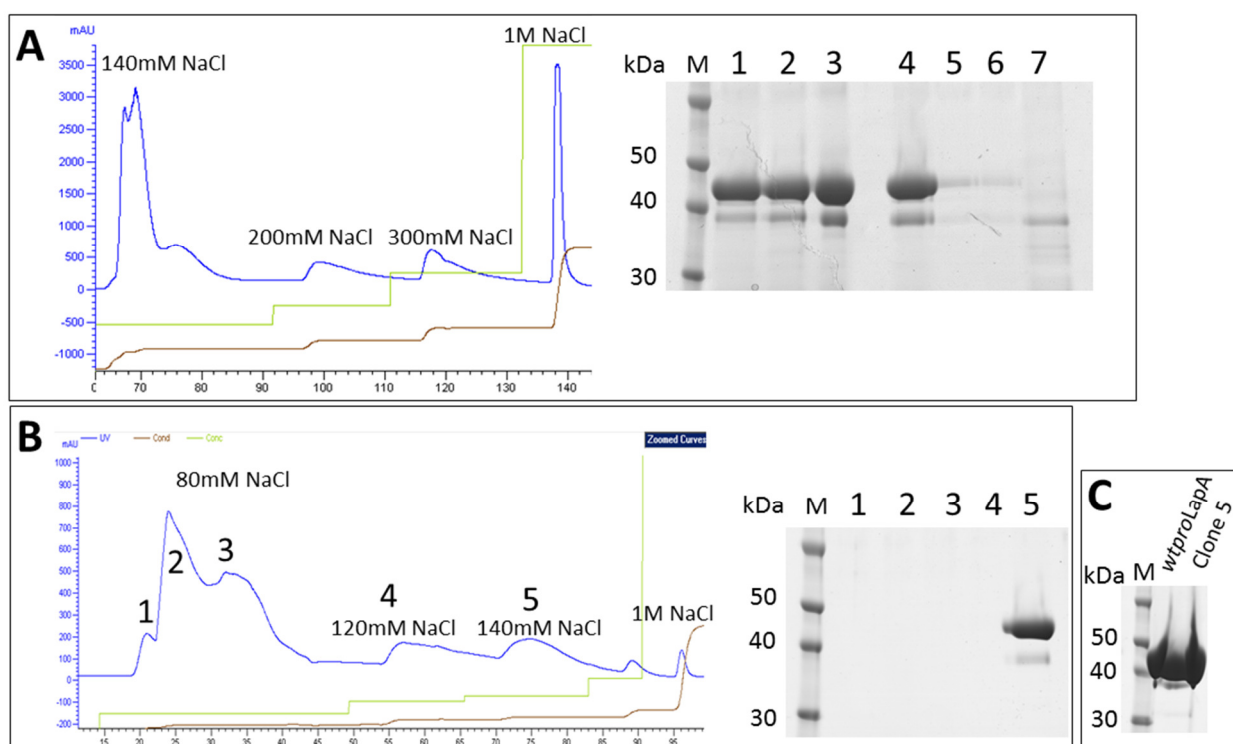
that this batch of LapA precursor was progressively processed to mature protein by either an autocatalytic mechanism or another co-purified protease in the *P. pastoris* supernatant. However, Clone 16 LapA precursor was not fully processed in either the SEC-eluted fraction or the following AEX purification step. A small portion of this enzyme still contained covalently linked propeptide domain (**Figure 4.29**, lane 5).

Since this purified LapA batch was of heterogeneous composition, with the bulk of the population consisting of mature processed protein, it was unlikely to yield crystals of LapA precursor. Therefore, it was decided to replicate the same purification protocol with a different *wtproLapA* batch (Clone 5).

#### **4.3.5.2 Purification of *wtproLapA* Clone 5**

Since the previous purification strategy gave rise to the unexplained maturation of LapA precursor, it was decided to alter the purification protocol yet again. Instead of the concentration of protein by AmSO<sub>4</sub> precipitation, followed by gel filtration, it was decided to purify *wtproLapA* Clone 5 solely by anion exchange chromatography. Also, this clone was selected due to the presence of a negligible amount of matured enzyme compared to other *wtproLapA* supernatants (**Figure 4.21**).

This time, *wtproLapA* Clone 5 supernatant was 4-fold concentrated by ultracentrifugation and subsequently dialyzed into AEX binding buffer overnight. Purification on a 5 mL Q HP column yielded 80 mM, 140 mM, 200 mM, 300 mM and 1 M NaCl peaks (**Figure 4.30**, A). Interestingly, the bulk of LapA precursor still eluted in the 140 mM peak, again suggesting the presence of prodomain did not affect binding on the positively charged AEX resin. Only the 140 mM peak was collected and repetitively applied onto an AEX column to remove traces of bound brown pigments. Following 3 rounds of elution at 140 mM NaCl from the AEX column, all of the associating pigments were completely eliminated and the resultant protein pool was SEC purified for final polishing and salt removal. The latter purification step yielded a single, symmetrical sharp *wtproLapA* peak at around 76 mL (**Figure 4.30**, C). This observation was in conjunction with an anticipated lower elution volume for LapA precursor.



**Figure 4.30** *wtproLapA* Clone 5 purification and SDS-PAGE analysis. **A.** Eluted fractions from 5 mL Q HP column and SDS-PAGE analysis: 1 – defrosted supernatant, 2 – supernatant after storing at RT overnight, 3 – supernatant after dialysis, 4 – 140mM NaCl peak, 5 – 200mM NaCl peak, 6 – 300mM NaCl peak, 7 – 1M NaCl peak; **B** 2<sup>nd</sup> AEX run: eluted fractions from 5 mL Q HP column and SDS-PAGE analysis: gel lanes agree with peaks eluted; **C** Final LapA precursor pool eluted at 76 mL from SEC.

**Figure 4.30 A** demonstrates that this time the majority of the recombinant LapA protein was in the form of the precursor in the defrosted supernatant. The zymogen/mature protein population ratio stayed the same throughout the purification process of *wtproLapA* Clone 5, and no signs of progressive propeptide cleavage were noticeable. In the 2<sup>nd</sup> round of AEX purification, an attempt to separate LapA zymogen from mature protein population was made. The NaCl concentration in the elution buffer was steadily increased in 1% increments. A triple peak at 80 mM NaCl was eluted, which exhibited no protein band on the SDS-PAGE (**Figure 4.30, B**: lanes 1, 2, 3). A mixture of two LapA populations still eluted at 140 mM NaCl concentration. The final SEC-eluted *wtproLapA* Clone 5 fraction (**Figure 4.30, C**) was concentrated to 9.4 mg/ml and subjected to subsequent crystallization trials (see section 2.8.2.2).

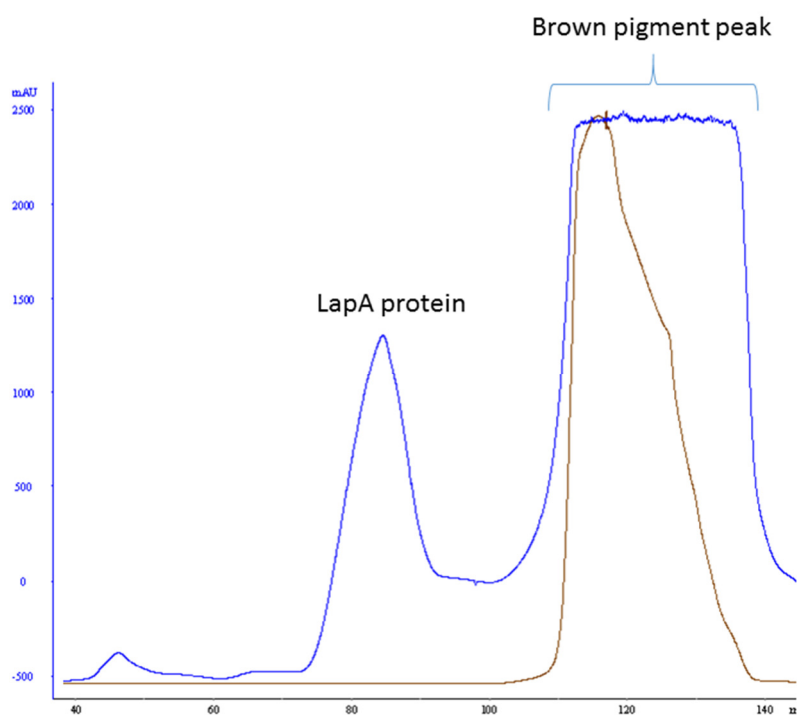
#### 4.3.6 Considerations of LapA protein purification from *P. pastoris* supernatant

LapA recombinant wild-type proteins were purified using the protocol described in section 2.13.4. However, later in the project, the standard low-resolution 5 mL Q HP AEX column was replaced with a MonoQ 4.6/100 PE high-resolution column. The latter allowed running the purification under high pressure (3.0 – 4.0 bar) due to the smaller particle beads, hence improving peak sharpness and separation. A separation of LapA precursor, from the majority of mature LapA population, was enabled on the MonoQ column, which aided the subsequent crystallisation process due to improved sample homogeneity. LapA precursor was successfully isolated from partially matured protein by elution at 80 mM NaCl concentration on MonoQ, whereas, typically, a mixture of mature and proenzyme LapA populations would only elute at 140 mM NaCl concentration on Q HP (**Figure 4.35**).

Consequently, a two-step purification protocol was almost sufficient to eliminate the majority of brown pigment from LapA sample. Firstly, recombinant LapA pellet from the AmSO<sub>4</sub> precipitation step was resuspended and loaded onto the SEC column in order to separate the protein of interest from the bulk of low molecular weight contaminants and brown pigment arising from the supernatant media (**Figure 4.31**). The second step involved loading the pooled active LapA fractions from the previous SEC onto an AEX column (Q HP/MonoQ). The eluted LapA precursor peak (140 mM NaCl) from this step would be almost pigment-free (**Figure 4.32** and **Figure 4.33**). LapA precursor was then treated with AP1 protease and the resultant protein mixture was loaded onto the MonoQ column again, to separate an active mature LapA protein population. Any residual pigment was completely eliminated in this final purification step. Residual brown pigment was found to have an adverse effect on LapA activity and interfere with spectrophotometric protein quantification. Therefore, removal of any remaining pigment was critical, to prevent experimental bias across LapA mutant samples. In this respect, the MonoQ column was superior to Q HP, as it was more discriminating and permitted improved separation of protein from pigment.

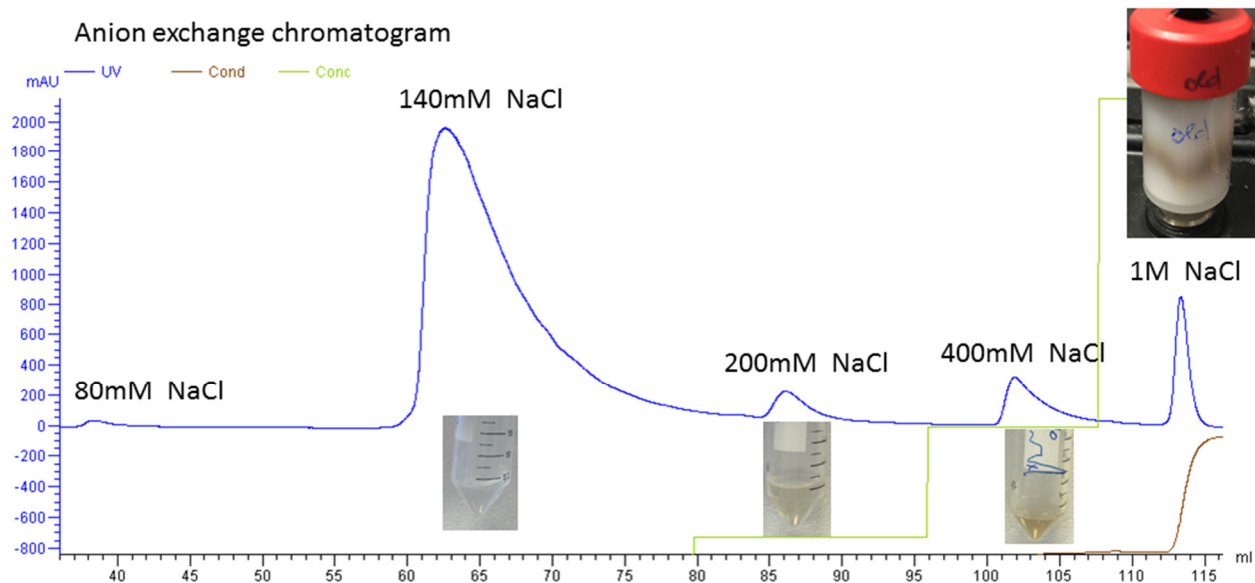
Extracellular secretion of recombinant protein in *P.pastoris* already acts as primary purification step. To further shorten the process, other purification routes were tested, in addition to the one described above. Since the recombinant LapA constructs harboured no affinity tags, the proteins could only be purified using their intrinsic physicochemical properties such as size, charge and hydrophobicity. It was not possible to load the harvested supernatant directly onto the AEX resin, due to the acidic pH of the *P.pastoris* expression culture (pH 5.0-6.0). For the protein to efficiently bind the anion exchange resin, the pH maintained in the binding buffer must be at least 1 unit above the theoretical pI of the target

protein (LapA precursor, pI ~4.9). LapA supernatant was concentrated and buffer-exchanged using ultrafiltration and the resultant protein solution was applied on the AEX resin using 5 mL Q HP columns. However, the 140 mM NaCl eluted fraction still contained a substantial amount of associating brown pigment. Even subsequent (twice repeated) re-applications on the AEX column did not separate all pigment from LapA precursor.



**Figure 4.31** Wild-type LapA purification on SEC column, following  $\text{AmSO}_4$  precipitation. LapA protein – main activity peak containing LapA precursor; Brown pigment peak – pigmented fraction, low molecular weight peptides from *P.pastoris* supernatant media.

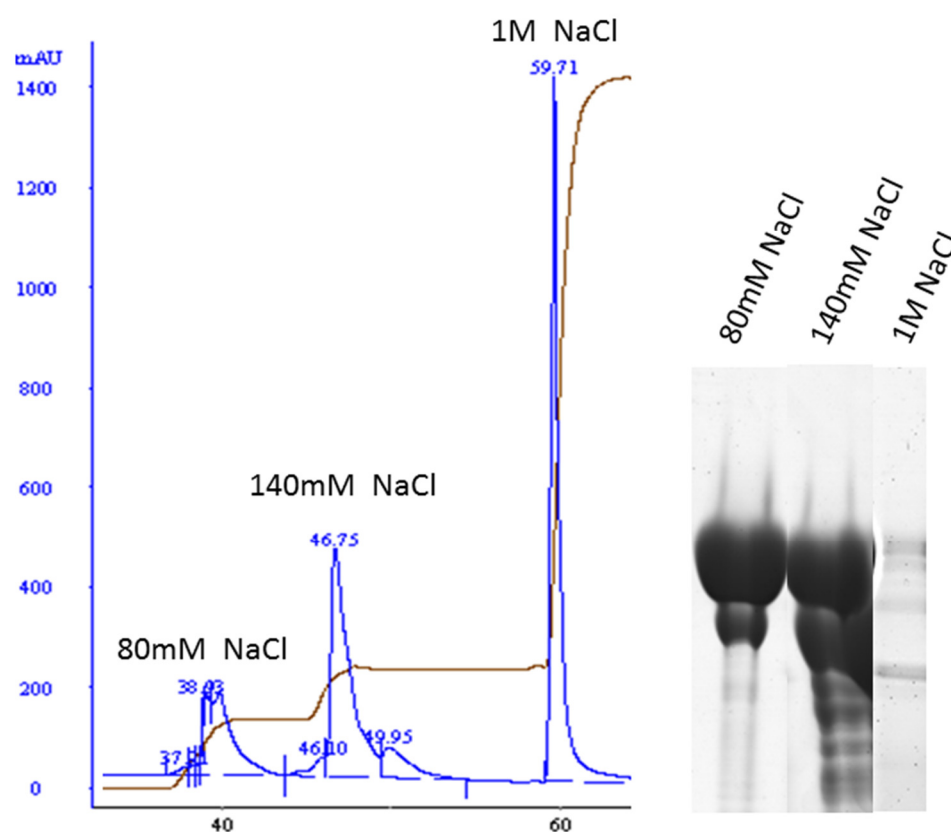
## Q HP column



**Figure 4.32** Typical peak elution pattern for LapA protein on a 5 mL Q HP AEX column. Higher concentration NaCl peaks contain increasing amounts of brown pigment, with 1 M NaCl fraction being the most pigmented (on-column appearance is also shown).

**Figure 4.32** summarizes a typical LapA elution pattern, observed on a 5 mL Q HP AEX column. The majority of LapA precursor elutes in the 140 mM NaCl fraction, which is the least pigmented. The LapA-bound pigment in this fraction can only be revealed by performing ultrafiltration (appearance of a concentrated brown layer towards the bottom of the membrane). The 200 mM and 400 mM eluted NaCl fractions were usually very pigmented and, therefore, discarded. The bulk of the pigment elutes in the 1 M NaCl wash, suggesting it contains highly negatively charged species binding to the resin with high affinity. This fraction usually contains small MW peptides from the media and LapA degradation products (**Figure 4.33**).

## MonoQ column



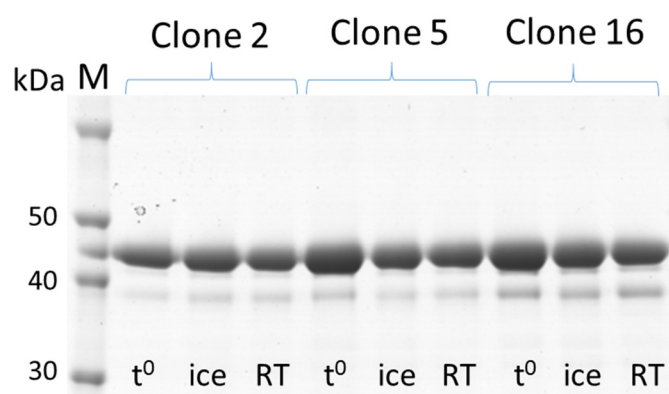
**Figure 4.33** Typical elution pattern of LapA protein from MonoQ AEX column. UV absorption trace is depicted by blue line, conductivity is depicted by brown line. SDS-PAGE of eluted fractions is given on the right. Approximately 50  $\mu$ g of total protein was loaded into each of the first two lanes of the gel.

In **Figure 4.33**, the SDS-PAGE (on the right) shows a typical distribution of LapA precursor and mature populations across the MonoQ-eluted fractions. The MonoQ AEX resin also enabled separation of an 80 mM NaCl peak, which mainly was comprised of LapA precursor. This fraction could then be concentrated and subjected to crystallisation screening. An equivalent 80 mM NaCl fraction could not be eluted from the Q HP column (**Figure 4.32**). The 140 mM NaCl fraction from the Mono Q column typically contained a 1:1 ratio of pro- and mature LapA populations. Interestingly, the majority of recombinant LapA, in the harvested supernatant, exists in a form of precursor (not shown). From a number of purifications performed on LapA precursor protein, it was evident that some form of precursor activation occurred during the purification process. However, the basis of this process could not be established. When partially purified solution, containing LapA

precursor, was left at room temperature for 24 h and SDS-PAGE analysis was performed the following day, no evidence of an autocatalytic process could be observed. Hence, the progressive cleavage of LapA propeptide, during the course of the purification process, must be due to the hydrolytic action of small amounts of co-purified exogenous protease in the *P.pastoris* supernatant.

#### 4.3.7 Exploring the cleavage of propeptide in *wtproLapA* proenzyme

The unexplained phenomenon of self-maturation of LapA precursor in the purification process of *wtproLapA* Clone 16 was also noted in the Clone 2 batch (results not shown). To investigate whether this was an autocatalytic process, supernatants of *wtproLapA* Clones 2, 5 and 16 were kept at RT and on ice (control) for 24h and changes in the distribution of protein populations were observed on SDS-PAGE (**Figure 4.34**). The maturation process observed could be a consequence of co-secreted, non-specific proteolytic activity in *P. pastoris* supernatant, which should become evident from prolonged incubation at room temperature.



**Figure 4.34** SDS-PAGE analysis of putative LapA autocatalytic maturation. t<sup>0</sup> – defrosted supernatant sample, ice – control sample following 24h incubation on ice, RT – sample following 24h incubation at room temperature.

However, no significant differences in LapA proenzyme processing were observed following incubation of selected supernatants on ice or at room temperature for 24h. Previously observed increases in the proportion of self-matured protein, in defrosted supernatants of Clone 2 and Clone 16, were not apparent in this experiment. Therefore, the recognized LapA self-maturation phenomenon in this context remained unclear. It was concluded, that LapA precursor requires treatment with exogenous protease for activation and this process is not autocatalytic. Also, there are no reported M28 family di-metallic

aminopeptidases with automaturation ability. The pro-form of the closest structural homologue known to date, Lap from *Vibrio proteolyticus*, required activation with Proteinase K (Bzymek et al., 2004).

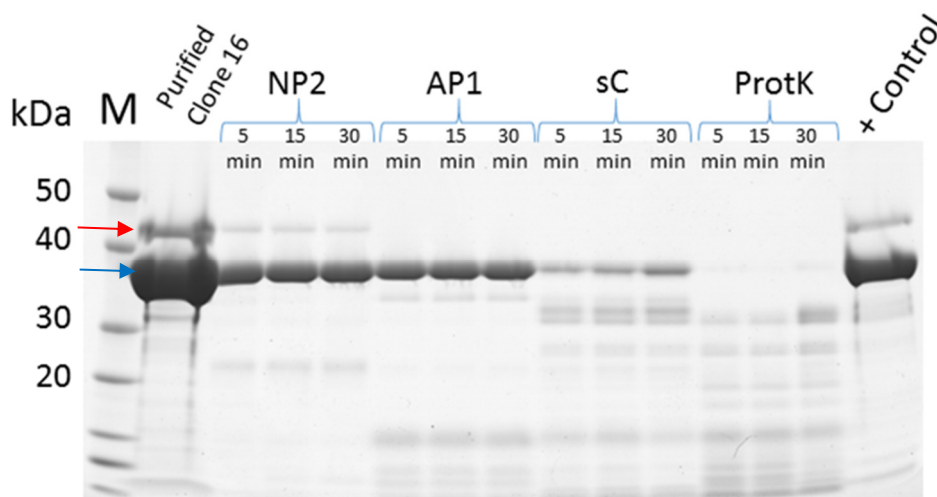
#### **4.3.7.1 Enzymatic processing of *wtproLapA* proenzyme with AP1 and NP2**

Previous attempts to use *fraction IV* from Flavourzyme to activate recombinant LapA precursor expressed in *P. pastoris* yielded promising results (sections 4.3.3 and 4.3.3.2). Therefore, the next rational step was to test individually isolated enzymes (AP1 and NP2) from *fraction IV* for their ability to process LapA pro-protein. The isolation protocol of AP1 and NP2 is described in section 3.3.

Purification of Clone 16 LapA precursor was attempted in section 4.3.5.1. This batch of LapA zymogen displayed a phenomenon of propeptide degradation in the purification process. However, a small fraction of this protein remained immature following the final AEX purification step. Thus, enzymatic activation of this remaining *wtproLapA* population was attempted, using isolated AP1 and NP2 pools from Flavourzyme to determine which enzyme is responsible for activation of native LapA pro-protein in *Aspergillus oryzae*. Also, Clone 16 was treated with commercially available Subtilisin Carlsberg and Proteinase K non-specific proteases.

15 µg of purified Clone 16 enzyme was combined with approximately 1 µg of each effector proteins and, while continuously incubating at 37 °C, samples were withdrawn at set time points to monitor the progress of LapA precursor processing.

AP1, Subtilisin Carlsberg, and Proteinase K immediately processed the Clone 16 zymogen population. No pro-protein was visible, even following 5 minute incubation (**Figure 4.35**). Subtilisin Carlsberg appears to have processed more than half of the initial content of Clone 16, whereas Proteinase K almost completely degraded the recombinant LapA pro-enzyme. Hence, the commercial non-specific proteases were not considered as potential candidate enzymes for activation of the LapA zymogen.



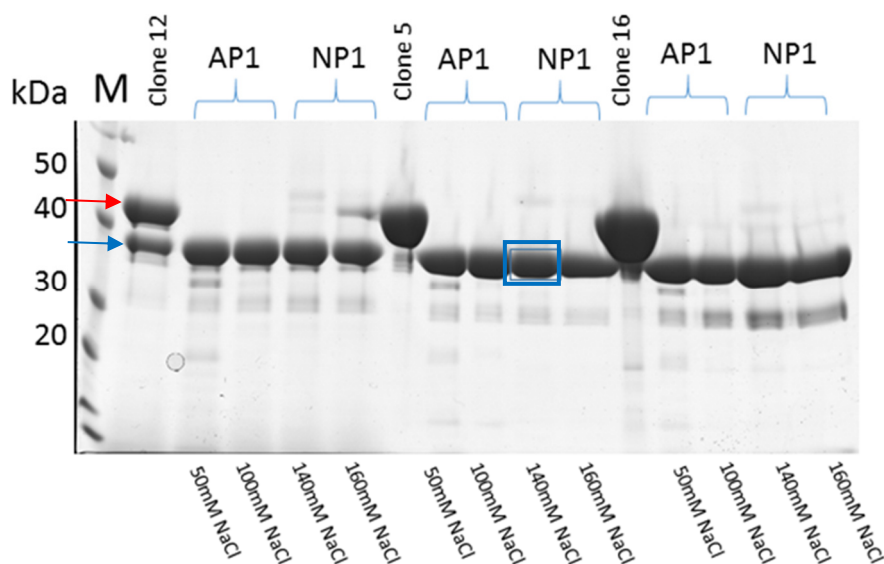
**Figure 4.35** SDS-PAGE analysis of *wtproLapA* Clone 16 enzymatic treatment with proteases. NP2 – added neutral protease 1 (190 mM NaCl peak); AP1 – added alkaline protease 1 (50mM NaCl peak); sC – added Subtilisin Carlsberg; ProtK – added Proteinase K; + Control – Clone 16 purified protein kept at 37 °C for 4h. Red and blue arrows indicate processed and unprocessed enzyme populations, respectively.

NP2 did not seem to have a significant propeptide cleavage ability. Therefore, it was concluded that AP1 protease is the most likely enzyme for future processing of recombinant LapA zymogens. At the same time, AP1 did not cause severe degradation of the Clone 16 enzyme and was proposed to primarily activate native LapA upon secretion into the extracellular environment of *Aspergillus oryzae*.

#### 4.3.7.2 Comparison of recombinant LapA precursor activation by NP1 and AP1 proteases

In the improved isolation protocol of Flavourzyme proteases, neutral protease 1 (NP1) was also purified from Flavourzyme (section 3.3.1). It was important to confirm whether NP1 protein could also be involved in LapA activation. For further elucidation of the mechanism of LapA N-terminal processing, it was deemed necessary to use both Flavourzyme-isolated NP1 and AP1 proteases in the same LapA zymogen activation trial and confirm any tentative differences in N-terminal processing.

Three different *wtproLapA* batches were utilized in this experiment: Clone 12, Clone 5 and Clone 16, which were formerly purified using the standard protocol (section 2.13.4). These LapA zymogens were combined with isolated AP1 and NP1 fractions from CAX – AEX purification protocol (see section 3.3.1). Following 60 minute incubation at 37 °C, samples were analyzed by SDS – PAGE (**Figure 4.36**).



**Figure 4.36** Analysis of enzymatic processing of different LapA zymogen proteins. Added fractions, which were eluted from a MonoQ column, in the ultimate AEX purification step, are indicated at the bottom of the gel. Activating proteases contained in these fractions are shown at the top. Red and blue arrows indicate processed and unprocessed enzyme populations, respectively. Blue rectangle indicates N – terminally sequenced protein band.

AP1 fractions effectively processed all three LapA precursors. NP1 also processed most of the LapA precursors, except in the Clone 12 sample at 160 mM NaCl, where some pre-protein was still present. The optimal pH of 7, reported for AP1 and NP1 proteins, was identical (Merz et al., 2015). The pH maintained in the LapA activation reactions was 1 pH unit above the optimal pH for these proteases. The reported optimal temperature for AP1 and NP1 was 40 and 50 °C, respectively. Therefore, both enzymes operated under almost ideal conditions. No other significant differences in processing efficacy were determined. Sampling at intermediate time points within a 60-minute interval could possibly reveal more differences between AP1 and NP1 proteases. Proteolytic degradation of LapA precursor proteins was also noticeable (to some extent) in all fractions.

To verify any differences in the resulting LapA N-terminus, following propeptide cleavage by the NP1 protease, the activated LapA sample (160 mM NaCl AEX fraction) was analysed. The first 5 residues identified at the LapA N – terminus were 'AVTYP', which was identical to that previously observed following treatment with isolated AP1 protein (see **Figure 4.26**). It is likely that both NP1 and AP1 proteases possess similar substrate specificity and, consequently, display an identical mechanism for *wtproLapA* processing. Another possibility is that the 160 mM NaCl NP1 fraction contained a small amount of co-purified AP1 protease, which was able to further modify the LapA N – terminus. In the CAX-AEX purification experiment, the last fraction of AP1 eluted at 120 mM NaCl concentration, whereas the first NP1 fraction appears at 140 mM NaCl concentration, suggesting a slight overlap in elution (**Figure 3.12**). Therefore, NP1 might not even be able to process LapA zymogen, and only trace amounts AP1 are required for efficient cleavage of LapA propeptide. This requires further experimental validation.

## **Chapter 5 STRUCTURAL STUDIES OF RECOMBINANT WILD - TYPE LAPA: MATURATION, INHIBITION AND SUBSTRATE BINDING**

---

## **5.1 Crystal structures of LapA precursor and mature enzyme: role of propeptide in inhibition and as intramolecular chaperone**

*This sub-section of Chapter 5 was written in the format of publication to the scientific journal. Therefore, the content is laid out in a slightly distinct manner to other chapters.*

### **5.1.1 Introduction**

Modification of the N-terminus is one of the most common polypeptide processing events, occurring in any cell. Aminopeptidases (EC 3.4.11) are a subset of hydrolase enzymes that either release a single N-terminal residue or dipeptides/tripeptides (in the case of dipeptidyl- and tripeptidyl-aminopeptidases). Metallo-aminopeptidases are a division in this class of enzymes, utilising one or two metal cofactors in the active site formation. Binuclear metallo-aminopeptidases are ubiquitous enzymes, commonly displaying a broad substrate specificity and involved in general degradation of proteins at the cellular level. Abnormal aminopeptidase activity and aberrant protein degradation was linked to inflammation, ageing, cancer and leukemia (Taylor, 1993b, Taylor, 1993a). Also, metallo- $\beta$ -lactamases (m $\beta$ ls), binding one or two zincs in the activate site, are active targets to combat multi-drug resistance in pathogenic bacteria. To date, no clinically useful inhibitor for m $\beta$ l's have been identified, with only a few potent compounds discovered recently (Wang et al., 2015). Thus, a mechanistic understanding of the principles behind the action of dizinc enzymes and subsequent design of novel inhibitors is critical (Tamilselvi and Muges, 2008, Tamilselvi et al., 2006).

It is well established, that prodomains of secretory endo- and exopeptidases play an important role in inhibition of their cognate catalytic domains, which undergo activation by removal of the prodomain either by autocatalytic process or external endopeptidase activity. By secreting protease precursors in a zymogen form, undesirable cytosolic activation and proteolysis events are prevented (Lazure, 2002). Normally the zymogen conversion does not involve conformational changes of catalytic residues and the active sites are sterically rendered inaccessible to substrate by the unique residues of prodomain or prosegment, thus preventing activity (Khan et al., 1999).

In addition, N-terminal propeptides are known to act as intramolecular chaperones (IMC), decreasing the energy barrier of transition from a molten globule intermediate to a kinetically trapped native state (Subbian, 2015). In this respect, a well-studied model enzyme, subtilisin, has been shown to be unable to escape the folding transition state between a molten-globule and a native conformation, without the presence of a prodomain; a rate-limiting step in the folding process of subtilisin (Shinde and Inouye, 2000). Remarkably, prodomains of numerous proteinases have been reported to assist as chaperones in both inter- and intramolecular manner, following expression in trans- and cis- form, respectively (Beggah et al., 2000, Tang et al., 2002, Safina et al., 2011). An analogous mode of action, for such prodomains, with no requirement for a covalent link between propeptide and catalytic domain, has been discovered in other enzymes (Yurimoto et al., 2004).

Subtle and rather complex relationships between propeptide and catalytic domain were found in subtilisin E, where unaltered, mature polypeptide chain of subtilisin E was shown to fold into a different conformation via a mutated intramolecular chaperone and to show altered substrate specificity; a phenomenon termed `protein memory` (Shinde et al., 1997). Normally, such intramolecular chaperones precede catalytic domain from the N-terminus, however, eukaryotic monozinc aminopeptidase A was found to recruit its C-terminal propeptide for this role (Rozenfeld et al., 2004). Such findings demonstrate the diverse and complex relationships between the essential prodomain and the cognate active conformation in the proteogenesis of proteases.

The aminopeptidase from *Vibrio proteolyticus* (AAP) is a model bimetallohydrolase enzyme, representing the M28 peptidase family, MH clan, whose extensive structural and functional studies have provided an in-depth understanding of the mechanism of action and the substrate preferences for these bi-metallic enzymes (Schalk et al., 1992, Chevrier et al., 1994, Chevrier et al., 1996, Bennett and Holz, 1997, Chen et al., 1997, Stamper et al., 2001, Stamper et al., 2004, Schurer et al., 2004, Desmarais et al., 2006, Kumar et al., 2007). Furthermore, the role of the propeptide in AAP and other homologous leucine aminopeptidases has been extensively studied *in vitro* (Zhang et al., 2000, Bzymek et al., 2004, Nirasawa et al., 1999). A number of studies have provided structural insights into AAP inhibition via different synthetic compounds, acting as competitive inhibitors (Chevrier et al., 1996, Stamper et al., 2004, Hanaya et al., 2012, Stamper et al., 2001). However, until now, no report exists to provide a mechanistic insight for leucine aminopeptidase inhibition by its natural inhibitor – the N-terminal propeptide - at a molecular level.

In this article, we present the crystal structures of recombinantly expressed LapA precursor (*wtproLapA*) and mature recombinant LapA (*wtLapA*), both purified from *P.pastoris* culture supernatant. These structures reveal a plausible molecular mechanism of LapA catalytic domain inhibition by propeptide and we propose a role for the intramolecular chaperone N-terminal domain in the process of maturation of LapA zymogen. We found that the mature catalytic domain of LapA does not achieve its native conformation, when produced in the absence of prosegment. This observation is reinforced by an expression and refolding study of 6xHis-tagged catalytic domain only (*refLapA*) from *E.coli* inclusion bodies, revealing a limited active conformation.

To date, the exact mechanism of leucine aminopeptidase inhibition and the role of an intramolecular chaperone (IMC) by the prodomain has not been detailed at a molecular level. To the best of our knowledge, this is the first crystal structure of a leucine aminopeptidase precursor, revealing structural aspects of mature domain inhibition by the cognate propeptide. Based on this work, we are able to show that known synthetic aminopeptidase inhibitors and substrates mimic key contacts between pro- and catalytic domain. These findings could aid future design of more effective inhibitors of bimetallic aminopeptidases and other dizinc enzymes, sharing an analogous reaction mechanism.

Several AAP crystal structures in complex with a natural substrate, LLL (Kumar et al., 2007) and substrate analog, 1-Butaneboronic acid (De Paola et al., 1999) have been reported. In this work, we also introduce the crystal structure of LapA with a single leucine trapped in the active site, termed *wtLapA-Leu*. We demonstrate that the binding mode of this leucine is analogous to the P1 moiety, observed in the AAP-LLL crystal structure.

## 5.1.2 Materials and Methods

### 5.1.2.1 Strains and plasmids

*Pichia pastoris* X33 strain was obtained from Biogrammmatics Inc. (Carlsbad, CA, USA). Secretory expression vector, pJ902, was provided by DNA 2.0 (USA). *Escherichia coli* DH5 $\alpha$  (*fhuA2*  $\Delta$ (*argF-lacZ*)*U169 phoA glnV44*  $\Phi$ 80  $\Delta$ (*lacZ*)*M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*) cloning strain were purchased from New England Biolabs (NEB,UK).

All primers used in this study were obtained from Sigma (UK). Zeocin<sup>TM</sup> was purchased from Invitrogen (USA). All other reagents were purchased from Sigma (UK) and were of analytical grade.

### 5.1.2.2 Construction of recombinant expression vectors and transformations in *Pichia pastoris*

Recombinant LapA open reading frames (Genebank accession: XP\_001825745.1) were codon-optimised and cloned into the pJ902 vector, by DNA 2.0, using the *EcoRI* and *NotI* restriction sites. In this way, an expression construct, harbouring the LapA native signal peptide and propeptide sequences (pJ-npro-LapA), was obtained. A second expression construct, containing a deletion of the LapA propeptide (pJ- $\Delta$ pro-LapA) was produced, utilising the QuickChange<sup>TM</sup> II site-directed mutagenesis kit from Stratagene (USA). Deletion of the propeptide sequence was achieved by performing PCR with designed primers 1; 5'-GGACGCTATCTGGATAAGCTAATGCGCTTGCC-3' and 2; 5'-GGCAAGCGCATTAGCTTATCCAGATAGCGTCC-3', using the pJ-npro-LapA construct as a template.

Following *SacI* linearization and clean-up, recombinant plasmids were transformed into an X33 expression strain by electroporation (a single 1500 V pulse in a 2 mm electroporation cuvette). Immediately following electroporation, the cells were resuspended in 0.5ml 1M sorbitol and 0.5ml YPD (1% yeast extract Y, 2% soy peptone P, 2% dextrose D) and incubated for 2h at 30°C under gentle agitation. The revived cells were then plated onto YPD agar (1% yeast extract, 2% soy peptone, 2% dextrose, 2% agar), containing either 100  $\mu$ g/ml or 2000  $\mu$ g/ml Zeocin<sup>TM</sup> and incubated at 30 °C 2-4 days. The resulting *P.pastoris* colonies were verified, by PCR, for appropriate genomic integration using the method previously described (Looke et al., 2011). In each case, 10 transformants were screened, picking the largest, single colonies. Each single colony was patch-plated onto a separate YPD agar plate and residual cells were re-suspended in 100 $\mu$ L lysis solution (200 mM lithium acetate, 1% SDS), followed by incubation at 70 °C for 15 min. Following 96% ethanol precipitation, the

DNA pellet was re-suspended in 100  $\mu$ L sterile water and 1 $\mu$ L (in each case) was used in each subsequent genomic PCR. Clone ID (Lucigen, USA), PCR pre-mix and a pair of AOX primers (AOX1-F; 5'-GACTGGTTCCAATTGACAAGC-3' and AOX1-R; 5'-GCAAATGGCATTCTGACATCC-3') were used, under the following conditions: 1 cycle of 98 °C for 2 min, 30 cycles at 98 °C for 30 s, 56 °C for 30 s, 72 °C for 1 min, finishing with 72 °C for 5 min. The PCR products of predicted size were visualised on a 1% agarose gel.

#### **5.1.2.3 Construction of recombinant expression vectors and transformations in *E.coli***

Mature LapA polypeptide (Genbank accession: XP\_001825745.1) was codon-optimised and recombinantly cloned into a pET-21a expression vector (Eurofins, UK), using *Nde*I and *Bam*HI restriction sites. In this way, a pET-6xHis $\Delta$ proLapA expression construct resulted, comprised of the T7 bacteriophage promoter, an N-terminal hexahistidine tag, mature LapA polypeptide (minus propeptide) and an ampicillin resistance marker. This LapA expression construct was transformed into BL21 (DE3) competent cells, according to the manufacturer's instructions.

#### **5.1.2.4 Over-expression of recombinant LapA in *P.pastoris***

Recombinant proteins were over-expressed in a *P. pastoris* X-33 strain (Mut<sup>+</sup>), under control of the *P<sub>AOX1</sub>* methanol inducible promoter. Shake flask expression trials were conducted in BMGY (1% yeast extract, 2% soy peptone, 1% glycerol, 1.34% YNB, 4E-5% biotin, 100mM potassium phosphate, pH 6.0) and BMMY (as for BMGY with 1% glycerol replaced by 0.5% methanol). Genomic PCR-verified *P.pastoris* clones were inoculated into 50 ml BMGY medium in 250 ml baffled conical flasks and incubated in a shaker (250 rpm) for 16h at 28 °C. Once the OD<sub>600</sub> reached 2-8, an equivalent amount of cells (corresponding to OD<sub>600</sub> of 1) were harvested and re-suspended in 100 ml of BMMY and shaken (250 rpm) in 500 ml conical baffled flasks at 28 °C for 4-5 days. Every 24h, the cultures were supplemented with 0.5% final concentration of methanol to maintain induction and OD<sub>600</sub> cell densities were recorded. After 120h of maintaining recombinant protein expression, the cell cultures were harvested by centrifugation at 6000  $\times$  g for 15 min. The supernatant was sampled and recombinant LapA protein detected by SDS-PAGE (NuPAGE® Novex® 4-12% Bis-Tris gels).

#### **5.1.2.5 Purification of recombinant LapA from *P.pastoris* supernatant**

Over-expressed *wtpro*LapA, *wt*LapA, *wt* $\Delta$ proLapA proteins were purified from *P.pastoris* supernatant employing 2 different protocols: 1) crude *P.pastoris* supernatant, containing

recombinant LapA, was dialysed against 20 mM Tricine pH 8.0, 1 mM ZnCl<sub>2</sub> for 16h at 4 °C, then filtered through a 0.22 µm steri-filter prior to loading onto a 5 mL HiTrap™ Q HP anion exchange column, pre-equilibrated with 20 mM Tricine pH 8.0, 1 mM ZnCl<sub>2</sub>. The column was washed (5 – 10 CV) with equilibration buffer until no trace of protein appeared in the eluate. Recombinant LapA was eluted using an increasing gradient (0 - 1M) of NaCl (20 mM Tricine pH 8.0, 1 mM ZnCl<sub>2</sub>, 1M NaCl). Active fractions were pooled and concentrated, using ultrafiltration membranes (10 kDa MWCO). In each case, the respective protein concentrate was loaded onto a HiLoad Superdex™ 16/60 200pg (GE Healthcare, UK) column, pre-equilibrated with 20 mM Tricine pH 8.0, 1 mM ZnCl<sub>2</sub>. Homogeneous, recombinant LapA fractions were concentrated and stored at -20 °C for further analysis; or 2) crude *P.pastoris* supernatant, containing the respective recombinant LapA constructs, was saturated to 90% ammonium sulphate by adding hard salt, followed by stirring for 1h at ambient temperature. LapA precipitant was pelleted by centrifugation at 16,000  $\times$  g for 20 minutes. The collected pellet was re-suspended in 20mM Tricine pH 8.0, 1 mM ZnCl<sub>2</sub>, followed by dialysis against the same buffer overnight at 4 °C. At this point, the purification procedure was equivalent to 1), starting at the anion exchange chromatography step. Recombinant proteins purified, using protocols 1) and 2), were analysed on SDS-PAGE.

#### **5.1.2.6 Over-expression of recombinant LapA in *E.coli***

All preliminary recombinant LapA over-expression trials were carried out using 10 mL cultures. A single BL21 (DE3) colony harbouring pET-6xHisΔproLapA expression construct was transferred to 2 mL LB broth and incubated at 37 °C (shaker at 220 rpm) overnight. 200 µL of the overnight inoculum was transferred to 10 mL of fresh LB media. The optical density (OD) at 600 nm was monitored regularly to reach a value between 0.5 and 1.0 absorbance units, after which IPTG (1 mM final concentration) was added to induce protein expression culture. The pre-induction sample was taken for further analysis on SDS – PAGE. The induced cultures were then incubated for 16-20h at either 37 °C or 20 °C. Following harvest at 16,000  $\times$  g, cells were lysed with BugBuster® Master mix (Novagen) and soluble versus insoluble protein fractions were analysed on SDS-PAGE.

#### **5.1.2.7 Purification and refolding of *refLapA* protein from *E.coli* inclusion bodies**

The pellet with inclusion bodies were re-solubilised in 8M urea buffer (50mM Tricine pH 8.0, 500 mM NaCl, 8M urea) gently agitating for 1 hour. The non-dissolved debris was separated by centrifugation for 20 min at 16,000  $\times$  g. The resulting supernatant was loaded onto a HisTrap™ FF 5 ml (GE Healthcare, UK) nickel immobilisation affinity column

(IMAC), pre-equilibrated with wash buffer (50 mM Tricine pH 8.0, 500 mM NaCl, 8M urea) and re-solubilised *ref*LapA was eluted in 50 mM Tricine pH 8.0, 500 mM NaCl, 8M urea and 200 mM imidazole. *ref*LapA was allowed to slowly refold by reducing the concentration of urea in the buffer: IMAC-eluted fraction was dialysed against refolding buffer (10 mM Tricine pH 8.0, 1 mM ZnCl<sub>2</sub>) for 20h at 4 °C. The *ref*LapA was further purified using a HiLoad Superdex™ 16/60 200pg (GE Healthcare, UK) gel filtration column pre-equilibrated with 20 mM Tricine pH 8.0, 1 mM ZnCl<sub>2</sub>. Sample collection (2 ml fractions) of fractions exhibiting the highest specific activity were pooled and concentrated on Vivaspin 15R – 10,000 Dalton Centrifugal Filter Units (Sartorius, UK) containing the HydroSart® membrane.

#### **5.1.2.8 Aminopeptidase assay**

LapA specific activity was determined spectrophotometrically, using a synthetic substrate (Leu-*p*NA) and monitoring the formation of the *p*-nitroanilide moiety (*p*NA), as absorbance, at 405nm ( $\epsilon = 9.9 \text{ mmol}^{-1}\text{cm}^{-1}$ ) in a cuvette with a 1 cm path length. Peptidase assays were performed on a CE1020 1000 Series spectrophotometer (Cecil Instruments). One unit of LapA activity was defined to catalyse the hydrolysis of one micromole of L-leucine-*p*-nitroanilide to L-leucine and *p*-nitroaniline, per minute at pH 7.2, 37 °C.

#### **5.1.2.9 Quantification of recombinant LapA**

LapA proteins were quantified either by Bradford assay (Bio-Rad, USA), using BSA as a standard, or spectrophotometrically, employing a NanoDrop® ND-1000 (Thermo Scientific), using the LapA absorbance extinction coefficient ( $23,505 \text{ M}^{-1}\text{cm}^{-1}$ ). The concentration of recombinant LapA in crude *P.pastoris* supernatant could only be determined by Bradford assay due to impurities interfering with spectrophotometric measurements.

#### **5.1.2.10 CD spectroscopy of recombinant LapA**

The circular dichroism spectra for all resultant proteins were recorded in the range 360 to 190 nm at 20 °C, using a Chirascan series 800 spectrophotometer. Measurements were taken using a 1 nm step size in a 0.1 mm path length quartz cell. CD data were averaged from 5 independent scans for each sample. Samples of *ref*LapA, *wt*LapA and *wt* $\Delta$ *pro*LapA proteins were analysed at 36  $\mu$ M, 21  $\mu$ M and 40  $\mu$ M concentrations, respectively, buffered in 20 mM HEPES pH 8.0.

### 5.1.2.11 Crystallisation and structure determination of recombinant LapA

Crystallisation screening was carried out using HTP robotics at Oxford Protein Production Facility (OPPF-UK) (96 well commercial sparse matrix screens, sitting-drop vapour diffusion method) and crystallisation optimisation was carried out in the home laboratory using 24 well, hanging-drop vapour diffusion. *wtproLapA*, *wtLapA* recombinant proteins were crystallised at concentrations of 13 mg/ml and 11 mg/ml, respectively. Both proteins were buffer exchanged into 20mM Tricine pH 8.0, 1mM ZnCl<sub>2</sub> prior to crystallisation. *wtproLapA* crystals with approximate dimensions 0.5 × 0.04 × 0.005 mm formed under F10 condition (0.1 M Bis-Tris, pH 5.5, 0.2 M NaCl, 25% PEG 3,350) in the Index HT™ screen (Hampton Research, USA) in 3 days at 21 °C. *wtLapA* crystals were produced in 24-well plate format at room temperature under previously identified F2 (0.1M Citrate, pH 5.0, 3.2 M Ammonium sulphate) crystallisation condition in JCSGPlus HT™ screen (Molecular Dimensions, UK).

Prior to data collection, crystals were cryo-protected by quick transfer through a mixture of mother liquor and 20% ethylene glycol (except *wtproLapA*, which had 25% PEG in the mother liquor) and flash cooled to 100K in liquid nitrogen. Diffraction data were collected at Diamond Light Source (UK), on beamlines I02 and I03. The diffraction data were recorded, by fine slicing (0.1° degree oscillation), on a Pilatus 6M detector. Indexing and integration were completed automatically by *XDS* (Kabsch, 2010) and *xia2* (Winter et al., 2013).

All the LapA structures were solved by maximum likelihood molecular replacement with PHASER (McCoy et al., 2005), using the coordinates of an established LapA homology model; LapA homology model was generated using the IntFold2 server (Roche et al., 2011), which identified aminopeptidase from *Vibrio sp.* (34% sequence identity) as a closely related structural homologue (PDBid: 1RTQ), as the template. Resulting LapA models were manually built, fitting to the resulting electron density maps using COOT (Emsley and Cowtan, 2004). All structures were refined using PHENIX (Adams et al., 2010). MolProbity4 was used for structure validation. Structural comparisons and superimpositions were generated using the TM-align server (Zhang and Skolnick, 2005). Structural images were prepared using PyMOL. Structural data was deposited in the PDB database with the following accession codes: 5NQJ (*wtproLapA*) and 5NQL (*wtLapA*).

### 5.1.3 Results

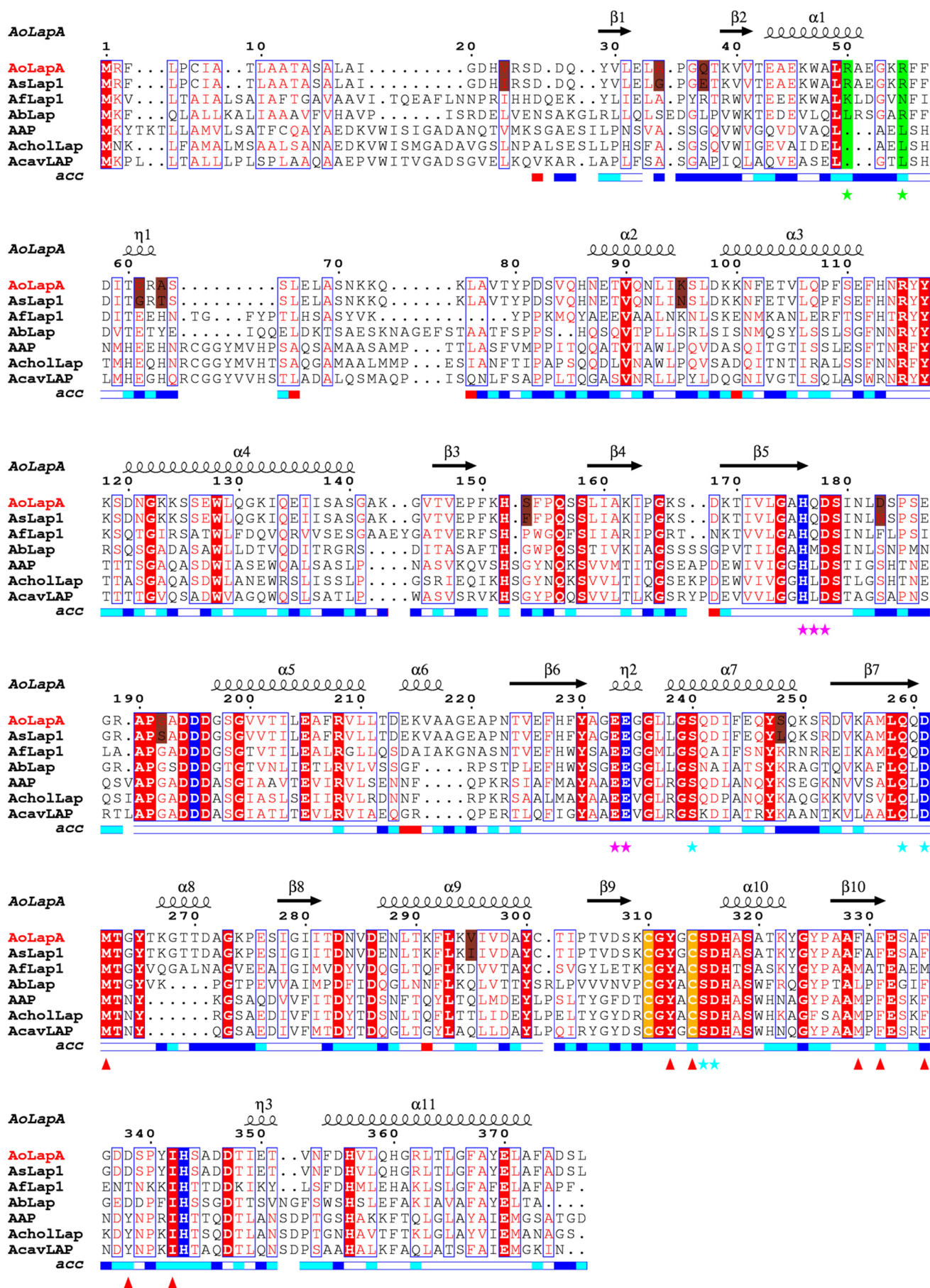
#### 5.1.3.1 General structural comparison of Lap peptidases

Extracellular aminopeptidases from the M28 family share a very similar architecture of their open reading frames (ORF`s). In general, a short secretion signal sequence is preceded by a moderate length propeptide (removed in the maturation process), and a central catalytic domain. It is convenient to compare their primary sequences and their tertiary structure folds to establish conservation and evolutionary divergence at both the sequence and structural level. Evolutionary relationships between Lap1 from *A.sojae* and eukaryotic, as well as microbial homologues, were previously investigated (Chien et al., 2002). In our study, full ORF`s of several of these homologous fungal and microbial extracellular Lap enzymes were aligned together with LapA from *A.oryzae* (AoLapA) and Lap1 from *A.sojae* (AsLapA) (**Figure 5.1**). Compared to highly conserved catalytic core residues, a varying degree of homology was observed within the signal peptide and propeptide sequences.

LapA from *A.oryzae*, Lap1 from *A.sojae* and AAP from *Aeromonas proteolytica* all belong to the M28 peptidase family; clan MH (MEROPS). AAP has been studied extensively and has been a model enzyme for this family, with regard to structural homology (Schalk et al., 1992, Chevrier et al., 1994). Lap enzymes are typical structural representatives of this family with two Zn(II) ions sequestered from the surrounding environment and coordinated in the active site by 5 conserved residues in sequence order His, Asp (coordinates both Zn(II) ions), Glu, Asp, Glu, His. Also, an activated water molecule, associated with both metal ions, acts as nucleophile during catalysis. Additional Asp and Glu residues, considered to be essential for catalysis, reside in close proximity within His-Xaa-Asp and Glu-Glu motifs. Lap enzymes contain an archaic alpha/beta hydrolase fold with a five-stranded beta sheet (1 strand antiparallel to the rest) sandwiched between a layer of helices on each side. The active site is parallel to the end of these beta strands.

The coordinated divalent zinc ions form a functional charge/dipole complex along with conserved five amino acid ligands (LapA numbering: His176, Asp195, Glu234, Asp261, His343). Another conserved Glu233 residue, within the Glu-Glu motif, was supposed to assist in deprotonation of the terminal catalytic water molecule and, subsequently, donating a proton to newly formed amino-terminus, described as a rate-limiting step in the catalysis (Schurer et al., 2004). Several mutant forms of this residue rendered AAP enzyme inactive (Bzymek and Holz, 2004, Bzymek et al., 2005). Also, a complementary second-shell hydrogen bond network was identified in AAP and another homologous aminopeptidase

from *Streptomyces griseus* (SGAP), whose crystal structures have been solved (Gilboa et al., 2000). We found that the Zn-D261-S315-Q259-D316-S240 network, identified in the AAP structure, also was conserved in *A. oryzae* LapA (AoLapA). Four residues (Met262, Tyr312, Cys314, Ile342), lining the hydrophobic specificity binding pocket, also were conserved in LapA and homologous extracellular aminopeptidases, including Lap1 from *A. sojiae* and AAP from *A. proteolytica* (**Figure 5.1**).



**Figure 5.1.** Multiple, full open reading frame alignment of homologous leucine aminopeptidases. LapA from *Aspergillus oryzae* (AoLapA, Q2U1F3), Lap1 from *Aspergillus sojae* (AsLap1, Q8J2N2), Lap1 from *Aspergillus fumigatus* (AfLap1, Q5VJG6), Lap from *Agaricus bisporus* (AbLap, Q8WZH8), Lap from *Vibrio proteolyticus* (AAP, Q01693), Lap from *Aeromonas cholerae* (AcholLap, P96152), Lap from *Aeromonas caviae* (AcavLAP, O82996) sequences were aligned. Secondary structure elements were assigned, according to structural data from the LapA precursor structure from this work. Conserved residues are displayed in red shading, Zn-coordinating residues – in blue shading, conserved His-Xaa-Asp and Glu-Glu motifs – labelled with pink stars, cysteine residues forming a single S-S bond are displayed in gold shading, and columns containing key LapA propeptide arginines - in green shading and green stars. Residues lining the hydrophobic specificity pocket and conserved second-shell hydrogen bond network are marked with red triangles and cyan stars, respectively. 11 different amino acids between AoLapA and AsLap1 open reading frames are displayed in brown shading. Residue accessibility (*acc*) bar depicts relative accessibility based on the LapA precursor structural coordinates (*dark blue* – very accessible, *light blue* – intermediate, *white* – inaccessible, *red* – not defined). Terminal parts of AAP C-terminal propeptide and the AcholLap PPC domain were truncated to achieve better alignment. Multiple sequence alignment produced using MAFF algorithm (Kato et al., 2002). The figure was prepared using ESPript 3.0 (Gouet et al., 2003).

The highly hydrophobic character of the specificity pocket allows such enzymes to favour bulky hydrophobic residues (Leu, Phe) in their active sites. The other four hydrophobic pocket residues varied among compared species; Phe329 in both AoLapA and AsLap1 was substituted with alternative hydrophobic residues – Leu in AbLap and Met in the rest of homologues, including AAP. Interestingly, two highly conserved residues Phe331 and Phe335 in AfLap1 were replaced by Thr and Met, respectively. The last distinct residue observed was Asp338 in AoLapA and AsLapA. At this position, a hydrophobic Tyr is placed in AAP, AcholLap, AcavLAP. Asp338 has been implicated in the hydrolysis of positively charged N-terminal amino acids by recombinant (rLap1) from *A. sojae* (Huang et al., 2015). LapA also contains a negatively charged Asp338. However, this does not render the enzyme capable of accommodating large positively charged residues (Arg or Lys), with the same efficiency as previously reported for rLap1 from *A. sojae* (Huang et al., 2015). LapA showed a preference for bulky hydrophobic residues in the active site and acted to remove N-terminal lysine with only 25% of the efficiency observed for the substrate with N-terminal leucine

(Matsushita-Morita et al., 2011). The reported differences in substrate specificity between LapA (AoLapA) and rLap1 (AsLapA) and wild-type Lap1 (AsLapA) must be associated with residue substitutions within the mature domain and propeptide (residues highlighted in brown in **Figure 5.1**).

### 5.1.3.2 LapA expression in *P.pastoris*

Coding regions of LapA precursor (*wtproLapA*) and prodomain-deleted version (*wtΔproLapA*) were codon-optimized for use by a *Pichia pastoris* host. Multiple gene dosage transformants were obtained for both (pJ-npro-LapA and pJ-Δpro-LapA) constructs. However, no real-time qPCR was attempted to quantify the number of integrated gene copies, in each case.

Preliminary expression trials of the wild-type recombinant LapA with an alpha factor signal sequence and the native LapA signal peptide, resulted in slightly higher secretion levels with the latter (results not shown). Thus, it was decided to study further the expression of recombinant LapA harbouring the native signal peptide.

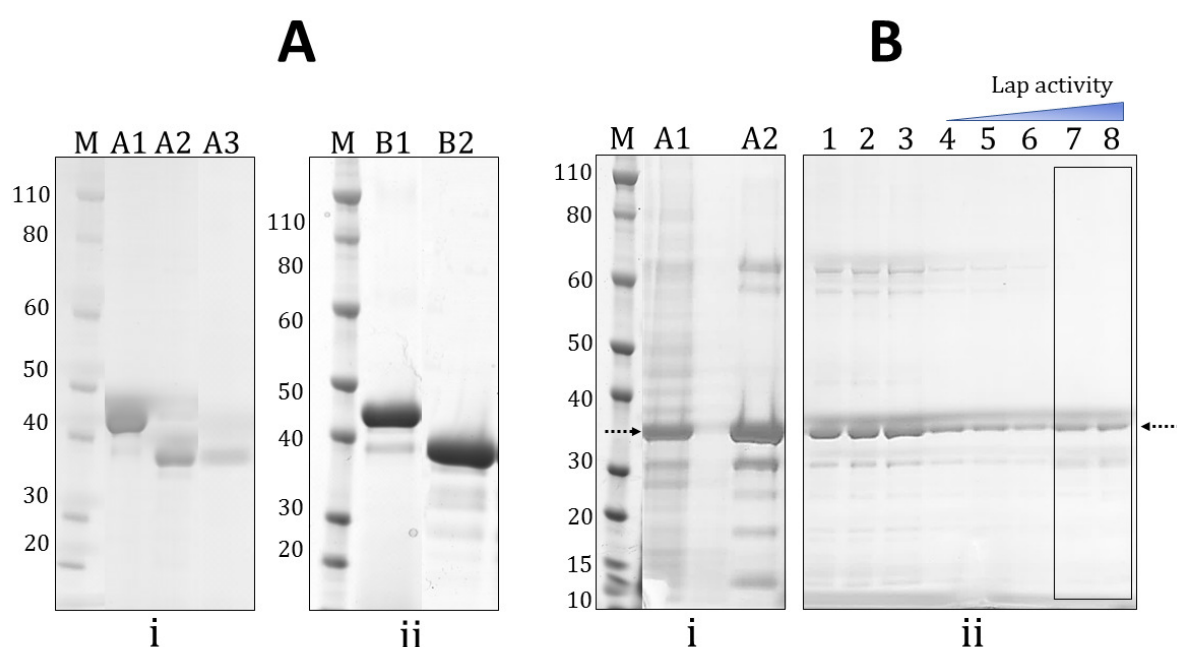
In the case of *wtLapA* and *wtproLapA*, expression levels reached approximately 0.3 g per litre. SDS-PAGE analysis revealed significantly lower expression levels for *wtΔproLapA* in *P.pastoris* supernatant (**Figure 5.2**, A). We speculate that low expression levels were observed, due to the absence of the prosequence in the pJ-Δpro-LapA construct. Alternatively, it may be due to the integration of a lower copy number of LapA genes. However, pJ-Δpro-LapA transformants were found to grow on YPD agar containing 2000 ug/ml Zeocin, which is very unlikely for the *P.pastoris* strain harbouring only a single copy of the recombinant gene. The lack of propeptide in the *wtΔproLapA* expression construct could have yielded mis-folded protein with exposed hydrophobic patches, associating with molecular chaperones, the mechanism by which unfolded proteins are targeted for degradation within the ER (McIver et al., 1995, Yurimoto et al., 2004). Hence, observed expression levels were significantly lowered compared to *wtproLapA*, which contained propeptide in the expression cassette.

