Structural and functional studies on lysostaphin, an antistaphylococcal endopeptidase

Sophie Rochette,

Msci (Fr. DEA)

"Environmental Sciences, Chemistry and Health, Major: Health"

Thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy

September 2008

Contents

Abstract

This PhD thesis describes research into the structure and function of lysostaphin (EC 3.4.24.75), a glycylglycine endopeptidase secreted by *Staphylococcus simulans* biovar *staphylolyticus* ATCC 1362. Lysostaphin is a member of the M23/M37 zinc metalloprotease family and is a pre-pro-enzyme. The mature form (after removal of the pro-region) contains two distinct domains, the *C*-terminal cell wall targeting domain of lysostaphin (termed LssTdom in the thesis) facilitates binding to *Staphylococcus aureus* cells. The endopeptidase domain (termed LssEdom in the thesis) cleaves the pentaglycine crosslinks in the peptidoglycan resulting in cell death through cell rupture of *S. aureus*.

Lysostaphin is a potential therapeutic antibiotic for Methicillin-resistant *Staphylococcus aureus* (MRSA) for which new antibacterials are required owing to the widespread occurrence of multi-drug resistant strains. To date, the structural requirements for enzymatic activity and the target in the cell wall for lysostaphin have not been fully elucidated. Thus a structure might enable rationally guided design of lysostaphin variants for the generation of new enzymes which would also cleave the non-canonical crosslinks and would thus overcome bacterial resistance.

Thus, one approach was to obtain the lysostaphin structure using homology modelling. Lysostaphin shares significant homology with the ALE-1 (83 % identity) and LytM (48 % identity) bacteriocins, and modelling the structure of lysostaphin was achieved using the recently released structures of LytM and ALE-1 derived from X-ray crystallography as templates. In

addition, we report the successful production of active recombinant lysostaphin (as well as the endopeptidase and the targeting domains) and initial characterisation of their secondary structure by Fourier-transform infrared spectroscopy.

Initially, advanced spectroscopic techniques including X-ray and NMR methods were investigated for molecular interaction studies between Lss and its putative ligands. Significant problems were encountered with these methods and mass spectrometry studies proved more amenable. Lysostaphin targeting domain-ligand complexes have been identified, along with their stoichiometry. The strength of the protein-ligand interactions has also been quantified. Lysostaphin was shown to bind *in vitro* Gly_5 (mimicking the pentaglycine cross-bridge) and Lys-D-Ala-D-Ala (mimicking the stem peptide) with low affinity, but not NAM-L-Ala-D-iGln-Lys. It was also shown that lysostaphin targeting domain affinity for Gly_5 was significantly reduced by addition of Gly-Gly-Ser-Gly-Ser (found in the host bacteria resistant to lysostaphin action – *S. simulans*) in solution. From these studies it could be concluded that resistance due to the incorporation of serine residues in the crossbridge were a result of the endopeptidase domain being unable to cleave this sequence and not due to the targeting domain being unable to bind it.

Acknowledgements

I dedicate this work to my parents. A maman, Alex, Fred et Manu pour leur amour.

I would like to thank several people who helped in completing this thesis. My supervisors past and present: Dr F. Villalba gave me fundamental skills and Dr S. Blanquet was exemplary in many ways, Prof N.R. Thomas and Prof R. James gave me the great opportunity to do a PhD in England. I appreciated their flexibility as they helped me go and learn from other institutes and groups.

I also would like to thank the University of Nottingham for funding my PhD and the School of Chemistry for awarding me a competitive BEST (Business, Engineering and Science, and Travel) scholarship to fund my participation in the Cold Spring Harbour Laboratory Workshop entitled Phage Display of Proteins and Peptides.

Dr M. Paoli and his student Dr S. Schneider, have contributed to the crystallisation studies and Dr N. Oldham and his student J. Hopper to the interaction studies. I am indebted to Dr J-P. Duneau and M. Mazanetz for their altruistic contributions to the structural studies.

A particular thank to Dr R. Warfield, for her kindness, for providing me with substrate, and for correcting my English, Dr J. Long for running preliminary NMR studies, J. Heep and G. Coxhill for technical support and Dr P. Bardelang, J. Sharpe, Dr M. Vankemmelbeke, Prof P. Soultanas, Prof M. Searle for useful advice. A special thank you to W.Chen, the M.Sci. student who worked on the project, for her dynamism and her curiosity.

My friends Anne, Aurélie, Delphine, Isabelle and Rebecca deserve great credit for sticking with me throughout my studies and they did not deny me their moral support. Anne-Laure, Cristina, Karin, Laetitia, Stephanie and Stephan deserve a special thank for helping me in my final year. My door will always be open for them.

Manuel kept me on track with his advice and ideas, his sense of humour and “art de vivre”. If it was not for his love and his care, I could not have gotten so far.

Abbreviations

AA	amino acid
Abs, A	absorbance
Abz	*N*-(2-aminobenzoyl)
ATR	attenuated total reflection
B. anthracis	*Bacillus anthracis*
C	concentration
CD	circular dichroism
(L-R) CPL	(left-right) circularly polarised light
Da	Dalton
DMSO	dimethyl sulfoxide
DN	scale down
DNA	deoxyribonucleic acid
DSC	differential scanning calorimetry
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDDnp	*N*-(2,4-dinitrophenyl) ethylene
E. faecalis	*Enterococcus faecalis* LMG2333
EnlA	enterolysin A
Epr	endopeptidase resitance factor
ESI	electrospray ionisation
FPLC	fast protein liquid chromatography
FRET	(Föster) fluorescence resonance energy transfer
FT-IR	Fourier-transform infrared

GCP	gel permeation chromatography
GISA	glycopeptide-resitant *Staphylococcus aureus*
GFP	green fluorescent protein
GST	glutathione S-transferase
HPLC	high performance liquid chromatography
HSQC	heteronuclear single quantum coherence
IC_{50}	half maximal inhibitory concentration
IMAC	immobilised metal affinity chromatography
IPTG	isopropyl-β-D-thiogalactopyranoside
ITC	isothermal titration calorimetry
K_D	dissociation constant
K_M	Michaelis constant
[L]	free ligand
L	ladder
LB	Luria-Bertani
Lif	lysostaphin immunity factor
Lss	lysostaphin
LssEdom	lysostaphin endopeptidase domain
LssTdom	lysostaphin targeting domain
LTA	lipoteichoic acid
M	molar
M. luteus	*Micrococcus luteus*
(M9)MM	M9 minimum medium
MRSA	methicillin-resistant *Staphylococcus aureus*
MS	mass spectrometry

NAG, G	*N*-acetylglucosamine
NAM, M	*N*-acetylmuramic acid
NMR	nuclear magnetic resonance
NTA	nitrile acetate agarose
OD	optical density (absorbance per unit distance)
(o-)ToF	(orthogonal) time of flight
[P]	free protein
[PL]	protein-ligand complex
PAGE	polyacrylamide gel electrophoresis
PBP	penicillin-binding protein
PCR	polymerase chain reaction
PDB	protein database
PEG	polyethylene glycol
PEP	phosphoenol pyruvate
pep3	tripeptide
pep5	pentapeptide
PGRP	peptidoglycan recognition protein
QMC	Queens medical centre
RFU	relative fluorescence unit
RMSD	root mean square deviation
RNA	ribonucleic acid
S. aureus	*Staphylococcus aureus*
S. capitis	*Staphylococcus capitis*
S. carnosus	*Staphylococcus carnosus*
SDS	Sodium Dodecyl Sulphate

S. epidermidis	*Staphylococcus epidermidis*
S. gordonii	*Streptococcus gordonii*
SH3	Src homology 3 domain
S. mutans	*Streptococcus mutans*
SPR	surface plasmon resonance
S. pyogenes	*Streptococcus pyogenes*
S. saprophyliticus	*Staphylococcus saprophylyticcus*
S. simulans	*Staphylococcus simulans* biovar *staphylolyticus* ATCC1362
S. zooepidemicus	*Streptococcus* equi subsp. *zooepidemicus*
TFA	trifluoroacetic acid
TLN	thermolysin
UDP	uridine diphosphate
UP	scale up
UV	ultra violet
VISA	vancomycin-intermediate *Staphylococcus aureus*
VRSA	vancomycin-resistant *Staphylococcus aureus*
Vmax	maximal velocity
v/v	volume per volume
w/v	weight per volume
WTA, TA	wall teichoic acid
Zif	zoocin immunity factor

List of amino acids and their code:

A, Ala	alanine
C, Cys	cysteine
D, Asp	aspartic acid
E, Glu	glutamic acid
F, Phe	phenylalanine
G, Gly	glycine
H, His	histidine
I, Ile	isoleucine
K, Lys	lysine
L, Leu	leucine
M, Met	methionine
N, Asn	asparagine
P, Pro	proline
Q, Gln	glutamine
R, Arg	arginine
S, Ser	serine
T, Thr	threonine
V, Val	valine
W, Trp	tryptophan
Y, Tyr	tyrosine

1. Introduction

1.1. Peptidoglycan structure and biosynthesis

1.1.1. Targeting bacterial cell wall

Peptidoglycan is a vital component of virtually all bacterial cells and accounts for approximately 50% of the weight of Gram-positive (staphylococci, streptococci) bacterial walls. Figure 1 shows a schematic of the cell envelope of a typical Gram-positive bacterium. The peptidoglycan polymer is responsible for the shape, mechanical strength and integrity of bacterial cells. If the peptidoglycan is disrupted or its synthesis is selectively blocked by antibiotic action, bacteria undergo a number of changes in shape and they ultimately die as a result of cell lysis. As mammalian cells do not possess a cell wall or macromolecular structure that ressembles peptidoglycan, antibiotics that act on peptidoglycan have excellent selective toxicity.

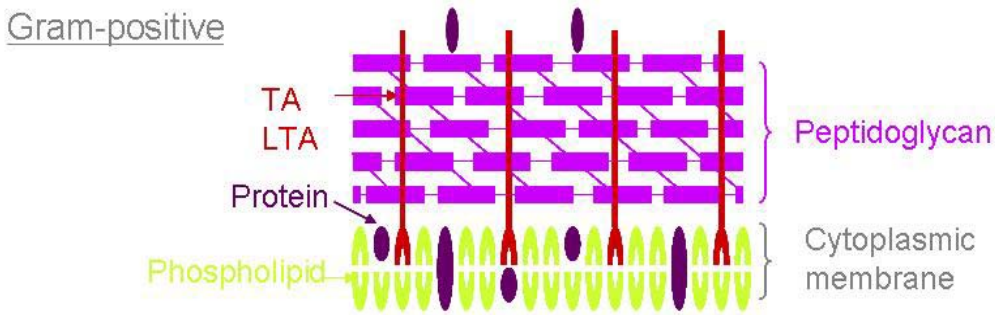


Figure 1: Schematic of the cell wall envelope of a typical Gram-positive bacterium[...].

In this study, we focused on staphylococci and particularly on *Staphylococcus aureus*, because it is a pathogenic Gram-positive bacterium that is a significant cause of healthcare infections. *S. aureus* is a spherical bacterium (coccus) which on microscopic examination appears in pairs, short

[...]Adapted from: http://student.ccbcmd.edu/courses/bio141/lecguide/uniH/prostruct/gpcw.html

chains, or bunched, grape-like clusters. It is found on the mucous membranes and on the skin of approximately a third of the human population and owes its name (aureus) to the golden colour of its colonies when grown on solid media. In susceptible hosts: neonates, immunologically suppressed and post-operative patients, *S. aureus* can cause serious hospital-acquired infections including: pneumonia, endocarditis, osteomyelitis, skin and soft tissue infections.

1.1.2. The main constituents of peptidoglycan

The peptidoglycan of *S. aureus* contains glycan chains, which are cross-linked by short peptide bridges. Alternating units of *N*-acetylglucosamine (NAG; G) and *N*-acetylmuramic acid (NAM; M) joined by β-1,4 glycosidic linkages constitute the glycan chains (20 disaccharides in length) (Schleifer and Kandler, 1972). Each M unit has a short "stem" peptide chain attached and this is cross-linked by a short bridging peptide chain. These peptide chains are species specific. In *S. aureus*, the short stem peptides are composed of L-alanine, D-isoglutamic acid, L-lysine and two D-alanine units (L-Ala-D-iGlu-L-Lys-D-Ala-D-Ala) and are linked via an amide bond to the D-lactyl moiety of *N*-acetylmuramic acid (Figure 2). Neighbouring wall peptides are cross-linked between the free ε-amino group of L-lysine in one peptide and the carboxyl group of the penultimate D-alanine in another via pentaglycine cross bridges (-Gly_5-), whereby the D-Ala in position 5 is split off during transpeptidation. Together, the glycan strands and cross-linked wall peptides generate the three-dimensional exoskeletal network of peptidoglycan. The recently reported NMR structure of peptidoglycan provided a right-handed helical conformation of the polysaccharide backbone producing a threefold

symmetry in the molecule (Meroueh *et al.*, 2006). This result led to the proposed honeycomb pattern for the cell wall.

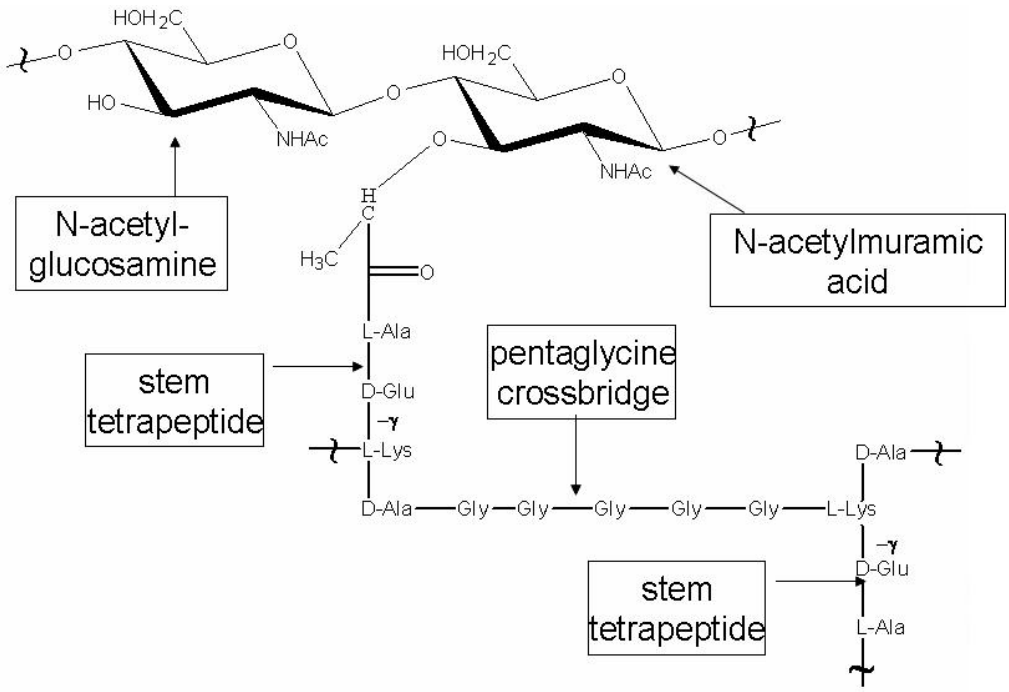


Figure 2: Structure of *S. aureus* peptidoglycan sheet [$].

The high cross-linkage (80%) of the peptides between the glycan chains gives the peptidoglycan macromolecule the mechanical strength needed to resist the high osmotic pressure that exists on bacterial cell membrane. It also acts as attachment site for various cell wall-sorted proteins, which are important in infection and virulence of this pathogen (Perry *et al.*, 2002, Roche *et al.*, 2003).

1.1.3. Stages of peptidoglycan biosynthesis

S. aureus peptidoglycan biosynthesis described by Schneider and coworkers (Schneider *et al.*, 2004) can be divided into three stages described

[$]Adapted from: Arbeloa *et al.*, 2004

in Figure 3. The first stage occurs in the cytoplasm starting with the nucleotide-linked sugar UDP-NAG, which is converted to UDP-NAM in two steps, catalysed by MurA and MurB. The pentapeptide is then sequentially elongated from the lactyl moiety of the UDP-NAM, one or two amino acids at a time, to produce the final cytoplasmic precursor, UDP-NAM- L-Ala-D-iGlu-L-Lys-D-Ala-D-Ala. These reactions are catalysed by MurC, MurD, MurE and MurF. The second stage occurs at the inner face of the cell membrane, starting with the transfer of phospho-NAM-L-Ala-D-iGlu-L-Lys-D-Ala-D-Ala from UDP to the lipid carrier undecaprenol-phosphate (C_{55}-P) by the integral membrane enzyme MraY, forming lipid-I. In the last intracellular enzymatic step, MurG catalyzes the transfer of NAG from UDP-NAG to lipid-I, forming lipid-II, which contains the basic monomeric unit of the cell wall, the disaccharide pentapeptide, linked to the undecaprenol carrier. To enter into the final stage of synthesis, lipid-II must be flipped so that the disaccharide-pentapeptide is exposed at the outer surface of the inner membrane and can be polymerised and cross-linked into the cell wall. This final stage involves two separate reactions: a transglycosylation reaction, in which linear non cross-linked glycan chains are formed, and a transpeptidation reaction, where the peptide substituents are cross-linked on the glycan chains by a pentaglycine.

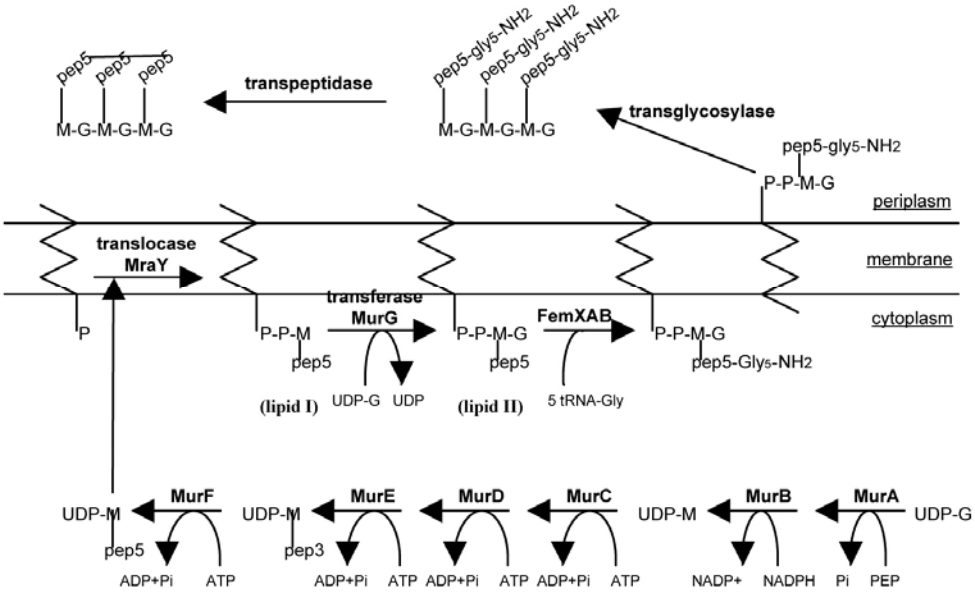


Figure 3: Overview of the pathway of cell wall synthesis♣. The jagged line in the membrane represents undecaprenol-P. Names for the enzymes catalysing each step are given above the arrows.

Extensive genetic analysis and characterisation of mutant cell walls suggested that the pentaglycine is synthesised in a sequential fashion by non-ribosomal peptidyltransferases (Hegde and Shrader, 2001, Rohrer and Berger-Bachi, 2003a). It has been shown that three proteins are involved in the formation of the pentaglycine crossbridge in *S. aureus* peptidoglycan. The factor catalysing the first step in its biosynthesis was found to be FemX, encoded by the essential gene *fmhB* (Rohrer *et al.*, 1999). Chain formation also depends on the two proteins FemA and FemB, which are encoded by the operon *femAB*. FemA was proposed to catalyse the addition of the second and third glycines (Maidhof *et al.*, 1991, Stranden *et al.*, 1997), whereas FemB is thought to add glycines four and five (Henze *et al.*, 1993) as shown in Figure 4.

♣Adapted from Ton-That *et al.*, 2004

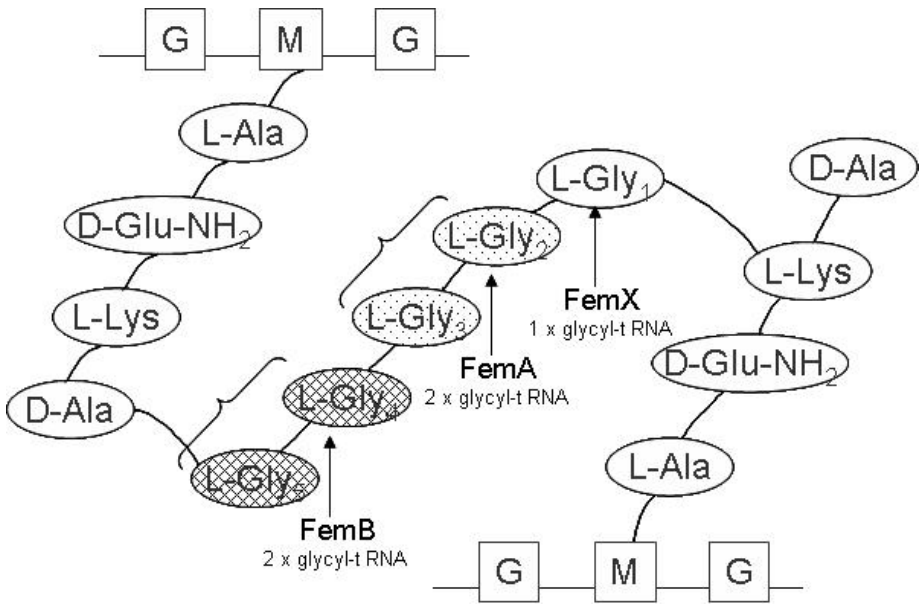


Figure 4: Pentaglycine bridge formation♥. Glycine is transferred to the free ε-amino group of lysine by the activity of glycyl-tRNA synthetase. FemX is essential for addition of the first glycine, while FemA and FemB are needed for addition of glycine 2, 3, and 4, 5, respectively.

Staphylococcal strains producing glycyl-glycine endopeptidases, such as lysostaphin (Thumm and Gotz, 1997) or ALE-1 (Sugai *et al.*, 1997b) (cf. sections 1.3 and 1.4), protect their cell walls with FemABX-like immunity factors, termed Lif or Epr respectively, which insert serine in place of glycine in positions three and five in the interpeptide cross-bridge (Ehlert *et al.*, 2000, DeHart *et al.*, 1995). This protects the cell wall as lysostaphin and ALE-1 cannot hydrolyse glycyl-serine or seryl-glycine bonds. It was found that, in order for Lif to operate, the functions of FemA and FemB were required, which is in contrast to the previously postulated theory that Lif complements FemB in *S. aureus femB* mutant (Tschierske *et al.*, 1997).

♥Adapted from Kopp *et al.*, 1996

1.1.4. Peptidoglycan recognition in innate immunity

Other major constituents of the bacterial cell wall are polyanionic wall teichoic acids (WTAs) and lipoteichoic acids (LTAs) (Ghuysen *et al.*, 1965), as well as proteins (Navarre and Schneewind, 1999). Many surface proteins are anchored by a reaction catalysed by a sortase enzyme via the free amino group of uncross-linked pentaglycine side chains to the *S. aureus* cell wall (Ton-That *et al.*, 2004). This ability to anchor to the cell wall requires a 35-residue sorting signal that consists of an LPXTG motif, followed by a *C*-terminal hydrophobic domain and a tail of mostly positively charged residues. Cell wall sorting consists of serial steps, which lead to the proteolytic cleavage of the polypeptide chain between threonine and glycine of the LPXTG motif. The carboxyl of threonine is subsequently amide-linked to the free amino-group in the pentaglycine crossbridge of the staphylococcal cell wall.

WTAs, LTAs and surface proteins are involved in the adhesion and pathogenesis of *S. aureus*. Indeed, proteins on the cell wall surface interact with host matrix proteins to conceal the bacterial surface from attack by the host defence system (Mengin-Lecreulx and Lemaitre, 2005).

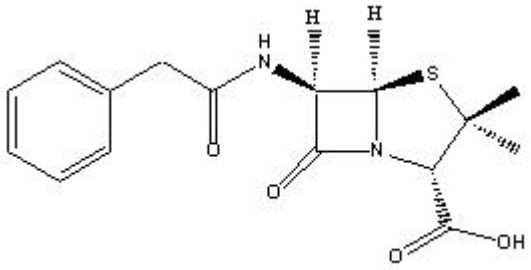
1.2. Antibiotics that work on bacterial cell wall and resistance to them

Since peptidoglycan plays a key role in maintaining the shape and structural integrity of the cell wall, interference with any stage in its synthesis and assembly is an effective means of inhibiting cell growth. Some of the most valuable antibiotics act in this way, including the β-lactams and the glycopeptides. However, use of antibiotics selects for growth of rare resistant organisms in an otherwise susceptible population. Large numbers of bacteria

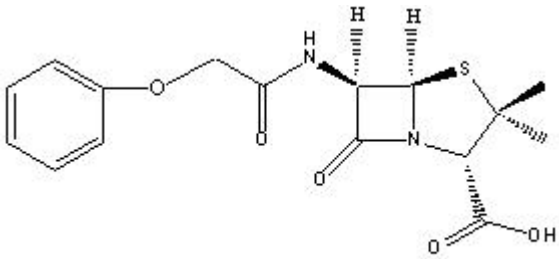
and the short generation time (approximately four hours) lead to emergence of resistance. Resistance genes collected on plasmids and transposable elements lead to rapid spread of multidrug resistance between bacterial species.

S. aureus was the first bacterium in which penicillins G and V resistance was found in 1947. Methicillin then became the antibiotic of choice. The structures of penicillins G and V and methicillin are shown in Figure 5. Penicillins G and V and methicillin are β-lactam antibiotics which irreversibly inhibit *S. aureus* transpeptidases by acting as alternative (suicide) substrates. They mimic the structure of D-alanyl-D-alanine residues and react with transpeptidases. Transpeptidases bind the β-lactam ring of the antibiotic, thus preventing the pentaglycine crosslinks from being incorporated into the peptidoglycan layer. The transpeptidase enzymes are usually referred to as penicillin-binding proteins (PBPs) (Spratt, 1994). Expression of β-lactamase enzymes is the most important mechanism by which organisms become resistant to β-lactams. Other mechanisms by which resistance to β-lactams can arise also exist. Methicillin-resistant strains of *S. aureus* (MRSA) display an altered pattern of PBPs in which an alternative PBP called 2a (or 2') is produced. This PBP, the product of *mecA* gene, is a transpeptidase with a very low susceptibility to inhibition by methicillin or other β-lactam antibiotics (Chambers *et al.*, 1990). Methicillin-resistant *S. aureus* (MRSA) is a major problem in the world, causing the most common hospital and community acquired infections that appeared in the 1980s. Wilkinson and coworkers reported that the peptidoglycan of some MRSA strains had slightly diminished glycine levels compared with that of normal *S. aureus* strains (Wilkinson *et*

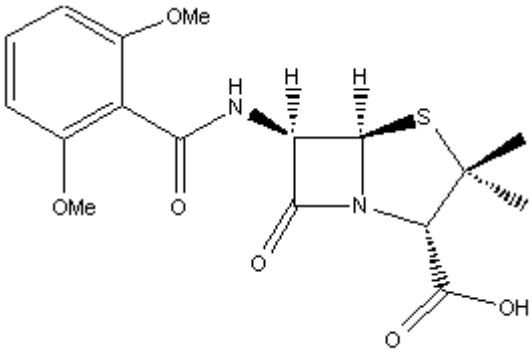
al., 1978). Thus a combination of modified peptidoglycan and expression of PBP2' is responsible for the overall level of resistance expressed by MRSA.



