

The  
University  
Of  
Sheffield.

# Functional and structural basis of peptidoglycan recognition by the SH3\_5 domain family

By:

**Luz Selene González Delgado**

A thesis submitted in partial fulfilment of the requirements for the degree of  
Doctor of Philosophy

The University of Sheffield  
Faculty of Science  
Department of Molecular Biology and Biotechnology  
Firth Court, Western Bank, Sheffield S10 2TN

August 2020



## Abstract

A large number of cell surface proteins contain an SH3 type 5 domain (SH3\_5) responsible for the cell wall binding of these proteins to peptidoglycan. SH3\_5 domains are involved in the recognition of peptidoglycan in different pathogenic bacteria including methicillin-resistant *Staphylococcus aureus* (MRSA) and vancomycin-resistant *Enterococcus faecalis* (VRE). This work sought to elucidate the mechanisms underpinning the recognition of peptidoglycan by SH3\_5 proteins. As a model system, we focused on the C-terminal SH3\_5 domain from the lysostaphin enzyme. Lysostaphin is a bacteriolytic enzyme produced by *S. simulans* that cleaves the pentaglycine crossbridge in *S. aureus* peptidoglycan. This potent enzyme can efficiently lyse MRSA strains and disrupt biofilms. The lysostaphin enzyme contains an N-terminal catalytic domain with glycyglycyl endopeptidase activity and a C-terminal SH3\_5 cell wall targeting domain. Nuclear Magnetic Resonance (NMR) titrations, X ray crystallography, and mutational analyses, revealed how this SH3\_5 domain efficiently recognises *S. aureus* peptidoglycan. Our NMR titrations showed that the pentaglycine crossbridge and the peptide stem are recognized by two independent binding sites located on opposite sides of the SH3\_5 domain. The 1.4 Å crystal structure of the SH3\_5 domain, in complex with a synthetic peptidoglycan fragment, revealed a dual SH3\_5 binding mechanism whereby the G5 ligand from a monomeric fragment was recognized by one SH3\_5 domain and the P4 peptide was bound to another SH3\_5 protein. We propose that this binding mechanism leads to a synergistic and structurally dynamic binding, which induces clustering of the SH3\_5 domains and increases the local concentration of the enzyme resulting in its potent bacteriolytic activity. We propose that this research will serve as a framework for understanding how bacteria control the enzymatic activity of PG hydrolases and will contribute to the development of antimicrobial approaches.

## Publications arising from this work

**Gonzalez-Delgado**, L.S., Walters-Morgan, H., Salamaga, B., Robertson A.J., Hounslow, A.M., Jagielska, E., Sabała, I., Williamson, M.P., Lovering A.L., & Mesnage, S. Two-site recognition of *Staphylococcus aureus* peptidoglycan by lysostaphin SH3\_5. *Nat Chem Biol* **16**, 24–30 (2020). <https://doi.org/10.1038/s41589-019-0393-4>.

## Acknowledgements

I've been told that to start a project takes courage, but to finish it requires perseverance. It has been quite a journey, but it has been worth it. Firstly, I must thank my supervisor Stéphane Mesnage for his constant drive to teach me how to generate quality science, develop critical thinking, and somewhat, to challenge authority. I will always be very grateful to Stéphane for his great support during these four years of my PhD, for fighting for us during the publication of our manuscript, and for caring when we've gone through difficult situations outside work. I would also like to thank my co-supervisor Mike Williamson, whose patience and love for science allowed me to enter the NMR world. Mike is always eager to talk about science and always made our meetings interesting. Likewise, I thank Andrea Hounslow for her infinite patience and willingness to help me with my NMR experiments and analyses. I have been very fortunate to have such great people helping me during the development of my project.

To my friends, I'm so thankful for the shared laughter, which always helped me enjoy the good times and cope with the difficult ones. I met Rob Smith on the first day of our PhDs and that has been, perhaps, one of the best coincidences of my life. Rob is probably one of the best people I've ever met, and his friendship has been key to my mental health. I hope that, even when we take different paths, our friendship always lasts. Jessie, thank you so much for your friendship and all the support you've given me during this year. Although there is not enough space to thank everyone, but to MariCarmen, María, Ximena, Mario, Sophie, Ruby, Joe, Cindy, Hannah, Shauna, Bartek, Kasia, Nicola, all of you, whether it was during this time or earlier, are a constant reminder of how important it is to enjoy life and good times.

It is so important for me to thank Laurence for his patience, help, and great support over the past three years. Laurence has been there for me to celebrate the wins and to lighten the load during the most difficult times. His constant effort to make our home a 'little Mexico', to make me laugh, to understand me, to value my independence, and, to encourage me into believing in myself have been indispensable for the completion of my PhD. Thank you so much for keeping me sane and for showing me the value of little things in life every day.

I find it difficult to express my appreciation towards my family. Since I opened my eyes and during every stage of my life they've been by my side, encouraging me to pursue what they believe should be my main goal: to be happy! Even when we're so far apart, I always feel close to them and think about them all the time. The completion of this project is dedicated to all my family, most importantly, my parents and my sister. Finally, this thesis was completed during the abrupt loss of Héctor González, a lifelong companion whom I loved so much. However, as he would have liked to sing 'true love waits in haunted attics, and true love lives on lollipops and crisps, so just don't leave'.

## Abbreviations

~	Approximately
°C	Degree Celsius
3D	Three dimensional
aa-tRNA	Aminoacyl-transfer ribonucleic acids
ACT	Acetic acid
AFM	Atomic force microscopy
Ami	<i>N</i> -acetylmuramyl-L-alanine amidase
Amp	Ampicillin
anhAmi	1,6-anhydroamidases
APS	Ammonium persulphate
ATP	Adenosine triphosphate
AvPCP	<i>Anabaena variabilis</i>
Bac41	Bacteriocin 41
BHI	Brain heart infusion
Big_9	Bacterial immunoglobulin-like (Ig) 9 domain
Bp	base pair
CD spectroscopy	Circular dichroism spectroscopy
CDB	Choline binding domain
CE	Cell envelope
CHAP	Cysteine, histidine-dependent amidohydrolases/peptidases
ChBS	Choline binding sites
ChW	Clostridial hydrophobic repeat
CM	Cytoplasmic membrane
CMP	Chloramphenicol
CPases	Carboxypeptidases
CPS	Capsular polysaccharide

CSP	Chemical shift perturbation
CWB1	Cell wall binding repeat type 1
CWS	Cell wall sorting signal
D-Ala	D-alanine
D-Glu	D-glutamic acid
dH <sub>2</sub> O	Distilled water
DMSO	Dimethyl sulphoxide
DTT	Dithiothreito
EDTA	Ethylenediamine tetra-acetic acid
eGFP	Enhanced green fluorescent protein
ELISA	Enzyme-linked immunosorbent assay
EPA	Enterococcal polysaccharide antigen
EPase	Endopeptidase
Ery	Erythromycin
FRET	Fluorescence resonance energy transfer
GFP	Green fluorescent protein
Glc2-DAG	Diglucosyl-diacylglycerol
GlcNAc	<i>N</i> -acetylglucosamine
GlcNAases	<i>N</i> -acetylglucosaminidases
GroP	1,3-L- $\alpha$ - rlycerol-3-phosphate
Gtases	Glycosyltransferases
IMAC	Immobilized metal affinity chromatography
InIB	Internalin B
ITC	Isothermal titration calorimetry
IWZ	Inner wall zone
$K_d$	Equilibrium dissociation constant
$k_{off}$	Disassociation rate constant
$k_{on}$	Association rate constant
L-Ala	L-alanine

L-Lys	L-lysine
Lss	Lysostaphin
LTAs	Lipoteichoic acids
LysM	Lysin motif
m-A <sub>2</sub> pm	<i>meso</i> -diaminopimelic acid
MBP	Maltose binding protein
MLA	Malonic acid
mNG	mNeonGreen fluorescent protein
MurNAc	<i>N</i> -acetylmuramic acid
MurNAases	<i>N</i> -acetylmuramidases
MWM	Molecular weight marker
ngH <sub>2</sub> O	Nuclease free water
Ni-Nta	Nickel metal affinity
NMR	Nuclear Magnetic Resonance
NpPCP	<i>Nostoc punctiforme</i>
OD <sub>600</sub>	Optical density measure at 600 nm
ORF	Open reading frame
PASTA	Penicillin-binding protein and serine/threonine kinase associated domain
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PG	Peptidoglycan
PG <sub>50</sub>	Peptidoglycan amount required for 50% binding
ppm	parts per million
RboP	1,5- <i>D</i> -ribitol-phosphate
RhapWPs	Rhamnose polysaccharides
RO deionized water	Reverse osmosis and deionized water
RP-HPLC	Reverse-High-performance liquid chromatography
SDS	Sodium dodecyl sulphate



SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
Sec	General secretory pathway
SH3	Src Homology 3 domain
SP	Signal peptide
SPOR	Sporulation-related domain
STPKs	Serine/threonine protein kinases
TAE	Tris base, acetate and EDTA
Taq	Thermostable DNA polymerase derived from <i>Thermus aquaticus</i>
TAs	Teichoic acids
TEMED	Tetramethylethylenediamine
Tetracycline	Tet
TEV	Tobacco Etch Virus
TPases	Transpeptidases
Tris	Tris (hydroxymethyl) aminomethane
UDP	Uridine diphosphate
UV	Ultraviolet
v/v	Volume for volume
VRE	Vancomycin resistance enterococci
w/v	Weight for volume
WPs	Wall polysaccharides
WT	Wild type
WTAs	Wall teichoic acids
$\delta$	Chemical shift
$\Delta\delta$	Chemical shift changes
$\Delta\delta_{\max}$	Maximum chemical shift changes for 100% saturation
$\Delta\delta_{\text{obs}}$	Observed chemical shifts

## Table of contents

Abstract .....	i
Publications arising from this work.....	ii
Abbreviations.....	iv
List of figures .....	xiv
List of tables.....	xx
CHAPTER I .....	1
Introduction .....	1
1.1 The cell wall of Gram-positive bacteria .....	1
1.1.1 Peptidoglycan.....	1
1.1.1.1 Peptidoglycan composition .....	1
1.1.1.2 Peptidoglycan biosynthesis .....	5
1.1.2. Anionic polymers .....	15
1.1.2.1 Teichoic acids. ....	15
1.1.2.2 Wall polysaccharides.....	16
1.1.3 Architecture .....	17
1.2. Cell surface proteins of Gram-positive bacteria .....	18
1.2.1 Cell surface proteins covalently bound to PG.....	19
1.2.1.1 Cell wall sorting mechanism.....	19
1.2.2 Cell surface proteins non-covalently bound to PG .....	23
1.2.2.1 LysM domains .....	23
1.2.2.2 WxL domains.....	26
1.2.2.3 Sporulation-related domain (SPOR) .....	27
1.2.2.4 Clostridial hydrophobic repeat (ChW).....	30
1.2.2.5 Choline binding proteins (CBD) .....	31
1.2.2.6 PASTA domains .....	33
1.2.2.7 SH3 domains .....	36
1.2.2.7.1 Structural comparison of SH3 domains targeting bacterial cell walls.....	40
1.2.2.7.1.1 SH3_3 domains.....	40
1.2.2.7.1.2 SH3_5 domains.....	44
1.2.2.7.1.3 SH3_6 and SH3_7 domains.....	47
1.2.2.7.1.4 GW (SH3_8) domains .....	49
1.2.2.7.2 Species distribution and domain organization of the SH3_5 family .....	51

1.2.2.7.3 Structural similarities between SH3_5 domains.....	54
1.3 Project Aims and Objectives .....	60
CHAPTER II .....	63
Materials and Methods.....	63
2.1 Chemicals and enzymes .....	63
2.2 Buffers and solutions .....	63
2.2.1 Buffers for DNA analysis.....	63
2.2.1.1 TAE (50x)(Tris-acetate-EDTA) .....	63
2.2.1.2 DNA loading buffer (10x) .....	63
2.2.2 Protein purification solutions.....	64
2.2.2.1 SDS-PAGE solutions.....	64
2.2.2.1.1 SDS-PAGE separating buffer (4x) .....	64
2.2.2.1.2 SDS-PAGE stacking buffer (4x).....	64
2.2.2.1.3 SDS-PAGE loading buffer (5x) .....	64
2.2.2.1.4 Coomassie Blue stain.....	64
2.2.2.1.5 Coomassie destaining solution .....	65
2.2.2.2 Protein purification buffers for affinity chromatography (His-tag fused proteins).....	65
2.2.2.2.1 Buffer A (Equilibration buffer).....	65
2.2.2.2.2 Buffer B (Elution buffer) .....	65
2.2.2.3 Protein purification buffers for size-exclusion chromatography (SEC) .....	65
2.2.2.3.1 His-SH3_5 proteins (alone or fused to mNeonGreen) .....	65
2.2.2.3.2 Lysostaphin proteins .....	66
2.2.2.3.3 His-TEV-SH3_5 and untagged SH3_5 proteins (SEC, TEV-cleavage and reverse-IMAC buffer) .....	66
Tris-HCl 50 mM.....	66
2.2.3 HPLC Muropeptide analysis buffers .....	66
2.2.3.1 Phosphate buffer system .....	66
2.2.3.1.1 Buffer A (Mobile phase buffer).....	66
2.2.3.1.2 Buffer B (Elution buffer) .....	66
2.2.3.2 Water-Acetonitrile System .....	67
2.2.3.2.1 Solution A (Desalting solution).....	67
2.2.3.2.2 Solution B (Elution).....	67
2.2.3.3 Buffers for enzymatic digestion.....	67

2.2.3.3.1 Mutanolysin and EnpA <sub>c</sub> digestion buffer (20x) .....	67
2.2.3.3.1 Lysostaphin digestion- Phosphate buffered saline (PBS) .....	67
2.2.3.4 Sodium borate buffer for reduction of muopeptides .....	68
2.3 Media .....	68
2.3.1 Brain heart infusion (BHI) .....	68
2.3.2 BHI agar .....	68
2.3.3 NMR minimal media M9 .....	68
2.3.4 Overnight Express™ Instant TB Medium .....	70
2.4 Antibiotics .....	70
2.5 Bacterial strains and plasmids .....	71
2.5.1 Bacterial strains .....	71
2.5.2 Plasmids .....	71
2.5.3 Growth conditions of <i>E. coli</i> and <i>S. aureus</i> .....	72
2.6 Construction of recombinant plasmids for protein production .....	73
2.6.1 DNA manipulation .....	73
2.6.1.1 Primer design .....	73
2.6.1.2 PCR amplification .....	75
2.6.1.2.1 Phusion polymerase .....	75
2.6.1.2.2 Agarose gel electrophoresis .....	76
2.6.1.2.3 Gel extraction .....	76
2.6.1.2.4 DNA quantification .....	77
2.6.1.3 Restriction digestion cloning .....	77
2.6.1.3.1 Restriction Endonuclease Digestion of DNA .....	77
2.6.1.3.2 Ligation of DNA .....	77
2.6.1.4 Gibson Assembly of DNA Fragments .....	78
2.6.1.5 Site-directed mutagenesis .....	79
2.6.1.6 <i>E. coli</i> transformation .....	79
2.6.1.6.1 Colony PCR screening .....	79
2.6.1.6.2 Plasmid DNA extraction .....	80
2.6.1.6.3 DNA Sequencing .....	80
2.7 Protein analysis .....	80
2.7.1 Protein expression .....	80
2.7.2 Affinity chromatography .....	82

2.7.3.1 Protein quantification .....	82
2.7.4 SDS-PAGE .....	83
2.8 Purification of <i>S. aureus</i> PG muropeptides .....	84
2.8.1 <i>S. aureus</i> PG isolation.....	84
2.8.1.1 Production of muropeptides .....	84
2.8.1.2 $\beta$ -elimination (to generate lactyl-peptide) .....	85
2.8.2 HPLC analysis and fractionation .....	85
2.8.2.1 Muropeptide reduction.....	85
2.8.2.2 Muropeptide separation by RP-HPLC.....	86
2.9 Production of PG fragments by chemical synthesis .....	86
2.10 Crystallography .....	86
2.11 NMR experiments .....	87
2.12 PG binding assays.....	87
2.13 Lysostaphin spot on-lawn assays .....	88
CHAPTER III .....	89
Production and characterisation of recombinant Lss SH3_5 domains and peptidoglycan fragments .....	89
3.1 Aims and Objectives.....	89
3.2 Results.....	89
3.2.1 Expression and purification of the SH3_5 domain for interaction studies and X- ray crystallography.....	89
3.2.1.1 Unlabelled His-tagged-SH3_5.....	91
3.2.1.2 Singly and doubly labelled His-SH3_5 proteins .....	94
3.2.1.3 Unlabelled and untagged SH3_5 proteins.....	94
3.2.2 Purification of peptidoglycan fragments .....	99
3.2.2.1 Optimization of the conditions to purify <i>S. aureus</i> PG disaccharide dimers.....	102
3.2.2.1.1 Large-scale digestion to produce the dimer [GM-P5-G5-GM-P4-G5].....	106
3.2.3 Quantitative <sup>1</sup> H NMR analyses of produced <i>S. aureus</i> peptidoglycan fragments.....	108
3.2.4 Discussion .....	117
CHAPTER IV .....	119
Molecular basis for substrate recognition by the Lss SH3_5 domain.....	119
4.1 NMR titrations as a mean to characterise protein-ligand interactions. ....	119
4.2 Aims and Objectives.....	121
4.3 Results.....	122

4.3.1 Complete resonance assignment of the doubly labelled SH3_5 domain .....	122
4.3.2 Analysis of Lss SH3_5-PG interactions by NMR .....	123
4.3.2.1 SH3_5-pentaglycine [G5] binding.....	123
4.3.2.1.1 Pentaglycine orientation on the SH3_5 binding cleft .....	133
4.3.2.1.1.1 Binding to YGGGGG (YG5) .....	133
4.3.2.1.1.2 Binding to GGGGGY [G5Y] .....	133
4.3.2.2 SH3_5-tetrasaccharide [GM-GM] interaction .....	138
4.3.2.3 SH3_5-Lactyl(+) A $\gamma$ QKA [P4] binding .....	138
4.3.2.4 SH3_5-Lactyl(+)-A $\gamma$ QK[GGGGG]A binding .....	142
4.3.2.5 SH3_5-A $\gamma$ QK[GGGGG]AA-A $\gamma$ QKA binding.....	145
4.3.2.6 SH3_5- DS-A $\gamma$ QK[GGGGG]A-DS-A $\gamma$ QK[GGGGG]AA binding .....	148
4.3.2.7 SH3_5 binding to a complex mixture of uncrosslinked (linear) PG fragments .....	148
4.3.2.8 SH3_5 binding to PG disaccharide peptides with various degrees of cross-linking .....	151
4.3.2.9 SH3_5 binding to a mixture of glycan chains and peptide stems .....	156
4.3.3 High resolution crystal structure of the SH3_5 domain in complex with the [P4-G5] fragment .....	156
CHAPTER V .....	170
Mutational analyses on the binding activity of lysostaphin SH3_5 domains .....	170
5.1 Aims and objectives .....	170
5.2 Results.....	171
5.2.1 Expression and purification of SH3_5 and full-length Lss mutants.....	171
5.2.1.1 Singly labelled His-SH3_5 derivatives .....	171
5.2.1.2 SH3_5-mNeonGreen fusion proteins .....	174
5.2.1.2.1 Construction of pET2818-SH3_5-NG expressing a SH3_5-mNeonGreen fusion .....	174
5.2.1.2.2 Construction of pET2818-mNG expressing the mNeonGreen protein .....	178
5.2.1.3 Lss proteins containing point mutations in the SH3_5 domain .....	185
5.2.2 Contribution of six identified SH3_5 residues to the recognition of pentaglycine and the peptide stem.....	185
5.2.2.1 Binding of the N405A-SH3_5 mutant to the G5 peptide .....	186
5.2.2.2 Binding of the M453A-SH3_5 mutant to the G5 peptide .....	191
5.2.2.3 Binding of the Y472S-SH3_5 mutant to the G5 peptide .....	191
5.2.2.4 Binding of the I425A-SH3_5 mutant to the P4 peptide .....	191
5.2.2.5 Binding of the R427M-SH3_5 mutant to the P4 peptide.....	196
5.2.2.6 Binding of the W489L-SH3_5 mutant to the P4 peptide .....	196

5.2.3 Binding of SH3_5 derivatives to purified PG sacculi .....	203
5.2.4 Impact of SH3_5 single-site mutations on Lss activity.....	208
5.3 Discussion .....	210
CHAPTER VI .....	214
Production of <i>E. faecalis</i> SH3_5 proteins to explore PG binding mechanisms across species.....	214
6.1 Aims and objectives .....	215
6.2 Results.....	216
6.2.1 Expression and purification of five unlabelled BacL <sub>1</sub> SH3_5-His proteins.....	216
6.2.2 Expression of EF1293 SH3_5 domain .....	233
6.3 Perspectives .....	243
CHAPTER VII .....	245
General discussion.....	245
7.1 The SH3 domain as a model to investigate functional and structural diversification during evolution.....	245
7.2 Does SH3_5 domain organization reflect the diversity of binding mechanisms?.....	246
7.2.1 Modularity of proteins containing SH3_5 domains.....	246
7.2.2 Modularity of the SH3_5 domain .....	247
7.3 Lss SH3_5 binding affinities.....	248
7.4 Lss SH3_5 binding specificity.....	249
7.5 Lss SH3_5 binding mechanism .....	249
7.6 Is the catalytic domain the major determinant of the Lss specificity? .....	251
7.7 Future work.....	253
7.7.1 Substrate specificity across SH3_5 domains .....	253
7.7.2 What is the role of the modularity of SH3_5 domains? .....	254
7.7.3 Could other properties regulate peptidoglycan recognition by SH3_5 domains? .....	254
7.7.4 Harnessing SH3_5 domains to generate antimicrobials.....	254
References .....	257
Appendix A.....	279
Appendix B.....	287

## List of figures

### Chapter I

<b>Figure 1.1</b>	Schematic representation of a Gram-positive cell envelope (CE)	2
<b>Figure 1.2</b>	Disaccharide pentapeptide subunit (monomer) of most Gram-positive bacteria	4
<b>Figure 1.3</b>	Peptidoglycan synthesis in <i>S. aureus</i> .	6
<b>Figure 1.4</b>	Peptidoglycan cross-linking in <i>C. difficile</i>	13
<b>Figure 1.5</b>	Cleavage sites of the different PG hydrolases	14
<b>Figure 1.6</b>	Covalent anchoring of surface proteins by SrtA	22
<b>Figure 1.7</b>	Structure of <i>E. faecalis</i> LysM AtIA module bound to GlcNAc <sub>5</sub>	24
<b>Figure 1.8.</b>	Structural basis of denuded glycan recognition by <i>P. aeruginosa</i> RlpA SPOR domains	28
<b>Figure 1.9</b>	Structure of <i>S. pneumoniae</i> LytA CBD proteins in complex with choline residues	32
<b>Figure 1.10</b>	Model of <i>M. tuberculosis</i> PknB activation promoted by PASTA-PG binding	35
<b>Figure 1.11.</b>	Structure and topology of eukaryotic and bacterial SH3 domains	39
<b>Figure 1.12</b>	Multiple amino acid sequence alignment between cell wall targeting SH3 bacterial domains (SH3-3,5,6,7,8) with a eukaryotic SH3 domain (c-Crk)	41
<b>Figure 1.13</b>	Structural similarities of CW binding SH3 domain families	42
<b>Figure 1.14.</b>	Hydrophobicity and surface charge of cell wall binding SH3 domain families.	45
<b>Figure 1.15</b>	Distribution of SH3_5 sequences across species	52
<b>Figure 1.16</b>	Modular organization of proteins containing SH3_5 domains across species	53
<b>Figure 1.17</b>	Sequence and structural similarities within four SH3 type 5 domains	55



<b>Figure 1.18</b>	Lysostaphin SH3_5 protein bound to a pentaglycine peptide (PDB 5LEO).	59
<b>Chapter III</b>		
<b>Figure 3.1</b>	Description of available Lss-SH3_5 constructs used for interaction studies and X-crystallography	90
<b>Figure 3.2</b>	Production, purification, and binding activity of recombinant Lss-SH3_5 protein	92
<b>Figure 3.3</b>	Production and purification of <sup>15</sup> N- and <sup>15</sup> N/ <sup>13</sup> C-SH3_5 labelled proteins	95
<b>Figure 3.4.</b>	Production and purification of the recombinant His-TEV-SH3_5 protein	96
<b>Figure 3.5</b>	Optimisation of the conditions for the TEV digestion of recombinant His-TEV-SH3_5	97
<b>Figure 3.6</b>	Purification of recombinant untagged SH3_5 protein for X-ray crystallography	98
<b>Figure 3.7</b>	<i>S. aureus</i> PG fragments used as SH3_5 ligands	100
<b>Figure 3.8</b>	RP-HPLC analysis of available lactyl (+) P4-G5 ligand	101
<b>Figure 3.9</b>	RP-HPLC analysis of available <i>S. aureus</i> PG complex mixtures	103
<b>Figure 3.10</b>	Muropeptide profile of <i>S. aureus</i> SH1000 PG obtained by RP-HPLC	104
<b>Figure 3.11</b>	Analysis of <i>S. aureus</i> PG digestion by mutanolysin	105
<b>Figure 3.12</b>	Analysis of small-scale mutanolysin-digestion assay to obtain the <i>S. aureus</i> PG dimer	107
<b>Figure 3.13</b>	RP-HPLC elution profiles to obtain the [GM-P5-G5-GM-P4-G5] PG dimer	109
<b>Figure 3.14</b>	Characterisation of the DS-dimer [GM-P5-G5-GM-P4-G5]	111
<b>Figure 3.15</b>	<sup>1</sup> H NMR analysis of the purified <i>S. aureus</i> S1000 [GM-P5-G5-GM-P4-G5] dimer	114

<b>Figure 3.16</b>	$^1\text{H}$ NMR spectra of pure <i>S. aureus</i> derived fragments used as SH3_5 ligands	115
<b>Chapter IV</b>		
<b>Figure 4.1</b>	Fully assigned $^1\text{H}$ - $^{15}\text{N}$ HSQC spectrum of the Lss SH3_5 apo domain (residues 402-493)	124
<b>Figure 4.2</b>	Mapping of SH3_5 residues interacting with the G5 peptide by NMR titrations	126
<b>Figure 4.3</b>	$^1\text{H}$ and $^{15}\text{N}$ chemical shifts from all SH3_5 residues upon titration of G5 peptide.	130
<b>Figure 4.4</b>	Calculated saturation curves for determining SH3_5-G5 binding affinities	131
<b>Figure 4.5</b>	Mapping of SH3_5 residues interacting with the YG5 peptide by NMR titrations	134
<b>Figure 4.6</b>	Mapping of SH3_5 residues interacting with the G5Y peptide by NMR titrations	136
<b>Figure 4.7</b>	SH3_5-tetrasaccharide [GM-GM] NMR titration.	139
<b>Figure 4.8</b>	Mapping of SH3_5 residues interacting with the P4 peptide by NMR titrations.	140
<b>Figure 4.9</b>	Mapping of SH3_5 residues interacting with the P4-G5 peptide by NMR titrations	143
<b>Figure 4.10</b>	Mapping of SH3_5 residues interacting with the P5-G5-P4 peptide by NMR titrations.	146
<b>Figure 4.11</b>	Mapping of SH3_5 residues interacting with the GM-P5-G5-MG-P4-G5 dimer by NMR titrations	149
<b>Figure 4.12</b>	Mapping of SH3_5 residues interacting with <i>S. aureus</i> linear PG fragments by NMR titrations	152

<b>Figure 4.13</b>	Mapping of SH3_5 residues interacting with <i>S. aureus</i> disaccharide peptides by NMR titrations	154
<b>Figure 4.14</b>	Mapping of SH3_5 residues interacting with <i>S. aureus</i> linear glycan chains and peptide stems by NMR titrations	157
<b>Figure 4.15</b>	Mapping of the SH3_5 protein surface interacting with distinct <i>S. aureus</i> PG fragments	160
<b>Figure 4.16</b>	2.5Å crystal structure of the Lss-SH3_5 domain forming Zn <sup>2+</sup> -bound heptamers	162
<b>Figure 4.17</b>	Crystallisation of Lss-SH3_5 proteins in complex with the P4-G5 ligand	163
<b>Figure 4.18</b>	Crystal structure of Lss-SH3_5 in complex with the P4-G5 ligand	167
 <b>Chapter V</b>		
<b>Figure 5.1</b>	Description of six SH3_5 single-site substitution mutant domains analysed by NMR	172
<b>Figure 5.2</b>	SDS-PAGE analysis of singly labelled SH3_5 derivatives used for NMR studies	173
<b>Figure 5.3</b>	Construction of pET-SH3_5-mNG fusion	175
<b>Figure 5.4</b>	Production and purification of the SH3_5mNeonGreen fusion	177
<b>Figure 5.5</b>	Construction of pET-2818-mNG.	179
<b>Figure 5.6</b>	Production and purification of mNeonGreen fluorescent protein	182
<b>Figure 5.7</b>	Location of the sixteen mutagenized residues at the SH3_5:P4-G5 interaction interface	183
<b>Figure 5.8</b>	SDS-PAGE analysis of the SH3_5-mNG derivative fusion proteins used for <i>in vitro</i> binding assays	185
<b>Figure 5.9</b>	Production and purification of full-length Lss proteins used to assess the impact of SH3_5 mutations on the enzyme activity	187
<b>Figure 5.10</b>	Residual binding of the N405A mutant to the G5 peptide	189
<b>Figure 5.11</b>	Residual binding of the M453A mutant to the G5 peptide	192

<b>Figure 5.12</b>	Residual binding of the Y472S mutant to the G5 peptide	194
<b>Figure 5.13</b>	Residual binding of the I425A mutant to the P4 peptide	197
<b>Figure 5.14</b>	Residual binding of the R427M mutant to the P4 peptide	199
<b>Figure 5.15</b>	Residual binding of the W489L mutant to the P4 peptide	201
<b>Figure 5.16</b>	Binding activity of the SH3_5-mNeonGreen fusion protein to WT <i>S. aureus</i> PG and <i>fem</i> mutants	204
<b>Figure 5.17</b>	Binding activity of the SH3_5-mNeonGreen derivative proteins to purified <i>S. aureus</i> PG	206
<b>Figure 5.18</b>	Enzymatic activity of Lss recombinant proteins harbouring mutations in the SH3_5 domain	209
<b>Figure 5.19</b>	Comparison of ligand-binding pocket to other SH3_5 structures and SH3 superfamily members	212
 <b>Chapter VI</b>		
<b>Figure 6.1</b>	PCR amplification of <i>E. faecalis</i> BacL <sub>1</sub> SH3 modules from pET22b-BacL <sub>1</sub> SH3	217
<b>Figure 6.2</b>	Amino acid sequence alignment of the three BacL <sub>1</sub> C-terminal SH3_5 repeats	219
<b>Figure 6.3</b>	Construction of pET2818-S2	221
<b>Figure 6.4</b>	Construction of pET2818-S2	223
<b>Figure 6.5</b>	Construction of pET2818-S3	225
<b>Figure 6.6</b>	Construction of pET2818-S1S2	227
<b>Figure 6.7</b>	Construction of pET2818-S1S3	229
<b>Figure 6.8</b>	Production and purification of BacL <sub>1</sub> SH3-5 domains	231
<b>Figure 6.9</b>	Amino acid sequence alignment between the C-terminal region of the <i>E. faecalis</i> EF1293 protein, the <i>E. faecalis</i> N-terminal BacL <sub>1</sub> SH3 domain, and the <i>S. aureus</i> Lss SH3_5 domain	234
<b>Figure 6.10</b>	Selection of <i>E. faecalis</i> EF1293 SH3_5 domain candidates	235
<b>Figure 6.11</b>	Construction of pET2818-EF1293 expression plasmids	237

<b>Figure 6.12</b>	PCR confirmation of pET2818 SH3_5 EF1293 (1-3) constructs	239
<b>Figure 6.13</b>	SDS-PAGE analysis of the expression level and solubility of the EF12931 and EF12932 recombinant proteins	241
<b>Chapter VII</b>		
<b>Figure 7.1</b>	Zymogram analysis of the lytic activity of EnpA <sub>c</sub> and EnpA <sub>c</sub> -SH3_5 fusion against <i>S. aureus</i> peptidoglycan	256
<b>Appendix A</b>		
<b>Figure 1A</b>	LC-MS of tetrasaccharide [GM-GM] ligand	280
<b>Figure 2A</b>	Characterisation of pentaglycine [G5] ligand	281
<b>Figure 3A</b>	Characterisation of pentaglycine derivative with an extra tyrosine residue at the N-terminus [YG5] ligand	282
<b>Figure 4A</b>	Characterisation of pentaglycine derivative with an extra tyrosine residue at the C-terminus [G5Y] ligand	283
<b>Figure 5A</b>	Characterisation of tetrapeptide stem with a lactyl group [P4] ligand	284
<b>Figure 6A</b>	Characterisation of tetrapeptide stem with no lactyl group linked to a pentaglycine peptide as a lateral chain [P4-G5] ligand	285
<b>Figure 7A</b>	Characterisation of pentapeptide stem crosslinked to a tetrapeptide stem through a pentaglycine bridge [P5-G5-P4] ligand	286

## List of tables

### Chapter I

<b>Table 1.1</b>	Domains involved in protein attachment to the Gram-positive cell wall	20
<b>Table 1.2</b>	Lineage distribution of the SH3 domain families	37

### Chapter II

<b>Table 2.1</b>	Antibiotic stock solutions and working concentrations	70
<b>Table 2.2</b>	Bacterial strains used in this study	71
<b>Table 2.3</b>	Plasmids used in this study	72
<b>Table 2.4</b>	Primers used in this study. Underlined are sites recognised by restriction enzymes	73
<b>Table 2.5</b>	Mass of insert used in each ligation reaction	78

### Chapter III

<b>Table 3.1.</b>	Final concentration of <i>S. aureus</i> -derived PG ligands determined by NMR	113
-------------------	---	-----

### Chapter IV

<b>Table 4.1</b>	CSPs associated with amino acid side chains	128
<b>Table 4.2</b>	Data collection and refinement statistics of the Lss SH3_5: P4-G5 co-crystal (PDB 6RK4).	166

# CHAPTER I

## Introduction

### 1.1 The cell wall of Gram-positive bacteria

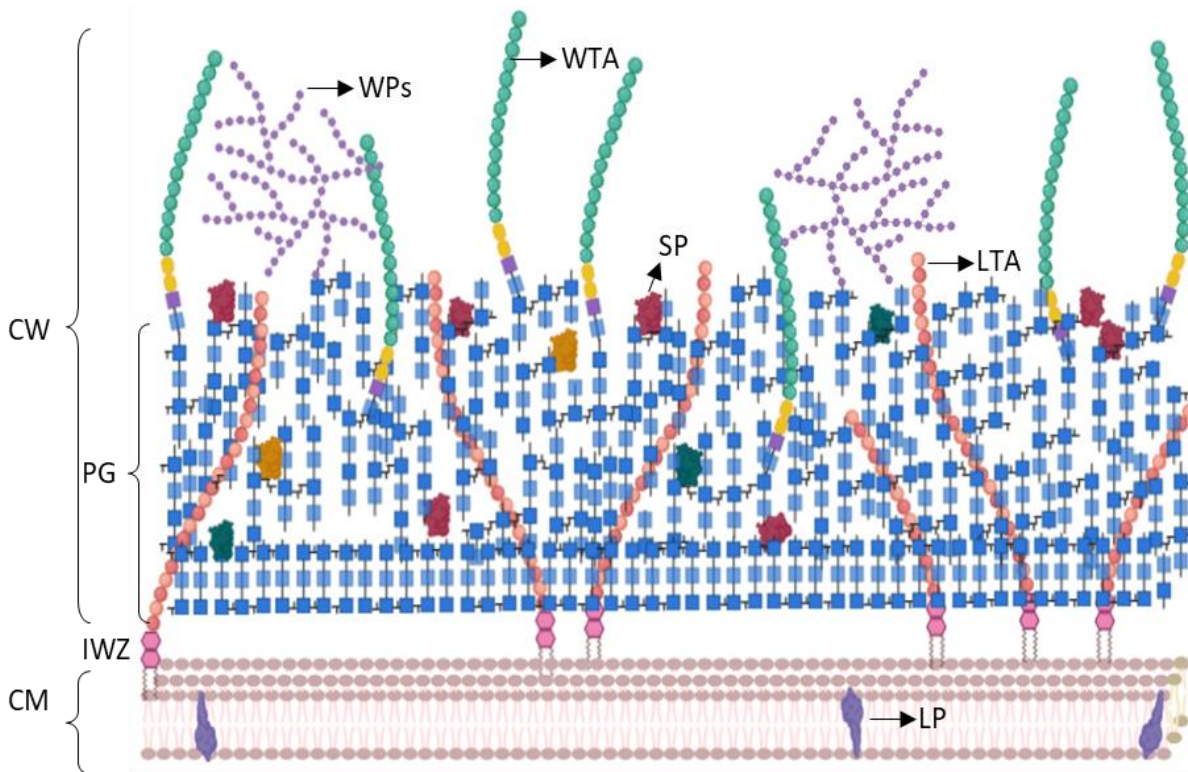
Gram-positive bacteria lack an outer membrane and their cell wall constitutes the outermost layer of the cell envelope (Rogers, 1980; Vollmer, 2008a). As a result, it plays a pivotal role in their interaction with the environment. The cell wall is a complex structure made of peptidoglycan (PG) and covalently attached anionic polymers such as teichoic acids (TAs), polysaccharides, and surface proteins (Fig. 1.1) (Dramsı S., 2008; Neuhaus, 2003; Vollmer, 2008a).

#### 1.1.1 Peptidoglycan

Peptidoglycan (PG) is an essential element of the bacterial cell wall ensuring cell shape and viability by withstanding the turgor pressure (Rogers, 1980; Turner, 2014; Vollmer, 2008a). PG is intimately involved in cell growth and division, as well as protective, physiological, and adaptative processes. In Gram-positive bacteria, it serves as a scaffold for both the covalent and non-covalent anchoring of polymers and surface proteins (Desvaux, 2018; Dramsı S., 2008; Shockman, 1983).

##### 1.1.1.1 Peptidoglycan composition

PG is a porous net-like macromolecule made of linear glycan chains crosslinked by short peptides. The glycan chains are formed of repeating disaccharide units of *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc) linked by  $\beta$  1-4 bonds (Rogers, 1980). The D-lactoyl group of each MurNAc is substituted by a short peptide stem constituted by amino acids with L- and D-configurations, most often: L-Alanine,  $\gamma$ D-Glutamate or iGlutamine, L-Lysine or *-meso*-Diaminopimelic acid (m-A<sub>2</sub>pm), D-Alanine and D-Alanine (Fig.1.2) (Schleifer, 1972; van Heijenoort, 2001)



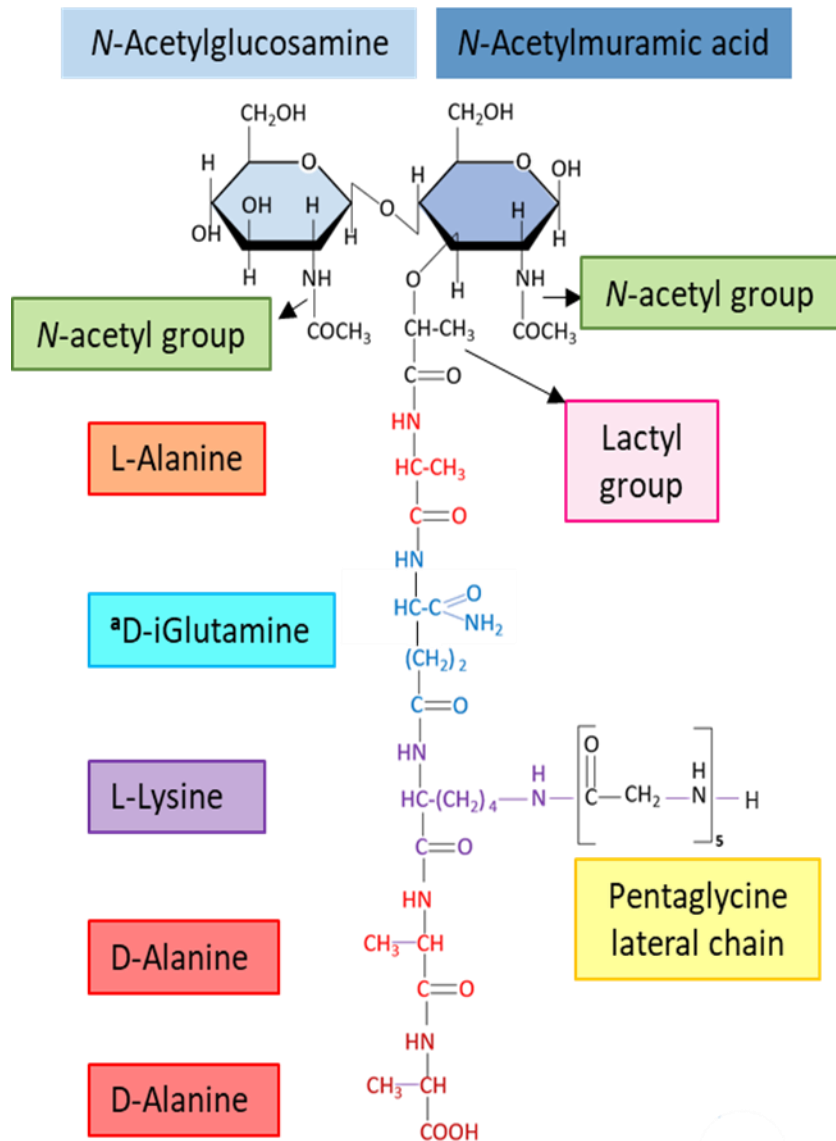
**Figure 1.1. Schematic representation of a Gram-positive cell envelope (CE).** The cell envelope of Gram-positive bacteria is constituted of the cytoplasmic membrane (CM) acting as a selective permeable barrier and the cell wall ensuring cell shape and viability. Lipoproteins in Gram-positive bacteria are anchored into the outer leaflet of the CM. The inner wall zone (IWZ) is a narrower zone of low density and devoid of cross-linked PG, the IWZ is not bound by two biological membranes like the case of Gram-negative bacteria and therefore is not considered as a periplasm although it presents some analogies (Desvaux, 2018). The cell wall is constituted of peptidoglycan (PG) which serves as a scaffold for the anchor of wall teichoic acids (WTA); wall polysaccharides (WPs); and surface proteins (SP). Lipoteichoic acids (LTA) are anchored to the CM and protrude from the cell wall. Adapted from (Silhavy, 2010). Created by BioRender.com.



The composition of the disaccharide glycan strands is highly conserved. However, the majority of bacterial species have developed modifications to their glycan units which are critical for cell shape, division, and pathogenesis (Vollmer, 2008b). In most bacterial species, glycan units become modified after insertion into the assembled cell wall. In Gram-positive bacteria the most common modifications are: (i) *N*-deacetylation of GlcNAc and MurNAc (Boneca, 2007; Vollmer, 2000), (ii) *N*-glycolylation of MurNAc (Raymond, 2005), (iii) *O*-acetylation of MurNAc and GlcNAc (Clarke, 2007; Sychantha, 2018), and (iv) attachments of surface polymers such as wall teichoic acids (WTA) and other highly diverse heteropolysaccharides like capsular and rhamnose polysaccharides (RhapWPs) (Brown, 2013; Deng L, 2000; Vollmer, 2008b). These alterations have been implicated in bacterial resistance to  $\beta$ -lactam antibiotics and host defence factors as in the case of lysozyme hydrolysis (Brown, 2013; Smith, 2019; Vollmer, 2008b).

Another important modification of the glycan strands is their average chain length (glycan chain length - GCL) which varies between species. Noticeably the GCL does not correlate with the thickness of the PG layer (Vollmer, 2008a). *Bacillus subtilis* was estimated to have long glycan chains of over 100 disaccharide units (Matias, 2005), whereas *Staphylococcus aureus* has a GCL of between 3-10 disaccharide units. A model across bacteria has been proposed in which initial PG is formed of dense stiff regions that undergo remodelling by PG hydrolysis. This allows the enlargement of the cell surface area with a more flexible PG. This model is independent of the specific mechanism of monomer insertion and mode of hydrolysis (Wheeler, 2011)

The stem peptides of PG can also be modified. The N-terminal L-Ala amino acid is conserved in most bacteria, however in the *Corynebacterium* and *Mycobacterium* genera Gly or L-Ser substitutions have been described (Hesse L, 2003; Mahapatra, 2000; van Heijenoort, 2001). In Gram-positive bacteria the  $\alpha$ -carboxyl of the D-glutamic acid is often amidated (D-iGln) and L-Lysine is often present in position 3, the exceptions being *Mycobacterium*, *Bacillus*, and *Clostridium* genera where A<sub>2</sub>pm is found instead (van Heijenoort, 2001; Vollmer, 2008a). The two C-terminal D-Alanine residues are highly conserved. D-Ala at position four is present in all bacteria. In vancomycin-resistant Enterococci strains the D-Ala at position five is substituted by a D-Lactate (D-Lac) or a D-Serine (D-Ser) (Healy VL, 2000).



**Figure 1.2. Disaccharide pentapeptide subunit (monomer) of most Gram-positive bacteria.** The PG building blocks consist of disaccharide units of alternating  $\beta$ -1-4 linked *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc) residues. The pentapeptide stem is covalently linked via a lactyl group to MurNAc. In most Gram-positive bacteria the epsilon-amino group of the side chain of the third amino acid is attached to a lateral chain of variable length and composition. In *S. aureus* the lateral chain is comprised of a pentaglycine peptide chain.

In most Gram-positives including pathogenic cocci, the  $\epsilon$ -NH<sub>2</sub> of the L-Lysine is substituted by a lateral chain of variable length and composition (Fig. 1.2). Different L- and D- amino acids have been identified: Gly, L-Ala, L- or D-Ser, D-Asx, L- or D-Glu, among others. For instance, in *S. aureus* the lateral chain is made of five glycines, whereas in *Enterococcus faecalis* it is composed of two L-Ala residues (Boniface A, 2006; Schleifer, 1972; Vollmer, 2008a).

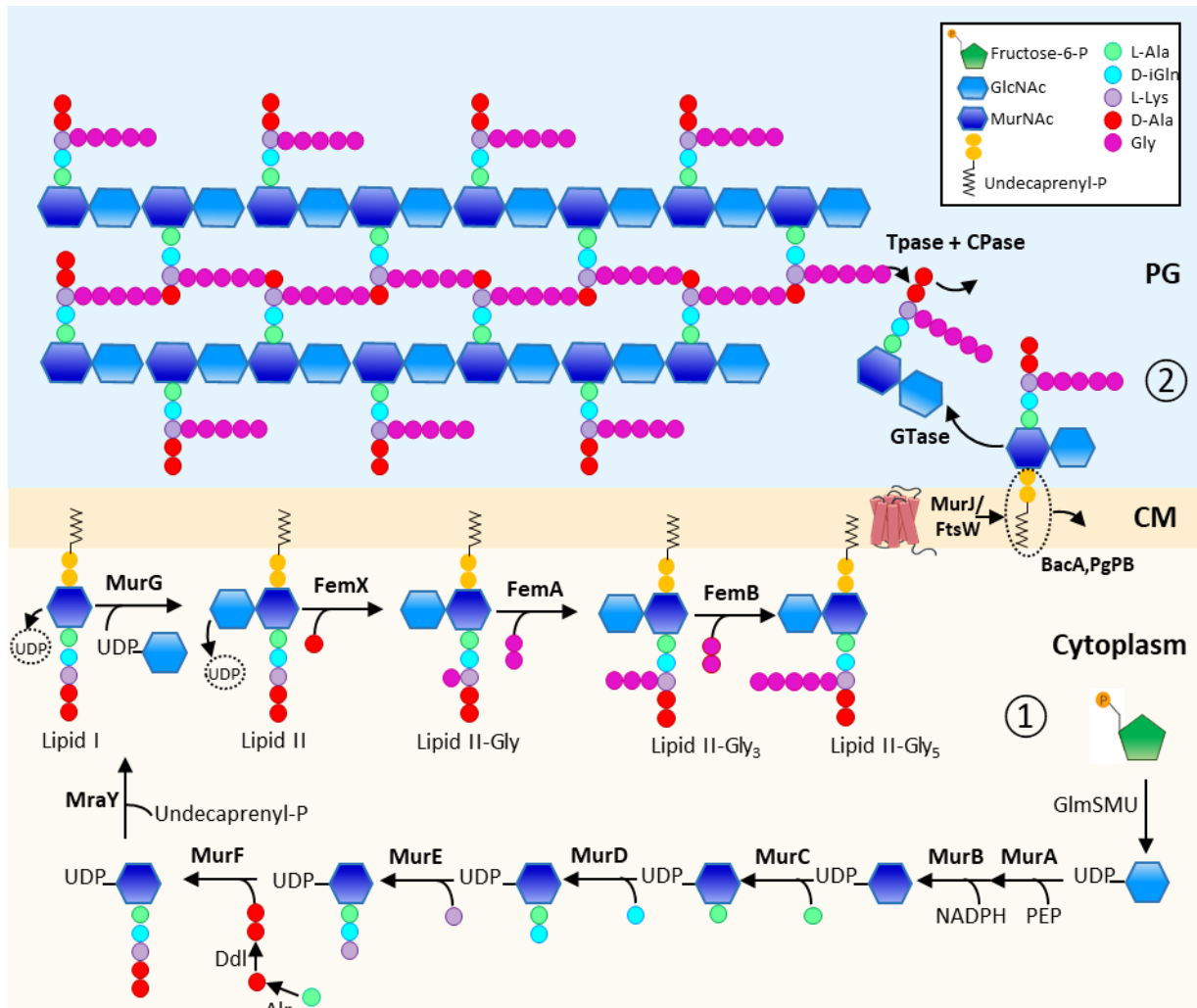
#### **1.1.1.2 Peptidoglycan biosynthesis**

The study of PG biosynthesis started nearly 70 years ago (Park, 1952) and a general overview of the biochemical pathway has been generated after the study of its different steps in various Eubacteria (van Heijenoort, 2001). PG biosynthesis is generally divided into a two-stage process: (i) assembly of the monomer unit and (ii) polymerization of the lipid II substrate (Fig.1.3). The assembly of the monomer unit can be further divided into a cytoplasmic, and cell membrane stage.

The characteristic features of the PG building blocks highlight the specificity and complexity involved during each step of the first-stage process. For instance, the unusual presence of muramic acid, the formation of a  $\gamma$ -linkage between D-iGln and a diamino acid (instead of the usual  $\alpha$ -carboxylic group), and the alternating L- and D- amino acid conformations.

Monomer assembly can be divided into four sequential reactions: (i) formation of UDP- GlcNAc, (ii) formation of UDP- MurNAc, (iii) formation of UDP- MurNAc -pentapeptide, and (iv) formation of the lipid intermediates. Steps (i) to (iii) occur in the cytoplasm whilst step (iv) occurs at the cell membrane.

The first step involves the conversion of fructose-6-P into UDP- GlcNAc by the GmIS/M/U transferases. The amidotransferase GmS is responsible for the transformation of fructose-6-P to an intermediate glucosamine-6-P; this step is followed by the interconversion of glucosamine-1-P by the phosphoglucosamine mutase GmM. During the last step, the bifunctional GmU synthase transforms glucosamine-1-P into UDP-GlcNAc by two consecutive reactions: first the GmU C-terminal domain acetylates the glucosamine-1-P molecule into



**Figure 1.3. Peptidoglycan synthesis in *S. aureus*.** The two-stage pathway is illustrated: ① assembly of the monomer unit and ② polymerization of the PG precursor lipid II. Cytosolic and membrane-associated enzymes lead to the synthesis of the lipid II-Gly<sub>5</sub>. Four successive groups of reactions are considered: formation of UDP-*N*-acetylglucosamine (UDP-GlcNAc) by GlmS, GlmM, and GlmU; formation of UDP-*N*-acetylmuramic acid by MurA and MurB; formation of the UDP-MurNAc-pentapeptides by MurC to MurF and Alr/Ddl; and formation of the lipid intermediates by MraY and MurG. The FemX transferase catalyses the addition of the first Glycyl residue of the lateral chain; the second and third glycylic residues are added by FemA and the fourth and fifth by FemB. Lipid II is flipped across the cytoplasmic to the outer leaflet in the inner wall zone (IWZ) by a flippase (MurJ/FtsW), where it is used as a substrate by glycosyltransferases

(Gtases) and DD-transpeptidases (DD-TPases). Peptides are trimmed by DD- and LD-carboxypeptidases (CPases). Adapted from (Egan, 2020; Typas, 2011).

*N*-acetylglucosamine-1-P, and finally its N-terminal domain catalyses the uridylation of UDP-GlcNAc. In Gram-positive bacteria, UDP-GlcNAc is an essential precursor for PG synthesis and GlcNAc containing polymers like teichoic acids and polysaccharides (Fig. 1.3) (Pooley, 1994; van Heijenoort, 2001).

The transformation of UDP-GlcNAc to UDP-MurNAc is a two-step process catalysed by the MurA and MurB enzymes. First, the transferase MurA places enolpyruvate from phosphoenolpyruvate (PEP) in position 3 of the GlcNAc residue yielding a UDP-GlcNAc-enolpyruvate intermediate. Next, the MurB reductase catalyses (in an NADPH and FAD-dependent reaction) the reduction of the enolpyruvate moiety into D-Lactoyl, to form the UDP-MurNAc precursor (Benson TE, 2001; Brown ED, 1995; Farmer, 1996).

The assembly of the UDP-MurNAc-pentapeptide is carried out by the consecutive addition of L-Ala, D-Glu, L-Lys or *m*-A<sub>2</sub>pm, and the D-Ala dipeptide onto the D-lactoyl group of MurNAc by the Mur C-F synthases. Mur C to F synthases are a family of highly specific cytoplasmic ADP/peptide-forming ligases (van Heijenoort, 2001). They all share a similar catalytic mechanism, driven by the hydrolysis of ATP into ADP, which promotes the formation of a peptide bond, and the loss of inorganic phosphate. The MurC ligase adds the first amino acid into the peptide stem. As mentioned in previous sections, MurC displays a preferential binding to L-Ala, and in rare cases Gly or L-Ser have been identified (Emanuele, 1996; Mahapatra, 2000). In *Mycobacterium* spp the substitution of Gly for L-Ala has been associated to specific growth conditions (Mahapatra, 2000). The next amino acid at position two is added by the MurD ligase. MurD has proven to be highly specific, since only D-Glu has been identified as the substrate for this enzyme. As such, most amino acid variations occur at the third position. Depending on the organisms L-Lys or *m*-A<sub>2</sub>pm are commonly found. However, other residues such as ornithine, diaminobutyric acid, homoserine, lanthionine, or 3-hydroxy-A<sub>2</sub>pm have been identified (Bertrand, 1997, 2000). Although MurE accepts a larger range of substrates, it has been shown that in most cases it can efficiently discriminate between different amino acids, and catalyse the addition of a specific residue into the peptide stem (Boniface A, 2006; Gordon E., 2001; van Heijenoort, 2001). The alanine racemase Alr catalyses the conversion of L-Ala to D-Ala which is then polymerised into

the dipeptide D-Ala-D-Ala, by the Ddl ligases (Fig. 1.3). This dipeptide is then added to the UDP-MurNAc-tripeptide by MurF. The specificity of MurF towards the D-Ala-D-Ala substrate has been studied using various analogues in different *in vivo* systems (Duncan K., 1990; Vollmer, 2008a). Although D-Ala is predominantly found at the fourth position in all bacterial species, the fifth D-Ala can be replaced by D-Lac or D-Ser which has little impact in transpeptidation reactions as observed in Vancomycin-resistant Gram-positive organisms (Bugg, 1991).

The final stage of assembly for the monomeric unit is the formation of the lipid intermediates I and II (Fig. 1.3). For these steps, the transferases MraY and MurG are involved. MraY is a transferase or translocase that catalyses the transfer of phospho-MurNAc-pentapeptide from the UDP-MurNAc-pentapeptide to the membrane acceptor-undecaprenyl phosphate resulting in the formation of the lipid I intermediate (MurNAc-pentapeptide-pyrophosphoryl undecaprenol). Subsequently, MurG, a *N*-acetylglucosamine transferase, catalyses the addition of UDP-GlcNAc into Lipid I, yielding the GlcNAc-MurNAc-pentapeptide-pyrophosphoryl undecaprenol (Lipid II) (Ikeda, 1991; van den Brink-van der Laan, 2003; van Heijenoort, 2001).

In most Gram-positive bacteria cross-linking between PG peptide stems occurs via a lateral peptide chain. Two enzyme families responsible for the 'branching' of interpeptide bridges have been described: the Fem transferases (for L-amino acids and glycine) and the ATP-grasp family enzymes (for D-amino acids) (Galperin, 1997; Mainardi, 2008; Vollmer, 2008a). Fem transferases are a family of non-ribosomal peptide bond-forming enzymes involved in the branching of aminoacyl-transfer ribonucleic acids (aa-tRNA). The aa-tRNAs are amino acids delivered to the ribosome for translation and they also participate as substrates for cell wall synthesis (Vollmer, 2008a). Gly and L-amino acids are activated as aa-tRNA and their amino acid moiety is transferred by the Fem transferases to the third amino acid in the pentapeptide stem of the lipid precursors to form branched peptide chains that will be linked during PG polymerization. The peptide bridge composition, and the precursor used by these enzymes, varies between bacterial species. It has been shown that lipid II is the precursor substrate for *S. aureus* Fem enzymes, whereas UDP-MurNAc-pentapeptide and lipid II are both substrates for *E. faecalis* Fem transferases, which are also known as BppA1 and BppA2 (Bouhss, 2002; Schneider, 2004). D-amino acids are

incorporated by the ATP-grasp family. These enzymes activate the amino acids as acyl phosphates at the expense of an ATP molecule and catalyse the ligation of the activated carboxyl group to different chemical groups, in this case an amino nitrogen (Aliashkevich, 2018; Bellais, 2006; Veiga, 2006).

The second stage in PG synthesis is the polymerization of the monomeric units, which requires the translocation of the lipid II located at the inner leaflet of the cytoplasmic membrane to the outer leaflet where is utilized as a substrate by PG synthases (Fig. 1.3). Owing to the recent development of dedicated enzymatic assays, the details of the PG synthesis regulation are being uncovered, particularly for the second stage of the pathway (Caveney NA., 2018; Egan, 2020).

The identity of the protein responsible for transporting or 'flipping' lipid II from the inner to the outer leaflet has remained elusive for decades. The coordination of proteins from two integral membrane protein families, FtsW and RodA from SEDS (shape, elongation, division and sporulation), and MurJ member of the MOP (multidrug, oligosaccharide-lipid, polysaccharide) transporter family have been suggested to fulfil this role (Egan, 2020; Liu, 2018). RodA and FtsW are two essential proteins for PG growth during elongation and division, respectively. The model rod-shaped *B. subtilis* possesses a third SEDS protein, SpoVE, involved in spore cortex synthesis. FtsW was shown to be responsible for transporting lipid II across a synthetic membrane when incorporated into proteoliposomes (Mohammadi T, 2011). From this, the authors assumed that the FtsW homologues, RodA and SpoVE, are also likely to participate in the translocation of lipid II during cell elongation and spore PG synthesis in *B. subtilis*. Moreover, no *in vitro* flippase activity was detected for MurJ in their synthetic system. However, neither FtsW nor RodA were implicated in the transport of lipid II in an *in vivo* assay in which colicin M (ColM) was added to cells to digest only the flipped (periplasmic) lipid II of *E. coli* (Sham, 2014). The protection of lipid II from ColM cleavage upon MurJ inactivation suggested that either lipid II is not flipped or that inhibiting MurJ somehow interferes with ColM import or activity. The results in both studies are not necessarily in opposition since key caveats, described below, remain for each set (Egan, 2020).



Recently, a high-resolution structure of MurJ was determined revealing different inward-facing and outward-facing conformations (Kuk, 2017). This alternate access configuration and the movement of the MurJ cargo seemed to be driven by the proton motive force (PMF) and the binding and release of Na<sup>+</sup> (Kuk, 2017; Kumar, 2019). For the *in vitro* experiment the absence of PMF in the artificial membranes could explain the lack of flipping activity by the MurJ proteins (Mohammadi T, 2011). For the *in vivo* study where FtsW and RodA did not appear to contribute to flipping, ColM might be a biased tool to explore this process as it inhibits PG biosynthesis by interfering with lipid carrier recycling (Harkness, 1988). Whether one of these proteins has a more direct involvement in the flipping activity of the lipid II, like the MurJ transition from an inward-open conformation to an outward-open conformation caused by substrate binding, it is likely that both MurJ and FtsW are required for transportation to occur. It has been suggested that the recruitment of MurJ at mid-cell during division requires both the presence of the precursor lipid II and functional FtsW, and that this cooperative activity might be necessary to fulfil the high demand for PG synthesis (Liu, 2018).

PG polymerisation is carried out by membrane-associated enzymes called Penicillin binding proteins (PBPs) (Goffin, 1998). PBPs can be divided into three groups: (i) Class A PBPs, which are bifunctional enzymes acting as glycosyltransferases (GTases) and DD-transpeptidases (TPases) for the polymerisation of glycan strands and cross-linking the stem peptides, (ii) Class B PBPs with monofunctional TPase activity, and (iii) Class C PBPs, a group of low molecular weight carboxypeptidases (Sauvage, 2008; Vollmer, 2008a). One of the difficulties in studying PG synthesis is the large number of PBPs with apparent redundancy: for example in both *S. pneumoniae* and *E. coli*, class A PBPs (PBP1A and PBP1B) have been described as partially redundant, as the bacterial cell only requires one of these for growth (Egan, 2020; Hoskins, 1999; Yousif, 1985). However, *S. aureus* PBP2 is the only bi-functional PBP (Pinho, 2001) and in *E. faecalis* all class A PBPs are dispensable, with deletions of all three enterococcal class A PBPs having a minor effects in PG cross-linking (Arbeloa, 2004).

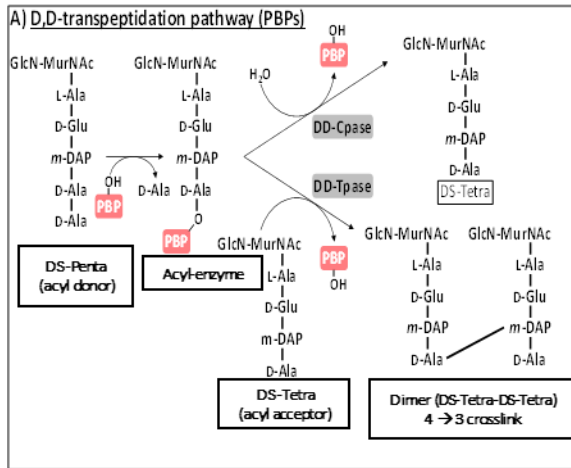
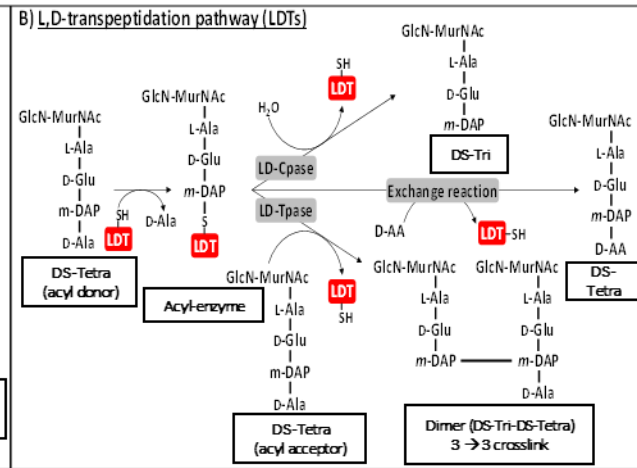
PBP2 (class B) is required for PG cross-linkage during the elongation process and its interaction with PBPA1 stimulates the GTase activity of the latter (Egan, 2020). SEDS proteins are another

type of monofunctional GTases and a recent study has shown that their function as PG polymerases requires the presence of its cognate PBP class B (Taguchi, 2019).

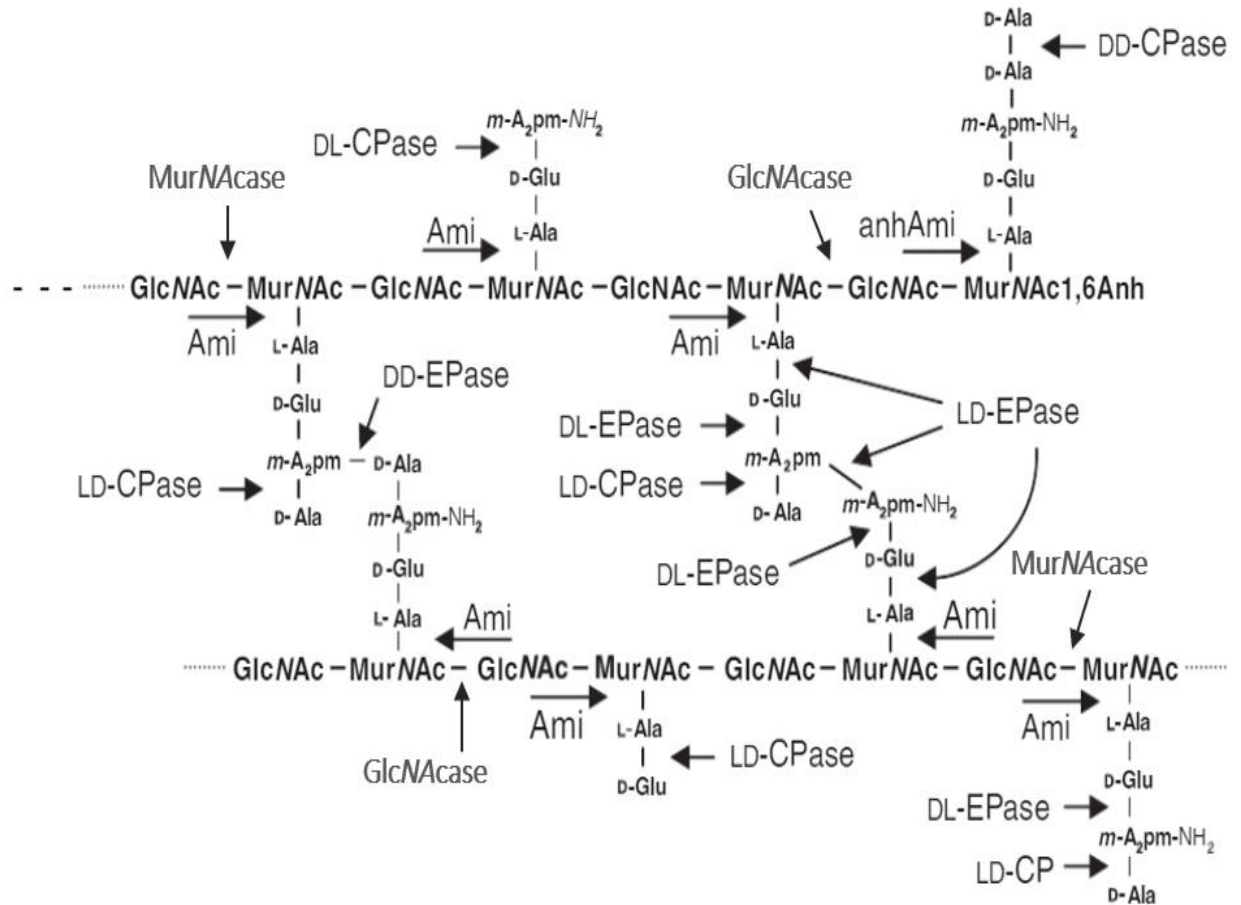
In many species, PBPs produce most of the crosslinkage in the PG. The pentapeptides are used as donors in DD-transpeptidation by catalysing peptide bond formation between the D-Alanine in position 4 (donor stem) and the amino group of a residue in position 3 (acceptor stem) (Vollmer, 2008a). PBP activity is inhibited by  $\beta$ -lactams and glycopeptides, two of the most clinically relevant antibiotics (Arbeloa, 2004; Boneca, 2003; Stapleton, 2002; Tipper, 1965). Class C PBPs are monofunctional enzymes with D,D-carboxypeptidase and endopeptidase activities involved in cell separation. These enzymes hydrolyse the D-Ala<sub>4</sub>-D-Ala<sub>5</sub> bond in the pentapeptide stem and the D-Ala<sub>4</sub>-L-Lys<sub>3</sub> bond of cross-linked PG, respectively (Vollmer, 2008c).

Finally, D-Ala is not the only possible acyl donor during the cross-linkage reaction. LD-transpeptidases (LDTs) use the carboxyl group of the amino acid at position 3 as donor to form 3-3 cross-links (Hugonnet, 2016; Wietzerbin, 1974). In most species L-D cross-links are of low abundance but increase during stationary phase or in  $\beta$ -lactam resistant strains (Hugonnet, 2016; Vollmer, 2008a). In the *Mycobacterium* and *Clostridium* genera, 3-3 cross-links are the most abundant type (Baranowski, 2018; Egan, 2020; Vollmer, 2008a; Wietzerbin, 1974), and some strains of *E. faecium* can bypass the need for DD-transpeptidation by exclusively performing LD-transpeptidation in the presence of  $\beta$ -lactams (Fig. 1.4) (Mainardi, 2005)

PG hydrolases play a key role in sculpting the shape, size, and thickness of the newly synthesized and matured sacculus, and contribute towards interbacterial interactions (Vollmer, 2008a). PG hydrolases can have redundant roles, and it has been shown that no single hydrolase knockout prevents growth or cell separation (Typas, 2011; Vollmer, 2008c). The major classes of PG hydrolases are *N*-acetylmuramidases, that hydrolyse the bond between GlcNAc and MurNAc residues and release a MurNAc residue at the reducing end; *N*-acetylglucosaminidases, cleaving between MurNAc and GlcNAc residues releasing a GlcNAc residue at the reducing end; transglycosylases, that cleave the same bond as *N*-acetylmuramidase but do not use a water molecule during cleavage; amidases, that cleave the bond between MurNAc and the *N*-terminal L-Ala residues of the stem peptide; and peptidases which are divided into endopeptidases, and

**A****B**

**Figure 1.4. Peptidoglycan cross-linking in *C. difficile*.** (A) Reactions catalysed by penicillin-binding proteins (PBPs) lead to the formation of 3-4 bonds; class C (low molecular weight) PBPs display D,D carboxypeptidase activity to hydrolyse the D-Ala<sub>4</sub>-D-Ala<sub>5</sub> bond in the pentapeptide stem. (B) Reactions catalysed by L,D-transpeptidases (LDTs) lead to the formation of 3-3 bonds. LDTs can also display L,D-carboxypeptidase activity and can exchange the C-terminal amino acid in position 4.



**Figure 1.5. Cleavage sites of the different PG hydrolases.** Hydrolysis of glycosidic, amide, and peptide bonds in *E. coli* or *B. subtilis* PG. *N*-acetylmuramidases (MurNAases) and *N*-acetylglucosaminidases (GlcNAases) cleave the glycosidic bonds between GlcNAc and MurNAc glycan chains. *N*-acetylmuramyl-L-alanine amidases (Ami) hydrolyse the amide bonds between the lactyl group of MurNAc and the L-alanine of the stem peptide. The anhAmi specifically cleaves at 1,6-anhydroMurNAc residues. Endopeptidases (DD-EPase, LD-EPase, DL-EPase) cleave amide bonds within the peptide stem or the cross bridges when present. The cleavage sites for LD- or DL-endopeptidases are indicated in a dimeric (cross-linked) peptide, but monomeric units can be hydrolysed. Carboxypeptidases (DD-CPase, LD-CPase, DL-CPase) cleave peptide bonds to remove C-terminal D- or L-amino acids (Adapted from Vollmer, 2008c).

carboxypeptidases. Endopeptidases cleave the bond within the peptide stem or cross bridge and carboxypeptidases remove the C-terminal D-or L-amino acids of peptide stems (Fig. 1.5)(Typas, 2011; Vollmer, 2008c).

### **1.1.2. Anionic polymers**

The cell wall of Gram-positive bacteria is densely decorated by linked heteropolymers. Teichoic acids and other diverse heteropolysaccharides such as capsules and rhamnopolysaccharides (RhapWPs) are intimately involved in cell division, cell shape (particularly in rod-shape microorganisms), resistance to host immune defences, and the modulation of resistance to cationic antibiotics and  $\beta$ -lactams (Brown, 2013; Desvaux, 2018; Percy, 2014; Smith, 2019).

#### **1.1.2.1 Teichoic acids.**

Teichoic acids (TAs) are a diverse family of cell surface glycopolymers containing phosphodiester-linked polyol repeat units (Ward, 1981). There are two types of TAs: (i) lipoteichoic acids (LTA) anchored in the cytoplasmic membrane and (ii) wall teichoic acids (WTA) covalently attached to PG (Brown, 2013; Caveney NA., 2018; Percy, 2014).

WTAs are covalently attached to PG via a phosphodiester bond to the C6 hydroxyl of MurNAc (Neuhaus, 2003; Ward, 1981). WTAs can be divided into two components: the disaccharide linkage unit and the main chain polymer. The disaccharide linkage unit is highly conserved across species and it is made of *N*-acetylmannosamine (ManNAc) and GlcNAc linked by a  $\beta$  1-4 bond, with one to two glycerol-3-phosphate (GroP) units attached to the C4 oxygen of ManNAc (Brown, 2013; Caveney NA., 2018). The anomeric phosphate of GlcNAc from the linkage unit forms the covalent bond with MurNAc in the PG. The main chain, made of phosphodiester-linked polyol repeats, extends from the GroP end of the linkage units (Ward, 1981). The two best-characterized WTA structures are comprised of repeating 1,5-D-ribitol-phosphate (RboP) or 1,3-L- $\alpha$ -glycerol-phosphate (GroP) units. However, WTA monomer structures can be highly diverse (Brown, 2013; Endl J, 1983). It has been shown that members of the same species display differences in their WTA composition. In *B. subtilis*, strains 168 and W23 contain GroP and RboP, respectively

(Brown, 2013), whilst in *E. faecalis*, the WTA repeating units of the clinical isolate 12030 were composed of D-glucose, D-galactose, 2-acetamido-2-deoxy-D-galactose, 2-acetamido-2-deoxy-D-glucose, D-ribitol, and phosphate in a molar ratio 1:2:1:1:1:1 (Theilacker, 2012). These modifications may represent adaptations to environmental pressure. It has been shown that WTAs are required for survival; a cell lacking WTAs has gross defects in cell division and cell morphology. WTAs comprise up to 60% of the cell wall mass (Brown, 2013). In *B. subtilis* and *S. aureus*, it has been estimated that every ninth PG MurNAc residue contains an attached WTA polymer made of 40 to 60 polyol repeats (Bera, 2007; Kojima, 1985). WTAs are also involved in biofilm formation, virulence, and antimicrobial resistance (Brown, 2013).

LTAs are defined as alditolphosphate-containing polymers which are linked via a lipid anchor to the cytoplasmic membrane. LTAs have been grouped into five different types based on their chemical structures. The best-characterized LTA structure is the polyglycerol phosphate, or type I LTA, which are the most common LTAs in Firmicutes. They have an unbranched 1–3 linked GroP backbone structure linked to the cytoplasmic membrane via a glycolipid anchor, which is often a diglucosyl-diacylglycerol (Glc2-DAG) unit. Type II to V LTAs have more complex structures which often contain glycosyl residues. In *S. pneumoniae*, type IV LTAs are decorated with phosphocholine (Percy, 2014). LTAs and WTAs are often modified with sugar moieties and D-Ala esters, that introduce positive charges to the negatively charged phosphate polymers. LTAs play an important role for bacterial growth and physiology, and contribute to membrane homeostasis and virulence (Percy, 2014; Reichmann, 2013, 2011).

#### **1.1.2.2 Wall polysaccharides**

Wall polysaccharides (WPs) are complex structures with various compositions e.g. capsule, RhapWPs, or teichuronic acids. The composition of WPs often varies and can be strictly strain specific. Capsules contribute to pathogenesis through specific interactions with opsonophagocytic antibodies. The capsular polysaccharide (CPS) is variable and non-ubiquitous, in *E. faecalis* V583 is made of diheteroglycan glucose and galactose and is involved in resistance to phagocytosis (Guerardel, 2020). The anchoring of CPs remains unclear, however, in *S. pneumoniae*, it has been shown to be covalently linked to the GlcNAc residues (Larson, 2017).

CPSs and RhapWPs are major virulence factors in many microorganisms (Geiss-Liebisch S, 2012; Teng, 2002; Thurlow, 2009). In Enterococci, the presence of these heteropolysaccharides allows bacteria to escape detection and clearance by the host immune system (Guerardel, 2020; Smith, 2019). The enterococcal polysaccharide antigen (EPA) has been associated with invasion of host tissues (Teng, 2002) and biofilm formation (Mohamed JA, 2005). The genes responsible for EPA production are encoded by a complex locus (*epa*) that contains 18 genes (*epaA* to *epaR*) (Palmer, 2012; Smith, 2019). Recent studies revealed that the decorations of the EPA polymer, encoded by genetic loci that are variable between isolates, underpin the biological activity of this surface polysaccharide by contributing to resistance towards phagocytosis and charged antimicrobials (Smith, 2019). The complete structure of *E. faecalis* strain V583 has been recently determined (Guerardel, 2020). The rhamnan backbone of EPA is composed of a hexasaccharide repeat unit of C2- and C3-linked rhamnan chains, partially substituted in the C3 position by  $\alpha$ -glucose ( $\alpha$ -Glc) and in the C2 position by  $\beta$ -N-acetylglucosamine ( $\beta$ -GlcNAc). The EPA decorations consist of phosphopolysaccharide chains corresponding to teichoic acids covalently bound to the rhamnan backbone. As previously mentioned, the composition of these complex heteropolysaccharides differs between species and it is likely to be strictly strain specific as determined for other *E. faecalis* EPA-containing strains (unpublished-Davis, J., Smith RE).

### 1.1.3 Architecture

PG is a single complex macromolecule that does not display a crystalline structure (Vollmer, 2008a). Different studies combining biochemical analyses and atomic force microscopy (AFM) on isolated PG from *B. subtilis* led to architectural models where the PG network is oriented parallel to the short axis of the cell (Hayhurst, 2008; Turner, 2014; Vollmer, 2010). A more recent study used high-resolution AFM of *S. aureus* and *B. subtilis* living cells and PG sacculi imaged in liquid (Pasquina-Lemonche, 2020). It showed that the mature cell wall is a disordered, mesh-like hydrogel with large (up to 60 nm diameter) and deep (up to 23 nm) pores. The cell wall thickness has been estimated at  $\sim$ 20 nm which implies that some pores could extend across most of the wall (Vollmer, 2010). It was also determined that the inner PG layer is mainly constituted of

recently synthesized PG. This surface appeared much denser with less than 7 nm of glycan strands spacing and it displayed two distinct architectures that are thought to be location dependent.

The two distinct inner surface architectures were attributed to different synthesis regimes. First, a highly ordered structure formed of circumferentially oriented glycan strands was proposed to be constituted of nascent septal material. The observations of long concentric rings support the key role of PG hydrolases in cell wall remodelling during growth and maturation. It was hypothesised that this material was deposited at the leading edge of the constricting cell membrane forming the post-synthesis core. The second inner architecture extends across the division septa to constitute the cytoplasm-facing side of the wall. This inner surface is the densest region with substantially smaller pores and randomly oriented material. These observations were consistent in *S. aureus* and *B. subtilis*, apart from the presence of concentrically oriented strands at the cylinder attributed to elongasome-associated PG synthesis in the latter (Pasquina-Lemonche, 2020; Turner, 2014).

It has been proposed that PG remodelling results in the reorientation of the glycan strands from the septal rings (*S. aureus*) or the cylinder (*B. subtilis*) to produce a porous surface architecture (Pasquina-Lemonche, 2020). The distinct architectural arrangements observed across the PG contribute to our understanding of this macromolecule as an elastic and dynamic structure. PG serves as a scaffold for the attachment of complex polymers likely protruding from its pores, as well as the transport and interaction of different surface proteins. It also controls the turgor pressure within the cell to prevent plasmolysis which requires a closer and denser PG structure of the inner, smoother surface.

## **1.2. Cell surface proteins of Gram-positive bacteria**

Bacterial surface proteins play a key role during cell growth and division. They interact with the host immune system, and in some cases, they are involved in the competition with other bacteria for a given ecological niche. The study of bacterial surface proteins dates back to the late 1950s with initial efforts to characterize predominant antigens in Staphylococci and Streptococci strains



(Jensen, 1958; Lancefield, 1962). Cell wall-associated proteins are classified into two main classes: (i) proteins covalently attached to PG and (ii) proteins non-covalently bound to either PG and/or cell wall associated polymers (Table 1.1).

### **1.2.1 Cell surface proteins covalently bound to PG**

#### **1.2.1.1 Cell wall sorting mechanism**

The covalent binding of proteins to PG is conserved in Gram-positive bacteria and is one of the best characterised surface protein anchoring mechanisms (Dramsi S., 2008; Fischetti, 1990; Navarre, 1999; Schneewind, 1995). It involves the recognition of a C-terminal cell wall sorting signal (CWS) characterised by a charged tail of a few residues followed by an LPXTG motif. LPXTG proteins are covalently linked to PG by a family of enzymes named sortases (Dramsi S., 2008). LPXTG proteins are present in all known Gram-positive bacteria and have been identified as colonising factors, toxins, proteases, amongst others (Navarre, 1999). *S. aureus* protein A has been the main subject of study to unravel the cell wall sorting mechanism (Navarre, 1999; Schneewind, 1995).

The C-terminal CWS is comprised of a highly conserved LPXTG motif, where X denotes any amino acid, followed by a hydrophobic domain and a positively charged tail (Figure 1.6) (Dramsi S., 2008; Fischetti, 1990). It has been determined that all elements from the CWS are necessary for the correct protein attachment to the cell wall. The hydrophobic domain and the charged tail are necessary for protein translocation where they are thought to hold the protein at the cytoplasmic membrane whilst the rest of the LPXTG protein is transported via the secretory pathway to the outer leaflet of the cell membrane (Dramsi S., 2017). Here the LPXTG sequence is recognized by sortase enzymes that attach the LPXTG protein on the nascent PG by cleaving between the T and G residues. Depending on the PG nature, sortases anchor this motif into the NH<sub>2</sub> group of the amino acid at position 3 in the PG peptide stem (Budzik JM, 2008) or on the N-terminal residue of the lateral chain (Marraffini, 2005).

**Table 1.1 Domains involved in protein attachment to the Gram-positive cell wall.**

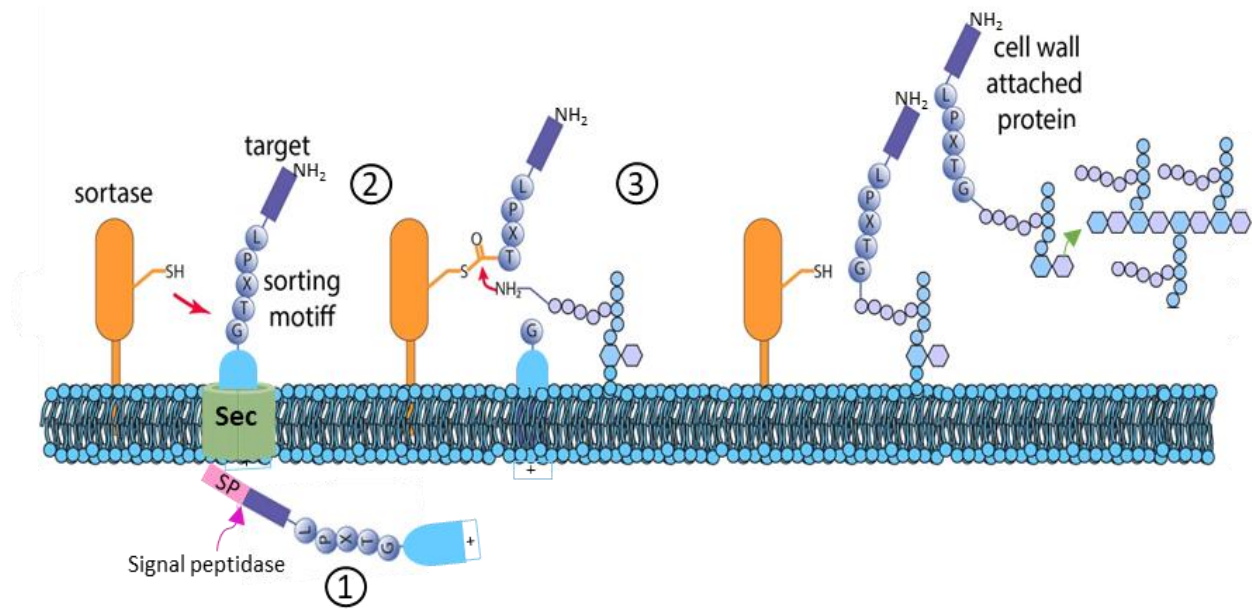
Name	Abbreviation	Other names <sup>a</sup>	InterPro	Other databases <sup>b</sup>	PDB <sup>c</sup>	CW ligand <sup>d</sup>
<b>Domain involved in covalent attachment to the CW</b>						
LPXTG domain	LPXTG		IPR019948	PF00746, PS50847, PR00015	3UXF	PG
<b>Domain involved in non-covalent attachment to the CW</b>						
Lysin motif	LysM		IPR018392	PF01476, SM00257, CD00118, PS51782, SSF54106	2MKX	PG
WXL domain	WXL		IPR027994	PF13731	-	PG
Sporulation-related domain	SPOR		IPR007730	PF05036, PS51724, SSF110997	6I0A	PG
Choline binding domain	CBP	ChBD	IPR018337	PF01473, PS51170	1HCX	Choline residues
Clostridial hydrophobic repeat (ChW)	ChW		IPR006637	PF07538, SM00728	-	n.d.
PASTA	PASTA		IPR005543	PF03793	5OAU	Uncross-linked PG
SH3 domain of type 3	SH3_3	SH3b	IPR003646	PF08239, SM00287, PS51781	4KRT	PG
SH3 domain of type 5	SH3_5	SH3b	IPR003646	PF08460	6RK4	PG
SH3 domain of type 6	SH3_6	SH3b1		PF12913	3M1U 3H41	PG
SH3 domain of type 7	SH3_7	SH3b2	IPR026864	PF12914	3M1U 3H41	PG
GW domain	GW	SH3 domain of type 8	IPR025987	PF13457, PS51780	1M9S	PG and/or LTAs

<sup>a</sup> Other names found in the literature. Name and abbreviation given in the first two columns are the most common.

<sup>b</sup> Member databases used to construct the Interpro entry, i.e. Pfam (PF), SMART (SM), Conserved Domain Database (CD), Prosite (PS), Prints (PR), SuperFamily (Fischer et al.).

<sup>c</sup> Accession number of structures available in the PDB.

<sup>d</sup> CW, cell wall; PG, peptidoglycan; LTAs, lipoteichoic acids; WTAs, wall teichoic acids; WPs, wall polysaccharides; n.d., not determined. Choline residues are found in WTAs and LTAs.



**Figure 1.6. Covalent anchoring of surface proteins by SrtA.** Most proteins covalently bound to PG have a C-terminal sorting signal comprised of: (i) the LPXTG sorting motif, where X denotes any amino acid, (ii) a hydrophobic sequence and (iii) a positively charged tail of 5-12 amino acids. The N-terminal sequence directs the full-length polypeptide (step 1) through the Sec system and, upon translocation, is cleaved by signal peptidases. The product of this reaction (step 2) is retained within the secretory pathway via its C-terminal hydrophobic domain (blue box) and positively charged tail (+). The sortase, a membrane-associated enzyme with an active cysteine group, cleaves the substrate between the threonine and glycine residues to form an acyl-enzyme, in which cysteine from the active site interacts with the carboxyl group of threonine. The carbonyl group of the threonine residue is then attached to the amine group of the last glycine residue of the lateral chain on a lipid II precursor. The attached protein to lipid II (step 3) reaches the cell surface at the lipid precursor and is polymerized into the existing network. (Adapted from BPS Bioscience, Inc).

There are six classes of sortases classified from A to F (Dramsi S., 2017). All sortase enzymes possess a similar structure, with an N-terminal signal peptide and a catalytic TLXTC motif. However, they display specific recognition patterns related to the chemical structure of PG and the variability within the LPTXG motif (Dramsi S., 2017; Siegel, 2017). Sortase A anchors a wide range of LPXTG proteins, whereas sortase B recognizes a related motif NP(Q/K)TN. The sortase class C is involved in pilus assembly whilst class D enzymes, predominantly identified in Bacilli, have been associated with the anchoring of sporulation proteins to the cell wall. Class E and F have been identified in Actinobacteria, but their functions remain unknown (Desvaux, 2018; Dramsi S., 2017; Siegel, 2017).

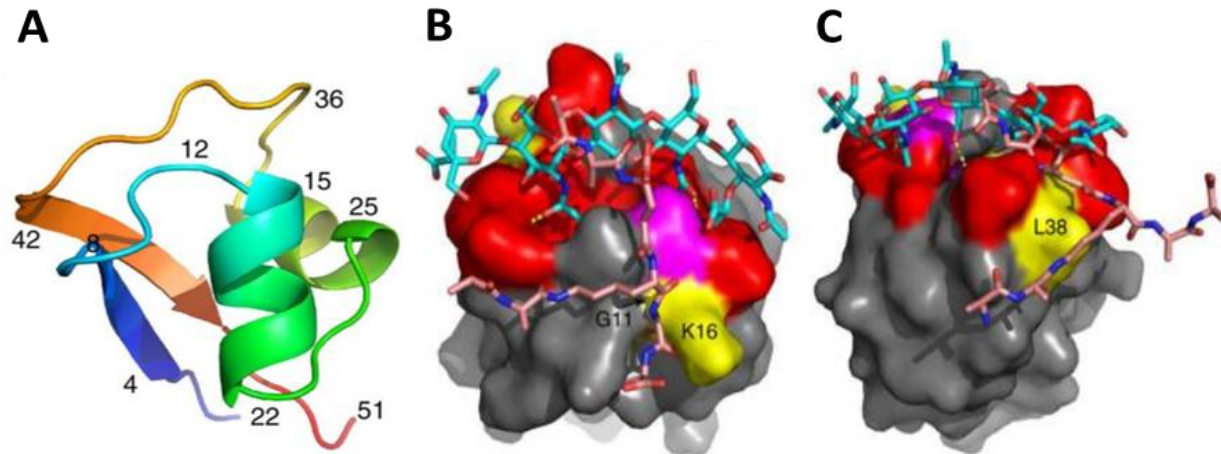
### **1.2.2 Cell surface proteins non-covalently bound to PG**

In Gram-positive bacteria, most surface proteins have a multi modular organization, with two or more domains. Many surface proteins are PG hydrolases with a general architecture consisting of a catalytic domain linked by its N- or C- terminal end to one or more cell wall binding modules (Vollmer, 2008c). Most cell wall binding domains consist of 30 to 100 amino acids that form non-covalent bonds enabling weak and dynamic interactions by the formation of hydrogen bonds, ionic bonds, van der Waals forces and/or hydrophobic interactions (Desvaux, 2018).

The discovery and classification of cell wall associated proteins has dramatically increased thanks to the availability of sequencing data and bioinformatic analyses. These data have highlighted the great structural and functional diversity of the cell wall binding domains (Table 1.1).

#### **1.2.2.1 LysM domains**

Lysin motif (LysM) domains are highly conserved carbohydrate-binding molecules present across all kingdoms (Buist, 2008; de Jonge, 2009; Kaku, 2006; Radutoiu, 2003). According to Pfam (June,2020) there are 28,700 LysM bacterial sequences from over 6,400 specie (El-Gebali, 2019). LysM domains consist of approximately 45 amino acids that adopt a  $\beta\alpha\alpha\beta$  fold with the two helices packing onto the same side of an anti-parallel beta sheet (Fig. 1.7 A) (Bateman, 2000; Mesnage, 2014; Wong, 2015). Up to 12 LysM domains can be found in a single protein



**Figure 1.7. Structure of *E. faecalis* LysM AtIA module bound to GlcNAc<sub>5</sub>.** (A) Cartoon representation of the NMR structure of the first AtIA LysM module.  $\beta$ -Sheets are formed of residues T4-V8 and G42-V47;  $\alpha$ -helices of L14-Y21 and V25-N32. (B,C) Model of interaction between one LysM domain to *E. faecalis* PG, including the peptide stem. Glycan strands MurNAC-(GlcNAc-MurNAC)<sub>2</sub> are depicted in blue. The peptide stem (pink) is represented in two possible orientations interacting with either the residue G11/K16 or L38 (Adapted from Mesnage, 2014).

and they are often separated by flexible linkers that are rich in S, T, and N residues (Mesnage, 2014).

LysM-carbohydrate interactions are highly diverse, and it has been shown that the binding mechanism varies within LysM domains. In eukaryotes, LysM proteins can form quaternary structures requiring posttranslational modifications involving the formation of disulphide bonds, that are necessary for carbohydrate recognition (Lefebvre, 2012; Radutoiu, 2007). In bacteria, LysM domains generally do not contain cysteine residues and have been shown to bind PG in a non-covalent manner via a cooperative binding mode (Mesnage, 2014).

Bacterial LysM domains bind to PG by interacting with GlcNAc polysaccharides (Mesnage, 2014). *E. faecalis* AtIA, an autolysin involved in cell division, is constituted of six C-terminal LysM repeats that recognise the GlcNAc-x-GlcNAc motif. One essential attribute of this LysM-PG interaction is the presence of a hydrogen bond network that creates a contact surface with four carbohydrates that are the minimal PG binding motif. It has been determined that the multiple AtIA LysM domains displayed an additive binding to either short PG polysaccharides or intact PG sacculi. Interestingly, AtIA binds to chitooligosaccharides (GlcNAc-containing oligomers) with a higher affinity, which highlights the adaptability of carbohydrate recognition across kingdoms (Mesnage, 2014). The optimal ligand for LysM appears to be a GlcNAc pentasaccharide (GlcNAc)<sub>5</sub> (Koharudin LM, 2011; Mesnage, 2014).

The crystal structure of a single fungal LysM module from the MoCVNH3 lectin of *Magnaporthe oryzae* was determined in its apo form and as a (GlcNAc)<sub>3</sub>- and (GlcNAc)<sub>4</sub>-bound protein (Koharudin LM, 2011). It was shown that the binding site is comprised of two loop regions that connect the secondary structures, i.e. the loop between  $\beta_1$ - $\alpha_1$  and the loop between  $\alpha_2$ - $\beta_2$ . Conformational changes caused by ligand recognition form a shallow groove that can accommodate the carbohydrates. These changes were especially pronounced for residues in the loop region connecting helix  $\alpha_2$  and strand  $\beta_2$ . It was also determined by NMR titrations that LysM domains can accommodate long GlcNAc chains in different fashions. These observations are in

agreement with the LysM structure from the putative NlpC/P60 D,L endopeptidase from *Thermus thermophilus* bound to a (GlcNAc)<sub>6</sub> polysaccharide (PDB, 4UZ3) where several alternative conformations were found for the last two GlcNAc oligomers (Wong, 2015). No dimerization of MoCVNH3 via its LysM domain was observed upon binding to GlcNAc<sub>6</sub>, unlike multiple eukaryotic LysM domain-containing proteins (Koharudin LM, 2011; Wong, 2015).

A recently determined co-crystal structure of *E. faecalis* AtIA bound to a GlcNAc pentasaccharide (GlcNAc)<sub>5</sub> (Galley N., F. Vincent, Mesnage S., unpublished) confirmed the published model made with the NMR structure of AtIA, docking the LysM-GlcNAc<sub>5</sub> oligosaccharide on the first LysM module (Mesnage, 2014). AtIA LysM exhibits the typical  $\beta\alpha\beta$  fold connected by two loops between  $\beta 1-\alpha 1$  and  $\beta 2-\alpha 2$ . NMR titrations with PG fragments ( $\beta$ -1,4- GlcNAc -MurNAc-peptide polymer) located the interaction site within the protein loops (Fig. 1.7 B). They formed a contiguous surface in accordance with the interaction shown with chitin (Koharudin LM, 2011; Wong, 2015). Moreover, it was suggested that LysM from AtIA has evolved to bind PG since NMR titrations indicated residues that only interacted in the presence of GlcNAc-MurNAc ligands containing both the lactoyl group and peptide stem (Fig. 1.7 B).

#### **1.2.2.2 WxL domains**

The C-terminal WxL family domain is comprised of two highly conserved sequences containing the Trp-x-Leu (WxL) motifs. The first WxL motif is followed by a well-conserved YXXX(L/I/V)TWXLXXXXP distal motif at the last 120 to 190 amino acid residues of the protein (Brinster, 2007; Siezen, 2006). WxL proteins have been identified in over 220 species of Gram-positive bacteria (June,2020) (El-Gebali, 2019), mainly in the orders of Lactobacillales and Bacillales (Brinster, 2007; Dumas, 2008; Galloway-Peña, 2015; Toh, 2013). To this date there are no structures available in the protein data base. WxL domains were shown to bind PG (Brinster, 2007). However, the PG binding motif has yet to be identified. In *E. faecalis*, recombinant fusion proteins including an N-terminal nuclease reporter and a C-terminal WxL domain were made from two WxL-containing proteins found in the V583 strain. These proteins were shown to bind



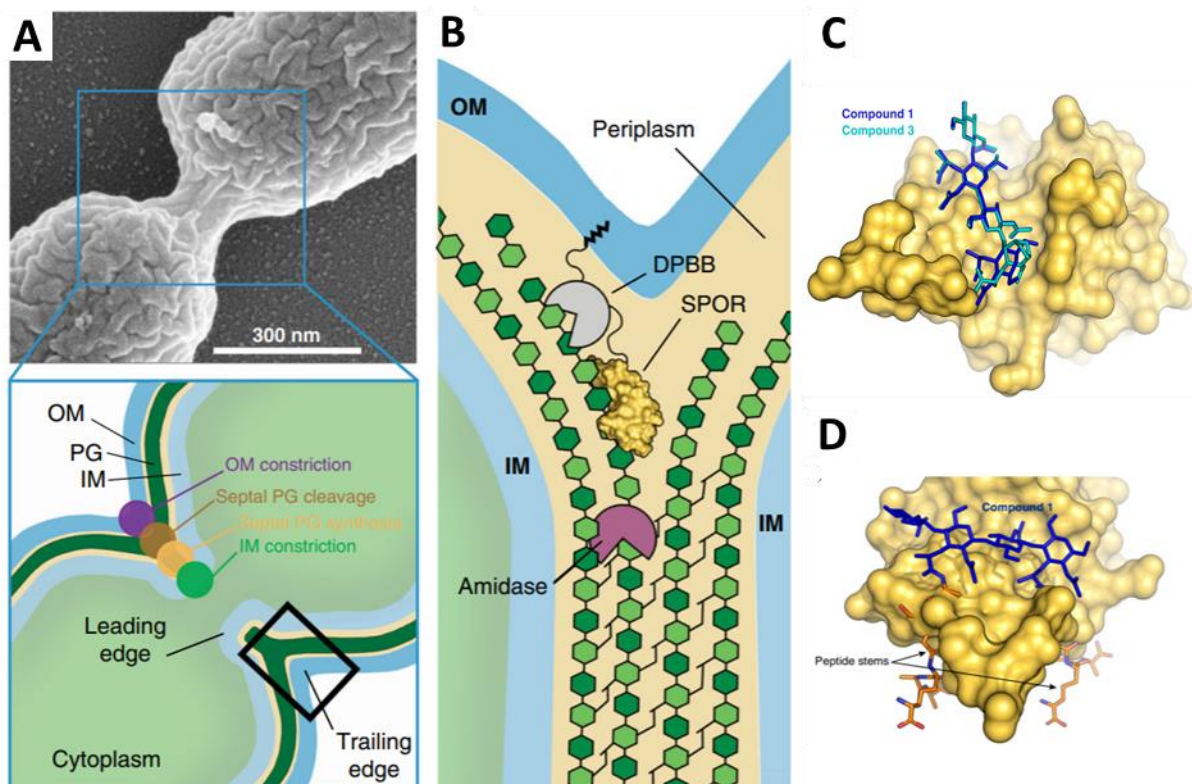
the bacterial cell wall of whole cells and isolated PG by Western Blot analyses. Moreover, deletions of the WxL domains proved to abolish binding to PG (Brinster, 2007).