*wtproLapA* and *wtLapA* recombinant proteins were obtained from the same expression cassette (pJ-npro-LapA) and were a result of clonal variation within *P.pastoris* transformants. In one of the clones, unresolved processing of expressed recombinant LapA precursor (*wtproLapA*) occurred, yielding mature active *wtLapA*. Most likely, cleavage of propeptide was a post-translational event since genomic PCR-based sequencing within the coding region showed no deletion of propeptide (results not shown). We propose that co-

secreted non-specific proteolytic activity by an endogenous *P.pastoris* peptidase was responsible for the processing of the LapA precursor protein.

N-terminal sequencing of *wt*LapA revealed three additional residues (AVT) preceding the N-terminal tyrosine, originally reported for LapA (Matsushita-Morita et al., 2011). Therefore, an unexplained endopeptidase activity, likely present in the *P.pastoris* supernatant, was different from that of native *A.oryzae*, which degrades the propeptide to N-terminal tyrosine (Y80).

As expected, *P.pastoris* supernatants with secreted *wt*LapA and *wtpro*LapA contained no other contaminant host proteins (**Figure 5.2**, A). This facilitated the subsequent purification of these recombinant proteins in the absence of artificial tags.



**Figure 5.2** Purification of Lap proteins. **A** – purification of *wt*LapA, *wtpro*LapA and *wtΔpro*LapA: i) *P.pastoris* supernatants sampled (10 μl) at 120 h, containing recombinant LapA proteins: A1 – *wtpro*LapA, A2 – *wt*LapA, A3 – *wtΔpro*LapA; ii) purified from supernatant: B1 – *wtpro*LapA, B2 – *wt*LapA; **B** – purification of *ref*LapA (indicated by dotted arrow): i) initial IMAC purification step: A1 – resolubilized *ref*LapA inclusion bodies in 8M urea, A2 – imidazole eluted fraction from IMAC column; ii) subsequent SEC purification of IMAC-eluted fraction and dialysis-refolded protein, each lane corresponds to individual fractions. Active (under blue triangular activity bar) and mostly active (black rectangle) fractions are indicated.

Purification of recombinantly expressed LapA (*refLapA*, *wtproLapA*, *wtΔproLapA*, *wtLapA*) enzymes was achieved as detailed in materials and methods (section 5.1.2.5). A dialysis step was required to liberate *refLapA* from 200 mM imidazole, present in the IMAC-eluted fraction. It is worth noting that only a small portion of refolded *refLapA* was significantly active, following a size exclusion chromatography step (fractions 7 and 8, **Figure 5.2 B(ii)**). These fractions were pooled and utilised in subsequent CD analysis (**Figure 5.3**). A large population of *refLapA* separated as inactive, potentially aggregated protein (fractions 1- 3, **Figure 5.2 B(ii)**) on gel filtration column. These high-MW aggregates were even visible under denaturing SDS-PAGE and reducing buffer conditions. The specific activity of non-aggregated and refolded *refLapA* (fractions 7-8) reached 59.5 U/mg (**Table 5.1**).

Also, repetitive AEX purification cycles were required to liberate *wtproLapA* and *wtLapA* from brown pigment existing in *P.pastoris* supernatant. Elimination of these amber pigments proved very challenging, which was deemed mandatory for crystallisation trials. These two mysterious brown pigments ( $M_r < 2\text{kDa}$ ) were reported previously to associate with other Lap peptidases and were briefly characterised by EPR to be spectroscopically active and contain an Fe(III) ligand (Schalk et al., 1992, Hartley et al., 2009, Aphale and Strohl, 1993, Bennett and Holz, 1997).

### 5.1.3.3 Specific activities of Lap enzymes

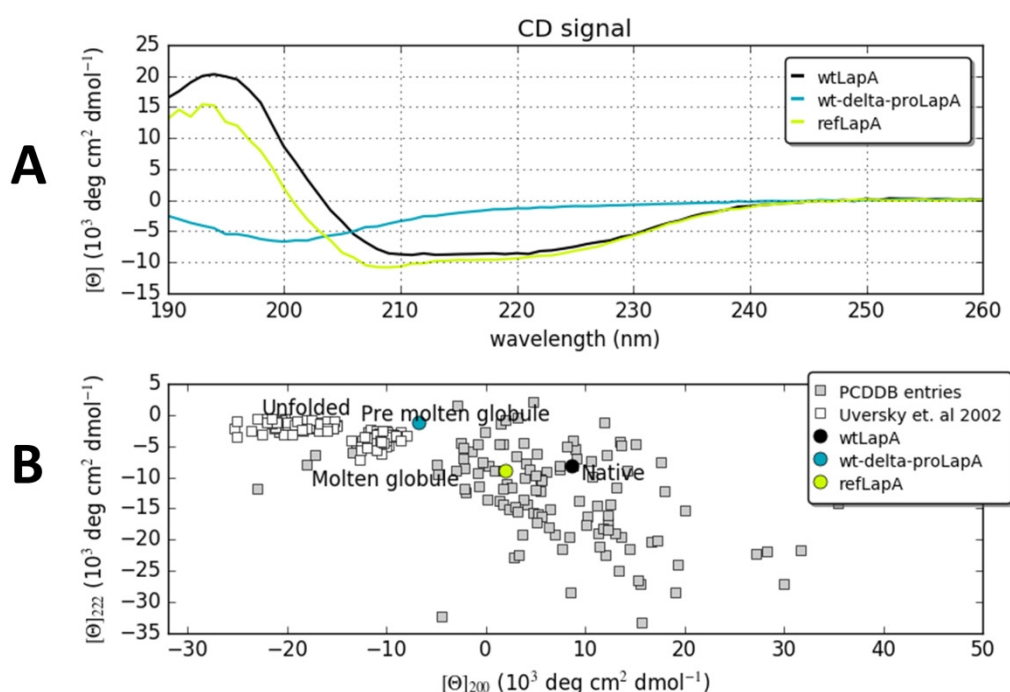
Standard leucine aminopeptidase assay was carried out to determine the differences in specific activity expressed by different LapA proteins, as demonstrated in **Table 5.1**. The importance of the observed specific activity differences is discussed in the following section (5.1.3.4) along with Circular Dichroism data.

Enzyme	Specific activity (U/mg)
<i>wtLapA</i>	140.15 ± 4.8
<i>wtΔproLapA</i>	0.34 ± 0.1
<i>refLapA</i>	59.5 ± 1.1

**Table 5.1** Specific activities of purified LapA proteins. Specific activity is averaged from 3 independent measurements with standard deviation determined.

#### 5.1.3.4 Circular Dichroism (CD) experiments of LapA proteins

LapA proteins were analysed by far-UV CD to estimate their secondary structure content and to study their folded state. The characteristic far-UV spectra of *refLapA*, *wtLapA* and *wtΔproLapA* proteins are shown in **Figure 5.3**.



**Figure 5.3** CD spectra and deduced folded state of LapA proteins. The observed CD signal from *wtLapA*, *wtΔproLapA* and *refLapA* **A**. Mean residue ellipticity -  $\Theta_{222}$  versus  $\Theta_{200}$  plot **B**. allows direct visualisation of the protein folded state, as described by Uversky *et al.* (Uversky, 2002). PCDDDB database entries of proteins with similar spectra are also plotted.

Ellipticities at 208 and 222nm relate to characteristic double minima of the alpha-helical spectrum. *wtLapA* displayed a typical double minimum at 222 and 208 nm and an intensive maximum near 195 nm, representing a protein with high  $\alpha$ -helical secondary structure content (**Figure 5.3**, A). Also, by comparison to existing protein CD databases, LapA was predicted to exist in a native folded state (**Figure 5.3**, B), which is in good agreement with its specific activity data. *refLapA* exhibited almost identical CD spectrum. However, the observed lower mean residue ellipticity at 200nm suggests a minor loss of secondary structure and the prominent close-to- molten globule state of this protein. *refLapA* also showed a higher content of irregular structure compared to *wtLapA*. These observations coincide with a measured >2-fold reduction in specific activity relative to *wtLapA* (**Table**

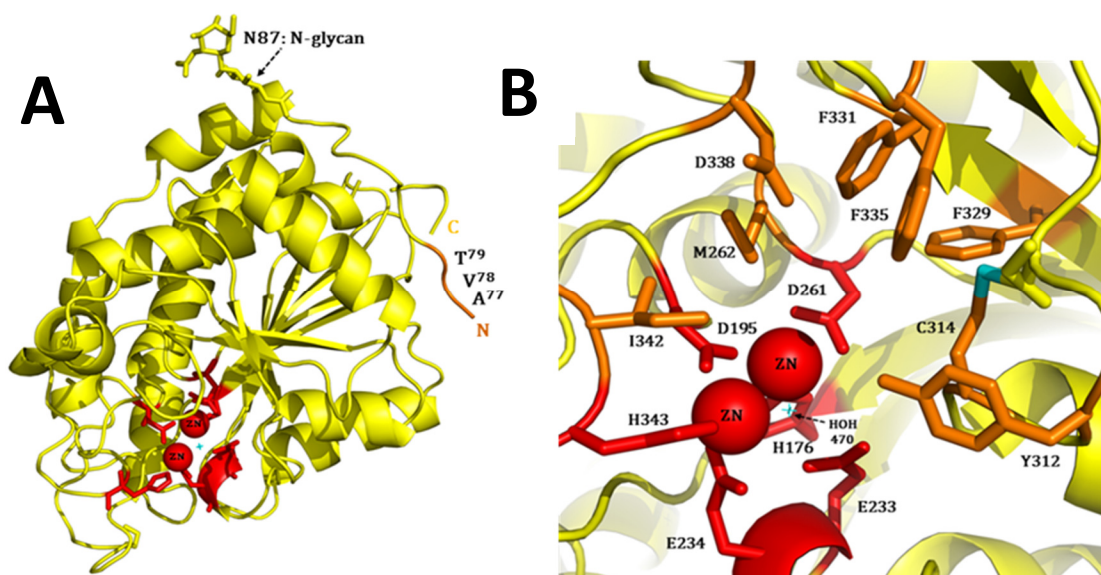
**5.1).** A lower content of secondary structure also was found in recombinant refolded aminopeptidase from *Vibrio proteolyticus* (AAP) compared to natively expressed wild-type enzyme (Hernandez-Moreno et al., 2014). Finally, a spectral minimum at 200 nm and close-to-zero ellipticity near 222 nm suggested a lack of secondary structure in the *wtΔproLapA* protein and a predominant pre-molten globule state, which explains its almost non-existent hydrolytic activity for this recombinant protein (**Table 5.1**). Importantly, this experiment, together with the specific activity data, demonstrates the negative effect of the absence of the propeptide in the expression construct of recombinant LapA in *P.pastoris* (*wtΔproLapA*) and *E.coli* (*refLapA*). Consequently, this reveals a potential role for the LapA propeptide in assisting the mature domain to gain a fully folded, functional conformation for its maximal activity.

Enzyme	Secondary structure content (%)		
	$\alpha$ -helix	$\beta$ -strand	irregular
<i>wtLapA</i>	10	36	37
<i>wtΔproLapA</i>	3	42	62
<i>refLapA</i>	20	27	49

**Table 5.2** Secondary structure prediction via CAPITO, based on the CD data analysis.

### 5.1.3.5 Wild-type LapA structure

The determined monomeric LapA structure (*wtLapA*) forms a classical  $\alpha/\beta$  globular domain. The hydrophobic core is comprised of a twisted 8-stranded  $\beta$ -sheet, sandwiched between  $\alpha$ -helices, representing a typical archaic hydrolase fold (**Figure 5.4, A**). A secondary structure assignment, using the DSSP (Kabsch and Sander, 1983) server, showed that 45%, 16% and 39% of the residues are involved in  $\alpha$ -helices,  $\beta$ -strands and coil structures, respectively. The fairly accessible active site pocket is located on the surface of the protein. The catalytic binding pocket is formed by 6 conserved residues, coordinating two zinc ions, embedded in the loop region close to the core parallel  $\beta_3$  and  $\beta_5$  sheets, and hydrophobic specificity pocket residues, mostly arranged at the C-terminal end of the protein (**Figure 5.4, B**). The two Zn atoms exist in a distorted tetrahedral geometry with an interatomic distance of 3.6 Å. We also identified a nucleophilic water molecule bridging the two zinc atoms. Since both zinc ions significantly decrease the  $pK_a$  of this water molecule (Desmarais et al., 2002), at the pH for optimal activity (8.0 – 9.0), it may exist as an  $OH^-$  ion.

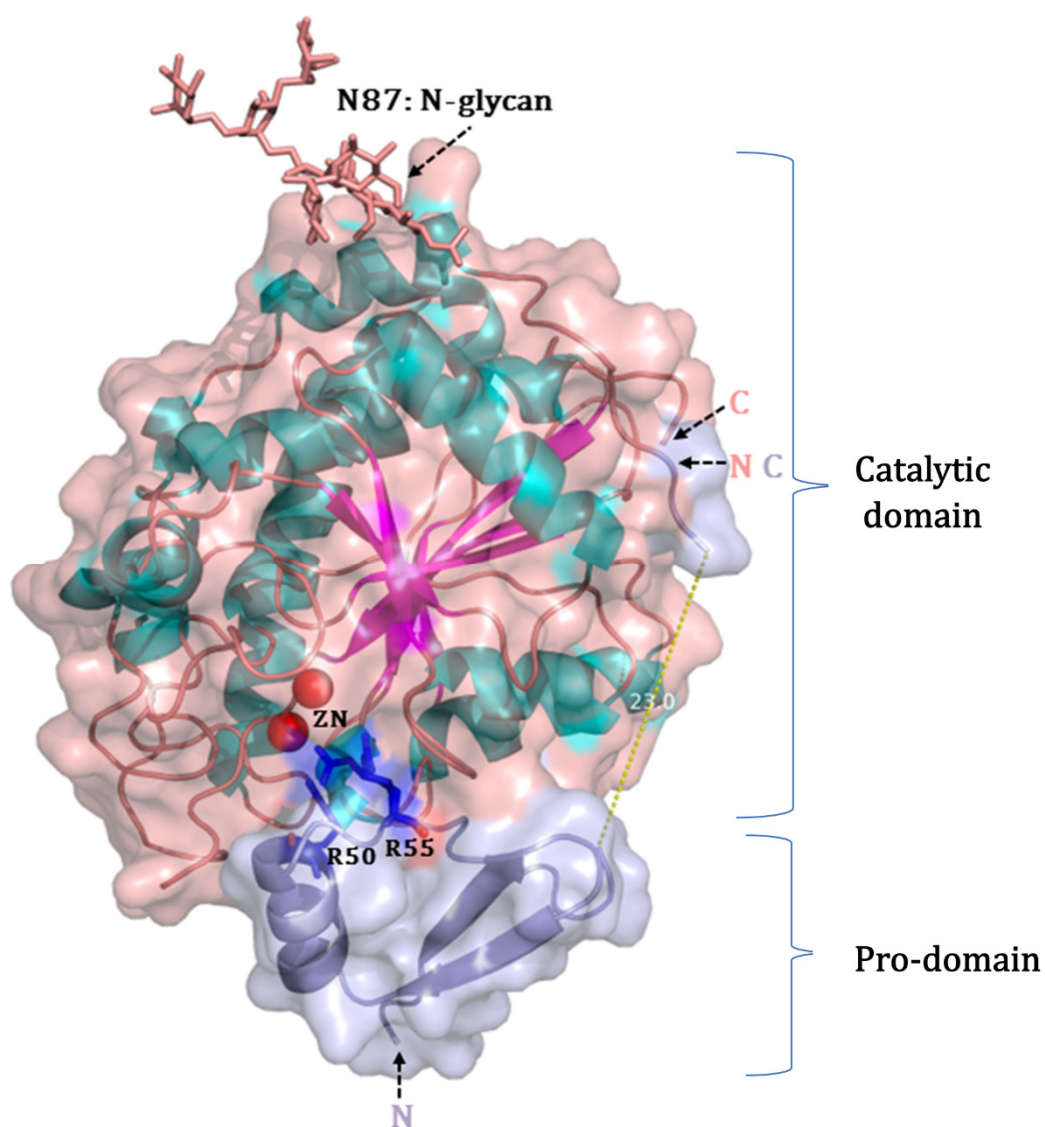


**Figure 5.4** Crystal structure of LapA catalytic domain (*wtLapA*) determined to 1.97Å resolution (A) and a detailed view of the active site (B). Catalytic residues and coordinated ZN atoms are coloured in red; nucleophilic water, HOH470, coordinated between the two ZN ions, indicated by cyan star; conserved hydrophobic binding pocket residues are shown in orange sticks; an essential S-S bond (C310 – C314) is depicted in cyan.

#### 5.1.3.6 LapA precursor structure

We determined the three-dimensional structure of LapA precursor. All residues were well located in the electron density, with the exception of 7 and 11 residues at N- and C-terminal region of the prodomain, respectively, due to inherent flexibility in these regions (not uncommon in crystal structures of proteins). The final model, determined to 1.61 Å resolution, consisted of 359 residues, 517 water molecules, including the catalytic bridging water, a single chloride ion and an ethylene glycol molecule (arising from cryo-protection).

From **Figure 5.5**, it is evident that the propeptide extends from the mature-prodomain junction, follows alongside the catalytic core (residues 66 – 76 are missing) until it enters and occludes the catalytic cleft, using its 3-turn  $\alpha$ -helix motif. In the crystal structure of *wtproLapA*, the propeptide forms a separate domain (residues 19 - 79) (**Figure 5.5**). Its overall structure is highly disordered and flexible: only 23% of total residue content within the propeptide forms secondary structure. The middle part of the propeptide sequence forms a 3-turn  $\alpha$ -helix (residues 43-52) and a short 2-stranded  $\beta$ -sheet (residues 29 - 31 and 39 - 41).



**Figure 5.5** Crystal structure of LapA precursor (*wtproLapA*) determined to 1.61 Å resolution. Cartoon and surface representation of the final protein model: prodomain is coloured in light blue, key arginines (R50', R55') are displayed as sticks in dark blue, catalytic ZN ions are depicted as red spheres, a single glycosylated asparagine (N87) is indicated by stick representation of the sugar moieties. N- and C-termini are indicated and coloured corresponding to the molecular surface of each domain. Also, missing residues (66 – 76; connecting the prodomain and the catalytic domain) are indicated by a dashed yellow line.

### 5.1.3.7 LapA precursor intradomain interactions

The propeptide buries 8.8% of mature domain solvent-accessible surface area (1005 Å<sup>2</sup> of 11474 Å<sup>2</sup>). Moreover, this does not take into account the missing 66 – 76 amino acid region of the propeptide. By *corollary*, the propeptide uses 29.7% of its surface area for the interface with the mature domain (**Table 5.3**).

Interface Summary				
	Mature domain		Propeptide	
<u>Number of atoms</u>				
interface	116	5.1%	102	30.3%
surface	1078	47.4%	241	71.5%
total	2281	100.0%	337	100.0%
<u>Number of residues</u>				
interface	37	12.4%	21	48.8%
surface	241	80.9%	42	97.7%
total	298	100.0%	43	100.0%
<u>Solvent-accessible area, Å</u>				
interface	1005.5	8.8%	1096	29.7%
total	11474.5	100.0%	3684.7	100.0%

**Table 5.3** Summary of the protein surface statistics calculated between mature and prodomains in the crystal structure of LapA precursor. The structure coordinate file was analysed using PDBePISA (Krissinel and Henrick, 2007).

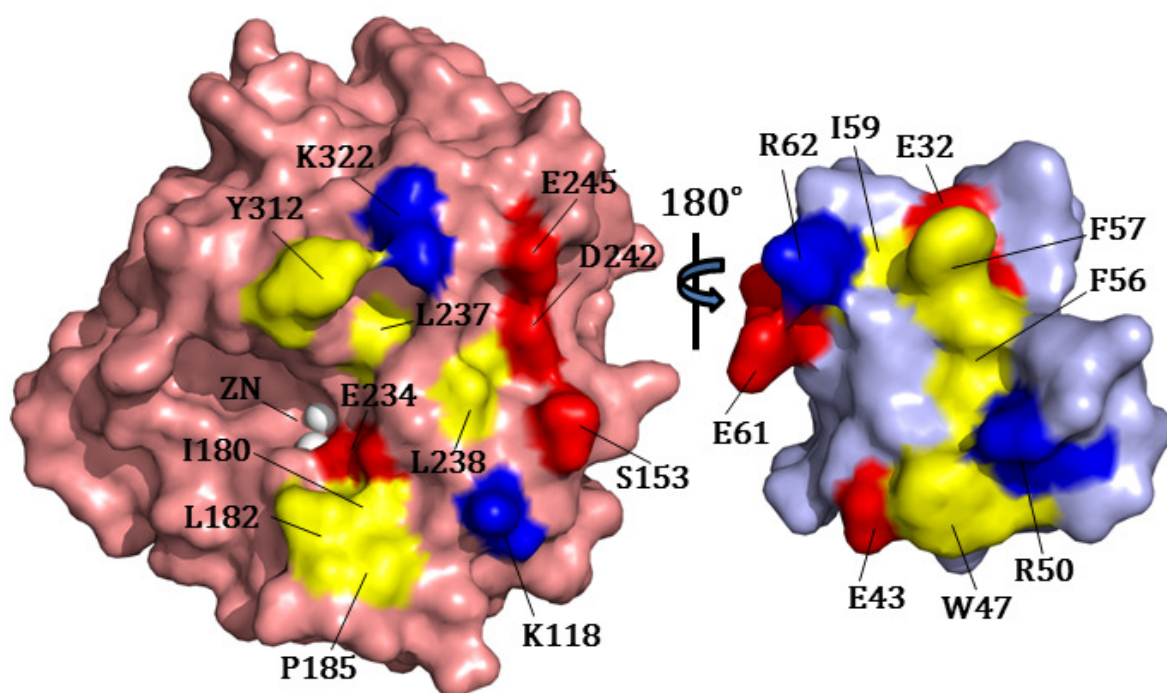
The inter-domain network of interactions involves 37 and 21 (almost a half) of the residues for the catalytic and prodomains, respectively. More specifically, the inter-domain interface consists of 12 short-range (2.2 - 3.3 Å) H-bonds, 3 long-range (3.3 - 3.8 Å) H-bonds, 6 side

chain-side chain salt bridges ( $< 3.3 \text{ \AA}$ ), and at least 10 aliphatic side-chain residues within the propeptide that make tight hydrophobic contacts with the mature domain (**Table 5.4**).

<b>Hydrogen bonds (2.2-3.3 Å)</b>			
	<u>Mature LapA domain</u>	<u>Dist. [Å]</u>	<u>LapA prodomain</u>
1	LYS 118[ NZ ]	3.20	TYR 29[ OH ]
2	LYS 322[ NZ ]	3.07	GLU 32[ OE1]
3	LYS 118[ NZ ]	2.74	GLU 43[ OE2]
4	GLN 241[ NE2]	3.02	ASP 58[ O ]
5	SER 153[ OG ]	2.62	GLU 61[ OE1]
6	ASN 288[ ND2]	3.21	VAL 67[ O ]
7	GLU 234[ O ]	2.84	ARG 50[ NH1]
8	GLU 234[ OE2]	2.99	ARG 50[ NH1]
9	GLU 234[ OE2]	2.86	ARG 50[ NH2]
10	GLY 236[ O ]	3.02	ASP 58[ N ]
11	GLU 245[ OE2]	2.59	ARG 62[ NH2]
12	ASP 242[ OD1]	2.69	ARG 62[ NH2]
<b>Long range H-bonds (<math>&gt; 3.3\text{\AA}</math>)</b>			
1	TYR 116[ OH ]	3.66	GLU 43[ OE2]
2	TYR 117[ OH ]	3.61	LYS 46[ NZ ]
3	GLN 241[ OE1]	3.59	ARG 62[ NH1]
<b>Salt bridges (<math>&lt; 3.3\text{\AA}</math>) side chain-side chain</b>			
1	LYS 322[ NZ ]	3.07	GLU 32[ OE1]

2	LYS 118[ NZ ]	2.74	GLU 43[ OE2]
3	GLU 234[ OE2]	2.99	ARG 50[ NH1]
4	GLU 234[ OE2]	2.86	ARG 50[ NH2]
5	GLU 245[ OE2]	2.59	ARG 62[ NH2]
6	ASP 242[ OD1]	2.69	ARG 62[ NH2]

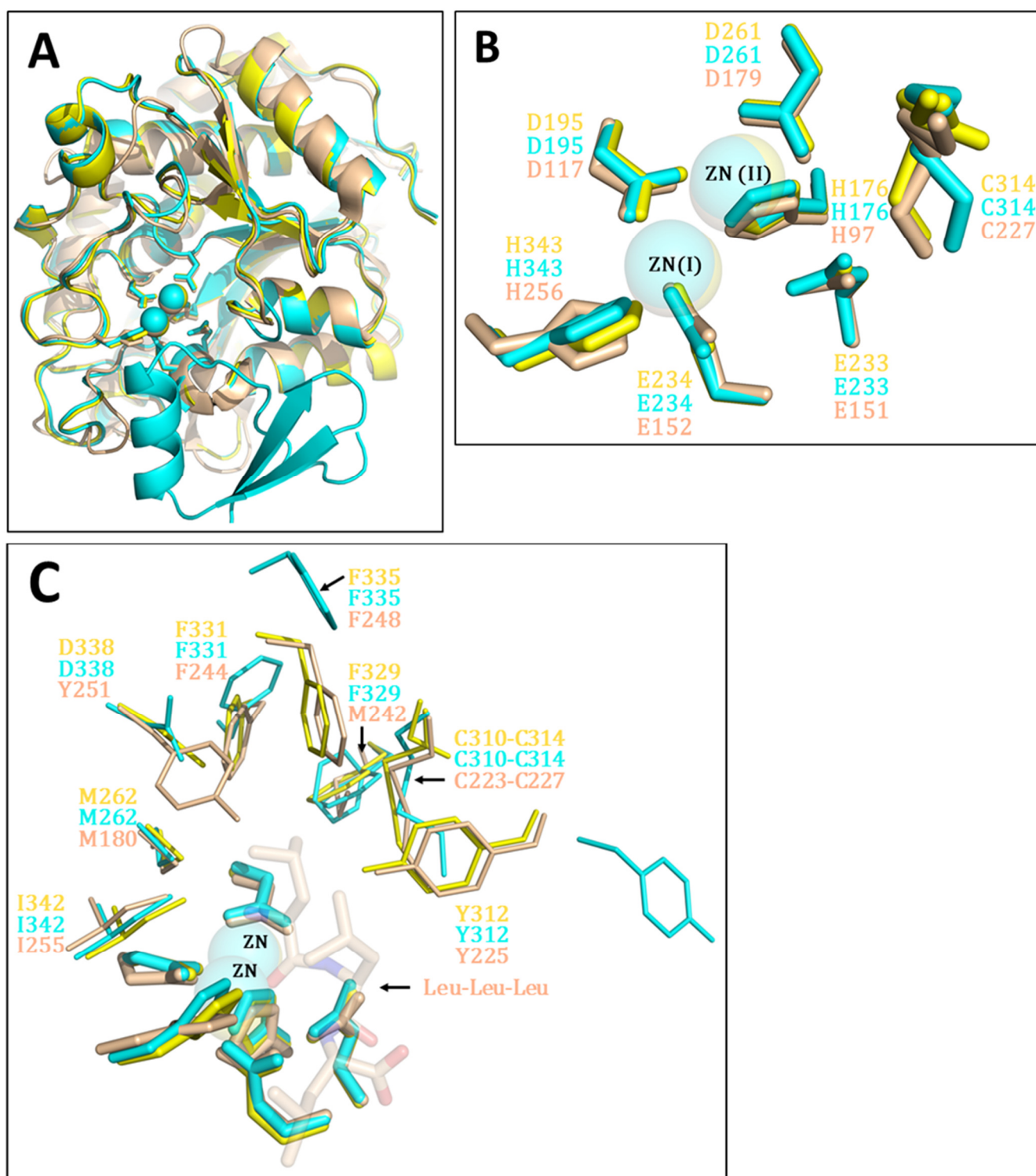
**Table 5.4** Contacts between the catalytic and prodomain. Contacts shaded in grey are visualised in **Figure 5.6**.



**Figure 5.6** Contacts between the catalytic (left) and prodomain (right). The prodomain is rotated 180 degrees to expose the corresponding surface contacts. The domain interface is stabilised mainly by salt bridges (positively charged residues – blue, negatively charged residues – red) and tight hydrophobic interactions (yellow). Dinuclear Zn site is shown as white spheres.

#### 5.1.3.8 Fold comparison of Lap structures

As expected, LapA precursor was found to deviate from mature LapA (*wt*LapA) only by 0.38 RMSD and by 1.23 RMSD from AAP (PDBid: 2IQ6) over an equivalent range of residues, which indicates the high similarity between these protein folds. The active site geometry was found almost identical in all three crystal structures, with highly conserved catalytic residues and dizinc metal cofactors superimposing, almost precisely (**Figure 5.7**). Despite only 34% sequence identity to AAP, *wt*LapA and *wtpro*LapA displayed an identical conserved fold, which is in agreement with proteins having greater conservation in tertiary structures and confirming their classification as M28 family members.

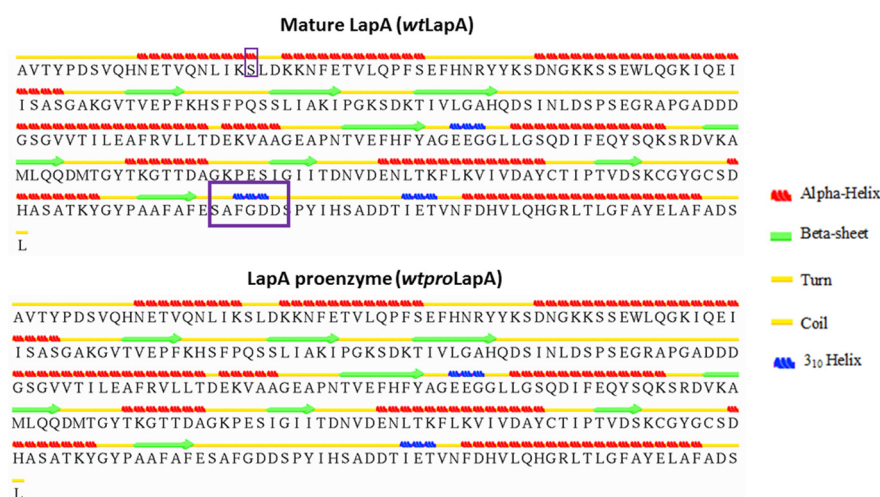


**Figure 5.7** Superimposition of LapA precursor (cyan), mature LapA (yellow) and AAP-LLL (PDBid: 2iq6) (wheat) crystal structures. **A.** – similarity of overall folds, **B.** – superimposition of active sites only, **C.** – superimposition of active sites and hydrophobic binding pocket residues (narrow sticks). Residue labels correspond to general colouring

scheme, ZN atoms and bound Leu-Leu-Leu substrate are shown transparent for clarity. The main chain atom representation is omitted.

### 5.1.3.9 Comparison of LapA precursor (*wtproLapA*) and mature (*wtLapA*) structures

Alignment of the structures for proenzyme and mature protein show high similarity. Structural coordinates of LapA prodomain were omitted in the superpositions to achieve a more accurate representation of the RMSDs between catalytic domains. Superposition of LapA precursor (propeptide omitted) and mature structures via TM-align yielded an RMSD of 0.38, over 301 equivalent residues. Minor differences in secondary structure content were observed between *wtproLapA* and mature enzyme (**Figure 5.8**). Using the STRIDE (Frishman and Argos, 1995) web-based tool, which utilises a combined H-bond and geometrical restraint approach to secondary structure elements (SSE) assignment, SSE were identified for both proteins.



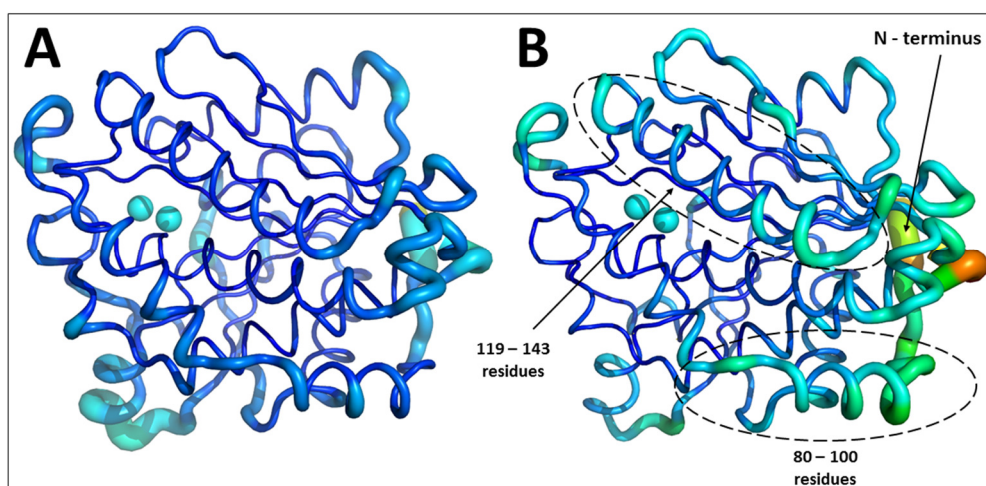
**Figure 5.8** Differences in the secondary structure of LapA proenzyme and mature proteins. Purple boxes indicate the differences in SSE.

Although additional differences in secondary structure elements were found in PyMol representations, upon superimposition of these two forms of LapA enzyme, the more sophisticated STRIDE algorithm only identified an additional, significant 3<sub>10</sub> Helix in *wtLapA*, appearing after cleavage of the propeptide.

#### 5.1.3.9.1 Comparison of overall B-factors

Following cleavage of the N – terminal propeptide, an overall increase in B-factor values was observed within the catalytic domain. Not surprisingly, the most significant differences were apparent at the N – terminal region of the mature LapA protein (80 – 150 residue stretch) (**Figure 5.9, B**). This suggests an improved flexibility of the N – terminal part of the protein, following removal of the covalently linked prodomain. The increased intramolecular flexibility of the LapA mature domain (*w/LapA*) may be essential to perform the catalytic function.

In fact, we determined an increase in thermal stability for the LapA precursor. The thermal melting midpoint value for the proenzyme increased by 10 °C, supporting the improved rigidity and overall thermal stability conferred by the covalently linked propeptide domain (**Table 6.30**). We concluded that, the higher temperature of unfolding is likely enthalpy driven, arising from the increased number of propeptide – mature domain interactions and providing more stability to the whole molecular edifice at elevated temperatures.



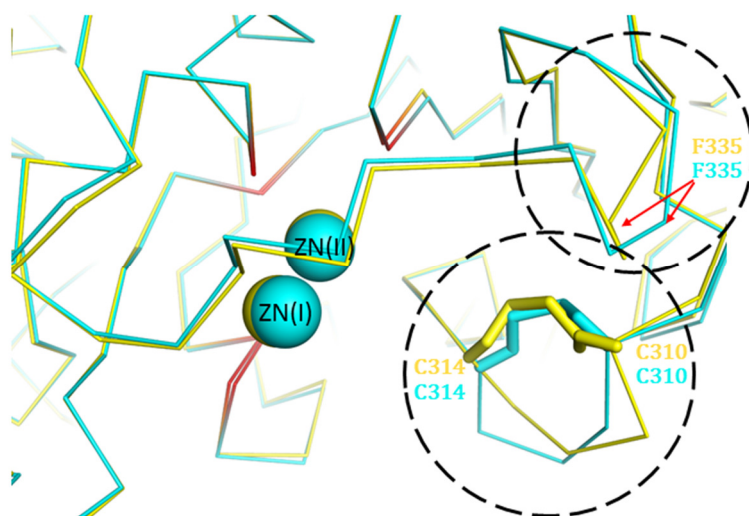
**Figure 5.9** Comparison of residue B-factor distribution in LapA proenzyme and activated protein. Lower to higher B-factor values are coloured from dark blue to red, respectively. **A.** LapA precursor, propeptide omitted for clarity; **B.** mature LapA protein, following cleavage of the propeptide. Residue ranges with the largest changes in B factor are indicated by dashes lines.

#### 5.1.3.9.2 Prodomain induced structural changes

The major structural rearrangements within the LapA catalytic domain, observed in the presence and absence of propeptide, are discussed below.

The presence of propeptide elicits a number of structural rearrangements in the catalytic cleft. Apart from the conformational changes in the hydrophobic binding pocket residues (discussed later), the architecture of the active site was not altered significantly. However, very subtle changes in the interatomic distances of the active site residues were detected. First, the distance between the two catalytic zinc atoms is longer (3.6 Å) in the mature enzyme compared to LapA precursor (3.4 Å). This stems from a minor change in coordination distances with catalytic residue ligands in LapA precursor.

The only disulphide bridge in all of these structures is located within the loop adjacent to the catalytic site (C310-C314). The conformation of this loop is slightly altered in LapA precursor (**Figure 5.10**).



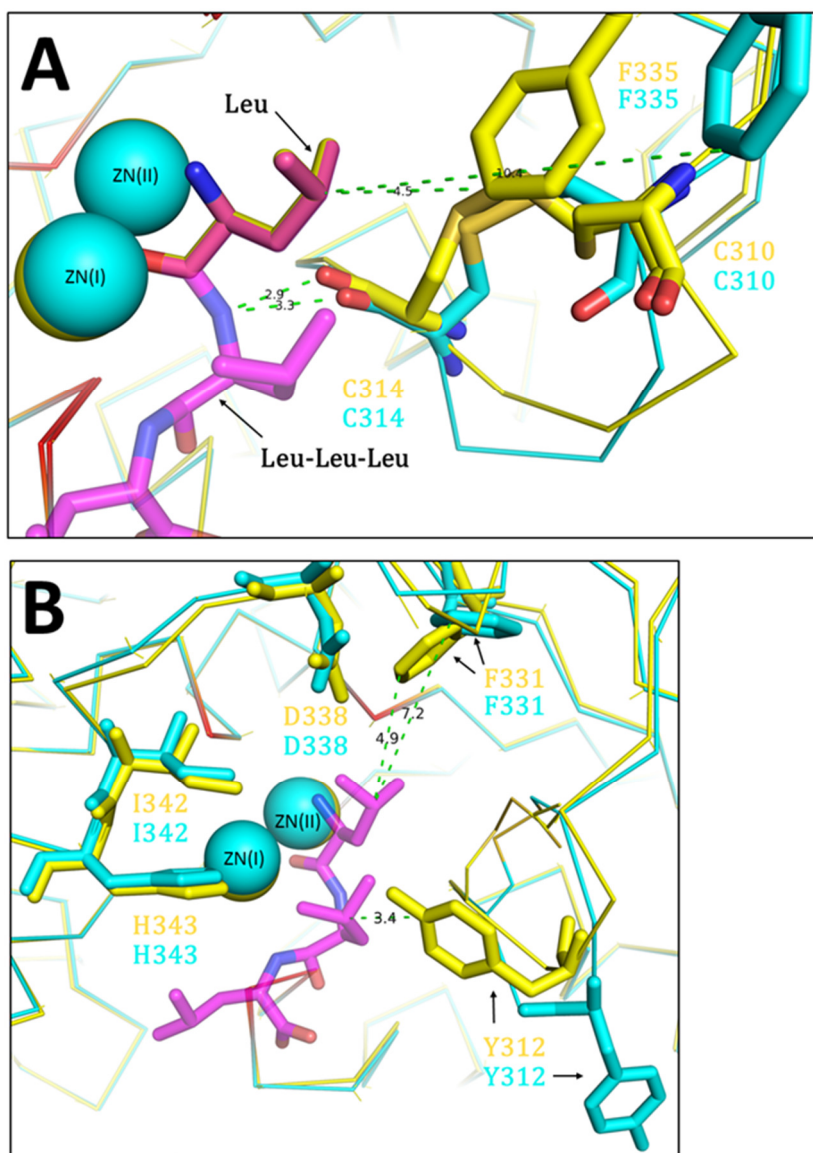
**Figure 5.10** Superimposition of *wtproLapA* (cyan) and mature *wtLapA* (yellow)  $\text{C}\alpha$  traces. Dashed circles indicate the affected loop regions. C310 and C314 residues, forming the disulphide bridge, are shown as sticks. Changes in the location of F335 ( $\text{C}\alpha$  atoms) are indicated by red arrows. Active site residues are coloured in red. Residue labels are coloured in agreement with their respective molecules.

Given the very high conservation in active site geometries for dimetallic zinc peptidases, the crystal structure of AAP in complexation with the *bona fide* LLL (L-leucyl-L-leucyl-L-leucine) substrate permits direct evaluation of propeptide-induced conformational changes within the substrate binding pocket and identification of the key residues involved. This conservation has been consolidated by our recent structure solution of *wt*LapA-Leu complex. With the aid of this structure, we demonstrate that the orientation of a single P<sub>1</sub> leucine in the *wt*LapA-Leu complex is identical to one of the leucine residues of the LLL tripeptide in AAP. The structural coordinates of the P<sub>1</sub> substrate moieties are directly superimposable (**Figure 5.11, A**).

Following cleavage of the propeptide, a movement in both C310-314s and F335s loops is apparent. The former appears to be a genuine loop conformational change, resulting in shortening the distance to the carbonyl oxygen of C314, which is involved in making direct contact with the amide nitrogen of the Leucine substrate (P<sub>1</sub>'). This H-bond shortening is subtle (from 3.3 Å to 2.9 Å), but sufficiently significant to strengthen this polar contact (**Figure 5.11, A**). Also, it is possible that the H-bond angle of this contact becomes more favourable in the mature *wt*LapA conformation. Therefore, the increased bond distance in LapA precursor might contribute to an impaired ability to bind substrate. Residue Cys314 has been identified as one of the key residues in the hydrophobic substrate binding pocket in Lap1 from *A. sojiae* (Huang et al., 2015). And, residue F335 is directly involved in substrate binding (making a hydrophobic contact with the sidechain of P<sub>1</sub> moiety). The transition of its side chain toward the active site, after cleavage of propeptide, might elicit a more catalytically competent conformation for the mature domain. In fact, from a more detailed analysis, it is apparent that the vdw distance between the side chains of F335 and the P<sub>1</sub> leucine moiety falls into the favourable region (4.5 Å), following the cleavage of propeptide (**Figure 5.11, A**).

Additionally, at least two more residues were significantly affected by the movement of the C310-314s and F335s loops. An approximate reduction of > 2Å in the distance to leucyl hydrophobic side chain (P<sub>1</sub>) was observed for the aromatic ring of F331 in *wt*LapA structure (**Figure 5.11, B**). Also, quite remarkably, the side chain of Y312 is flipped, following the conformational change in the C310-314s loop. Finally, the side chains of other residues (D338, I342, H343), participating in stabilising the substrate (LLL), also move slightly deeper into the pocket following this transition.

A special consideration has to be made for the conformational change of Tyr312 in the LapA precursor structure. An conserved residue, Tyr225, in an equivalent location is found in the AAP protein. In the initial model of peptide hydrolysis, Tyr225 was proposed to interact with the N-terminus of the incoming substrate (Chen et al., 1997). However, this concept was later challenged by the analysis of AAP-bestatin and AAP-LPA complexes. It was concluded that Tyr225 may not be involved mechanistically, but it provides yet another stabilising hydrophobic contact (Stamper et al., 2004). Additionally, the phenolic oxygen of Tyr225 shares a H-bond with BES, Tris, LeuP and IDH. In contrast, we demonstrate that the N-terminus of the P1' leucyl moiety is too distant to form a H-bond with the hydroxide of the Tyr312 side chain, but, rather it could make a favourable hydrophobic contact with the side chain of the P1' leucyl moiety (**Figure 5.11, B**). In any scenario, this highlights the importance of the Tyr312 side chain in stabilising the hydrophobic substrate, which is impaired in the flipped orientation of Tyr312 seen in the LapA precursor.

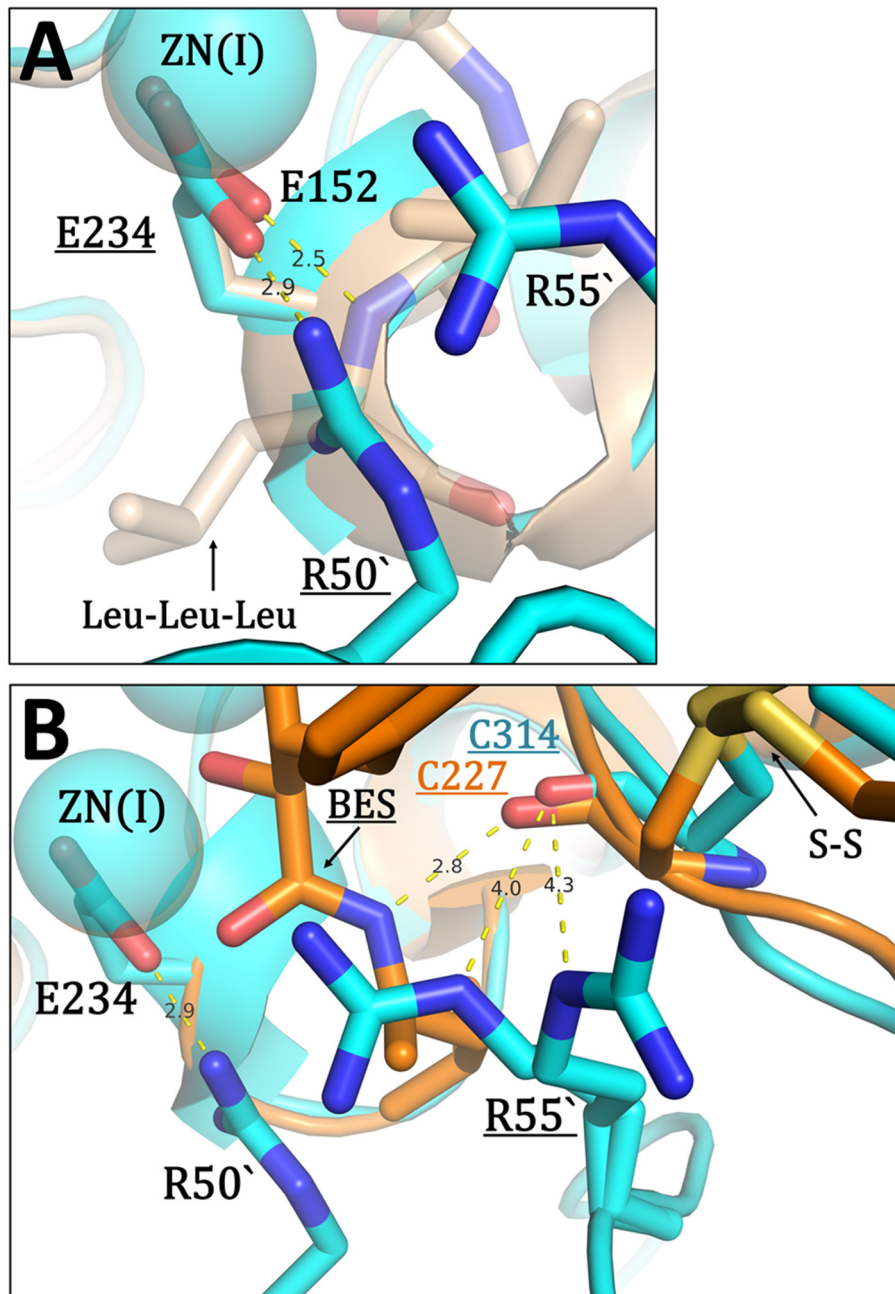


**Figure 5.11** Detailed analysis of the Leu substrate binding modes in *wproLapA* (cyan) and mature *wLapA* (yellow) structures. Active site residues are coloured red. Residue labels are coloured in agreement with their respective molecules. **A.** Leu (yellow) – leucine ligand observed in the crystal structure of *wLapA*-Leu complex, sitting in the identical position to Leu-Leu-Leu (magenta) – leucine substrate from the complex with AAP (PDBid: 2IQ6); **B.** Side chains of F329 and F335 are omitted for clarity.

Differences between the *wtp*roLapA and *w*tLapA structures indicate key structural rearrangements within the catalytic domain upon activation. Alternative side chain conformations, of the key conserved substrate binding residues, were observed in the activated structure of *w*tLapA, as a consequence of C310-314s and F335s loop movements. Several conserved hydrophobic binding pocket residues were seen to make closer contacts with the leucyl substrate, which are absent in the inhibited structure of the LapA precursor. However, these differences (in the *w*tLapA structure) were only associated with substrate binding residues and no major C $\alpha$  backbone movements were observed in other parts of protein, following the transition to mature protein.

#### **5.1.3.10 Elucidation of the mechanism of inhibition by propeptide**

The crystal structure of LapA proenzyme enabled elucidation of the core inhibitory interactions exerted by the propeptide domain. A key residue in the propeptide domain is Arg50, which forms a tight 2.9 Å hydrogen bond with OE1 of the catalytic Glu234 in the main chain (**Figure 5.12, A**). This interaction directly mimics the analogous H-bond (2.5 Å) of Glu152 OE1 and the amide group of P3' LLL ligand – one of three essential H-bonds in forming the AAP-LLL pre-transition state complex (Kumar et al., 2007). In the novel proposed peptide hydrolysis mechanism for AAP, where a nucleophile was not delivered by a metal ion (Kumar et al., 2007), Glu152 was the most likely candidate to deprotonate a nucleophilic water molecule, so that the released metal-bound hydroxide could attack the activated scissile carbonyl carbon of the peptide substrate, resulting in formation of a *gem*-diolate reaction intermediate and, ultimately, returning the proton to the nascent amino terminus of the substrate. While our data does not directly indicate a role for Glu234 in catalysis, it does show that inhibition by the propeptide is via blocking of one of the principal H-bonding sites for substrate binding (**Figure 5.12, A**).



**Figure 5.12** Mechanism of propeptide inhibition in LapA precursor (*wtproLapA*). Superimposition of three crystal structures: AAP with Leu-Leu-Leu ligand bound (wheat), AAP bound aminopeptidase inhibitor – bestatin BES (orange) and LapA precursor (cyan). Catalytic Zn(II) ions correspond to colouring of polypeptide chains. Key propeptide arginines (R50', R55'), **A.** Leu-Leu-Leu substrate complex with AAP (PDBid: 2IQ6), **B.** bestatin (BES) substrate complex with AAP (PDBid: 1TXR), catalytic glutamates – E234 (*wtproLapA*) and E152 (AAP-LLL, equivalent disulphide forming C314 (*wtproLapA*) and C227 (AAP) are labelled.

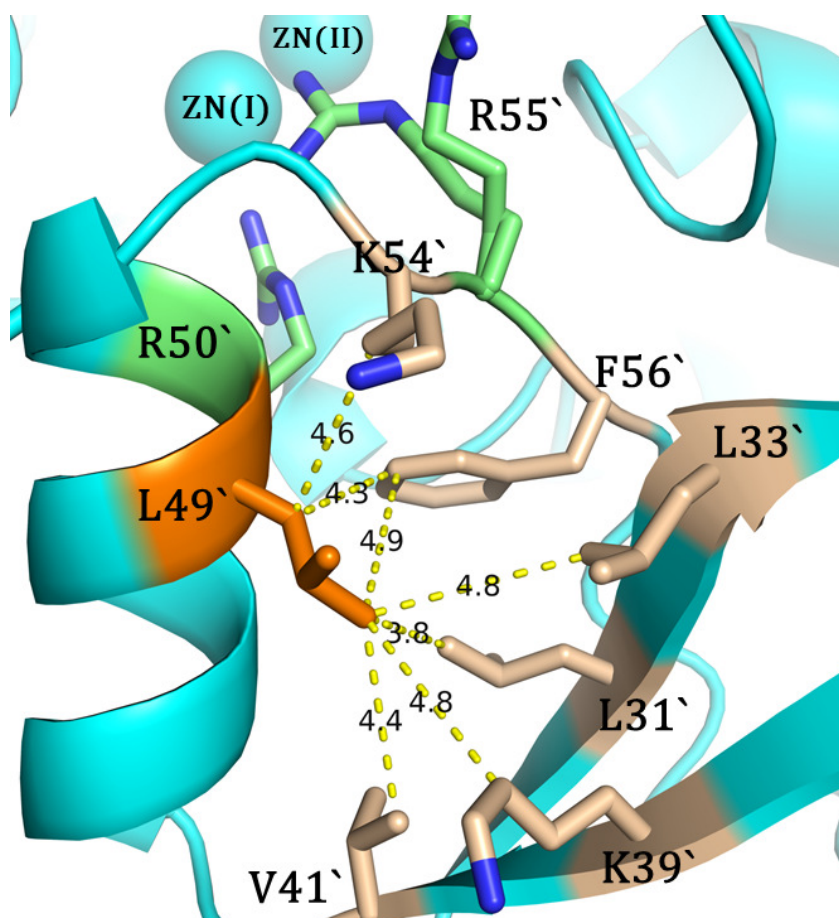
The mechanism of Glu234 inhibition by recruitment of the Arg50 side chain in the LapA prodomain provides additional support to the newly proposed carboxylate-mediated nucleophile delivery hypothesis presented by Kumar et al (2007). It is very likely that Glu234, rather than Glu233 (Glu151 in AAP), is more mechanistically important in the initiation step of the catalysis (deprotonation of the nucleophilic water). Therefore, from an evolutionary standpoint, Glu234 likely was preferentially targeted to achieve complete inhibition by LapA propeptide.

In the case where Glu233 acts as a proton acceptor, from the bridging water species, the dangling oxygen of Glu234 was proposed to stabilise the activated hydroxide ion, or the protonated Glu233 carboxylate side chain, by participating in the H-bond, in either case (Desmarais et al., 2006). The Arg50 – Glu234 interaction had a significant effect on the coordination of this residue to the Zn1 atom. The Zn – O distance of Glu234 OE1 (dangling oxygen) atom is 2.64 Å in the mature LapA structure, while in the propeptide-inhibited structure this increases to 2.90 Å due to a strong electrostatic interaction with both the NH1 and NH2 nitrogens delivered by Arg50. However, the lengthened Zn-O distance does not alter the distance of the Glu234 dangling oxygen to the reactive OH<sup>-</sup> species in the LapA precursor structure. The first coordination sphere for Glu234 OE2 was less affected with a marginal change in the Zn – O distance from 2.03 to 2.07 Å.

Arg55 also contributes to the inhibitory nature of LapA propeptide by providing a flexible terminal guanidinium moiety, which (based on the electron density maps) we found to have three distinct orientations. Each of the three nitrogens in the guanidinium moiety can potentially form favourable van der Waals contact with the carbonyl oxygen of the adjacent Cys314 (**Figure 5.12, B**). This oxygen is involved in one of three key H-bond contacts and the only H-bond present in the AAP complex with the LLL substrate (PDBid: 2iq6) and aminopeptidase inhibitor – bestatin (PDBid: 1txr), respectively. We propose that, in a similar fashion, the NE atom of the Arg55 mobile side chain is mimicking an essential hydrogen bond contact between the backbone carbonyl oxygen of Cys227 and the N1 atom of bestatin (BES) (**Figure 5.12, B**). Also, partial occupancy of Arg55 in our electron density maps suggests dynamic behaviour of the side chain of this residue, possibly distorting the dinuclear ZN site by intervening with the terminal nitrogen on the guanidinium moiety. In this respect, Arg50 and likely together with Arg55, prevents binding and initiation of substrate hydrolysis in LapA. Thus, we propose two additional key inhibitory propeptide residues, Arg50 and Arg55, in LapA (AoLapA) and Lap1 (AsLap1) whereby their long side chains enter the catalytic cleft and block activation (**Figure 5.12**).

Also, in the homologous (57% identity to AoLapA) aminopeptidase from *Aspergillus fumigatus*, Arg50 is substituted for another positively charged residue, Lys60, and Arg55 is replaced by a less structurally similar residue, Asn66 (**Figure 5.1**). Interestingly, Leu59 was found in the equivalent position to Arg50 in the homologous aminopeptidase (AbLap) from *Agaricus bisporus* (43% identity to LapA). However, there is an Arg at position 60, next to Leu59, in this homolog, suggesting it could be orientated to the identical location (to Arg50 in AoLapA) in its tertiary structure (as yet, undetermined). Importantly, Arg55 is not only recruited by LapA and Lap1, but also by the latter AbLap aminopeptidase. Taken together, these observations support a plausible role for such a positively charged side chain in mimicking the substrate-like interactions in the active site of these aminopeptidases. Therefore, we propose that LapA, Lap1 and some other Lap precursors within the M28 family of homologous extracellular aminopeptidases recruit positively charged side chain moieties, which, by making key contacts with conserved active site and substrate binding residues, resemble the interactions of known synthetic inhibitors (bestatin) and natural ligands (leucine). This emerged as a potential generic mechanism utilised by the precursors of eukaryotic fungal leucine aminopeptidases.

Lastly, another interesting residue was identified in the LapA prodomain sequence, Leu49, which was found to be highly conserved amongst the aligned species (**Figure 5.1**). In the crystal structure of LapA precursor, Leu49 was found favourably interacting with at least six residues from distant parts of the prodomain (**Figure 5.13**). These multiple hydrophobic interfaces between the aliphatic side chain of Leu49 and side chains of interconnecting residues could be essential for stabilising the overall fold of the prodomain. Therefore, we conclude that the strategically placed Leu49 is making contacts with residues flanking both ends of the prodomain helix that occludes the active site and holds the correct orientation of adjacent loops connecting to/from this helix, stabilising the overall fold of the prodomain. The strategic orientation of the prodomain helix could be essential for steric occlusion of the active site cleft, also assisting the crucial orientation of the key inhibitory arginines (Arg50 and Arg55).

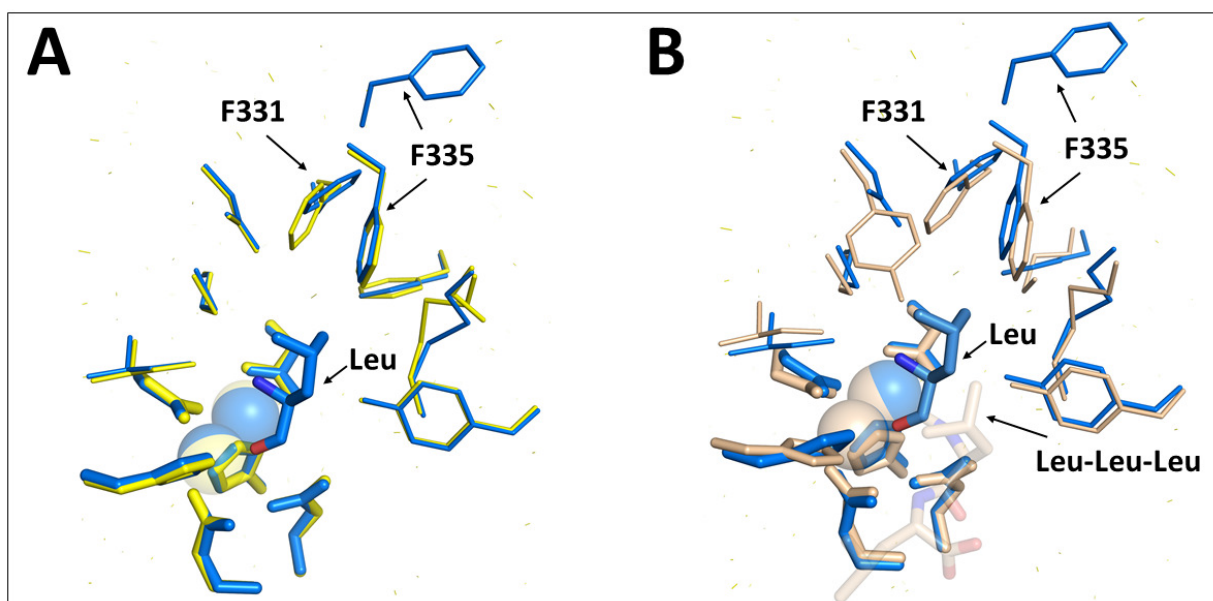


**Figure 5.13** Multiple hydrophobic contacts of prodomain Leu49' (orange). Target residues making hydrophobic contacts with Leu49' are coloured in wheat. Distances to Leu49' are indicated in Angstroms. Key inhibitory prodomain arginines (R50', R55') are shown in lime.

## 5.2 Crystal structure of LapA with natural substrate – Leucine (*wtLapA*-Leu)

Crystal structure of the recombinant mature LapA (*wtLapA*) in complex with a single leucyl moiety was determined to 1.9 Å resolution. The crystallisation process involved briefly soaking the previously formed *wtLapA* crystals in leucine tripeptide (Leu-Leu-Leu). However, only a single leucine could be reliably traced in the resulting electron density map. Most likely, the enzyme exhibited hydrolytic activity in a crystalline state, cleaving the scissile bond between P<sub>1</sub> and P<sub>1</sub>' residues of the Leu-Leu-Leu substrate.

The coordination geometry of the leucyl moiety was identical to that observed in the analogous crystal structure of AAP-LLL (PDBid: 2iq6). This observation reflects the conservation of the; active site residues, substrate binding pocket geometries and reaction mechanism for these homologous leucine aminopeptidases, which display only 34% sequence identity (**Figure 5.14**).