A: Structure of penicillin G



B: Structure of penicillin V



C: Structure of methicillin

Figure 5: Penicillins G (structure A) and V (structure B) and methicillin (structure C) are β-lactam antibiotics.

Since 1997 there have been reports of strains carrying varying levels of resistance to glycopeptide antibiotics (glycopeptide intermediate *S. aureus* or GISA), in particular vancomycin (Koehl *et al.*, 2004). Vancomycin, the structure of which is shown in Figure 6, has a different mode of action to that

of penicillins as it binds to the terminal D-alanyl-D-alanine on each pentapeptide side chain by formation of five hydrogen bonds, rather than interacting with the transpeptidase itself (Loll and Axelsen, 2000). The presence of the bulky glycopeptides on each D-alanyl-D-alanine residue prevents extension of the linear glycan peptide in the cell wall by denying access of the transpeptidases to the substrate. Boyle-Vavra and coworkers have described a spectrum of changes in peptidoglycan composition in different GISA strains (Boyle-Vavra *et al.*, 2001). These changes in peptidoglycan composition include a decrease in the proportion of highly cross-linked peptidoglycan species. It has been proposed that GISA strains have an increased number of D-alanyl-D-alanine residues in their cell walls that form false targets that sequester vancomycin in the cell wall, preventing it from reaching its target D-alanyl-D-alanine in membrane-anchored lipid disaccharide-pentapeptide peptidoglycan precursors (Zygmunt *et al.*, 1967). Increased levels of PBP2 have also been reported in VISA (vancomycin-intermediate *S. aureus*) isolates (Moreira *et al.*, 1997), leading to the conclusion that the bacteria synthesise more of these proteins which then compete more effectively with vancomycin for their substrate.

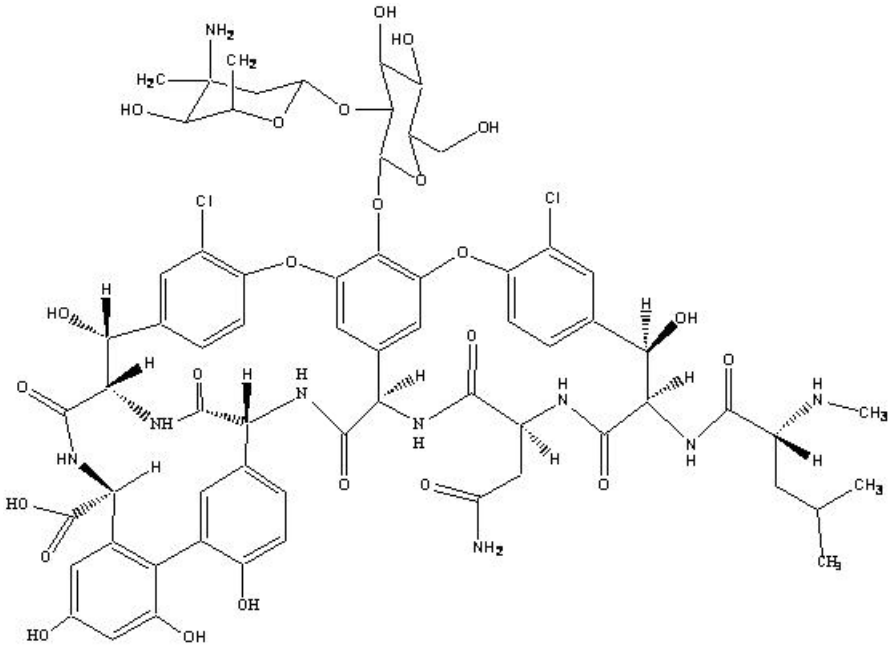


Figure 6: Structure of Vancomycin.

Lysostaphin, a lysin expressed by *Staphylococcus simulans* biovar *staphylolyticus* (Schindler and Schuhardt, 1964), has the capability to digest the cell wall of certain staphylococci when applied exogenously (cf. section 1.3). It has gained renewed interest as an anti-staphylococcal therapeutic agent because of the growing emergence of multi-drug resistant *S. aureus*. The literature is inconclusive on lysostaphin susceptibility of MRSA. Zygmunt and coworkers found that over 400 strains of multiply resistant MRSA were lysed by lysostaphin (Zygmunt *et al.*, 1967), while Sabath and coworkers reported diminished lysis of MRSA with lysostaphin (Sabath *et al.*, 1969). Huber and Huber studied this question and concluded that lysostaphin susceptibility seems to be a reliable taxonomic criterion even in strains which have acquired methicillin resistance (Huber and Huber, 1989). Nevertheless, resistance to lysostaphin has been described among *femAB* null mutants of MRSA strains

with alteration of the pentaglycine cross-bridge (Stranden *et al.*, 1997). The interpeptide bridge is essential for the full expression of methicillin resistance in MRSA. Shortening of the pentaglycine chain impairs growth, leading to aberrant septum formation and lowered methicillin resistance level (Maidhof *et al.*, 1991). The interpeptide is essential for *S. aureus*; survival with a shortened bridge reduced to only one glycine requires as yet uncharacterised compensatory mutations (Ling and Berger-Bachi, 1998). One possible solution to the problem of *S. aureus* becoming resistant to antibiotics is the co-administration of a β-lactam antibiotic with lysostaphin (Kiri *et al.*, 2002). It has been observed that resistance to lysostaphin is usually coupled with an increase in the susceptibility to β-lactams, which the strain was previously resistant to (Climo *et al.*, 2001). Thus, the right combination of enzyme and antibiotic could help in the control of certain antibiotic resistant bacteria.

Lysostaphin and the phage K lytic enzyme LysK were also tested to determine whether their simultaneous use was competitive or synergistic to kill MRSA (Becker *et al.*, 2008). The results revealed a clear synergistic effect in the efficiency when both enzymes where used. Thus, in addition to more effective killing, the use of two different lysins to kill resistant *S. aureus* could retard the emergence of enzyme-resistant mutants significantly.

1.3. Lysostaphin

1.3.1. The characteristics of lysostaphin

Lysostaphin was originally isolated by Schindler and Schuhardt in 1964 from cell-free filtrates of the Gram-positive coccus, *Staphylococcus simulans* biovar *staphylolyticus* ATCC1362 (Schindler and Schuhardt, 1964). The name was given to a preparation that included a glycylglycine

endopeptidase, a glucosaminidase and an amidase, but the term was later used just to describe the endopeptidase (Iversen and Grov, 1973). Lysostaphin was reported as a potential anti-staphylococcal agent against *S. aureus* and, to a lesser extent, *S. epidermidis* (Wu *et al.*, 2003). It was active against both live and autoclaved cells, indicating that it did not disrupt a biosynthetic pathway (Schindler and Schuhardt, 1964). The rate at which lysostaphin lysed bacterial cells was found to be dependent on the temperature, pH and ionic strength of the assay medium, and also on the concentration of the enzyme (Schindler and Schuhardt, 1965).

The glycinase action of the endopeptidase was demonstrated further by Browder and coworkers by instant thin layer chromatographic separation and UV detection of the post-hydrolysis dansylated products from glycine peptides (Browder *et al.*, 1965). This activity indicated that lysostaphin catalyses the hydrolysis of the pentaglycine crossbridge in the bacterial peptidoglycan, leading to osmotic rupture of the cell membrane and cell death. It has been suggested that lysostaphin cleaves specifically between the third and fourth glycine residues of the pentaglycine cross-bridge based on the release of staphylococcal surface proteins by lysostaphin (Schneewind *et al.*, 1995), however analysis of lysostaphin cleaved muropeptides by mass spectrometry revealed that cleavage occurs between glycine residues 2 and 3, 3 and 4 and 4 and 5, with preference for the latter cleavage site (Xu *et al.*, 1997). Moreover lysostaphin was reported to bind tightly to elastin and to contain elastolytic activity (Park *et al.*, 1995). It was also shown to catalyse transpeptidation of glycine peptides (Sloan *et al.*, 1977).

Slightly different genes encoding lysostaphin, residing on the largest of the five plasmids of *S. simulans*, pACK1 (Recsei *et al.*, 1987, Heath *et al.*, 1991), were cloned into *Escherichia coli* by two groups: Recsei and coworkers (Recsei *et al.*, 1987) and Heinrich and coworkers (Heinrich *et al.*, 1987) in 1987. Using DNA sequencing, the open reading frame was determined by the first group to consist of 1167 base pairs, corresponding to a precursor of 389 amino acids, with a calculated molecular weight of 42.2 kDa (Recsei *et al.*, 1987) and by the second group, to consist of 1440 base pairs, corresponding to a precursor of 480 amino acids, with a calculated molecular weight of 51.7 kDa (Heinrich *et al.*, 1987). Figure 7 shows the lysostaphin nucleotide and amino acid sequences of the gene and encoding gene product, respectively, in accordance with Recsei and coworkers (Recsei *et al.*, 1987).

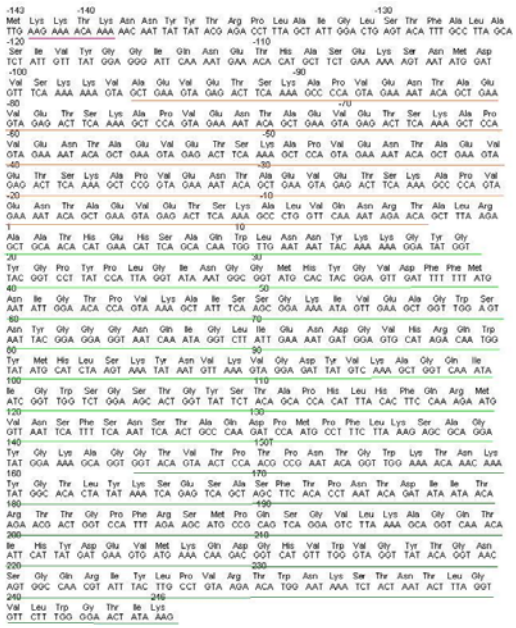


Figure 7: Preprolysostaphin gene from *S. simulans* and amino acid sequence of the encoded gene product♣. The *N*-terminal formylmethionine of the preprolysostaphin is designated as residue –143 and the *N*-terminal alanine of mature lysostaphin is designated as residue +1. Are underlined: in purple, the first residues of the signal peptide; **in orange,** the seven tandem repeat sequences; **in light green (+1/+139) and in dark green (+154/+246),** the endopeptidase and targeting domain respectively, of the mature lysostaphin.

♣Adapted from: Recsei *et al.*, 1987

It was observed by amino acid sequence analysis of the encoding product that the preprolysostaphin is organised in three distinct regions. Commonly observed in preprolysostaphin, the *N*-terminal region contains a cluster of four positively charged residues followed by an uncharged, largely hydrophobic sequence. The *N*-terminal has the properties characteristic of a signal peptide directing the protein to the hydrophobic cell membrane. The Ala-Ser bond at position –121 and –120 or that at position –108 and –107, in the alternative gene, is the likely signal cleavage site that helps to transform preprolysostaphin into prolysostaphin. The second region is composed of 7 tandem repeats of a 13-amino acid pattern (from the sequence obtained by Recsei and coworkers) (Recsei *et al.*, 1987) or composed of 15 tandem repeats if the sequence obtained by Heinrich and coworkers is considered instead (Heinrich *et al.*, 1987). The first repeats (six or fourteen) are identical, while the last one contains three amino acid substitutions. This portion of the molecule, called prolysostaphin, is highly ionic and has a net negative charge. The third region is the mature lysostaphin, which is made up of 246 amino acids. Prolysostaphin is secreted, and then activated (converted to the mature enzyme) extracellularly by a cysteine (sulfhydryl) protease, via cleavage of the *N*-terminal portion of the proenzyme (Neumann *et al.*, 1993, Bunn *et al.*, 1998). The protease that converts prolysostaphin to the mature form is itself secreted as a proenzyme and then processed autocatalytically to provide the active form (Bunn *et al.*, 1998).

The mature lysostaphin is a zinc-containing metalloenzyme (it contains one zinc ion per molecule of protein), consisting of a single polypeptide chain, with a molecular weight, when determined by sedimentation equilibrium of 25

kDa, a calculated pI of 9.52 and a pH optimum for catalytic activity of 7.5 (Browder *et al.*, 1965, Trayer and Buckley, 1970). Lysostaphin contains no cysteine residues. The enzyme activity is inhibited by excess zinc cations (Zygmunt and Tavormina, 1972, Park *et al.*, 1995). Figure 8 shows a representation of the arrangement of preprolysostaphin (Thumm and Gotz, 1997).

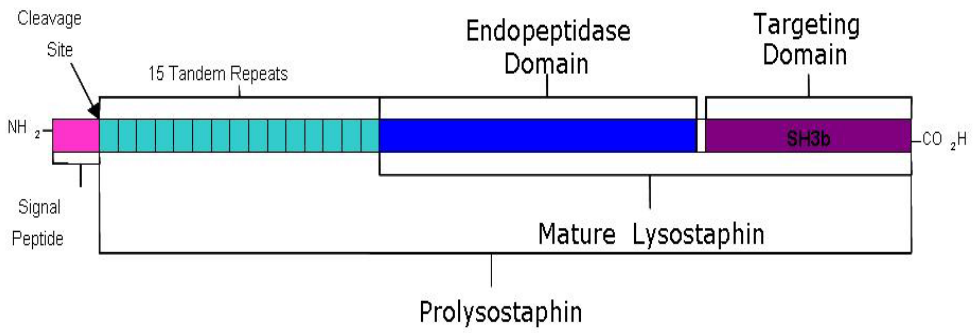


Figure 8: Schematic diagram of the amino acid regions of preprolysostaphin.

The biologically active structure sufficient for the binding of mature lysostaphin to the cell wall of *S. aureus* consists of the 92 *C*-terminal residues lysostaphin. A lysostaphin molecule lacking the *C*-terminal targeting domain was enzymatically active, but had lost its ability to distinguish between *S. aureus* and *S. simulans* cells, indicating that this domain functions to confer target cell specificity to the bacteriolytic molecule (Baba and Schneewind, 1996). The targeting domain has been classified as belonging to the SH3b family, a prokaryotic homologue of the eukaryotic Src homology 3 (SH3) domain mediating protein-protein interactions (Koch *et al.*, 1991). The mature lysostaphin lacking the targeting domain is called the endopeptidase domain and is classified as part of the M23/37 protease family in the MEROPS database (Rawlings *et al.*, 2002).

1.3.2. Zinc metalloprotease mechanism

Zinc-containing metalloproteases constitute an expanding list of structurally related proteases that are widely distributed in Nature. Thermolysin (TLN) is a typical example from the zinc-metalloprotease superfamily. Isolated from *Bacillus thermoproteolyticus* (Endo, 1962), this extracellular endopeptidase catalyses the hydrolysis of the peptide bond specifically on the amino side of large hydrophobic residues (Rao *et al.*, 1998). The X-ray crystal structure of the enzyme showed that the active site of the wild-type TLN had two histidines (His 142, His 146) and one glutamic acid (Glu 166) residue coordinating a zinc cation (Colman *et al.*, 1972). The glutamic acid 143 formed, together with the above-mentioned two histidines, the consensus HExxH motif of the enzyme amino-acid sequence (Lipscomb and Strater, 1996). Hall and coworkers suggested an analogy between cryptic peptidase sonic hedgehog and thermolysin reaction mechanism (Hall *et al.*, 1995) and Bochtler and coworkers reported similarities between lysostaphin-type enzymes, D-Ala-D-Ala metallopeptidases and the *N*-domain of cryptic peptidase sonic hedgehog ("LAS" proteins) (Bochtler *et al.*, 2004). Although the active-site geometry might be different in lysostaphin, LytM (a member of the M23/37 family) (Firczuk *et al.*, 2005, Bochtler *et al.*, 2004) and thermolysin, we suggest some similarity of lysostaphin with thermolysin in terms of their hydrolysis of the peptide bond.

A single-step reaction mechanism for thermolysin was obtained following a proposal by Pangburn and Walsh in 1975 (Pangburn and Walsh, 1975), and refined by Kester and Matthews in 1977 (Kester and Matthews, 1977). This proposal proceeds as shown in Figure 9 (Hangauer *et al.*, 1984,

Holden *et al.*, 1987). Starting from the unligated TLN **1**, with a zinc-coordinated water molecule, the wide groove separating the enzyme in two large domains accommodates an extended peptide chain, forming the Michaelis complex **2**. The glutamic acid 143 activates the incoming nucleophilic water molecule, and the protonated histidine 231 that is located across the hypothetical position of the substrate, together with the zinc cation, enhances the nucleophilicity of the water oxygen that polarises the substrate peptide bond for nucleophilic attack on the peptide carbonyl carbon. The proton accepted by glutamic acid 143 is immediately shuttled to the peptide nitrogen. As a result, a tetrahedral *gem*-diolate intermediate **3** is formed, where zinc is penta-coordinated and the peptide carbon is tetrahedral. In the next step, a direct cleavage of the peptide C-N bond takes place, with the amine product proposed to be released in its protonated form **4**, fitting the neutral pH conditions. Here, the glutamic acid 143 serves as a proton shuttle, abstracting the second proton from the water oxygen and delivering this proton to the amine nitrogen. The catalytic cycle is completed by carboxyl product release, probably mediated by incoming water molecule uptake (Pelmenschikov *et al.*, 2002).

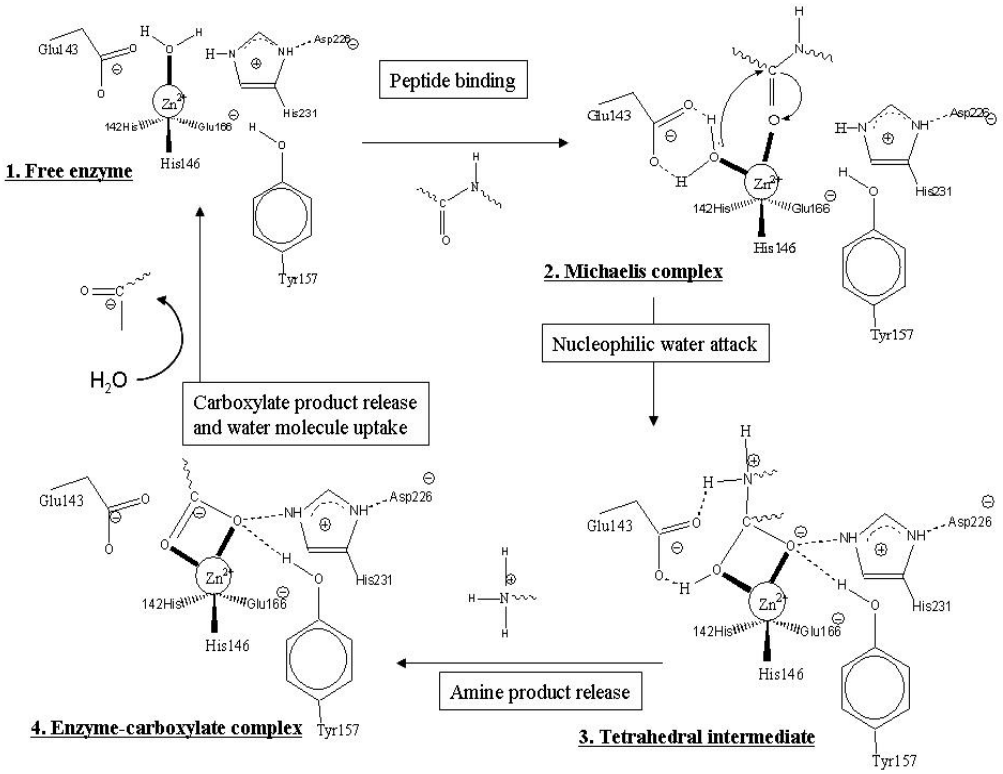


Figure 9: Experimentally proposed reaction for thermolysin♦.

1.3.3. Therapeutic applications

Lysostaphin has been investigated as an alternative to conventional antibiotics since the 1960s. The drug company Mead Johnson tried to develop a product based on a natural extract from *S. simulans*, containing a glycylglycine endopeptidase, a glucosaminidase and an amidase (Zygmunt and Tavormina, 1972). It appeared, after experimental studies of lysostaphin's effect on renal staphylococcal infection models in mice, that lysostaphin warranted further consideration as a potential antistaphylococcal agent for systemic use (Harrison and Zygmunt, 1967). Other parallel studies in animal models of oxacillin-susceptible *S. aureus*-

♦Adapted from: Lipscomb *et al*., 1996

induced infection indicated that lysostaphin was an effective treatment agent against experimental aortic valve endocarditis in dogs caused by *S. aureus* (Goldberg *et al.*, 1967). It was also found that lysostaphin could be a potentially useful agent for the treatment of epidemiologically important nasal carriers of *S. aureus* (Quickel *et al.*, 1971).

However, the lysostaphin preparations used were only available in partially purified form, and as there were other anti-staphylococcal antibiotics widely available (Zygmunt and Tavormina, 1972), further development of lysostaphin for clinical use was halted until the recent introduction of recombinant lysostaphin (that is more than 90% pure) in the 1980s enabled more extensive investigation. In 2000 lysostaphin was placed into clinical trials by the American company Biosynexus and large scale production of the protein was carried out by a partner company, Avecia (McCoy, 2004). The availability of recombinant lysostaphin has provided an opportunity to assess the efficacy of lysostaphin in an animal model of *S. aureus* endocarditis and to compare it to the efficacy of standard therapy with vancomycin (Climo *et al.*, 1998, Patron *et al.*, 1999). It was found that lysostaphin was effective whereas vancomycin performed poorly. A more extensive investigation of the clinical use of this form of lysostaphin to treat humans was reported in the case of a formulation of nasal cream containing lysostaphin (von Eiff *et al.*, 2003). This cream was spread in the anterior nostrils: these are the primarily ecologic niche for *S. aureus*, to stop the nasal colonisation by this opportunist pathogen that increases the risk of development of a *S. aureus* infection (Kokai-Kun *et al.*, 2003). From these experiments it appeared that lysostaphin is not

inactivated by body fluids. The development of lysostaphin as an intravenous antibiotic has also been investigated, showing an increased effectiveness of lysostaphin over penicillin G in augmenting the survival rate of mice infected by staphylococci (Schaffner *et al.*, 1967). Zygmunt and coworkers showed in an *in vitro* study that serum exhibited a minimal antagonizing effect on the antibiotic activity of lysostaphin (Zygmunt *et al.*, 1966). However, clinical use of bacterial enzymes systematically has been associated with some toxicity and hypersensitivity problems. It was found that conjugation by polyethylene glycol (PEG) significantly enhanced lysostaphin's therapeutic value as an intravenous antibiotic against *S. aureus*, by increasing its tissue distribution and decreasing its immunogenicity (Walsh *et al.*, 2003). It was found that the half life of lysostaphin in blood serum increased from one hour (when unconjugated) to a maximum of 24 hours upon conjugation with PEG.

Another example of clinical use of lysostaphin was found by coating a catheter with lysostaphin to prevent catheter colonisation by several strains of *S. aureus* (Shah *et al.*, 2004). Staphylococcal biofilms can lead to infections in biomedical implants, which tend to be difficult to treat and are usually much more resistant to antibiotics than those in normal colonies; this may make it necessary to remove a pacemaker or an artificial joint and replace it. To solve the problem, the activity of lysostaphin on biofilms made up of colonies of *S. aureus* and *S. epidermidis* has been examined. Lysostaphin not only killed *S. aureus* in biofilms, which were resistant to very high concentrations of oxacillin, vancomycin and clindamycin, but also disrupted the extracellular matrix of biofilms *in vitro* on plastic and

glass surfaces (Wu *et al.*, 2003). Unlike most antibiotics that interfere with the growth of bacteria, lysostaphin is highly effective in lysing both actively growing and dormant *S. aureus* cells (Schindler and Schuhardt, 1964), which gives it an additional therapeutic value.

1.4. Endopeptidases related to lysostaphin: structure and function

Bacteriocins are proteinaceous antibiotics produced by one bacterium that principally have killing activity against closely related bacteria (Jack et al., 1995, Joerger, 2003). Bacteria produce various peptidoglycan hydrolases that include *N*-acetyl-glucosaminidases, *N*-acetyl-muramidases, *N*-acetyl-muramyl-L-alanine amidases, endopeptidases, and transglycosylases (Ghuysen *et al.*, 1966). Sites of hydrolysis of various muralytic enzymes on the peptidoglycan of *S. aureus* are shown in Figure 10. Enzymes secreted by a number of different bacteria have been found to have similar activity to that of lysostaphin as described below.

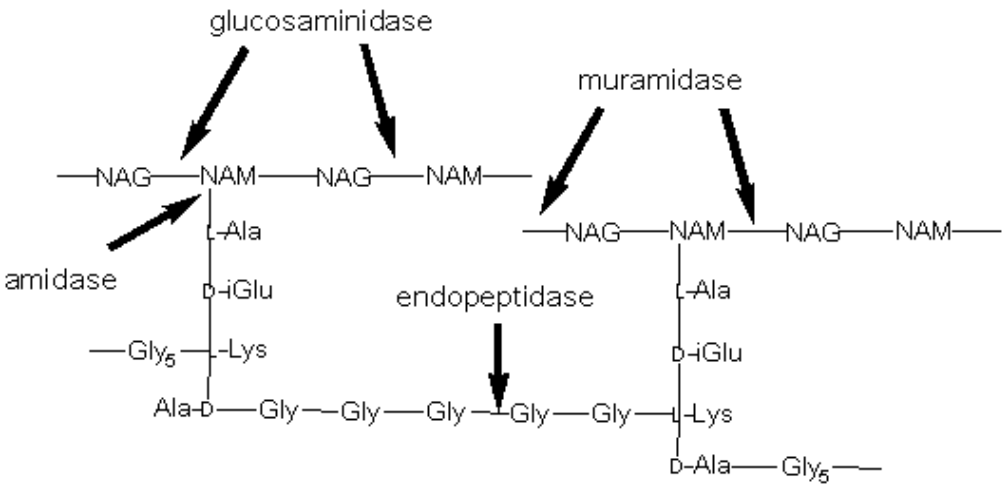


Figure 10: Sites of hydrolysis of various muralytic enzymes on the peptidoglycan of *S. aureus*∴.