It has been shown that WxL proteins are encoded by genes organised in clusters and from bioinformatic and transcriptome data in *L. plantarum* and *E. faecium* it has been hypothesized that these loci are likely to form cell surface protein complexes involved in virulence or colonization-related functions (Brinster, 2007; Galloway-Peña, 2015; Siezen, 2006). Further biochemical and structure-function analyses are required to elucidate the architectural and biological properties of these proteins.

### **1.2.2.3 Sporulation-related domain (SPOR)**

The sporulation-related (SPOR) domains are PG binding proteins found in a large number of bacteria. Amongst Gram-positive bacteria these domains are mainly found in Firmicutes, especially in the Clostridiales and Bacillales orders (Alcorlo, 2019; Yahashiri, 2017). Currently, the Pfam database includes 4127 bacterial species harbouring SPORs and five structures have been determined (June,2020) (El-Gebali, 2019). One to five SPOR domains can be found in a single protein (Mishima, 2005). They are approximately 70 amino acids long ,constituted of two 35-residue repeats. Overall, SPOR proteins share low sequence similarity, with less than 20% amino acid identity. However, at a structural level they share a conserved fold made by a four stranded antiparallel  $\beta$ -sheet flanked on one side by two  $\alpha$ -helices. These domains have been classified as part of the ribonucleoprotein (RNP) fold superfamily which are also present in eukaryotic proteins, although RNP folds are not associated with the recognition of a particular ligand (Yahashiri, 2017).

Originally, SPOR domains were found in proteins associated with sporulation (Mishima, 2005). The identification of the first C-terminal SPOR protein was in the *B. subtilis* Cwlc amidase, which is involved in PG hydrolysis of the mother cell that allows the release of the spore. SPOR proteins are quite diverse and are also involved in the remodelling of PG, primarily acting as PG septal



**Figure 1.8. Structural basis of denuded glycan recognition by *P. aeruginosa* RlpA SPOR domains.**

(A) Top, scanning-electron micrograph showing the *P. aeruginosa* daughter-cell separation process. Inset shows a schematic cartoon representing the cell-division septum and coordinated processes involved in cell-envelope constriction. (B) Diagram of the division site (corresponding to the blue boxed area in (A)), showing the coordinated action of PG amidases (purple) and the glycosylation of RlpA catalytic domain (gray) upon recognition of the denuded glycan chains by its SPOR domain (yellow). OM, outer membrane; IM inner membrane. The protein has a covalently attached lipid at the amino terminus. (C) Crystal structure of the SPOR-RlpA in complex with (MurNAc-GlcNAc)<sub>2</sub> (compound 1) and tetrasaccharide containing anhMurNAc at position 1 (compound 3). The structure of compound 1 (blue sticks) is superimposed (protein omitted for clarity). (D) Model to represent the steric clashes of PG peptide stems in the SPOR-RlpA:1 complex. The peptide conformation is according to PDB 6FCS. The modelled peptide stems are in orange for the carbon atoms. Adapted from (Alcorlo, 2019; Gray, 2015).

targeting domains that recruit proteins to the cell division site (Ursinus, 2004; Yahashiri, 2017, 2015). SPOR proteins bind 'denuded' PG, i.e. PG devoid of peptide stems (Gerding, 2009; Ursinus, 2004; Yahashiri, 2015). Denuded glycan strands are located at septal PG and are generated by amidases to allow the separation of the daughter cells (Fig. 1.8 A-B) (Alcorlo, 2019; Vollmer, 2008c; Yahashiri, 2017).

According to Pfam, these domains are present in multiple domain organizations associated with proteins involved in different cellular functions. In a simplified overview of the domain organization the most abundant architecture, around 60% of the SPOR sequences, is associated with proteins containing an N-terminal transmembrane anchor with a large unstructured linker region of low sequence complexity. The best-studied proteins from this group are the *E.coli* FtsN, DamX, and DedD proteins, which are all part of the apparatus that remodels PG during cell division (Yahashiri, 2017; Yang, 2004). Previous observations using GFP-SPOR fusions of the mentioned proteins (FtsN, DamX, and DedD) showed no binding to isolated PG sacculi from an *E.coli* mutant lacking the three main amidases responsible for daughter-cell separation ( $\Delta$ *amiABC*) (Heidrich C, 2001; Priyadarshini, 2007). As mentioned above, these enzymes remove peptide stems from glycan strands, so denuded glycans presumably are absent in the  $\Delta$ *amiABC* mutant. This suggested that, not only do SPORs bind denuded PG, they are incapable of binding to a 'complete' monomeric PG unit. Although it is likely that SPOR domains share an overall similar PG recognition, regardless of their modular organization, a structure of the SPOR-glycan complex of cell-division proteins is necessary, in addition to the existing biophysical studies (Mishima, 2005; Williams, 2013; Yang, 2004) to confirm the nature of their binding mode as for the recently solved co-crystal structure of the SPOR RlpA PG hydrolase (Fig. 1. 8) (Alcorlo, 2019).

The second most abundant domain architecture found in SPORs (~20%) corresponds to SPOR-containing PG hydrolases, e.g. CwlC in *B. subtilis* and RlpA in *E.coli* and *Pseudomonas aeruginosa*. RlpA is a lytic transglycosylase needed for the efficient separation of daughter cells and the maintenance of rod shape (Alcorlo, 2019). A solved structure of the RlpA SPOR domain in complex with synthetic tetrasaccharide (MurNAc - GlcNAc)<sub>2</sub> ligands corroborated that the peptide stems

in the monomeric PG unit abrogate binding. The PG-binding site in the SPOR domain is located within the concave face of the  $\beta$ -sheet forming a cavity where the tetrasaccharide sits in an extended conformation (Fig. 1.8 C-D). It was proposed that the minimal ligand was constituted of two disaccharide (MurNAc-GlcNAc)<sub>2</sub> units devoid of their peptide stems. Although all four glycans directly interact with the protein, the strongest polar interactions were with the carboxylate moieties of the MurNAc rings, highlighting their crucial involvement in binding. These carboxylate moieties become available only after the removal of the peptide stems by the actions of PG amidases. The exposed D-lactyl group of MurNAc 1 interacts with a conserved Q residue identified in previous studies, and a positively charged patch made of three R residues, whereas the second D-lactyl group interacts with an F residue. These interactions create a bent conformation of the glycan chain.

SPOR proteins were also capable of binding modified polysaccharides (containing the anhMurNAc residues) which suggests that SPORs are able to bind at the end of the glycan chains (peripheral PG) (Fig. 1.8 C). However, fewer polar interactions were observed between the lactyl carboxylate and the Q and R residues. Computationally modelled complexes with *O*-acetylated glycans allowed these interactions, whereas *N*-deacetylated glycan chains seemed unfavourable for binding. The ligand did not remain in the binding pocket since the *N*-deacetylated PG was not able to form a hydrogen bond with the conserved Q residue compared to other tolerated modifications.

The remaining SPOR domain architectures are found in proteins that fall into different categories, most of which have unknown functions.

#### **1.2.2.4 Clostridial hydrophobic repeat (ChW)**

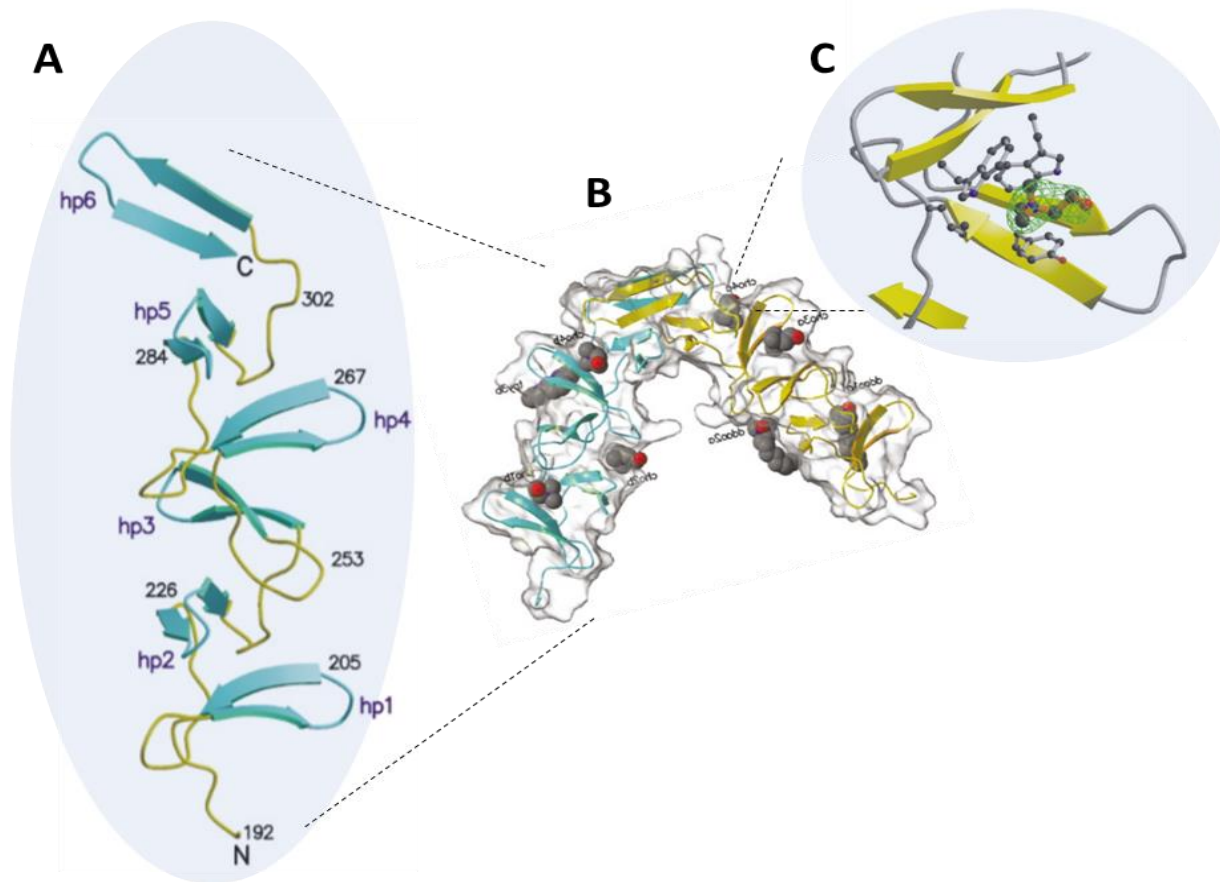
The ChW (clostridial hydrophobic W (tryptophan)) domains are approximately 50 amino acids long and contain a highly conserved Gly–Trp dipeptide motif. A single protein can harbour up to 12 domains but there are no structures yet available. They were originally identified in *Clostridium acetobutylicum* and proposed to be involved in cell wall binding based on bioinformatic assignment (Desvaux, 2005; Nölling, 2001). Although they are mainly associated

with Clostridia, they are present in some Gram-positive bacteria, including *E. faecalis*, *S. agalactiae* and some members of the phyla Actinobacteria, specially *Streptomyces* (Desvaux, 2018). ChW domains are mostly found in bacteriophage endolysins where they could be involved in targeting the cell wall, and very little is known about their structure or function (Oliveira, 2013).

#### **1.2.2.5 Choline binding proteins (CBD)**

The choline-binding proteins (CBD) are comprised of conserved regions of approximately 20 amino acids long. This conserved motif is also known as the cell wall binding repeat type 1 (CWB1). CBDs are commonly found in tandem or in association with other PG binding domains (Galán-Bartual, 2015; García, 1998; Varea J, 2000). They have been primarily described as TA binding modules in Streptococci, as well as in some of their associated bacteriophages (Eldholm, 2009; Fernández-Tornero, 2001; Maestro, 2016). The TA polymers (WTAs and LTAs) from the members of the Mitis group in the Streptococcus genus (including *S. pneumoniae*) are decorated with phosphatidyl choline (P-Cho) groups. In *S. pneumoniae*, P-Cho moieties appear essential for several physiological functions such as colonization of the upper respiratory tract in mice models (Kharat AS, 2006).

LytA is the major autolysin in *S. pneumoniae*. It is made of an N-terminal *N*-acetylmuramoyl l-alanine amidase domain with six C-terminal choline binding repeats (Fig. 1.9 A) (Fernández-Tornero, 2001; Maestro, 2016). CBDs bind to the TAs at the cell wall and their deletion from LytA prevents the localization of the amidase at the cell surface and therefore the autolysis of pneumococcal cells (Giudicelli, 1984). The overall shape of the six terminal repeats is approximately cylindrical. Each repeat encompasses two structural units: a  $\beta$ -hairpin formed by two antiparallel  $\beta$ -strands, and an 8-10-residue connecting loop. The six  $\beta$ -hairpins adopt a solenoid fold that stack to form a left-handed superhelix with a boomerang-like shape (Fig. 1.9 A). LytA CBDs bind to choline groups at hydrophobic cavities present in the protein surfaces (Fig. 1.9 B) (Fernández-Tornero, 2001). The choline binding sites (ChBS) are formed by the interface



**Figure 1.9. Structure of *S. pneumoniae* LytA CBD proteins in complex with choline residues. (A)** Cartoon representation of the six C-terminal LytA CBD domains.  $\beta$ -hairpins ('hp') are colored cyan, whereas the loops connecting them are colored yellow. **(B)** Partially transparent protein surface of the choline binding sites in a crystallized CBD-LytA dimer. Monomers are highlighted in different colors: yellow and cyan. CBD binding site (ChBS) 1 and 2 of monomer 'a' (yellow) are occupied by DDAO molecules from crystallization solutions. ChBS3 of monomer 'b' (cyan) is occupied by the (2,2':6',2''-terpyridine)-platinum(II) used for MAD phasing. The hydrophobic components of choline (labeled 'cho'), DDAO (labeled 'ddao') and terpyridin (labeled 'tpy') molecules, schematized as CPK, occupy small hydrophobic cavities on the surface of the protein. **(C)** Enlarged image of ChBS4, where choline is highlighted in orange. The side chains of the hydrophobic conserved residues forming the cavity are shown in the 'ball-and-stick' format. The  $2F_o - F_c$  omit map (green) of the choline molecule was contoured at  $1.0 \sigma$ . (Adapted from Fernández-Tornero C, 2001).

between consecutive  $\beta$ -hairpin pairs, i.e. hairpins 1 and 2 form the first ChBS (ChBS1), followed by hairpins 2 and 3 (ChBS2), hairpins 3 and 4 (ChBS3), and hairpins 4 and 5 (ChBS4) (Fig. 1.9 C) (Fernández-Tornero, 2001).

Other structures of CBD proteins bound to choline or analogues shared a similar binding mode. The CBD repeats from the Cpl1-1 bacteriophage endolysin bind to the choline groups in the pneumococcal cell wall between its C-terminal hairpin 1-2, and 2-3 of the superhelical domain (Hermoso, 2003).

#### **1.2.2.6 PASTA domains**

PASTA domains (penicillin-binding protein and serine/threonine kinase-associated domains) have been described in PBPs and serine/threonine kinases (STPK) of Firmicutes and Actinobacteria (Mir, 2011; Ogawara, 2016; Pares, 1996). They exhibit a modular structure comprised of 1 to 5 modules. Each module consists of approximately 70 amino acids. Although PASTA domains share low sequence similarity (<25%) they display a strong structural conservation (Ogawara, 2016).

PASTA domains in PBPs have a globular fold consisting of three  $\beta$ -strands and an  $\alpha$ -helix, with a loop region of variable length between the first and second strands (Gordon, 2000; Yeats, 2002). They were first identified in the C-terminus of PBP2x proteins in *S. pneumoniae* (Pares, 1996). The solved structure showed van de Waals interactions between the  $\beta$ -lactam ring of the cephalosporin cefuroxime to one of the two PASTA domains present in the protein. The structural similarities of the cefuroxime  $\beta$ -lactam ring to D-Ala-D-Ala dipeptides in the PG led to the proposed binding recognition of PASTA domains to uncrosslinked PG (Yeats, 2002). The localization of PBP2x in the mid-cell was shown to be dependent on PASTA domains which in turn is reliant on the presence of its substrate (PG precursors) (Jones, 2006; Peters, 2014). However, their biological function is still in debate since other PASTA domains do not seem to bind PG. In *M. tuberculosis*, PBP PonA2 which has a single PASTA domain does not bind muropeptides nor  $\beta$ -lactams, or polymeric PG (Calvanese, 2014). In *B. subtilis*, PASTA domains are only present in two

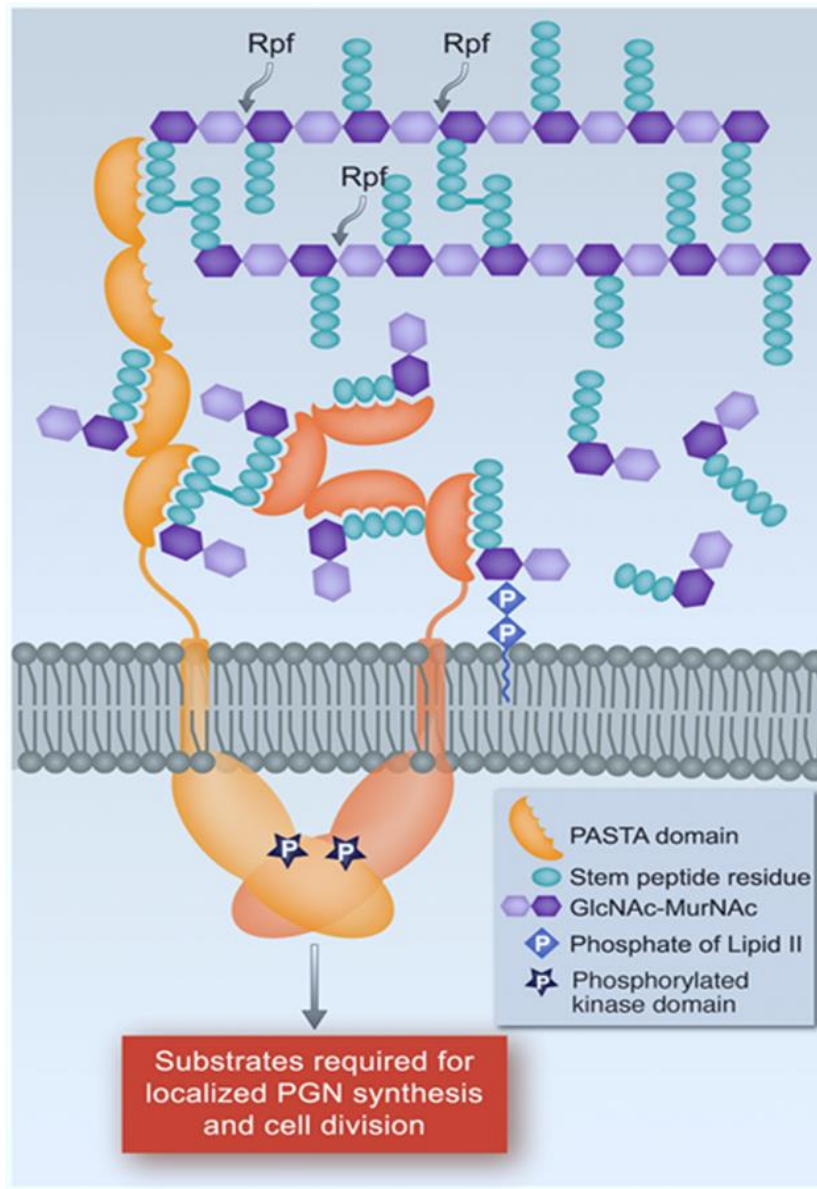
of the sixteen different PBPs expressed, of them: PBP2b which has 2 PASTA domains and SpoVD which has only one. The PASTA domain in SpoVD however, is not essential for the endospore cortex PG assembly (Bukowska-Faniband, 2015).

PASTA domains are also present in serine/threonine protein kinases (STPKs) (Ruggiero, 2011). The STPK in *S. pneumoniae* (StkP) is comprised of an N-terminal phosphatase PhpP and four C-terminal PASTA domains (Osaki, 2009). Since the StkP PASTA domains were shown to bind PG and  $\beta$ -lactams, uncrosslinked PG are thought to act as signals for StkP localization at the division site. The activation of StkP in response to the binding of PASTA domains is proposed to cause the phosphorylation of the cell-division proteins DivIVA and FtsZ (Maestro, 2011).

The PknB STPK in *M. tuberculosis* is another example of a PASTA-containing kinase that binds to muropeptides (Mir, 2011). It regulates many aspects of mycobacterial physiology and is a key component of a signal transduction pathway that regulates cell growth, cell shape, and cell division via phosphorylation of targeted proteins like the DivIVA ortholog Wag31 (Barthe P, 2010; Kang, 2005; Ogawara, 2016). PknB contains four C-terminal PASTA domains for which their binding to PG is dependent on the presence of both D-iGln and m-A<sub>2</sub>pm at the second and third position in the peptide stem, and on the presence of the MurNAc sugar moiety. A proposed model suggested that PG fragments produced at the mid-cell and cell poles act as signalling molecules for the PknB localization via their PASTA domains (Barthe P, 2010; Mir, 2011). Moreover, a close homologue of PknB, PknC in *B. subtilis*, induces germination of dormant spores through interactions with its PASTA domains (Shah, 2008).

PASTA domains from PBPs and STPK display different structural arrangements following ligand binding. In *S. pneumoniae* PBP2x, the two PASTA domains interact with each other to form a compact globular structure (Yeats, 2002), compared to the linear organization of the four PASTAs in the PknB NMR structure (Mir, 2011). The presence of a  $\beta'$ / $\beta$  brace in the place of a  $\alpha$ -helix prevents the interaction between the individual PASTA domains and the formation of a compact structure in a single PknB protein. In PBP2x the close configuration of its two C-terminal domains





**Figure 1.10. Model of *M. tuberculosis* PknB activation promoted by PASTA-PG binding.** In this model, the PknB C-terminal PASTA domains bind to PG precursors or PG hydrolysed fragments produced at the mid-cell and poles. This interactions lead to PknB phosphoryl kinase activation. Authors depicted two PASTA domains in two possible arrangements following ligand binding. Rpf, resuscitation promoting factor (Adapted from Mir M, 2011).

were shown necessary for ligand binding. A model of the activation of the PknB kinase based on the binding of PASTA domains to muropeptides acting as signalling molecules was proposed (Fig. 1.10). This model works on the basis that the mid-cell and poles are active sites of PG synthesis and hydrolysis, and that the high local PG concentration would result in the localization of the PknB PASTA domains. The accumulation of PknB proteins would then lead to dimerization of the kinase domains and its activation, which in turn causes the phosphorylation of cell-division proteins and cell wall synthesis at the cytoplasm. Further functional and structural analyses are needed to corroborate the aforementioned model and to determine whether the number of domains present in a single protein is necessary for PG recognition.

#### **1.2.2.7 SH3 domains**

SH3 [(src) homology 3] domains were first identified in eukaryotic proteins involved in cell signalling, division, and migration. The core ~60-residue region is present in many proteins such as the Src kinase family, the Crk adaptor protein, and phospholipase C- $\gamma$  (Mayer, 1998; Yu, 1994). Eukaryotic SH3 proteins have been characterized as protein-protein interaction modules involved in the recognition of proline-rich sequences, PxxP being defined as the minimal consensus target for binding. Based on the relative positioning of a positively charged residue (R,K), two classes of peptide ligands have been described: class I peptide ligands with the consensus sequence (R/K)xxPxxP and class II ligands xPxxPx(R/K) with the consensus sequence in the opposite orientation. The existence of a negatively charged pocket in the SH3 binding surface defines the orientation of the Class I and II ligands (Feng, 1994; Fernandez-Ballester, 2004; Kaneko, 2008).

The evolution of the protein data base has allowed the identification of a large subset of structurally related SH3-like proteins in bacteria and viruses (Ponting, 1999; Whisstock, 1999). However, they share low sequence similarity to eukaryotic domains. Their sequence and structural conformation will be described in detail below. According to Pfam (May 2020), the SH3 clan is present in 8,500 species, and contains 36 families and a total number of 121, 150 domain sequences (Table 1.2) (El-Gebali, 2019).

**Table 1.2. Lineage distribution of the SH3 domain families.**

SH3 clan				
Family	Eukaryotic	Prokaryotic	Viral	Unique protein structures <sup>a</sup>
SH3_1	+			
hSH3	+			
SH3_2	+			
SH3_9	+			
SH3_10	+			
SH3_15	+			
SH3_19	+			
CAP_GLY	+			
Gemin6	+			
Gemin7	+			
SH3_12	+			
DUF4648	+			
Myosin_N	+			
MLVIN_C	+		+	
SH3_11			+	
SH3_14			+	
DUF1653	+	+	+	1
Ndhs	+	+	+	1
SH3_4	+	+		-
SH3_3	+	+		5
SH3_5		+	+	6
SH3_6		+		1
SH3_7		+		1
SH3_8 (GW)	+	+		3
DUF3104	+	+		-
SH3_16		+		-
SH3_17		+		1
SH3_18		+		1
DUF150_C		+		2
DUF1541		+		2
DUF3247		+		1
DUF4453		+		-
PhnA		+		2
KapB		+		1
DUF3601		+		1
SH3_13		+	+	1

SH3 families with members proposed to bind bacterial CW are highlighted in blue.

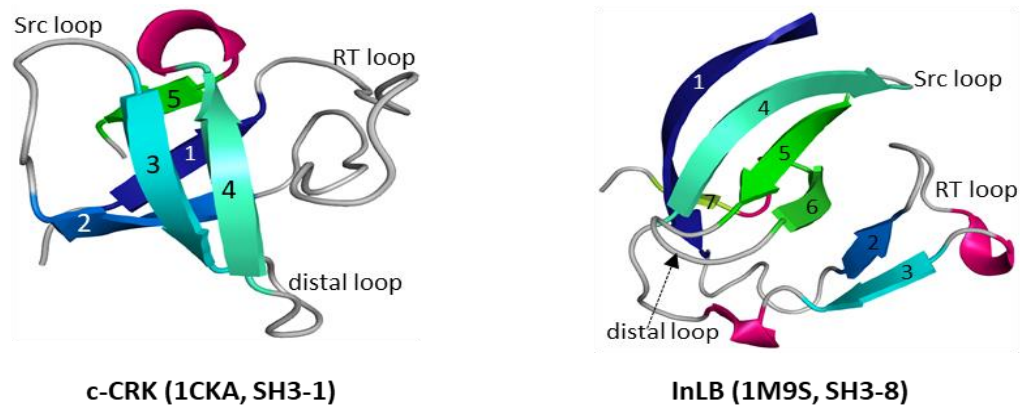
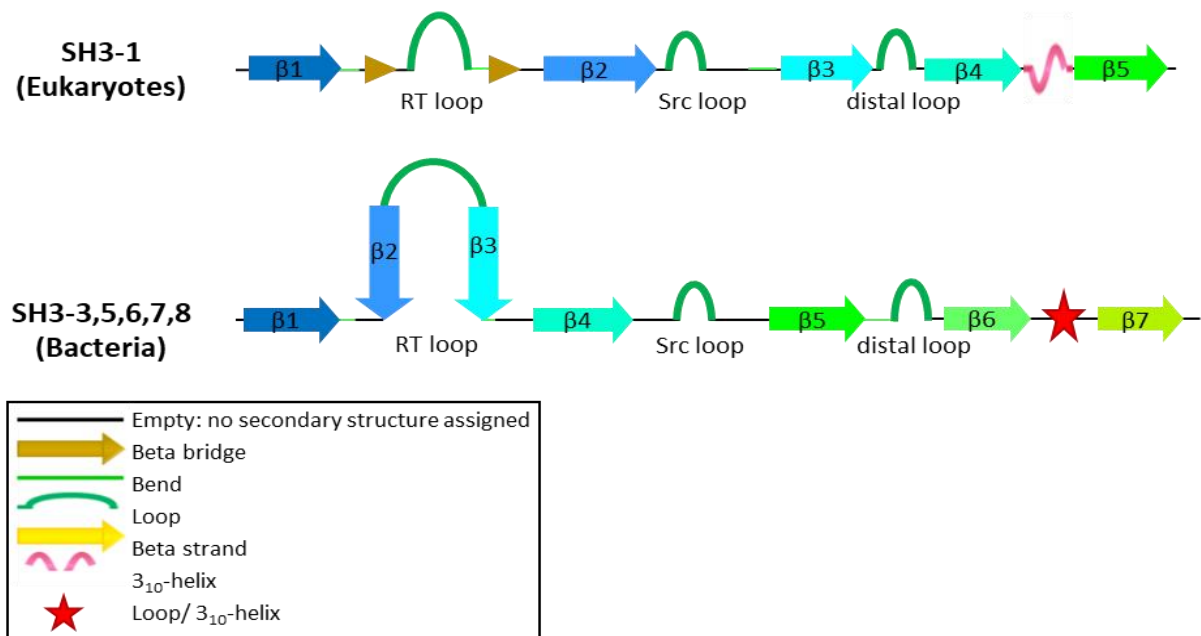
<sup>a</sup> Resolved structures in PDB (protein data bank). Only unique protein sequences were considered.

To date, around thirty bacterial SH3 structures have been determined, and in several families their function remains unknown. Twenty SH3 families have been found in bacteria, five of which have been identified as protein domains involved in non-covalent binding to the cell wall of different species: SH3\_3, SH3\_5, SH3\_6, SH3\_7, and GW (SH3\_8).

The canonical SH3 structure comprises five  $\beta$ -strands arranged into two antiparallel  $\beta$ -sheets or in a  $\beta$ -barrel.  $\beta$ -strands are connected by three loops: the RT, Src, and distal loops, and a short  $3_{10}$  helix. These play important roles in the recognition of binding partners (Fig. 1.11) (Kaneko, 2008; Kurochkina, 2013). In eukaryotes, PxxP peptides bind to a relatively flat hydrophobic groove made of three shallow pockets flanked by the RT loop and the distal  $3_{10}$  helix. The Src loop is at the margin of the substrate binding groove (Fig. 1.11 A) (Feng, 1994; Fernandez-Ballester, 2004; Kurochkina, 2013).

The five  $\beta$ -strands which represent the protein core are relatively conserved in bacterial SH3 domains. The most pronounced structural differences are in the RT loop. Bacterial SH3 have long insertions of amino acids that form secondary structures. Most bacterial SH3 domains have two extra  $\beta$ -strands in the RT loop and, in some cases, additional  $\alpha$ -helices are also present (Fig. 1.11 B). The additional  $\beta$ -strands form  $\beta$ -sheets with the five conserved strands. In many bacterial SH3 structures, strands  $\beta_2$  and  $\beta_6$  interact by hydrogen bonds between the main chain atoms giving the RT loop a more closed conformation. This greatly reduces the accessible surface and creates a geometric incompatibility for the recognition of PxxP ligands (Fig. 1.11) (Kamitori, 2015; Lu JZ, 2006).

There are also other subtle differences between regions in the SH3 cell wall targeting domains that alter the recognition of PxxP ligands. The distal loop is slightly tilted away from the RT loop in prokaryotes when compared to the canonical position in eukaryotes. The extended position of strands  $\beta_4$  and  $\beta_5$  causes the tip of the Src loop to be positioned nearly 180 degrees away from its site in SH3\_1 proteins. These conformational changes have formed surface regions that allow for interactions with common features of the bacterial cell wall (Fig. 1.11).

**A****B**

**Figure 1.11. Structure and topology of eukaryotic and bacterial SH3 domains.** (A) Cartoon representation of the eukaryotic SH3 type 1 domain, c-CRK (PDB entry 1CKA) (left) and the bacterial SH3 type 8, InLB distal C-terminal domain (PDB entry 1M9S) (right). Secondary structures are labelled:  $\beta$ -strands are numbered and rainbow coloured according to the position in the sequence from N- to C-terminus;  $3_{10}$  helices are coloured in dark pink. (B) Topology of the canonical eukaryotic SH3 proteins (Kelley et al.) and a general representation of the topology of bacterial SH3 families (bottom). Secondary structures and loops are depicted.

Despite their structural similarities, eukaryotic and prokaryotic SH3 proteins show low sequence identity (>30%). A multiple sequence alignment between various cell wall binding SH3 proteins (from families 3 to 8) with a eukaryotic SH3\_1 domain is shown in Figure 1.12. In eukaryotic domains, the open conformation of the binding groove is caused by hydrophobic contacts between the highly conserved F7, F9, W35, P49, and Y52 residues (Kaneko, 2008; Kurochkina, 2013). These residues are rarely conserved in bacterial SH3 domains. Glycine 46 is the most conserved residue in eukaryotic and prokaryotic proteins and is part of the Gly-Trp sequence that gives name to the GW (SH3\_8) domains. Whilst there is a high diversity between bacterial SH3 proteins, there are conserved regions in bacterial families that allow favourable contacts with PG peptide structures as shown by the results obtained in this study.

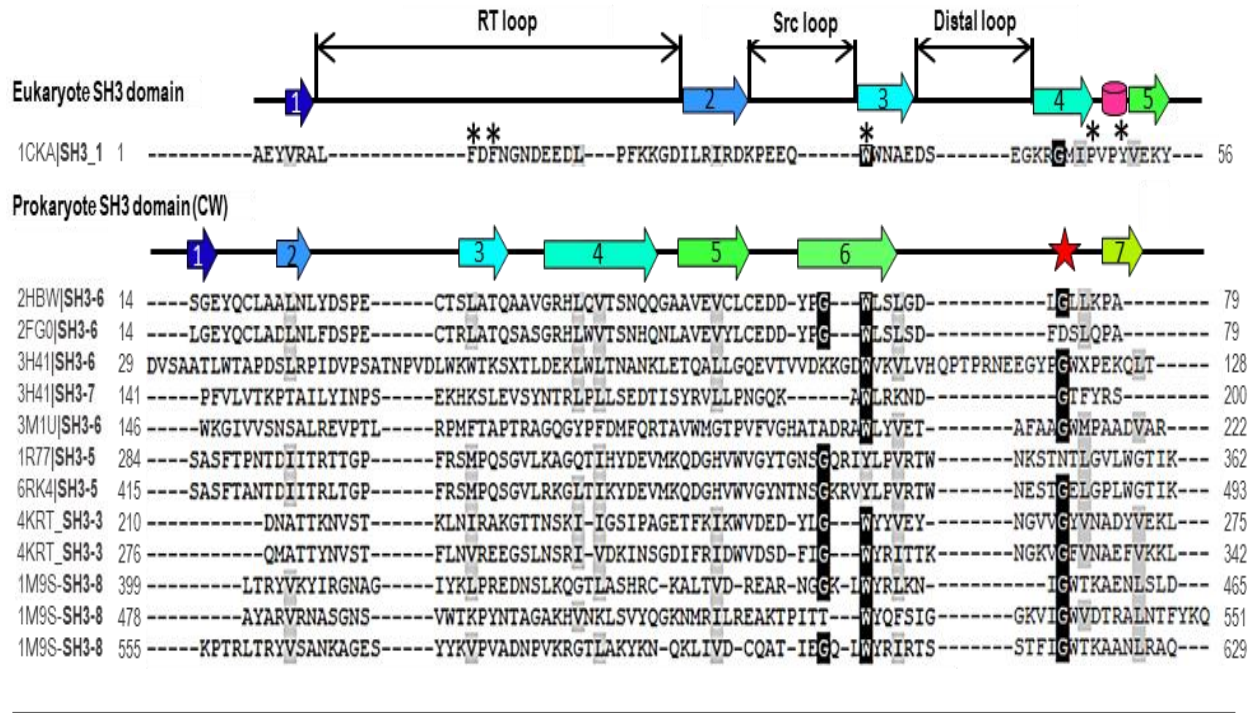
#### **1.2.2.7.1 Structural comparison of SH3 domains targeting bacterial cell walls**

The secondary and tertiary structural arrangements of five families of SH3 domains targeting bacterial cell walls are shown in Figure 1.13. This section provides an overview of the structure-function relationship between these families as described in the literature. The topology of the described SH3 structures was created using the sequence display mapped in the PDB data base.

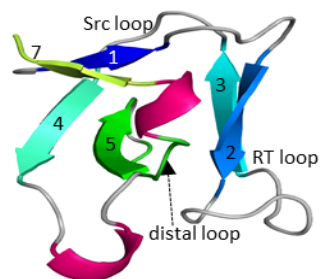
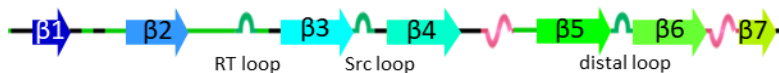
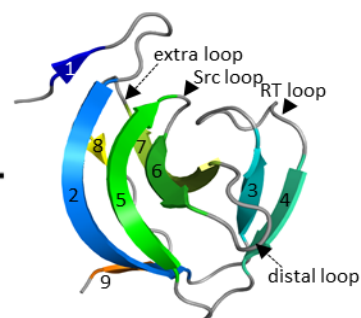
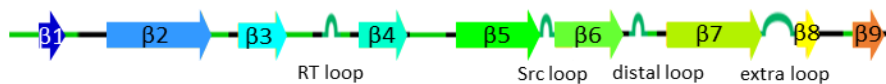
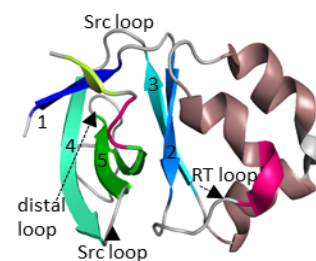
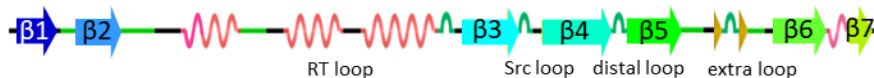
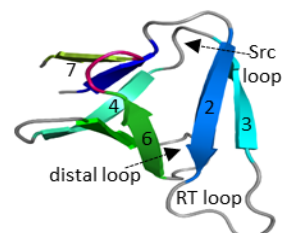
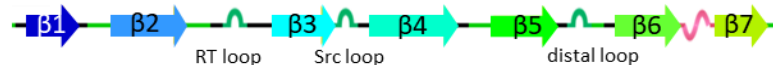
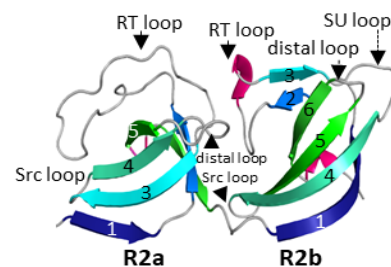
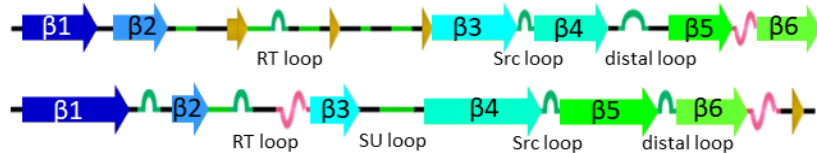
##### **1.2.2.7.1.1 SH3\_3 domains**

SH3\_3 domains were first described as part of bacteriophage endolysins. Endolysins are expressed in the final stage of a lytic cycle. One of the best characterised endolysins harbouring an SH3\_3 protein is the Psm enzyme encoded by the episomal phage phiSM101. Psm displays a specific muramidase activity against *Clostridium perfringens* strains (Nariya, 2011).

Psm harbours an N-terminal catalytic domain belonging to the glycoside hydrolase family 25 and two C-terminal tandem repeats corresponding to the SH3\_3 family domain (Tamai, 2014). The structures of these two SH3\_3 modules are highly similar, despite sharing only 51% sequence identity. Both domains are formed of seven  $\beta$ -strands and the strands  $\beta$ 2 and  $\beta$ 3 are inside the



**Figure 1.12. Multiple amino acid sequence alignment between cell wall targeting SH3 bacterial domains (SH3-3,5,6,7,8) with a eukaryotic SH3 domain (c-Crk).** c-Crk, Mouse (PDB entry 1CKA); AvPCP, *Anabaena variabilis* (PDB entry 2HBW ); NpPCP, *Nostoc punctiforme* (PDB entry 2FG0 ); BcYkfc, *B. cereus* (PDB entry 3H41); DVU-0896, *Desulfovibrio vulgaris* (PDB entry 3M1U); Ale-1, *S. capitis* (PDB entry 1R77); Lysostaphin, *S. aureus* (PDB entry 6RK4); phiSM101, *Clostridium perfringens* type A strain SM101(PDB entry 4KRT); InlB, *Listeria monocytogenes* (PDB entry 1M9S). Conserved residues among all proteins are highlighted in black and residues in light grey indicate conservative substitutions. The positions of  $\beta$ -strands and helices are shown by arrows and cylinders, respectively.  $\beta$ -strands are numbered and coloured in rainbow;  $3_{10}$  helices are coloured in dark pink; the red star represents either a loop or a  $3_{10}$  helix. The position of conserved hydrophobic amino acid residues (F7, F9, W35, P49, and Y552) among eukaryotic SH3 protein is indicated by a (\*).

**A****Psm (4KRT, SH3-3)****B****Lss (6RK4, SH3-5)****C****Ykfc (3H41, SH3-6)****D****Ykfc (3H41, SH3-7)****E****Atl (4EPC, SH3-8)**

	Empty: no secondary structure assigned
	Beta bridge
	Bend
	Loop
	Beta strand
	3 <sub>10</sub> -helix
	Alpha helix



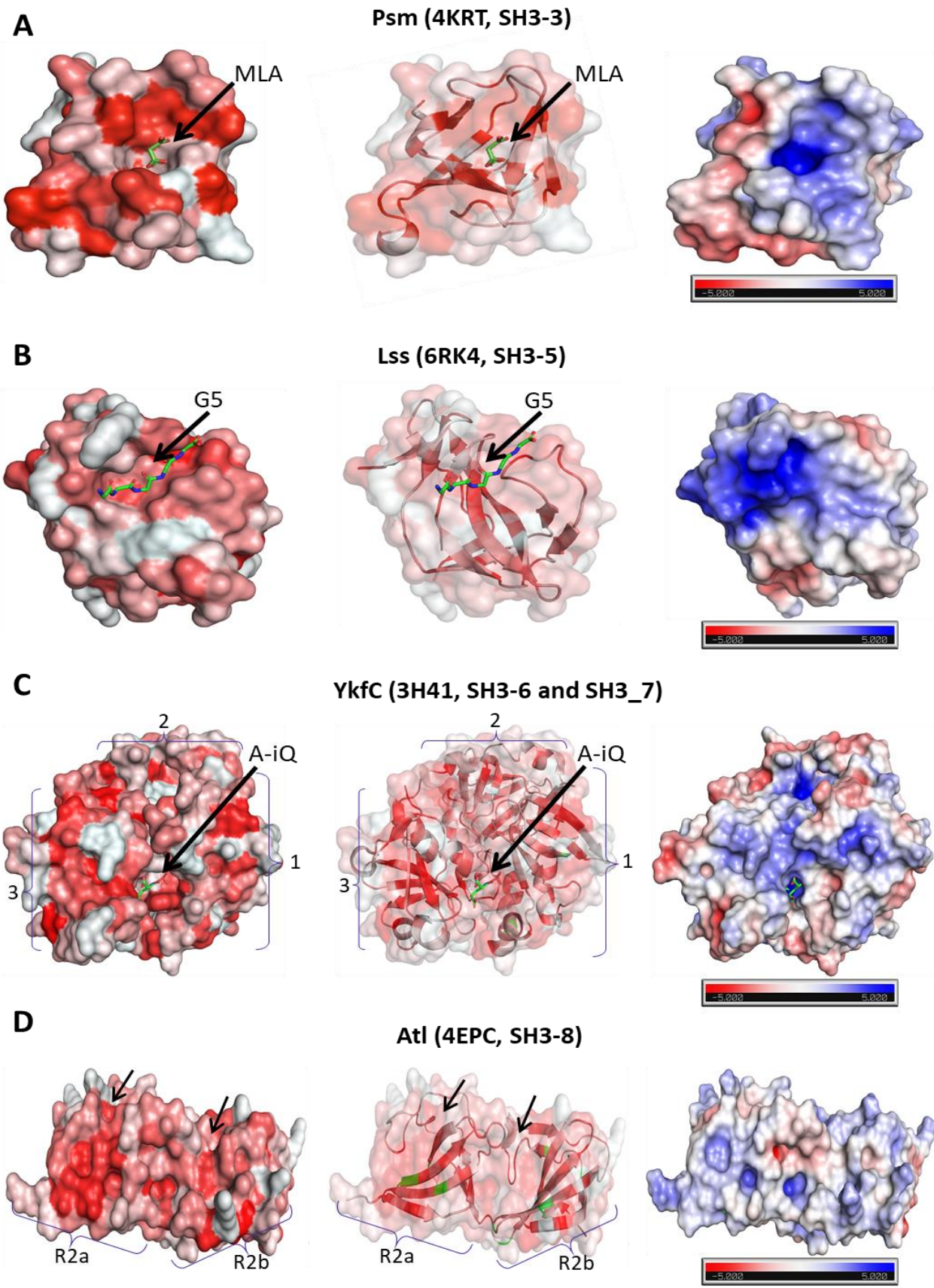
**Figure 1.13. Structural similarities of cell wall binding SH3 domain families.** One protein from each of the five SH3 domain families targeting bacterial cell walls is described; from top to bottom: **(A)** SH3 type 3, Psm endolysin (PDB entry 4KRT); **(B)** SH3 type 5, Lss endopeptidase (PDB entry 6RK4); **(C)** SH3 type 6, YkfC endopeptidase (PDB entry 3H41); **(D)** SH3 type 7, YkfC endopeptidase (PDB entry 3H41) both domains (SH3\_6 and SH3\_7) are found in the same PBD structure from the YkfC endopeptidase; **(E)** SH3 type 8, Atl autolysin (PDB entry 4EPC). The topology map of each protein structure is shown on the left side. Secondary-structure elements and loops are depicted as indicated in the key below the structures. The coloured codes of  $\beta$ -strands and  $3_{10}$  helices are in agreement with Figure 1.11. The corresponding cartoon representation of each protein is shown on the right side. Three dimensional images of folds were prepared with PyMOL (DeLano, 2002) and the graphical representation of the protein architecture was built based on sequence entries as reported in UniProtKB.

RT loop. The topology and structure of the first SH3<sub>3</sub> domain is shown in Figure 1.13 A. A closed hydrophobic RT loop. The topology and structure of the first SH3<sub>3</sub> domain is shown in Figure 1.13 A. A closed hydrophobic surface is formed by  $\beta$ 2 and the distal  $3_{10}$  helix which extends to a positively charged cavity between strands  $\beta$ 2 and  $\beta$ 6 (Fig.1.14). Although the motif recognized by the SH3<sub>3</sub> domains has not been identified, the presence of acetic (ACT) and malonic (MLA) acid from crystallization solutions within the positive cavity of each SH3<sub>3</sub> repeat has been reported (Tamai, 2014). The ACT was bound to the SH3N and MLA to SH3C, both efficiently forming salt-bridge interactions with R224 and R290, respectively (Fig. 1.14). These positively charged cavities have been proposed to accommodate a negatively charged PG peptide sidechain, which is likely to be D-Glu since in *C. perfringens* the L-L-diaminopimelate in position 3 is amidated. *C. perfringens* PG is crosslinked by the L,L-diaminopimelate in position 3 through a glycine interpeptide bridge to a D-alanine in position 4 (Leyh-Bouille, 1970).

#### 1.2.2.7.1.2 SH3<sub>5</sub> domains

SH3<sub>5</sub> domains are formed of 60 to 100 amino acids adopting a characteristic SH3 fold. The SH3<sub>5</sub> domains from lysostaphin (Lss) and its close homologue Ale-1 have been extensively studied. These domains are found in hydrolases that display specific glycylglycine endopeptidase activity against members of the *Staphylococcus* genus, including methicillin-resistant *S. aureus* (MRSA). These C-terminal SH3<sub>5</sub> or SH3b domains have been proposed to drive a preferential binding to the PG pentaglycine cross-links which would confer the specificity of the bacteriolytic enzyme (Gründling, 2006; Lu JZ, 2006). Despite being discovered in 1964 (Schindler, 1964), the recognition of the PG by the Lss SH3<sub>5</sub> domain has not yet been fully characterized. The elucidation of this binding mechanism is the focus of Chapter IV.

SH3<sub>5</sub> domains comprise 7 to 9  $\beta$ -strands connected by the RT, Src and distal loops. Amongst the SH3<sub>5</sub> domains from staphylococcal hydrolases, additional conserved residues at the N-terminus form a unique  $\beta$ -strand ( $\beta$ 1) and the additional N-terminal part of  $\beta$ 2 (Fig.1.13 B) (Lu JZ, 2006). This region has been identified as interacting with the pentaglycine cross-bridges (Lu JZ, 2006; Sabała, 2014) (Fig. 1.14). The SH3<sub>5</sub> domains from Lss and Ale-1 share 84% sequence identity and



**Figure 1.14. Hydrophobicity and surface charge of cell wall binding SH3 domain families.** One protein from each of the five SH3 domain families targeting bacterial cell walls is shown in hydrophobic (red) to non-hydrophobic (white) gradient in a surface and cartoon representation, and on the right their surface electrostatic potential ; from top to bottom: **(A)** SH3 type 3, Psm endolysin (PDB entry 4KRT) with the presence of malonic acid (MLA) located in the proposed binding groove; **(B)** SH3 type 5, Lss endopeptidase (PDB entry 6RK4) bound to a pentaglycine peptide (G5); **(C)** YkfC endopeptidase (PDB entry 3H41) made of two N-terminal SH3 proteins, an SH3\_6 domain (1), an SH3\_7 domain (2), and a C-terminal NlpC/P60 domain (3). The proposed binding pocket for the L-Ala-iQ ligand is located at the SH3\_6-NlpC/P60 interface; **(D)** SH3 type 8, Atl autolysin (PDB entry 4EPC). The two subunits, R2a and R2b, that form the SH3\_8 domain are shown. Their proposed binding grooves are indicated by arrows. No ligand has been identified for any of the SH3\_8 domains.

their structures are almost identical. The Lss SH3\_5 domain is comprised of two  $\beta$ -sheets packed at right angles against each other.  $\beta$ -sheet I is built from strands  $\beta$ 5-  $\beta$ 7 and the N-terminus of  $\beta$ 2, whereas  $\beta$ -sheet II is made from  $\beta$ 3- $\beta$ 4,  $\beta$ 8 and the C-terminus of  $\beta$ 2. Strands  $\beta$ 3 and  $\beta$ 4 are inside the RT loop, in contrast with SH3\_5 proteins from the *Streptococcus*, *Enterococcus*, and *Lactococcus* genera that exhibit the canonical insertion of  $\beta$ 2 and  $\beta$ 3 strands in their RT-loop. In the staphylococcal SH3\_5 domain there is a relatively long loop called the 'extra loop' between strands  $\beta$ 7 and  $\beta$ 8 in the place occupied by a short  $3_{10}$  helix in eukaryotic SH3 domains (Fig. 1.13-1.14).

The structure of the Lss SH3\_5 domain in complex with a pentaglycine peptide has been determined (Mitkowski, 2019). As previously suggested the pentaglycine (G5) ligand sits in a groove built by the side chains of the N-terminal extra conserved residues ( $\beta$ 1-  $\beta$ 2) (Hirakawa, 2009; Lu JZ, 2006). The bound G5 adopts an extended conformation between these strands to the RT loop side (Fig. 1.14). Based on distance, the geometry of the binding site is not compatible with backbone hydrogen bond formation. Instead the G5 interacts with the side chains of residues from both  $\beta$ 1 and  $\beta$ 2 and from the RT loop. The carbonyl oxygen atom of the first G from G5 accepts a hydrogen bond (donor-acceptor distance 3.1 Å) from the side chain carboxamide of residue N405, located at the end of  $\beta$ 1.

Many staphylococcal SH3\_5 domains have been described in the literature, however most SH3\_5 domains are found in *Lactococcus*, *Streptococcus*, and *Enterococcus* genera (El-Gebali, 2019). The *Staphylococcus* genus only represents 10% of the species harbouring SH3\_5 domains. There are six unique structures available in Pfam and four of them are *Staphylococcus* related proteins. Further structural and functional studies are required to determine what is the consensus and the specificity across different SH3\_5 domains.

#### **1.2.2.7.1.3 SH3\_6 and SH3\_7 domains**

The presence of two distinct N-terminal tandem SH3 domains, previously named SH3\_51 and SH3\_52, in a family of  $\gamma$ -D-glutamyl-L-diamino acid endopeptidases has been described (Xu,

2010). These SH3 domains are structurally related and have been identified as SH3 type 6 and type 7, respectively. These two domains (SH3\_6 and SH3\_7) are found in the YkfC endopeptidase from *Bacillus cereus* (Fig. 1.13 C-D and Fig. 1.14), in its ortholog (BcYkfC) from *Bacillus subtilis*, and in a putative  $\gamma$ -D-glutamyl-L-diamino acid endopeptidase from *Desulfovibrio vulgaris* strain *Hildenborough* (PDB 3M1U). In all three enzymes the N-terminal SH3 domains are linked to a C-terminal NlpC/P60 cysteine peptidase that has been proposed to cleave the linkage between D-Glu and m-A<sub>2</sub>pm (Xu, 2010).

Although these domains adopt a typical SH3 fold with a similar five stranded core, they display much larger structural differences. The SH3\_6 domains are approximately 90 amino acids long; they conserve the structural insertion of  $\beta$ 2 and  $\beta$ 3 in the RT loop region but lack the typical  $\beta$ -hairpin motif compared to other members of the bacterial SH3 clan. Instead, *Bacillus* endopeptidases have 30 additional amino acids forming a novel helical insertion ( $\alpha$ 1-3) between strand  $\beta$ 2 and  $\beta$ 3 (Fig.1.13 C), whilst *Desulfovibrio vulgaris* also presents a long insertion in the RT loop but no additional secondary structures.

The SH3\_7 domain, composed of 60 amino acids, has a more conserved SH3-like structure with seven  $\beta$ -strands. Strands  $\beta$ 2 and  $\beta$ 3 are inserted in the RT but no other significant insertions are present. SH3\_7 sits between the SH3\_6 and the catalytic domain (Fig.1.13 D). They are arranged in a triangular shape such that each of them interacts with the other two (Fig. 1.14). The interface between the two SH3 domains is mostly hydrophobic and is centred on interactions between the RT and Src loops of SH3\_6 and the  $\beta$ 3- $\beta$ 4 strands and the distal  $\alpha$ 10 helix of SH3\_7.

The presence of an SH3\_6 domain has also been reported in two  $\gamma$ -D-glutamyl-L-diamino acid endopeptidases from cyanobacteria. *Anabaena variabilis* (AvPCP) and *Nostoc punctiforme* (NpPCP) (Xu, 2009) AvPCP and NpPCP have almost identical structures and share 80% sequence similarity. These proteins have N-terminal SH3 domains with the characteristic 30 amino acid insertion forming three helical structures in the RT loop. The SH3\_6 domain from the YkfC endopeptidase is located at an equivalent position to the SH3\_6 domains from cyanobacterial

AvPCP and NpPCP proteins. However, SH3\_6 is linked to the catalytic NlpC/P60 peptidase by a long loop occupying the space of the SH3\_7 in YkfC.

No ligand has been identified for any of the SH3 domains. However, SH3\_6 has been suggested to recognize the PG crossed-linked stem peptide. In YkfC, the active site is located at the SH3\_6–NlpC/P60 interface providing a favourable pocket for the binding of L-Ala where the carbonyl group of E83 and the side chain of a conserved Y residue (Y118 in YkfC and Y64 in AvPCP) are predicted to form hydrogen bonds with the amino group of L-Ala (Xu, 2015) (Fig. 1.14).

The function of SH3\_7 is unknown. In YkfC, this domain is distal to the active site. It has been proposed that the nonessential SH3\_7 domain was evolutionarily lost over time in the cyanobacterial endopeptidases (Xu, 2010).

#### **1.2.2.7.1.4 GW (SH3\_8) domains**

The SH3\_8 family is one of the best studied bacterial SH3 families. The SH3\_8 domains, better-known as GW domains, are cell wall binding modules of about 80-90 amino acids named after a conserved Glycine-Tryptophan (GW) dipeptide (Braun, 1997). Over the last 20 years a large number of GW domains have been described. In prokaryotes, these domains are mainly found in Firmicutes of the *Lactobacillus*, *Bacillus*, *Enterococcus*, *Listeria*, and *Staphylococcus* genera.

In *S. aureus* and *S. epidermidis*, the autolysin Atl plays a major role in cell separation. The Atl enzymes from both species have the same modular organization: signal peptide (SP), pro peptide (PP), *N*-acetylmuramyl-L-alanine amidase (Ami), three major repeats (R1-R3, each one built of two subunits (a) and (b)) which have been identified as GW domains, and an *N*-acetylglucosaminidase (GlcNAse). Following cleavage of the SP, external Atl is proteolytically cleaved at two positions, after the PP and after repeat R2, leading to the formation of the mature AM-R1-R2 and R3-GL enzymes (Heilmann, 1997; Zoll, 2012).

The crystal structure of the R2 domain has been determined by X-ray crystallography (Zoll, 2012). Each subunit carries a conserved GW motif exhibiting the typical SH3 fold (Fig.1.13 E and Fig.

1.14)). Both domains resemble a half-open  $\beta$ -barrel formed by a semi-circular  $\beta$ -sheet made of strands  $\beta$ 1,  $\beta$ 3- $\beta$ 5 of R2a and strands  $\beta$ 1,  $\beta$ 4-  $\beta$ 6 of R2b. Their loops were named in accordance to the SH3 nomenclature, except for an extra loop named the SU loop. The most conserved residues from both subunits are located around the GW motif. However, these regions exhibit significant structural differences. In R2a the conserved cluster around the GW motif is centre around a wide and shallow hydrophobic groove formed with the RT loop (Fig. 1.14). Both RT loops are especially prominent. In R2a it adopts a parallel orientation relative to the central  $\beta$ -sheet which creates the shallow cavity. R2b presents a deep cavity on the opposite side of the protein which is formed between the tip of its RT loop and the conserved GW motif (R2a, amino acids 762 to 763; R2b, amino acids 838 to 839) (Fig. 1.14).

GWs are often part of multi modular proteins with additional enzymatic domains which exhibit PG hydrolytic activity. The internalin protein B (InlB) from *Listeria monocytogenes* is an exception. InlB mediates the entry of endothelial, epithelial, and hepatocytic cells. InlB contains three C-terminal GW domains. A search for homologous proteins using the *DALI* server (Holm, 2008) identified the GW domain of *L. monocytogenes* InlB as the closest homologue of the *S. aureus* and *S. epidermidis* AtIA R1ab and R2ab repeats (Zoll, 2012).

The LTA cell wall polymers have been implicated as the binding receptors for the GW domains (Jonquière, 1999; Zoll, 2012). In both studies, based on *L. monocytogenes* and *S. aureus*, respectively, InlB and AtIA were proposed to recognise LTA polymers by their GW domains using fixed purified polymers in ELISA experiments. Binding to LTA polymers has also been suggested based on experimental data using fluorescent-labelled GW repeats on the surface of *S. aureus* WT cells and the LTA-deficient strain (RN4220) (Zoll, 2012). However, the initial *S. aureus* experiments indicated a direct interaction between AtIA GW repeats and isolated PG (Biswas). In this study the AtIA Ami-R1abR2ab and the R1ab-R3ab repeats were incubated with WT PG, PG lacking O-acetylation, and PG lacking WTA. Both recombinant proteins (Ami-R1abR2ab and R1ab-R3ab) displayed a dose-dependent binding to all isolated PG sacculi with no evidence of preferential binding to either of the modified PG structures. A more recent study in *L.*

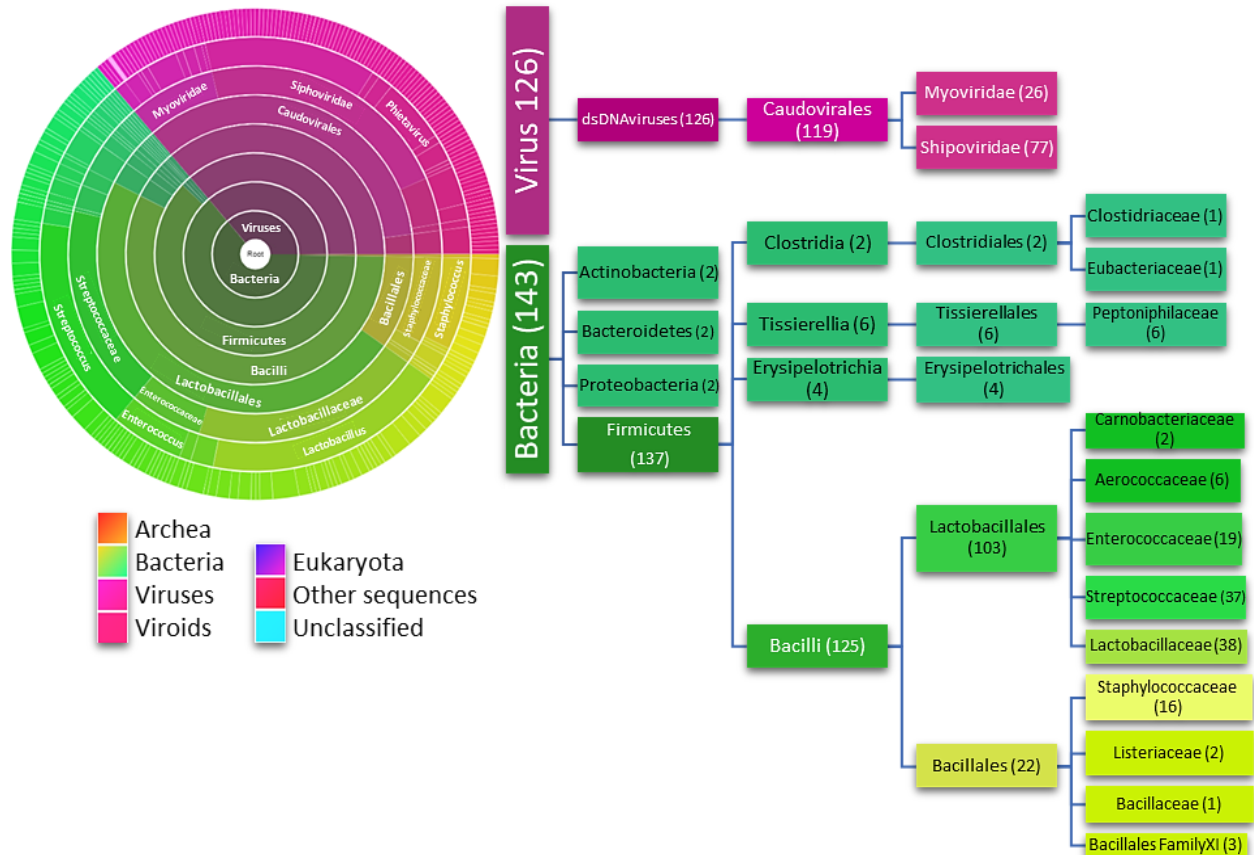


*monocytogenes* determined direct binding to PG by the InIB GW domains (GW<sub>1-3</sub>) and the GW repeats of the autolysin Ami (GW<sub>1-8</sub>) (Percy, 2016). This study used Western blot analyses of InIB proteins and a variant of InIB (with replaced native GWs for the eight GW repeats of the autolysin Ami: InIB-GW<sub>Ami</sub>), from cells and supernatant fractions of *L. monocytogenes*. This showed that GW-containing proteins are retained in the cell wall of *L. monocytogenes* mutant strains lacking D-Ala and galactose LTA modifications or the complete LTA polymer, as well as in WTA-negative strains. In the same study, the eight Ami GW repeats were produced as MBP-fusions and their binding to purified PG isolated from WT or WTA deficient *L. monocytogenes* was also determined (Percy, 2016). Although there was no direct evidence of the recognition of LTA by the GW domains in this study, a dual binding recognition mode was suggested by the authors. It is important to note that these studies used synthetic LTA polymers for their binding tests, so a dual-binding hypothesis should therefore be treated with caution. Overall, the recognition of LTA by GWs therefore remains an open question due to limited experimental data.

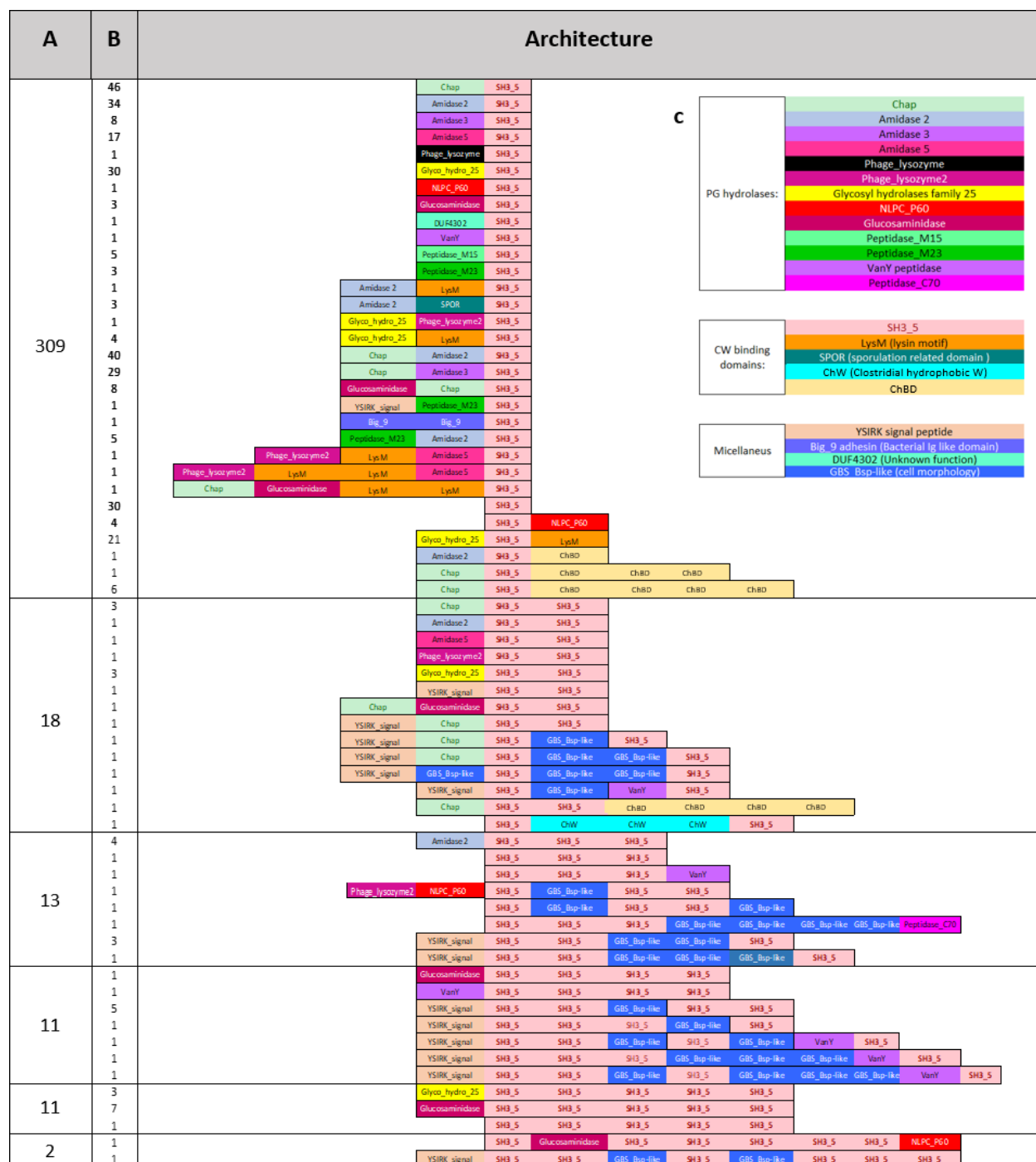
#### **1.2.2.7.2 Species distribution and domain organization of the SH3\_5 family**

SH3 domains from the type 5 family are found in bacteria and dsDNA viruses (Fig. 1.15). According to Pfam (June,2020) there are over 270 species potentially producing SH3\_5 domain proteins (El-Gebali, 2019). The distribution between bacteria and viruses is almost equivalent (53% are present in bacteria and 47% in bacteriophages).

From 143 SH3\_5 sequences identified in bacteria, over 95% are found in Firmicutes. Of these, 91% are found in members of the Bacilli class, with Lactobacillales being the most abundant order. Members of the *Lactobacillaceae*, *Streptococcaceae*, and *Enterococcaceae* families harbour the highest number of identified SH3\_5 sequences i.e. 38, 37, and 19, respectively. In the order of Bacillales there are 22 identified SH3\_5 sequences, of which 16 belong to the *Staphylococcaceae* family. All phage-encoded proteins harbouring SH3\_5 domains have been associated to PG hydrolases exclusively targeting Firmicutes.



**Figure 1.15. Distribution of SH3\_5 sequences across species.** Sunburst chart illustrating the phylogenetic distribution of the SH3 type 5 family (May 2020). The SH3\_5 family is present in bacteria (342 sequences, 143 species) and viruses (126 species and sequences). In bacteria, this protein family is predominantly found in the Firmicutes phylum. The sunburst image was obtained from the Pfam database (El-Gebali, 2019). The tree was built by considering the taxonomic lineage of each sequence that has a match to this family and coloured according to the assignment in Pfam.



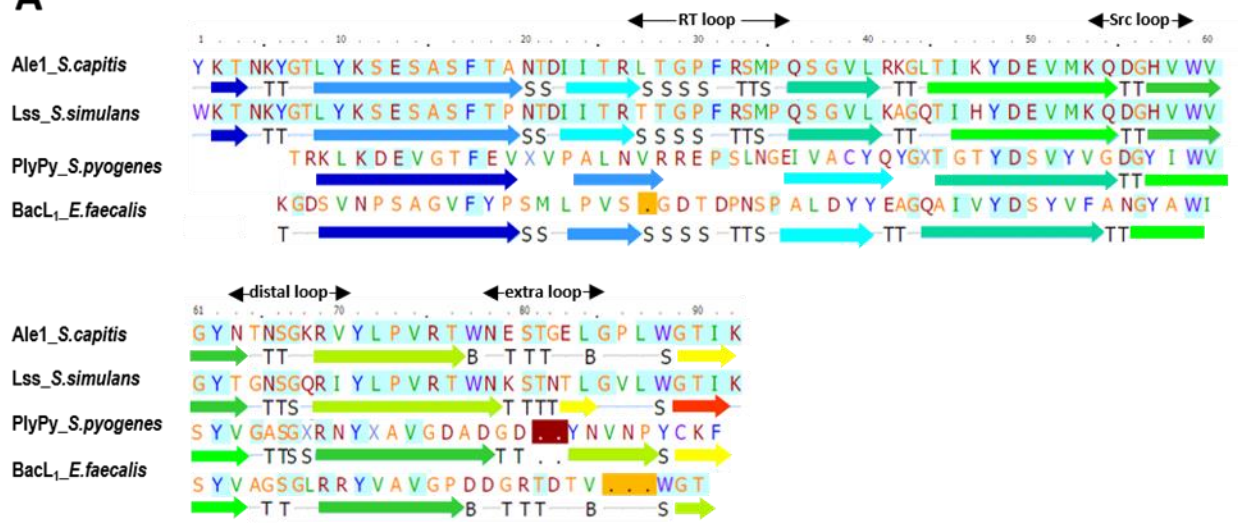
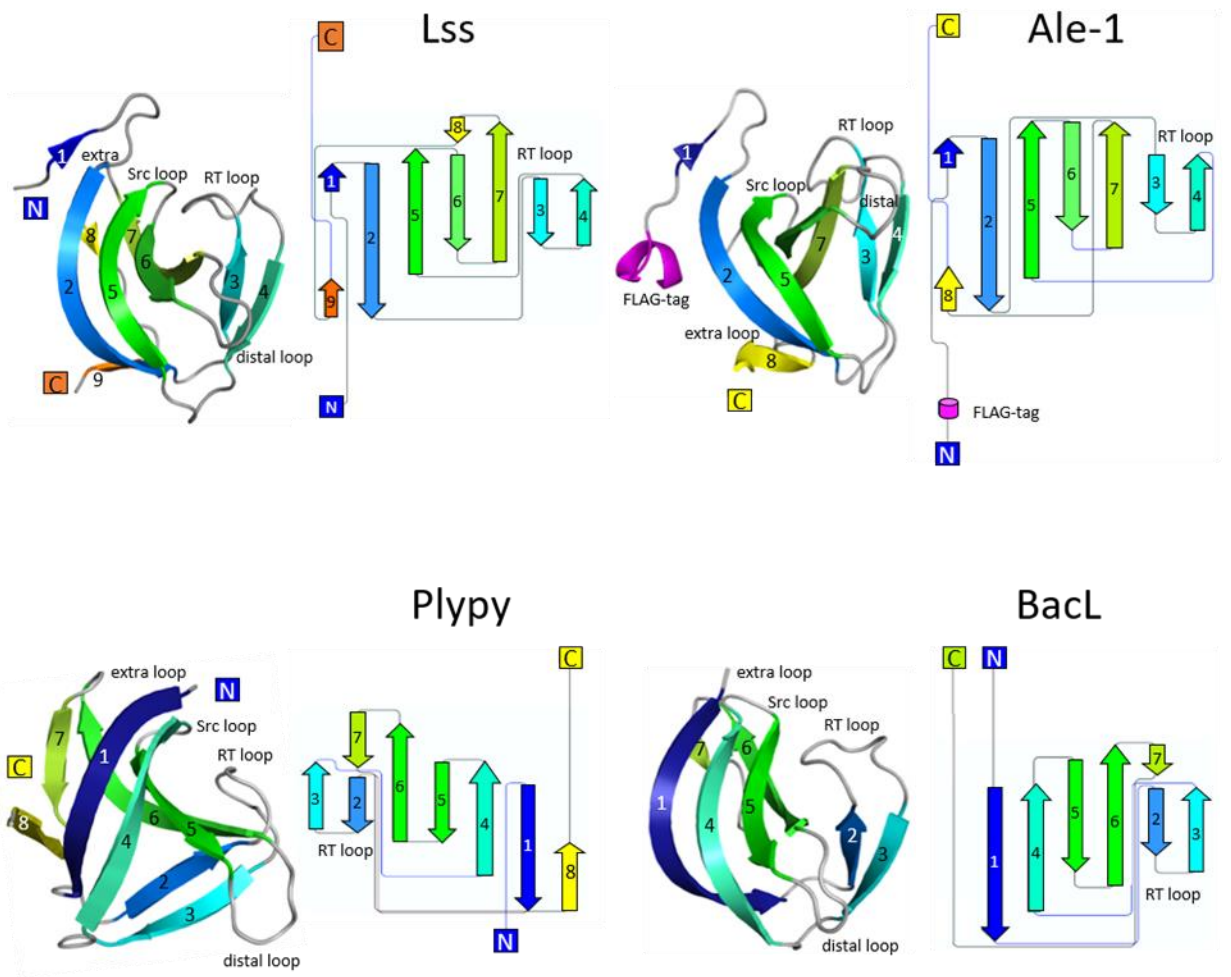
**Figure 1.16. Modular organization of proteins containing SH3\_5 domains across species.** Modular proteins were divided based on the number of SH3\_5 repeats. Column (A) shows the total number of sequences with an identical number of SH3\_5 repeats. Column (B) shows the distribution of sequences following a specific modular organisation. A schematic representation of proteins with SH3\_5 domains is shown.

Figure 1.16 shows the domain organization in which the SH3\_5 domain is found. The most common architecture is made of one or more N-terminal PG hydrolases followed by a single C-terminal SH3\_5 domain. As the number of SH3\_5 repeats increase, the number of identified sequences gradually decrease. Only two sequences with a maximum of six SH3\_5 repeats have been identified; a multi modular protein from *Lactobacillus lindner* containing two PG hydrolases, and a protein from group B *Streptococcus* sp. HSISS2 (GBS) that has been implicated in the control of cell morphology.

SH3\_5 domains are commonly found in PG hydrolases that have additional types of cell wall binding domains. This would allow multiple interactions with different PG substrates. YSIRK is a Gram-positive signal peptide, mainly present in *Streptococcus* and *Staphylococcus* genera, which appears to be involved in secretion of surface proteins to the cell wall, however the mechanisms supporting this type of protein precursor trafficking are unknown (DeDent A, 2008; Yu, 2018). SH3\_5 domains have also been identified in proteins involved in cell adhesion, like Big\_9 from the intimin/invasin family. Big\_9 proteins have a similar topology to the eukaryotic immunoglobulin superfamily. In most intimin/invasin proteins, Big-9 domains appear in tandem followed by a C-terminal C-type lectin-like domain, however they can also occur in combination with cell wall targeting domains like LysM and/or SH3\_5 domains (Heinz, 2016).

#### **1.2.2.7.3 Structural similarities between SH3\_5 domains.**

An amino acid sequence alignment of four members of the SH3 type 5 family, generated in the phyre2 web portal (Kelley, 2015) is shown in Figure 1.17 A. As mentioned previously, Ale-1 and Lss are hydrolases targeting members of the *Staphylococcus* genus, and their SH3\_5 domains typically have additional N-terminal amino acid residues when compared to the SH3\_5 domains of other genera. Plypy is a phage-associated cell wall hydrolase from *S. pyogenes* M1. This hydrolase has an N-terminal CHAP domain (cysteine, histidine-dependent amidohydrolases/peptidases) linked to a C-terminal SH3\_5 domain. CHAP domains are commonly associated with SH3\_5 domains (Fig. 1.16). Many proteins with CHAP domains remain uncharacterized but in LysK and Twort staphylolytic phages it has been shown that the CHAP

**A****B**

**Figure 1.17. Sequence and structural similarities within four SH3 type 5 domains. (A)** Amino acid sequence alignment of four members of the SH3\_5 family produced by Phyre2 web portal (Kelley, 2015): Ale-1, *S. capitis* (PDB entry 1R77); Lysostaphin, *S. aureus* (PDB entry 6RK4); Plypy, *S. pyogenes* (PDB entry 5UDM); BacL<sub>1</sub>, *E. faecalis* (UniProtKB entry B1B1Z5). Conserved residues among all proteins are highlighted in blue. The positions of  $\beta$ -strands are shown below each sequence.  $\beta$ -strands are coloured coded in agreement with Figure 1.11. The position of the loops in the protein structures is indicated at the top of each sequence. **(B)** Structure and topology of SH3\_5 domains. The cartoon representation with secondary structure elements is shown on the left side of each SH3\_5 protein. The architecture of the folds is illustrated schematically on the right of each structure.  $\beta$ -strands are shown as arrows. The numbering and coloured codes of secondary structures are consistent with section A. Loops are labelled in the cartoon representations. 3D images of folds were prepared with PyMOL Molecular Graphics System and 2D protein topology diagrams were built using Pro-origami bioinformatics (Stivala A, 2011).

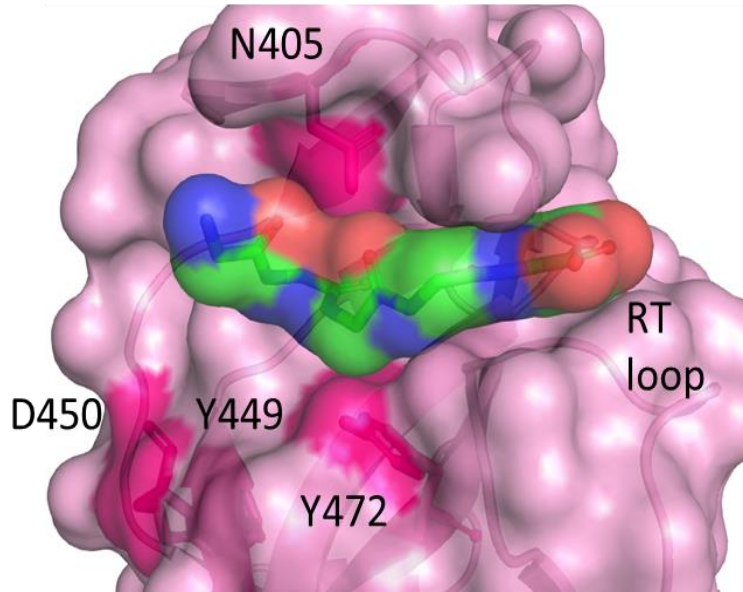
domain displays amidase activity, cleaving the chemical bond between MurNAc and L-Ala at the N-terminus of the stem peptides (Becker, 2009). In the Twort phage endolysin the lytic activity of the CHAP domain is enhanced by the C-terminal SH3\_5 domain (Becker, 2015).

In *S. pyogenes* PG the amino acid L-lysine in position 3 is crosslinked by an L-ala-L-ala interpeptide bridge to a D-alanine in position 4. Recent studies using a combination of LC-MS spectrometry and FRET-based assays with PG-like fluorophore-quencher synthetic peptides showed that *S. pyogenes* Plypy displays a D-alanyl-L-alanine endopeptidase activity (Lood, 2014). It has been determined that Plypy displays a specific activity towards a limited number of bacterial species with D-Ala–L-Ala bonds (Lood, 2014). However, other determinants are required for the cleavage of Plypy since it has shown that species such as *S. mutans* and *E. faecalis* harbouring the same interpeptide composition as *S. pyogenes* proved to be insensitive to Plypy activity. This phenomenon was also observed in the phage lysin GBS from *S. agalactiae* bacteriophage, which displays a D-alanyl-L-alanine endopeptidase activity and was not active against *S. mutans* PG. It has been speculated that high levels of polymers like WTA could block the access of these hydrolases (Pritchard, 2004).

In *E. faecalis*, the bacteriocin 41 (Bac41) produced by clinical isolates consists of six open reading frames (ORFs). Two of them, *bacL1* and *bacA*, are co-expressed as surface proteins and confer the bactericidal activity of Bac41; BacA is presumed to be an accessory factor (Kurushima, 2013). BacL<sub>1</sub> is a  $\gamma$ -D-glutamyl-L-lysine PG endopeptidase and comprises two distinct N-terminal domains associated with PG hydrolysis and three C-terminal SH3\_5 domains (Kurushima, 2015). The first N-terminal domain shows homology to the bacteriophage-type PG hydrolase followed by a homologue of the NlpC/P60 PG hydrolase family. The description of the SH3\_5 repeats from the BacL<sub>1</sub> endopeptidase is addressed in greater detail in Chapter VI. Figure 1.16 shows the sequence and structure of the first BacL<sub>1</sub> SH3\_5 domain (closest to the N-terminal region). Binding of BacL<sub>1</sub> to Gram-positive bacteria harbouring an L-Ala-L-Ala crosslinked PG structure was shown in *E. faecalis*, *S. pyogenes*, and *S. pneumoniae* (Kurushima, 2015). However, the bactericidal activity of Bac41 is strictly specific to *E. faecalis*, including VanB-type vancomycin-resistant strains, and

requires the presence of the two extracellular components, BacL<sub>1</sub> and BacA proteins. The SH3\_5 domains from *E. faecalis* BacL<sub>1</sub> exhibit a higher structural identity to the Plypy SH3\_5 domain from *S. pyogenes* compared to the SH3\_5 domains from staphylolytic hydrolases. (Fig. 1.17 B). The binding mechanisms of the SH3\_5 domains from BacL<sub>1</sub> and Plypy have not been investigated. Amongst the four SH3\_5 proteins shown in Figure 1.17, the major structural elements are well maintained: five  $\beta$ -strands ( $\beta$ 3 -  $\beta$ 7) forming the core of the two antiparallel  $\beta$ -sheets, the C-terminus of strand  $\beta$ 2, and the  $\beta$ -hairpin conformation between the RT loop (Fig. 1.17 B). Studies investigating Lss and Ale-1 SH3\_5 domains have shown that some of the most conserved residues within the proteins (shown in Figure 1.16 A) are either near or form part of regions identified as interacting with the peptide moieties of *S. aureus* PG (Lu JZ, 2006; Mitkowski, 2019; Tossavainen, 2018). According to the numbering in the sequence alignment, the amino acid residues Y48 (Lss Y449; Ale-1 Y318), D49 (Lss D450; Ale-1 D319), and Y71 (Lss Y472; Ale-1 Y341) are located around the pentaglycine binding groove of the staphylococcal SH3\_5 domains (Fig. 1.17 and 1.18). Residues Y48 and D49 are part of strand  $\beta$ 5, which forms one of the side walls of the groove. Polar and charged residues are involved in direct interactions that stabilize  $\beta$ -sheet I, for example D49, which forms a salt-bridge with the conserved R residue in position 69 (Lss R470; Ale-1 R339), (Lu JZ, 2006). In the staphylococcal SH3\_5 domains, the tyrosine amino acid residue in position 71 is part of strand  $\beta$ 7. This strand creates the base of the pentaglycine groove with the RT loop (Fig.1.16 B). In Plypy and BacL<sub>1</sub> SH3\_5 domains, the Y71 residue is part of the N-terminal region of strand  $\beta$ 6 which is in close contact to their RT loops. The importance of some conserved residues like F17 (Lss F418; Ale-1 F286), R69 (Lss R470; Ale-1 339), and W88 (Lss W489 ; Ale-1 W358) and their role in the recognition of PG stem peptides is the centre of discussion of Chapters IV and V.





**Figure 1.18. Lysostaphin SH3\_5 protein bound to a pentaglycine peptide (PDB 5LEO).** Close up view of the pentaglycine binding groove from the lysostaphin SH3\_5 domain. The RT loop and residues interacting with the pentaglycine ligand are indicated.

### 1.3 Project Aims and Objectives

The identification of new SH3\_5 domains has undergone a drastic expansion in the last decade. As mentioned before, these cell wall binding domains are part of a diverse range of PG hydrolases that play key roles in cell wall growth and regulation, and in other cases displaying bacteriolytic activities against other bacterial species in competition for a niche. The presence of SH3\_5 domains in diverse organisms with different PG compositions raises questions about the specificity of these proteins. However, as it has been mentioned in eukaryotic SH3 domains, this complexity exists within certain moderation (Saksela, 2012). Sequence and structure alignments between members of the type 5 family suggests the existence of conserved features necessary for the recognition of canonical ligand binding moieties. This phenomenon is not specific to SH3 proteins, as it has been shown in the LysM carbohydrate binding modules present across all kingdoms (Mesnage, 2014).

To contribute to the understanding of how bacteria control the enzymatic activity of PG hydrolases, during this project we aimed to elucidate the mechanisms underpinning ligand recognition by bacterial SH3\_5 domains. As a model system, the C-terminal SH3\_5 protein of the lysostaphin (Lss) bacteriocin was studied. In addition, we also aimed to explore SH3\_5 binding recognition across species and the role of protein modularity in the enzymatic activity of proteins harbouring multiple SH3\_5 domains.

To gain insight into the binding mechanism of the Lss SH3\_5 to PG, chapter III, IV and V describe the characterisation of the protein-substrate recognition through functional and structural analyses using NMR titrations, X-ray crystallography and mutational analyses. Briefly, chapter III covers the development of the first two objectives of this research: i) to produce the required recombinant Lss SH3\_5 proteins and ii) to generate a panel of ligands with increasing complexity derived from *S. aureus* PG composition.

As little was known about the SH3\_5 binding specificity and affinity to PG, and the minimal ligand recognised by Lss SH3\_5 proteins, in chapter IV two main objectives were explored: i) to identify and map the residues interacting with the PG fragments generated in chapter III and ii) to measure affinities to determine the minimal PG motif recognised by the SH3\_5 domains.

The affinity of an eukaryotic SH3 domain for peptide ligands typically ranges from 1 to 100  $\mu$ M (Lim, 1994; Mayer, 2001; Saksela, 2012). This weak affinity may be desirable under certain circumstances where protein complexes need to assemble and disassemble to allow processive catalysis. Homodimerization between SH3 proteins has been previously described in eukaryotic SH3 domains (Kaneko, 2008; Kristensen, 2006). On the other hand, the presence of proteins harbouring multiple SH3 domains highlights the complexity of the interactions governing the SH3\_5-ligand binding. Here we combined NMR and X-ray crystallography to elucidate binding affinities and specificity of the Lss SH3\_5 proteins.

In chapter V, we sought to determine the contribution of the SH3\_5 residues identified by NMR and X-ray crystallography to the recognition of two PG peptide motifs, the pentaglycine crossbridges and the peptide stems. Site-directed mutagenesis was used to explore three main objectives: i) to confirm the role of six chosen residues to the recognition of the pentaglycine (G5) and peptide stem (P4) ligands by NMR titrations; ii) to investigate the contribution of sixteen residues in binding to the natural substrate of the enzyme, the *S. aureus* PG sacculi, using a quantitative *in vitro* binding assay; and iii) to explore the impact of the same SH3\_5 mutations on the enzymatic activity of the mature Lss.

Our work on the characterisation of the Lss SH3\_5 proteins shed light into a two-site recognition of *S. aureus* PG by a single protein domain. The two binding surfaces are located on opposite sides of the domain, thus allowing only one SH3\_5 domain to bind to the pentaglycine crossbridge and a second SH3\_5 domain to bind to the contiguous peptide stem. This binding mechanism would permit the clustering of proteins on the PG. The work presented in chapter VI aimed to answer two main questions: i) Is the Lss binding mechanism conserved across other

species? and ii) does multimodular proteins i.e. proteins harbouring multiple SH3\_5 domains recognise PG independently or do they form quaternary structures? To explore these questions, the aim of chapter VI was to investigate the binding mechanism of two distinct *E. faecalis* proteins containing a single and multiple SH3\_5 domains. *E. faecalis* PG is crosslinked by two L-Ala amino acids instead of a pentaglycine like in the case of *S. aureus*. The selected proteins were: i) the single SH3\_5 domain from the predicted EF1293 hydrolase, and ii) the three SH3\_5 repeats from the BacL<sub>1</sub> bacteriocin. The specific objectives described in this chapter were i) to produce: the single C-terminal SH3\_5 domain from the putative PG endolysin EF1293 protein, and five variants of the three modular BacL<sub>1</sub> SH3\_5 repeats, i.e. SH3<sub>1</sub>; SH3<sub>2</sub>; SH3<sub>3</sub>; SH3<sub>1-2</sub>; SH3<sub>1-3</sub>; ii) to perform NMR structural and functional analyses using a set of ligands corresponding to *E. faecalis* PG: the L-Ala-L-Ala, cross-links, the AQKAA, peptide stem and the AQK(L-Ala-L-Ala)AA, peptide stem with lateral chain.

Finally, chapter VII discusses the implications of this research. Our published paper “Two-site recognition of *Staphylococcus aureus* peptidoglycan by lysostaphin SH3\_5” is also included in the Appendices.

## CHAPTER II

### Materials and Methods

#### 2.1 Chemicals and enzymes

All chemicals and enzymes used in this study were of analytical grade and were purchased from Sigma-Aldrich, Fisher Scientific, MP Biomedical or Roche. Restriction enzymes, ligases, polymerases, DNase and appropriate buffers were purchased from New England Biolabs.

#### 2.2 Buffers and solutions

Stock buffers were autoclaved and stored at room temperature. All 1x buffers and solutions were prepared with distilled water (dH<sub>2</sub>O), except for the HPLC buffers which were made with reverse osmosis and deionized water (RO deionized water), or equivalent. All 1x buffers were filtered (0.2µm), unless otherwise stated.

##### 2.2.1 Buffers for DNA analysis

###### 2.2.1.1 TAE (50x)(Tris-acetate-EDTA)

Tris	242 g/L
Glacial acetic acid	5.7% (v/v)
Na <sub>2</sub> EDTA pH 8.0	0.05 M

1x solution adjusted with dH<sub>2</sub>O.

###### 2.2.1.2 DNA loading buffer (10x)

Bromophenol blue	0.25% (w/v)
Glycerol	50% (v/v)

## 2.2.2 Protein purification solutions

### 2.2.2.1 SDS-PAGE solutions

#### 2.2.2.1.1 SDS-PAGE separating buffer (4x)

Tris-HCl 1.5 M

SDS 0.4%

A final pH of 8.5 adjusted with HCl. Filtered and stored at room temperature.

#### 2.2.2.1.2 SDS-PAGE stacking buffer (4x)

Tris-HCl 0.5 M

SDS 0.4%

A final pH of 6.8 adjusted with HCl. Filtered and stored at room temperature.

#### 2.2.2.1.3 SDS-PAGE loading buffer (5x)

Tris-HCl pH 6.8 250 mM

SDS 10% (w/v)

Bromophenol blue 0.5 % (w/v)

Glycerol 50% (v/v)

DTT 0.5 M

#### 2.2.2.1.4 Coomassie Blue stain

Coomassie Blue R-250 0.25% (w/v)

Methanol 50% (v/v)

Acetic acid glacial 10% (v/v)

#### **2.2.2.1.5 Coomassie destaining solution**

Methanol	40% (v/v)
Glacial acetic acid	10% (v/v)

#### **2.2.2.2 Protein purification buffers for affinity chromatography (His-tag fused proteins)**

##### **2.2.2.2.1 Buffer A (Equilibration buffer)**

Tris-HCl	50 mM
NaCl	0.5M

Final pH dependant on the  $pK_a$  of the protein. A final pH of 8.0 was used for SH3\_5 proteins (alone or fused to mNeonGreen) and a pH of 7.0 for lysostaphin proteins. All buffers were filtered and stored at room temperature.

##### **2.2.2.2.2 Buffer B (Elution buffer)**

Tris-HCl	0.05M
NaCl	0.5M
Imidazole	0.5M

Final pH of 8.0 for SH3\_5 proteins (alone or fused to mNeonGreen) and a final pH of 7.0 for lysostaphin proteins. All buffers were filtered and stored at room temperature.

#### **2.2.2.3 Protein purification buffers for size-exclusion chromatography (SEC)**

##### **2.2.2.3.1 His-SH3\_5 proteins (alone or fused to mNeonGreen)**

$\text{NaH}_2\text{PO}_4$	0.05 M
---------------------------	--------

Final pH of 6.0. All buffers were filtered, degassed, and stored at room temperature.

#### **2.2.2.3.2 Lysostaphin proteins**

NaOAc 0.02 M

NaCl 0.015 M

Final pH of 5.5. All buffers were filtered, degassed, and stored at room temperature.

#### **2.2.2.3.3 His-TEV-SH3\_5 and untagged SH3\_5 proteins (SEC, TEV-cleavage and reverse-IMAC buffer)**

Tris-HCl 50 mM

NaCl 0.15 M

Adjusted to pH 8.35 with HCl.

### **2.2.3 HPLC Muropeptide analysis buffers**

#### **2.2.3.1 Phosphate buffer system**

##### **2.2.3.1.1 Buffer A (Mobile phase buffer)**

NH<sub>4</sub>H<sub>2</sub>PO<sub>4</sub> 10 mM

NaN<sub>3</sub> 5 mM

Final pH of 5.5. All buffers were filtered and store at room temperature.

##### **2.2.3.1.2 Buffer B (Elution buffer)**

NH<sub>4</sub>H<sub>2</sub>PO<sub>4</sub> 10 mM

NaN<sub>3</sub> 5 mM

Methanol 30% (v/v)

Final pH of 5.5.



Final pH of 5.5. All buffers were filtered and store at room temperature.

### **2.2.3.2 Water-Acetonitrile System**

#### **2.2.3.2.1 Solution A (Desalting solution)**

Formic acid	0.1% (v/v)
-------------	------------

All buffers were prepared with reverse osmosis (RO) deionized water to avoid filtering.

#### **2.2.3.2.2 Solution B (Elution)**

Acetonitrile	99.9% (v/v)
--------------	-------------

Formic acid	0.1% (v/v)
-------------	------------

All buffers were prepared with RO deionized water to avoid filtering.

### **2.2.3.3 Buffers for enzymatic digestion**

#### **2.2.3.3.1 Mutanolysin and EnpA<sub>c</sub> digestion buffer (20x)**

NaH <sub>2</sub> PO <sub>4</sub>	0.5 M
----------------------------------	-------

Final pH of 5.5.

#### **2.2.3.3.1 Lysostaphin digestion- Phosphate buffered saline (PBS)**

NaCl	8 g/L
------	-------

Na <sub>2</sub> HPO <sub>4</sub>	1.4 g/L
----------------------------------	---------

KCl	0.2 g/L
-----	---------

KH <sub>2</sub> PO <sub>4</sub>	0.2 g/L
---------------------------------	---------

Final pH of 7.4 adjusted with NaOH.

#### 2.2.3.4 Sodium borate buffer for reduction of muropeptides

##### Solution A:

H<sub>3</sub>BO<sub>3</sub> 0.25 M

##### Solution B:

Na<sub>2</sub>[B<sub>4</sub>O<sub>5</sub>(OH)<sub>4</sub>]·8H<sub>2</sub>O 0.0625 M

Note: Solution A and B were prepared separately.

To prepare sodium borate buffer at pH 9:

50 ml of solution A (0.25M boric acid) + 115 ml of solution B (0.0625 M borax). The final solution was diluted to a total of 200 ml.

### 2.3 Media

All media were prepared in distilled water (dH<sub>2</sub>O) and sterilised by autoclaving (121°C, 20 min) unless otherwise stated.

#### 2.3.1 Brain heart infusion (BHI)

Brain heart infusion (Oxoid) 37 g/L

#### 2.3.2 BHI agar

Brain heart infusion (Oxoid) 37 g/L

Oxoid agar No. 1 1.5 % (w/v).

#### 2.3.3 NMR minimal media M9

##### 1<sup>st</sup> solution:

Per litre: (to 950ml of RO deionized water):

Na<sub>2</sub>HPO<sub>4</sub> 6.0 g/L

$\text{KH}_2\text{PO}_4$  3.0 g/L

0.5g NaCl 0.5 g/L

pH was adjusted to 7.4 and autoclave sterilised.