**Figure 5.14** Comparison of the substrate binding pockets in *wtLapA*-Leu and AAP-LLL crystal structures. **A.** Superimposition of the resting LapA molecule (*wtLapA*) and leucine-bound state. **B.** Superimposition of *wtLapA*-Leu and AAP-LLL coordinates. Colour code: *wtLapA* – yellow, *wtLapA*-Leu – marine, AAP-LLL – wheat; Narrow sticks – HBP (hydrophobic binding pocket) residues; wide sticks – catalytic residues; ZN (I,II) are depicted as spheres. Leu and Leu-Leu-Leu substrates are shown as solid and transparent sticks, respectively.

The distance between the two Zn atoms in *wt*LapA-Leu structure is 3.3 Å, which is slightly less than that observed in the resting LapA molecule (3.6 Å).

The coordination geometry of the P<sub>1</sub> leucine residue was directly superimposable with the P<sub>1</sub> moiety of the Leu-Leu-Leu ligand observed in the AAP-LLL crystal structure (PDBid: 2iq6) (**Figure 5.14**, B). The isobutyl sidechain of the P<sub>1</sub> ligand interacts with an expansive, hydrophobic S<sub>1</sub> subsite, formed by the intervening sidechains of residues Y312, C314, F335, F329, F331, M262 and I342, which all are within 5 Å of the ligand sidechain. Accommodation of the leucine ligand elicited the conformational changes of the key hydrophobic residues in the HBP pocket. Mainly, the phenyl rings of F331 and F335 displayed altered conformations, whereby the distance to the P<sub>1</sub> sidechain exceeded 5 Å. It is possible that we have trapped a transient state conformation of *wt*LapA in the substrate hydrolysis mechanism, since no equivalent sidechain movements were apparent in the AAP-LLL complex (**Figure 5.14**, B) or the resting *wt*LapA molecule. The hydrophobic isobutyl sidechain of the P<sub>1</sub> leucine could result in the observed movement of F331 and F335 phenyl rings (as a result of substrate hydrolysis) leading to temporarily induced conformations and product release. Interestingly, the movement of the equivalent residues is seen in the propeptide-inhibited LapA structure, which may be linked to the increased flexibility of this HBP region during catalysis. Leucine is known to be a poor, purely competitive inhibitor of AAP (Kumar et al., 2007). Thus, the determined substrate-inhibited state of LapA might mirror the conformational changes already present in natively inhibited LapA precursor. This transient state is the first of its type to be observed in the family of M28 aminopeptidases.

Ligand	Hydrogen bonds		Non-bonded interactions	
	Interacting partners	Distance (Å)	Interacting partners	Distance (Å)
Leucine	N --- Met262 SD	3.40	O --- ZN1	2.20
	N --- Asp195 OD1	3.32	O --- ZN2	1.87
	N --- Asp195 OD2	3.37		
	N --- Asp261 OD2	3.28		
	N --- Asp261 OD1	3.16		
	O --- His176 NE2	3.15		

**Table 5.5** Hydrogen bond and non-bonded distances between polar groups of *w*LapA and leucine ligand, in the complexed state.

### 5.3 Discussion

LapA from *A. oryzae* was recombinantly produced as proenzyme (approx. molecular weight 39 kDa). The mature enzyme has a molecular weight of 32.5 kDa. The final molecular weight is dependant on the extent of N-linked glycosylation at Asn87. The enzyme is not known to be capable of autoprocessing. The expression of precursor proteins is well established in other metalloenzymes, such as subtilisin, thermolysin and homologous aminopeptidases from *Aeromonas caviae* T-64 and *Vibrio proteolyticus* (Ikemura et al., 1987, Gao et al., 2010, Nirasawa et al., 1999, Bzymek et al., 2004). Usually, these are extracellularly secreted enzymes, where propeptide acts as a competitive inhibitor, including subtilisin E (Hu et al., 1996), carboxypeptidase A (Segundo et al., 1982), a metalloprotease from *Brevibacillus brevis* (Serkina et al., 1999). This could be an evolutionary adaptation to prevent non-specific intracellular proteolysis. The catalytic nature of LapA can induce toxic affects to the host cell, therefore, most probably the activation and processing of the LapA precursor occurs following secretion into the extracellular environment.

A more detailed understanding of the molecular action of these propeptide domains is required to ascertain the role of their cognate active domains in folding processes. Also, the design of more specific inhibitors for dimetallal proteins can be facilitated by mirroring the pre-determined interactions with these native prodomains.

#### 5.3.1.1 LapA protein purification and crystallisation

Two major difficulties to crystallising recombinant LapA proteins were (1) potential heterogeneity of protein populations presented by native host glycosylation and (2) elimination of brown pigments associated with recombinant LapA. The former was not so

problematic, as only Asn87 appeared glycosylated and crystals were obtained in 24 - 72 hours. However, long branched glycan chains can have a major effect on crystal packing and, we could expect improved molecular packing and better resolution quality data in the absence of glycosylation. On the other hand, liberating LapA proteins from bound brown pigment was a major challenge. Previous efforts to obtain crystals of Lap proteins suggested that the presence of this impurity could hinder crystallisation. Successive removal of these two pigments on AEX column gave rise to the rapid formation of LapA crystals. It is worth noting that, in later crystallisation attempts of LapA precursor, the enzyme preparation still contained moderate levels of brown pigments but this did not hinder single crystal formation.

#### **5.3.1.2 LapA precursor**

It is interesting to note, that unobserved electron density for C-terminal prodomain residues is a result of mobility, which is likely required for better accessibility to activating an endopeptidase involved in maturation of LapA precursor. This observation concurred with the unexplained N-terminal processing of LapA precursor in the *P.pastoris* supernatant. Specifically, the end of the missing C-terminal propeptide region (A<sup>77</sup>V<sup>78</sup>T<sup>79</sup>) is the beginning of the verified N-terminus of the processed precursor (*wtLapA*). Therefore, we speculate that LapA precursor was processed by a co-expressed, non-specific host endopeptidase. This would seem to be the only viable explanation since, in the other clones, the vast majority of recombinant protein was found in the form of the precursor in the culture supernatant (**Figure 4.21**).

#### **5.3.1.3 Mode of inhibition by LapA propeptide**

Since the active site geometry is almost identical in the crystal structures analysed, AAP complexed with Leu-Leu-Leu (PDBid: 2IQ6) and bestatin (PDBid: 1TXR) can be used as a guide to understanding the mechanism of LapA inhibition by propeptide and its implications for substrate binding. Therefore, superimposition of LapA precursor and AAP-LLL crystal structures enabled us to determine the mode of inhibition by propeptide and decipher the key residues involved.

Given the conservation of the active site composition and geometry between LapA and AAP proteins, structure coordinates of LLL (substrate) and BES (inhibitor) ligands from AAP complex solutions could be incorporated into the active site of LapA to study the mimicry by prodomain residues. Also, we have solved the structure of *wtLapA* enzymes with a single bound leucyl moiety, which demonstrated binding of the leucyl moiety in the exact orientation to the P1' ligand in the AAP-LLL complex (**Figure 5.11, A**).

Together, with the observed alterations of side chain conformation of the hydrophobic binding pocket residues in LapA precursor (**Figure 5.11**), a model of LapA inhibition can be proposed. In addition to sterically occluding the active site cleft, the propeptide acts by neutralising the key Glu234 and Tyr312 residues, thereby preventing any possible incidence of peptide hydrolysis. Therefore, the LapA prodomain utilises an obvious mimicry of binding the substrate (as seen in known synthetic inhibitor binding) as a mode of inhibiting the catalytic domain.

In conclusion, our LapA precursor structure provides a motive to revisit the mechanism of catalysis in these dinuclear enzymes and, also to design more effective inhibitors for dimetallic peptidases aimed at the conserved Glu234 residue.

#### **5.3.1.4 LapA propeptide – role of intramolecular chaperone (IMC)**

In this study, the role of propeptide in the expression construct of LapA precursor was investigated, using a *Pichia pastoris* recombinant expression platform. Firstly, due to the propeptide deletion *wtΔproLapA*, recombinant LapA was secreted at significantly reduced levels compared to full LapA precursor (*wtproLapA*). Lack of the propeptide in *wtΔproLapA* expression cassette could have yielded inappropriately folded protein with exposed hydrophobic patches, associating with molecular chaperones, the mechanism by which unfolded proteins are targeted for degradation within the ER (McIver et al., 1995, Yurimoto et al., 2004). Hence, the observed expression levels of *wtΔproLapA* protein were significantly reduced compared to *wtproLapA*, which contained prodomain in the expression construct; in turn escaping the degradation mechanism. Secondly, the CD spectrum of *wtΔproLapA* exhibited a lack of secondary structure and showed a predominant pre-molten globule state, while LapA was able to acquire a native fold in the presence of propeptide. This experiment alone demonstrated the requirement of prosequence, which may act as a chaperone in an intramolecular manner.

Also, in the *in vitro* refolding study, the N-terminally His-tagged mature LapA was unable to regain its native fold and exhibited decreased specific activity levels. The ability of *refLapA* to refold from *E.coli* inclusion bodies was not complete, as demonstrated by specific activity measurements and CD experiments. The specific activity of *refLapA* was only ~43% of that observed for *wtLapA* and it was predicted to exist in a molten globule state, according to CD analysis. This implies that absence of propeptide, in the expression construct of *refLapA*, gives rise to an improperly folded protein, which was unable to regain its full native conformation under 20 h of refolding conditions. The refolding process of Lap

from *Vibrio proteolyticus* (AAP) was demonstrated to be under kinetic control, as a significant increase in specific activity was achieved following incubation of refolded AAP at 37 °C for 72 h (Hernandez-Moreno et al., 2014). Therefore, it is possible that prolonged dialysis against refolding buffer could increase the population of molten globule state *refLapA* or yield native-like molecular conformation. However, this requires further experimental validation.

It is important to note, that *refLapA* carried a hexahistidine tag on the N-terminus, which was implicated to potentially act as an intramolecular chaperone in a similar refolding study of Lap from *Vibrio proteolyticus* (Hernandez-Moreno et al., 2014). Similarly, the observed specific activity of rLap from *V. proteolyticus* was somewhat reduced compared to the wild-type enzyme. These findings and our experiments highlight the importance of the propeptide region and its potential role as an intramolecular chaperone in the folding process of these aminopeptidases, leading to a completely functional conformation of the catalytic domain.

To further validate the role of the propeptide as an intramolecular chaperone, it would be useful to study the unfolding/refolding process of recombinant LapA, in the presence and absence of propeptide, using urea or guanidinium chloride. The ability to fully regain activity in completely unfolded pro-LapA polypeptide would indicate the chaperonin effect of LapA propeptide, provided the mature unfolded segment would fail to refold into an active conformation. Also, addition of propeptide, as a separate peptide to *refLapA* refolding mixture, may induce the transition from a molten globule state to a native-like conformation. In a similar fashion, molten globule subtilisin could be converted to native molecule (Fu et al., 2000). Nevertheless, these experiments highlight the importance of prodomain in the expression constructs of wild-type LapA and suggest its potential role as an intramolecular chaperone, accelerating the rate of folding and secretion in the native and recombinant hosts.

We have demonstrated that LapA propeptide acts as an intramolecular chaperone and is essential for the maturation of the catalytic domain when expressed in *cis*. Further, it is evident that LapA prodomain may be able to function in an intermolecular manner and assist with secretion and folding of the mature polypeptide when expressed in *trans*. Several successful studies of prodomain expression in non-covalently linked state (*trans*) were reported for homologous and more distantly related peptidases. Expression of propeptide and mature domains, as separate polypeptides, resulted in the recovery of activity for membrane Type 1-Matrix Metalloproteinase (Cao et al., 2000). Impeded secretion of the mature domain of aspartic proteinase Sap1p from *C. albicans* in *P. pastoris* was restored

upon co-expression of the pro-polypeptide in *trans* (Beggah et al., 2000). Unlinked propeptide co-expression in *P.pastoris* also was shown to rescue the appropriate secretion and activity levels of *Streptomyces mobaraensis* Transglutaminase (Yurimoto et al., 2004). We are currently investigating an expression platform for co-production of covalently unlinked LapA domains. Proven successful, this could overcome the requirement of downstream enzymatic processing of LapA proenzyme. This, in turn, would provide ready access to highly active, appropriately folded and relatively pure recombinant enzyme from host culture supernatant, which is desirable in commercial applications.

## **Chapter 6 STRUCTURAL AND FUNCTIONAL STUDIES OF LAPA MUTANTS: CHARACTERISATION OF THERMOLABILITY**

---

## 6.1 LapA thermolabile mutants M1 – M6: 1st round mutagenesis

### 6.1.1 Rationale and structural bioinformatics

The inherent thermolability of cold-adapted (psychrophilic) proteins encouraged the exploration of protein sequence databases in an effort to identify key sequence and structural differences that could lead to improved thermolability of LapA. Equally, it was essential to understand hot-adapted (thermophilic) homologues to clearly establish the structural links between increased rigidity and thermostability in thermophiles, which should be avoided in LapA mutagenesis.

Homologous sequences of psychrophilic and thermophilic aminopeptidases were obtained by performing a BLAST search (of non-redundant sequences) using the mature sequence of LapA as the query sequence. Only sequences above 60 - 70% query coverage were chosen for subsequent analysis.

Multiple sequence alignment was performed with selected psychrophilic and thermophilic sequences, which were manually inspected, to determine the habitat of their respective organisms (**Figure 10.5**, Appendix II). The most homologous peptidase (38% identity, 98% sequence cover) was aminopeptidase from the Antarctic fungus *Glaciozyma antarctica*. Other sequences from genuine psychro- and thermophilic organisms shared a significantly lower degree of homology to LapA mature sequence (<25% identity, 50-60% sequence cover). The region showing the most sequence homology, across selected 15 sequences, was between residues 78(156) – 293(372), where the number in brackets corresponds to a full reading frame. This region spans the conserved  $\alpha/\beta$  sandwich fold in M28 peptidases, harbouring the catalytic and substrate binding pocket residues. The least homologous region was found at the N-terminus from residues 60(139) – 77(155), where some of the psychrophilic homologues contained an additional long insert, which formed a separate domain in the predicted structure model (**Figure 10.6**, Appendix II).

Structural models, using a template-based approach, of select sequences (M28 peptidases from *Psychroserpens burtonensis* and *Rhodonellum psychrophilum*) were obtained using the online structure prediction server Phyre 2 (Bennett-Lovsey et al., 2008). The homology models obtained were superimposed with the newly acquired LapA crystal structure (**Figure 10.6**, Appendix II) for comparison.

The highest sequence and structural disparity between LapA and temperature-adapted homologues were found in variable loop regions and mainly the protein surface. Therefore,

it was postulated that one of the routes to temperature adaptation in psychrophilic proteins was via optimisation of residue content in these flexible loop regions and modulation of interactions with the solvent by surface residues. This observation is in agreement with evidence in the literature, which suggests that additional flexibility leads to cold adaptation and thermolability in psychrophilic proteins (as discussed in sections 1.5 and 1.6). These observations helped guide our mutagenesis strategy, whereby the protein core was left intact, and residue substitutions were targeted to the loop regions and ends of secondary structure elements of release conformational entropy - favouring increased flexibility and increased entropy.

The first round of LapA thermolability mutations (**Table 6.1**) were designed based on the multiple sequence alignments of the identified psychro- and thermophilic homologues (**Figure 10.5**, Appendix II) and the newly acquired structure of native LapA (*flavLapA*) from Flavourzyme (section 3.2). The *flavLapA* crystal structure provided detailed information regarding secondary structure elements, precise interatomic distances and key salt bridges stabilising the protein, as well as localised residue rigidity, as a mean of B-factor (a crystallographic thermal motion parameter) - features that could be exploited during mutagenesis to impart thermolability.

For tertiary structure prediction, the selected sequences containing the M1 – M6 mutations were submitted to the SWISS-MODEL server (Biasini et al., 2014), using the *flavLapA* structural coordinates as a template. However, no major structural rearrangements or protein backbone movements, in the vicinity of the mutation sites, were predicted in the resulting structural models, suggesting no major structural changes should be exerted by the M1 – M6 point mutations.

In this round, also the effort was directed toward investigation of the importance of the glycosylation site, N87, previously confirmed in the *flavLapA* crystal structure. This N-linked sugar moiety, in the recombinant LapA protein, could hinder subsequent crystallisation attempts of thermolabile mutants. However, more importantly, glycosylation has been shown to have a thermostabilizing effect on numerous protein structures (Guo et al., 2016, Shental-Bechor and Levy, 2008). Thus, removal of this post-translational modification could result in the desired reduction in thermostability of LapA. On the other hand, some studies suggested a possible influence of glycosylation on the folding process and protein stability (Imperiali and O'Connor, 1999, Xu and Ng, 2015). Therefore, this warranted further study.

An alternative mutagenesis strategy also was available, whereby the lowest B-factor residues would be targeted (referred to as rigid 'hot spots'). This approach has lead to destabilisation and significantly increased thermolability of lipase from *Pseudomonas aeruginosa* (Reetz et al., 2009). Using the *flav*LapA crystal structure, key rigid 'hot spot' residues were identified, as follows: Q246 (Bfactor: 16.3 Å<sup>2</sup>), G324 (Bfactor: 19.7 Å<sup>2</sup>), G174 (Bfactor:20.2 Å<sup>2</sup>). However, these residues are conserved and located within the protein core, forming essential secondary structure elements. Destabilising LapA in the conserved fold regions might have a detrimental effect on tertiary structure and overall stability. A high-throughput mutant screening platform would be required to take a directed protein evolution approach, which was available in the Reetz et al. study. Considering the complexity of recombinant cloning and expression in *P.pastoris*, this latter strategy of LapA mutagenesis was not feasible.

Mutant	Residue/Substitution	Putative effect	B-factor ( $\text{\AA}^2$ )	Conservation	Location	Rationale
M1	P155/G	Increase local flexibility	30.1	Not conserved	$\beta$ turn	Very few psychrophilic homologues have prolines at this position. Proline to glycine mutation has rendered Subtilisin Carlsberg more thermolabile (Fuchita et al., 2012).
M2	P304/A	Increase local flexibility	37	Not conserved	Loop	Proline was found in LapA and GAP, but no other psychrophilic homologues. However, thermophilic counterparts contain proline at this position.
M3	R252/K	Weaken the salt bridge	31.3	Not conserved	Loop	No arginine present in other psychrophilic homologues, but in thermophilic counterparts. Forms a very strong salt bridge with E226.
M4	Q91/A	Brake the salt bridge	35.5	Not conserved	Helix	Forms a strong salt bridge with R362, which is part of a rigid helical secondary structure element.
M5	N87/A	Eliminate glycosylation site	32.2	Not conserved	Helix	Glycosylation might contribute to added thermostability of LapA protein, therefore elimination of the sugar moieties should render LapA more thermolabile.
M6	C310/A	Brake the disulphide bond	55.5	Not conserved	Loop	Strong covalent disulphide bond in C310 – C314 loop largely contributes to overall stability of LapA structure. Removal of this interaction should significantly destabilise the whole structure.

**Table 6.1** Library of M1 – M6 LapA thermolability mutants. B-factor analysis was based on the crystal structure of native LapA enzyme from Flavourzyme ( $B_{av} = 33.9 \text{ \AA}^2$ ).

### 6.1.2 Recombinant cloning and expression of M1 – M6 mutants

Due to the complex and laborious nature of recombinant LapA expression and purification from *P.pastoris* supernatant, only a limited number of LapA variants could be screened in the first round of mutagenesis. In this initial round of mutagenesis, six LapA thermolability point mutations were generated (**Table 6.2**).

All LapA thermolability mutants were generated using the Quickchange site – directed mutagenesis protocol (section 2.2.6). The wild – type LapA precursor expression construct, pD-npro-LapA, served as a template in the mutagenic PCR reaction using the primers pairs given in **Table 6.2**.

Mutant name	Mutation	Mutagenic primers
M1	P155G	5'-AATCTTTGCGATCAAGGATGATTGACCAAACGAATGTTTGAAAGGCTCTAC-3' 5'-GTAGAGCCTTTCAAACATTCGTTTGGTCAATCATCCTTGATCGCAAAGATT-3'
M2	P304A	5'-CGCCTACTGTACTATCGCTACTGTCGATTCTAAGT-3' 5'-ACTTAGAATCGACAGTAGCGATAGTACAGTAGGCG-3'
M3	R252K	5'-GTTGTAACATAGCTTTAACATCTTTTGACTTCTGAGAATACTGTTTCG-3' 5'-CGAACAGTATTCTCAGAAGTCAAAAGATGTTAAAGCTATGTTACAAC-3'
M4	Q91A	5'-TTTTTTGTCCAACGATTTGATAAGATTAGCCACAGTTTCATTGTGTTGGACGCTATC-3' 5'-GATAGCGTCCAACACAATGAAACTGTGGCTAATCTTATCAAATCGTTGGACAAAAA-3'
M5	N87A	5'-TTGTCCAACGATTTGATAAGATTATTCACAGTTTCATTGTGTTGGACGC-3' 5'-GCGTCCAACACAATGAAACTGTGAATAATCTTATCAAATCGTTGGACAA-3'
M6	C310A	5'-GTGATCACTACATCCATAGCCAGCCTTAGAATCGACAGTAGGGATA-3' 5'-TATCCCTACTGTCGATTCTAAGGCTGGCTATGGATGTAGTGATCAC-3'

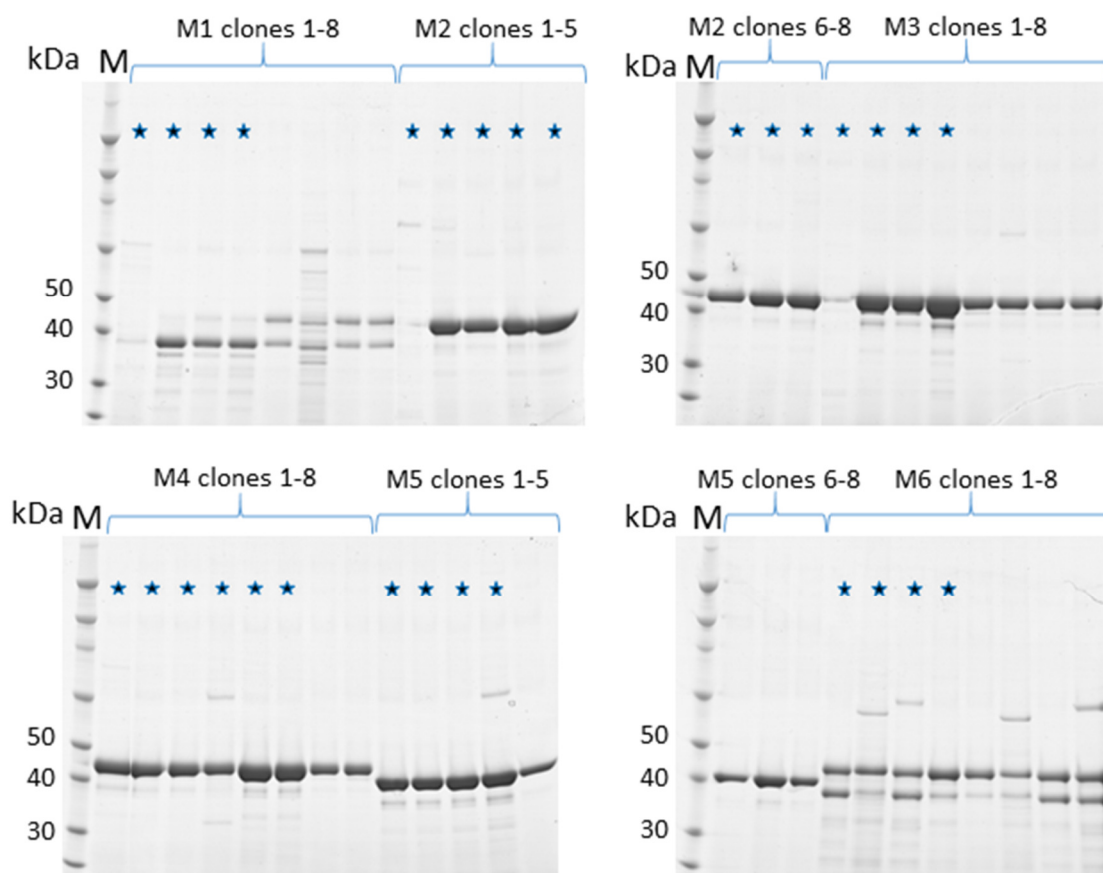
**Table 6.2** Primers used to generate M1 – M6 point mutations

Five colonies, of each mutant *E.coli* clones, were tested to verify that these carried the gene, by colony PCR using AOX1 primers. All *E.coli* clones were found to harbor the expected mutant LapA plasmids. Subsequent plasmid DNA purification and sequence verification showed successful integration of site – directed mutations (results not shown).

Following integration of all LapA mutant expression cassettes into the *P. pastoris* X33 strain, approximately 20 colonies were observed on high concentration ('Bonus') zeocin plates for each mutant strain. It was decided to omit transformant verification by colony PCR at this stage, since the very high concentration of antibiotic only selects for multiple genomic integration events of the expression cassette.

Using a high-throughput micro-fermentor, 1 mL cultures were set up to determine the secretion profiles of M1 – M6 mutants (2.13.3). Variation of 'Bonus' and regular transformants was picked for inoculation to test any differences in the secretion levels.

After 115h of continuous induction and a final concentration of 0.5% MeOH, supernatants were harvested and analyzed by SDS – PAGE (**Figure 6.1**).



**Figure 6.1** Analysis of the secretion profiles of the M1 – M6 LapA thermolability mutants. All lanes represent a 10  $\mu$ L supernatant load. ‘Bonus’ transformants are indicated by a star.

All six thermolability LapA mutants yielded extracellularly secreted recombinant protein in the *P. pastoris* supernatant. ‘Bonus’ transformants showed higher secretion levels in all mutants, except clones of the M6 mutant. This suggests that screening for multiple homologous recombination events was an effective strategy to increase secretion levels of the LapA mutants of interest.

The M1 and M6 *P. pastoris* clones secreted partially mature LapA protein. The remaining LapA mutants were secreted primarily in the form of precursor. The M5 (N87A) precursor migrated at a slightly lower molecular weight as expected, due to the successful elimination of the N – glycosylation site. Thus, M5 served as a reference, leading to the conclusion that the remaining higher – migrating LapA mutants were glycosylated.

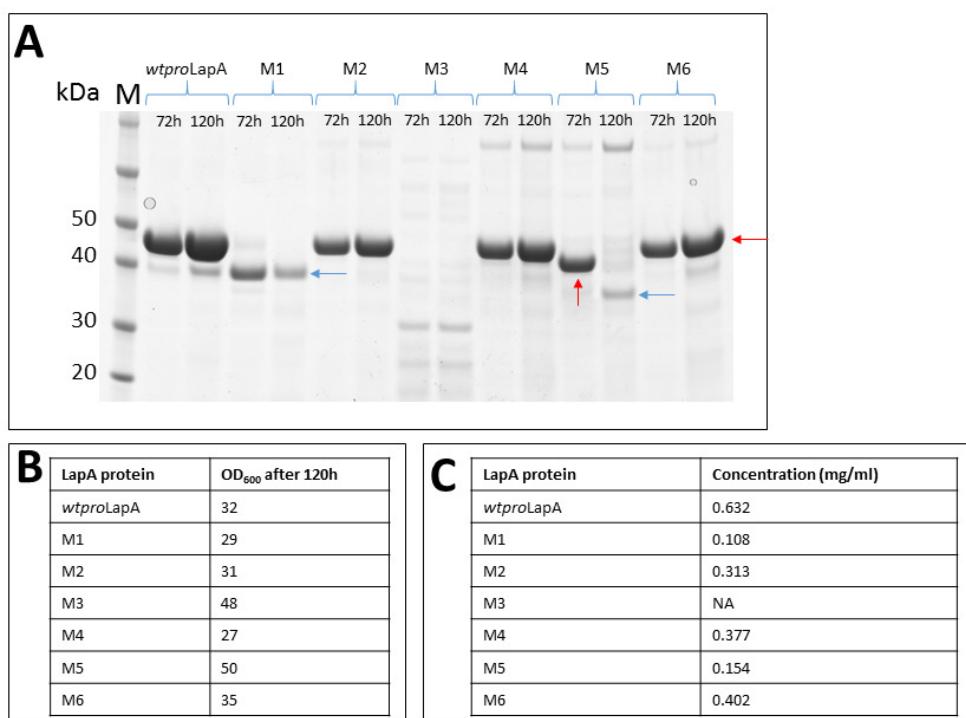
Based on SDS-PAGE analysis, clones of M2, M3, M4, and M5 showed comparable levels of secretion. M1 and M6 mutations negatively affected secretion of the recombinant protein, possibly due to adverse implications for structural integrity and polypeptide folding. Interestingly, only M1 and M6 mutants showed indicative maturation events, which could render the precursor more susceptible to proteolytic processing by co-secreted *P. pastoris* non-specific proteases.

#### 6.1.2.1 Scale-up expression of selected LapA mutant clones

400 mL shake flask cultures were set up with selected LapA mutant clones (**Table 6.3**). Wild – type LapA protein also was expressed, as a positive control. After 120h of continuous induction, with a final concentration of 0.5% MeOH, supernatants were harvested and analyzed by SDS – PAGE (**Figure 6.2**).

LapA mutant	Clones selected
M1	2 and 4
M2	2 and 8
M3	2 and 4
M4	5 and 6
M5	1 and 3
M6	1 and 4

**Table 6.3** LapA mutant clones selected for subsequent large-scale expression.



**Figure 6.2** Recombinant expression data from large – scale trial of wild-type and M1-M6 LapA mutants. **A** – analysis of supernatants sampled after 72 and 120h; **B** – OD<sub>600</sub> values after 120h; **C** – protein quantification, by Bradford assay, after 120h. LapA proenzyme and mature protein populations are indicated by red and blue arrows, respectively.

**Figure 6.2** (A) shows that *wtproLapA*, M2, M4 and M6 proteins were secreted in a pro – form (precursor protein). M1 mutant appeared in a mature form, but suffered from degradation within 72 – 120h time interval. Maturation and concomitant degradation of M5 mutant also seemed to occur within a similar time interval. Removal of the N – linked glycosylation site (N87) may have rendered the M5 mutant more susceptible to proteolytic degradation during the 5 – day culture incubation period. The M3 mutant showed no full – length secreted target protein, but this was due to culture contamination, which was indicative from the smell of the culture.

In this trial, *wtproLapA* secreted the largest amount of protein (0.63 mg/mL). Other mutants showed moderate levels of secretion (0.3 – 0.4 mg/mL), except M1 and M5, both of which yielded very low amounts of recombinant protein after 120 h (**Figure 6.2, C**).

### 6.1.3 Purification and AP1 – mediated enzymatic activation of M1 – M6 mutant precursors

Mutants M1 – M6 were purified and their active forms were obtained, using the workflow described in **Figure 2.1**. Briefly, all LapA mutant proteins were concentrated and separated from bulk media by  $\text{AmSO}_4$  precipitation, and loaded onto a gel filtration resin to isolate the recombinant protein from the pigmented small molecular weight peptides. Active fractions were pooled and loaded onto a Q HP AEX column, to eliminate residual brown pigments. Typically, the 140 mM NaCl eluted fraction contained pigment – free LapA precursor. However, in some cases, the brown pigment could not be completely separated from recombinant LapA protein, and repetitive AEX cycles were required.

The established AP1 – mediated enzymatic processing protocol (section 2.13.4.1) was employed for activation of all purified mutant LapA precursor proteins. Approximately 1  $\mu\text{g}$  of Flavourzyme – purified AP1 protease was able to fully process 1 mg of purified LapA precursor within 60 minutes at 37 °C. Following the propeptide cleavage reaction, AP1 – LapA mixture was re-applied onto an AEX column. The 140 mM NaCl eluted fraction typically contained activated mature LapA protein, which was subsequently used in thermolability assays (as described in section 6.5.1).

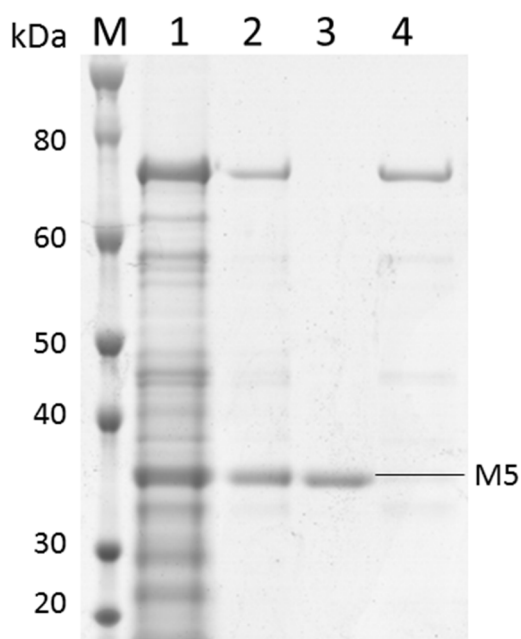
Following 120 h incubation, M1 and M5 mutants were already in the active mature form (**Figure 6.2**). However, purified protein batches were still processed to ensure consistent successive comparisons of LapA specific activity. *wtp<sub>pro</sub>LapA*, M2, M3 and M4 proteins showed 6 – 7 fold activation, following treatment with AP1 (Table 6.4). Interestingly, M1 still encountered 4.5-fold activation, suggesting that the protein was not secreted in fully mature form and was still inhibited by the propeptide domain, although showing a descent on the SDS-PAGE gel (**Figure 6.2**, A). M5 and M6 were degrading throughout the course of the purification process.

LapA protein	Specific activity, pre-treatment by AP1 (U/mg)	Specific activity, post-treatment by AP1 (U/mg)	Activation -fold
<i>wtproLapA</i>	32.0	165.0	-5.15
M1	16.0	71.4	-4.46
M2	21.5	150.0	-6.97
M3	12.5	87.14	-6.97
M4	26.7	159.0	-5.95
M5	39.0	35.0	degradation
M6	1.36	0.64	degradation

**Table 6.4** Specific activity values of LapA proteins before and after enzymatic treatment with AP1 protease.

M5 (N87A substitution at the N – terminal glycosylation site) is worthy of special consideration, since reduced thermal stability of recombinant LapA enzyme was anticipated. It was vital also to evaluate a role for this post – translational modification on enzyme folding and specific activity. Also, high-mannose hyper-linked oligosaccharides, attached upon secretion in *P. pastoris*, are quite distinct from usually shorter and less complex *A. oryzae* N-linked sugars (Kainz et al., 2008). Therefore, natively expressed LapA might possess a modified N-glycosylation profile, with subsequent affects that should be determined.

Surprisingly, expression of M5 in *P. pastoris* triggered the release of other ER proteins into the culture supernatant (**Figure 6.3**, lane 1), suggesting compromised folding and secretion processing of this heterologous protein. The active pool of M5 was eluted at slightly higher (200 mM) NaCl concentration, implying a stronger association with the AEX resin, which might be a consequence of an increase in exposure of charged residues in unfolded M5 protein. The N87A mutation also severely reduced the catalytic ability of LapA enzyme, with only 21 – 25% remaining activity in relation to the wild-type protein (**Table 6.4**). Although the N87 residue is very distant from the active site (located approximately 32 Å away from the active site and surface exposed), it must be indirectly involved in activity, most likely exhibiting its effects via the folding mechanism. Interestingly, another *A. oryzae* protease showed no change in specific activity following the elimination of its N-linked sugar and expression in *P. pastoris* (Lei et al., 2013). Substantially decreased secretion levels of M5 protein also indicates the importance of this modification in processing.



**Figure 6.3** SDS-PAGE analysis of the purification of M5 LapA mutant. Lanes: 1 – M5 crude supernatant; 2 – active fractions pooled from gel filtration step; 3 – AEX purification step: 200 mM NaCl eluted fraction; 4 - AEX purification step: 400 mM NaCl eluted fraction. The active fraction of M5 is indicated.

The introduction of an N-glycosylation site has been shown to improve the secretion of numerous heterologous proteins in yeasts (Sagt et al., 2000). The process of N-glycosylation is known to occur in two steps in yeast. Firstly, asparagine modification begins in the endoplasmic reticulum, as a co-translational event involving attachment of pre-assembled sugar moieties. The process is finalised in the golgi, where glycans are restructured into more complex assemblies (Bretthauer and Castellino, 1999). Negatively affected secretion and specific activity in the LapA M5 mutant, suggested that this post-translational modification is somehow mediating the folding of LapA enzyme. Successfully obtained crystals of *wtLapA* and *wtproLapA* (section 5.1) confirmed that N87-linked glycan was not interfering with crystallisation process. As a result, the N87 site was left unmodified in the 2<sup>nd</sup> round of LapA thermolability mutants.

## 6.2 LapA thermolabile mutants M7 – M13: 2nd round mutagenesis

### 6.2.1 Rationale and structural bioinformatics

In this round of mutagenesis, more emphasis was given to the sequence and structural comparison with the closest genuine psychrophilic homologue – GAP M28 peptidase, identified during the bioinformatics search (see section 6.1.1). A GAP homology model (see section 8.1) was superimposed on the recombinant LapA crystal structure (*w*LapA) and the main differences were deduced (see **Figure 8.2**). The key structural differences involved a missing K267 – A272 helix and a 6-residue insertion in P164 – G165 region. This comparison suggested targeting of the K267 – A272 helix in an effort to destabilise the helix, contributing to enhanced thermolability (M8 mutant, **Table 6.6**). Furthermore, in this round of mutagenesis, it was decided to exploit the concept of more flexible and heat – labile active sites, as observed in other psychrophilic proteins. Consequently, the GAP homology model and the *w*LapA crystal structure were analysed using the CASTp (Liang et al., 1998) server to determine the difference in the number and size of cavities and pockets within each of their respective molecular scaffolds. One unique pocket was identified to be substantially distinct in GAP homology model (POC 57, **Table 6.5**). This pocket was in the active site crevice and contained residue substitutions with notably less bulky side-chains (**Table 6.5**). In fact, the CASTp server identified 37 cavities and pockets in the *w*LapA structure, compared to 57 cavities found in the GAP homology model. This difference could give rise to greater overall plasticity in the GAP protein and be one of the major cold-adaptation strategies for this enzyme.

Pocket	Area <sub>ms</sub> *	Vol <sub>ms</sub> *	Length	Residues					
POC 57 GAP	310.99	361.26	133.34	V271	C33 4	A33 6	G33 8	S343	M34 7
POC 37 LapA	165.83	216.44	88.25	<u>M26</u> <u>2</u>	<u>F329</u>	<u>F331</u>	<u>S333</u>	<u>D33</u> <u>8</u>	I342

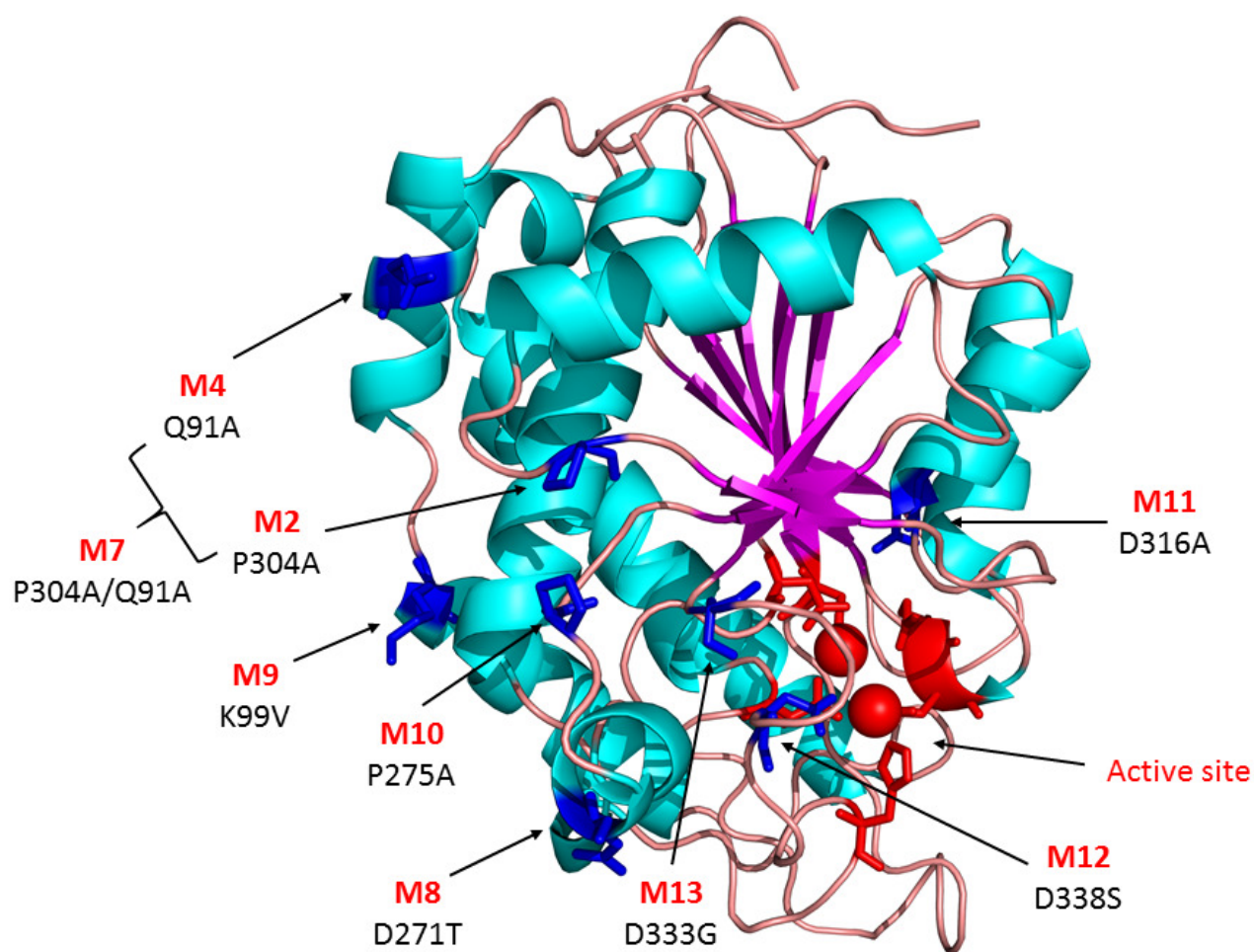
\*molecular surface area in Ångstroms; \*\*molecular surface volume in Ångstroms

**Table 6.5** CASTp pocket analysis of LapA and GAP proteins. Residues in LapA containing more bulky side-chains are underlined.

This analysis led to the design of M12 (D338S) and M13 (S333G) mutations (**Table 6.6**). Moreover, these observations coincided with an analysis of rigid clusters in *wtLapA* conducted using the KINARI-Web server (Fox et al., 2011), whereby D338 formed part of the identified D338 – Y341 rigid cluster. And finally, residue content analysis (using Prot Param server (Gasteiger et al., 2003)) indicated that GAP possessed a significantly lower number of lysines (21 versus 14), but a greater number of bulky hydrophobic tryptophan (3 versus 1), methionine (8 versus 2) and phenylalanine (19 versus 16) residues. Therefore, reducing the property of core hydrophobicity seemed counter-intuitive in *wtLapA*.

Other possible mutations identified were as follows: LPESKT stretch of residues could be inserted between residues P164 and G165 to mimic an analogous insert in the GAP structure; P164, which is part of the loop interconnecting  $\beta$ 2 and  $\beta$ 3 strands, could be mutated to E or K, as observed in psychrophilic homologues at this position; Q241 and A299 could be substituted to E, D or K residues, also present in psychrophilic homologues. More negatively charged and surface exposed conserved residues were previously identified in psychrophilic proteins (Feller et al., 1999). Alternatively, non-conserved P185, P222, P191 residues, located in loop regions, could be targeted to confer more flexibility to neighboring residues. The introduction of more cavities in the protein core, by the elimination of bulky side-chains, is another feasible mutagenesis strategy, which could be achieved by I94A substitution in *wtLapA*. It is reported that psychrophilic proteins soften the hydrophobic compactness effect by reducing the size and relative hydrophobicity of nonpolar residue clusters (Saelensminde et al., 2009). The S333G mutation could confer localised flexibility in the vicinity of the active site. Strategically placed glycine residues also have been reported in psychrophilic proteins (Kulakova et al., 2004, Mavromatis et al., 2002).

In summary, the LapA mutagenesis library in this round included a selection of different approaches to achieve thermolability: elimination of salt bridges, unraveling helical elements, replacing prolines in loops and conferring flexibility to the active site (**Table 6.6**). The location of these residues in the recombinant *wtLapA* tertiary structure is shown in **Figure 6.4**.



**Figure 6.4** Proposed LapA thermolability mutations mapped onto the *wt*LapA crystal structure. Targetted residues and active site residues are shown as sticks and are coloured in blue and red, respectively.

Mutant	Residue/Substitution	Putative effect	B-factor (avge for residue)	Conservation	Location	Rationale
M7	P304A + Q91A (double mutant)	Increase local flexibility	33.4 + 44.4	Not conserved	Loop + Helix	M2 (P304A) and M4 (Q91A) single point mutants were combined to test the additive effect on thermolability.
M8	D271T	Unravel the helix	47.3	Not conserved	Helix	D271 is part of the K267 – A272 helix, which is absent in GAP. Also, D271 is involved in the formation of stabilising a 3.1 Å salt bridge with K267. Threonine was chosen to weaken the side chain – side chain interaction and increase the number of surface exposed nonpolar residues, which is an entropy-driven destabilising factor (Aghajari et al., 1998, Russell et al., 1998, Georlette et al., 2003). This substitution also could unravel the helix.
M9	K99V	Increase overall flexibility	31.6	Not conserved	Helix	K99 forms a strong salt bridge with D335, connecting two distant long helical elements. Valine is identified in pairwise alignment in GAP at this position.
M10	P275A	Increase local flexibility	48.6	Not conserved	Loop	Located in the variable loop connecting helical and $\beta$ strand secondary structure elements. Expected to be as successful as P304A mutation.
M11	D316A	Increase active site plasticity	27.7	Strictly conserved	Close to the active site	Quite rigid residue, possessing multiple electrostatic interactions surrounding the active site, potentially involved in 2 <sup>nd</sup> shell electrostatic network.
M12	D338S	Increase active site plasticity	31.3	Not conserved	Active site crevice	Determined by CASTp analysis of pockets and cavities. Substitution was analogous to the residue in GAP present at this position.
M13	S333G	Increase active site plasticity	29.1	Not conserved	Active site crevice	Determined by CASTp analysis of pockets and cavities. Substitution was analogous to the residue in GAP present at this position.

**Table 6.6** Library of M7 – M13 LapA proposed thermolability mutants. B-factors were obtained from the *w*LapA crystal structure ( $B_{av} = 36.8$ ).

## 6.2.2 Recombinant cloning and expression of M7 – M13 mutants

All LapA thermolability mutants were generated using Quickchange site – directed mutagenesis (section 2.2.6). The *wt*LapA precursor expression construct pD-npro-LapA served as a template in the mutagenic PCR reaction, using the primer pairs given in **Table 6.7**, except for generation of the M7 double-mutant construct, whereby an expression cassette containing P304 (M2) mutation was used as a template.

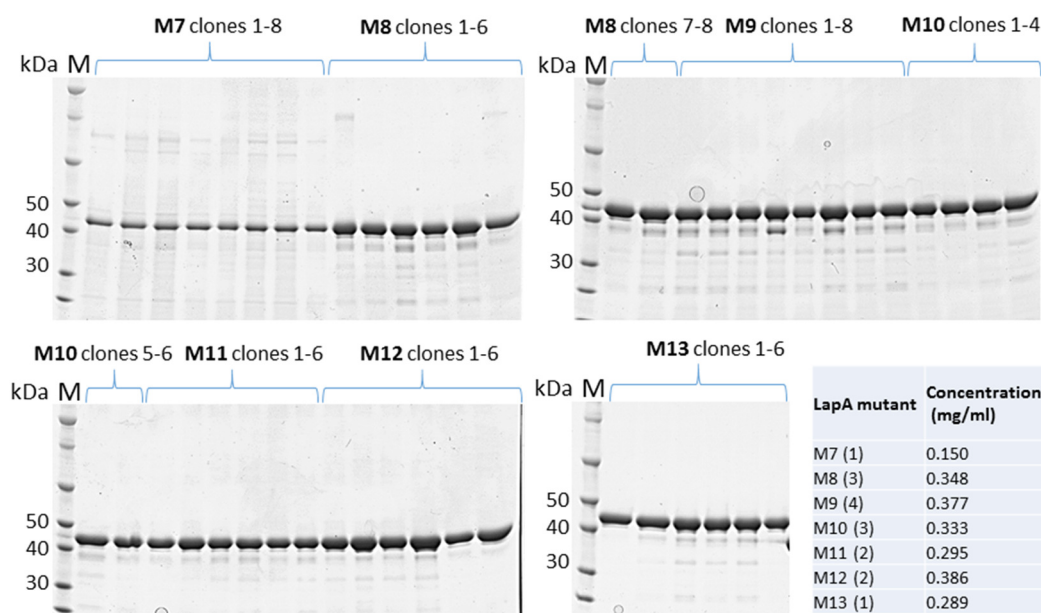
Mutant name	Mutation	Mutagenic primers
M7	P304/Q91	5'-TTTTTTGTCCAACGATTTGATAAGATTAGCCACAGTTTCATTGTGTGGACGCTATC-3' 5'-GATAGCGTCCAACACAATGAAACTGTGGCTAATCTTATCAAATCGTTGGACAAAAA-3'
M8	D271T	5'-TCTGGTTTACCAGCAGTTGTAGTACCCTTTGTGTAACCAGTCA-3' 5'-TGACTGGTTACACAAAGGGTACTACAACCTGCTGGTAAACCAGA-3'
M9	K99V	5'-CGGTTGCAAAACAGTTTCGAAATTTTAACGTCCAACGATTTGATAAGATTTTGCAC-3' 5'-GTGCAAAATCTTATCAAATCGTTGGACGTTAAAAATTCGAAACTGTTTTGCAACCG-3'
M10	P275A	5'-TATACCAATGGACTCAGCTTTACCAGCATCTGTAGTACCC-3' 5'-GGGTACTACAGATGCTGGTAAAGCTGAGTCCATTGGTATA-3'
M11	D316A	5'-GGTGGCAGAAGCGTGAGCACTACATCCATAGCC-3' 5'-GGCTATGGATGTAGTGCTCACGCTTCTGCCACC-3'
M12	D338S	5'-CGCATTTGAGAGCGCTTTCGGTGATTCTTCCCCATACATTCATTCT-3' 5'-AGAATGAATGTATGGGGAAGAATCACCGAAAGCGCTCTCAAATGCG-3'
M13	S333G	5'-GAGTCATCACCGAAAGCACCTCAAATGCGAAAGCCG-3' 5'-CGGCTTTCGCATTTGAGGGTGCTTTCGGTGATGACTC-3'

**Table 6.7** Primers used to generate M7 – M13 point mutations.

Five colonies of each mutant *E.coli* clones were tested to carry the gene by colony PCR using AOX1 primers. All *E.coli* clones were found to harbor the expected mutant LapA plasmids. Subsequent plasmid DNA purification and sequence verification showed successful integration of site – directed mutations (data not shown).

Following integration of all LapA mutant expression cassettes into the *P.pastoris* X33 strain, >100 colonies were observed for each mutant strain on high concentration ('Bonus') zeocin plates. It was decided to omit transformant verification, by colony PCR, since the very high concentration of antibiotic only selects for multiple genomic integration events of the expression cassette. Therefore, directly 1 mL cultures were set up in a high-throughput micro-fermentation experiment to determine the secretion profiles of the M7 – M13 mutants (section 2.13.3). This time only 'Bonus' transformants were picked for inoculation, since these should give rise to higher levels of secretion. After 96h of continuous induction, with

a final concentration of 0.5% MeOH, supernatants were harvested and analyzed by SDS – PAGE (**Figure 6.5**).



**Figure 6.5** SDS-PAGE analysis of the secretion profiles of M7 – M13 LapA mutants. 10  $\mu$ L supernatant load, after 96h of induction, was used. Quantified LapA protein is provided for each mutant, with chosen clone number in brackets.

All seven LapA mutants yielded extracellularly secreted recombinant protein in *P. pastoris* supernatant. Clonal variation was almost absent among different transformants of LapA mutants. Not a single mutant gave rise to significant levels of mature LapA protein in the supernatants analyzed, with LapA precursor dominating in all samples. All mutants, in this round, secreted comparable levels of protein (0.29 – 0.39 mg/ml), except the M7 double-mutant, which yielded ~ 50% reduction in the level of secretion. Also, the M7 supernatants tested contained one or two co-secreted higher molecular weight proteins, either implying release of other ER proteins in *P.pastoris* or higher associations (oligomers) of LapA recombinant protein. In either case, this could suggest that the M7 double-mutant might possess a less stable molecular conformation and that cooperativity of single point mutations has significantly influenced its structural integrity.

LapA mutant	Clones selected
M7	1 and 8
M8	3 and 7
M9	4 and 6
M10	3 and 4
M11	2 and 4
M12	2 and 4
M13	1 and 6

**Table 6.8** LapA mutant clones selected for subsequent large-scale expression.

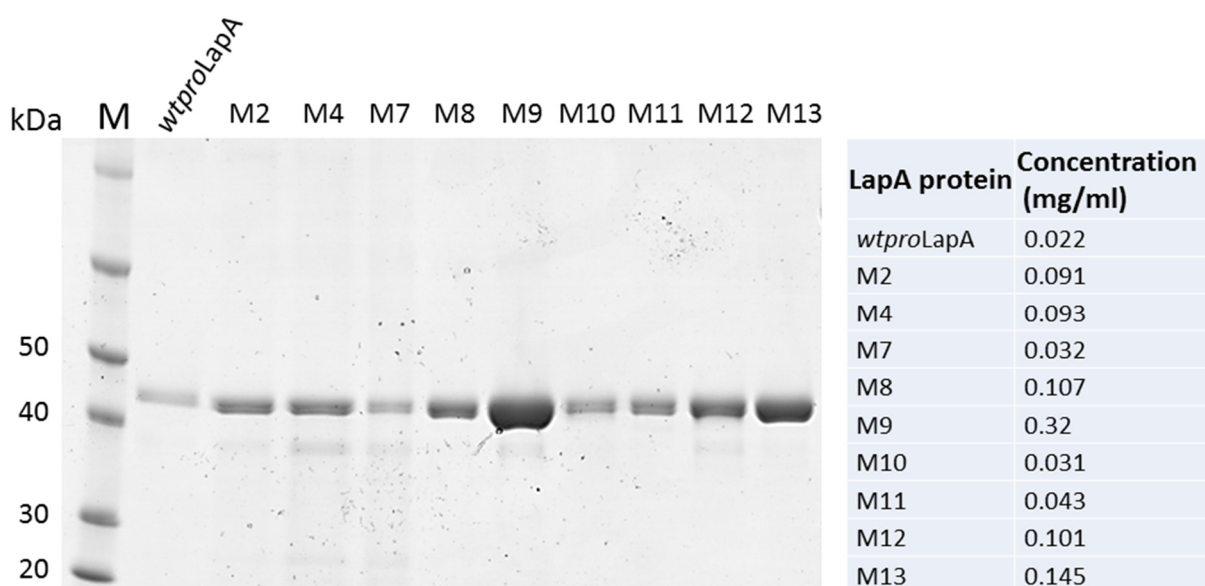
### 6.2.3 M7 – M13 recombinant expression in YNB synthetic media

Standard YP – based (yeast extract, soy peptone) medium, utilized for the expression of recombinant LapA proteins, contains unidentified brown pigments associating with secreted LapA protein. The underlying rationale of conducting this experiment (in YNB synthetic media) was to try to eliminate the pigments arising, potentially from the YP ingredients. Complete segregation of these pigments from recombinant LapA was problematic (at the time) when using Q HP AEX resin for purification. It was hypothesised further, that the residual pigments could hinder the crystallisation process for successive LapA mutants and prohibit crystal structure determination. Therefore, a new formulation of YP – free media was established, whereby the yeast nitrogen base was supplemented with amino acids and the YP component was substituted with sterilized water. However, a significant reduction in *P. pastoris* biomass was anticipated, due to diminished nutrient complexity in the growth medium. Indeed, compared to YP – based expression, on average ~ 50% decrease in biomass was observed, when measuring optical density at 600 nm for the wild – type and mutant LapA proteins (**Table 6.9**).

LapA proteins	OD <sub>600</sub> 48h	OD <sub>600</sub> 96h
WT LapA	13	18
M2	15	26
M4	13	21
M7	13	21
M8	14	21
M9	10	18
M10	8	8
M11	9	15
M12	8	5
M13	13	19

**Table 6.9** Cell density measured in the secondary culture of LapA supernatants.

500 mL shake flask cultures were set up with selected LapA mutant clones, using YNB synthetic media. M2 and M4 mutants were also added to the trial, in an effort to obtain pigment-free preparations for crystallization. Wild – type LapA protein also was expressed as a positive control. M10 and M12 mutant cultures did not grow well in YNB synthetic media. In general, the OD's reached by all cultures (after 96h) was variable, yet smaller than those routinely observed in YP rich cultures (typically reaching 30 – 40 OD<sub>600</sub> units).



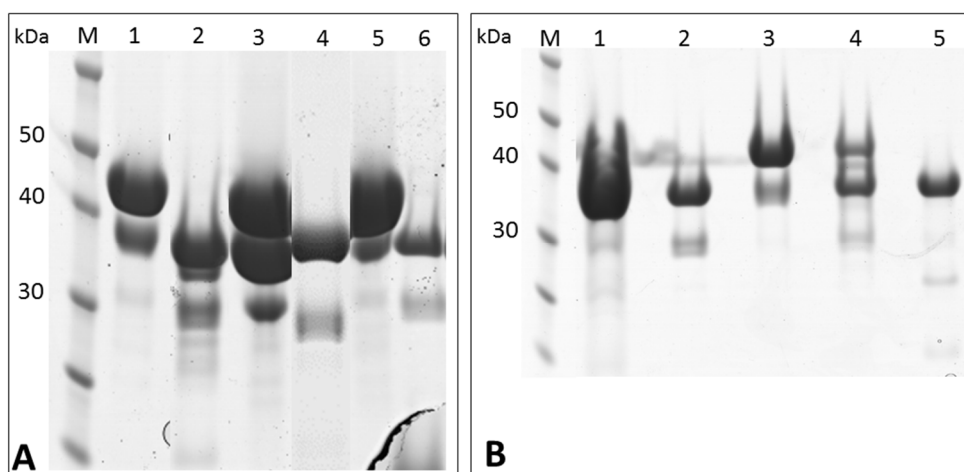
**Figure 6.6** SDS-PAGE analysis of LapA protein supernatants, following 96h induction with MeOH. 10  $\mu$ L of each supernatant was loaded for analysis.

The secretion level, typically observed for wild – type LapA protein grown in YP rich media, reaches 0.3 – 0.4 mg/ml after 96h. In this trial, possibly due to culture contamination, *wtproLapA* only revealed 0.022 mg/ml of protein (**Figure 6.6**). Final recombinant protein content was severely reduced for all mutant proteins, except M9, which displayed almost equivalent to rich media secretion levels. Therefore, it was concluded that growth of *P. pastoris* and subsequent production of heterologous protein was restricted by the limited availability of nutrients in the expression media.

The M9 mutant precursor was successfully purified and crystallized from its resulting supernatant (according to section 2.13.4). No associating brown pigments were observed for this particular batch of protein. However, in general, significantly lower yields of target protein, using this expression protocol in YNB synthetic media, prevented further exploration of other proteins. Fortunately, alternative crystallization strategies were later discovered for LapA precursor proteins, expressed in YP rich media and containing residual pigments.

#### **6.2.4 Purification and AP1 – mediated enzymatic activation of M7 – M13 mutant precursors**

LapA M7 – M13 mutants were purified and their active forms were obtained (**Figure 6.7**), using the equivalent procedure for the M1 – M6 mutants, as described in section 6.1.3. However, the previously encountered problem of incomplete pigment separation on a Q HP column was resolved by use of a MonoQ AEX column. Consequently, all LapA mutant proteins could be expressed in complex BMMY medium with no further need to eliminate the brown pigments during downstream processing (as discussed in the previous section 6.2.3).