∴Adapted from: Navarre *et al.*, 1999

1.4.1. LytM

A gene encoding an autolytic activity was identified in *S. aureus* by Mani and coworkers, in 1993 (Mani *et al.*, 1993). Its open reading frame, designated LytM, is made of 948 base pairs that encode a polypeptide of 316 amino acids. The calculated molecular mass of LytM corresponds to that of the autolytic activity protein detected: 36 kDa. A putative signal sequence, as well as a putative signal peptide cleavage site, is present within the first 26 amino acids. The signal sequence is strongly hydrophobic and characteristic of secreted proteins. Significant amino acid sequence homology was found with lysostaphin when compared with LytM: the *C*-terminal region of LytM showed significant homology (48%) with *N*-terminal region of the mature lysostaphin (Ramadurai and Jayaswal, 1997, Nilsen *et al.*, 2003). Based on its similarity with lysostaphin, the enzyme would be expected to be specific for glycine rich sequences, a conclusion that is supported by experimental data (Ramadurai *et al.*, 1999). It degrades the cell walls from *S. carnosus* and the Lyt$^-$ mutant of *S. aureus*, but not those from *Micrococcus luteus*, leading the authors to speculate that the enzyme is a glycyl-glycine endopeptidase (Ramadurai *et al.*, 1999).

They also suggested that LytM could be a member of the zinc-metalloprotease enzyme group as it contains a 38 amino acid motif Y-x-H-x(11)-V-x(12/20)-G-x(5-6)-H, homologous to other members (Sugai *et al.*, 1997a, Bochtler *et al.*, 2004), and has conserved histidine residues, which are believed to bind the zinc cofactor (Ramadurai *et al.*, 1999). Recently the X-ray crystal structure as been presented by Odintsov and coworkers in 2004 as the first structure of a lysostaphin-type peptidase (Odintsov *et al.*, 2004). It

demonstrated that LytM is a two domain β-protein. The *N*-terminal domain was found to make limited contact with the *C*-terminal domain (Figure 11).

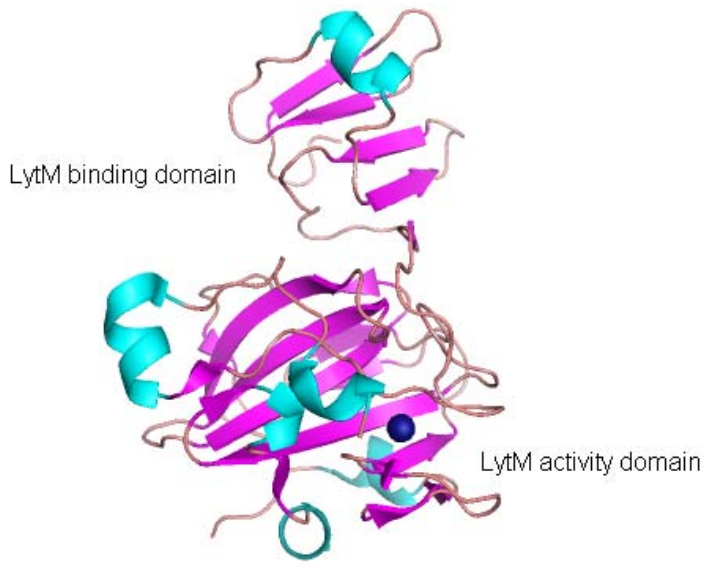


Figure 11: Ribbon representation of latent LytM structure (PDB code 1QWY)•.

In the active $LytM_{185-316}$ structure presented by Firczuk and coworkers in 2005, the zinc cation is tetrahedrally coordinated by the side-chains of His 210, Asp 214 and His 293 and presumably the water molecule as the fourth ligand as shown in Figure 12 (Firczuk *et al.*, 2005). They are part of two motifs, H-x(3)-D and H-x-H. The N^{ε} of histidine 210 and the N^{δ} of histidine 293 imidazole rings donate their free electron pairs to the zinc ion (Firczuk *et al.*, 2005, Bochtler *et al.*, 2004). The N^{ε} of histidine 210 and the N^{δ} of histidine 293 are therefore protonated and donate a hydrogen bond to a carbonyl oxygen of the main chain and a glutamine of the side chain, respectively. Histidine 260 is assigned as the catalytic base that activates the water molecule together with the zinc cation (Bochtler *et al.*, 2004). All active-site residues are located on or

•Adapted from: Odintsov *et al.*, 2004

around a β-sheet scaffold. The core of this central β-sheet consists of four antiparallel β-strands *(Firczuk et al.*, 2005).

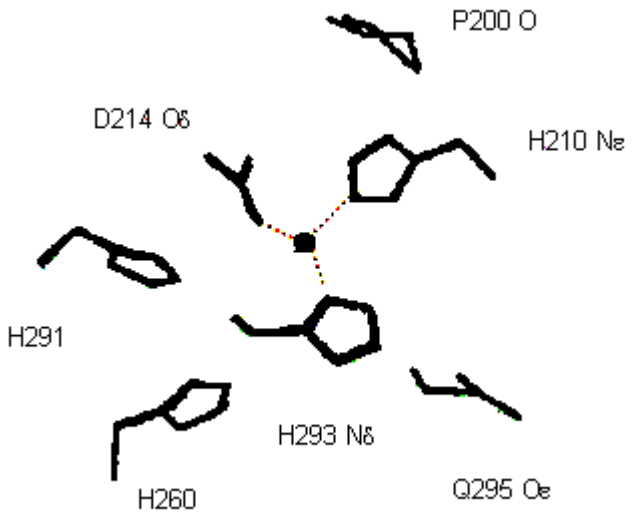


Figure 12: Representation of the selected residues of the active site of LytM$_{185-316}$▪. The zinc cation is shown as a black ball. All zinc cation coordinating interactions are represented by dotted lines. H210, D214, H260, H291, H293 are residues essential for activity. Q295 and P200 are orientated key residues.

Interestingly, three consecutive glycine residues, 206 to 208, in one of the most exposed loops of the LytM structure, appeared to fill different parts of the LytM$_{185-316}$ groove in two different crystal forms as a surrogate substrate analogue (Firczuk *et al.*, 2005). It was shown that a tetraglycine chain, labelled with a dinitrophenyl group, had moderate (millimolar) binding affinity for the enzyme (Firczuk *et al.*, 2005).

1.4.2. ALE-1

In 1997, Sugai and coworkers reported the purification of a new staphylolytic enzyme from the supernatant of a culture of *Staphylococcus capitis* EPK1, which was named ALE-1 (Sugai *et al.*, 1997a). It was identified as a glycylglycine endopeptidase metalloenzyme that contained one atom of

▪Adapted from: Firczuk *et al.*, 2005

zinc cation per molecule. The optimum pH range was from 7 to 9 and its pI was 9. Like lysostaphin, ALE-1 is synthesised as a precursor protein that is cleaved after the alanine residue at position 35. Mature ALE-1 consists of 362 amino acid residues and has a molecular weight of 35.6 kDa. The primary structure was very similar to that of the lysostaphin proenzyme form (Nilsen *et al.*, 2003), as it contains an *N*-terminal domain with tandem repeats of a 13 amino acid sequence followed by a central catalytic domain and a *C*-terminal targeting domain of 92 amino acids. As with lysostaphin, the targeting domain of ALE-1 has been classified as SH3b, a prokaryotic homologue of the eukaryotic SH3 domain. One difference observed between ALE-1 and lysostaphin was that the *N*-terminal repeat domain in ALE-1 is not extracellularly processed as happens to prolysostaphin.

The 1.75 Å crystal structure of ALE-1 targeting domain shown in Figure 13 as been presented recently (Lu *et al.*, 2006). Like LytM, this is distinct from the catalytic domain that precedes it in the enzyme structure. It consists of eight β-strands. Two anti-parallel multiple-stranded β-sheets pack at approximately right angles: β5- β7 and the *N*-terminus of β2 in sheet I; β3-β4, β8, and the *C*-terminus of β2 in sheet II, and give a narrow groove that can potentially accommodate an extended pentaglycine by forming β-sheet-type interactions with strands β1 and β3. Lu and coworkers compared the binding of the ALE-1 targeting domain to various staphylococcal peptidoglycans with genetically modified interpeptide bridges and their isogenic parents (Lu *et al.*, 2006). The targeting domain also showed a strong preference for Gly_5 in this study. Forms of ALE-1 that lacked the targeting domain had significantly decreased binding activity to *S. aureus* cells, confirming that this portion was

responsible for targeting the enzyme to the bacteria. The first nine residues in this domain are likely to be responsible for the specificity for *S. aureus*, as the truncated version of the targeting domain, lacking these residues, had drastically reduced binding to the bacterial peptidoglycan.

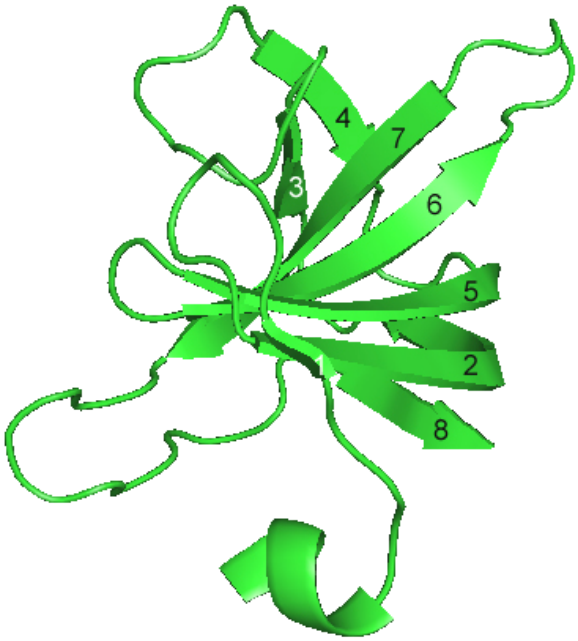


Figure 13: ALE-1 targeting domain crystal structure with strands numbered sequentially[&].

S. capitis also carries a gene that is homologous to the *femAB* operon for the production of endopeptidase resistance factor, *epr* (Sugai *et al.*, 1997b). This was found to have similarities to the amino acid sequence of FemA (Berger-Bachi *et al.*, 1989) and FemB (Henze *et al.*, 1993). *Epr* is similar to *lif*, in that it confers resistance to attack on *S. capitis* by serine-incorporation at position three and five in the pentaglycine crossbridge (Ehlert *et al.*, 2000).

[&]Adapted from: Lu *et al.*, 2006

1.4.3. zoocinA

Simmonds and coworkers purified a proteinaceous bacteriocin-like inhibitory substance, named zoocin A, from cultures of *Streptococcus equi* subsp. *zooepidemicus*, in 1995 (Simmonds *et al.*, 1995). Analysis of the nucleotide sequence revealed an open reading frame of 858 nucleotides. Zoocin A is synthesized as a 285 amino acid pre-peptide that is cleaved to release a 262 amino acid peptide into the environment (Simmonds *et al.*, 1997). The deduced molecular mass of the predicted peptide is 27.9 kDa, which is similar to the molecular mass of the purified peptide, as estimated by SDS-PAGE.

The pre-peptide sequence is made of four characteristic regions. The 23 amino acid leader sequence, as mentioned before, is a signal peptide to locate the protein at the cell surface. The *N*-terminal region shows strong homology to the *N*-terminal domain of lysostaphin (Simmonds *et al.*, 1997) and in both cases contains the catalytic sites that bring about hydrolysis of the peptidoglycan. The mid-sequence region is a short threonine and proline-rich linker, and is attached to the *C*-terminal region, which is the cell wall recognition site, allowing zoocin A to be specific to its target bacteria. Evidence for this domain structure has also been found using NMR spectroscopy (Liang *et al.*, 2004). The targeting domain was shown to still be able to bind sensitive cells when expressed on its own, but the catalytic domain, when treated similarly, only retained 2 % of the activity of the native enzyme. It was concluded that the loss of activity for the catalytic domain was likely therefore to be due to loss of targeting ability.

Zoocin A treated cells revealed that *Streptococcus zooepidemicus* strains other than the producer: *Streptococcus pyogenes*, *Streptococcus gordonii* and *Streptococcus mutans*, are sensitive to zoocin A (Simmonds *et al.*, 1996). These are closely related to the producing strain, and include some of the main causative agents of group A streptococcal sore throat and dental caries. The *C*-terminal portion was found to be unique to zoocin A, explaining the difference in targets in between this enzyme and lysostaphin. In 1996, studies of the mode of action of zoocin A demonstrated that it was able to cleave a hexaglycine substrate with an efficiency comparable to that of lysostaphin, suggesting that like lysostaphin, zoocin A functions enzymatically as an endopeptidase (Simmonds *et al.*, 1996). In 2004, Heath and coworkers demonstrated that purified zoocin A bound penicillin (ressembles D-alanyl-D-alanine) covalently and that its peptidoglycan hydrolase activity was inhibited by penicillin (Heath *et al.*, 2004). These results suggested that zoocin A is presumably a D-alanyl endopeptidase that controls the extent of peptidoglycan cross-linking.

Reported by Beatson and coworkers, flanking the gene encoding zoocin A (*zooA*) in a back-to-back orientation was a gene zoocin A immunity factor (*zif*), responsible for protecting the producer strain from the otherwise lethal action of its own products (Beatson *et al.*, 1998). *zif* has sequence homology to *femA* (factor essential for methicillin resistance) and *lif* (lysostaphin immunity factor). No differences were observed in amino acid or amino sugar compositions of peptidoglycan purified between *zif* containing and non-*zif* containing *Streptococcus gordonii* DL1 transformants. These results suggested that unlike FemA, FemB and Lif, Zif is not involved in

modifying the biosynthesis of the peptidoglycan precursor crosslink. Additionally, they suggested that the homology of zoocin A to lysostaphin and of Zif to Lif, the similarity in position and orientation of their encoding genes, and the similarity in the arrangement of the flanking insertion sequences, may be due to the horizontal transfer of large segments of DNA between staphylococci and streptococci (Beatson *et al.*, 1998).

1.4.4. Peptidoglycan Recognition Proteins

Peptidoglycan recognition proteins (PGRPs) are innate immunity molecules conserved from insects to mammals. PGRPs have at least one *C*-terminal PGRP domain which is highly conserved and homologous to bacteriophage and bacterial type 2 amidases (Kang *et al.*, 1998). Insects have up to 19 PGRPs, classified into short (S) and long (L) forms (Werner *et al.*, 2000). Mammals have four PGRPs, which were initially named PGRP-S, PGRP-L, PGRP-Iα and PGRP-Iβ (for "short", "long", or "intermediate" transcripts, respectively) (Liu *et al.*, 2001). Subsequently, their symbol was changed to PGLYRP-1, PGLYRP-2, PGLYRP-3, PGLYRP-4 respectively. All mammalian PGLYRPs are secreted, usually as disulphide-linked homo- and heterodimers.

Almost all PGRPs have two closely spaced conserved cysteines in the middle of the PGRP domain that form a disulfide bond. These residues are thought to be essential as a mutation in one of these cysteines in human PGLYRP-2 (C419A) abolished its amidase activity (Wang *et al.*, 2003). PGLYRP-2 is an amidase with a conserved zinc cation binding site in the peptidoglycan-binding groove which consists of two histidines, one tyrosine, and one cysteine. Few PGRPs structures have been released recently and have

revealed a general design similar to type 2 bacteriophage amidases (Guan *et al.*, 2005, Guan *et al.*, 2006). Crystallographic analysis of human PGLYRP-1 and the *C*-terminal PGRP domain of PGLYRP-3 showed that the ligand binding groove accommodated *N*-acetylmuramic acid bound to three peptide-bond amino acids (muramyl tri-peptide), which was the minimum peptidoglycan fragment hydrolysed by PGLYRP-2 (Wang *et al.*, 2003). Those amidases also bound muramyl tetra-peptide and muramyl penta-peptide with higher affinity than muramyl tri-peptide (Kumar *et al.*, 2005, Guan *et al.*, 2006). Moreover, binding of muramyl-pentapeptide (but not muramyl tri-peptide) to the *C*-terminal fragment of PGLYRP-3 induced a conformational change in the molecule. In addition, PGLYRP-1 seemed to have high affinity for lipoteichoic acid and lipopolysaccharide (Tydell *et al.*, 2006). It was not clear however, whether these other ligands bound specifically to the peptidoglycan-binding groove or unspecifically to hydrophobic regions of the molecule.

1.4.5. Enterolysin A

Enterolysin A, a class III bacteriocin of lactic acid bacteria, was originally isolated by Nilsen and coworkers from cell free filtrate of *Enterococcus faecalis* LMG 2333 (Nilsen *et al.*, 2003). This bacteriocin is a heat-labile protein with a broad inhibitory spectrum, which breaks down the cell wall of lactobacilli, lactococci, pedicocci and enterococci. The bacilli, listeria and staphylococci strains tested showed zones of growth inhibition only in the overlay assay. It is believed that the bacteriolytic activity of the bacteriocin is based on its ability to degrade the cell wall by hydrolysing the peptide bond of the sugar unit side chain.

Sequence analysis of *enlA* encoding enterolysin A revealed that the enzyme is translated as a 343-amino acid preprotein with a typical signal peptide of 27 amino acids directing the protein to the hydrophobic cell membrane and allowing its release in the periplasm, followed by a sequence corresponding to the *N*-terminal part of the purified protein. The expression of *enlA*, the gene of enterolysin A amplified from *E. faecalis* II/1, has been achieved in *Escherichia coli* (Nigutova *et al.*, 2008). The mature enzyme, product of activation of the preprotein, consisted of 316 amino acids, its molecular weight was close to 35 kDa on electrophoretic gel and it was overall a basic protein. Enteroysin A was predicted to exhibit a typical domain structure organisation with two separate domains. The *N*-terminal domain of the bacteriocin contains an amino acid motif HxxxD and HxH typical of metallopeptidase family M23 (Rawlings *et al.*, 2008) and thus was predicted to have a catalytic activity similar to lysostaphin, zoocin A, LytM and ALE-1 (Nilsen *et al.*, 2003). The *C*-terminal domain is linked by a threonine and proline-rich region to the former one and has significant sequence homology to a lysine from bacteriophage A2 and an *N*-acetomuramoyl-L-alanine amidase from bacteriophage PL-1, both of which are bacteriophages of *Lactobacillus casei* (Nilsen *et al.*, 2003). Both of those domains where reported to cooperate to yield a functional protein (Nigutova *et al.*, 2008). A deletion mutant of enterolysin A lacking 58 amino acid *C*-terminal domain had lost biological activity and was presumed to have an important role in the cell-wall binding of enterolysin A.

1.4.6. Bacteriophage endolysins

Nonfilamentous bacteriophages (bacterial viruses) have developed specific lytic enzymes for the release of their progeny from bacterial cells. Enzymatic cleavage of peptidogolycan is conducted by a holin-endolysin system (Young *et al.*, 2000, Tabac *et al.*, 2005). Many lysins are produced as a single polypeptide but contain two functional domains, a catalytic domain, which cleaves peptidoglycan bonds, and a cell wall binding module, which may bind specific features in the cell wall (Loessner, 2005). The modular structure of endolysins enabled alteration of their binding specificity and enzymatic activity independently by replacing either domain with the corresponding domain from another enzyme (Sheehan *et al.*, 1996), but there are several other strategies for engineering enzymes, for instance, random mutagenesis and targeted mutagenesis (Cedrone *et al.*, 2000, Bloom *et al.*, 2005). There are many data supporting the hypothesis that host and phage murein hydrolases share common ancestry and moreover, in some organisms the proteins have co-evolved by interchanging their functional domains (Lopez and Garcia, 2004). Such postulated evolutionary relationship between endolysins and autolysins may be the consequence of horizontal gene transfer between bacteriophages and their host bacterial cells (Hambly and Suttle, 2005).

Very recently the catalytic module of PlyB, an endolysin encoded by phage BcpI, which displays potent lytic activity against the *Bacillus anthracis*-like strain ATCC 4342 has been reported (Porter *et al.*, 2007). PlyB comprises an *N*-terminal catalytic module that shares sequence similarity with the members of the glycosyl hydrolase family 25 (GH-25) and cell wall binding

module homologous to bacterial SH3b domains. It was suggested that PlyB cleaved the β-1,4-glycosidic bond between *N*-acetylmuramic acid and *N*-acetylglucosamine units. It has also been shown that the SH3b domain played an important role in PlyB activity. It is unlikely that SH3b from PlyB recognises inter-peptide pentaglycine bridges as was shown for ALE-1 *(Lu et al.*, 2006) because its target, the bacilli, do not possess pentaglycine cross-bridges. The crystal structure of another *B. anthracis* phage endolysin, PlyL, has also been described (Low *et al.*, 2005). PlyL, the endolysin of the λ prophage Ba02, is also composed of an *N*-terminal catalytic module and a *C*-terminal cell wall binding domain. The structure of the catalytic domain revealed that its fold is similar to the Drosophila PGRP-LB (Kim *et al.*, 2003) and may have a NAM-amidase function. Remarkably, the presence of the cell wall binding domain within the full-length PlyL inhibits the lytic activity of the catalytic module (Low *et al.*, 2005). A possible explanation could be that the *C*-terminal domain might inhibit the catalytic activity of the *N*-terminal domain when not bound to the cell wall. This inhibitory effect might be relieved on high-affinity binding to the cell wall by the *C*-terminal domain.

1.5. Aims and objectives

Eighty years ago a mouldy Petri dish changed the face of medicine, and Alexander Fleming's discovery of penicillin opened a golden age of antibiotic research. But today, so-called superbugs are front page news and the antibiotic development pipeline is very poor. Some of the largest pharmaceutical companies have simply given up on what has become a tricky market. Experts now say that without improved approaches, along with better regulation and

controlled use of antimicrobials, the future of antibiotic development could be in jeopardy.

The industry has been looking at the genomes of these bacteria and has been screening potential antibacterial agents against targets essential for bacterial survival. This has failed for all types of targets, because bacteria have had billions of years of evolution to cope with harsh environments. The limited number of antibiotic classes means that all drugs in use continue to target the same group of mechanisms, attacking four chinks in the bacterial armour: cell wall synthesis; protein synthesis; DNA/RNA precursor synthesis and DNA/RNA synthesis. Since some bacterial strains have developed resistance to all these methods of attack, new tactics are desperately needed.

Lysostaphin is considered a good alternative to antibiotics, although it is unlikely to replace it, because it has the properties of cleaving specifically the cross-linking pentaglycine bridges in the cell walls of staphylococci and thus display a narrow spectrum of lytic activity. Unlike most antibiotics that interfere with the growth of bacteria, lysostaphin is highly effective in lysing both actively growing and dormant *S. aureus* cells. Nonetheless, *S. simulans* produces and protects itself from lysostaphin. Production of the endopeptidase appears to act as a selective pressure that removes the susceptible cells in the population, as the enzyme appears in the medium. The FemAB-like factors: lysostaphin immunity factor (Lif) and lysostaphin endopeptidase resistance factor (Epr), confer resistance to lysostaphin by replacing a glycine by a serine in the peptidoglycan crosslink. Thus, we can predict that other staphylococcal strains (not to mention the example of *S. aureus*) could acquire, by horizontal

transfer, the genes (*lif* or related gene) that confer *S. simulans* resistance to lysostaphin.

The aim of this project was to investigate the mechanism by which lysostaphin interacts with its substrate. With this knowledge, in future studies, we wished to evolve this bacteriocin to give it new binding and catalytic properties to thwart the *lif* mechanism of resistance. By anticipating the resistance of staphyloccocal cells to wild-type lysostaphin and providing mutants able to bypass the mechanisms of resistance, we would tackle superbugs.

Thus, the approach described in this thesis was to obtain the lysostaphin catalytic domain (termed LssEdom in the thesis) structure, so that in future studies, by using structure-based enzyme engineering, we could modify the amino acids in interaction with the substrate to expend its substrate specificity. An additional approach described in this thesis was to define the targets of the lysostaphin targeting domain (termed LssTdom in the thesis) on the bacterial cell wall. In future studies, by directing its evolution using the phage display technique, we hoped that lysostaphin would acquire altered binding properties.

One strategy adopted was to perform crystallisation trials on active LssEdom with the aim of producing highly ordered crystals that diffract X-rays at high resolution to determine peptide structure by X-ray crystallography. To conduct crystallisation trials, the endopeptidase domain of lysostaphin was sub-cloned and expressed as a His-tagged peptide in order to facilitate its purification. Optimisations of both, LssEdom activity and LssEdom crystallisation process were investigated by modifying the protocol of

purification and by removing the His-tag of the peptide, respectively. As well, a structure model of LssEdom *in silico* was investigated. Considering the difficulties encountered to crystallise LssEdom, another strategy to obtain its structure emerged. Fourier-transform infrared (FT-IR) spectroscopy technique was explored for the determination of protein secondary structure in solution. Finally, to define the targets of the lysostaphin targeting domain (LssTdom) on the bacterial cell wall, we looked at protein-ligand interactions using mass spectrometry (ESI-ToF).

2. Materials and methods

2.1. Techniques of molecular biology

2.1.1. Bacterial strains and media

Strains used with their relevant genotypes and sources are listed in Table 1. Chemically competent *E. coli* were transformed by the heat-shock transformation technique (Sambrook *et al.*, 1989).

Table 1: List of strains with their relevant genotypes and sources.

***E.coli* strains**	**Relevant genotypes**	**Source**
DH5α	lacZΔM15 recA1	Invitrogen
BL21 (DE3)	araB::T7	Invitrogen

Recombinant clones that acquired resistance to ampicillin were selected on LB or 2YT agar plates supplemented with ampicillin (75 $\mu g.ml^{-1}$, final concentration). Bacterial cultures were grown in LB or 2YT broth (Sambrook *et al.*, 1989) supplemented with ampicillin (50 $\mu g.ml^{-1}$, final concentration). Isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM to induce expression of T7 RNA Polymerase in BL21 cultures. Protein for NMR studies was obtained by expression in M9 Minimum Medium (Na_2HPO_4 (10 g), KH_2PO_4 (3 g), NaCl (0.5 g), $MgSO_4.7H_2O$ (0.3 g), $CaCl_2$ (0.015 g), NH_4Cl (1 g), glucose (4 g), biotin (10 mg), thiamine (10 mg) in 1 L milliQ grade H_2O). Protein uniformly labelled with isotope ^{15}N was obtained by expression in ^{15}N-enriched minimum media (^{15}N NH_4Cl, 1 $g.l^{-1}$).

2.1.2. Plasmids

The pET21a(+) vector (Novagen) contains a multiple restriction site, shown in Figure 14, in which the protein sequence of recombinant lysostaphin (Lss) and lysostaphin endopeptidase domain (LssEdom) were inserted at the *Nde*I/*Xho*I restriction sites by Dr E. Antoniadou and Dr P. Bardelang (Centre for Biomolecular Sciences, University of Nottingham, U.K.). The vectors built in this way were called respectively pEA3 (pET21a(+)-$Lss_{248-493}$) and pETedom (pET21a(+)-$Lss_{248-385}$). The protein sequence of recombinant lysostaphin targeting domain (LssTdom) was sub-cloned in pET21d(+) vector (Novagen) at the *Nco*I/*Xho*I restriction sites by Dr E. Antoniadou. The vector built in this way was called pEA7 (pET21d(+)-$Lss_{401-493}$). Expression was under control of strong bacteriophage T7 transcription and translation signals. The *C*-terminal His-tag fused to the recombinant proteins enabled rapid purification by metal-affinity chromatography. The calculated molecular weight* of this *C*-terminal His-tagged Lss is 28,141 Da and its pI is 9.6. For the endopeptidase domain the calculated molecular weight* is 16,386 Da and the pI is 8.8 and for the targeting domain the calculated molecular weight* is 11,623 Da and the pI is 9.7.