**2<sup>nd</sup> solution:**

The following were added to the above solution following autoclaving:

Autoclaved trace elements (**Note 1**) 650 $\mu\text{L}$

$^{13}\text{C}$ -glucose (**Note 2**) 2g/ 8mL

Thiamine (**Note 3**) 1.0 mL

$(^{15}\text{NH}_4)_2\text{SO}_4$  or  $^{15}\text{NH}_4\text{Cl}$  (**Note 4**) 4.0ml (0.25g/mL)

Autoclaved  $\text{MgSO}_4$  1.0ml (1M)

Autoclaved  $\text{CaCl}_2$  (**Note 5**) 0.1ml (1M)

**Note 1:** Trace elements (100 mL):

The following were dissolved in 70 mL of RO deionized water:

$\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$  550 mg

$\text{MnSO}_4 \cdot \text{H}_2\text{O}$  140 mg

$\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$  40 mg

$\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$  220 mg

$\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$  45 mg

$\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$  26 mg

$\text{H}_3\text{BO}_4$  40 mg

KI 26 mg

pH was adjusted to 8.0 and subsequently added:

EDTA 500 mg

pH was re-adjusted to 8.0 and lastly added:

$\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$  375mg

The final volume was adjusted to 100 mL and autoclaved.

**Note 2:** When  $^{13}\text{C}$  was not needed 12 mL of 25% glucose was added.

**Note 3:** Stock solutions (1mg/mL) of thiamine were filter sterilised. Aliquots were stored at  $-20^\circ\text{C}$ .

**Note 4:** At higher concentrations the solubility decreased.

**Note 5:**  $\text{CaCl}_2$  was added last. As a white precipitate formed, flasks were rapidly mixed to avoid precipitation.

### 2.3.4 Overnight Express<sup>TM</sup> Instant TB Medium

Overnight Express Instant TB medium                          60 g/L

Glycerol    10 mL/L

1<sup>st</sup> The necessary amount of dH<sub>2</sub>O was autoclaved in a flask at least twice the size of the final volume and was left to cool down.

2<sup>nd</sup> The Overnight Express Instant TB medium and glycerol were added. The flask was swirl gently until the medium was dissolved. The appropriate antibiotics were added prior to inoculation.

## 2.4 Antibiotics

All stock solutions were filtered sterilised (0.2  $\mu\text{m}$ ) and stored at  $-20^\circ\text{C}$ .

**Table 2.1. Antibiotic stock solutions and working concentrations.**

<b>Antibiotics</b>	<b>Stock concentration (mg/mL)</b>	<b><i>E. coli</i> working concentration (<math>\mu\text{g}/\text{mL}</math>)</b>	<b><i>S. aureus</i> isogenic Fem mutants (<math>\mu\text{g}/\text{mL}</math>)</b>	<b>Solvent</b>
Ampicillin (Amp)	100	100	-	dH <sub>2</sub> O
Erythromycin (Ery)	30	200	<i>fem B</i> 64	95% (v/v) ethanol
Tetracycline (Tet)	10	-	<i>fem AB</i> 2	50% (v/v) ethanol
Chloramphenicol	35	35		98% ethanol

## 2.5 Bacterial strains and plasmids

### 2.5.1 Bacterial strains

Bacterial strains used in this study are described in Table 2.2. Strains were grown at 37°C in BHI or M9 minimal medium and supplemented with the suitable antibiotics where necessary to maintain the selection of resistance markers. Strains were stored as glycerol stocks (20% glycerol in BHI broth) at -80°C.

**Table 2.2. Bacterial strains used in this study.**

Strains	Relevant properties or genotype	Source or reference
<b><i>Staphylococcus aureus</i></b>		
SH1000	8325-4 derivative with a restored <i>rsbU</i> allele	(Horsburgh, 2002)
NCTC8325	Wild type strain allele	(Berger-Bächi, 1983)
<i>femB</i> mutant BB815	8325 derivative ( <i>mec</i> $\Omega$ 2006 <i>femB</i> ::Tn551), Erm <sup>R</sup>	(Henze, 1993)
<i>femAB</i> null mutant AS145	8325 derivative ( <i>mec</i> , $\Delta$ <i>femAB</i> : <i>tetK</i> ), Tet <sup>R</sup>	(Strandén, 1997)
<b><i>Escherichia coli</i></b>		
Lemo21(DE3)	BL21 derivative for protein production, Cmp <sup>R</sup>	NEB
NEB5 $\alpha$	Host strain for DNA cloning	NEB

Cmp<sup>R</sup>, resistant to ampicillin; Erm<sup>R</sup>, resistant to erythromycin; Tet<sup>R</sup>, tetracycline resistant.

### 2.5.2 Plasmids

Plasmids used in this study are listed in Table 2.3. All plasmid DNAs were purified using the GeneJET Plasmid Mini kit (Thermo Scientific).

**Table 2.3. Plasmids used in this study.**

Plasmid	Relevant properties or genotype	Source or reference
pET15b	Plasmid to produce proteins with an N-terminal His-tag <sup>a</sup>	Novagen
pET21a	Plasmid to produce proteins with a C-terminal His-tag <sup>a</sup>	Novagen
pET2818	Plasmid to produce proteins with a C-terminal His-tag <sup>a</sup>	Lab stock
pET2817-TEV	Plasmid to produce proteins with an N-terminal cleavable His-tag <sup>a</sup>	Lab stock
pET15b-His- SH3_5	pET15b derivative encoding Lss SH3_5 domain for NMR experiments <sup>a</sup>	(Jagielska, 2016)
pET21a-Lss	pET21a derivative for the expression of the full length Lss lysostaphin <sup>a</sup>	(Sabała, 2014)
pET22b(+)-BacL1-SH3	pET22b(+) derivative encoding BacL <sub>1</sub> SH3 repeats <sup>a</sup>	(Kurushima, 2013)
pET-SH3_5-TEV	pET2817-TEV derivative for the expression of the Lss-SH3_5 domain for X-ray crystallography <sup>a</sup>	Lab stock
pET-SH3_5-mNG	pET2818 derivative for the expression of Lss SH3_5-mNeonGreen fusions <sup>a</sup>	This study
pET-mNG	pET2818 derivative for the expression of mNeonGreen protein control <sup>a</sup>	This study
pTetH-1293	pAT18 derivative encoding TetR for tetracycline-inducible expression in <i>E. faecalis</i>	Lab stock
pET2818-EF1293	pET2818 derivative for the expression of EF1293 SH3_5 protein candidates <sup>a</sup>	This study

<sup>a</sup> Ampicillin resistance.

### 2.5.3 Growth conditions of *E. coli* and *S. aureus*

Strains were streaked from the glycerol stocks and grown on BHI agar plates at 37°C. The following day, a single colony from the agar plate was used as a starter culture in BHI. Cells were

grown overnight with agitation (200 rpm) at 37°C, unless otherwise started. All growth media were supplemented with antibiotics where appropriate.

## 2.6 Construction of recombinant plasmids for protein production

The plasmid expressing the full-length lysostaphin protein (pET21a-Lss) and the Lss SH3\_5 domain with a non-cleavable N-terminal His-tag (pET15b-His- SH3\_5) have been previously described (Jagielska, 2016; Sabata, 2014). The following section describes the construction of the pET-SH3\_5-mNG, pET2818-mNG and pET2818-EF1293 plasmids.

### 2.6.1 DNA manipulation

#### 2.6.1.1 Primer design

Primers were synthesised by Eurofins MWG Operon. Primers were resuspended in nuclease free H<sub>2</sub>O (nfH<sub>2</sub>O) and stored as 100 µM master stocks or 10 µM working stocks at -20°C. Primers used in this study are listed in Table 2.4.

**Table 2.4. Primers used in this study.**

Oligonucleotide	Sequence 5'→ 3'	Enzyme
<b>Lss_SH3 domain</b>		
mNG_for	CTTTTGTTTAACTTTAAGAAGGAGATATACCATGGTGAGCAAGG GCGAAGAGGA	NcoI
mNG_rev	CTAGTCAGTTAATGATGATGATGATGATGGGATCCCTTGTATAA CTCATCCATGCCCATCACG	BamHI
D450N_for	CAAACAATTCATTATAATGAAGTGATGAAACAAGAC	
D450N_rev	GTTTCATCACTTCATTATAATGAATTGTTTGACCTG	
E451M_for	ACAATTCATTATGATATGGTGATGAAACAAGACGGTCATG	
E451M_rev	GTCTTGTTTCATCACCATATCATAATGAATTGTTTGAC	
F418V_for	CAGAGTCAGCTAGCGTCACACCTAATACAGATATAATAAC	

F418V_rev	ATCTGTATTAGGTGTGACGCTAGCTGACTCTGATTTATATAGTG
I425A_for	ACCTAATACAGATATAGCAACAAGAACGACTGGTCCATTTAG
I425A_rev	CCAGTCGTTCTTGTTGCTATATCTGTATTAGGTGTGAAGCTAG
R427M_for	ACAGATATAATAACAATGACGACTGGTCCATTTAGAAGC
R427M_rev	ATGGACCAGTCGTCATTGTTATTATATCTGTATTAGGTG
L473A_for	GGCCAACGTATTTACGCGCCTGTAAGAACATGGAATAAATC
L473A_rev	CATGTTCTTACAGGCGCGTAAATACGTTGGCCACTGTTAC
M453A_for	CATTATGATGAAGTGGCAAAACAAGACGGTCATGTTTG
M453A_rev	ATGACCGTCTTGTTTTGCCACTTCATCATAATGAATTG
N405A_SH3_5_	CATGGGATGGAAAACAGCCAAATATGGCACACTATATAAATC
N405A_SH3_5_r	TAGTGTGCCATATTTGGCTGTTTTCCATCCCATGGTATATC
N421L_for	GCTAGCTTCACACCTCTTACAGATATAATAACAAGAACGAC
N421L_rev	GTTATTATATCTGTAAGAGGTGTGAAGCTAGCTGACTC
T409V_for	ACAAACAAATATGGCGTGCTATATAAATCAGAGTCAG
T409V_rev	CTCTGATTTATATAGCACGCCATATTTGTTTGTTC
T429V_for	ATAATAACAAGAACGGTTGGTCCATTTAGAAGCATG
T429V_rev	CTTCTAAATGGACCAACCGTCTTGTTATTATATCTG
V452A_for	TTCATTATGATGAAGCGATGAAACAAGACGGTCATG
V452A_rev	ACCGTCTTGTTTCATCGCTTCATCATAATGAATTGTTTG
V461A_for	ACGGTCATGTTTGGGCAGGTTATACAGGTAACAGTG
V461A_rev	TTACCTGTATAACCTGCCCAAACATGACCGTCTTGTTTC
W489L_SH3_5_f	ACTTTAGGTGTTCTTCTGGGAECTATAAAGGGATCCGGAG
W489L_SH3_5_r	TCCCTTTATAGTTCCCAGAAGAACACCTAAAGTATTAGTAG
Y411S_for	AAATATGGCACACTATCAAATCAGAGTCAGCTAGCTTC
Y411S_rev	AGCTGACTCTGATTTTGATAGTGTGCCATATTTGTTTG
Y427S_for	AGTGGCCAACGTATTTCTTGCTGTAAGAACATGGAAT
Y472S_rev	TGTTCTTACAGGCAAGGAAATACGTTGGCCACTGTTAC
<b>Lss</b>	
N405A_Lss_for	TACAGGTTGGAAAACAGCCAAATATGGCACACTATATAAATC



N405A_Lss_rev	AGTGTGCCATATTTGGCTGTTTTCCAACCTGTATTCGGCGTTG	
W489L_Lss_f	TACTTTAGGTGTTCTTCTGGGAACTATAAAGCTCGAGCAC	
W489L_Lss_r	GAGCTTTATAGTTCCCAGAAGAACACCTAAAGTATTAGTAG	
<b>BacL<sub>1</sub> SH3 domain</b>		
pET_up (T7 promoter)	TAATACGACTCACTATAGGG	
pET_dn (T7 term)	GCTAGTTATTGCTCAGCGGT	
FWD S1	GGAGATT <u>CCATGGG</u> TTCAAAGGAGATTCAGTGAATCCT	NcoI
FWD S2	TACCC <u>CCATGGG</u> ATCGGGTTCAAATACGGGAAGTGCA	NcoI
FWD S3	GGTGGAG <u>CCATGGG</u> TTACAGGCACACCCTAATTCT	NcoI
RVS S1	GGG <u>CTCGAG</u> GGTATTATTTAAAAATCCTGTTC	XhoI
RVS S2	GGG <u>CTCGAG</u> ATCTCCACCATTATCAAAAAATCCTGT	XhoI
<b>EF1293 SH3 domain</b>		
FWD 1_EF1293	GGG <u>CCATGGG</u> AAACGATGGCGACATTGCTGAAC	NcoI
FWD 2_EF1293	TTT <u>CCATGGG</u> ACATTTGTTAGGCTATGTGAAC	NcoI
FWD 3_EF1293	GGG <u>CCATGGG</u> GAAATATCAAGTAGGACAAGCAATTCGT	NcoI
RVS pTetH (sm_341)	GTG <u>GGATCCG</u> CTAGCAAATGATCCCCATGTattctgtg	BamHI

---

Underlined are sites recognised by restriction enzymes.

## 2.6.1.2 PCR amplification

### 2.6.1.2.1 Phusion polymerase

PCR amplification reactions were performed using the commercial Phusion® High-Fidelity PCR Master Mix (2x) (New England Biolabs).

Reaction mix:

Phusion High Fidelity Master Mix (2x)	25 µl
Forward primer (10 µM)	2.5 µl
Reverse primer (10 µM)	2.5 µl

Template DNA	50-100 ng
Sterile nfH <sub>2</sub> O	up to 50 $\mu$ l

PCR amplifications were carried out in a Bio-Rad T100 thermal cycler using the following program:

Initial denaturation	98°C / 30 s
Denaturation	98°C / 10 s
Annealing	55-65°C / 10 s
Extension	72°C / 15-30 s/kb, go to denaturation step, 32x
Final extension	72°C / 2-5 min

#### **2.6.1.2.2 Agarose gel electrophoresis**

DNA samples were resolved in an agarose gel at a concentration dependent on the expected fragment size, generally prepared at 1% (w/v). Before polymerization, agarose gels were stained with 20 ng ml<sup>-1</sup> SYBR safe (Invitrogen) prepared in DMSO. Analytical DNA samples were mixed with 6x DNA loading buffer (Thermo Scientific) before loading them into the gel. DNA samples that were used for purification were mixed with the 6x UView<sup>TM</sup> loading dye (BioRad). Fragments were separated for 30 min at 100 V. DNA was visualized using an UV transilluminator at 305 nm.

The size of the DNA fragments was determined using a molecular-weight size marker of 1kb (New England Biolabs) loaded adjacent to all DNA samples.

#### **2.6.1.2.3 Gel extraction**

DNA samples were resolved on agarose gels (see above). DNA fragments were visualized using a UV transilluminator (305 nm) and bands were cut out using a scalpel. DNA was extracted from the gel using the GeneJET Gel Extraction kit (Thermo Fisher), as per the manufacturer's instructions. DNA was eluted in nuclease free water (nfH<sub>2</sub>O) and quantified via A<sub>260</sub> spectrometry.

#### **2.6.1.2.4 DNA quantification**

All DNA concentrations were determined using a NanoDrop DeNovix® spectrophotometer/fluorimeter. Blank measurements were taken with 1 µl solution of DNA eluent (nfH<sub>2</sub>O). The same volume (1 µl) of DNA sample was then used to measure the concentration at 260 nm.

#### **2.6.1.3 Restriction digestion cloning**

The plasmids encoding the SH3\_5 *E. faecalis* (BacL<sub>1</sub> and EF1293) proteins were constructed by restriction endonuclease digestion and ligation. BacL<sub>1</sub> DNA fragments were cut with NcoI and XhoI and cloned into a pET2818 similarly digested. EF1293 DNA fragments were cut with NcoI and BamHI and were also cloned into a similarly digested pET2818 vector.

##### **2.6.1.3.1 Restriction Endonuclease Digestion of DNA**

Endonuclease digestion of DNA was performed using restriction digestion enzymes (New England Biolabs) as per manufacturer's instructions. Restriction enzymes and their appropriate buffers were added to solutions containing the purified DNA. Digestion reactions were incubated at 37°C for 1h. For cloning purposes, the digested fragments were purified (Sections 2.6.1.2.2 and 2.6.1.2.3).

##### **2.6.1.3.2 Ligation of DNA**

The digested DNA fragments (see above) were ligated using T4 DNA ligase (New England Biolabs) as followed:

Vector	25 ng
Insert	3-fold molar excess of vector DNA
T4 DNA ligase	0.5 µL (100 U)
T4 ligase buffer (10X)	1 µL

nfH<sub>2</sub>O

up to 10  $\mu$ L

Ligation reactions were performed with a 3:1 molar ratio of insert to vector. The following references, as shown in Table 2.5 were used to calculate the mass of the insert required.

**Table 2.5. Mass of insert used in each ligation reaction.**

	<b>Insert 500 bp</b>	<b>Insert 1000 bp</b>	<b>Insert 1500 bp</b>	<b>Insert 2000 bp</b>	<b>Insert 2500 bp</b>
<b>pET2818 ~5300bp</b>	7 ng	14 ng	21 ng	28 ng	35 ng

Ligation reactions were incubated at 16°C overnight and heat inactivated at 65°C for 10 minutes before *E. coli* transformation.

#### **2.6.1.4 Gibson Assembly of DNA Fragments**

The plasmid encoding the Lss SH3\_5–mNeonGreen fusion (pET-SH3\_5-mNG) and the mNeonGreen control protein (pET-mNG) were constructed by Gibson assembly using a DNA synthetic fragment (Integrated DNA Technology) cloned into the vector pET2818 cut with NcoI and BamHI.

PCR amplified fragments were ligated in one step by Gibson assembly (Gibson, 2009). This was performed using the NEBuilder HiFi DNA Assembly Cloning Kit (New England Biolabs). The inserted fragments had ~30 bp of overlapping sequence with the adjacent DNA fragment. After PCR amplification, inserts were gel extracted (Section 2.6.1.2.3) and added into a solution containing 25 ng of vector in a two-fold molar excess of insert. HiFi assembly master mix was added to a final concentration of 1x and the reactions were taken to a final volume of 10  $\mu$ L using nfH<sub>2</sub>O. Reactions were then incubated at 50°C for 1 h before being used in *E. coli* transformations.

### 2.6.1.5 Site-directed mutagenesis

Mutagenesis of plasmids pET-SH3\_5-mNG and pET-Lss was performed using the GeneArt® Site-Directed Mutagenesis System (Thermo Fisher Scientific). All primers containing the target mutations are described in Table 2.4. The same pair of oligonucleotides were used to introduce mutations in both plasmids, except for the N405A and W489L substitutions, which required distinct pairs of oligonucleotides to build the mNeonGreen fusions and lysostaphin mutants. In summary, the plasmid template was added to an AccuPrime™ Pfx Reaction mix which includes a DNA methyl transferase. PCR reactions were carried out as per manufacturer's instructions. After the reactions, an *in vitro* recombination reaction was performed at room temperature. This system uses *McrBC* endonuclease to cleave the template methylated plasmid. After 10 min the recombination reaction was stopped by adding 1 µL 0.5 M EDTA and the sample was incubated on ice for 15 min before transformation in *E. coli* NEB5α.

### 2.6.1.6 *E. coli* transformation

2 µl of ligation or HiFi assembly, or 1 µl of a plasmid (including the above recombination reaction), was added to 25 µl of chemically competent *E. coli* NEB5α and incubated on ice for 30 min. Cells were then heat-shocked at 42°C for 30 sec and incubated on ice for a further 2 min. 950 µl of BHI medium was added to the mixture and incubated at 37°C for 1 h before being spread on BHI agar containing the appropriate antibiotic(s).

#### 2.6.1.6.1 Colony PCR screening

For colony PCR reactions a PCR BIO Taq Red Mix (Ref PB10.13-10, Insight Biotechnology Ltd) was used. Using a sterile tip, a single transformed colony (see above) was transferred into a PCR tube. A final 25 µL reaction was performed as follow:

Taq Master Mix (2x)	12.5µl
---------------------	--------

Forward primer (10 $\mu$ M)	1.25 $\mu$ l
Reverse primer (10 $\mu$ M)	1.25 $\mu$ l
nfH <sub>2</sub> O	10 $\mu$ l

PCR amplifications were carried out using the following program:

Cell lysis	90°C	5 min
Denaturation	94°C	10 s
Annealing	50-60°C	20 s
Extension	72°C	1 min/kb, go to denaturation step, 30-35x
Final extension	72°C	5 min

#### **2.6.1.6.2 Plasmid DNA extraction**

Positive colonies were grown overnight (Section 2.5.3). The following day, plasmid DNA was extracted using the GeneJET Plasmid Miniprep kit (Thermo Scientific) as per the manufacturer's instructions. DNA was eluted in 60  $\mu$ l nuclease free water. Plasmids used in this project are outlined in Table 2.3.

#### **2.6.1.6.3 DNA Sequencing**

Plasmids were sequenced by Eurofins Genomics. Sequencing results were analysed using SnapGene® v.5.0.8.

### **2.7 Protein analysis**

#### **2.7.1 Protein expression**

For the purification of all recombinant proteins, *E. coli* Lemo21 (DE3) were transformed (Section 2.6.1.6) and plated with the appropriate antibiotic(s) at 37°C overnight. For Lemo21 (DE3), plates

were prepared with chloramphenicol at a final concentration of 35 µg/mL and ampicillin at a concentration of 100 µg/mL for all expression plasmids used in this work (Table 2.3).

During day 2, a single colony was used to inoculate ~10 mL of BHI medium with the corresponding antibiotics in a universal tube or falcon tube (1 mL starter culture per 100 mL large scale culture). Cultures were grown overnight at 37°C (200 rpm).

At day 3, large scale cultures with the necessary antibiotics were inoculated with the starter cultures, using a ratio of 1 mL of the starter culture per 100 mL of the large-scale culture. To ensure an adequate aeration, 800 mL of medium was prepared in a 3L flask, or 1.6 L in 5 L flasks. Cells were grown at 37°C shaking at 180-200 rpm to an optical density at 600 nm (OD<sub>600</sub>) of 0.7 in BHI (Section 2.3.1), or M9 medium for NMR analyses (Section 2.3.3).

To optimize the expression of *E. faecalis* EF1293 proteins described in Chapter VI, Lemo21 (DE3) cells were also grown in Overnight Express™ Instant TB medium (autoinduction medium) (Section 2.3.4). All cultures in autoinduction media were initially grown for 4h at 37°C (200 rpm) and changed to 25°C (200 rpm) for overnight. For all the above media: BHI, M9, and Overnight Express™ Instant TB, the starter cultures were grown in BHI medium.

To induce protein expression in cultures with BHI or M9 media, IPTG was added to a final concentration of 1 mM. After 4 h at 37°C or 25°C for overnight, all induced cells were harvested by centrifuging at 5000 rpm for 15 mins, and collected in 50 mL falcon tubes either by scrapping the pellet with the help of a spatula or by resuspending the cells in 10-15 mL of the affinity chromatography buffer A (Section 2.2.2.2.1). The cell pellets were either taken directly to purification at this stage or frozen at -80°C.

Protease inhibitor cocktail tablets [25x] (EDTA free) were mixed into the pellet previously resuspended in buffer A (Section 2.2.2.2.1). Crude lysates were then obtained by sonication (3 × 30 s, 20% output; Branson Sonifier 450) and used for affinity purification.

### **2.7.2 Affinity chromatography**

The sonicated crude extract was centrifuged at 45,000 g for 25 min at 4°C. Soluble proteins were loaded onto a HiTrap IMAC column (GE Healthcare) charged with 0.1M NiSO<sub>4</sub> for SH3\_5 proteins (alone or fused to mNeonGreen) or 0.1M ZnCl<sub>2</sub> for the full-length lysostaphin. His-tagged proteins were eluted with a 20-column-volume linear gradient of buffer B (Section 2.2.2.2.2).

### **2.7.3 Size-exclusion chromatography**

Recombinant His-tagged proteins were concentrated using Amicon® Ultra Centrifugal filter tubes and purified by size-exclusion chromatography at room temperature on a HiLoad™ 26/60 Superdex™ 75 prep grade (GE Healthcare) or Superdex™ 75 10/300 (GE Healthcare). His-SH3\_5 alone or fused to mNeonGreen were purified using SEC buffer (Section 2.2.2.3.1), full-length lysostaphin with SEC buffer (Section 2.2.2.3.2), and His-TEV-SH3\_5 with buffer (Section 2.2.2.3.3), this buffer was used for the SEC purification, TEV digestion, and reverse-IMAC steps. All purified proteins were analyzed by SDS-PAGE.

As mentioned, for crystallography experiments the SH3\_5 domain was produced using plasmid pET-SH3\_5-TEV described in Table 2.3. The N-terminal tag of the IMAC/SEC-purified proteins was removed using recombinant TEV protease (0.5 mg of TEV per mg of SH3\_5 protein). Digestions were performed at 37 °C overnight in buffer (Section 2.2.2.3.3). Following digestion, proteins were loaded onto a HiTrap IMAC column equilibrated in buffer (Section 2.2.2.3.3). Cleaved SH3\_5 proteins were recovered in the flow through. Elution buffer B (Section 2.2.2.3.3) was used to elute the cleaved His-tag bound to the column.

#### **2.7.3.1 Protein quantification**

All protein concentrations were determined using a NanoDrop DeNovix® spectrophotometer/fluorimeter. Blank measurements were taken with 1 µl of eluent buffer. The same volume of sample was then used to measure the concentration at 280 nm.



#### 2.7.4 SDS-PAGE

Cell lysates, soluble fractions, and each purification step were analyzed by the SDS-PAGE method (Laemmli, 1970). The acrylamide concentration was dependent on the size of the proteins to be resolved. Standard 16% gels were prepared as follows:

##### SDS-PAGE 16% (w/v) resolving gel:

30% (w/v) acrylamide/bis (37.5:1)	6.75 mL
4x separating buffer (Section 2.2.2.1.1)	3.13 mL
dH <sub>2</sub> O	3.65 mL
10% (w/v) APS 100 µl	125 µL
TEMED	25 µl

##### SDS-PAGE 16% (w/v) stacking gel:

30% (w/v) acrylamide/bis (37.5:1)	750 µl
4x separating buffer (Section 2.2.2.1.2)	1.5 mL
dH <sub>2</sub> O	3.65 mL
10% (w/v) APS 100 µl	12 µL
TEMED	30 µl

Samples were mixed with 5x SDS-PAGE loading buffer (Section 2.2.2.1.3) and incubated for 10 min at 95°C. An appropriate volume was loaded into the gel. Electrophoresis was performed at a constant voltage (190 V) for 1 h or until the dye front approached the end of the gel. When required, gels were stained with Coomassie blue stain (Section 2.2.2.1.4) and subsequently

destained for analysis (Section 2.2.2.1.5). For fluorescence gel documentation, SDS-PAGE gels were directly scanned using a ChemiDoc XRS<sup>+</sup> System.

## **2.8 Purification of *S. aureus* PG muropeptides**

### **2.8.1 *S. aureus* PG isolation**

PG sacculi were isolated from exponentially growing *S. aureus* cells at an OD<sub>600</sub> of 0.7 as previously described (Mesnage, 2008). Cells were centrifuged (6,500 rpm, 10 min) and washed with RO deionized water. The cell pellet was snap-frozen in liquid nitrogen (using 50 mL centrifuge tubes). The frozen pellets were then boiled inside a glass beaker. 8 mL of 8% v/v SDS was then added and left for 30 min. The tubes were then left to cool at room temperature. 4 to 6 washes with RO deionized water was then carried out at room temperature (45,000 *g* for 10 min) using 30 mL Beckman Polyallomer (25 x 89 mm) tubes. The pellets were resuspended in 2 mL Tris-HCl (10 mM) pH 7.4 and treated with pronase (Sigma-Aldrich) (2 mg/mL) for 3 h at 60°C. The volume was adjusted with RO deionized water to 15 mL, and 15 mL of 8 % v/v SDS was added. Samples were boiled for 30 min. Six more washes with RO deionized water was performed. To eliminate secondary polymers, the sample was treated with 1 mL of HF (48% v/v) and incubated for 48h at 4°C. After HF treatment, the pure PG was extensively washed with RO deionized water. Pure PG was freeze-dried and resuspended at a final concentration of 25 mg/mL.

#### **2.8.1.1 Production of muropeptides**

*S. aureus* PG was digested with mutanolysin (Sigma-Aldrich). To purify PG dimers (GM–P5–G5–GM–P4–G5), 180 mg of PG was digested with 4.5 mg of mutanolysin in a final volume of 7 mL using buffer (Section 2.2.3.3.1). After overnight incubation, the enzyme was heat inactivated. Around 60 % of the previous digestion was used to obtain disaccharide peptides. A sequential digestion with EnpA<sub>c</sub> (1 mg EnpA<sub>c</sub> / 50 mg PG) was performed. The pH was adjusted to 7.5 and after overnight incubation, the enzyme was heat inactivated (Reste de Roca, 2010).

### 2.8.1.2 $\beta$ -elimination (to generate lactyl-peptide)

Soluble muropeptides 200  $\mu$ L

Ammonia solution 32% 63.7  $\mu$ L

The reaction was mixed (vortex). After incubation for 5 h at 37°C, the following was added to neutralise:

Acetic acid 100% (98%) 60.3  $\mu$ L

Samples were freeze-dried and resuspended in buffer A (Section 2.2.3.1.1) .

## 2.8.2 HPLC analysis and fractionation

### 2.8.2.1 Muropeptide reduction

Soluble muropeptides were reduced before reverse-phase HPLC fractionation. MurNAc is reduced to NAc-muraminitol with sodium borohydride ( $\text{NaBH}_4$ ) to eliminate double peaks corresponding to the  $\alpha$ - and  $\beta$ -anomers. The soluble muropeptides were mixed with an equal volume of sodium borate buffer at pH 9.0 (Section 2.2.3.4). The sample was transferred to 15 mL centrifuge tubes and approximately 2 mg of  $\text{NaBH}_4$  was added using a small spatula. The sample was incubated for 20 min at room temperature. The final pH was adjusted to 3.0 with 20 % phosphoric acid. The reduced muropeptides were centrifuged at 17, 000 g for 2 min. The supernatant with the reduced muropeptides was recovered and injected into the RP-HPLC system or stored at -20 °C.

### **2.8.2.2 Muropeptide separation by RP-HPLC**

Fractionation of material corresponding to the mutanolysin digestion was carried out on a preparative Hypersil GOLD aQ column (C18; 20 × 250 mm; Thermo Scientific Fisher) and separated at a flow rate of 10 mL/min using buffer A (Section 2.2.3.1.1) as a mobile phase. After a short isocratic step (two column volumes), PG fragments were eluted with a 15-column-volume methanol linear gradient (0–30%) in buffer B (Section 2.2.3.1.2). Individual peaks were collected, freeze-dried and analysed by mass spectrometry. The fractions corresponding to the major dimer (GM–P5–G5–GM–P4–G5) were desalted by HPLC using a water–acetonitrile gradient (Section 2.2.3.2), freeze-dried and resuspended in RO deionized water. Separation of mutanolysin/EnpA<sub>c</sub> digested fragments was performed as previously described (Reste de Roca, 2010).

### **2.9 Production of PG fragments by chemical synthesis**

The production of the tetrasaccharide (GMGM) was described previously (Mesnage, 2014). All peptides and branched peptides (>95% purity) were purchased from Peptide Protein Research. Their purity was determined by HPLC and mass spectrometry. The characterization of all ligands is described in the following Chapter III. The pentaglycine peptide (G5) was purchased from Sigma Aldrich (cat. no. G5755).

### **2.10 Crystallography**

Crystallization of the lysostaphin SH3\_5 was carried out in collaboration with Dr Andrew Lovering and Hannah Walters at the University of Birmingham. Experiments were performed using standard screening in a sitting drop 96-well clover-leaf crystallography tray at 18 mg ml<sup>-1</sup> with 3.41 mM AyQK[GGGGG]A (P4-G5) in a 1:2 drop ratio of screening agent to protein solution. The trays were incubated at 18 °C. Tetragonal bipyramidal crystals formed within the first 48 h in 100 mM Bis-Tris, pH 5.5, 25% (wt/vol) poly(ethylene) glycol 3350- and 200-mM ammonium sulphate. Crystals were cryo-protected using the above conditions including P4-G5 and additional 20% (vol/vol) ethylene glycol. Two datasets were collected: a high-resolution set at the I03 beamline, Diamond Light Source, Oxford and a second set on a Rigaku Micromax home source.

All data was processed with Xiall/XDS (Kabsch, 2010) at the University of Birmingham as described in (Gonzalez-Delgado, 2020).

### 2.11 NMR experiments

NMR experiments were conducted on Bruker Avance I 800 and DRX-600 spectrometers at 298 K. Two-dimensional  $^{15}\text{N}$ HSQC experiments were carried out using the `b_hsqcetf3gpsi` pulse program (Bruker) with a relaxation delay of 1 s, 128 complex increments (approximately 1 h 18 m per spectrum). Lss SH3\_5 proteins were quantified by measuring the absorbance at 280 nm (Section 2.7.3.1) and adjusted to a concentration of 60  $\mu\text{M}$  in SEC buffer (Section 2.2.2.3.1). All ligands were quantified by one-dimensional  $^1\text{H}$ -NMR spectra with a relaxation delay of 10 s. Quantifications were carried out based on the intensity of methyl protons, using trimethylsilylpropanoic acid (TSP) as a standard.  $^{15}\text{N}$  HSQC experiments and CSP analysis were performed as previously described (Mesnage, 2014) using Bruker TOPSPIN (versions 4.0.7) and FELIX 2007 software (Felix NMR, Inc., San Diego, CA).  $K_d$  values were obtained through fitting to standard saturation curves using Microsoft Excel. Chemical shift changes were analysed as a weighted sum of  $^1\text{H}$  and  $^{15}\text{N}$  shift changes:  $\Delta\delta = \sqrt{(\Delta\delta\text{H}^2 + (0.154 \times \Delta\delta\text{N})^2)}$  (Williamson, 2013).

### 2.12 PG binding assays

The PG binding activity of the Lss SH3\_5-mNeonGreen fusion proteins was determined using in-gel fluorescence. Protein amounts equivalent to 3  $\mu\text{g}$  of the recombinant wild-type and derivative proteins were adjusted based on fluorescence intensity of the bands corresponding to the full-length proteins. Fusion proteins were incubated in the presence of increasing amounts of PG (0–400  $\mu\text{g}$ ) for 20 min at room temperature in a final volume of 40  $\mu\text{L}$ . PG and bound proteins were then pelleted at 17,000  $g$  for 5 min. 20  $\mu\text{L}$  of supernatant corresponding to unbound proteins were loaded onto an SDS–PAGE gel (12%) and scanned using a BioRad Chemidoc XRS<sup>+</sup> system. Fluorescence intensity was quantified using ImageJ software. The percentage of binding was determined using the signal intensity measured in the absence of PG (control) as a reference.

### **2.13 Lysostaphin spot on-lawn assays**

*S. aureus* SH1000 (Horsburgh, 2002) was grown to an optical density OD<sub>600</sub> of 1.0. Cells were collected, resuspended in dH<sub>2</sub>O water, autoclaved (121°C, 20 min), and incorporated onto agar plates at a final OD<sub>600</sub> of 0.5. An arbitrary lysostaphin activity unit (50 ng) was defined as the greatest dilution used for the wild-type and derivative proteins. 5 µL corresponding to serial dilutions of the recombinant lysostaphin proteins were spotted onto plates coated in an autoclaved bacterial lawn (the substrate) and incubated overnight at 37 °C. Lytic activities were detected as clearing zones and compared by determining the lowest amount of enzyme displaying a detectable digestion of the substrate.

## CHAPTER III

### **Production and characterisation of recombinant Lss SH3\_5 domains and peptidoglycan fragments**

Previous studies have suggested that *S. aureus* PG recognition by the SH3\_5 domain is critical for the substrate specificity of Lss. Therefore, the molecular mechanisms underpinning this PG recognition were characterised and the impact of this process on Lss activity explored.

#### **3.1 Aims and Objectives**

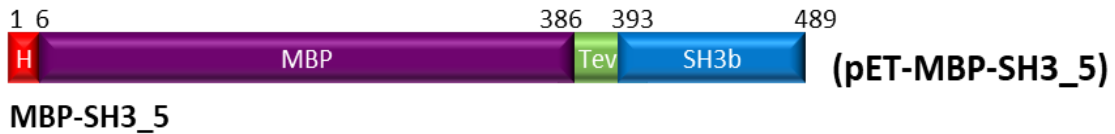
To elucidate the binding mechanisms of the SH3\_5 domain to PG, we first tried to identify the minimal PG motif recognised by the SH3\_5 domain. To carry out functional and structural analyses using NMR titrations and X-ray crystallography, we defined two main objectives:

- To produce a series of recombinant SH3\_5 proteins
- To generate a set of ligands with increasing complexity derived from *S. aureus* PG

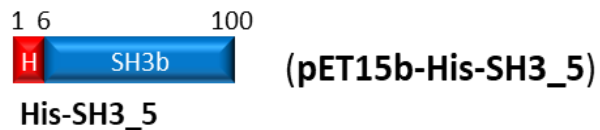
#### **3.2 Results**

##### **3.2.1 Expression and purification of the SH3\_5 domain for interaction studies and X-ray crystallography**

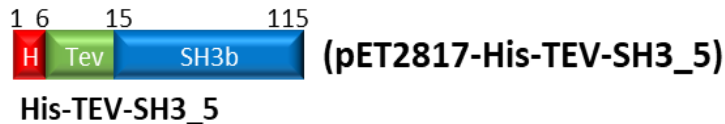
Before the start of my project, a former PhD student (Barker S, personal communication) had built an expression vector (pET-MBP-SH3\_5) to produce the Lss SH3\_5 domain fused to the maltose binding protein (MBP). The SH3\_5-MBP fusion protein has a 6x-His tag at its N-terminus and a TEV (Tobacco Etch Virus) cleavage site between the 2 domains (Fig 3.1 A).

**A**

MKIHHHHHHEEGKLVWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIIF  
 WAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTW  
 EEIPALDKELKAKGKSALMFNLQEPYFTWPLIADGGYAFKYENGYDIKDVGVNAGAKAGLTFLLVD  
 LIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSA  
 GINAASPKNELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELVKDPRIATMENAQKGEIMPNI  
 PQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNSSNNNNNNNNNGG**ENLYFQSMGWKTNK**  
 YGTYKSESASFTPNTDIITRTTGPFRSMPQSGVLKAGQTIHYDEVMKQDGHVWVGTYGNSGQRIYL  
 PVRTWNKSTNTLGVLWG**TIKGS**

**B**

MGHHHHHHEFWKTNKYGTYKSESASFTPNTDIITRTTGPFRSMPQSGVLKAGQTIHYDEVMKQDG  
 HVWVGTYGNSGQRIYLPVRTWNKSTNTLGVLWG**TIK**

**C**

HHHHHHAMG**ENLYFQSMGWKTNKYGTYKSESASFTPNTDIITRTTGPFRSMPQSGVLKAGQTIHYDEV**  
 MKQDGHVWVGTYGNSGQRIYLPVRTWNKSTNTLGVLWG**TIK**

**Figure 3.1. Description of available Lss-SH3\_5 constructs used for interaction studies and X-ray crystallography. (A)** Domain organization and amino acid sequence of the recombinant fusion protein MBP-SH3\_5 encoded by the pET-MBP-SH3\_5. **(B)** Recombinant His-SH3\_5 protein organization and amino acid sequence from pET15b-His-SH3\_5 vector. **(C)** Domain organization and amino acid sequence of the recombinant His-TEV-SH3\_5 protein encoded by pET2817-His-TEV-SH3\_5. Numbering corresponds to the transition between domains. Amino acids corresponding to the N-terminal 6x His-tag are coloured in red; Lss-SH3\_5 domain in blue (94aa); Maltose-binding protein (MBP) in purple (380aa); Tobacco Etch Virus (TEV) cleavage site in green.



To purify this MBP-SH3\_5 fusion, nickel-metal affinity (Ni-Nta) and size exclusion chromatography were performed, followed by an overnight TEV protease digestion. The products were then purified again using size exclusion chromatography to purify the cleaved SH3\_5 domain. This recombinant SH3\_5 protein was then used as a template for the NMR structural assignment.

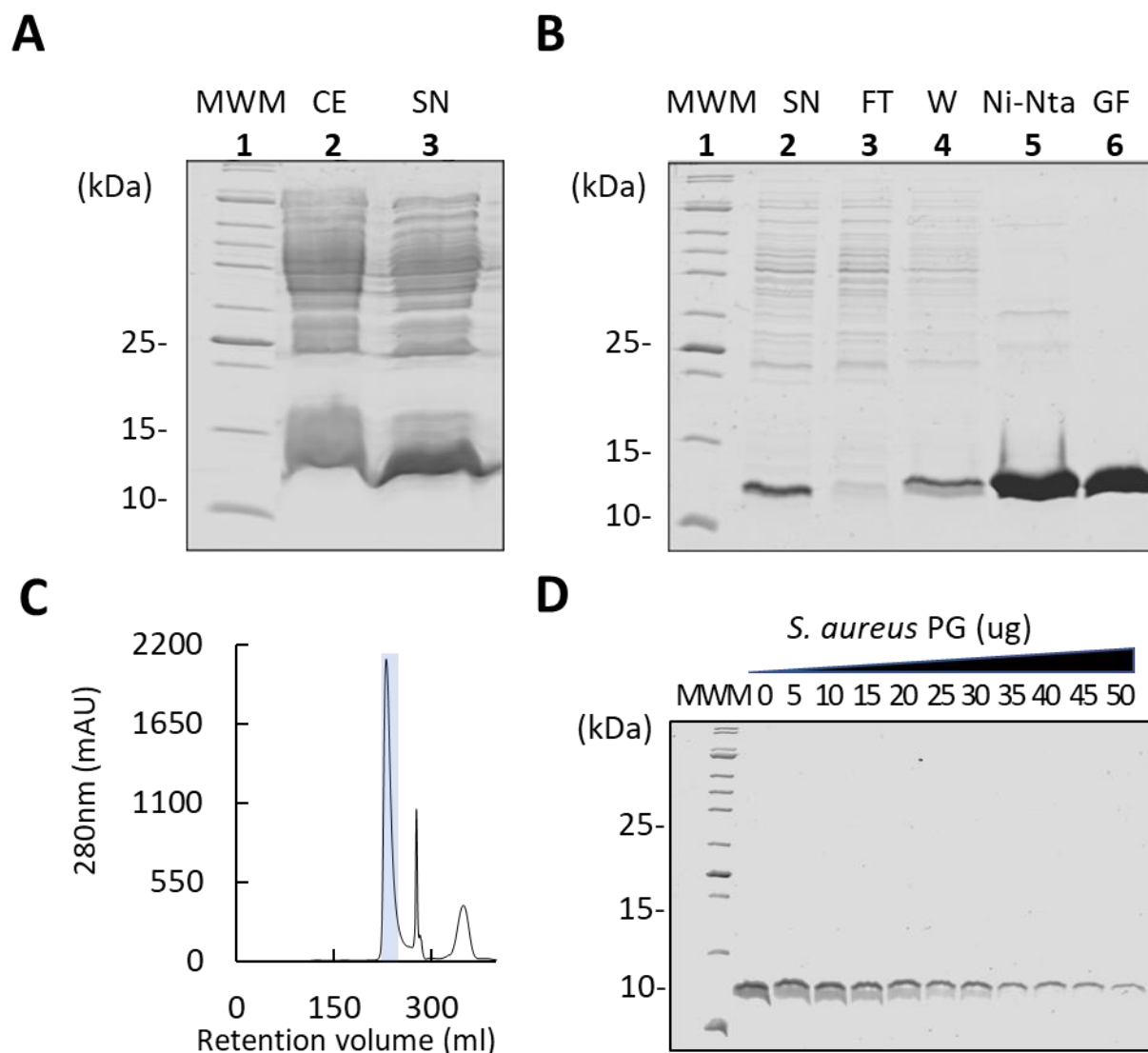
At the start of my project, one of our collaborators, Dr. Izabela Sabała, sent a new expression vector (pET15b-His-SH3\_5) allowing the expression of an N-terminally 6xHis SH3\_5 protein (Fig 3.1 B). Given the more straightforward purification strategy, it was decided to use this plasmid for all my NMR studies.

For X-ray crystallography experiments, an expression vector available in the laboratory collection (pET2817-His-TEV-SH3\_5) was used. This vector encodes a protein with an N-terminal cleavable His-tag that can be removed by TEV cleavage. (Fig 3.1 C).

### **3.2.1.1 Unlabelled His-tagged-SH3\_5**

The recombinant 11.7 kDa His-tagged SH3\_5 protein (Fig. 3.1 B) was produced as described in the material and methods, and purified by affinity chromatography, followed by gel filtration (Fig. 3.2 A-C). The purity of the purified protein was estimated to be over 95% with a yield of 15 mg/L.

To test the binding activity of the recombinant His-SH3\_5 protein, a pull-down assay was carried out using increasing amounts of *S. aureus* PG sacculi (0-50 µg) with 10 µg of His-SH3\_5 protein. After incubation, samples were centrifuged and the unbound His-SH3\_5 protein and supernatants were loaded onto an SDS-PAGE. These results showed that the recombinant SH3\_5 protein bound to PG in a dose-response manner (Fig. 3.2 D).



**Figure 3.2. Production, purification, and binding activity of recombinant Lss-SH3\_5 protein. (A)** Production and solubility of the 11.7 kDa Lss-SH3\_5 protein produced in *E. coli* Lemo21(DE3) cells. Lane 1 (MWM), molecular-weight makers; lane 2 (CE), crude extract; lane 3 (SN), soluble fraction. **(B)** SDS-PAGE analysis of Lss-SH3\_5 purification steps. Lane 1 (MWM), molecular weight makers; lane 2 (SN), soluble fraction; lane 3 (FT), flow-through; lane 4 (W), wash; lane 5 (Ni-Nta), elution pool; lane 6 (GF), pooled fractions from gel filtration step. **(C)** Gel filtration chromatogram of the purified His-SH3\_5 protein. Gel filtration chromatography was performed using a HiLoad 26/60 Superdex 75 size exclusion column as described in material and methods. Fractions eluting at the expected volume (highlighted in blue) were pooled and the purity of the recombinant His-SH3\_5

protein was confirmed by SDS-PAGE. Purification was followed measuring UV absorbance at 280nm. **(D)** SH3\_5-PG pull-down assay. The binding activity of the purified His-SH3\_5 domain was assessed by incubating increasing amounts of *S. aureus* PG sacculi [0-50µg] with 10µg of protein. After 30-min incubation at room temperature , all samples were centrifuged and the protein present at the supernatant was loaded on an SDS-PAGE. The amount of His-SH3\_5 protein present at the supernatant decreased as the PG concentration increased.

### **3.2.1.2 Singly and doubly labelled His-SH3\_5 proteins**

The His-SH3\_5 construct described in Fig. 3.1 B was singly labelled with  $^{15}\text{N}$  (Fig. 3.3 A) and doubly labelled with  $^{15}\text{N}$  and  $^{13}\text{C}$  (Fig. 3.3 B) to investigate Lss SH3\_5-PG interactions using NMR titration experiments. In both cases, proteins were purified by affinity and gel filtration chromatography.

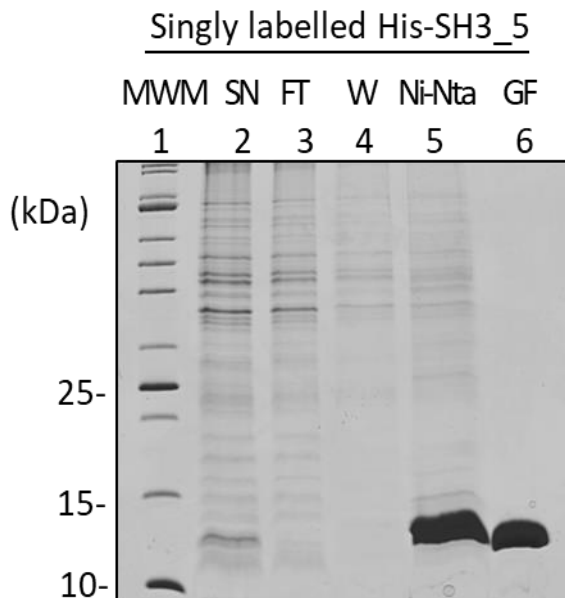
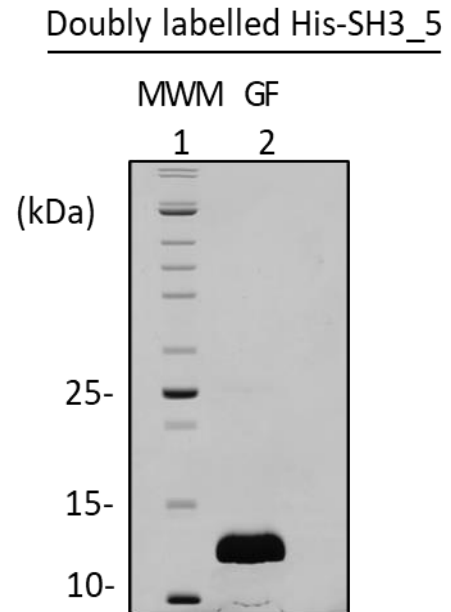
Protein purity was greater than 95% (SDS-PAGE) and the yields were 12 mg/L and 10 mg/L for the  $^{15}\text{N}$ - and  $^{15}\text{N}/^{13}\text{C}$ -labelled proteins, respectively.

### **3.2.1.3 Unlabelled and untagged SH3\_5 proteins**

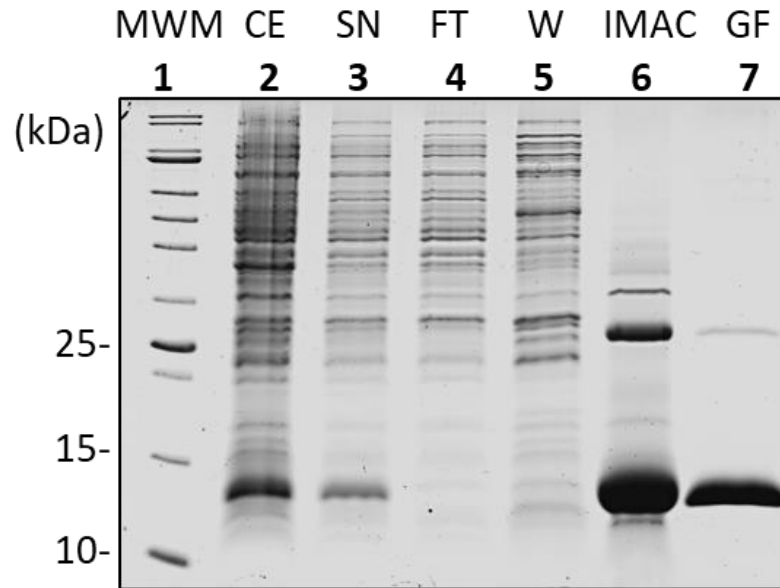
For X-ray crystallography studies, a recombinant untagged SH3\_5 protein was used to avoid potential crystallisation issues caused by the unstructured tag. The 12.3 kDa His-tagged-TEV-SH3\_5 protein described (Fig. 3.1 C) was produced and purified by affinity chromatography followed by gel filtration (Fig. 3.4). The purity of the protein estimated on SDS-PAGE was >90 % with a yield of 60 mg/mL.

A digestion trial was next performed to define the best conditions for the TEV protease cleavage of the His-tag. The standard buffer (150 mM NaCl, 50 mM Tris-HCl, pH 8.35) was tested with and without 0.2 mM DTT and 25 mM EDTA. For each buffer, increasing amounts of TEV protease (0-1  $\mu\text{g}$ ) were incubated with 10  $\mu\text{g}$  of His-TEV-SH3\_5 protein per reaction. It was found the addition of DTT and EDTA had no impact on the activity of the TEV protease. As shown in Figure 3.5, nearly 100% of the N-terminal His-tag was cleaved in the presence of 1  $\mu\text{g}$  of TEV per 10  $\mu\text{g}$  of His-TEV-SH3\_5 substrate after 1-hour digestion at 37 °C, irrespective of the buffer condition.

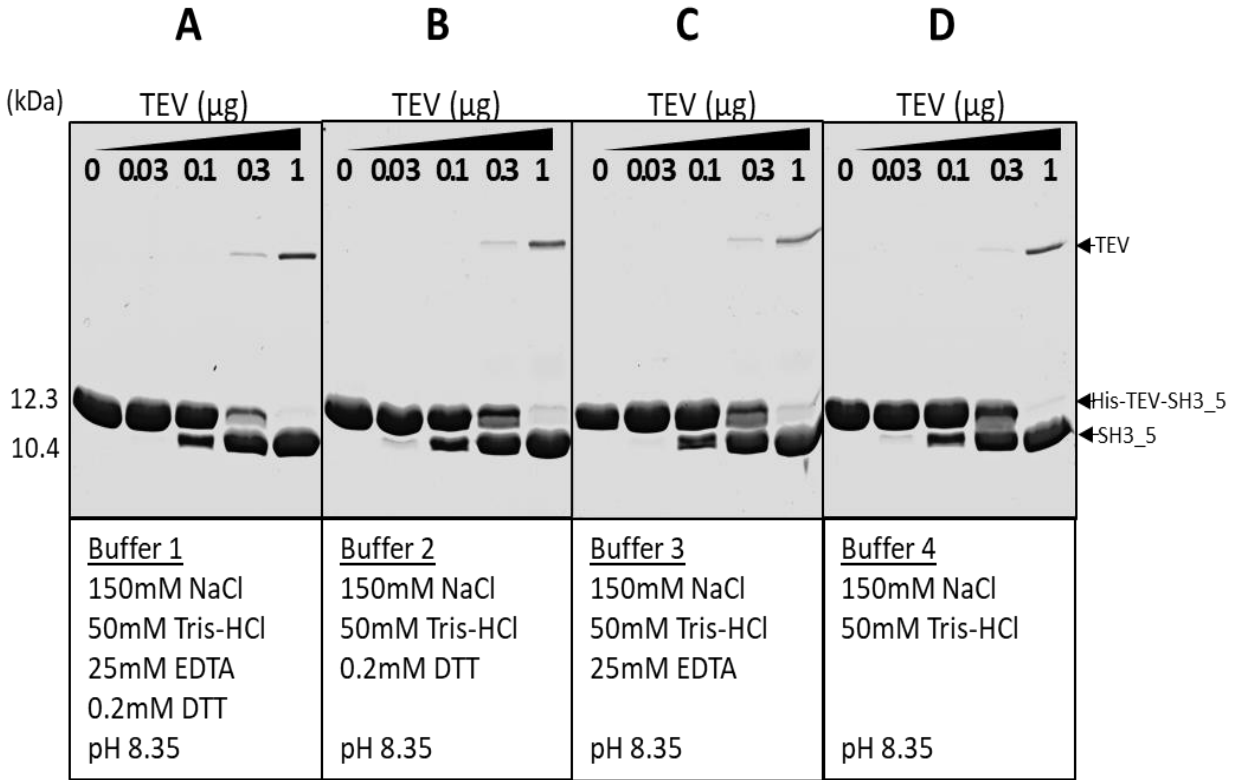
For large scale production, 1 mg of the tagged protein was incubated in the presence of 50  $\mu\text{g}$  of TEV overnight using buffer 4 (150 mM NaCl, 50 mM Tris.HCl, pH 8.35). Following cleavage, a reverse affinity chromatography was carried out to recover the 10.4 kDa untagged SH3\_5 protein in the flow through (Fig 3.6). The purity of the protein estimated on SDS-PAGE was >95 % with a yield of 20 mg/mL (Fig. 3.6).

**A****B**

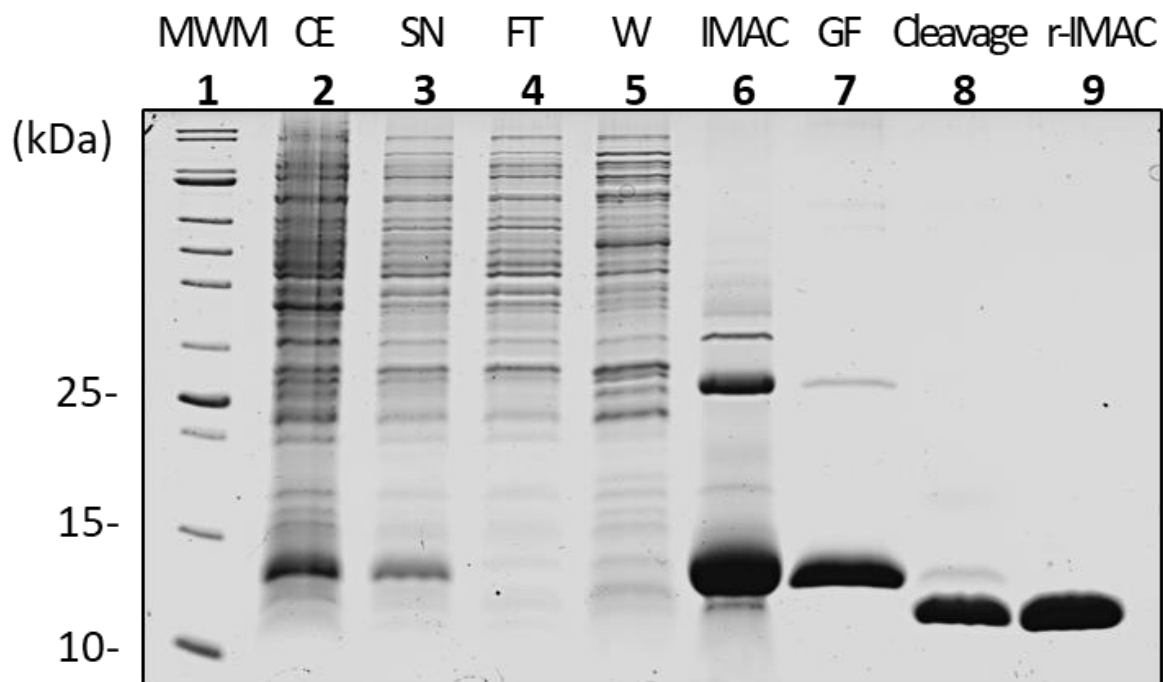
**Figure 3.3. Production and purification of  $^{15}\text{N}$ - and  $^{15}\text{N}/^{13}\text{C}$ -SH3\_5 labelled proteins. (A)** SDS-PAGE analysis of the purification steps to obtain the 11.7 kDa singly labelled His-SH3\_5 protein expressed in *E. coli* Lemo21 (DE3) cells. Lane 1 (MWM), molecular weight makers; lane 2 (SN), soluble fraction; lane 3 (FT), flow through; lane 4 (W), wash; lane 5 (Ni-Nta), elution pool; lane 6 (GF), pooled fractions resulting from the gel filtration step. **(B)** SDS-PAGE of the 11.7 kDa doubly labelled His-SH3\_5 protein expressed in *E. coli* Lemo21 (DE3) cells. Lane 1 (MWM), molecular weight makers; lane 2 (GF), pooled fractions resulting from the gel filtration step.



**Figure 3.4. Production and purification of the recombinant His-TEV-SH3\_5 protein.** SDS-PAGE analysis of the purification steps to produce the 12.3kDa His-TEV-SH3\_5 protein expressed in *E. coli* Lemo21 DE3 cells. Lane 1 (MWM), molecular-weight makers; lane 2 (CE), crude extract; lane 3 (SN), soluble fraction; lane 4 (FT), flow through; lane 5 (W), wash; lane 6 (IMAC), affinity chromatography elution pool; lane 7 (GF), pooled fractions resulting from the gel filtration step.



**Figure 3.5. Optimisation of the conditions for the TEV digestion of recombinant His-TEV-SH3\_5.** (A) Cleavage of the His-tag was analysed by incubating increasing amounts of recombinant TEV protease (0-1 µg) with 10 µg of His-TEV-SH3\_5 protein using buffer 1 (150 mM NaCl, 50 mM Tris-HCl, 25 mM EDTA, 0.2 mM DTT, pH 8.3). (B) Same experimental conditions were tested using buffer 2 (150 mM NaCl, 50 mM Tris.HCl, 0.2 mM DTT, pH 8.35). (C) Buffer 3 (150 mM NaCl, 50 mM Tris.HCl, 25 mM EDTA, pH 8.35). (D) Buffer 4 (150 mM NaCl, 50 mM Tris.HCl, pH 8.35). After 1-hour incubation at 37 °C, all samples were loaded and analysed by SDS-PAGE. A concentration of 1 µg TEV protease/10 µg His-TEV-SH3\_5 protein cleaved approximately 100 % of the His-TEV fusion tag in all buffers.



**Figure 3.6. Purification of recombinant untagged SH3\_5 protein for X-ray crystallography.** SDS-PAGE of all purification steps to obtain the final 10.4 kDa untagged SH3\_5 protein expressed in *E. coli* Lemo21 (DE3) cells. Lane 1 (MWM), molecular weight makers; lane 2 (CE), crude extract; lane 3 (SN), soluble fraction; lane 4 (FT), flow through; lane 5 (W), wash; lane 6 (IMAC), elution pool; lane 7 (GF), pooled fractions from gel filtration elution; lane 8 (Cleavage), verification of TEV cleavage after O/N incubation; lane 9 (r-IMAC), recovered SH3\_5 protein after reverse-IMAC.



### 3.2.2 Purification of peptidoglycan fragments

To investigate PG-recognition by the Lss-SH3\_5 domain, nine ligands with increasing complexity derived from *S. aureus* PG were generated. The most complex ligand was purified from PG sacculi via an *N*-acetyl muramidase (mutanolysin) digestion and the rest were synthetic peptides produced by solid-phase synthesis (Fig 3.7).

The ligands used in this study were:

- A linear tetrasaccharide (GlcNAc-MurNAc)<sub>2</sub>, **[GM-GM]** (Fujimoto Y, 2007)
- A pentaglycine cross-bridge [GGGGG], **[G5]**.
- A G5 derivative with an extra tyrosine residue at the N-terminus [YGGGGG], **[YG5]**
- A G5 derivative with an extra tyrosine residue at the C-terminus [GGGGGY], **[G5Y]**
- A tetrapeptide stem with a lactyl group [L(+)-A $\gamma$ QKA], **[P4]**.
- A lactyl (+) tetrapeptide stem bound to a pentaglycine lateral chain [A $\gamma$ QK(GGGGG)A], **[P4-G5]**
- A tetrapeptide stem with no lactyl group linked to a pentaglycine peptide as a lateral chain [A $\gamma$ QK(GGGGG)A], **[P4-G5]**.
- A pentapeptide stem crosslinked to a tetrapeptide stem through a pentaglycine bridge [A $\gamma$ QK[GGGGG]AA-A $\gamma$ QKA], **[P5-G5-P4-G5]**.
- A dimer made of a disaccharide pentapeptide crosslinked to a disaccharide tetrapeptide via a pentaglycine crosslink plus a lateral pentaglycine chain [GlcNAc-MurNAc-A $\gamma$ QK(GGGGG)AA-GlcNAc-MurNAc-A $\gamma$ QK(GGGGG)A], **[GM-P5-G5-GM-P4-G5]**.

One of the key ligands used in this study corresponded to a monomeric peptide stem [P4-G5]. Depending on the experiment carried out (NMR or X-ray), this ligand was obtained by different strategies. For NMR titrations, only a limited amount of ligand was required, so it was purified from *S. aureus* PG using a combination of two enzymes (mutanolysin + EnpA<sub>c</sub>). The disaccharide-peptide [GM-P4-G5] was  $\beta$ -eliminated to generate the lactyl-peptide [P4-G5]. This strategy was only able to produce a yield of less than 1 mg of compound (Fig. 3.8). For X-ray crystallography, a large amount of the equivalent peptide (5-10 mg) was required. Therefore, a synthetic branched peptide was purchased. Due to added complexity of the synthesis and significant cost associated with this modification, this peptide did not contain a lactyl group.

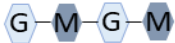







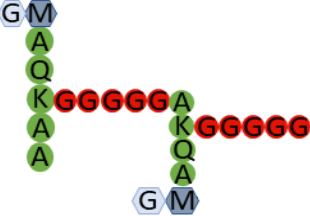
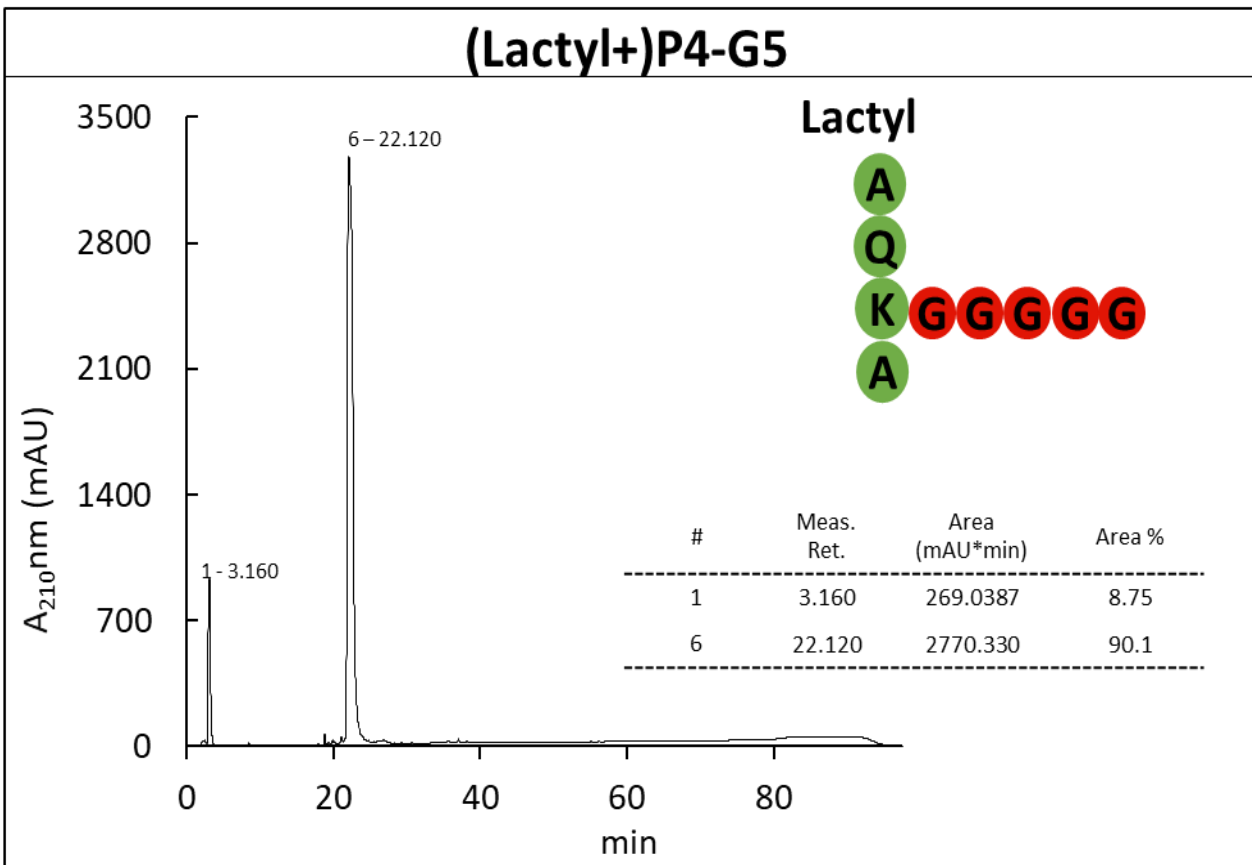
Ligand	Source	Use	
[GM-GM]		Synthetic compound	NMR
[G5]		Synthetic peptide	NMR
[YG5]		Synthetic peptide	NMR
[G5Y]		Synthetic peptide	NMR
[P4]	Lactyl 	Synthetic peptide	NMR + X-ray crystallography
[P4-G5]	Lactyl 	$\beta$ -elimination of disaccharide- peptides after mutanolysin and EnpA <sub>c</sub> digestion	NMR
[P4-G5]		Synthetic peptide	X-ray crystallography
[P5-G5-P4]		Synthetic peptide	NMR + X-ray crystallography
[GM-P5-G5- GM-P4-G5]		Mutanolysin digestion	NMR

Figure 3.7. *S. aureus* PG fragments used as SH3<sub>5</sub> ligands. The name and schematic representation of the structure of each ligand are indicated.

Column: Hypersil GoldaQ C18 column (1.9 $\mu$ m, 2.1 x 200mm).  
 Flow Rate: 0.300ml/min  
 Injection Volume: 15 $\mu$ l  
 Method Info: Analysis carried out using a 1.9 $\mu$ m, 2.1 x 200mm column, gradient from 0% - 80% Acetonitrile, in 97.5 minutes. At 60°C.



**Figure 3.8. RP-HPLC analysis of available lactyl (+) P4-G5 ligand.** RP-HPLC chromatogram of the tetrapeptide stem with a lactyl group [L(+)-A $\gamma$ QKA], [P4] ligand used in NMR titration experiments. From a sequential digestion of *S. aureus* PG by mutanolysin followed by EnpA, a DS-AiQK(GGGGG)A monomer was produced and purified, to generate the Lactyl (+) P4G5 peptide the glycan strands from the monomer were  $\beta$ -eliminated.

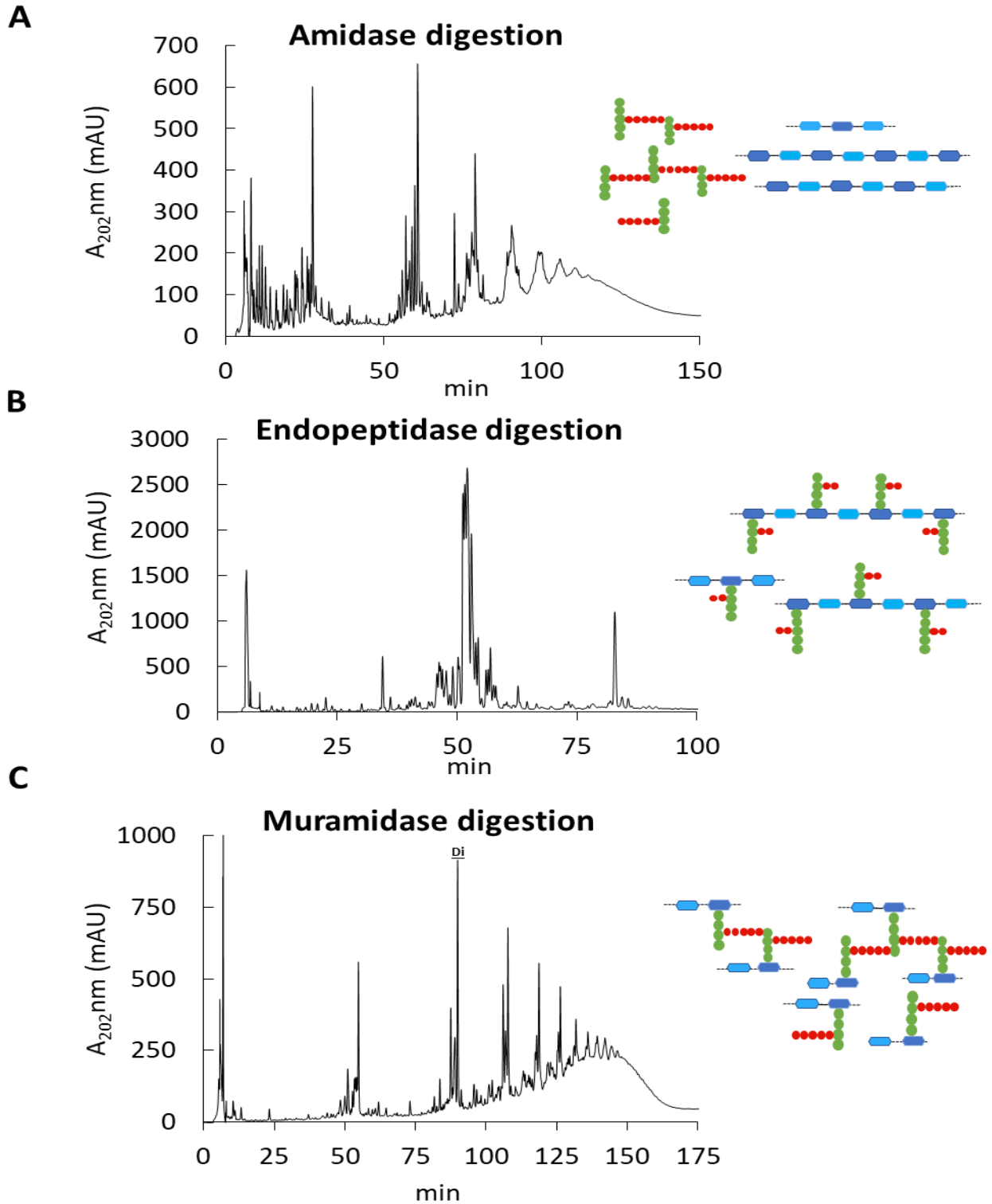
Three additional *S. aureus* PG fragments available in the laboratory were added to the list of ligands to be analysed by NMR titrations. The ligands corresponded to complex mixtures derived from the hydrolysis of PG sacculi by three distinct enzymes (Mesnage, 2014). Digestion by an amidase (*S. aureus* Atl autolysin) generated glycan chains and peptide stems (Fig. 3.9 A); digestion by Lss generated a mixture of linear PG (Fig. 3.9 B); finally, digestion by mutanolysin resulted in soluble disaccharide peptides with various degrees of crosslinking (Fig. 3.9 C).

### **3.2.2.1 Optimization of the conditions to purify *S. aureus* PG disaccharide dimers**

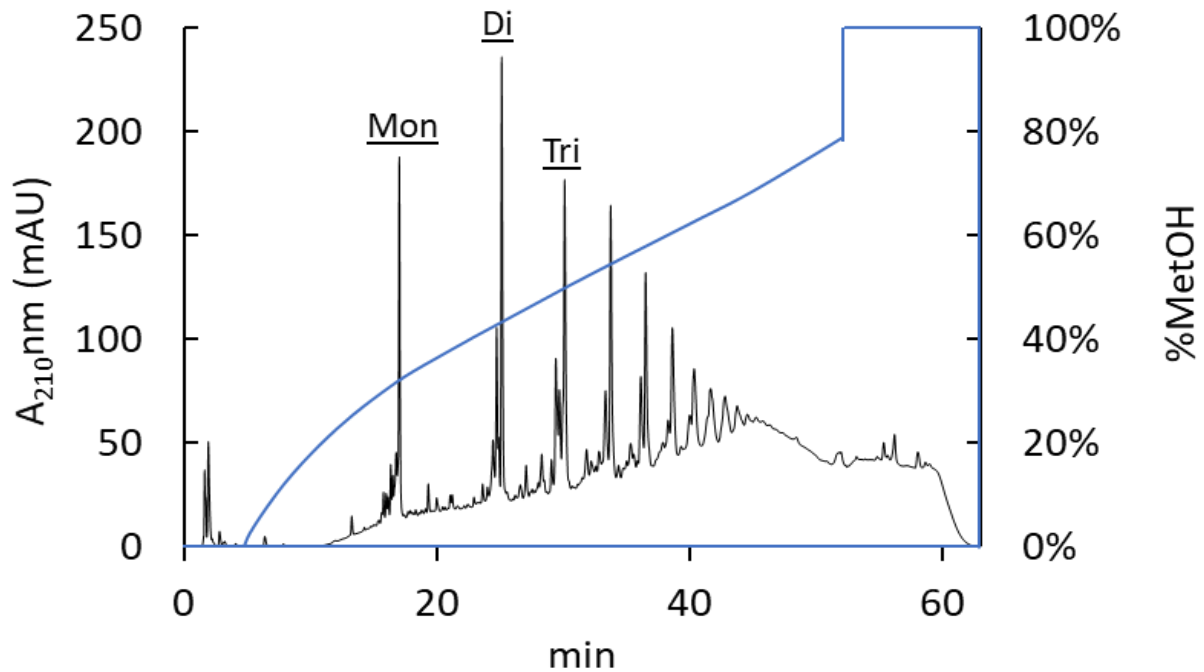
The PG dimer [GM-P5-G5-GM-P4-G5] can be isolated following digestion of sacculi by the *N*-acetyl muramidase (mutanolysin) (Fig. 3.10). The purification of a dimer at a mg scale requires the use of large amounts of mutanolysin. Given the cost of this enzyme, PG digestion conditions were first optimized to minimise the amount of mutanolysin used.

A fixed amount of PG (75 µg) was digested with serial 2-fold dilutions of mutanolysin (between 100-1.55 µg/mg) using a standard digestion volume of 100 µL/reaction. Following overnight incubations at 37 °C, all samples were reduced and analysed by RP-HPLC. As expected, higher mutanolysin concentrations led to an increasing amount of multimeric muropeptides (Fig 3.11 A).

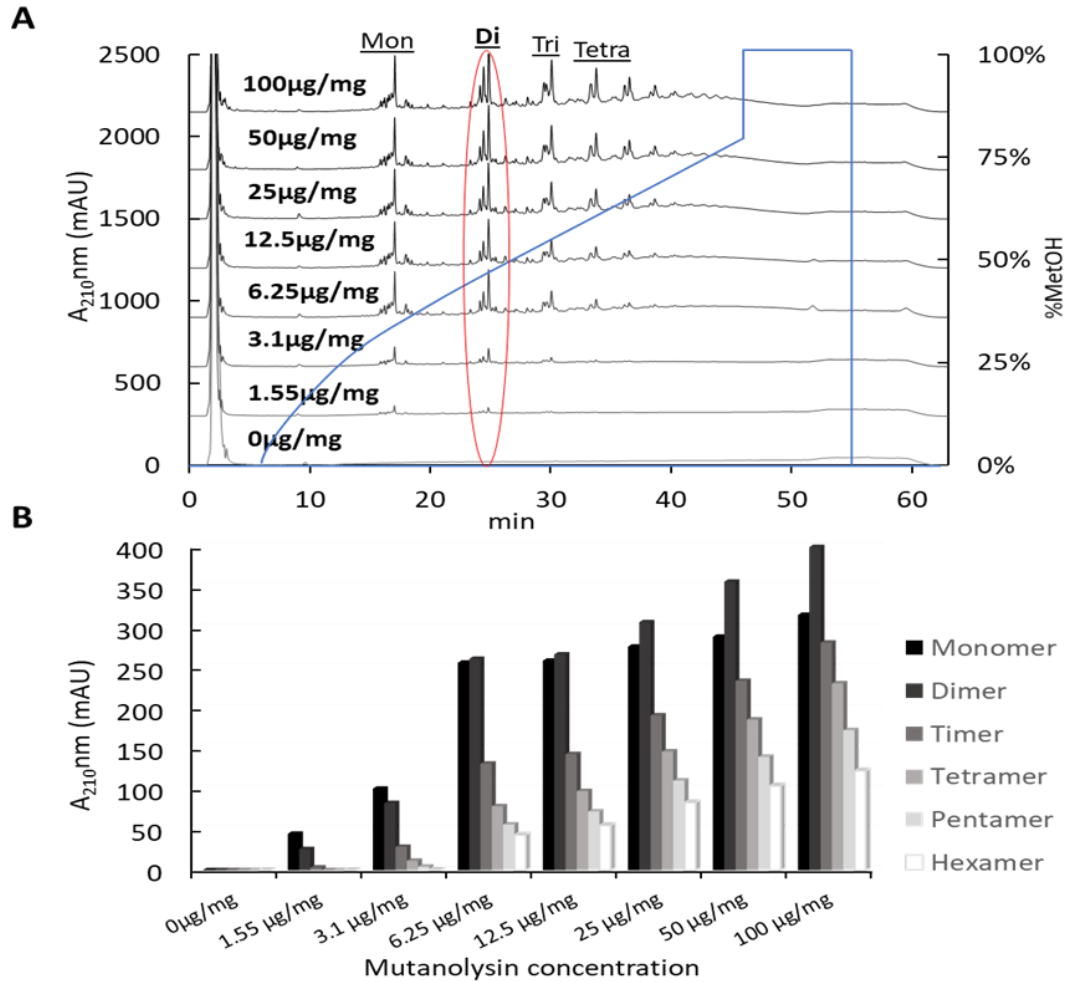
To determine the optimal concentration of mutanolysin required, each muropeptide was quantified using UV absorbance at 210 nm. The abundance of monomers and dimers almost reached a plateau after 6.25 µg/mg of mutanolysin (Fig. 3.11 B), with dimers plateauing around 25 % of total muropeptides. This was expected because low complexity muropeptides are not produced from the cleavage of highly cross-linked multimers. Incubations at concentrations of 6.25 and 12.5 µg/mg displayed a similar multimeric composition with almost equal absorbance intensities. The three highest enzyme concentrations (25, 50, and 100 µg/mL) exhibited comparable muropeptide profile with a relatively high amount of crosslinked multimers.



**Figure 3.9. RP-HPLC analysis of available *S. aureus* PG complex mixtures.** The enzyme used and schematic representation of the structures produced are indicated. Adapted from (Mesnage, 2014).



**Figure 3.10. Muropeptide profile of *S. aureus* SH1000 PG obtained by RP-HPLC.** Mutanolysin-solubilized muropeptides analysed by RP-HPLC. Peaks corresponding to monomeric (Mon), dimeric (Di), and trimeric (Tri) muropeptides are indicated. Reduced multimers were separated on a 63-min methanol gradient (0 to 30 % linear gradient in buffer A), as described in the material and methods.



**Figure 3.11. Analysis of *S. aureus* PG digestion by mutanolysin. (A)** Muropeptide profile of *S. aureus* SH1000. Small scale digestions with fixed 75 µg of PG sacculi were incubated with increasing concentrations of mutanolysin [1.55-100 µg/mg] using a standard digestion volume of 100 µL/reaction. Reduced soluble muropeptides were analysed by RP-HPLC using a standard methanol separation gradient. Peaks corresponding to the monomeric (Mon), dimeric (Di), trimeric (Tri), and tetrameric (Tetra) muropeptides are indicated on the top chromatogram **(B)**. Analysis of the DS-multimer profile generated after each mutanolysin digestion trial. The integrated areas of the peaks corresponding to monomeric and multimeric muropeptides were measured and plotted. The muropeptide profile pattern showed an increased production of highly cross-linked muropeptides as the enzyme concentration augmented. Overnight incubations with the three highest concentrations exhibited the typical muropeptide profile where the DS-dimers are the most abundant species of multimers (>25 %).

A second mutanolysin-digestion assay was performed using a low-reaction volume (10  $\mu\text{L}$  per reaction instead of 100  $\mu\text{L}$ ) to establish the predicted conditions for a large-scale digestion (180 mg PG/7 mL). Incubations with 2-fold serial dilutions of mutanolysin were performed with fixed amounts of PG sacculi (250  $\mu\text{g}$ ). After overnight digestions all samples were adjusted to a 100  $\mu\text{L}$  volume, reduced with sodium borohydride ( $\text{NaBH}_4$ ) to 200  $\mu\text{L}$ , and analysed by RP-HPLC by injecting 100  $\mu\text{L}$  (125  $\mu\text{g}$ ) of each sample (Fig. 3.12 A). In agreement with our previous observations, monomers plateaued after 6.25  $\mu\text{g}/\text{mg}$  (Fig. 3.12 B). Higher enzyme concentrations led to a progressive but rather limited increase in the production of dimers and highly crosslinked muropeptides.

Experiments presented in Fig. 3.11 and 3.12 indicated that the amount of dimer released (as quantified by UV absorbance) did not change markedly beyond 6.25  $\mu\text{g}$  of mutanolysin/mg of PG. We therefore decided to use 25  $\mu\text{g}/\text{mg}$  of PG as a standard condition. We assumed that this concentration was a reasonable compromise between using as much enzyme as possible to ensure optimal release of dimers, whilst keeping the cost of the experiment low.

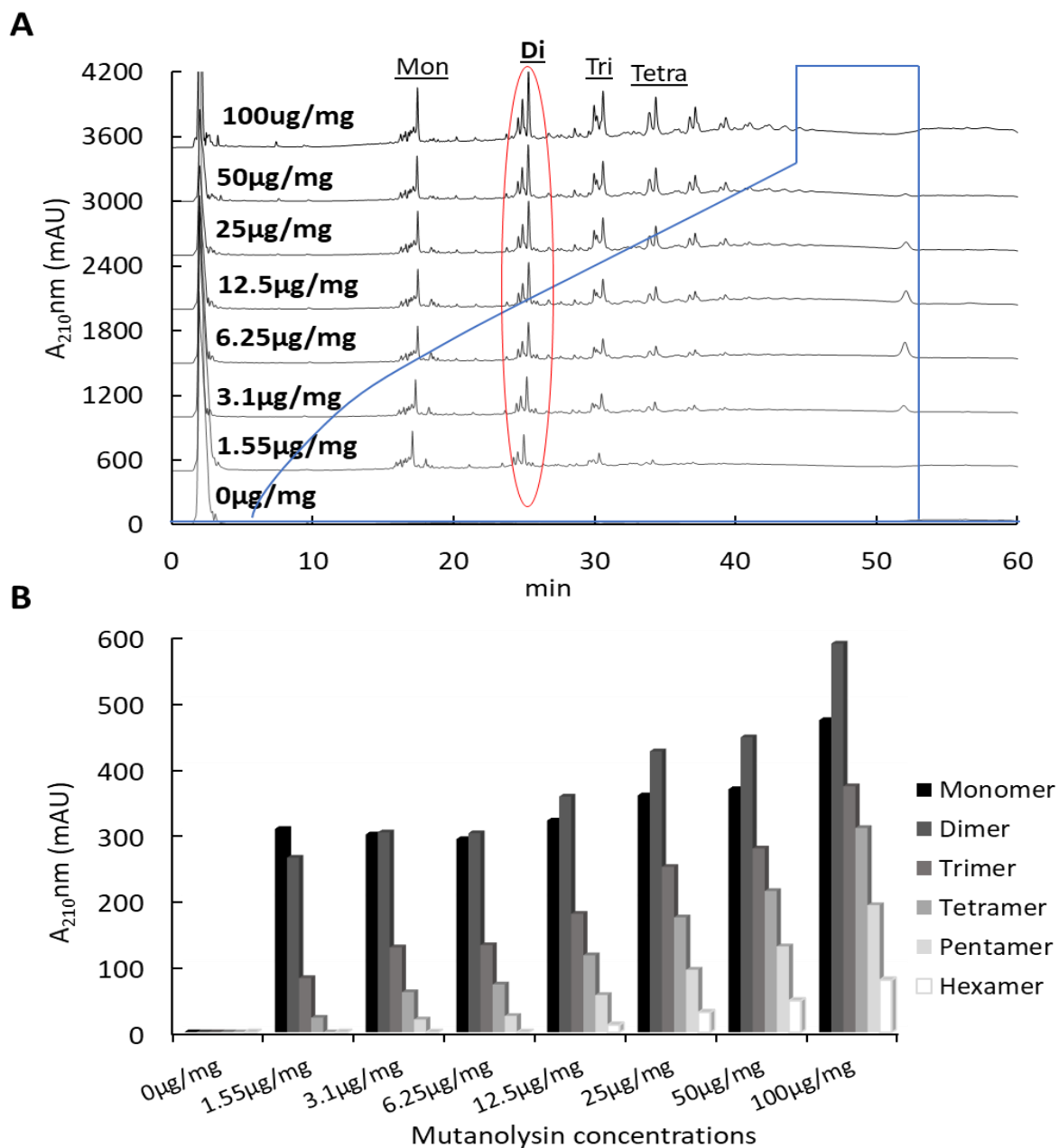
#### **3.2.2.1.1 Large-scale digestion to produce the dimer [GM-P5-G5-GM-P4-G5]**

For our large-scale digestion, 180 mg of PG sacculi were digested with 4.5 mg of mutanolysin in a 7 mL volume reaction. As a quality control step, 150  $\mu\text{g}$  of digested PG was analysed by RP-HPLC. The chromatogram showed the expected *S. aureus* muropeptide profile (Fig 3.13 A).

As an initial strategy to obtain additional DS-multimers, 40 % of the total digestion was used to purify the DS-dimer [GM-P5-G5-GM-P4-G5]. Soluble muropeptides were divided into five batches of  $\sim 15$  mg of digested PG. Samples were then reduced and separated by RP-HPLC on a preparative C18 column (5 $\mu\text{m}$ , 250 x 20 mm). Dimers were eluted at a methanol concentration of between 37 and 43 % (Fig. 3.13 B), and all desired fractions were collected, pooled, and freeze-dried. Approximately 25 mg of material were recovered and resuspended in 500  $\mu\text{L}$  of water.

To desalt and further purify the collected fraction, the gradient was adjusted to an analytical C18 column (3  $\mu\text{m}$ , 250 x 4.6 mm). Five  $\sim 5$  mg injections were performed in a 45-minute linear acetonitrile gradient in water with 0.1 % (v/v) formic acid. Despite a large amount of sample





**Figure 3.12. Analysis of small-scale mutanolysin-digestion assay to obtain the *S. aureus* PG dimer.** (A) Muropeptide profile of *S. aureus* SH1000. 250  $\mu\text{g}$  of PG sacculi were incubated with increasing concentrations of mutanolysin [1.55-100  $\mu\text{g}/\text{mg}$ ] in a 10  $\mu\text{L}$  volume reaction. Soluble muropeptides were reduced and half of each digestion was analysed by RP-HPLC. Peaks corresponding to the monomers (Mon), dimers (Di), trimers (Tri), and tetramers (Tetra) are labelled. (B) Quantification of muropeptides by UV absorbance. Areas corresponding to peaks were measured and plotted.

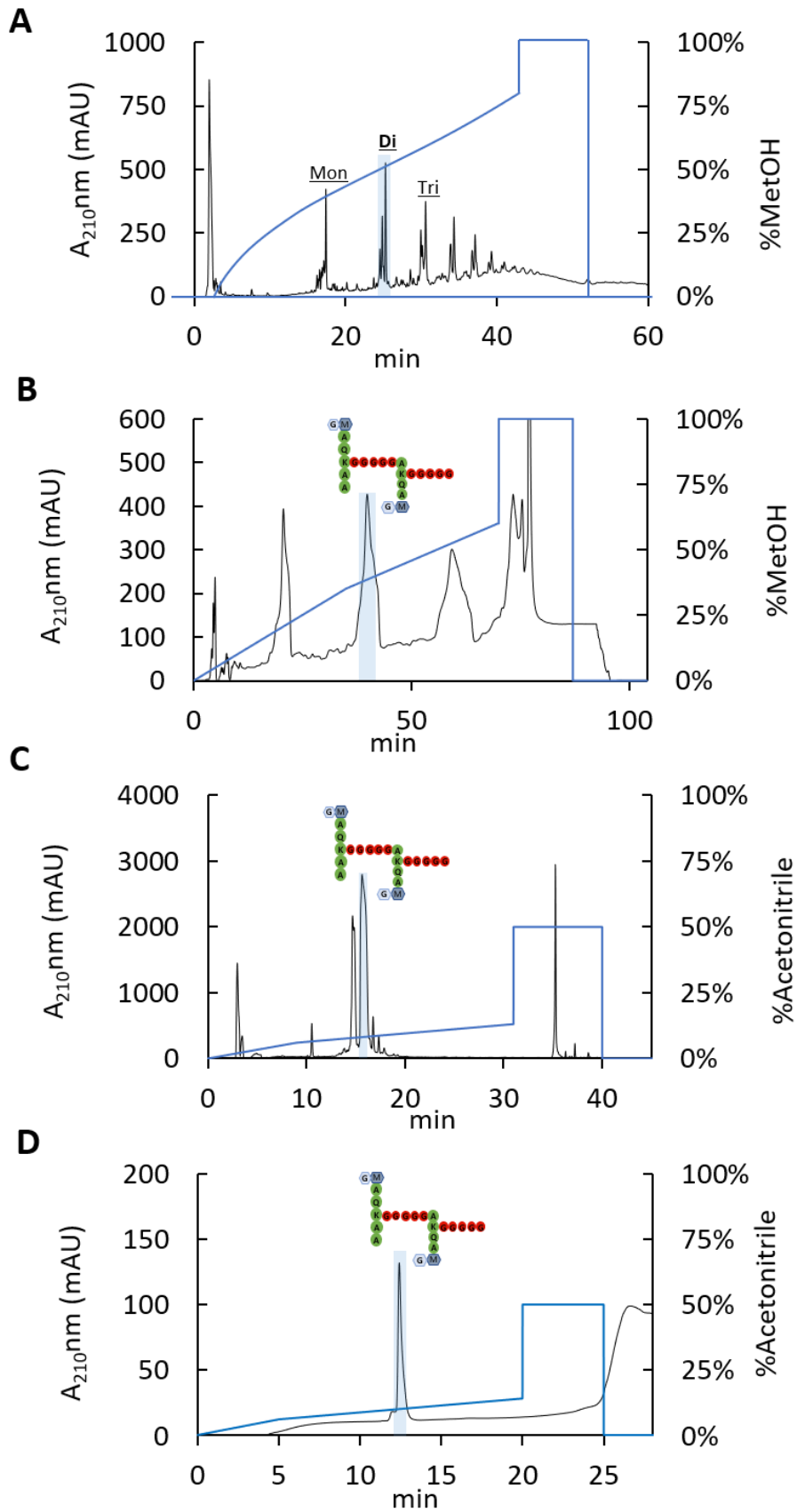
per purification, the RP-HPLC profile revealed a well-separated peak corresponding to the dimer [GM-P5-G5-GM-P4-G5] (Fig. 3.13 C).

All previous desalted fractions were pooled and freeze-dried. Approximately 0.65 mg of sample were recovered and resuspended in 250  $\mu$ L of water. The final purification product was analysed by RP-HPLC by injecting  $\sim$ 35  $\mu$ g of the sample on a 28-min acetonitrile gradient in water with 0.1 % (v/v) formic acid using an analytical Hypersil Gold aQ C18 column (1.9  $\mu$ m, 2.1 x 200 mm) (Fig. 3.13 D). The purified dimer was analysed by RP-HPLC and LC-MS. Both experiments suggested that this ligand was >93% pure (Fig. 3.14).

The production of the P4-G5 peptides from the remaining mutanolysin digestion was performed by a sequential digestion with the endopeptidase EnpA<sub>c</sub> (1 mg enzyme/ 50 mg PG). The generated P4-G5 was used for NMR analyses. EnpA<sub>c</sub> cleaves between the fourth D-Ala of the donor stem and the N-terminal residue of the sidechain, generating a monomeric disaccharide tetrapeptide with the full-length pentaglycine sidechain [GM-P4-G5]. The disaccharide-peptide was then  $\beta$ -eliminated to generate the lactyl-peptide [P4-G5]. The purification by RP-HPLC was performed using the standardized method described by (Reste de Roca, 2010). Approximately 1 mg of P4G5 was obtained. The purity of the P4-G5 fraction (>95 %) was estimated by RP-HPLC and NMR (Fig. 3.8).

### **3.2.3 Quantitative <sup>1</sup>H NMR analyses of produced *S. aureus* peptidoglycan fragments**

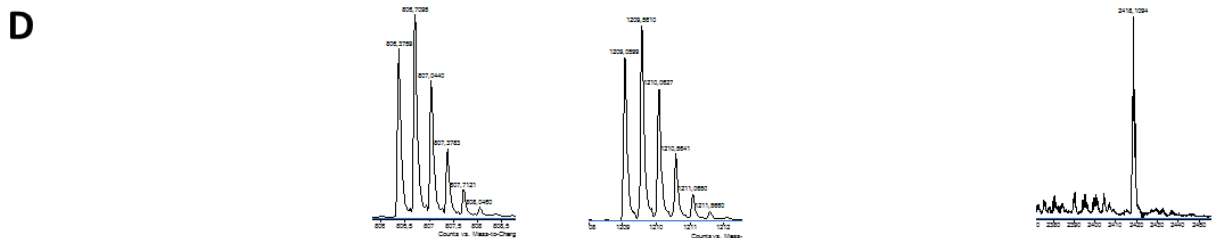
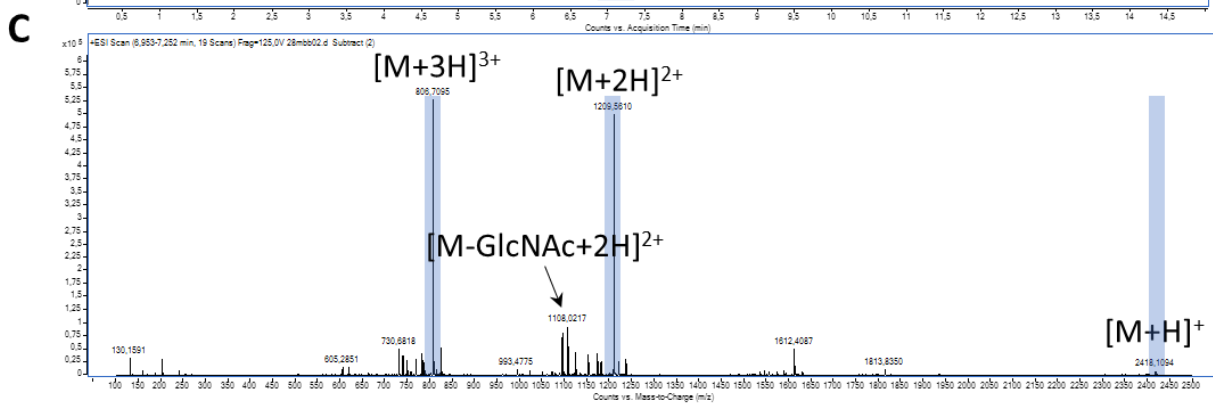
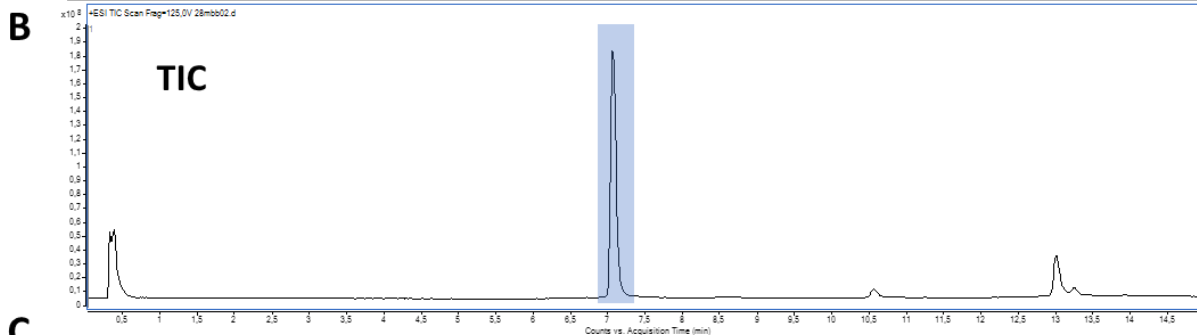
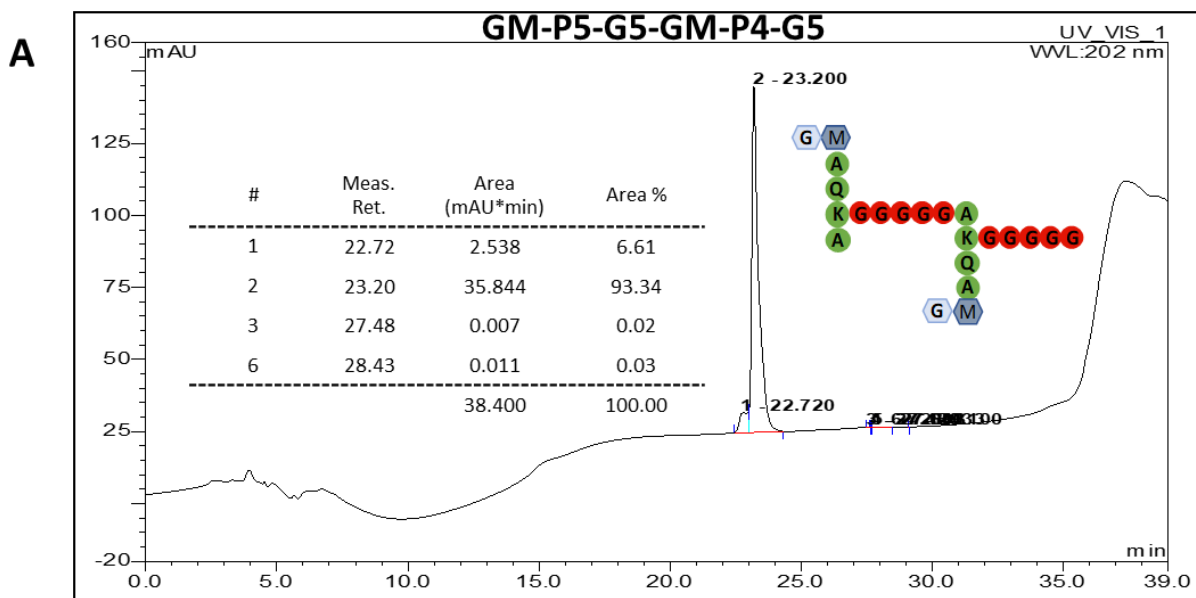
Prior to studying SH3\_5-PG interactions, a quantitative NMR analysis was performed on the complete series of produced and synthesized *S. aureus*-derived PG ligands. <sup>1</sup>H NMR is a direct and widely used method to determine metabolite concentrations based on absolute quantifications against internal reference standards. Acquiring 1D NMR spectra also allowed us to evaluate purity of the compounds. 1D NMR experiments were carried out with 10-second relaxation delays to ensure a reliable quantification by obtaining maximum magnetisation recovery.