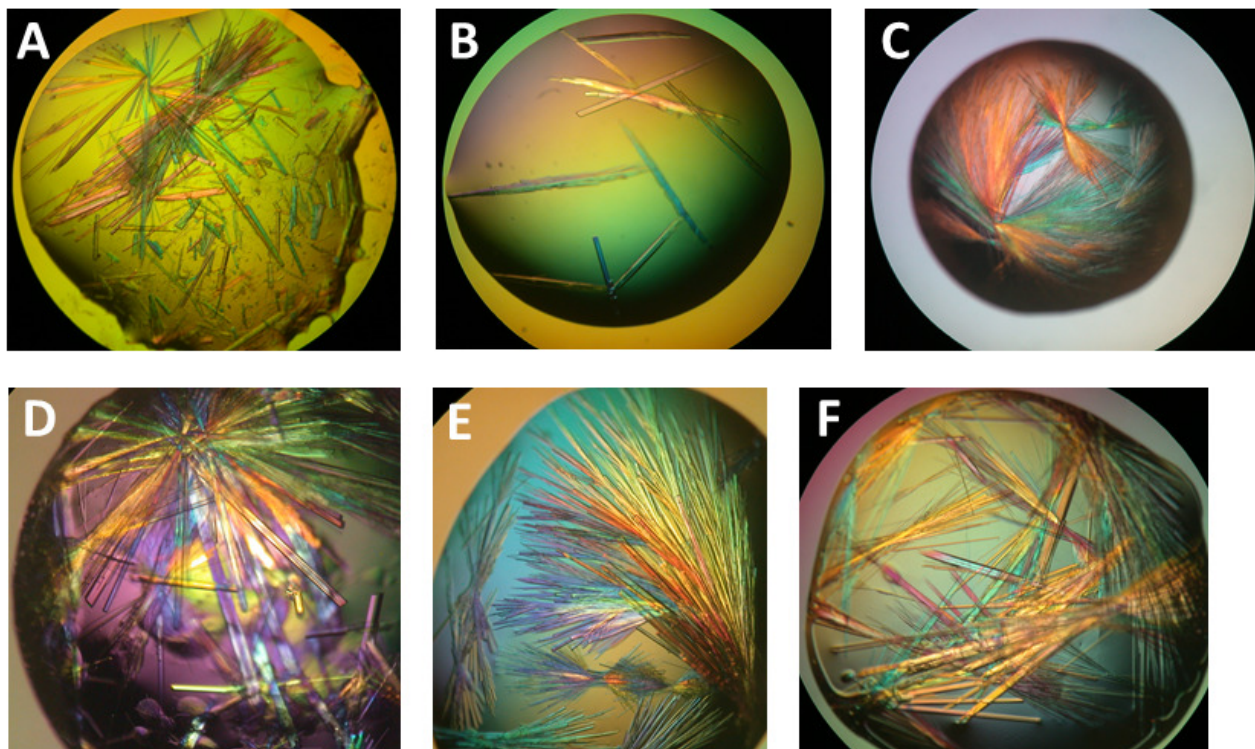


**Figure 6.7** SDS-PAGE analysis for purification and enzymatic activation of LapA mutants. **A** Lanes: 1 – M8 prior to AP1 protease treatment, 80 mM NaCl elution peak; 2 – M8 after AP1 protease treatment, 140 mM NaCl elution peak; 3 – M9 prior to AP1 protease treatment, 80 mM NaCl elution peak; 4 – M9 after AP1 protease treatment, 140 mM NaCl elution peak; 5 – M10 prior to AP1 protease treatment, 80 mM NaCl elution peak; 6 – M10 after AP1 protease treatment, 140 mM NaCl elution peak; **B** Lanes: 1 – M12 prior to AP1 protease treatment, 80 mM NaCl elution peak; 2 – M12 after AP1 protease treatment, 140 mM NaCl elution peak; 3 - M13 prior to AP1 protease treatment, 80 mM NaCl elution peak; 4 - M13 prior to AP1 protease treatment, 200 mM elution peak; 5 - M13 after AP1 protease treatment, 140 mM NaCl elution peak;

In the established purification protocol, a size exclusion chromatography step was followed by MonoQ AEX column chromatography to separate the resultant 80 mM and 200 mM NaCl peaks. The former contained a larger population of LapA precursor, whereas the latter was usually dominated by a mature LapA population (e.g. **Figure 6.7, B** lanes 3 and 4). These AEX fractions still contained traces of brown pigment. The 80 mM NaCl elution peak was concentrated (to 40 – 50 mg/ml) and aliquots of this protein preparation were used in subsequent crystallisation trials (see section 2.8.2.3). Remaining 80 mM and 200 mM NaCl fractions were then recombined, and 1 µg of AP1 was added to 1 mg of mutant protein to enable cleavage of the propeptide. Following the propeptide cleavage reaction, the AP1 – LapA mixture was re-applied onto the MonoQ column. The eluted 140 mM NaCl fraction contained fully activated pigment-free LapA mutant protein (**Figure 6.7 A** lanes 2, 4, 6; **B** lanes 2, 5). The AP1 activated and purified mutant fractions were subsequently used in characterisation of their thermal properties (see section 6.5.2).

### 6.3 Crystallisation of LapA thermolabile mutant structures

Structure solution and refinement, from the 1<sup>st</sup> and 2<sup>nd</sup> round of mutagenesis, are described herein. LapA thermolability mutants were purified and crystallised in the proenzyme form, harbouring the propeptide domain, according to section 2.8.2.3. An alternative route to obtaining overnight crystals was accidentally discovered during the course of this experiment, consequently, the previously described prerequisite of having a homogenous pigment-free LapA sample prior to crystallisation was no longer required. Furthermore, attempts to overexpress and purify LapA proteins from YNB synthetic media, lacking the pigmenting YP components, was found unnecessary (section 6.2.3). LapA mutant proenzymes were produced in BMMY complex media and purified according to the workflow in section 2.13.4, **Figure 2.1**. Following separation of the 80 mM and 200 mM peak fractions on a MonoQ AEX column, LapA pools still contained a considerable amount of brown pigments. However, rather surprisingly, crystal formation was not hindered by these pigments. Also, the 80 mM pool typically contained a significantly larger population of LapA proenzyme, whereas the 200 mM pool showed a substantial amount of mature protein present. Nevertheless, both peaks gave rise to similar quality of LapA proenzyme crystals (evident from similar size and shape of mutant crystals arising from 80 mM and 200 mM NaCl pools, **Figure 6.8**). Since the concentration of LapA mutant proteins in the crystal drop reached 30 – 50 mg/ml, multiple populations of crystals were observed (**Figure 6.8**).



**Figure 6.8** Crystal images of LapA thermolability mutants. **A** – M2 (from 80 mM NaCl pool); **B** – M7 (from 200 mM NaCl pool); **C** – M8 (from 80 mM NaCl pool); **D** – M9 (from 80 mM NaCl pool); **E** – M10 (from 200 mM NaCl pool); **F** – M12 (from 80 mM NaCl pool).

## **6.4 Analysis and comparison of wild-type LapA and thermolabile mutant structures**

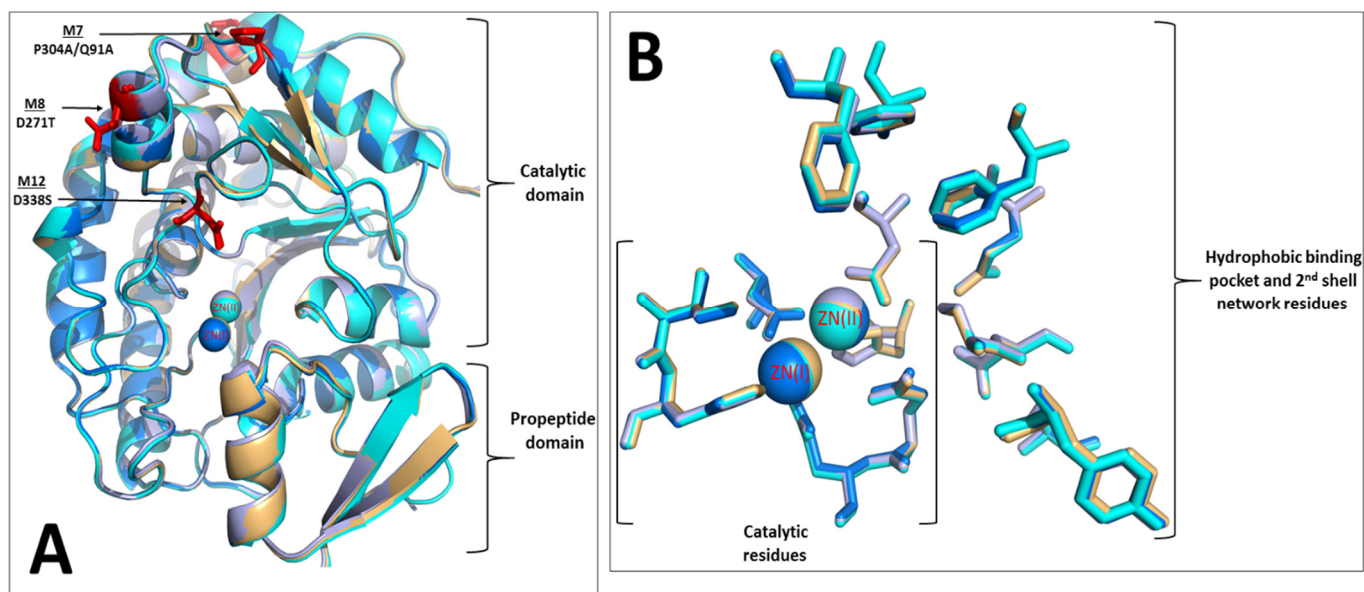
It was essential to identify the key structural differences between wild-type and mutant proteins to decipher the extent of the mutation-induced changes that occurred in the tertiary structure of LapA, which might contribute to thermolability.

Structures of LapA thermolabile mutants were solved in the form of proenzyme, bearing the covalently linked propeptide domain. This was due to a simplified crystallisation protocol (described in section 2.8.2.3), avoiding enzymatic processing of prodomain and requirement of successive additional purification cycles and purity checks.

A structural comparison was carried out by structural superimposition of two - domain LapA precursors, whereas the functional thermolability assays were performed on AP1-activated single – domain LapA mutant enzymes.

### **6.4.1 Comparison of LapA precursor (*wtproLapA*) and M7, M8, M12 structures**

Interesting (M12) and relevant (M7, M8) mutants identified in thermolability assays (section 6.5) were structurally compared to the wild-type protein (*wtproLapA*). Structures of most thermolabile (M7 – double mutant, M8) and the most thermostable (M12) mutants were further compared to each other and to the wild-type enzyme.



**Figure 6.9** Superimposition of LapA wild-type and mutant proenzyme structures. *wtproLapA* – cyan, M7 – light orange, M8 – light blue, M12 – marine; **A** Overall view of overlaid structures, mutated residues are shown in red sticks; **B** Detailed superposition of the active site, 2<sup>nd</sup> shell electrostatic network, and hydrophobic specificity pocket residues. Atomic coordinates were superimposed using PyMOL.

As demonstrated in **Figure 6.9**, the tertiary structures of *wtproLapA* and the thermolabile mutants are almost identical. This is even more apparent when observing functionally important residues (**Figure 6.9, B**). Site-directed thermolability mutation did not appear to induce any visible conformation changes within the active site, second shell electrostatic network, or hydrophobic binding pocket residues. This finding was very promising because it provided firm evidence that no adverse effects on substrate binding or impairment of hydrolytic activity would likely be observed for these mutants.

#### 6.4.1.1 Alignment of secondary structure elements (SSE)

Analysis using the STRIDE algorithm (Frishman and Argos, 1995) only highlighted a few minor differences between wild-type LapA precursor and mutant precursor proteins. The P304A mutation caused lengthening of the  $\beta_6$  strand in the M7 mutant, as S277 was able to form a side chain to backbone H-bond to A304. By contrast, SSE visualisation in PyMOL suggested that this strand is shorter by 2 residues in the M7 mutant. Also, STRIDE predicted

2 new beta bridges over a single residue in M7. M8 showed identical to wild-type secondary structural elements. Finally, STRIDE prediction of M12 secondary structure elements identified E111 residue to extend the  $\alpha 3$  helix by a single residue. Based on the SSE analysis using STRIDE, a lack of significant secondary structure differences between wild-type LapA proenzyme and mutant structures was observed, suggesting a subtle effect on the overall fold and secondary structure elements of LapA protein.

#### 6.4.1.2 Structural alignment using TM-align

For comparison of structural similarity between LapA wild-type and mutant proteins, the TM-align server (Zhang and Skolnick, 2005) was utilised.

Superimposition of the full proenzyme structures of *wtp*roLapA and M7 mutant yielded an RMSD value of 0.21 over 338 residues. Whereas, if only the residue range corresponding to the catalytic domain after activation is superimposed for these proteins, the RMSD value drops to 0.12 over 301 residues. This indicates half of the differences (change in atomic positions/distances) between *wtp*roLapA and M7 structures reside within the propeptide domain, consisting of only 37 fully defined residues. However, we are primarily interested in differences within the catalytic domain since these could provide an explanation for the observed modulated thermal properties. Overall, the RMSD scores indicated only very subtle differences between *wtp*roLapA and the thermolabile mutant structures. The TM-align results for the superimposed catalytic domains are provided in **Table 6.10**.

Superimposition (catalytic domains)	RMSD score	TM-align score	Number of residues aligned
<i>wtp</i> roLapA versus M7	0.12	0.99964	301
<i>wtp</i> roLapA versus M8	0.16	0.99936	301
<i>wtp</i> roLapA versus M12	0.13	0.99623	300

**Table 6.10** TM-align results from the superimposition of *wtp*roLapA and mutant structures for the catalytic domain residues. TM-score > 0.9 shows high structural identity.

### 6.4.1.3 Superimposition of tertiary structures

No major structural rearrangements were observed in the LapA structure as a result of the thermolabile mutations M7, M8, and M12. **Figure 6.10** illustrates the effect of the site-directed thermolability mutations on the C $\alpha$  backbone. With respect to the C $\alpha$  trace, only minor differences in the following regions were found: 81 - 95, 140 - 144, 210 - 220, 270 - 275, 310 - 314 and 335 - 360, all of which appear to cluster on one side of the molecule adjacent to the active site and the mutations (**Figure 6.10, A**).

Further detailed analyses of individual juxtaposed structures are provided below.

#### *wtproLapA* vs. M7

Residue Q91 makes an electrostatic interaction with the side chain carbonyl of N87, which is glycosylated and has a covalently linked NAG to its amide nitrogen. This interaction is tight and measures 3.0Å in the *wtLapA* structure. The absence of this contact in M7 causes the N87 - attached glycan to change orientation. Further, the Q91A mutation caused a backbone movement of a loop of residues 86 - 95, including markedly altered orientation of their respective side chains. In addition, the second mutation, P304A, caused local and long-distance backbone movements, which translated to the C310-C314 loop, which is involved in a disulphide bridge adjacent to the active site and could have implications for substrate affinity, as Y312 (in this loop) is involved in binding the substrate. **Figure 6.10, B** illustrates the effect of the M7 mutation on the catalytically important 310 - 314 loop, adjacent to Y312. Atomic transitions, of 0.4 - 0.5 Å, were observed throughout this loop, which could result in altered substrate affinity and altered thermolabile behaviour.

#### *wtproLapA* vs. M8

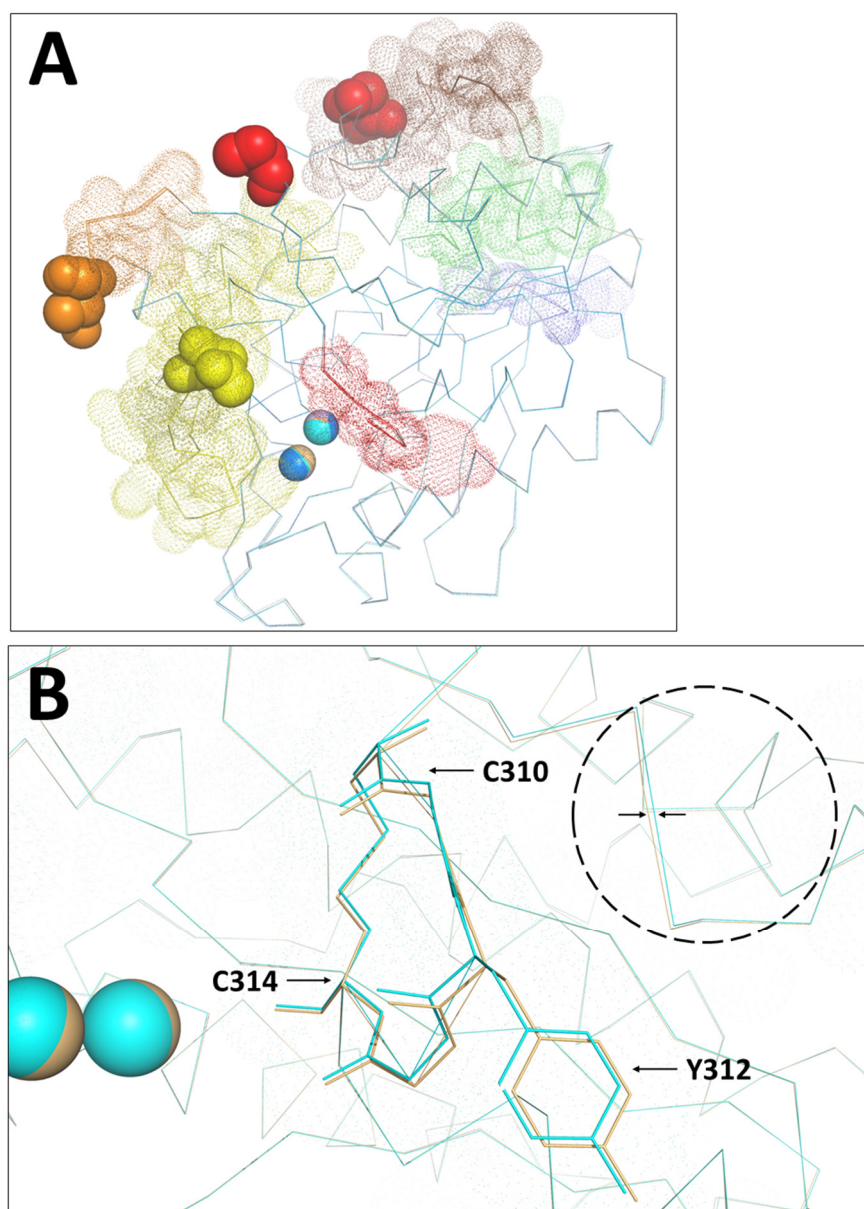
The expectation of the D271T substitution was to weaken the 3.2Å electrostatic interaction with K267. In the M8 mutant structure, we observed the complete abrogation of this electrostatic interaction, which caused 0.4Å, 0.3Å and 0.2Å movements of the C $\alpha$  atoms of the adjacent A272, G273, K274 residues, respectively. This ripple effect was commonly observed in other regions flanking the mutation sites. Also, movement of the nascent T271 residue added to the pre-existing T270 and T269, forming a surface-exposed triplet of threonine side-chains, which would contribute to an additional entropy-driven destabilising factor of surface-exposed non-polar side-chains.

Interestingly, this mutation also affected the C310 - C314s loop, altering the position of Y312 C $\alpha$  atom by 0.3Å. The D271T mutation in the M8 structure caused 0.1 - 0.3Å

positional displacement for the atoms of residues going in both directions from the mutation. Towards the C-terminus, in particular, these backbone and side chain adjustments propagated to the 360 – 365 residue region. And, close to the N-terminus, shifts in the 134 – 144 and 210 – 220 residue regions were apparent. In fact, from this detailed analysis, all differences were more apparent compared to other mutants (M8 showed the highest RMS difference score among all mutants, **Table 6.10**).

#### ***wtproLapA* vs. M12**

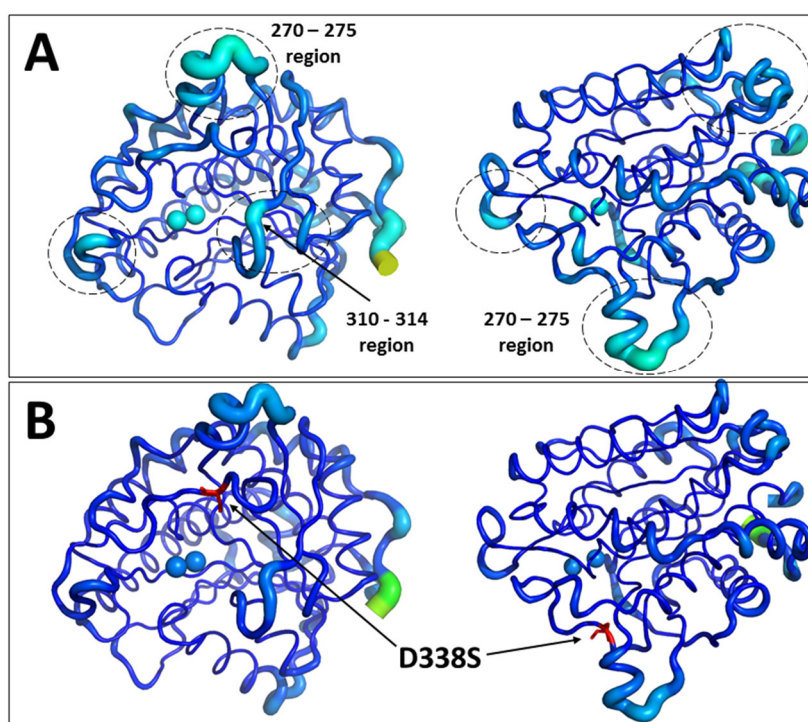
The D338S mutation caused movement of the adjacent hydrophobic binding pocket residue F335 by 0.3 – 0.4 Å, which could have implications in substrate binding, as this residue has been observed to rotate upon substrate binding (**Figure 5.14**). No other major alterations in structure towards the C-terminus were observed in M12. However, significant changes to the protein backbone and side chains were detected nearer to the N-terminus in M12 mutant, particularly in the 81 – 86 and 180 – 189 residue regions, whereby the differences in atom positions were equal to 0.3 – 0.4 Å.



**Figure 6.10** Detailed structural comparison of LapA wild-type and thermolabile mutant structures. M7, M8, M12 and wtproLapA structures are superimposed. **A** Shows the common protein regions, where backbone and side chain movements were observed. M7 (P304A, Q91A), M8 (D271T), M12 (D338S) mutations are depicted in red, orange and yellow spheres, respectively. Moving residue regions: 81 – 95 (brown dots), 140 – 144 (blue dots), 210 – 220 (green dots), 270 – 275 (orange dots), 310 – 314 (red dots), 335 – 360 (yellow dots). Propeptides are omitted for clarity. **B** Overlay of wtproLapA (cyan) and M7 (light orange). Closer view of the moving 310 – 314 loop (sticks) and 81 – 95 region (dashed circle, arrows indicate movement in C $\alpha$  backbone). Zn atoms are depicted as spheres.

#### 6.4.1.4 Analysis of molecular flexibility via structural B - factors

Experimentally determined B factors (a parameter of thermal motion) can give an indication of changes in the local backbone or side-chain flexibility. The most thermostable mutant, M12, displayed quite remarkable changes in overall structural rigidity. **Figure 6.11** demonstrates significantly lowered overall B-factors for M12. The differences are seen in previously flexible surface loops compared to *wtproLapA*. For example, the average whole chain (main chain + side chain) B factor in M12 drops from 33.9 Å<sup>2</sup> to 26.23 Å<sup>2</sup> and 28.2 Å<sup>2</sup> to 20.5 Å<sup>2</sup> in 270 – 275 and 310 -314 residue regions, respectively (**Figure 6.11**, A). The overall reduction in average whole chain B-factor value across all residues was from 21.7 Å<sup>2</sup> (*wtproLapA*) to 16.9 Å<sup>2</sup> (M12) (**Table 6.11**). The experimentally observed reduction in B values for the M12 mutant suggests an overall increase in structural rigidity, which is likely responsible for the significantly improved thermostability of this mutant.



**Figure 6.11** Comparison of overall B factor representation for *wtproLapA* and M12 structures. Average B-factor in each residue is depicted by a 'putty representation'. Lower to higher B-factor values go from dark blue to red, respectively. Both sides (180° rotation) of the molecule are represented. Propeptide is omitted for clarity. **A.** B-factor putty of *wtproLapA* protein, significantly higher B-factor areas are circled (dashed lines); **B.** B-factor putty of M12 mutant protein. Location of the D338S mutation is shown in red sticks.

Interestingly, the thermolability mutants M7 and M8 showed no significant changes in overall B-factor compared to *wtproLapA* (**Table 6.11**). In the structure of M8, the most substantial increase in flexibility was observed in the 182 – 188 residue region, with an average whole chain B-factor shift from 25.8 Å<sup>2</sup> to 32.3 Å<sup>2</sup>. This was accompanied by a 0.2 – 0.45Å residue movement. This loop is adjacent to the active site pocket and connects the β3 strand to the α4 helix. The B-factor data (summarised in **Table 6.11**) suggests that the overall increase in thermal motion and flexibility for M7 and M8 mutant structures could not be determined from simply all-atom temperature factors and, importantly, more localised differences need to be deduced.

<b>Protein</b>	<b>T_atoms*</b>	<b>B_main chain</b>	<b>B_side chain</b>	<b>B_whole chain</b>
<i>wtproLapA</i>	2369	19.6	23.9	21.7
M7	2395	18.7	22.6	20.7
M8	2428	18.9	24.3	21.5
M12	2372	14.9	19.2	16.9

\* T\_atoms – total number of atoms measured;

**Table 6.11** Temperature factor (B) analysis of LapA crystal structures. The analysis was conducted using dedicated temperature factor analysis software in CCP4 (Winn et al., 2011).

## 6.5 Functional studies of LapA

### 6.5.1 High-temperature inactivation assays for M1 – M6 thermolabile mutants

The wild-type and mutant recombinant enzymes were expressed and secreted as pro-proteins, which were subsequently activated with Flavourzyme-derived AP1 protease, as detailed in section 2.13.4.1. The observed N-terminal sequence for each of these activated processed wild-type and mutant proteins was identical to *wLapA*, described previously, namely, possessing an 'AVT' sequence extension to the original N-terminus.

#### 6.5.1.1 Full temperature range comparison

A LapA thermolability assay was designed to determine the thermostability of LapA variants at various temperatures, relevant to industry. Aliquots of LapA enzymes were incubated at 5 different temperatures: 37, 50, 60, 70 and 80 °C. Measurements of residual specific activity at 37 °C served as a reference, where no loss of activity was anticipated. Measurements in the 50 – 60 °C range were key to the identification of potential thermolabile LapA variants. Finding a LapA mutant that could be completely inactivated >60 °C would be considered successful, according to the industrial requirement from Biocatalysts Ltd. Incubation in the 70 – 80 °C range was expected to completely denature and inactivate LapA proteins. As a starting point, complete inactivation conditions for the LapA enzyme were as follows: 80 °C for 30 – 60 min. Therefore, to be considered successful, a mutation should induce a significant reduction in LapA thermal stability and should maintain 80 – 100% specific activity equivalent to *wLapA*.

Full temperature range profiles were obtained for LapA proteins, incubating each enzyme at 5 distinct temperatures and withdrawing aliquots for measurement of residual specific activity, promptly, every 15 minutes for 60 – 90 minutes. The thermolability assay protocol is described in section 2.10.3. An initial assay, to assess thermostability of LapA mutants, was implemented in a 3ml cuvette format and using a conventional spectrophotometer. Once initiated, the reaction was allowed to proceed, and the difference in absorption was manually recorded after 60 seconds. Subsequently, in the second round of mutagenesis, this manual cuvette assay was replaced by a more automated microtiter plate assay. Specific activities, measured at 37 °C, are presented in **Table 6.12**.

The M1 (P155G) mutation caused a reduction in activity of ~ 65%. It is likely that glycine-induced flexibility at this position was potentially detrimental to fold integrity. This can be supported by the observed concomitant reduction in secretion levels of the M1 LapA mutant (section 6.1.2, **Figure 6.2**). A more conservative P155A mutation could be tried in the future to see whether a glycine substitution caused excessive backbone plasticity. Also, the observed M1 thermolability profile was not promising, as it retained 77% and 46% of initial activity at 60 °C following 30 and 60-minute incubation, respectively.

The M3 (R252K) mutation showed a substantial loss of specific activity, while maintaining a high degree of thermostability at 60 °C; 81% and 74% of initial activity following 30 and 60-minute incubation, respectively. Substitution of Arg252 with another positively charged residue – lysine – may not have resulted in complete elimination of R252 – E226 salt bridge. Hence, the altered (but not abolished) electrostatic interaction at this position only had an adverse effect on LapA activity and not thermostability.

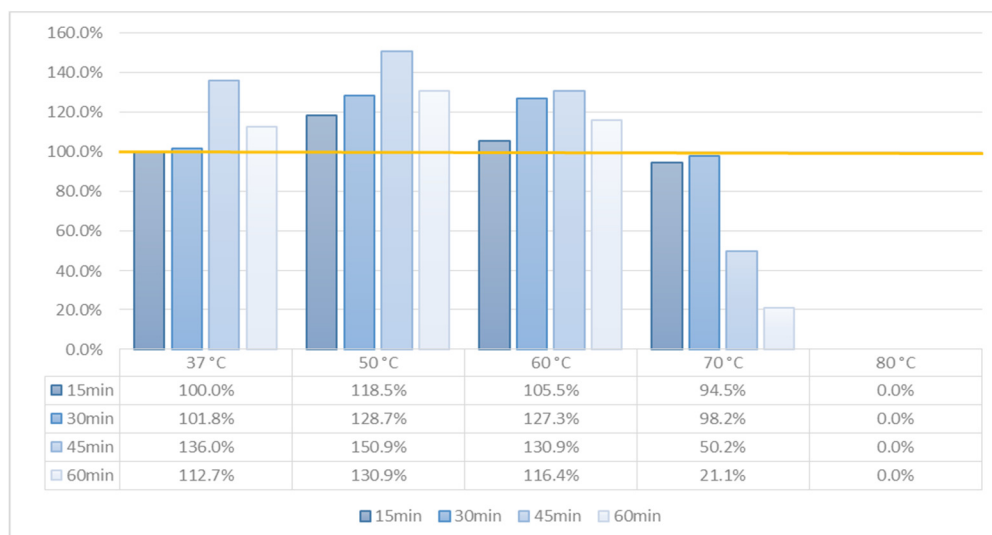
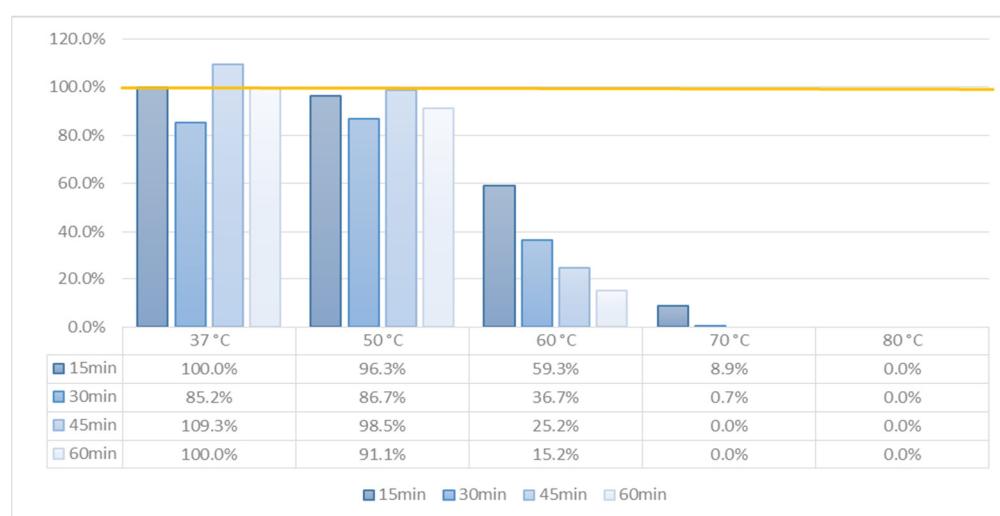
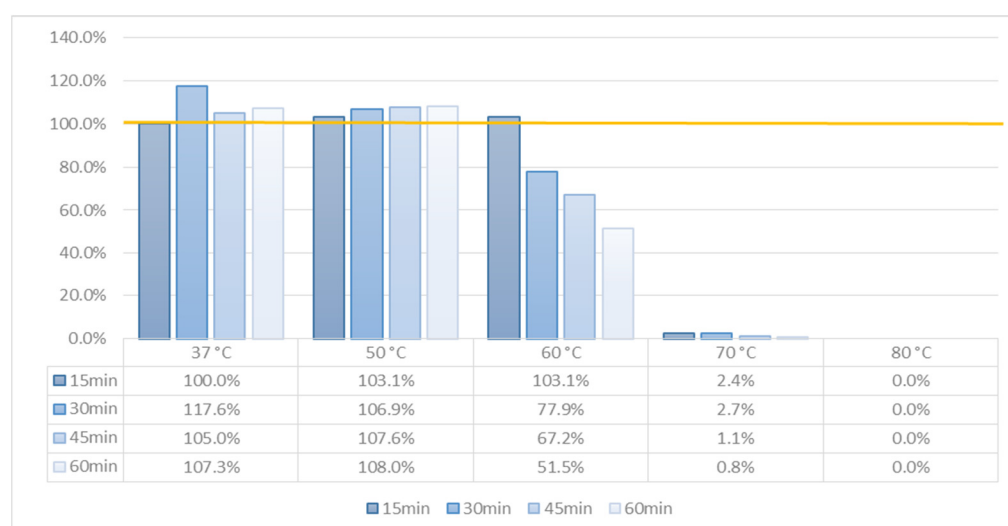
The M5 (N87A) mutation resulted in a major loss of LapA activity and the remaining specific activity of 25% did not meet qualifying criteria. Therefore, a full thermolability profile was not obtained for this mutant. From the structure solution of *wt*LapA, we found the N87 residue to be glycosylated. Glycosylation is known to have a thermodynamic stabilization effect on protein structures (Shental-Bechor and Levy, 2008). Therefore, decreased thermostability of the M5 mutant was expected, following the N87A mutation. However, the significantly reduced activity and lowered secretion levels of M5 (section 6.1.2) highlight the importance of glycosylation in proteogenesis and folding of the LapA enzyme.

The M6 (C310A) mutation had a detrimental effect on LapA specific activity. M6 only retained 0.45% of *wt*LapA activity. Hence, a full thermolability profile was not completed for this protein either. C310 forms an intramolecular disulphide bond with C314. This post-translational modification is considered highly stabilizing and necessary for proper folding of LapA. However, the high expression levels acquired for M6 (section 6.1.2) suggest that formation of the C310 – C314 disulphide bond may not be rate limiting step in the LapA folding pathway, although, appropriate positioning of the C314 carbonyl oxygen likely is crucial for enzyme ability to bind substrate. The equivalent carbonyl oxygen of cysteine residue C227, in the Lap structure from *Vibrio proteolyticus*, was shown to be involved in interaction with the Leu-Leu-Leu substrate (Kumar et al., 2007). Thus, elimination of the Cys-Cys bond in LapA could compromise its catalytic function.

In conclusion, both M2 and M4 mutations had no major effect on LapA specific activity. Initial thermolability profiles showed a significant reduction in thermostability (**Figure 6.12**) for both mutants. Taken together, with similar to wild-type expression levels (section 6.1.2, **Figure 6.2**), M2 and M4 appeared to be the best candidate mutants from the first round of matagenesis.

LapA variant	Specific activity (U/mg)	Activity retained (%)
WT LapA	140.15 $\pm$ 4.8	-
M1	48.48 $\pm$ 4.4	34.60%
<b>M2</b>	<b>128.46 <math>\pm</math> 13</b>	<b>91.40%</b>
M3	87.14 $\pm$ 13	62.10%
<b>M4</b>	<b>143.59 <math>\pm</math> 9.9</b>	<b>102.10%</b>
M5	35.01 $\pm$ 5.5	25%
M6	0.64	0.45%

**Table 6.12.** Specific activity values observed for LapA variants in the 1<sup>st</sup> round of mutagenesis. Retained percentage activities for mutant proteins were calculated in relation to wild-type LapA.

**A****B****C**

**Figure 6.12** Temperature inactivation profiles of LapA variants. **A** – wild-type LapA, **B** – M2 mutant, **C** – M4 mutant. 100% reference (yellow line) is specific activity measured at 37 °C, following 15 minute incubation.

In Figure 6.12, the remaining percentage activity was plotted against the full temperature range for *wtLapA*, M2, and M4 mutants. The *wtLapA* enzyme exhibited very thermostable behavior at 60 and 70 °C. Additionally, *wtLapA* maintained consistently higher specific activity values across 50 – 60 °C range, which was not unexpected since native LapA, produced in *Aspergillus oryzae*, also was shown to have a temperature optimum at 60 °C (Matsushita-Morita et al., 2011). M2 and M4 showed a progressive decline in specific activity within a 60 minute incubation period, at 60 and 70 °C (Figure 6.12, B and C). Also, no overall increase in activity was observed when incubating these enzymes at 50 °C, suggesting negatively affected thermostability, even at this temperature.

Further detailed thermolability analysis of selected mutants, M2 and M4, is provided in the following sections.

#### 6.5.1.2 Analysis of thermal inactivation of M2 and M4 mutants at 60 °C

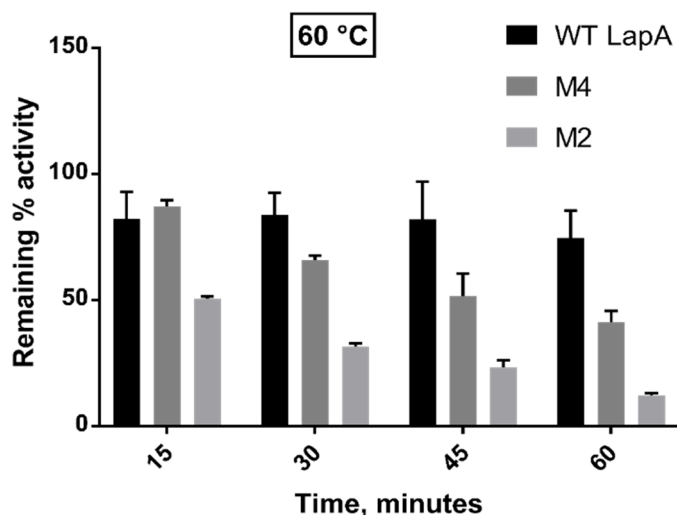
Remaining activities (as percentage) for *wtLapA* and mutants were expressed, taking 155 U/mg specific activity as 100% reference (Table 6.13). This value was the highest activity measured for the wild-type enzyme, after 15-minute incubation at 37 °C. Data normalization was necessary to perform a non-linear regression fit to the inactivation data and to compare the remaining percentage activities at different temperatures.

#### *Inactivation data at 60 °C*

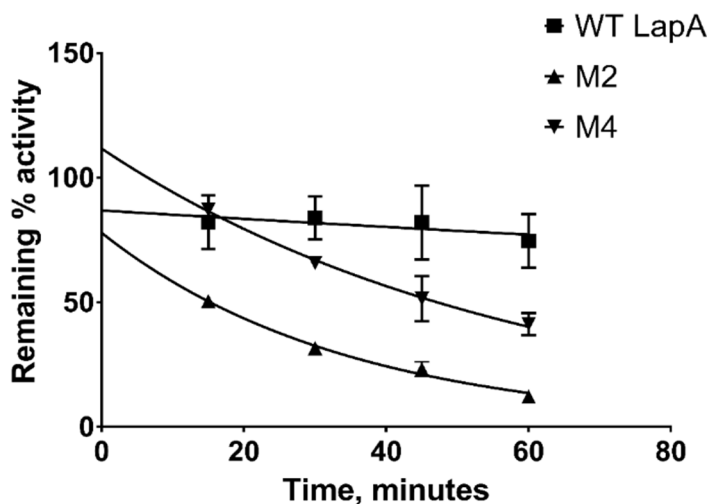
	Minutes			
	15	30	45	60
<b><i>wtLapA</i></b>	82.20%	83.90%	82.10%	74.65%
<b>M2</b>	50.50%	31.65%	23.25%	12.15%
<b>M4</b>	87.25%	65.80%	51.50%	41.25%

**Table 6.13** Remaining percentage activities of LapA proteins, following incubation at 60 °C. Data normalised to 155 U/mg (100%) activity value. Given percentage values are means of at least three independent measurements. The data are visualised in **Figure 6.13**.

The inactivation data plot in **Figure 6.14** and percentage values in **Table 6.13** clearly demonstrate that WT LapA retains substantially more activity at 60 °C as opposed to mutant enzymes (reflected in the height of the fitted curve on Y-axis). M2 LapA mutant retained the lowest amount of activity at this temperature.



**Figure 6.13** Comparison of remaining percentage activity of LapA proteins at 60 °C, within the first 60 minutes of incubation. Standard deviation (SD) of each measurement is presented by an error bar. All measurements were carried out in triplicate or quadruplicate.



**Figure 6.14** A single exponential rate model fitted to LapA protein inactivation data at 60 °C by non-linear regression. Thermal inactivation rate constant ( $k$ ) and half-life were obtained from best-fit to inactivation data of LapA wild-type and mutant enzymes.

According to the fitted exponential rate decay model the *wt*LapA enzyme exhibited the longest (351 min) half-life, compared to the thermolability mutants M2 and M4 (**Table 6.14**). Both single point mutations decreased the inactivation time of LapA, at 60 °C, with 40.7 and 23.9 minute half-lives for M4 and M2 mutants, respectively. The determined inactivation rate constants showed a substantial increase in inactivation rate, with 8.62- and 14.71-fold for M4 and M2 mutants, respectively. This is represented also by significant differences in  $\Delta\Delta G$  values between wild-type LapA and mutants, at this temperature.

**Table 6.14** Kinetic parameters for thermal inactivation of wild-type and mutant LapA enzymes.

Enzyme	$t_{1/2}$ <sup>a</sup> (min)	Rate constant of inactivation $k$ (min <sup>-1</sup> ) <sup>a</sup>	Increase in inactivation rate	Difference in activation free energy of inactivation $\Delta\Delta G$ (kJmol <sup>-1</sup> ) <sup>b</sup>
<i>wt</i> LapA	351	0.0019		
M2	23.9	0.0291	14.71	-7.45
M4	40.7	0.017	8.62	-5.97

<sup>a</sup> Thermolability profiles of LapA proteins were determined from 3-4 independent measurements at 60 °C. Protein half-lives, inactivation rate constants and standard errors were calculated using nonlinear regression to fit a single exponential rate model to the inactivation data (using Graph-Pad Prism software).

<sup>b</sup>  $\Delta\Delta G = \Delta G_{\text{mutant}} - \Delta G_{\text{wild-type}} = -RT\ln(k_{\text{mutant}}/k_{\text{wild-type}})$

Based on the remaining percentage activity and calculated half-life values at 60 °C, M2 emerged as the most thermolabile mutant, from the first round of mutagenesis. A 15-fold increase in inactivation rate for M2 clearly demonstrates the negative effect of the P304A single point mutation on LapA thermostability.

## **6.5.2 High-temperature inactivation assays for M7 – M13 thermolability mutants**

A new microtiter plate enzymatic LapA assay format was introduced to expedite the analysis. Also, 300  $\mu$ L total assay volume allowed enzyme handling using an electronic multi-channel pipette, which in turn permitted more time points of thermal inactivation to be recorded. This way, 80 and 90-minute data points were added to the analysis, ensuring a more accurate fit of a single exponential rate model to the inactivation data.

The LapA inactivation data, presented herein, corresponds to the thermolability assays performed at 50 and 60 °C, since this represents a desirable inactivation temperature range for industrial application.

Remaining percentage activities for *w*tLapA and mutants were expressed, taking 155 U/mg specific activity as 100% reference. This value was the highest activity measured for the wild-type enzyme after 15-minute incubation at 37 °C. This data normalisation was necessary to perform non-linear regression fits to the inactivation data and to compare remaining percentage activities at different temperatures appropriately.

### **6.5.2.1 Thermal inactivation profile for M7 double mutant**

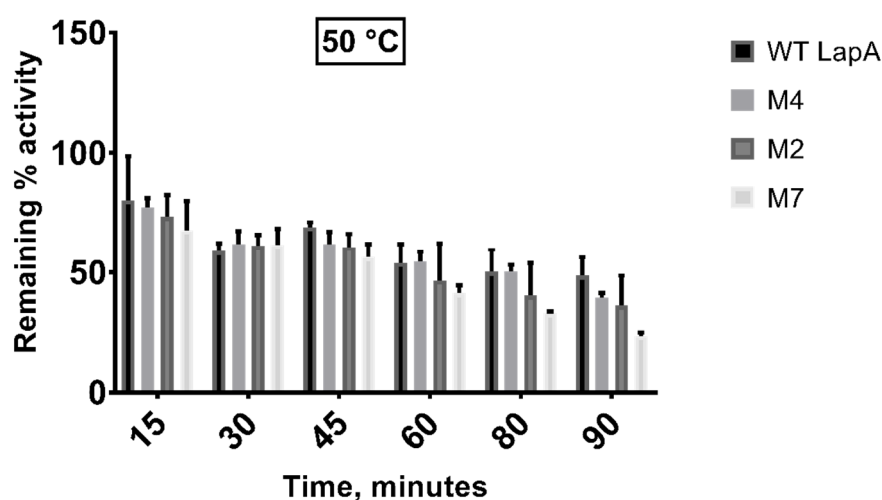
The previously identified most thermolabile LapA variants (M2 and M4) were combined to yield a double mutant, M7, to explore whether or not the thermostability of LapA could be decreased further via the additive effect of the two single point mutations. Fresh protein batches of *w*tLapA, M2 and M4 were prepared for analysis.

The inactivation data of M2/M4 (single) and M7 (double) mutants is presented in the following sections (**Table 6.15**, **Table 6.16**, **Table 6.17**, **Table 6.18**).

### *Inactivation data at 50 °C*

	Minutes					
	15	30	45	60	80	90
<b>wtLapA</b>	80.30%	59.03%	69.10%	53.97%	50.30%	48.80%
<b>M4</b>	77.40%	62.05%	61.82%	54.59%	50.42%	39.45%
<b>M2</b>	73.40%	61.20%	60.73%	46.63%	40.50%	36.30%
<b>M7</b>	67.87%	61.73%	56.40%	41.33%	33.00%	23.50%

**Table 6.15** Remaining % activities of LapA proteins following incubation at 50 °C. Data normalised to 155 U/mg (100%) activity. Given percentage values are means of multiple independent measurements. The data are visualised in **Figure 6.15**.



**Figure 6.15** Comparison of remaining percentage activity of LapA proteins, at 50 °C, within the first 90 minutes of incubation. SD of each measurement is presented by error bar. All measurements were carried out in triplicate or quadruplicate.

**Table 6.16** Kinetic parameters for thermal inactivation of wild-type and mutant LapA enzymes.

Enzyme	$t_{1/2}$ (min) <sup>a</sup>	Rate constant of inactivation $k$ (min <sup>-1</sup> ) <sup>a</sup>	Increase in inactivation rate	Difference in activation free energy of inactivation $\Delta\Delta G$ (kJmol <sup>-1</sup> ) <sup>b</sup>
<i>wt</i> LapA	110.7	0.0063		
M4	93.69	0.0074	1.18	-0.49
M2	75.67	0.0091	1.46	-1.02
M7	57.14	0.0121	1.98	-1.78

<sup>a</sup> Thermolability profiles of LapA proteins were determined from 3-4 independent measurements at 50 °C. Protein half-lives, inactivation rate constants and standard errors were calculated using nonlinear regression to fit a single exponential rate model to the inactivation data (using Graph-Pad Prism software).

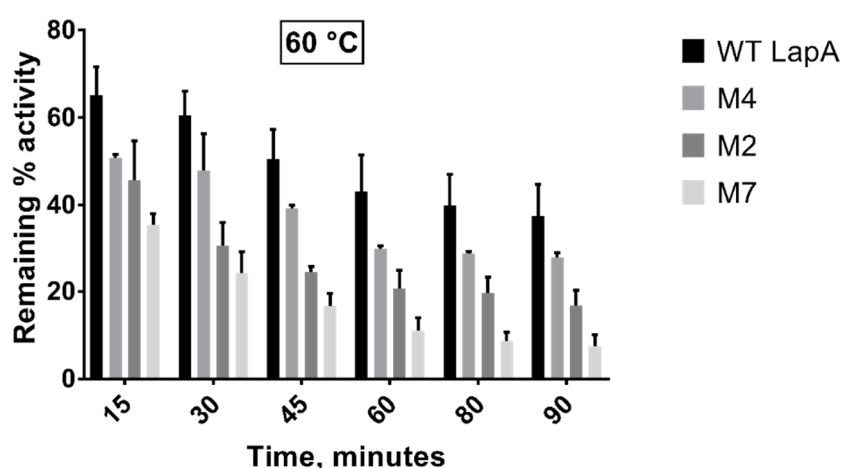
<sup>b</sup>  $\Delta\Delta G = \Delta G_{\text{mutant}} - \Delta G_{\text{wild-type}} = -RT\ln(k_{\text{mutant}}/k_{\text{wild-type}})$

The M7 mutation increased the inactivation rate of the LapA enzyme by 1.98-fold at 50 °C. At this temperature, differences in the inactivation profiles were not very significant across LapA enzymes. The M7 double mutant showed the shortest half-life, as expected. This is also reflected in only a modest difference in  $\Delta\Delta G$  values between *wt*LapA and the mutants, at this temperature.

### *Inactivation data at 60 °C*

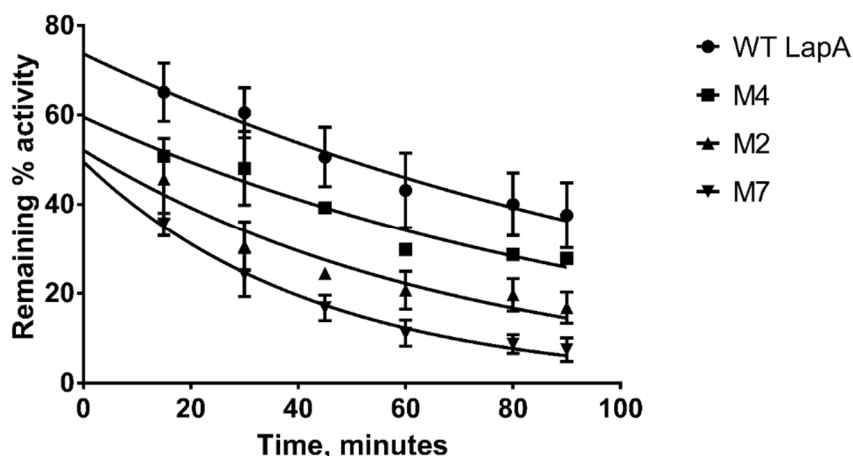
	Minutes					
	15	30	45	60	80	90
<b>WT LapA</b>	65.07%	60.47%	50.57%	43.13%	39.90%	37.50%
<b>M4</b>	50.83%	48.00%	39.23%	29.77%	28.65%	27.82%
<b>M2</b>	45.67%	30.53%	24.43%	20.67%	19.60%	16.80%
<b>M7</b>	35.50%	24.13%	16.70%	11.07%	8.60%	7.40%

**Table 6.17** Remaining % activities of LapA proteins, following incubation at 60 °C. Data are normalised to 155 U/mg (100%) activity value. Given percentage values are means of multiple independent measurements. The data are visualised in **Figure 6.16**.



**Figure 6.16** Comparison of remaining % activity of LapA proteins at 60 °C within first 90 minutes of incubation. SD of each measurement is presented by error bar. All measurements carried out in triplicate or quadruplicate.

The obtained inactivation profiles of LapA proteins at 60 °C exhibited more substantial differences. The inactivation data plots (**Figure 6.17**) clearly indicate that *wt*LapA protein retains significantly more activity at 60 °C, compared to the mutant enzymes (height of the fitted curve on Y-axis, **Figure 6.17**). M7 double mutant retained the lowest amount of activity at this temperature.



**Figure 6.17** A single exponential rate model fitted to LapA protein inactivation data, at 60 °C by non-linear regression. Thermal inactivation rate constant ( $k$ ) and half-life were obtained from best-fit to the inactivation data of LapA wild-type and mutant enzyme.

**Table 6.18** Kinetic parameters for thermal inactivation of wild-type and mutant LapA enzymes.

Enzyme	$t_{1/2}$ (min) <sup>a</sup>	Rate constant of inactivation $k$ (min <sup>-1</sup> ) <sup>a</sup>	Increase in inactivation rate	Difference in activation free energy of inactivation $\Delta\Delta G$ (kJmol <sup>-1</sup> ) <sup>b</sup>
wtLapA	87.85	0.0079		
M4	74.55	0.0093	1.18	-0.45
M2	48.61	0.0143	1.81	-1.63
M7	29.68	0.0234	2.96	-3.01

<sup>a</sup> Thermolability profiles of LapA proteins were determined from 3-4 independent measurements at 60 °C. Protein half-lives, inactivation rate constants and standard errors were calculated using nonlinear regression to fit a single exponential rate model to the inactivation data (using Graph-Pad Prism software).

<sup>b</sup>  $\Delta\Delta G = \Delta G_{\text{mutant}} - \Delta G_{\text{wild-type}} = -RT \ln(k_{\text{mutant}}/k_{\text{wild-type}})$

The M7 mutation increased the inactivation rate of LapA by nearly 3-fold, at 60 °C ( **Table 6.18**). In fact, at 60 °C, the inactivation temperature differences were very significant across LapA enzymes. The M7 double mutant showed the shortest half-life (29.68 min), with more than half of that observed for wtLapA (87.85 min). This is represented also by substantial differences in  $\Delta\Delta G$  values between wtLapA and mutants at this temperature ( **Table 6.18**).

In conclusion, the LapA thermolability assay data showed the varying extent of thermal instability introduced by the selected mutations. The two most thermolabile mutants, M4 and M2, (identified in the first round of mutagenesis) were combined in the following round to yield a double mutant, M7, with further enhanced thermolability at 60 °C. The additive effect of M2 (P304A) and M4 (Q91A) single point mutations was apparent.

### 6.5.2.2 Validation of the non-linear regression model to the inactivation data

To validate the fit between the plotted data points and the fitted exponential one-phase decay model (**Figure 6.17** and **Table 6.18**), visual plots of residuals (**Figure 6.18**) and fitted vs. observed data points (**Figure 6.19**) were used. Statistical parameters of the best-fit values also are provided in **Table 6.19**.

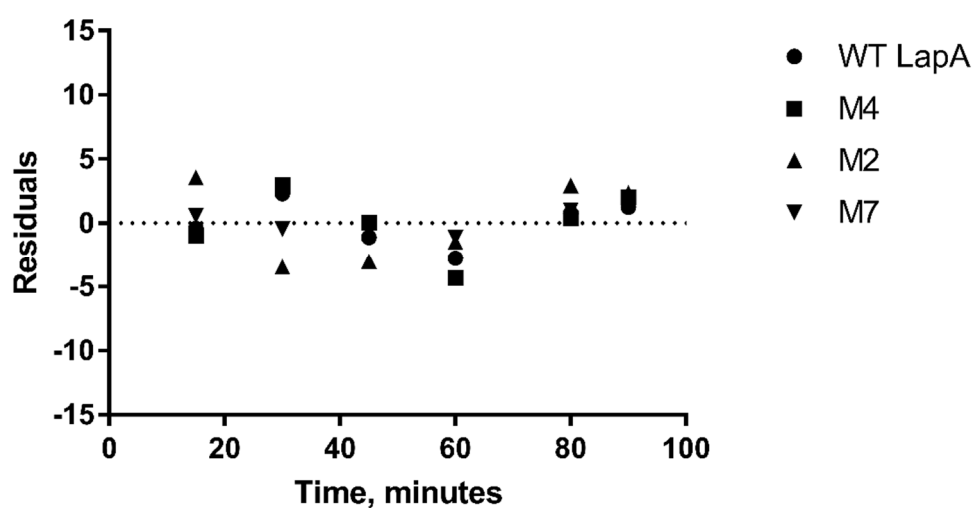
**Table 6.19** Statistical evaluation of exponential one-phase decay model fitted to the inactivation data of select LapA thermolability variants.

	<i>wt</i> LapA	M4	M2	M7
$k^a$	0.005425 to 0.01044	0.006806 to 0.01189	0.009899 to 0.0191	0.01936 to 0.02782
Half Life <sup>a</sup>	66.38 to 127.8	58.3 to 101.8	36.28 to 70.02	24.91 to 35.79
Degrees of Freedom <sup>b</sup>	16	10	16	16
R square <sup>b</sup>	0.7474	0.8783	0.769	0.9291

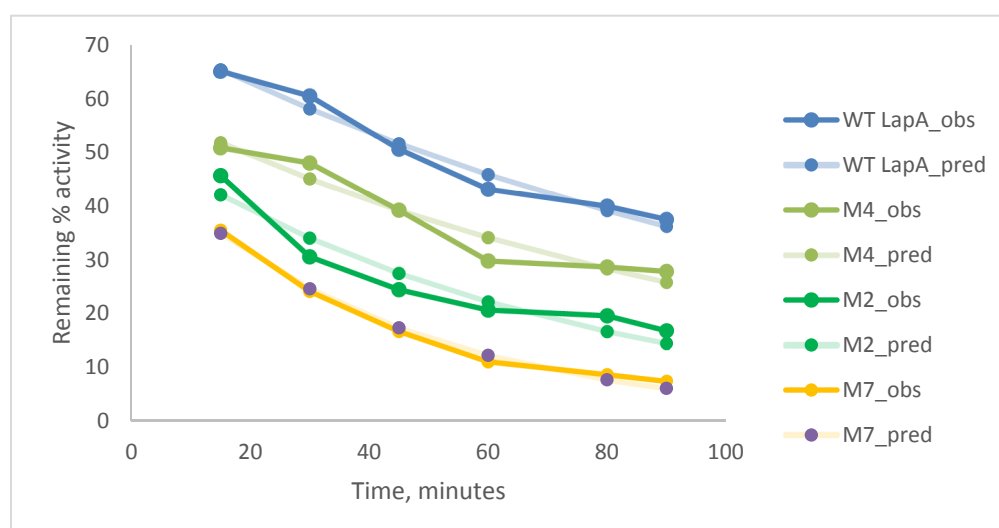
<sup>a</sup> 95% CI (profile likelihood)

<sup>b</sup> Goodness of Fit

In **Figure 6.18**, apart from M2, the data points are quite tightly distributed along the zero dotted line, suggesting a good model-to-data fit. Similarly, **Figure 6.19** shows a reasonably good fit of the model to the inactivation data, particularly for the M7 mutant. The almost ideal model-to-data fit of M7 also can be seen from the narrow confidence interval of protein half—life parameter in **Table 6.19**.



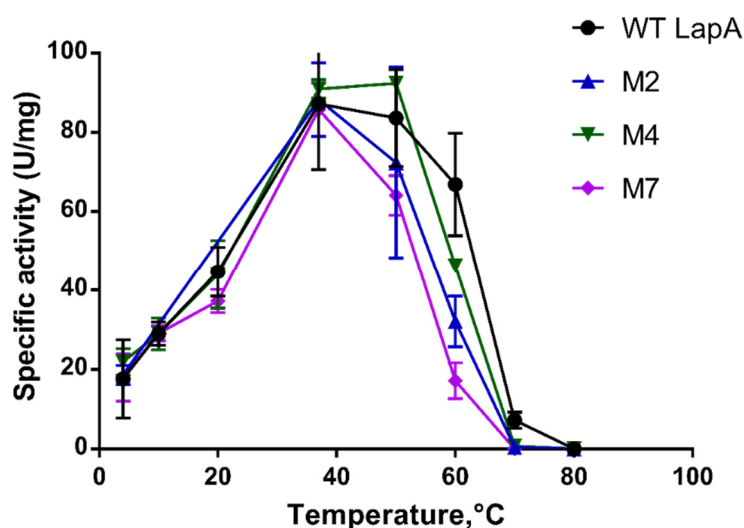
**Figure 6.18** Plot of residuals linked to LapA inactivation data fit in **Figure 6.17**.



**Figure 6.19** Overlay of observed versus predicted remaining activity values relative to the LapA inactivation data fit in **Figure 6.17**. Solid curves represent observed data (\_obs), and transparent curves correspond to values predicted by fitting the exponential decay model to the inactivation data (\_pred).

### 6.5.2.3 Temperature-induced decay for specific activity of LapA proteins

In this analysis, the observed absolute specific activity (U/mg) for *wt*LapA and M2, M4, M7 thermolability mutants were compared (**Figure 6.20**). It was important to perform this comparison to further validate the identified differences in thermostability across the different mutants. The specific activity values herein are not expressed as a relative percentage (by 155 U/mg value) but are represented as absolute numbers, calculated according to section 2.10.1.



**Figure 6.20** Comparison of determined LapA specific activity values, following 60 min incubation at increasing temperatures. Error bars for each data point are calculated from 3 – 4 independent measurements.

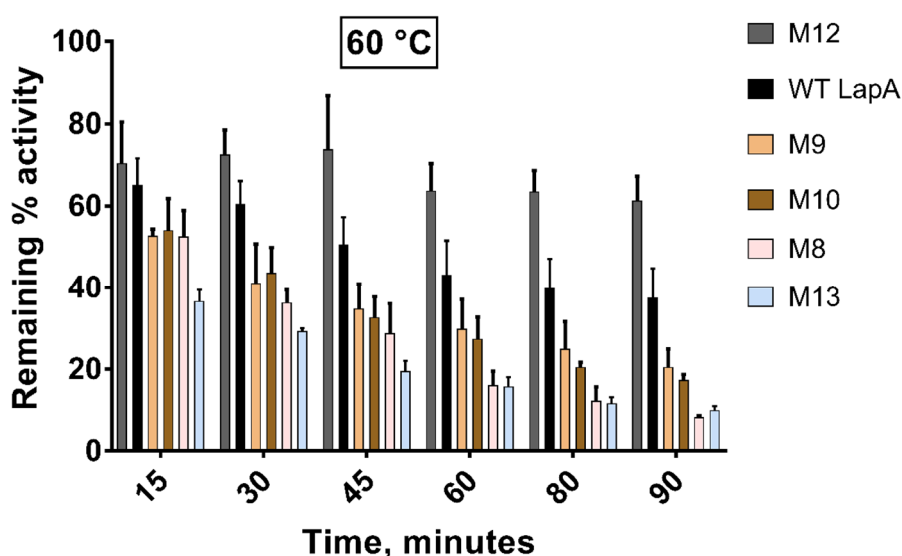
As shown in **Figure 6.20**, within the 4 - 37 °C temperature interval the specific activity is similar for each of the *wt*LapA and M2, M4, M7 mutants. Specific activity profiles begin to diverge at 40 °C and are completely distinct by 50 °C. In agreement with the results from the regression analysis (see **Figure 6.17**), the M7 double mutant displays the lowest specific activity at temperatures above 40 °C, with the most rapid activity decay of the mutants compared.

Based on these activity profiles, it is clear that the optimal activity temperature range for *wt*LapA is between 37 – 40 °C. This is quite different from reported optimal activity temperature of 60 °C, first described by Matsushita-Morita et al. (2011). This *wt*LapA protein batch displayed significantly increased thermolabile behaviour compared to the batch from the 1<sup>st</sup> round of mutagenesis (section 6.5.1.1). These differences were most

evident following incubation at 70 °C. This can be clearly seen in **Figure 6.12** (A), where *wtLapA* batch 1 retained over 90% of its original activity after 15 minute incubation at 70 °C. One possible explanation for the batch-to-batch thermostability difference between *wtLapA* proteins could be due to the additional cooling step at 4 °C before the activity measurement in the 2<sup>nd</sup> round of mutagenesis (section 2.10.3). In the latter thermostability 300 µL assays, the enzyme samples were incubated at 4 °C for 5 minutes before initiating the reaction, compared to the 3 mL initial assay protocol, whereby the incubation step was carried out on ice for 3 minutes. This step was necessary to normalize the reaction kinetics for LapA samples incubated at a wide range of temperatures (37 – 70 °C). Possibly, batch 1 of *wtLapA* achieved a lower degree of cooling, hence showed an increase in residual activity, following 70 °C pre-incubation. However, even more importantly, we noticed a possible degeneration of the developed recombinant wild-type LapA *P.pastoris* strain. Numerous subsequent growth attempts of this strain, from glycerol stock, resulted in significantly reduced biomass levels and reduced specific activity of purified LapA protein. Therefore, it is very likely that batch 1 was the most representative of wild-type LapA and batch 2 already had slightly altered thermostability due to the latterly identified expression strain repercussions. Importantly, no significant changes in growth or biomass were observed for the expression strains of M2 and M4 LapA mutants and subsequent batches of purified protein.