* Calculated with Expasy software: http://www.expasy.org/tools/pi_tool.html

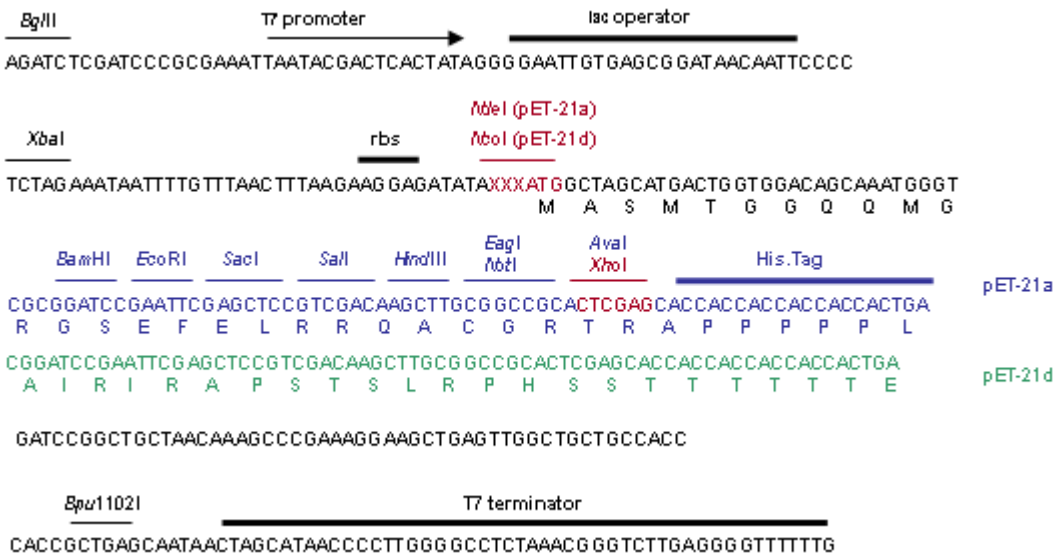


Figure 14: pET-21a-d(+) cloning/expression region.

The pGAT2 plasmid (originally constructed by Dr. M. Hyvönen, University of Cambridge, U.K.) (Peranen *et al.*, 1996), shown in Figure 15, carries a T7 promoter inducible with IPTG and a sequence coding for a His-tag, which is fused to the *N*-terminal region of the protein when expressed. Notably, this vector contains a sequence coding for a thrombin cleavage site in between the sequence coding for a His-tag and the multiple restriction site. pSGAT2 is a pGAT2 vector for which the GST sequence has been removed by restriction with *Spe*I. The third glycine (Gly+18 downstream of Met+1) in the pentaglycine sequence has been mutated into a serine to give a pSGAT2$_{Gly \rightarrow Ser}$ vector. The protein sequences of recombinant Lss and LssEdom have been inserted in this vector, in two parallel experiments, at the *Bam*HI/*Hind*III restriction sites. For *N*-terminal His-tagged Lss the calculated molecular weight* is 29,668 kDa and the pI is 9.7 and for the endopeptidase domain the calculated molecular weight* is 17,914 kDa and the pI is 9.3.

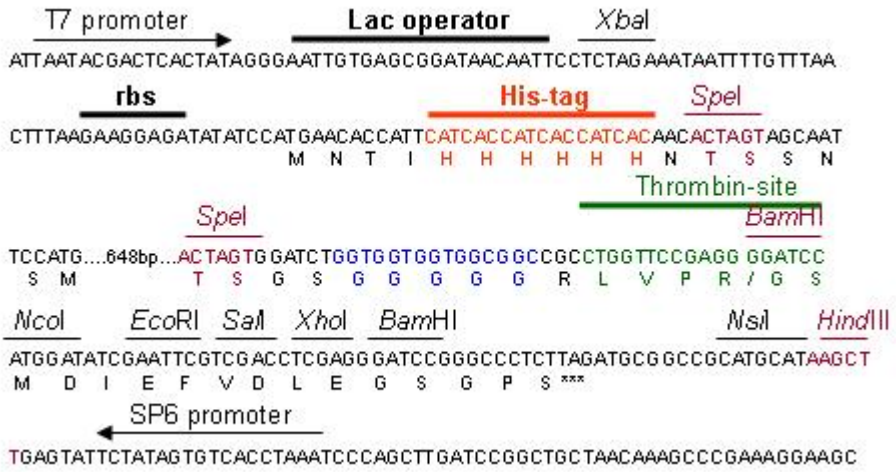


Figure 15: pGAT2 cloning/expression region.

2.1.3. Electrophoresis, DNA extraction and quantification

Agarose gel electrophoresis was completed as detailed by Sambrook and coworkers (Sambrook *et al.*, 1989). DNA ladder standards (New England Biolabs) were loaded alongside DNA samples to enable estimation of their length and concentration.

Extraction of DNA from agarose gels was completed using the QIAEX II kit (Qiagen) as directed by the manufacturer's instructions.

To determine DNA concentrations, a sample of 1 µl was analysed by NanoDrop™ 1000 spectrophotometer (Thermo Scientific). The resulting spectral diagram showed the absorbance of the solution in the 220-350 nm range of wavelength. Concentration was confirmed by UV absorbance at 260 nm. An approximate conversion factor estimates that duplex DNA is about 50 $\mu g/A_{(260nm)}$.

2.1.4. Plasmid restriction, ligation, transformation

Endonuclease digestion of DNA was completed as detailed by Sambrook and coworkers (Sambrook *et al.*, 1989). Ligation of DNA fragments into plasmids was performed using the Rapid DNA Ligation Kit (Roche Diagnostics GmbH). Transformation of competent *E. coli* cells was completed as detailed by Sambrook and coworkers (Sambrook *et al.*, 1989).

2.1.5. Polymerase Chain Reaction (PCR)

PCR was completed using an Eppendorf Mastercycler Personal. Taq DNA Polymerase (Roche, Penzberg, Germany) was used for standard reactions and Pwo DNA Polymerase (Roche, Penzberg, Germany) was used for high fidelity reactions. The PCR cycles were realised according to a standard program that consisted of a hot start at 94°C for 2 minutes, followed by 30 cycles of denaturation (94°C for 30 seconds), hybridisation (T_M[†] for 30 seconds), extension (72°C for n[‡] seconds) and finished with a final extension at 72°C for 10 minutes. Products of reaction were conserved at 4°C. Details of the oligonucleotides used (synthesised by MWG-Biotech, U.K.) are shown in Table 2. Oligonucleotide secondary structures were predicted utilising software found in http://mfold.bioinfo.rpi.edu/.

[†] Melting temperature (TM) is the temperature at which an oligonucleotide duplex is 50% in single-stranded form and 50% in double-stranded form, which depends on its nucleotide content.

[‡] n corresponds ideally to 2 minutes extension for the amplification of 1 DNA kilobase with high fidelity polymerase.

Oligonucleotide ID (restriction site)	**Sequence (5'→3')**
LssForw (*Bam*HI)	GAGATATAGGATCCGCTGCAACACAT
LssRev (*Hind*III)	GTGGTGAAGCTTACTTTATAGTTCC
Use: Lss sequence amplification in pEA3 for sub-cloning in pSGAT2	
LssForw (*Bam*HI)	GAGATATAGGATCCGCTGCAACACAT
LssEdomRev (*Hind*III)	GTGGTGAAGCTTACTTTATAGTTCC
Use: LssEdom sequence amplification in pEA3 for sub-cloning in pSGAT2	
T7PP	TAATACGACTCACTATAGGG
Gly3mutpSGAT2 (*Bam*HI)	AGCGGATCCCCTCGGAACCAGGCGGCCGCCAGAA CCACCAGATCC
Use: Mutate G18 to S in pSGAT2-LssEdom and pSGAT2-Lss	
pSGAT2EdomFor (*Bam*HI)	CCGAGGGGATCCGCTGCAACACAT
pSGAT2EdomRev (*Hind*III)	ATACTCAAGCTTATCCTGCGCTCTT
Use: Insert a Stop codon in pSGAT2-Ser-LssEdom	
T7PP	CCGAGGGGATCCGCTGCAACACAT
T7TP	ATACTCAAGCTTATCCTGCGCTCTT
Use: Sequencing	
WC1	GCATAGACAATTCTATATGCATCTAAGTAAAT
WC2	GATGCATATAGAATTGTCTATGCACTCCATC
T7For (*Nde*I)	GCATATGGCTGCAACACATGAACATTC
T7Rev (*Xho*I)	GCTCGAGTCCTGCGCTCTTTAAGAAAGG
Use: Mutate W80 to F in pETedom	
WC3	GTTGAAGCTGGTTTCAGTAATTACGGAGGAGGT
WC4	GTAATTACTGAAACCAGCTTCAACTATTTTTCC
T7For (*Nde*I)	GCATATGGCTGCAACACATGAACATTC
T7Rev (*Xho*I)	GCTCGAGTCCTGCGCTCTTTAAGAAAGG
Use: Mutate W59 to F in pWC3	

Table 2: List of oligonucleotide primers used in cloning.

2.1.6. Site-directed mutagenesis

Site-directed mutagenesis was completed in a two-step PCR modified method (Sarkar and Sommer, 1990). The method utilised four oligonucleotides (cf. section 2.1.5), two flanking oligonucleotides and two internal oligonucleotides carrying the mutation of interest, to perform two parallel rounds of polymerase chain reaction (cf. section 2.1.5). The two megaprimers generated were used to polymerise the complete DNA sequence with the desired mutation. Mutation W80 to F in pWC3 and mutations W59/80F in pWC5 were performed by W. Cheng (Centre for Biomolecular Sciences, University of Nottingham, U.K.) under my supervision. Mutation H114 to A in pEA3 was performed by Dr E. Jarvis (Centre for Biomolecular Sciences, University of Nottingham, U.K.).

2.1.7. DNA sequencing and sequence analysis

To prepare the sequencing reaction, 1 µl of the template, DNA plasmid, was used at a concentration of 100 $ng.\mu l^{-1}$. To this template was added 5 µl of forward (T7PP) or reverse (T7TP) primers, 30 $pmol.\mu l^{-1}$. DNA sequencing reactions were completed using the Applied Biosystems BigDye® Terminator v3.1 Cycle Sequencing Kit and 96-Well GeneAmp® PCR System 9700. The reaction involved 25 cycles of 96ºC for 10 seconds, 50ºC for 5 seconds and 60ºC for 4 minutes. The DNA was ethanol precipitated and reactions were run on a 3130xl Genetic Analyzer (Applied Biosystems) by sequencing technician, S. Ouadi, at the Queens Medical Centre (QMC) University DNA sequencing laboratory (Nottingham, U.K.).

The results of the sequencing were analysed with ClustalW software[§] (Thompson *et al.*, 1994) for alignment studies of biopolymer sequences.

2.1.8. Protein expression

2.1.8.1. Experimental plan for expression

A volume of 4 ml of a 3 hour 2YT broth culture of *E. coli* BL21 cells, transformed with the relevant plasmid, was used to inoculate 400 ml 2YT broth. When an OD_{600nm} of 0.6 was reached, the cells were transferred from a 37°C to a 30°C incubator and induced by addition of IPTG to 1mM. After two hours of induction at 30°C the cells were harvested by centrifugation at 10,000g for 2 x 10 minutes in a Beckman Avanti J-20 centrifuge. Cells were stored at -20°C.

Expression of LssEdom for NMR studies was realised at an analytical scale according to an outlined experimental plan (Abergel *et al.*, 2003). Three temperatures (20ºC, 25ºC, 30ºC) and three media (M9MM (section 2.1.1), M9MM with added 1mM zinc chloride, M9MM with added 200 mM sodium chloride) were tested as shown in Table 3. A volume of 150 µl of a 4 ml over-night LB broth culture of *E. coli* BL21 cells, transformed with the relevant plasmid, was used to inoculate 4 ml of the media cited.

Table 3: Experimental plan for expression of LssEdom for NMR studies.

Bacterial strains	Medium		
	20ºC	**25ºC**	**30ºC**
BL21 (DE3)	M9MM + NaCl	M9MM + $ZnCl_2$	M9MM
BL21 (DE3)	M9MM + $ZnCl_2$	M9MM	M9MM + NaCl
BL21 (DE3)	M9MM	M9MM + NaCl	M9MM + $ZnCl_2$

[§] http://www.ebi.ac.uk/clustalw/

2.1.8.2. Cell lysis

Frozen cells were resuspended in 30 ml buffer A (20 mM sodium phosphate (pH 7.4), 5 mM imidazole, 0.5 M sodium chloride, 5% (v/v) glycerol) with 300 µl phenyl methyl sulfonyl fluoride (1 mM) and disrupted in a French Press. Cell lysates was separated from insoluble cell debris after two successive 30 minute centrifugations at 10,000 rpm.

2.2. Analytical methods

2.2.1. Protein purification

2.2.1.1. Metal affinity chromatography

A standard metal-affinity chromatography protocol was employed for the purification of His-tagged Lss, LssEdom and LssTdom and their mutants. Proteins were purified using HiTrap^{TM} chelating HP columns (Amersham Biosciences) connected to a BiologicLP low-pressure liquid chromatography system (BioRad). The column was charged with nickel or zinc ions using a 12 ml injection of 50 mM nickel sulfate or a 24 ml injection of 50 mM zinc chloride, washed in water and equilibrated in buffer A (20 mM sodium phosphate (pH 7.4), 5 mM imidazole, 0.5 M sodium chloride, 5% (v/v) glycerol). The sample was applied to the column at a flow-rate of 1 ml.min^{-1}. The loaded column was washed in 5% buffer B (95% buffer A) (20 mM sodium phosphate (pH 7.4), 0.5 M imidazole, 0.5 M sodium chloride, 5% (v/v) glycerol) until no further unbound proteins were eluted from the column. Bound proteins were eluted from the column using a linear gradient of 5-85% buffer B in a total volume of 40 ml. The change in absorbance at 280 nm of the protein eluted was recorded over time. Fractions containing the required protein, identified by SDS-PAGE, were pooled and dialysed overnight (4°C) in

the appropriate buffer. His-tagged Lss and LssEdom were diluted 10-fold into cold buffer iexA (10 mM Tris-HCl (pH 7.4), 50 mM sodium chloride, 5% (v/v) glycerol) and dialysed in buffer iexA overnight (4°C), ready for ion-exchange chromatography.

2.2.1.2. Ion-exchange chromatography

Lss (pI~9) or LssEdom (pI~9), with or without His-tag, were subjected to cation-exchange chromatography. This step allowed thrombin (pI~6) and contaminants to be washed away. A 5 ml HiTRAP SP column (Pharmacia Biotech) connected to a BiologicLP low-pressure liquid chromatography system (BioRad) or a Mono S 5/50 GL (Tricorn^{TM}) column connected to an ÄKTA™ purifier liquid chromatography system (Amersham Biosciences) were used. All the columns were equilibrated in iexA buffer (10 mM Tris-HCl (pH 7.4), 50 mM sodium chloride, 5% (v/v) glycerol). Lss or LssEdom in solution in iexA buffer were loaded on to the column at a flow-rate of 1 ml.min^{-1} and the column washed with 25 ml buffer iexA. Proteins bound to the column were eluted using a linear gradient of 0-100% buffer iexB (10 mM Tris-HCl (pH 7.4), 1 M sodium chloride, 5% (v/v) glycerol) in a volume of 60 ml. The change in absorbance at 280 nm of the protein eluted was recorded over time. Fractions containing the required protein, identified by SDS-PAGE, were pooled and dialysed overnight (4°C) in the appropriate buffer.

2.2.1.3. Gel filtration chromatography

A Superdex^{TM} 75 10/300 GL or Superdex^{TM} 200 10/300 GL (Tricorn^{TM}, Amersham Biosciences) column with an optimum separation range of 3,000 – 70,000 Da or 10,000 – 600,000 Da respectively, attached to an ÄKTA Explorer FPLC system (Amersham Biosciences), were equilibrated

with two column volumes of 50 mM sodium phosphate (pH 7.4), 150 mM sodium chloride (0.2 μm filtered and degassed with helium) at a flow rate of 0.8 $ml.min^{-1}$. Sample protein was then loaded onto the column at a flow rate of 0.5 ml min^{-1}. The change in absorbance at 280 nm of the protein eluted was recorded over time. Fractions containing the required protein, identified by SDS-PAGE, were pooled and dialysed overnight (4°C) in the appropriate buffer.

2.2.1.4. His-tag removal

Cleavage of proteins at the excision site Leu-Val-Pro-Arg↓Gly-Ser for His-tag removal was complete following 16 h treatment with thrombin, at 19°C, with 3 units of thrombin per mg of protein, in iexA buffer (10 mM Tris-HCl (pH 7.4), 50 mM sodium chloride, 5% (v/v) glycerol).

2.2.2. Electrophoretic analysis

Sodium Dodecyl Sulfate-PolyAcrylamide Gel Electrophoresis (SDS-PAGE) was completed using a mini gel system (Atto). Stacking (6%) and resolving (16%) gels were prepared as described by Sambrook and coworkers (Sambrook *et al.*, 1989). Protein samples were loaded onto a gel after 1 minute denaturation at 95°C in an equal volume of 2 x loading buffer (100 mM Tris-HCl (pH 6.8), 4% (w/v) SDS, 0.2% (w/v) bromophenol blue dye, 200 mM DTT). Protein molecular weight standards (New England Biolabs) were loaded alongside protein samples to enable estimation of molecular weights.

Gels were stained over 30 minutes using Coomassie blue stain (0.2% bromophenol blue dye, 45% methanol, 10% acetic acid) and destained twice in fast destain solution for 10 minutes (40% methanol, 10% acetic acid) followed by slow destain solution over-night (10% methanol, 10% acetic acid).

2.2.3. Dynamic Light Scattering (DLS)

Dynamic light scattering (DLS) measurements were carried out on a Zetasizer Nano S from Malvern Instruments. Each sample was filtered through a 0.2 µm filter (Millipore) prior to measurements, which were made at room temperature. The particle size was taken as the mean value of three measurements. LssEdom, LssEdomW80F and LssEdomW59/80F (15 µM), buffer exchanged in 1 mM sodium phosphate (pH 7.0) buffer, were measured.

2.2.4. Mass Spectrometry

2.2.4.1. Principles of Mass Spectrometry

Mass spectrometry (MS) was the analytical technique chosen to identify proteins and protein/ligand interactions on the basis of the mass-to-charge ratio of the charged particles. The mass spectrometer was designed with three essential modules (Figure 16): an electrospray ion source, which generated intact, positively charged ions, in the form of $[M+nH]^{n+}$; an orthogonal-time-of-flight (o-ToF) analyser, which focused and sorted the ions along their intended flight path and separated them according to their mass-to-charge (*m/z*) ratio; and a detector, which measured the value of indicator quantity and thus provided data for calculating the abundances of respective ions.

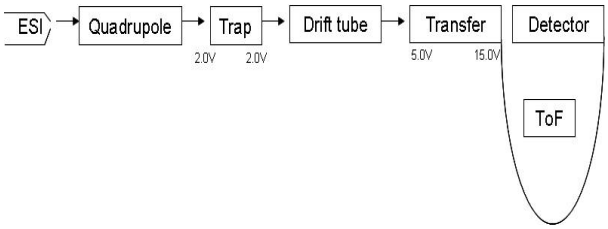


Figure 16: Schematics of a mass spectrometer with an ESI source combined with a TOF analyser, using a combination of quadrupole filters orthogonally coupled to a TOF tube, for the transmission and analysis of protein complexes.

Electrospray ionisation (ESI) was the technique chosen to produce ions from proteins and protein non-covalent complexes without disrupting them in the gas-phase. ESI is a commonly used ionisation method for MS studies of protein complexes (Fenn *et al.*, 1989, Kebarle, 2000, Wilm and Mann, 1996). Pressure is applied to deliver an appropriate flow of liquid through a needle maintained at kilovolt potential. By passing through a high electric field density the emerging liquid droplets become charged. This charging leads to droplets with an elongated minuscus (Taylor cone) (Mora *et al.*, 2000, Kebarle, 2000). These residual charges are subsequently transferred to proteins during their desolvation under a stream of dry gas in the sample chamber, often in the form of inorganic ions (e.g. triethylammonium ions) (Peschke *et al.*, 2002, Lemaire *et al.*, 2001). These excess charges are transferred to basic sites on the surface of globular proteins. Gas-phase ions are produced when the coulombic repulsion in the reduced droplet size generates droplet fission into even smaller droplets (Rayleigh, 1882). This ionisation process takes place at atmospheric pressure and is, therefore, very gentle. The most efficient electrospray are carried out in a certain proportion of organic solvent at relatively high temperatures to aid the spray, charging and desolvation processes.

Once the molecular species ($[M+nH]^{n+}$) are produced they are transferred through the quadrupole, trap, drift tube and transfer regions to the mass analyser by a series of rf and dc voltages. A discrete pulse of molecular ions is injected into a field-free “drift tube”, measuring the time taken for ions to reach the detector. With instruments incorporating an orthogonal time-of-flight (o-ToF) mass analyser, the trajectories of ions are determined not only

by the acceleration imparted in the orthogonal extraction source but also by the axial velocity of the ions as they enter this region. Direct measurement of the velocity (υ) of the charged molecules is used in the time-of-flight (ToF) technique (Tang *et al.*, 1994). Non-covalent complexes can be examined as a result of the coupling of time-of-flight analysis with electrospray ionisation, achieved through the use of an orthogonal acceleration process with a mass range up to 20,000 *m/z*. The detector records ions as a function of *m*/*z* where *m* is the mass of the ion and *z* is the ion charge

When representing data, we used the dimensionless m/z, where z is the number of elementary charges (e) on the ion (z=Q/e). This quantity represents the ratio of the mass number and the charge number, z.

2.2.4.2. Analysis

Electrospray mass spectrometry (ESI-MS) of protein samples was carried out on a Bruker MicroTOF instrument using direct infusion of the sample solutions via a syringe pump at a flowrate of 4 $\mu l.min^{-1}$ by G. Coxhill (School of Chemistry, University of Nottingham, Nottingham). Thoroughly desalted proteins were supplied as 5 μM solutions in the appropriate buffer, containing methanol (buffer:methanol 1:1 (v/v)) and acetic acid, 1%.

2.2.4.3. Interaction studies

Of the different buffers tested, triethylammonium acetate (25 mM, pH 7.4), was the buffer chosen to dissolve LssTdom. The following ligands: Gly_5 (G_5), Gly_4 (G_4), Gly_3 (G_3), Gly_2 (G_2) Gly-L-Ser (GS), L-Ser (S), *N*-acetylmuramyl-Ala-D-isoglutaminyl-N^{ε}-stearoyl-Lys (NAM-Ala-iGlu-Lys), acetyl-Lys-D-Ala-D-Ala (KA_2), were obtained from Sigma-Aldrich (Poole, UK). The ligand Gly-Gly-L-Ser-Gly-L-Ser (G_2SGS) was synthesised at the

Queen Medical Centre University synthesis laboratory (Nottingham, UK). Gly_5 (14.53 mg) was dissolved in 5 µl trifluoroacetic acid (TFA), evaporated by Speed Vacuum. The sample was resuspended in 1 ml milliQ water and lyophilised three times in the presence of 0.1 M HCl, directly replacing TFA counterions with chloride ions (Andrushchenko *et al.*, 2007). All the ligands cited were resuspended in milliQ water. For K_D measurements, the thoroughly desalted LssTdom protein was used at a concentration of 5 µM in buffer with added methanol (10%). Ligands interacting non-covalently with LssTdom were titrated into the protein solution to give final (total) concentrations in the range of 10-40 µM, whilst the protein concentration was held constant. Upon mixing, samples were allowed to equilibrate for 30 min before measurement.

ESI-MS of LssTdom in interaction with small ligands was carried out on a SYNAPT™ High Definition Mass Spectrometry™ (HDMS™) System (Waters, MA) equipped with an ESI source and an orthogonal acceleration time-of-flight (o-ToF) mass analyser. The conditions within the mass spectrometer were adjusted to preserve non-covalent interactions. Samples were infused into the Mass Spectrometer from a gas-tight syringe at a flowrate of 5 $\mu l.min^{-1}$. MS experiments were performed at a capillary voltage of 2.5 kV, desolvation gas flow of 587 $L.h^{-1}$ and desolvation gas temperature of 50ºC. The quadrupole allowed the detection of ions at over 8000 *m/z* mass range. Data were analysed using MassLynx™ software.

Spectra were acquired in the range *m/z* 500 to 4000 at a rate of 1 $scan.s^{-1}$ for 116 scans. Averaged spectra were smoothed (smooth window ± 8; number of smooths 4) and the signal intensities of the two largest charge states ($[M+4H]^{4+}$ and $[M+5H]^{5+}$) combined. The apparent concentrations of free

proteins ([P]), protein-ligand complex ([PL]) and free ligand ([L]) were calculated from the concentration of total protein and ligand used, and the signal intensities. It was assumed that P and PL gave the same linear detector response. Plots of [PL]/[L] vs. [L] were used to determine K_D for the ligands employed. Plots of [PL] vs. [inhibitor] were used to determine IC_{50} in the competitive binding experiments.

2.2.5. Determination of protein concentration

To determine protein concentrations, a 2 μl sample was analysed by NanoDrop™ 1000 spectrophotometer (Thermo Scientific). The resulting spectral diagram showed the absorbance of the solution in the 220-350nm range of wavelength. The concentration was confirmed by UV absorbance at λ 280 nm. The expression of Beer's law: $A = \varepsilon_{280nm}.l.C$, was applied to deduce the concentration (C) of the sample from the absorbance value (A), with a pathlength for the cell of 1 cm. The molar absorption coefficients of the proteins were determined using their tryptophan (W), tyrosine (Y) and cysteine (C) amino acid composition, as described in the following equation (Gill and vonHippel, 1989, Pace, 1995):

$$\varepsilon = (nW \times 5500) + (nY \times 1490) + (nC \times 125)$$

where n is the number of each residue and the stated values are the amino acid molar absorptivities at 280 nm. The calculated molar extinction coefficients of the purified proteins used in this thesis are listed in Table 4.

Table 4: List of the molecular weights, isolelectric points and molar extinction coefficients of purified proteins used in this thesis.