**Figure 3.13. RP-HPLC elution profiles to obtain the [GM-P5-G5-GM-P4-G5] PG dimer. (A)** Muropeptide profile of *S. aureus* SH1000 PG obtained after a large-scale digestion. As a verification step 150 µg of mutanolysin-solubilized muropeptides were analysed by RP-HPLC. Peaks corresponding to monomeric (Mon), dimeric (Di), and trimeric (Tri) muropeptides are indicated. Reduced multimers were separated on a 63-min methanol gradient. **(B)** RP-HPLC chromatogram profile of the initial large-scale purification to obtain the dimeric fraction [GM-P5-G5-GM-P4-G5]. Approximately 75 mg of reduced mutanolysin-solubilized muropeptides were split into 5 injections of ~15 mg to separate the desired dimeric fraction (highlighted in blue). The desired fraction eluted at a concentration between 37-43% methanol gradient on a preparative C18 column (5 µm, 250 x 20 mm). The collected fractions were pooled and freeze-dried. **(C)** Desalting and re-purification of the collected dimer [GM-P5-G5-GM-P4-G5]. Five injections of approximately 5 mg of the collected material were separated by RP-HPLC, the gradient was adjusted to acetonitrile in water with 0.01 % formic acid using an analytical C18 column (3 µm, 250 x 4.6 mm). The peak corresponding to [GM-P5-G5-GM-P4-G5] is highlighted in blue. All fractions were pooled and freeze-dried. **(D)** Chromatogram profile of the final pooled fraction. 35 µg of the sample were eluted in acetonitrile gradient using an analytical Hypersil Gold aQ C18 column (1.9 µm, 2.1 x 200 mm).

Column: Hypersil GoldAQ C18 column (1.9µm, 2.1 x 200mm).  
 Flow Rate: 0.300ml/min  
 Injection Volume: 40ul  
 Method Info: Analysis carried out using a 1.9µm, 2.1 x 200mm column,  
 gradient from 0% - 80% Acetonitrile, in 39 minutes. At 60°C.

(20416.0613 Da)



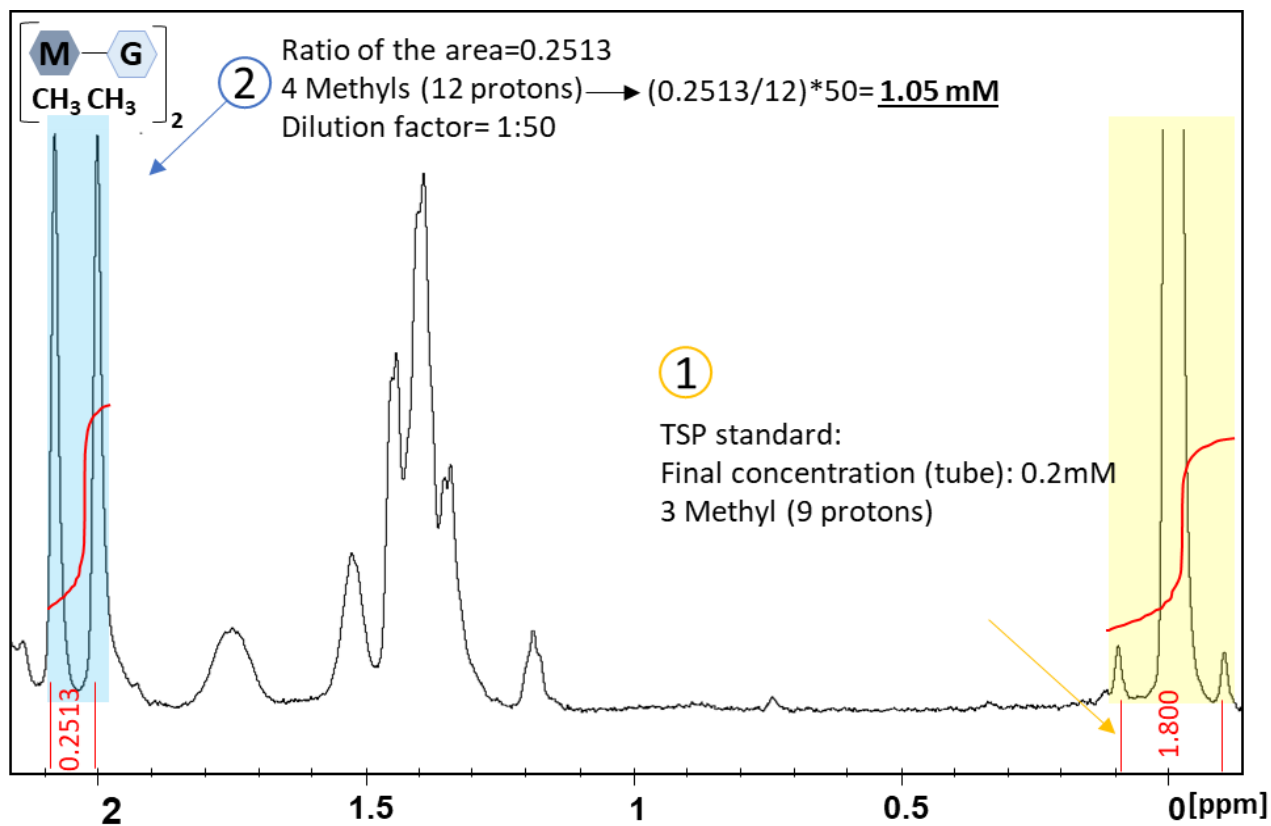
**Figure 3.14. Characterisation of the DS-dimer [GM-P5-G5-GM-P4-G5].** **(A)** Analysis of the purified [GM-P5-G5-GM-P4-G5] by RP-HPLC. A purity of >93% was estimated for the major DS-dimer (peak 2, eluted at 23.2 min) on an analytical Hypersil Gold aQ C18 column (1.9  $\mu\text{m}$ , 2.1 x 200 mm) column. **(B)** Total Ion Chromatogram (TIC) of the DS-dimer [GM-P5-G5-GM-P4-G5] ligand. Peak of interest highlighted in blue. **(C)** The mass spectrum from all scans within the highlighted area of the TIC (6.953-7.252 minutes) was extracted and the background was removed. Background was defined as ions appearing in an area of the TIC lacking peaks. **(D)** Isotopic series of peaks highlighted in panel C used to determine ion charge.

To determine the concentration of the PG ligands, the peak integrals of interest were compared to trimethylsilyl propanoic acid (TSP), a known concentration standard. Ligands comprised of the saccharide units GlcNAc and MurNAc were quantified using the methyl group from their *N*-acetyl moiety as there are no neighbouring hydrogens and their signal remains as a singlet (single peak). For those compounds lacking saccharide units, their concentration was measured by integrating methyl groups corresponding to the lactyl group and their peptide moieties (D-Ala and L-Ala). These methyl groups appeared as two peaks of equal size, known as doublets, due to the presence of one hydrogen on the adjacent atom. Doublets were individually integrated and compared to the total methyl region with the appropriate proton correction.

The quantification of the [GM-P5-G5-GM-P4-G5] dimer is shown in Figure 3.15 as an example. TSP was used at a final concentration of 0.2 mM. The area under the curve corresponding to TSP was determined. This area corresponds to a methyl resonance equivalent to 9 protons, that was then used as a reference to quantify the signal associated with the four methyl groups (12 protons) from the dimer sugars (between 1.85-2 ppm). Considering the dilution factor of the purified dimer in the NMR tube, a concentration of 1.05 mM was determined for the purified [GM-P5-G5-GM-P4-G5] ligand. Final volumes and concentrations are summarized in Table 3.1 and Figure 3.16.

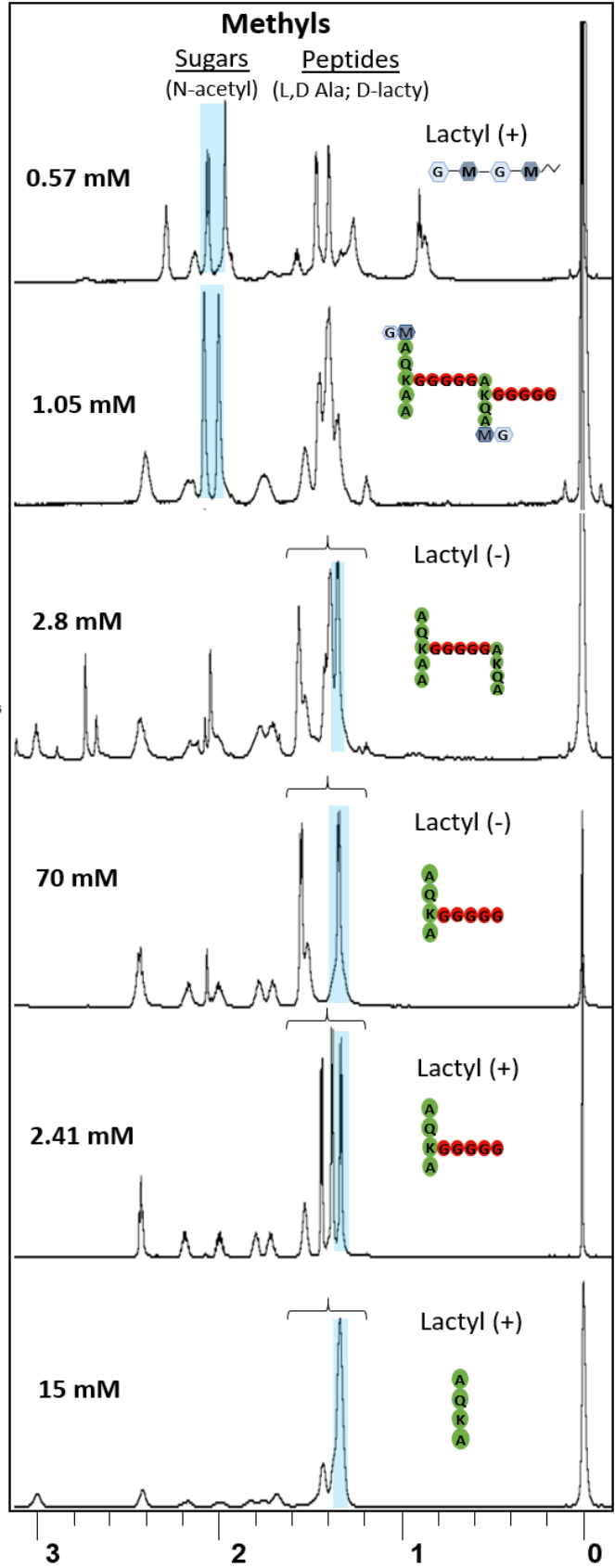
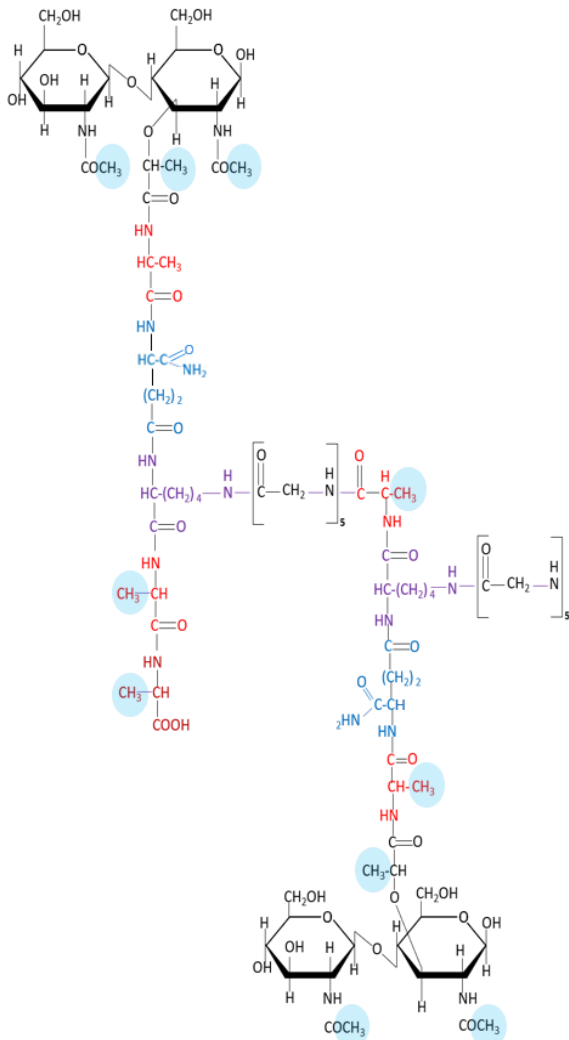
**Table 3.1. Final concentration of *S. aureus*-derived PG ligands determined by NMR.**

Ligands	Concentration (mM)	Volume (mL)
Lactyl (+) [GM-GM]	0.57	0.500
[GM-P5-G5-GM-P4-G5]	1.05	0.250
[P5-G5-P4-G5]	2.8	1.200
[P4-G5]	70	0.300
Lactyl (+) [P4-G5]	2.41	0.250
Lactyl (+)[P4]	15	0.500



**Figure 3.15.**  $^1\text{H}$  NMR analysis of the purified *S. aureus* S1000 [GM-P5-G5-GM-P4-G5] dimer. 1D NMR spectrum of the purified [GM-P5-G5-GM-P4-G5] dimer was acquired after long relaxation delays (10 seconds), using TSP as an internal calibration standard. The value of the integrated area corresponding to the methyl groups from the TSP standard (yellow) was corrected for the number of protons. The concentration of the dimer [1.05 mM], is calculated from the integrated area of interest (blue), which corresponds to methyl groups from the *N*-acetyl moieties, when corrected for the number of protons (12H) and dilution factor after TSP calibration.





**Figure 3.16.**  $^1\text{H}$  NMR spectra of pure *S. aureus* derived fragments used as SH3\_5 ligands. 1D NMR was used to quantify seven chemically defined *S. aureus* PG fragments using TSP as an internal calibration standard. The name, integrated area (blue), and schematic representation of every compound is indicated on each spectrum. The chemical structure of *S. aureus* DS-dimer [GM-P5-G5-GM-P4-G5] with methyl groups highlighted in blue is illustrated on the left.

### 3.2.4 Discussion

To study PG recognition by Lss, we produced a set of SH3\_5 recombinant proteins for NMR and X-ray crystallography analysis. NMR experiments were performed using singly and doubly labelled His-tagged SH3\_5 proteins. The N-terminal His-tag was kept since it allowed a more straightforward purification process and based on the distance from the described G5 interaction region, it was not expected to interfere with binding. A distinct expression vector was used to remove the His-tag for X-ray crystallography experiments to avoid the potential interference of the unstructured tag during the crystallisation process.

Highly pure compounds were required for both experimental approaches. The G5 ligand was purchased from SIGMA Aldrich. The rest of the synthetic peptides were purchased from Peptide Protein Research Ltd. Peptide purity, estimated >95%, was assessed by HPLC and mass spectrometry (Appendix A). The production of the GM-GM ligand is described in (Mesnage, 2014). Due to its complexity, the DS-dimer [GM-P5-G5-GM-P4-G5] was produced from pure *S. aureus* PG sacculi by *N*-acetylmuramidase (mutanolysin) digestion and purified by RP-HPLC.

In addition to RP-HPLC and MS analyses, all our chemically defined PG fragments were assessed by NMR spectroscopy. NMR is a highly reliable and commonly used technique to measure the concentration of different compounds. NMR relaxation processes are key for the correct measurement and interpretation of NMR spectra. Data acquisition consists of inducing a signal by intense and short radio frequency (rf) pulses resulting in an excitation of all resonances. Application of a short rf pulse at the appropriate frequency will rotate the bulk magnetization by 90° (into an x-y plane). A few  $\mu$ s after each pulse, an extremely sensitive receiver is turned on and the signal recorded, this signal will then decay and return to equilibrium (z-plane).  $T_1$  determines how fast can you pulse, in other words, for the recovery of the excited magnetic state to return to equilibrium, the relaxation delay should be  $t = 5xT_1$ , i.e. 5 times the longitudinal relaxation time ( $T_1$ ) of a metabolite present in the sample (Evilia, 2007). For larger proteins, this occurs in a time in the order of  $T_1 = 1$  s. All our 1D NMR experiments were performed using 10 s relaxation delays.

The purity and concentration of the produced dimer was comparable with our most complex synthetic fragments as shown by 1D NMR and LC-MS analyses. Compounds comprised of the disaccharide units GlcNAc and MurNAc displayed two signals corresponding to *N*-acetyl moieties. *N*-acetyl groups normally resonate at 1.94 ppm. The signal corresponding to the *N*-acetyl group from the muramic acid was shifted on the spectra (~2.04 ppm), which corroborates the modification of the muramic acid to an anhydrous group (after reduction with sodium borohydride), as previously described (Atrih, 1999). Peptide analysis was conducted using the signal of methyls from the L-Ala, D-Ala, or D-lactyl groups as reference. A clear display of methyl resonances from each methyl is shown for the [Lactyl (+) P4-G5], where no overlapping of the signal is shown; each peak appeared as a doublet due to the hydrogen attached to the neighbouring carbon. To verify the accuracy of our quantification, doublets were individually integrated and compared to the total methyl region with the appropriate proton correction.

Finally, our NMR quantifications highlighted slight overestimations on the concentration of our synthetic PG fragments. In peptide synthesis, counterions such as TFA are used for stability and HPLC separation; overestimation on the concentration can be explained due to the presence of TFA salts and residual water.

Altogether, the results presented in this chapter showed the production of highly pure recombinant SH3\_5 domains and PG fragments to establish the molecular basis for PG recognition by the Lss SH3\_5 domain.

## CHAPTER IV

### Molecular basis for substrate recognition by the Lss SH3\_5 domain

#### 4.1 NMR titrations as a mean to characterise protein-ligand interactions.

NMR titration experiments can be used to study biomolecular interactions. A common experimental approach is the study of chemical shift perturbations (CSPs) following the addition of increasing amounts of a binding partner. CSPs are measured in two-dimensional NMR spectra and quantified for every peak of the protein during each titration point, providing structural and kinetic information to characterise its interaction (Williamson, 2013; Zuiderweg, 2002). The detection of CSPs for individual residues shows a change in the chemical environment and can be either due to the interaction with the ligand or due to local structural rearrangements.

During titration experiments, relative chemical shifts of both  $^1\text{H}$  and  $^{15}\text{N}$  nuclei from each residue are measured. If a spectrum assignment is available, specific residues involved in the ligand recognition can be identified. A standard practice is to measure the values obtained from the  $^{15}\text{N}$  HSQC spectra, then to calculate the geometrical distance travelled by each peak, weighting  $^{15}\text{N}$  shifts by a factor of 0.15 compared to  $^1\text{H}$  shifts, and select those residues for which the weighted shift change is larger than the standard deviation of the shift for all residues (Becker, 2018; Williamson, 2013).

To determine binding affinities, the thermodynamic equilibrium that occurs when a protein (P) binds reversibly to a ligand (L) at a single binding site:  $\text{P} + \text{L} \rightleftharpoons \text{PL}$  needs to be considered. Kinetic experiments are measured at incremental time points and analysed to estimate the dissociation constant ( $K_d$ ) and the rate constants for free and bound forms,  $k_{\text{on}}([\text{P}][\text{L}]) = k_{\text{off}}([\text{PL}])$ , respectively.  $K_d$  equals  $[\text{P}][\text{L}]/[\text{PL}]$ , where [P], [L], and [PL] represent concentrations of free protein, free ligand, and complex. At equilibrium,  $k_{\text{on}}$  and  $k_{\text{off}}$  are equal which means that  $K_d$  also equals  $k_{\text{off}}/k_{\text{on}}$  (Fersht, 1999; Maity Sanhita 2019; Williamson, 2013).

The appearance of spectra during titration experiments is dependent on the exchange rate between free and bound states relative to the frequency (Hz) difference between them.

Exchange rates are classified as slow, fast, or intermediate. During slow exchange,  $k_{off}$  is slower than the difference in Hz between the chemical shifts of free and bound protein, causing the signal of both states to be detected in the spectra; during incremental additions of ligand, the free signal starts disappearing as the bound signal appears. Opposite to slow exchange rate, during fast exchange rate,  $k_{off}$  is greater (at least 10x faster) than the chemical shift difference between free and bound states. A single resonance is observed as its chemical shift is the weight average of the chemical shifts of the two individual states. During fast exchange a progressive change in the peak position is detected as the free signal moves to the bound spectrum. Between the limits of these two rates there is an intermediate exchange rate displaying a more complex behaviour on the spectra, since the exchange rate between both states is close to the shift difference (Hz) between them, peaks shift and broaden at the same time, and in some cases signals also disappear (Kleckner I.R., 2011; Waudby, 2016; Williamson, 2013).

During a titration, the total ligand and protein concentrations are known. These are the sum of free and bound forms and can be expressed as:

$$(1) [L]_t = [L] + [PL]$$

$$(2) [P]_t = [P] + [PL]$$

Chemical shifts ( $\delta$ ) are reported in ppm units, and during fast exchange the weight average of shifts in the free and bound states can be observed. In other words,  $\delta_{obs} = \delta_f f_f + \delta_b f_b$ ; where,  $f_f$  and  $f_b$  are fractions of free and bound states. Because  $f_f + f_b = 1$ , the fraction of ligand bound can be expressed as:

$$(3) f_b = (\delta_{obs} - \delta_f) / (\delta_b - \delta_f)$$

By knowing this, a  $K_d$  can be fitted using the following equation (Williamson, 2013):

$$(4) \Delta\delta_{obs} = \Delta\delta_{max} \{ ([P]_t + [L]_t + K_d) - \sqrt{([P]_t + [L]_t + K_d)^2 - 4[P]_t[L]_t} / 2[P]_t$$

The ( $\Delta\delta$ ) are the chemical shift changes, where  $\Delta\delta_{obs}$  is the change in the observed shift from free state, and  $\Delta\delta_{max}$  is the maximum shift change for 100 % saturation.  $\Delta\delta_{max}$  is commonly calculated as part of the fitting process since experimentally enough ligand would need to be added to

saturate binding i.e. a ligand concentration at least 10x the  $K_d$  which is only achievable through tightly bound ligands. This formula (4) can be set up in a spreadsheet by using test values of  $\Delta\delta_{\max}$  and  $K_d$  to obtain the expected shift change, and then find the best fit for the dataset by using the least-square fit method, providing a visual demonstration of the relationship between these data points. As a rule of thumb, several different peaks are fitted into individual saturation curves from which the results are averaged to obtain a mean value of the  $K_d$  (Becker, 2018; Williamson, 2013).

Different measurements are required to obtain affinities in slow and intermediate regimes. In the limit of slow exchange, the observed shift ( $\Delta\delta_{\text{obs}}$ ) does not move and the free signal simply disappears. If the exchange rate is genuinely slow, affinities are determined by plotting changes in the intensity of the signal as a function of the ligand concentration. Like chemical shifts, intensities are proportional to the concentration of free and bound protein and it is perfectly valid to fit a  $K_d$  using the same equation (4) (Williamson, 2013). In intermediate rates, the  $\Delta\delta_{\text{obs}}$  starts to move less than the predicted chemical shift weight average; peaks will move, broaden and might disappear. During intermediate limits, peaks should reappear as saturation is reached. If the peaks are sufficiently sharp to measure their position in an HSQC spectrum, the error of fitting equation (4) is small enough to be acceptable. However, intermediate rates are extremely complex, and signals could be affected differently introducing systematic errors in the fitted  $K_d$ . NMR line shape analysis also known as dynamic NMR, is a quantitative and well-established method that can simulate line shapes for different  $k_{\text{on}}$  and  $k_{\text{off}}$  rates and correctly accounts for the effects of intermediate exchange rates (Waudby, 2016).

## 4.2 Aims and Objectives

We sought to investigate the mechanisms underpinning SH3\_5-PG interaction. This involved two objectives:

- Mapping of the residues interacting with PG fragments
- Determine the minimal motif by obtaining binding affinities

## 4.3 Results

### 4.3.1 Complete resonance assignment of the doubly labelled SH3\_5 domain

Before the start of my project, an SH3\_5-MBP fusion described in chapter III (Fig. 3.1 A) was expressed as a doubly labelled protein. After cleavage of the MBP protein, the labelled SH3\_5 protein was used for the assignment of the SH3\_5 backbone amide groups using standard triple resonance experiments.

Triple resonance experiments are employed to obtain a sequence-specific assignment of the  $H^N$ , N,  $C\alpha$ ,  $C\beta$ , and when necessary the  $C'$  resonances of a protein (Reed, 2003; Whitehead, 1997). The strategy is to observe cross-peaks between sequential neighbouring amino acids through multidimensional NMR spectra and link them via fragments (sequential strings). The standard experiments are primarily based on the correlation of the HNCACB and HN(CO)CACB spectra (Higman, 2019; Reed, 2003). The HNCACB experiment correlates the NH group of a determined residue with its own  $C\alpha$  and  $C\beta$  chemical shifts and of the preceding residue ( $pC\alpha$  and  $pC\beta$ ). In each sequential strip, four peaks may be visible, two of them corresponding to the same residue and 2 from the preceding one. Peaks from the preceding residue are usually weaker and are also identified using the accompanied HN(CO)CACB experiment. The HN(CO)CACB only correlates with the NH group of the preceding  $pC\alpha$  and  $pC\beta$  chemical shifts, which provides verification of the linkage of one NH group to another contiguous one, forming a long chain. In some residues, the chemical shifts of the  $C\alpha$  and  $C\beta$  adopt characteristic values specific to its amino acid type e.g. alanine, threonine, and serine display very different  $C\beta$  chemical shift from other amino acids making them easier to identify. Another clear example is Glycine where there is no  $C\beta$  (Higman, 2019; Reed, 2003).

Once a long sequential string has been built, i.e. a chain of NH groups with its corresponding  $C\alpha$  and  $C\beta$  chemical shifts, it can be matched to the known sequence. To do so, a common strategy is to identify specific amino acid sequential patterns e.g. SxSASxT and to match them into the sequence. Other experiments can be performed when the quality of the spectra is low, like in the case of large proteins, where it is possible to study the chemical shift of the  $C'$  instead of the  $C\beta$



which might not be visible above noise. These HNCO and HNCA-based experiments can be more sensitive than the HNCACB and HN(CO)CACB spectra however they provide less information about the amino acid type (Reed, 2003; Yang, 1999).

Most standard experiments determine all H-N correlations, mainly from the backbone amide groups as described above, but also from the side chains where present e.g. tryptophan, asparagine, glutamine. The arginine N $\epsilon$ -H $\epsilon$  peaks are, in principle, visible, but their chemical shift is outside the region usually recorded (the spectrum width). Instead, their chemical shift has to be calculated so as to fold the peaks into the recorded spectrum (Cavanagh, 2007; Higman, 2019).

As I started my project, a doubly labelled His-tagged SH3\_5 protein was produced using a construct provided by collaborators. This allowed a more straightforward purification strategy (Fig. 3.1 B, Chapter III). The previous assignment was used as template for our new construct. To get a complete assignment of the protein, the tryptophan sidechain (N $\epsilon$ -H $\epsilon$  groups), as well as the asparagine and glutamine sidechain (N $\delta$ -H $\delta$ 2/N $\epsilon$ -H $\epsilon$ 2 groups) were also identified (Fig. 4.1).

### **4.3.2 Analysis of Lss SH3\_5-PG interactions by NMR**

#### **4.3.2.1 SH3\_5-pentaglycine [G5] binding**

The interaction between the singly labelled SH3\_5 proteins and the synthetic G5 peptide was studied by  $^{15}\text{N}$  HSQC titrations. NMR spectra were recorded after the addition of 0.5, 1, 2.5, 4.5, 7.5, 11.5, 16.5, 22.5 and 30 equivalents of peptide (red to purple) (Fig. 4.2 A). The CSP of the SH3\_5-G5 interaction displayed a fast exchange equilibrium; peaks that interacted with the peptide progressively moved maintaining the same shape throughout the titration.

As mentioned above, to give equal weight to both  $^1\text{H}$  and  $^{15}\text{N}$  CSP per residue,  $^{15}\text{N}$  chemical shifts were scaled by a factor of 0.154. Once the individual  $\Delta\delta_{\text{obs}}$  of each amino acid was obtained, their CSP values were divided by their average (all residues were considered). Normalised CSP were plotted and an arbitrary threshold of two was chosen. The largest peak shifts,  $\Delta\delta/\text{average } \Delta\delta \geq 2$ ,

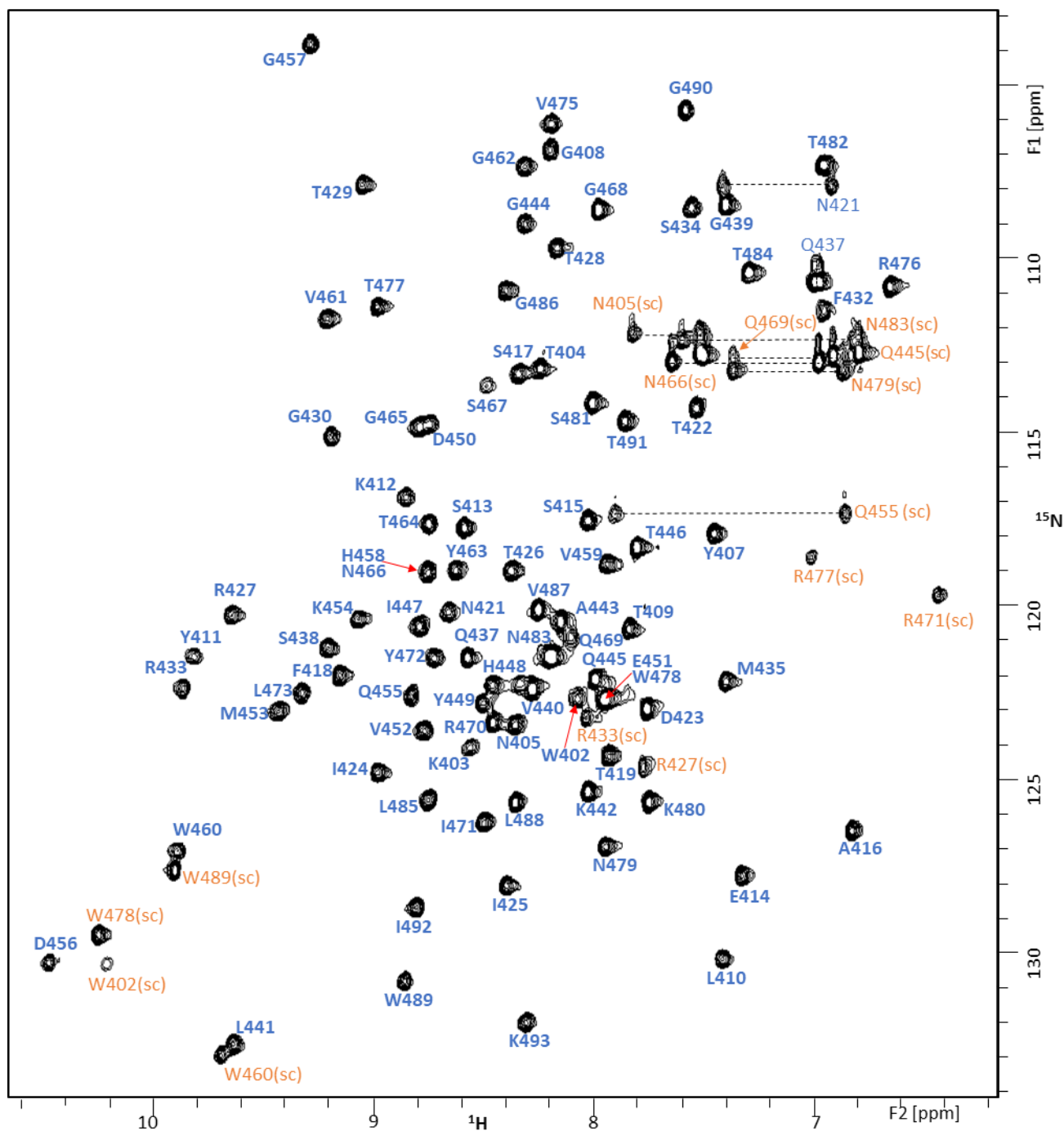


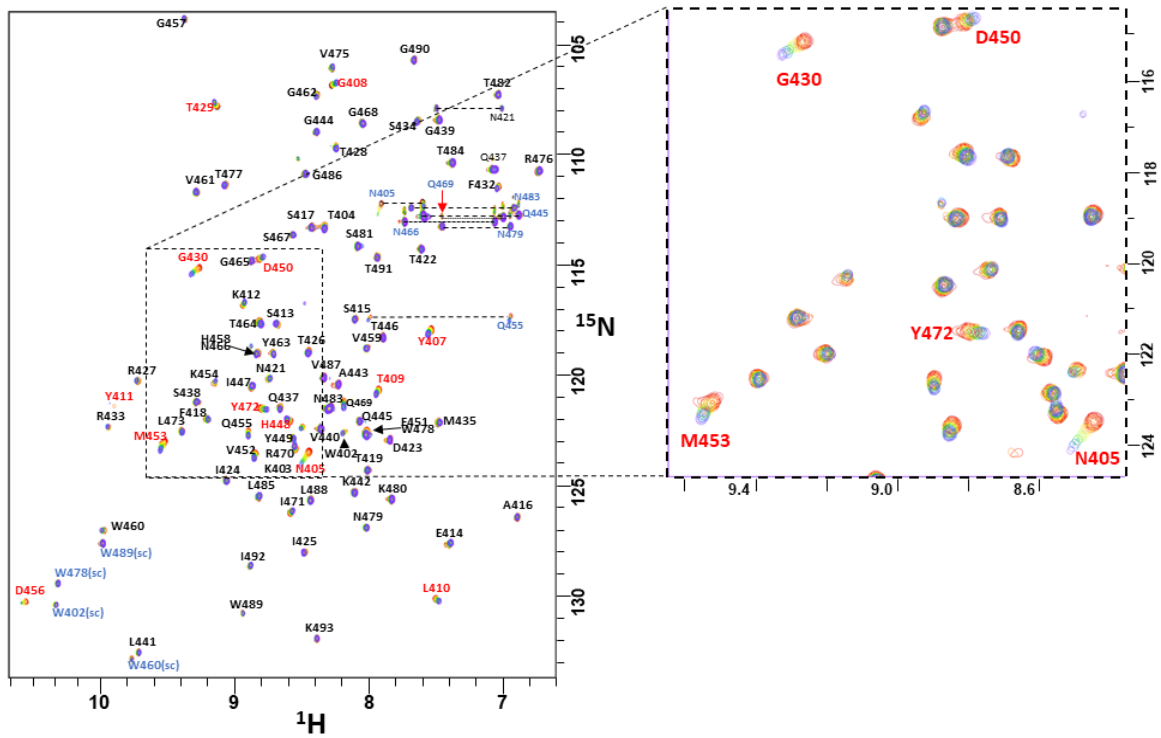
Figure 4.1. Fully assigned  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of the Lss SH3\_5 apo domain (residues 402-493). Assigned backbone amide resonances are shown in blue. Side-chain resonances of tryptophan, arginine, glutamine, and asparagine residues are in orange; the asparagine/glutamine  $\text{N}\delta\text{-H}\delta_2/\text{N}\epsilon\text{-H}\epsilon_2$  groups are indicated with dotted lines.

were observed for the residues N405 to Y411 located in strands  $\beta$ 1 and  $\beta$ 2; T429 and G430 at the RT loop between strands  $\beta$ 3 and  $\beta$ 4; H448, D450, V452, M453, D456 in strand  $\beta$ 5 and the Src loop; and Y472 in strand  $\beta$ 7 (Fig. 4.2 B). The N405 side chain also exhibited a CSP value above the established threshold (Table 4.1). Our NMR data is in accordance with the localisation of the G5 binding cleft for ALE-1, a homolog of Lss, produced by *Staphylococcus capitis* (Lu JZ, 2006) and the crystal structure of Lss-SH3\_5-pentaglycine complex (PDB 5LEO), which was used to map the residues identified by the CSP analysis in all our NMR titrations (Fig. 4.2 B).

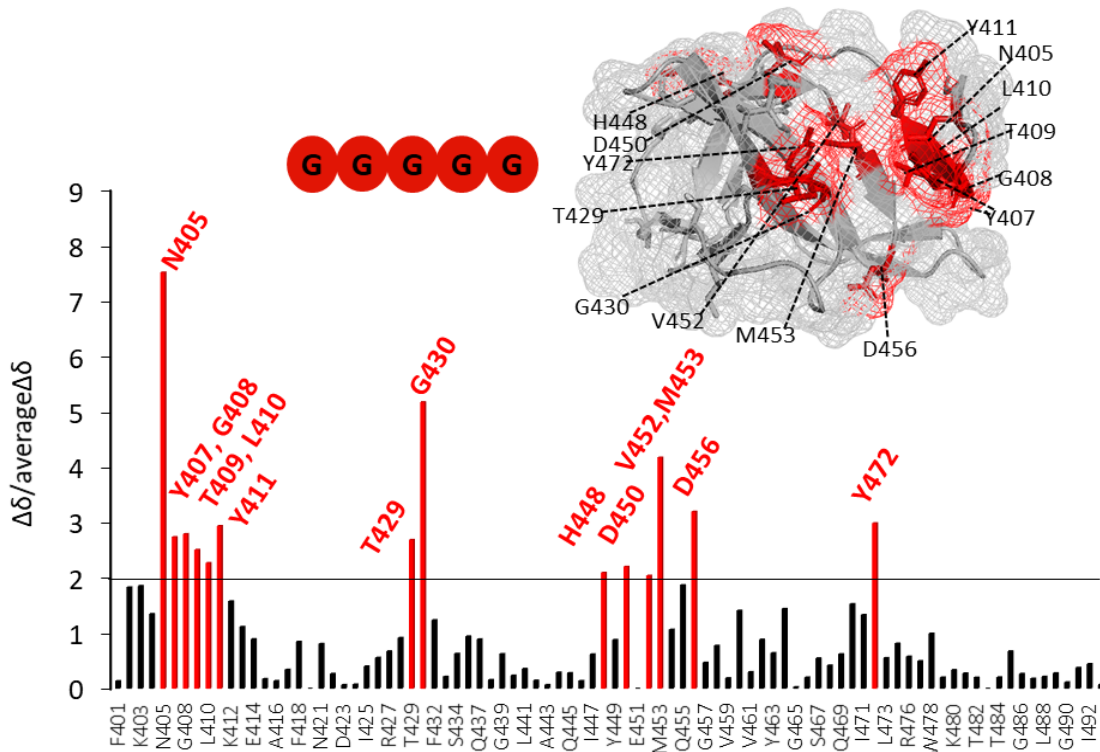
To obtain the G5 binding affinity, the value of individual  $^1\text{H}$  and  $^{15}\text{N}$  chemical shifts were measured from all residues after the incremental addition of the ligand. The  $\Delta\delta_{\text{obs}}$  were plotted and those residues that could be fitted into saturation curves were selected (Fig 4.3). In a spreadsheet, test values of  $K_d$  and  $\Delta\delta_{\text{max}}$  were assigned using the previously described equation ( $\Delta\delta_{\text{obs}} = \Delta\delta_{\text{max}} \left\{ \frac{([P]_t + [L]_t + K_d) - \sqrt{([P]_t + [L]_t + K_d)^2 - 4[P]_t[L]_t}}{2[P]_t} \right\}$ ), then individual saturation curves were fitted by the least-square method (Fig. 4.4 A). The calculated  $K_d$  values were then averaged to a mean  $K_d$  in the range of millimolar ( $889 \pm 163 \mu\text{M}$ ), which indicates a weak binding affinity (Fig. 4.4 B). This observation is in line with other G5-SH3\_5 domain interactions described by (Benešik M., 2018; Gu, 2014; Tossavainen, 2018).

To validate our method, binding affinities were estimated by fitting many saturation curves simultaneously. In theory, this is a better practice though less common. Instead of calculating  $K_d$  values separately, the same  $K_d$  but different  $\Delta\delta_{\text{max}}$  were used, and by using the same equation as before, a single fitting was calculated as the sum of squares for all resonances (22 resonances were fitted at the same time). The shift changes were re-scaled so that the maximum shifts of the N were roughly in the same range as the H, otherwise the N shift changes would have been much larger than the H shift changes, and effectively only the N differences in the fit would be considered. A weak binding affinity was estimated in agreement with the values calculated by averaging individual fits ( $K_d = 852 \pm 120 \mu\text{M}$ ) (Fig. 4.4 C). Although this is a better fitting procedure, it has the disadvantage that the error value of fitting all the affinities in a single equation is harder to calculate. The method of choice for fitting global parameters and the corresponding errors is the Monte Carlo error estimation (Arai, 2012).

**A**



**B**



**Figure 4.2. Mapping of SH3\_5 residues interacting with the G5 peptide by NMR titrations. (A)** Overlaid  $^1\text{H}^{15}\text{N}$  HSQC spectra of the SH3\_5-G5 titration acquired in the absence (dark red contour) and presence (light red to purple contour) of 0.5 to 30 equivalents of peptide. Peaks from the backbone NH assignment are labelled in black and sidechains in blue. The enlarged view illustrates a region of the spectra with peaks interacting with the G5 peptide in fast exchange. The largest peak shifts are indicated in red. **(B)** Histogram of the observed CSP values calculated as  $\Delta\delta = \sqrt{(\Delta\delta\text{H}^2 + (0.154 \times \Delta\delta\text{N})^2)}$ , as a function of the amino acid sequence. The y-axis represents the ratio between individual CSPs and the average CSP (taking all residues into account). An arbitrary threshold of 2 was chosen. Residues associated with CSPs above the threshold are in red. The SH3\_5 protein (PDB 5LEO) was used to map the residues identified to interact with G5 (N405, Y407-Y411, T429, G430, H448, D450, V452, M453, D456, and Y472).

**Table 4.1. CSPs associated with amino acid side chains.**

<b>G5</b>			<b>P4</b>		
Side chain	Normalized $\Delta\delta^a$ (ppm)	$\Delta\delta/\text{average}\Delta\delta^b$ (0.016) <sup>c</sup>	Side chain	Normalized $\Delta\delta^a$ (ppm)	$\Delta\delta/\text{average}\Delta\delta^b$ (0.079) <sup>c</sup>
Ws402	0.006	0.383	Ws402	0.039	0.489
N405s	0.122	7.643	N405s	0.028	0.350
Ws460	0.005	0.288	Rs427	0.168	2.122
Ws478	0.004	0.258	Rs433	0.048	0.610
Ws489	0.001	0.068	Ws460	0.143	1.796
			Rs470	0.038	0.484
			Rs476	0.015	0.194
			Ws478	0.014	0.175
			Ws489	0.157	1.982
<b>P4-G5</b>			<b>P5-G5-P4-G5</b>		
Side chain	Normalized $\Delta\delta^a$ (ppm)	$\Delta\delta/\text{average}\Delta\delta^b$ (0.069) <sup>c</sup>	Side chain	Normalized $\Delta\delta^a$ (ppm)	$\Delta\delta/\text{average}\Delta\delta^b$ (0.119) <sup>c</sup>
Ws402	0.010	0.143	Ws402	0.059	0.499
N405s	0.180	2.497	N405s	0.078	0.653
Rs427	*	*	Rs427	0.244	2.059
Rs433	0.024	0.348	Rs433	0.132	1.108
Ws460	0.129	1.879	Ws460	0.189	1.597
Rs470	0.231	3.367	Rs470	0.083	0.703
Rs476	0.012	0.175	Rs476	0.032	0.272
Ws478	0.004	0.058	Ws478	0.172	1.452
Ws489	0.097	1.417	Ws489	0.095	0.802
<b>GM-P5-G5-GM-P4-G5</b>					
Side chain	Normalized $\Delta\delta^a$ (ppm)	$\Delta\delta/\text{average}\Delta\delta^b$ (0.057) <sup>c</sup>			
Ws402	0.015	0.256			
N405s	0.191	3.343			
Rs427	0.071	1.247			
Rs433	0.013	0.235			
Ws460	0.066	1.165			
Rs470	0.019	0.334			
Rs476	0.029	0.513			
Ws478	0.053	0.925			
Ws489	0.039	0.689			

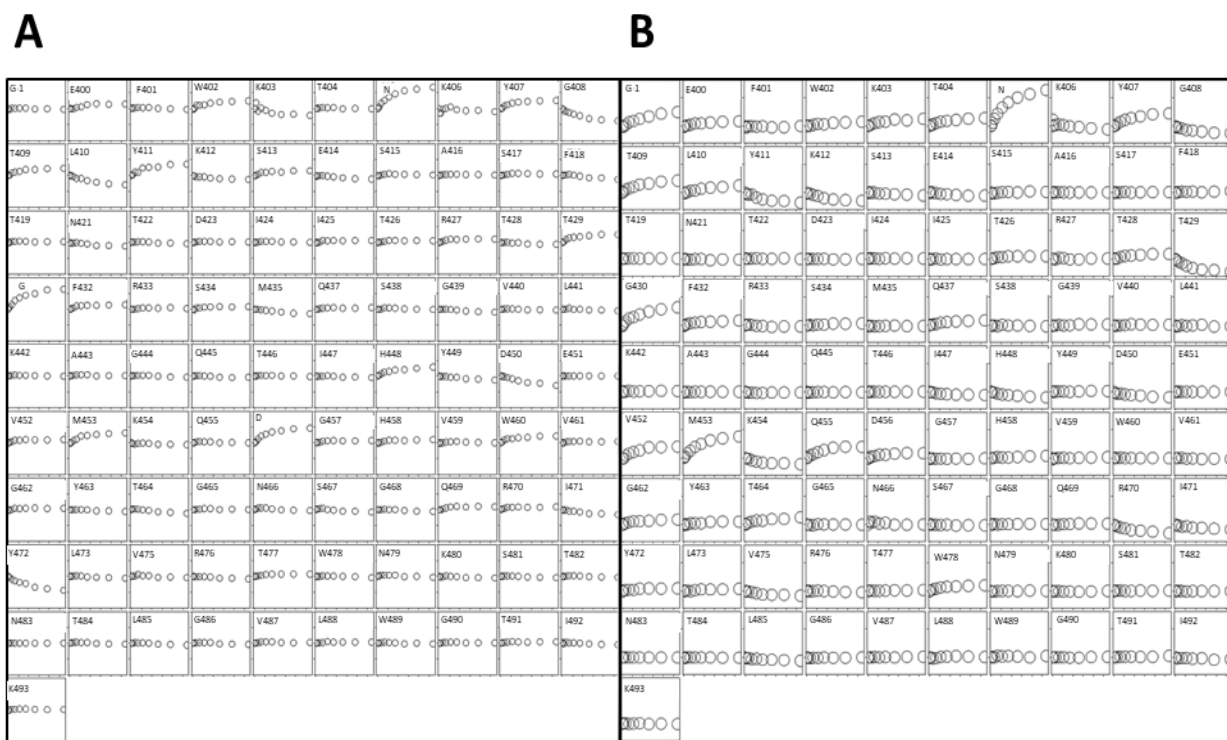
Side chain signals above the threshold are highlighted in pink.

<sup>a</sup> Normalized <sup>1</sup>H and <sup>15</sup>N chemical shifts in ppm.

<sup>b</sup> Value for the <sup>1</sup>H <sup>15</sup>N chemical shift divided by the mean.

<sup>c</sup> Average chemical shift value from all residues when titrated with the corresponding ligand.

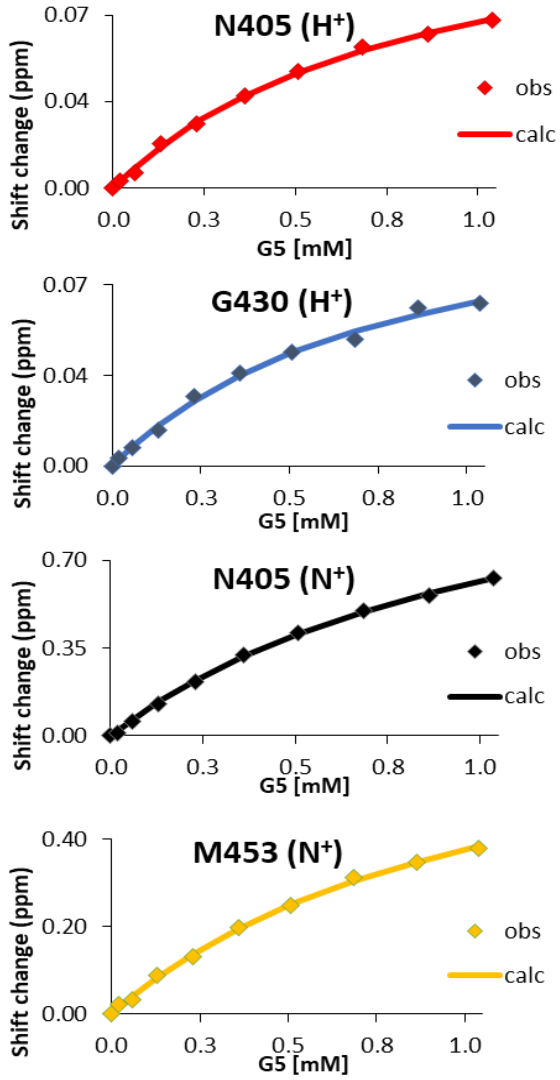
\*Intermediate conformational exchange was reported for the backbone and side chain of this residue, showing line broadening and disappearance of signal when adding the highest ligand concentrations.



**Figure 4.3.**  $^1\text{H}$  and  $^{15}\text{N}$  chemical shifts from all SH3\_5 residues upon titration of G5 peptide. **(A)** Individual  $^1\text{H}$  chemical shift charts from all SH3\_5 residues after nine titration points with the G5 peptide. The y-axis represents the relative chemical shift of the  $^1\text{H}$  nuclei from each residue (0.10 to 0.10 ppm) and the x-axis the ligand concentration (0-1 mM). **(B)** Individual  $^{15}\text{N}$  chemical shift charts from SH3\_5 residues after nine titration points with the G5 peptide. The y-axis corresponds to the relative chemical shift of the  $^{15}\text{N}$  nuclei from each residue (-0.30 to 0.80 ppm) and the x-axis represents the ligand concentration (0-1 mM). The plotted titrations were used to select  $^1\text{H}$  and  $^{15}\text{N}$  chemical shifts that could be fitted into saturation curves to measure binding affinities ( $K_d$ ).



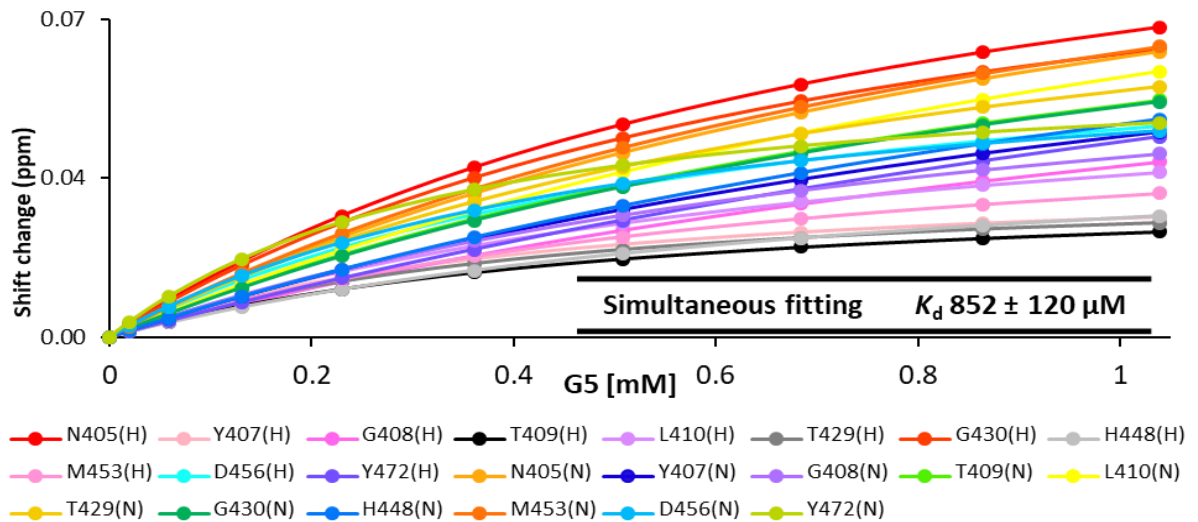
**A**



**B**

Individual nuclei	$K_d$ ( $\mu\text{M}$ )
N405(H <sup>+</sup> )	739
G430(H <sup>+</sup> )	760
N405(N <sup>+</sup> )	1060
M453(N <sup>+</sup> )	996
<b>Average <math>K_d</math></b>	<b>889 <math>\pm</math> 163</b>

**C**



**Figure 4.4. Calculated saturation curves for determining SH3\_5-G5 binding affinities. (A)** Binding affinity estimation by averaging individual  $K_d$  values. Four nuclei with the largest chemical shifts from the G5 titration were assigned with  $K_d$  and  $\Delta\delta_{\max}$  test values to fit their CSP data to equation:  $(\Delta\delta_{\text{obs}} = \Delta\delta_{\text{max}} \{ ([P]_t + [L]_t + K_d) - \sqrt{([P]_t + [L]_t + K_d)^2 - 4[P]_t[L]_t} \} / 2[P]_t)$ . Using the least-square fit method, calculated saturation curves were obtained and individual  $K_d$  values were determined: N405 (H) in red, G430 (H) in blue, N405 (N) in black, and M453 (N) in yellow. An average  $K_d$  of  $889 \pm 163 \mu\text{M}$  was estimated. **(B)** Binding affinity estimation by fitting different CSP data simultaneously. A common  $K_d$  with different  $\Delta\delta_{\max}$  is used to fit individual CSP data into the same equation as before but fitting all data simultaneously. A single fitting is used as the sum of squares for all resonances (22 resonances). The N chemical shifts were re-scaled to give comparable shift changes to the H nuclei. A global binding affinity of  $K_d = 852 \mu\text{M}$  was calculated.

#### **4.3.2.1.1 Pentaglycine orientation on the SH3\_5 binding cleft**

To further investigate the SH3\_5-G5 interaction in solution, the orientation of the G5 peptide in the cleft was analysed by NMR. The reason for this study was that although the crystal structure shows only a single orientation, it looks possible to fit a pentaglycine into the binding cleft in both orientations, and there are well-established examples of protein binding sites being able to bind peptides in both orientations (Fernandez-Ballester, 2004; Saksela, 2012). To test this hypothesis, the binding of SH3\_5 was analysed by titrating two synthetic pentaglycine peptides with an extra tyrosine residue either at the N-terminus (YGGGGG; YG5) or the C-terminus [GGGGGY; G5Y]. <sup>15</sup>N HSQC titrations were carried out under the same conditions as previously described for G5.

##### **4.3.2.1.1.1 Binding to YGGGGG (YG5)**

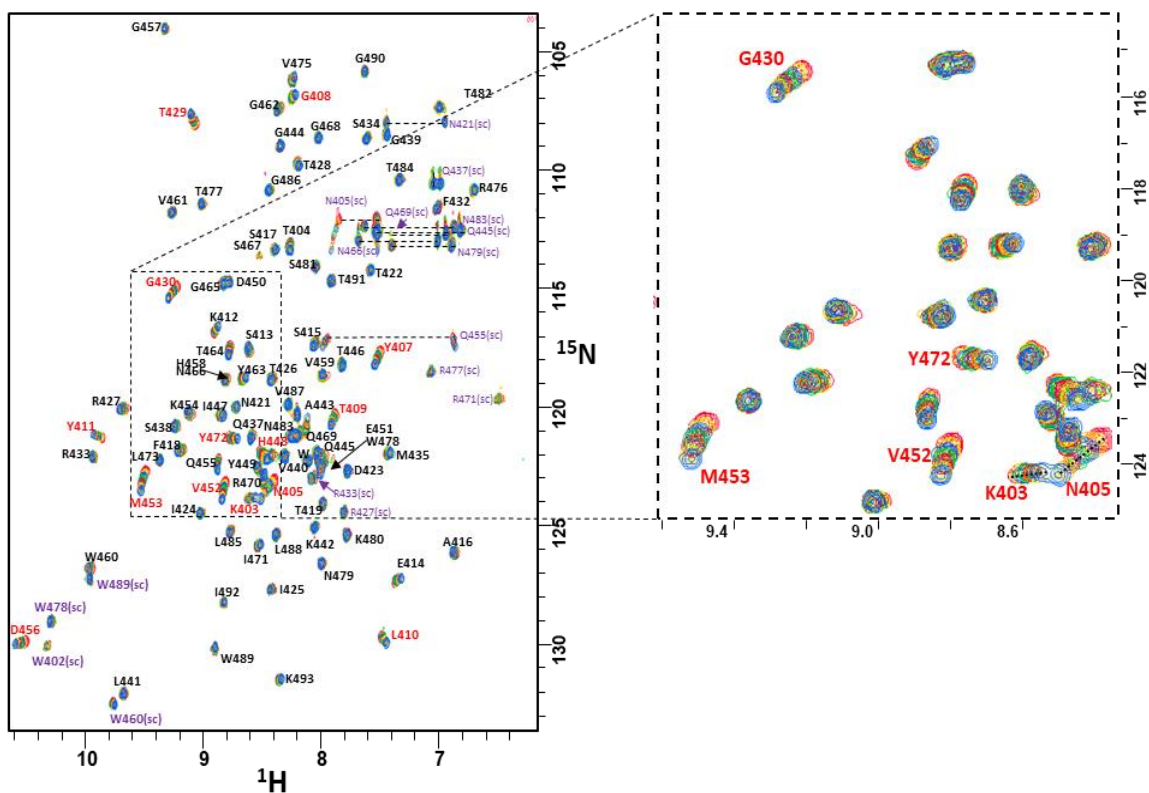
SH3\_5-YG5 NMR titrations were recorded after the addition of 1, 2.5, 4.5, 7.5, 11.5, 16.5, 22.5 and 30 equivalents of peptide. The CSP of the SH3\_5-YG5 interaction displayed a fast exchange rate as observed for the G5 titration (Fig. 4.5 A). The largest peak shifts above the established threshold ( $\Delta\delta/\text{average}\Delta\delta \geq 2$ ) were residues K403, N405, Y407, T409, Y411, T429, G430, H448, V452, M453, D456, and Y472. Only one residue (K403) was not identified in the G5 titration and appeared to be specifically interacting with the tyrosine located at the N-terminus of the peptide. This amino acid is in the  $\beta$ 1 strand at the top of the binding cleft, where the N-terminal end of the G5 peptide is found in the crystal structure of SH3\_5-pentaglycine complex (PDB 5LEO) (Fig. 4.5 B).

YG5 binding affinities were calculated by fitting individual values as described previously. The average  $K_d$  from fitting individual nuclei was  $345 \pm 60 \mu\text{M}$ .

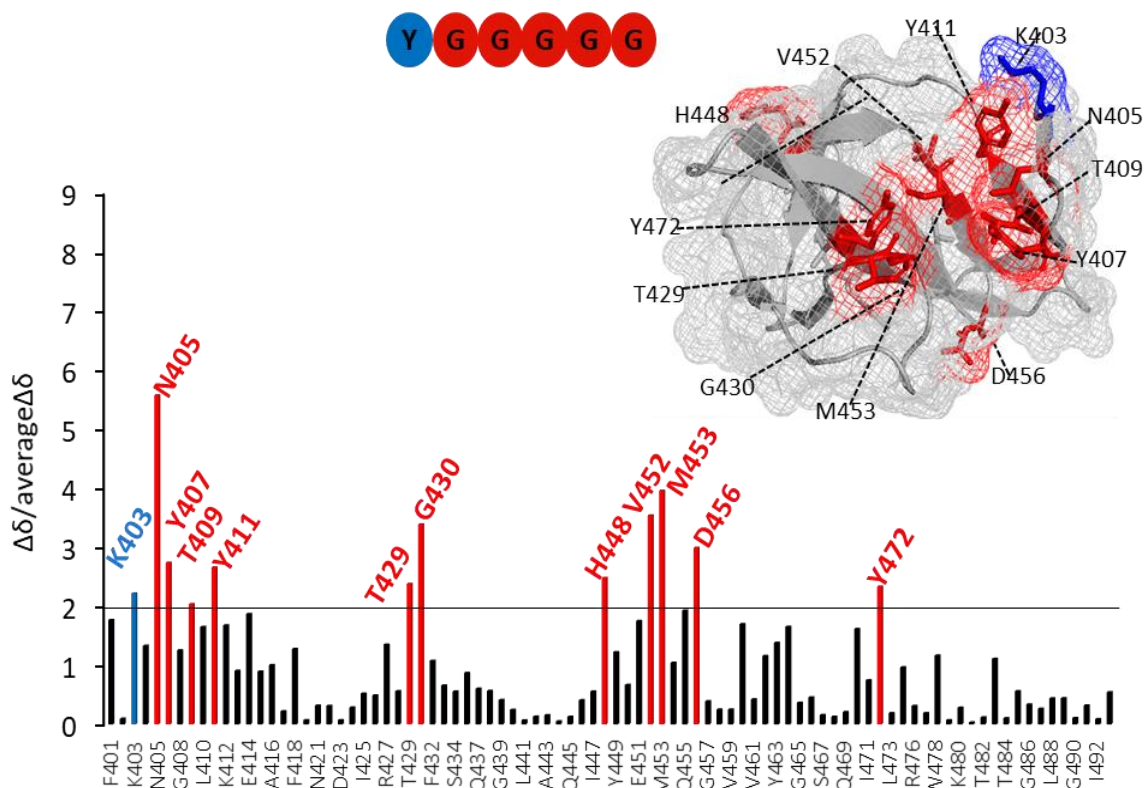
##### **4.3.2.1.1.2 Binding to GGGGGY [G5Y]**

The G5Y NMR titrations were performed as described for the YG5 peptide. In agreement with G5 and YG5, SH3\_5 binding to G5Y exhibited a fast exchange rate (Fig. 4.6 A). Residues with the largest peak shifts were N405, Y407, T409, L410, Y411, T429, G430, R433, S434, H448, V452, M453, and D456. Two residues, R433 and S434, located in the RT loop between  $\beta$  strands 3 and 4 at the

**A**

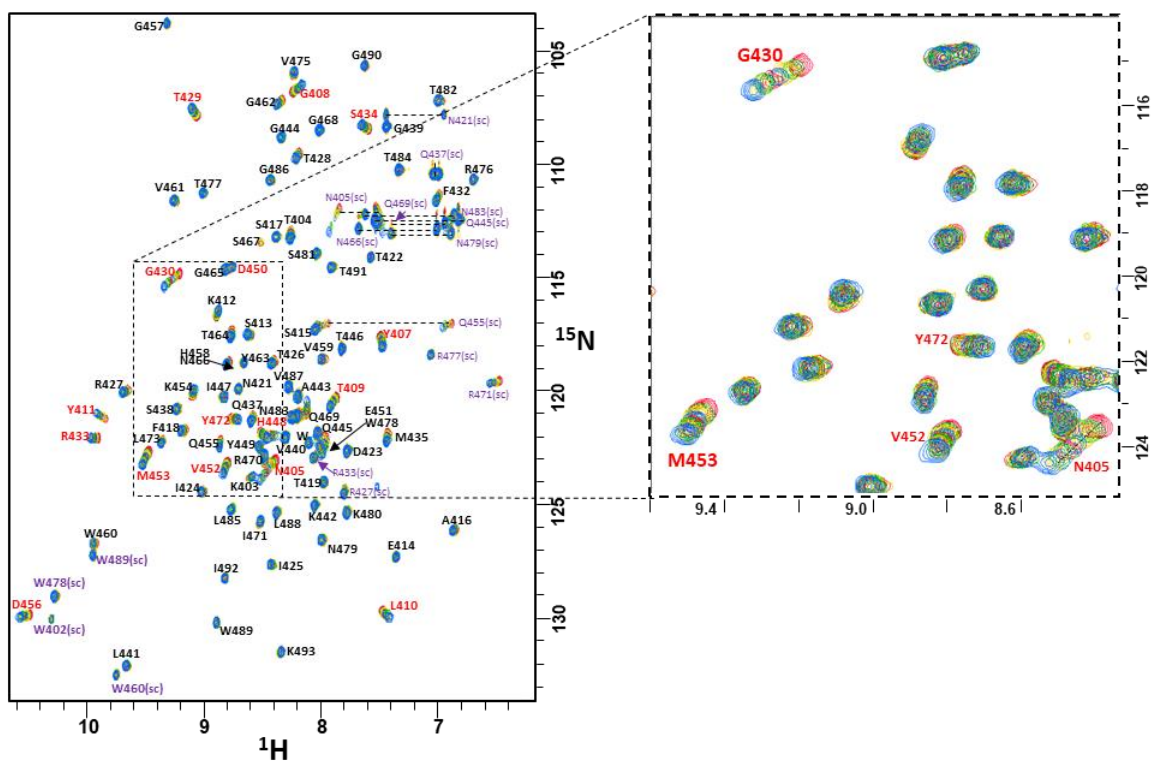


**B**

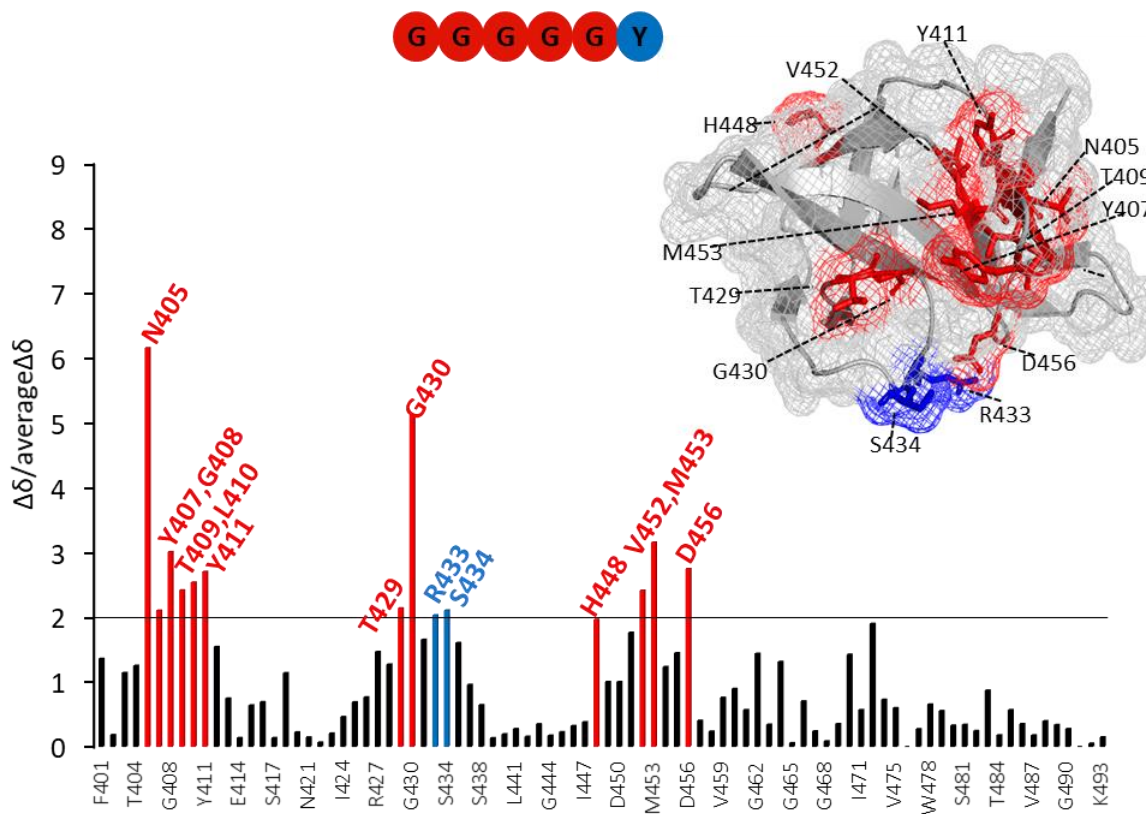


**Figure 4.5. Mapping of SH3\_5 residues interacting with the YG5 peptide by NMR titrations. (A)** Superposition of the  $^1\text{H}^{15}\text{N}$  HSQC spectra from the SH3\_5-YG5 titration acquired in the absence (red contour) and presence (yellow to blue contour) of 0.5 to 30 equivalents of peptide. Peaks from the backbone NH assignment are labelled in black and sidechains in purple. The enlarged view illustrates a region of the spectra with peaks interacting with the YG5 peptide in fast exchange. The largest peak shifts are indicated in red. **(B)** Histogram of the observed CSP values calculated as  $\Delta\delta = \sqrt{(\Delta\delta\text{H}^2 + (0.154 \times \Delta\delta\text{N})^2)}$ , as a function of the amino acid sequence. The  $y$ -axis represents the ratio between individual CSPs and the average CSP (taking all residues into account). An arbitrary threshold of 2 was chosen. Residues above the threshold associated with G5 CSPs are in red (N405, Y407, T409, Y411, T429, G430, H448, V452, M453, D456, and Y472). Residue K403 (in blue), appeared to be specifically interacting with the tyrosine at the N-terminus of the peptide. The SH3\_5 protein (PDB 5LEO) was used to map the residues identified to interact with the YG5 peptide.

**A**



**B**



**Figure 4.6. Mapping of SH3\_5 residues interacting with the G5Y peptide by NMR titrations. (A)** Overlay of  $^1\text{H}^{15}\text{N}$  HSQC spectra of the SH3\_5-G5Y titration acquired in the absence (red contour) and presence (yellow to blue contour) of 0.5 to 30 equivalents of peptide. Peaks from the backbone NH assignment are labelled in black and sidechains in purple. The enlarged view shows a region of the spectra with peaks interacting with the G5Y peptide in fast exchange. The largest peak shifts are indicated in red. **(B)** Histogram of the observed CSP values calculated as  $\Delta\delta = \sqrt{(\Delta\delta\text{H}^2 + (0.154 \times \Delta\delta\text{N})^2)}$ , as a function of the amino acid sequence. The y-axis represents the ratio between individual CSPs and the average CSP (taking all residues into account). An arbitrary threshold of 2 was chosen. Residues above the threshold associated with G5 CSPs are in red (N405, Y407- Y411, T429, G430, H448, V452, M453, and D456). Residues R433 and S434 (in blue), appeared to be specifically interacting with the tyrosine at the C-terminus of the peptide. The SH3\_5 protein (PDB 5LEO) was used to map the residues identified to interact with the G5Y peptide.

bottom of the G5 binding groove were identified to interact with the C-terminal tyrosine (Fig. 4.6 B). Similar binding affinities were observed for both YG5 and G5Y peptides. The averaged  $K_d$  of the SH3\_5-G5Y interaction was  $494 \pm 48 \mu\text{M}$ .

Titration experiments with YG5 and G5Y peptides therefore indicated the expected preferential binding orientation of the pentaglycine motif in the cleft, with no evidence that the peptide binds in the opposite orientation.

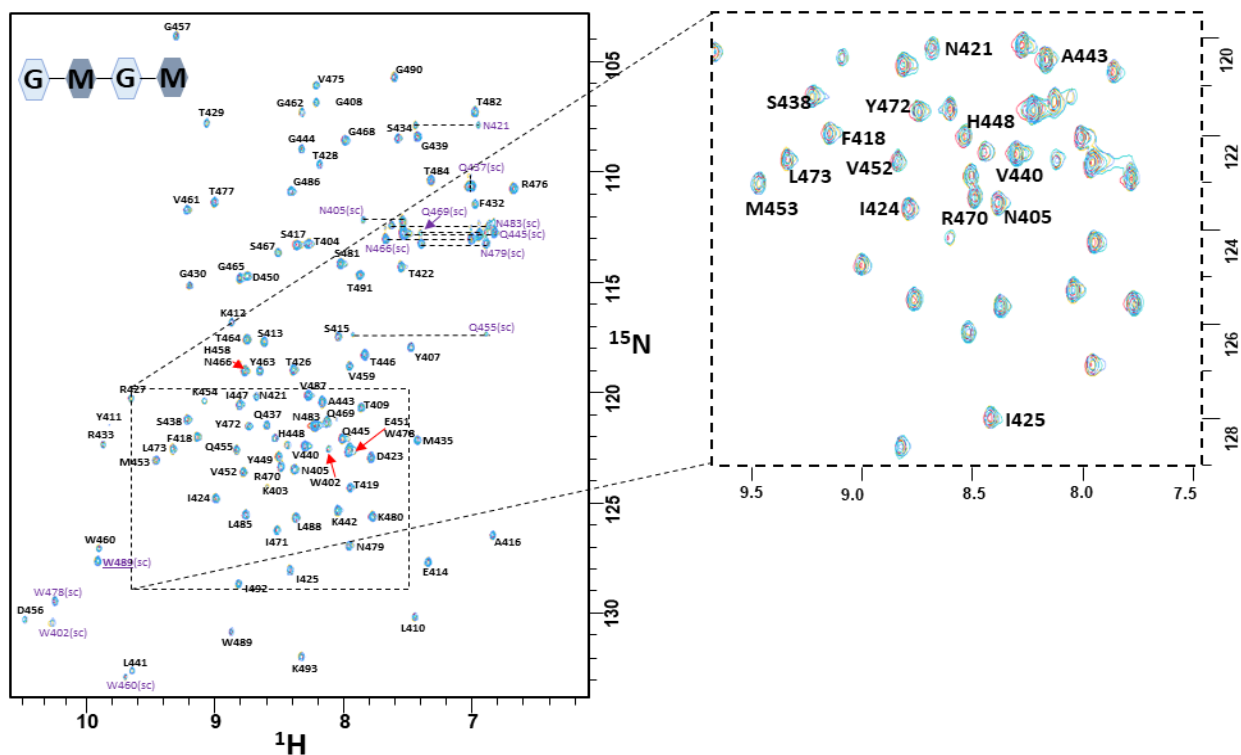
#### **4.3.2.2 SH3\_5-tetrasaccharide [GM-GM] interaction**

NMR titration experiments were performed under the same conditions as the previous ligands. No CSPs (peak movements or intensity changes) were identified on the spectra after the addition of 0.5, 1, and 2 equivalents of tetrasaccharide. All peaks were in line with our blank spectrum (no ligand added) revealing no interaction between the SH3\_5 domain and the saccharide moieties (Fig. 4.7).

#### **4.3.2.3 SH3\_5-Lactyl(+) AyQKA [P4] binding**

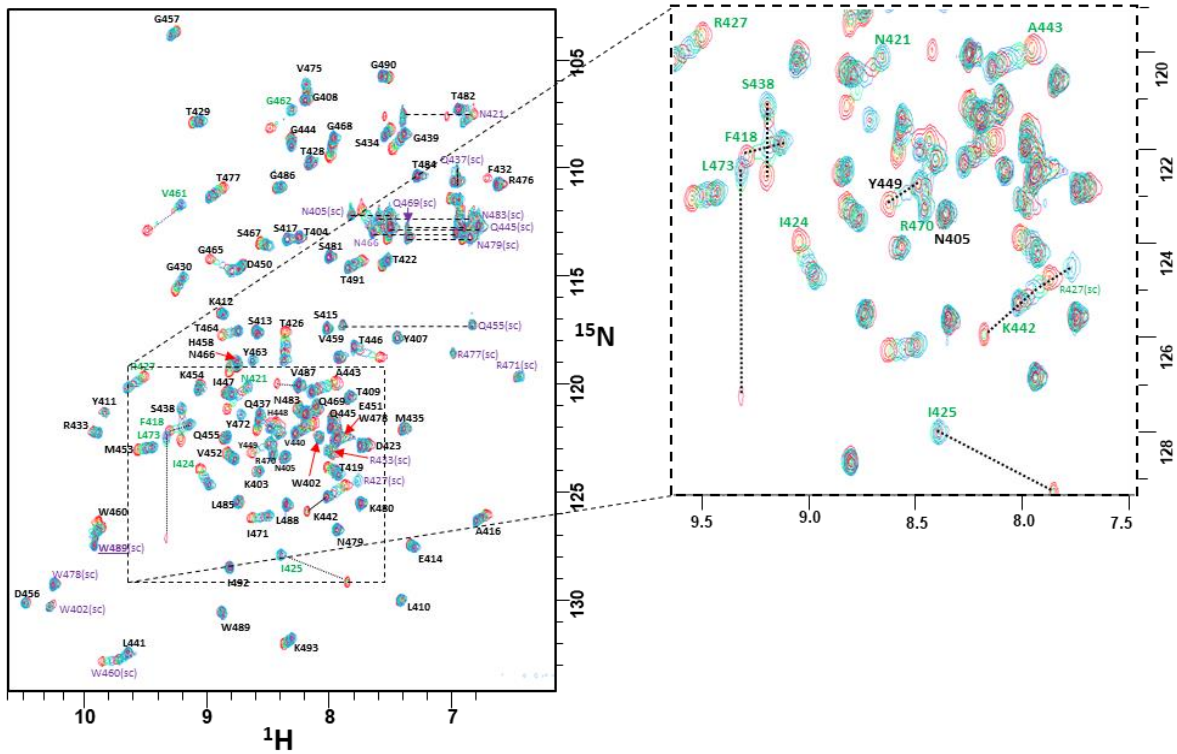
The weak affinity of the SH3\_5 domain for the G5 peptide suggested that other residues may contribute to the recognition of PG. The interactions of the SH3\_5 protein with synthetic P4 peptides were investigated. Due to the high level of cross-linkage in *S. aureus* PG, the P4 peptides resemble the most abundant peptide stems. NMR spectra were acquired after the addition of 0.33, 0.66, 1, 2, 4, 8, 16, and 32 equivalents of peptide. Interestingly, SH3\_5 proteins bound to the P4 ligand and their binding displayed a fast exchange equilibrium in a similar manner to G5 (Fig. 4.8 A). Residues interacting with P4 were located on the opposite side of the protein surface ( $\sim 180^\circ$  rotation), with the largest chemical shifts observed for N421 located between strands  $\beta 2$  and  $\beta 3$ ; I425 and T426, in strand  $\beta 3$ ; S438 in strand  $\beta 4$ ; K442 and A443 between strands  $\beta 4$  and  $\beta 5$ ; V461, G462, and G465 in strand  $\beta 6$ ; and R470 and L473 in strand  $\beta 7$  (Fig. 4.8 B). The sidechain of residue R427 also displayed a CSP above the established threshold and is in the same region as the identified P4 interactions (Table 4.1).



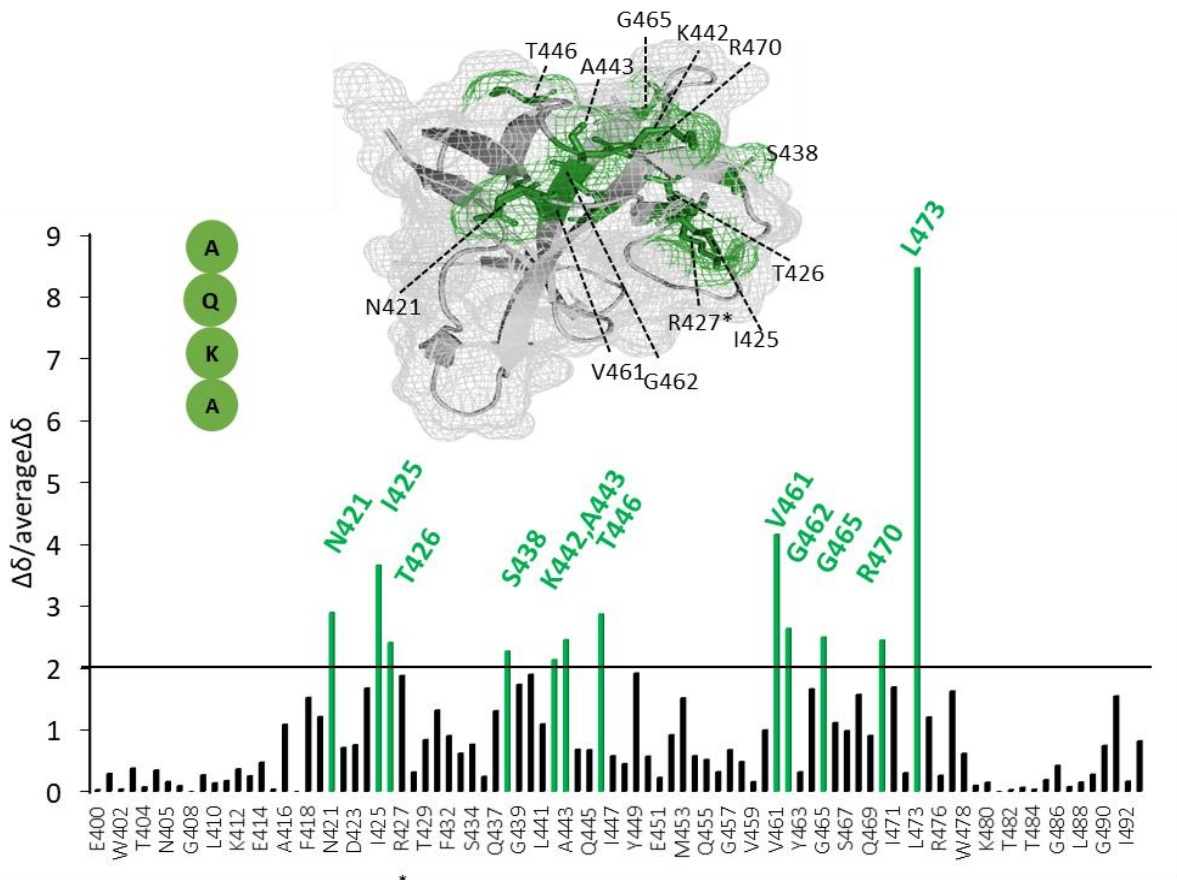


**Figure 4.7.** SH3\_5-tetrasaccharide [GM-GM] NMR titration. Overlay of  $^1\text{H}^{15}\text{N}$  HSQC spectra of the SH3\_5-GM-GM titration acquired in the absence (red contour) and presence (yellow to blue contour) of 0.5 to 2 equivalents of peptide. Peaks from the backbone NH assignment are labelled in black and sidechains in purple. The enlarged view shows in better detail the superposition of peaks due to the absence of interaction between the SH3\_5 domain and the GM-GM ligand.

**A**



**B**



**Figure 4.8. Mapping of SH3\_5 residues interacting with the P4 peptide by NMR titrations. (A)** Overlaid  $^1\text{H}^{15}\text{N}$  HSQC spectra of the SH3\_5-P4 titration acquired in the absence (red contour) and presence (yellow to blue contour) of 0.33 to 32 equivalents of peptide. Peaks from the backbone NH assignment are labelled in black and sidechains in purple. The enlarged view illustrates a region of the spectra with peaks interacting with the P4 peptide in fast exchange. The largest peak shifts are indicated in green. **(B)** Histogram of the observed CSP values of all residues calculated as  $\Delta\delta = \sqrt{(\Delta\delta\text{H}^2 + (0.154 \times \Delta\delta\text{N})^2)}$ . The y-axis represents the ratio between individual CSPs and the average CSP (taking all residues into account). An arbitrary threshold of 2 was chosen. Residues associated with CSPs above the threshold are in green. The SH3\_5 protein (PDB 5LEO) was used to map the residues identified to interact with P4 (N421, I425, T426, S438, K442, A443, T446, V461, G462, G465, R470, and L473). The (\*) in residue R427 indicates that the sidechain crossed the threshold.

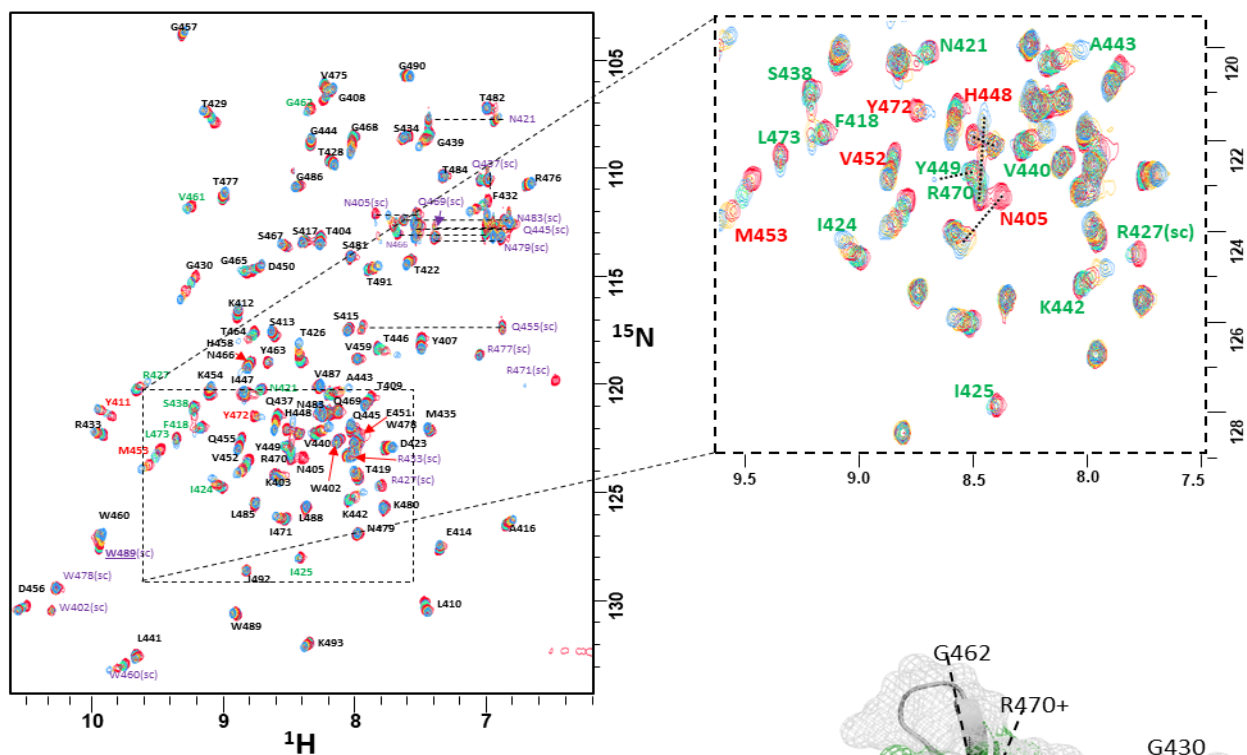
Residues V461 and L473 displayed the most pronounced chemical shift changes yet are buried in the structure suggesting no direct contact with the ligand. NMR chemical shift changes are very sensitive and can be measured very accurately. This implies that almost any genuine interaction will produce a CSP, however binding could also affect adjacent residues. It is possible to hypothesize that given the small size of the P4 peptide, it possess a dynamic structure that would allow it to fit and facilitate contact with a region in close proximity to these residues, and it is also likely that the protein changes its conformation upon binding to P4. This conformational change, however, must be subtle since there are no obvious rearrangements in the bound crystal structure as discussed later.

The average  $K_d$  displayed a low affinity in the millimolar range ( $963 \pm 198 \mu\text{M}$ ). Like the G5 peptide, SH3\_5 domains showed a weak binding to the P4 ligand. These results showed that the SH3\_5 domain recognises both PG peptide moieties but on opposite faces of the protein.

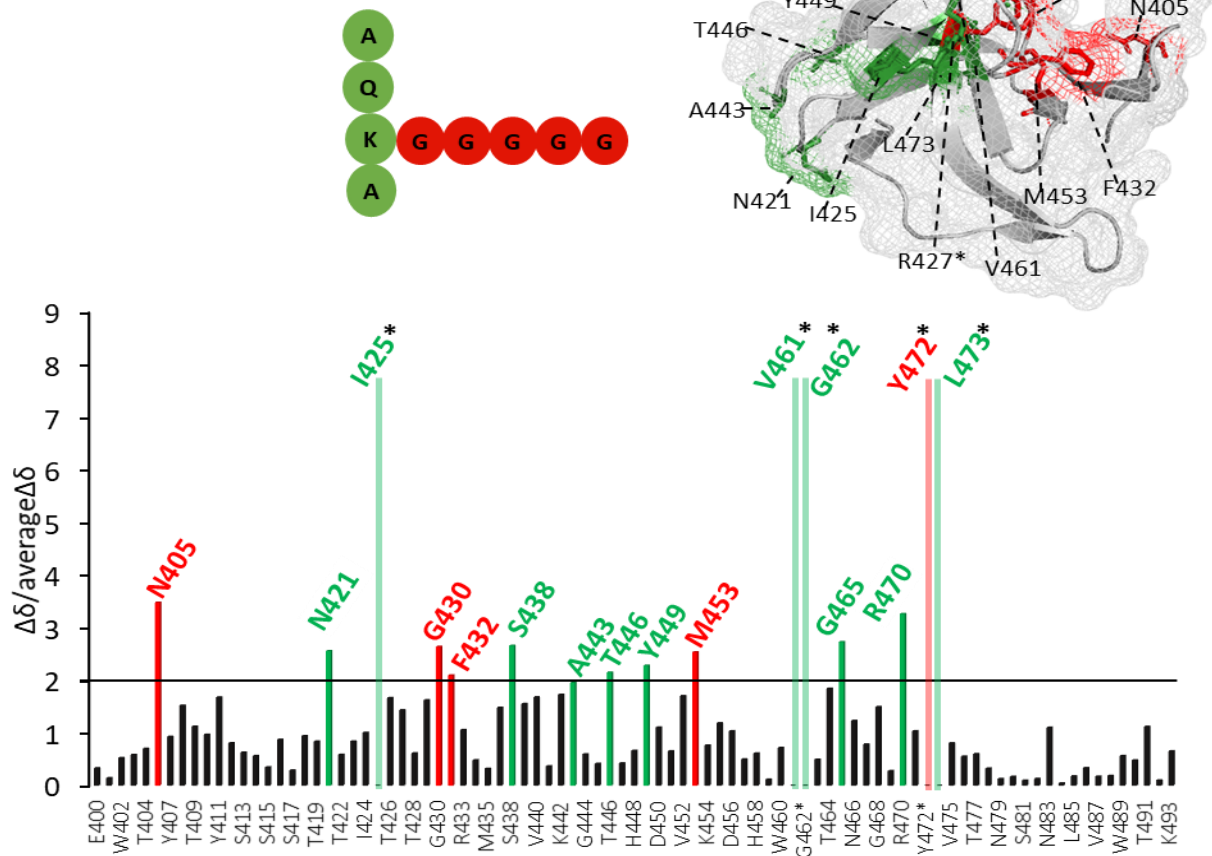
#### **4.3.2.4 SH3\_5-Lactyl(+)-AyQK[GGGGG]A binding**

To expand our studies on SH3\_5 binding mechanisms and to confirm specific interactions to both peptide moieties, binding to a monomeric P4-G5 ligand was analysed. Titrations were acquired after the addition of 0.33, 0.66, 1, 1.5, 3, 5, and 9 equivalents of peptide (Fig. 4.9 A). During the first titration points a larger number of peaks shifted throughout the spectra in comparison to the simpler ligands. After the addition of 1 equivalent of peptide, seven peaks that were originally shifting broadened and disappeared (N405, I425, V461, G462, Y472, L473, and the side chain of R427) indicating stronger interactions; N405 was the only signal to re-appear for the last two titration points (Fig. 4.9 B; Table 4.1). In total, sixteen residues were identified to interact with the ligand, fourteen of them corresponding to residues previously determined in the G5 and P4 titrations (N405, N421, I425, G430, S438, A443, T446, M453, V461, G462, R470, Y472, L473, and R427 sidechain) (Fig. 4.9 B). The two remaining residues, F432 and Y449, are also localised in the binding regions (Fig. 4.9 B). These observations are characteristic of an intermediate exchange equilibrium, and as expected SH3\_5 exhibited a tighter binding to the monomer P4-G5 with a  $K_d$  of  $98 \pm 42 \mu\text{M}$ , almost ten times higher than the individual G5 and P4 peptides.

**A**



**B**



**Figure 4.9. Mapping of SH3\_5 residues interacting with the P4-G5 peptide by NMR titrations.**

**(A)** Overlaid  $^1\text{H}^{15}\text{N}$  HSQC spectra of the SH3\_5-[P4-G5] titration acquired in the absence (red contour) and presence (yellow to blue contour) of 0.33 to 9 equivalents of peptide. Peaks from the backbone NH assignment are labelled in black and sidechains in purple. The enlarged view illustrates a region of the spectra with peaks interacting with the P4-G5 peptide exhibiting intermediate exchange rate characterised by shifting, broadening, and in some peaks disappearance of signals. The largest CSP are indicated in green for those residues previously identified to interact with the P4 peptide and in red for G5. **(B)** Histogram of the observed CSP values of all residues calculated as  $\Delta\delta = \sqrt{(\Delta\delta\text{H}^2 + (0.154 \times \Delta\delta\text{N})^2)}$ . The  $y$ -axis represents the ratio between individual CSPs and the average CSP (taking all residues into account). An arbitrary threshold of 2 was chosen. Residues associated with CSPs above the threshold associated with the G5 chain are in red, and those with CSPs above the threshold associated with P4 are in green. The SH3\_5 protein (PDB 5LEO) was used to map the residues identified in the P4-G5 titration (N405, N421, I425, G430, F432, S438, A443, T446, Y449, M453, V461, G462, G465, R470, Y472, and L473). Seven residues (\*) were shifting and disappeared. The (\*) in residue R470 indicates that its sidechain peak also crossed the threshold.

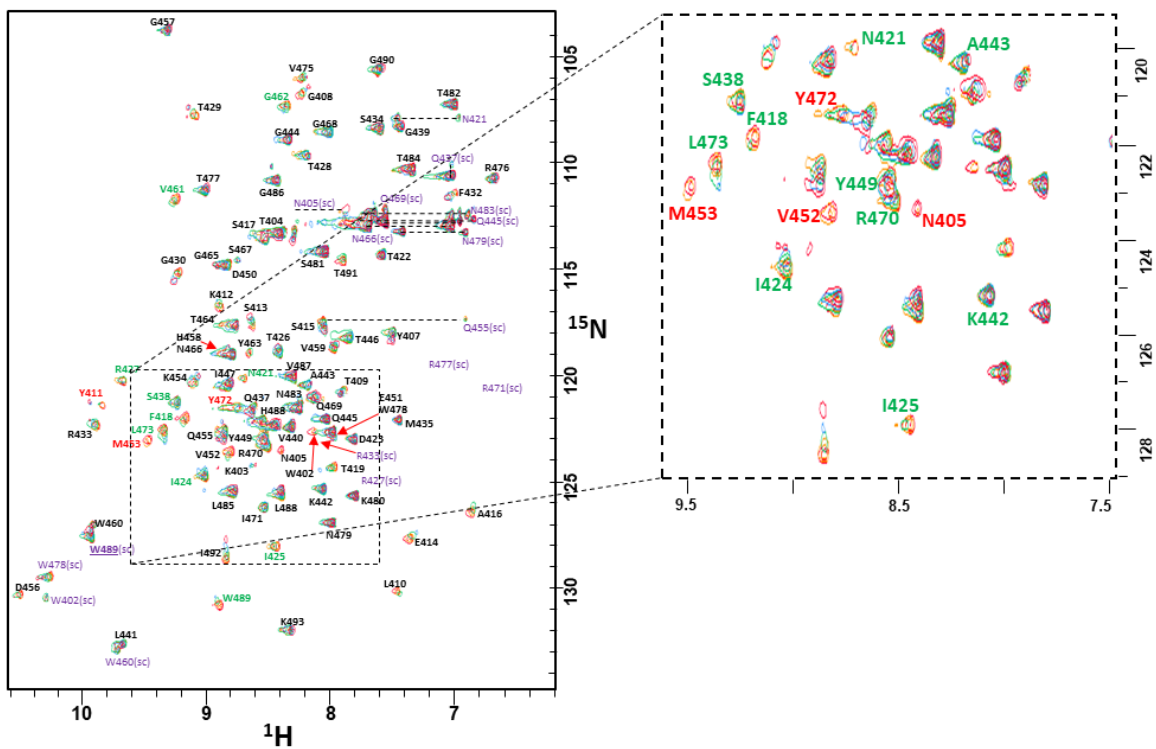
#### 4.3.2.5 SH3\_5-A $\gamma$ QK[GGGGG]AA-A $\gamma$ QKA binding

We next analysed SH3\_5 interactions with branched PG fragments. SH3 proteins were titrated with a synthetic dimer containing two peptide stems connected by a pentaglycine crossbridge [P5-G5-P4]. NMR spectra was acquired after the addition of 0.3, 1, 2, 4, 8, and 16 equivalents of peptide. During the first titration points, peaks that interacted with the ligand were seen to shift, however higher concentrations of ligand led to the aggregation of the protein. In general, the formation of aggregates significantly decreases the intensity of the NMR signals; usually all the signals of the labelled protein are affected and very often they do not recover (this is therefore different from the intermediate exchange described in the previous section, where only a subset of signals broadened and disappeared.) Unlike the monomer P4-G5, signals decreased but did not disappear and it was possible to follow changes in residues interacting with the protein (Fig. 4.10 A).