#### 6.5.2.4 Functional characterisation of M8 – M13 site-directed mutants

Temperature inactivation data for M8 – M13 thermolability mutants were collected over the full temperature range (37 – 80 °C), however, only the 60 °C segment is presented below (**Figure 6.21**), since the most significant changes occurred at this temperature (as previously determined). At 60 °C, inactivation temperature differences were very significant across M8 – M13 LapA mutant enzymes. The bar plot, shown in **Figure 6.21**, clearly demonstrates that M12 protein exhibited increased thermostability, while M8 and M13 mutants showed the most rapid decline in specific activity at 60 °C.



**Figure 6.21** Comparison of remaining % activity of LapA proteins at 60 °C, within the first 90 minutes of incubation. SD of each measurement is presented by an error bar. All measurements are carried out in triplicate or quadruplicate.

As demonstrated in **Table 6.20**, the M8 double mutant showed the shortest half-life (29.55 min), with almost 1/3 of that observed for wild-type LapA (87.85 min) at this temperature. Interestingly, the M8 single site – directed mutant exhibited a comparable thermolability profile to the M7 double mutant, which showed 29.68 min half-life. M7 and M8 LapA mutants also displayed an almost identical increase in the inactivation rate, 2.96- and 2.97-fold, respectively, as well as similar differences in their activation free energy of inactivation ( $\Delta\Delta G$ ) at 60 °C.

M10 displayed a lower degree of thermolability (45.49 min half-life). However, this estimate is close to the M2 mutant (48.61 min half-life) ( **Table 6.18**). Structurally, both M2 (P304A)

and M10 (P275A) involve elimination of proline from the distant-to-active-site loop. Hence, the exerted effect on thermolability may be very similar.

M11 (D316A mutation) almost completely abolished the hydrolytic activity of LapA.

M12 (D338S mutation) showed an opposite effect with regards to thermostability. This mutation resulted in an impressive 3.5-fold increase in half-life at 60 °C. The exact mechanism of induced thermostability for this mutant is not clear.

M13 appeared as a comparatively short-lived enzyme (35.78 min half-life). However, the S333G substitution had an adverse effect on protein specific activity, retaining only 60 – 70 % that of *w*LapA, therefore, this mutation was not considered successful.

**Table 6.20** Kinetic parameters for thermal inactivation of wild-type and mutant LapA enzymes. The M7 double mutant from the previous analysis was included for comparison.

Enzyme	$t_{1/2}$ (min) <sup>a</sup>	Rate constant of inactivation $k$ (min <sup>-1</sup> ) <sup>a</sup>	Increase in inactivation rate	Difference in activation free energy of inactivation $\Delta\Delta G$ (kJmol <sup>-1</sup> ) <sup>b</sup>
<i>w</i> LapA	87.85	0.0079		
M7	29.68	0.0234	2.96	-3.01
M8	29.55	0.0235	2.97	-3.02
M9	57.57	0.012	1.53	-1.17
M10	45.49	0.0152	1.93	-1.82
M12	311	0.0022	-	3.50
M13	37.84	0.0183	2.32	-2.33

<sup>a</sup> Thermolability profiles of LapA proteins were determined from 3-4 independent measurements at 60 °C. Protein half-lives, inactivation rate constants and standard errors were calculated using nonlinear regression to fit a single exponential rate model to the inactivation data (using Graph-Pad Prism software).

<sup>b</sup>  $\Delta\Delta G = \Delta G_{\text{mutant}} - \Delta G_{\text{wild-type}} = -RT \ln(k_{\text{mutant}}/k_{\text{wild-type}})$

### 6.5.3 Comparison of the most thermolabile mutants by additional parameters

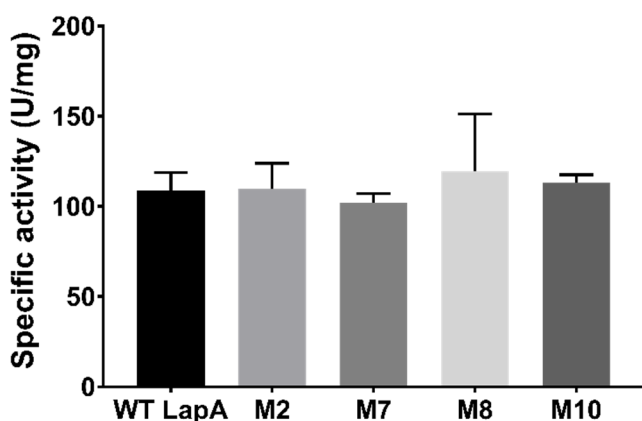
Based on an analysis of the kinetic parameters of LapA mutant thermal inactivation (from the two rounds of mutagenesis), the most thermolabile mutants meeting all the qualifying criteria were M2, M8, and M10 (single point mutations) and M7 (double mutant).

#### 6.5.3.1 Comparison of LapA enzyme activity at 37 °C

The most thermolabile mutants identified from two rounds of mutagenesis were evaluated using a standard industrial LapA assay to verify any potential differences in specific activity, as a result of the point mutations. It is important to note that the calculated specific activity values (**Table 6.21**) are slightly underestimated due to the differences in the path length provided by 300 µL well volume in a microtitre plate. Instead of 1 cm path length, in the standard 3 ml cuvette, 300 µL reaction volume gives rise to a 0.865 cm pathlength. Adjusting the calculation of the observed specific activity (116.7 U/mg, **Table 6.21**) of the *wt*LapA to 1 cm pathlength would result in 132.5 U/mg activity value, which closely resembles the one observed in 3 mL assay (140.15 U/mg) (**Table 6.12**).

	<i>wt</i> LapA	M2	M7	M8	M10
U/mg	<b>116.6698</b>	<b>109.547</b>	<b>100.086</b>	<b>119.3595</b>	<b>112.9799</b>
SD	9.935978	14.38779	4.975062	31.76096	4.593031

**Table 6.21** Determined absolute specific activity values of LapA proteins. Data represents measurements from 3 - 4 independent experiments with standard deviations (SD).



**Figure 6.22** Visual representation of the determined specific activity differences of LapA proteins. SD is depicted by error bars.

The specific activity of LapA enzymes, from the standard LapA assay, showed no major differences between *wt*LapA and the thermolability mutants analysed. Site - directed mutations in M2, M7, M8 and M10 LapA enzymes did not significantly alter specific activity towards the synthetic LPNA substrate. All aforementioned LapA mutants retained comparable levels of specific activity to *wt*LapA, under the conditions of a standard industrial LapA assay. Arguably, in the case of the M8 mutant, a slight overall increase in specific activity was noted.

#### 6.5.3.2 Comparison of best LapA thermolability mutants by $T_{60/30}$ parameter

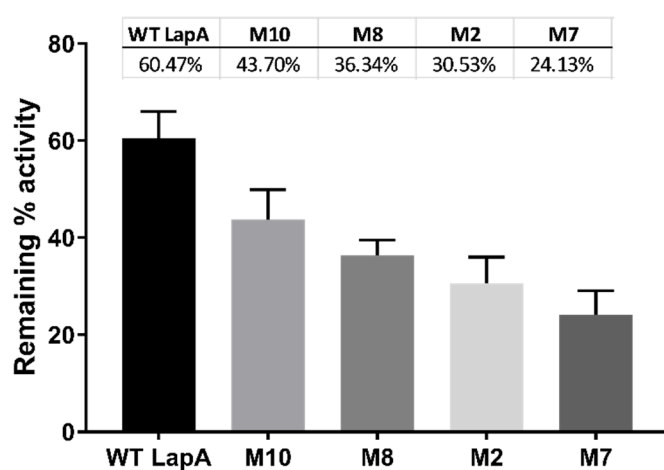
##### Industrial relevance of $T_{60/30}$ estimate:

The industrial hydrolysis reaction was allowed to proceed and peptidase inactivation was performed by heating the reaction mixture at 60 °C for 30 minutes to prevent adverse heating effects on the reaction products. Data were normalised to 155 U/mg (100%) activity value (**Table 6.22**).

The M7 LapA mutant displayed the lowest  $T_{60/30}$  value (24.13%; **Table 6.22**, **Figure 6.23**) among the most thermally labile LapA mutant proteins, with more than a half of that displayed by *wt*LapA (60.47%; **Table 6.22**, **Figure 6.23**). M2 and M8 mutants displayed comparable  $T_{60/30}$  values, however, in this instance, M7 was superior to M8, with respect to industrial hydrolysis and inactivation at 60 °C.

	Minutes			
	15	30	45	60
<b><i>wt</i>LapA</b>	65.07%	60.47%	50.57%	43.13%
<b>M7</b>	35.50%	24.13%	16.70%	11.07%
<b>M8</b>	52.54%	36.34%	28.77%	16.10%
<b>M2</b>	45.67%	30.53%	24.43%	20.67%
<b>M10</b>	54.10%	43.70%	32.73%	27.46%

**Table 6.22** Remaining percentage activities of LapA proteins following incubation at 60 °C.



**Figure 6.23** Comparison of *wt*LapA and thermolability mutants by  $T_{60/30}$  estimate. The data represent an average remaining percentage activity for different enzymes following incubation at 60 °C for 30 minutes, based on 3-4 independent measurements (SD is shown by error bars).

#### 6.5.4 Stability of recombinant LapA proteins at the room temperature (RT)

Protein stability at room temperature is a considerably important property regarding industrial application. Very stable proteins are likely to have longer shelf-lives in industrial formulations. In this regard, the observed LapA and mutant stability data at room temperature (**Table 6.23**), showed that the recombinant proteins retained a large proportion of their initial specific activity after prolonged incubation periods.

**Table 6.23** shows that *wt*LapA retains 85.37% and 73.37% of its original activity after RT incubation at 72h and 168h, respectively. Most of the thermolability mutants remained highly active, even after 7-day incubation at RT, suggesting that the single point mutations did not severely affect fold integrity and stability of the LapA enzyme.

**Table 6.24** provides a quantitative measurement of LapA ability to remain active at room temperature. High long-term stability of LapA proteins is even more evident from the protein half-lives calculated by non-linear regression. *wt*LapA, M2, M4, M7 and M12 appeared to be the most stable variants, exhibiting (extrapolated) half-lives of >250-300h at room temperature.

	Duration of incubation at RT			
	0h	24h	72h	7 days (168h)
<b>WT LapA</b>	100.00%	94.71%	85.37%	73.37%
<b>M2</b>	100.00%	90.74%	82.03%	79.53%
<b>M4</b>	100.00%	95.27%	87.41%	73.66%
<b>M7</b>	100.00%	90.34%	87.68%	60.93%
<b>M8</b>	100.00%	90.90%	80.09%	57.47%
<b>M9</b>	100.00%	80.74%	72.92%	76.08%
<b>M10</b>	100.00%	64.42%	57.74%	53.33%
<b>M12</b>	100.00%	85.33%	85.91%	69.94%
<b>M13</b>	100.00%	73.04%	56.01%	33.21%

**Table 6.23** Remaining percentage activities of WT and mutant LapA enzymes following prolonged incubation at room temperature. 0h, 100% activity – average specific activity observed for each LapA protein at 37 °C, following incubation for 15 minutes.

	<b>WT LapA</b>	<b>M2</b>	<b>M4</b>	<b>M7</b>	<b>M8</b>
<b><i>k</i> (h<sup>-1</sup>)</b>	0.001926	0.00169	0.001835	0.002724	0.003269
<b>Half life (h)</b>	359.9	410.2	377.8	254.5	212
<b>Days</b>	15.0	17.1	15.7	10.6	8.8

	<b>M9</b>	<b>M10</b>	<b>M12</b>	<b>M13</b>	<b>LapA Kex2</b>
<b><i>k</i> (h<sup>-1</sup>)</b>	0.002367	0.00557	0.002249	0.007685	0.003213
<b>Half life (h)</b>	292.8	124.5	308.2	90.2	215.7
<b>Days</b>	12.2	5.2	12.8	3.8	9.0

**Table 6.24** Room temperature inactivation profiles for *wt*LapA and mutants. Protein half-lives and inactivation rate constants were calculated using nonlinear regression to fit a single exponential rate model to the inactivation data (using Graph-Pad Prism software).

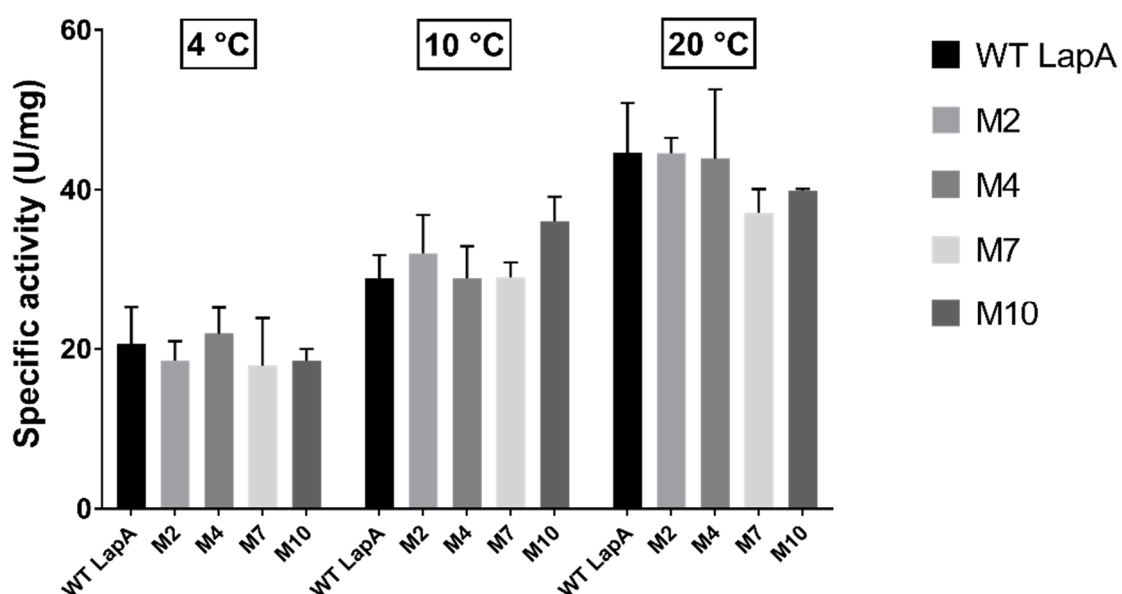
### 6.5.5 Exploring low-temperature activity of recombinant LapA proteins

To determine whether or not the LapA thermolability mutants gained any increase in specific activity at lower temperatures, due to the acquired structural flexibility, the protein samples were assayed in triplicate at 4, 10, 20 °C (**Table 6.25**), using the standard Lap assay protocol (section 2.10.2), following pre-incubation at the given temperature for 60 minutes. Due to insufficient sample, data for the M8 mutant was not collected.

At 4 °C, all LapA proteins exhibited severely reduced specific activity (**Figure 6.24**), with only one quarter of the activity levels normally observed at 37 °C (~100 - 130 U/mg). The increase in reaction temperature between 10 - 20 °C led only to a slight improvement in activity values. It is clear that temperature is a significant factor for the activity of *wt*LapA and mutant proteins, hence, the reaction is kinetically driven.

	U/mg		
	4°C	10°C	20°C
<b>WT LapA</b>	17.59 ± 4.65	20.8 ± 2.87	34.22 ± 6.24
<b>M2</b>	18.57 ± 2.4	21.27 ± 4.83	34.66 ± 1.91
<b>M4</b>	21.91 ± 3.32	20.77 ± 4.04	33.96 ± 8.63
<b>M7</b>	17.93 ± 5.95	20.5 ± 1.89	28.35 ± 2.95
<b>M10</b>	18.5 ± 1.51	23.04 ± 3.1	30.7 ± 0.23

**Table 6.25** Measured specific activity values for WT LapA and mutants at lower temperature range.



**Figure 6.24** Specific activity values for LapA mutants over a lower temperature range.

The expected result of this experiment was enhanced ability of LapA thermolability mutants to perform the reaction at low temperatures compared to the wild - type enzyme. It was assumed that the potential increase in molecular flexibility, for these mutations, would allow M2, M4, M7 and M10 mutants to exhibit more specific activity under higher viscosity and reduced kinetic energy environment, however, this was not the case.

### 6.5.6 Industrial pasteurisation for LapA protein inactivation

Industrial pasteurisation can be utilised to thermally inactivate the enzymes performing the hydrolysis of whey protein. The standard process used in the food industry is 72 °C exposure for <1 minute. As a result of the short exposure time to heat, the quality of the reaction products can be preserved and, in this case, gelation effects of whey protein hydrolysates are less likely to be induced. The previously obtained thermal inactivation data suggested that the M7 thermolabile mutant could be completely inactivated, following incubation at 70°C for 15 minutes. Therefore, industrial pasteurisation was predicted to be effective in rendering M7 and, possibly, M8 mutants inactive.

The full procedure for this method is described in section 2.11.

#### 6.5.6.1 Inactivation of native LapA in Flavourzyme

In brief, 1 mg/ml Flavourzyme solution in ultrapure water was used to test whether the native LapA (*flavLapA*), present in the commercial fungal extract, could be completely inactivated by industrial pasteurisation. Commercially available Flavourzyme was used to evaluate the effectiveness of this method for thermal inactivation, prior to testing the recombinant LapA proteins.

A 15 second holding time (duration of exposure) was selected for inactivation of Lap proteins (*flavLapA* and Lap2) in Flavourzyme. The remaining activity measurements were carried out within a 60-minute period, following pasteurisation. The hydrolytic contribution from Lap2 was considered negligible, in subsequent residual activity measurements of post-pasteurised samples. The remaining percentage activities are summarised in **Table 6.26**. Pasteurisation of Flavourzyme effectively inactivated native LapA in the 75 – 80 °C temperature range, within 15 seconds.

Enzyme	Untreated	Pasteurisation temperature (°C)			
		70	72	75	80
<i>flavLapA</i> (in Flavourzyme)	100.00%	26.70%	12.50%	3.30%	<b>inactivated</b>

**Table 6.26** Remaining percentage activities of *flavLapA* in post-pasteurisation samples.

#### 6.5.6.2 Inactivation of LapA thermolability mutants

Wild-type LapA and the two previously determined thermolabile mutants, M7 and M8, were used to determine the thermal inactivation profiles by industrial pasteurisation. These

recombinant LapA proteins were expressed as proenzymes and had undergone enzymatic activation by AP1 protease, prior to the pasteurisation experiment. The homogeneity of the mature population of LapA was confirmed by SDS–PAGE analysis (not shown). It was essential to eliminate the pro-protein, exhibiting propeptide-induced higher thermostability, prior to analysis.

In this experiment, the holding time (temperature exposure) was increased to 30 seconds in an effort to improve the enzyme inactivation rate. The remaining activity measurements were carried out within a 2 – 3 hour period, following pasteurisation. The post-pasteurisation inactivation profiles of *wt*LapA, M7 and M8 proteins are summarised in **Table 6.27**.

<b>A</b>		<b>Pasteurisation temperature (°C)</b>				
<b>Enzyme</b>	<b>Untreated</b>	<b>65</b>	<b>70</b>	<b>72</b>	<b>75</b>	<b>80</b>
<i>wt</i> LapA	88.3	81.7	77.0	NA	48.4	35.7
M7	23.4	14.0	9.8	6.6	3.5	2.1
M8	55.0	34.5	25.3	20.5	15.3	12.2

<b>B</b>		<b>Pasteurisation temperature (°C)</b>				
<b>Enzyme</b>	<b>Untreated</b>	<b>65</b>	<b>70</b>	<b>72</b>	<b>75</b>	<b>80</b>
<i>wt</i> LapA	100.00%	92.55%	87.23%	NA	54.79%	40.43%
M7	100.00%	59.70%	41.79%	28.36%	14.93%	8.96%
M8	100.00%	62.70%	46.03%	37.30%	27.78%	22.22%

**Table 6.27** Inactivation profiles of *wt*LapA, M7 and M8 proteins by industrial pasteurisation. **A** Remaining specific activity (U/mg); **B** Remaining percentage activity with reference to *Untreated* samples (100%). NA – not available.

Interestingly, *wt*LapA and the thermolabile mutant enzymes exhibited a degree of thermostability in the 65 - 80 °C temperature range. Complete (100%) inactivation could not be achieved, even at 80 °C. However, a progressive temperature inactivation was observed in all proteins. The most thermosensitive enzyme was the M7 mutant, with only a negligible amount of specific activity (2.1 U/mg / 8.96%) remaining, following pasteurisation at 80 °C for 30 seconds. As expected, *wt*LapA exhibited the greatest thermostability, at all temperatures.

It is important to note that the specific activity measurements were conducted after a 2 - 3 hour time lapse, post-pasteurisation. Unfortunately, due to restricted access to a

spectrophotometer, all samples (from every temperature) were collected first, which was followed by a LapA activity assay in a different laboratory, to quantify the remaining activity. These experiments could suggest that the LapA enzymes were not completely unfolded within the short 30 second exposure time and, therefore, were capable of retaining (even regaining) an active conformation. This would impact the final amount of measured residual activity. We confirmed this by a measurement of *wt*LapA activity, immediately following pasteurisation at 80 °C, where the enzyme displayed complete momentary inactivation (no evidence of hydrolytic activity). However, after 3 hour incubation at room temperature, the same sample exhibited 40% activity increase (**Table 6.27, B**), supporting the possibility of refolding to regain an active conformation.

Therefore, we conclude that very short, high temperature exposure periods are not sufficient to completely disrupt the highly compact and thermostable hydrolase fold. The temporarily induced inactive conformation of LapA could involve unravelling surface sub-structures, loop regions, stripping the catalytic dinuclear zinc, but it fails to completely abolish the stable hydrophobic core. Significantly longer exposure times or increased temperature (kinetic energy) are likely required to break these enthalpy-driven intramolecular interactions. In addition, in this experiment, LapA proteins were diluted in 20 mM Tricine pH 8.0, 1 mM ZnCl<sub>2</sub> buffer, which is equivalent to the buffer used for successful *in vitro* refolding *ref*LapA protein (section 4.1.2.3). It is plausible that the additional zinc ions in this buffer highly promote and accelerate the formation of active LapA conformers.

With respect to native LapA, the observed increase in thermostability of recombinant LapA proteins may have a number of plausible explanations. For example, more time was allowed for recovery of recombinant proteins, following pasteurising. And, there are apparent minor structural differences between native and recombinant LapA proteins, which could have a profound effect on thermostability. Specifically, the AP1-activated *wt*LapA, M7 and M8 proteins all were demonstrated to possess an additional N-terminal ‘AVT’ extension (section 4.3.7.2), which could alter the thermostability profile. Finally, and probably most importantly, the highly branched mannosylated N-glycan pattern, as present in extracellular yeast proteins, may contribute to the improved thermostability of these recombinant LapA proteins, whereas the fungal *flav*LapA from Flavourzyme, is likely to possess less hyperglycosylated N-glycan structures (Deshpande et al., 2008) and, therefore, exhibit lower thermostability by comparison.

## 6.6 Kinetic characterisation of wild-type LapA and mutant enzymes

### 6.6.1 Determination of kinetic parameters of LapA hydrolytic reaction

To investigate whether any of the LapA thermolability mutants led to alteration of the reaction kinetics, a steady-state kinetic analysis was performed. LapA wild-type and mutant enzyme velocities were determined, with increasing concentrations of LPNA substrate at 37 °C (section 2.10.4). A Michaelis-Menten model was implemented and analysed by non-linear regression, using GraphPrism software package. Non-linear regression is superior to Lineweaver-Burk plot linear regression analysis due to less internal error and no requirement of data transformation.  $V_{\max}$ ,  $K_m$  and  $k_{\text{cat}}$  values were calculated for each enzyme.

**Table 6.30** summarises the determined kinetic constants for LapA enzymes from steady-state kinetic data, using non-linear regression. The *wt*LapA enzyme displayed a  $k_{\text{cat}}$  of 76.53 s<sup>-1</sup>. Overall, M2 – M13 thermolability mutations had a slightly negative effect on LapA turnover number with a modest effect showed by the M8 mutant (65.4 s<sup>-1</sup>). All of the mutants showed a retention to convert a comparable quantity of substrate molecules per unit of time.

The Michaelis-Menten constant ( $K_m$ ) was affected significantly across all thermolability mutants. All of the LapA mutants showed a trend toward increased substrate affinity ( $K_m$ ), which in turn increased their respective catalytic efficiency ( $k_{\text{cat}}/K_m$ ) towards LPNA substrate. M10 and M12, in particular, displayed 1.59- and 1.81-fold decrease in  $K_m$ , respectively. The most thermolabile mutants M2, M7 and M8 showed no significant increase in catalytic efficiency compared to *wt*LapA. Their apparent increase in substrate affinity was outweighed by the equivalent reduction in turnover number, hence, no substantial change was observed in the  $k_{\text{cat}}/K_m$  ratio.

This experiment showed that the most catalytically efficient LapA mutants were M10 and M12 owing to their improved substrate affinity values. These enzymes were able to achieve their maximum reaction velocity at lower substrate concentration compared to *wt*LapA. However, the substrate turnaround rates ( $k_{\text{cat}}$ ) were slightly reduced in these enzymes compared to *wt*LapA. Using LPNA, as a substrate, the determined  $K_m$  (120 µM) and  $k_{\text{cat}}$  (76.53 s<sup>-1</sup>) values for recombinant wild-type enzyme were substantially different from the values (244 µM, 50.9 s<sup>-1</sup>) originally reported for native LapA expressed in *Aspergillus oryzae* (Matsushita-Morita et al., 2011).

**Table 6.28** shows that overall catalytic efficiency, for recombinant LapA, was 3-fold higher under the conditions used in our assay (37 °C, pH 7.2). However, the kinetic constants for native LapA were determined at significantly higher reaction temperature and pH (50 °C, pH 8.5). This could also be attributed to the differences in LapA precursor processing (*wt*LapA had 3-residue 'AVT' addition to its N-terminus) in relation to *flav*LapA, although further work (deletion of the AVT residues) would be needed to confirm this conclusion. However, this modification alone could have enhanced *wt*LapA catalytic capabilities. It is reasonable to speculate that, due to higher kinetic energy at 50 °C, the enzyme-substrate collisions would be more frequent and *flav*LapA would display higher turnover rate ( $k_{\text{cat}}$ ) per unit of time. However, this should be validated experimentally, looking for an increase in  $k_{\text{cat}}$  for *wt*LapA at 50 °C. Nevertheless, this observation further supports the notion, that *wt*LapA is genuinely able to turn over more LPNA molecules per second than native LapA.

	<b>Recombinant LapA</b> <b>(<i>wt</i>LapA)</b>	<b>Native LapA</b> <b>(<i>flav</i>LapA)</b>
$k_{\text{cat}}$ ( $\text{s}^{-1}$ )	76.53	50.9
$K_{\text{m}}$ ( $\mu\text{M}$ )	120	244
$k_{\text{cat}}/K_{\text{m}}$ ( $\text{s}^{-1}/\text{mM}^{-1}$ )	637.75	209

**Table 6.28** Comparison of kinetic constants between recombinant LapA and native LapA enzymes.

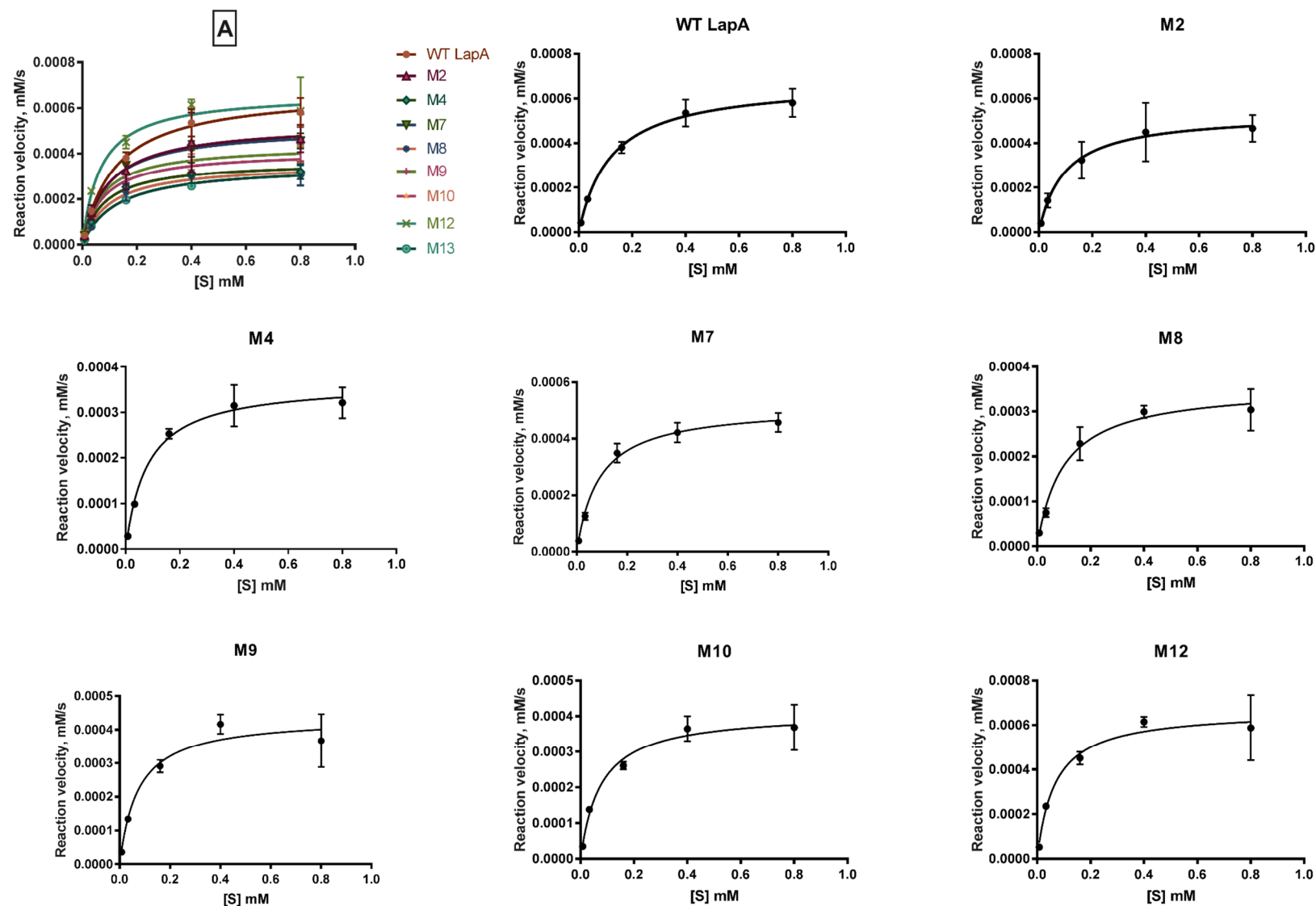
The underlying rationale, for determining the steady-state kinetics of LapA mutants, was two-fold. Firstly, it was necessary to confirm that point mutations did not severely reduce either the affinity for LPNA substrate or the ability to turn over the substrate. Secondly, via an increase in molecular flexibility, LapA thermolability mutants were expected to display higher  $k_{\text{cat}}$  values; a phenomenon observed with cold-adapted proteins at low temperatures (Feller and Gerday, 1997). Conversely, an overall minor reduction in turnover rates and a tendency toward improved affinity for LPNA substrate suggests that more measurements would be required to determine the full extent of these mutations on LapA enzyme kinetics, particularly at <37 °C, whereby the likely gain in flexibility of the thermolability mutants would permit an outperformance of the wild-type enzyme on substrate turnover rate (ie. an increase in  $k_{\text{cat}}$ ).

A		WT LapA	M2	M7	M4	M8	M9	M10	M12
	$E_t$ (mM)	8.86E-06	8.94E-06	8.57E-06	6.39E-06	5.45E-06	8.23E-06	6.89E-06	1.28E-05
	$k_{cat}$ ( $s^{-1}$ )	76.53	59.75	60.18	57.31	65.4	53.11	59.69	51.85
	$K_m$ ( $\mu M$ )	120	95.56	88.66	79.78	98.75	72.64	75.44	66.2
	$V_{max}$ (mM/s)	0.0006778	0.0005339	0.0005159	0.0003661	0.0003564	0.0004372	0.0004111	0.0006656
	Std. Error								
	$k_{cat}$	3.519	5.611	2.2	2.137	3.784	3.446	3.063	3.152
	$K_m$	0.0196	0.03444	0.01273	0.01206	0.02169	0.01963	0.01595	0.01717
	95% CI (profile likelihood)								
	$k_{cat}$	69.59 to 84.89	49.33 to 74.88	55.84 to 65.12	53.26 to 61.87	58.24 to 74.19	46.49 to 61.26	53.52 to 67.19	45.66 to 59.44
	$K_m$	84 to 170	44 to 210	66 to 119	59 to 107	63 to 155	42 to 127	47 to 121	38 to 116
	Goodness of Fit								
	Degrees of Freedom	13	13	13	23	13	13	13	13
	R square	0.9746	0.8776	0.9784	0.9567	0.9525	0.9209	0.9489	0.9233

B		WT LapA	Change (-fold)	M2	Change (-fold)	M4	Change (-fold)	M7	Change (-fold)	M8	Change (-fold)
	$k_{cat}$ ( $s^{-1}$ )	76.53	NA	59.75	-1.28	57.31	-1.34	60.18	-1.27	65.4	-1.17
	$K_m$ ( $\mu M$ )	120	NA	95.56	-1.26	79.78	-1.50	88.66	-1.35	98.75	-1.22
	$k_{cat}/K_m$ ( $s^{-1}/mM^{-1}$ )	637.75	NA	625.26	-1.02	718.35	+1.13	678.77	+1.06	662.28	+1.04
		M9	Change (-fold)	M10	Change (-fold)	M12	Change (-fold)	M13	Change (-fold)		
	$k_{cat}$ ( $s^{-1}$ )	53.11	-1.44	59.69	-1.28	51.85	-1.48	50.88	-1.50		
	$K_m$ ( $\mu M$ )	72.64	-1.65	75.44	-1.59	66.2	-1.81	124.2	+1.04		
	$k_{cat}/K_m$ ( $s^{-1}/mM^{-1}$ )	704.00	+1.10	821.72	+1.29	783.23	+1.23	409.66	-1.56		

**Table 6.29** LapA enzyme steady-state kinetic parameters calculated from non-linear regression fits to reaction data. Kinetic reaction constants along with statistical confidence calculations, goodness of fit (part A) are provided. Additionally, enzyme kinetic constants were separated with determined catalytic efficiency values ( $k_{cat}/K_m$ ) and fold change in every constant compared to the wild-type LapA (part B).



**Figure 6.26** Reaction velocity plots versus LPNA substrate concentration for LapA enzymes. A – representation of individual enzyme data on a single-plot.

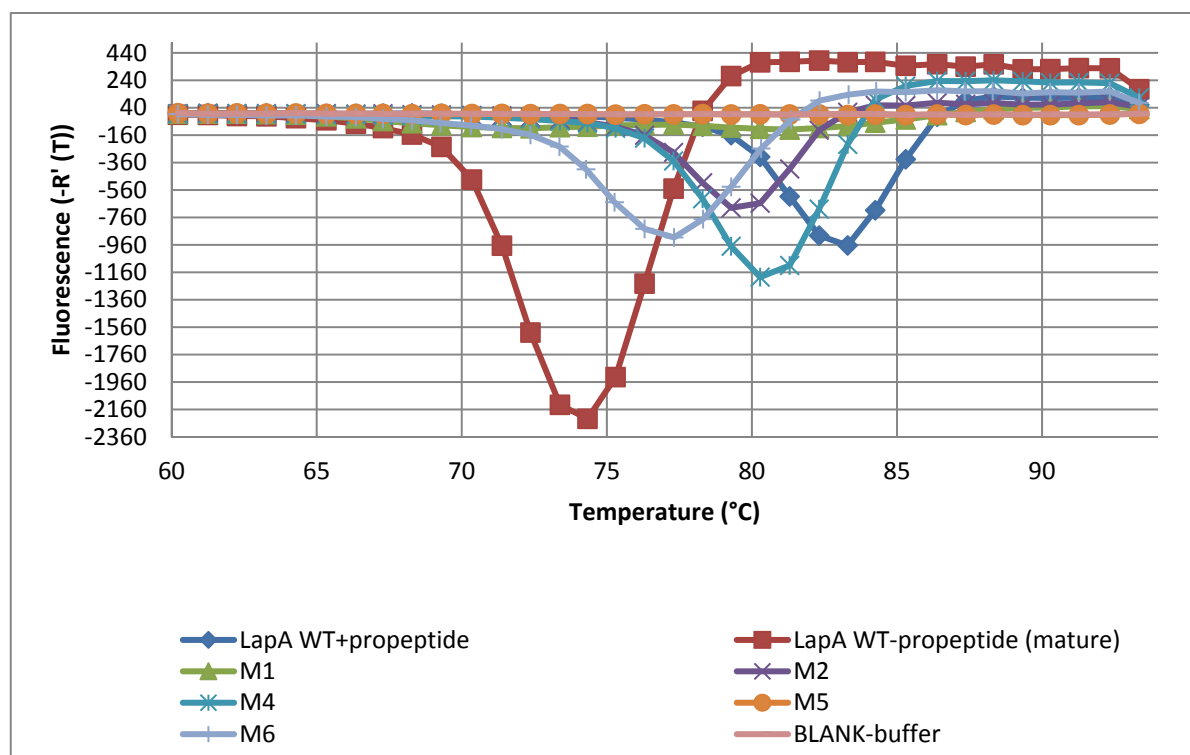
## 6.7 Thermodynamic characterisation of wild-type LapA and mutant enzymes

### 6.7.1 LapA characterisation by differential scanning fluorimetry (DSF)

#### 6.7.1.1 DSF of M1 – M6 mutants

A 96-well plate format thermofluor assay (section 2.9.1) allowed us to quickly obtain the melting temperature of each LapA variant, in a high – throughput format, without consuming significant amounts of protein sample. This served as a rapid screening protocol for LapA thermolability mutants.

In the first round of mutagenesis,  $T_m$  values of purified M1 - M6 mutant precursors were obtained together with *wtproLapA* and *wLapA* (mature) enzymes for direct comparison (Figure 6.27).



**Figure 6.27** Plot showing the derivative of the fluorescence signal versus temperature. ‘LapA WT+propeptide’ corresponds to *wtproLapA* enzyme, ‘LapA WT-propeptide’ corresponds to *wLapA* (mature) enzyme. Signal intensity is positively correlated with the enzyme concentration.

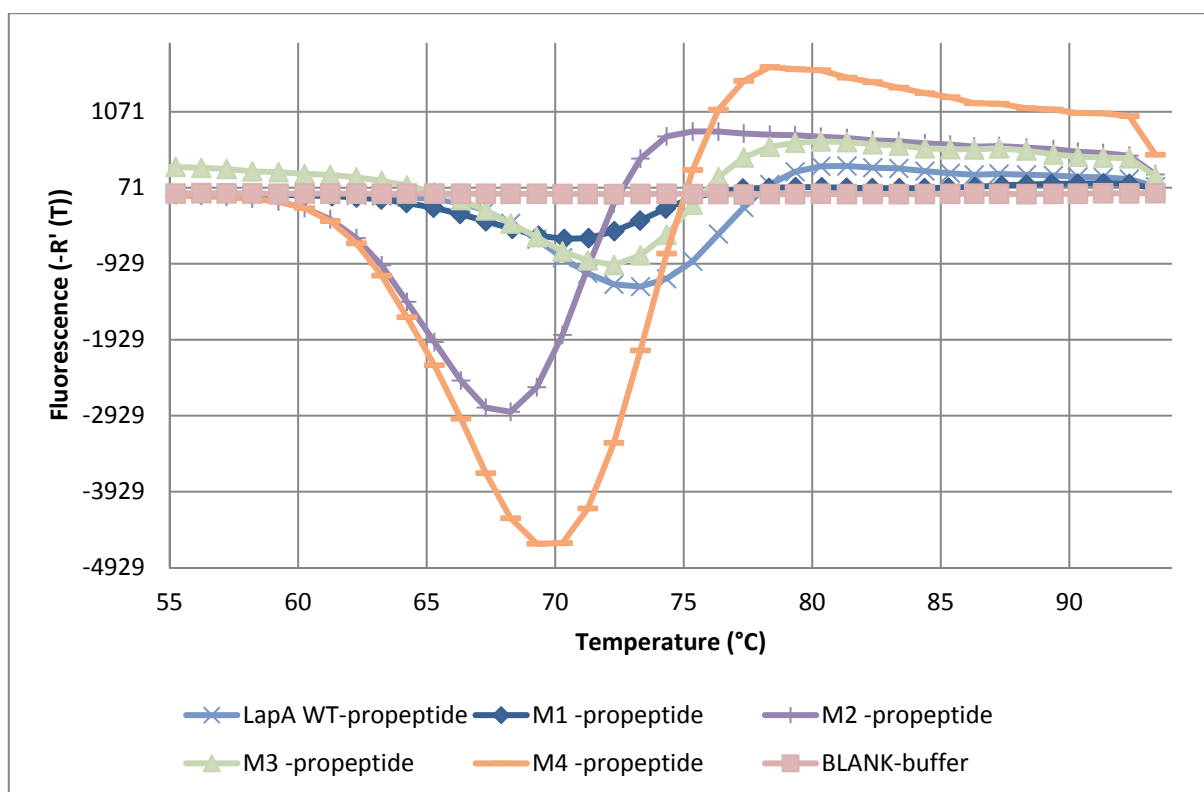
The presence of propeptide had a significant thermostabilising effect on the *wtproLapA* enzyme ( $T_m = 83.3$  °C, Table 6.30). Following cleavage of propeptide, *wLapA* displayed a 9 °C lower melting temperature ( $T_m = 74.3$  °C, Table 6.30). M2, M4 and M6 mutant

precursors showed lower than wild-type  $T_m$  values. A temperature shift for M6 was the most significant, possibly due to elimination of the stabilising disulphide bond. Due to insufficient protein quantities,  $T_m$  values for M1, M3 and M5 mutants were not determined in this instance.

Enzyme	Sample concentration (mg/ml)	Melting temperature, $T_m$ (°C)	$\Delta T_m$ (°C)
<i>wtpro</i> LapA (+propeptide)	6.5	83.3	
<i>wt</i> LapA (-propeptide)	11	74.3	-9.0
M1 (+propeptide)	1.2	ND	ND
M2 (+propeptide)	5.5	79.3	-4.0
M3 (+propeptide)	0.5	ND	ND
M4 (+propeptide)	7.0	80.3	-3.0
M5 (+propeptide)	0.4	ND	ND
M6 (+propeptide)	5.8	77.3	-6.0

**Table 6.30** Differences in  $T_m$  values for wild-type LapA and thermolability mutants.  $\Delta T_m$  represents a difference in comparison to *wtpro*LapA enzyme. ND – not determined.

In the following experiment, the LapA precursors were enzymatically processed and a thermofluor assay was performed for the second time (**Figure 6.28**).  $T_m$  values were obtained for the activated LapA mutant proteins and compared to the *wt*LapA enzyme (**Table 6.31**). The most thermolabile mutants, M2 and M4, displayed a 5.1 and 3.8 °C decrease in melting temperature, respectively, in comparison to the wild-type enzyme. Melting temperature profiles, for M5 and M6 mutants, were not collected due to the previously determined severe reduction in specific activity arising from these specific site-directed mutations.



**Figure 6.28** Plot of the derivative of the fluorescence signal versus temperature. In this instance, all LapA enzymes were analysed in the mature state. Signal intensity is positively correlated with enzyme concentration.

Enzyme	Sample concentration (mg/ml)	Melting temperature, $T_m$ (°C)	$\Delta T_m$ (°C)
<i>wt</i> LapA (-propeptide)	7.8	73.3	
M1 (-propeptide)	4.5	70.4	-2.9
M2 (-propeptide)	17.7	68.2	-5.1
M3 (-propeptide)	ND	72.2	-1.1
M4 (-propeptide)	24	69.5	-3.8

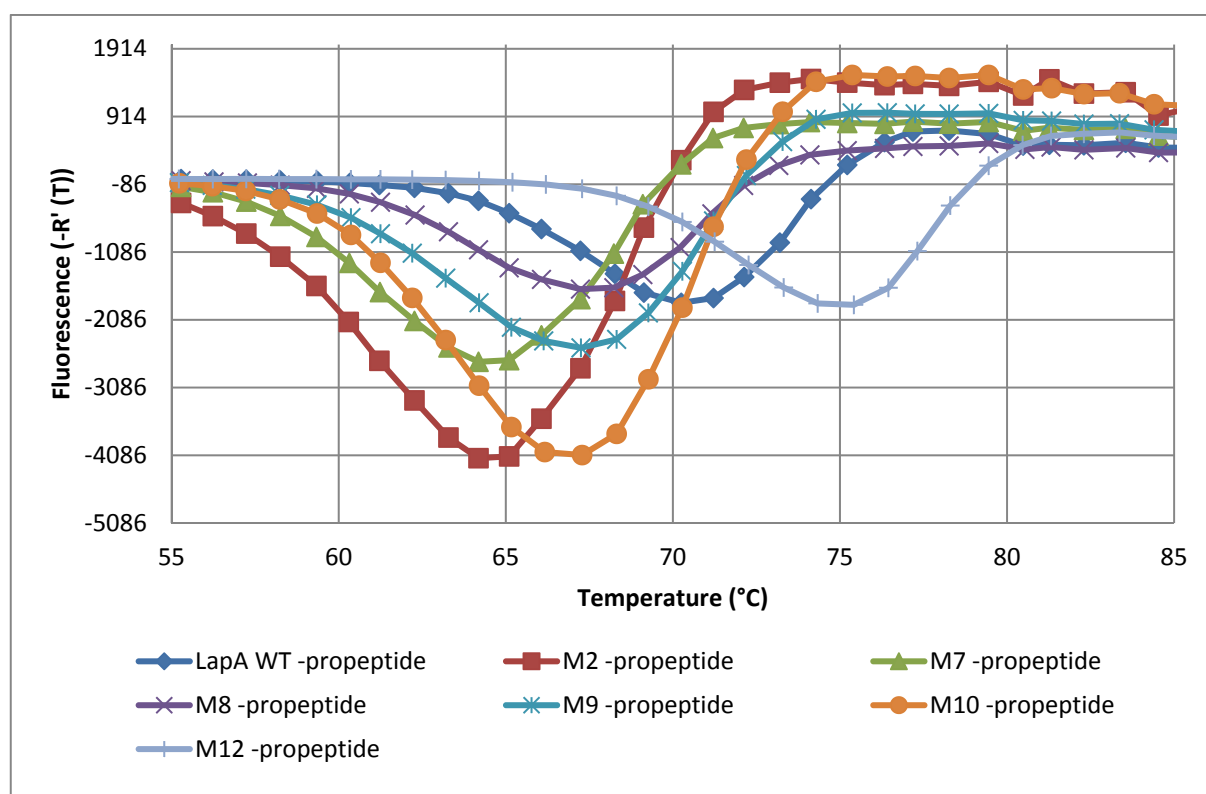
**Table 6.31** Differences in  $T_m$  values of wild-type LapA and thermolability mutants.  $\Delta T_m$  represents a difference in comparison to *wt*LapA enzyme. ND – not determined.

Findings from this experiment were in agreement with the temperature incubation assay data (see section 6.5.1), which confirmed that both the M2 and M4 mutants, from the first round of mutagenesis, exhibited the most thermolabile profiles while retaining equivalent to wild-type enzyme activity levels.

### 6.7.1.2 DSF of M7 – M13 mutants

The remaining LapA thermolability mutants also were characterised by DSF, following the second round of mutagenesis. This time all LapA samples were measured in triplicate, to determine the extent of experimental/sample handling error (**Figure 6.29**).

All the LapA thermolability mutant samples have undergone the same enzymatic activation procedure, prior to performing the thermal shift assay.



**Figure 6.29** Plot of the derivative of the fluorescence signal versus temperature. All LapA enzymes were analysed in the mature state. Signal intensity is positively correlated with enzyme concentration.

In this round of analysis, a new batch of M2 (the most thermolabile mutant from round one) also was included. Based on the DSF data (**Table 6.32**), M2 still displayed the most significant thermal shift ( $\Delta T_m = -5.78$  °C). The M7 double mutant showed an almost identical  $\Delta T_m$  value ( $-5.48$  °C, Table 6.32). The thermal shifts of other mutants were less substantial, with the exception of the M12 mutant, which exhibited higher thermostability ( $\Delta T_m = +4.98$  °C) in relation to wild-type LapA. The M13 mutant was not characterised, due to problems experienced during the purification process.

The standard deviations obtained from triplicate measurements were found to be very narrow (**Table 6.32**), confirming the accuracy and high data redundancy in the DSF experiment.

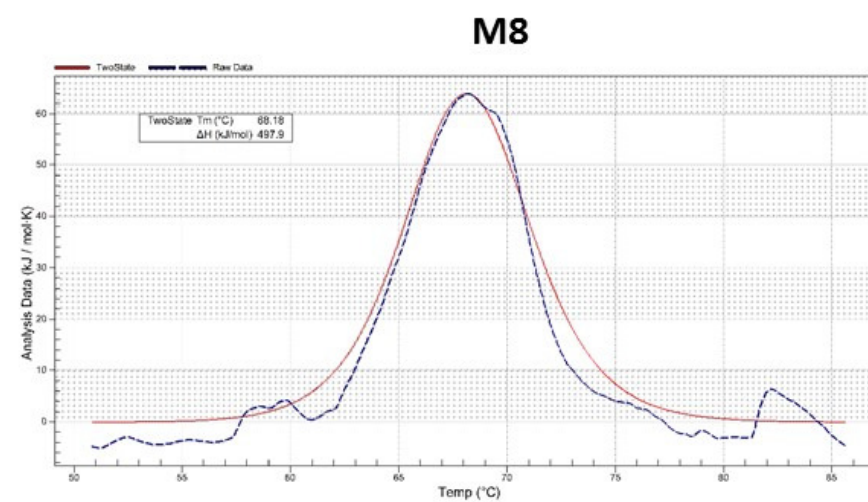
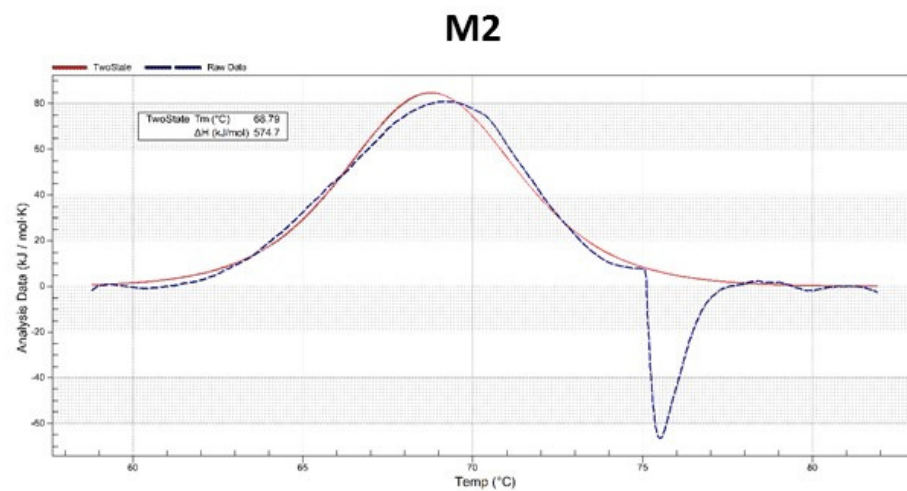
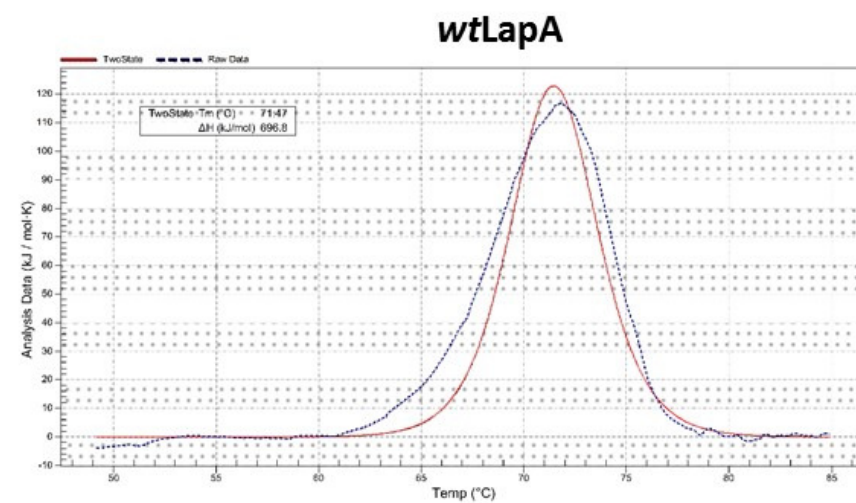
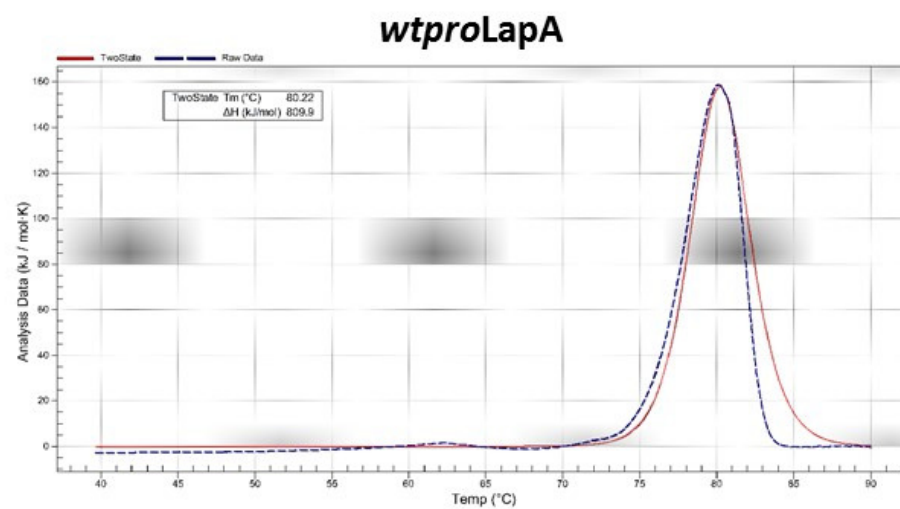
Enzyme	Sample concentration (mg/ml)	Melting temperature, $T_m$ (°C) (SD)	$\Delta T_m$ (°C)
<i>wt</i> LapA (-propeptide)	8.93	$70.27 \pm 0.07$	
M2 (-propeptide)	13.88	$64.49 \pm 0.27$	-5.78
M7 (-propeptide)	7.19	$64.79 \pm 0.27$	-5.48
M8 (-propeptide)	5.52	$67.93 \pm 0.61$	-2.34
M9 (-propeptide)	3.85	$66.71 \pm 0.54$	-3.56
M10 (-propeptide)	12.7	$67.38 \pm 0.10$	-2.89
M12 (-propeptide)	5.76	$75.25 \pm 0.30$	+4.98

**Table 6.32** Differences in  $T_m$  values of wild-type LapA and thermolability mutants.  $\Delta T_m$  value represents the difference compared to *wt*LapA enzyme.

### 6.7.2 LapA characterisation by differential scanning calorimetry (DSC)

In the DSC experiment, the difference in energy uptake between protein sample solution and sample buffer, under increasing temperature, was measured. In this way, it was possible to establish the original thermodynamic parameters for the endothermic transition of wild-type LapA protein. In turn, it was possible to characterise how these parameters are altered, following targeted thermolability mutations. DSC allowed direct tracking of the energetic changes occurring in the unfolding process for the LapA mutants.

Protein endothermic curves, expressed as molar heat capacity, were obtained in a 20 – 90 °C temperature range (**Figure 6.30**). Transition peak ( $T_m$ ) values were determined from a dynamic 2-state equilibrium, at which half of the protein population is unfolded. Thermodynamic data, for each transition, were obtained by fitting a simple 2-state unfolding model following normalisation by protein concentration and baseline correction. The calorimetric enthalpy ( $\Delta H$ ) was calculated as the area under the endothermic peak.



**Figure 6.30** Analysis of LapA proteins by DSC. Dashed line and solid red line represent raw data curve and fitted two-state model curve, respectively. Thermodynamic transition midpoint ( $T_m$ ) and enthalpy ( $\Delta H$ ) values were obtained from these curves. The 'bumps' and 'dents' in the traces for M2 and M8 were caused by impurities.

Endothermic peak data, obtained for the LapA variants, were fitted to the two-state transition model, available in the analysis software (**Table 6.33**). Due to some difficulties encountered with the experimental equipment, DSC traces were determined only for wild-type recombinant LapA proteins (mature and precursor) and two thermolability mutants (M2 and M8).

Protein	T <sub>m</sub> (°C)	ΔH (kJmol <sup>-1</sup> )	ΔS (kJmol <sup>-1</sup> K)
<i>wtpro</i> LapA	80.22	809.9	2.08
<i>wt</i> LapA	71.47	696.8	2.40
M2	68.79	574.7	1.40
M8	68.18	497.9	1.24

**Table 6.33** Thermodynamic parameters for LapA variants, as analysed by DSC.

LapA proenzyme appeared as the most thermostable variant, owing to its added prodomain at the N-terminus. Mature LapA protein showed a T<sub>m</sub> value of 71.47 °C, suggesting the prodomain contributes an additional 9 °C to the midpoint transition temperature. This is reflected also in the difference in enthalpic contribution, adding another 113 kJmol<sup>-1</sup> to ΔH for the *wtpro*LapA proenzyme. M8 showed the lowest enthalpic difference of 497.9 kJmol<sup>-1</sup>, suggesting a reduction in the number of intramolecular interactions in this protein. This was in agreement with the observed thermolabile behaviour of M8 mutant and the determined low half-life values (see section 6.5.2.4).

## **Chapter 7** EXPLORATION OF THE ALTERNATIVE ROUTES TO RECOMBINANT LAPA SECRETION IN *P.PASTORIS*

---

## 7.1 Introduction

Post – translational activation of recombinant LapA precursor presents significant challenges. It is time consuming, with many additional steps in the purification protocol, to obtain mature, active and pure LapA protein. Also, LapA enzymatic processing, via addition of an exogenous protease, is highly undesirable and cost-ineffective for industry, which may not be directly applicable to Biocatalysts enzyme production pipeline.

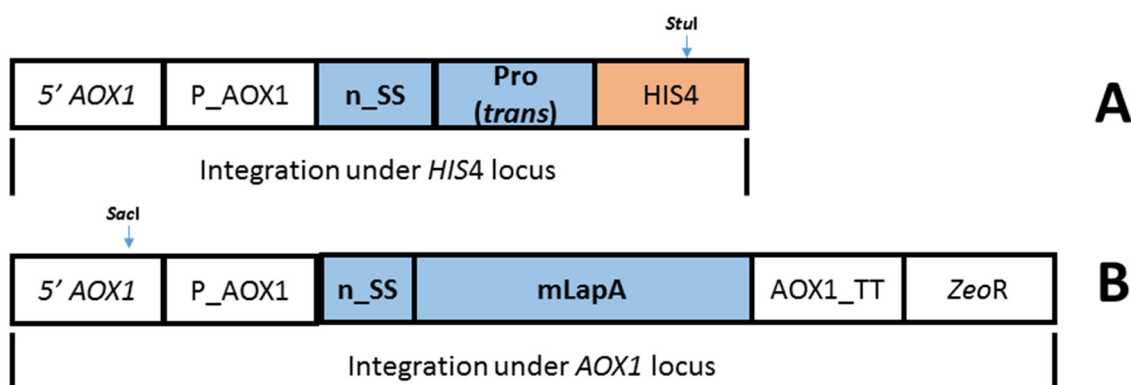
Alternative ways of directly acquiring activated LapA domain in *P.pastoris* supernatant were sought after. Attempts to secrete solely mature LapA domain were somewhat successful, however, the resultant polypeptide was found to exist in a molten globule – like state, with negligible residual specific activity (section 4.3.2.1). It was obvious that LapA propeptide sequence cannot be omitted in the *P. pastoris* expression cassette to obtain functional enzyme.

## 7.2 LapA *trans* – acting propeptide

Several successful studies of prodomain protein expression in a non-covalently linked state (*trans*) have been reported for homologous and more distantly related peptidases. Expression of propeptide and mature domains, as separate polypeptides, resulted in the recovery of active membrane Type 1-Matrix Metalloproteinase (Cao et al., 2000). Impeded secretion of the mature domain of aspartic proteinase, Sap1p, from *C.albicans* in *P. pastoris* was restored upon co-expressing the pro-polypeptide in *trans* (Beggah et al., 2000). Unlinked propeptide co-expression in *P. pastoris* was also shown to rescue the appropriate secretion and activity levels of *Streptomyces mobaraensis* Transglutaminase (Yurimoto et al., 2004).

We have already shown that LapA propeptide acts as an intramolecular chaperone and is essential for the maturation of the catalytic domain, when expressed in *cis*. From the literature search, it was evident that LapA prodomain may be able to function in an intermolecular manner and to assist with secretion and folding of mature polypeptide, if expressed in *trans*. Most importantly, this could solve subsequent downstream processing problems of LapA pro-protein.