Protein	Calculated molecular weight (Da)	pI	Molar extinction coefficient ($M^{-1}.cm^{-1}$)
Lss-H6 (pEA3)	28,141	9.6	67,840
LssEdom-H6 (pETedom)	16,386	8.8	36,900
LssTdom-H6 (pEA7)	11,623	9.7	29,450
H6-Lss (pSGAT2-Lss)	29,668	9.7	67,840
H6-LssEdom (pSGAT2-LssEdom)	17,914	9.3	36,900
LssEDW80F-H6 (pWC3)	16,347	8.8	31,400
LssEDW59/80F-H6 (pWC5)	16,308	8.8	25,900
LssH114A-H6 (pEJ1)	28,074	9.5	67,840

To obtain a higher concentration of proteins, centrifugation of protein solutions using 5ml Vivaspin Vertical filter tubes (VivaScience) with a molecular weight cut off of 3,000 Da or 10,000 Da when appropriate was performed.

2.2.6. Determination of protein activity

2.2.6.1. Substrate for protein activity determination

A figure of the protein sequence MV11 is shown in Figure 17. Dr M. Vankemmelbeke (Centre for Biomolecular Sciences, University of Nottingham, Nottingham) designed this protein. The 57 residue *N*-terminal domain of MV11 was originated from the translocation domain of colicin E9 and the *C*-terminal domain was made of a green fluorescent protein (GFP), fused to it by a 7 residue linker. Additionally, MV11 contained a 6 residue *C*-terminal His-tagged sequence and a lysostaphin cleavage site (pentaglycine sequence) introduced by mutating the colicin E9 sequence. The introduction of

a N44C mutation in the recombinant protein allowed derivatisation with AlexaFluor 546 to result in a biological FRET substrate in which the AlexaFluor 546 would quench the fluorescence emission of GFP. The relative molecular mass of this biological substrate, in which a lysostaphin cleavage site has been introduced, is 33,800 Da (not included AlexaFluor 546).

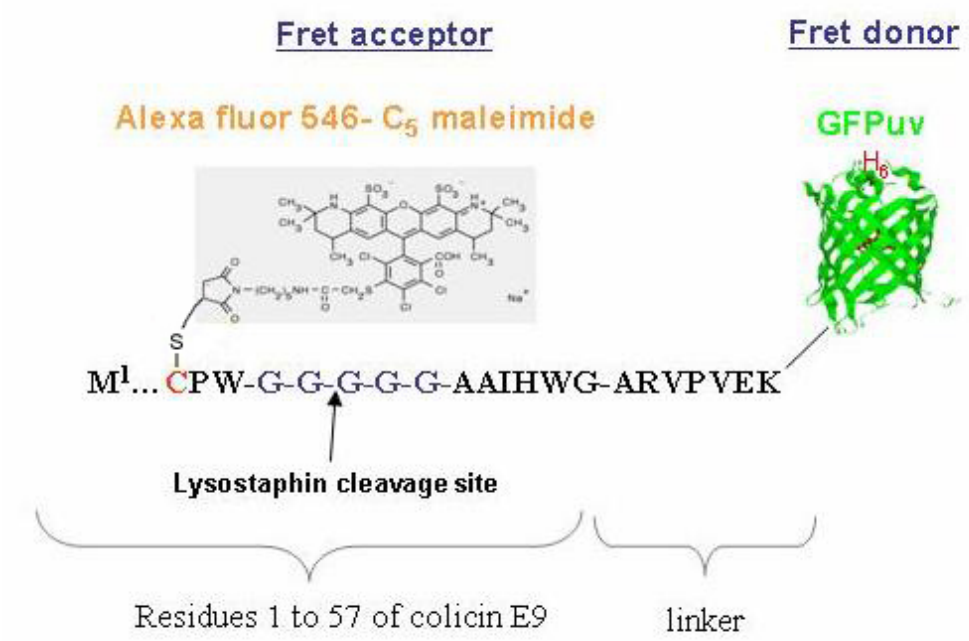


Figure 17: Internally quenched MV11 substrate for peptide activity determination.

2.2.6.2. Pentaglycine cleavage assay

Lss and LssEdom activity was tested on MV11 substrate. The enzyme, 1 μM, was mixed with MV11, 1 μM, in buffer (25 mM sodium phosphate (pH 7.0), 150 mM sodium chloride). The cleavage reaction was stopped after 1 h, 2 h and 3 h by chilling the reaction vessel in ice. The cleavage reaction was analysed by SDS-PAGE.

2.2.7. Kinetic parameters

2.2.7.1. Principles of Föster Resonance Energy Transfer (FRET)

Föster Resonance Energy Transfer (FRET) describes an energy transfer mechanism between two chromophores. A donor chromophore in its excited state transfers energy by a non-radiative, long-range dipole-dipole coupling mechanism to an acceptor chromophore in close proximity. The FRET efficiency (E) is described by the Föster equation and depends on the donor-to-acceptor separation distance *r* with an inverse 6th power law due to a dipole-dipole coupling mechanism (Förster, 1948):

$$E = \frac{1}{1 + (r / R_0)^6}$$

with R_0 being the Föster distance of this pair of donor and acceptor at which the FRET efficiency is 50%.

The Föster distance depends on the overlap integral of the donor emission spectrum with the acceptor absorption spectrum, highlighted in Figure 18, and their mutual orientation as expressed by the following equation:

$$R_0^{\ 6} = 8.8 \times 10^{-28}\, \kappa^2\, n^{-4}\, Q_0\, J$$

where κ^2 is the dipole orientation factor, *n* is the refractive index of the medium, Q_0 is the fluorescence quantum yield of the donor and *J* is the spectral overlap integral.

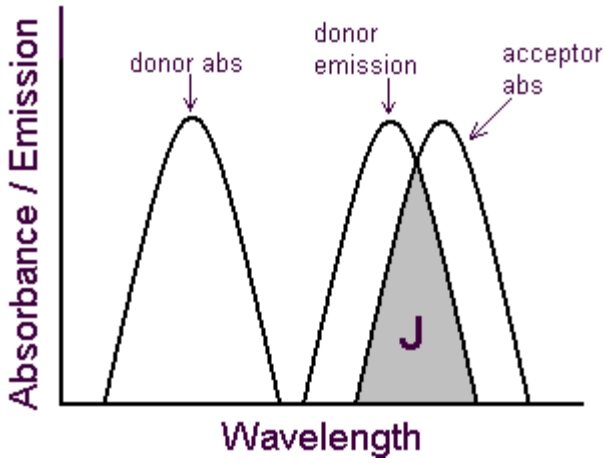


Figure 18: Overlap (J) of the donor emission with the absorbance spectrum of the acceptor in a FRET pair.

The FRET effect was employed as a technique for monitoring the rate of hydrolysis of the pentaglycine bond by lysostaphin and its endopeptidase domain and thus for determining the kinetic constants for hydrolysis, V_{max} and K_M. Figure 19 outlines the mechanism by which the rate of increase in donor fluorescence is proportional to the rate of cleavage of the linker.

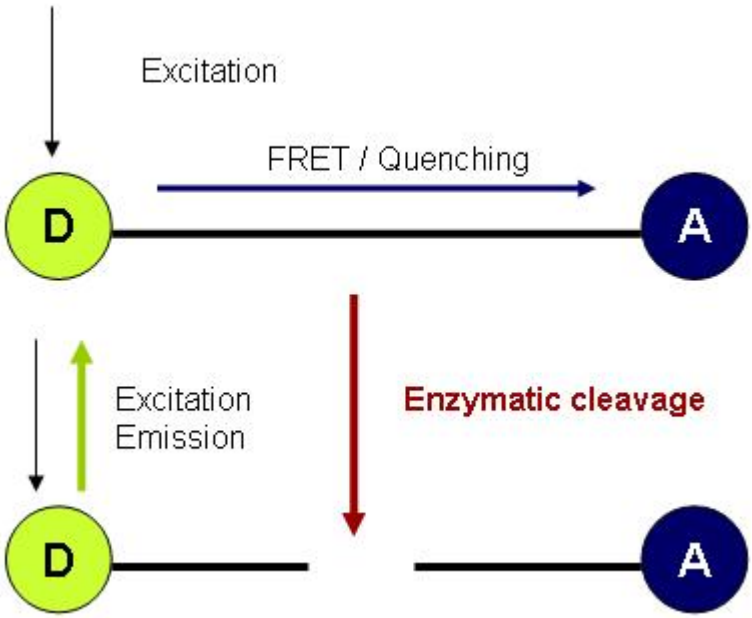


Figure 19: Fluorescence Resonance Energy Transfer prior and following cleavage of a linker chain.

In the first part of the diagram, a donor fluorophore (D) is held within 100 Å of an acceptor group (A) by a linker chain. Internal quenching occurs when the acceptor is absorbing energy at the donor emission wavelength, which results in limited fluorescence from the donor group across the spectrum. Enzymatic cleavage of the linker enables the two groups to move further apart so that internal quenching no longer occurs, and emission from the donor is restored.

2.2.7.2. FRET substrate

The substrate used to investigate the cleavage by Lss of the pentaglycine sequence is shown in Figure 20. It consists of a pentaglycine chain that has a *N*-(2-aminobenzoyl) (Abz) group on one end, and *N*-(2,4-dinitrophenyl) ethylene diamine (EDDnp) group on the other. It was synthesised by Dr R. Warfield (Centre for Biomolecular Sciences, University of Nottingham, Nottingham) (Warfield *et al.*, 2006). The Abz group is a fluorophore that acts as a donor (λ_{max} (absorption) = 325nm, λ_{max} (emission) = 420 nm) and the EDDnp group acts as a quencher (λ_{max} (absorption) = 365 nm), for the FRET studies.

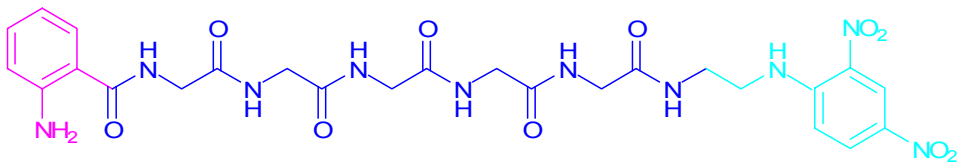


Figure 20: FRET assay substrate Abz-Gly$_5$-EDDnp: pentaglycine chain with an *N*-(2-aminobenzoyl) (Abz) group on one end and *N*-(2,4-dinitrophenyl) ethylene diamine (EDDnp) group on the other

The substrate used to investigate the cleavage of the pentaglycine sequence by LssEdom and its mutants, LssEdomW80F and LssEdomW59/80F, was MV11 (section 2.2.5.1).

2.2.7.3. Initial study into kinetic parameters

Fluorescence experiments were carried out using a Perkin Elmer LS50B luminescence spectrophotometer with a 5 mm quartz cuvette pathlength under the control of FLWinlab software. The temperature was maintained at 37˚C throughout these experiments, using a circulating water bath. The buffer/substrate solution was allowed to equilibrate for 5 min before addition of protein.

For the initial studies into V_{max} for the cleavage of Abz-Gly_5-EDDnp by Lss, the excitation wavelength used in each case was 325 nm (obtained with a xenon discharge lamp excitation source) and scans were run to measure the emission at 420 nm (excitation slit 4.0 nm). Runs that were 20 min in length were performed with 10 scans at 2 min intervals for each concentration of substrate. The substrate concentrations used were 10 μM, 20 μM, 30 μM, 40 μM, 50 μM and 70 μM, taken from a 0.39 mM stock solution in dimethyl sulfoxide (DMSO). The concentration was confirmed by UV absorbance at 20°C at λ 359 nm (ε_{359nm}=17700 $M^{-1}.cm^{-1}$) and 365 nm (ε_{365nm}=17100 $M^{-1}.cm^{-1}$). The Beer-Lambert equation: A=ε.l.C, was applied to deduce the concentration (C) of the substrate from the absorbance value, with a calculated pathlength for the spectrophotometer cell of 1 cm. The protein concentration used in each case was 0.833 μM in 25 mM sodium phosphate (pH 7.0) and 150 mM sodium chloride buffer.

For the initial studies into V_{max} for the cleavage of MV11 by the lysostaphin endopeptidase domain and mutants W80F and W59/80F, the excitation wavelength used in each case was 475 nm and scans were run to measure the emission at 508 nm (excitation slit 5.0 nm). Runs that were 10

min in length were performed with 4 scans at 10 sec intervals for each concentration of substrate. The substrate concentrations used were 0.2 μM, 0.4 μM, 0.5 μM, 1.0 μM, 1.5 μM and 2.0 μM taken from a 0.19 mM stock solution in 1 mM sodium phosphate (pH 7.0) buffer. The protein concentration used in each case was 0.100 μM in 1 mM sodium phosphate (pH 7.0) buffer.

The Grafit© [**]data-fitting program was used to calculate K_M and $V_{max.}$

2.2.7.4. Inner filter effect correction

For Abz-Gly_5-EDDnp, absorbance readings were taken at 37˚C, at 325 nm and 420 nm, for the different substrate concentrations in assay buffer. The correction factor was calculated from the optical density data (OD), measured for a 5 mm pathlength cuvette, by the equation (Bagui *et al.*, 1996, Lakowicz, 1999):

$$Corr.\ fact. = 10\wedge \frac{OD_{325nm} + OD_{420nm}}{2}$$

2.2.8. Determination of physico-chemical parameters

2.2.8.1. Differential Scanning Calorimetry (DSC)

The VP-DSC high sensitivity scanning calorimeter (MicroCal, Inc.) was used to study the thermal stability of untagged LssEdom. Prior to the injection in the calorimeter cell, all the solutions were filtered and degassed. The reference cell was filled with an identical solution to that in the sample cell, but without the protein. The protein concentration was 12.7 μM in 10 mM HEPES (pH 7.0) with added sodium chloride, 150 mM. The up-scanning experiments were performed under an extra constant pressure of 2 atm and with a scan rate of 20ºC.h^{-1}. The thermograms were processed with Origin 7.0.

[**] http://www.erithacus.com/grafit

The calorimetric enthalpy was determined by integrating the area under the peak after subtraction of the buffer-buffer baseline and by integrating the area under the peak after adjusting the pre- and post-transition baselines.

2.2.8.2. Circular Dichroism (CD) spectroscopy

CD measurements were performed on an Applied Photophysics Pi-Star-180 spectrophotometer, employing constant N_2 flushing. The temperature was regulated using a Neslab RTE-300 circulating programmable water bath and a thermoelectric temperature controller. CD spectra were recorded at 25 or 85ºC using a 1 mm quartz cuvette. Protein samples were prepared at 26 µM concentration. The secondary structure was studied in 1 mM sodium phosphate (pH 7.0) buffer. Spectra were recorded from 210 to 260 nm, at 0.5 nm intervals with an integration time of 2 s, and are the averages of three scans, with the appropriate background buffer spectrum subtracted. CD measurements were converted into mean residue molar ellipticity [Θ] reported in deg.cm^2.dmol^{-1}, to reflect the fact that the peptide bond is the absorbing species in this case, using the formula:

$$[\Theta] = \Theta obs \div (10 \times l \times C \times n)$$

where Θobs is the observed ellipticity in millidegrees, l is the optical pathlength in centimetres, c is the molar protein concentration, and n is the number of peptide bonds in the protein.

Thermal denaturation curves were recorded over the temperature range 25-85ºC using a 26 µM protein solution in a 1 mm quartz cuvette. Single wavelength data were recorded at 224 nm over a single accumulation. The sample was required to reach thermal equilibrium at each temperature for a period of at least 30 s with a tolerance of ± 0.2ºC before recording each data

point. The temperature was increased every 4ºC and decreased every 6ºC. Ellipticity data were corrected to mean residue molar ellipticity using the above formula.

2.2.8.3. Attenuated Total Reflection (ATR) spectrometry

Mid-infrared (IR) spectroscopy is one of the classical methods for structure determination of molecules. The absorption of infrared radiation excites vibrational transitions of molecules. Vibrational frequency depends on the strength and polarity of the vibrating bonds, the masses of the vibrating atoms and is influenced by intra- and inter-molecular effects. The infrared spectrum is plotted against the inverse of the wavelength, the *wavenumber ũ*, which is proportional to the transition energy and has the unit cm^{-1}.

In Fourier transform infrared (FT-IR) spectrometers, the light emitted from the light source is split between two beams that generate a variable optical path difference; it gives a detector signal that contains the spectral information. The optical path difference generates an interference measured by the instrument, called an interferogram, which is the Fourier transform of the spectrum. A second Fourier transform performed by a computer converts the measured data back into a spectrum.

Attenuated total reflection (ATR) is a sampling technique used in conjunction with infrared spectroscopy which enables samples to be examined directly in the solid or liquid state. ATR uses the property of total internal reflection called the evanescent wave (Harrick, 1967). An attenuated total reflection accessory is illustrated Figure 21 (Barth, 2007). It operates by

measuring the changes that occur in a totally internally reflected infrared beam when the beam comes into contact with the sample.

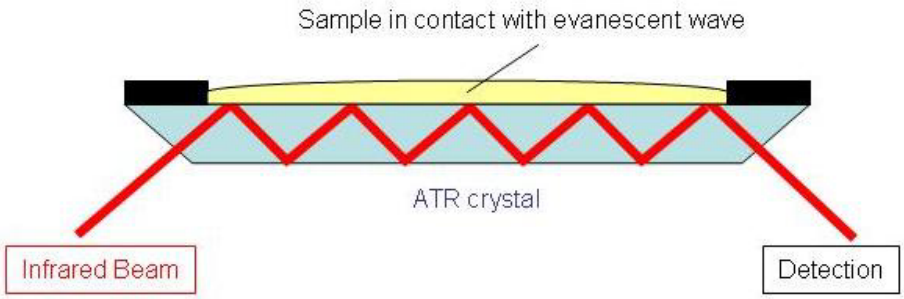


Figure 21: A typical ATR setup.

An infrared beam is directed onto an optically dense diamond crystal with a high reflective index at a certain angle. This internal reflectance creates an evanescent wave that extends beyond the surface of the crystal into the sample held in contact with the crystal. In regions of the infrared spectrum where the sample absorbs energy, the evanescent wave will be attenuated or altered. The attenuated energy from each evanescent wave is passed back to the IR beam, which then exits the opposite end of the crystal and is passed to the detector in the IR spectrometer. The system then generates an infrared spectrum.

The protein samples (100 μl) dialysed in 1 mM sodium phosphate (pH 7.0) buffer were deposited on the attenuated total reflection (ATR) crystal for FT-IR measurements. All IR spectra were collected at room temperature with an Equinox 55 Fourier transform infraRed (FT-IR) spectrometer (Bruker Optics, Inc., Billerica, MA) with a Bio-ATR II attachment (Bruker Optics, Inc.) equipped with a diamond ATR crystal. For each protein spectrum, 100 interferograms repeated 10 times were accumulated and Fourier transformed to yield infrared spectra with a nominal resolution of 2 cm^{-1}. ATR spectra were

automatically converted to absorption spectra using the Opus 4.2 software (Bruker Optics, Inc.). Solvent spectra obtained under the same conditions were subtracted from all sample spectra. Baseline correction was performed by the Opus 4.2 software. All IR spectra were normalised for easier comparison. The amide I region (1700-1600 cm^{-1}) was used for protein structure determination. The regions of the Tyr and His (1518 and 1602 cm^{-1} and 1594 cm^{-1}, respectively) were also used to monitor changes in protein conformation (Noguchi *et al.*, 1999, Venyaminov and Kalnin, 1990). Due to the composite character of the amide I band, overlapping infrared bands were resolved according to Fourier self-deconvolution procedures as described previously (Kauppinen *et al.*, 1981). The presence of certain protein secondary structure elements was determined based on the positions of the component bands (Byler and Susi, 1986, Goormaghtigh *et al.*, 1994). In order to quantify the approximate amount of each protein secondary structure, a curve fitting of deconvolved amide I band contour was performed with Gaussian function (Surewicz *et al.*, 1993). The resulting fractional areas of the bands assigned to different types of secondary structure (i.e., α-helix, β-sheet, turn and irregular structure) were assumed to represent the percentages of these structures in a given protein (Byler and Susi, 1986, Surewicz and Mantsch, 1988, Goormaghtigh *et al.*, 1990).

2.3. Crystallogenesis and attempts at structure resolution

2.3.1. Principles of protein crystallisation

Crystallisation is a phase transition phenomenon as illustrated in Figure 22 (Chayen, 2004). Nuclei form from an aqueous protein solution when the

solution is brought into supersaturation, which is achieved by varying certain parameters such as protein concentration, nature of precipitant and solvent, pH, temperature, and the presence of additives such as ligands and ions. Once nuclei have formed, which can be studied by electron microscopy, actual crystal growth can begin.

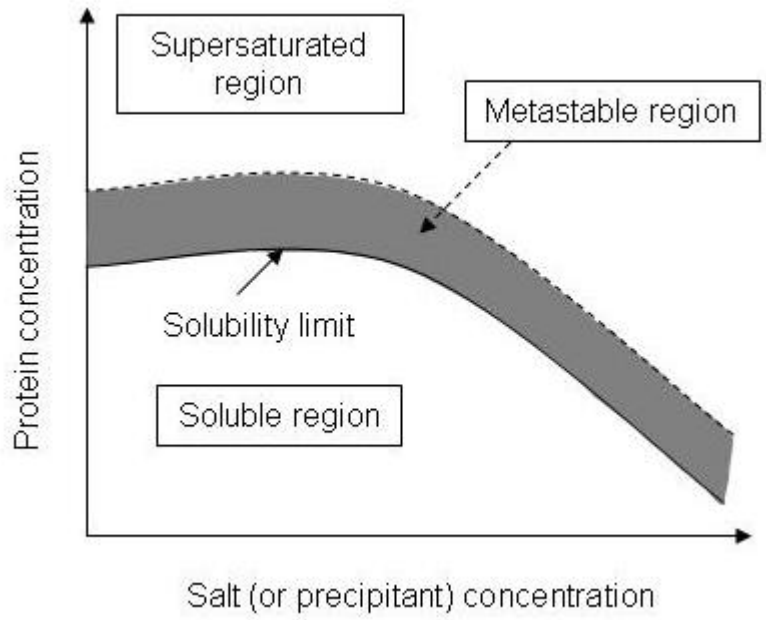


Figure 22: Phase diagram illustrating the effects of salt or other precipitant concentration on protein solubility.

In the supersaturated region of the phase diagram, nucleation is easy but haphazard, and subsequent crystal growth rapid and hard to control. However, in the "metastable" region, at concentrations slightly above the saturation limit, small nuclei are unstable but larger nuclei may grow at the expense of smaller ones.

The most popular method that has been developed to aid macromolecular crystal formation is that of vapour diffusion, which has two major variants: sitting and hanging drop, pictured in Figure 23. The small

drop, containing protein, stabilizing buffer, precipitant and other additives, is placed in vapour equilibrium in a closed system with a much larger reservoir volume. The reagent concentration in the reservoir is much higher than in the drop and therefore pulls water from the drop into a vapour phase until equilibrium between them is reached. During this process, the sample in the drop is concentrated, increasing its relative super saturation. Once nuclei have formed, the concentration of protein in the solute will drop, thereby leading the system into the metastable zone, where a few single crystals will grow.

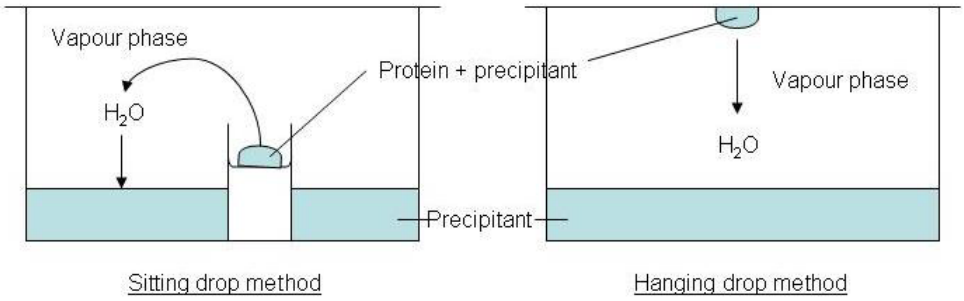


Figure 23: Method for protein crystallisation.

To obtain good structural data, crystals need to be single and have dimensions of at least 10 μm^3. Crystallisation experiments are based mainly on trial and error. The ability to dispense crystallisation trials by way of automating and miniaturising, with the HYDRA II robot dispenser (Matrixtechcorp), reduce the time needed to set up the series of experiments. Such procedures are very valuable for finding the initial conditions for crystallisation.

In addition, as seed-grown crystals avoid some of the problems inherent to spontaneously nucleated crystals, this method was tested. Indeed, it is said to be profitable to stimulate nucleation by seeding with crystals of the

same protein grown under other conditions. This seeding method is called epitaxial jump and is used to obtain crystals under conditions which may be significantly different from those in which the original seed was obtained. In our case, streak seeding shown on Figure 24 was utilised (Bergfors, 1999).

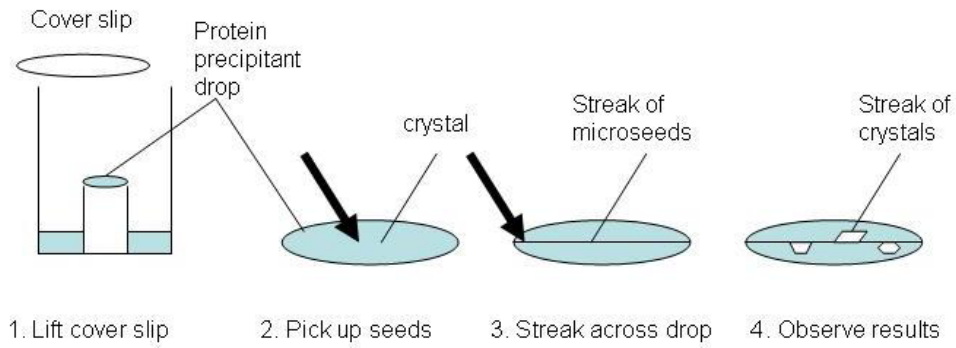


Figure 24: Streak seeding method.

The well is gently opened and the crystal seeds are picked up with a whisker. This is streaked across a pre-equilibrated drop which allows growth of well-formed crystals.

2.3.2. Crystal screening trials

Protein crystallisation studies on cleaved His-tag LssEdom were carried out in collaboration with Dr Max Paoli (Centre for Biomolecular Sciences, University of Nottingham, Nottingham) after unsuccessful trials on Lss and LssEdom domain *C*-terminally His-tagged.

High throughput crystallisation screens were carried out using the sitting-drop vapour-diffusion method in 96-well Intelliplates (Hampton Research) and the Hydra II micro dispensing system (Robertson Scientific). The crystallisation screens tested from NextalBiotech™ were: the Classics™, the $AmSO_4$™ (ammonium sulphate), the anions™, the

cations™, the PEGs™, the pHClear™, and the MPD™ (3-methyl-1,5-pentadiol). The volume of the reservoir solution was 60 µl and for the three drops its volume was 1.0 µl. The robot placed, in each well, different concentrations (27, 55, 66 µM) of the same preparation of LssEdom dialysed in buffer Tris-HCl, 10 mM, pH 7.0 with sodium chloride, 150 mM. It was placed on the corresponding reservoir solution droplet at a ratio of 1:1 (v/v). The plates were sealed carefully with tape (Henkel) and placed in the incubator at 19ºC. Every 5 days the droplets were examined under a microscope for the presence of crystals and the "Hit conditions" were marked.