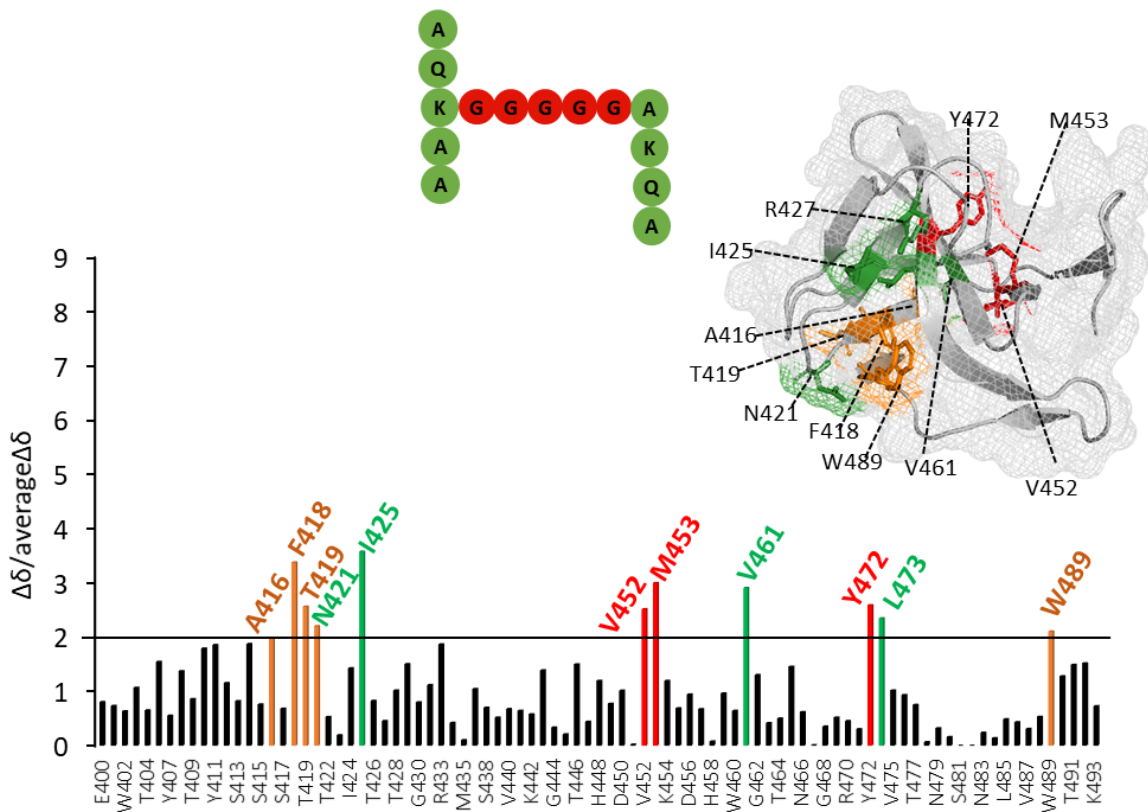
Most CSPs were observed in residues already identified using simpler ligands (N421, I425, V452, M453, V461, Y472, L473, and the sidechain of R427). One surface residue (W489) and three residues buried in the structure (A416, F418, T419) had not been previously identified. It was established that they were more likely to interact with the peptide stem based on their proximity to residues found in the P4 binding region (Fig. 4.10 B). The CSPs from P5-G5-P4 interactions exhibited fewer and less pronounced chemical changes (Fig. 4.10 B). As previously seen for peptides harbouring a P4 moiety, the side chain of residue R427 also displayed a CSP above the established threshold (Table 4.1).

SH3\_5-P5-G5-P4 binding affinities were approximately 10 times higher than for the individual peptides P4 and G5, as determined for the P4-G5 peptide. An average  $K_d$  of  $100 \pm 34 \mu\text{M}$  was calculated. It is worth noting that compared to P4-G5, the fitted affinity is roughly the same, but the chemical changes were smaller and protein aggregation was observed. These features are discussed later.

**A**



**B**





**Figure 4.10. Mapping of SH3\_5 residues interacting with the P5-G5-P4 peptide by NMR titrations. (A)** Overlaid  $^1\text{H}^{15}\text{N}$  HSQC spectra of the SH3\_5-[P5-G5-P4] titration acquired in the absence (red contour) and presence (yellow to green contour) of 0.3 to 16 equivalents of peptide. Peaks from the backbone NH assignment are labelled in black and sidechains in purple. The enlarged view illustrates a region of the spectra where peaks that interacted with the ligand shifted during the first titrations, but at higher concentrations of ligand, protein aggregation caused significant loss of signal intensity. The largest CSP are indicated in green for those residues previously identified to interact with the P4 peptide and in red for G5. **(B)** Histogram of the observed CSP values of all residues calculated as  $\Delta\delta = \sqrt{(\Delta\delta\text{H}^2 + (0.154 \times \Delta\delta\text{N})^2)}$ . The y-axis represents the ratio between individual CSPs and their average CSP. An arbitrary threshold of 2 was chosen. Most residues associated with CSPs above the threshold had already been identified with simpler ligands, residues associated with the G5 chain are in red (V452, M453, and Y472), those associated with P4 are in green (N421, I425, V461, and L473). Four residues (A416, F418, T419, and W489) had not been previously identified (orange), however they were more likely to interact with the peptide stem based on their proximity to the P4 binding region. The SH3\_5 protein (PDB 5LEO) was used to map the residues identified in the P5-G5-P4 titration.

#### **4.3.2.6 SH3\_5- DS-AγQK[GGGGG]A-DS-AγQK[GGGGG]AA binding**

Our most complex defined ligand, the dimer GM-P5-G5-GM-P4-G5, was titrated by adding 0.25, 0.5, 1, 2, 3.5, and 5 equivalents of peptide (Fig. 4.11 A). All CSP corresponded to peaks previously identified (N405, N405 sidechain, Y411, A416, F418, V452, M453, V461, L473, and W489) (Fig. 4.11 B; Table 4.1). Interestingly, interactions with residues A416, F418, and W489 were only determined for our two most complex ligands.

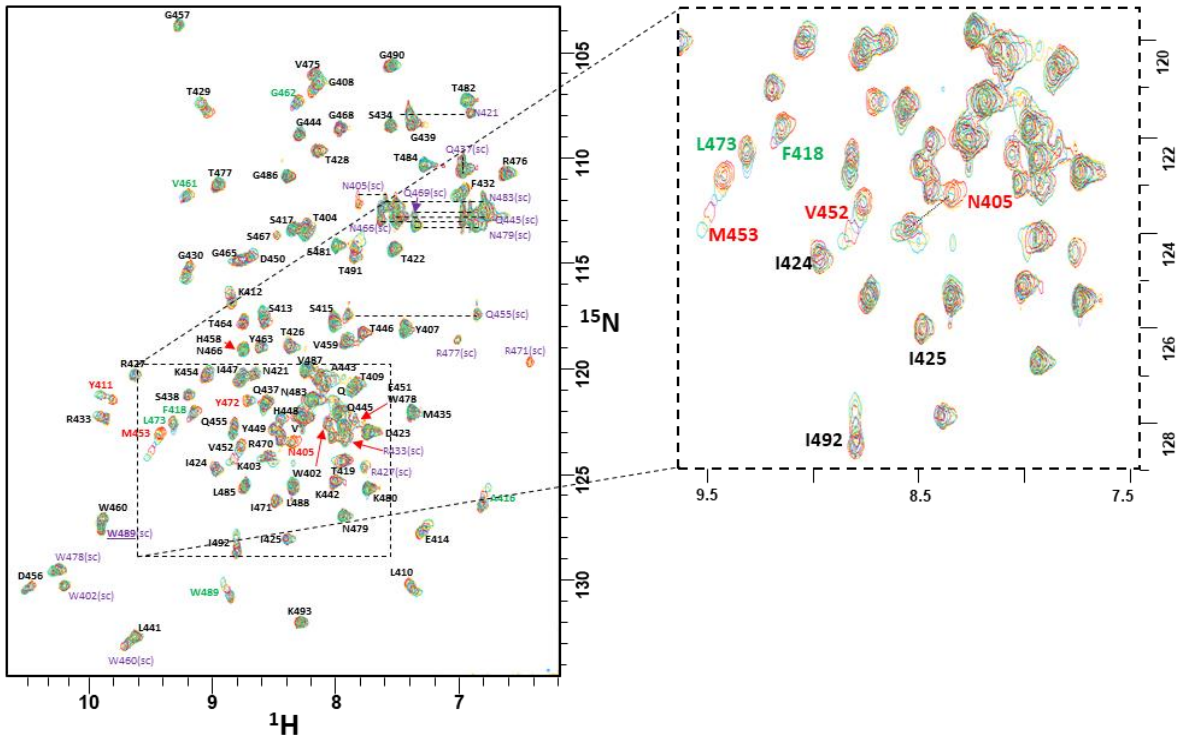
Binding affinities were not determined since the protein precipitated after the addition of the fourth titration point, resulting in line broadening and general loss of signal intensity. The unusual changes in the line shape during titrations with the most complex ligands indicated that more complex interactions were happening, such as a possible structural rearrangement of the protein/ligand complex, or protein oligomerisation due to binding. As discussed later, we believe that the correct interpretation is protein oligomerisation arising from interactions at multiple sites on the protein.

#### **4.3.2.7 SH3\_5 binding to a complex mixture of uncrosslinked (linear) PG fragments**

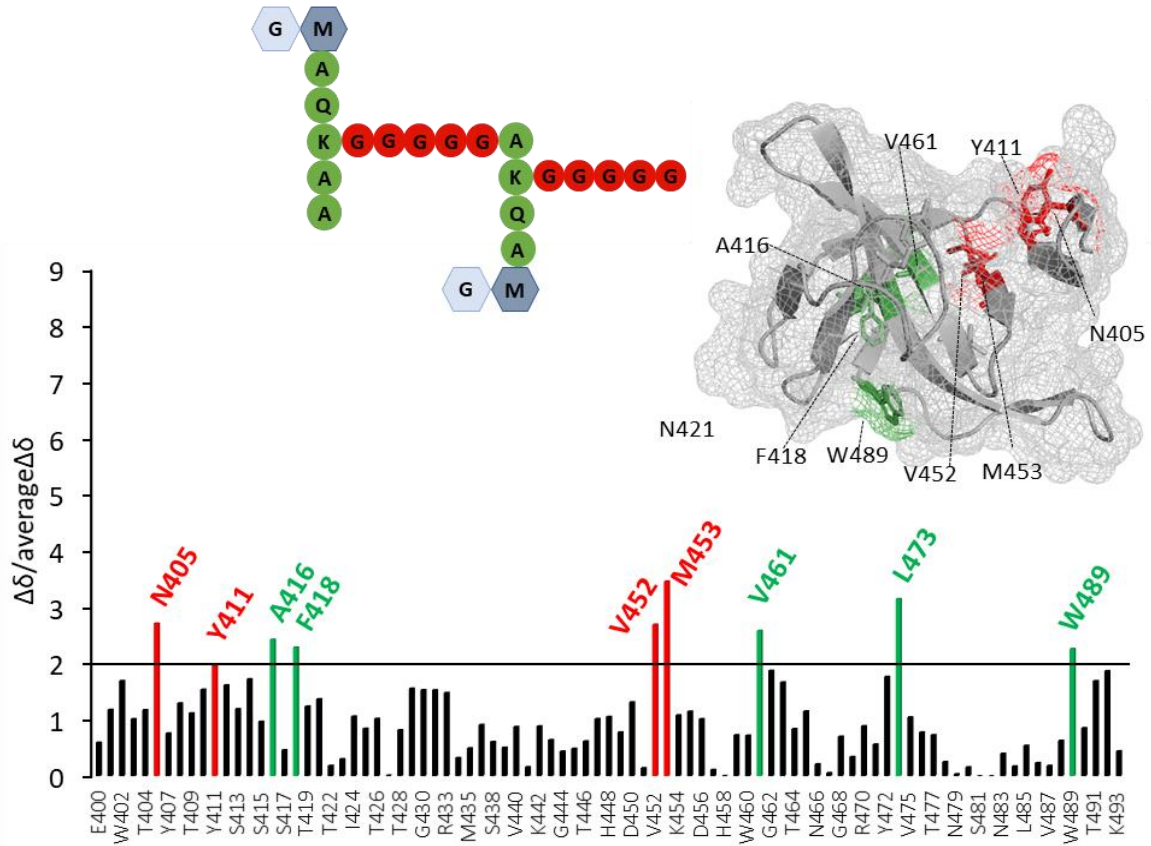
To study the binding of the SH3\_5 domain to complex PG fragments, we solubilised PG sacculi using Lss. The resulting fragments correspond to uncrosslinked material (linear glycan strands substituted by heptapeptide stems). Titrations were performed by adding 8, 16, 24, and 32 μL of the digestion mixture. Peaks interacting with the protein hardly moved, instead they broadened, and their signal intensity decreased; some peaks disappeared during the last titrations, presumably due to a tight binding (Fig. 4.12 A). Protein precipitation was evident during the last titration points.

CSP analyses in the limits of the slow exchange regime can be determined by plotting changes in the intensity of the signal. First, to adjust the rate of signal decrement for all residues the intensity of the signals that disappeared during the last two titrations was set to zero. The residues that disappeared at first were Y411, I425, R427, and W489, followed by N405, V452, and S438. To correct for the presence of protein precipitation all intensities were plotted, and 18 unaffected peaks were selected (peaks not interacting). These intensities were collected and averaged to get

**A**



**B**



**Figure 4.11. Mapping of SH3\_5 residues interacting with the GM-P5-G5-MG-P4-G5 dimer by NMR titrations. (A)**  $^1\text{H}^{15}\text{N}$  HSQC spectra overlaid of the SH3\_5-[GM-P5-G5-GM-P4-G5] titration acquired in the absence (red contour) and presence (yellow to blue contour) of 0.25 to 5 equivalents of peptide. Peaks from the backbone NH assignment are indicated in black and sidechains in purple. The enlarged view shows a group of peaks that were originally shifting but after the addition of 1 equivalent of peptide the protein precipitated causing line broadening and loss of signal intensity. The largest CSP are indicated in red for those residues previously identified to interact with the G5 peptide and in green for P4. **(B)** Histogram of the observed CSP values of all residues calculated as  $\Delta\delta = \sqrt{(\Delta\delta\text{H}^2 + (0.154 \times \Delta\delta\text{N})^2)}$ . The  $y$ -axis represents the ratio between individual CSPs and their average CSP. An arbitrary threshold of 2 was chosen. Residues associated with the G5 chain are in red (N405, Y411, V452, and M45), those associated with P4 are in green (A416, F418, V461, L473, and W489). The SH3\_5 protein (PDB 5LEO) was used to map the residues identified in this titration.

the mean of 'unaffected' intensities, with these values the relative average and the standard deviation of each titration point was calculated. Then, all values were divided by each of their corresponding means to obtain a normalised table to plot our data. Due to the slow exchange equilibrium and the presence of precipitation it was not possible to fit a  $K_d$ , but the obtained values were fitted to a straight line and gradients were considered as a measure of signal decay. The biggest changes were seen for residues N421, V461, I425, R427, W489, V452, N405, S438, and M453 (Fig. 4.12 B).

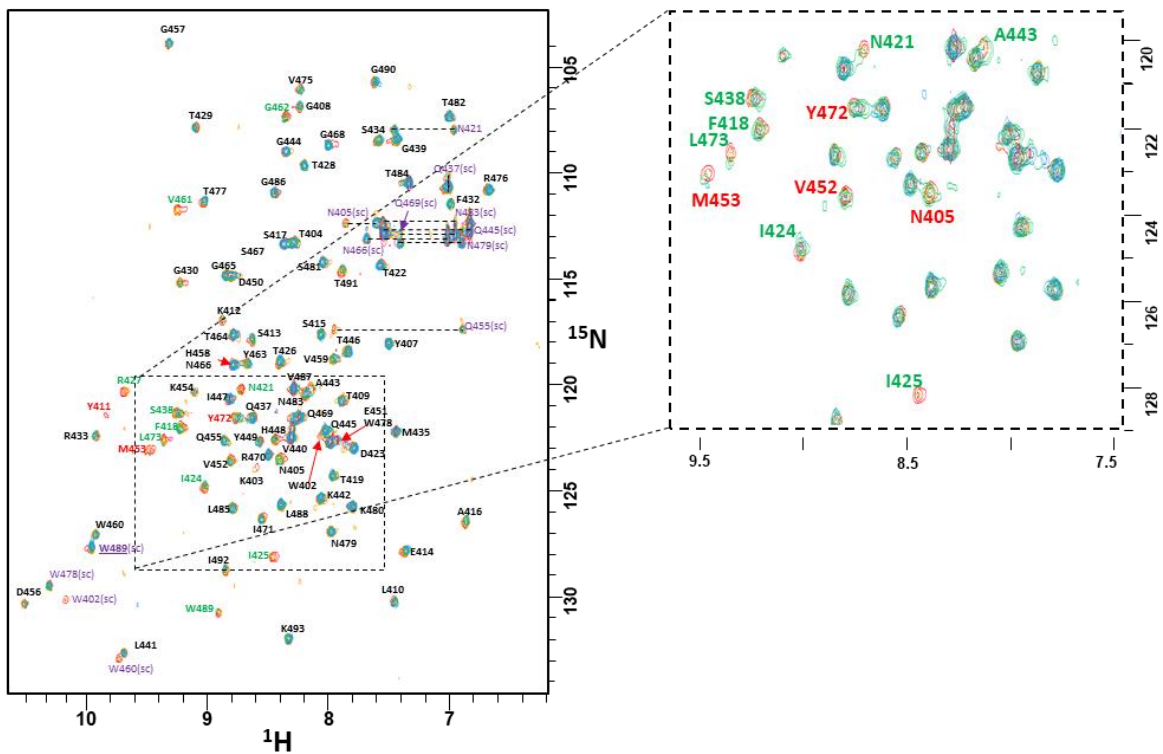
The strongest interactions were consistent with our identification of a defined group of residues in the G5 and P4 binding sites (Fig. 4.12 B). The sidechains of residues I425, R427, and W489 are exposed forming a crevice in the surface where SH3\_5 potentially fits the peptide stems.

#### **4.3.2.8 SH3\_5 binding to PG disaccharide peptides with various degrees of cross-linking**

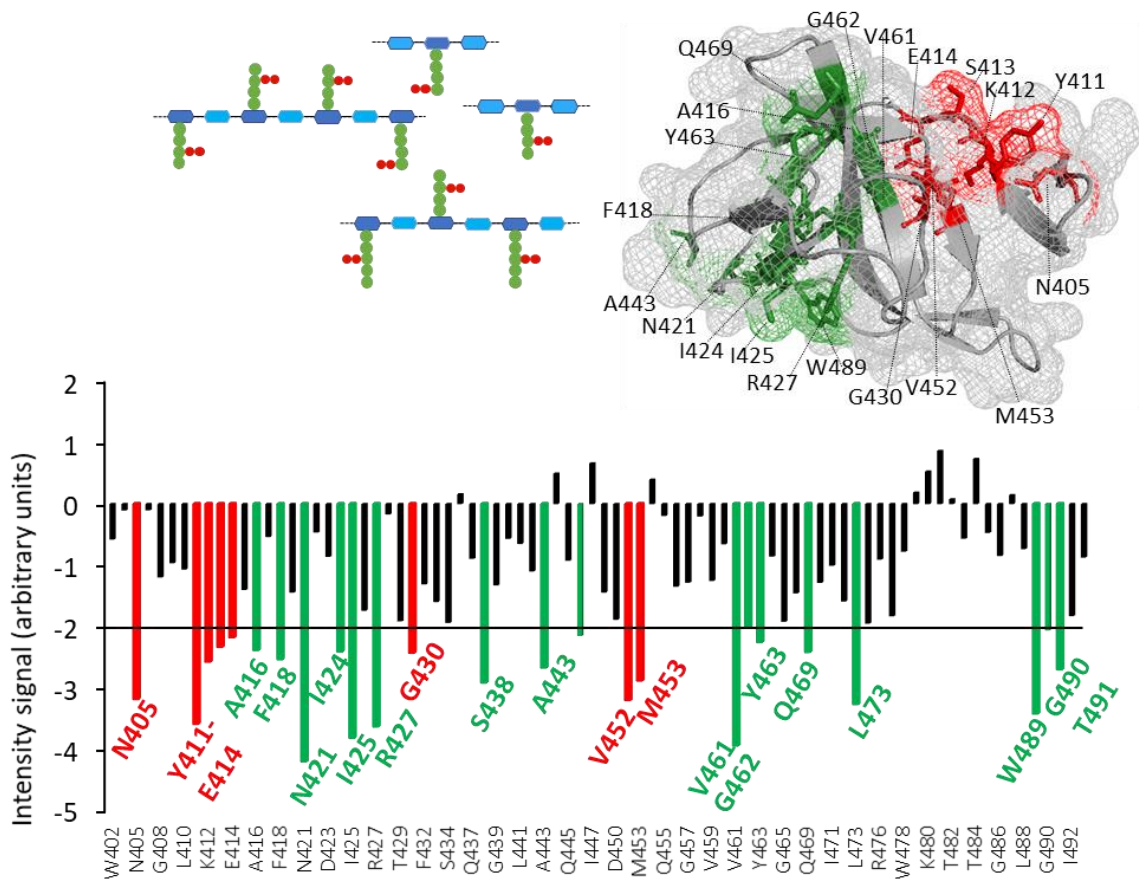
Another complex mixture of PG fragments was prepared by mutanolysin digestion of *S. aureus* PG sacculi (Fig. 4.13 A). The resulting soluble PG fragments correspond to disaccharide peptides with various degrees of cross-linking. After the addition of 10 and 30  $\mu$ L of sample, signals of residues interacting with the digestion mixture rapidly decreased and disappeared. Like in the titration with Lss products, CSP analysis exhibited a slow exchange rate reflecting a tighter binding (Fig. 4.13 B). No further titrations were carried out due to the precipitation of the protein.

CSP analysis was performed by measuring signal intensities throughout the titrations. The signal intensity of residues that disappeared in the last titrations was adjusted to zero and the signal decay from all residues was normalised to correct for the presence of protein precipitation. As explained for the previous titration, values were fitted to a straight line to obtain gradients as a measure of signal decrement. The rate of signal decay was similar to the previous titration. The strongest interactions were seen in residues K406, N405, K412, M453, V452, Y411, and V461; followed by F432, R427, I425, T491, N421, L473, T429, Y472, G430, and A416. All residues had been previously identified (Fig. 4.13 B).

**A**

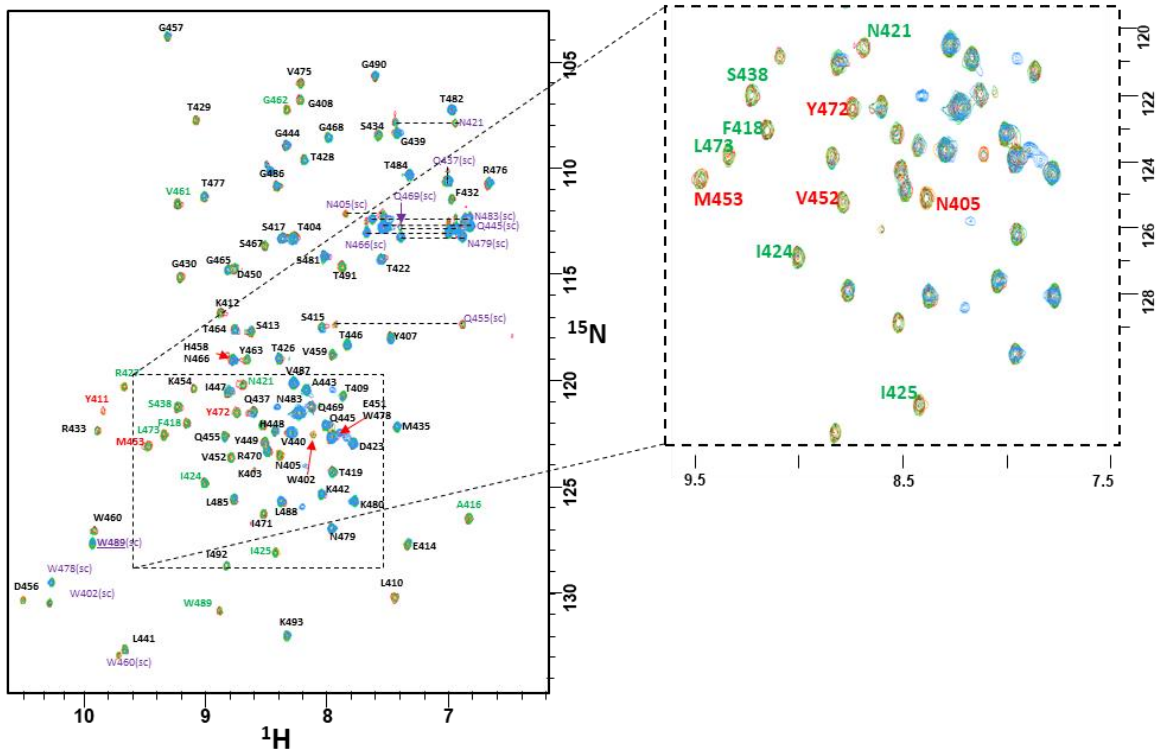


**B**

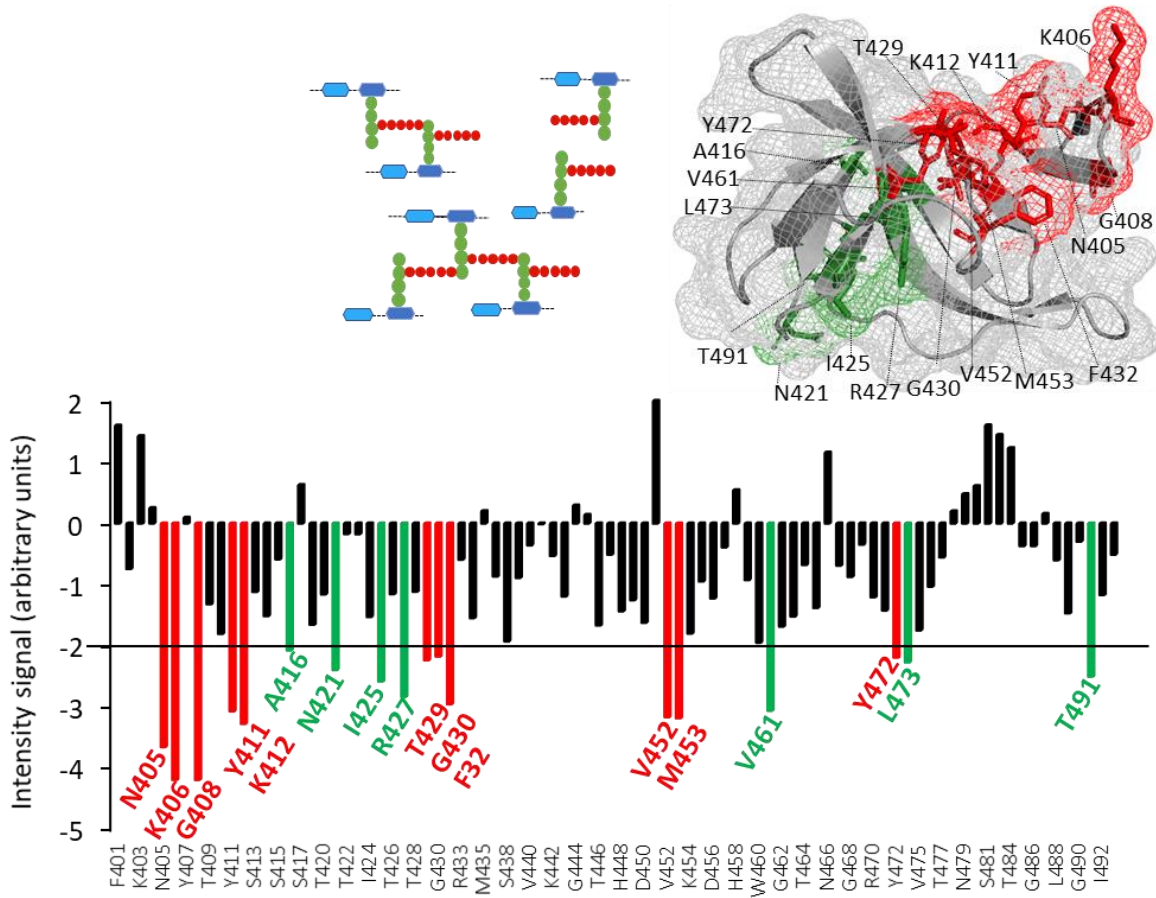


**Figure 4.12. Mapping of SH3\_5 residues interacting with *S. aureus* linear PG fragments by NMR titrations. (A)** Overlaid  $^1\text{H}^{15}\text{N}$  HSQC spectra of the SH3\_5 interaction with *S. aureus* linear PG fragments derived from Lss digestion. Titrations were acquired in the absence (red contour) and presence (yellow to blue contour) of 8 to 32  $\mu\text{L}$  of the digestion mixture. Peaks from the backbone NH assignment are labelled in black and sidechains in purple. The enlarged view shows a region of the spectra with peaks that disappeared during the last titrations (absence of the last peak contours). The largest CSP are indicated in green for those residues previously identified to interact with the P4 peptide and in red for G5. **(B)** Histogram of signal intensities as a function of protein-ligand interaction. The intensities of residues that disappeared during the last two titrations were set to 0. Values of the rate of signal decay were adjusted to correct for the presence of protein precipitation. Individual intensities were divided by the average intensity of unaffected residues calculated for each titration point. Normalised values were fitted to a straight line and the gradient was considered as a measure of signal decay ( $y$ -axis). An arbitrary threshold of -2 was chosen. Most residues associated with CSPs that decreased beyond the threshold had already been identified. Residues N421, V461, I425, R427, W489, V452, N405, S438, and M453 displayed the biggest chemical changes. The SH3\_5 protein (PDB 5LEO) was used to map the residues identified in this titration.

**A**



**B**





**Figure 4.13. Mapping of SH3\_5 residues interacting with *S. aureus* disaccharide peptides by NMR titrations.** (A) Overlaid  $^1\text{H}^{15}\text{N}$  HSQC spectra of the SH3\_5 interaction with *S. aureus* disaccharide fragments derived from mutanolysin digestion. Titrations were acquired in the absence (red contour) and presence (green to blue contour) of 10 and 30  $\mu\text{L}$  of the digestion mixture. Peaks from the backbone NH assignment are labelled in black and sidechains in purple. The enlarged view shows a region of the spectra with peaks that disappeared during the titration (absence of the last peak contours). The largest CSP are indicated in green for those residues previously identified to interact with the P4 peptide and in red for G5. (B) Histogram of signal intensities as a function of protein-ligand interaction. The signals that disappeared during the ligand addition were set to 0. Values of the rate of signal decay were adjusted to correct for the presence of protein precipitation. Individual intensities were divided by the average intensity of unaffected residues calculated for each titration point. Normalised values were fitted to a straight line and the gradient was considered as a measure of signal decay (y-axis). An arbitrary threshold of -2 was chosen. All residues associated with CSPs that decreased beyond the threshold had already been identified. Residues K406, N405, K412, M453, V452, Y411, and V461 displayed the biggest chemical changes; followed by F432, R427, I425, T491, N421, L473, T429, Y472, G430, and A416. The SH3\_5 protein (PDB 5LEO) was used to map the residues identified in this titration.

#### **4.3.2.9 SH3\_5 binding to a mixture of glycan chains and peptide stems**

Finally, the last NMR titration was performed using soluble PG fragments resulting from the digestion of *S. aureus* sacculi by an amidase. The digestion products correspond to a mixture of linear glycan strands and peptide stems. 10, 18, 26, and 34  $\mu\text{L}$  of this mixture were used for titration experiments (Fig. 4.14 A). CSPs were characteristic of slow exchange equilibrium (Fig. 4.14 B). Fewer residues exhibited loss of signal beyond our established threshold (N405, Y411, I425, G430, V452, M453, and V461), but precipitation and signal intensities were roughly at the same rate as previously seen in titrations containing a mixture of PG fragments (Fig. 4.14).

Altogether, our NMR analyses revealed a distinct subset of residues interacting with both PG peptide moieties, with the pentaglycine cleft and the proposed peptide stem pocket located on opposite sides of the surface. (Fig. 4.15 A-H).

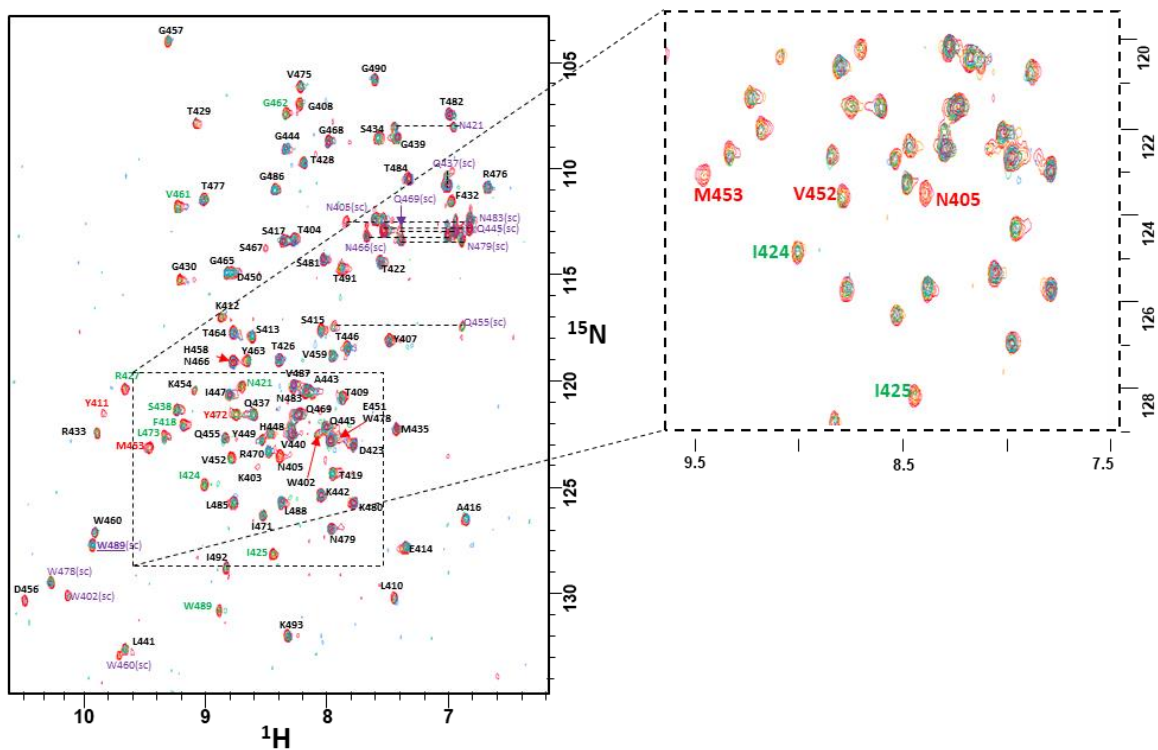
#### **4.3.3 High resolution crystal structure of the SH3\_5 domain in complex with the [P4-G5] fragment**

Our results clearly showed that the G5 and P4 PG moieties interact in two distinct binding sites of the SH3\_5 domain. Around the same time, another research group obtained similar results in a study using NMR titrations with a P4 peptide (Tossavainen, 2018). From these results two proposed models tried to explain the SH3\_5 binding mechanism to a monomeric P4-G5 fragment (Mitkowski, 2019; Tossavainen, 2018).

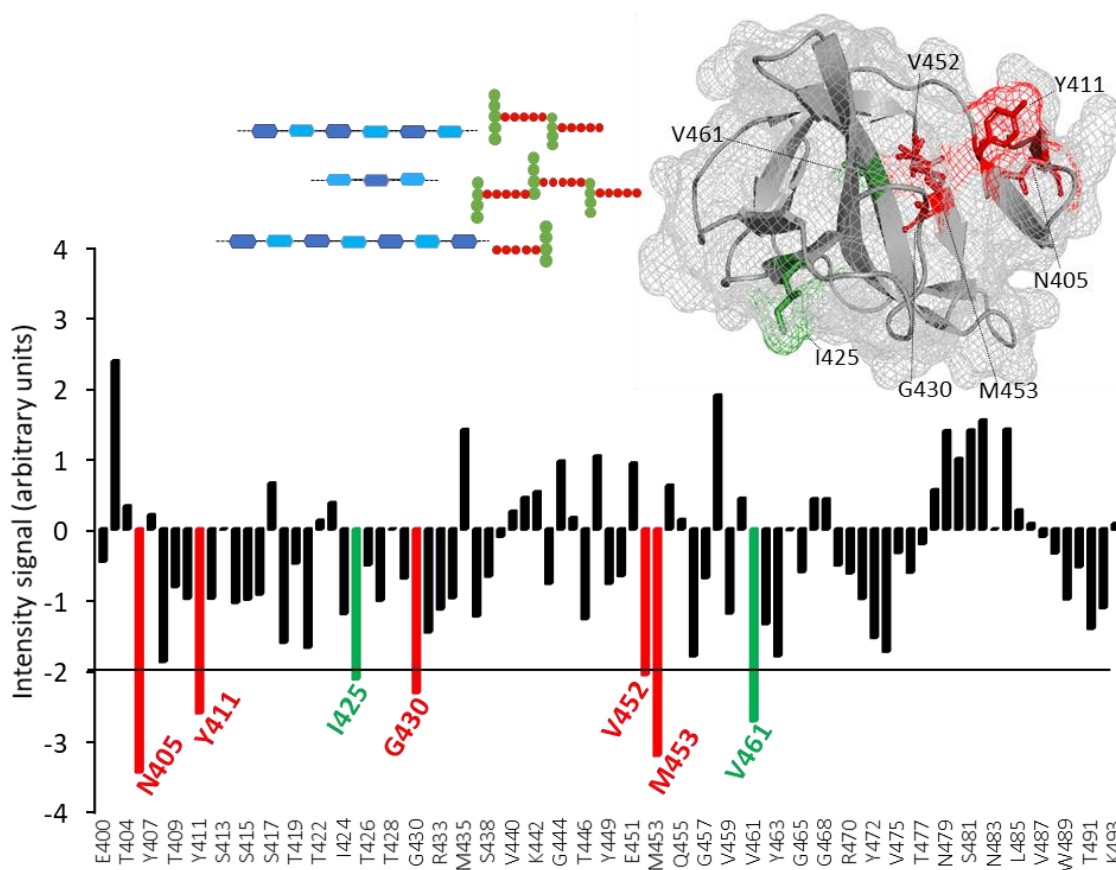
In all the NMR titrations reported above, most of the CSP are associated with residues situated in the G5 cleft or the stem peptide binding site. No residues located between the two sites undergo CSP. We therefore hypothesized that an adjoining crosslinked peptide cannot bind to both sites at the same time, and the interaction must be more complex. Changes in signal intensity also imply that there are dynamic processes occurring on binding. To fully characterise binding to both substrates, the co-crystallisation of the SH3\_5 domain in complex with our P4-G5 ligand was attempted.

For the X-ray crystallography studies, a recombinant untagged SH3\_5 protein was produced and concentrated to 20mg/mL (Fig. 3.6, chapter III). Our collaborators at the University of

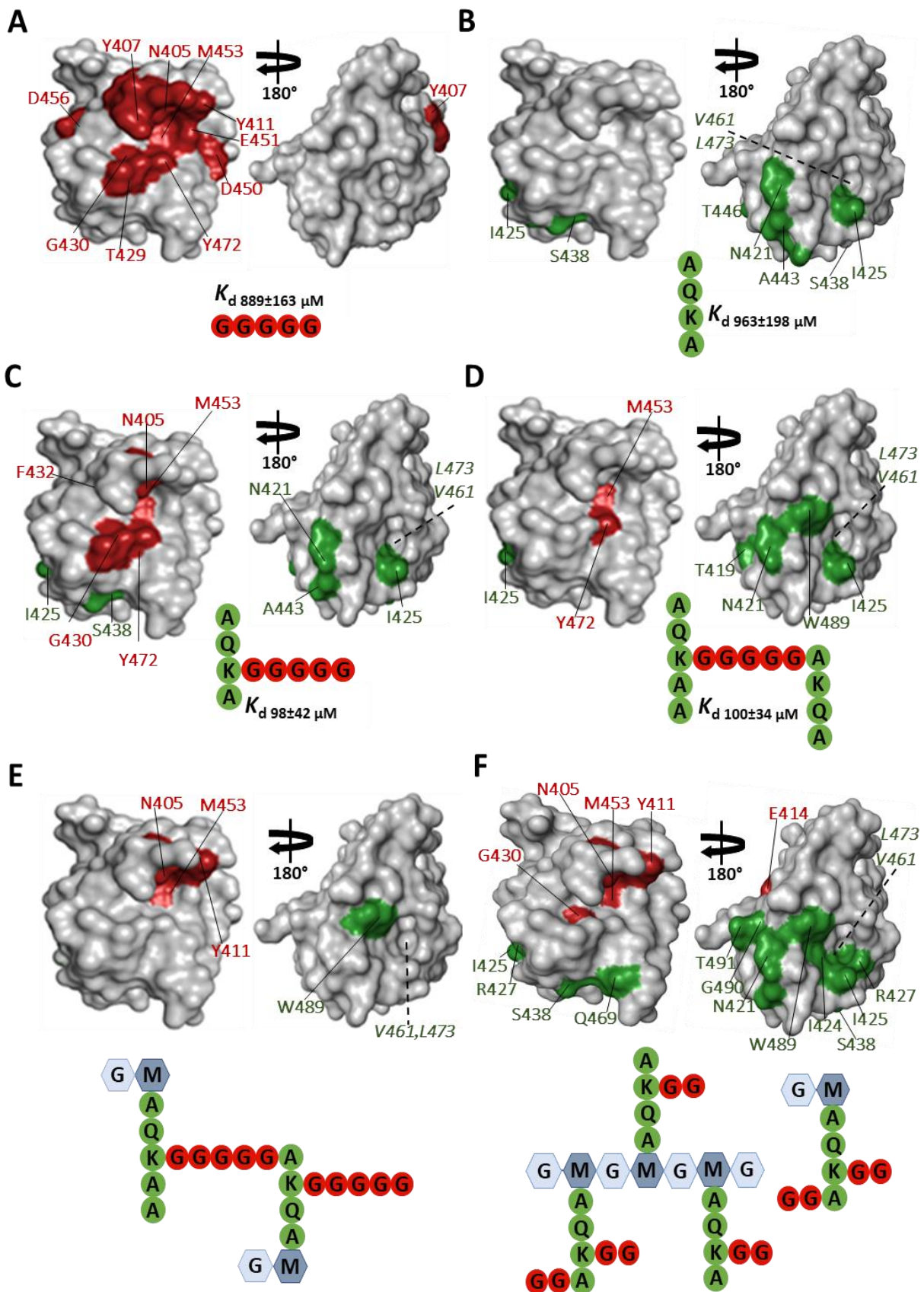
**A**

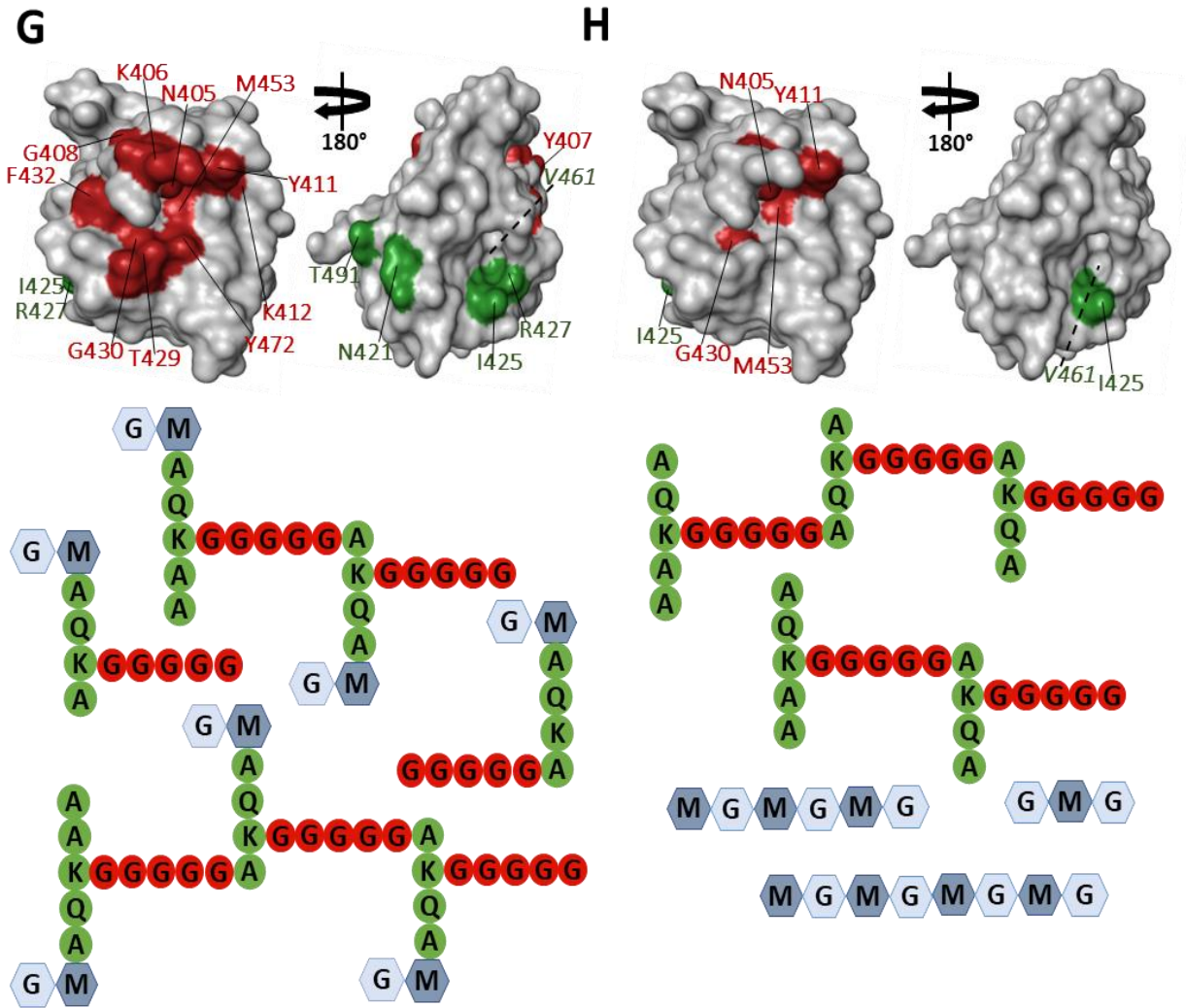


**B**



**Figure 4.14. Mapping of SH3\_5 residues interacting with *S. aureus* linear glycan chains and peptide stems by NMR titrations.** (A) Overlaid  $^1\text{H}^{15}\text{N}$  HSQC spectra of the SH3\_5 interaction with *S. aureus* linear glycan chains and peptide stems derived from amidase digestion. Titrations were acquired in the absence (red contour) and presence (yellow to blue contour) of 10 and 34  $\mu\text{L}$  of the digestion mixture. Peaks from the backbone NH assignment are labelled in black and sidechains in purple. The enlarged view shows a region of the spectra with peaks that disappeared during the titration (absence of the last peak contours). The largest CSP are indicated in green for those residues previously identified to interact with the P4 peptide and in red for G5. (B) Histogram of signal intensities as a function of protein-ligand interaction. The signals that disappeared during the ligand addition were set to 0. Values of the rate of signal decay were adjusted to correct for the presence of protein precipitation. Individual intensities were divided by the average intensity of unaffected residues calculated during every titration point. Normalised values were fitted to a straight line and the gradient was considered as a measure of signal decay ( $y$ -axis). An arbitrary threshold of -2 was chosen. Residues associated with CSPs that decreased beyond the threshold were N405, Y411, I425, G430, V452, M453, and V461. The SH3\_5 protein (PDB 5LEO) was used to map the residues identified in this titration.





**Figure 4.15. Mapping of the SH3\_5 protein surface interacting with distinct *S. aureus* PG fragments. (A-H)** Interaction maps of the two opposite sides of the Lss SH3\_5 protein (PDB 5LEO) with the identified binding sites for the pentaglycine [G5] and tetrapeptide [P4] ligands. **(A)**, G5 peptide; **(B)**, P4; **(C)**, P4-G5; **(D)**, P5-G5-P4; **(E)**, GM-P5-G5-GM-P4-G5; **(F)**, Lss-digested PG fragments; **(G)**, Mutanolysin-digested PG fragments; **(H)**, Amidase-digested PG fragments. Black lines show the location of key residues interacting with G5 (highlighted in red) and those for P4 (green); dotted lines indicate the location of residues buried in the structure.

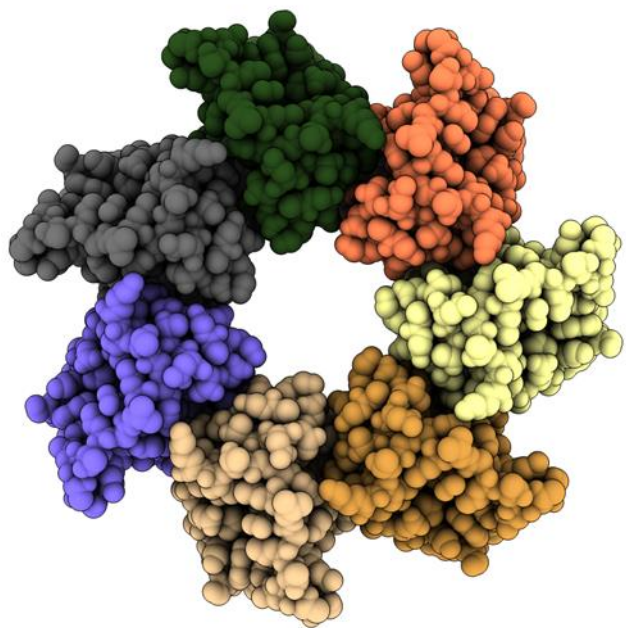
Birmingham, Dr Andrew L. Lovering and Hannah Walters-Morgan carried out the crystallographic experiments and analyses.

To determine the conditions for SH3\_5 crystallisation, standard screen conditions with and without ligand were run in parallel. The JCSG+ screen (Molecular dimensions) was performed using a 2  $\mu$ L drop volume and a 1:1 ratio of screening agent to protein solution in 200  $\mu$ L wells.

Crystals were formed in 0.2 M zinc acetate dihydrate, 0.1 M imidazole pH 8.0 and 20% (w/v) PEG 3000. Four crystals were taken from this condition and three of them were soaked with 20% (v/v) ethylene glycol for 2, 3, or 4 minutes for cryo-protection. One crystal was frozen without any cryo-protection. The cryo-protected crystals diffracted poorly, the non-cryo crystal diffracted to about 2.5 Å, allowing the structure to be solved. Interestingly, the SH3\_5 protein formed a heptamer with the presence of Zn<sup>+</sup> at the interface of monomers (Fig. 4.16). These crystals were optimised with a home screen using similar drops as previously described. Needles formed in 0.1 M imidazole pH 7.75, 30% PEG 3000 and 0.2 M zinc acetate dihydrate. The first approach to generate protein-ligand complex crystals was soaking the P4 and P5-G5-P4 peptides, however, after testing different conditions the diffraction showed that the crystals had no ligand.

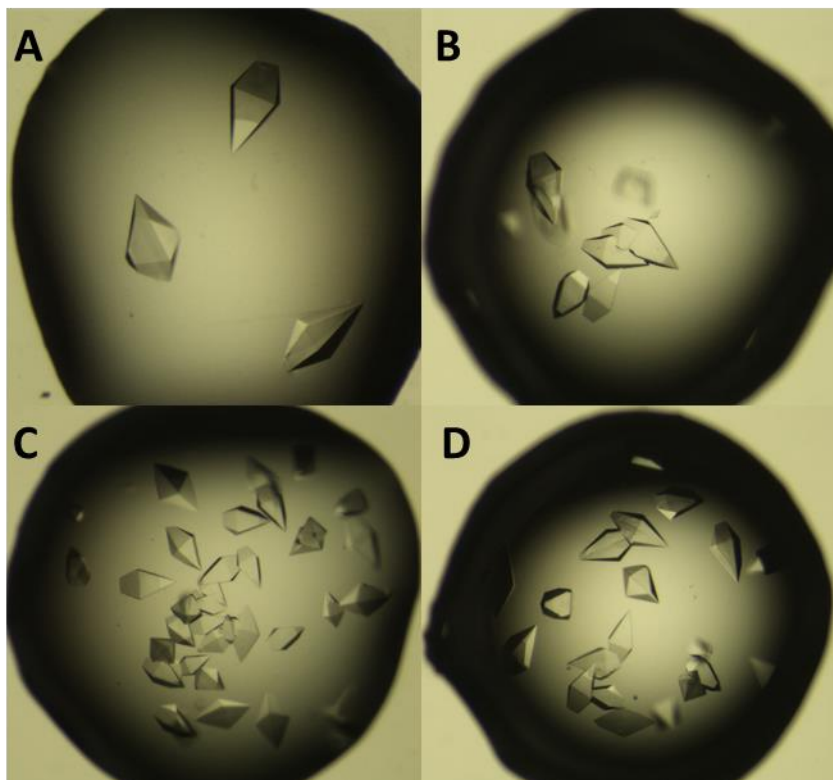
For co-crystallisation experiments, SH3\_5 proteins were crystallised in the presence of a synthetic peptide ([A $\gamma$ QK(GGGGG)A], [**P4-G5**]) a key ligand in our NMR studies. For these experiments, a highly concentrated compound was required. 14 mg of peptide were resuspended in water at a final concentration of 70 mM. Crystallization was initiated via JCSG+ screening in sitting drop 96-well clover-leaf crystallography trays at 18 mg/mL SH3\_5 (1.7 mM) with 3.41 mM [P4-G5] in a 1:2 drop ratio of screening agent to protein solution. Trays were incubated at 18°C. Tetragonal bipyramidal crystals formed within the first 48 h in 100 mM Bis-Tris, pH 5.5, 25% (w/v) poly(ethylene) glycol 3350 and 200 mM ammonium sulphate. Crystals were cryo-protected using the above conditions (including P4-G5 to maintain the ligand–protein complex) and additional 20% (v/v) ethylene glycol (Fig. 4.17).

Two datasets were collected, a high-resolution set at the I03 beamline, Diamond Light Source, Oxford and a second set on a Rigaku Micromax home source (PDB 6RJE). Both datasets were used



**Figure 4.16. 2.5Å crystal structure of the Lss-SH3\_5 domain forming Zn<sup>2+</sup>-bound heptamers.** Each SH3\_5 monomer is coloured separately; Zn<sup>2+</sup> was present at the interface of the monomers. Crystals were formed in JSCG+ crystallisation screening with 0.2 M zinc acetate dihydrate, 0.1 M imidazole pH 8.0 and 20% PEG 3000. Non-cryoprotected crystals diffracted at 2.5 Å. Crystals were optimised with a home screen; all drops were 2 µL: 2 µL protein: well. Needles formed in 0.1 M imidazole pH 7.75, 30% PEG 3000 and 0.2 M zinc-acetate.





**Figure 4.17. Crystallisation of Lss-SH3\_5 proteins in complex with the P4-G5 ligand.** Optimised crystallisation of the SH3\_5-[P4-G5] complex from initial hit used for diffraction. Crystallography trays at 18mg/mL with 3.41 mM P4-G5 in a 1:2 drop ratio of screening agent to protein solution. (A) 25% PEG 3350, 0.1 M Bis tris pH 5.75, 0.2 M Ammonium sulphate (JCSG+). From this tray two datasets, a Rigaku Micromax home source and a high resolution set at the 103 beamline, Diamond Light Source, Oxford were collected; (B) 35 % PEG 3350, 0.1 M Bis tris pH 6.0, 0.2 M Ammonium sulphate (JCSG+) ; (C) 25% PEG 3350, 0.1 M Bis tris pH 5.75, 0.2 M Lithium sulphate (JCSG+); (D) 35 % PEG 3350, 0.1 M Bis tris pH 5.0, 0.2 M Lithium sulphate (JCSG+). All crystals were cryo-protected in 20% ethylene glycol (v/v) for 3 minutes.

to solve the final structure. A high-resolution diffraction pattern at 1.4 Å (PDB 6RK4) was obtained (Table 4.2). All data was processed with Xiall/ XDS40. The B factors for the high-resolution set were higher than expected, but match that of the Wilson B, and the dataset has a normal intensity distribution. An initial model was solved using the existing apo structure (PDB 5LEO, 1.6 Å) as a molecular replacement model in PHASER (McCoy, 2007), and the corresponding structure was auto built using PHENIX (Zwart, 2008), with the ligand added via visual inspection of the difference map. The structure was updated and refined using COOT (Emsley, 2004), PHENIX (Zwart, 2008) and PDB-redo (Joosten, 2011), resulting in a final structure with an R/Rfree of 19.9%/23.0% .

The 1.4 Å Lss SH3\_5 crystal structure in complex with the P4-G5 fragment revealed an unexpected dual binding mechanism. The pentaglycine chain is identically bound to the SH3\_5 protein as the PDB 5LEO structure, but contrary to recently published models (Mitkowski, 2019; Tossavainen, 2018), the bridge-linked stem peptide (K to G5) is directed towards a second protein domain that fits the tetrapeptide stem [P4] in a pocket located at the opposite side to the G5 binding cleft (Fig. 4.18 A). As shown in Figure 4.18 B, the SH3\_5 domains do not interact with one another but are brought in close proximity through interactions with the shared ligand acting as a bridge.

The P4 binding site displays a more open space for L-Ala1,  $\gamma$ Q2, and K3 interactions; the L-Ala 1 location would place the linear saccharide chains at the edge of both domains (Fig. 4.18 C-D). These three peptide units adopt a linear  $\beta$ -like conformation with their peptide bonds hydrogen-bonding to residues from both SH3\_5 domains. Figure 4.18 D shows the bond between  $\gamma$ Q2 NH to the carbonyl of K406, the K3 NH to the carbonyl of D423', and K3 CO to the NH of I425' (prime indicates opposing SH3\_5). The carbon atoms of the K3 sidechain make favourable contacts with the hydrophobic sidechains of Y407, T422', I424' and W489'.

Finally, the D-Ala 4 unit sits in a defined hydrophobic pocket made of I424', I425', R433', H458', W460', P474', and W489' which encloses the methyl sidechain of the ligand; with the COO<sup>-</sup> forming a salt bridge to the R427' sidechain. The conformational arrangement adopted by the P4

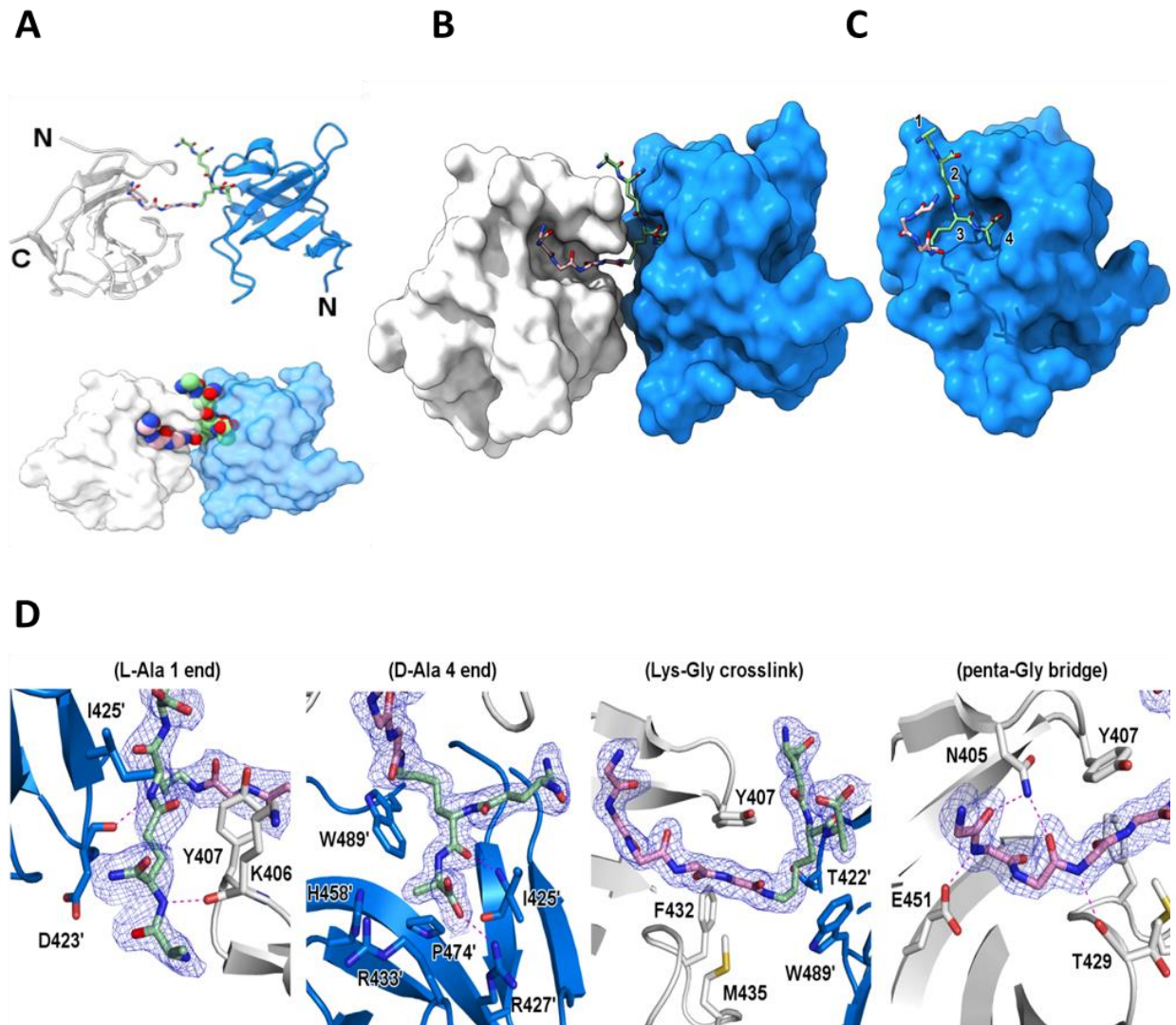
chain generates stronger contacts to the third and fourth units of the stem peptide. The G5 chain adopts an extended conformation and sits into the pentaglycine cleft in a more rigid lock-and key type of binding. As previously reported, the walls of the cleft are formed of the sidechains of residues N405, T429, G430, F432, E451, M453, and Y472, between strands  $\beta$ 1,  $\beta$ 2, and the RT loop.

**Table 4.2 Data collection and refinement statistics of the SH3\_5: P4-G5 co-crystal (PDB 6RK4)**

	Home source set	Synchrotron se
<b>Data collection</b>		
Space group	P4 <sub>1</sub> 2 <sub>1</sub> 2	P4 <sub>1</sub> 2 <sub>1</sub> 2
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	47.2, 47.2, 123.1	47.1, 47.1, 122.4
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	2.5(2.6-2.5)*	1.43 (1.47-1.43)
R <sub>merge</sub>	12.8(21.5)	5.9 (341.3)
<i>I</i> / $\sigma$ <i>I</i>	14.3(6.0)	21.6(1.1)
Completeness (%)	99.6(98.7)	98.9(98.4)
Redundancy	11.9(8.9)	24.3(24.7)
<b>Refinement</b>		
<b>Resolution</b>		
(Å)	2.5	1.43
No. reflections	5270	26342
R <sub>work</sub> / R <sub>free</sub>	25.4/29.0	19.9/23.0
<b>No. atoms</b>		
Protein	741	741
Ligand/ion	49	52
Water	3	24
<b>R.m.s. deviations</b>		
Protein	27.7	43.4
Ligand/ion	29.7	44.2
Water	17.9	44.3
<b>R.m.s. deviations</b>		
Bond lengths (Å)	0.003	0.008
Bond angles (°)	1.18	1.49

A single crystal was used for both data collections.

\*Values in parentheses are for highest-resolution shell



**Figure 4.18. Crystal structure of Lss-SH3\_5 in complex with the P4-G5 ligand.** The SH3\_5 domain and a symmetry-related partner are coloured white and blue, respectively; ligands are coloured by atom type, with P4 C atoms green and G5 C atoms pink. **(A)**, cartoon protein fold with termini labelled; proteins shown in two representations (ribbon/stick and surface formats). **(B)**, Co-crystal structure showing the SH3\_5:ligand shape complementarity. **(C)**, Rotated view of a single SH3\_5 domain showing the interaction with the P4 ligand; peptide units labelled 1-4: L-A1- $\gamma$ Q2-K3-A4. **(D)**, experimental 2fo-fc electron density map contoured at  $1\sigma$  for different regions of the bound ligand, with selected interacting residues in stick form. The L-Ala 1 end panel is from the 2.5 Å dataset, others are from the 1.4 Å form. Residues from the symmetry-related SH3\_5 domain are denoted by use of a prime ('), and hydrogen bonds represented as a dashed line.

#### 4.4 Discussion

NMR and X-ray crystallography are two of the major techniques to analyse protein structure and function. The combined use of these methods allowed the identification of two distinct binding sites for *S. aureus* PG. The co-crystal structure determined in this study corroborated our NMR observations revealing an unexpectedly complex binding mechanism. The pentaglycine binding was found to rely on the very tight positioning of the glycine residues within a narrow cleft previously described (Lu JZ, 2006; Mitkowski, 2019), and at the opposite face, a more open site for the peptide stem binding. Previously the Lss-SH3\_5 domain had been thought to confer the substrate specificity of the enzyme by its G5 recognition, but our NMR results showed that the SH3\_5 domain adopted a dynamic binding mechanism by interacting with two distinct PG substrates.

Individually, G5 and P4 display weak binding affinities with some large CSPs clearly defining the location of each site, whilst more complex ligands comprised of both G5 and P4 moieties displayed higher affinities but lower CSPs. Although the affinities to more complex ligands are approximately 10x higher, this suggests a cooperative but suboptimal binding due to the distance between both binding pockets preventing a simultaneous binding of both peptide moieties from the same molecule. The CSPs from the increasingly complex ligands highlighted a dynamic binding mechanism rather than an overall weaker binding, as could be interpreted from the small CSPs observed.

The analysis and interpretation of the CSPs, and of the affinities determined by fitting CSPs to a saturation curve exhibited the intricacy of individual molecules interacting together to form active complexes. NMR exploits the inherent property of some nuclei that exist in specific nuclear spin states when exposed to an external magnetic field. The electronic environment of each nucleus slightly modifies its exact resonance frequency through a process called chemical shielding. The local chemical environment surrounding a given nucleus determines the extent of shielding of that nucleus, causing different nuclei to resonate at slightly different frequencies. A perturbation of an  $^{15}\text{N}$  or  $^1\text{H}$  nuclear shielding is typically caused by a change in the magnetic field

at the nucleus; for example, a change in hydrogen bonding or a change in the position of neighbouring functional groups. Since hydrogen bonds are highly directional, a dynamic interaction with increased mobility of complex ligands could result in smaller CSPs. As mentioned before, this does not necessarily imply weaker binding as observed in the P4-G5 and P5-G5-P4 peptides, since the loss of enthalpy caused by spending less average-time hydrogen bonding can be compensated by the gain of entropy. Furthermore, titrations with complex mixed PG fragments were in the slow exchange limits which implies tighter interactions along with loss of signal due to protein aggregation.

SH3\_5 domains are structurally related to the eukaryotic SH3 domains. SH3 domains are involved in a wide variety of biological processes such as increasing local concentration of proteins, altering their subcellular location, and mediating the assembly of large multiprotein complexes (Mayer, 2001). The solved crystal structure of the SH3\_5 in complex with the P4-G5 peptide is consistent with the formation of large protein aggregates observed during titration with complex PG fragments. Recognition of PG peptide stems by independent SH3\_5 domains leads to the formation of “clusters” that would effectively increase the enzyme concentration at the cell surface.

## CHAPTER V

### Mutational analyses on the binding activity of lysostaphin SH3\_5 domains

Previous studies have proposed that the SH3\_5 domain of Lss is responsible for the high specificity of the enzyme to staphylococcal PG by guiding it to pentaglycine interpeptide bridges, the target substrate of the catalytic domain. The NMR titrations and X-crystallography results described in Chapter IV led to the identification of two independent binding sites located on opposite sides of the SH3\_5 domain. Moreover, the interpretations of the CSP perturbations of the most complex defined PG fragments exhibited a larger number of residues interacting with the P4 ligand compared to the G5 peptide.

#### 5.1 Aims and objectives

Site-directed mutagenesis was used to explore the individual contribution of SH3\_5 residues to the binding activity of the Lss-SH3\_5 domain. The specific objectives were:

- To confirm the role of chosen SH3\_5 residues to the recognition of pentaglycine and stem peptides using NMR titrations.
- To study the contribution of several residues to the binding of SH3\_5-mNeonGreen fusions to PG sacculi.
- To investigate the impact of SH3\_5 mutations on the activity of the full-length mature Lss enzyme.



## 5.2 Results

### 5.2.1 Expression and purification of SH3\_5 and full-length Lss mutants

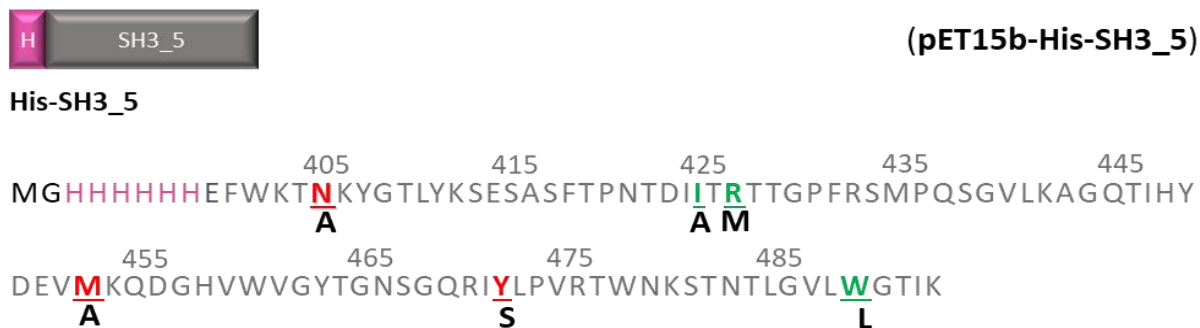
#### 5.2.1.1 Singly labelled His-SH3\_5 derivatives

The contributions of six residues identified by NMR and X-ray crystallography to the binding of G5 and P4 peptides were analysed by  $^{15}\text{N}$  HSQC titrations. The pET15b-His-SH3\_5 vector, used to produce the singly labelled WT SH3\_5 domain, was mutagenized to generate three recombinant proteins harbouring single-site substitutions in residues associated with the G5 peptide (N405A, M453A, and Y472S), and three SH3\_5 mutants harbouring substitutions in residues interacting with P4 peptide (I425A, R427M, and W489L) (Figure 5.1).

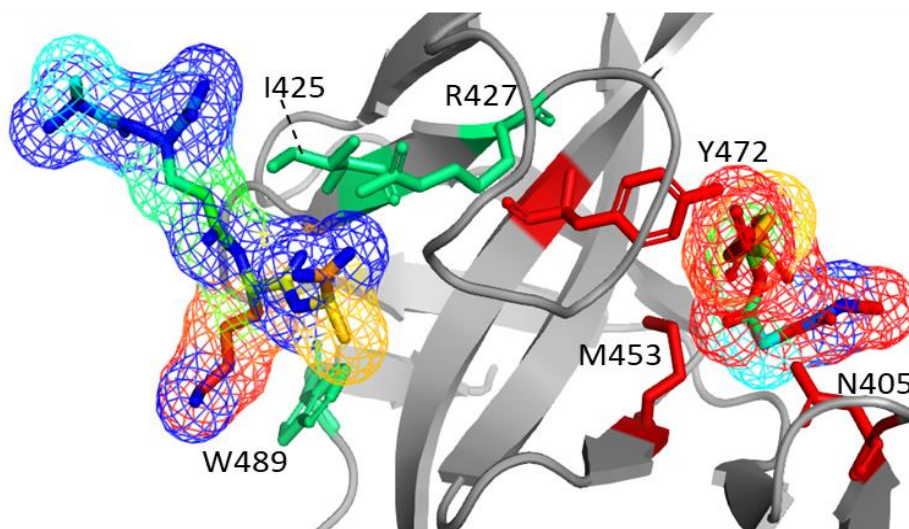
The aromatic ring current effects, observed in amino acids with aromatic side chains, can considerably influence chemical shifts in molecules. Ring current effects are largely dominant and most of the direct effects on protein chemical shifts from ligand titrations are due to either ring current effects or, for those amino acids without an aromatic ring, from hydrogen bonding to amide protons. The largest HN shifts generally arise from hydrogen bonding interactions. In fact, chemical shift effects arising from hydrogen through-space interactions are just as important as through-bond interactions. Thus, the reasoning behind the selected amino acid substitutions for all the amino acid residues investigated in this research, was to disrupt distances to alter ring current effects and hydrogen-bonding, through non-conservative amino acid substitutions of the side chains with different physicochemical properties.

The six His-SH3\_5 constructs were expressed as singly  $^{15}\text{N}$ -labelled His-SH3\_5 mutant domains. The 11.7 kDa proteins were purified by affinity and gel filtration chromatography. A typical purification corresponding to the mutant R427M is shown as an example in Figure 5.2 A. The purity of the produced proteins was greater than 90% (SDS-PAGE) (Fig. 5.2 B). The obtained yield of the  $^{15}\text{N}$  His-SH3\_5 proteins was  $\sim 10\text{mg/L}$ .

**A**

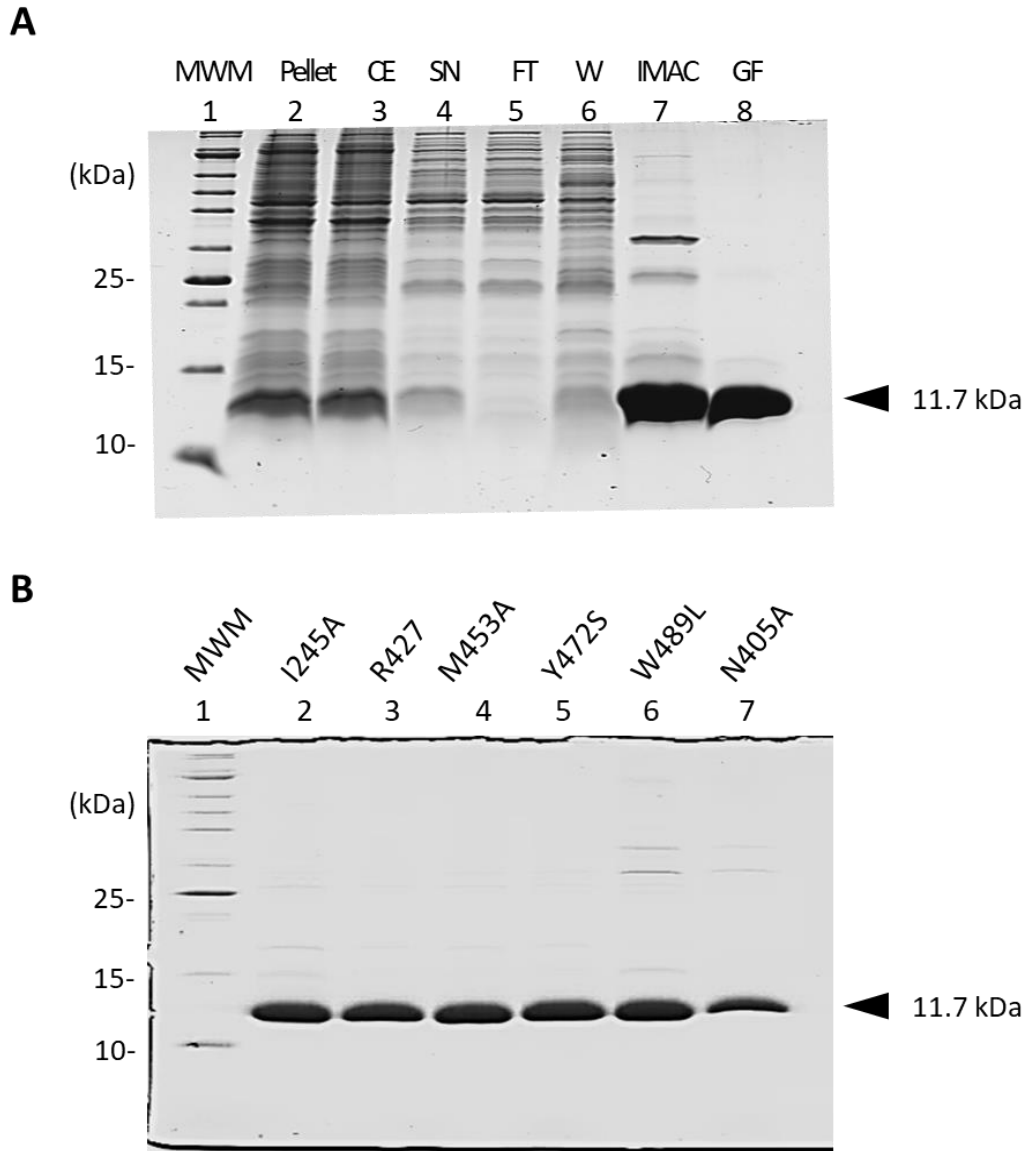


**B**



**Figure 5.1. Description of six SH3\_5 single-site substitution mutant domains analysed by NMR.**

**(A)** Schematic representation of the domain organization and amino acid sequence of the recombinant His-SH3\_5 protein encoded by the pET15b-His-SH3\_5 vector used for site-directed mutagenesis. Amino acids corresponding to the N-terminal 6x His-tag are coloured in pink; Lss-SH3\_5 domain in grey; mutagenized amino acids interacting with the G5 ligand are underlined and coloured in red, those interacting with the P4 peptide are coloured in green. The residues that were introduced by site-directed mutagenesis in constructing the SH3\_5 derivatives are shown below the WT SH3\_5 sequence **(B)** Region of the bound ligands from the Lss SH3\_5 co-crystal structure displaying the six residues selected for mutation in stick form and coloured according to their G5 (red) or P4 (green) peptide recognition. The G5 and P4 ligand mesh surfaces are coloured by atom type.



**Figure 5.2. SDS-PAGE analysis of singly labelled SH3\_5 derivatives used for NMR studies. (A)** SDS-PAGE analysis of the purification steps to obtain the 11.7 kDa singly labelled R427M His-SH3\_5 mutant protein expressed in *E. coli* Lemo21 (DE3) cells. Lane 1 (MWM), molecular weight makers; lane 2, pellet ; lane 3 (CE), crude extract; lane 4 (SN), soluble fraction; lane 5 (FT), flow through; lane 6 (W), wash; lane 7 (IMAC), elution pool; lane 8 (GF), pooled fractions resulting from the gel filtration step. **(B)** SDS-PAGE of the six 11.7 kDa singly labelled His-SH3\_5 derivative proteins expressed in *E. coli* Lemo21 (DE3) cells. Lane 1 (MWM), molecular weight makers; lane 2-7, pooled fractions resulting from the gel filtration step to obtain the I425A, R427M, M453A, Y472S, W489L, and N405A His-SH3\_5 mutant proteins.

## 5.2.1.2 SH3\_5-mNeonGreen fusion proteins

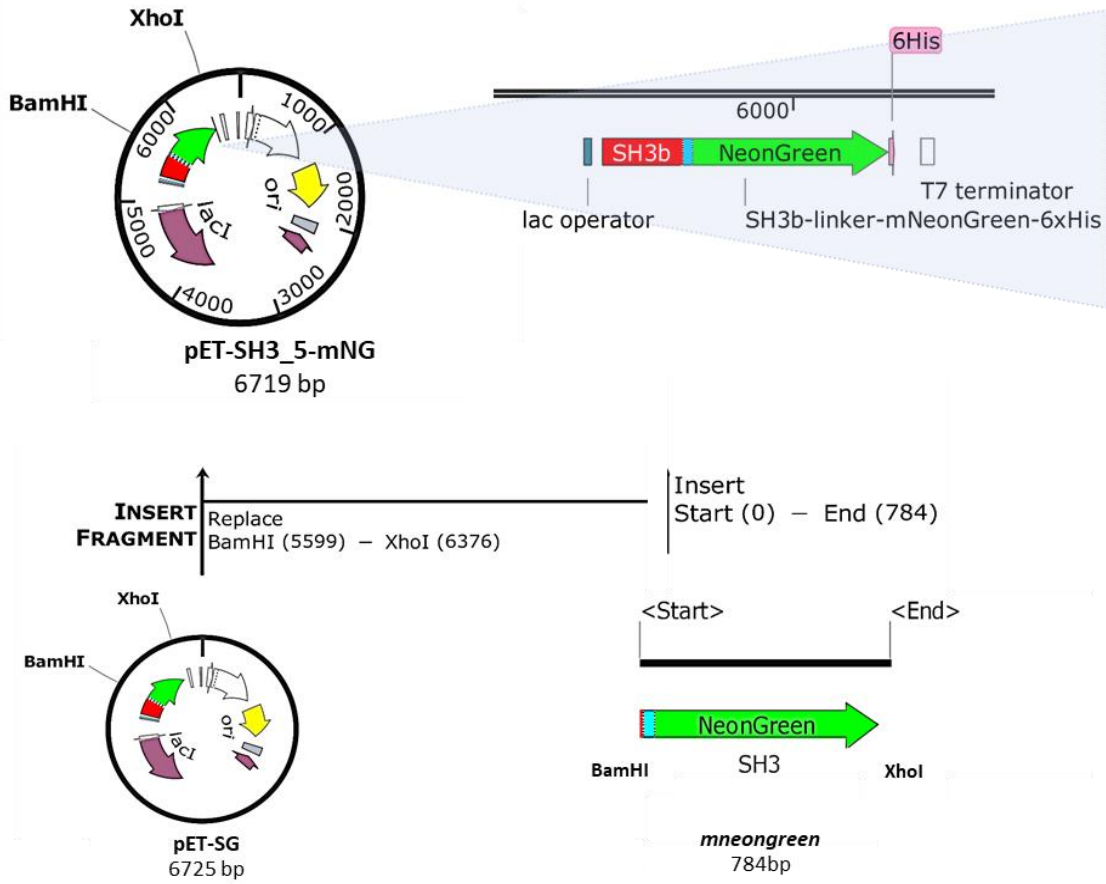
### 5.2.1.2.1 Construction of pET2818-SH3\_5-NG expressing a SH3\_5-mNeonGreen fusion

A quantitative *in vitro* binding assay was designed to investigate the impact of SH3\_5 mutations on binding to PG sacculi, which represents the natural substrate of the Lss enzyme. The SH3\_5 domain was fused to the fluorescent protein mNeonGreen. The mNeonGreen fluorescent protein, described by (Shaner, 2013), is a monomeric protein brighter than any other monomeric fluorescent protein (green and yellow). It has become a better alternative to GFP and eGFP by being systematically brighter by 5 to 3 times, respectively. This has also been shown when tested for *in vivo* imaging in *C. elegans* tissues (Hostettler, 2017).

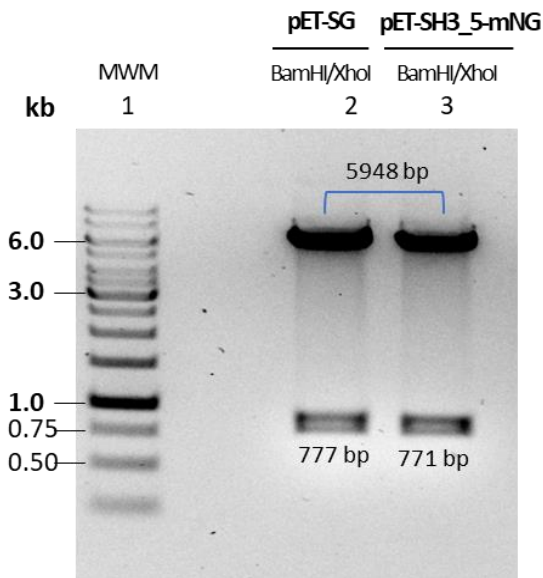
The plasmid encoding the WT SH3\_5-mNeonGreen fusion was constructed by Gibson assembly using a 784 bp DNA synthetic fragment flanked with BamHI and XhoI sites (Fig. 5.3 A). It encoded the last C-terminal residues of the SH3\_5 domain, a 14-residue linker, followed by the mNeonGreen fluorescent protein and a C-terminal 6x-His tag. The synthetic DNA fragment was digested with BamHI and XhoI and cloned into a similarly digested pET2818-SH3\_5-GFP (pET-SG). The pET-SG plasmid was constructed during my first year of PhD using a similar strategy. Recombinant plasmids were screened by PCR and putative positive clones were sequenced. The final 6.7 kb pET-SH3\_5-mNG plasmid containing the 784 bp insert was digested with BamHI and XhoI and the expected 771 bp mNeonGreen fragment was obtained (Fig. 5.3 B). Since the previous pET-SH3\_5-GFP plasmid had a similar digestion pattern, a second plasmid restriction analysis with KpnI and BspHI was also performed to confirm the replacement of the desired fragment (Fig. 5.3 C).

The 39.1 kDa WT SH3\_5-mNeonGreen fusion was produced and purified by affinity chromatography followed by gel filtration (Fig. 5.4). The purity of the protein estimated on SDS-PAGE was >90%, with a yield of 12mg/mL.

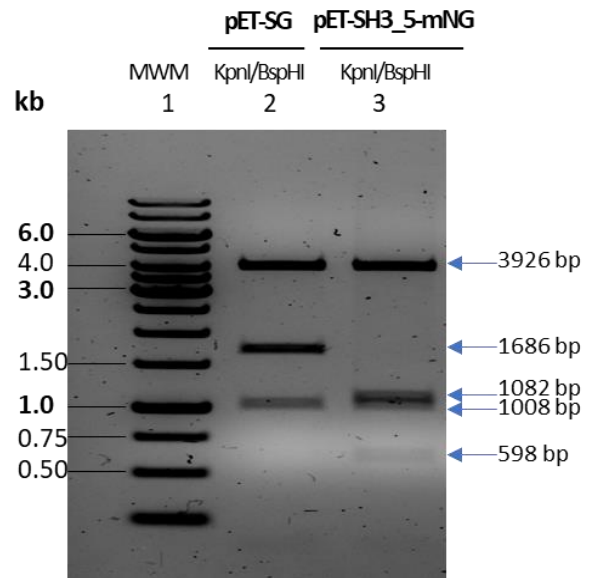
**A**



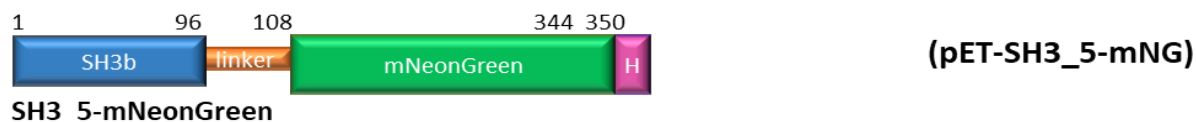
**B**



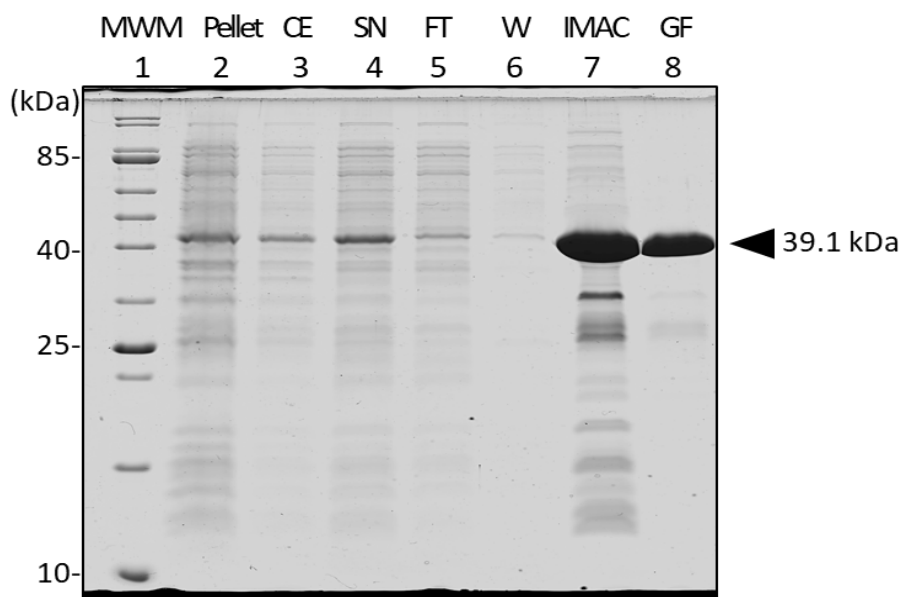
**C**



**Figure 5.3. Construction of pET-SH3\_5-mNG fusion. (A)** Schematic representation of the cloning strategy. A 784 bp DNA synthetic fragment encoding the last C-terminal 10 bp of the SH3\_5 domain, a 36 bp linker, and the 708 bp of the mNeonGreen fluorescent protein, flanked with BamHI and XhoI sites, was cloned by Gibson assembly into the similarly digested 6.7 kb pET-SG plasmid. **(B)** Comparison of the restriction profile from pET-SG and pET-SH3\_5-mNG after digestion with BamHI and XhoI. (MW), molecular weight marker; lane 2 (pET-SG), doubly digested plasmid with BamHI/XhoI, showing the 777 bp liberated fragment encoding the GFP-linker-SH3\_5 fusion and the 5948 bp plasmid; lane 3 (pET-SH3\_5m-NEG), doubly digested plasmid with BamHI/XhoI, showing the 771 bp liberated fragment encoding the mNeonGreen-linker-SH3\_5 fusion and the 5948 bp plasmid. **(C)** Restriction profile confirmation of the mNeonGreen replacement. (MW), molecular weight marker; lane 2 (pET-SG), doubly digested plasmid with KpnI/BspHI, showing the expected DNA bands sizes of 3926 bp, 1686 bp, and 1008 bp; lane 3 (pET-SH3\_5-mNG), doubly digested plasmid with KpnI/BspHI, showing the predicted excised DNA bands of 3926 bp, 1082 bp, 1008 bp, and 598 bp. The expected sizes of the liberated DNA fragments are indicated with blue arrows on the right hand of the gel.

**A**

MGWKTNKYGTLYKSESASFTPNTDIITRTTGPFRSMPQSGVLKAGQTIHYDEVKQDGHVWVGYTGNSGQ  
 RIYLPVRTWNKSTNTLGVWGTIKGSGGSGSGSNNSGMVSKGEEDNMA SLPATHELHIFGSINGVDFDMV  
 GQGTGNPNDGYEELNLKSTKGD LQFSPWILVPHIGYGFHQYLPYPDGMSPFQAAMVDGSGYQVHRTMQFE  
 DGASLTVNYRYTYEGSHIKGEAQVKGTGFPADGPMVMTNSLTAADWCRSKKTYPNDKTIISTFKWSYTTGNGKR  
 YRSTARTTYTFAKPMAANYLKNQPMYVFRKTELKHSKTELNFKEWQKAFTDVMGMDELYKHHHHH

**B**

**Figure 5.4. Production and purification of the SH3\_5mNeonGreen fusion. (A)** Domain organization and amino acid sequence of the recombinant SH3\_5-mNeonGreen fusion protein encoded by the pET-SH3\_5-mNG. Numbering corresponds to the amino acid positions. Amino acids corresponding to the Lss-SH3\_5 domain are coloured in blue (94aa); linker in orange (14 aa); mNeonGreen protein in green (236aa); C-terminal 6x His-tag in pink. **(B)** SDS-PAGE analysis of the purification steps to obtain the 39.1 kDa SH3\_5-mNeonGreen fusion expressed in *E. coli* Lemo21 (DE3) cells. Lane 1 (MWM), molecular weight makers; lane 2, pellet ; lane 3 (CE), crude extract; lane 4 (SN), soluble fraction; lane 5 (FT), flow through; lane 6 (W), wash; lane 7 (IMAC), elution pool; lane 8 (GF), pooled fractions resulting from the gel filtration step.

#### 5.2.1.2.2 Construction of pET2818-mNG expressing the mNeonGreen protein

To produce the mNeonGreen control, the pET2818-mNG plasmid was constructed by Gibson assembly (Fig. 5.5). A 774 bp DNA fragment encoding the mNeonGreen protein was PCR amplified from the pET2818-SH3\_5-NG plasmid using the oligonucleotides mNG-FWD and mNG-RVS, introducing NcoI and BamHI flanking restriction sites and a sequence encoding 6 histidine residues (material and methods). The PCR product was cloned into a similarly digested pET2818. The 6.4 kb pET2818-mNG plasmid containing the *mNeonGreen* insert was digested with BamHI and NcoI and the expected 710 bp *mNeonGreen* fragment was obtained. The final pET2818-mNG plasmid was also sent for sequencing to verify the absence of mutations in the insert.

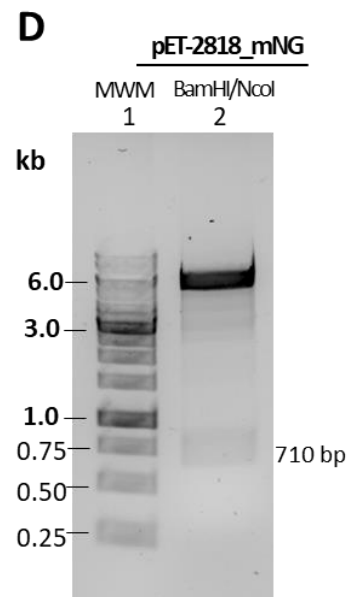
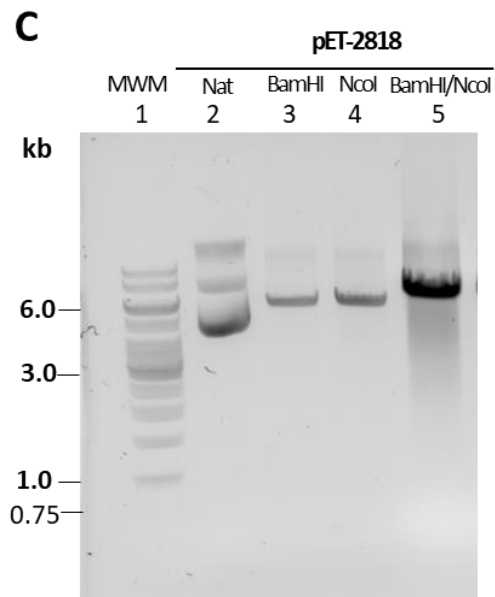
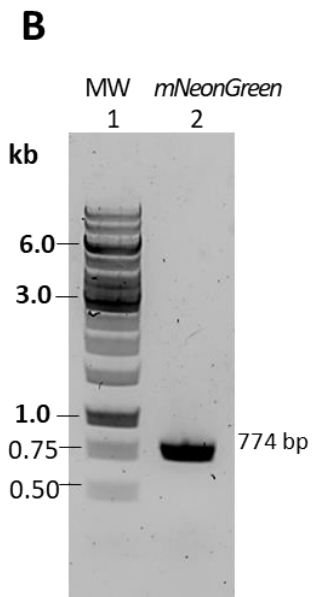
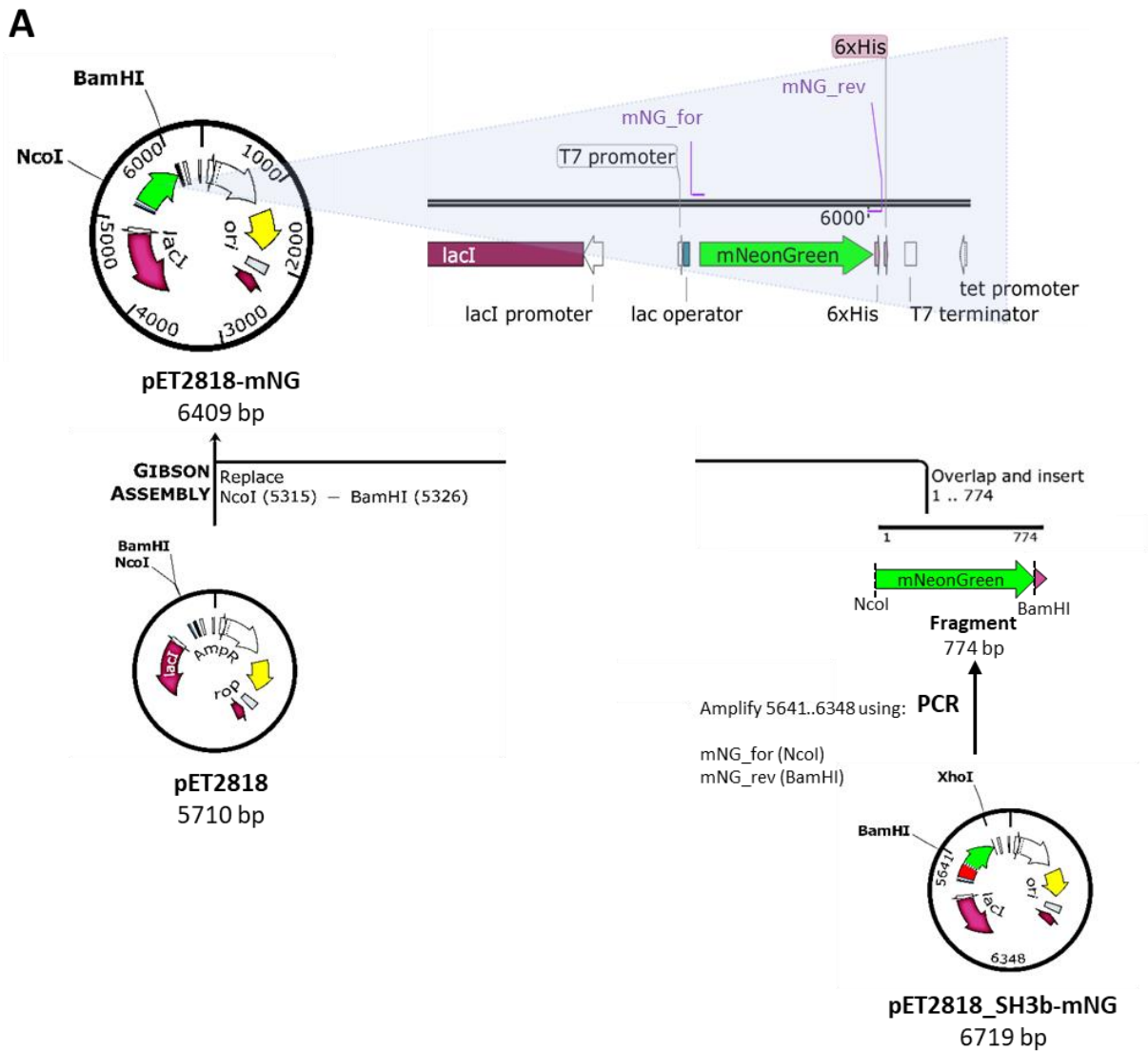
The 27.6 kDa mNeonGreen protein was produced and purified by affinity chromatography followed by gel filtration (Fig. 5.6). The purity of the protein estimated on SDS-PAGE was >90%, with a yield of 20mg/mL.