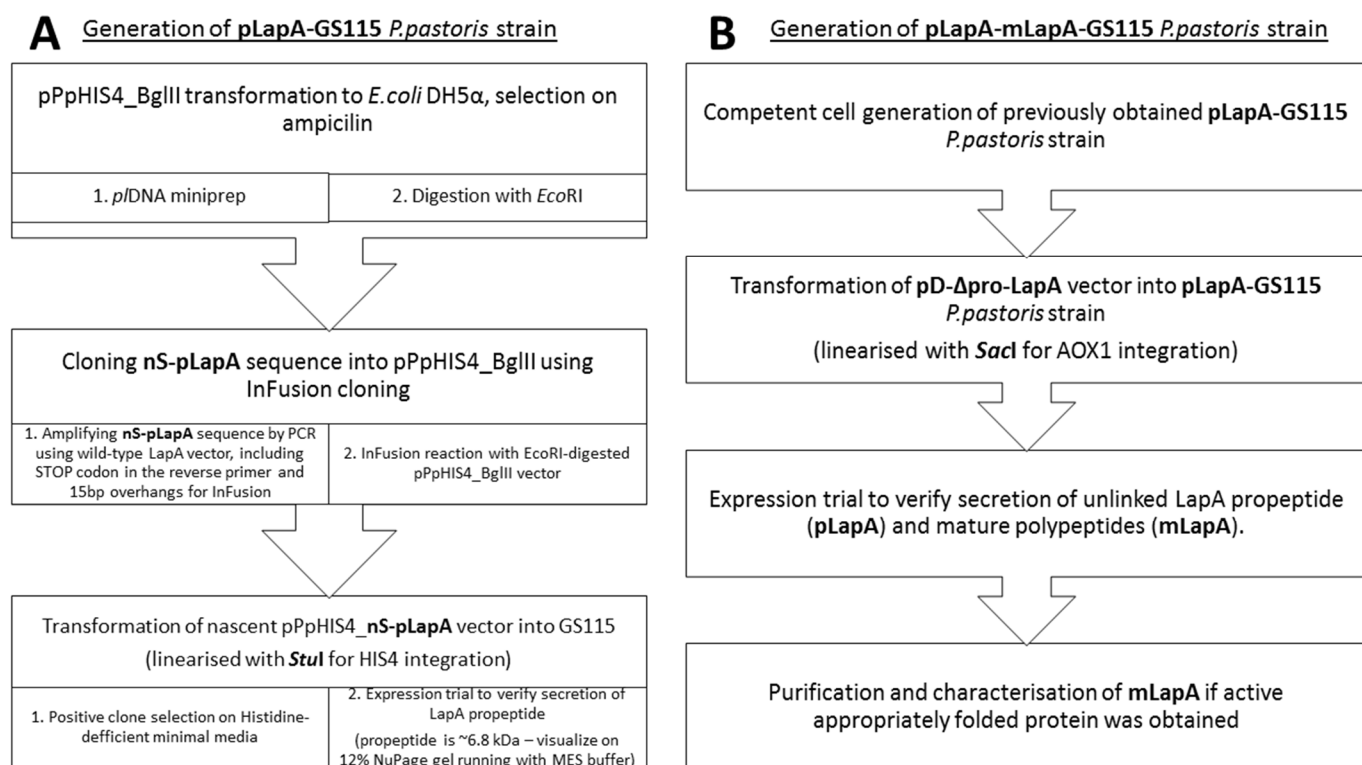
For this purpose, a recombinant system of LapA propeptide expression in *trans* in *P. pastoris* was designed (**Figure 7.1**).



**Figure 7.1** *P. pastoris* cassettes for LapA **A.** pro- and **B.** mature domain expression in *trans*. 5' AOX1, P\_AOX1, AOX\_TT: promotor and terminator regions; n\_SS: native LapA secretion signal; Pro: LapA propeptide sequence; mLapA: mature domain sequence; HIS4 – histidine auxotrophy marker; ZeoR: zeocin resistance marker; *SacI* and *StuI* linearization sites are indicated by an arrow.

LapA propeptide and mature domain were treated as two individual genes in this experiment. Due to the requirement of positive selection of *P. pastoris* transformants for the two separate genes, two different chromosomal integration loci had to be utilized, namely HIS4 and AOX1. HIS4 auxotrophy marker allowed screening of positive transformants of LapA propeptide expression cassette (**Figure 7.1, A**), whereas zeocin resistance marker allowed screening of mature LapA domain (mLapA) transformants (**Figure 7.1, B**).

As illustrated in **Figure 7.2**, pPpHIS4\_BgIII vector was propagated in *E.coli* and linearized with *EcoRI* to enable subsequent InFusion cloning reaction. DNA fragment equivalent to LapA native signal sequence and prodomain (nS-pLapA) was amplified by PCR using the pD-npro-LapA vector as a template and the following set of primers: 5' – ATTATTGAAAGAATTCAAACGATGAGATTCCTACCAT – 3' and 5' – CGCCACTAGTGAATTTTAGGTCACAGCAAGCTTTTGCTTTT – 3'.

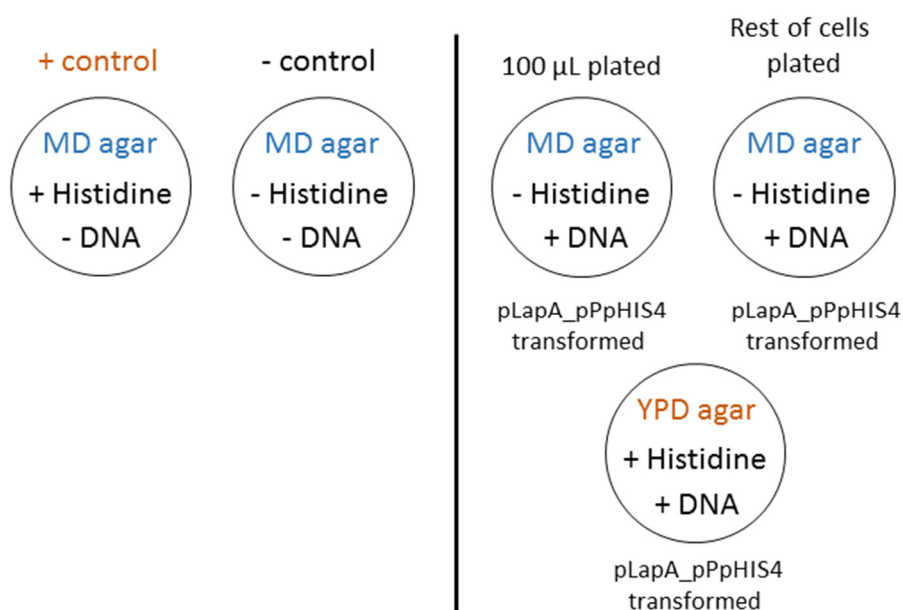


**Figure 7.2** Experimental plan to generation of LapA propeptide (**pLapA**) and mature domain (**mLapA**) *P.pastoris* GS115 expression strains.

Amplified nS-pLapA DNA fragment was inserted into pPpHIS4\_BglII expression cassette by InFusion cloning to yield the nS-pLapA\_pPpHIS4 vector. Following scale-up, nS-pLapA\_pPpHIS4 DNA was linearized with *StuI* and transformed into competent *P. pastoris* GS115 strain. The transformed cells were plated on histidine-deficient minimal-dextrose (MD) agar. However, due to unidentified reasons, no transformants were obtained after 3-day incubation at 28 °C.

It was found that GS115 cells were not able to grow on minimal agar due to the lack of essential amino acids. Therefore, in the next *P. pastoris* transformation experiment, MD agar was supplemented with all essential amino acids, except histidine (to see if histidine auxotrophy was restored in pLapA-GS115 transformants). However, yet again no transformants were obtained on MD agar (- histidine) plates after a 3-day incubation at 28 °C, suggesting histidine auxotrophy was not restored. Nevertheless, GS115 colonies were visible on the YPD agar plate. No further PCR-based clone identification was performed, using these colonies, since these could simply be empty GS115 cell colonies, which failed

to uptake pLapA\_pPpHIS4 DNA. Colonies were also visible on the positive control MD agar plate, as expected (**Figure 7.3**).



**Figure 7.3** GS115 agar plates generated in the transformation experiment. Positive and negative controls are shown on the left. Variants transformed with pLapA\_pPpHIS4 DNA are shown on the right.

One of the reasons why GS115 *P. pastoris* failed to uptake the recombinant DNA could be due an old competent cell batch. A new transformation should be attempted in the future using a freshly prepared GS115 cell batch.

### 7.3 Kex2/Ste13 – mediated cleavage of LapA

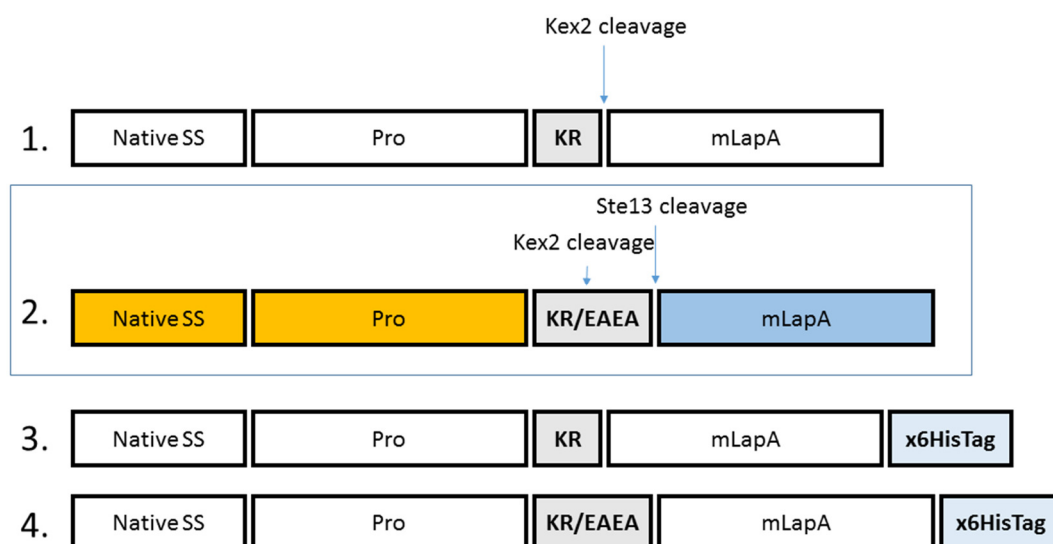
Another approach to successful secretion of mature LapA polypeptide into the culture supernatant of *P. pastoris* was undertaken. Yeast mating pheromone  $\alpha$  – factor (MAT $\alpha$ ) pre-pro leader polypeptide has been known to be able to direct the secretion of a target protein to the yeast culture supernatant; first observed in budding yeast *Saccharomyces cerevisiae* (Brake et al., 1984).

The MAT $\alpha$  secretion leader sequence is the most commonly used sequence in *P. pastoris* secretory expression trials. The sequence is comprised of a 19 amino acid long signal pre-sequence, followed by 60 amino acid propeptide sequence. Cleavage of this pre-pro leader sequence is mediated by a Kex2 protease in the Golgi, recognizing a di-basic (Lys-Arg) site,

which is present at the C-terminal end of the MAT $\alpha$  polypeptide sequence. The Kex2 cleavage site is followed by two Glu-Ala repeats, which are further trimmed by type IV dipeptidyl aminopeptidase, Ste13, also present in the Golgi (Daly and Hearn, 2005).

Therefore, it was desirable to engineer a dibasic Kex2 recognition site (Lys-Arg) between LapA propeptide and its mature sequence, thereby mimicking the processing of the MAT $\alpha$  leader sequence. In this way, it was hypothesised that LapA propeptide cleavage could occur, as part of the natural secretion process in *P. pastoris*. This would be highly favourable for industrial processing, as no subsequent downstream enzymatic treatments would be necessary to obtain fully functional, mature LapA.

Insertion of the Glu-Ala repeats, between the N-terminus of the target protein and the Kex2 cleavage site, would increase the proportion of correctly processed propeptide by preventing steric hindrance of the Kex2 cleavage site. However, due to the inability of Ste13 to cope with large amounts of recombinant protein produced (Daly and Hearn, 2005), the target protein could possess N-terminal Glu-Ala adducts, which could affect both structural and functional properties of the resultant enzyme. Therefore, in anticipation of potential processing problems (arising from the additional Glu-Ala N-terminal repeats), two separate LapA expression constructs were engineered. One possessing only Kex2 recognition site (Lys-Arg) and the other - Kex2/Ste13 site (KR/EAEA), as outlined in **Figure 7.4**. Additionally, two Kex2/Ste13 processing variants were modified to possess a C-terminal hexahistidine tag to facilitate downstream purification of expressed protein. Thus, four different constructs were made in total (**Figure 7.4**).



**Figure 7.4** Engineered LapA expression cassettes for integration into *P. pastoris* X33 strain. The corresponding Kex2 and Ste13 cleavage sites are indicated by arrows. Construct number 2 (blue rectangle) yielded secretion of mature active LapA in the culture supernatant.

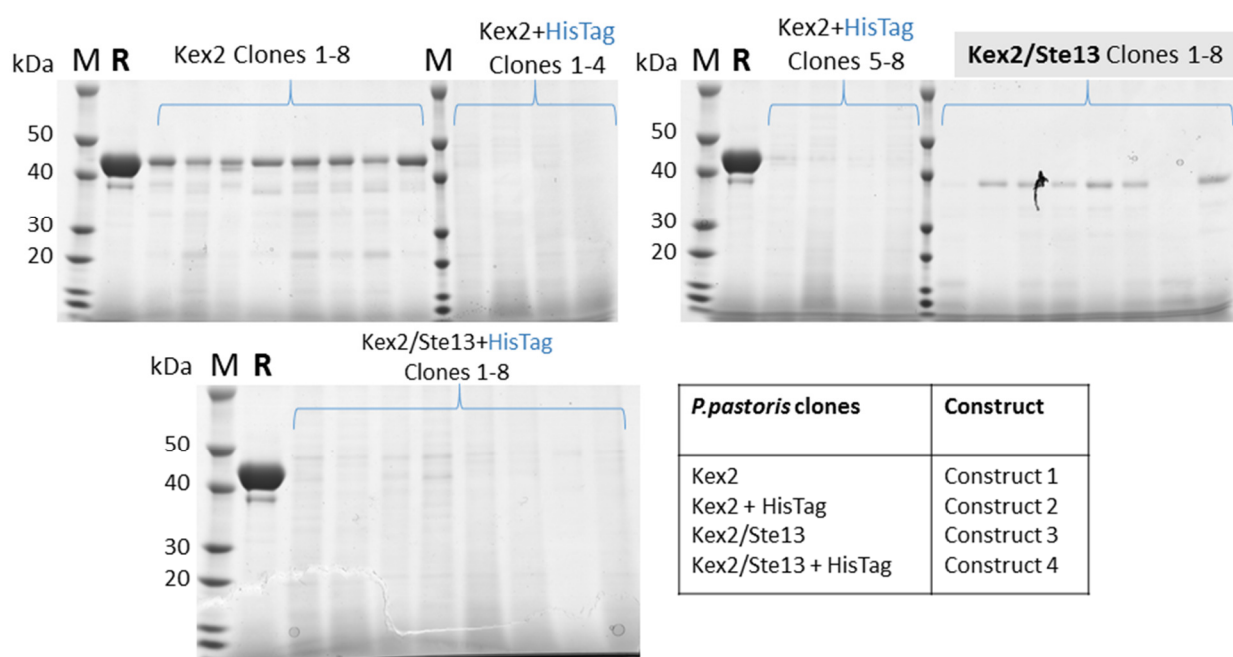
To introduce the Kex2/Ste13 recognition sites, a Quickchange™ mutagenesis protocol was implemented using the pD-npro-LapA vector as a template and the primer pairs shown in **Table 7.1**.

Constructs	Primers
Constructs 1 and 3	5'-AAAAGCTTGCTGTGACCAAAAGATATCCAGATAGCGTCCA-3' 5'-TGGACGCTATCTGGATATCTTTTGGTCACAGCAAGCTTTT-3'
Constructs 2 and 4	5'-CAAAAGCTTGCTGTGACCAAAAGAGAGGC TGAAGCTTATCCAGATAGCGTCCAA-3'  5'-TTGGACGCTATCTGGATAAGCTTCAGCC TCTCTTTTGGTCACAGCAAGCTTTTG-3'

**Table 7.1** Primers used in Quickchange™ cloning PCR reaction.

Colony PCR was carried out on randomly selected *E. coli* transformants using the following primers: 5' – GACTGGTTCCAATTGACAAGC – 3' and 5' – GCAAATGGCATTCTGACATCC – 3'. Three colonies were picked for each of the four constructs, and all gave a positive band when analysed on an agarose gel (not shown). Thereon, the established *P. pastoris* expression strain generation protocol (see section 2.13.1) was followed for all four constructs. Following transformation of *P. pastoris* X33

strain, around 30 ‘Bonus’ transformants were obtained for each of the four constructs. A 48-well high-throughput expression plate was set up using the Biolector micro-fermentation protocol (section 2.13.3). Eight ‘Bonus’ transformants were tested for expression of each construct. *P. pastoris* cultures were harvested after 70 h post-induction in secondary culture and the supernatants were analysed on SDS-PAGE (**Figure 7.5**).



**Figure 7.5** Recombinant protein secretion analysis of LapA Kex2/Ste13 constructs (1 – 4). M - molecular weight marker, R – *wtp<sub>pro</sub>LapA* protein as reference. Engineered LapA proteins, labeled to correspond to the designed expression constructs as in **Figure 7.4**.

Based on analysis of the expression of Kex2/Ste13 engineered LapA constructs in *P. pastoris*, only the expression strain harbouring LapA with both cleavage sites (Kex2/Ste13) secreted active mature LapA into the culture supernatant. Clonal variation was also visible in this strain, with only 6 out of 8 clones secreting the target protein and 2 yielding no secretion. *P. pastoris* clones with only the Kex2 cleavage site engineered, primarily secreted unprocessed LapA precursor. Therefore, the Kex2 site alone was not sufficient for appropriate cleavage of LapA propeptide, possibly due to inaccessibility and steric hindrance, which was reduced by adding the 'EAEA' Ste13 recognition site. Interestingly, almost no secretion of recombinant LapA was observed for constructs containing the additional C-terminal hexahistidine tag, which probably impeded correct folding of the

resulting LapA polypeptide chain and potentially triggered ER – associated protein degradation (ERAD) or unfolded protein response (UPR) in *P. pastoris*.

N – terminal sequencing (Appendix I, **Figure 10.4**) of secreted mature LapA protein showed an 'EAEA' extension to protein N – terminus, hence this recombinant protein was named -  $_{EA}$ LapA. As expected, Ste13 aminopeptidase was not able to cope with the processing of highly secreted LapA protein and, following Kex2 cleavage, no further modification of the N – terminus occurred. Inefficient processing likely affected the observed secretion level of  $_{EA}$ LapA protein, which was 0.05 mg/ml – six-fold reduction compared to *wtpro*LapA protein.

### **7.3.1 Purification of $_{EA}$ LapA protein**

Maturely secreted  $_{EA}$ LapA was purified using the newly established 3 – step protocol (section 2.13.4). The protein displayed apparent stronger affinity for the AEX resin (MonoQ column) and eluted at 250 mM NaCl, compared to *wt*LapA, which typically eluted at 140mM NaCl. This could be explained by the additional highly negatively charged 'EAEA' extension to the N – terminus, which is probably exposed on the protein surface, contributing to a net negative charge for LapA. Also, this increased negative charge appeared to render  $_{EA}$ LapA more prone to precipitation on an ultrafiltration membrane. Approximately 30 – 40% loss, in total protein quantity, was observed after each concentration step, using a typically inert Hydrosart membrane. Another distinguishing feature of  $_{EA}$ LapA was its strong association with media – derived brown pigments. Three purification passes on a MonoQ column still were not sufficient to completely eliminate the pigments, possibly because  $_{EA}$ LapA eluted at a higher concentration of NaCl, where pigmented LapA populations are typically released.

The measured specific activity of purified  $_{EA}$ LapA protein was  $130 \pm 1.2$  U/mg and was not significantly different from purified *wt*LapA protein ( $140.15 \pm 4.8$  U/mg). A slight reduction in activity may be linked to residual pigment – induced inhibition, as previously observed. Thus, it was concluded that the N – terminal 'EAEA' adduct had no significant effect on the specific activity of LapA.

## 7.4 MAT $\alpha$ -mediated secretion of mature LapA

The  $\alpha$ -mating factor (MAT $\alpha$ ) from *Saccharomyces cerevisiae* is the most commonly utilised leader sequence in *P.pastoris* expression cassettes, directing the polypeptide into culture supernatant. It is comprised of 19-aa signal sequence followed by 67-aa pro-region containing three glycosylation sites and a dibasic Kex2 recognition site (Kurjan and Herskowitz, 1982). Three-step processing of MAT $\alpha$  leader prepro sequence includes removal of signal peptide by signal peptidase in ER followed by Kex2-mediated cleavage of pro-region and, finally, N-terminal elimination of Glu-Ala repeats mediated by Ste13 aminopeptidase in Golgi (Brake et al., 1984).

Within MAT $\alpha$  leader sequence the pre-peptide is deemed to be important for translocation to ER by either co-translational or post-translational mechanism (Hegde and Bernstein, 2006). The role mostly hydrophobic propeptide region has not been established yet. It is speculated, that MAT $\alpha$  prodomain may act as molecular chaperonin to slow down and ensure proper folding process of secreted protein, because secretion of some heterologous proteins is abolished in the absence of pro-sequence (Chaudhuri et al., 1992, Kjeldsen et al., 1997).

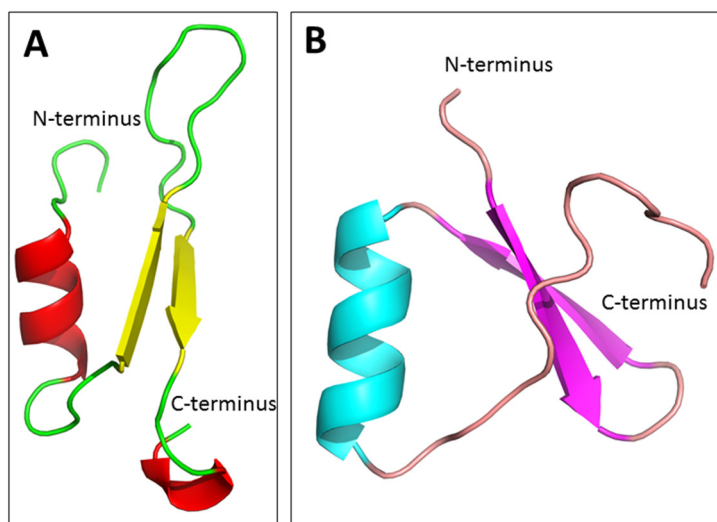
### 7.4.1 Background rationale

Lower than expected secretion levels, of  $_{EA}$ LapA recombinant protein, prompted us to explore alternative ways to produce mature LapA protein in the *P. pastoris* supernatant. Previous work, by Biocatalyst Ltd, to successfully express recombinant LapA in eukaryotic yeast *P. pastoris* strain X33 with a MAT $\alpha$  leader sequence, yielded comparable amounts of secreted LapA polypeptide harbouring the native secretion signal. However, due to the presence of the native LapA prodomain, located between the MAT $\alpha$  and LapA catalytic domains, the nascent secreted precursor protein still required enzymatic activation to reveal functional, mature protein.

Within the MAT $\alpha$  leader sequence, the pre-peptide is deemed to be important for translocation to the ER by either a co-translational or a post-translational mechanism (Hegde and Bernstein, 2006). The role of the hydrophobic propeptide region has not been established yet. It is speculated that the MAT $\alpha$  prodomain may act as a molecular chaperonin to slow down and ensure proper folding of secreted protein because secretion of some heterologous proteins is abolished in the absence of the pro-sequence (Chaudhuri et al., 1992, Kjeldsen et al., 1997).

Interestingly, as illustrated in **Figure 7.6**, a striking structural similarity was discovered between the MAT $\alpha$  leader sequence prodomain region and the native LapA prodomain. This

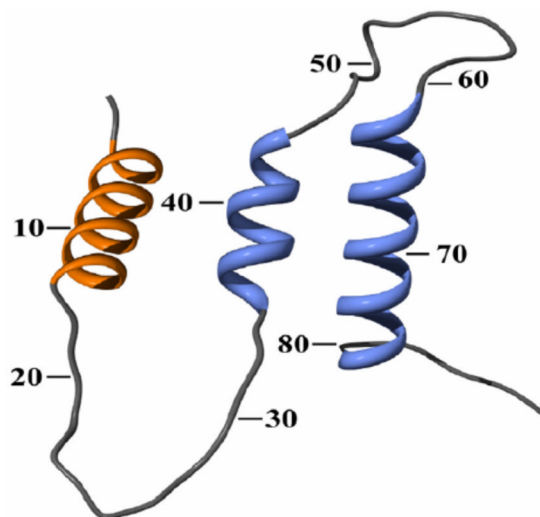
structural comparison was only possibly following determination of the crystal structure of full LapA precursor, containing the pro-region obtained during the course of this thesis work. MAT $\alpha$  prodomain was modeled using the SWISS-MODEL server, and the resulting structural coordinates of the two pro-regions in question were visually compared (**Figure 7.6**).



**Figure 7.6** Structural comparison of MAT $\alpha$  prodomain (**A**) and LapA prodomain (**B**). MAT $\alpha$  prodomain was modeled using SWISS-MODEL server (Biasini et al., 2014); the structure of LapA prodomain was determined by X-ray crystallography (see section 5.1).

The structural comparison led to the hypothesis that the MAT $\alpha$  prodomain alone might be able to assist in molecular folding of LapA prodomain, thus omitting the LapA propeptide in the expression construct. Consequently, we would anticipate efficient cleavage of MAT $\alpha$  leader polypeptide and high-level secretion of a correctly folded, active form of LapA – the result, which could not be achieved completely with <sub>EA</sub>LapA protein, as it was hampered by the design of an optimal Kex2/Ste13 recognition site that prevented processing of LapA prodomain.

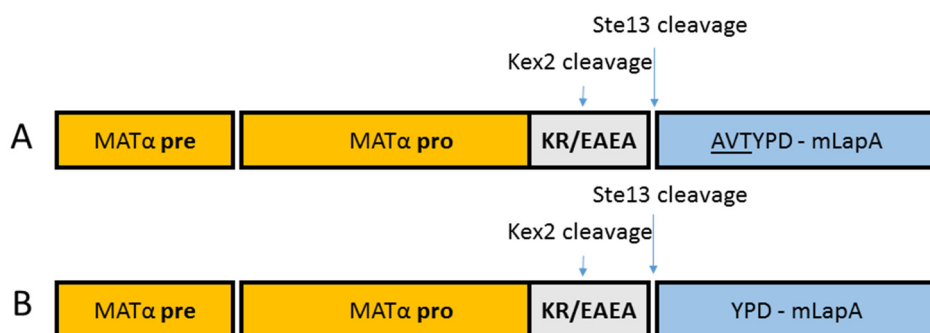
Similarly, Lin-Cereghino and colleagues attempted modeling of a MAT $\alpha$  prodomain region (Lin-Cereghino et al., 2013). However, in their molecular model, obtained using JPRED3 secondary structure prediction server followed by Modeller structure modeling, the MAT $\alpha$  propeptide was predicted to form two aligned helices spanning the C-terminal region (**Figure 7.7**), contrary to two antiparallel  $\beta$ -sheets predicted in our model following template based structure prediction approach.



**Figure 7.7** Model of MAT $\alpha$  propeptide obtained in the study by Lin-Cereghino et al., 2013.

#### 7.4.2 MAT $\alpha$ -LapA cloning and expression in *P. pastoris*

All LapA recombinant protein precursors, processed by the AP1 protease containing fraction from Flavourzyme, contained an additional 3-residue extension ('AVT') to the original LapA N-terminus observed in the crystal structure of *flav*LapA. The effect of omitting these three residues, on protein thermolability, was not experimentally determined. Therefore, given the successful secretion of appropriately folded LapA from this experiment, any characterized thermolability mutants would have to be reconstructed in the expression cassette, harbouring the 'AVT' extension. In this way, the resulting thermolability profiles would be internally consistent for any mutants produced. For this reason, two expression constructs, namely M $\alpha$ -AVT-LapA and M $\alpha$ -YPD-LapA (M $\alpha$  corresponds to MAT $\alpha$  prepro leader), were designed, possessing distinct N-termini, as shown in **Figure 7.8**.



**Figure 7.8** Engineered LapA expression cassettes for integration into *P. pastoris* X33 strain. **A** - M $\alpha$ -AVT-LapA protein with 3-residue extension; **B** - M $\alpha$ -YPD-LapA protein equivalent

to native LapA from *Aspergillus oryzae* (*flav*LapA). MAT $\alpha$  pre pro leader sequence is shown in yellow, with a Kex2/Ste13 cleavage site in grey; released LapA catalytic domain (mLapA) is depicted in blue.

Deletion mutagenesis of LapA propeptide was performed by a Quickchange™ protocol, using a previously constructed MAT $\alpha$ -pro-LapA vector from Biocatalysts Ltd. **Table 7.2** gives the primer pairs used in the PCR reaction.

Constructs	Primers
Construct A	5'-AAAGAGAGGCCGAAGCTGCTGTGACCTATCCA-3' 5'-TGGATAGGTCACAGCAGCTTCGGCCTCTCTTT-3'
Construct B	5'-AAGAGAGGCCGAAGCTTATCCAGATAGCGTCC-3' 5'-GGACGCTATCTGGATAAGCTTCGGCCTCTCTT-3'

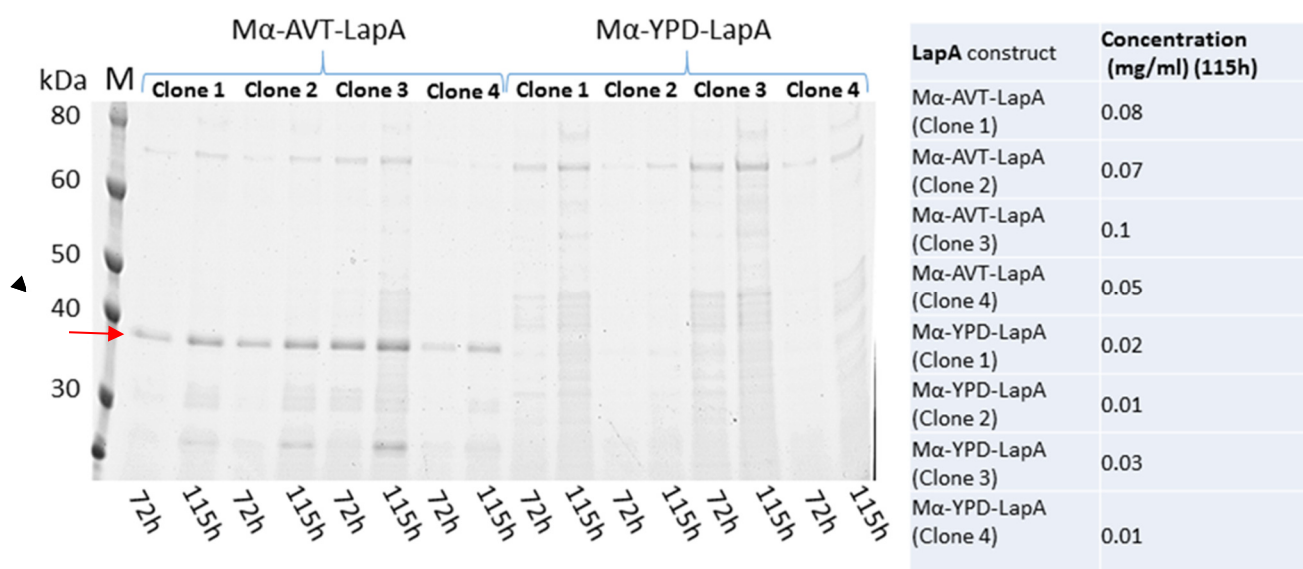
**Table 7.2** Primers used in Quickchange™ deletion mutagenesis reaction.

Positive *E.coli* transformants were verified by Sanger sequencing (section 2.12.2). Approximately 10  $\mu$ g of linearized constructs A and B were transformed into *P. pastoris* X33 strain. After 3 days, >50 colonies were observed on high concentration Zeocin ('Bonus') plates for both recombinant strains. 200 mL shake flask expression trial was performed, providing the *P.pastoris* strains with 0.5% final concentration of MeOH. Monitored optical cell density is reported in **Table 7.3**.

M $\alpha$ -AVT-LapA	OD <sub>600</sub> 72h	OD <sub>600</sub> 115h	M $\alpha$ -YPD-LapA	OD <sub>600</sub> 72h	OD <sub>600</sub> 115h
Clone 1	28	32	Clone 1	23	28
Clone 2	34	36	Clone 2	29	34
Clone 3	27	32	Clone 3	23	33
Clone 4	25	28	Clone 4	28	29

**Table 7.3** Cell density measured in the secondary culture of different clones of M $\alpha$ -AVT-LapA and M $\alpha$ -AVT-LapA strains.

*P. pastoris* cultures were harvested after 115h from induction in secondary culture and the supernatants were analysed by SDS-PAGE (**Figure 7.9**).



**Figure 7.9** Recombinant protein secretion analysis of Mα-AVT-LapA and Mα-AVT-LapA constructs. Protein secretion levels were quantified by Bradford assay and summarized in the shaded box (right). POI is indicated by red arrow.

As seen in **Figure 7.9**, protein secretion products were obtained from all clones tested. A protein migrating at approximately 37 kDa corresponds to mature LapA, presumably containing the 'AVT' extension to its N-terminus (Mα-AVT-LapA). Rather surprisingly, exclusion of these three residues, in the Mα-YPD-LapA construct, completely abolished secretion of the 37 kDa species. Low specific activity (0 – 5 U/mg) towards the LPNA substrate was observed in all supernatants (only qualitative measurements carried out). Furthermore, for the Mα-YPD-LapA supernatants, activity was extremely low (approximately 10-fold reduced) compared to the Mα-AVT-LapA supernatants. This finding supports the absence of the 37 kDa LapA protein in the Mα-YPD-LapA supernatants. However, residual LapA specific activity in these supernatants could be attributed to an observed 70 kDa species. The presence of high molecular weight bands, under reducing conditions, indicates that recombinant LapA is prone to formation of compact higher order oligomeric states, which could suggest the existence of populations of improperly folded protein in solution. It is not clear why the Mα-YPD-LapA construct only produced mostly

inactive oligomers. The identity of these oligomers should be experimentally confirmed by Western Blot using LapA antibody.

This experiment indicates that substituting the LapA prodomain with the MAT $\alpha$  leader sequence may not be a feasible solution, to yield secreted, correctly folded LapA in *P. pastoris*. Purification and subsequent quantification of M $\alpha$ -AVT-LapA activity may provide a some clarity and insight toward the actual potential of this enzyme for industrial use.

#### **7.4.2.1 Purification and specific activity of M $\alpha$ -AVT-LapA**

Maturely secreted M $\alpha$ -AVT-LapA protein was purified using the newly established 3 – step protocol (section 2.13.4). A typical separation from pigmented fractions, during SEC purification, was compromised for this protein. Specifically, the peak exhibiting high LapA activity was very broad and overlapped the brown pigment fractions with higher than usual elution volume. Additionally, the isolated fractions (from SEC) showed stronger affinity to the AEX resin of the Q HP column and eluted at 500 mM NaCl compared to *wt*LapA typically eluting at 140 mM NaCl. These observations, during purification of this construct, are highly suggestive of non-specific association with the chromatographic resin by partially unfolded (or mis-folded) M $\alpha$ -AVT-LapA protein. The specific activity measured for purified M $\alpha$ -AVT-LapA protein was only  $7 \pm 0.5$  U/mg, indicating almost inactive LapA, which would corroborate the presence of unfolded/mis-folded protein.

This experiment negated the hypothesis that a MAT $\alpha$  prodomain could replace the native LapA propeptide to act as an intramolecular chaperone and yield active mature LapA.

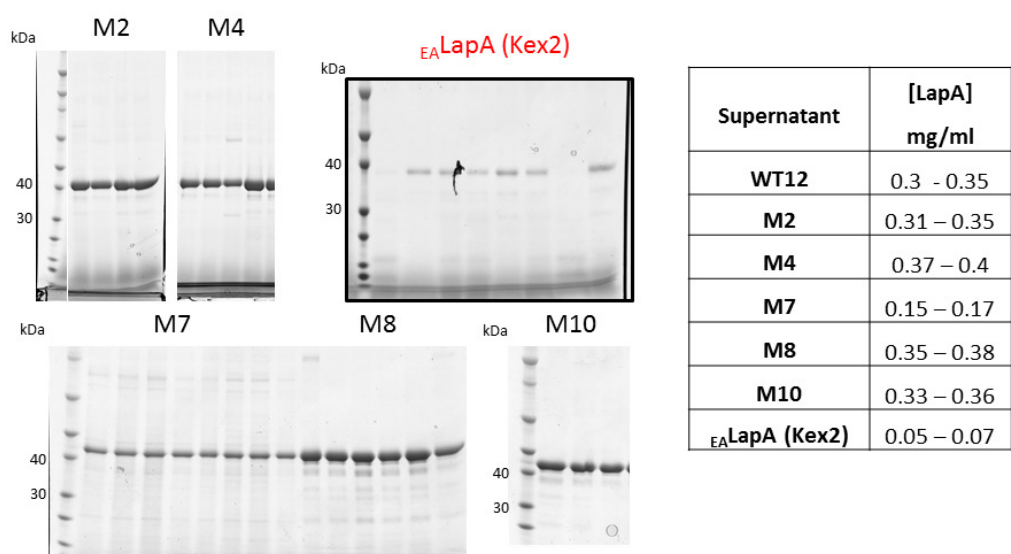
## 7.5 Considerations for pilot scale production of selected LapA wild-type and thermolabile mutant proteins

*This section slightly deviates from the main story line of this chapter. However, it was incorporated here, at the end of all result chapters, relating to production of LapA wild-type and mutant proteins in *P. pastoris*. It was necessary to integrate all the secreted expression data, in light of the pilot scale production.*

In this part of the project, a series of LapA mutants with improved thermolability properties were over-expressed in food-grade yeast – *Pichia pastoris*. The apparent clean secretion profiles of these mutant proteins would enable Biocatalysts to carry out application trials in whey hydrolysis with crude culture supernatant samples, thus, bypassing subsequent downstream purification steps. However, at the current stage of this project, WT LapA and thermolability mutants are secreted as inactive precursors, limiting the direct transformation of *P.pastoris* supernatant into product. In the established LapA precursor processing protocol we utilise Flavourzyme-derived non-specific alkaline protease 1 (AP1). Co-expression of this proteolytic enzyme with recombinant LapA precursor in a single *P.pastoris* strain could potentially yield an interesting endo/exopeptidase blend to the culture supernatant, which would not only contain an active LapA population, but also non-specific endoproteolytic activity, both of which could be exploited for efficient whey hydrolysis. Also, separately over-expressed LapA and AP1 could be recombined in pre-determined proportions at the latest stage of product preparation. Other commercially available proteolytic enzymes (e.g. trypsin) could be added directly to bioreactor upon LapA production to enable processing, however, recombinant LapA precursor was demonstrated to be susceptible to complete degradation by some other non-specific endopeptidases (**Figure 4.35**).

### *Secreted expression levels of LapA variants*

Since recombinant protein expression levels are one of the key factors for industrial scale production, comparisons of the observed expression levels in shake flask fed-batch culture of the most thermolabile LapA mutants and WT enzyme are necessary, to identify the ‘best producing’ *P.pastoris* clones for each protein.



**Figure 7.10** Secretion levels of different LapA thermolability mutants. Table (right) represents the measured range of secreted protein depending on the *P.pastoris* clone (left).

Expression levels of all LapA thermolability mutants are comparable after harvesting the supernatants at 120 h, except M7 double mutant showing 2-fold reduction compared to the WT LapA (**Figure 7.10**). In this double mutant, the two site-directed mutations could have caused fold disruptions. Thus, potential explanation could be a larger distribution of unfolded M7 protein populations across *P.pastoris* secretion pathway. This could decrease the secretion levels of M7 due to an elimination of misfolded protein by *P. pastoris* secretion machinery.

Also,  $_{EA}$ LapA (Kex2 processed) expression construct, which gave access to mature LapA polypeptide in *Pichia* supernatant, showed 6-fold reduction in expression levels. Introduction of site-directed thermolability mutations in  $_{EA}$ LapA protein might give a further reduction of secreted protein, which may hamper utilisation of this recombinant protein.

The requirement of post-translational processing of LapA mutants by exogenous proteases may further decrease the end yield of recombinant protein. However, large-scale fermentation in a bioreactor (available at Biocatalysts) under tightly controlled growth conditions (T, pH, CO<sub>2</sub>, carbon source) would permit to obtain greater biomass and, hence, recombinant protein yields.

The determination of the integrated copy number of recombinant LapA gene might be essential in order to enhance the productivity of *P.pastoris* strain. All the thermolability

mutant strains were generated under high antibiotic (zeocin) concentrations (‘‘Bonus’ transformants’), suggesting multiple LapA gene integration events. However, it was reported that the assumption ‘higher gene dosage equals higher production’ was not necessarily true due to detrimental effects on cell metabolism and survival by certain genes (Aw and Polizzi, 2013). Therefore, determination of gene copy number by qPCR in mutant strains might be the first step towards improving secretion levels. The batch culture mode (shake-flasks or deep-well plates) can be used for screening the secretion levels of different gene dosage clones.

Lastly, considering the thermolability aspect of all recombinantly produced proteins in this project, LapA lacking the N-terminal propeptide in the expression construct, termed *wtΔproLapA*, displayed the greatest instability and heat sensitivity, as discussed in section 5.1. However, the main pitfalls with this enzyme was that (i) it was secreted as folding intermediate in pre-molten/molten globule state due to the absence of propeptide-mediated chaperonin effect, with (ii) consequential severely reduced specific activity. Also, the secretion levels of *wtΔproLapA* protein were very low in *P.pastoris*, nonetheless comparable to M7 mutant. Therefore, if appropriate ways to stabilise this polypeptide, immediately after large-scale production, could be identified, Biocatalysts could consider exploiting this already partially unfolded, heat labile peptidase. Importantly, *wtΔproLapA* protein is secreted in its mature active form and would not require additional enzymatic activation.

**Chapter 8 AMINOPEPTIDASE FROM *GLACIOZYMA*  
*ANTARCTICA* (GAP): A PSYCHROPHILIC HOMOLOGUE OF LAP**

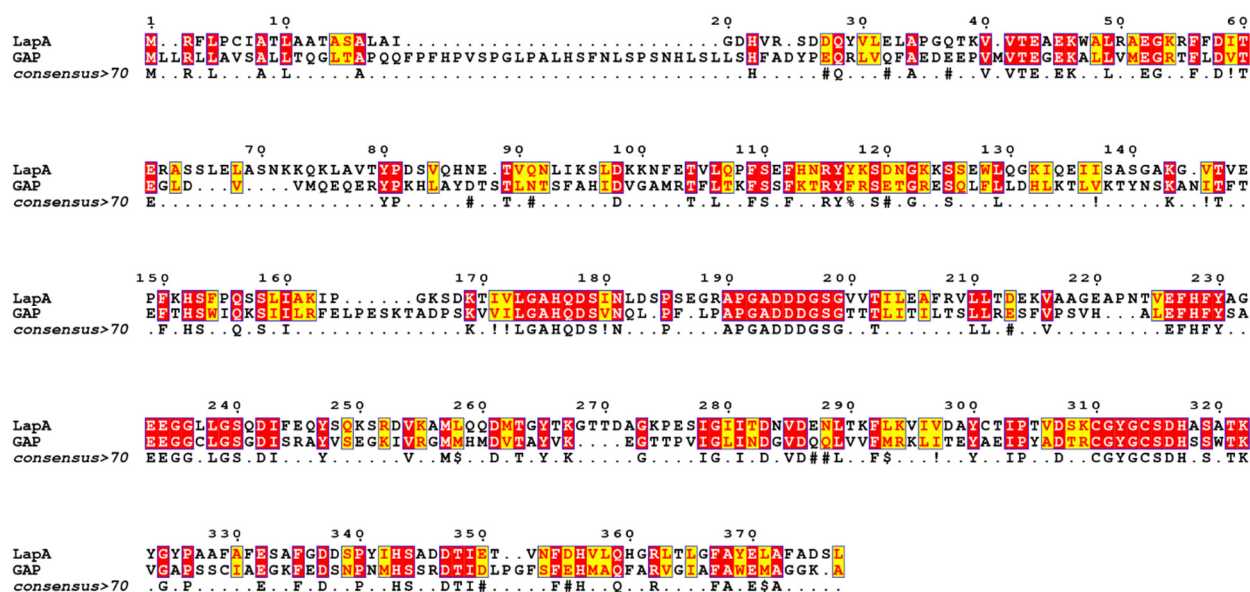
---

## 8.1 Introduction and preliminary bioinformatics study

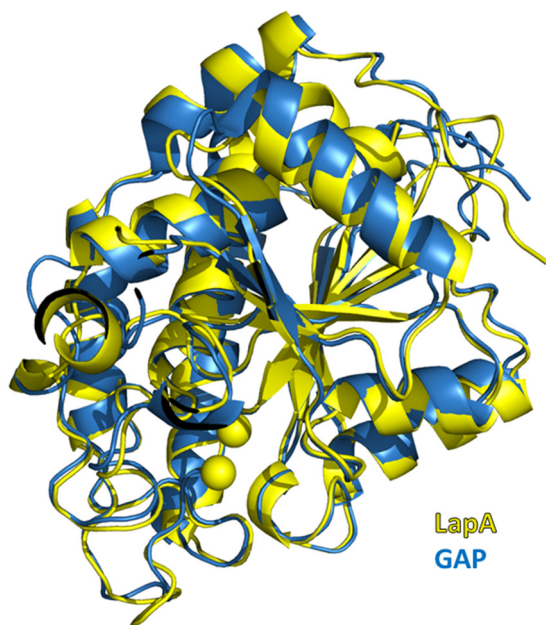
During the latter stages of this thesis work, *Glaciozyma antarctica* leucine aminopeptidase (GAP), an obligatory psychrophilic Antarctic yeast protein, was identified as having 40% sequence identity to LapA, following release of the GAP sequence in 2015 (**Figure 8.1**). Biochemical characterisation and structure determination of this Antarctic protein could provide a better understanding of M28 peptidase cold adaptation. If GAP is naturally a cold - adapted peptidase, then a crystal structure of this enzyme would allow determination of potential structural features of cold adaptivity and thermolability, by extension to other members of the M28 peptidase domain family, which would be completely novel discovery.

Indeed, it is very likely that GAP is genuine cold adapted aminopeptidase, since other isolated and characterised enzymes from this Antarctic fungus showed optimal activity temperatures around 15 - 20 °C, instability and low inactivation temperature below 40 °C, traits characteristic of psychrophilic proteins (Alias et al., 2014, Turkiewicz et al., 2003).

An homology model of GAP was constructed using the IntFold2 modeling server (Roche et al., 2011), and the newly determined structure (from this thesis work) of *wt*LapA. Despite only 40% sequence identity to LapA, GAP showed almost an identical tertiary fold (**Figure 8.2**), confirming the canonical model of observing a higher degree of conservation at the structural level. The striking similarity of these proteins provided a strong impetus to pursue characterisation of GAP aminopeptidase in an effort to study natural structural adaptations to cold environments and tentative thermolability of this new aminopeptidase homologue.



**Figure 8.1** Pairwise alignment of LapA and GAP full open reading frame sequences. Identical and similar residues are shaded in red and yellow, respectively.



**Figure 8.2** Structural superimposition of the crystal structure of *wtLapA* (yellow) and the homology model of GAP (blue). RMSD, based on the  $\alpha$  backbone, between the two models is 0.865.

## 8.2 Recombinant cloning and expression of GAP in *P.pastoris*

Given the experience of recombinant expression of Lap, presented herein in working leading up to this discovery, and since GAP originates from a fungal organism, recombinant expression was only attempted in *P. pastoris*.

Codon optimized and synthesized GAP open reading frame (**Table 8.1**), containing the N-terminal native signal sequence and propeptide, was placed under control of the P<sub>AOX</sub> methanol-inducible promoter in the pD902 expression vector. In this way, an expression cassette, harbouring the native GAP signal peptide and propeptide sequences (pD-npro-GAP), was obtained. Another variant of GAP expression cassette also was produced in pD912 vector, sharing the same features, except the native GAP signal sequence was replaced by the MAT $\alpha$  secretion signal. The rationale behind making the latter construct was to evaluate the potential difference in secretion levels between the two constructs.

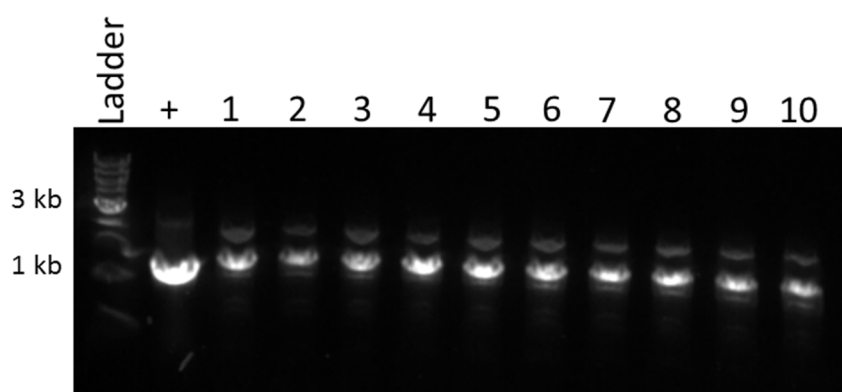
Expression construct	Proteins	Open reading frame sequence
pD902-npro-GAP	<i>wtproGAP</i>	MLLRLLAVSALLTQGLTAPQQFPFHPVSPGLPALHSFNLSP SNHLSLLSHFADYPEQRLVQFAEDEEPPVMVTEGEKALLVM EGRTFLDVTEGLDVVMQEQUERYPKHLAYDTSTLNTSFAHI DVGAMRTFLTKFSSFKTRYFRSETGRESQLFLLDHLKTLVK TYNSKANITFTEFTHSWIQSIILRFELPESKTADPSKVILG AHQDSVNQLPFLPAPGADDDGSGTTTLITLTSLLRESFVPS VHALEFHFYSAEEGGCLGSGDISRAYVSEGKIVRGMMMD VTAYVKEGTTPVIGLINDGVDQQLVVFMRKLITEYAEIPYA DTRCGYGCS DHSSWTKVGAPSSCIAEGKFEDSNPNMHSSR DTIDLPGFSFEHMAQFARVGIAFAWEMAGGKA
pD912-MAT $\alpha$ pro-GAP	<i>wtproGAP</i>	MRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYS DLEGDFDVAVLPFSNSTNNGLLFINTTIAAIAKEEGVSLEK REAEAAPQQFPFHPVSPGLPALHSFNLSPSNHLSLLSHFAD YPEQRLVQFAEDEEPPVMVTEGEKALLVMEGRTFLDVTEGL DVVMQEQUERYPKHLAYDTSTLNTSFAHIDVGAMRTFLTKF SSFKTRYFRSETGRESQLFLLDHLKTLVKTYNSKANITFTEF THSWIQSIILRFELPESKTADPSKVILGAHQDSVNQLPFL PAPGADDDGSGTTTLITLTSLLRESFVPSVHALEFHFYSAE EGGCLGSGDISRAYVSEGKIVRGMMMDVTAYVKEGTTPVI GLINDGVDQQLVVFMRKLITEYAEIPYADTRCGYGCS DHSS SWTKVGAPSSCIAEGKFEDSNPNMHSSRDTIDLPGFSFEHM AQFARVGIAFAWEMAGGKA

**Table 8.1** Open reading frame sequences of LapA expression cassettes synthesized in pD902 and pD912 *P. pastoris* expression vectors. Native GAP and MAT $\alpha$  secretion signals, propeptide, mature sequences are depicted in red, blue and black, respectively.

7.4 µg and 7 µg of SacI – linearized DNA was transformed into competent *P. pastoris* cells for pD902-npro-GAP and pD912-MAT $\alpha$ pro-GAP constructs, respectively. The latter construct produced ‘Bonus’ transformants with potential multiple - copy gene recombination events. The expression cassette, harbouring the native signal sequence (pD902-npro-GAP), gave no ‘Bonus’ transformants and colonies were only visible on low antibiotic plates.

Therefore, colony PCR was carried out to determine genuine transformants, using the standard AOX1 primer pair: 5’ – GACTGGTTCCAATTGACAAGC – 3’ and 5’ – GCAAATGGCATTCTGACATCC – 3’. All ten randomly selected pD-npro-GAP transformants produced a positive signal in the colony PCR experiment (**Figure 8.3**), suggesting these all contained at least one copy of the GAP target gene. Clones 1 and 2 were selected for subsequent shake flask expression trial.

200 mL shake flask cultures were used for recombinant expression of pD902-npro-GAP and pD912-MAT $\alpha$ pro-GAP in the *P. pastoris* strains. The rest of the heterologous GAP expression protocol was according to the procedure described in section 2.13.3. As illustrated in **Table 8.2**, the observed growth of both recombinant GAP strains was quite slow (typically the OD<sub>600</sub> exceeds 35 absorbance units after 96 h). After 96 h of MeOH induction, the supernatants were harvested and analysed by SDS – PAGE (**Figure 8.4**).

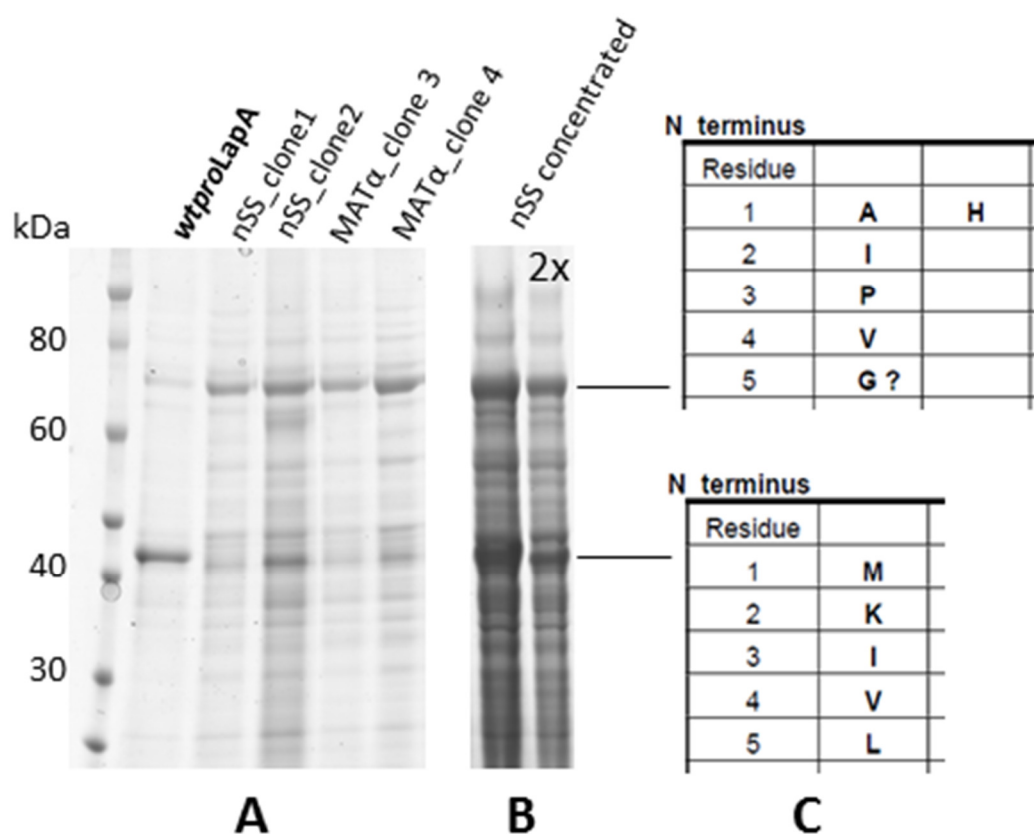


**Figure 8.3** 1% DNA agarose gel analysis of PCR amplicons of pD-npro-GAP transformants. Positive (+, PCR using pD-npro-LapA as a template) control is indicated. The expected size of GAP amplicon is 1361 bp.

pD902-npro-GAP	OD <sub>600</sub> 24h	OD <sub>600</sub> 48h	OD <sub>600</sub> 72h	OD <sub>600</sub> 96h
Clone 1	12	21.5	22	20
Clone 2	11	18.85	21	25

pD912-MAT $\alpha$ pro-GAP	OD <sub>600</sub> 24h	OD <sub>600</sub> 48h	OD <sub>600</sub> 72h	OD <sub>600</sub> 96h
Clone 3	10	12	25	30
Clone 4	12	16.85	21	27.5

**Table 8.2** Cell density measured in the secondary culture of different GAP expression clones.



**Figure 8.4** Expression analysis of secreted putative GAP protein. **A** – *wtproLapA* loaded for reference, nSS – GAP clones harbouring native signal peptide, MAT $\alpha$  – GAP clones harbouring MAT $\alpha$  secretion leader; **B** – concentrated nSS\_clone 2 supernatant, 2x – 50% dilution of the same supernatant; **C** – results from N – terminal sequencing of indicated bands.

The expected molecular weight of GAP precursor protein is 42.5 kDa (383 aa), which is quite similar to 39.3 kDa (359 aa) theoretical molecular weight of LapA precursor. However, due to N – linked glycosylation *wtpro*LapA migrates around 42 – 43 kDa. In GAP, N115 and N169 are predicted to be glycosylated, which should increase the molecular weight of the final protein product to approximately 45 – 47 kDa, assuming the extent of glycosylation is similar.

From an analysis of the GAP supernatants, it was apparent that the secretion profile was heterogeneous. Two main protein bands were predominant in all GAP supernatants: one migrating at 42 – 45 kDa and another at approximately 70 kDa. Both proteins bands were excised and N – terminally sequenced (**Figure 8.4**, C part). The first five residues expected at the N- terminus of GAP precursor is 'APQQF'. However, both N – termini observed for the unknown secreted proteins were very distinct from that expected, namely 'MKIVL' and 'AIPVG', for the 42 – 45 kDa protein and 70 kDa protein, respectively (**Figure 8.4**, C part).

Also, there was very low activity observed towards LPNA (leucine *p*-nitroanilide) substrate, which might indicate low levels of expression of GAP precursor, which could not be readily identified in this experiment. In addition, despite the sequence and structural homology to LapA, the substrate specificity of GAP peptidase may be quite distinct. Therefore, activity towards LPNA substrate may not be the best indicator of active aminopeptidase in GAP *P.pastoris* supernatants.

In general, *P. pastoris* does not secrete detectable levels of native proteins into culture supernatant, therefore recombinantly produced protein usually constitutes vast majority of extracellular protein content (Macauley-Patrick et al., 2005). In particular, **Figure 8.4** B part shows the presence of secreted multiple unidentified protein bands in supernatant of GAP clones harbouring native signal peptide. It was concluded, that integration of GAP precursor gene triggered aberrant protein secretion response in *P.pastoris*.

### 8.3 Conclusions and future possibilities

The optimal growth temperature (28 °C) of the *P. pastoris* X33 strain sets a considerable challenge for the correct folding and structural integrity of this Antarctic aminopeptidase. The next step in successful heterologous expression of GAP protein would be a low temperature expression trial, which time did not permit. A thermolabile aspartic proteinase from *Mucor mucedo* DSM 809 showed improved activity in *P. pastoris* expression upon

incubating the strain at 20 – 24 °C (Yegin and Fernandez-Lahore, 2013). Similarly, the optimal hydrolytic activities at 20 °C were observed for cold-adapted serine protease from *G. antarctica* expressed in *P. pastoris* (Alias et al., 2014). It is reasonable to assume that lowering the post-induction temperature in BMMY medium might have a positive effect on GAP expression. To date, there are no available yeast expression systems, specifically tailored to the expression of cold-adapted proteins. However, several such systems are available for *E.coli* host: i) based on co-expression with cold-adapted chaperonins, allowing an operational folding system at 4 – 12 °C (Ferrer et al., 2003); and low temperature inducible promoter utilising a cold shock protein (*cspA*) (Vasina and Baneyx, 1996). Therefore, GAP over-expression in these cold-temperature targeted *E.coli* host systems is another possible option.

The absence of recombinant yeast expression hosts, specifically for psychrophilic proteins, could encourage Biocatalysts Ltd (or other industry leaders) to explore alternative cold-tolerant fungal yeasts in order to exploit the ever-increasing number of psychrophilic protein targets of interest to industry.

## Chapter 9 GENERAL DISCUSSION AND FUTURE DIRECTIONS

---

*This chapter presents an overall discussion and suggestions for future work based on the experimental results obtained throughout this thesis work. It highlights the main obstacles and identifies potential solutions, on the pathway toward engineering thermolabile industrial peptidases. Also, the chapter reiterates the novelty of the thesis work, presented in Chapters 3 -8, with regard to an unconventional engineering approach for peptidase cold-adaptation and thermal sensitivity via destabilisation and mimicry of structural features observed in homologous psychrophilic enzymes.*

## **9.1 Industrial relevance of cold-adapted proteases: the driving force**

Industrial biocatalysis can benefit from enzymes that remain active at low temperatures. Reduction of the reaction temperature, at no expense of catalytic rates, provides an apparent advantage to cold-active enzymes (over warm-adapted counterparts) and permits significant energy savings in industrial enzymatic processing (Feller, 2013). The widespread chemical elimination of biocatalysts is a special concern in the food industry, where high-temperature inactivation is detrimental to the reaction products. In this respect, psychrophilic enzymes present another useful property; complete inactivation at low-to-moderate temperatures, termed structural thermolability (Georlette et al., 2004, Gerday et al., 2000). The inherent intramolecular flexibility of such enzymes results in increased thermal sensitivity and permits superior kinetic characteristics in a low entropy environment (cold).

Industrial processes, sensitive to the occurrence of secondary reactions, heat-sensitive substrates, or increased risk of microbial contamination, all benefit from reduced temperature biocatalysis (Jeon et al., 2009). In food and feed industries, it is imperative to avoid harming or changing the nutritional value and flavour of the heat-sensitive substrates and products. Furthermore, selective inactivation at moderate temperatures by heat-labile enzymes permits controllable inhibition of sequential reactions at intermediate steps in the food industry (Cavicchioli et al., 2011).

Enzymatic hydrolysis is a preferred method in the production of functional foods, where soybean and whey protein hydrolysates (WPH) are used as hypoallergenic ingredients. Due to the simultaneous action of endo- and exopeptidases such hydrolysates display improved techno-functional properties, such as solubility, emulsifying capacity along with sensory perception properties: (i) reduced bitterness and (ii) more pleasant taste, which in turn favours customer acceptance (Meinlschmidt et al., 2016). Commercial enzyme blends, including endo-acting Subtilisin Carlsberg (Alcalase) and exo-acting leucine aminopeptidase A (Flavourzyme) are often used to achieve the aforementioned characteristics in whey and soy peptide hydrolysates (Ma et al., 2013).