Crystallisation conditions were then reproduced manually with the sitting drop vapour diffusion method in Libro plates (Hampton Research). Various ammonium sulphate concentrations (1.4 M to 2.4 M) were screened against different PEG 400 concentrations (1% to 11%) with 0.1 M MES sodium salt buffer pH 6.5. The drops were obtained by mixing 2.0 µl of protein at different concentrations (37, 100 and 219 µM) with 2.0 µl of solvent, with a reservoir volume of 1 ml.

As no reproducibility was observed, the effect of the buffer solution was tested. The same conditions were reproduced, with the exception of the replacement of MES sodium salt buffer by Tris-HCl buffer (0.1 M, pH 6.5). The salt concentration potential effect was investigated as well by applying the protein against the same conditions in which was added 100 mM or 200 mM sodium chloride. To test whether or not the lack of reproducibility was due to a problem in the preparation of the solutions used, NextalBiotech™ solutions of MES sodium salt pH 6.5, ammonium sulphate, PEG 400 were

used. The drops were obtained by mixing 1 μl of 60, 113 and 169 μM of the protein with 1 μl of the reservoir solution.

In parallel, seeding experiments were performed. A crystal wand was made by cutting a cat's whisker, threading it into a white pipette tip and fixing it with glue. Three pre-equilibrated drops were prepared with 1.70 M ammonium sulphate, 3% PEG 400 and 0.1 M MES sodium salt pH 6.5, mixed 2 μl:2 μl with 37, 100 and 219 μM of LssEdom. The seed source well containing needles grown under the same conditions was opened gently. Seeds were picked up with the whisker and the crystal growth wells were streaked with the whisker across the pre-equilibrated drop. In addition, the seed source well, containing crystals grown under the following conditions: 2.0 M ammonium sulphate, 5% PEG 400 and 0.1 M MES sodium salt pH 6.5, was opened gently for mounting experiments and then used for streak seeding experiments. Seeds were picked up with the whisker. The crystal growth wells were streaked with the whisker across the pre-equilibrated drop.

2.3.3. Crystal cryo-mounting

The crystal cryo-mounting experiment was carried out in collaboration with Dr Max Paoli (Centre for Biomolecular Sciences, University of Nottingham, Nottingham). The crystals were suspended in a film of an "antifreeze" solution (a "cryo-protected" solution forming a frozen vitrified water in which the crystal is stable) held by surface tension across a small diameter loop of fiber (Hampton Research) and quickly plunged into a liquid cryogen.

Different “cryo-protectant” solutions were tested:

- 100 mM MES pH 6.5, 150 mM NaCl, 10 mM Tris-HCl pH 6.5, 2.08 M ammonium sulphate, 6% (w/v) PEG 400;
- 100 mM MES pH 6.5, 2.08 M ammonium sulphate, 6% (w/v) PEG 400, 320 mM sodium malonate;
- 100 mM Tris-HCl pH 6.5, 2.08 M ammonium sulphate, 6% (w/v) PEG 400, 740 mM sodium malonate;

When observed under the microscope, those solutions showed that there was a problem of phase separation. Thus, the amount of PEG 400 was lowered and a loop was dipped with a “cryo-protectant” solution made of 100 mM MES pH 6.5, 2.08 M ammonium sulphate, 2.5% (w/v) PEG 400, 0.96 M sodium malonate. It was flash-frozen and placed on the gionometer. The solution was glassy and was used to immerse the crystal for 30 seconds prior to being speedily flash-frozen. Two crystals were mounted.

2.3.4. Nuclear Magnetic Resonance (NMR) spectroscopy preliminary studies

All NMR spectra were collected by Dr J. Long (Centre for Biomolecular Sciences, University of Nottingham, U.K.) at 298 K on a Brüker AV-600 spectrometer operating at 600 MHz. The spectrometer was equipped with a triple-resonance probe using the standard library of pulse sequences available with XWINNMR 3.51 (Bruker). The LssEdom sample (uniformly ^{15}N-labelled) was buffered in 10% D_2O solution 10 mM sodium phosphate (pH 7.4), 150 mM sodium chloride, 0.04% sodium azide, to give a final protein concentration of 75 μM. The 2D ^{15}N-HSQC spectrum was collected over-night using XWINNMR version 3.51.

2.3.5. Computational techniques

2.4. Hierarchical clustering

Multiple alignments with hierarchical clustering were performed using Multalin[††]. Multalin created a multiple sequence alignment from a group of related sequences using progressive pairwise alignments. The method used is described by Corpet (Corpet, 1988). Protein sequences (Fasta format) were rearranged with postulated gaps so that similar residues were juxtaposed. A positive score was attached to identities and a penalty to gaps was subtracted from the alignment score of two clusters each time a new gap was inserted in one cluster. The program maximised the total score, taking account of all possible alignments and allowing for any length gap at any position.

2.5. General homology procedure

The amino acid sequence of the target protein, lysostaphin, was obtained from GenBank (Accession AAB53783) in which 239 residues were involved. The template proteins used were LytM (PDB code 2B44) (Odintsov *et al.*, 2004, Firczuk *et al.*, 2005), and ALE-1 (PDB code 1R77) (Lu *et al.*, 2006), deposited in Protein Data Bank.

The BLAST search algorithm was used for the search online[‡‡] (Altschul *et al.*, 1997). Program Modeller 8v2 (Sali and Blundell, 1993), which is an implementation of an automated approach to comparative modelling by satisfaction of spatial restraints, was employed to build the 3D structure of lysostaphin. All sequences were imported in ClustalW program[§§] (Thompson *et al.*, 1994). The ribbon illustration was made using PyMol

[††] http://bioinfo.genopole-toulouse.prd.fr/multalin/multalin.html
[‡‡] http://www.ncbi.nlm.nih.gov
[§§] http://www.ebi.ac.uk/clustalw/

software. The structure with the lowest energy was checked using MolProbity software (Davis *et al.*, 2007). SYBYL 7.3 software[§] was used to calculate the root mean square deviation (RMSD) comparing backbone structures with each others.

[§] http://www.tripos.com/data/SYBYL

3. Lysostaphin secondary structure prediction

3.1. Secondary structure

Proteins are amino acid-based compounds and the composition of amino acids in a protein gives it its unique structure. The backbone geometry of a protein is defined by the dihedral angles Phi (Φ) and Psi (Ψ) angles of individual residues (Figure 25). Certain combinations of these angles in consecutive residues result in a specific secondary structure motif (Karplus, 1996). These structures are stabilised by intermolecular hydrogen bonds.

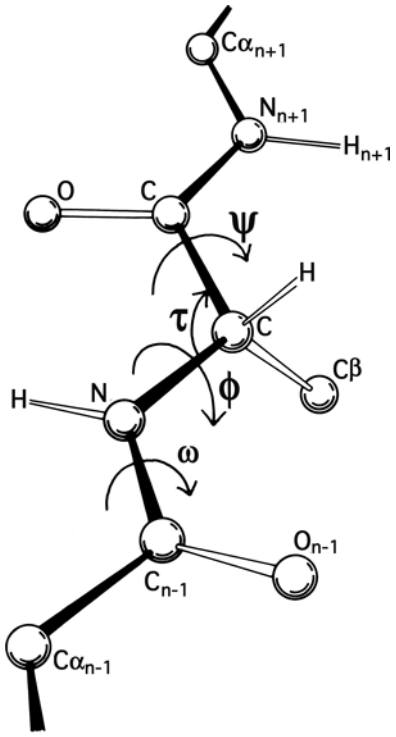


Figure 25: Peptide torsion angle. Definition of the Phi (Φ) and Psi (Ψ) angles*.

*http://cmgm.stanford.edu/biochem201

Protein chains form frequently observed structural motifs in their native state. The most frequently observed secondary structure elements are α-helices and parallel and anti-parallel β-sheets. α-Helices (Figure 26) are helix-like structures and β-sheets (Figure 27) are pleated sheet-like structures connected by β-turns (in the case of anti-parallel β-sheets), as their names suggest. Every region remaining in the protein after all the α-helix, β-sheet and β-turn regions are assigned is called a loop region.

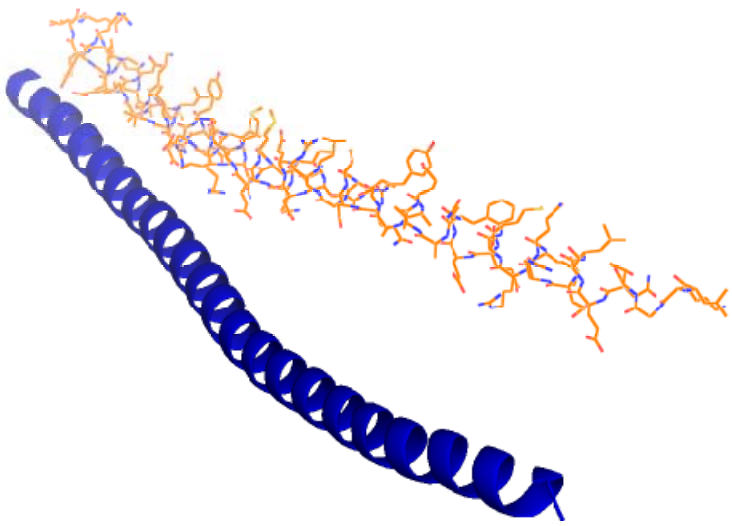


Figure 26: The right-handed α-helical secondary structure of human vimentin coil 2B fragment (PDB code 1GK4).

The right-handed α-helix results when the nth peptide unit forms hydrogen bonds between its C-O and the N-H of the $(n+4)^{th}$ peptide unit and between its N-H and the $(n-4)^{th}$ C-O; there is a 1.5 Å translation and 100º turn between two consecutive peptide units, giving 3.6 amino acid residues per turn.

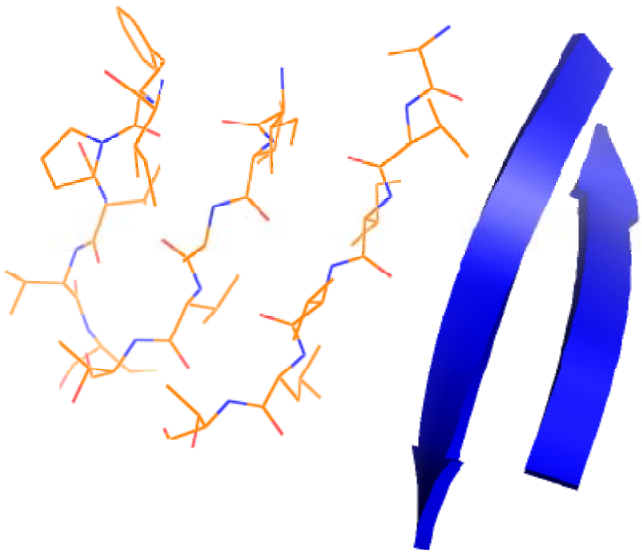


Figure 27: Anti-parallel β-sheet secondary structure from pyruvate kinase (PDB code 1PKN).

Typically the strands of an anti-parallel β-sheet are linked by β-turns where the nth peptide unit forms hydrogen bonds with the $(n+3)^{rd}$ peptide unit.

3.2. Prediction strategies

Determination of protein secondary structure is not an exact science. There are a number of methods of varying resolution to determine the structure of proteins. There are both laboratory based methods and theory based methods for the determination of secondary structure. Laboratory based methods can be used to guess the secondary structure content without knowing the tertiary structure. The percentage of common motifs in a protein can be determined by means of circular dichroism spectroscopy (CD) (Fasman, 1996) or Fourier transform infrared spectroscopy (FT-IR) (Byler and Susi, 1986). In another method, the scientist first determines the tertiary structure of the protein using a method such as X-ray crystallography (Rupp and Wang, 2004) and then specifies the regions of different secondary structure by careful

inspection of the structure. If the structure of a protein is known, the secondary structure determination is a pattern recognition problem using algorithms developed for this purpose.

3.3. Protein expression, purification and quality assessment

LssEdom was expressed and purified to near homogeneity by means of immobilised metal-affinity chromatography (IMAC) as described in the Materials and Methods section (cf. sections 2.1.8. and 2.2.1.). In order to obtain the structure of the active form of LssEdom it was necessary to assess its activity. A biological substrate, MV11, containing a pentaglycine cleavage site was used to primarily observe the cleavage activity of LssEdom. In addition, a quantitative measurement to compare the activity of different Lss preparations (zinc- or nickel- IMAC purified) was performed by Fluorescence Resonance Energy Transfer (FRET) assay.

The recombinant proteins were found to be active on the MV11 biological substrate, cleaving MV11 at the Gly_5 sequence, as determined by *in vitro* activity assay and observed for LssEdom on SDS-PAGE (Figure 28).

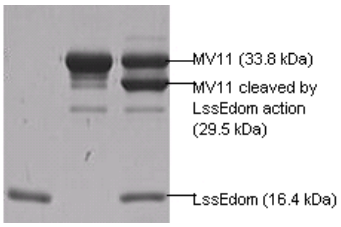


Figure 28: LssEdom activity test on MV11 substrate observed on SDS-PAGE.

Quantitative measurement of Lss activity was performed by (FRET) assay. An internally quenched pentaglycine with an *ortho*-aminobenzoic acid

fluorophore (Abz) and a (2,4-dinitrophenyl)ethylene diamine (EDDnp) quenching group was used as a substrate (Figure 29) (Warfield *et al.*, 2006).

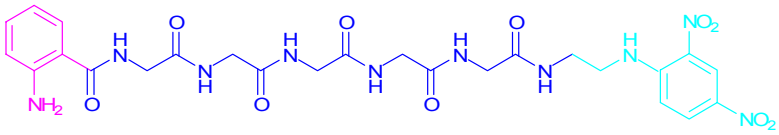


Figure 29: FRET substrate: internally quenched pentaglycine with *ortho*-aminobenzoic acid fluorophore (Abz) and a (2,4-dinitrophenyl)ethylene diamine (EDDnp) quenching group.

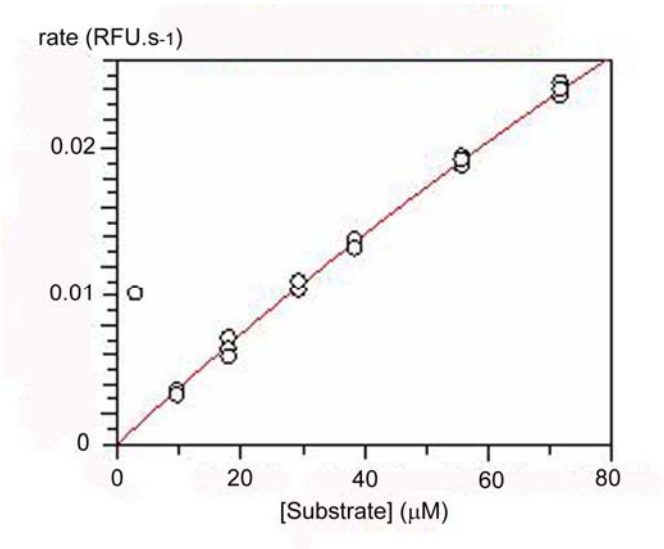
The samples of lysostaphin used in the assays up to this point had been purified using a nickel(II)-affinity chromatography column, charged with 50 mM $NiSO_4$. It was suggested that this might lead to replacement of the zinc(II) ion by a nickel(II) ion in the active site of the metallo-enzyme, and hence lead to a decrease in enzyme activity. This theory was tested by comparing activities obtained after nickel(II) or zinc(II)-affinity chromatography purification of lysostaphin samples.

Titration of the Abz-Gly_5-EDDnp FRET substrate with lysostaphin resulted in a concentration dependent cleavage of the pentaglycine and quantitative emission of the fluorescence. At high substrate concentrations, initial rates appeared to decrease due to the inner filter effect. In order to overcome this effect a correction factor was calculated as described in the Materials and Methods section (cf. section 2.2.7.4.) and applied for each concentration of substrate. Three assays were performed at each substrate concentration and the mean value of the reaction rates was determined for each set. The corrected rates could be fitted to the following Michaelis-Menten equation:

$$V_0 = V_{max} [S_0] / ([S_0] + K_M)$$

where: V_0 = intial rate
V_{max} = maximum rate possible at a given enzyme concentration
$[S_0]$ = initial substrate concentration
K_M = Michaelis-Menten constant

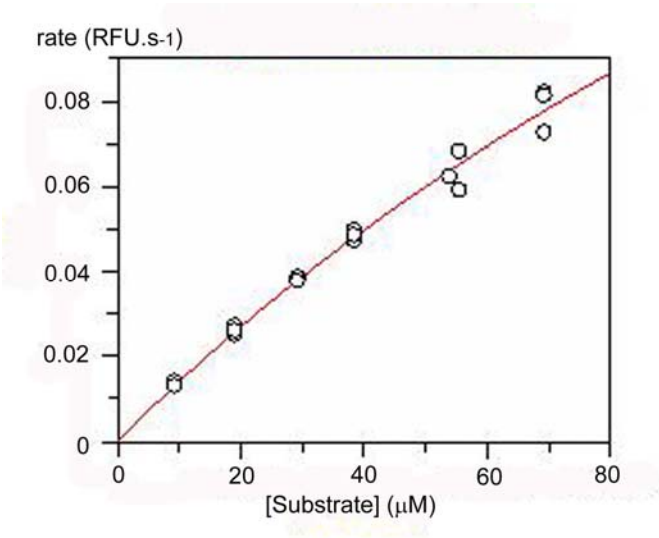
The Grafit©*** data-fitting program was used to calculate K_M and V_{max}. Figure 30 displays the Michaelis-Menten plot from the lysostaphin purified on nickel(II) results and Figure 31 displays the Michaelis-Menten plot from the lysostaphin purified on zinc(II) results.



nickel-purified Lss	Value	Std. Error
Vmax (RFU.s^{-1})	0.259	0.063
K_M (mM)	0.7	0.2

Figure 30: Grafit© Michaelis-Menten plot for data collected using nickel-purified lysostaphin.

*** http://www.erithacus.com/grafit



zinc-purified Lss	Value	Std. Error
Vmax ($RFU.s^{-1}$)	0.330	0.064
K_M (mM)	0.2	0.05

Figure 31: Grafit© Michaelis-Menten plot for data collected using zinc-purified lysostaphin.

Grafit© estimated the values of V_{max} to be 0.26 (± 0.06) $RFU.s^{-1}$ for Lss purified on nickel(II) and 0.33 (± 0.06) $RFU.s^{-1}$ for Lss purified on zinc(II). Estimated values for K_M were 0.7 (± 0.2) mM and 0.2 (± 0.05) mM respectively for Lss purified on nickel(II) and on zinc(II). The values calculated for the V_{max} and K_M are only approximate as the usable substrate concentration range is not sufficiently wide enough to adequately cover the range required to obtain accurate kinetic constants with small errors associated with them.

At a substrate concentration of 10 μM, the relative rates were calculated to be 0.0179 (± 0.0002) $RFU.s^{-1}$ for Lss purified on zinc(II) and 0.0051 (± 0.0002) $RFU.s^{-1}$ for Lss purified on nickel(II). Thus, the observed cleavage ability was greater than 3.5 fold for the zinc-purified sample. These results indicated that protein activity was enhanced when purification was performed with zinc(II) rather than nickel(II). Indeed Lss is a zinc

metalloprotease that contains a zinc ion in its activity site which is important for the cleavage reaction.

3.4. Circular Dichroism (CD) spectroscopy

Circular dichroism (CD) is defined as the differential absorbance of left circularly polarized light (LCPL) and right circularly polarized light (RCPL): CD = Abs (LCPL) – Abs (RCPL). To be "CD optically active", a molecule must interact differently with left- and right-handed photons. Asymmetry can come from chiral molecules such as the peptide backbone of proteins or their aromatic amino acid side chains (all amino acids except glycine contain a chiral carbon, thus are asymmetrical). In the far-UV region, the peptide bond is the major absorbing group.

Five common secondary structure motifs (α-helix, parallel β-sheet, antiparallel β-sheet, β-turns, and other (e.g. random coil conformation)) exhibit distinctive CD spectra in the far-ultraviolet region (170-260 nm) (Johnson and Tinoco, 1972). Based on these and the spectra of standard proteins, there are many algorithms currently available for protein secondary structure analysis (Compton and Johnson, 1986, Johnson, 1999). These algorithms use a basic set of model protein structures. The shape of the spectrum is compared to the basis spectra and five structural contributions are extracted (α-helix, parallel β-sheet, antiparallel β-sheet, β-turns, and other (e.g. random coil conformation)). However, it has limited value due to over-simplification.

CD is reported in units of absorbance or ellipticity. CD data are typically reported in ellipticity (θ), which is related to absorbance by a factor of $32.98\theta = 33.98\Delta Abs$. Ellipticity is usually reported in millidegrees (mdeg), which are one thousandth of a degree. Molar ellipticity ([θ]) is CD corrected

for concentration. Conversion from molar extinction (absorbance corrected for concentration) to molar ellipticity uses a factor 3298 ([θ] = 3298Δε).

3.4.1. Exciton coupling phenomenon

CD spectra of active LssEdom were obtained in order to predict its secondary structure to assess if it was a suitable protein for crystallisation studies by determining amount of random coil. Shown in Figure 32 is the LssEdom CD spectra recorded at 25ºC and then 85ºC and again 25ºC after denaturation at 85ºC. Activities of LssEdom before and after denaturation were assessed by FRET assay (cf. Appendix 1). The thermal curves on denaturation and attempt at renaturation for LssEdom are provided in section 4.3.2. CD data have been corrected for concentration and are reported as the mean residue molar ellipticity. All solutions of protein used exhibited similar spectra with maxima near 224 nm.

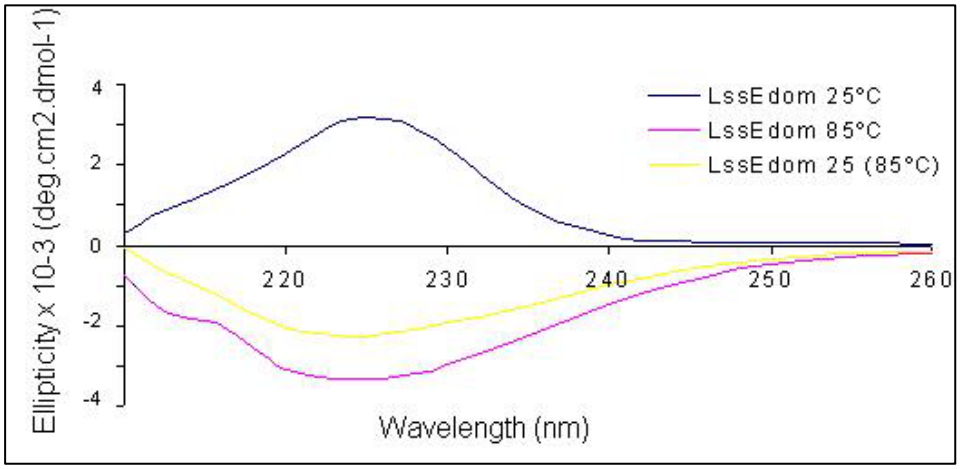


Figure 32: Far-UV CD spectra of LssEdom recorded at 25ºC (blue), 85ºC (pink) and 25ºC after denaturation (orange). The protein concentration was 0.30 mg.ml^{-1}. The buffer employed was 1 mM sodium phosphate, pH 7.0.

LssEdom is 48% identical to $LytM_{186-314}$. Thus, the predicted amount of secondary structure elements for LssEdom should be in agreement with $LytM_{186-314}$ (PDB codes 2B44, 2B0P, 2B13) content, knowing that a given sequence of amino acids always folds into almost the same structure under the same environmental conditions. The main part of the secondary structure of $LytM_{186-314}$ is dominated by a β-sheet (42%) that extends throughout the entire molecule; some α-helical conformation (13%) is also present in the structure (Firczuk *et al.*, 2005). The CD spectra of β-proteins having similar secondary structures and little α-helix vary widely (Manning and Woody, 1987), which complicates estimations of the amounts of the various types of secondary structure. Contributions from aromatic side chains, tryptophan residues in particular, are considered to be an important source of these variations in the far-UV region. LssEdom contains a large fraction of tryptophan residues (4 Trp, 2.7%). An average protein has a tryptophan content of 1.7% (Singh and Thornton, 1992), which suggests that the tryptophan residues in LssEdom may make important contributions to the far-UV CD spectra recorded. Some of the tryptophan residues are located in aromatic clusters as shown on the LssEdom model (cf. section 4.6.3) displayed in Figure 33.

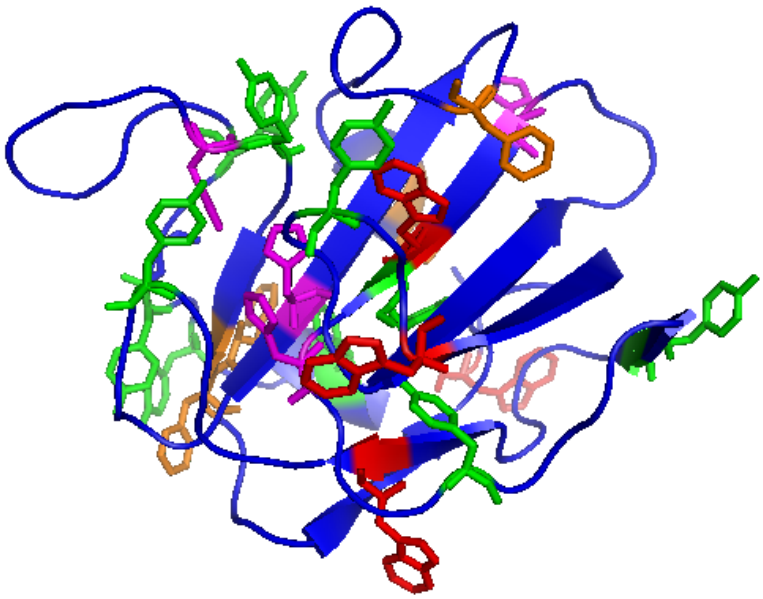


Figure 33: Polypeptide backbone of LssEdom including the side chains of aromatic residues. Tryptophan are shown in red, tyrosine in green, phenylalanine in orange and histidine in pink.

Several proteins have strong positive CD bands in the 225-230 nm region, probably attributable to aromatic side chains. For β-proteins possessing a CD spectrum of low intensity, tryptophan residue CD bands can markedly influence the far-UV region, especially in proteins containing several tryptophan residues (Grishina and Woody, 1994). The complexity of the far-UV spectrum most probably arises from exciton splitting of transitions from tryptophans. If two identical chromophores are sufficiently close, their excited states will interact. This phenomenon is referred to as exciton effect (Davydov, 1971). The CD bands overlap, giving rise to a characteristic band called couplet (Schellman, 1968). The exciton coupling observed probably arises due to the highly asymmetrical environment of the aromatic residues in LssEdom. Since exciton coupling between aromatic groups increases when they are in close proximity, perturbation of the CD spectra thwarts attempts to estimate

the amounts of the various types of secondary structure in LssEdom. These problems might also be emphasized in this protein because of its relatively low contents of α-helix. Together these characteristics make this protein a good model for studying contributions to the CD spectrum made by structural elements other than secondary structure.