#### 5.2.1.2.3 Single-site mutagenesis of sixteen SH3\_5-NG fusion proteins

The pET2818-SH3\_5-NG vector was mutagenized to produce 16 SH3\_5-mNeonGreen fusions harbouring single-site substitutions in residues identified to interact with pentaglycine crossbridges and the peptide stems. Figure 5.7 shows a representation of two SH3\_5 domains displaying the location of the mutagenized residues interacting with G5 peptide (N405A, T409V, Y411S, T429V, D450N, E451M, V452A, M453A, and Y472S), and the P4 peptide (F418V, N421L, I425A, R427M, V461A, L473A, and W489L). The oligonucleotides used to introduce the single-site substitutions are described in the Materials and Methods section (Table 2.4). Two putative clones per mutant were sent for sequencing to verify for the presence of the single-site substitution.

The sixteen 39.1 kDa SH3\_5-mNeonGreen fusions were produced and purified by affinity and gel filtration chromatography. Figure 5.8 shows the final purification step of the 18 fusion proteins produced for the quantitative binding analyses (mNG, WT and 16 mutants). Despite the addition of protease inhibitors during the purification process, monomeric mNG protein was detected in all fusion protein purifications, indicating a cleavage of the full-length protein. The amount of

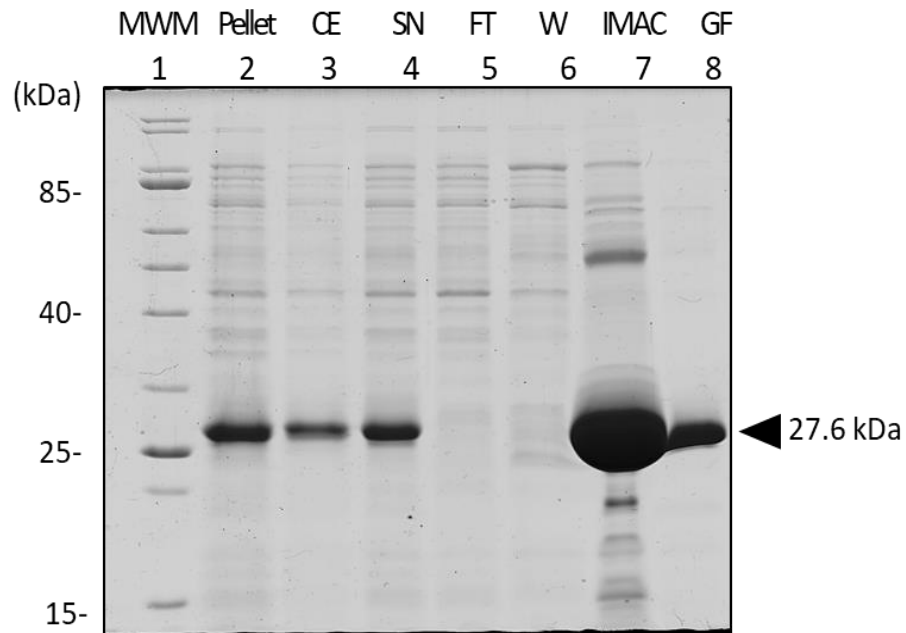




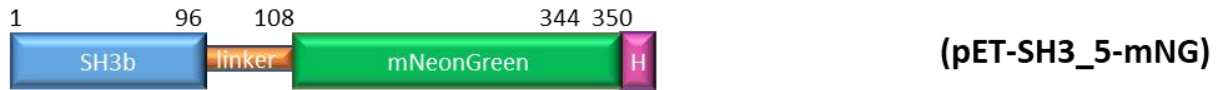
**Figure 5.5. Construction of pET-2818-mNG. (A)** Schematic representation of the cloning strategy. A 774 bp DNA fragment encoding the mNeonGreen fluorescent protein, flanked with BamHI and NcoI sites was PCR amplified from the pET2818-SH3\_5-mNG plasmid. The amplified fragment was cloned by Gibson assembly into the 5.7 kb pET-2818 plasmid which was similarly digested. **(B)** PCR amplified *mNeonGreen* fragment. (MW), molecular weight marker; lane 2, 774 bp *mNeonGreen* DNA sequence. **(C)** Restriction digestion of pET2818 with BamHI and NcoI. (MW), molecular weight marker; lane 2 (Nat), undigested 5710 bp pET2818 plasmid; lane 3 (BamHI) 5710 bp singly digested pET-2818 plasmid with BamHI restriction enzyme; lane 4 (NcoI), 5710 bp singly digested pET-2818 plasmid with NcoI restriction enzyme; lane 5 (BamHI/NcoI), doubly digested plasmid with BamHI/NcoI restriction enzymes. **(D)** Restriction profile confirmation of pET2818-mNG. (MW), molecular weight marker; lane 2 (pET-2818-mNG), doubly digested plasmid with BamHI/NcoI, showing the expected excised DNA bands of 5699 bp plasmid and 710 bp *mNeonGreen* fragment.

**A****(pET2818-mNG)**

MVSKGEEDNMA<sup>1</sup>SLPATHELHIFGSINGVDFDMVGQGTGNPN<sup>236</sup>DGYEELNLKSTKGDLQFSPWILVPHIGYGFH  
 QYLPYPDGMSPFQAAMVDGSGYQVHRTMQFEDGASLTVNRYTYEGSHIKGEAQVKGTFPADGPVMTN  
 SLTAADWCRSKKTYPNDKTIISTFKWSYTTGNGKRYRSTARTTYTFAKPMAANYLKNQPMYVFRKTELKHSKT  
 ELNFKEWQKAFTDVMGMDELYKGS<sup>244</sup>HHHHHH

**B**

**Figure 5.6. Production and purification of mNeonGreen fluorescent protein. (A)** Domain organization and amino acid sequence of the recombinant mNeonGreen protein encoded by the pET2818-mNG. Numbering corresponds to the transition between domains. Amino acids corresponding to the mNeonGreen protein are coloured in green (236aa); N-terminal 6x His-tag in pink. **(B)** SDS-PAGE analysis of the purification steps to purify the 27.6 kDa mNeonGreen protein expressed in *E. coli* Lemo21 (DE3) cells. Lane 1 (MWM), molecular weight makers; lane 2, pellet ; lane 3 (CE), crude extract; lane 4 (SN), soluble fraction; lane 5 (FT), flow through; lane 6 (W), wash; lane 7 (IMAC), elution pool; lane 8 (GF), pooled fractions resulting from the gel filtration step.

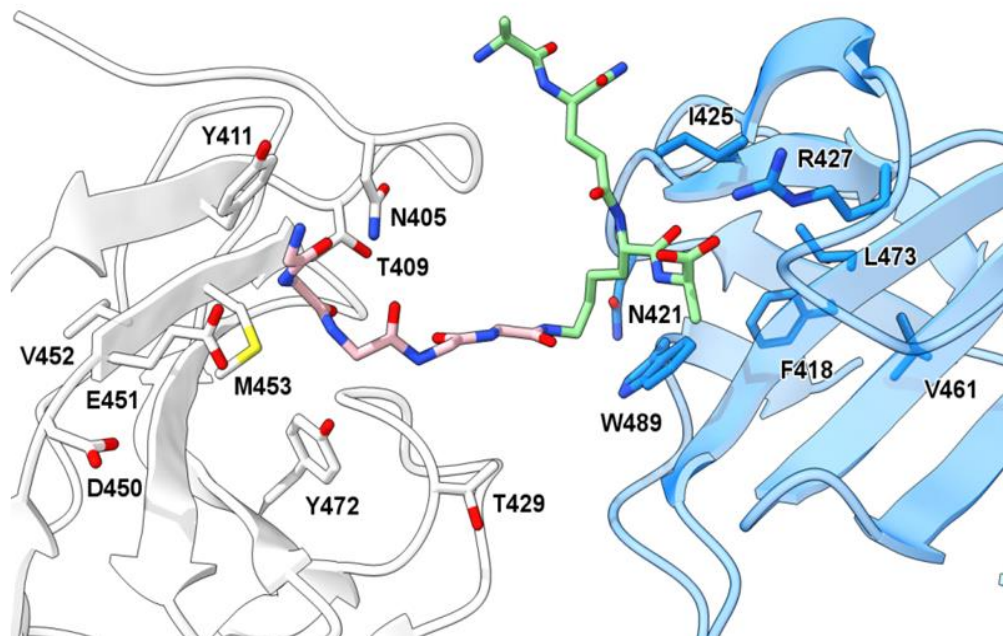
**A**

SH3\_5-mNeonGreen

405 415 425 435 445 455 465  
 MGWKTNKYGTLYKSESASFTPNTDIIRTTGPFRSMPQSGVLKAGQTIHYDEVMKQDGHVVGYTGNS  
 A V S V L A M V N M A A A

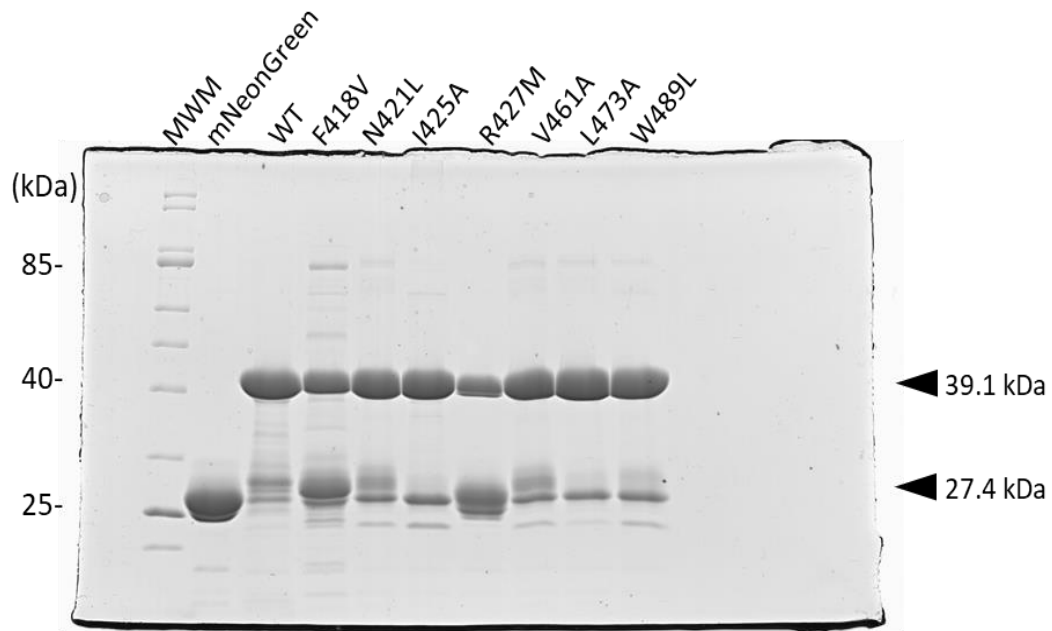
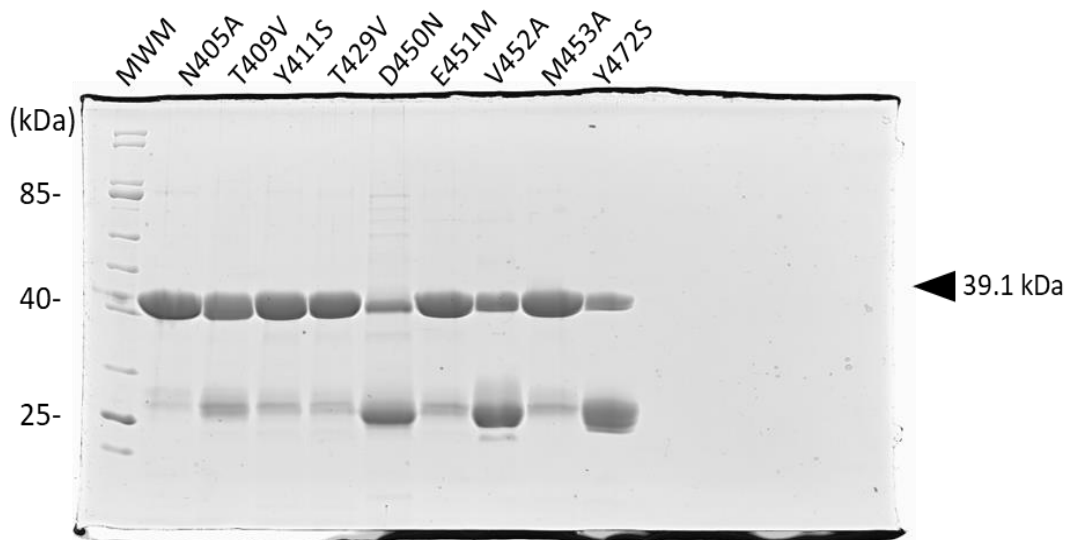
475 485  
 GQRIYLPVRTWNKSTNTLGVLWGTIKGSGGSGSGSNSGMVSKGEEDNMASLPATHELHIFGSINGVDFD  
 SA L

MVGQGTGNPNDGYEELNLKSTKGDLQFSPWILVPHIGYGFHQYLPYPDGMSPFQAAMVDGSGYQVHRTM  
 QFEDGASLTVNRYRYTEGSHIKGEAQVKGTFPADGPVMTNSLTAADWCRSKKTYPNDKTIISTFKWSYTTGN  
 GKRYRSTARTTYTFAKPMAANYLNQPMYVFRKTELKHSKTELNFKEWQKAFTDVMGMDELYKHHHHH

**B**

**Figure 5.7. Location of the sixteen mutagenized residues at the SH3\_5:P4-G5 interaction interface. (A)** Domain organization and amino acid sequence of the recombinant SH3\_5-mNeonGreen fusion protein encoded by the pET-SH3\_5-mNG used for site-directed mutagenesis. Numbering corresponds to the amino acid positions. Amino acids corresponding to the Lss-SH3\_5

domain are coloured in blue (94aa); linker in orange (14 aa); mNeonGreen protein in green (236aa); C-terminal 6x His-tag in pink. SH3\_5 mutagenized amino acids are underlined; the residues that were introduced by site-directed mutagenesis in constructing the SH3\_5-mNG derivatives are shown below the WT SH3\_5 sequence. **(B)** Two symmetry-related SH3\_5 domains coloured in blue and white with the P4-G5 ligand bound cooperatively. The P4-G5 ligand is coloured by atom type, with the P4 C atoms in green and G5 C atoms in pink. For clarity, all mutated residues are shown in stick form and displayed only once on the SH3\_5 domain that places them closest to the ligand.

**A****B**

**Figure 5.8. SDS-PAGE analysis of the SH3\_5-mNG derivative fusion proteins used for *in vitro* binding assays. (A)** Pooled fractions resulting from the gel filtration step to obtain the 27.4 kDa mNeonGreen control and 39.1 kDa SH3\_5-mNG proteins harbouring mutations in the residues involved in the interaction with G5 ligands. **(B)** Pooled fractions resulting from the gel filtration step to obtain the 39.1 kDa SH3\_5-mNG mutants harbouring mutations in the residues involved

in the interaction with P4 ligands. The amount of protein per binding assay was adjusted using the fluorescent signal intensity of the full-length protein.

protein per binding assay was adjusted using the fluorescent signal intensity of the full-length protein.

### **5.2.1.3 Lss proteins containing point mutations in the SH3\_5 domain**

To investigate the impact of SH3\_5 mutations on the activity of the Lss enzyme, an expression vector available in the laboratory collection (pET21a-Lss; Fig. 5.9 A) was used to produce 13 Lss derivatives harbouring the previously described single-site substitutions in their C-terminal SH3\_5 domains. The same pair of oligonucleotides used to introduce mutations in the SH3\_5 fusion domains were used for the full-length enzyme, except for N405A and W489L, which required a distinct pair of oligonucleotides to build the Lss derivatives (Materials and Methods, Table 2.4).

Three Lss mutants were not analysed since they did not bind to the HiTrap IMAC column charged with Zn<sup>2+</sup> ions (F418V, R427M, and M453A). The 28.1 kDa full-length Lss enzymes were produced and purified by affinity chromatography followed by gel filtration. Figure 5.9 B shows the purification steps for the WT-Lss protein. The purity of the proteins estimated on SDS-PAGE was >95%, with a yield of 8mg/mL (Fig. 5.9 C).

### **5.2.2 Contribution of six identified SH3\_5 residues to the recognition of pentaglycine and the peptide stem.**

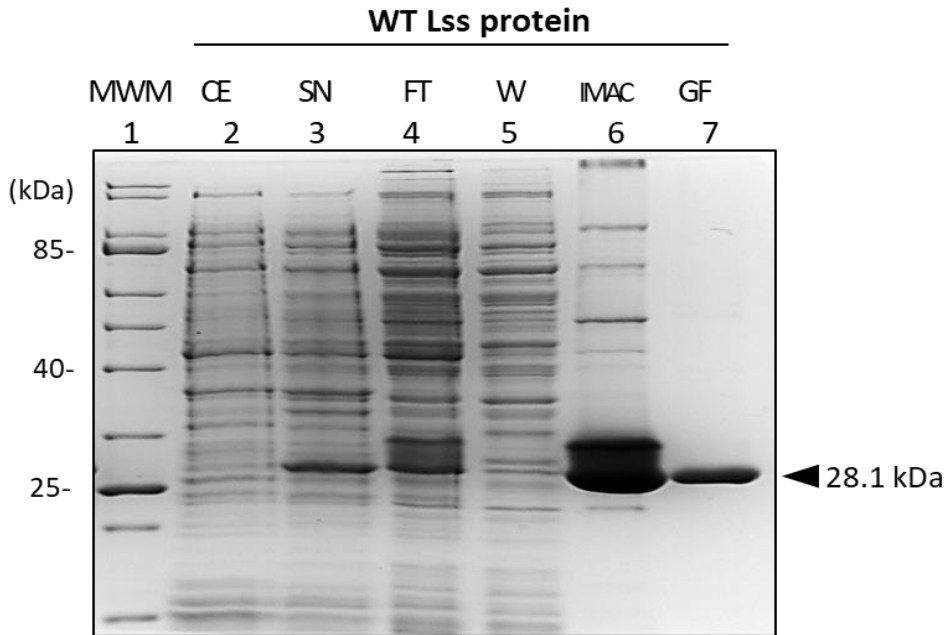
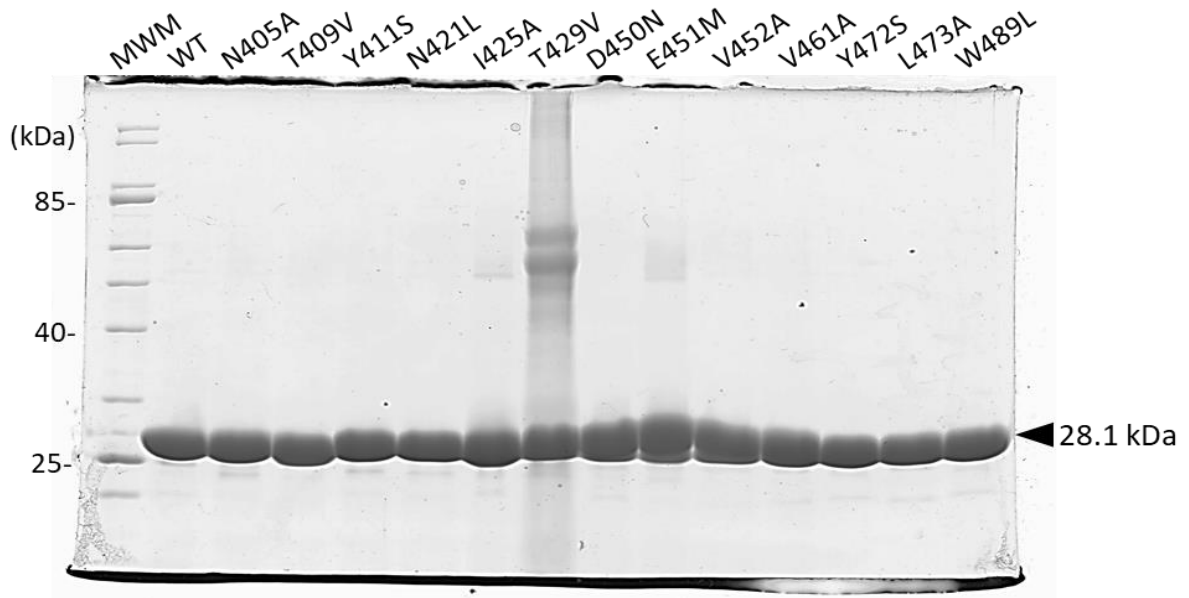
To confirm the contribution of SH3\_5 residues to the recognition of G5 and P4 peptides identified by NMR and X-ray crystallography, 6 His-SH3\_5 constructs described in Figures 5.1 and 5.2 were expressed as a singly <sup>15</sup>N labelled recombinant proteins and analysed by NMR titrations. The <sup>15</sup>N HSQC spectra of all mutant domains revealed that these were properly folded.

### 5.2.2.1 Binding of the N405A-SH3\_5 mutant to the G5 peptide

The interaction between the singly labelled N405A SH3\_5 mutant and the synthetic G5 peptide was studied by <sup>15</sup>N HSQC titrations. NMR spectra were recorded before and after the addition of the maximum concentration of ligand used for the previous WT SH3\_5 analysis (30 equivalents of peptide, black to red on Fig. 5.10 A). The CSP of the N405A SH3\_5-G5 interaction displayed a fast exchange rate with an evident reduction of the chemical shift changes on the residues interacting with G5 compared to the WT SH3\_5-G5 titration. The largest effects were seen in residues G430, F432, M435, M453, D456, and Y472 (Fig. 5.10 B).

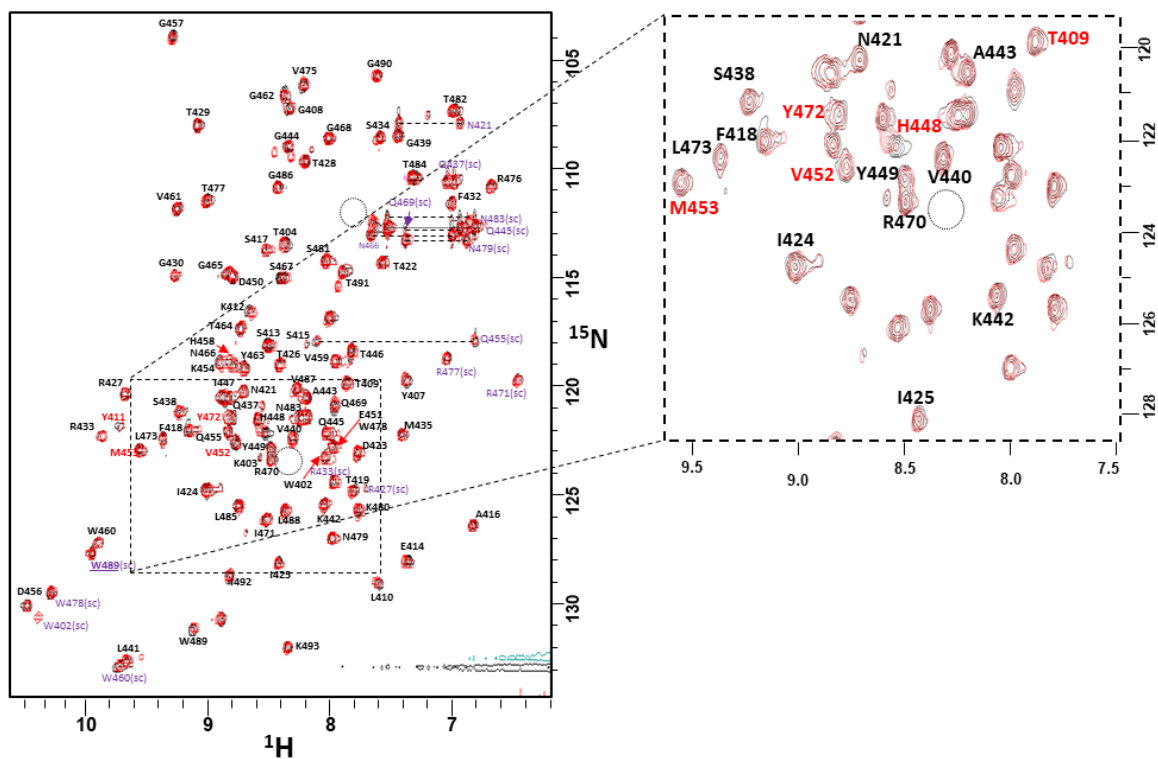
To determine the reduction on the binding affinity, the percentage of residual binding was established as a measurement of the ratio of chemical shift changes of mutant to WT for every residue and then averaged over all amino acids. The N405A SH3\_5 domain retained 16.1% of residual binding and had the most pronounced effect from the group of domains with mutations in residues involved in the interaction with the G5 peptide



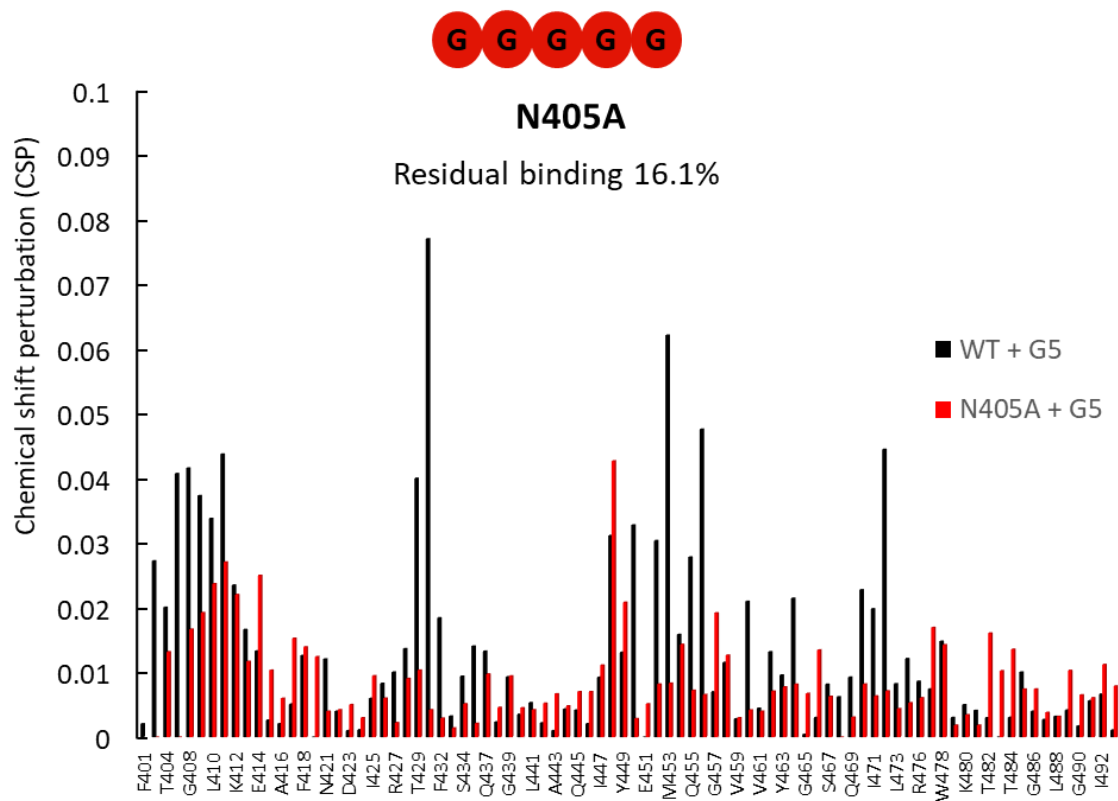
**A****B****C**

**Figure 5.9. Production and purification of full-length Lss proteins used to assess the impact of SH3\_5 mutations on the enzyme activity. (A)** Domain organization and amino acid sequence of the mature full-length Lss protein encoded by the pET-21a-Lss plasmid. Numbers correspond to amino acid positions. Amino acids corresponding to the N-terminal catalytic domain are coloured in red (136aa); linker in yellow (15 aa); SH3\_5 protein in blue (92aa); C-terminal 6x His-tag in pink. SH3\_5 mutagenized amino acids are underlined; the residues that were introduced by site-directed mutagenesis in constructing the Lss derivatives are shown below the WT sequence. **(B)** SDS-PAGE analysis of the purification steps to obtain the 28.1 kDa WT Lss protein expressed in *E. coli* Lemo21 (DE3) cells. Lane 1 (MWM), molecular weight makers; lane 2 (CE), crude extract; lane 3 (SN), soluble fraction; lane 4 (FT), flow through; lane 5 (W), wash; lane 6 (IMAC), elution pool; lane 7 (GF), pooled fractions resulting from the gel filtration step. **(C)** SDS-PAGE of the thirteen 28.1 kDa Lss derivative proteins expressed in *E. coli* Lemo21 (DE3) cells. Lane 1 (MWM), molecular weight makers; lane 2 (WT), pooled fractions resulting from the gel filtration step to obtain the full-length WT Lss protein; lane 3-15, pooled fractions resulting from the gel filtration step to obtain the thirteen Lss derivatives with mutagenized residues involved in the interaction with the G5 and P4 peptide.

**A**



**B**



**Figure 5.10. Residual binding of the N405A mutant to the G5 peptide. (A)** Overlaid  $^1\text{H}^{15}\text{N}$  HSQC spectra of the N405A SH3\_5-G5 titration acquired in the absence (black) and presence (red) of the maximum concentration of G5 ligand used in the WT SH3\_5 analysis (30 equivalents of peptide). The distribution of the peaks revealed a similar spectrum to that of the WT. Peaks from the backbone NH assignment are labelled in black and side chains in purple. The enlarged view illustrates a region of the spectra with peaks that interacted with the G5 peptide in fast exchange in the WT protein (labelled in red). Dotted circle indicates the original location of the mutagenized residue. **(B)** Comparison of chemical shift perturbations (CSP) in WT and N405A SH3\_5 mutant domains associated with the binding to G5 ligand. Histograms show individual CSP values from the  $^{15}\text{N}$ -HSQC titrations of N405A (in red) compared to the CSP values from the WT protein (in black) following addition of 30 equivalents of G5. The percentage of residual binding activity deduced from the average CSP values was 16.1% with the largest effects observed in residues G430, F432, M435, M453, D456, and Y472.

### **5.2.2.2 Binding of the M453A-SH3\_5 mutant to the G5 peptide**

<sup>15</sup>N HSQC titrations of the M453A SH3\_5-G5 interaction were recorded before and after the addition of 30 equivalents of peptide (black to red on Fig. 5.11 A). Like mutant N405A, the CSP of the M453A SH3\_5-G5 binding displayed a fast exchange rate; however, the peaks that displayed the most pronounced chemical shift changes in the WT had a similar CSP rate in the M453A mutant domain. The main differences were seen in residues K403, N421, and M435. (Fig. 5.11 B). The M453A-SH3\_5 mutant exhibited the least effect from the group of domains with mutations in residues involved in the interaction with the G5 peptide with a 55.5% of residual binding.

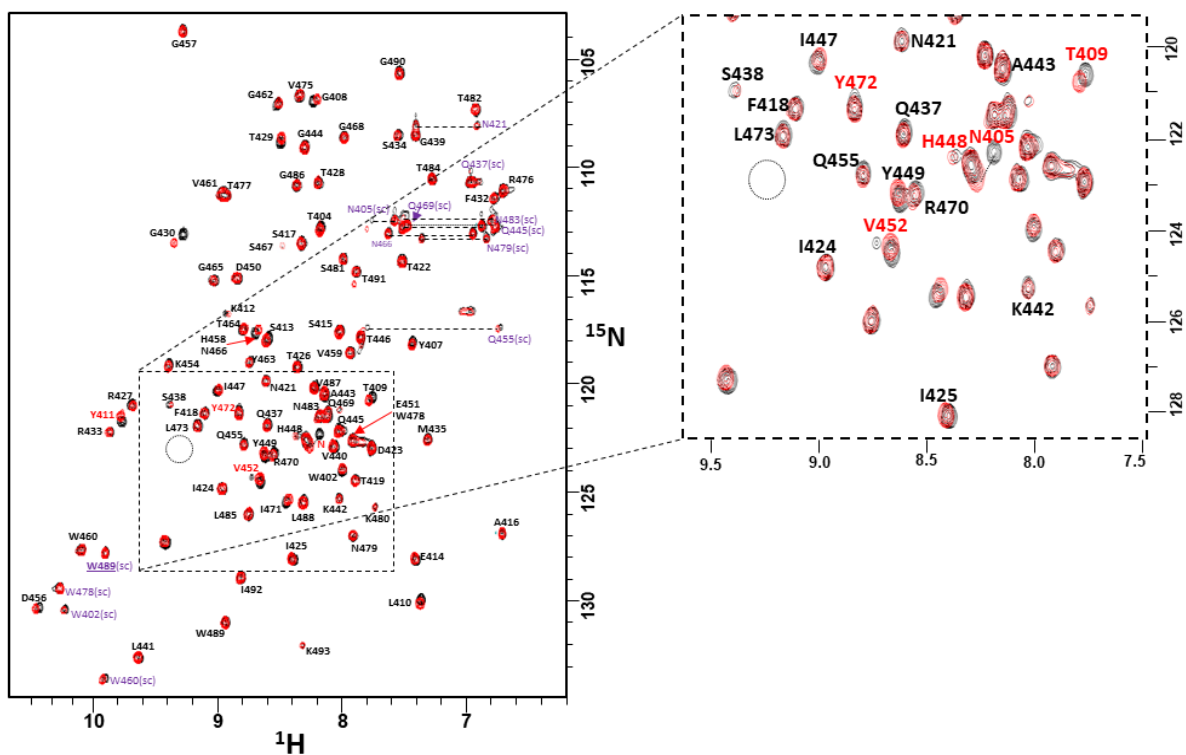
### **5.2.2.3 Binding of the Y472S-SH3\_5 mutant to the G5 peptide**

<sup>15</sup>N HSQC spectra of the Y472S SH3\_5 mutant revealed amide signals similarly positioned as those of the WT protein. Only a few residues were affected (R427-G430), suggesting that the protein was properly folded and did not have major conformational differences. The Y472S SH3\_5-G5 interaction was analysed as previously described by adding a single titration point with the maximum concentration of ligand used in the WT Lss-SH3\_5 titration (black and red on Fig. 5.12A). The Y472S mutant protein bound to the ligand in a fast exchange equilibrium with noticeable differences in the chemical shift changes of residues identified to interact with the ligand, the main differences compared to the WT were observed for residues K412, F432, D450, and M453. The Y472S mutant domain retained 24.6% of residual binding (Fig. 5.12 B).

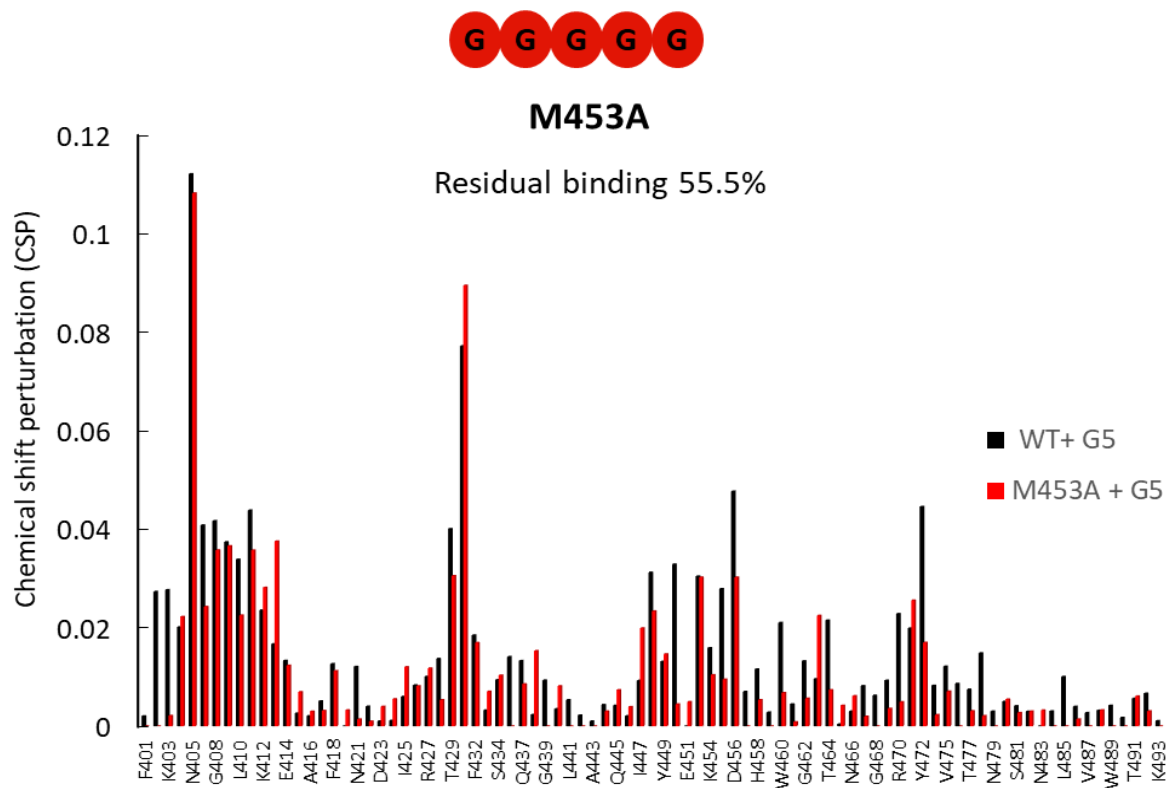
### **5.2.2.4 Binding of the I425A-SH3\_5 mutant to the P4 peptide**

The interaction between the I425A SH3\_5 derivative and the P4 peptide was analysed by NMR following the same strategy as the previous mutant domains. NMR spectra were recorded before and after the addition of the maximum concentration of P4 ligand used for the previous WT SH3\_5 analysis (32 equivalents of peptide, black to green on Fig. 5.13 A). The same amount of ligand was used for all mutant domains harbouring mutations in residues interacting with the P4 peptide.

**A**

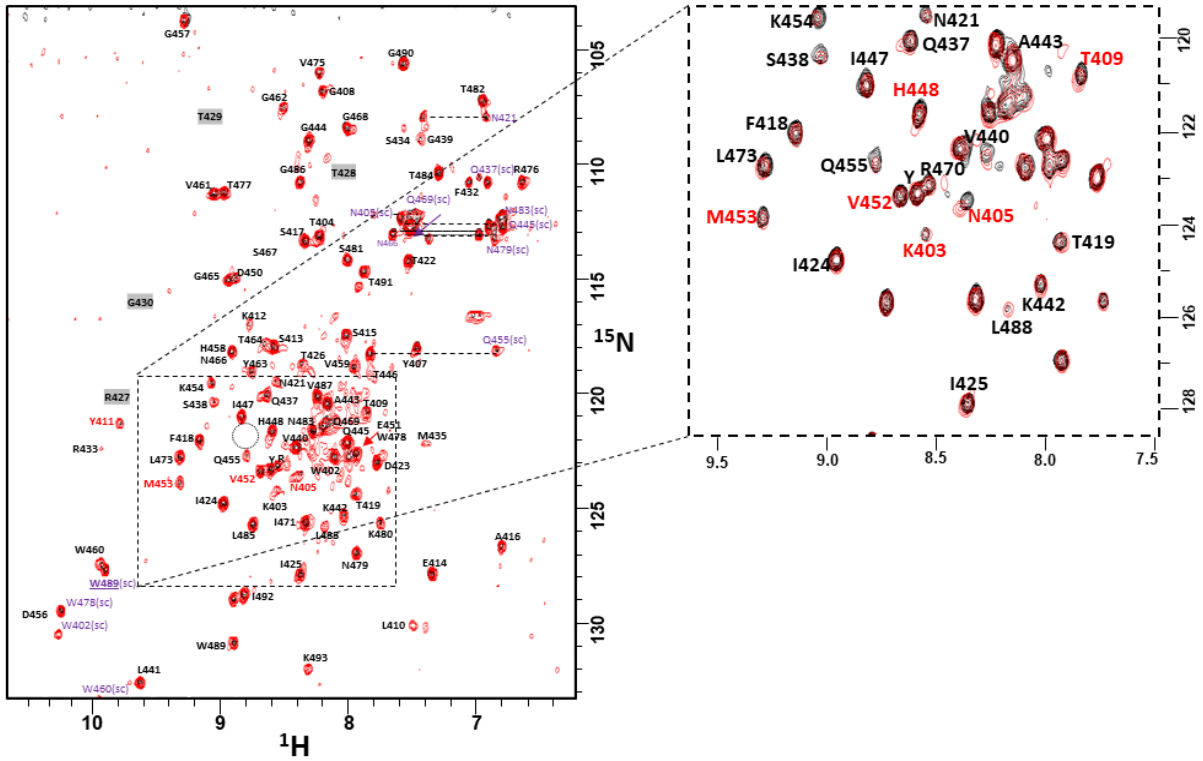


**B**



**Figure 5.11. Residual binding of the M453A mutant to the G5 peptide. (A)** Overlaid  $^1\text{H}^{15}\text{N}$  HSQC spectra of the M453A SH3\_5 mutant titrated with the G5 peptide. Spectra were acquired in the absence (black) and presence (red) of the maximum concentration of G5 ligand used in the WT SH3\_5 analysis (30 equivalents of peptide). The distribution of the peaks revealed that the protein was properly folded. Peaks from the backbone NH assignment are labelled in black and side chains in purple. The enlarged view illustrates a region of the spectra with peaks that interacted with the G5 peptide in fast exchange in the WT protein (in red). Dotted circle indicates the original location of the mutagenized residue. **(B)** Comparison of chemical shift perturbations (CSP) in WT and M453A SH3\_5 mutant domain associated with the binding to G5 ligand. Histogram shows individual CSP values from the  $^{15}\text{N}$ -HSQC titrations of M453A in red compared to the CSP values from the WT protein (in black) following addition of 30 equivalents of G5. The percentage of residual binding activity deduced from the average CSP values was 55.5% with the largest effects observed in residues K403, N421, and M435.

**A**

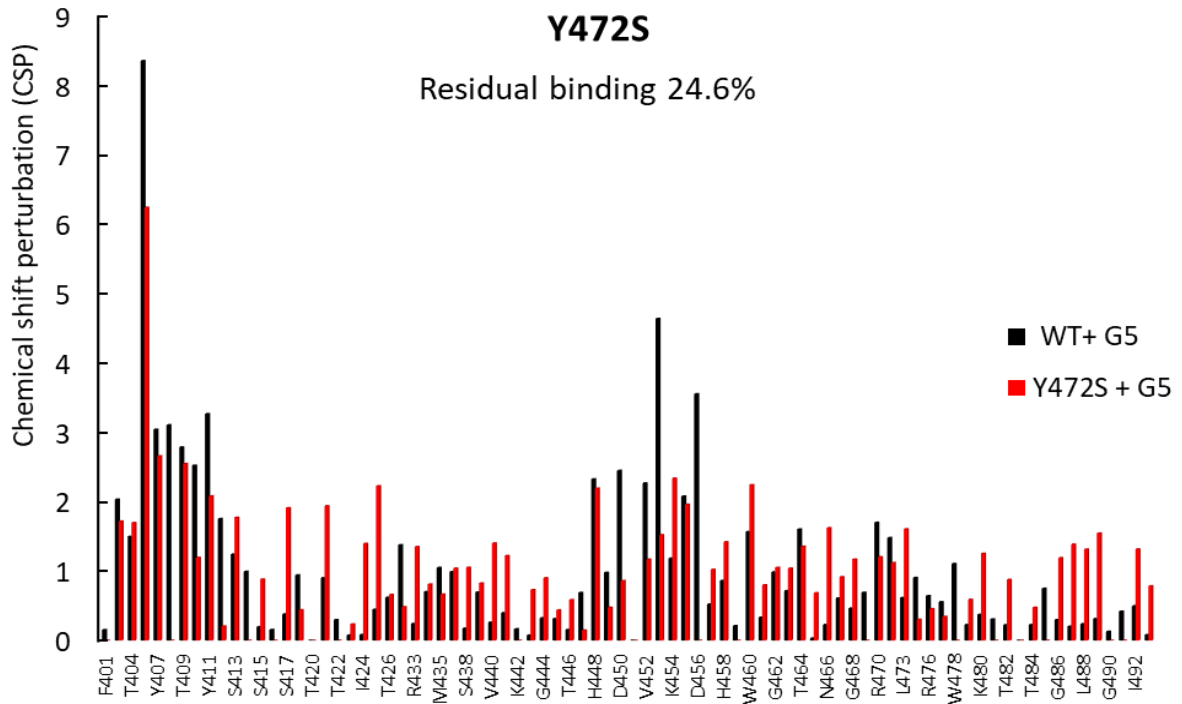


**B**



**Y472S**

Residual binding 24.6%





**Figure 5.12. Residual binding of the Y472S mutant to the G5 peptide. (A)** Overlaid  $^1\text{H}^{15}\text{N}$  HSQC spectra of the Y472S SH3\_5 -G5 titration acquired in the absence (black) and presence (red) of the maximum concentration of G5 ligand used in the WT SH3\_5 analysis (30 equivalents of peptide). The distribution of the peaks revealed that peaks were similarly distributed as those of WT indicating that the protein was properly folded. Peaks from the backbone NH assignment are labelled in black and side chains in purple. Dotted circle indicates the original location of the mutagenized residue. The enlarged view illustrates a region of the spectra with peaks that interacted with the G5 peptide in fast exchange in the WT protein (labelled in red). Residues highlighted in grey were not assigned to the Y472S mutant spectrum since they did not resonate at the expected position (ppm) as compared to the WT protein. **(B)** Comparison of chemical shift perturbations (CSP) in WT and Y472S SH3\_5 mutant domain associated with the binding to G5 ligand. Histograms show individual CSP values from the  $^{15}\text{N}$ -HSQC titrations of Y472S in red compared to the CSP values from the WT protein (in black) following addition of 30 equivalents of G5. The percentage of residual binding activity deduced from the average CSP values was 24.6% with the largest effects observed in residues K412, F432, D450, and M453.

The I425A SH3\_5 mutant bound to the P4 ligand in a fast exchange equilibrium. During the CSP analysis, a reduction of the chemical shift changes in residues previously identified to interact with the P4 ligand was determined, particularly for residues R470 and L473. However, the I425A SH3\_5 mutant displayed the least effect on binding to the P4 ligand, with 29.4% of residual binding (Fig. 5.13 B).

#### **5.2.2.5 Binding of the R427M-SH3\_5 mutant to the P4 peptide**

<sup>15</sup>N HSQC titrations were recorded before and after the addition of 32 equivalents of peptide. A marked impact on binding to the P4 peptide was observed during the spectra acquisition. All peaks maintained the same shape throughout the titration and no apparent chemical shift changes were observed after the addition of the maximum concentration of ligand, as seen in the almost complete superposition of the bound spectrum (SH3\_5+ligand) over the blank spectrum (protein with no ligand) (Fig. 5.14 A).

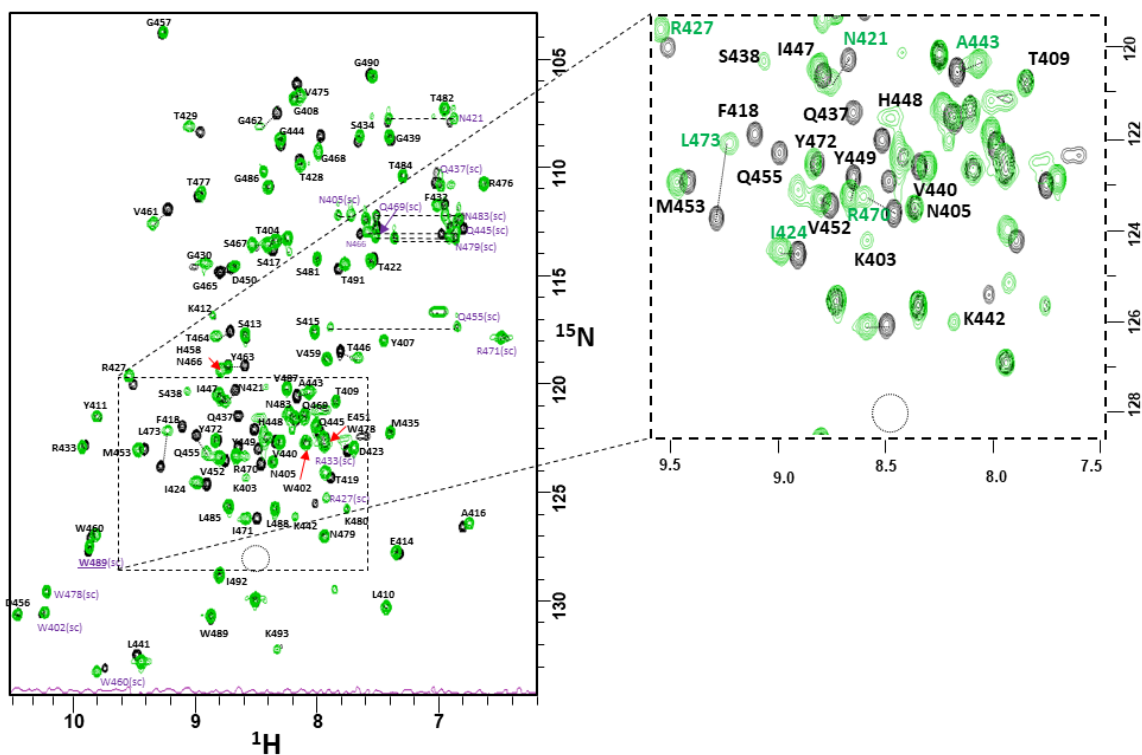
As expected from the co-crystal structure, the CSP analysis of the R427M SH3\_5 mutant revealed the largest impact in binding to the P4 peptide with 1.6% residual binding compared to the WT protein (Fig. 5.14 B).

#### **5.2.2.6 Binding of the W489L-SH3\_5 mutant to the P4 peptide**

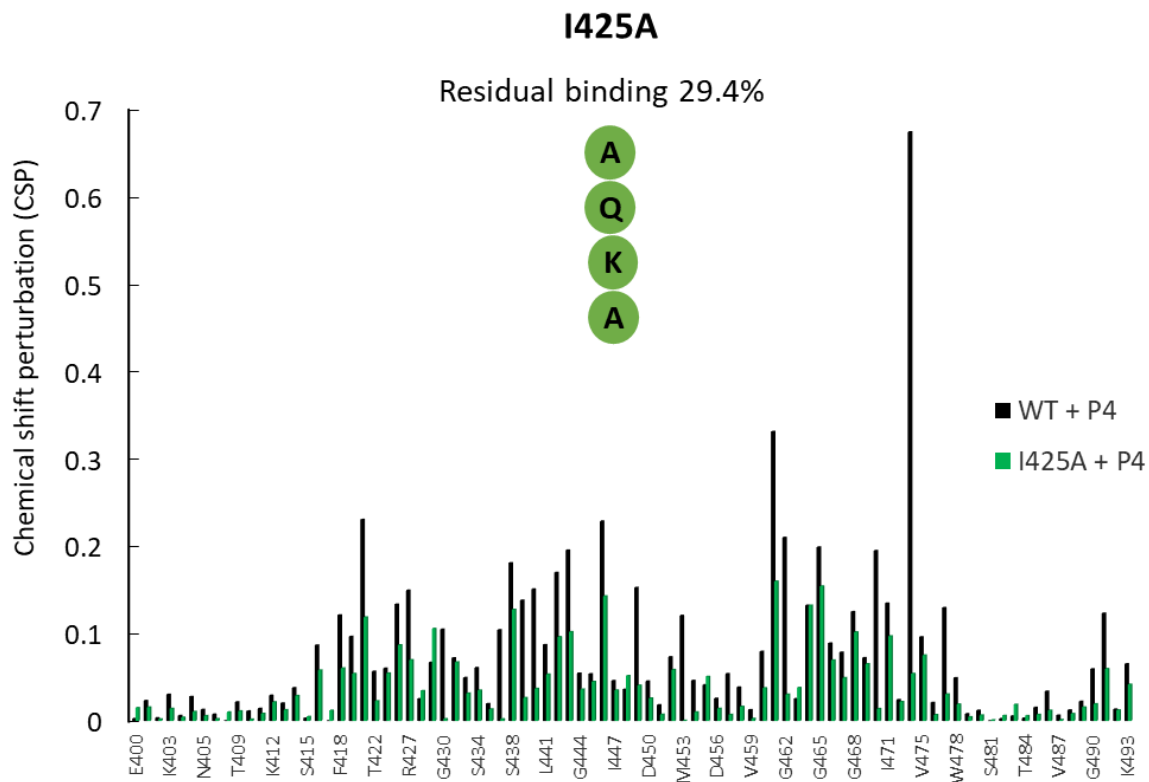
The last NMR titration was performed assessing the W489L SH3\_5 mutant. After the addition of 32 equivalents of P4 peptide, the CSP of the W489L SH3\_5-P4 interaction displayed a fast exchange rate with a noticeable decrease in the chemical shift changes from residues identified to interact with the P4 peptide. As observed for all previous titrations, all peaks maintained the same shape after the addition of the maximum ligand concentration (Fig. 5.15 A). Like the R427M mutant domain, the W489L mutation had a major impact on binding to the P4 ligand, and the mutant protein only retained 6.6% residual binding compared to the WT protein (Fig. 5.15 B).

Collectively, the NMR titration experiments confirmed the contribution of 6 SH3\_5 residues identified by NMR and X-ray crystallography to the recognition of the *S. aureus* PG peptide

**A**

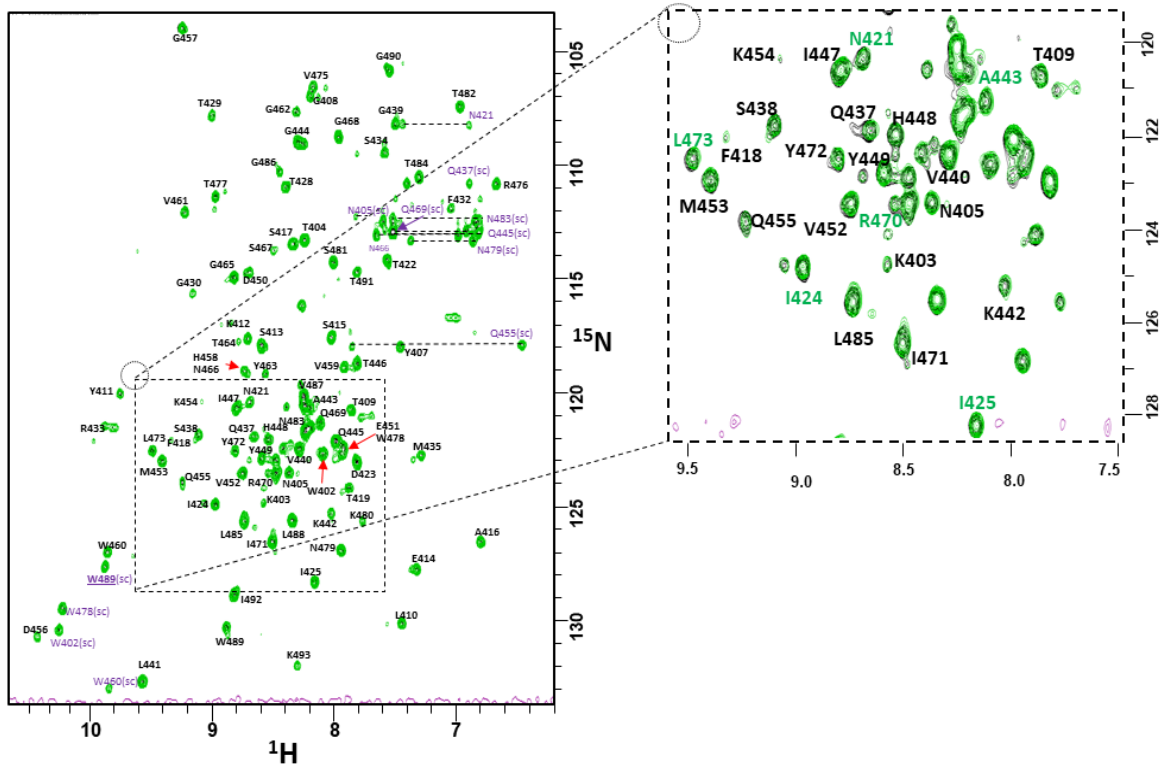


**B**

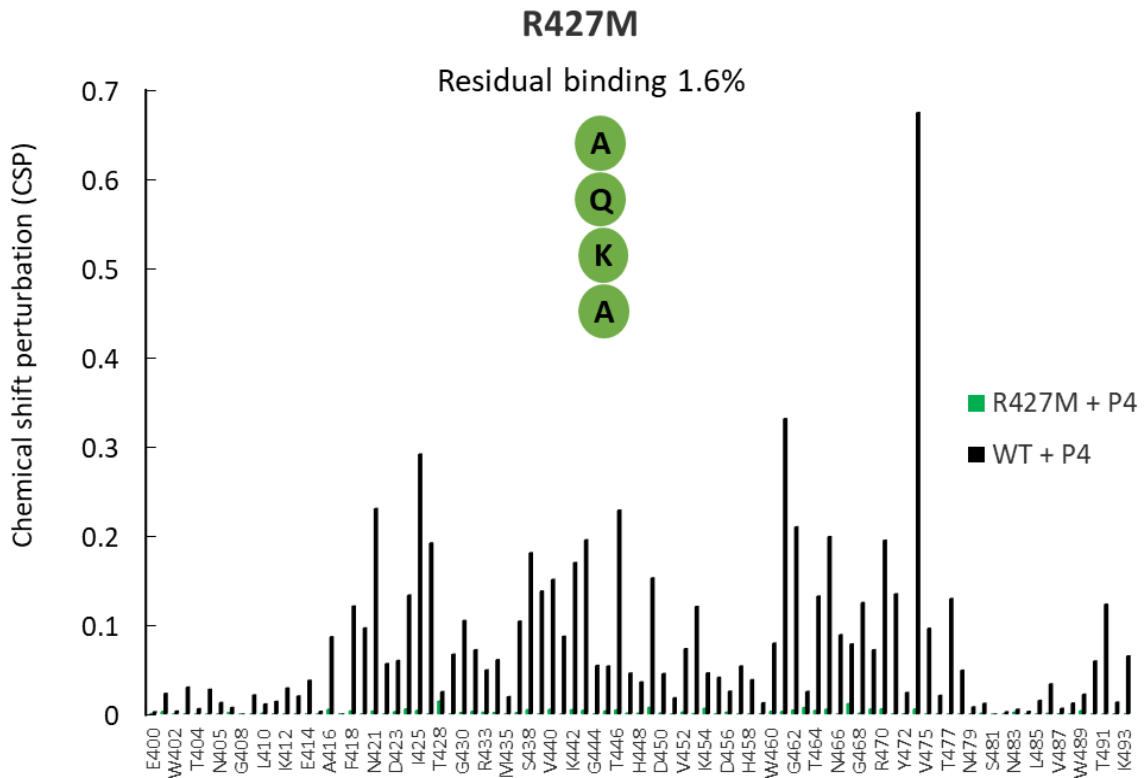


**Figure 5.13. Residual binding of the I425A mutant to the P4 peptide. (A)** Overlaid  $^1\text{H}^{15}\text{N}$  HSQC spectra of the I425A SH3\_5 -P4 titration acquired in the absence (black) and presence (green) of the maximum concentration of P4 ligand used in the WT SH3\_5 analysis (32 equivalents of peptide). The distribution of the peaks revealed that the protein was properly folded. Peaks from the backbone NH assignment are labelled in black and side chains in purple. The dotted circle indicates the original location of the mutagenized residue. The enlarged view illustrates a region of the spectra with peaks that interacted with the P4 peptide in fast exchange in the WT protein (labelled in green). **(B)** Comparison of chemical shift perturbations (CSP) in WT and I425A SH3\_5 mutant domain associated with the binding to P4 ligand. Histograms show individual CSP values from the  $^{15}\text{N}$ -HSQC titrations of I425A (in green) compared to the CSP values from the WT protein (in black) following addition of 32 equivalents of P4. The percentage of residual binding activity deduced from the average CSP values was 29.4% with the largest effects observed in residues R470 and L473.

**A**

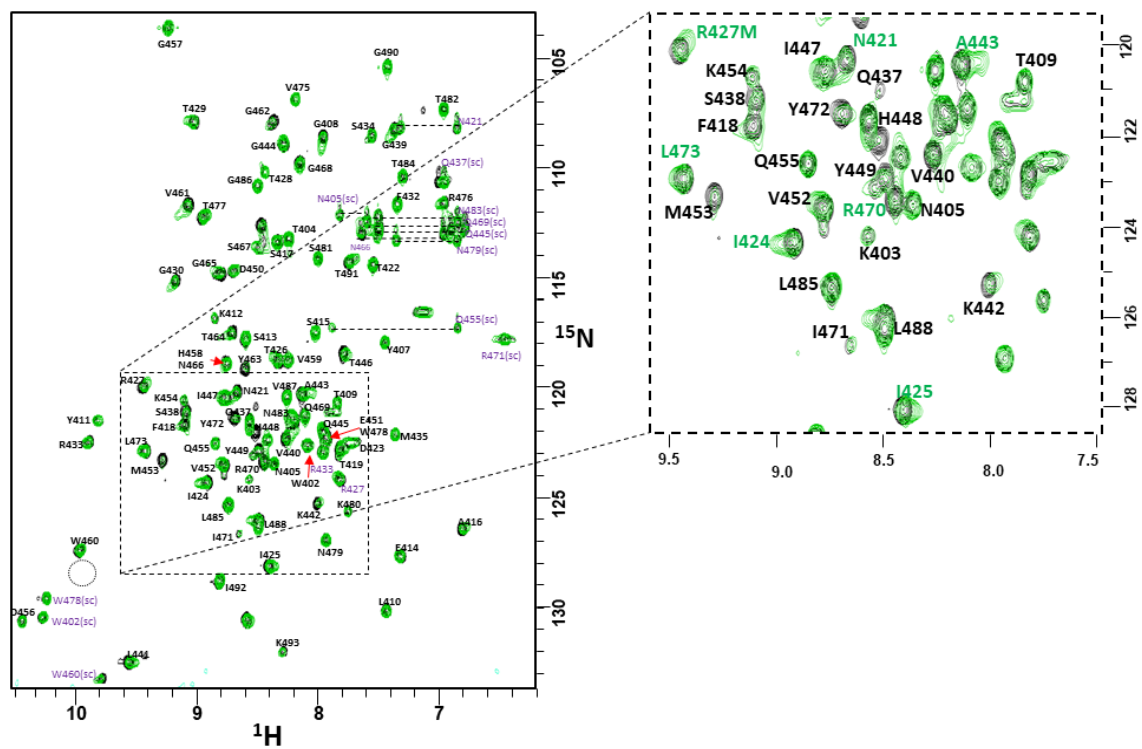


**B**

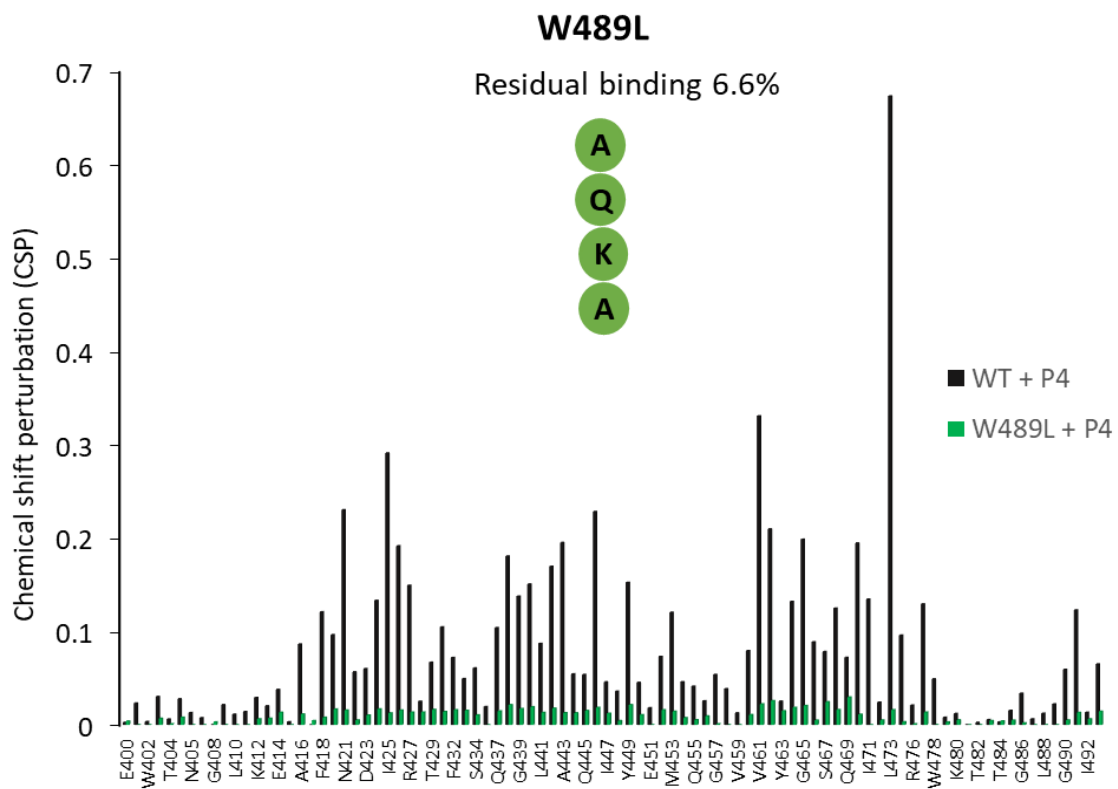


**Figure 5.14. Residual binding of the R427M mutant to the P4 peptide. (A)** Overlaid  $^1\text{H}^{15}\text{N}$  HSQC spectra of the R427M SH3\_5 mutant titrated with the P4 peptide. Spectra were acquired in the absence (black) and presence (green) of the maximum concentration of P4 ligand used in the WT SH3\_5 analysis (32 equivalents of peptide). The distribution of the peaks revealed a similar spectrum distribution to that of the WT. Peaks from the backbone NH assignment are labelled in black and side chains in purple. Dotted circle indicates the original location of the mutagenized residue. The enlarged view illustrates a region of the spectra with peaks that interacted with the P4 peptide in fast exchange in the WT protein (labelled in green). **(B)** Comparison of chemical shift perturbations (CSP) in WT and R427M SH3\_5 mutant domain associated with the binding to P4 ligand. Histograms show individual CSP values from the  $^{15}\text{N}$ -HSQC titrations of R427M (in green) compared to the CSP values from the WT protein (in black) following addition of 32 equivalents of P4. From all the SH3\_5 mutants analysed by NMR, R427M exhibited the largest effect in binding with an average residual binding of 1.6% compared to the WT SH3\_5 protein.

**A**



**B**



**Figure 5.15. Residual binding of the W489L mutant to the P4 peptide. (A)** Overlaid  $^1\text{H}^{15}\text{N}$  HSQC spectra of the W489L SH3\_5 -P4 titration acquired in the absence (black) and presence (green) of the maximum concentration of P4 ligand used in the WT SH3\_5 analysis (32 equivalents of peptide). The distribution of the peaks revealed that the protein was properly folded. Peaks from the backbone NH assignment are labelled in black and side chains in purple. Dotted circle indicates the original location of the mutagenized residue. The enlarged view illustrates a region of the spectra with peaks that interacted with the P4 peptide in fast exchange in the WT protein (labelled in green). **(B)** Comparison of chemical shift perturbations (CSP) in WT and W489L SH3\_5 mutant domain associated with the binding to P4 ligand. Histograms show individual CSP values from the  $^{15}\text{N}$ -HSQC titrations of W489L in green compared to the CSP values from the WT protein (in black) following addition of 32 equivalents of P4. The percentage of residual binding activity deduced from the average CSP values was 6.6%, the second lowest activity after the R427M mutant.



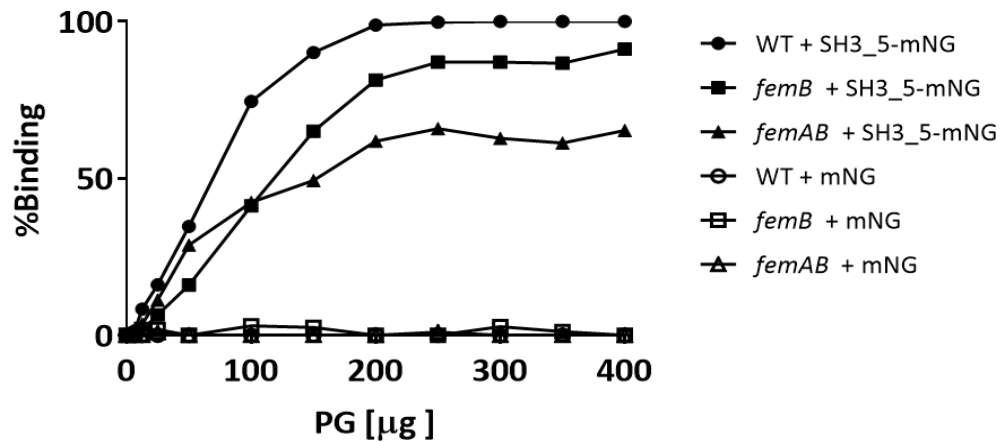
moieties. All mutant domains exhibited a decrease in binding activity to the peptide ligands tested. Domains harbouring mutations with residues identified to interact with the G5 peptide maintained a relatively high residual binding. The N405A derivative displayed the largest impact on binding (16.1% residual binding), whilst the M453A mutant retained 55.5% of binding to G5 compared to the WT. Mutations in residues associated with the P4 peptide displayed major effects in binding. The I425A had the least impact maintaining 29.4% of residual binding, whereas R427M and W489L substitutions caused a major impact in the recognition of the P4 peptide (1.6 and 6.6% residual binding, respectively).

Since the affinity reduction was not associated with any major structural rearrangements, the limited impact in binding for most of the mutations is in agreement with the X-ray crystal structure and the WT NMR analyses, in which it is shown that the recognition of both peptide moieties relies on a complex network of interactions, particularly for the mutations in residues involved in interactions with G5, where specific interaction with the pentaglycine stem seems to be reliant on the very tight positioning of the glycine residues within the narrow binding cleft.

### **5.2.3 Binding of SH3\_5 derivatives to purified PG sacculi**

A quantitative *in vitro* binding assay using *S. aureus* PG sacculi was designed to further investigate the contribution of residues involved in the recognition of both PG peptide moieties. The binding of the WT SH3\_5-mNG fusion was first assessed by incubating the protein with three distinct PG sacculi, the WT *S. aureus* SH1000 (Horsburgh, 2002), *femB* and *femAB* mutants (containing five, three and one glycine residue in their crossbridges, respectively). Binding was followed in the presence of increasing amounts of PG using in-gel fluorescence. Protein amounts equivalent to 3 µg of the WT SH3\_5-mNG fusion were adjusted based on the fluorescence intensity of the bands corresponding to the full-length protein (Fig. 5.16 A). Binding curves were used to determine the amount of PG required for 50% binding (PG<sub>50</sub>; Fig. 5.16 B). This value was used as a proxy to compare binding affinities to various PGs and to compare the relative binding activities of mutants (Fig. 5.17).

**A**



**B**

SH3b	PG <sub>50</sub> <sup>a</sup>	Fold change <sup>b</sup>
WT <sup>c</sup>	69.2	1.00
<i>femB</i>	118.3	1.71
<i>femAB</i>	152.6	2.20

**Figure 5.16. Binding activity of the SH3\_5-mNeonGreen fusion protein to WT *S. aureus* PG and *fem* mutants.** (A) Analysis of the PG-binding activity of the WT SH3\_5-mNeonGreen fusions (SH3\_5-mNG) to three distinct PG sacculi, the WT *S. aureus* SH1000 and two mutants with altered PG crossbridges, *femB* mutant (GGG) and *femAB* (G), respectively. Binding was followed in the presence of increasing amounts of PG (0-400 µg). Protein amounts equivalent to 3 µg of the WT SH3\_5-mNG fusion were adjusted based on the fluorescence intensity of the bands corresponding to the full-length protein. Fluorescence of the supernatant was measured using in-gel fluorescence. The graph shows dose binding responses of the SH3\_5-mNG fusions to all purified PG. As a positive control, no binding activity of the mNeonGreen (mNG) protein alone was determined in the presence of (0-400 µg PG); as a negative control, the maximum amount of PG sacculi (400 µg) was incubated without the presence of protein to verify for autofluorescence, and finally the SH3\_5-mNG fusion was also incubated without the presence of PG (B) The amount of PG required for 50% binding (PG<sub>50</sub>) was determined from the dose-response binding curves shown in (A) and the corresponding fold change as compared to the amount of PG required for 50% binding of the SH3\_5-mNG to the WT PG.

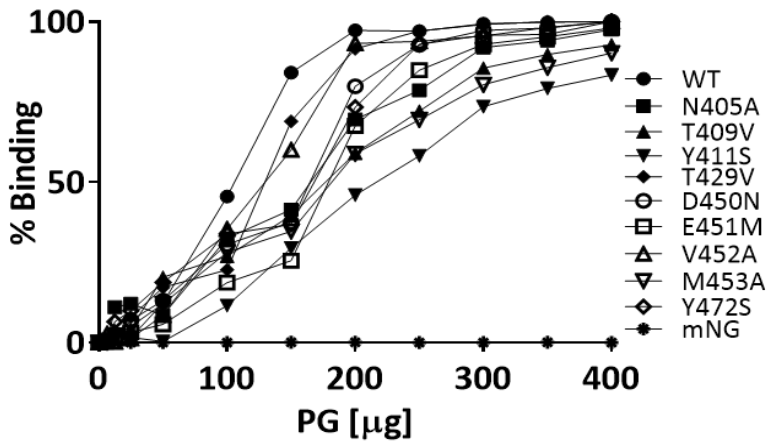
As expected, the WT SH3\_5-mNG fusion bound in a dose-response manner to the *S. aureus* PG sacculi (GGGGG) (Fig. 5.16 A). The SH3\_5-mNG protein also bound to the *femB* mutant (GGG) in a dose-response manner but its binding was reduced by a 1.71-fold change. A larger impact was observed for the *femAB* mutant with a binding reduction of 2.20-fold binding reduction. The mNeonGreen protein alone did not bind to any of the purified PG sacculi. The residual binding observed in both *fem* mutant PGs supports the existence of a second binding recognition site for the peptide stem. This observation was also described by (Lu JZ, 2006), where they determined that the SH3\_5 domain from Ale 1, a close homologue of Lss, exhibited residual binding to both *femB* and *femAB* mutants.

Sixteen SH3\_5-mNeonGreen fusions with mutations in residues previously identified were purified (Fig. 5.8), and their binding to WT *S. aureus* S1000 PG sacculi was measured (Fig. 5.17). Two independent series of protein purifications were carried out, the first included the group of SH3\_5 derivatives harbouring residues identified to interact with the G5 peptide, and the second set was comprised of mutant domains associated with the P4 peptide. Each purification batch included a WT SH3n-mNG fusion. The binding activity was assessed using the same strategy as the previously described WT SH3\_5-NG fusion. All proteins were adjusted on the basis of fluorescence intensity of the bands corresponding to the full-length proteins.

Most of the mutants displayed reduction in binding to PG, and the amount of PG required for 50% binding was similar to that of the WT protein to the *femB* PG. Like the SH3\_5 derivatives studied by NMR, the major impact in binding was observed with the R427M and W489L mutants exhibiting a 2.47 and 2.18-fold reduction, respectively. In the set of proteins harbouring substitutions in residues interacting with the G5 peptide, the Y411S SH3\_5-mNG fusion displayed the largest effect with a reduction of 1.83-fold-change.

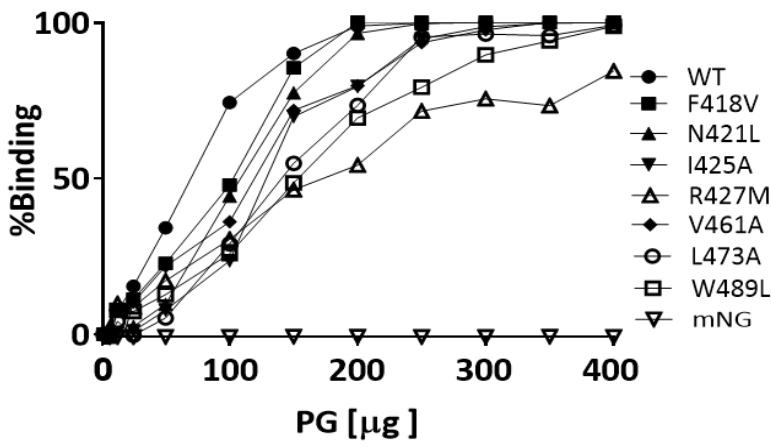
No single substitution abolished binding to the PG sacculi confirming the involvement of multiple amino acid residues in the recognition of both peptide moieties as determined by NMR and X-ray crystallography. To further corroborate the dual recognition by the Lss-SH3\_5 domain, the binding of the R427M SH3\_5-mNG mutant to the *fem AB* PG was assessed (Fig. 5.17 C).

A



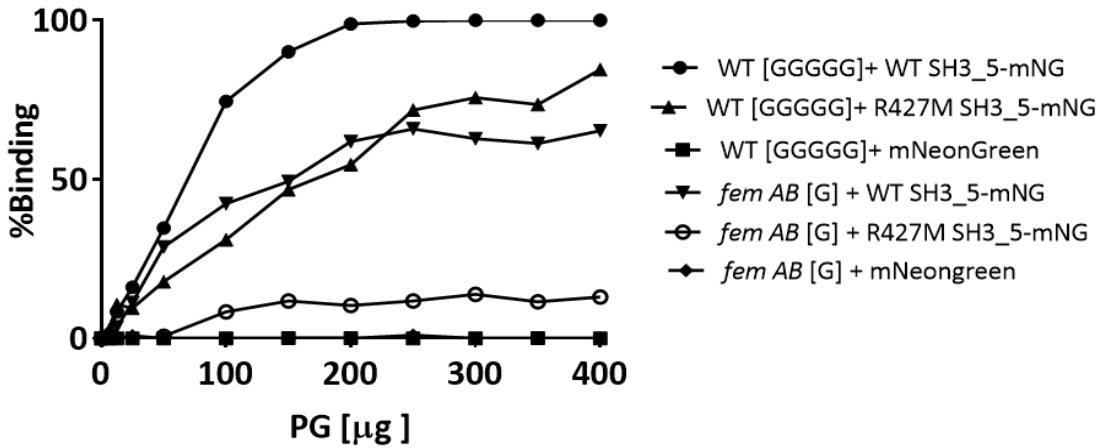
SH3_5	PG <sub>50</sub>	Fold change
WT	105.8	1.00
N405A	163.0	1.54
T409V	167.3	1.58
Y411S	193.5	1.83
T429V	129.5	1.22
D450N	164.8	1.56
E451M	177.9	1.68
V452A	129.7	1.23
M453A	171.5	1.62
Y472S	168.0	1.59
mNG		

B



SH3_5	PG <sub>50</sub>	Fold change
WT	69.2	1.00
F418V	102.4	1.48
N421L	107.9	1.56
V461A	118.9	1.72
I425A	128.1	1.85
R427M	170.6	2.47
L473A	139.0	2.01
W489L	151.0	2.18
mNG		

C



**Figure 5.17. Binding activity of the SH3\_5-mNeonGreen derivative proteins to purified *S. aureus* PG.** Analysis of the PG-binding activity of sixteen SH3\_5-mNeonGreen derivatives (SH3\_5-mNG) to *S. aureus* WT (SH1000) PG. For each SH3\_5-mNG fusion, protein amounts equivalent to 3  $\mu$ g were adjusted based on the fluorescence intensity of the bands corresponding to the full-length protein. Binding was followed in the presence of increasing amounts of *S. aureus* PG (0-400 $\mu$ g). Fluorescence of the supernatant was measured using in-gel fluorescence. **(A)** PG-binding activities of WT SH3\_5-mNG and derivatives with substitution of residues involved in the interaction with the G5 ligand. **(B)** PG-binding activities of WT SH3\_5-mNeonGreen and derivatives with substitution of residues involved in the interaction with the P4 ligand. The graphs show dose binding responses of all SH3\_5-mNG fusions to *S. aureus* PG. No binding activity to PG was determined for the mNeonGreen (mNG) protein control. The table next to each graph indicates the amount of PG required for 50% binding ( $PG_{50}$ ) and the corresponding fold change as compared to the amount of PG required for 50% binding of the WT protein as a reference. **(C)** Comparison of the binding activity of R427M SH3\_5-mNeonGreen fusion to purified WT *S. aureus* PG and *femAB* mutant. PG-binding activity of WT SH3\_5-mNG and the R427M SH3\_5-mNG derivative to the *femAB* mutant in comparison with their binding to WT *S. aureus* SH1000 PG. The graph shows dose binding responses of the WT SH3\_5-mNG fusion to both purified PG sacculi. The R427M SH3\_5-mNG fusion displayed a similar binding dose response to WT *S. aureus* PG as the WT SH3\_5-mNG protein to the *femAB* mutant. The major impact in binding was observed between the R427M SH3\_5-mNG derivative and the *femAB* mutant. No binding activity was determined for the mNeonGreen (mNG) protein control to any of the PG sacculi.

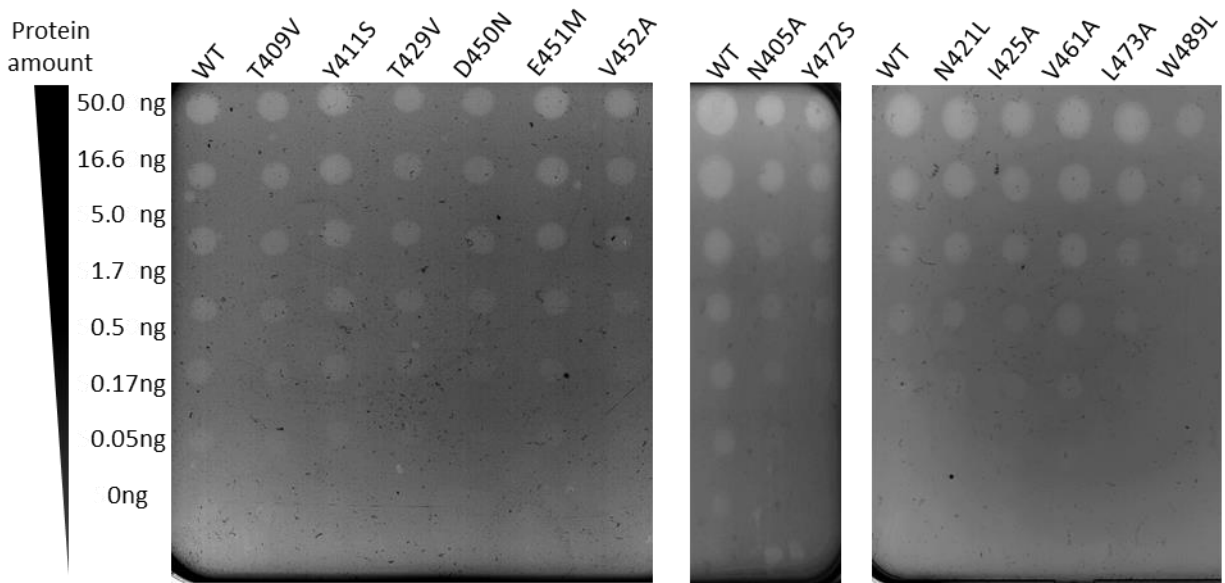
As expected, binding was not completely abolished but a major impact was observed. As a reference, the binding curves of the (i) WT SH3\_5-mNG to *fem AB* PG) and the (ii) R427M SH3\_5-mNG to WT *S. aureus* PG were also included. These results showed that, when only one recognition side was impaired either by the mutant SH3\_5 protein or *fem AB* PG, comparable binding curves were observed with the previously described PG<sub>50</sub> reduction of 2.20 and 2.47-fold, respectively.

#### **5.2.4 Impact of SH3\_5 single-site mutations on Lss activity**

To study the impact of SH3\_5 mutations on the activity of the full-length Lss enzyme, 13 Lss derivatives were produced (Fig. 5.9 C). Two-fold serial dilutions of each recombinant protein were spotted on LB agar plates containing *S. aureus* autoclaved cells (Fig. 5.18). An arbitrary Lss activity unit was defined as the greatest dilution of each sample. The enzymatic activity was detected as clear lytic zones on the agar plate as a result of the solubilization of the cell walls. The lytic zones were compared by determining the lowest amount of enzyme with a detectable digestion of the substrate. Three independent series of protein purifications were carried out, each including a WT Lss protein. In all purifications the WT Lss proteins exhibited comparable activity.

The spot assay results showed no major differences between the activity of the Lss derivatives compared to the WT proteins. Two Lss mutants showed the largest effect when compared to the WT: the Y472S and W489L mutants both exhibited a nine-fold decrease in activity. These observations are in line with the limited impact on binding to PG by the SH3\_5-NG derivatives. Unfortunately, the R427M Lss mutant did not bind to the nickel column and had to be excluded from the study.

The results from the mutational analyses described in this chapter showed that the recognition of the pentaglycine peptide and the peptide stem are equally important, it could even be suggested that recognition of the peptide stem is the most critical, as will be discussed later.



**Figure 5.18. Enzymatic activity of Lss recombinant proteins harbouring mutations in the SH3\_5 domain.** Three independent series of purifications were carried out (left, middle and right panels), each including a WT Lss protein as a control. Five  $\mu\text{L}$  corresponding to two-serial dilutions of recombinant Lss proteins were spotted on agar plates containing autoclaved *S. aureus* cells (final  $\text{OD}_{600}$  of 1) as a substrate. Lytic activities were detected as clearing zones and compared by determining the lowest amount of enzyme giving a detectable digestion of the substrate.

### 5.3 Discussion

NMR mutational analyses confirmed the role of six residues identified to interact with the G5 and P4 peptides by NMR and X-ray crystallography. Interestingly, substitutions in residues involved in the recognition of the peptide stems displayed the biggest effect in binding, especially R427M and W489L which almost prevented binding to the ligand in solution.

However, when a larger set of SH3\_5 derivatives were tested using *S. aureus* PG sacculi as a substrate, only a moderate impact in binding was determined. Nonetheless, the SH3-mNG derivatives R427M and W489L also exhibited the most pronounced effect. These residues have been identified in the homologous SH3\_5 domain from the Ale-1 hydrolase produced by *S. capitis* EPK1. Mutagenized Ale-1 SH3\_5 domain harbouring single-site substitutions in residues R296A and W358A, corresponding to R427 and W489 in the Lss SH3\_5 protein, displayed 3 to 2-fold reduction in binding to *S. aureus* PG, respectively, as compared to the WT protein (Lu JZ, 2006).

Sequence alignments between the Lss SH3\_5 domain (SH3\_5\_5) and other proteins have shown that residues I425, R427 and W489 are also conserved amongst various cell wall binding domains, including phage endolysins targeting PG that differ in their interpeptide bridge composition (Benešík M., 2018). These results indicate that the P4 D-Ala–carboxylate pocket is likely a conserved feature of wider bacterial SH3 family members (Fig. 5.19), with relevance for the SH3\_3 family. The SH3\_3 domains of *Clostridium* phage lysin phiSM101 have a carboxylate ligand bound at this position, suggesting an important role for the residues equivalent to Lss R427 (Tamai, 2014).

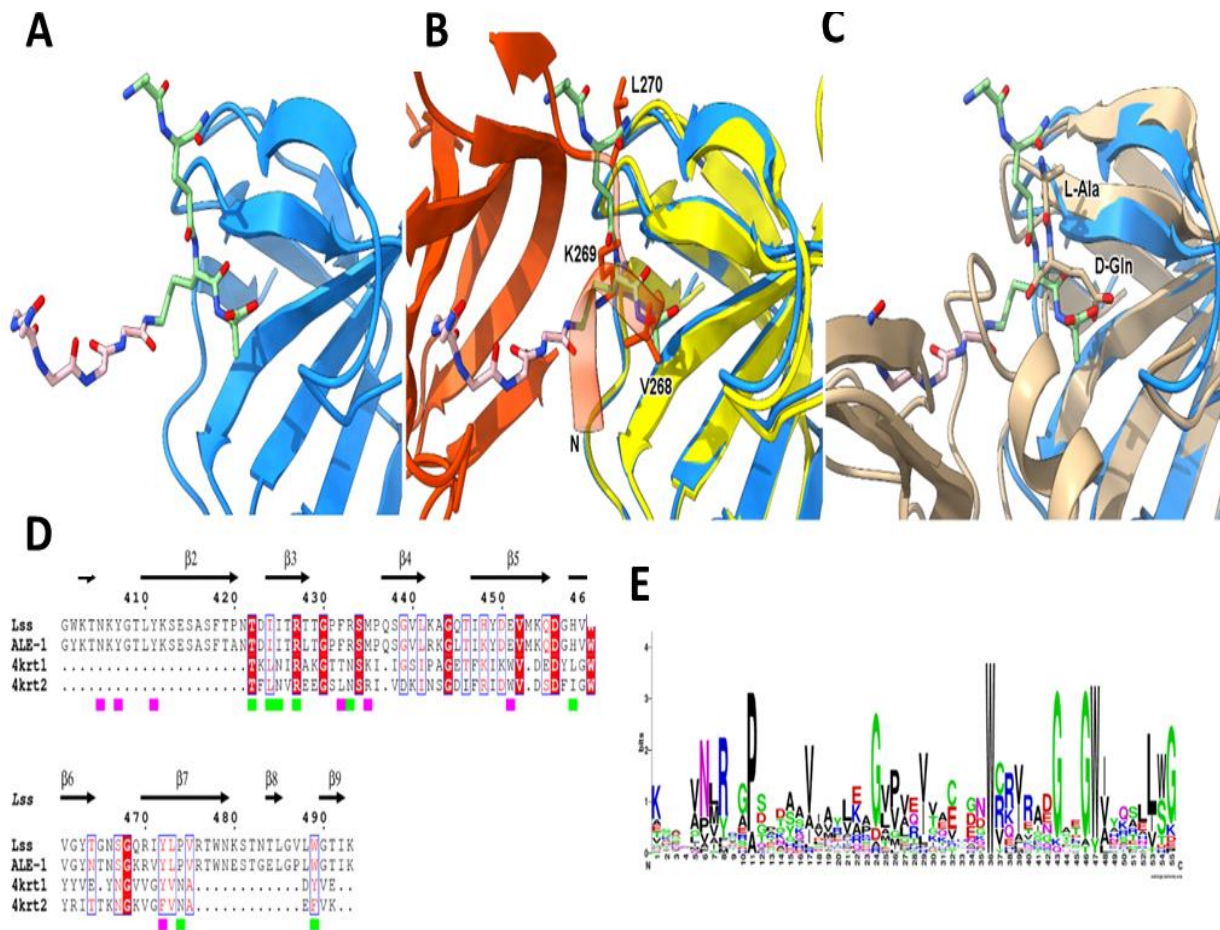
The Lss-SH3\_5 recognition of both peptide moieties appears to provide a biological advantage to the SH3\_5 protein as determined for the moderate binding impact in the protein derivatives and PG mutants (*fem* mutants). Binding of the WT SH3\_5-mNG fusion to the *fem B* (GGG) mutant had a similar effect to those observed for the SH3\_5 mutations to WT *S. aureus* PG. The WT SH3\_5-mNG fusion exhibited limited binding to the *fem AB* mutant (G) with a 2.20-fold change compared to the WT PG. This limited binding has also been described by (Gargis S.R., 2010), and the result resembled the decrease in binding observed with the R427M SH3\_5-mNG fusion to WT PG, implying that impaired binding to the pentaglycine or stem peptide moieties could be moderately



compensated by the recognition of the second site. Moreover, when the binding of the R427M SH3\_5-mNG derivative to the *fem AB* mutant was tested, a major decrease in binding was determined (Fig. 5.17 C). These observations, along with the limited effect in the lytic activity of the full-length Lss derivatives, support the NMR results reported for the WT SH3\_5 protein (Chapter IV) and the determined crystal structure of the SH3\_5 domain in complex with the P4-G5 peptide. It appears that the Lss-resistance to modified cross bridges in staphylococcal PG is likely to be associated with a deficient enzymatic activity of the catalytic domain.

Single-site substitutions in residues involved in the recognition of the peptide stem [P4] displayed a higher impact in binding. These observations are in agreement with the results obtained from the NMR titrations of the WT SH3\_5 domain using complex PG fragments (more cell wall-like ligands), in which lesser CSP were determined for residues interacting with the G5 ligand compared to the P4 peptide. These results appeared to be a mechanism adopted by the Lss enzyme to favour binding of its SH3\_5 domain to the peptide stems as long as some recognition of the pentaglycine remains. By not binding too tightly to either of its ligands, especially the pentaglycine cross bridges, it would allow access for the catalytic domain to cleave its substrate, whilst always keeping at least one site bound to avoid being detached from the PG surface so the SH3\_5 domain could continue binding to more non-contiguous PG peptide moieties.

Therefore, the recognition of non-contiguous PG fragments determined in this study reveals the mechanisms adopted by the cell targeting domain to effectively bind the PG lattice and thus enhance the activity of the catalytic domain.



**Figure 5.19. Comparison of ligand-binding pocket to other SH3\_5 structures and SH3 superfamily members. (A)** Lss SH3\_5 domain in complex with the P4-G5 ligand. **(B)** Superimposition of the structure shown in A onto the ALE-1 structure (PDB 1R77, two symmetry-related monomers orange and yellow). The P4-G5 ligand occupies the same space as an affinity purification tag (helical turn, N-terminus labelled). **(C)** Superimposition of the structure shown in A onto the phi7917 structure (PDB 5D76, tan). Ligand P3(K)-P4(A) are sterically equivalent to tag residues K269 and V268, respectively. The phi7917 ligand (L-Ala-D-Gln) is positioned with its peptide bond over the P3-P4 peptide. **(D)** Sequence alignment of Lss SH3\_5 (SH3\_5 subfamily) with ALE-1 (SH3\_5) and the two tandem domains of *Clostridium* phage phiSM101 (PDB 4krt, SH3\_3 subfamily), G5 and P4-ligating residues annotated with magenta and green block below text, respectively. **(E)** Weblogo (Crooks *et al.*, 2004) plot of sequence consensus of SH3\_4 subfamily, identifying features with likely equivalence to SH3\_3 and SH3\_5 alignment: NxR (position 6-8, match Lss I425-R427), W (36, match to W460), GxxGW (43-47, match to G468-Y472)

and LWG (53-55, match to L488-G490). No structures are currently available for SH3\_4 proteins, but structural comparison between Lss/ALE-1 and 4krt confirms conservation of the P-stem D-Ala(4)-carboxylate pocket.

## CHAPTER VI

### **Production of *E. faecalis* SH3\_5 proteins to explore PG binding mechanisms across species**

Our work on lysostaphin has shed light on a very unusual PG recognition mechanism and several questions remain to be answered: is the dual recognition of the peptide stem and the crossbridge shared across bacteria? Do multimodular SH3\_5 domains recognise PG independently or do they form quaternary structures? In this chapter, we tried to address these two questions using two proteins produced by *E. faecalis*. One of these proteins is a PG hydrolase with a single SH3\_5 repeat (EF1293) and the other protein is a bacteriocin that contains an SH3\_5 domain made of three repeats.

*E. faecalis* is a Gram-positive commensal bacterium present in the gastro-intestinal tract of humans. This opportunistic pathogen can cause life-threatening nosocomial infections associated with antimicrobial resistance, such as the vancomycin-resistant enterococci (VRE) strains (Gilmore, 2013; Paulsen, 2003; Polidori, 2011; Wisplinghoff, 2004).

Plasmid-encoded bacteriocins are often present in infection-derived *E. faecalis* strains (del Campo, 2001; Jack, 1995; Ness, 2014). These antimicrobial proteins are thought to provide a competitive advantage to the producer bacteria in an ecological niche where a wide variety of species are present. Bacteriocin-producing bacteria also possess immunity factors to protect them from their lytic effect (Kurushima, 2013, 2015; Ness, 2014). A study based on 636 VRE isolates showed that 44% of them were bacteriocin-producing strains (Inoue, 2006). One of these bacteriocins called “bacteriocin 41” or Bac41 is frequently found amongst clinical isolates (in up to 50% of the isolates analysed) (Ke, 2010)

The bactericidal activity of Bac41 is conferred by two extracellular proteins BacL<sub>1</sub> and BacA. Although both proteins are shown to be necessary for cell lysis, BacL<sub>1</sub> possess the enzymatic

machinery to bind and hydrolyse *E. faecalis* PG, whilst BacA has been proposed to act as an accessory factor. The BacL<sub>1</sub> C-terminal region is made of three nearly identical SH3\_5 repeats of approximately 90 amino acids. A previous study determined that fluorescent labelled BacL<sub>1</sub> can be localized in species harbouring an L-Ala-L-Ala crosslinked structure, such as *E. faecalis*, *S. pyogenes*, and *S. pneumoniae* (Kurushima, 2015). Although BacL<sub>1</sub> recognises these 3 different species, it can only lyse *E. faecalis* cells. The mechanism underpinning the specific lysis of *E. faecalis* by this bacteriocin is unknown.

Almost a quarter of the genome of *E. faecalis* V583 clinical isolates consists of mobile and/or exogenously acquired DNA, including integrated phage regions (Paulsen, 2003). The chromosome of V583 harbours seven prophage-like elements (pp1 to pp7), one of which (pp2) is found in all *E. faecalis* isolates and is part of the core genome (Matos, 2013; Paulsen, 2003; Solheim, 2011). Within this pp2 region, a well-conserved gene encoding for a putative PG hydrolase named EF1293 (accession number **Q835S9**) is present. The putative PG hydrolase is made of an N-terminal amidase domain and a single C-terminal SH3\_5 repeat. Interestingly, this is the only amidase described in *E. faecalis*. To our knowledge no further studies have been performed to characterise the binding of the EF1293 SH3\_5 domain.

## 6.1 Aims and objectives

The aim of the work described in this chapter was to explore (i) the binding mechanism of the SH3\_5 domain from the hydrolase EF1293, recognising a PG with a composition distinct from *S. aureus* (*E. faecalis* PG contains 2 L-Ala-L-Ala residues instead of 5 glycine residues), and (ii) to investigate the binding mechanism of a multimodular domain containing three SH3\_5 domain repeats (BacL<sub>1</sub>).