To prevent excessive bitterness, or when the desired degree of hydrolysis (DH) and characteristics of the hydrolysate have been achieved, the active endoprotease and/or exopeptidase needs to be deactivated, preventing further proteolysis (Conesa and FitzGerald, 2013). This is usually achieved by heat treatment (>85 – 90 °C for >20 minutes) causing thermal denaturation of previously active peptidases. Yet, upon heating, various undesirable

chemical (Maillard reaction products) and physical (peptide denaturation) effects have been observed. Hence, the biological, functional and nutritional properties of hydrolysates are affected by heat processing, for which energy input must be minimised (Nicoli et al., 1997, Lijun et al., 2012). For example, gelation of peptide hydrolysates upon elevated ( $>70^{\circ}\text{C}$ ) temperatures is a well-recognised problem in the whey hydrolysis industry, resulting in spoiling the end product and clogging heat exchange vessels.

Therefore, there is a high demand by Industry for thermolabile endo- and exo-peptidases that can be inactivated by mild heat treatment such that the beneficial properties and quality of the vegetable and whey protein hydrolysates are preserved. In this project, the industrial partner – Biocatalysts Ltd – identified a commercial need for both endo- and exo-peptidases that can be inactivated  $<50^{\circ}\text{C}$ . The common utilisation of Subtilisin Carlsberg (sC) and leucine aminopeptidase A (LapA), in industrial vegetable and whey protein hydrolysis, determined the selection of these enzymes for the development of cold-activation and thermolability, via protein engineering. Consequently, the overall aim of this study was to engineer two commercially utilised mesophilic peptidases, namely sC and LapA, to confer psychrophilic-like behaviour (improved cold-activity) and, in particular, thermolability (complete low-to-moderate temperature inactivation). To address this aim, a number of specific objectives were identified, as follows: (i) identification of structurally viable mutations, using the known crystal structure of Subtilisin Carlsberg, which may lead to improved thermolability, (ii) production and characterisation of identified sC mutants in *E.coli*, (iii) isolation of native LapA from Flavourzyme and subsequent structure determination using X-ray crystallography, (iv) determination of structurally viable mutations, using newly elucidated LapA structure, (v) kinetic and thermodynamic characterisation of the most thermolabile sC and LapA mutants, and (vi) production of the most promising mutant protein(s), on a pilot scale at Biocatalysts, for evaluation of such mutants in whey hydrolysis application.

To date, no alternative industrial formulations (containing either cold-adapted or thermolabile peptidases) are available to the food protein hydrolysis market. The novelty of engineering cold adaptivity in mesophilic proteases, lack of cold-adapted and thermolabile proteins in multiple industries, as well as the paucity of structural information concerning cold-adapted proteins, signifies the importance of this project from both: (i) the perspective of fundamental understanding of protein structural properties governing thermal stability/lability and temperature adaptation and, (ii) the commercial perspective for the

industrial partner (Biocatalysts Ltd), which, upon achieving the targets of this project, could use the obtained sC and LapA mutants in formulating unique enzyme blends.

## 9.2 Project highlights: from inception to completion

This project is focused toward altering the thermal properties of industrially relevant proteases, namely, increasing the thermolability of select endo- and exo-peptidases. As this is an industrial case study, this project involved close collaboration with Biocatalysts Ltd, who previously had identified an industrial need for an exopeptidase and/or endoprotease that could be inactivated at temperatures <50 °C. At the time of this project, the two most relevant proteases for Biocatalysts Ltd were leucine aminopeptidase (LapA) from *Aspergillus sp.* and Subtilisin Carlsberg (sC) from *Bacillus sp.* These proteases display mesophilic-like thermal adaptation. Thus, the initial overall aim of this project was to alter their thermal properties in order to confer thermolability and potentially increase low temperature activity, thus achieving psychrophilic thermal adaptation.

The project was initiated by performing a thorough literature review of structural and functional determinants of protein thermal properties, with a focus on understanding the existing strategies for mutational analysis of Subtilisin Carlsberg (sC) to achieve the desired thermolability. A three dimensional structure for Subtilisin Carlsberg was already available, which enabled identification and exploration of potential mutational sites. Once a list of potential mutations was agreed with Biocatalysts, the project proceeded with cloning and expression of wild - typesC gene in *E.coli* (4.1).

Concomitantly, Biocatalysts had received a set of potential thermal mutants for LapA (20.3) and Lap2 (20.1), as a result of a separate project effort. After approximately 18 months, these mutant constructs were released to us, for optimisation of soluble expression of these mutants in *E.coli* (section 4.1.2.1). Additionally, wild - type mature LapA (10\_20.3) was recombinantly expressed in *E.coli*, omitting the propeptide in the original construct (section 4.1.2.2), followed by successful *in vitro* refolding studies of *ref*LapA protein from inclusion bodies (section 4.1.2.3). Based on the lack of solubility for recombinant wild - type mature LapA, the construct was amended to include the N-terminal propeptide sequence flanking the mature sequence of LapA (section 4.1.2.4).

Since there was no three-dimensional structure available for LapA, it was decided in parallel to purify native LapA and Lap2 (which were abundantly present) from crude *Aspergillus*

*oryzae* cell extract – Flavourzyme, in order to expedite the crystallisation and structure determination. A crystal structure for native Lap proteins would facilitate a rational mutagenesis approach to achieve thermolability, in a similar manner to that utilised for sC. The entire purification, crystallisation and structure determination process is laid out in Chapter 3. These experiments were carried out along with recombinant expression of sC and LapA proteins in *E.coli*.

These experiments were completed in approximately 2 years from project initiation. The challenging expression of sC in *E.coli* proved that a more suitable expression host is required for this protein. At the same time the determined crystal structure of native LapA from Flavourzyme enabled the design of structure-based, rational mutation library to render this protein thermolabile (section 6.1). Also, the insoluble and difficult to purify recombinant LapA present in the intracellular *E.coli* fraction encouraged us to explore alternative expression hosts for this eukaryotic natively glycosylated protein. The preliminary expression trials in yeast *Pichia pastoris* proved to be successful, yielding relatively pure, properly folded LapA proenzyme in the culture supernatant. As a result, in the following years the project proceeded with studies of heterologous over-production of different forms of LapA in *P. pastoris*.

Optimisation of heterologous expression and purification protocols for LapA is discussed in Chapter 4 (section 4.3). X-ray crystallographic analysis of a LapA precursor, mature and liganded structures enabled us to provide detailed structural insights into processes of LapA maturation, inhibition, substrate binding (Chapter 5). Subsequently, two generations of LapA thermolability mutants were characterised structurally and functionally in an effort to allow complete thermal inactivation of this exopeptidase at moderated temperatures, culminating with evaluation of the best mutants in industrial pasteurisation trials (Chapter 6). At the same time, the acquired knowledge of the role propeptide in the folding and maturation process of LapA enabled exploration of the alternative routes to secretion of the desired mature LapA population in *P. pastoris* supernatant, as opposed to the secretion of inactive proenzyme (Chapter 7). Finally, following an exhaustive analysis of cold-adaptive and thermolabile determinants in psychrophilic proteins, an homologous sequence of a related leucine aminopeptidase from Antarctic fungus *Glaciozyma antarctica* (GAP) was identified. This was followed by an attempt of recombinant expression in *P. pastoris*, however, these initial efforts yielded no distinguishable GAP protein, suggesting further optimisation was needed to obtain this potentially thermolabile and desirable peptidase (Chapter 8).

It is important to note that all the work, commencing from LapA over-expression through to GAP over-expression in *P. pastoris* (Chapters 4 – 8), was completed in less than 1.5 years, highlighting the importance and urgency of the selection of an appropriate expression host for Industrial use. The duration was not affected by significantly increased cultivation times of *P. pastoris* (120h growth for a full expression cycle), which was repeated multiple times throughout this project. Therefore, it cannot be stressed enough that, identification of an appropriate over-expression strategy for these target proteins was clearly a rate-limiting factor in this project and a crucial driving force for the Industrial partner.

The experimental timeline for this 4-year PhD project is summarised in **Table 9.1**. The experimental overview of this project, in its entirety, is graphically presented in **Figure 9.1**.

Experimental workflow	Timeline	Year in the project
Literature review and study of protein thermal properties and bioinformatics of Subtilisin Carlsberg and LapA	10.2013 – 02.2014	1 <sup>st</sup>
Initial cloning and expression attempts in <i>E.coli</i> : <ul style="list-style-type: none"> <li>• Recombinant expression trials with sC</li> <li>• Recombinant expression trials with LapA: set of thermolability mutants, optimisation of soluble expression of wild - type mature sequence; refolding from inclusion bodies; expression of LapA proenzyme</li> </ul>	02.2014 – 06.2015	1 <sup>st</sup> and 2 <sup>nd</sup>
Purification and structural studies of native LapA ( <i>flavLapA</i> ) from Flavourzyme (in parallel to <i>E.coli</i> work)	02.2014 – 12.2015	1 <sup>st</sup> and 2 <sup>nd</sup>
Optimisation of recombinant LapA expression in yeast <i>Pichia pastoris</i> : <ul style="list-style-type: none"> <li>• Structural studies of wild-type proenzyme (<i>wtproLapA</i>) and mature enzyme (<i>wtLapA</i>)</li> </ul>	06.2015 – 01.2016	2 <sup>nd</sup> and 3 <sup>rd</sup>
Structural and functional studies of LapA thermolability: <ul style="list-style-type: none"> <li>• 1<sup>st</sup> round of LapA thermolability mutants M1 –M6</li> <li>• 2<sup>nd</sup> round of LapA thermolability mutants M7 –M13</li> <li>• X-ray structure determination</li> <li>• Thermal inactivation assays</li> <li>• Kinetic and thermodynamic characterisation of mutant proteins</li> </ul>	03.2016 – 01.2017	3 <sup>rd</sup> and 4 <sup>th</sup>
Exploration of alternative routes to LapA mature secretion in <i>P.pastoris</i> supernatant: <ul style="list-style-type: none"> <li>• ‘LapA trans-acting propeptide’ project</li> <li>• ‘Kex2/Ste13 – mediated propeptide cleavage’ project</li> <li>• ‘Mata prepro – mediated secretion’ project</li> </ul>	01.2016 – 03.2017	3 <sup>rd</sup> and 4 <sup>th</sup>
Expression of homologous cold – adapted aminopeptidase from <i>Glaciozyma antarctica</i>	06.2016 – 03.2017	3 <sup>rd</sup> and 4 <sup>th</sup>
Summarising results in thesis format; Publication writing	03.2017 – 09.2017	4 <sup>th</sup>

**Table 9.1** Experimental timeline for the industrial peptidase thermolability 4-year PhD project

### Isolation of native LapA from Flavourzyme

- Optimisation of purification protocol
- 2.71 Å crystal structure of *flavLapA*

### **Chapter 3** Overexpression of LapA in *E.coli*

- Absence of propeptide in the expression construct
- Refolding from inclusion bodies
- Moderately active, partially folded *refLapA* protein

### **Chapter 4** Overexpression of LapA in *P.pastoris*

### **Chapter 5**

- Presence of propeptide in the expression construct
- Extracellular secretion of appropriately folded protein
- Optimisation of purification protocol
- Crystal structures of *wtLapA* (mature) and *wtproLapA* (precursor) enzymes

### Industrial application experiments

- LapA mutant inactivation by pasteurisation
- Performance of LapA mutants in whey hydrolysis
- Pilot scale production of identified LapA mutants

### **Chapter 6**

#### Generation of the 1<sup>st</sup> round of LapA thermolability mutants

- M1 – M6 cloning and expression in *P.pastoris*
- Structural and functional characterisation
- M2 ( $T_{1/2} = 48.61$ ) and M4 ( $T_{1/2} = 74.55$ ) – thermolability at 60 °C
- Increased affinity towards the substrate from reaction kinetic studies

### **Chapter 6**

#### Generation of the 2<sup>nd</sup> round of LapA thermolability mutants

- M7 – M13 cloning and expression in *P.pastoris*
- Structural and functional characterisation
- M7 ( $T_{1/2} = 29.68$ ) and M8 ( $T_{1/2} = 29.55$ ) – thermolability at 60 °C
- Increased affinity towards the substrate from reaction kinetic studies

### **Chapter 7**

#### Alternative routes to mature secretion of LapA

- LapA trans – acting propeptide: not completed
- Kex2/Ste13 – mediated propeptide cleavage: **mature and active secretion**
- MAT $\alpha$  - mediated propeptide cleavage: mature and almost inactive secretion

### **Chapter 8**

#### Recombinant expression of cold – adapted LAP from *G.antarctica*

- Successful cloning and genomic integration in *P.pastoris*
- No secretion of target protein
- Optimisation of expression conditions is required

**Figure 9.1** Experimental synopsis and highlights of the LapA thermolability project.

### 9.3 Isolation of native peptidases from Flavourzyme

At the start of the project it was imperative to obtain the crystal structure of LapA to enable a rational design of thermolability mutant libraries. Considering the challenging over-expression of recombinant LapA constructs in *E.coli*, Flavourzyme (a crude mixture of *Aspergillus oryzae* extracellular enzymes) offered putatively ready access to native LapA and other relevant proteases, in abundance. Therefore, in parallel to targetting heterologous protein expression in *E.coli*, isolation of LapA, Lap2 and other non-specific proteases (AP1, NP1, NP2) was carried out from Flavourzyme. Importantly, an automated purification protocol of all Flavourzyme proteins was reported in the literature (Merz et al., 2015), during the course of this work. However, our purification procedure was quite distinct in that, initially we were only interested in isolation of LapA and Lap2. However, an equivalent protocol to the Merz paper was later utilised to facilitate purification of alkaline protease 1 (AP1) and neutral protease 1 (NP1), both of which subsequently were used in enzymatic processing of LapA precursor proteins (section 4.3.7). The isolation of native LapA (termed *flavLapA*) presented a number of obstacles due to the absence of artificial affinity tags and the highly negatively charged nature of the protein (molecular adhesiveness to other proteins and ultrafiltration membranes). The inability to separate *flavLapA* from TAKA  $\alpha$ -amylase even resulted in obtaining crystals and high quality structural data for the latter (section 3.2.3.1). In addition, repetitive purification cycles on AEX and HIC resins were required to liberate *flavLapA* from two unidentified brown pigments, which have been reported to associate with Lap peptidases in a number of studies (Schalk et al., 1992, Hartley et al., 2009, Aphale and Strohl, 1993, Bennett and Holz, 1997). The characteristic high thermal stability of *flavLapA* permitted the final separation from NP1 protease, but only in the presence of a high ionic strength environment (section 3.2.1). Overnight growth of *flavLapA* crystals yielded a structure solution to 2.7 Å resolution. The refined model of *flavLapA* showed quite distinct molecular packing in the crystal lattice to the later determined structures of recombinant enzymes *wtLapA* and *wtproLapA*. Subsequent detailed structural analysis indicated novel primary unfolding states in which the *flavLapA* molecules were trapped by X-ray crystallography (section 3.2.6). These thermally unfolded transition states appeared likely as a result of thermal heating of the enzyme preparation to 67 °C for 20 minutes as a penultimate purification step. Thus, the observed distinct molecular packing and medium resolution quality X-ray diffraction probably was a consequence of structural non-uniformity of the different *flavLapA* unfolded populations (as shown in **Figure 3.8**). This incidentally obtained, partially unfolded native LapA structure provided important insights to the

initiation of a thermal unfolding event of this otherwise highly thermostable peptidase. The surface solvent exposed regions (corresponding to residues 180 – 190, 215 – 220, 265 – 275) were mostly affected by heating. Importantly, the active site geometry and coordination of the dizinc catalytic center was altered substantially in *flavLapA*, as observed in each of the monomers in the asymmetric unit. No catalytically essential nucleophilic water/hydroxide was found in the active site of *flavLapA*. Despite that, the enzyme retained 67% of the specific activity determined for the folded recombinant protein, raising intriguing questions about the actual requirement for stringent ZN coordination geometry and the role of the nucleophilic water in the postulated reaction mechanism of dimetallic zinc aminopeptidases (Chen et al., 1997). In fact, Zn1, deemed to be involved in the nucleophile delivery in the catalysis, is stripped out in half of the *flavLapA* monomers, suggesting the hydrolysis may proceed via alternative mechanism.

Lastly, isolation of native AP1, NP1 and NP2 proteases enabled the study of processing the LapA zymogen with the inference to native *Aspergillus* and recombinantly secreted protein from *P. pastoris*. We were not able to conclude exactly which (AP1 or NP1) protease is likely to be responsible for processing the LapA precursor. These proteases may work simultaneously and may be able to cleave distinct sites within the LapA propeptide sequence. However, the N-termini of both AP1- and NP1-processed polypeptide displayed identical terminal 'AVT' amino acids, suggesting yet another activity (secreted by native *A. oryzae*) is responsible for complete N-terminal truncation to 'YPD' residues, as observed in the *flavLapA* structure and originally reported by Morita et al. (Matsushita-Morita et al., 2011). NP2 might be responsible for this ultimate modification, however, no attempt to incubate the AP1/NP1-treated LapA population with NP2 was made.

## 9.4 Exploration of suitable routes to recombinant production of active mature target proteases

The initial attempts of sC endopeptidase expression in *E.coli* proved very challenging. After numerous efforts, sC proenzyme was successfully expressed intracellularly and in the periplasmic fraction (section 4.1). However, the majority of sC recombinant enzyme appeared in inclusion bodies and only a small population of soluble mature protein could be detected by Western Blot following 6xHis-tag – mediated affinity purification from 1.6 litre starting culture volume (**Figure 4.5**). Only marginal improvement in solubility was observed following the reduction of post-induction temperature to facilitate the folding process of sC proenzyme. These findings were indicative of a severe autocatalytic degradation, further supported by disappearance of soluble protein band after 16 hours from induction (**Figure 4.5**, lane 6), suggesting the activity and autolysis of recombinant sC in *E.coli* cytoplasm. In addition, over-expression of sC serine protease caused an acute cell toxicity in *E.coli*. This was apparent from descending cell density (OD) readings following overnight incubation of induced *E.coli* culture. Interestingly, the N-terminally His-tagged (sC-pOPINF) construct showed the most severe cell death rates, suggesting the release of even more potent toxic sC population following autocatalytic self-cleavage of 6xHis-prodomain motif. The formation of inclusion bodies and increased cell toxicity is well documented for some subtilisin-like serine proteases (Parry, 2002). Therefore, our experiments only confirmed these predictions. Altogether, we demonstrated that it would be overly complicated to utilise *E.coli* – based recombinant expression system to study the thermal mutant properties of Subtilisin Carlsberg non-specific endoprotease. The only viable solution was to employ a *Bacillus* expression host for extracellular for over-production of this protein, which could offer more native folding and extracellular secretion machinery to this protease.

Heterologous expression of eukaryotic LapA exopeptidase in *E.coli* also presented significant challenges. At the start of the project a set of supposedly thermolabile LapA and Lap2 constructs were received from our industrial partner, Biocatalysts. These constructs were pre-designed by another collaborating party (J. Mullins, University of Swansea) using proprietary bioinformatics methods able to predict the thermostability/lability of target protein. The attempts of over-expression of these constructs employing SoluBL21 solubility-enhancing *E.coli* strain are presented in section 4.1.2.1. None of the tested Lap constructs expressed soluble intracellular protein. Careful analysis of the constructs indicated engineering of extensive alterations in the core regions of secondary structure of these proteins, possibly targeting the resulting polypeptides towards misfolding and subsequent

aggregation. The only partially soluble protein obtained in this experiment was the wild-type N-terminally His-tagged LapA construct (20.3-10), lacking the native N-terminal propeptide. Access to 20.3-10 construct was the only useful outcome from this endeavour, which was followed by optimisation of soluble expression in various *E.coli* strains and media conditions (**Figure 4.7**), culminating in *in vitro* refolding studies (section 4.1.2.3) and contribution towards revealing the role of intramolecular chaperone of LapA propeptide domain (section 5.1). Subsequent cloning and over-expression attempts of LapA constructs harbouring N-terminal propeptide encountered solubility problems under standard induction and growth conditions. This problem was somewhat alleviated by dropping the post-induction temperature to 15 °C (**Figure 4.11**). However, the resultant soluble pro-LapA population was prone to aggregation (evident from higher MW oligomers on Western Blot), suggesting improper folding pattern in *E.coli* cytoplasm. Importantly, one of the contributing factors to misfolding of pro-LapA could be the engineered C-terminal hexahistidine tag, which was later demonstrated to completely abolish secretion of LapA precursor in *P.pastoris* (**Figure 7.5**), suggesting the artificial tag likely interferes with folding mechanism of LapA polypeptide. Isolation of intracellularly produced recombinant LapA precursor would be improbable from *E.coli* cell lysate in the absence of recombinant tag. Therefore, to circumvent folding and solubility problems, series of alternative tags could be tried (GST, MBP, STREP). Nevertheless, this would not guarantee the successful over-production of this eukaryotic protein in *E.coli*, since we later demonstrated the potential requirement of N-linked glycosylation in obtaining fully active conformation of LapA in *P.pastoris* (discussed in section 6.1.3). This post-translational modification pathway is absent in *E.coli*, thus making this prokaryotic host unsuitable for *in vitro* functional studies of LapA fungal exopeptidase.

In summary, due to the challenges encountered with recombinant over-expression of both bacterial endopeptidase – sC and fungal exopeptidase – LapA in *E.coli*, the most time-efficient alternative seemed switching to eukaryotic expression host *Pichia pastoris* for over-production of LapA. In fact, *P.pastoris* was a food grade organism, preliminarily shown to successfully secrete LapA proenzyme into the culture supernatant in the expression trials conducted by Biocatalysts. In addition, *P.pastoris* offered a significantly more homologous secretion and post-translational modification machinery to LapA protein in relation to the native *Aspergillus* host. As a result, further studies and mutant development proceeded in eukaryotic *Pichia pastoris* strain X33.

#### 9.4.1 Secretion and post – translational activation of recombinant LapA precursor

Recombinant proteins can be made either intracellularly or directed into the extracellular environment of *P.pastoris*. This organism secretes almost no endogenous protein, therefore, recombinantly produced protein usually constitutes the vast majority of extracellular protein content. In this way, secretion of heterologous protein (into the medium) acts as a first purification step in the pipeline (Macauley-Patrick et al., 2005).

In this project, the requirement of the prosequence, in the expression construct of LapA, was clearly demonstrated. As a result, LapA in *P.pastoris* was secreted as inactive precursor polypeptide, which required treatment by exogenous protease (AP1 – purified from Flavourzyme). Post – translational activation of recombinant LapA precursor presented significant challenges. It was time – consuming additional step in the protocol of obtaining mature, active and pure LapA protein. Also, enzymatic processing of LapA, by added exogenous protease activity, is a highly undesirable and a cost-ineffective process in industry, which may not be directly applicable to Biocatalysts enzyme production pipeline. Therefore, direct secretion of mature fully functional LapA polypeptide was desirable.

We have demonstrated already that the LapA propeptide acts as an intramolecular chaperone and is essential for the maturation of catalytic domain when expressed in *cis* (section 5.1). From available evidence, in the literature, it was apparently that the LapA propeptide may function also in *trans* (covalently detached from the mature domain). Therefore, recombinant cloning and expression experiment of mature LapA polypeptide and covalently unlinked propeptide in histidine auxotrophic *P.pastoris* GS115 strain was designed (see section 7.2). However, due to problems encountered in the transformation step this experiment could not be completed. Nevertheless, one important discovery was made in this process: supplementation with amino acids was shown to be essential, to ensure optimal GS115 cell growth on minimal MD agar, which likely would facilitate future transformations utilising this *P.pastoris* strain.

A more successful attempt was the Kex2/Ste13 – mediated cleavage of the LapA propeptide (see section 7.3). Generation of several different propeptide processing constructs (**Figure 7.4**) led to mature secretion of the polypeptide, termed <sub>EA</sub>LapA, containing an unprocessed N-terminal ‘EAEA’ dipeptidyl adduct. This exogenous extension only caused a 6-fold reduction in secretion levels, but affected the specific activity of <sub>EA</sub>LapA. Importantly, preliminary LapA thermolability assays showed an increase in thermal inactivation rate, by

this variant of the wild-type protein. Cloning the currently determined best thermolability mutants (M7 and M8) into the  $_{EA}$ LapA expression cassette could yield a mutant with even further enhanced thermal lability properties. Also, the Kex2/Ste13 recognition site (KR/EAEA) could be further optimised, in future work, for more efficient processing.

Compromised post-translational processing and secretion levels of  $_{EA}$ LapA protein have led to subsequent exploration of alternative routes to mature LapA secretion in appropriate folded state and comparable to wild-type quantities. It was recognised that the pre-pro sequence of the  $\alpha$ -mating factor (MAT $\alpha$ ), used to direct recombinant protein to the extracellular environment, could be placed upstream of the mature LapA sequence. The role of mostly hydrophobic propeptide region has not been established yet. It is speculated, that the MAT $\alpha$  prodomain may act as a molecular chaperonin to slow down and to ensure proper folding of secreted protein, because secretion of some heterologous proteins is abolished in the absence of pro-sequence (Chaudhuri et al., 1992, Kjeldsen et al., 1997). In this way, the hydrophobic pro-region of MAT $\alpha$  was predicted to exhibit the analogous function of intramolecular chaperonin by the LapA prodomain. Subsequent structural modelling of the MAT $\alpha$  prodomain identified a striking structural similarity to the LapA prodomain, determined in our studies (**Figure 7.6**). These observations encouraged us to pursue a full recombinant cloning and expression study in *P.pastoris* to verify predictions (section 7.4). However, the secreted LapA polypeptide, with N-terminal 'AVT' extension, exhibited poor levels of secretion and almost negligible specific activity, suggesting that, despite structural resemblance to LapA prodomain, the MAT $\alpha$  propeptide was unable to mirror the essential interactions with the catalytic domain required to induce the rate-limiting transition from molten globule state to native-like conformation.

In summary, our experiments in Chapter 7 demonstrated that the most plausible way of obtaining properly folded, mature LapA enzyme was via engineering the Kex2/Ste13 recognition site, which also may have further implications for LapA thermolability. However, the potential of propeptide production in *trans* should not be underestimated and could be exploited in the future.

## **9.5 LapA proenzyme structure: elucidation of the mechanism of prodomain-mediated inhibition and its role as intramolecular chaperone**

Heterologous secretion of LapA precursor protein using *P.pastoris* expression platform provided access to semi-purified, which could subsequently be readily purified and studied by X-ray crystallography. High resolution crystal structures of LapA proenzyme (*wtproLapA*) and mature protein (*wtLapA*), determined to 1.61 Å and 1.97 Å, respectively, permitted in-depth analysis, elucidation of the inhibition mechanism and determination of interdomain contacts that may be essential to elicit the proposed role of intramolecular chaperone by prodomain (section 5.1). The role of their respective propeptides was extensively studied for the multitude of dimetalloaminopeptidase enzymes (Zhang et al., 2000, Bzymek et al., 2004, Nirasawa et al., 1999). However, no report existed to provide a mechanistic insight on leucine aminopeptidase inhibition by its natural inhibitor – N-terminal propeptide - on a structural level. We also demonstrated that mature domain of LapA is not able to attain its native conformation when produced in the absence of prosegment. This was reinforced by the expression and refolding study of 6xHis-tagged catalytic domain (*refLapA*) from *E.coli* inclusion bodies, revealing the limited active conformation (section 4.1.2.3). For the first time, we are able to show that known synthetic aminopeptidase inhibitors and substrates actually mimic key contacts between pro- and catalytic domains in the herein determined crystal structure of LapA precursor. These findings might aid future design of more effective inhibitors for bimetallic aminopeptidases. To date, the exact mechanism of leucine aminopeptidase inhibition and the role of intramolecular chaperone (IMC) by prodomain have not been elucidated at a structural level. Therefore, with an aid of the first of its kind crystal structure of the leucine aminopeptidase precursor, we revealed structural aspects of mature domain inhibition by the cognate propeptide.

## 9.6 Achieving thermolability

An abundance of relevant mutational studies facilitated the creation of a thermolability mutant library for endopeptidase – Subtilisin Carlsberg (sC) (**Table 4.1**). However, the scarcity of engineering studies towards cold adaptivity/thermolability of leucine aminopeptidases have provided additional challenges to more accurate and rational generation of similar mutant libraries. The only available crystal structure of a cold-adapted aminopeptidase (deposited in the PDB) was from *Colwellia psychrerythraea* strain 34H (PDBid: 3CIA). A cold-active peptidase, termed ColAP, belonging to the family of M1 peptidases, harbours a conserved ‘HEXXHX(18)H’ motif (Bauvois et al., 2008). Synergistic action of multiple structural elements was attributed to cold-adaptation and consequential thermolability of ColAP. Given very low sequence and structural homology, LapA, also a family member of M28 peptidases, could not be directly compared to ColAP. Therefore, any subsequent analysis of the determinants of cold adaptivity and thermolability within the family of M28 peptidases, relied on multiple sequence alignments, using manually inspected sequences of psychrophilic and thermophilic homologues, and homology modelling of the most identical sequences (sections 6.1.1 and 6.2.1). The identification of a genuine psychrophilic homologue – family M28 leucine aminopeptidase from *Glaciozyma Antarctica* (termed GAP) – subsequently led to attempts to recombinantly clone and secrete this protein in *P.pastoris*, discussed in Chapter 8.

Initial feasibility studies of mutational strategies included utilisation of high-throughput approaches to laboratory evolution of protein thermolability: iterative saturation mutagenesis (ISM) in the structurally predetermined rigid ‘hot spots’. However, these approaches were only cost- and time-effective utilising an *E.coli* expression platform, and physically could not be implemented utilising *P.pastoris*, given the laboratory resources available and the time constraints of the project. Despite the fact that a 48-well Biolector microfermentation system was accessible at the facility of our industrial partner (Biocatalysts), the clonal variation of *P.pastoris* transformants impeded screening of an excessive number of variants. Therefore, the implemented mutagenesis strategy involved a highly rationalised, site-directed approach based on accumulated knowledge from the literature and, LapA sequence and structural comparisons to psychrophilic homologues. The highest sequence and structural disparity between LapA and temperature-adapted homologues was found in the variable loop regions located mainly on the protein surface. Therefore, it was assumed that one of the possible routes to temperature adaptation for psychrophilic proteins was via optimisation of residue content in flexible loop regions and

modulation of interactions to solvent by the surface residues. This observation coincided with evidence in the literature regarding cold adaptation and thermolability of psychrophilic proteins, and in turn guided our further mutagenesis strategy, whereby the protein core was left intact and any residue substitutions were targeted to the loop regions and ends of secondary structure elements of lower conservation.

Once the recombinant LapA expression in *P.pastoris* was optimised (Chapter 4, section 4.3), two rounds of thermolability mutant generation were performed. Structural and functional characterisation of these mutants is presented in Chapter 6. The most significant differences in thermolability were determined following incubation in the 50 – 60 °C range. All of the characterised mutants displayed very poor thermostability at 70 °C, resulting in almost complete inactivation (following incubation for 15 minutes), contrary to the control wild-type LapA retaining almost 100% of initial specific activity under equivalent conditions. The specific activity was only adversely affected in 5 of the 13 thermolability mutants studied. The remaining mutants maintained comparable to wild-type levels of specific activity, with added thermolabile behaviour. Based on an analysis of the thermal inactivation data, by non-linear regression, and obtained kinetic parameters of the inactivation process of each M1 – M13 mutant, the most thermolabile mutants were M7 (M2/M4 double mutant) from the 1<sup>st</sup> round and M8 (single point mutant) from the 2<sup>nd</sup> round of mutagenesis, displaying almost equivalent half-lives of 29.68 and 29.55 minutes at 60 °C, respectively.

Although thermolability mutations were selected in an effort to destabilise the overall LapA fold, to increase molecular flexibility and to disrupt LapA molecular edifice, most of the single point mutations did not induce severe effects on LapA structural integrity, as studied by X-ray crystallography (section 6.4). However, some of the site-directed mutations (M1, M5) were quite disruptive and detrimental to the overall fold, as deduced from adversely affected expression profiles in *P.pastoris*. Other mutants (M3, M6, M11) only compromised the specific activity and showed no significant differences in secretion levels, indicative of an unaffected overall structure. Therefore, carefully chosen single site substitution is unlikely to force such a stable and compact polypeptide fold as LapA to fall apart, however, a combination of two residue substitutions (M7) could have an additive, more pronounced disruptive effect on fold, as deduced from the observed 2 – 3-fold reduction in secretion levels.

Structural factors leading to the altered thermal properties of the most heat labile LapA mutants include:

M7: in this double mutant the synergistic destabilising affect of single Q91A (M4) and P304A (M2) point mutations expressed in the most thermolabile behaviour of all mutants. The additive effect of elimination of the salt bridge (M4) and improved local flexibility by proline substitution (M2) led to increased plasticity within the secondary structure elements. It was reported that flexibility between secondary structure elements is achieved by reducing the number of proline residues in loops joining these elements (Sakaguchi et al., 2007).

M8: one of the possible explanations to the exerted thermolability effect of D271T mutation could be the elimination of a stabilizing salt bridge with K267 and substitution with threonine increases the number of surface exposed nonpolar residues, which is an entropy-driven destabilising factor (Aghajari et al., 1998, Russell et al., 1998, Georlette et al., 2003).

All the mutations, except M12 (D338S), had a destabilising effect and elicited a desired reduction in thermostability. This substitution displayed overall increased structural rigidity in M12 molecule (**Figure 6.11**), indicative of observed 3.5-fold improvement of thermostability at 60 °C. The exact mechanism of this rigidifying mutation could not be deduced, however, a serine at this position is found in the homology model of psychrophilic GAP protein. Thus, amino acid context must play a role in this thermostability mechanism of M12 and a less bulky sidechain of S338 was not a determining factor favouring thermolability in this mutant. Also, D338 residue is a part of hydrophobic substrate binding pocket in LapA. Y251 is found in AAP enzyme at this position. D338 was implicated in the ability to hydrolyse positively charged N-terminal amino acids by recombinant (rLap1) from *A.sojae* (Huang et al., 2015). Therefore, one could speculate that, in LAP proteins, residue substitutions at this position is not only responsible for substrate specificity, but also is subtly involved in determining the thermostability.

A comparison of the most successful mutants by thermolability (M2, M7, M8 and M10) showed a very similar specific activity to the wild-type enzyme, under standard reaction conditions, indicating that the chosen mutations had not altered the ability of LapA to hydrolyse the substrate (**Table 6.21**). M8 arguably displayed elevated levels of activity at 37 °C, which may be linked to improved local flexibility in the vicinity of the active site pocket, rendering this mutant even more attractive for industrial application. M7 was superior to other mutants, according to the industrially relevant  $T_{60/30}$  estimate (**Figure 6.23**), showing the most thermolabile character following incubation at 60 °C for 30 minutes. The remarkable stability of the LapA hydrolase fold was further demonstrated by incubation of WT and mutant variants at the room temperature (**Table 6.23**). Extrapolated half-lives of

M7 and M8 mutants exceeded 200 h (**Table 6.24**), where the  $t_{1/2}$  of WT protein was almost twice that, suggesting a negative impact on fold integrity by these destabilising mutations. However, the projected shelf-life of most of the thermolability mutants could still be extraordinary if one considers adding protein stabilising aids to liquid formulations or spray drying the enzyme.

Selected LapA thermolability mutants also were evaluated for low temperature activity (cold adaptivity) (section 6.5.5). Specific activity measurements indicated no improvement in hydrolytic ability by M2, M4, M7 and M10 mutants, following incubation at 4, 10 and 20 °C. However, determination of kinetic catalytic parameters ( $k_{cat}$  and  $K_m$ ) would probably be more appropriate in this instance. The emphasis throughout this thesis work was toward obtaining a LapA mutant that could be inactivated at significantly lower temperature (i.e more thermolabile). Thus, the chosen mutants were not kinetically characterised to determine substrate turnover rates at low temperatures. Also, distant-to-active-site mutations (M7/M8/M9/M10) were not predicted to particularly effect low temperature activity. It is more likely that the substitutions targeted to the proximity of the active site (M11/M12/M13) would show more cold adaptive behaviour. However, this requires further experimental validation.

In the final experimental stages of this work, our industrial partner Biocatalysts suggested utilisation of an industrial pasteurisation process as a method of inactivating LapA mutant proteins (section 6.5.6). Importantly, this approach would resemble the industrial procedure of protease thermal inactivation, following hydrolysis reaction. A standard pasteurising process used in the food industry is 72 °C exposure for <1 minute. As a result of a short exposure time to heat, the quality of reaction products can be preserved and, in this case, gelation effects of whey protein hydrolysates are less likely to be induced. However, the previously determined most thermolabile LapA mutants (M7 and M8) still exhibited 28% and 37% residual activity, respectively, following pasteurisation at 72 °C for 30 seconds (**Table 6.27**). Longer (~1 min.) pasteurisation times may be required to fully inactivate these mutants at this temperature. More importantly, this experiment highlighted the ability of LapA to rapidly regain an active conformation, following a short-duration heat exposure and incubation at room temperature. Therefore, we speculate that very short, high temperature exposure periods are not sufficient to completely disrupt the highly compact and thermostable hydrolase fold. The temporarily induced inactive conformation of LapA might involve unravelling surface sub-structures, loop regions, stripping off catalytic dinuclear zinc, but it fails to completely abolish the stable hydrophobic core. Significantly longer

exposure times or increased temperature (kinetic energy) are likely required to break these enthalpy-driven intramolecular interactions.

Thermodynamic measurements of wild-type and mutant LapA enzymes also indicated significant changes in melting temperature ( $T_m$ ) and enthalpic-entropic parameters (section 6.7). All the thermolability mutants tested displayed a varying degree of reduction in  $T_m$  in both DSF and DSC experiments. A slight variation in determined transition midpoint values (seen in both methods) could be a result of the differences in final sample concentration or the rate of heating. It is important to note that, although the difference in transition midpoint values between *wt*LapA and thermolability mutants never exceeded more than 6 °C in the DSF experiment, the most thermolabile mutants (M7 and M8) exhibited a 3-fold reduction in half-life determined in the thermal inactivation assays (**Table 6.20**). This observation suggests that the LapA thermal inactivation process can be uncoupled from a thermal unfolding event. Therefore, an elevated temperature may primarily induce substantial alterations to the active site structure (zinc coordination geometry) even before thermal unfolding affects other structural regions of the protein. This notion would explain the observed higher thermal lability of LapA at 60 °C based on activity assays, where no sign of unfolding can be detected in either DSF or DSC thermograms.

## **9.7 Future structure-to-function evolution of LapA thermolability and cold adaptivity**

It is well established in the literature that different packing of external surface residues play a major role in thermal adaptation of proteins from a multitude of functional families (Glyakina et al., 2007). Recently, more generally accepted concept of protein thermal stability is not only based on the number of ion pairs (single short-range salt bridges), but also on long-range interactions with charged or polar residues, referred to as electrostatic context (Pezzullo et al., 2013). The ability of cold adapted proteins to operate in a low kinetic energy environment emerged from improved backbone flexibility and subsequent trade-off in thermal stability. Any effort to mimic the properties of cold-active enzymes should focus on mainly increasing the overall flexibility of the molecular edifice and improving interactions with solvent. Therefore, one can utilise the enthalpic destabilisation effect to achieve the desired thermolability in target enzymes and entropic compensation (flexibility effect) to ensure conformational mobility, leading to improved catalytic efficiency at low temperatures (Arnold et al., 2001).

The chosen mutagenesis strategy for LapA, targeting residues situated within distant-to-active-site surface loops and small helical regions, was reasonably successful. Surveying the crystal structures of the LapA thermolability mutants, the protein backbone movements were mainly observed in the loop regions, clustered on one side of the LapA molecule (**Figure 6.10, A**). Several reports indicated that the thermal unfolding process might be initiated from the disordered surface loop (Deng et al., 2014, Anbar et al., 2012). Thus, the successfully chosen destabilizing point mutations on the surface of LapA could expedite the thermal unfolding process. This was further evidenced by an analysis of the *flav*LapA crystal structure, which supposedly represents primary unfolding states of LapA protein, exhibiting structural alterations in surface loops and helical regions, following the implemented heating step (67 °C, 20 minutes) in the purification protocol (section 3.2.6).

However, to further engineer LapA to enable it to operate at significantly lower temperatures (cold adaptivity), one needs to consider targeting hydrophobic clusters within the very compact hydrolase fold, in order to confer core plasticity to LapA. This could be achieved by reducing the number of intramolecular interactions by substitution of hydrophobic bulky amino acids (Leu, Trp, Ile), in a similar fashion to the reported study of cold-adaptation of *B.subtilis* lipase (LipJ), which was altered to a significantly lower  $T_{opt}$  of 20 °C (from 37 °C) by a single point mutation (I137M) targeted to the vicinity of the substrate binding pocket (Goomber et al., 2016). Such mutations, aimed at the active site, usually have the purpose of lowering the activation energy by providing the required local plasticity to operate at cold temperatures. Also, the required core flexibility may be attained via engineering of cavities in a hydrophobic cluster (removal of bulky hydrophobic sidechains). The engineering of low crystallographic B-factor ('rigid hot-spot') residues is another viable strategy, successfully implemented in thermolability and cold adaptivity evolution of bacterial lipase, containing an analogous stable hydrolase fold (Reetz et al., 2009). Not surprisingly, the lowest temperature factors were observed for residues forming the conserved  $\beta$ -strand regions in the crystal structure of LapA. In this respect, our chosen mutagenesis strategy was more predictable, since a drastic substitution of the key residue involved in stabilisation of the hydrophobic core might lead to unforeseeable and detrimental disruptive effects within the conserved hydrolase fold.

Lastly, towards the end of this project an established collaboration with a research group from Bath University led to a rigidity analysis of our LapA crystal structure, using a proprietary algorithm that enabled identification of specific destabilising mutations (N114M, N181W, T170P, E226P). All the proposed mutations reside within the hydrophobic core and

were predicted to increase the overall flexibility of the LapA macromolecule without directly affecting the active site residues. Experimental validation of these mutants, in future work, might result in further and/or additional cold adaptivity and thermolability of LapA.

## **9.8 Cold adaptation of a psychrophilic homologue – GAP aminopeptidase**

To date, there is only a single cold-adapted aminopeptidase crystal structure deposited on Protein Data Bank (Bauvois et al., 2008), which indicates a large paucity of such structural data and limits our understanding of temperature adaptation of this important class of metalloaminopeptidases. Therefore, an effort to overexpress the naturally, cold-adapted GAP aminopeptidase sequence in *P.pastoris* was made. GAP represents a close sequence (40% identity) and structural (RMSD 0.865) homologue of LapA, making it an ideal target for structural studies of cold adaptivity and likely thermal lability for M28 class of peptidases, which is not reported in the literature to date. A successful over-expression and subsequent structure – function characterisation of GAP aminopeptidase would help deduce natural molecular adaptations of this enzyme to operate at low temperatures. Also, structure solution of GAP would permit an in-depth understanding of its likely thermal lability. In turn, this knowledge could guide future engineering of LapA for enhanced thermolability. It is plausible that, identification of subtle water networks and a single residue substitution, in the vicinity of the active site, could lead to heat sensitivity of LapA, as demonstrated in the comparative structural study of psychrophilic  $\beta$  – glucosidase, where a loop insertion and unique binding pocket His299 residue were shown to confer thermolability in a cooperative fashion (Miao et al., 2016). However, to gain an equivalent level of detailed insight, a crystal structure of GAP is required.

One of the plausible explanations as to why GAP secretion was impeded in *P.pastoris* could be the cultivation temperature (28 °C), which was not compatible with an optimal folding temperature of such a naturally, cold-adapted protein. This is a commonly encountered obstacle in the heterologous production of temperature sensitive proteins in conventional expression systems, such as *Pichia pastoris* (Bjerga et al., 2016). Given the close evolutionary relationship to LapA, the GAP sequence also should be readily secreted in *P.pastoris*. In addition, prior to the synthesis of GAP expression cassette, the sequence was analysed using a proprietary secretion prediction algorithm (DFM), developed by

Biocatalysts, which identified a high likelihood of secretion of the GAP sequence, with a high confidence score.

Given these considerations, optimisation of the over-expression of GAP aminopeptidase in *P.pastoris* should be a priority target, in the future, since this could provide direct access to a cold-active, potentially thermolabile leucine aminopeptidase, exhibiting analogous substrate specificity to LapA.

## 9.9 Concluding remarks

This project has provided scope for unconventional engineering of cold adaptivity and thermolability of commercially relevant peptidases – Subtilisin Carlsberg (sC) and Leucine aminopeptidase A (LapA). Innate cytotoxic effects of sC recombinant protein have prevented further moderation of its thermoproperties. However, successive heterologous expression of LapA in *E.coli* and *P.pastoris* provided an immense amount of structural and functional data, which has contributed significantly to elucidation of the complex mechanisms of; (i) recombinant secretion, (ii) folding, and (iii) thermal lability of this fungal aminopeptidase.

The initial aims and objectives, identified at the start of this project, have been achieved. Development of an engineered enzyme, with reduced thermal stability, without compromising catalytic efficiency, initially was successful due to the rational design of LapA variants with information obtained from extensive bioinformatics and literature analysis. Additionally, structure elucidation of native *flav*LapA peptidase provided a more accurate prediction of the effect of individual destabilising mutations on the precise 3D fold. The characterised LapA thermolability mutants showed a significant reduction in thermostability at 60 °C, which delivered a proof of concept of the chosen mutagenesis strategy. Continued experimental work, along these lines, will contribute to a deeper understanding of the precise structural determinants that govern the thermal properties of these enzymes exhibiting a stable hydrolase fold.

To date, there are no studies reporting thermal engineering of dimetallic aminopeptidases. In fact, the deliberate engineering of thermolability in any type of enzyme is largely an unexplored area, with only a single such study, unsurprisingly concerning an industrial hydrolase (Reetz et al., 2009).

Scientific findings from this project should lead to three manuscripts in peer-reviewed journals: (i) reporting the isolation and structure solution of native LapA from Flavourzyme (Chapter 3), (ii) studying the inhibitory and intramolecular chaperone role of LapA propeptide (Chapter 5), and (iii) discussing the structural and functional studies of LapA thermolability (Chapter 6).

## Appendices

---

# Appendix I

## Protein View: gi|169782566

leucine aminopeptidase 1 [Aspergillus oryzae RIB40]

Database: NCBIInr  
Score: 275  
Monoisotopic mass (M<sub>r</sub>): 41308  
Calculated pI: 5.03  
Taxonomy: [Aspergillus oryzae RIB40](#)

This protein sequence matches the following other entries:

- [gi|121798407](#) from [Aspergillus oryzae RIB40](#)
- [gi|183774489](#) from [Aspergillus oryzae RIB40](#)
- [gi|1391865176](#) from [Aspergillus oryzae 3.042](#)
- [gi|1635505916](#) from [Aspergillus oryzae 100-8](#)

Sequence similarity is available as [an NCBI BLAST search of gi|169782566 against nr](#).

### Search parameters

MS data file: C:\Data from 4800\Nov14\07Nov14\Ged Lap1 rec well e1.mgf  
Enzyme: Trypsin: cuts C-term side of KR unless next residue is P.  
Fixed modifications: [Carbamidomethyl \(C\)](#)  
Variable modifications: [Deamidated \(NQ\)](#), [Oxidation \(M\)](#)

### Protein sequence coverage: 14%

Matched peptides shown in **bold red**.

```
1 MRFLPCIATL AATASALAIG DHVRSDDQYV LELAPGQTKV VTEAEKWALR
51 AEGKRFFDIT ERASSLELAS NKKQKLAVTY PDSVQHNETV QNLIKSLDKK
101 NFETVLQPFSEFHNRYYKSD NGKKSSEWLQ GRIQEIIISAS GAKGVITVEPF
151 KHSFPQSSLI AKIPGKSDKT IVLGAHQDSINLDSPSSEGRA PGADDDGSGV
201 VTILEAPFRVL LTDEKVAAGE APNTVEFHFY AGEEGGLLGS QDIFEQYSQK
251 SRDVKAMLQQ DMTGYTKGTT DAGKPEISIGI ITDNVDENLT KFLKVIVDAY
301 CTIPTVDSKC GYGCSDHASA TRYGYPAFA FESAFGDDSP YIHSADDTIE
351 TVNFDHVLQH GRLTLGFAYE LAFADSL
```

Unformatted sequence string: [377 residues](#) (for pasting into other applications).

Sort by ☒ residue number ☐ increasing mass ☐ decreasing mass  
Show ☒ matched peptides only ☐ predicted peptides also

Query	Start – End	Observed	Mr (expt)	Mr (calc)	Delta M	Score	Expect	Rank	U	Peptide
<a href="#">44</a>	101 – 115	1864.9165	1863.9092	1863.9268	-0.0176	0 101		1		K.NFETVLQPFSEFHNRY + Deamidated (NQ); [-0.9476 at E3]
<a href="#">57</a>	170 – 189	2109.0837	2108.0764	2108.0498	0.0266	0 100	1.7e-05	1	U	K.TIVLGAHQDSINLDSFSEGR.A
<a href="#">48</a>	190 – 208	1889.9426	1888.9353	1888.9167	0.0186	0 99	2.4e-05	1	U	R.APGADDDGSGVVVTILEAFR.V

Figure 10.1 MS identification of native LapA enzyme in Flavourzyme

## Protein View: gi|169770151

### leucine aminopeptidase 2 [Aspergillus oryzae RIB40]

**Database:** NCBItr  
**Score:** 95  
**Monoisotopic mass (M<sub>r</sub>):** 53694  
**Calculated pI:** 4.94  
**Taxonomy:** [Aspergillus oryzae RIB40](#)

This protein sequence matches the following other entries:

- [gi|238487472](#) from [Aspergillus flavus NRRL3357](#)
- [gi|121804190](#) from [Aspergillus oryzae RIB40](#)
- [gi|83767404](#) from [Aspergillus oryzae RIB40](#)
- [gi|220699853](#) from [Aspergillus flavus NRRL3357](#)
- [gi|391864054](#) from [Aspergillus oryzae 3.042](#)
- [gi|635506109](#) from [Aspergillus oryzae 100-8](#)

Sequence similarity is available as [an NCBI BLAST search of gi|169770151 against nr](#).

### Search parameters

**MS data file:** C:\Data from 4800\July14\15July14\Ged 20\_1.mgf  
**Enzyme:** Trypsin: cuts C-term side of KR unless next residue is P.  
**Fixed modifications:** [Carbamidomethyl \(C\)](#)  
**Variable modifications:** [Deamidated \(NQ\)](#), [Oxidation \(M\)](#)

### Protein sequence coverage: 3%

Matched peptides shown in **bold red**.

```
1 MRSLLWASLL SGVLAGRALV SPDEFPEDIQ LEDLLEGSQQ LEDFAYAYPE
51 RNRVFGGKAH DDTVNYLYEE LKKTGYDVY KQPQVHLWSN ADQTLKVGDE
101 EIEAKTMTYS PSVEVTADVA VVKNLGCSEA DYPSDVEGKV ALIKRGECPF
151 GDKSVLAAGA KAAASIVYNN VAGSMAGTLG AAQSDKGPYS AIVGISLEDG
201 QKLIKLAEEG SVSVDLWVDS KQENRTTYNV VAQTKGGDPN NVVALGGHTD
251 SVEAGPGIND DSGGIISNLV IAKALTQYSV KNAVRFLFWT AEEFGLLGSN
301 YYVSHLNATE LNKIRLYLNF DMIASPNYAL MIYDGDGSAF NQSGPAGSAQ
351 IEKLFEDYYD SIDLPHIPTQ FDGRSDYEAF ILNGIPSGGL FTGAEGIMSE
401 ENASRWGGQA GVAYDANYHA AGDNMTNLNH EAFLINSKAT AFAVATYAND
451 LSSIPKRNTT SSLHRRARTM RPFGRAPKT HAHVSGSGCW HSQVEA
```

Unformatted sequence string: [496 residues](#) (for pasting into other applications).

Sort by ☒ residue number ☐ increasing mass ☐ decreasing mass  
Show ☒ matched peptides only ☐ predicted peptides also

Query	Start – End	Observed	Mr(expt)	Mr(calc)	Delta M	Score	Expect	Rank	U	Peptide
<a href="#">20</a>	59 – 73	1837.9272	1836.9199	1836.8894	0.0306	1 95	4.8e-05	1	U	K.AHDDTVNYLYEELKK.T

Figure 10.2 MS identification of native Lap2 enzyme in Flavourzyme

## Protein View: gi|169782566

### leucine aminopeptidase 1 [Aspergillus oryzae RIB40]

Database: NCBI  
Score: 275  
Monoisotopic mass (M<sub>r</sub>): 41308  
Calculated pI: 5.03  
Taxonomy: [Aspergillus oryzae RIB40](#)

This protein sequence matches the following other entries:

- [gi|1121798407](#) from [Aspergillus oryzae RIB40](#)
- [gi|183774483](#) from [Aspergillus oryzae RIB40](#)
- [gi|1391865176](#) from [Aspergillus oryzae 3.042](#)
- [gi|1635505916](#) from [Aspergillus oryzae 100-8](#)

Sequence similarity is available as [an NCBI BLAST search of gi|169782566 against nr](#).

### Search parameters

MS data file: C:\Data from 4800\Nov14\07Nov14\Ged Lap1 rec well e1.mgf  
Enzyme: Trypsin: cuts C-term side of KR unless next residue is P.  
Fixed modifications: [Carbamidomethyl \(C\)](#)  
Variable modifications: [Deamidated \(NQ\)](#), [Oxidation \(M\)](#)

### Protein sequence coverage: 14%

Matched peptides shown in **bold red**.

```

1 MRFLPCIATL AATASALAIG DHVRSDDQYV LELAPGQTKV VTEAEKWALR
51 AEGKRFFDIT ERASSLELAS NKKQKLAVTY PDSVQHNETV QNLIKSLDKK
101 NFETVLQPPS EFHNRYYKSD NGKKSSEWLQ GKIQEIISSA GAKGVTVEPF
151 KHSFPQSSLI AKIPGKSDKT IVLGAHQDSI NLDSPSEGRA PGADDDGSGV
201 VTILEAFRVL LTDEKVAAGE APNTVEFHFY AGEEGGLGS QDIFEQYSQK
251 SRDVKAMLQQ DMTGYTKGTT DAGKPESIGI ITDNVDENLT KFLKVIVDAY
301 CTIPTVDKSC GYGCSDHASA TRYGYPAFA FESAFGDDSP YIHSADDTIE
351 TVNFDHVLQH GRITLGFAYE LAFADSL

```

Unformatted sequence string: [377 residues](#) (for pasting into other applications).