The complex band pattern observed in the far-UV CD spectrum of the native protein has changed in the corresponding spectrum of denatured LssEdom recorded at 85ºC and 25ºC after denaturation. This indicates that the asymmetric environment around the various aromatic amino acid residues in the native state was disrupted upon unfolding, resulting in a CD spectrum typical for an unordered polypeptide chain.

3.4.2. Strategy for mutations

In order to study the contribution of tryptophan residues to the CD spectrum and to predict the secondary structure of LssEdom, mutations were performed on residues that were predicted to contribute to the exciton coupling phenomenon observed. Of the four tryptophan and ten tyrosine residues in LssEdom sequence, only tryptophan residues were considered, as they were thought to make the major contribution to the exciton coupling phenomenon (Grishina and Woody, 1994). In the model of LssEdom structure (cf. section 4.6.3.), tryptophan residues were highlighted to evaluate their distances and respective orientations (Figure 34). The side chains of tryptophan at position 59 and 80 in the LssEdom model were in the most parallel orientation, with one ring rotated by less than 90º relative to the other and within the shortest distance (13 Å) of the all four tryptophan residues. Tryptophan 80 is located on a β-sheet buried in the interior of the enzyme molecule and is close to the

essential amino acid histidine 83 (Warfield *et al.*, 2006). It is surrounded by tyrosine 34 and 62, histidine 83 and 116 and phenylalanine 124. In this case, coupling of tryptophan transitions is not necessarily dominant over Tyr-Trp, Tyr-Tyr, Phe-Trp, etc. transitions. Tryptophan 59 is located in a β-turn and is exposed to the solvent. It is unknown which residues are responsible for such spectroscopic properties and to what extent individual tryptophan residues affect the tertiary structure and function of LssEdom, because no mutation studies have been performed on any of the tryptophan residues.

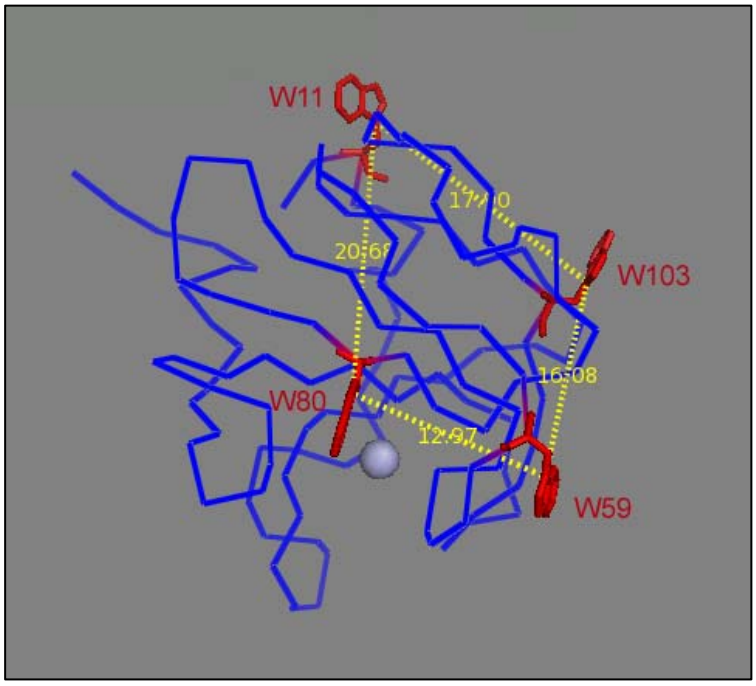


Figure 34: Model of LssEdom structure (cf. Section 4.6.3.). Ribbon representation with **Trp** shown as sticks and highlighted in red. Relative **distances** between Trp residues are shown in yellow: 13.0, 16.0, 17.4, 20.7 Å for W80-W59, W59-W103, W103-W11, W11-W80 distances, respectively.

Looking at substitution preferences in the literature (Betts and Russell, 2003, Livingstone and Barton, 1993), it was reported that replacing tryptophan by phenylalanine by site directed mutagenesis was the most favoured substitution in order to minimise the destabilisation of the native state as a consequence of the substitution. Based on the above information, tryptophan

residues W80, on one hand, and W80 with W59, on the other hand, were chosen to be mutated into phenylalanine residues by site-directed mutagenesis to lower the exciton coupling phenomenon observed by circular dichroism and thus allow secondary structure prediction of LssEdom.

3.4.3. LssEdom mutants

Ms W. Chen, a final year project student under my supervision, constructed a W80F mutant and a W59/80F mutant of LssEdom. The W80F substitution in LssEdom was achieved by site-directed mutagenesis in pETedom. Two megaprimers were generated by PCR with primers T7F and WC2 for the first reaction, and T7R and WC1 for the second reaction (sequences for each of the primers are displayed in the Material and Methods, section 2.1.5.). The two megaprimers generated were used to polymerise the complete DNA sequence with the desired mutation. The products of amplification are shown in Figure 35.

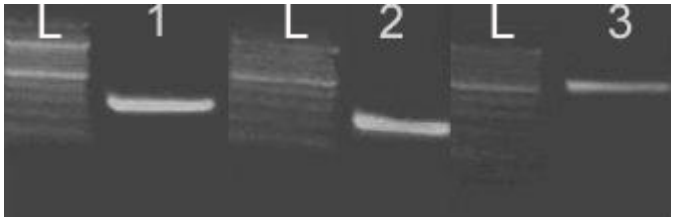


Figure 35: PCR products: (1) megaprimer generated with primers T7F and WC2 on pETedom, **(2)** megaprimer generated with primers T7R and WC1 on pETedom, **(3)** LssEdom DNA sequence with W80F substitution generated with megaprimer 1 and 2.

Cutting the pETedom DNA template with NdeI and XhoI resulted in the excision of a 600 bp insert which was replaced by the above PCR product purified and digested with the same enzymes. The clones were confirmed by PCR on the recombinant plasmid pWC3 and by DNA sequencing; using T7F (forward) and T7R (reverse) primers. DNA sequencing data returned from the QMC DNA sequencing laboratory (E1T7F) were translated and compared to

the LssEdom sequence with the ClustalW program. The results confirmed the presence of W80F substitution in the pWC3 plasmid (Figure 36).

Translation of pWC3 sequencing data compared with LssEdom sequence

```
E1T7F       MAATHEHSAQWLNNYKKGYGYGPYPLGINGGMHYGVDFFMNIGTPVKAISSGKIVEAGWS 60
Edom        MAATHEHSAQWLNNYKKGYGYGPYPLGINGGMHYGVDFFMNIGTPVKAISSGKIVEAGWS 60
            ************************************************************

E1T7F       NYGGGNQIGLIENDGVHRQFYMHLSKYNVKVGDYVKAGQIIGWSGSTGYSTAPHLHFQRM 120
Edom        NYGGGNQIGLIENDGVHRQWYMHLSKYNVKVGDYVKAGQIIGWSGSTGYSTAPHLHFQRM 120
            ******************* ****************************************

E1T7F       NSFSNSTAQDPMPFLKSA 138
Edom        NSFSNSTAQDPMPFLKSA 138
           ******************
```

Figure 36: ClustalW homology analysis showing sequence alignment of LssEdom and E1T7F sequence, identified by nucleotide sequencing and translated. Stars indicate a perfect match in between the theoretical sequence and the sequencing data.

The W59/80F substitution in LssEdom was achieved by site-directed mutagenesis in pWC3. Two megaprimers were generated by PCR with primers T7F and WC4 for the first reaction, and T7R and WC3 for the second reaction. The two megaprimers generated were used to polymerise the complete DNA sequence with the desired mutation. The products of amplification are shown in Figure 37.

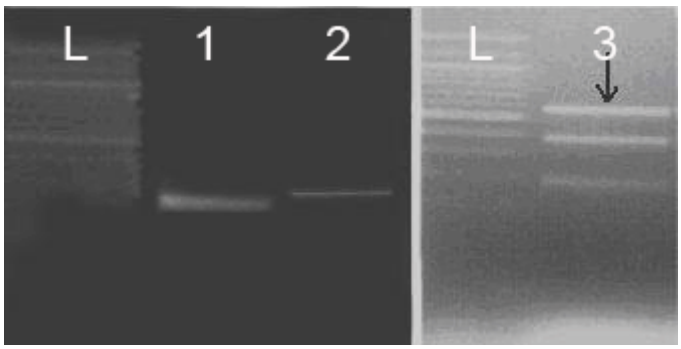


Figure 37: PCR products: (1) megaprimer generated with primers T7F and WC4 on pWC3, **(2)** megaprimer generated with primers T7R and WC3 on pWC3, **(3)** LssEdom DNA sequence with W59/80F substitutions generated with megaprimer 1 and 2.

Cutting pWC3 with the restriction enzymes NdeI and XhoI resulted in the excision of a 600 bp insert which was replaced by the above PCR product

purified and digested with the same enzymes. The clones were confirmed by PCR on the recombinant plasmid pWC5 and by DNA sequencing; using T7F (forward) and T7R (reverse) primers. DNA sequencing data returned from the QMC DNA sequencing laboratory (E2T7F) were translated and compared to the LssEdom sequence with the ClustalW program. The results confirmed the presence of W59/80F substitution in the pWC5 plasmid (Figure 38).

Translation of pWC3 sequencing data compared with LssEdom sequence

```
E1T7F       MAATHEHSAQWLNNYKKGYGYGPYPLGINGGMHYGVDFFMNIGTPVKAISSGKIVEAGFS 60
Edom        MAATHEHSAQWLNNYKKGYGYGPYPLGINGGMHYGVDFFMNIGTPVKAISSGKIVEAGWS 60
            ************************************************************

E1T7F       NYGGGNQIGLIENDGVHRQFYMHLSKYNVKVGDYVKAGQIIGWSGSTGYSTAPHLHFQRM 120
Edom        NYGGGNQIGLIENDGVHRQWYMHLSKYNVKVGDYVKAGQIIGWSGSTGYSTAPHLHFQRM 120
            ******************* ****************************************

E1T7F       NSFSNSTAQDPMPFLKSA 138
Edom        NSFSNSTAQDPMPFLKSA 138
            ******************
```

Figure 38: ClustalW homology analysis showing sequence alignment of LssEdom and E1T7F sequence, identified by nucleotide sequencing and translated. Stars indicate a perfect match in between the theoretical sequence and the sequencing data.

The recombinant plasmids: pWC3 or pWC5, were transformed into the host, *E.coli* BL21 (DE3), and addition of isopropyl β-thiogalactoside (IPTG) induced the overexpression of bands of approximately 16 kDa molecular weight, corresponding to the LssEdomW80F-H6 and LssEdomW59/80F-H6 fusion protein. The proteins were overexpressed in soluble form in the cytoplasm.

LssEdom, LssEdomW80F and LssEdomW59/80F were purified to near homogeneity, observed on SDS-PAGE, using twice the same affinity chromatographic step on zinc agarose (Figure 39).

Figure 39: SDS-PAGE showing the fractions containing LssEdom (1), LssEdomW80F (2) and LssEdomW59/80F (3) purified by metal affinity chromatography.

Protein size studies on LssEdom, LssEdomW80F and LssEdomW59/80F (C = 13 µM, [protein] = 0.2 mg.ml^{-1}) were conducted by dynamic light scattering on a Malvern Nanosizer (Figures 40 and 41).

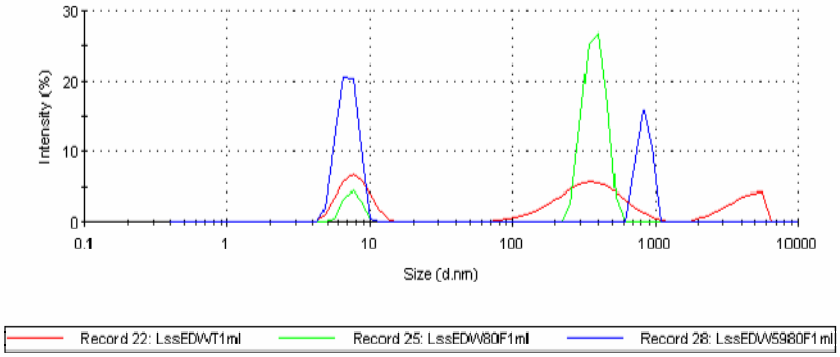


Figure 40: Dynamic light scattering size distribution by intensity of LssEdom (red), LssEdomW80F (green) and LssEdomW59/80F (blue).

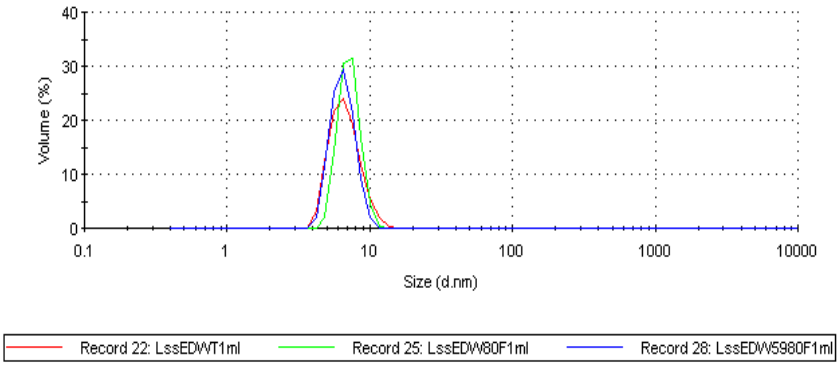


Figure 41: Dynamic light scattering size distribution by volume of LssEdom (red), LssEdomW80F (green) and LssEdomW59/80F (blue).

All samples showed the same size distribution. Peaks were centred on 6.53 nm. The Z average was 659 nm, 722 nm, 616 nm in diameter for LssEdom, LssEdomW80F and LssEdomW59/80F, respectively. The substitutions performed on LssEdom tryptophan residues 80 and 59 did not lead to an increased state of aggregation, which generally occurs when proteins are unfolded.

3.4.4. Kinetic data

In principle, substitution of a specific tryptophan residue might influence not only the point of mutation but also other parts of the protein structure, and if this is the case, such substitution might also alter the native environment of the remaining tryptophan residues. In order to determine if the mutated lysostaphin endopeptidase domains (LssEdomW80F and LssEdomW59/80F) were correctly folded and maintained the structure of the active form of LssEdom, it was necessary to assess their activities. A biological FRET substrate, MV11, containing a pentaglycine cleavage site was used to measure quantitatively the cleavage activities of LssEdom, LssEdomW80F and LssEdomW59/80F.

The mutation of tryptophan (Trp, W) into phenylalanine (Phe, F) at position 80 in the LssEdom sequence was believed to lead to a certain loss of activity as it was just three residues from histidine 83, an essential residue for lysostaphin activity (Warfield *et al.*, 2006). This theory was tested by comparing the relative activities of LssEdom and its mutants LssEdomW80F and LssEdomW59/80F.

The recombinant proteins were found to cleave the MV11 biological substrate at the Gly_5 sequence, as determined by *in vitro* activity assay and observed on SDS-PAGE (Figure 42).

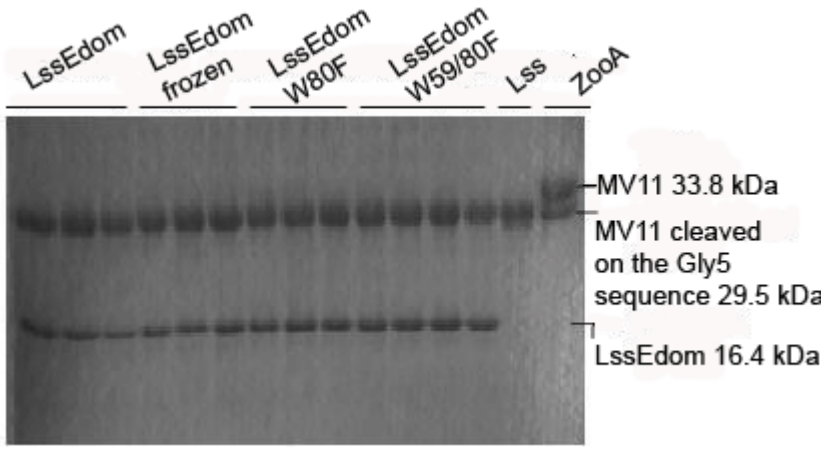


Figure 42: LssEdom, LssEdomW80F , LssEdomW59/80F and Lss activity test on MV11 substrate observed on SDS-PAGE. Negative control showing MV11 partly cleaved by zoocin A.

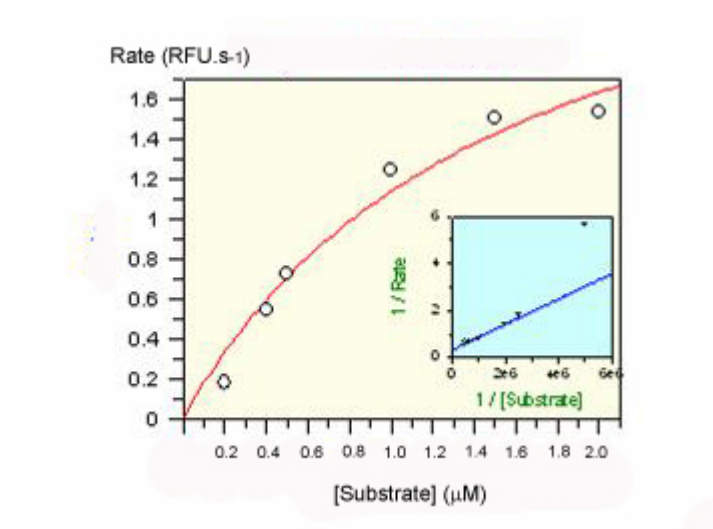
Titration of the MV11 FRET substrate with LssEdom, LssEdomW80F and LssEdomW59/80F resulted in a concentration dependent cleavage of the pentaglycine and quantitative emission of the fluorescence. Three assays were performed at each substrate concentration and the mean value of the reaction rates was determined for each set (Table 5).

Table 5: Relative rates of cleavage for different concentration of MV11 for LssEdom, LssEdomW80F and LssEdomW59/80F.

[MV11] (µM)	Rate ($RFU.s^{-1}$) LssEdom	Rate ($RFU.s^{-1}$) LssEdomW80F	Rate ($RFU.s^{-1}$) LssEdomW59/80F
0.2	0.18	0.06	/
0.4	0.55	0.28	0.11
0.5	0.72	0.33	0.12
1.0	1.24	/	0.16
1.5	1.50	/	0.18
2.0	1.54	0.55	/

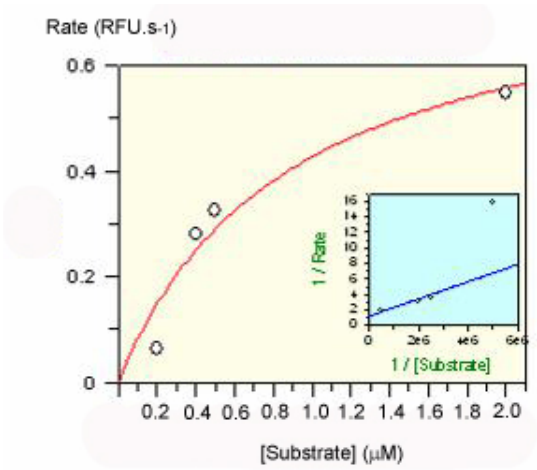
The rates obtained could be fitted to the Michaelis-Menten equation (cf. section 3.3.). The Grafit© data-fitting program was used to calculate K_M

and V_{max}. The Michaelis-Menten plot for LssEdom is shown in Figure 43, the plot for LssEdomW80F is shown Figure 44 and the plot for LssEdomW59/80F is shown Figure 45.



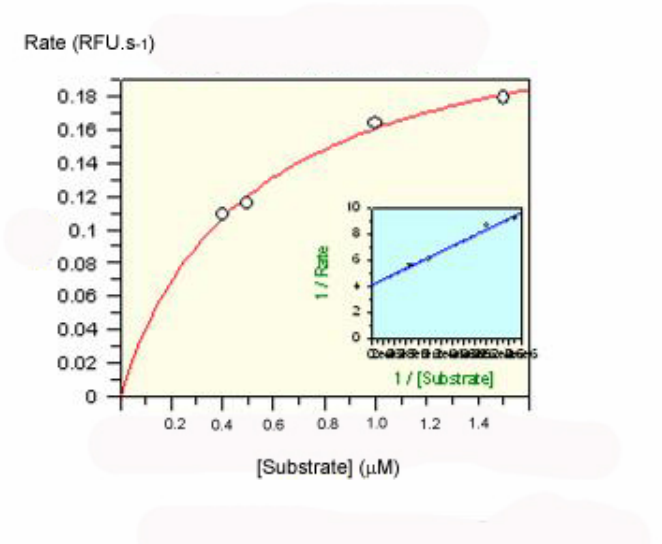
LssEdom	Value	Std. Error
V_{max} (RFU.s-1)	2.9	0.6
K_M (μM)	1.6	0.6

Figure 43: Grafit© Michaelis-Menten plot for data collected using LssEdom.



LssEdomW80F	Value	Std. Error
Vmax (RFU.s-1)	0.8	0.2
K_M (μM)	0.9	0.5

Figure 44: Grafit© Michaelis-Menten plot for data collected with LssEdomW80F.



LssEdomW80F	Value	Std. Error
Vmax (RFU.s-1)	0.24	0.01
K_M (mM)	0.51	0.07

Figure 45: Grafit© Michaelis-Menten plot for data collected with LssEdomW59/80F.

Grafit© estimated the values of V_{max} to be 2.9 (± 0.6) RFU.s^{-1} for LssEdom, 0.8 (± 0.2) RFU.s^{-1} for LssEdomW80F and 0.24 (± 0.01) RFU.s^{-1} for LssEdomW59/80F. Estimated values for K_M were 1.6 (± 0.6) µM, 0.9 (± 0.5) µM and 0.51 (± 0.07) µM respectively for LssEdom, LssEdomW80F and LssEdomW59/80F. The values calculated for the V_{max} and K_M are only approximate as the usable substrate concentration range is not sufficiently wide enough to adequately cover the range required to obtain valid kinetic constants with small errors associated with them.

These results indicated that the mutations performed on LssEdom were affecting the protein activity. Among the mutants, W80F has the highest activity, about 28% of that of the wild-type, and W59/80F has the lowest activity, about 8% of that of the wild-type. Fulfilment of the activity criterion by LssEdomW80F mutant was a strong indication that the specific native

conformation was formed and maintained. In order to determine if LssEdomW59/80F was an inactive mutant its structure had to be assessed. It was possible that the loss of activity was owing to an unfolding of the mutant's structure, caused by the mutation of tryptophan into phenylalanine in position 59 and 80.

3.4.5. Circular Dichroism spectra

As a first step in the assignment of the contribution of the tryptophan residues to the native CD spectrum, spectra in the far-UV region were recorded for LssEdomW80F and LssEdomW59/80F at 25ºC (Figure 46). The CD data have been corrected for concentration and are reported as the mean residue molar ellipticities.

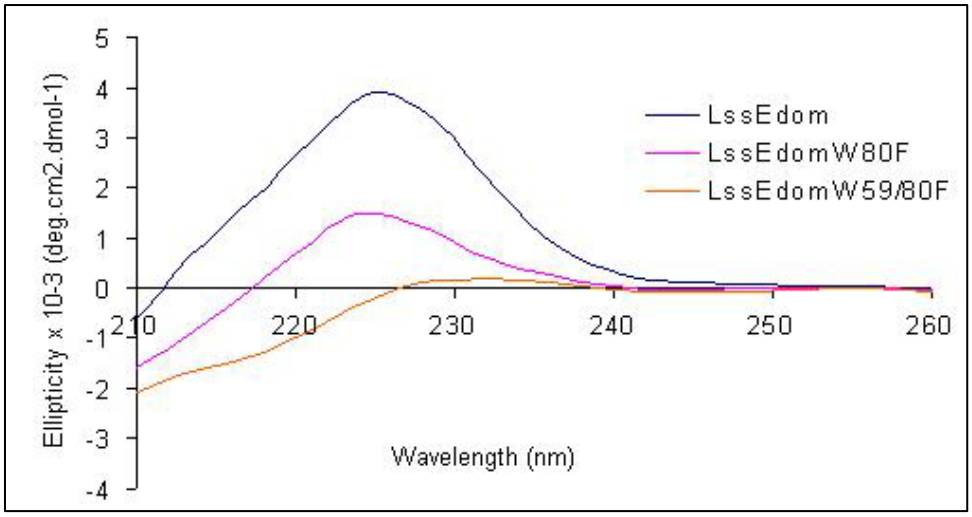


Figure 46: The far-UV CD spectra of LssEdom (blue), LssEdomW80F (pink) and LssEdomW59/80F (orange) recorded at 25ºC. The protein concentration was 0.30 mg.ml^{-1}. The buffer employed was 1 mM sodium phosphate, pH 7.0**.**

LssEdomW80F exhibited a similar spectrum to LssEdom, with a maximum near 224 nm. The spectrum of LssEdomW59/80F in the far-UV region has a comparatively low intensity, which is to be expected from a

protein with a low α-helical content (Kabsch and Sander, 1983). There is a negative band at 212 nm ([θ] = - 1803 deg.cm^2.dmol^{-1}) that is in accordance with β-sheet structure (Manning and Woody, 1987). In addition, the intensities of the bands are most likely affected by both positive and negative contributions from other secondary structures, as well as contributions from aromatic residues. The intensity of the CD band below 210 nm was not informative. This and the contribution from aromatic residues to the CD spectra, even in the double mutant LssEdomW59/80F, hindered attempts to estimate the amounts of the various types of secondary structure elements in LssEdom.

The complex band pattern observed in the far-UV CD spectra of the native proteins has changed in the corresponding spectra of denatured LssEdomW80F and LssEdomW59/80F recorded at 85ºC and 25ºC after denaturation (Figure 47). This indicates that the asymmetric environment around the various aromatic amino acid residues in the native state was disrupted upon unfolding, resulting in CD spectra typical for an unordered polypeptide chain.