The specific objectives were:

- To produce *E. faecalis* SH3\_5 recombinant proteins with different domain architectures:

- (i) Five variants of the three modular BacL<sub>1</sub> SH3\_5 repeats, i.e. SH3<sub>1</sub>; SH3<sub>2</sub>; SH3<sub>3</sub>; SH3<sub>1-2</sub>; SH3<sub>1-3</sub>.
  - (ii) The single C-terminal SH3\_5 domain from the putative PG endolysin EF1293 protein.
- To perform NMR structural and functional analyses using a set of ligands corresponding to *E. faecalis* PG: the L-Ala-L-Ala, cross-links, the AQKAA, peptide stem and the AQK(L-Ala-L-Ala)AA, peptide stem with lateral chain.

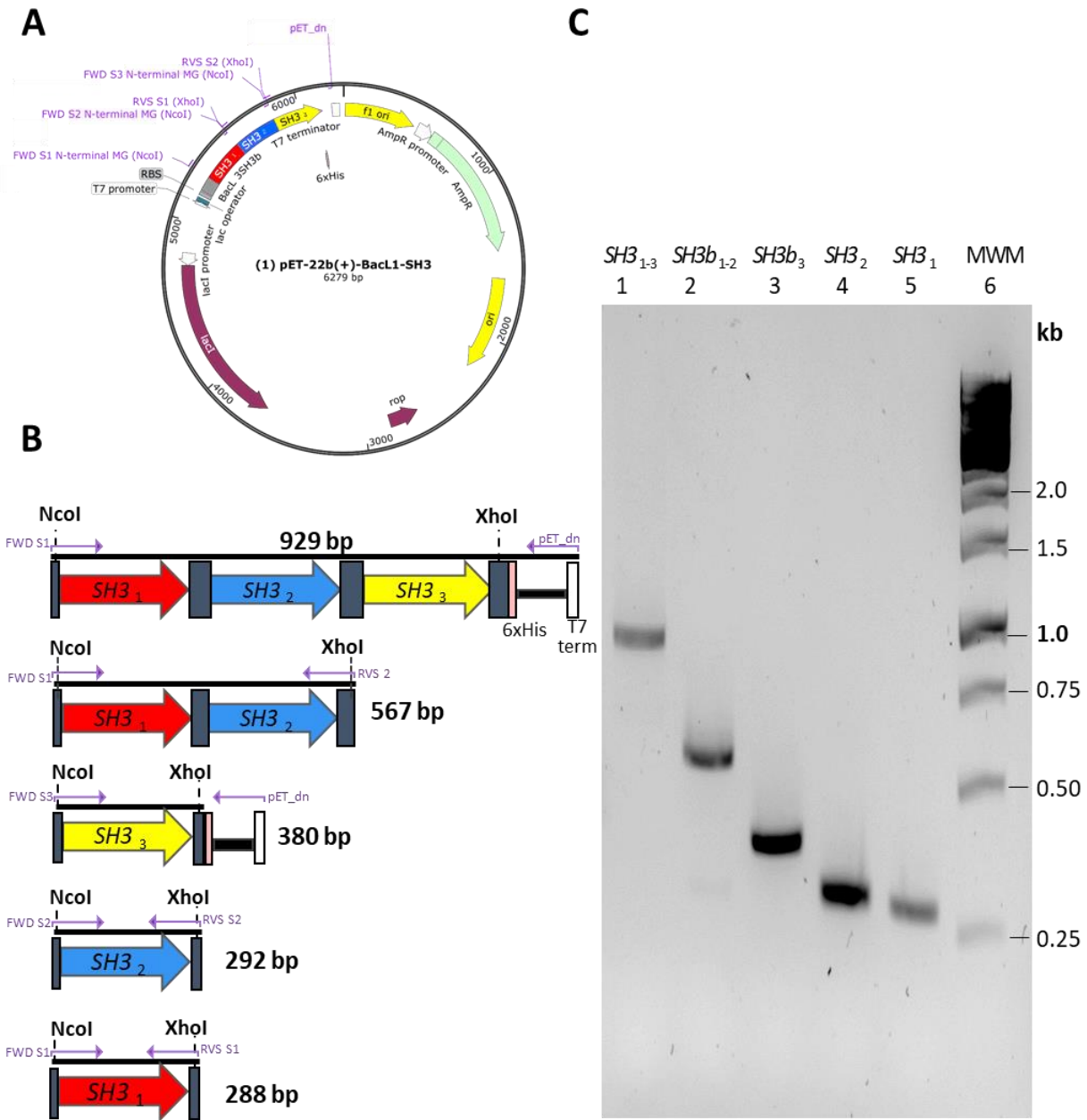
## 6.2 Results

### 6.2.1 Expression and purification of five unlabelled BacL<sub>1</sub> SH3\_5-His proteins

Five variants of the BacL<sub>1</sub> SH3\_5 domain were constructed using the plasmid pET22-BacL-SH3 previously described as a template (Kurushima, 2013) (Fig. 6.1 A). This plasmid encodes the BacL<sub>1</sub> SH3\_5 domain made of three repeats connected by their linker regions. The region encoding the entire domain is flanked by NcoI and XhoI sites followed by a C-terminal histidine-tag (Fig. 6.1 A). An alignment of the three BacL<sub>1</sub> SH3\_5 repeats with their linker regions is shown in Figure 6.2.

Five DNA fragments flanked with NcoI and XhoI sites were PCR amplified from the pET22-BacL-SH3 plasmid (Fig. 6.1 B-C):

- (i) a 929 bp fragment encoding the complete BacL<sub>1</sub> SH3\_5 domain (SH3<sub>1-3</sub>);
- (ii) a 567 bp fragment encoding the first two SH3\_5 repeats (SH3<sub>1-2</sub>);
- (iii) a 380 bp fragment encoding the most distal SH3\_5 repeat (SH3<sub>3</sub>)
- (iv) a 292 bp fragment encoding the second SH3\_5 repeat (SH3<sub>2</sub>);
- (v) a 288 bp fragment encoding the first SH3\_5 repeat (SH3<sub>1</sub>).

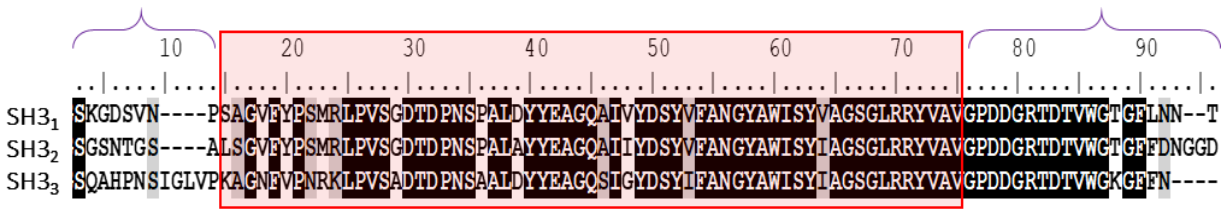


**Figure 6.1. PCR amplification of *E. faecalis* BacL<sub>1</sub> SH3 modules from pET22b-BacL<sub>1</sub> SH3.** (A) Schematic representation of the pET22b-BacL<sub>1</sub> SH3 plasmid encoding the three *E. faecalis* BacL<sub>1</sub> SH3<sub>5</sub> repeats (Kurushima, 2015) used as a PCR template. The SH3<sub>1</sub> (183 bp), SH3<sub>2</sub> (180bp) and SH3<sub>3</sub> (186bp) DNA fragments are coloured in red, blue and yellow, respectively; linkers in grey (from 3' to 5') are 30, 93, 102, and 57 bp, respectively. (B) SH3<sub>5</sub> fragments amplified for protein expression. All fragments were PCR amplified from pET22b-BacL<sub>1</sub> SH3 and flanked with

NcoI and XhoI sites: From top to bottom: **(i)** a 929 bp fragment encoding the complete BacL<sub>1</sub> SH3\_5 domain (SH3<sub>1-3</sub>), **(ii)** a 567 bp fragment encoding the first two SH3\_5 repeats (SH3<sub>1-2</sub>), **(iii)** a 380 bp fragment encoding the most distal SH3\_5 repeat (SH3<sub>3</sub>), **(iv)** a 292 bp fragment encoding the second SH3\_5 repeat (SH3<sub>2</sub>); **(v)** a 288 bp fragment encoding the first SH3\_5 repeat (SH3<sub>1</sub>).

**(C)** Agarose gel electrophoresis of *SH3\_5* PCR products encoding SH3<sub>1-3</sub> (lane 1), SH3<sub>1-2</sub> (lane 2), SH3<sub>3</sub> (lane 3), SH3<sub>2</sub> (lane 4), SH3<sub>1</sub> (lane 5); lane 6, molecular weight marker (MWM).



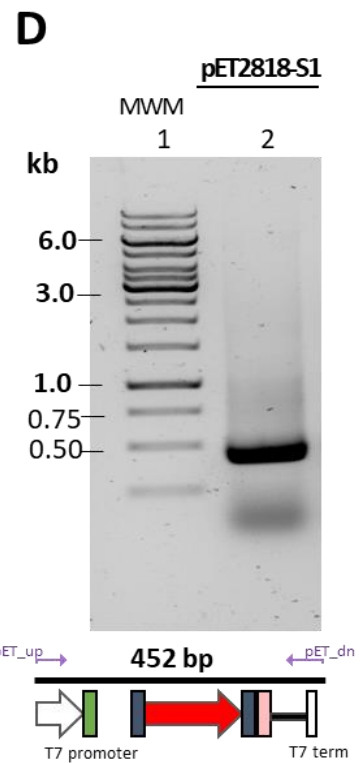
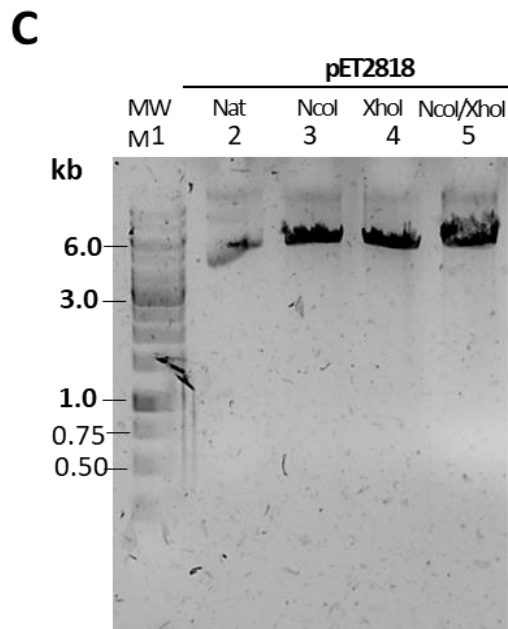
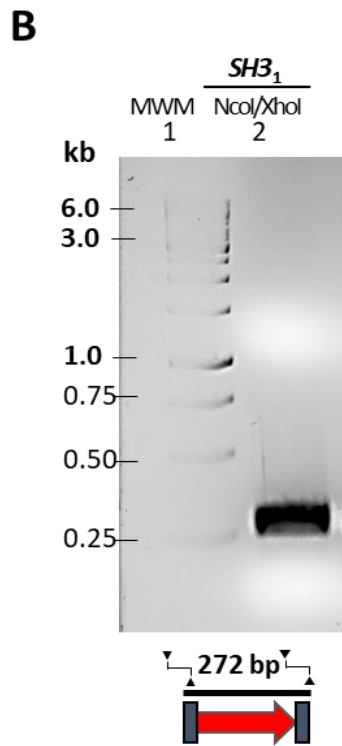
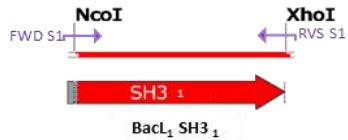
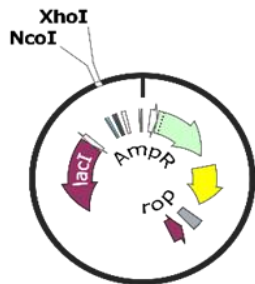
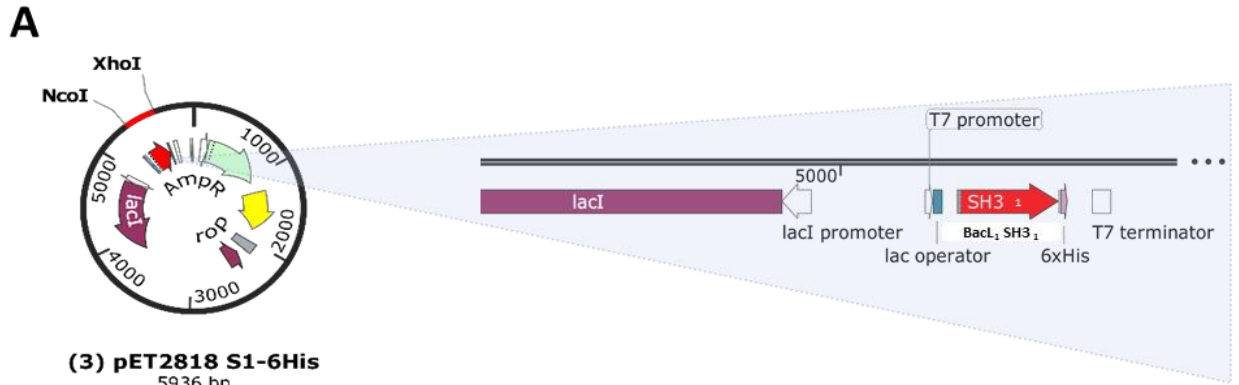


**Figure 6.2. Amino acid sequence alignment of the three BacL<sub>1</sub> C-terminal SH3<sub>5</sub> repeats.** The residues corresponding to the three SH3<sub>5</sub> proteins (SH3<sub>1</sub> to SH3<sub>3</sub>) are boxed in red (60 residues); linker sequences flanking the N- and C-terminal regions of each protein repeat are indicated by purple brackets. Identical residues are highlighted in black and residues in light grey indicate conservative substitutions.

All PCR products were cloned into pET2818 using XhoI and NotI restriction sites to produce proteins with a C-terminal His-tag encoded by the plasmid. The cloning strategy for each construct is described in Figures 6.3 to 6.7. The final plasmids: pET2818-S1, pET2818-S2, pET2818-S3, pET2818-S1S2, and pET2818-S1S3 were confirmed by colony PCR screens and sent to sequencing to check for the absence of mutations in the inserts.

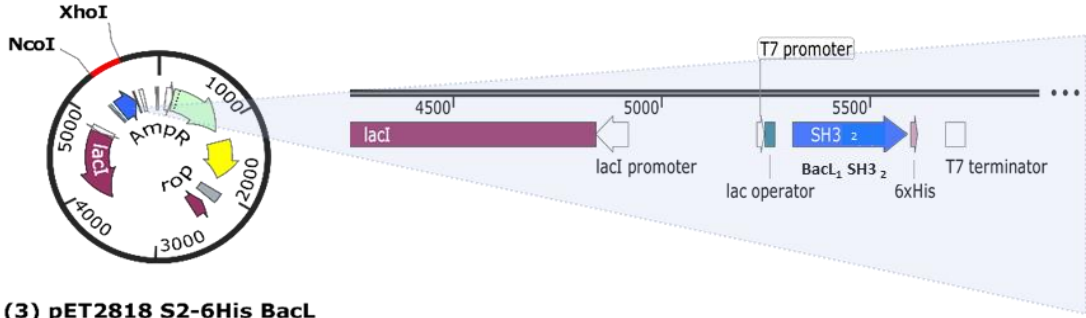
Small-scale expression trials were performed to determine the expression level and solubility of the five recombinant SH3 constructs (Fig. 6.8 A-C). Proteins were produced in *E. coli* Lemo21(DE3) cells and IPTG induced at 25°C for overnight. Cells and supernatant samples were harvested and analysed by SDS-PAGE. All recombinant SH3\_5 proteins were highly expressed and soluble under standard expression conditions (Fig. 6.8 C). The expected molecular weights of the produced proteins are: SH3<sub>1-3</sub>, 30.2 kDa; SH3<sub>1-2</sub>, 20.4 kDa; SH3<sub>1</sub>, 10.7 kDa; SH3<sub>2</sub>, 10.7 kDa; SH3<sub>3</sub>, 10.9 kDa (Fig 6.8 B-C).

A subsequent large-scale expression (1L) was performed for each protein . Proteins were purified by affinity chromatography followed by gel filtration (Fig. 6.8 D). The overall purity was estimated on SDS-PAGE (>90%) with a yield of 20mg/mL.

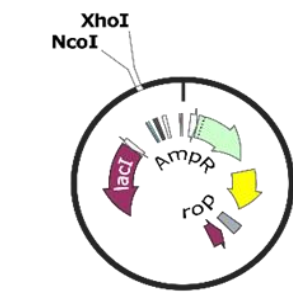
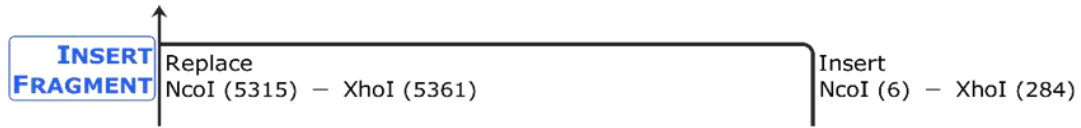


**Figure 6.3. Construction of pET2818-S1. (A)** Schematic representation of the cloning strategy. A 288 bp PCR fragment encoding the first BacL<sub>1</sub> SH3<sub>1</sub> repeat (SH3<sub>1</sub>) was cloned into pET2818 using NcoI and XhoI sites. **(B)** Restriction digest of the SH3<sub>1</sub> PCR fragment using NcoI and XhoI. Lane 1 (MW), molecular weight marker; lane 2, 277bp SH3<sub>1</sub> fragment digested with NcoI and XhoI. **(C)** Restriction digest of pET2818 with NcoI and XhoI. Lane 1, (MW), molecular weight marker; lane 2 (Nat), undigested 5710 bp pET2818 plasmid; lane 3 (NcoI), 5710 bp pET-2818 digested with NcoI; lane 4 (XhoI), 5710 bp pET-2818 digested with XhoI; lane 5 (NcoI/XhoI), pET-2818 doubly digested with NcoI and XhoI. **(D)** Identification of pET2818-S1 by colony PCR using oligonucleotides pET\_up and pET\_dn. Lane 1, (MW), molecular weight marker; lane 2, 452 bp PCR fragment corresponding to SH3<sub>1</sub>.

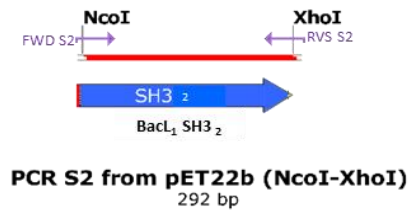
**A**



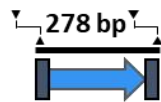
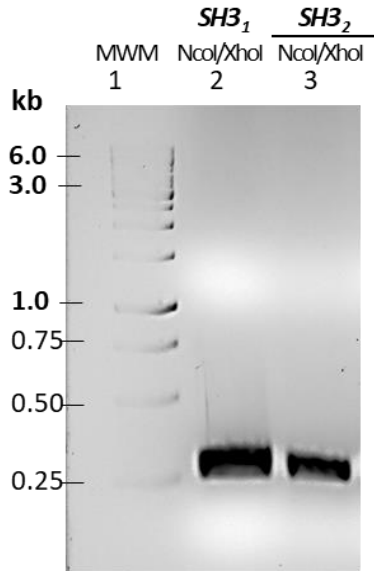
**(3) pET2818 S2-6His BacI**  
5942 bp



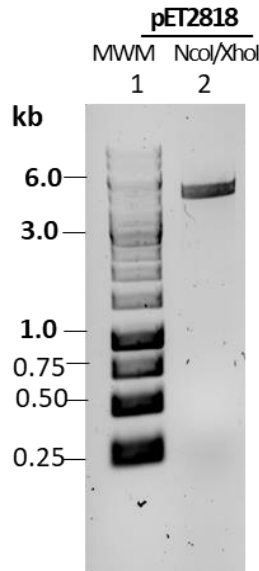
**pET2818**  
5710 bp



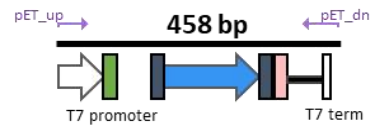
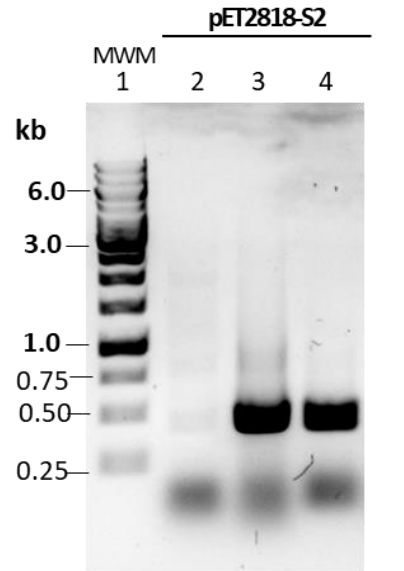
**B**



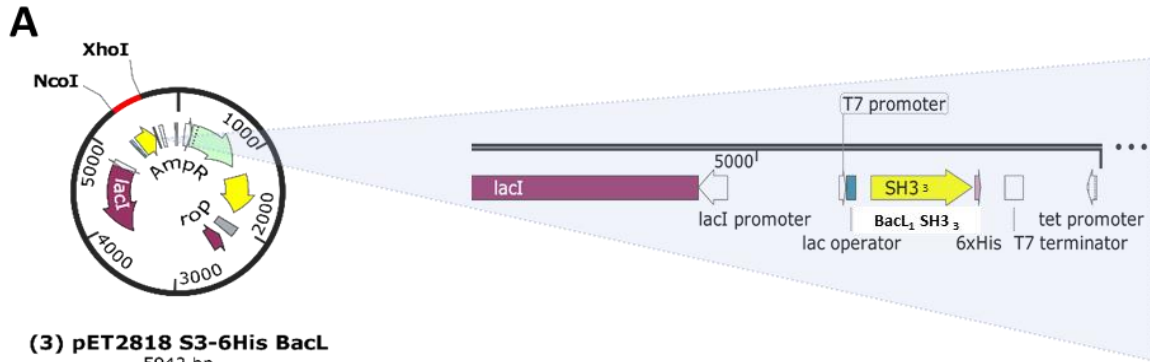
**C**



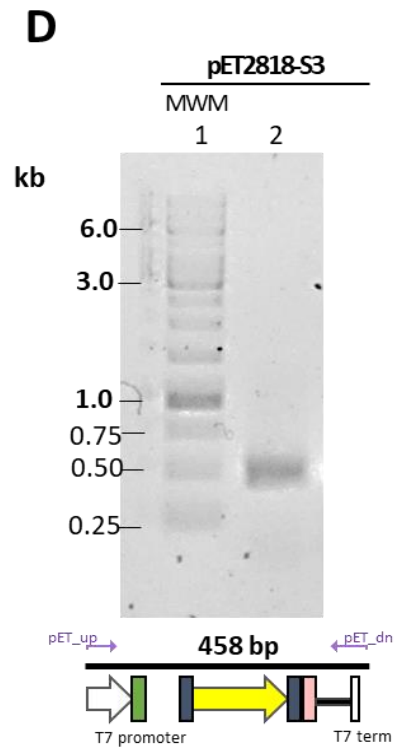
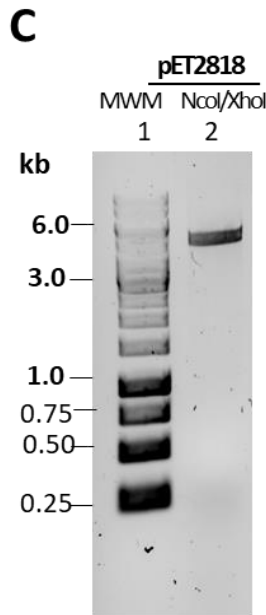
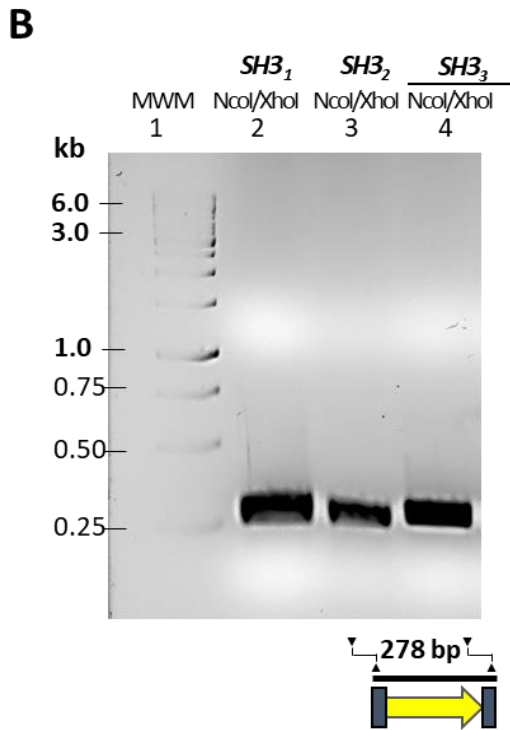
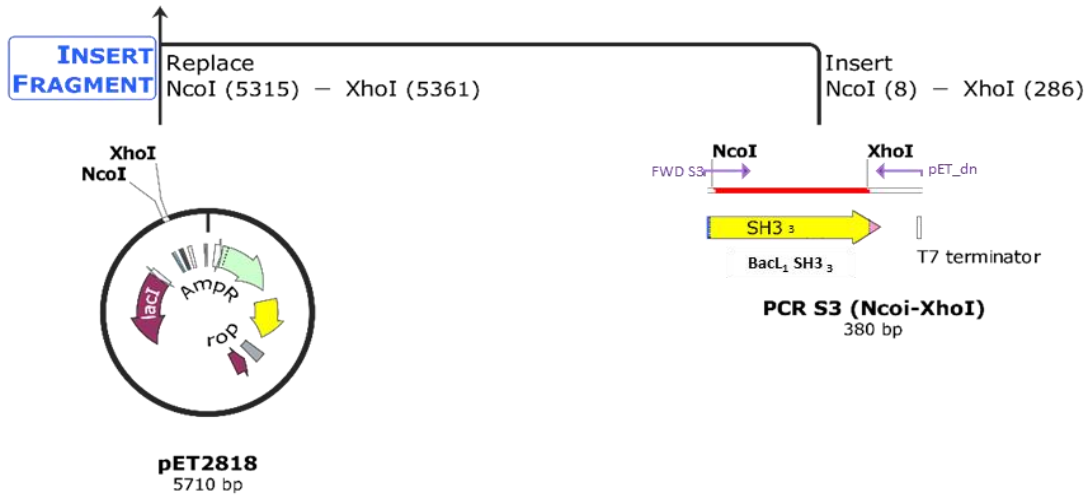
**D**



**Figure 6.4. Construction of pET2818-S2. (A)** Schematic representation of the cloning strategy. A 292 bp PCR fragment encoding the second BacL<sub>1</sub> SH3<sub>5</sub> repeat (SH3<sub>2</sub>) was cloned into pET2818 using NcoI and XhoI sites. **(B)** Restriction digest of the SH3<sub>2</sub> PCR fragment using NcoI and XhoI. Lane 1 (MW), molecular weight marker; lane 2, 272bp SH3<sub>1</sub> fragment digested with NcoI and XhoI; lane 3, 278bp SH3<sub>2</sub> fragment digested with NcoI and XhoI. **(C)** Restriction digest of pET2818 with NcoI and XhoI. Lane 1, (MW), molecular weight marker; lane 2 (Nat), undigested 5710 bp pET2818 plasmid; lane 3 (NcoI), 5710 bp pET-2818 digested with NcoI; lane 4 (XhoI), 5710 bp pET-2818 digested with XhoI; lane 5 (NcoI/XhoI), pET-2818 doubly digested with NcoI and XhoI. **(D)** Identification of pET2818-S2 by colony PCR using oligonucleotides PET\_up and pET\_dn. Lane 1, (MW), molecular weight marker; lane 2 to 4 (PET2818-S2), 458 bp PCR fragment corresponding to SH3<sub>2</sub>.

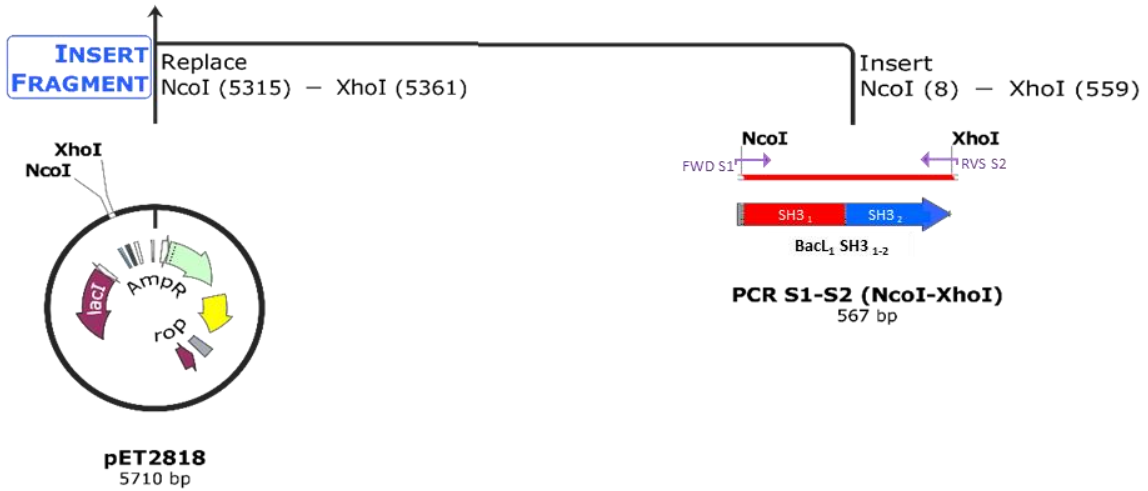
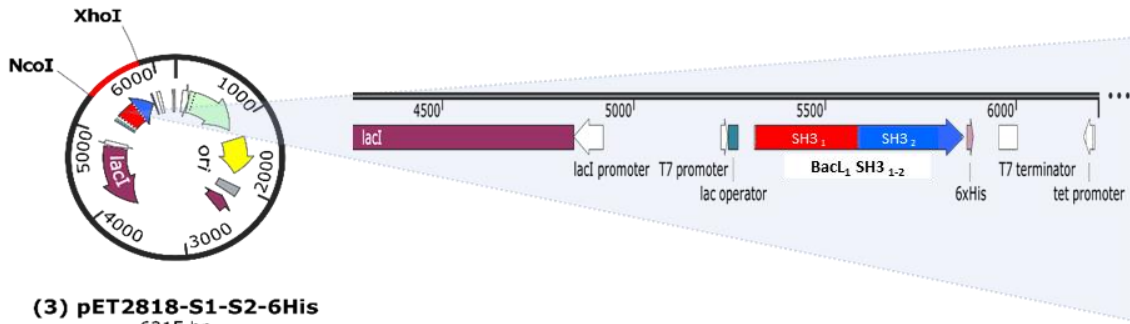
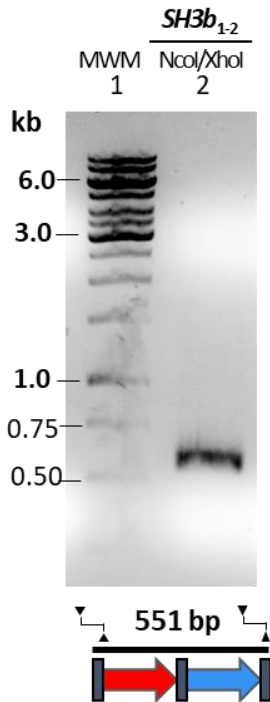
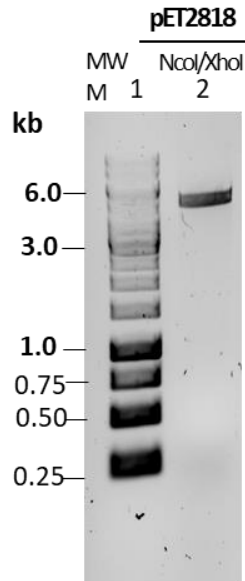
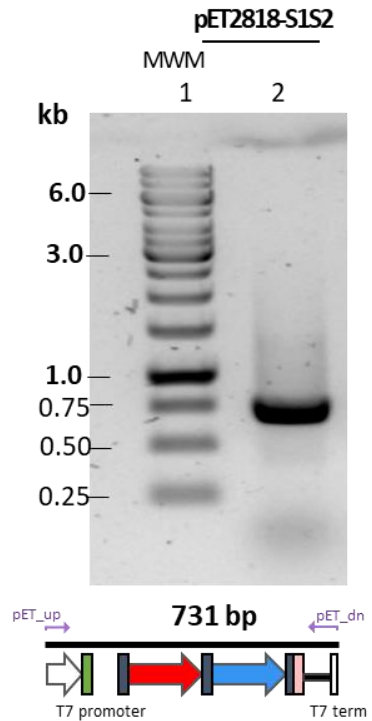


**(3) pET2818 S3-6His BacL**  
5942 bp

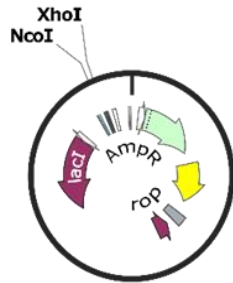
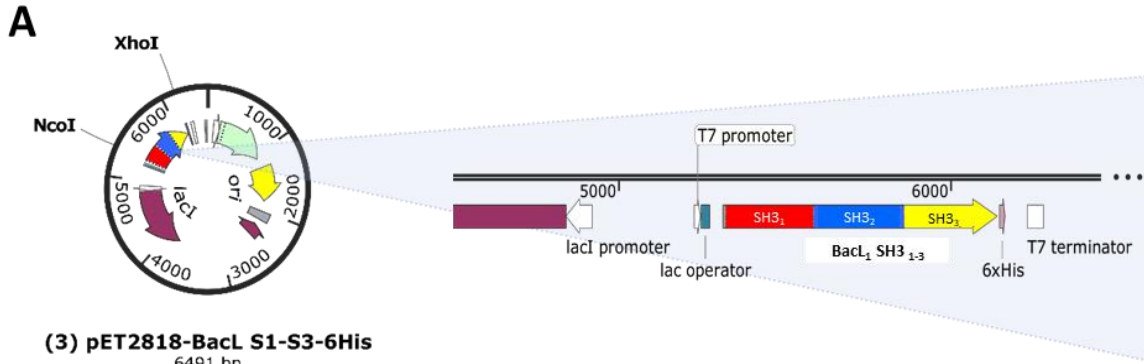


**Figure 6.5. Construction of pET2818-S3. (A)** Schematic representation of the cloning strategy. A 380 bp PCR fragment encoding the third BacL<sub>1</sub> SH3<sub>5</sub> repeat (SH3<sub>3</sub>) was cloned into pET2818 using NcoI and XhoI sites. **(B)** Restriction digest of the SH3<sub>3</sub> PCR fragment using NcoI and XhoI. Lane 1 (MW), molecular weight marker; lane 2, 272bp SH3<sub>1</sub> fragment digested with NcoI and XhoI; lane 3, 278bp SH3<sub>2</sub> fragment digested with NcoI and XhoI; lane 4, 278bp SH3<sub>3</sub> fragment digested with NcoI and XhoI. **(C)** Restriction digest of pET2818 with NcoI and XhoI. Lane 1, (MW), molecular weight marker; lane 2, (NcoI/XhoI), pET-2818 doubly digested with NcoI and XhoI. **(D)** Identification of pET2818-S3 by colony PCR using oligonucleotides PET<sub>up</sub> and pET<sub>dn</sub>. Lane 1, (MW), molecular weight marker; lane 2 (PET2818-S3), 458 bp PCR fragment corresponding to SH3<sub>3</sub>.

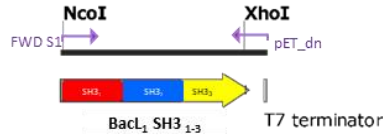


**A****B****C****D**

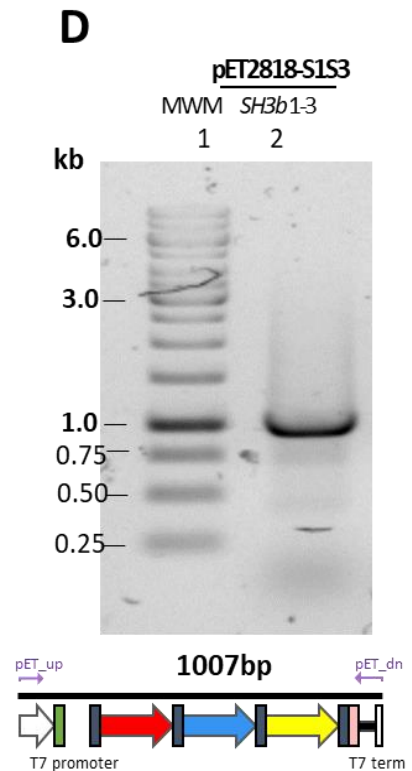
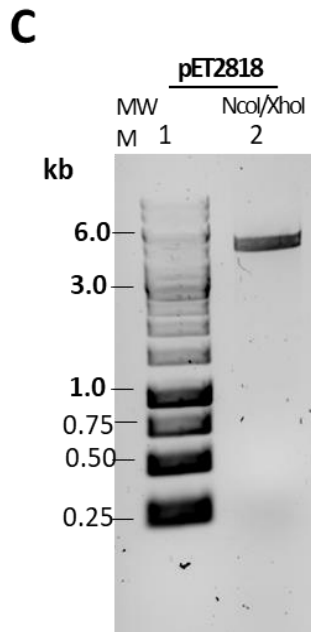
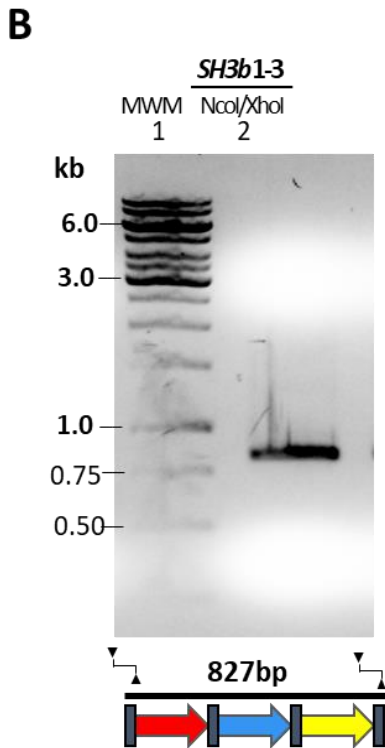
**Figure 6.6. Construction of pET2818-S1S2. (A)** Schematic representation of the cloning strategy. A 567 bp PCR fragment encoding the first and second BacL<sub>1</sub> SH3\_5 repeats (SH3<sub>1-2</sub>) was cloned into pET2818 using NcoI and XhoI sites. **(B)** Restriction digest of the SH3<sub>1-2</sub> PCR fragment using NcoI and XhoI. Lane 1 (MW), molecular weight marker; lane 2, 551bp SH3<sub>1-2</sub> fragment digested with NcoI and XhoI. **(C)** Restriction digest of pET2818 with NcoI and XhoI. Lane 1, (MW), molecular weight marker; lane 2, (NcoI/XhoI), pET-2818 doubly digested with NcoI and XhoI. **(D)** Identification of pET2818-S1S2 by colony PCR using oligonucleotides pET\_up and pET\_dn. Lane 1, (MW), molecular weight marker; lane 2 (pET2818-S3), 731 bp PCR fragment corresponding to SH3<sub>1-2</sub>.



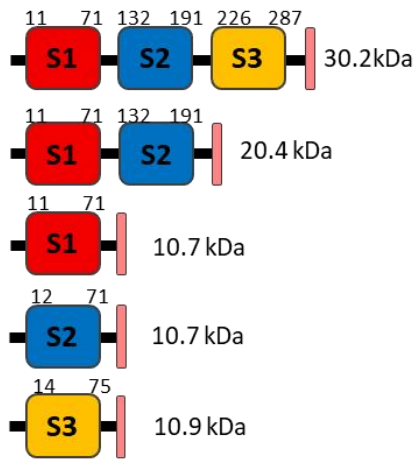
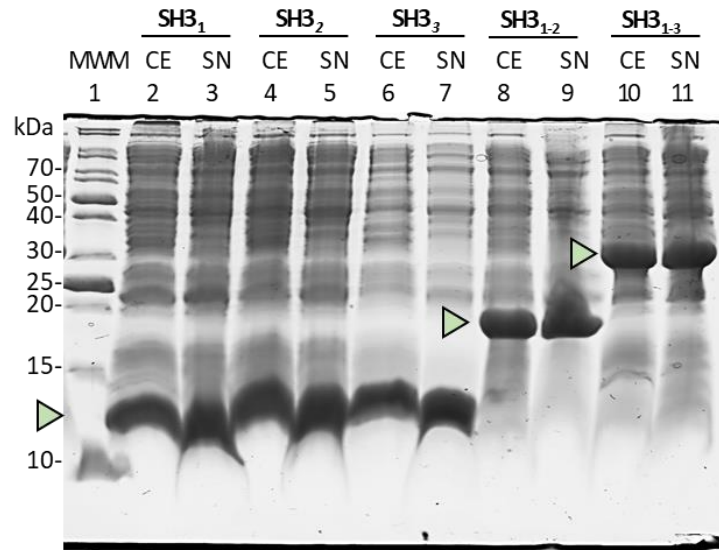
pET2818  
5710 bp



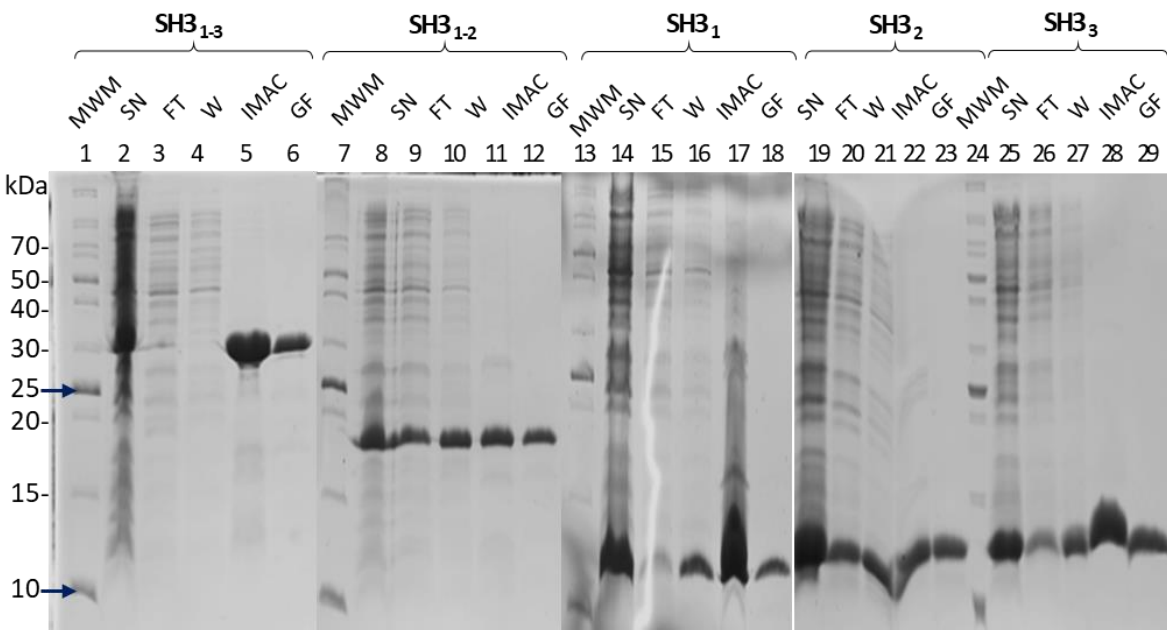
PCR from pET22b [S1-S3 (NcoI-XhoI)]  
929 bp



**Figure 6.7. Construction of pET2818-S1S3. (A)** Schematic representation of the cloning strategy. A 929 bp PCR fragment encoding the three BacL<sub>1</sub> SH3<sub>5</sub> repeats (SH3<sub>1-3</sub>) was cloned into pET2818 using NcoI and XhoI sites. **(B)** Restriction digest of the SH3<sub>1-3</sub> PCR fragment using NcoI and XhoI. Lane 1 (MW), molecular weight marker; lane 2, 827 bp SH3<sub>1-3</sub> fragment digested with NcoI and XhoI. **(C)** Restriction digest of pET2818 with NcoI and XhoI. Lane 1, (MW), molecular weight marker; lane 2, (NcoI/XhoI), pET-2818 doubly digested with NcoI and XhoI. **(D)** Identification of pET2818-S1S3 by colony PCR using oligonucleotides pET\_up and pET\_dn. Lane 1, (MW), molecular weight marker; lane 2 (pET2818-S3), 1007 bp PCR fragment corresponding to SH3<sub>1-3</sub>.

**A****C****B**

10 20 30 40 50 60 70 80 90 100  
 S1 MGSKGD SVN---PSACVVEYPSMRLPVS GDTDPNSPALDYYEAGQAIIVYDSYVFANGYAWI SYVAGSGLRRYVAVGPDDGR TD TVWGGGFLNN---LEHHHHHH  
 S2 MGSNTGS---ALSCVYFYSMRLPVS GDTDPNSPALDYYEAGQAIIVYDSYVFANGYAWI SYVAGSGLRRYVAVGPDDGR TD TVWGGGFLNNGDLEHHHHHH  
 S3 MGSQAH PNSIGLVPKACNFVFNRLPVSADTDPNSAALDYYEAGQSIGVDSYVFANGYAWI SYVAGSGLRRYVAVGPDDGR TD TVWGGGFFN---LEHHHHHH

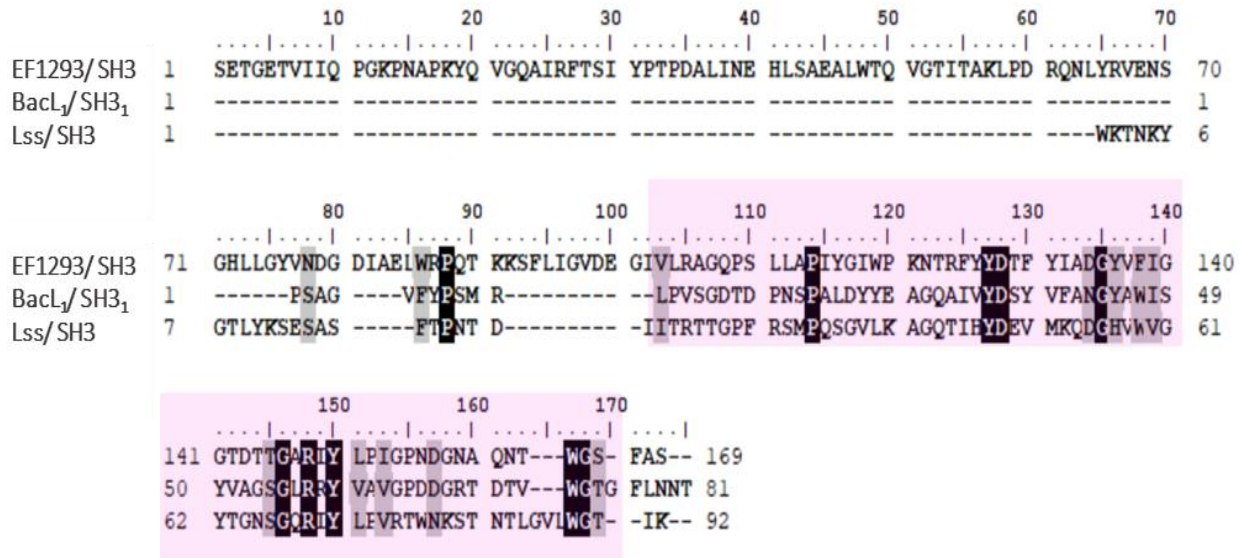
**D**

**Figure 6.8. Production and purification of BacL<sub>1</sub> SH3-5 domains.** **(A)** Domain organization of *E. faecalis* BacL<sub>1</sub> recombinant SH3\_5 domains produced in this study. Amino acid numbers refer to the transition between SH3\_5 domains. The expected molecular weight of each protein is shown on the right. **(B)** Sequence alignment of the three SH3\_5 modules present in the C-terminal domain of the bacteriocin protein BacL<sub>1</sub>. Identical residues are highlighted in black and residues in light grey indicate conservative substitutions. **(C)** SDS-PAGE analysis of the expression level and solubility of five BacL<sub>1</sub> SH3\_5 recombinant proteins described in **A**; proteins were produced in *E. coli* Lemo21(DE3) cells. Lane 1 (MWM), molecular-weight makers; lane 2 and 3, SH3<sub>1</sub>; lane 4 and 5, SH3<sub>2</sub>; lane 6 and 7, SH3<sub>3</sub>; lane 8 and 9, SH3<sub>1-2</sub>; lane 10 and 11, SH3<sub>1-3</sub>; (CE), crude extract; (SN), soluble fraction. **(D)** SDS-PAGE analysis of the purification steps to obtain the five BacL<sub>1</sub> SH3\_5 recombinant proteins. Lane 1 (MWM), molecular weight makers; lane 2 to 6, SH3<sub>1-3</sub>; lane 7 (MWM), molecular weight makers; lane 8 to 12, SH3<sub>1-2</sub>; lane 13 (MWM), molecular weight makers; lane 14 to 18, SH3<sub>1</sub>; lane 19 to 23, SH3<sub>2</sub>; lane 24 (MWM), molecular weight makers; lane 25 to 29, SH3<sub>3</sub>; (SN), soluble fraction; (FT), flow-through; (W), wash; (IMAC), elution pool; lane 6 (GF), pooled fractions from gel filtration step.

### 6.2.2 Expression of EF1293 SH3\_5 domain

The EF1293 protein consists of an N-terminal amidase domain and C-terminal SH3\_5 domain. It is unclear where the precise N-terminus (start) of the SH3\_5 protein is located within the C-terminal region of the PG hydrolase since the predicted SH3-like fold is preceded by a  $\beta$ -rich region. Protein homology predictions indicate a strong sequence similarity between the Lss and the EF1293 SH3\_5 domains corresponding to the last seven  $\beta$ -strands of these proteins. This sequence similarity starts about 20 amino acids into the Lss SH3\_5 domain (Fig. 6.9). This was somewhat expected since the N-terminal extension of the staphylococcal SH3\_5 domain is involved in binding to the pentaglycine cross-bridges (Lu, 2006; Sabala, 2019). In the *E. faecalis* BacL<sub>1</sub> SH3<sub>1</sub> protein (first repeat), this region of homology started more proximally, about 10 amino acids into the BacL<sub>1</sub> protein (Fig. 6.9). All factors considered, it seemed likely that the more distal seven  $\beta$ -strands of the EF1293 protein formed the SH3\_5 structure (Fig. 6.10 A), although the possibility that the C-terminal domain is adorned with extra secondary structures from the preceding  $\beta$ -rich region cannot be excluded.

Three alternative N-terminal sequences were selected as candidates to express and produce the EF1293 SH3\_5. A plasmid encoding the full length EF1293 protein (pTetH-1293 plasmid, available in the laboratory collection, (Fig. 6.10 B) was used as a PCR template to build expression vectors. The selected sequences were: (i) A 296 bp fragment encoding the last eight  $\beta$ -strands of the EF1293 protein (EF1293<sub>1</sub>; residues N274-S365); (ii) a 314 bp fragment encoding the last nine  $\beta$ -strands of the EF1293 protein (EF1293<sub>2</sub>; residues H268-S365); and (iii) a 474 bp fragment encoding the complete  $\beta$ -rich sequence (EF1293<sub>3</sub>; residues K214-S365, (Fig. 6.10 C). The three fragments were digested with NcoI and BamHI and cloned into a similarly digested pET2818 vector. All plasmids encode proteins with a C-terminal 6x-his-tag (Fig. 6.11). The recombinant expression plasmids (pET2818-EF1293<sub>1</sub>, pET2818-EF1293<sub>2</sub>, and pET2818-EF1293<sub>3</sub>) were identified by colony PCR and sent for sequencing to ensure the absence of mutations in the inserts (Fig. 6.12).



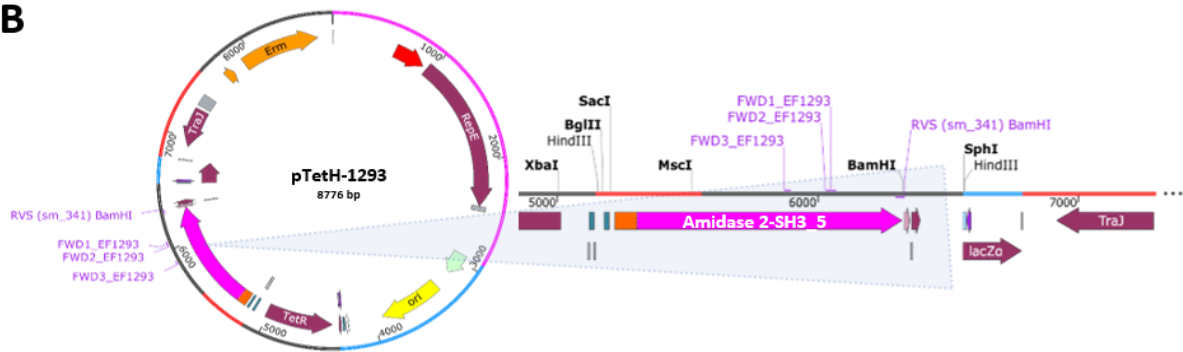
**Figure 6.9. Amino acid sequence alignment between the C-terminal region of the *E. faecalis* EF1293 protein, the *E. faecalis* N-terminal BacL<sub>1</sub> SH3<sub>1</sub> repeat, and the *S. aureus* Lss SH3<sub>5</sub> domain.** The region of homology between the EF1293 C-terminal region; BacL<sub>1</sub> SH3<sub>1</sub>; and Lss SH3<sub>5</sub> is highlighted in pink. Conserved residues among all proteins are highlighted in black and conservative substitutions are in light grey.



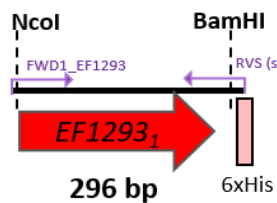
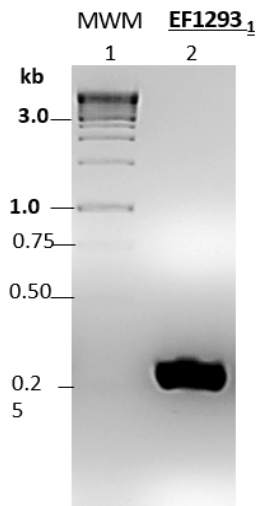
**A**



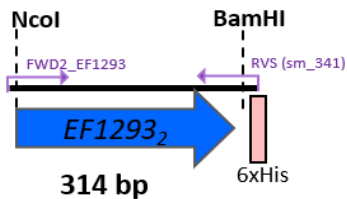
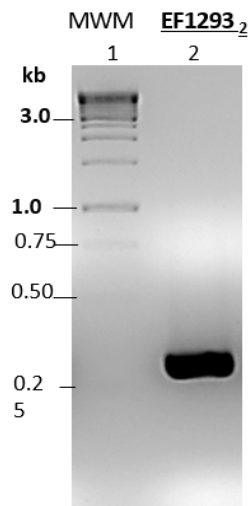
**B**



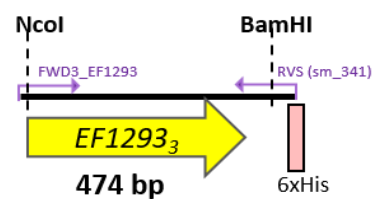
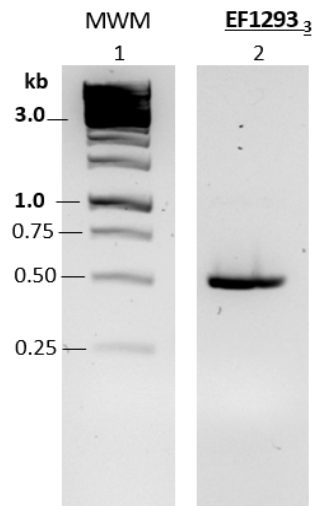
**C**



**D**

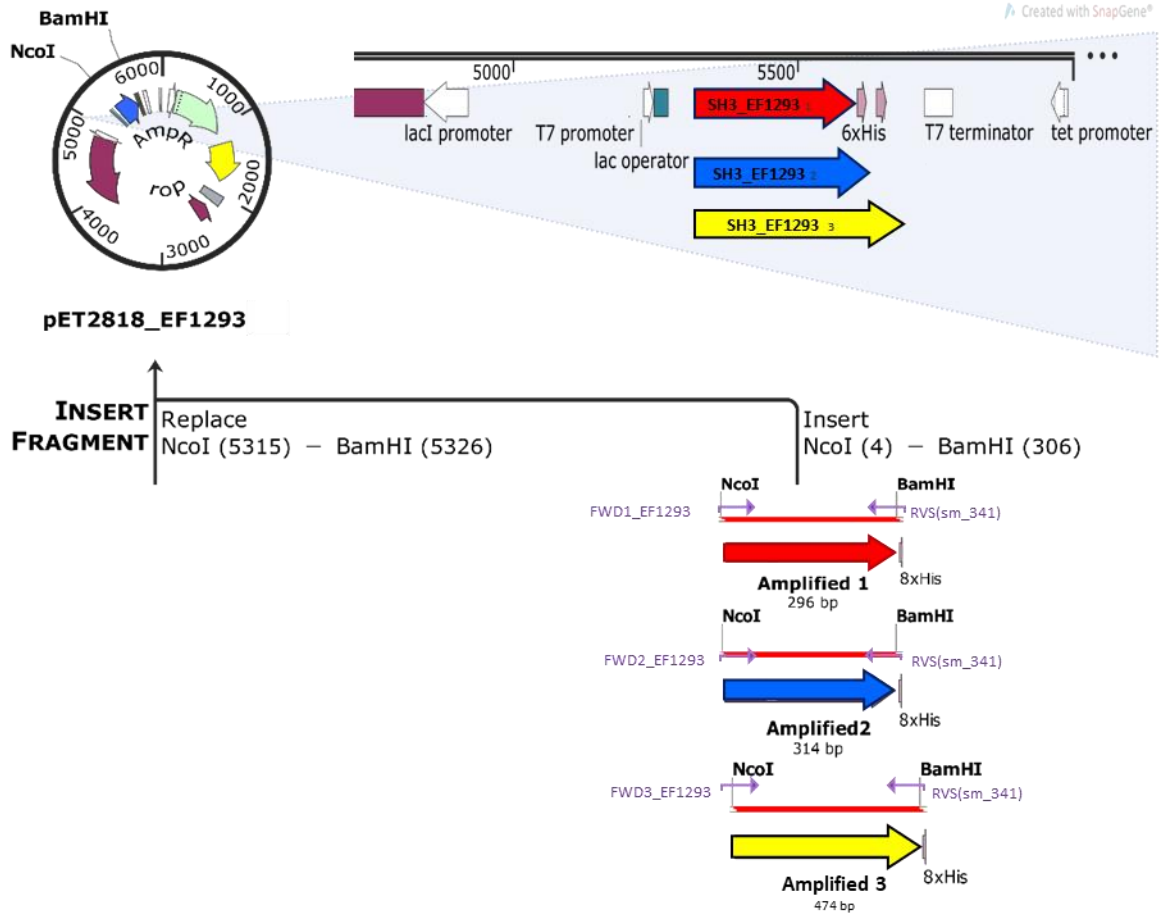


**E**

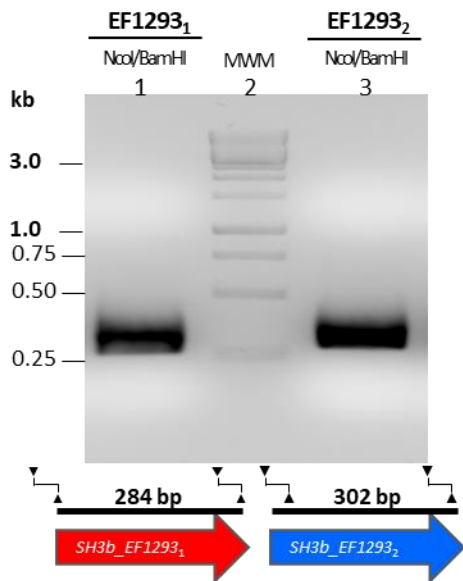


**Figure 6.10. Selection of *E. faecalis* EF1293 SH3\_5 domain candidates.** **(A)** Secondary structure prediction of the EF1293 endolysin from *E. faecalis*. The sequence analysis was generated using the Phyre2 web portal for protein modelling, prediction and analysis. Secondary structures are indicated by arrows ( $\beta$ -strands) and ribbons ( $\alpha$ -helix). Coloured boxes correspond to the three N-terminal candidates of the EF1293 SH3\_5 domain: EF1232<sub>1</sub> red; EF1293<sub>2</sub>, blue, and EF1293<sub>3</sub>, yellow **(B)** Schematic representation of pTetH-1293 plasmid encoding the EF1293 endolysin used for the amplification of three SH3\_5 domains with alternative N-terminal sequences. **(C)** PCR product encoding EF1293<sub>1</sub>; lane 1, (MW), molecular weight marker; lane 2, 296 bp *EF1293*<sub>1</sub> DNA fragment flanked with NcoI and BamHI sites. **(D)** PCR product encoding EF1293<sub>2</sub>; lane 1, (MW), molecular weight marker; lane 2, 314 bp *EF1293*<sub>2</sub> DNA fragment flanked with NcoI and BamHI sites **(E)** PCR product encoding EF1293<sub>3</sub>; lane 1, (MW), molecular weight marker; lane 2, 474 bp *EF1293*<sub>3</sub> DNA fragment flanked with NcoI and BamHI sites.

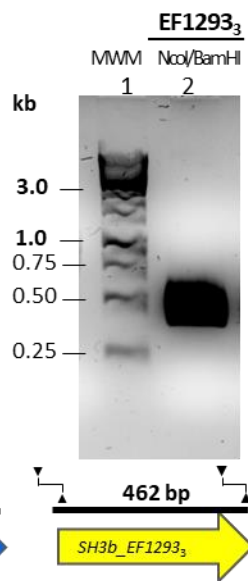
**A**



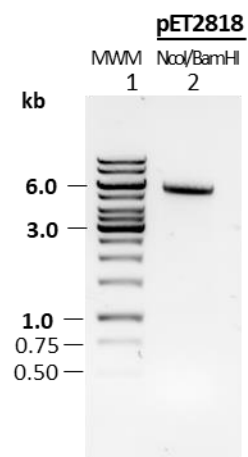
**B**



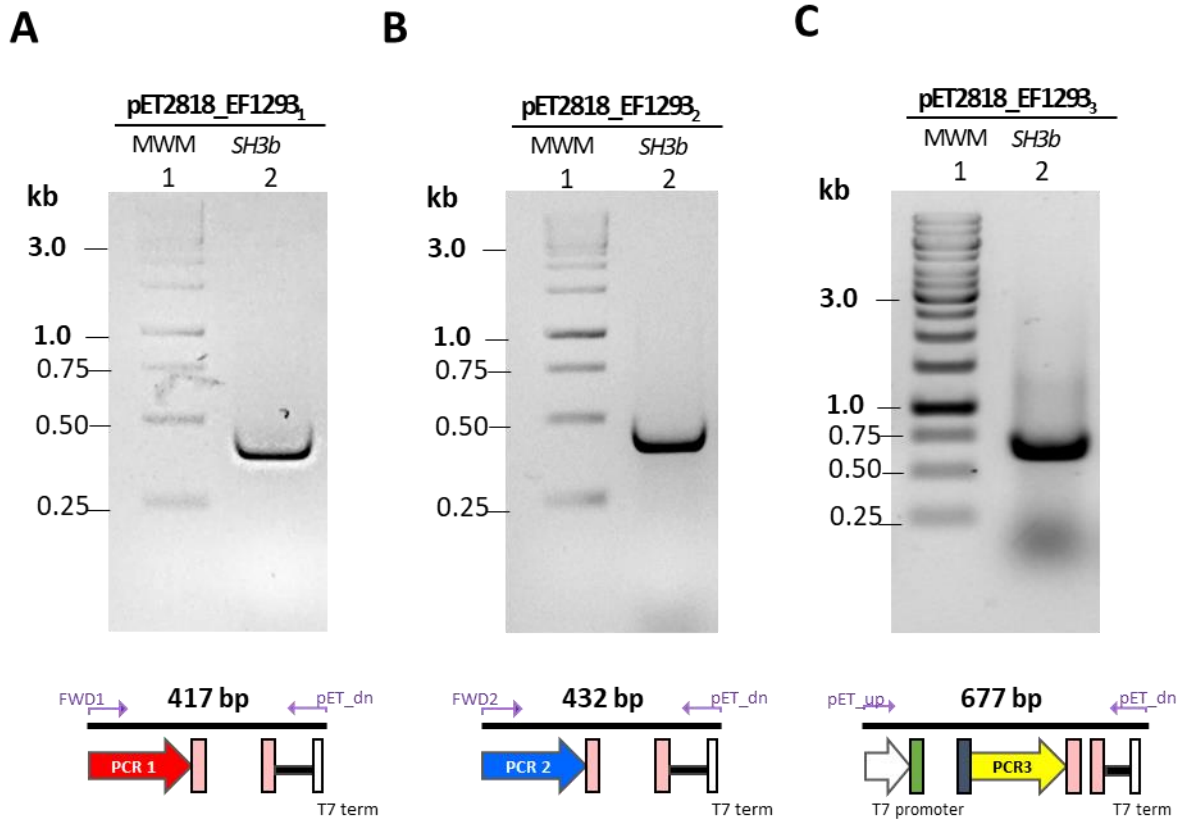
**C**



**D**



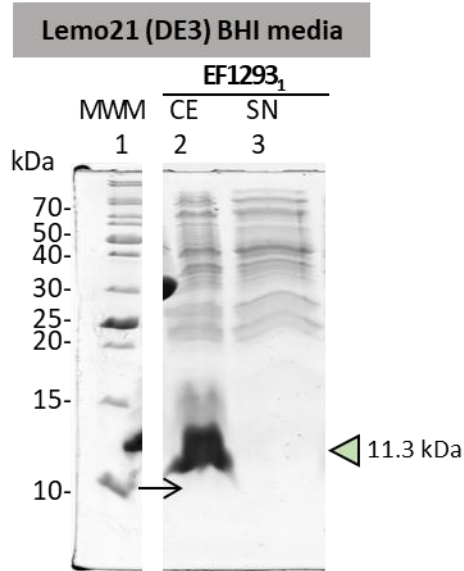
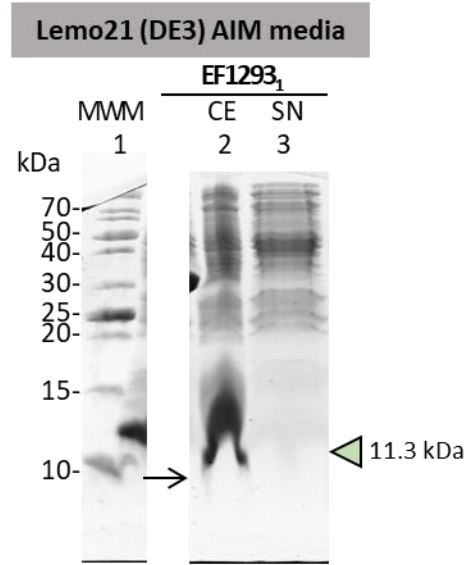
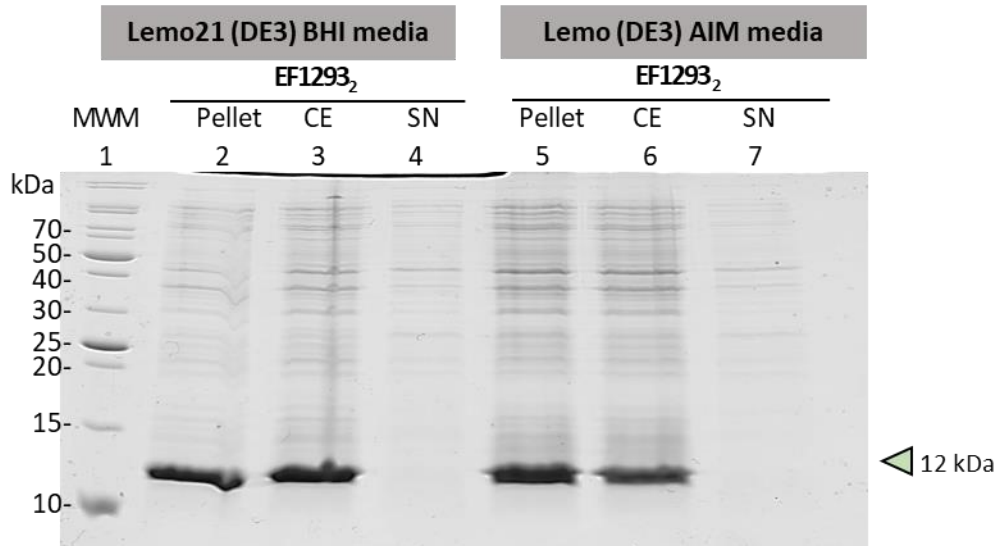
**Figure 6.11. Construction of pET2818-EF1293 expression plasmids. (A)** Schematic representation of the cloning strategy. Three distinct DNA fragments encoding the EF1293 SH3\_5 domain flanked with NcoI and BamHI sites were cloned into pET2818. **(B)** Restriction digest of the PCR fragments EF1293<sub>1</sub> and EF1293<sub>2</sub>. Lane 1, 248bp EF1293<sub>1</sub> fragment digested with NcoI and BamHI; lane2 (MW), molecular weight marker; lane 3, 302 bp EF1293<sub>2</sub> fragment digested with NcoI and BamHI. **(C)** Restriction digest of the PCR fragment EF1293<sub>3</sub>. Lane 1 (MW), molecular weight marker; lane 2, 462 bp EF1293<sub>3</sub> fragment digested with NcoI and BamHI. **(D)** Restriction digest of pET2818. Lane 1, (MW), molecular weight marker; lane 2, 5710 bp pET-2818 plasmid digested with NcoI and BamHI.



**Figure 6.12. PCR confirmation of pET2818-EF1293 expression plasmids. (A)** Confirmation of pET2818-EF1293<sub>1</sub> by colony PCR. Lane 1, (MW), molecular weight marker; lane 2, 417 bp PCR product amplified with oligos FWD1\_EF1293 and pET\_dn . **(B)** Confirmation of pET2818-EF1293<sub>2</sub> by colony PCR. Lane 1, (MW), molecular weight marker; lane 2, 432 bp PCR product amplified using oligos FWD2\_EF1293 and pET\_dn. **(C)** Confirmation of pET2818-EF1293<sub>3</sub> by colony PCR. Lane 1, (MW), molecular weight marker; lane 2, 677 bp PCR product using pET-up and pET\_dn oligos.

Small-scale expression trials were performed to determine the expression level and solubility of the EF1293 constructs (Fig. 6.13). Due to the lockdown caused by Covid-19 all experiments were stopped and only the production of EF1293<sub>1</sub> and EF1293<sub>2</sub> proteins was assessed. Both recombinant proteins were produced in *E. coli* Lemo21(DE3) using two different media: BHI medium and Overnight Express™ Instant TB medium. Cultures grown in BHI broth were IPTG-induced (1 mM) at OD<sub>600</sub>=0.6 and grown alongside the cultures in AIM at 25°C for 16 h. Cells and supernatant samples were harvested and analysed by SDS-PAGE (Fig. 6.13 A-C). None of the SH3\_5 recombinant proteins were soluble under standard expression conditions in either of the conditions tested (Fig. 6.13). Proteins at the expected molecular weight (11.3 and 12 kDa for EF1293<sub>1</sub> and EF1293<sub>2</sub>, respectively), were only observed in the crude extract of each culture (Fig 6.13).

As previously mentioned, owing to the circumstances it was not possible to further analyse the expression of the EF1293<sub>3</sub> recombinant protein.

**A****B****C**

**Figure 6.13. SDS-PAGE analysis of the expression level and solubility of the EF1293<sub>1</sub> and EF1293<sub>2</sub> recombinant proteins. (A)** Small-scale expression of EF1293 (11.3 kDa) produced in *E. coli* Lemo21(DE3) cells grown in BHI medium at 25°C for 16 h after IPTG induction. Lane 1 (MWM), molecular-weight makers; lane 2 (CE), crude extract; lane 3 (SN), soluble fraction. **(B)** Small-scale expression of EF1293<sub>1</sub> (11.3 kDa) produced in *E. coli* Lemo21(DE3) cells grown in Overnight Express™ Instant TB medium at 25°C for 16 h. Lane 1 (MWM), molecular-weight markers; lane 2

(CE), crude extract; lane 3 (SN), soluble fraction. **(C)** Small-scale expression of EF1293<sub>2</sub> (12 kDa). Lane 1 (MWM), molecular-weight markers; lane 2 to 4, EF1293<sub>2</sub> produced in *E. coli* Lemo21(DE3) cells grown in BHI medium at 25°C for 16 h after IPTG induction; lane 5 to 7, EF1293<sub>2</sub> produced in Lemo21(DE3) cells grown in Overnight Express™ Instant TB medium at 25°C for 16 h; (CE), crude extract; lane 4 (SN), soluble fraction.



### 6.3 Perspectives

To explore *E. faecalis* PG recognition by the BacL<sub>1</sub> SH3\_5 domain and investigate the role of domain modularity, a combination of NMR, isothermal titration calorimetry (ITC), circular dichroism (CD) spectroscopy, and crystallography experiments will be required. A first step will be to express doubly labelled proteins. Several experiments will have to be performed. These include standard heteronuclear triple resonance experiments (HNCACB and HN(CO)CACB) to assign the spectra corresponding to the different proteins analysed. Protein-ligand titrations (<sup>15</sup>N HSQC) will also be required to map the residues involved in substrate recognition

NMR titrations with labelled SH3\_5 domains and four key peptide ligands already available in our laboratory could have been sufficient to determine the PG motif(s) recognised by the BacL<sub>1</sub> SH3\_5 domain. The ligands planned to be used in this study were:

- An L-Ala dipeptide cross-bridge [L-Ala-L-Ala], [**L-A2**].
- A D-Ala dipeptide cross-bridge [D-Ala-D-Ala], [**D-A2**].
- A pentapeptide stem [-AQKAA], [**P5**].
- A pentapeptide stem bound to a dipeptide L-Ala lateral chain [A $\gamma$ QK(L-Ala-L-Ala)AA], [**P5-L-A2**]

The time-length and number of NMR titrations depends on the protein-ligand biomolecular interactions: a standard <sup>15</sup>N HSQC titration takes approximately 2.5 hours per titration point and typically requires eight titrations. A combination of ITC and CD spectroscopy would allow the individual contribution of SH3\_5 repeats to PG binding to be explored and enable the study of protein conformational behaviour in solution alongside the NMR analyses. This could have been done by performing a comparative analysis between the SH31, SH3\_51-2, and SH31-3 proteins, both in the presence and in the absence of a PG ligand identified by NMR and/or ITC. Despite the high sensitivity of ITC, this method might not be suitable if, like in the case of the Lss SH3\_5 domain, the BacL<sub>1</sub> SH3\_5 repeats exhibit relatively low binding affinities to their ligands. This is because weak binding interactions lead to small enthalpy changes (released heat), resulting in ITC outputs having relatively low signal to noise.

Altogether, the outlined experiments would have taken approximately four to six months and fitted within the established time frame of my PhD. With most of the experimental methods already standardized, it seemed plausible to have provided enough information about the contribution of individual modules to the binding of *E. faecalis* PG as well as the minimal PG ligand. This could be recognized by either individual modules that do not interact with each other and contribute to additive binding as seen for *E. faecalis* AtIA LysM modules and *M. tuberculosis* PknB PASTA domains (Mesnage, 2014; Mir, 2011) or a quaternary structure necessary to generate a functional protein as determined for *S. pneumoniae* PBP2x PASTA domains (Yeats, 2002).

## CHAPTER VII

### General discussion

Protein domains are classified into fold families which are thought to share functional and structural features often resulting from a common origin (Hartwell, 1999; Wang, 2009). The independent evolution of protein domains is a major drive for the diversification of protein functions (Hartwell, 1999; Vogel, 2004) and how protein domain organization contributes to the creation, recruitment, and diversification of biological functions that underpin life is a fundamental question (Caetano-Anollés, 2009; Murzin, 1995; Vogel, 2004; Wang, 2009). One of the most studied domains to answer this question are SH3s. SH3 domains are found in proteins distributed across different kingdoms and are part of multimodular proteins exhibiting a wide range of functions.

#### **7.1 The SH3 domain as a model to investigate functional and structural diversification during evolution**

Eukaryotic SH3 domains represent a model system for the study of protein-ligand interactions. They are found in proteins playing critical roles in a wide variety of biological processes such as cell proliferation, migration, and cytoskeletal modifications (e.g. tyrosine kinases, myosin, and cortactin) (Mayer, 1998). SH3 domains are also found in a large number of prokaryotic proteins and represent one of the most common domains. Since SH3 proteins are structurally conserved across kingdoms (Lu JZ, 2006; Ponting, 1999; Saksela, 2012; Whisstock, 1999), their evolutionary timelines provide important information about their structure/function relationships (te Velthuis, 2007; Wang, 2009).

Over the last decade, proteomic evolutionary timelines have been generated based on 3D structures since fold family domains generally share only moderate sequence similarity (Caetano-Anollés, 2003; Wang, 2009; Yang, 2005). Atomic structures and domain organization are far more

conserved than amino acid sequences and have remained well-preserved for long periods of time (Fukami-Kobayashi K, 2007). 3D architectures carry evolutionary information that has revealed three main eras in proteomics evolution: (i) architectural diversification, (ii) super kingdom speciation, (iii) organismal diversification (Caetano-Anollés, 2009; Wang, 2009).

SH3 proteins are considered highly promiscuous because they are found in >132 domain combinations in eukaryotes (Wang, 2009). Evolutionary timelines place SH3 proteins in early architectural diversifications and show their involvement in several combinatorial events during super kingdom speciation (Wang, 2008; te Velhuis, 2007). It is likely that their ability to establish multiple inter and intramolecular interactions has allowed them to evolve moderate selectivity towards a large variety of ligands, as determined for the numerous eukaryotic and bacterial SH3-like families (Desvaux, 2018; Gonzalez-Delgado, 2020; Saksela, 2012). Interestingly, a massive domain combinatorial event coincides with the establishment of organismal lineages. This event marked the period where multifunctional architectures were replaced by highly specialized counterparts, a process that is clearly seen for SH3s (Caetano-Anollés, 2009; Kim, 2006; Wang, 2009).

## **7.2 Does SH3\_5 domain organization reflect the diversity of binding mechanisms?**

According to Pfam (as of June, 2020), 364 sequences with an SH3\_5 domain are organised into 65 protein domain organizations (El-Gebali, 2019). Both the domain organisation of SH3\_5 proteins and the organisation of the SH3\_5 domains themselves are very diverse, suggesting that during evolution, proteins have used the SH3\_5 domain as a building block to diversify protein functions.

### **7.2.1 Modularity of proteins containing SH3\_5 domains**

Most SH3\_5 domains are associated with one or two other domains (54% and 36% of the proteins, respectively). Only five proteins are made of four domains including an SH3\_5 domain (Fig. 1.16, Chapter I). This domain organisation amongst SH3\_5 proteins is in agreement with the results published in a study searching for the number of domains present in 4636 protein

combinations. These proteins were analysed at a family fold level and it was shown that 64% of them were comprised of two domains, 18% had three domains, and the rest were similarly distributed either by single domains or more than three domains (Wang, 2009).

The domain organization of SH3\_5 proteins matches evolutionary models suggesting the selective advantages of preserving short, multimodular proteins (Brocchieri, 2005; Kurland, 2007). Relatively short proteins (with fewer domains and hence a limited modularity) seem less costly to preserve than exceedingly long protein sequences that could enhance the likelihood of events such as fission or intragenomic shuffling. Although bacteria have acquired domains from other kingdoms by horizontal gene transfer, it is through the duplication, insertion, and deletion of genes that proteins mostly acquire or lose domains to generate new combinations. These processes are referred to as domain shuffling (Gomperts, 2009; Wang M, 2009).

### **7.2.2 Modularity of the SH3\_5 domain**

SH3\_5 domains can contain one to six repeats, but a clear preference for a single SH3\_5 repeat is observed. From the 364 total protein sequences, 85% of them harbour a single SH3\_5 repeat. Only 5% of the protein sequences have two SH3\_5 repeats, 3.5% have three repeats, 3% have 4 repeats, 3% have 5 repeats, and only 2 sequences (0.5%) have 6 repeats.

The fact that most proteins contain a single SH3\_5 repeat indicates that the binding mechanism mediated by these domains must be highly efficient. The protein evolution of Lss has preserved this situation to allow for a more dynamic binding to PG (as detailed below). On the other hand, the existence of multiple repeats in SH3\_5 domains may reflect distinct cell wall binding mechanisms in a similar way to what has been described for the LysM domain. It is possible that multiple repeats provide a stronger binding activity as in the case of the LysM domain from *E. faecalis* AtIA (Mesnage, 2014). The AtIA LysM domain is made of 6 LysM repeats all able to bind to PG independently. They have been proposed to be organised as “beads on a string”, contributing to substrate recognition in a cooperative manner. Another possibility is that in some proteins, SH3\_5 domains form a quaternary structure (e.g. a dimeric, trimeric or tetrameric

SH3\_5 structure) that binds optimally to PG. It has been shown that *Cladosporium fulvum* LysM domain effector, Ecp6, can dimerize and generate an ultrahigh affinity binding site ( $K_d=280$  pM) recognising chitin (Sanchez-Vallet, 2013). In some eukaryotic LysM proteins, disulphide bonds between LysM repeats form a trimeric structure that have been proposed to maintain the structural integrity of the LysM domain (Liu, 2012).

Interestingly, a relatively large number of proteins are exclusively made of SH3\_5 repeats (30 proteins contain a single SH3\_5 repeat and 3 proteins were found to contain 2, 3 and 5 repeats). This result seems surprising, but a similar phenomenon has been described with proteins exclusively made of LysM repeats (Willmann, 2011). These LysM proteins are produced by fungi and bind soluble chitin. They have been shown to bind fungal cell wall fragments generated by host chitinases in response to a phytopathogen infection. These LysM binding proteins have been proposed to “hide” the fungal cell wall fragments from the host immune system. It is therefore possible that bacteria have evolved a similar mechanism with SH3\_5 domains that could bind PG fragments and hide them from the host immune system to limit the inflammatory response.

### **7.3 Lss SH3\_5 binding affinities**

Lss is a powerful lytic enzyme highly specific to staphylococci with a pentaglycine crossbridge (Baba, 1996; Lu JZ, 2006; Thumm, 1997). The presence of a single SH3\_5 domain in this protein might seem counterintuitive since one might predict it to have a weak binding affinity for PG (Gonzalez-Delgado, 2020; Tossavainen, 2018). Low binding affinities have been previously described in the SH3 eukaryotic counterparts (Mayer, 2001; Tossavainen, 2018). At the start of this work, two possibilities were envisaged: (i) a high binding affinity and a high specificity or (ii) like in eukaryotic SH3 proteins, a weak affinity and a moderate specificity. Based on the very powerful activity of Lss and its high specificity for PG containing pentaglycine crossbridges, we expected that the SH3\_5 targeting domain would have a very high affinity and specificity for the PG pentaglycine crossbridges.

Interestingly, the CSP analyses revealed rather low binding affinities for soluble monomeric PG fragments (millimolar range) and as the complexity of the ligands increased, a 10x improvement in binding affinities was determined. We were also able to observe that an increased level of ligand complexity led to the formation of protein aggregates. Together, these results were indicative of the ability of a single domain to mediate highly efficient binding interactions with cross-linked PG fragments.

#### **7.4 Lss SH3\_5 binding specificity**

As previously mentioned, Lss displays a restrictive lytic activity against *Staphylococcus* spp. strains with a PG crosslinked by pentaglycine crossbridges. Besides the strength of the Lss SH3\_5-PG binding affinities, another unanswered question that we investigated was the role of the SH3\_5 domain in the enzyme specificity. Since the discovery of the Lss PG substrate, it had been largely assumed that Lss exclusively recognised pentaglycine crosslinks. The results obtained in this study were unexpected since they showed that the SH3\_5 domain has evolved moderate selectivity to its substrate. This domain was shown to bind to a PG feature common to most Gram-positive bacteria, the tetrapeptide stem (AiQKA, P4) in addition to a specific feature of the staphylococcal PG (the pentaglycine crossbridge, G5). Moreover, the results obtained here also show that SH3\_5 can efficiently bind to PG lacking an intact pentaglycine crossbridge like those found in the *S. aureus* mutants *fem B* and *fem AB*, containing three and one glycine residues in their PG crossbridges, respectively.

#### **7.5 Lss SH3\_5 binding mechanism**

The NMR analyses of the SH3\_5-PG titrations revealed that the Lss SH3\_5 domain recognises both PG peptide moieties via distinct sets of residues located at opposite sides of the protein surface. The results were initially hard to reconcile with (i) the assumption that this domain was primarily binding to pentaglycine crosslinks via a binding groove made of the N-terminal extension (Lu JZ, 2006; Sabała, 2014) and (ii) the fact the contiguous peptide stem is too short to reach the identified P4 binding cleft located at the opposite side of the protein. Another puzzling observation we could not immediately explain was the fact that the G5 and P4 binding sites

represented two completely independent and discrete binding sites, with no other CSPs being observed between them. Another group published similar results but could not explain how the two PG motifs recognised (P4 and G5) could bind the SH3\_5 domain simultaneously (Tossavainen, 2018). At this point, we sought to determine the structure of the SH3\_5 domain in complex with a PG fragment comprised of both PG moieties (P4-G5).

The 1.4 Å co-crystal structure of the SH3\_5-[P4-G5] complex shed light onto the dual SH3\_5 binding mechanism whereby the G5 ligand from a monomeric fragment was recognized by one SH3\_5 domain and the contiguous P4 peptide was bound to another SH3 protein. Altogether, the NMR and crystallography results provided an explanation of the efficient mechanisms adopted by this potent enzyme. They allow us to explain an interesting phenomenon occurring in the CSPs from the PG-like fragments and to propose a model for the molecular mechanisms underpinning the Lss-PG binding, both of which are outlined below.

The greater SH3\_5 affinities to complex PG ligands carrying both P4-G5 moieties exhibited smaller CSPs. We hypothesized that the off-rates are fast and that the binding partners can be rapidly exchanged upon re-localization (Gonzalez-Delgado, 2020). This dual ligand recognition of the SH3\_5 domain is at the cost of suboptimal geometry i.e. having two binding pockets at opposite sides of the protein prevents simultaneous binding by the same molecule (monomeric fragment). This reduced time-average hydrogen bonding can be compensated for by acquiring a more dynamic binding (more entropy). Furthermore, single-site mutations displayed modest impact on PG binding and on the enzymatic activity, suggesting that the weaker enthalpy caused by this suboptimal geometry is compensated for by greater entropy (increased ligand mobility) governed by a large network of interactions.

The mature PG lattice is therefore more entropically favourable to satisfy the SH3\_5 binding requirements and highlights the physiological relevance of the results determined in this study. In other words, such plasticity is advantageous since it allows the SH3\_5 proteins to 'walk' across the PG without detaching from the surface by always having one site bound. A similar strategy



has been described in some celluloses that potentiate their activity by having two different cellulose-binding domains (Bolam, 1998; Gill, 1999; Nagy, 2007; Raghothama, 2001). Moreover, two distinct peptide-binding sites have been identified in the eukaryotic SH3 protein from the Ser/Thr mixed-lineage kinase 3 (MLK3) (Kokoszka, 2018). The binding pockets are located at opposite faces of the SH3 protein and were shown to interact with different viral peptides. The MLK3 SH3 domain binds the proline rich NS5A peptide from the Hepatitis C virus inhibiting the apoptosis of the infected cells (Macdonald, 2004; Tan, 1999). The non-canonical binding pocket is located between the Src loop and the edge of the  $\beta$ 3 strand, whilst the canonical proline-rich peptide from the hepatitis C virus binds in the groove made from three loops: RT, n-Src, and  $3_{10}$  helix loop (Kokoszka, 2018).

The binding mechanism proposed here is consistent with the aggregation observed during the titration of complex PG fragments. These titrations were within the slow exchange limits indicating a tighter binding. Taken together, these results suggest that protein clusters are formed and lead to an increase in the concentration of the enzyme at the PG surface, thereby enhancing the activity of the catalytic domains. In agreement with this hypothesis, a study using atomic force microscopy in *S. aureus* cells exposed to Lss revealed the creation of nanoscale perforations that preceded lysis (Francius, 2008). Interestingly, a similar mechanism has been recently described for the single LysM domain from the fungal wheat pathogen *Zymoseptoria tritici*. The crystal structure of this LysM domain revealed their ability to form a supramolecular structure through the induction of homodimers that are 'bridged' by their ligand (chitin). This property confers protection of *Z. tritici* from the host by secreting LysM effectors that scavenge the chitin oligomers that trigger the plant immune system (Sánchez-Vallet, 2020).

#### **7.6 Is the catalytic domain the major determinant of the Lss specificity?**

On the basis of the results obtained in this research, it is necessary to reconsider the catalytic activity rather than the binding as the major factor in the restricted selectivity of Lss to *Staphylococcal* species. Firstly, CSP analyses of the SH3-PG interactions showed that, although the SH3 domain displays similarly weak affinities to simpler G5 and P4 peptides, fewer

interactions occurred at the G5 site as the complexity of the ligands increased. These results suggest that SH3\_5 proteins display weaker interactions with ligands comprised of a more cell wall-like structure at the G5 site. Furthermore, the CSP analyses of the titrations with the single-site mutant domains revealed that the residues involved in the interaction with P4 exhibited the most pronounced decrease in binding affinities, as compared to the wild-type SH3\_5. These results were also consistent when assessing binding to purified PG sacculi and the enzymatic activity. We hypothesized that the SH3\_5 domain from Lss has evolved to favour binding to P4 peptide stems to allow for the catalytic domain to cleave its substrate (G5) as long as it still recognised both ligands and does not bind too tightly to either so it can remain attached to the PG lattice.

As previously mentioned for the moderate selectivity of the Lss SH3\_5 domain, most of the studied SH3\_5 mutations had a limited impact on binding to *S. aureus* WT PG, similar to that seen for the WT SH3\_5-mNG fusion to the PG *fem B* [GGG] mutant. In *S. carnosus*, the *fem B* mutation confers high resistance to the Lss enzyme with a 3000-fold increase in the MIC values from 0.01 to 32  $\mu\text{g}\cdot\text{mL}^{-1}$  (Nega, 2015). Furthermore, the WT SH3\_5-mNG fusion still bound in a dose-dependent manner to the PG *fem AB* [G] mutant.

Site-directed mutagenesis of three key amino acid residues of the Lss catalytic domain (H33,N37,H116) were shown to inhibit endopeptidase activity in turbidity assays (Bardelang, 2009; Sabała, 2014). Additionally, using modified fluorescence resonance energy transfer (FRET) substrates, the catalytic domain was not capable of cleaving fragments with an any amino acid substitution at position 3 of the pentaglycine peptide (GGXGG). However, a competition FRET assay showed that binding to a modified substrate (GGSGG) was not affected in the full-length Lss (Bardelang, 2009).

Altogether, these observations suggest that the restricted lytic activity of the Lss enzyme is conferred by its catalytic domain. However, previous experimental approaches using synthesized FRET peptides have highlighted the complexity of generating soluble complex fragments and the

need for more sensitive assays that lead to shorter experimental times to avoid the precipitation of substrates. The low affinities of Lss to its substrates have required the use of high concentrations of ligands leading to a significant decrease in fluorescent emissions caused by the inner filter effect. Although different strategies have been applied to compensate for these problems, the search for peptides with lower  $K_m$  values should be further studied (Bardelang, 2009).

## 7.7 Future work

### 7.7.1 Substrate specificity across SH3\_5 domains

SH3\_5 domains are made by bacteria producing PG with distinct compositions. Two major questions remain to be answered: how conserved is the recognition mechanism described for the Lss SH3\_5 domain? Do all SH3\_5 domains recognise both the peptide stem and the crosslinks? Sequence and structure alignments of the SH3\_5 family suggests the existence of conserved features necessary for the recognition of canonical ligand binding moieties (Fig 1.14, Chapter 1). Residues identified to interact with P4 are conserved across SH3\_5 domains present in the Protein Data Bank (PDB) despite their low sequence similarity (e.g. 32% similarity with *E. faecalis* BacL<sub>1</sub> or 30% similarity with *Streptococcus pyogenes* Plypy). The recognition of the pentaglycine crossbridge in the Lss SH3\_5 domain is mediated via the N-terminal extension that does not appear to be conserved in other domains. Although we cannot exclude that some residues upstream of the SH3\_5 consensus are involved in the recognition of the lateral chain in Gram-positive cocci, it is possible that the only PG motif recognised is the pentapeptide stem. Further structural and functional studies are required to explore the binding mechanism of SH3\_5 from bacteria with distinct PG structures. Towards the end of my PhD, I have started to investigate the recognition of *E. faecalis* PG by the SH3\_5 domain of the bacteriocin BacL<sub>1</sub>. Unfortunately, I have not been able to carry out the NMR experiments planned to test if these SH3-5 domains were recognising the pentapeptide stem and the L-Ala-L-Ala crossbridges via distinct binding sites.

### **7.7.2 What is the role of the modularity of SH3\_5 domains?**

As mentioned previously, the majority of SH3-5 domains are made of a single binding module but several proteins (55 in total) contain up to five modules. It will be important to investigate the binding properties of proteins with a distinct number of modules to explore if, like with the LysM domains, the diversity in the modularity of the SH3-5 domain reflects the diversity of binding mechanisms. It has been shown that other SH3 domains can undergo homodimerization (IB1 protein) (Kristensen, 2006). As previously mentioned, homodimerization is important for the binding activity of another binding domain, LysM (Sánchez-Vallet, 2020). An interesting model system will be the BacL<sub>1</sub> domain which contains three SH3\_5 binding modules. One of my objectives was to compare the binding activity of domains containing a variable number of repeats to PG fragments and whole sacculi. My preliminary results indicate that combination of BacL<sub>1</sub> SH3\_5 domains can be expressed at high levels and as soluble polypeptides (single domains, tandem domains) so it is tempting to assume that the three modules are not required to form a quaternary structure and are therefore likely to bind independently to PG.

### **7.7.3 Could other properties regulate peptidoglycan recognition by SH3\_5 domains?**

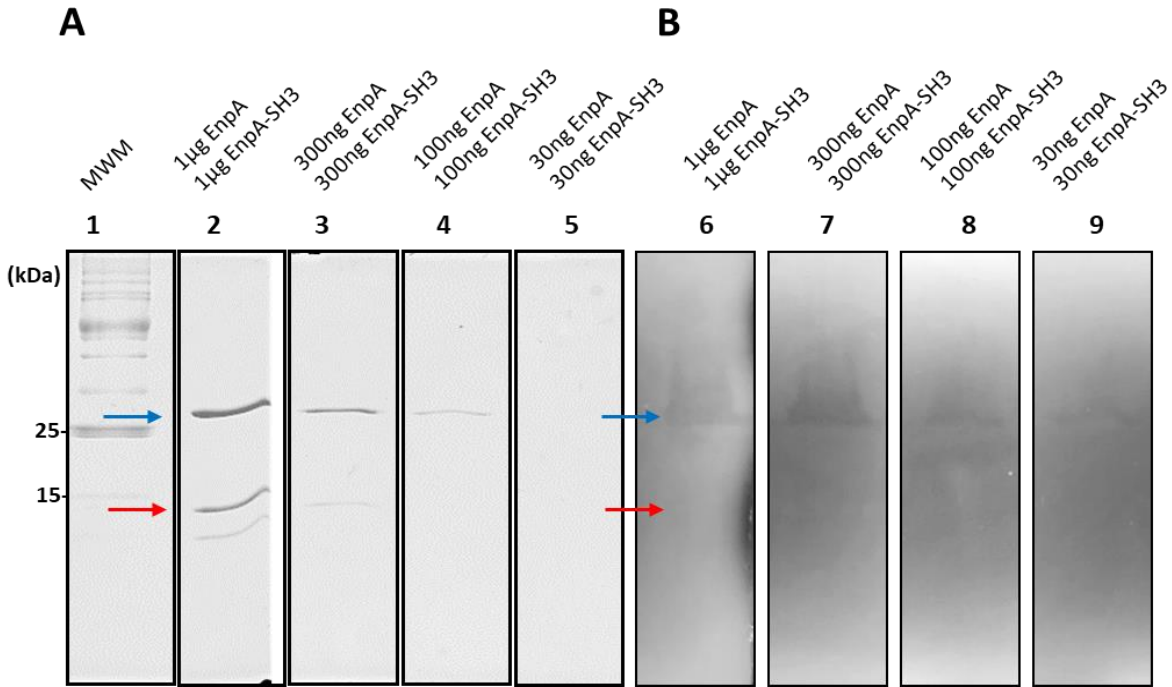
Several properties of SH3 domains have been shown to modulate the recognition of their ligands. SH3 proteins regulate distinct processes by associating with other proteins, through phosphorylation (WW2, SH3, SH2, and PDZ) (Akiva, 2012), heterodimerization (VAV N-terminal and GRB2 C-terminal SH3 domains complex) (Nishida, 2001), and *cis-trans* isomerization (Sarkar, 2011). None of these processes have been investigated in prokaryotic domains and it would therefore be interesting to investigate whether these modifications modulate PG recognition by SH3\_5 domains.

### **7.7.4 Harnessing SH3\_5 domains to generate antimicrobials**

The concept of exploiting bacteriolytic enzymes as therapeutic agents to eradicate pathogens has been contemplated for decades. The presence of serine amino acid substitutions in the interpeptide bridge prevents autolysis in the Lys producing staphylococci (Thumm, 1997;

Tschierske, 1997). This work showed that the Lss SH3\_5 domain can recognise staphylococcal PGs which do not have a pentaglycine crossbridge. To target all staphylococci, including those with modified pentaglycine crossbridges in their PG, we could therefore use the Lss SH3\_5 domain as a binding module, fusing it to a catalytic domain able to cleave a broader range of PG structures. To explore this possibility, the Lss SH3\_5 domain was fused to the catalytic domain of the *E. faecalis* EnpA (EnpA<sub>c</sub>) endopeptidase which can cleave a wide range of bacterial species harbouring distinct interpeptide bridges (Reste de Roca, 2010). EnpA<sub>c</sub> cleaves between the fourth D-Ala of the donor stem and the N-terminal L-Ala residue of the sidechain of the acceptor stem. So long as the PG is not directly crosslinked it can cleave different chemotypes irrespective of their length and amino acid sequence (Reste de Roca, 2010). EnpA<sub>c</sub> has been shown to be poorly active against intact PG and requires low ionic strength conditions (I. Sabala, personal communication). Previous work indicated that EnpA<sub>c</sub> has no detectable lytic activity against *S. aureus* (M. Molloy and S. Mesnage, unpublished). When fused to the Lss SH3\_5 domain, the EnpA<sub>c</sub>-SH3\_5 fusion protein displayed a clear lytic activity against *S. aureus* (Fig. 7.1). These preliminary results indicate that fusing the Lss SH3 domain to the catalytic domain of PG hydrolases could lead to chimeric proteins with potent antimicrobial activity.