Sort by ☒ residue number ☐ increasing mass ☐ decreasing mass

Show ☒ matched peptides only ☐ predicted peptides also

Query	Start - End	Observed	Mr (expt)	Mr (calc)	Delta M	Score	Expect	Rank	U	Peptide
<a href="#">44</a>	101 - 115	1864.9165	1863.9092	1863.9268	-0.0176	0	101	1	1	K.NFETVLQPPSEPHNR.Y + Deamidated (NQ); [-0.9476 at E3]
<a href="#">57</a>	170 - 189	2109.0837	2108.0764	2108.0498	0.0266	0	100	1	U	K.TIVLGAHQDSINLDSPSEGR.A
<a href="#">48</a>	190 - 208	1889.9426	1888.9353	1888.9167	0.0186	0	99	2	U	R.APGADDDGSGVVVTILEAFR.V

**Figure 10.3** MS identification of native TAKA  $\alpha$ -amylase enzyme in Flavourzyme

## Protein sequence report

Alta Bioscience code: S6791

Customer sample code: EE IN

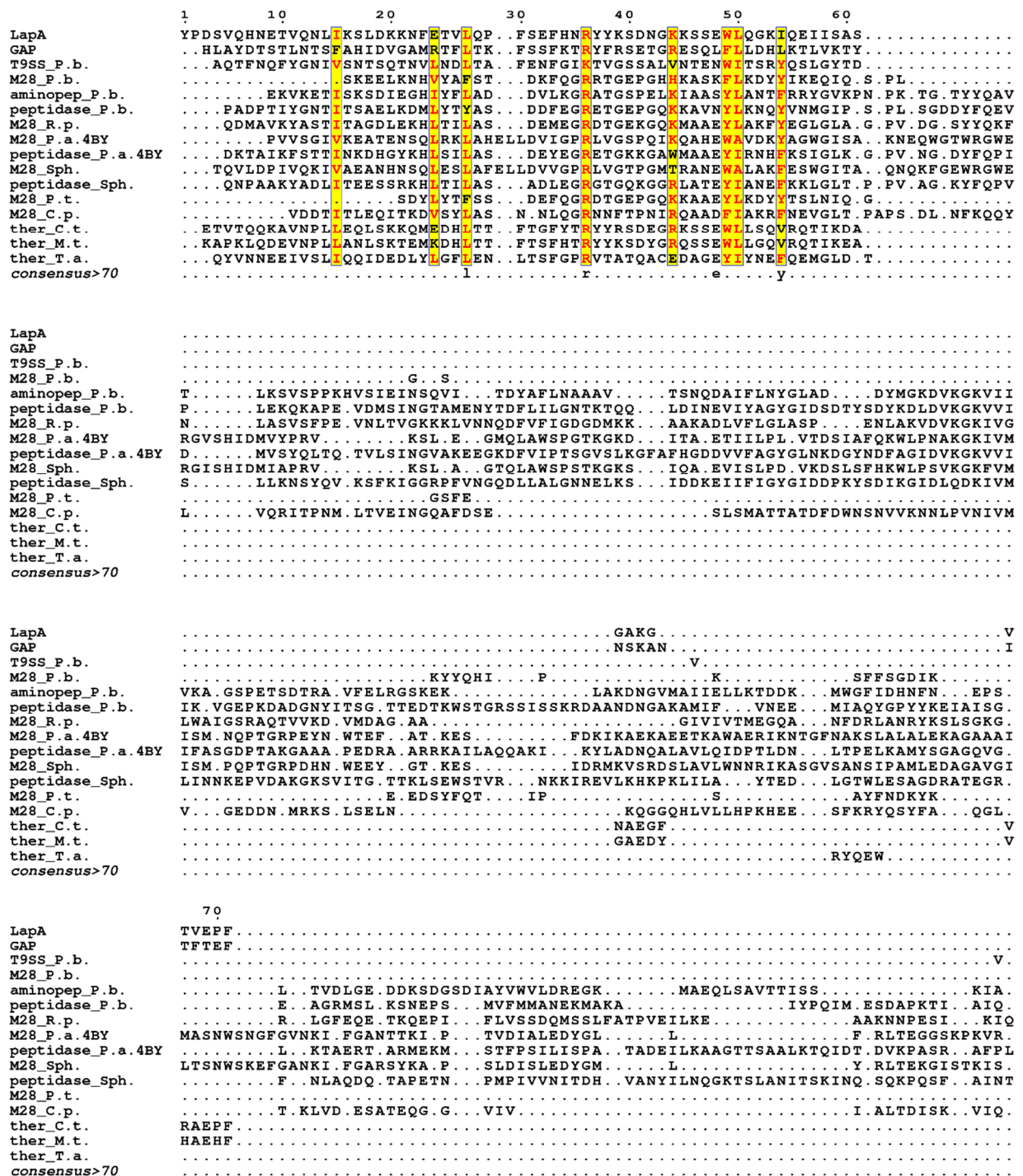
### N terminus

Residue			
1	<b>E</b>	<b>X</b>	
2	<b>A</b>		
3	<b>E</b>		
4	<b>A</b>		
5	<b>Y</b>		
6			

**Figure 10.4** N-terminal sequencing result of <sup>EA</sup>LapA protein. Sequencing performed by Alta Biosciences, UK.

## Appendix II

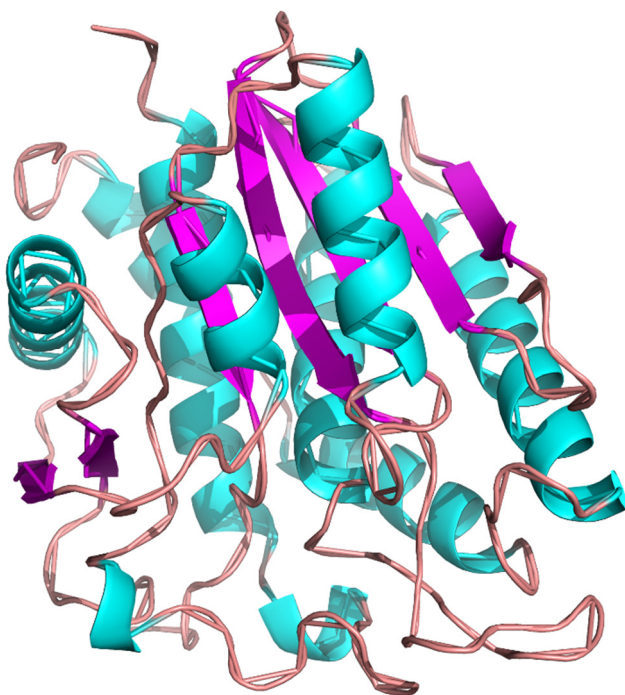
LapA	.....KHSFPQSSIIA..KIPG.....KSDKTIVIGAHQDSINLSDSP.....S.EGRAPGADDDG
GAP	.....THSWIQKSIILRFELPESKTA..DPSKVVIIIGAHQDSVNQL.....P.FLPAPGADDDG
T9SS_P.b.	..L...QPFTYSAGTSNNIIV..TKTG..T.L..YPNTFLIIIDAHYDTIN.....GP.GTNDNG
M28_P.b.	.....NSQNVIA..YVKG..N.E..FPPEVLIISAHSDHEGY.....T...D.TEINYNGADDDG
aminopep_P.b.	..I...GDKQEEAVISQNVIG..VVEG..T.DETLKNEYIISAHYDHVGIGEP....D.E.T..G.DTIYNGARDNA
peptidase_P.b.	..ISCAIKSQAEAVSSENVVA..FLKG..T.E..KPDEILVISAHLDEHGI.....K...D.GQVYNGADDDG
M28_R.p.	..KASFLVKKSKKLVPTENVMA..YLEG..T.D..KKEEVLVISAHYDHTGI.....D.S..E.GRVNNGADDDG
M28_P.a.4BY	..IK.AESKELGAVPTFNTIAS..IKG..SE...KPNEYVLIISAHFDSDWG.....GTGATDNG
peptidase_P.a.4BY	..TVSGTAVKEEVKIRAEVNLG..FLEG..S.DPVLKNEVLVITGHYDHIGL.....T.LSGE.DKVNNGADDDG
M28_Sph.	..VE.TASKDLGNVPSFNTIAE..IKG..SE...FPNEYVLIISAHLDSWEG.....GS.GATDNG
peptidase_Sph.	..SFNAQLGMVAEQFTDPNVLG..LEEG..S.D..LKDEILVIGGHYDHDGK.....S.S..N.GTIFFGADDDNA
M28_P.t.	.....ASENVIS..IKG..S.V..LPPEYLVITSHYDHVGM.....K...D.GEINYNGADDDG
M28_C.p.	..FNVKGNNAKVSTSALSNNVG..ILPG..K.S..KANEIVLISAHYDHLGVKPSIDGAPNSNAQ..QSSDIFNGADDDA
ther_C.t.	.....QHSWGNSTIA..TPPG.....QTNSTIVIGAHQDSINLFFF.....S.FLAAPGADDDG
ther_M.t.	.....KHPWGOHSIIA..TIPG.....KTNSTVVIIGAHQDSINLWLP.....S.ILAAPGADDDG
ther_T.a.	.....SNGNLYGNVEA..SLPG..I.NE.TSEIYIICAHYDSVYG.....SPGADDDG
consensus>70	.....nvi.....g.....n...v..aH.D.....GadD#g
LapA	SGVVTILEAFRVLITDEKVAAGEAPNVEFHFYAGEEGGLCSQDIFEQY.SQKS.RDVKAMQQD.MTGYSKGTTD
GAP	SGTTLTILITLTSLLRESFVP...SVHALEHFYSAGEEGCLSGDISRAY.VSEG.KIVRGMMHMD.VTAYKEGT
T9SS_P.b.	TGVVLLELARLLKNVE...T.EKSIKFHFSGEEDGVVGSNYVNN.TVIPQNLDIRLVFNID.EVGGVNG...
M28_P.b.	SGTAAVMEMAQGFKKAALEGHGP.KRSILVHLTGEEVGRGSRYSYAKY.PVFSM.KNTVANNID.MIGRDDAHK
aminopep_P.b.	VGTTTILSMAENLAKYP...T.KRSALFILFTGEEKGLGSSYVEN.PVPL.NQMVCYFNID.NAGYNDT...
peptidase_P.b.	SGTVAILEIAEAFKAAADKGDGP.KRSILFLHVTGEEKGLGSRHYSNDNPPIFPL.KNTVADNID.MIGRTPKRE
M28_R.p.	SGTIVSIMEIAEAFAMAAKDGIKP.RRSILFLNVTGEEKGLGSEYSDH.PIFPL.ENTVNNINID.MVGRDIEYQ
M28_P.a.4BY	TGTVIMMEAMRILKSVYPN...PKRTILVGHWSSEEGQLNGSRAFVEDH.PEVV.NNLQALFNQDNGTGRVNIISG
peptidase_P.a.4BY	SGTIGVMIADAFKAKKAGKGP.KRSILFMTVVGECKGLGSEWYSEH.PVFPL.NKTIANNID.MIGRDKHEH
M28_Sph.	TGTVIAMEVARILKKILPN...PKRTILGLWGSEEGQLNGSRFVLDN.PEII.ENTQAVFNIDNGTGRVERING
peptidase_Sph.	SGTAVILELAKAFSQAAGEKGP.RRSILFLHVTGEEKGLGSKYITEN.PVFPL.ENTVTCINID.MIGRIDDKHL
M28_P.t.	SGTVAIMEIAEAFQKAVSEGYRP.KRSILFLHVTGEEKGLGSKYISEN.PIYPL.ENTVANNID.MIGRDEAHQ
M28_C.p.	SGVSAIINLANHFAKLGN...EKTILMAAFSAEEIGGFGSKHFTSTN..LEPT..TITAMINIE.MVGRDIAVFGD
ther_C.t.	SGVVTILEAFRVLQSQDVLQGAQNTIEFHWYSAGEEGGLCSQATFSAY.EKQG.RDVKAMQQD.MTGYSKRTLD
ther_M.t.	SGVVTILEAFRVLQSQDVLQGNHQNTHIEFHWYSAGEEGGLCSQATFSAY.EKEG.RDVKAMQQD.MTGYSKRTLD
ther_T.a.	SGVAAVLSAAKVMSSQSFN...HTIREVTFSGEEGGLYGSFYVEE.AYINN.ENIIALNAD.MIGFAENEDD
consensus>70	sgt..ile.....i.f.....EE.GI.GS..y.....n...n.#.....g.....
LapA	AG..KPESIGIITD.....NDENITKFLKVIYDAYC.TIPT.VDSKCGYGCSDHASATKYGYAAFAF...
GAP	T...PVIGLIND.....GVQQQLVVFMRKILITEYA.EIPIY.ADTRCGYGCSDHSWTKVGAFASSCIA...
T9SS_P.b.	..M..TNDTVVCERDL..TSPOQNNAAASNAATNSLAILIEL..YSNL.ETEISNAYASDYVPFENNGETITGLY...
M28_P.b.	NN..PNYIYILGADR..LSTQLHYISEAANAQFTNLELDYK.LNSD.KDPNRYFRSDYINFAQEGGVVIFYF...
aminopep_P.b.	..S..LIS..IIGLRTTAENNIFSAETFGAIKAI.....EDP.AKEQGLDRSDYNNFAKKGIAPTFSLGFT
peptidase_P.b.	GD..RNVYVILGSDK..LSTDLHNISEENANKYTNILDYT.YNDE.NDPNRFYRSDYINFAKNNIIVIFYF...
M28_R.p.	DAENKDYVYVIGSEM..LSSQLKTINEYNNITHTNILDYR.YDAE.DDPNRFYRSDYINFAKNNIIVIFYF...
M28_P.a.4BY	QG..FAKSETYILTR...WLNRR...IPEKTRAEIK.T.SF..PGSPGAGGSDTASFVAAGAFGSLS...
peptidase_P.a.4BY	QN..PDFIYIIGSDM..LSSDLNAIIGKANETYTTLNLOMT.YNNR.TDPNQYRSDYINFAKHGIIVIFYF...
M28_Sph.	SG..FVHAYDFISR.....WMSA...VPAHTTKIK.T.EF..PGSPGGGGSDTSFVAAGVAFMLS...
peptidase_Sph.	NG.NHNYIHSIGADK..LSSELARINKEANDKSSKLELDFM.YDNP.KDPMRLYRSDYINFAASKGIIIFIFYF...
M28_P.t.	DD..PNYVYILGSDK..LSTDLHEISEKANSTYTNNLDYT.YNDE.NDPNRFYRSDYINFAKFDIIFIFYF...
M28_C.p.	..G..TVWMTGMDRS.....NLGEQLNQALAPQNLKVYA.DPY..PKQNLFRSDYINFAKFDIIFIFYF...
ther_C.t.	AG..KPESVGVIVD.....YVDKNLTETFIKKVIKCYC.TIPF.VETKCGYACSDHASASKAGYSAFVI...
ther_M.t.	AG..KPESVGVIVD.....YVDPDLTQFIKVIDEYC.DIPY.VETECGYACSDHASASKAGYSAFVI...
ther_T.a.	AS..KINVY..ED.....DNSEWLAETFRIVSQEYNEYLELEVIFSGYTWSGSDHYRWEAGYHAIFYA...
consensus>70	.....d.....d.....d.....y..Sdh.....g.p.....
LapA	E..SAFGDDSPYIHSADDTIE..TVNFDHVLQHGRI.TLGFAYELAFAADS.....
GAP	E..GKFEDSNPNMHSRRDTIDLPGFSEVHMAQFARV.GIAFAWEMAGGK.....
T9SS_P.b.	..ES...NVSPVNHNTINDLA..NMVPPYAYEVIKGSLGAVLEFAV.....
M28_P.b.	N..G...EHEDYHOPTDTEPE..KIDVALLEKRTKILIFATAWYLANSENRIIVDQ.D
aminopep_P.b.	SFDG...DVTKYHYHRPGDEAH..TLDYDYLKFFKSYVLAKRKIANDPVFPVWNS.GDKY
peptidase_P.b.	N..G...THDDYHOPSDTPD..KIQDLDLERNRRLVFFYTAWEIANRVRDLVVDK.ADT
M28_R.p.	N..G...VHDDYHOHTDTPD..KIEFPLMTKRAQIIFHTAWDLANRMNRTPVDR.TST
M28_P.a.4BY	S..LGWSYGNYTWHNTNRTFD..KIVFDDVQSNAIVAAIMAYAASEDPERSTEK.SILPNNRTGTQPGVWPTPVKP
peptidase_P.a.4BY	N..G...VHEDYHOPGDEVE..KIHFDLAKRAQIIVYFTAWDLANRAKRPVADK.NED
M28_Sph.	S..LSWGYFSNTWHTNLDTYD..KLVFDDIKYNVILITAVMAYKASEEAKLVNREQ.RVLPAGT.DGKPGIWWPIRQP
peptidase_Sph.	S..G...LHPHYHTPDDTPD..KIDFPIVVKREKIFQTAWDVANSAATKPAVDS.KK
M28_P.t.	N..G...VHEDYHOPTDTEPE..KIEYDLEKRAKILVFFTAWEIANRDERPFVVDK.PT
M28_C.p.	STQL...DKDQHYHOVTDTPD..SILNLPMLQVVKMLAIATTPLVGSGITPSRIDPNKV
ther_C.t.	E..SDFDSDPHIHTTDDTIS..HLSEFDDHMLQHARMTLAFAYELAFT
ther_M.t.	E..SAFEYSDNHIHTTDDLIK..YLSFDDHMLQHARMTLAFAYELAFT
ther_T.a.	E..Y...NFNDYYHSPEDTIE..HMDISYATRASKIATLAELSEI
consensus>70	.....H...D..d.....f..m.....a...a.....



**Figure 10.5** Multiple sequence alignment of LapA with homologous psychrophilic and thermophilic sequences. Psychrophilic peptidases: [GAP] – aminopeptidase from *Glaciozyma Antarctica*; [T9SS\_P.b., M28\_P.b., aminopep\_P.b., peptidase\_P.b.] – four different peptidases from *Psychroserpens burtonensis*; [M28\_R.p.] – peptidase from *Rhodonellum psychrophilum*; [M28\_P.a.4BY, peptidase\_P.a.4BY] – two different peptidases from *Pedobacter antarcticus* 4BY; [M28\_Sph., peptidase\_Sph.] – two different peptidases from *Sphingobacterium* sp. IITKGP-BTPF85; [M28\_P.t.] – peptidase from *Psychroflexus torquis*; [M28\_C.p.] – peptidase from *Colwellia psychrerythraea*; Thermophilic peptidases: [ther\_C.t.] – aminopeptidase-like protein from *Chaetomium thermophilum* DSM 1495; [ther\_M.t.] – hypothetical protein from *Myceliophthora thermophila* ATCC 42464; [ther\_T.a.] – peptidase from *Thermoplasmatales archaeon* SCGC AB-539-C06.



**Figure 10.6** Superimposition of LapA crystal structure and homologous psychrophilic proteins. LapA is shown in green, M28 peptidase from *Psychroserpens burtonensis* – cyan, M28 peptidase from *Rhodonellum psychrophilum* – purple.



**Figure 10.7** Homology model of LapA from *Aspergillus oryzae*. Generated using IntFold2 server using the model of Lap from *Vibrio proteolyticus* (PDBid: 1RTQ) as a template.

## Appendix III

	<i>flavLapA</i>
Resolution (Å)	75.55 - 2.71 (2.807 - 2.71)
Space group	C 1 2 1
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	138.697 150.16 87.562
$\alpha$ , $\beta$ , $\gamma$ (°)	90 125.84 90
Total reflections	69962 (3291)
Unique reflections	39074 (943)
Multiplicity	3.3 (3.5)
Completeness (%)	99.06 (98.73)
Mean I/sigma(I)	5.69 (0.88)
Wilson B-factor	38.38
R-merge	0.2212 (1.359)
R-meas	0.2646
CC1/2	0.974 (0.363)
CC*	0.993 (0.73)
R-work	0.2663 (0.3683)
R-free	0.3496 (0.4192)
Number of non-hydrogen atoms	8981
macromolecules	8718
ligands	97
water	166
Protein residues	1158
RMS(bonds)	0.014
RMS(angles)	1.88
Ramachandran favored (%)	88
Ramachandran outliers (%)	2.8
Clashscore	31.7
Average B-factor	33.9
macromolecules	33.9
ligands	48.9
solvent	28.7

**Table 10.1** Data collection and refinement statistics of *flavLapA*. Statistics for the highest-resolution shell are shown in parentheses.

	<b><i>wt</i>LapA</b>
Resolution (Å)	73.81 - 1.97 (2.041 - 1.97)
Space group	P 63 2 2
Unit cell	153.92 153.92 88.64 90 90 120
Total reflections	574615 (56213)
Unique reflections	43994 (4297)
Multiplicity	13.1 (13.1)
Completeness (%)	1.00 (0.99)
Mean I/sigma(I)	10.35 (1.18)
Wilson B-factor	33.91
R-merge	0.2203 (2.178)
R-meas	0.2293 (2.267)
CC1/2	0.997 (0.653)
CC*	0.999 (0.889)
Reflections used in refinement	43993 (4297)
Reflections used for R-free	2125 (197)
R-work	0.1672 (0.3374)
R-free	0.1913 (0.3791)
CC(work)	0.941 (0.390)
CC(free)	0.939 (0.388)
Number of non-hydrogen atoms	2759
macromolecules	2334
ligands	83
Protein residues	303
RMS(bonds)	0.007
RMS(angles)	0.93
Ramachandran favored (%)	98
Ramachandran allowed (%)	1.3
Ramachandran outliers (%)	0.66
Rotamer outliers (%)	0.79
Clashscore	4.7
Average B-factor	40.13
macromolecules	36.83
ligands	75.98
solvent	53.91
Number of TLS groups	1

**Table 10.2** Data collection and refinement statistics of *wt*LapA. Statistics for the highest-resolution shell are shown in parentheses.

	<i>wtproLapA</i>
Resolution (Å)	42.61 - 1.61 (1.668 - 1.61)
Space group	P 21 21 21
Unit cell	43.43 94.39 99.11 90 90 90
Total reflections	231145 (22301)
Unique reflections	53338 (5269)
Multiplicity	4.3 (4.2)
Completeness (%)	99.48 (99.55)
Mean I/sigma(I)	11.36 (1.59)
Wilson B-factor	20.19
R-merge	0.07303 (0.8915)
R-meas	0.08292
CC1/2	0.998 (0.55)
CC*	1 (0.842)
Reflections used for R-free	
R-work	0.1527 (0.2876)
R-free	0.1739 (0.2899)
CC(work)	
CC(free)	
Number of non-hydrogen atoms	3281
macromolecules	2698
ligands	101
water	482
Protein residues	341
RMS(bonds)	0.007
RMS(angles)	1.25
Ramachandran favored (%)	98
Ramachandran allowed (%)	
Ramachandran outliers (%)	0
Clashscore	2.56
Average B-factor	27
macromolecules	23.1
ligands	56.7
solvent	42.5

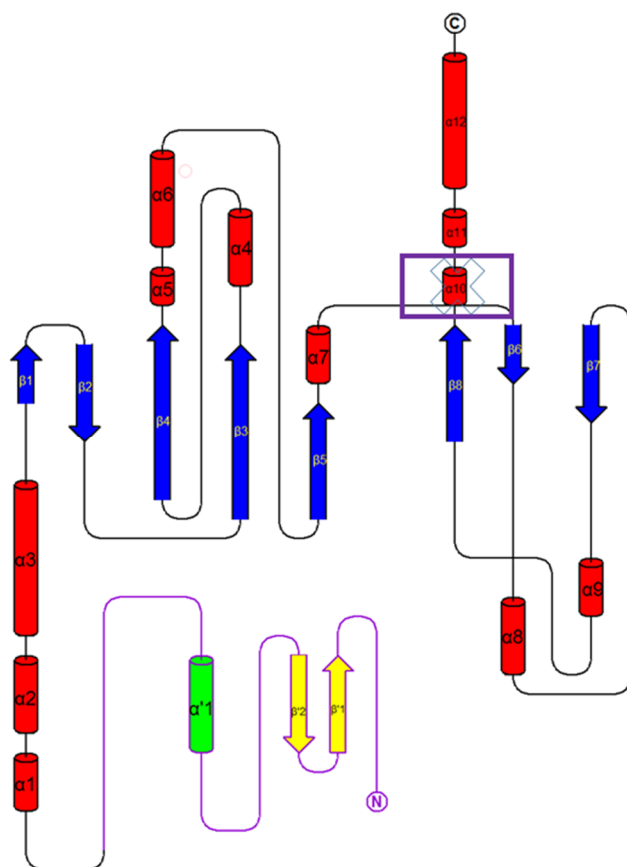
**Table 10.3** Data collection and refinement statistics of *wtproLapA*. Statistics for the highest-resolution shell are shown in parentheses.

<b>LapA M7</b>	
<b>Resolution (Å)</b>	45.14 - 1.43 (1.481 - 1.43)
<b>Space group</b>	P 21 21 21
<b>Unit cell</b>	43.15 90.28 99.15 90 90 90
<b>Total reflections</b>	140274 (12041)
<b>Unique reflections</b>	71455 (6528)
<b>Multiplicity</b>	2.0 (1.8)
<b>Completeness (%)</b>	98.66 (91.08)
<b>Mean I/sigma(I)</b>	11.32 (2.07)
<b>Wilson B-factor</b>	15.43
<b>R-merge</b>	0.04158 (0.3396)
<b>R-meas</b>	0.05881 (0.4803)
<b>R-pim</b>	0.04158 (0.3396)
<b>CC1/2</b>	0.996 (0.622)
<b>CC*</b>	0.999 (0.876)
<b>Reflections used in refinement</b>	71434 (6526)
<b>Reflections used for R-free</b>	3545 (331)
<b>R-work</b>	0.1571 (0.3159)
<b>R-free</b>	0.1801 (0.3327)
<b>CC(work)</b>	0.972 (0.789)
<b>CC(free)</b>	0.958 (0.763)
<b>Number of non-hydrogen atoms</b>	3233
<b>macromolecules</b>	2757
<b>ligands</b>	71
<b>solvent</b>	405
<b>Protein residues</b>	339
<b>RMS(bonds)</b>	0.005
<b>RMS(angles)</b>	0.84
<b>Ramachandran favored (%)</b>	98.51
<b>Ramachandran allowed (%)</b>	1.49
<b>Ramachandran outliers (%)</b>	0
<b>Rotamer outliers (%)</b>	1.36
<b>Clashscore</b>	5.77
<b>Average B-factor</b>	25.91
<b>macromolecules</b>	21.98
<b>ligands</b>	59.78
<b>solvent</b>	46.67
<b>Number of TLS groups</b>	1

**Table 10.4** Data collection and refinement statistics of LapA M7 mutant. Statistics for the highest-resolution shell are shown in parentheses.

<b>LapA M8</b>	
<b>Resolution range</b>	43.13 - 1.67 (1.73 - 1.67)
<b>Space group</b>	P 21 21 21
<b>Unit cell</b>	42.99 87.76 99.05 90 90 90
<b>Total reflections</b>	84937 (6805)
<b>Unique reflections</b>	42807 (3594)
<b>Multiplicity</b>	2.0 (1.9)
<b>Completeness (%)</b>	96.51 (81.72)
<b>Mean I/sigma(I)</b>	13.74 (2.86)
<b>Wilson B-factor</b>	15.9
<b>R-merge</b>	0.03682 (0.2529)
<b>R-meas</b>	0.05207 (0.3577)
<b>R-pim</b>	0.03682 (0.2529)
<b>CC1/2</b>	0.999 (0.694)
<b>CC*</b>	1 (0.905)
<b>Reflections used in refinement</b>	42806 (3594)
<b>Reflections used for R-free</b>	2211 (216)
<b>R-work</b>	0.1630 (0.2821)
<b>R-free</b>	0.2087 (0.3248)
<b>CC(work)</b>	0.976 (0.820)
<b>CC(free)</b>	0.954 (0.761)
<b>Number of non-hydrogen atoms</b>	3224
macromolecules	2764
ligands	113
solvent	347
<b>Protein residues</b>	342
<b>RMS(bonds)</b>	0.008
<b>RMS(angles)</b>	1.06
<b>Ramachandran favored (%)</b>	97.34
<b>Ramachandran allowed (%)</b>	2.37
<b>Ramachandran outliers (%)</b>	0.3
<b>Rotamer outliers (%)</b>	0.34
<b>Clashscore</b>	5.85
<b>Average B-factor</b>	25.75
macromolecules	23.53
ligands	44.27
solvent	37.36
<b>Number of TLS groups</b>	3

**Table 10.5** Data collection and refinement statistics of LapA M8 mutant. Statistics for the highest-resolution shell are shown in parentheses.



**Figure 10.8** LapA precursor topology diagram. Generated with TopDraw tool in CCP4 program suite. Alpha 10 helix (purple rectangle) is not present in the proenzyme and only appears in the mature enzyme, following activation.

Structural solution	Resolution (Å)	PDB accession
<i>flav</i> LapA	2.71	5NRJ
LapA proenzyme	1.61	5NQJ
LapA mature enzyme	1.97	5NQL

**Table 10.6** Resolution and accession codes for LapA structures deposited in the public PDB database.

## Literature references

- ACEVEDO, J. P., RODRIGUEZ, V., SAAVEDRA, M., MUNOZ, M., SALAZAR, O., ASENJO, J. A. & ANDREWS, B. A. 2013. Cloning, expression and decoding of the cold adaptation of a new widely represented thermolabile subtilisin-like protease. *J Appl Microbiol*, 114, 352-63.
- ADAMS, P. D., AFONINE, P. V., BUNKOCZI, G., CHEN, V. B., DAVIS, I. W., ECHOLS, N., HEADD, J. J., HUNG, L. W., KAPRAL, G. J., GROSSE-KUNTLEVE, R. W., MCCOY, A. J., MORIARTY, N. W., OEFFNER, R., READ, R. J., RICHARDSON, D. C., RICHARDSON, J. S., TERWILLIGER, T. C. & ZWART, P. H. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*, 66, 213-21.
- AGHAJARI, N., FELLER, G., GERDAY, C. & HASER, R. 1998. Crystal structures of the psychrophilic alpha-amylase from *Alteromonas haloplanctis* in its native form and complexed with an inhibitor. *Protein Sci*, 7, 564-72.
- AGHAJARI, N., FELLER, G., GERDAY, C. & HASER, R. 2002. Structural basis of alpha-amylase activation by chloride. *Protein Sci*, 11, 1435-41.
- ALIAS, N., AHMAD MAZIAN, M., SALLEH, A. B., BASRI, M. & RAHMAN, R. N. 2014. Molecular Cloning and Optimization for High Level Expression of Cold-Adapted Serine Protease from Antarctic Yeast *Glaciozyma antarctica* PI12. *Enzyme Res*, 2014, 197938.
- ALTSCHUL, S. F., MADDEN, T. L., SCHAFFER, A. A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, 25, 3389-402.
- ALVAREZ, M., ZEELLEN, J. P., MAINFROID, V., RENTIER-DELRUE, F., MARTIAL, J. A., WYNS, L., WIERENGA, R. K. & MAES, D. 1998. Triose-phosphate isomerase (TIM) of the psychrophilic bacterium *Vibrio marinus*. Kinetic and structural properties. *J Biol Chem*, 273, 2199-206.
- ANTHONY, J. C., ANTHONY, T. G., KIMBALL, S. R. & JEFFERSON, L. S. 2001. Signaling pathways involved in translational control of protein synthesis in skeletal muscle by leucine. *J Nutr*, 131, 856S-860S.
- APHALE, J. S. & STROHL, W. R. 1993. Purification and Properties of an Extracellular Aminopeptidase from *Streptomyces-Lividans* 1326. *Journal of General Microbiology*, 139, 417-424.
- ARAKI, T. 1991. The effect of temperature shifts on protein synthesis by the psychrophilic bacterium *Vibrio* sp. strain ANT-300. *J Gen Microbiol*, 137, 817-26.
- ARNORSDOTTIR, J., HELGADOTTIR, S., THORBJARNARDOTTIR, S. H., EGGERTSSON, G. & KRISTJANSSON, M. M. 2007. Effect of selected Ser/Ala and Xaa/Pro mutations on the stability and catalytic properties of a cold adapted subtilisin-like serine proteinase. *Biochim Biophys Acta*, 1774, 749-55.
- BALDWIN, R. L. 2008. The search for folding intermediates and the mechanism of protein folding. *Annu Rev Biophys*, 37, 1-21.
- BARRETT, A. J. & RAWLINGS, N. D. 1995. Families and clans of serine peptidases. *Arch Biochem Biophys*, 318, 247-50.
- BAUVOIS, C., JACQUAMET, L., HUSTON, A. L., BOREL, F., FELLER, G. & FERRER, J.-L. 2008. Crystal structure of the cold-active aminopeptidase from *Colwellia psychrerythraea*, a close structural homologue of the human bifunctional leukotriene A4 hydrolase. *The Journal of biological chemistry*, 283, 23315-25.

- BEGGAH, S., LECHENNE, B., REICHARD, U., FOUNDLING, S. & MONOD, M. 2000. Intra- and intermolecular events direct the propeptide-mediated maturation of the *Candida albicans* secreted aspartic proteinase Sap1p. *Microbiology-Uk*, 146, 2765-2773.
- BENNETT, B. & HOLZ, R. C. 1997. Spectroscopically distinct cobalt(II) sites in heterodimetallic forms of the aminopeptidase from *Aeromonas proteolytica*: characterization of substrate binding. *Biochemistry*, 36, 9837-46.
- BIASINI, M., BIENERT, S., WATERHOUSE, A., ARNOLD, K., STUDER, G., SCHMIDT, T., KIEFER, F., GALLO CASSARINO, T., BERTONI, M., BORDOLI, L. & SCHWEDE, T. 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res*, 42, W252-8.
- BLINKOVSKY, A. M., BYUN, T., BROWN, K. M., GOLIGHTLY, E. J. & KLOTZ, A. V. 2000. A non-specific aminopeptidase from *Aspergillus*. *Biochim Biophys Acta*, 1480, 171-81.
- BOS, C., GAUDICHON, C. & TOME, D. 2000. Nutritional and physiological criteria in the assessment of milk protein quality for humans. *J Am Coll Nutr*, 19, 191S-205S.
- BOWMAN, J. P. & MCCUAIG, R. D. 2003. Biodiversity, community structural shifts, and biogeography of prokaryotes within Antarctic continental shelf sediment. *Applied and environmental microbiology*, 69, 2463-83.
- BOZA, J. J., MOENNOZ, D., VUICHOUD, J., JARRET, A. R., GAUDARD-DE-WECK, D. & BALLEVRE, O. 2000. Protein hydrolysate vs free amino acid-based diets on the nutritional recovery of the starved rat. *Eur J Nutr*, 39, 237-43.
- BRAKE, A. J., MERRYWEATHER, J. P., COIT, D. G., HEBERLEIN, U. A., MASIARZ, F. R., MULLENBACH, G. T., URDEA, M. S., VALENZUELA, P. & BARR, P. J. 1984. Alpha-factor-directed synthesis and secretion of mature foreign proteins in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A*, 81, 4642-6.
- BROCK, T. D. & EDWARDS, M. R. 1970. Fine structure of *Thermus aquaticus*, an extreme thermophile. *J Bacteriol*, 104, 509-17.
- BZYMEK, K. P., D'SOUZA, V. M., CHEN, G., CAMPBELL, H., MITCHELL, A. & HOLZ, R. C. 2004. Function of the signal peptide and N- and C-terminal propeptides in the leucine aminopeptidase from *Aeromonas proteolytica*. *Protein Expr Purif*, 37, 294-305.
- CAO, J., HYMOWITZ, M., CONNER, C., BAHOU, W. F. & ZUCKER, S. 2000. The propeptide domain of membrane type 1-matrix metalloproteinase acts as an intramolecular chaperone when expressed in trans with the mature sequence in COS-1 cells. *J Biol Chem*, 275, 29648-53.
- CASANUEVA, A., TUFFIN, M., CARY, C. & COWAN, D. A. 2010. Molecular adaptations to psychrophily: the impact of 'omic' technologies. *Trends in Microbiology*, 18, 374-381.
- CAVICCHIOLI, R., CHARLTON, T., ERTAN, H., OMAR, S. M., SIDDIQUI, K. S. & WILLIAMS, T. J. 2011. Biotechnological uses of enzymes from psychrophiles. *Microbial Biotechnology*, 4, 449-460.
- CAVICCHIOLI, R., SIDDIQUI, K. S., ANDREWS, D. & SOWERS, K. R. 2002. Low-temperature extremophiles and their applications. *Current Opinion in Biotechnology*, 13, 253-261.
- CHAUDHURI, B., STEUBE, K. & STEPHAN, C. 1992. The pro-region of the yeast prepro-alpha-factor is essential for membrane translocation of human insulin-like growth factor 1 in vivo. *Eur J Biochem*, 206, 793-800.
- CHAYEN, N. E. 1998a. Comparative studies of protein crystallization by vapour-diffusion and microbatch techniques. *Acta Crystallogr D Biol Crystallogr*, 54, 8-15.

- CHAYEN, N. E. 1998b. Comparative studies of protein crystallization by vapour-diffusion and microbatch techniques. *Acta crystallographica Section D, Biological crystallography*, 54, 8-15.
- CHEN, X. L., XIE, B. B., LU, J. T., HE, H. L. & ZHANG, Y. 2007. A novel type of subtilase from the psychrotolerant bacterium *Pseudoalteromonas* sp. SM9913: catalytic and structural properties of deasein MCP-01. *Microbiology*, 153, 2116-25.
- CHERRY, J. R., LAMSA, M. H., SCHNEIDER, P., VIND, J., SVENDSEN, A., JONES, A. & PEDERSEN, A. H. 1999. Directed evolution of a fungal peroxidase. *Nat Biotechnol*, 17, 379-84.
- CIPOLLA, A., D'AMICO, S., BARUMANDZADEH, R., MATAGNE, A. & FELLER, G. 2011. Stepwise Adaptations to Low Temperature as Revealed by Multiple Mutants of Psychrophilic alpha-Amylase from Antarctic Bacterium. *Journal of Biological Chemistry*, 286, 38348-38355.
- CLAVERIE, P., VIGANO, C., RUYSSCHAERT, J. M., GERDAY, C. & FELLER, G. 2003. The precursor of a psychrophilic alpha-amylase: structural characterization and insights into cold adaptation. *Biochim Biophys Acta*, 1649, 119-22.
- CLEMENTE, A. 2000. Enzymatic protein hydrolysates in human nutrition. *Trends in Food Science & Technology*, 11, 254-262.
- COLLINS, T., MEUWIS, M. A., GERDAY, C. & FELLER, G. 2003. Activity, stability and flexibility in Glycosidases adapted to extreme thermal environments. *Journal of Molecular Biology*, 328, 419-428.
- CONESA, C. & FITZGERALD, R. J. 2013. Total Solids Content and Degree of Hydrolysis Influence Proteolytic Inactivation Kinetics Following Whey Protein Hydrolysate Manufacture. *Journal of Agricultural and Food Chemistry*, 61, 10135-10144.
- D'AMICO, S., CLAVERIE, P., COLLINS, T., GEORLETTE, D., GRATIA, E., HOYOUX, A., MEUWIS, M. A., FELLER, G. & GERDAY, C. 2002a. Molecular basis of cold adaptation. *Philos Trans R Soc Lond B Biol Sci*, 357, 917-25.
- D'AMICO, S., GERDAY, C. & FELLER, G. 2001. Structural determinants of cold adaptation and stability in a large protein. *J Biol Chem*, 276, 25791-6.
- D'AMICO, S., GERDAY, C. & FELLER, G. 2002b. Dual effects of an extra disulfide bond on the activity and stability of a cold-adapted alpha-amylase. *J Biol Chem*, 277, 46110-5.
- D'AMICO, S., MARX, J. C., GERDAY, C. & FELLER, G. 2003. Activity-stability relationships in extremophilic enzymes. *J Biol Chem*, 278, 7891-6.
- D'AMICO, S., SOHIER, J. S. & FELLER, G. 2006. Kinetics and energetics of ligand binding determined by microcalorimetry: insights into active site mobility in a psychrophilic alpha-amylase. *J Mol Biol*, 358, 1296-304.
- DALY, R. & HEARN, M. T. 2005. Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. *J Mol Recognit*, 18, 119-38.
- DAVAIL, S., FELLER, G., NARINX, E. & GERDAY, C. 1994. Cold adaptation of proteins. Purification, characterization, and sequence of the heat-labile subtilisin from the antarctic psychrophile *Bacillus* TA41. *J Biol Chem*, 269, 17448-53.
- DEMING, J. W. 2002. Psychrophiles and polar regions. *Current opinion in microbiology*, 5, 301-9.
- DEMIRJIAN, D. C., MORIS-VARAS, F. & CASSIDY, C. S. 2001. Enzymes from extremophiles. *Current Opinion in Chemical Biology*, 5, 144-151.
- DONG, D., IHARA, T., MOTOSHIMA, H. & WATANABE, K. 2005. Crystallization and preliminary X-ray crystallographic studies of a psychrophilic subtilisin-like protease Apa1 from Antarctic *Pseudoalteromonas* sp. strain AS-11. *Acta Crystallogr Sect F Struct Biol Cryst Commun*, 61, 308-11.

- EDGAR, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32, 1792-1797.
- EMSLEY, P. & COWTAN, K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*, 60, 2126-32.
- ERMAK, D. L. & MCCAMMON, J. A. 1978. BROWNIAN DYNAMICS WITH HYDRODYNAMIC INTERACTIONS. *Journal of Chemical Physics*, 69, 1352-1360.
- FELLER, G. 2010. Protein stability and enzyme activity at extreme biological temperatures. *Journal of Physics-Condensed Matter*, 22.
- FELLER, G. 2013. Psychrophilic enzymes: from folding to function and biotechnology. *Scientifica*, 2013, 512840.
- FELLER, G., D'AMICO, D. & GERDAY, C. 1999. Thermodynamic stability of a cold-active alpha-amylase from the Antarctic bacterium *Alteromonas haloplanctis*. *Biochemistry*, 38, 4613-9.
- FELLER, G. & GERDAY, C. 2003. Psychrophilic enzymes: hot topics in cold adaptation. *Nat Rev Microbiol*, 1, 200-8.
- FELLER, G., PAYAN, F., THEYS, F., QIAN, M., HASER, R. & GERDAY, C. 1994. Stability and structural analysis of alpha-amylase from the antarctic psychrophile *Alteromonas haloplanctis* A23. *Eur J Biochem*, 222, 441-7.
- FIELDS, P. A. 2001. Review: Protein function at thermal extremes: balancing stability and flexibility. *Comp Biochem Physiol A Mol Integr Physiol*, 129, 417-31.
- FIELDS, P. A. & SOMERO, G. N. 1998. Hot spots in cold adaptation: localized increases in conformational flexibility in lactate dehydrogenase A4 orthologs of Antarctic notothenioid fishes. *Proc Natl Acad Sci U S A*, 95, 11476-81.
- FUCHITA, N., ARITA, S., IKUTA, J., MIURA, M., SHIMOMURA, K., MOTOSHIMA, H. & WATANABE, K. 2012. Gly or Ala substitutions for Pro(210)Thr(211)Asn(212) at the beta8-beta9 turn of subtilisin Carlsberg increase the catalytic rate and decrease thermostability. *Biochim Biophys Acta*, 1824, 620-6.
- GEORLETTE, D., BLAISE, V., COLLINS, T., D'AMICO, S., GRATIA, E., HOYOUX, A., MARX, J. C., SONAN, G., FELLER, G. & GERDAY, C. 2004. Some like it cold: biocatalysis at low temperatures. *Fems Microbiology Reviews*, 28, 25-42.
- GEORLETTE, D., DAMIEN, B., BLAISE, V., DEPIEREUX, E., UVERSKY, V. N., GERDAY, C. & FELLER, G. 2003. Structural and functional adaptations to extreme temperatures in psychrophilic, mesophilic, and thermophilic DNA ligases. *J Biol Chem*, 278, 37015-23.
- GHASEMI, Y., DABBAGH, F. & GHASEMIAN, A. 2012. Cloning of a fibrinolytic enzyme (subtilisin) gene from *Bacillus subtilis* in *Escherichia coli*. *Mol Biotechnol*, 52, 1-7.
- GIOVANNONI, S. J., BRITSCHGI, T. B., MOYER, C. L. & FIELD, K. G. 1990. Genetic diversity in Sargasso Sea bacterioplankton. *Nature*, 345, 60-3.
- GIVER, L., GERSHENSON, A., FRESKGARD, P. O. & ARNOLD, F. H. 1998. Directed evolution of a thermostable esterase. *Proc Natl Acad Sci U S A*, 95, 12809-13.
- GOODCHILD, A., RAFTERY, M., SAUNDERS, N. F., GUILHAUS, M. & CAVICCHIOLI, R. 2004. Biology of the cold adapted archaeon, *Methanococcoides burtonii* determined by proteomics using liquid chromatography-tandem mass spectrometry. *J Proteome Res*, 3, 1164-76.
- HARTLEY, M., YONG, W. & BENNETT, B. 2009. Heterologous expression and purification of *Vibrio proteolyticus* (*Aeromonas proteolytica*) aminopeptidase: a rapid protocol. *Protein Expr Purif*, 66, 91-101.
- HE, H. L., CHEN, X. L., LI, J. W., ZHANG, Y. Z. & GAO, P. J. 2004. Taste improvement of refrigerated meat treated with cold-adapted Protease. *Food Chemistry*, 84, 307-311.

- HEBRAUD, M. & POTIER, P. 1999. Cold shock response and low temperature adaptation in psychrotrophic bacteria. *J Mol Microbiol Biotechnol*, 1, 211-9.
- HEGDE, R. S. & BERNSTEIN, H. D. 2006. The surprising complexity of signal sequences. *Trends Biochem Sci*, 31, 563-71.
- HOLM, L. & ROSENSTROM, P. 2010. Dali server: conservation mapping in 3D. *Nucleic acids research*, 38, W545-9.
- HUSTON, A. L. 2008. *Biotechnological aspects of cold-adapted enzymes*.
- HUSTON, A. L., HAEGGSTROM, J. Z. & FELLER, G. 2008. Cold adaptation of enzymes: Structural, kinetic and microcalorimetric characterizations of an aminopeptidase from the Arctic psychrophile *Colwellia psychrerythraea* and of human leukotriene A(4) hydrolase. *Biochimica Et Biophysica Acta-Proteins and Proteomics*, 1784, 1865-1872.
- IKEMURA, H., TAKAGI, H. & INOUE, M. 1987. REQUIREMENT OF PRO-SEQUENCE FOR THE PRODUCTION OF ACTIVE SUBTILISIN-E IN *ESCHERICHIA-COLI*. *Journal of Biological Chemistry*, 262, 7859-7864.
- IMPERIALI, B. & O'CONNOR, S. E. 1999. Effect of N-linked glycosylation on glycopeptide and glycoprotein structure. *Curr Opin Chem Biol*, 3, 643-9.
- JAENICKE, R. 1991a. Protein folding: local structures, domains, subunits, and assemblies. *Biochemistry*, 30, 3147-61.
- JAENICKE, R. 1991b. Protein stability and protein folding. *Ciba Found Symp*, 161, 206-16; discussion 217-21.
- JEON, J. H., KIM, J. T., KANG, S. G., LEE, J. H. & KIM, S. J. 2009. Characterization and its potential application of two esterases derived from the arctic sediment metagenome. *Mar Biotechnol (NY)*, 11, 307-16.
- KABSCH, W. 2010. Xds. *Acta Crystallogr D Biol Crystallogr*, 66, 125-32.
- KAINZ, E., GALLMETZER, A., HATZL, C., NETT, J. H., LI, H., SCHINKO, T., PACHLINGER, R., BERGER, H., REYES-DOMINGUEZ, Y., BERNREITER, A., GERNGROSS, T., WILDT, S. & STRAUSS, J. 2008. N-glycan modification in *Aspergillus* species. *Appl Environ Microbiol*, 74, 1076-86.
- KALLBERG, M., WANG, H., WANG, S., PENG, J., WANG, Z., LU, H. & XU, J. 2012. Template-based protein structure modeling using the RaptorX web server. *Nat Protoc*, 7, 1511-22.
- KANO, H., TAGUCHI, S. & MOMOSE, H. 1997. Cold adaptation of a mesophilic serine protease, subtilisin, by in vitro random mutagenesis. *Appl Microbiol Biotechnol*, 47, 46-51.
- KAWAMOTO, J., KURIHARA, T., KITAGAWA, M., KATO, I. & ESAKI, N. 2007. Proteomic studies of an Antarctic cold-adapted bacterium, *Shewanella livingstonensis* Ac10, for global identification of cold-inducible proteins. *Extremophiles*, 11, 819-26.
- KELLEY, L. A. & STERNBERG, M. J. 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc*, 4, 363-71.
- KJELDSSEN, T., PETTERSSON, A. F., HACH, M., DIERS, I., HAVELUND, S., HANSEN, P. H. & ANDERSEN, A. S. 1997. Synthetic leaders with potential BiP binding mediate high-yield secretion of correctly folded insulin precursors from *Saccharomyces cerevisiae*. *Protein Expr Purif*, 9, 331-6.
- KOMA, D., YAMANAKA, H., MORIYOSHI, K., OHMOTO, T. & SAKAI, K. 2007. Over-expression and characterization of thermostable serine protease in *Escherichia coli* encoded by the ORF TTE0824 from *Thermoanaerobacter tengcongensis*. *Extremophiles*, 11, 769-79.
- KU, T. H., LU, P. Y., CHAN, C. H., WANG, T. S., LAI, S. Z., LYU, P. C. & HSIAO, N. W. 2009. Predicting melting temperature directly from protein sequences. *Computational Biology and Chemistry*, 33, 445-450.

- KUDDUS, M. & RAMTEKE, P. W. 2009. Cold-active extracellular alkaline protease from an alkaliphilic *Stenotrophomonas maltophilia*: production of enzyme and its industrial applications. *Can J Microbiol*, 55, 1294-301.
- KUDDUS, M. & RAMTEKE, P. W. 2011. Production optimization of an extracellular cold-active alkaline protease from *Stenotrophomonas maltophilia* MTCC 7528 and its application in detergent industry. *African Journal of Microbiology Research*, 5, 809-816.
- KUDDUS, M. & RAMTEKE, P. W. 2012. Recent developments in production and biotechnological applications of cold-active microbial proteases. *Critical Reviews in Microbiology*, 38, 330-338.
- KULAKOVA, L., GALKIN, A., NAKAYAMA, T., NISHINO, T. & ESAKI, N. 2004. Cold-active esterase from *Psychrobacter* sp. Ant300: gene cloning, characterization, and the effects of Gly-->Pro substitution near the active site on its catalytic activity and stability. *Biochimica et biophysica acta*, 1696, 59-65.
- KUMAR, A., PERIYANNAN, G. R., NARAYANAN, B., KITTELL, A. W., KIM, J.-J. & BENNETT, B. 2007. Experimental evidence for a metallohydrolase mechanism in which the nucleophile is not delivered by a metal ion: EPR spectrokinetic and structural studies of aminopeptidase from *Vibrio proteolyticus*. *The Biochemical journal*, 403, 527-36.
- KUMAR, C. G. & TAKAGI, H. 1999. Microbial alkaline proteases: from a bioindustrial viewpoint. *Biotechnol Adv*, 17, 561-94.
- KUMAR, S., MA, B., TSAI, C. J., SINHA, N. & NUSSINOV, R. 2000. Folding and binding cascades: dynamic landscapes and population shifts. *Protein Sci*, 9, 10-9.
- LAHL, W. J. & BRAUN, S. D. 1994. ENZYMATIC PRODUCTION OF PROTEIN HYDROLYSATES FOR FOOD USE. *Food Technology*, 48, 68-71.
- LEI, D., XU, Y., HE, Q., PANG, Y., CHEN, B., XIONG, L. & LI, Y. 2013. Glycosylation analysis of recombinant neutral protease I from *Aspergillus oryzae* expressed in *Pichia pastoris*. *Biotechnol Lett*, 35, 2121-7.
- LIJUN, Y., LIN, Z., REGENSTEIN, J. M., MOUMING, Z. & DONG, L. 2012. Effect of thermal treatment on the characteristic properties of loach peptide. *International Journal of Food Science & Technology*, 47, 2574-2581.
- LIN-CEREGHINO, G. P., STARK, C. M., KIM, D., CHANG, J., SHAHEEN, N., POERWANTO, H., AGARI, K., MOUA, P., LOW, L. K., TRAN, N., HUANG, A. D., NATTESTAD, M., OSHIRO, K. T., CHANG, J. W., CHAVAN, A., TSAI, J. W. & LIN-CEREGHINO, J. 2013. The effect of alpha-mating factor secretion signal mutations on recombinant protein expression in *Pichia pastoris*. *Gene*, 519, 311-7.
- LONHIENNE, T., ZOIDAKIS, J., VORGAS, C. E., FELLER, G., GERDAY, C. & BOURIOTIS, V. 2001. Modular structure, local flexibility and cold-activity of a novel chitinase from a psychrophilic Antarctic bacterium. *J Mol Biol*, 310, 291-7.
- LOOKE, M., KRISTJUHAN, K. & KRISTJUHAN, A. 2011. Extraction of genomic DNA from yeasts for PCR-based applications. *BioTechniques*, 50, 325-8.
- LOW, P. S., BADA, J. L. & SOMERO, G. N. 1973. Temperature adaptation of enzymes: roles of the free energy, the enthalpy, and the entropy of activation. *Proc Natl Acad Sci U S A*, 70, 430-2.
- MA, Y., WANG, L., SUN, X., MA, B., ZHANG, J., GAO, F. & LIU, C. 2013. Study on Hydrolysis Conditions of Flavourzyme in Soybean Polypeptide Alcalase Hydrolysate. In: JIANG, Z. Y., LIU, X. H., JIAO, S. H. & HAN, J. T. (eds.) *Advances in Materials and Materials Processing, Pts 1-3*.
- MARGESIN, R. & FELLER, G. 2010. Biotechnological applications of psychrophiles. *Environ Technol*, 31, 835-44.
- MARGESIN, R. & SCHINNER, F. 1994. Properties of Cold-Adapted Microorganisms and Their Potential Role in Biotechnology. *Journal of Biotechnology*, 33, 1-14.

- MATSUSHITA-MORITA, M., TADA, S., SUZUKI, S., HATTORI, R., MARUI, J., FURUKAWA, I., YAMAGATA, Y., AMANO, H., ISHIDA, H., TAKEUCHI, M., KASHIWAGI, Y. & KUSUMOTO, K. I. 2011. Over-expression and Characterization of an Extracellular Leucine Aminopeptidase from *Aspergillus oryzae*. *Current Microbiology*, 62, 557-564.
- MAVROMATIS, K., TSIGOS, I., TZANODASKALAKI, M., KOKKINIDIS, M. & BOURIOTIS, V. 2002. Exploring the role of a glycine cluster in cold adaptation of an alkaline phosphatase. *European journal of biochemistry / FEBS*, 269, 2330-5.
- MCCOY, A. J., GROSSE-KUNSTLEVE, R. W., STORONI, L. C. & READ, R. J. 2005. Likelihood-enhanced fast translation functions. *Acta Crystallogr D Biol Crystallogr*, 61, 458-64.
- MCCOY, M. 2001. Novozymes emerges. *Chemical & Engineering News*, 79, 23-25.
- MERZ, A., YEE, M. C., SZADKOWSKI, H., PAPPENBERGER, G., CRAMERI, A., STEMMER, W. P., YANOFSKY, C. & KIRSCHNER, K. 2000. Improving the catalytic activity of a thermophilic enzyme at low temperatures. *Biochemistry*, 39, 880-9.
- MERZ, M., EISELE, T., BERENDS, P., APPEL, D., RABE, S., BLANK, I., STRESSLER, T. & FISCHER, L. 2015. Flavourzyme, an Enzyme Preparation with Industrial Relevance: Automated Nine-Step Purification and Partial Characterization of Eight Enzymes. *Journal of agricultural and food chemistry*, 63, 5682-93.
- MICHEL, V., LEHOUX, I., DEPRET, G., ANGLADE, P., LABADIE, J. & HEBRAUD, M. 1997. The cold shock response of the psychrotrophic bacterium *Pseudomonas fragi* involves four low-molecular-mass nucleic acid-binding proteins. *J Bacteriol*, 179, 7331-42.
- MORITA, R. Y. 1975. Psychrophilic bacteria. *Bacteriological reviews*, 39, 144-67.
- MURYOI, N., SATO, M., KANEKO, S., KAWAHARA, H., OBATA, H., YAISH, M. W., GRIFFITH, M. & GLICK, B. R. 2004. Cloning and expression of *afpA*, a gene encoding an antifreeze protein from the arctic plant growth-promoting rhizobacterium *Pseudomonas putida* GR12-2. *J Bacteriol*, 186, 5661-71.
- NAJAFI, M. F., DEOBAGKAR, D. & DEOBAGKAR, D. 2005. Potential application of protease isolated from *Pseudomonas aeruginosa* PD100. *Electronic Journal of Biotechnology*, 8, 197-203.
- NARINX, E., BAISE, E. & GERDAY, C. 1997. Subtilisin from psychrophilic antarctic bacteria: characterization and site-directed mutagenesis of residues possibly involved in the adaptation to cold. *Protein Eng*, 10, 1271-9.
- NICOLI, M. C., ANESE, M., PARPINEL, M. T., FRANCESCHI, S. & LERICI, C. R. 1997. Loss and/or formation of antioxidants during food processing and storage. *Cancer Letters*, 114, 71-74.
- NOTREDAME, C., HIGGINS, D. G. & HERINGA, J. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of molecular biology*, 302, 205-17.
- OKUBO, Y., YOKOIGAWA, K., ESAKI, N., SODA, K. & KAWAI, H. 1999. Characterization of psychrophilic alanine racemase from *Bacillus psychrosaccharolyticus*. *Biochem Biophys Res Commun*, 256, 333-40.
- PARRY, M. A. 2002. Over-expression and purification of active serine proteases and their variants from *Escherichia coli* inclusion bodies. *Curr Protoc Protein Sci*, Chapter 21, Unit 21 11.
- PIETTE, F., D'AMICO, S., STRUVAY, C., MAZZUCHELLI, G., RENAUT, J., TUTINO, M. L., DANCHIN, A., LEPRINCE, P. & FELLER, G. 2010. Proteomics of life at low temperatures: trigger factor is the primary chaperone in the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125. *Mol Microbiol*, 76, 120-32.

- PIETTE, F., STRUVAY, C. & FELLER, G. 2011. The protein folding challenge in psychrophiles: facts and current issues. *Environ Microbiol*, 13, 1924-33.
- POMEROY, L. R. & WIEBE, W. J. 2001. Temperature and substrates as interactive limiting factors for marine heterotrophic bacteria. *Aquatic Microbial Ecology*, 23, 187-204.
- QIU, Y., KATHARIOU, S. & LUBMAN, D. M. 2006. Proteomic analysis of cold adaptation in a Siberian permafrost bacterium--*Exiguobacterium sibiricum* 255-15 by two-dimensional liquid separation coupled with mass spectrometry. *Proteomics*, 6, 5221-33.
- RAO, M. B., TANKSALE, A. M., GHATGE, M. S. & DESHPANDE, V. V. 1998. Molecular and biotechnological aspects of microbial proteases. *Microbiol Mol Biol Rev*, 62, 597-635.
- RAWLINGS, N. D. & BARRETT, A. J. 1994. Families of serine peptidases. *Methods Enzymol*, 244, 19-61.
- REETZ, M. T. & CARBALLEIRA, J. D. 2007. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nature protocols*, 2, 891-903.
- ROBERT, X. & GOUET, P. 2014. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res*, 42, W320-4.
- ROCHE, D. B., BUENAVISTA, M. T., TETCHNER, S. J. & MCGUFFIN, L. J. 2011. The IntFOLD server: an integrated web resource for protein fold recognition, 3D model quality assessment, intrinsic disorder prediction, domain prediction and ligand binding site prediction. *Nucleic Acids Res*, 39, W171-6.
- RODRIGUES, D. F., IVANOVA, N., HE, Z., HUEBNER, M., ZHOU, J. & TIEDJE, J. M. 2008. Architecture of thermal adaptation in an *Exiguobacterium sibiricum* strain isolated from 3 million year old permafrost: a genome and transcriptome approach. *BMC Genomics*, 9, 547.
- RUSSELL, N. J. 1997. Psychrophilic bacteria--molecular adaptations of membrane lipids. *Comparative biochemistry and physiology Part A, Physiology*, 118, 489-93.
- RUSSELL, R. J., GERIKE, U., DANSON, M. J., HOUGH, D. W. & TAYLOR, G. L. 1998. Structural adaptations of the cold-active citrate synthase from an Antarctic bacterium. *Structure*, 6, 351-61.
- SAELENMINDE, G., HALSKAU, O., JR. & JONASSEN, I. 2009. Amino acid contacts in proteins adapted to different temperatures: hydrophobic interactions and surface charges play a key role. *Extremophiles*, 13, 11-20.
- SAGT, C. M., KLEIZEN, B., VERWAAL, R., DE JONG, M. D., MULLER, W. H., SMITS, A., VISSER, C., BOONSTRA, J., VERKLEIJ, A. J. & VERRIPS, C. T. 2000. Introduction of an N-glycosylation site increases secretion of heterologous proteins in yeasts. *Appl Environ Microbiol*, 66, 4940-4.
- SAHA, B. C. & HAYASHI, K. 2001. Debittering of protein hydrolyzates. *Biotechnology Advances*, 19, 355-370.
- SAKAGUCHI, M., MATSUZAKI, M., NIIMIYA, K., SEINO, J., SUGAHARA, Y. & KAWAKITA, M. 2007. Role of proline residues in conferring thermostability on aqualysin I. *J Biochem*, 141, 213-20.
- SCHALK, C., REMY, J. M., CHEVRIER, B., MORAS, D. & TARNUS, C. 1992. Rapid purification of the *Aeromonas proteolytica* aminopeptidase: crystallization and preliminary X-ray data. *Arch Biochem Biophys*, 294, 91-7.
- SHENTAL-BECHOR, D. & LEVY, Y. 2008. Effect of glycosylation on protein folding: a close look at thermodynamic stabilization. *Proc Natl Acad Sci U S A*, 105, 8256-61.
- SHINDE, U. & INOUE, M. 1996. Propeptide-mediated folding in subtilisin: the intramolecular chaperone concept. *Adv Exp Med Biol*, 379, 147-54.
- SIDDIQUI, K. S. & CAVICCHIOLI, R. 2006. Cold-adapted enzymes. *Annual Review of Biochemistry*.

- SIDDIQUI, K. S., PARKIN, D. M., CURMI, P. M., DE FRANCISCI, D., POLJAK, A., BARROW, K., NOBLE, M. H., TREWHELLA, J. & CAVICCHIOLI, R. 2009. A novel approach for enhancing the catalytic efficiency of a protease at low temperature: reduction in substrate inhibition by chemical modification. *Biotechnol Bioeng*, 103, 676-86.
- SIDDIQUI, K. S., POLJAK, A., GUILHAUS, M., DE FRANCISCI, D., CURMI, P. M., FELLER, G., D'AMICO, S., GERDAY, C., UVERSKY, V. N. & CAVICCHIOLI, R. 2006. Role of lysine versus arginine in enzyme cold-adaptation: modifying lysine to homo-arginine stabilizes the cold-adapted alpha-amylase from *Pseudoalteromonas haloplanktis*. *Proteins*, 64, 486-501.
- SIDDIQUI, K. S., POLJAK, A., GUILHAUS, M., FELLER, G., D'AMICO, S., GERDAY, C. & CAVICCHIOLI, R. 2005. Role of disulfide bridges in the activity and stability of a cold-active alpha-amylase. *J Bacteriol*, 187, 6206-12.
- SIDDIQUI, K. S., WILLIAMS, T. J., WILKINS, D., YAU, S., ALLEN, M. A., BROWN, M. V., LAURO, F. M. & CAVICCHIOLI, R. 2013. Psychrophiles. In: JEANLOZ, R. (ed.) *Annual Review of Earth and Planetary Sciences, Vol 41*.
- SIEMENSMA, A. D., WEIJER, W. J. & BAK, H. J. 1993. THE IMPORTANCE OF PEPTIDE LENGTHS IN HYPOALLERGENIC INFANT FORMULAS. *Trends in Food Science & Technology*, 4, 16-21.
- SIEZEN, R. J. 1996. Subtilases: subtilisin-like serine proteases. *Adv Exp Med Biol*, 379, 75-93.
- SMALAS, A. O., HEIMSTAD, E. S., HORDVIK, A., WILLASSEN, N. P. & MALE, R. 1994. Cold adaption of enzymes: structural comparison between salmon and bovine trypsins. *Proteins*, 20, 149-66.
- SMALAS, A. O., LEIROS, H. K., OS, V. & WILLASSEN, N. P. 2000. Cold adapted enzymes. *Biotechnology annual review*, 6, 1-57.
- SOMERO, G. N. 2004. Adaptation of enzymes to temperature: searching for basic "strategies". *Comp Biochem Physiol B Biochem Mol Biol*, 139, 321-33.
- TAGUCHI, S., KOMADA, S. & MOMOSE, H. 2000. The complete amino acid substitutions at position 131 that are positively involved in cold adaptation of subtilisin BPN'. *Appl Environ Microbiol*, 66, 1410-5.
- TAGUCHI, S., OZAKI, A. & MOMOSE, H. 1998a. Engineering of a cold-adapted protease by sequential random mutagenesis and a screening system. *Applied and Environmental Microbiology*, 64, 492-495.
- TAGUCHI, S., OZAKI, A. & MOMOSE, H. 1998b. Engineering of a cold-adapted protease by sequential random mutagenesis and a screening system. *Appl Environ Microbiol*, 64, 492-5.
- TEHEI, M., FRANZETTI, B., MADERN, D., GINZBURG, M., GINZBURG, B. Z., GIUDICI-ORTICONI, M. T., BRUSCHI, M. & ZACCAI, G. 2004. Adaptation to extreme environments: macromolecular dynamics in bacteria compared in vivo by neutron scattering. *EMBO Rep*, 5, 66-70.
- TINDBAEK, N., SVENDSEN, A., OESTERGAARD, P. R. & DRABORG, H. 2004. Engineering a substrate-specific cold-adapted subtilisin. *Protein Engineering Design & Selection*, 17, 149-156.
- TSAL, C. J., MA, B. & NUSSINOV, R. 1999. Folding and binding cascades: shifts in energy landscapes. *Proc Natl Acad Sci U S A*, 96, 9970-2.
- TURKIEWICZ, M., PAZGIER, M., KALINOWSKA, H. & BIELECKI, S. 2003. A cold-adapted extracellular serine proteinase of the yeast *Leucosporidium antarcticum*. *Extremophiles*, 7, 435-42.
- VAITILINGOM, M., AMATO, P., SANCELME, M., LAJ, P., LERICHE, M. & DELORT, A.-M. 2010. Contribution of microbial activity to carbon chemistry in clouds. *Applied and environmental microbiology*, 76, 23-9.

- VANZ, A. L., NIMTZ, M. & RINAS, U. 2014. Decrease of UPR- and ERAD-related proteins in *Pichia pastoris* during methanol-induced secretory insulin precursor production in controlled fed-batch cultures. *Microbial cell factories*, 13, 23.
- VETRIANI, C., MAEDER, D. L., TOLLIDAY, N., YIP, K. S., STILLMAN, T. J., BRITTON, K. L., RICE, D. W., KLUMP, H. H. & ROBB, F. T. 1998. Protein thermostability above 100 degreesC: a key role for ionic interactions. *Proc Natl Acad Sci U S A*, 95, 12300-5.
- VIEILLE, C., BURDETTE, D. S. & ZEIKUS, J. G. 1996. Thermozyms. *Biotechnol Annu Rev*, 2, 1-83.
- WATANABE, T., KOJIMA, H. & FUKUI, M. 2012. Draft genome sequence of a psychrotolerant sulfur-oxidizing bacterium, *Sulfuricella denitrificans* skB26, and proteomic insights into cold adaptation. *Appl Environ Microbiol*, 78, 6545-9.
- WINTER, G., LOBLEY, C. M. & PRINCE, S. M. 2013. Decision making in xia2. *Acta Crystallogr D Biol Crystallogr*, 69, 1260-73.
- WINTRODE, P. L. & ARNOLD, F. H. 2000. Temperature adaptation of enzymes: lessons from laboratory evolution. *Adv Protein Chem*, 55, 161-225.
- WU, J., BIAN, Y., TANG, B., CHEN, X., SHEN, P. & PENG, Z. 2004. Cloning and analysis of WF146 protease, a novel thermophilic subtilisin-like protease with four inserted surface loops. *FEMS Microbiol Lett*, 230, 251-8.
- XIE, B. B., BIAN, F., CHEN, X. L., HE, H. L., GUO, J., GAO, X., ZENG, Y. X., CHEN, B., ZHOU, B. C. & ZHANG, Y. Z. 2009. Cold adaptation of zinc metalloproteases in the thermolysin family from deep sea and arctic sea ice bacteria revealed by catalytic and structural properties and molecular dynamics: new insights into relationship between conformational flexibility and hydrogen bonding. *J Biol Chem*, 284, 9257-69.
- XUE, T., QIUPING, W., GUOWEI, L. & YONGHUI, S. 2012. Effects of heat treatment on structural modification and in vivo antioxidant capacity of soy protein. *Nutrition*, 28, 1180-1185.
- YIP, K. S., STILLMAN, T. J., BRITTON, K. L., ARTYMIUK, P. J., BAKER, P. J., SEDELNIKOVA, S. E., ENGEL, P. C., PASQUO, A., CHIARALUCE, R., CONSALVI, V. & ET AL. 1995. The structure of *Pyrococcus furiosus* glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. *Structure*, 3, 1147-58.
- YURIMOTO, H., YAMANE, M., KIKUCHI, Y., MATSUI, H., KATO, N. & SAKAI, Y. 2004. The propeptide of *Streptomyces mobaraensis* transglutaminase functions in cis and in trans to mediate efficient secretion of active enzyme from methylotrophic yeasts. *Bioscience, biotechnology, and biochemistry*, 68, 2058-69.
- ZHANG, Y. & SKOLNICK, J. 2005. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res*, 33, 2302-9.
- ZHENG, S., PONDER, M. A., SHIH, J. Y., TIEDJE, J. M., THOMASHOW, M. F. & LUBMAN, D. M. 2007. A proteomic analysis of *Psychrobacter articus* 273-4 adaptation to low temperature and salinity using a 2-D liquid mapping approach. *Electrophoresis*, 28, 467-88.