Understanding the mechanisms underpinning ligand recognition by the SH3\_5 domains serves as a framework for deciphering the evolutionary strategies adopted by distinct cell wall binding domains and facilitates the development of future antimicrobial approaches. Our work raises questions about the significance of protein specificity for meaningful biological outputs.



**Figure 7.1 Zymogram analysis of the lytic activity of EnpA<sub>c</sub> and EnpA<sub>c</sub>-SH3\_5 fusion against *S. aureus* peptidoglycan. (A) SDS-PAGE analysis of the EnpA<sub>c</sub> and EnpA<sub>c</sub>-SH3\_5 proteins and (B) zymogram analysis of their lytic activity against *S. aureus* PG. The blue and red arrows indicate the bands corresponding to the EnpA<sub>c</sub>-SH3\_5 fusion (27 kDa) and the EnpA<sub>c</sub> domain (15.1 kDa), respectively. Lane 1, (MWM), molecular weight makers from the SDS-PAGE Coomassie stained gel; lane 2-5, four serial dilutions with a mixture of the EnpA<sub>c</sub> and EnpA<sub>c</sub>-SH3\_5 proteins were loaded and separated by SDS-PAGE; lane 6-9, zymography of the four serial dilutions of the EnpA<sub>c</sub> and EnpA<sub>c</sub>-SH3\_5 proteins separated by SDS-PAGE using gels with *S. aureus* PG. The haloes indicate the lysis of intact PG substrate by the enzyme on the zymogram (lane 6-9) located at the same position of the over-expressed protein bands (lane 2-5).**

## References

- Akiva, E., Friedlander, G., Itzhaki, Z., Margalit, H (2012). A dynamic view of domain-motif interactions. *PLoS Comput Biol* 8, e1002341.
- Alcorlo, M., Dik, D.A., De Benedetti, S., Mahasanen, K.V., Lee, M., Domínguez-Gil, T., Heseck, D., Lastochkin, E., López, D., Boggess, B., Mobashery, S., Hermoso, J.A (2019). Structural basis of denuded glycan recognition by SPOR domains in bacterial cell division. *Nat Commun* 10, 5567.
- Aliashkevich, A., Alvarez, L., & Cava, F (2018). New Insights Into the mechanisms and biological roles of D-amino acids in complex eco-systems. *Front Microbiol* 9, 683.
- Arai, M., Ferreon, J.C., Wright, P.E (2012). Quantitative analysis of multisite protein-ligand interactions by NMR: binding of intrinsically disordered p53 transactivation subdomains with the TAZ2 domain of CBP. *J Am Chem Soc* 134, 3792-3803.
- Arbeloa, A., Segal, H., Hugonnet, J. E., Josseaume, N., Dubost, L., Brouard, J. P., Gutmann, L., Mengin-Lecreulx, D., and Arthur, M (2004). Role of class A penicillin-binding proteins in PBP5-mediated beta-lactam resistance in *Enterococcus faecalis*. *J Bacteriol* 186, 1221–1228.
- Atrih, A., Bacher, G., Allmaier, G., Williamson, M.P., Foster, S.J (1999). Analysis of peptidoglycan structure from vegetative cells of *Bacillus subtilis* 168 and role of PBP 5 in peptidoglycan maturation. *J Bacteriol* 181, 956–3966.
- Baba, T., & Schneewind, O (1996). Target cell specificity of a bacteriocin molecule: a C-terminal signal directs lysostaphin to the cell wall of *Staphylococcus aureus*. *EMBO J* 15, 4789-4797.
- Baranowski, C., Welsh, M.A., Sham, L.T., Eskandarian, H.A., Lim, H.C., Kieser, K.J., Wagner, J.C., McKinney, J.D., Fantner, G.E., Ioerger, T.R., Walker, S., Bernhardt, T.G., Rubin, E.J., Rego, E.H (2018). Maturing *Mycobacterium smegmatis* peptidoglycan requires non-canonical crosslinks to maintain shape. *eLife* 7, e37516.
- Bardelang, P., Vankemmelbeke, M., Zhang, Y., Jarvis, H., Antoniadou, E., Rochette, S., Thomas, N.R., Penfold, C.N., James, R (2009). Design of a polypeptide FRET substrate that facilitates study of the antimicrobial protease lysostaphin. *Biochem J* 418, 615-624.
- Barthe P, M.G., Roumestand C, Cohen-Gonsaud (2010). The structure of PknB extracellular PASTA domain from *Mycobacterium tuberculosis* suggests a ligand-dependent kinase activation. *Structure* 185, 606-615.
- Bateman, A., and Bycroft, M (2000). The structure of a LysM domain from *E. coli* membrane-bound lytic murein transglycosylase D (MltD). *J Mol Biol* 299, 1113-1119.

- Becker, S., Swift, S., Korobova, O., *et al* (2015). Lytic activity of the staphylolytic Twort phage endolysin CHAP domain is enhanced by the SH3b cell wall binding domain. *FEMS Microbiol Lett* 362, 1-8.
- Becker, S.C., Foster-Frey, J., Stodola, A.J., Anacker, D., Donovan D.M (2009). Differentially conserved staphylococcal SH3b\_5 cell wall binding domains confer increased staphylolytic and streptolytic activity to a streptococcal prophage endolysin domain. *Gene* 443, 32-41.
- Becker, W., Bhattiprolu, K.C., Gubensäk, N., Zangger, K (2018). Investigating protein-ligand interactions by solution nuclear magnetic resonance spectroscopy. *Chemphyschem* 19, 895-906.
- Bellais, S., Arthur, M., Dubost, L., Hugonnet, J. E., Gutmann, L., Van Heijenoort, J., *et al* (2006). Aslfm, the D-aspartate ligase responsible for the addition of D-aspartic acid onto the peptidoglycan precursor of *Enterococcus faecium* *J Biol Chem* 281, 11586–11594.
- Benešík M., N.J., Janda L., Dopitová R., Pernisová M., Melková K., Tišáková L., Doškař J., Žídek L., Hejátko J., Pantůček R (2018). Role of SH3b binding domain in a natural deletion mutant of Kayvirus endolysin LysF1 with a broad range of lytic activity. *Virus Genes* 54, 130-139.
- Benson TE, H.M., Choi GH, *et al* (2001). A structural variation for MurB: X-ray crystal structure of *Staphylococcus aureus* UDP-N-acetylenolpyruvylglucosamine reductase (MurB). *Biochemistry* 40, 2340-2350.
- Bera, A., Biswas, R., Herbert, S., Kulauzovic, E., Weidenmaier, C., Peschel, A., Götz, F (2007). Influence of wall teichoic acid on lysozyme resistance in *Staphylococcus aureus*. *J Bacteriol* 189, 280-283.
- Berger-Bächli, B., & M. L. Kohler (1983). A novel site on the chromosome of *Staphylococcus aureus* influencing the level of methicillin resistance: genetic mapping. *FEMS Microbiol Lett* 20, 305-309.
- Bertrand, J.A., Auger, G., Fanchon, E., Martin, L., Blanot, D., van Heijenoort, J., Dideberg, O (1997). Crystal structure of UDP-N-acetylmuramoyl-L-alanine:D-glutamate ligase from *Escherichia coli*. *EMBO J* 16, 3416-3425.
- Bertrand, J.A., Fanchon, E., Martin, L., Chantalat, L., Auger, G., Blanot, D., van Heijenoort, J., Dideberg, O (2000). "Open" structures of MurD: domain movements and structural similarities with folylpolyglutamate synthetase. *J Mol Biol* 301, 1257-1266.
- Biswas, R., Voggu, L., Simon, U.K., Hentschel, P., Thumm, G., Götz, F (2006). Activity of the major staphylococcal autolysin Atl. *FEMS Microbiol Lett* 2006, 260-268.
- Bolam, D.N., Ciruela, A., McQueen-Mason, S., Simpson, P., Williamson, M. P., Rixon, J. E., Gilbert, H. J (1998). Pseudomonas cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. *Biochem J* 331 ( Pt 3)(Pt 3), 775–781.



Boneca, I., & Chiosis, G (2003). Vancomycin resistance: occurrence, mechanisms and strategies to combat it. *Expert Opin Ther Targets* 7, 311-328.

Boneca, I., Dussurget, O., Cabanes, D., *et al* (2007). A critical role for peptidoglycan N-deacetylation in *Listeria* evasion from the host innate immune system. *Proc Natl Acad Sci USA* 104, 997-1002.

Boniface A, B.A., Mengin-Lecreux D, Blanot D (2006). The MurE synthetase from *Thermotoga maritima* is endowed with an unusual D-lysine adding activity. *J Biol Chem* 281, 15680-15686.

Bouhss, A., Josseaume, N., Severin, A., *et al* (2002). Synthesis of the L-alanyl-L-alanine cross-bridge of *Enterococcus faecalis* peptidoglycan. *J Biol Chem* 277, 45935-45941.

Braun, L., Dramsi, S., Dehoux, P., Bierne, H., Lindahl, G., Cossart, P (1997). InlB: an invasion protein of *Listeria monocytogenes* with a novel type of surface association. *Mol Microbiol* 25, 285-294.

Brinster, S., Furlan, S., Serror, P (2007). C-terminal WxL domain mediates cell wall binding in *Enterococcus faecalis* and other gram-positive bacteria. *J Bacteriol* 189, 1244-1253.

Brocchieri, L., Karlin, S (2005). Protein length in eukaryotic and prokaryotic proteomes. *Nucleic Acids Res* 33, 3390-3400.

Brown ED, V.E., Walsh CT, Kolter R (1995). MurA (MurZ), the enzyme that catalyzes the first committed step in peptidoglycan biosynthesis, is essential in *Escherichia coli*. *J Bacteriol* 177, 4194-4197.

Brown, S., Santa Maria, J. P., Jr., Walker, S (2013). Wall teichoic acids of gram-positive bacteria. *Annu Rev Microbiol* 67, 313-336.

Budzik JM, M.L., Souda P, Whitelegge JP, Faull KF, Schneewind O (2008). Amide bonds assemble pili on the surface of bacilli. *Proc Natl Acad Sci USA* 105, 10215–10220.

Bugg, T., Wright, GD., Dutka-Malen, S., Arthur, M., Courvalin, P., & Walsh, CT (1991). Molecular basis for vancomycin resistance in *Enterococcus faecium* BM4147: biosynthesis of a depsipeptide peptidoglycan precursor by vancomycin resistance proteins VanH and VanA. *Biochemistry* 30, 10408-10415.

Buist, G., Steen, A., Kok, J., Kuipers, O. P (2008). LysM, a widely distributed protein motif for binding to (peptido)glycans. *Mol Microbiol* 68, 838-847.

Bukowska-Faniband, E., & Hederstedt, L (2015). The PASTA domain of penicillin-binding protein SpoVD is dispensable for endospore cortex peptidoglycan assembly in *Bacillus subtilis*. *Microbiology* 161, 330-340.

Caetano-Anollés, G., & Caetano-Anollés, D (2003). An evolutionarily structured universe of protein architecture. *Genome Res* 13, 1563-1571.

Caetano-Anollés, G., Wang, M., Caetano-Anollés, D., Miententhal, JE (2009). The origin, evolution and structure of the protein world. *Biochem J* 417, 621-637.

Calvanese, L., Falcigno, L., Maglione, C., *et al* (2014). Structural and binding properties of the PASTA domain of PonA2, a key penicillin binding protein from *Mycobacterium tuberculosis*. *Biopolymers* 101, 712-719.

Cavanagh, J., Fairbrother, W., Palmer, AG & Skelton, NG (2007). Protein NMR Spectroscopy – Principles and Practice (Academic Press) (Elsevier Inc. All rights reserved).

Caveney NA., L.F., Strynadka NC (2018). Enzyme structures of the bacterial peptidoglycan and wall teichoic acid biogenesis pathways. *Curr Opin Struct Biol* 53, 45-58.

Clarke, A., Strating, H., and Blackburn, N (2007). Pathways for the O-Acetylation of Bacterial Cell Wall Polysaccharides (New York: Plenum Publishing Co. Ltd). 10.1007/0-306-46821-2\_7.

de Jonge, R., & Thomma, B. P (2009). Fungal LysM effectors: extinguishers of host immunity? *Trends Microbiol* 17, 151-157.

DeDent A, B.T., Missiakas DM, Schneewind O (2008). Signal peptides direct surface proteins to two distinct envelope locations of *Staphylococcus aureus*. *EMBO J* 27, 2656-2668.

del Campo, R., Tenorio, C., Jiménez-Díaz, R., *et al* (2001). Bacteriocin production in vancomycin-resistant and vancomycin-susceptible *Enterococcus* isolates of different origins. *Antimicrob Agents Chemother* 45, 905-912.

Deng L, K.D., Krick TP, Wessels MR (2000). Characterization of the linkage between the type III capsular polysaccharide and the bacterial cell wall of group B *Streptococcus*. *J Biol Chem* 275, 7497-7504.

Desvaux, M. (2005). The cellulosome of *Clostridium cellulolyticum*. *Enzyme Microb Technol* 37, 373-385.

Desvaux, M., Candela, T., & Serror, P (2018). Surfaceome and Proteosurfaceome in Parietal Monoderm Bacteria: Focus on Protein Cell-Surface Display. *Front Microbiol* 9, 100.

Dramsi S., M.S., Davison S., Arthur M (2008). Covalent attachment of proteins to peptidoglycan. *FEMS Microbiol Rev* 32, 307-320.

Dramsi S., B.H. (2017). Spatial Organization of Cell Wall-Anchored Proteins at the Surface of Gram-Positive Bacteria. *Curr Top Microbiol Immunol* 404. 10.1007/82\_2016\_4.

Dumas, E., Meunier, B., Berdagué, JL., Chambon, C., Desvaux, M., Hébraud, M (2008). Comparative analysis of extracellular and intracellular proteomes of *Listeria monocytogenes* strains reveals a correlation between protein expression and serovar. *Appl Environ Microbiol* 74, 7399-7409.

Duncan K., v.H.J., Walsh CT (1990). Purification and characterization of the D-alanyl-D-alanine-adding enzyme from *Escherichia coli*. *Biochemistry* 29, 2379-2386.

Egan, A., Errington, J., Vollmer, W (2020). Regulation of peptidoglycan synthesis and remodelling. *Nat Rev Microbiol* 18, 446-460.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, SR., Luciani, A., Potter, SC., Qureshi, M., Richardson, LJ., Salazar, GA., Smart, A., Sonnhammer, ELL., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, SCE., Finn, RD (2019). The Pfam protein families database in 2019. *Nucleic Acids Res* 47, D427–D432.

Eldholm, V., Johnsborg, O., Haugen, K., Ohnstad, HS., Håvarstein, LS (2009). Fratricide in *Streptococcus pneumoniae*: contributions and role of the cell wall hydrolases CbpD, LytA and LytC. *Microbiology* 155, 2223-2224.

Emanuele, J.J., Jr, Jin, H., Jacobson, B.L., Chang, C.Y., Einspahr, H.M., & Villafranca, J.J (1996). Kinetic and crystallographic studies of *Escherichia coli* UDP-N-acetylmuramate:L-alanine ligase. *Protein Sci* 5, 2566-2574.

Emsley, P., and Cowtan, K (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60, 2126-2132.

Endl J, S.H., Fiedler F, Schleifer KH (1983). Chemical composition and structure of cell wall teichoic acids of staphylococci. *Arch Microbiol* 135, 215-223.

Evilia, R. (2007). Quantitative NMR spectroscopy. *Analytical Lett* 34, 2227-2236.

Farmer, B., Constantine, KL., Goldfarb, V., *et al* (1996). Localizing the NADP<sup>+</sup> binding site on the MurB enzyme by NMR. *Nat Struct Biol* 12, 995-997.

Feng, S., Chen, JK., Yu, H., Simon, JA., Schreiber, SL (1994). Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3-ligand interactions. *Science* 266, 1241-1247.

Fernandez-Ballester, G., Blanes-Mira, C., Serrano, L (2004). The Tryptophan Switch: Changing Ligand-binding Specificity from Type I to Type II in SH3 Domains. *J Mol Biol* 335, 619–629.

Fernández-Tornero, C., López, R., García, E., Giménez-Gallego, G., Romero, A (2001). A novel solenoid fold in the cell wall anchoring domain of the pneumococcal virulence factor LytA. *Nat Struct Biol* 8, 1020-1024.

Fersht, A.R. (1999). Structure and Mechanism in Protein Science (W.H. Freeman, New York) 631 pp.

Fischer, W., Mannsfeld, T., and Hagen, G. (1990). On the basic structure of poly(glycerophosphate) lipoteichoic acids. *Biochem Cell Biol* 68, 33-43.

- Fischetti, V., Pancholi, V., Schneewind, O (1990). Conservation of a hexapeptide sequence in the anchor region of surface proteins from gram-positive cocci. *Mol Microbiol* 4, 1603-1605.
- Francius, G., Domenech, O., Mingeot-Leclercq, M. P., & Dufrêne, Y. F (2008). Direct observation of *Staphylococcus aureus* cell wall digestion by lysostaphin. *J Bacteriol* 190, 7904–7909.
- Fujimoto Y, I.S., Kawasaki A, Shiokawa Z, Shimoyama A, Hashimoto T, Kusumoto S, Fukase K (2007). Chemical synthesis of peptidoglycan fragments for elucidation of the immunostimulating mechanism. *J Endotoxin Res* 13, 189-196.
- Fukami-Kobayashi K, M.Y., Tateno Y, Nishikawa K (2007). A tree of life based on protein domain organizations. *Mol Biol Evol* 24, 1181-1189.
- Galán-Bartual, P.-D., I., García, P., Hermoso, J.A (2015). Structure and function of choline-binding proteins. In *Streptococcus Pneumoniae: Molecular Mechanisms of Host-Pathogen Interactions*, Vol 1st ed (Academic Press: Cambridge, MA, USA,).
- Galloway-Peña, J., Liang, X., Singh, KV., *et al* (2015). The identification and functional characterization of WxL proteins from *Enterococcus faecium* reveal surface proteins involved in extracellular matrix interactions. *J Bacteriol* 197, 882-892.
- Galperin, M., Koonin, EV (1997). A diverse superfamily of enzymes with ATP-dependent carboxylate-amine/thiol ligase activity. *Protein Sci* 6, 2639-2643.
- García, J., Sánchez-Beato, AR., Medrano, FJ., López R (1998). Versatility of choline-binding domain. *Microb Drug Resist* 4, 25-36.
- Gargis S.R., H.H.E., LeBlanc P.A., Dekker L., Simmonds R.S., Sloan G.L (2010). Inhibition of the Activity of Both Domains of Lysostaphin through Peptidoglycan Modification by the Lysostaphin Immunity Protein. *Appl Environ Microbiol* 76 (20), 6944-6946.
- Geiss-Liebisch S, R.S., Beczala A, Sanchez-Carballo P, Kruszynska K, Repp C, *et al* (2012). Secondary cell wall polymers of *Enterococcus faecalis* are critical for resistance to complement activation via mannose-binding lectin. *J Biol Chem* 287, 37769–37777.
- Gerding, M., Liu, B., Bendezu, FO., Hale, CA., Bernhardt, TG., de Boer, PA (2009). Self-enhanced accumulation of FtsN at division sites and roles for other proteins with a SPOR domain (DamX, DedD, and RlpA) in *Escherichia coli* cell constriction. *J Bacteriol* 191, 7383–7401.
- Gibson, D., Young, L., Chuang, R. *et al* (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6, 343-345.
- Gill, J., Rixon, J. E., Bolam, D. N., McQueen-Mason, S., Simpson, P. J., Williamson, M. P., Gilbert, H. J (1999). The type II and X cellulose-binding domains of *Pseudomonas xylanase A* potentiate

catalytic activity against complex substrates by a common mechanism. *Biochem J* 342 ( Pt 2)(Pt 2), 473–480.

Gilmore, M., Lebreton, F., van Schaik, W (2013). Genomic transition of enterococci from gut commensals to leading causes of multidrug-resistant hospital infection in the antibiotic era. *Curr Opin Microbiol* 16, 10-16.

Giudicelli, S., Tomasz, A (1984). Attachment of pneumococcal autolysin to wall teichoic acids, an essential step in enzymatic wall degradation. *J Bacteriol* 158, 1188-1190.

Goffin, C.G., J. M (1998). Multimodular penicillinbinding proteins: an enigmatic family of orthologs and paralogs. *Microbiol Mol Biol Rev* 62, 1079–1093.

Gonzalez-Delgado, L., Walters-Morgan, H., Salamaga, B., Robertson, A., Hounslow, A., Jagielska, E., Sabała, I., Williamson, M.P., Lovering, A., Mesnage, S (2020). Two-site recognition of *Staphylococcus aureus* peptidoglycan by lysostaphin SH3b. *Nat Chem Biol* 16, 24-30.

Gordon, E., Mouz, N., Duée, E., Dideberg, O (2000). The crystal structure of the penicillin-binding protein 2x from *Streptococcus pneumoniae* and its acyl-enzyme form: implication in drug resistance. *J Mol Biol* 299, 477-485.

Gordon E., F.B., Chantalat L., van Heijenoort J., Mengin-Lecreulx D., Dideberg O (2001). Crystal structure of UDP-N-acetylmuramoyl-L-alanyl-D-glutamate: meso-diaminopimelate ligase from *Escherichia coli*. *J Biol Chem* 276, 10999-11006.

Gray, A., Egan, AJ., Van't Veer, IL., *et al* (2015). Coordination of peptidoglycan synthesis and outer membrane constriction during *Escherichia coli* cell division. *eLife* 4, e07118.

Gründling, A., & Schneewind, O (2006). Cross-linked peptidoglycan mediates lysostaphin binding to the cell wall envelope of *Staphylococcus aureus*. *J Bacteriol* 188, 2463-2472.

Gu, J., Feng, Y., Feng, X., Sun, C., Lei, L., Ding, W., Niu, F., Jiao, L., Yang, M., Li, Y., Liu, X., Song, J., Cui, Z., Han, D., Du, C., Yang, Y., Ouyang, S., Liu, Z. J., Han, W (2014). Structural and biochemical characterization reveals LysGH15 as an unprecedented "EF-hand-like" calcium-binding phage lysin. *PLoS Pathog* 10, e1004109.

Guerardel, Y., Sadovskaya, I., Maes, E., Furlan, S., Chapot-Chartier, M-P., Mesnage, S., Rigottier-Gois, L., Serron, P (2020). Complete structure of the enterococcal polysaccharide antigen (EPA) of vancomycin-resistant *Enterococcus faecalis* V583 reveals that EPA decorations are teichoic acids covalently linked to a rhamnopolysaccharide backbone. *mBio* 11, e00277-00220.

Harkness, R.E., and Braun, V (1988). Colicin M inhibits peptidoglycan biosynthesis by interfering with lipid carrier recycling. *J Biol Chem* 264, 6177–6182.

- Hartwell, L.H., Hopfield, J.J., Leibler, S., and Murray, A.W (1999). From molecular to modular cell biology. *Nature* 402, C47–C52.
- Hayhurst, E., Kailas, L., Hobbs, J.K., Foster S.J (2008). Cell wall peptidoglycan architecture in *Bacillus subtilis*. *Proc Natl Acad Sci USA* 105, 14603-14608.
- Healy V.L., M.L., Li X, Hall S.E., Raushel F.M., Walsh C.T (2000). D-Ala-D-X ligases: evaluation of D-alanyl phosphate intermediate by MIX, PIX and rapid quench studies. *Chem Biol* 7, 505-514.
- Heidrich C, T.M., Ursinus A, et al (2001). Involvement of N-acetylmuramyl-L-alanine amidases in cell separation and antibiotic-induced autolysis of *Escherichia coli*. *Mol Microbiol* 41, 167-178.
- Heilmann, C., Hussain, M., Peters, G., Götz, F (1997). Evidence for autolysin-mediated primary attachment of *Staphylococcus epidermidis* to a polystyrene surface. *Mol Microbiol* 24, 1013-1024.
- Heinz, E., Stubenrauch, C.J., Grinter, R., et al (2016). Conserved Features in the Structure, Mechanism, and Biogenesis of the Inverse Autotransporter Protein Family. *Genome Biol Evol* 8, 1690-1705.
- Henze, U., Sidow, T., Wecke, J., Labischinski, H. and Berger-Bächi, B (1993). Influence of femB on methicillin resistance and peptidoglycan metabolism in *Staphylococcus aureus*. *J Bacteriol* 175, 1612–1620.
- Hermoso, J., Monterroso, B., Albert, A., et al (2003). Structural basis for selective recognition of pneumococcal cell wall by modular endolysin from phage Cp-1. *Structure* 11, 1239-1249.
- Hesse L, B.J., Dementin S, Blanot D, Mengin-Lecreux D and Chopra I (2003). Functional and biochemical analysis of *Chlamydia trachomatis* MurC, an enzyme displaying UDP-N-acetylmuramate: amino acid ligase activity. *J Bacteriol* 185, 6507–6512.
- Higman, V. (2019). Protein NMR A Practical Guide (<https://www.protein-nmr.org.uk/>).
- Hirakawa, H., Akita, H., Fujiwara, T., Sugai, M., Kuhara, S (2009). Structural insight into the binding mode between the targeting domain of ALE-1 (92AA) and pentaglycine of peptidoglycan. *Protein Eng Des Sel* 22, 385-391.
- Holm, L., Kääriäinen, S., Rosenström, P., Schenkel, A (2008). Searching protein structure databases with DaliLite v.3. *Bioinformatics (Oxford, England)* 24, 2780-2781.
- Horsburgh, M.J., Aish, J. L., White, I. J., Shaw, L., Lithgow, J. K., & Foster, S. J (2002).  $\sigma$ B modulates virulence determinant expression and stress resistance: characterization of a functional rsbU strain derived from *Staphylococcus aureus* 8325-4. *J Bacteriol* 184(19), 5457–5467.
- Hoskins, J., Matsushima, P., Mullen, D.L., et al (1999). Gene disruption studies of penicillin-binding proteins 1a, 1b, and 2a in *Streptococcus pneumoniae*. *J Bacteriol* 181, 6552-6555.

- Hostettler, L., Grundy, L., Käser-Pébernard, S., Wicky, C., Schafer, W. R., & Glauser, D. A (2017). The Bright Fluorescent Protein mNeonGreen Facilitates Protein Expression Analysis *in vivo*. *G3 (Bethesda, Md)* 7(2), 607–615.
- Hugonnet, J.E., Mengin-Lecreux, D., Monton, A., den Blaauwen, T., Carbonnelle, E., Veckerlé, C., Brun, Y. V., van Nieuwenhze, M., Bouchier, C., Tu, K., Rice, L. B., & Arthur, M (2016). Factors essential for l,d-transpeptidase-mediated peptidoglycan crosslinking and  $\beta$ -lactam resistance in *Escherichia coli*. *eLife* 5, e19469
- Ikeda, M., Wachi, M., Jung, H.K., Ishino, F., Matsushashi, M (1991). The *Escherichia coli* mraY gene encoding UDP-N-acetylmuramoyl-pentapeptide: undecaprenyl-phosphate phospho-N-acetylmuramoyl-pentapeptide transferase. *J Bacteriol* 173, 1021-1026.
- Inoue, T., Tomita, H., Ike, Y (2006). Bac 32, a novel bacteriocin widely disseminated among clinical isolates of *Enterococcus faecium*. *Antimicrob Agents Chemother* 50, 1202-1212.
- Jack, R., Tagg, J.R., Ray, B (1995). Bacteriocins of gram-positive bacteria. *Microbiol Rev* 59, 171-200.
- Jagielska, E., Chojnacka, O., Sabała, I (2016). LytM Fusion with SH3b-Like Domain Expands Its Activity to Physiological Conditions. *Microb Drug Resist* 22, 461–469.
- Jensen, K. (1958). A normally occurring staphylococcus antibody in the human serum. *Acta Pathol Microbiol Scand* 44, 421–428.
- Jones, G., & Dyson, P (2006). Evolution of transmembrane protein kinases implicated in coordinating remodeling of gram-positive peptidoglycan: inside versus outside. *J Bacteriol* 188, 7470-7476.
- Jonquière, R., Bierne, H., Fiedler, F., Gounon, P., Cossart, P (1999). Interaction between the protein InlB of *Listeria monocytogenes* and lipoteichoic acid: a novel mechanism of protein association at the surface of gram-positive bacteria. *Mol Microbiol* 34, 902-914.
- Joosten, R.P., Joosten, K., Cohen, S. X., Vriend, G. and Perrakis, A. (2011). Automatic rebuilding and optimization of crystallographic structures in the Protein Data Bank. *Bioinformatics* 27, 3392–3398.
- Kabsch, W. (2010). Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr* 66, 133-144.
- Kaku, H., Nishizawa, Y., Ishii-Minami, N., Akimoto-Tomiyama, C., Dohmae, N., Takio, K., Minami, E., Shibuya, N (2006). Plant cells recognize chitin fragments for defense signaling through a plasma membrane receptor. *Proc Natl Acad Sci USA* 103, 11086-11091.

- Kamitori, S., & Yoshida, H (2015). Structure-Function Relationship of Bacterial SH3 Domains. In *SH Domains: Structure, Mechanisms and Applications* (Springer), 71-89 pp.
- Kaneko, T., Li, L., Li, S.S (2008). The SH3 domain--a family of versatile peptide- and protein-recognition module. *Front Biosci* 13, 4938-5.
- Kang, C., Abbott, DW., Park, ST., Dascher, CC., Cantley, LC., Husson, RN (2005). The *Mycobacterium tuberculosis* serine/threonine kinases PknA and PknB: substrate identification and regulation of cell shape. *Genes Dev* 19, 1692-1704.
- Ke, Y., Inoue, T., Yamashita, H., Tomita, H (2010). Abstract S6:1. Paper presented at: 3rd International ASM Conference on Enterococci (American Society for Microbiology).
- Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* 10, 845-858.
- Kharat AS, T.A. (2006). Drastic reduction in the virulence of *Streptococcus pneumoniae* expressing type 2 capsular polysaccharide but lacking choline residues in the cell wall. *Mol Microbiol* 60, 93-107.
- Kim, H., Mittenthal, JE., Caetano-Anollés, G (2006). MANET: tracing evolution of protein architecture in metabolic networks. *BMC Bioinformatics* 7, 351.
- Kleckner I.R., a.F.M.P. (2011). An introduction to NMR-based approaches for measuring protein dynamics. *Biochim Biophys Acta* 1814, 942–968.
- Koharudin LM, V.A., Montanini B, *et al* (2011). Structure-function analysis of a CVNH-LysM lectin expressed during plant infection by the rice blast fungus *Magnaporthe oryzae*. *Structure* 19, 662-674.
- Kojima, N., Araki, Y., Ito, E. (1985). Structural studies on the linkage unit of ribitol teichoic acid of *Lactobacillus plantarum*. *Eur J Biochem* 148, 29-34.
- Kokoszka, M., Kall, S., Khosla, S., McGinnis, J., Lavie, A., and Kay, B. (2018). Identification of two distinct peptide-binding pockets in the SH3 domain of human mixed-lineage kinase 3. *J Biol Chem* 293.
- Kristensen, O., Guenat, S., Dar, I., *et al* (2006). A unique set of SH3-SH3 interactions controls IB1 homodimerization. *EMBO J* 25, 785-797.
- Kuk, A.C., Mashalidis, E.H., and Lee, S.Y (2017). Crystal structure of the MOP flippase MurJ in an inward facing conformation. *Nat Struct Mol Biol Evol* 24, 171–176.
- Kumar, S., Rubino, F.A., Mendoza, A.G., & Ruiz, N (2019). The bacterial lipid II flippase MurJ functions by an alternating-access mechanism. *J Biol Chem* 294, 981–990.



Kurland, C., Canbäck, B., Berg, OG (2007). The origins of modern proteomes. *Biochimie* 89, 1454-1463.

Kurochkina, N., Guha, U (2013). SH3 domains: modules of protein-protein interactions. *Biophys Rev* 5, 29-39.

Kurushima, J., Hayashi, I., Sugai, M., Tomita, H (2013). Bacteriocin protein BacL1 of *Enterococcus faecalis* is a peptidoglycan D-isoglutamyl-L-lysine endopeptidase. *J Biol Chem* 288, 36915-36925.

Kurushima, J., Nakane, D., Nishizaka, T., & Tomita, H (2015). Bacteriocin protein BacL(1) of *Enterococcus faecalis* targets cell division loci and specifically recognizes L-Ala(2)-cross-bridged peptidoglycan. *J Bacteriol* 197, 286–295.

Laemmli, U. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680-685.

Lancefield, R. (1962). Current knowledge of type-specific M antigens of group A streptococci. *J Immunol* 89, 307-313.

Larson, T., . & Yother, J (2017). *Streptococcus pneumoniae* capsular polysaccharide is linked to peptidoglycan via a direct glycosidic bond to  $\beta$ -D- N-acetylglucosamine. *PNAS* 114, 2016-20431

Lefebvre, B., Klaus-Heisen, D., Pietraszewska-Bogiel, A, *et al* (2012). Role of N-glycosylation sites and CXC motifs in trafficking of medicago truncatula Nod factor perception protein to plasma membrane. *J Biol Chem* 287, 10812-10823.

Leyh-Bouille, M., Bonaly, R., Ghuysen, JM., Tinelli, R., Tipper, D (1970). LL-diaminopimelic acid containing peptidoglycans in walls of *Streptomyces* sp. and of *Clostridium perfringens* (type A). *Biochemistry* 9, 2944-2952.

Lim, W., Fox, RO., Richards, FM (1994). Stability and peptide binding affinity of an SH3 domain from the *Caenorhabditis elegans* signaling protein Sem-5. *Protein Sci* 3, 1261-1266.

Liu, T., Liu, Z., Song, C., Hu, Y., Han, Z., She, J., Fan, F., Wang, J., Jin, C., Chang, J., Zhou, J. M., Chai, J (2012). Chitin-induced dimerization activates a plant immune receptor. *Science* 336, 1160-1164.

Liu, X., Meiresonne, N. Y., Bouhss, A. & den Blaauwen, T (2018). FtsW activity and lipid II synthesis are required for recruitment of MurJ to midcell during cell division in *Escherichia coli*. *Mol Microbiol* 109, 855-884.

Lood, R., Raz, A., Molina, H., Euler, CW., Fischetti, VA (2014). A highly active and negatively charged *Streptococcus pyogenes* lysin with a rare D-alanyl-L-alanine endopeptidase activity protects mice against streptococcal bacteremia. *Antimicrob Agents Chemother* 58, 3073-3084.

Lu JZ, F.T., Komatsuzawa H, Sugai M, Sakon J (2006). Cell wall-targeting domain of glycyglycine endopeptidase distinguishes among peptidoglycan cross-bridges. *J Biol Chem* 281, 549-558.

- Macdonald, A., Crowder, K., Street, A., McCormick, C., & Harris, M (2004). The hepatitis C virus NS5A protein binds to members of the Src family of tyrosine kinases and regulates kinase activity. *J Gen Virol* 85, 721-729.
- Maestro, B., & Sanz, JM (2016). Choline Binding Proteins from *Streptococcus pneumoniae*: A Dual Role as Enzybiotics and Targets for the Design of New Antimicrobials. *Antibiotics (Basel)* 5, 21.
- Maestro, B., Novaková, L., Heseck, D., *et al* (2011). Recognition of peptidoglycan and  $\beta$ -lactam antibiotics by the extracellular domain of the Ser/Thr protein kinase StkP from *Streptococcus pneumoniae*. *FEBS Lett* 585, 357-363.
- Mahapatra, S., Crick, DC., Brennan, PJ (2000). Comparison of the UDP-N-acetylmuramate:L-alanine ligase enzymes from *Mycobacterium tuberculosis* and *Mycobacterium leprae*. *J Bacteriol* 182, 6827-6830.
- Mainardi, J., Fourgeaud, M., Hugonnet, JE., *et al* (2005). A novel peptidoglycan cross-linking enzyme for a beta-lactam-resistant transpeptidation pathway. *J Biol Chem* 280, 38146-38152.
- Mainardi, J., Villet, R., Bugg, TD., Mayer, C., Arthur, M (2008). Evolution of peptidoglycan biosynthesis under the selective pressure of antibiotics in Gram-positive bacteria. *FEMS Microbiol Rev* 32, 386-408.
- Maity Sanhita , K.G.R., and Suresh Kumar Thallapuram Krishnaswamy (2019). NMR Methods to Characterize ProteinLigand Interactions. *Natural Prod Comm*, 1–17.
- Marraffini, L., Schneewind, O (2005). Anchor structure of staphylococcal surface proteins. V. Anchor structure of the sortase B substrate IsdC. *J Biol Chem* 280, 16263-16271.
- Matias, V., Beveridge, TJ (2005). Cryo-electron microscopy reveals native polymeric cell wall structure in *Bacillus subtilis* 168 and the existence of a periplasmic space. *Mol Microbiol* 56, 240-251.
- Matos, R., Lapaque, N., Rigottier-Gois, L., *et al* (2013). *Enterococcus faecalis* prophage dynamics and contributions to pathogenic traits. *PLoS Genet* 9, e1003539.
- Mayer, B.J. (2001). SH3 domains: complexity in moderation. *J of Cell Sci* 14, 1253-1263.
- Mayer, B.J., & Hamaguchi M (1998). A novel viral oncogene with structural similarity to phospholipase C. *Nature* 332, 272-275.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read R.J (2007). Phaser crystallographic software. *J Appl Crystallogr* 40, 658-674.
- Mesnage, S., Chau, F., Dubost, L., Arthur, M (2008). Role of *N*-acetylglucosaminidase and *N*-acetylmuramidase activities in *Enterococcus faecalis* peptidoglycan metabolism. *J Biol Chem* 283, 19845-19853.

- Mesnager, S., Dellarole, M., Baxter, N. J., Rouget, J. B., Dimitrov, J. D., Wang, N., Fujimoto, Y., Hounslow, A. M., Lacroix-Desmazes, S., Fukase, K., Foster, S. J., Williamson, M. P (2014). Molecular basis for bacterial peptidoglycan recognition by LysM domains. *Nat Commun* 5, 4269.
- Mir, M., Asong, J., Li, X., Cardot, J., Boons, GJ., Husson, RN (2011). The extracytoplasmic domain of the *Mycobacterium tuberculosis* Ser/Thr kinase PknB binds specific muropeptides and is required for PknB localization. *PLoS Pathog* 7, e1002182.
- Mishima, M., Shida, T., Yabuki, K., Kato, K., Sekiguchi, J., Kojima, C (2005). Solution structure of the peptidoglycan binding domain of *Bacillus subtilis* cell wall lytic enzyme CwC: characterization of the sporulation-related repeats by NMR. *Biochemistry* 44, 10153-10163.
- Mitkowski, P., Jagielska, E., Nowak, E., Bujnicki, J., Stefaniak, F., Niedzialek, D., Bochtler, M., Sabała, I (2019). Structural bases of peptidoglycan recognition by lysostaphin SH3b domain. *Sci Rep* 9, 5965.
- Mohamed JA, H.W., Nallapareddy SR, Teng F, Murray BE. (2005). Influence of Origin of Isolates, Especially Endocarditis Isolates, and Various Genes on Biofilm Formation by *Enterococcus faecalis*. *Infect Immun* 73, 7075.
- Mohammadi T, v.D.V., Sijbrandi R, *et al* (2011). Identification of FtsW as a transporter of lipid-linked cell wall precursors across the membrane. *EMBO J* 30, 1425–1432.
- Murzin, A., Brenner, S.E., Hubbard, T., & Chothia, C (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* 247, 536–540.
- Nagy, T., Tunnicliffe, RB., Higgins, LD., Walters, C., Gilbert, HJ., Williamson, M.P (2007). Characterization of a double dockerin from the cellulosome of the anaerobic fungus *Piromyces equi*. *J Mol Biol* 373, 612–622.
- Nariya, H., Miyata, S., Tamai, E., Sekiya, H., Maki, J., Okabe, A (2011). Identification and characterization of a putative endolysin encoded by episomal phage phiSM101 of *Clostridium perfringens*. *Appl Microbiol Biotechnol* 90, 1973-1979.
- Navarre, W., & Schneewind, O (1999). Surface proteins of Gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. *Microbiol Mol Biol Rev* 63, 174-229.
- Nega, M., Dube, L., Kull, M., Ziebandt, A. K., Ebner, P., Albrecht, D., Götz, F (2015). Secretome analysis revealed adaptive and non-adaptive responses of the *Staphylococcus carnosus* femB mutant. *Proteomics* 15, 1268–1279.
- Ness, I., Diep, DB., Ike, Y (2014). Enterococcal bacteriocins and antimicrobial proteins that contribute to niche control (Boston: Massachusetts: Eye and Ear Infirmary).

- Neuhaus, F.C., Baddiley, J. (2003). A continuum of anionic charge: structures and functions of D-alanyl-teichoic acids in gram-positive bacteria. *Microbiol Mol Biol Rev* 67, 686-723.
- Nishida, M., Nagata, K., Hachimori, Y., *et al* (2001). Novel recognition mode between Vav and Grb2 SH3 domains. *EMBO J* 20, 2995-3007.
- Nölling, J., Breton, G., Omelchenko, MV., *et al* (2001). Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *J Bacteriol* 183, 4823-4838.
- Ogawara, H. (2016). Distribution of PASTA domains in penicillin-binding proteins and serine/threonine kinases of *Actinobacteria*. *J Antibiot (Tokyo)* 69, 660-685.
- Oliveira, H., Melo, LD., Santos, SB., *et al* (2013). Molecular aspects and comparative genomics of bacteriophage endolysins. *J Virol* 87, 4558-4570.
- Osaki, M., Arcondéguy, T., Bastide, A., Touriol, C., Prats, H., Trombe, MC (2009). The StkP/PhpP signaling couple in *Streptococcus pneumoniae*: cellular organization and physiological characterization. *J Bacteriol* 19, 4943-4950.
- Palmer, K., Godfrey, P., Griggs, A., Kos, VN., Zucker, J., Desjardins, C., *et al* (2012). Comparative genomics of enterococci: variation in *Enterococcus faecalis*, clade structure in *E. faecium*, and defining characteristics of *E. gallinarum* and *E. casseliflavus*. *MBio* 3, e00318–00311.
- Pares, S., Mouz, N., Pétilot, Y., Hakenbeck, R., Dideberg, O (1996). X-ray structure of *Streptococcus pneumoniae* PBP2x, a primary penicillin target enzyme. *Nat Struct Biol* 3, 284-289.
- Park, J. (1952). Uridine-5'-pyrophosphate derivatives. II. Isolation from *Staphylococcus aureus*. *J Biol Chem* 194, 877-884.
- Pasquina-Lemonche, L., Burns, J., Turner, RD., *et al* (2020). The architecture of the Gram-positive bacterial cell wall. *Nature* 582, 294-297.
- Paulsen, I., Banerjee, L., Myers, GS., *et al* (2003). Role of mobile DNA in the evolution of vancomycin-resistant *Enterococcus faecalis*. *Science* 299, 2071-2074.
- Percy, M., & Gründling, A (2014). Lipoteichoic acid synthesis and function in gram-positive bacteria. *Annu Rev Microbiol* 68, 81-100.
- Percy, M., Karinou, E., Webb, AJ., Gründling, A (2016). Identification of a Lipoteichoic Acid Glycosyltransferase Enzyme Reveals that GW-Domain-Containing Proteins Can Be Retained in the Cell Wall of *Listeria monocytogenes* in the Absence of Lipoteichoic Acid or Its Modifications. *J Bacteriol* 198, 2029-2042.

- Peters, K., Schweizer, I., Beilharz, K., *et al* (2014). *Streptococcus pneumoniae* PBP2x mid-cell localization requires the C-terminal PASTA domains and is essential for cell shape maintenance. *Mol Microbiol* 92, 733-755.
- Pinho, M.G., De Lencastre, H., Tomasz, A (2001). An acquired and a native PBP cooperate in building the cell wall of drug-resistant staphylococci. *Proc Natl Acad Sci USA* 98, 10886–10891.
- Polidori, M., Nuccorini, A., Tascini, C., *et al* (2011). Vancomycin-resistant *Enterococcus faecium* (VRE) bacteremia in infective endocarditis successfully treated with combination daptomycin and tigecycline. *J Chemother* 234240-241.
- Ponting, C., Aravind, L., Schultz, J., Bork, P., Koonin, EV (1999). Eukaryotic signalling domain homologues in archaea and bacteria. Ancient ancestry and horizontal gene transfer. *J Mol Biol* 289, 729-745.
- Pooley, H.M., Karamata, D (1994). Teichoic acid synthesis in *Bacillus subtilis*: genetic organization and biological roles ( Amsterdam Elsevier) 27, 187-198.
- Pritchard, D., Dong, S., Baker, JR., Engler, JA (2004). The bifunctional peptidoglycan lysin of *Streptococcus agalactiae* bacteriophage B30. *Microbiology* (Reading, England) 150, 2079-2087.
- Priyadarshini, R., de Pedro, MA., Young, KD (2007). Role of peptidoglycan amidases in the development and morphology of the division septum in *Escherichia coli*. *J Bacteriol* 189, 5334-5347.
- Radutoiu, S., Madsen, L., Madsen, E., Felle, H., Umehara, Y., Grønlund, M., Sato, S., Nakamura, Y., Tabata, S., Sandal, N., Stougaard, J (2003). Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature* 425, 585-592.
- Radutoiu, S., Madsen, L., Madsen, E., Jurkiewicz, A., Fukai, E., Quistgaard, E., Albrektsen, A. S., James, E., Thirup, S., & Stougaard, J (2007). LysM domains mediate lipochitin-oligosaccharide recognition and *Nfr* genes extend the symbiotic host range. *EMBO J* 26, 3923-3935.
- Raghothama, S., Eberhardt, RY., Simpson, P., Wigelsworth, D., White, P., Hazlewood, GP., Nagy, T., Gilbert, HJ., Williamson, M.P (2001). Characterization of a cellulosome dockerin domain from the anaerobic fungus *Piromyces equi*. *Nat Struct Biol* 8, 775–778.
- Raymond, J.B., Mahapatra, S., Crick, D.C., Pavelka, M.S. Jr (2005). Identification of the namH gene, encoding the hydroxylase responsible for the N-glycolylation of the mycobacterial peptidoglycan. *J Biol Chem* 280, 326-333.
- Reed, M., Hounslow, A., Sze, K., Barsukov, I., Hosszu, L., Clarke, A., Craven, J., Waltho, J (2003). Effects of Domain Dissection on the Folding and Stability of the 43 kDa Protein PGK Probed by NMR. *J Mol Biol* 330(5), 1189-1201.

- Reichmann, N., Cassona, C., Gründling, A (2013). Revised mechanism of D-alanine incorporation into cell wall polymers in gram-positive bacteria. *Microbiology* (Reading, England) *159*, 1868-1877.
- Reichmann, N., Gründling, A (2011). Location, synthesis and function of glycolipids and polyglycerolphosphate lipoteichoic acid in gram-positive bacteria of the phylum Firmicutes. *FEMS Microbiol Lett* *319*, 97-105.
- Reste de Roca, F., Duché, C., Dong, S., Rincé, A, Dubost, L., Pritchard, D G., Baker, J., Arthur, M., and Mesnage, S (2010). Cleavage Specificity of Enterococcus faecalis EnpA (EF1473), a Peptidoglycan Endopeptidase Related to the LytM/Lysostaphin Family of Metallopeptidases. *J Mol Biol* *398*, 507–517.
- Rogers, H.J., Perkins, H.R., & Ward, J.B (1980). Microbial cell walls and membranes (London).
- Ruggiero, A., Squeglia, F., Marasco, D., Marchetti, R., Molinaro, A., Berisio, R (2011). X-ray structural studies of the entire extracellular region of the serine/threonine kinase PrkC from *Staphylococcus aureus*. *Biochem J* *435*, 33-41.
- Sabała, I., Jagielska, E., Bardelang, PT., et al (2014). Crystal structure of the antimicrobial peptidase lysostaphin from *Staphylococcus simulans*. *FEBS J* *281*, 4112-4122.
- Saksela, K., & Permi, Perttu (2012). SH3 domain ligand binding: What's the consensus and where's the specificity?. *FEBS Lett* *586*.
- Sanchez-Vallet, A., Saleem-Batcha, R., Kombrink, A., Hansen, G., Valkenburg, D. J., Thomma, B. P., Mesters, J. R (2013). Fungal effector Ecp6 outcompetes host immune receptor for chitin binding through intrachain LysM dimerization. *eLife* *2*, e00790.
- Sánchez-Vallet, A., Tian, H., Rodriguez-Moreno, L., et al (2020). A secreted LysM effector protects fungal hyphae through chitin-dependent homodimer polymerization. *PLoS Pathog* *16*, e1008652.
- Sarkar, P., Saleh, T., Tzeng, SR., Birge, RB., Kalodimos, CG (2011). Structural basis for regulation of the Crk signaling protein by a proline switch. *Nat Chem Biol* *7*, 51-57.
- Sauvage, E., Kerff, F., Terrak, M., Ayala, J.A., Charlier, P (2008). The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. *FEMS Microbiol Rev* *32*, 234-258.
- Schindler, C.A., & Schuhardt, V. T (1964). Lysostaphin: a new bacteriolytic agent for the *Staphylococcus*. *Proc Natl Acad Sci USA* *51*(3), 414–421.
- Schleifer, K., Kandler, O (1972). Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriol Rev* *36*, 407-477.
- Schneewind, O., Fowler, A., Faull, KF (1995). Structure of the cell wall anchor of surface proteins in *Staphylococcus aureus*. *Science* *268*, 103-106.

Schneider, T., Senn, M.M., Berger-Bächi, B., Tossi, A., Sahl, H.G., Wiedemann, I (2004). *In vitro* assembly of a complete, pentaglycine interpeptide bridge containing cell wall precursor (lipid II-Gly5) of *Staphylococcus aureus*. *Mol Microbiol* 53, 675-685.

Shah, I., Laaberki, MH., Popham, D., Dworkin, J. (2008). A eukaryotic-like Ser/Thr kinase signals bacteria to exit dormancy in response to peptidoglycan fragments. *Cell* 135, 486-496.

Sham, L.T., Butler, E.K., Lebar, M.D., Kahne, D., Bernhardt, TG., Ruiz, N (2014). MurJ is the flippase of lipid-linked precursors for peptidoglycan biogenesis. *Science* 345, 220–222.

Shaner, N.C., G. G. Lambert, A. Chammass, Y. Ni, P. J. Cranfill *et al* (2013). A bright monomeric green fluorescent protein derived from *Branchiostoma lanceolatum*. *Nat Methods* 10, 407–409.

Shockman, G.D., & Barrett, J.F. (1983). Structure, function, and assembly of cell walls of Gram-positive bacteria. *Annu Rev Microbiol* 37, 501–527.

Siegel, S.D., Reardon, M.E., Ton-That, H (2017). Anchoring of LPXTG-Like Proteins to the Gram-Positive Cell Wall Envelope. *Curr Top Microbiol Immunol* 404, 159-175.

Siezen, R., Boekhorst, J., Muscariello, L., Molenaar, D., Renckens, B., Kleerebezem, M (2006). *Lactobacillus plantarum* gene clusters encoding putative cell-surface protein complexes for carbohydrate utilization are conserved in specific gram-positive bacteria. *BMC Genomics* 7, 126.

Silhavy, T.J., Kahne, D., Walker, S (2010). The bacterial cell envelope. *Cold Spring Harb Perspect Biol* 2, a000414.

Smith, R.E., Salamaga, B., Szkuta, P., Hajdamowicz, N., Prajsnar, T. K., Bulmer, G. S., Fontaine, T., Kołodziejczyk, J., Herry, J. M., Hounslow, A. M., Williamson, M. P., Serror, P., & Mesnage, S (2019). Decoration of the enterococcal polysaccharide antigen EPA is essential for virulence, cell surface charge and interaction with effectors of the innate immune system. *PLoS Pathog* 15, e1007730.

Solheim, M., Brekke, MC., Snipen, LG., Willems, RJ., Nes, IF., Brede D (2011). Comparative genomic analysis reveals significant enrichment of mobile genetic elements and genes encoding surface structure-proteins in hospital-associated clonal complex 2 *Enterococcus faecalis*. *BMC Microbiol* 11, 3.

Stapleton, P.D., and Taylor, P. W (2002). Methicillin resistance in *Staphylococcus aureus*: mechanisms and modulation. *Sci Prog* 85, 57-72.

Stivala A, W.M., Wirth A, Whisstock J, Stuckey P (2011). Automatic generation of protein structure cartoons with Pro-origami Bioinformatics (Oxford, England) 27, 3315-3316.

Strandén, A.M., Ehlert, K., Labischinski, H. and Berger-Bächi, B (1997). Cell wall monoglycine cross-bridges and methicillin hypersusceptibility in a femAB null mutant of methicillin-resistant *Staphylococcus aureus*. *J Bacteriol* 179, 9–16.

- Sychantha, D., Brott, AS., Jones, CS., Clarke, AJ (2018). Mechanistic Pathways for Peptidoglycan O-Acetylation and De-O-Acetylation. *Front Microbiol* 9, 2332.
- Taguchi, A.e.a. (2019). FtsW is a peptidoglycan polymerase that is functional only in complex with its cognate penicillin-binding protein. *Nat Microbiol* 4 587–594.
- Tamai, E., Yoshida, H., Sekiya, H., Nariya, H., Miyata, S., Okabe, A., Kuwahara, T., Maki, J., Kamitori, S (2014). X-ray structure of a novel endolysin encoded by episomal phage phiSM101 of *Clostridium perfringens*. *Mol Microbiol* 92, 326–337.
- Tan, S., L., Nakao, H., He, Y., Vijaysri, S., Neddermann, P., Jacobs, B. L., Mayer, B. J., and Katze, M. G. (1999). NS5A, a nonstructural protein of hepatitis C virus, binds growth factor receptor-bound protein 2 adaptor protein in a Src homology 3 domain/ligand-dependent manner and perturbs mitogenic signaling. *Proc Natl Acad Sci USA* 96, 5533–5538.
- te Velthuis, A.J.W., Admiraal, J.F., and Bagowski, C.P (2007). Molecular evolution of the MAGUK family in metazoan genomes. *BMC Evol Biol* 7, 129.
- Teng, F., Jacques-Palaz, KD., Weinstock, GM., Murray, BE (2002). Evidence that the enterococcal polysaccharide antigen gene (epa) cluster is widespread in *Enterococcus faecalis* and influences resistance to phagocytic killing of *E. faecalis*. *Infect Immun* 70, 20102015.
- Theilacker, C., Holst, O., Lindner, B., Huebner, J., Kaczyński, Z (2012). The structure of the wall teichoic acid isolated from *Enterococcus faecalis* strain 12030. *Carbohydr Res* 354, 106-109.
- Thumm, G., & Götz, F (1997). Studies on prollystaphin processing and characterization of the lysostaphin immunity factor (Lif) of *Staphylococcus simulans* biovar *staphylolyticus*. *Mol Microbiol* 23, 1251-1265.
- Thurlow, L., Thomas, V., Hancock, L (2009). Capsular Polysaccharide Production in *Enterococcus faecalis* and Contribution of CpsF to Capsule Serospecificity. *J Bacteriol* 191, 6203-6210.
- Tipper, D.J., & Strominger, J. L. (1965). Mechanism of action of penicillins: a proposal based on their structural similarity to acyl-D-alanyl-D-alanine. *Proc Natl Acad Sci USA* 54, 1133–1141.
- Toh, H., Oshima, K., Nakano, A., et al (2013). Genomic adaptation of the *Lactobacillus casei* group. *PLoS One* 8, e75073.
- Tossavainen, H., Raulinaitis, V., Kauppinen, L., Pentikäinen, U., Maaheimo, H., & Permi, P (2018). Structural and Functional Insights Into Lysostaphin-Substrate Interaction. *Front Mol Biosci* 5, 60.
- Tschierske, M., Ehlert, K., Strandén, AM., Berger-Bächli, B (1997). Lif, the lysostaphin immunity factor, complements FemB in staphylococcal peptidoglycan interpeptide bridge formation. *FEMS Microbiol Lett* 153, 261-264.



- Turner, R.D., Vollmer, W., Foster, S.J (2014). Different walls for rods and balls: the diversity of peptidoglycan. *Mol Microbiol* 9, 862-874.
- Typas, A., Banzhaf, M., Gross, C.A., & Vollmer, W (2011). From the regulation of peptidoglycan synthesis to bacterial growth and morphology. *Nat Rev Microbiol* 100, 123–136.
- Ursinus, A., van den Ent, F., Brechtel, S., de Pedro, M., Höltje, J.V., Löwe, J., Vollmer, W (2004). Murein (peptidoglycan) binding property of the essential cell division protein FtsN from *Escherichia coli*. *J Bacteriol* 186, 6728-6737.
- van den Brink-van der Laan, E., Boots, J.W., Spelbrink, R.E., *et al* (2003). Membrane interaction of the glycosyltransferase MurG: a special role for cardiolipin. *J Bacteriol* 185, 3773-3779.
- van Heijenoort, J. (2001). Recent advances in the formation of the bacterial peptidoglycan monomer unit. *Nat Prod Rep* 18, 503–519.
- Varea J, S.J., López-Zumel C, *et al* (2000). Do sequence repeats play an equivalent role in the choline-binding module of pneumococcal LytA amidase? *J Biol Chem* 275, 26842-26855.
- Veiga, P., Piquet, S., Maisons, A., Furlan, S., Courtin, P., Chapot-Chartier, M. P., *et al* (2006). Identification of an essential gene responsible for D-Asp incorporation in the *Lactococcus lactis* peptidoglycan crossbridge. *Mol Microbiol* 62, 1713–1724.
- Vogel, C., Bashton, M., Kerrison, N.D., Chothia, C., and Teichmann, S.A (2004). Structure, function and evolution of multidomain proteins. *Curr Opin Struct Biol* 14, 208-216.
- Vollmer, W. (2008b). Structural variation in the glycan strands of bacterial peptidoglycan. *FEMS Microbiol Rev* 32, 287–306.
- Vollmer, W., & Seligman, S.J (2010). Architecture of peptidoglycan: more data and more models. *Trends Microbiol* 18, 59-66.
- Vollmer, W., & Tomasz, A (2000). The *pgdA* gene encodes for a peptidoglycan N-acetylglucosamine deacetylase in *Streptococcus pneumoniae*. *J Biol Chem* 275, 20496-20501.
- Vollmer, W., Blanot, D., & de Pedro, M.A (2008a). Peptidoglycan structure and architecture. *FEMS Microbiol Rev* 32, 149-167.
- Vollmer, W., Joris, B., Charlier, P., & Foster, S.J (2008c). Bacterial peptidoglycan (murein) hydrolases. *FEMS Microbiol Rev* 32, 259-286.
- Wang, M., Caetano-Anollés, G (2009). The evolutionary mechanics of domain organization in proteomes and the rise of modularity in the protein world. *Structure* 17, 66-78.
- Ward, J. (1981). Teichoic and teichuronic acids: biosynthesis, assembly, and location. *Microbiol Rev* 45, 211-43.

Waudby, C.A., Ramos, A., Cabrita, L. D. & Christodoulou, J (2016). Two-Dimensional NMR Lineshape Analysis. *Sci Rep* 6, 24826.

Wheeler, R., Mesnage, S., Boneca, IG., Hobbs, JK., Foster, SJ (2011). Super-resolution microscopy reveals cell wall dynamics and peptidoglycan architecture in ovococcal bacteria. *Mol Microbiol* 82, 1096-1109.

Whisstock, J., & Lesk, AM (1999). SH3 domains in prokaryotes. *Trends Biochem Sci* 24, 132–133.

Whitehead, B., Craven, CJ., Waltho, JP (1997). Double and triple resonance NMR methods for protein assignment. *Methods Mol Biol* 60, 29-52.

Wietzerbin, J., Das, BC., Petit, JF., Lederer, E., Leyh-Bouille, M. & Ghuysen, JM (1974). Occurrence of D-alanyl-(D)-meso-diaminopimelic acid and meso-diaminopimelyl-meso-diaminopimelic acid interpeptide linkages in the peptidoglycan of Mycobacteria. *Biochemistry* 13, 3471–3476.

Williams, K., Yahashiri, A., Arends, SJ., Popham, DL., Fowler, CA., Weiss, DS (2013). Nuclear magnetic resonance solution structure of the peptidoglycan-binding SPOR domain from *Escherichia coli* DamX: insights into septal localization. *Biochemistry* 52, 627-639.

Williamson, M.P. (2013). Using chemical shift perturbation to characterise ligand binding. *Prog Nucl Magn Reson Spectrosc* 73, 1–16.

Willmann, R., Lajunen, H. M., Erbs, G., Newman, M. A., Kolb, D., Tsuda, K., Katagiri, F., Fliegmann, J., Bono, J. J., Cullimore, J. V., Jehle, A. K., Götz, F., Kulik, A., Molinaro, A., Lipka, V., Gust, A. A., Nurnberger, T (2011). Arabidopsis lysin-motif proteins LYM1 LYM3 CERK1 mediate bacterial peptidoglycan sensing and immunity to bacterial infection. *Proc Natl Acad Sci USA* 108, 19824-19829.

Wisplinghoff, H., Bischoff, T., Tallent, SM., Seifert, H., Wenzel, RP., Edmond, MB (2004). Nosocomial bloodstream infections in US hospitals: analysis of 24,179 cases from a prospective nationwide surveillance study *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America* 39, 309-317.

Wong, J., Midtgaard, SR., Gysel, K, *et al* (2015). An intermolecular binding mechanism involving multiple LysM domains mediates carbohydrate recognition by an endopeptidase. *Acta Crystallogr D Biol Crystallogr* 71, 592-605.

Xu, Q., Abdubek, P., Astakhova, T., *et al* (2010). Structure of the gamma-D-glutamyl-L-diamino acid endopeptidase YkfC from *Bacillus cereus* in complex with L-Ala-gamma-D-Glu: insights into substrate recognition by NlpC/P60 cysteine peptidases. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 66, 1354-1364.

Xu, Q., Mengin-Lecreulx, D., Liu, X. W., Patin, D., Farr, C. L., Grant, J. C., Chiu, H. J., Jaroszewski, L., Knuth, M. W., Godzik, A., Lesley, S. A., Elsliger, M. A., Deacon, A. M., & Wilson, I. A (2015).

Insights into Substrate Specificity of NlpC/P60 Cell Wall Hydrolases Containing Bacterial SH3 Domains. *mBio* 6, e02327-02314.

Xu, Q., Sudek, S., McMullan, D., Miller, M. D., Geierstanger, B., Jones, D. H., Krishna, S. S., Spraggon, G., *et al* (2009). Structural basis of murein peptide specificity of a gamma-D-glutamyl-L-diamino acid endopeptidase. *Structure* 17, 303-313.

Yahashiri, A., Jorgenson, MA., Weiss, DS (2017). The SPOR Domain, a Widely Conserved Peptidoglycan Binding Domain That Targets Proteins to the Site of Cell Division. *J Bacteriol* 199, e00118-00117.

Yahashiri, A., Jorgenson, MA., Weiss, DS (2015). Bacterial SPOR domains are recruited to septal peptidoglycan by binding to glycan strands that lack stem peptides. *Proc Natl Acad Sci USA* 112, 11347-11352.

Yang, D.W.K., L. E (1999). Improved (HN)-H-1-detected triple resonance TROSY-based experiments. *J Biomol NMR* 13, 3-10.

Yang, J., Van, Den., Ent, F., Neuhaus, D., Brevier, J., Löwe J (2004). Solution structure and domain architecture of the divisome protein FtsN. *Mol Microbiol* 52, 651-660.

Yang, S., Doolittle, R.F., and Bourne, P.E (2005). Phylogeny determined by protein domain content. *Proc Natl Acad Sci USA* 102, 373-378.

Yeats, C., Finn, RD., Bateman, A (2002). The PASTA domain: a beta-lactam-binding domain. *Trends Biochem Sci* 27, 438.

Yousif, S.Y., Broome-Smith, J. K. & Spratt, B. G (1985). Lysis of *Escherichia coli* by  $\beta$ -lactam antibiotics: deletion analysis of the role of penicillin-binding proteins 1A and 1B. *J Gen Microbiol* 131, 2839–2845.

Yu, H., Chen, JK., Feng, S., Dalgarno, DC., Brauer, AW., Schreiber, SL (1994). Structural basis for the binding of proline-rich peptides to SH3 domains. *Cell* 76, 933-945.

Yu, W., Missiakas, D., Schneewind, O (2018). Septal secretion of protein A in *Staphylococcus aureus* requires SecA and lipoteichoic acid synthesis. *eLife* 7, e34092.

Zoll, S., Schlag, M., Shkumatov, AV., *et al* (2012). Ligand-binding properties and conformational dynamics of autolysin repeat domains in staphylococcal cell wall recognition. *J Bacteriol* 194, 3789-3802.

Zuiderweg, E.R.P. (2002). Mapping Protein-Protein Interactions in Solution by NMR Spectroscopy. *Biochemistry* 41, 1-7.

Zwart, P.H., Afonine, P.V., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., McKee, E., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K., Storoni, L.C., Terwilliger, T.C., Adams, P.D (2008). Automated structure solution with the PHENIX suite. *Methods Mol Biol* 426, 419-435.

## **Appendix A**

### **Characterisation of synthetic PG ligands**

# GMGM

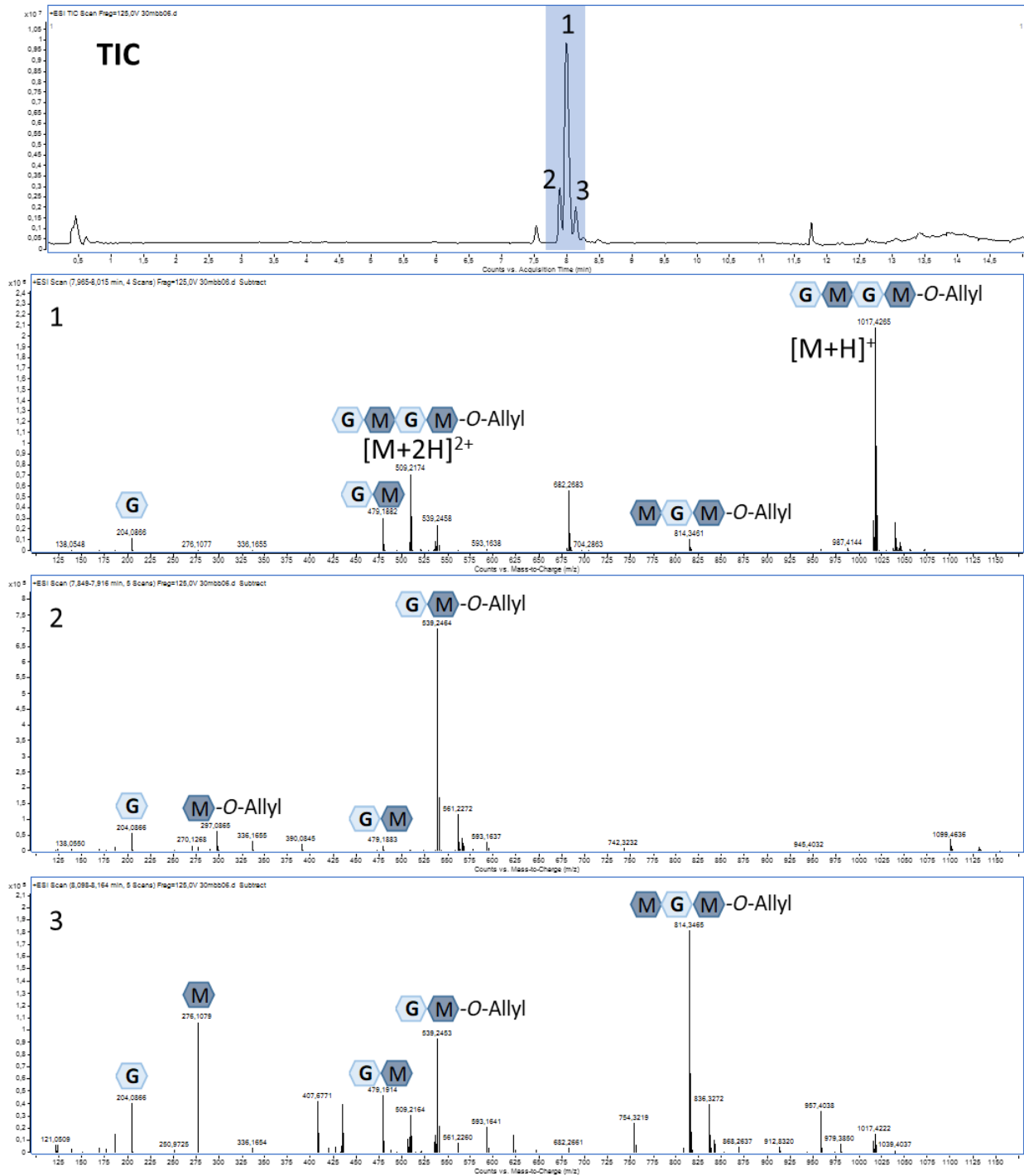


Figure 1A. LC-MS of tetrasaccharide [GM-GM] ligand.

# G5



## Gly-Gly-Gly-Gly-Gly

1 Product Result | Match Criteria: Keyword

Properties ▾

**Gly-Gly-Gly-Gly-Gly** Synonym: Glycyl-glycyl-glycyl-glycyl-glycine, Pentaglycine  
Linear Formula:  $\text{NH}_2\text{CH}_2\text{CO}(\text{NHCH}_2\text{CO})_3\text{NHCH}_2\text{COOH}$  | Molecular Weight: 303.27 | CAS Number: 7093-67-6

**G5755** Sigma-Aldrich [SDS](#) [Close](#)

SKU-Pack Size	Availability	Price (GBP)	Quantity
G5755-25MG	<span style="color: green;">✔</span> Available to ship on 31.07.19 - FROM	24.50	0 <span style="color: blue;">★</span> <span style="color: blue;">i</span>

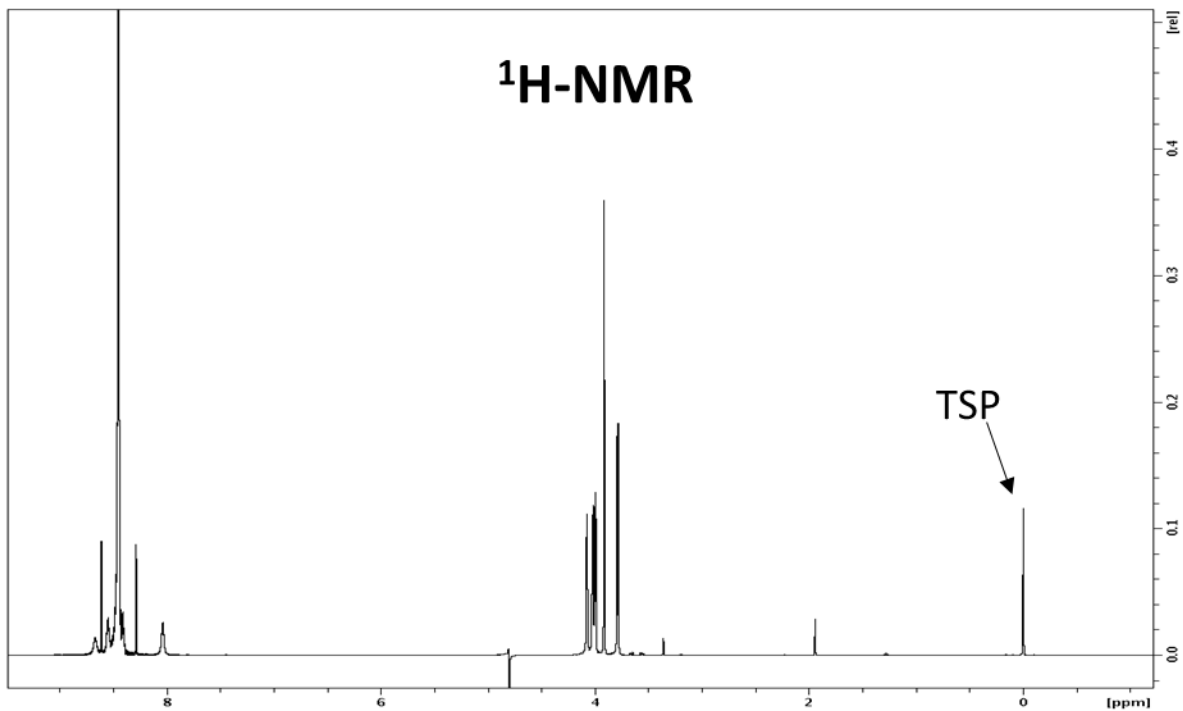
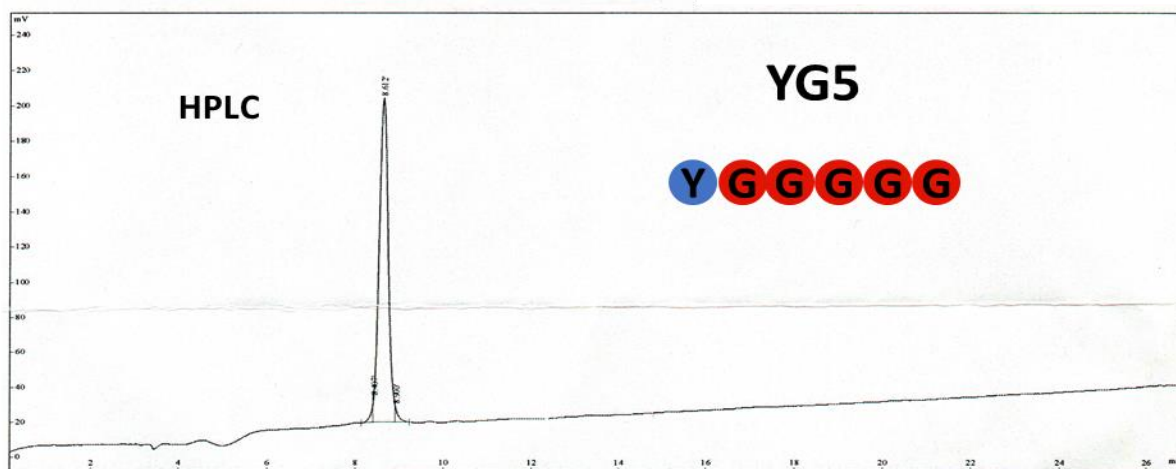


Figure 2A. Characterisation of pentaglycine [G5] ligand.

Column : 4.6\*250mm, Boston Crest ODS  
 Solvent A : 0.1% trifluoroacetic in 100% acetonitrile  
 Solvent B : 0.1% trifluoroacetic in 100% water  
 Gradient :
 

	A	B
0.01min	2%	98%
25min	27%	73%
25.1min	100%	0%
30min		STOP

Flow rate : 1.0ml/min  
 Wavelength : 220nm  
 Volume : 10  $\mu$ l



Peak No.	Ret Time	Conc	Area	Height
1	8.407	2.06	55559	15947
2	8.612	96.61	2609403	184906
3	8.900	1.33	36010	7560

100

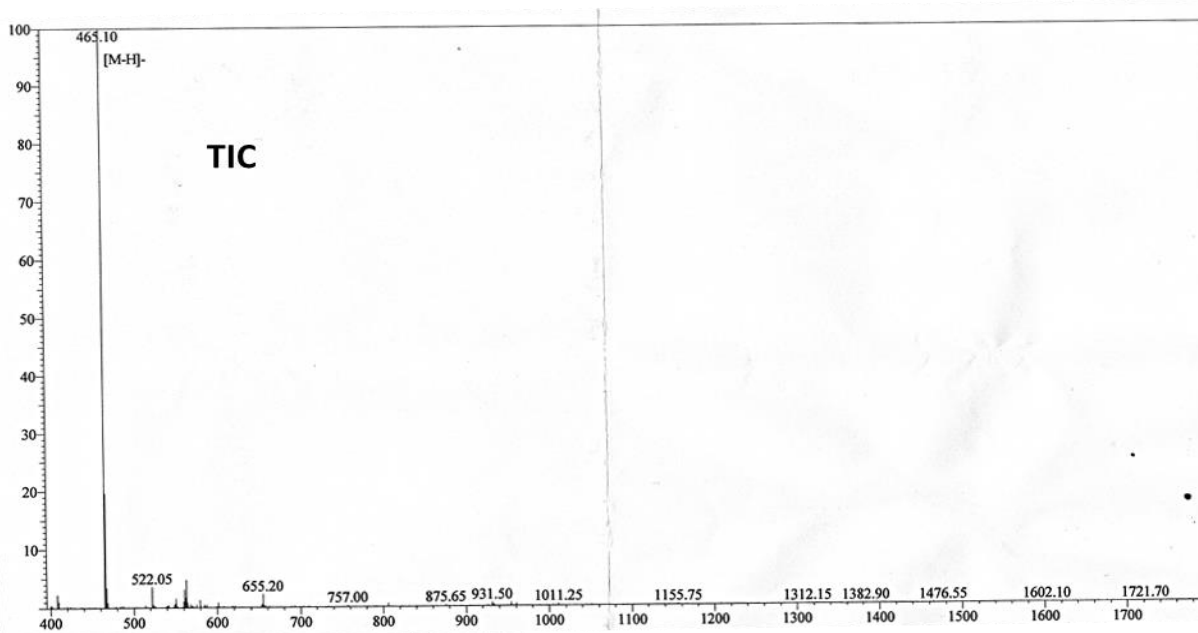


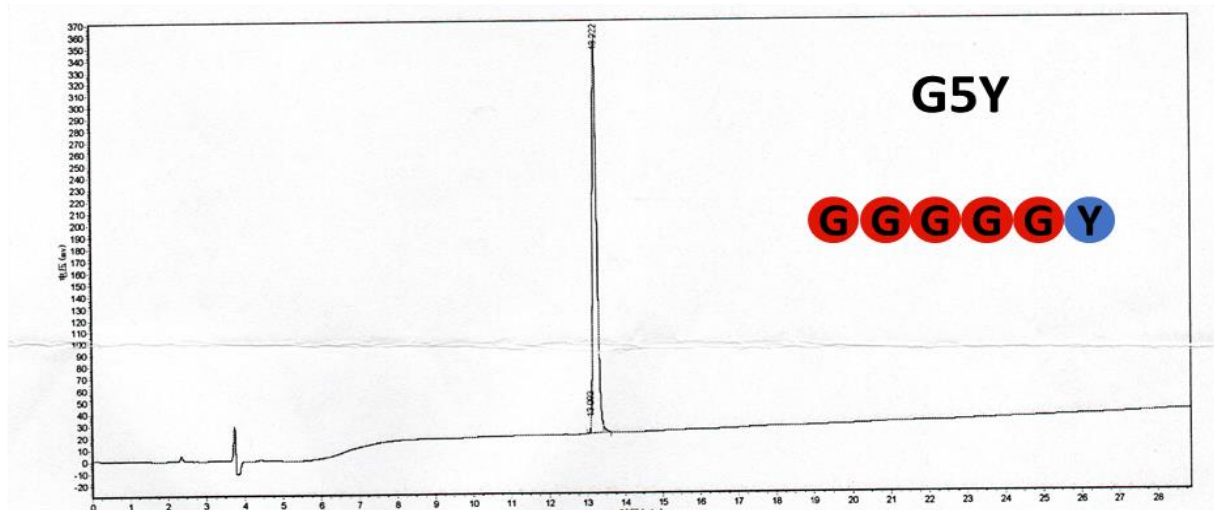
Figure 3A. Characterisation of pentaglycine derivative with an extra tyrosine residue at the N-terminus [YG5] ligand.



Column : 4.6\*250mm, 5um, 100A, Agela  
 Solvent A : 0.1% trifluoroacetic in 100% acetonitrile  
 Solvent B : 0.1% trifluoroacetic in 100% water  
 Gradient
 

	A	B
0.01min	4%	96%
25min	29%	71%
25.1min	100%	0%
30min	STOP	

Flow rate : 1.0ml/min  
 Wavelength : 220nm  
 Volume : 10ul



Peak No.	Ret Time	Height	Area	Conc.
1	13.093	6713.476	7116.408	0.2671
2	13.222	332271.531	2630591.000	98.7410
3	13.222	7623.025	26424.465	0.9919

100

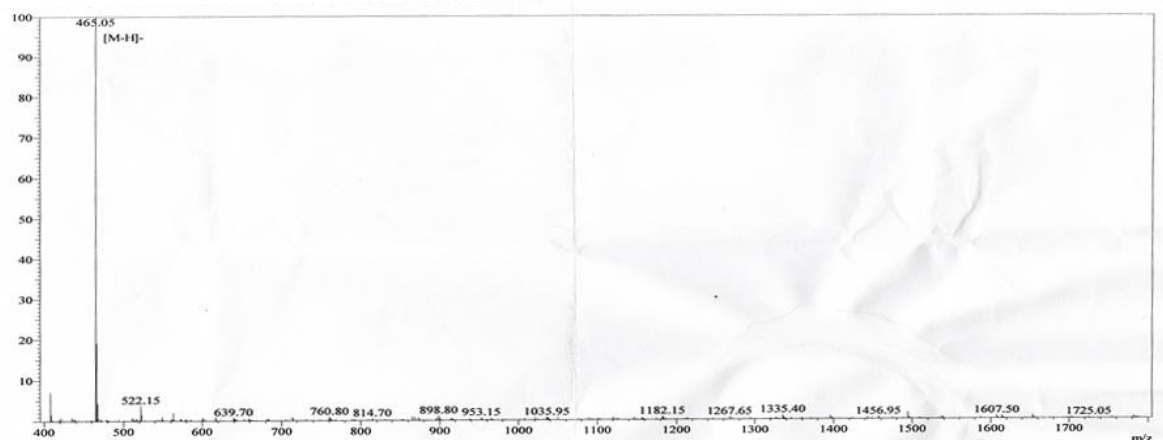


Figure 4A. Characterisation of pentaglycine derivative with an extra tyrosine residue at the C-terminus [G5Y] ligand.

Column: Aeris peptide 3.6u XB-C18 100A  
 Method: C:\CHEM32\2\METHODS\C2\_FA0-80\_AERIS\_20.M  
 Instrument: Instrument 2  
 Flow Rate: 1.0mls/min  
 Injection Volume: 20ul  
 Method Info: Analysis carried out using a 100A 4.6 x 150mm column, gradient from 0% - 80% Acetonitrile, in 15 minutes. At 20C.

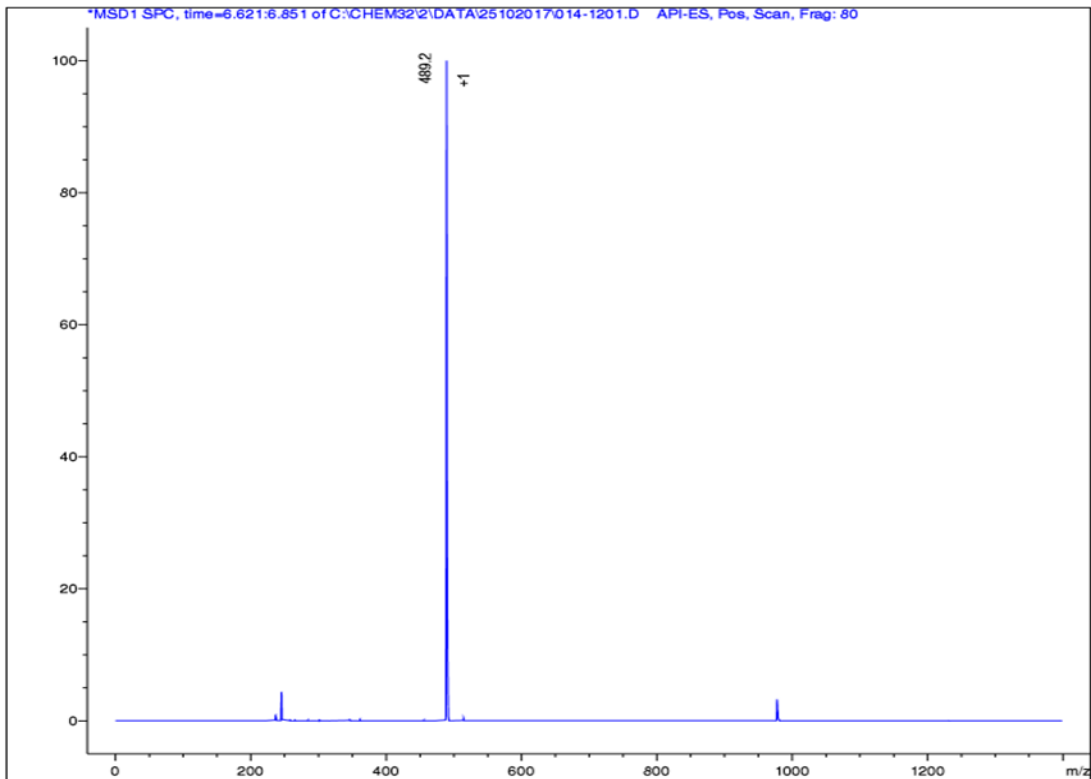
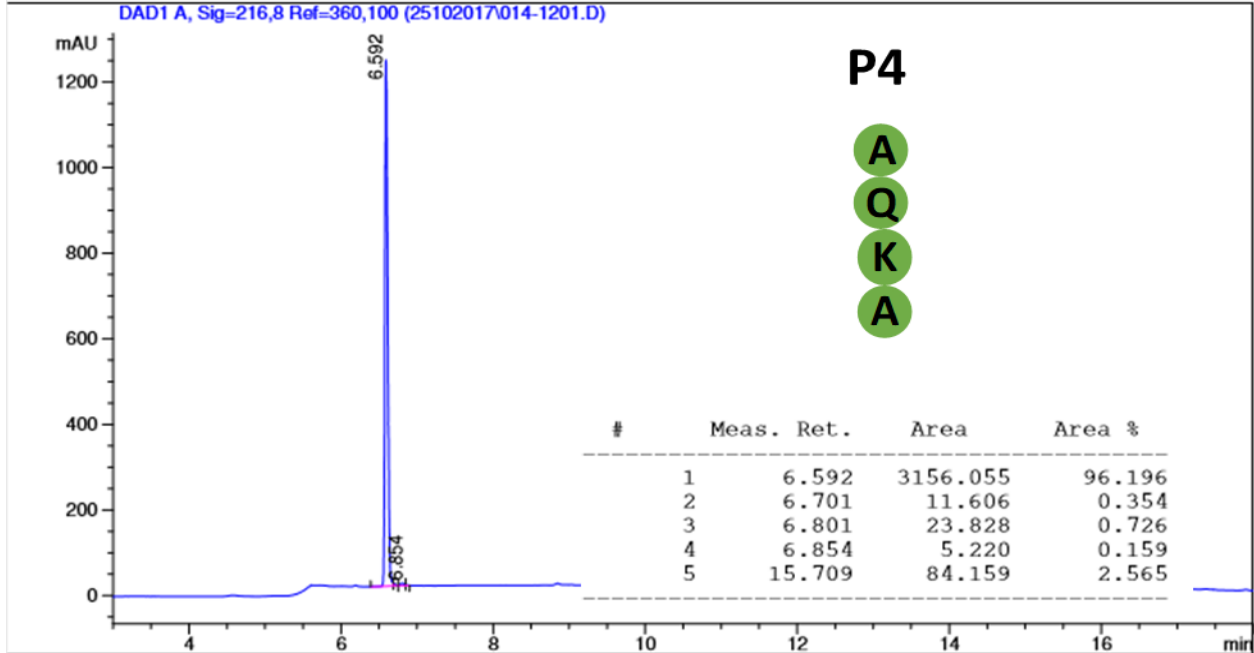


Figure 5A. Characterisation of tetrapeptide stem with a lactyl group [P4] ligand.

Column: Aerie peptide 3.6u XB-C18 100A  
 Method: C:\CHEM32\1\METHODS\C2\_FA0-80\_AERIS\_20.M  
 Instrument: Instrument 2  
 Flow Rate: 1ml/min  
 Injection Volume: 20ul  
 Method Info: Analysis carried out using a 100A 4.6 x 50mm column, gradient from 0% - 80% Acetonitrile, in 8 minutes.

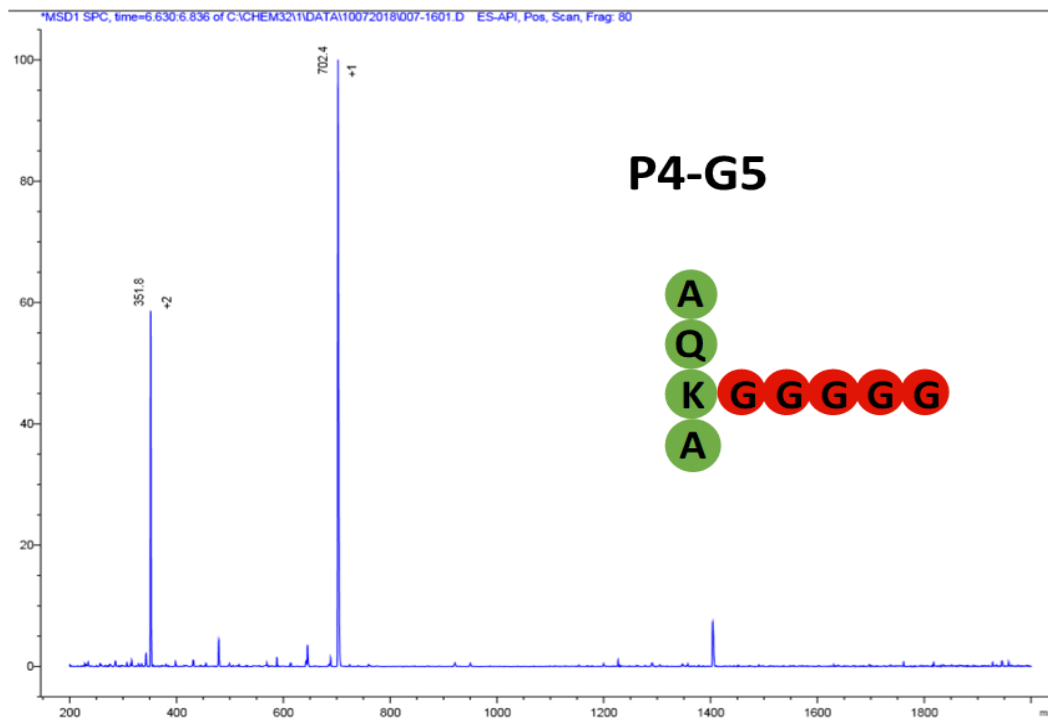
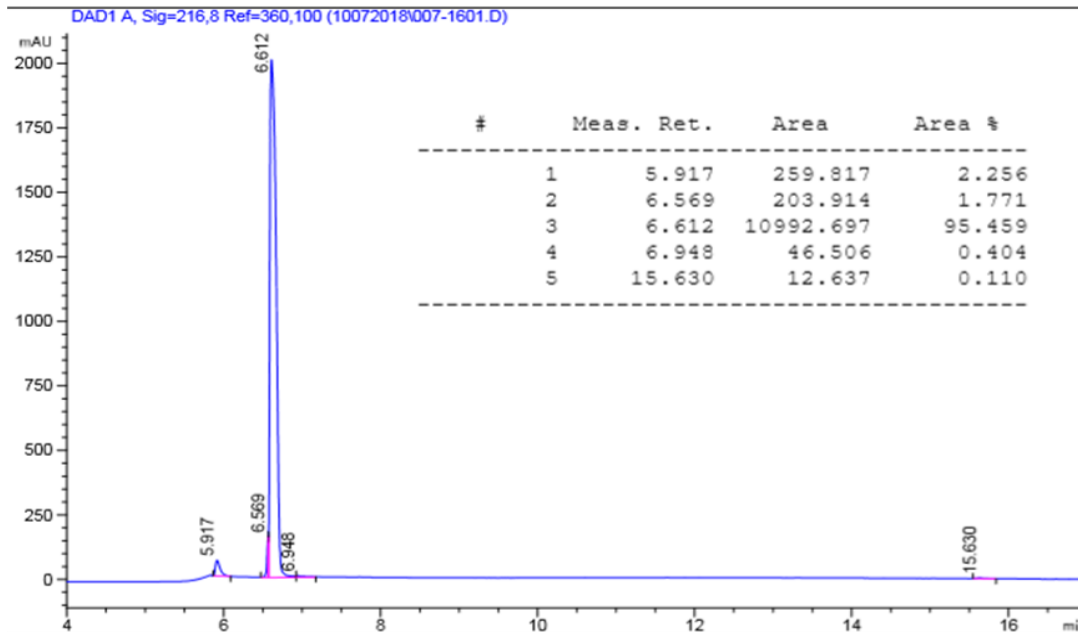


Figure 6A. Characterisation of tetrapeptide stem with no lactyl group linked to a pentaglycine peptide as a lateral chain [P4-G5] ligand.

Column: Kinetex XB-C18 2.6u 100A  
 Method: C:\CHEM32\1\METHODS\C1\_FA0-80\_60.M  
 Instrument: Instrument 2  
 Flow Rate: 1.5mls/min  
 Injection Volume: 20ul  
 Method Info: Analysis carried out using a 100A 4.6 x 50mm column, gradient from 10% - 90% Acetonitrile, in 8 minutes. At 60C. Sample dissolved in DMSO.

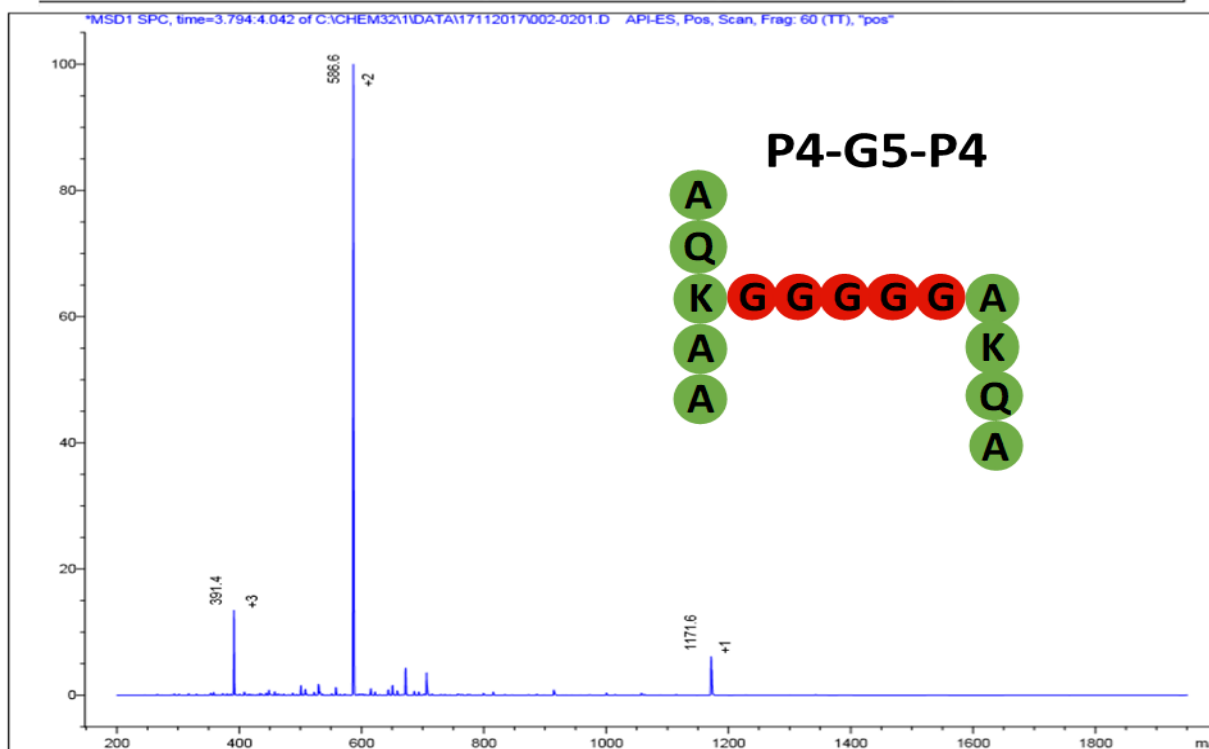
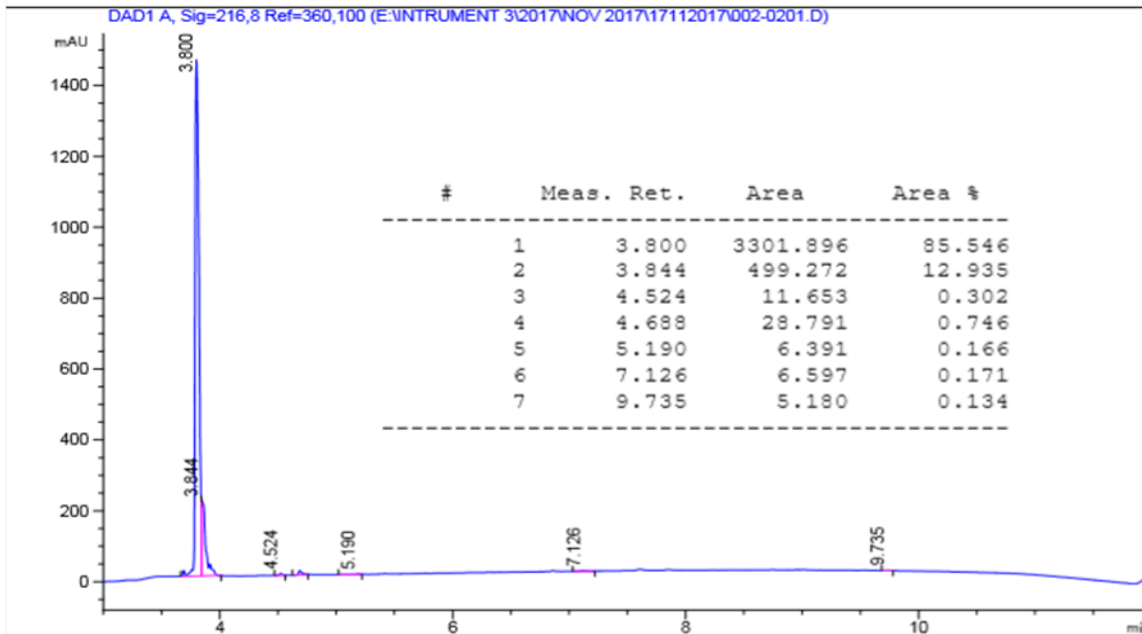


Figure 7A. Characterisation of pentapeptide stem crosslinked to a tetrapeptide stem through a pentaglycine bridge [P5-G5-P4] ligand.

## Appendix B

Presented here is the accepted manuscript “Two-site recognition of *Staphylococcus aureus* peptidoglycan by lysostaphin SH3\_5” deposited in White Rose Research Online (WRRO) open access repository (04 Nov 2019) originally published in *Nature Chemical Biology*.