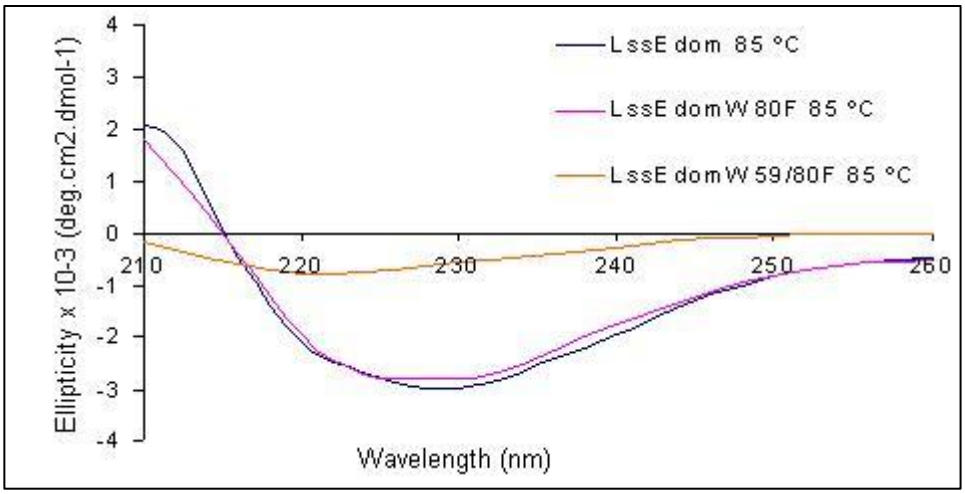


Figure 47: The far-UV CD spectra of LssEdom (blue), LssEdomW80F (pink) and LssEdomW59/80F (orange) recorded at 85ºC. The protein concentration was 0.30 mg.ml^{-1}. The buffer employed was 1 mM sodium phosphate, pH 7.0.

All the tryptophan residues in LssEdom contribute to the CD spectrum over essentially the entire far-UV spectral range. The contribution of the deleted tryptophan residue can be obtained by subtracting the tryptophan mutant spectra from the wild type spectra. The results from such calculations are shown in Figure 48.

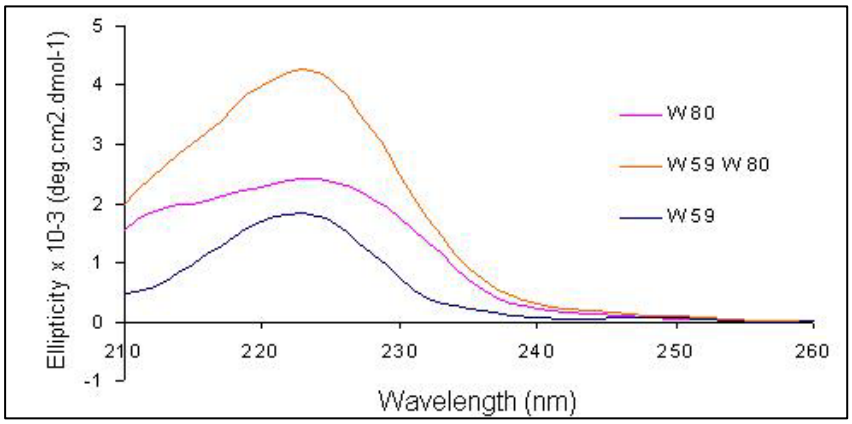


Figure 48: Contribution of individual Trp residues to the far-UV CD spectrum of LssEdom. These spectra were obtained as difference spectra: LssEdom minus LssEdomW80F **(W80, pink),** LssEdom minus LssEdomW59/80F **(W59 W80, orange)** and LssEdomW80F minus LssEdomW59/80F **(W59, blue).**

The wavelength maximum of the CD spectra of the individual tryptophan residues W80 and W59 is 224nm. The accompanying ellipticities are respectively 2407, 1804 and 4211 deg.cm^2.dmol^{-1} for W80, W59 and both W59/80. The magnitude of those bands indicates that both tryptophan contribute very largely to the far-UV CD spectrum of LssEdom. This means that tryptophan residues will interfere with predictions of the contents of various types of secondary structure based on algorithms using ellipticity data in the wavelength region 210-260 nm. Interestingly, tryptophan 59, which is exposed to the solvent, contributes least to the CD spectrum. For the phenylalanine-substituted mutants, putative contributions from the inserted phenylalanine residues are of course included in the corresponding tryptophan CD spectrum.

3.5. Attenuated Total Reflection (ATR) spectrometry

3.5.1. Lss versus LytM

Fourier Transform infrared spectroscopy (FT-IR) experiments were carried out to investigate the secondary structure of LssEdom$_{248-385}$. FT-IR was also performed on the control protein LytM$_{186-314}$, which is 50% homologous in sequence to LssEdom$_{248-385}$ (Figure 49). Indeed, the LytM$_{186-314}$ structure has been solved (Firczuk *et al.*, 2005) and thus could serve to determine the accuracy of the secondary structure element assignments for LssEdom$_{248-385}$.

```
LytM          186 -------ASWLTSRKQLQPYGQYHG---GGAHYGVDYAMPENSPVYSLTDGTVVQAGWSN
Lss           248 AATHEHSAQWLNNYKKGYGYGPYPLGINGGMHYGVDFFMNIGTPVKAISSGKIVEAGWSN
                      *.**.. *:   ** *     ** *****: *  .:** :::.*.:*:*****

LytM          YGGGNQVTIKEANSNNYQWYMHNNRLTVSAGDKVKAGDQIAYSGSTGNSTAPHVHFQRMS
Lss           YGGGNQIGLIENDGVHRQWYMHLSKYNVKVGDYVKAGQIIGWSGSTGYSTAPHLHFQRMV
              ******: : * :. : ***** .: .*..** ****: *.:***** *****:*****

LytM          GGIGNQYAVDPTSYLQSR 314
Lss           NSFSNSTAQDPMPFLKSA 385
              ..:.*. * ** .:*:*
```

Figure 49: Sequence alignment of LytM and LssEdom using Clustal W.

The proteins, $LssEdom_{248-385}$ and $LytM_{186-314}$, were analysed at 112 µM concentration in 1 mM sodium phosphate (pH 7.0) buffer. Infrared spectra of the amide I and II bands were obtained for both proteins. The spectral region that is most sensitive to the protein secondary structural components is the amide I band (1700-1600 cm^{-1}), which is due almost entirely to the C=O stretch vibrations of the peptide linkages (approximately 80%). The amide II band, in contrast, derives mainly from in-plane N-H bending (40-60% of the potential energy) and from the C-N stretching vibration (18-40%) (Krimm and Bandekar, 1986), showing much less protein conformational sensitivity than its amide I counterpart. As with many other proteins, the amide I contour is a composite of overlapping component bands that represent different elements of the protein secondary structure. Some of these components are usually resolvable by the computational procedure of band narrowing by Fourier self-deconvolution. The deconvolved spectra of $LssEdom_{248-385}$ and $LytM_{186-314}$ in solution are shown in Figure 50.

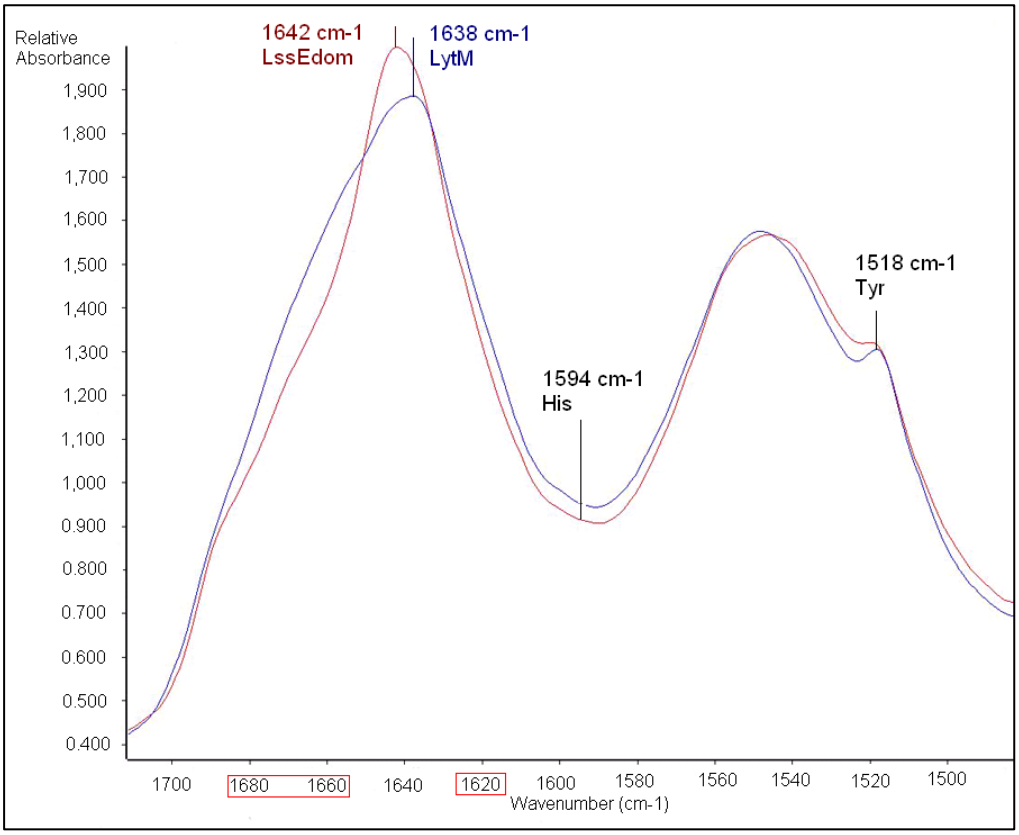


Figure 50: Infrared deconvolved spectra of LytM$_{186-314}$ (blue) and LssEdom$_{248-385}$ (red). The spectra are overlaid to show the similarities between them.

The spectra displayed the characteristics of a predominantly β-sheet structured protein, with a maximum at 1638 cm^{-1} for LytM$_{186-314}$ and 1642 cm^{-1} for LssEdom$_{248-385}$. For both spectra there is a distinctive peak at 1518 cm^{-1}, with a width of 8 cm^{-1}, representative of the high level of tyrosine residues in both proteins (11 tyrosines in LytM$_{186-314}$ and 10 tyrosines in LssEdom$_{248-385}$). The peak for the histidine (1594 cm^{-1}) is less evident as it is broader. The LytM$_{186-314}$ and LssEdom$_{248-385}$ spectra do not superimpose. The difference seen at 1600 cm^{-1} is not significant as this region of the spectrum is very variable between proteins. This could be due to the contribution of lateral chains. Nonetheless, the differences observed at 1620 and 1670 cm^{-1} are significant and are certainly related to differences in β-sheets conformation.

The presence of certain peptide secondary structure elements was determined based on the positions of the component bands. In order to quantify the approximate amount of each protein secondary structure, a curve fitting of deconvolved amide I band contour was performed with Gaussian function. The assignments of the main infrared absorption bands are given in Table 6, where the percentage of the area of a component band assigned to a certain conformation in the total area of all bands was considered to reflect the approximate amount of each structure.

Table 6: Infrared absorption band positions (in wavenumbers, cm^{-1}) determined with the aid of a Fourier self deconvolution procedure for amide I for $LssEdom_{248-385}$ and $LytM_{186-314}$. Assignments are given based on the literature data (Byler and Susi, 1986, Goormaghtigh *et al.*, 1994). Italic numbers in brackets show tentative percentage of each structure obtained from curve fitting.

Protein	Turns	β-sheets	Unordered
$LssEdom_{248-385}$	1668 (*29*) 1689 (*3*)	1629 (*47*)	1642 (*21*)
$LytM_{186-314}$	1662 (*51*) 1689 (*4*)	1616 (*3*) 1634 (*42*)	

According to the Protein Database (PDB) the structure of $LytM_{186-314}$ (PDB code 2B44, 2B0P, 2B13) contains 13% helical feature (3 helices, 17 residues) and 42% of β-sheet (12 strands, 55 residues). As shown by the fitting, the FT-IR spectra have curves at 1616 and 1634 cm^{-1}, representative of β-sheet backbone conformation, and curves at 1662 and 1689 cm^{-1}, which can be assigned to ordered features like turns (Byler and Susi, 1986, Goormaghtigh *et al.*, 1994). The proportion of β-sheet (45%) is in agreement with the X-ray structural data. The same fitting process was applied to $LssEdom_{248-385}$ spectra using Gauss curves. The curve at 1629 cm^{-1} is representative of β-sheet backbone conformation, those at 1668 and 1689 cm^{-1}, are representative of

ordered features such as turns. As shown on this fitting, there is a band at 1642 cm^{-1} that contributes for 21% and may originate from unordered features.

We can observe a net difference between the spectra of $LytM_{186-314}$ and $LssEdom_{248-385}$. There is a change in the conformation backbone, which, according to the fitting, is due to a weakening of the β-sheet structure in $LssEdom_{248-385}$. Its contribution could be explained by the presence of a flexible tail at the extremity of the activity domain (LssEdom) that is not present in the fragment of LytM studied.

3.5.2. Is LssEdomW59/80F an inactive mutant?

LssEdomW80F and LssEdomW59/80F are mutants of the lysostaphin endopeptidase domain. LssEdomW59/80F activity was dramatically reduced compared to LssEdom, as shown in the different activity test run on MV11 (cf. section 3.4.4.). LssEdomW80F and LssEdomW59/80F conformations were investigated employing infrared. The native conformation of LssEdom served as a control to measure any change in conformation in the mutants LssEdomW80F and LssEdomW59/80F. The proteins were analysed at 112 µM concentration in 1 mM sodium phosphate (pH 7.0) buffer. Infrared deconvolved spectra of the amide I and II bands for the three proteins are displayed in Figure 51.

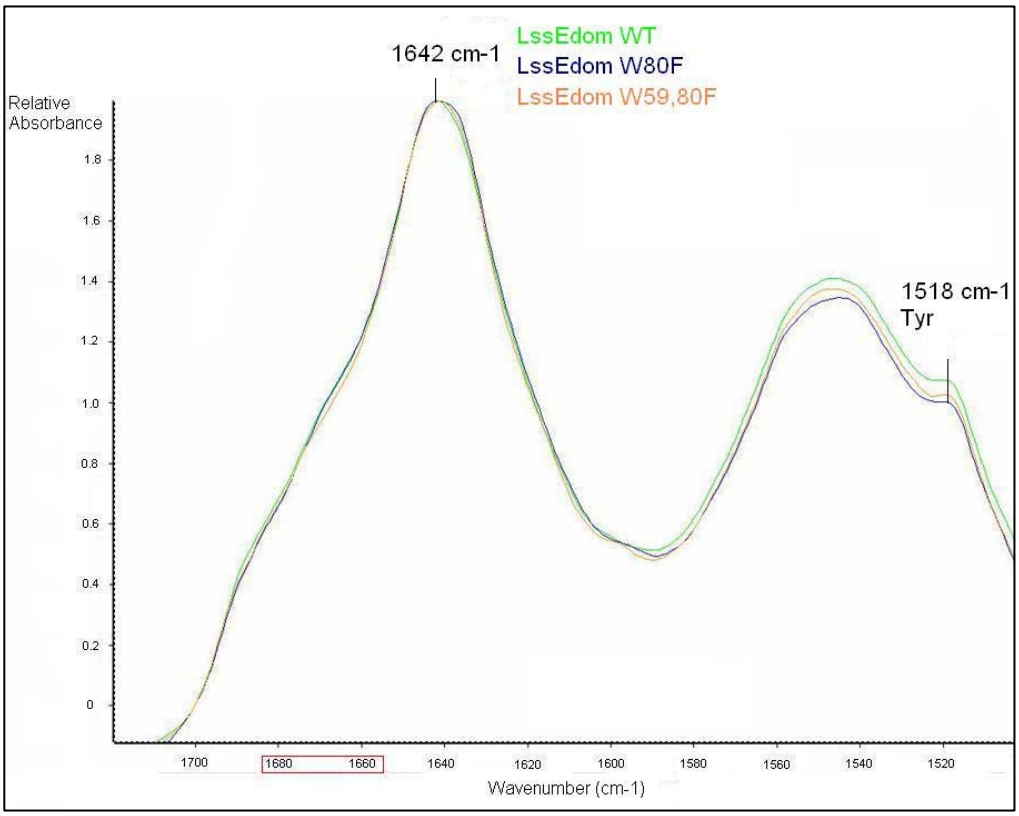


Figure 51: Infrared deconvolved spectra of LssEdomWT (green), LssEdomW80F (blue) and LssEdomW59/80F (orange). The spectra are overlaid for comparison.

A visual comparison of the infrared spectra of LssEdomWT, LssEdomW80F, LssEdomW59/80F indicates the absence of major changes in the protein secondary structure upon mutation of tryptophan into phenylalanine. All traces represent a single amide I band with a maximum at 1642 cm^{-1}. A significant absorbance can also be observed at 1518 cm^{-1}. It is indicative of the presence of 10 tyrosines in all the protein sequences.

The difference observed at 1670 cm^{-1} is representative of a weakening in the β-sheets conformation. Also, this could imply LssEDW59/80F is not an inactive mutant as its structure is slightly different from the WT, showing signs of denaturation.

The amide I region was used for peptide structure determination. The assignments of the main infrared absorption bands are given in Table 7.

Table 7: Infrared absorption band positions (in wavenumbers, cm^{-1}) determined with the aid of Fourier self deconvolution procedure for amide I for LssEdomWT, LssEdomW80F and LssEdomW59/80F. Assignments are given based on the literature data (Byler and Susi, 1986, Goormaghtigh *et al.*, 1994). Italic numbers in brackets show the tentative percentage of each structure obtained from curve fitting.

Protein	Turns	β-sheets	Unordered
LssEdomWT	1668 (*29*) 1689 (*3*)	1629 (*47*)	1642 (*21*)
LssEdomW80F	1670 (*18*) 1688 (*8*)	1616 (*7*)	1641 (*67*)
LssEdomW59/80F	1670 (*24*) 1688 (*4*)	1616 (*7*)	1641 (*65*)

The same fitting process was applied to LssEdomWT, LssEdomW80F and LssEdomW59/80F spectra using Gauss curves. The comparison of the band positions for the three traces show a shift from 1629 cm^{-1} for the wild type protein to 1616 cm^{-1} for the mutants. This could reflect an aggregated state with intermolecular β-sheets (Surewicz *et al.*, 1990). In addition, the dominance of the 1641 cm^{-1} band in LssEdomW80F and LssEdomW59/80F spectra points to a relatively high proportion of poorly defined structure. This is in clear contrast to the LssEdomWT which gives rise to a weaker component band in the spectral region characteristic of an unordered structure.

Although it can be contested, the fitting show more unordered features for the mutants compared to the wild-type, which leads to the conclusion that the reduction of activity might be due to a fraction of the proteins in solution that are aggregated after mutation. Consequently, it can be argued that LssEdomW59/80F is an inactive mutant as the mutations performed on LssEdom might have led to destabilisation of the structure.

3.5.3. Is LssEdom suited for crystallisation studies?

Fourier Transform infrared spectroscopy (FT-IR) experiments were carried out to investigate the conformation of LssTdom, LssEdom and full

length Lss to determine which candidate was best for crystallisation studies. For each protein spectrum (112 µM for LssEdom, 30 µM for Lss and LssTdom) 100 accumulations repeated 10 times were collected at room temperature with a resolution of 2 cm^{-1}. Infrared deconvolved spectra of the amide I and II bands for the three proteins are displayed in Figure 52.

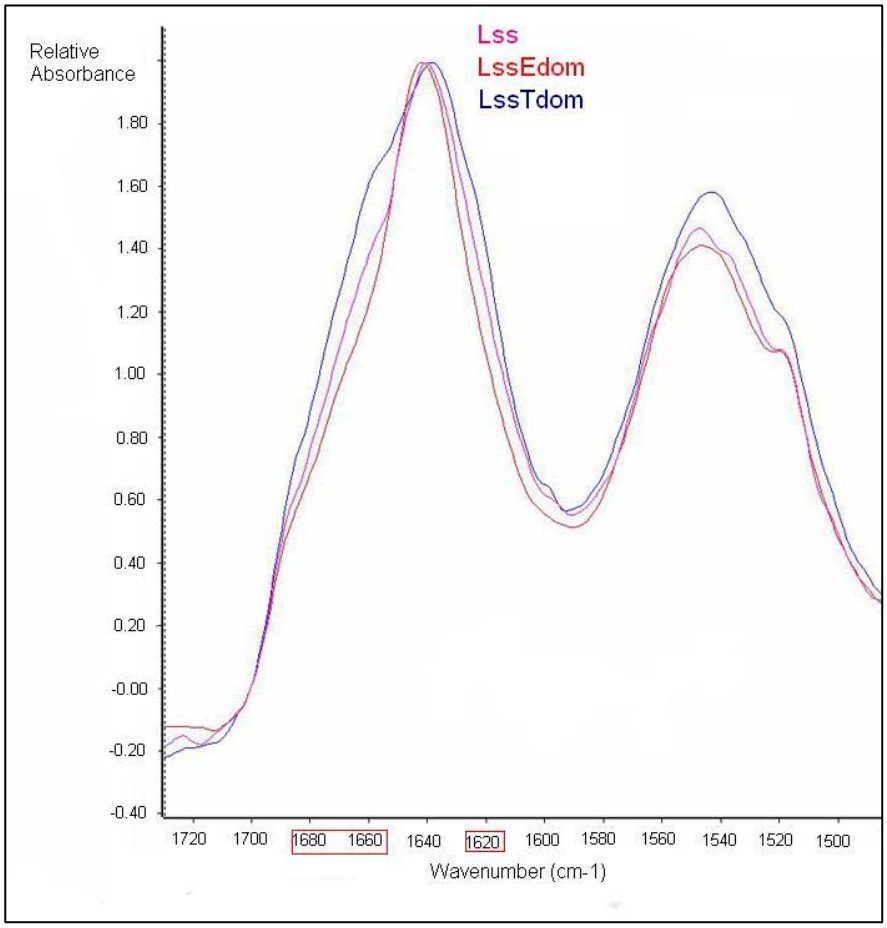


Figure 52: Infrared deconvolved spectra of Lss (pink), LssEdom (red) and LssTdom (blue). The spectra are overlaid for comparison.

The Lss, LssEdom and LssTdom spectra do not superimpose. The difference seen at 1600 cm^{-1} is not significant as this region is very variable. This could be due to the contribution of lateral chains. However, the differences observed at 1620 and 1670 cm^{-1} are significant and are certainly related to differences in β-sheet conformation.

The amide I region was used for peptide structure determination. The assignments of the main infrared absorption bands are given in Table 8.

Table 8: Infrared absorption band positions (in wavenumbers, cm^{-1}) determined with the aid of Fourier self deconvolution procedure for amide I for Lss, LssEdom, LssTdom. Assignments are given based on the literature data (Byler and Susi, 1986, Goormaghtigh *et al.*, 1994). Italic numbers in brackets show tentative percentage of each structure obtained from curve fitting.

Protein	Turns	β-sheets	Unordered
Lss	1662 (*56*) 1689 (*4*)	1630 (*35*)	1641 (*5*)
LssEdom	1668 (*29*) 1689 (*3*)	1629 (*47*)	1642 (*21*)
LssTdom	1662 (*39*) 1687 (*2*)	1620 (*2*) 1627 (*49*) 1635 (*8*)	

The same fitting process was applied to Lss, LssEdom and LssTdom spectra using Gauss curves. For Lss spectra, the curve at 1630 cm^{-1} is representative of β-sheet backbone conformation, those at 1662 and 1689 cm^{-1}, are representative of ordered features such as turns. As shown on this fitting, there is a contribution of 5% of unordered features, which could be explained by the presence of the linker in between the endopeptidase domain (LssEdom) and the targeting domain (LssTdom) in full length Lss. For LssEdom spectra, the curve at 1629 cm^{-1} is representative of β-sheet backbone conformation, those at 1667 and 1688 cm^{-1}, are again representative of ordered features such as turns. As shown on this fitting, there is a contribution of 21% of unordered features. Its contribution could be explained by the presence of a flexible tail at the extremity of LssEdom. The fitting of LssTdom spectra showed curves at 1620, 1627 and 1635 cm^{-1}, representative of β-sheet backbone conformation, and curves at 1661 and 1686 cm^{-1}, representative of ordered features such as turns. ALE-1 and LssTdom are 80% homologous, as shown by the alignment

in Figure 53. The structure of ALE-$1_{260-361}$ deposited at the Protein Database (PDB code 1R77) contains 5% helical feature (1 helice, 5 residues out of 101) and 50% of β-sheet (8 strands, 50 residues out of 101). The proportions are in agreement with the X-ray structural data.

```
ALE-1     260 DYKTNKYGTLYKSESASFTANTDIITRLTGPFRSMPQSGVLRKGLTIKYDEVMKQDGHVW
LssTdom   401 GWKTNKYGTLYKSESASFTPNTDIITRTTGPFRSMPQSGVLKAGQTIHYDEVMKQDGHVW
               :*****************.******* *************: * **:************

ALE-1         VGYNTNSGKRVYLPVRTWNESTGELGPLWGTIK 361
LssTdom       VGYTGNSGQRIYLPVRTWNKSTNTLGVLWGTIK 493
              ***. ***:*:********:**. ** ******
```

Figure 53: Sequence alignment of ALE-1 and LssTdom using Clustal W.

Curve fitting analysis shows limitations for the quantitative assessment of the "absolute" content of protein secondary structure. However it provides a sensitive tool for monitoring, in relative terms, the nature of changes in the conformation of protein backbone (Surewicz *et al.*, 1993). It seems, according to the fitting, that the conformation of full length Lss is more ordered than that of the LssEdom. This could be due to cooperation in between the two domains, LssEdom and LssTdom, maintaining a certain conformation for substrate docking.

It can be argued that the difference in shape of the spectra for LssTdom is due to the difference in concentration of the proteins used. Nevertheless, LssTdom fitting is in good agreement with ALE-1 X-ray structure and the prediction of backbone structure show an ordered protein.

3.6. Conclusions

Circular dichroism (CD) spectroscopy was used empirically to estimate the amounts of various types of secondary structure in the LssEdom, in order to evaluate the suitability of this protein for crystallisation studies. However, the CD spectra of β-proteins having little or no α-helix is reported to vary widely (Manning and Woody, 1987). Despite both the wealth of information available on protein structures and the increasing use of CD, the contribution of aromatic side chains to the CD spectrum has not yet been fully resolved. The sign, the magnitude and the wavelength are aspects of the aromatic CD bands that are poorly understood. This complicated estimations on the amounts of various types of secondary structure in the LssEdom.

Therefore, an empirical approach to allow estimations of the effects of tryptophan side chains on the far-UV CD spectrum of LssEdom was to use site-directed mutagenesis and replace the tryptophan residues by phenylalanine. The first crucial step toward this goal was to construct tryptophan mutants of LssEdom. It was important to assess the stability and enzymatic activity of these mutants to ensure that the native conformation of the three-dimensional structure had not been changed. Results showed that contributions from tryptophan residues were an important source of variations in the far-UV region. It was demonstrated that there is a considerable interference of CD bands from the tryptophan residues in the far-UV region spectrum of LssEdom making it difficult to assign changes in ellipticity to changes in secondary structure content.

Fourier transform infrared (FT-IR) spectroscopy was a good alternative to circular dichroism (CD) spectroscopy for the detection of

conformational changes in related proteins. The IR spectral data of proteins are interpreted in terms of the vibrations of a structural repeat unit and are highly sensitive. The computerised FT-IR instrumentation improved the signal-to-noise ratio and allowed data manipulation. The band-narrowing method, Fourier self-deconvolution, has not only enriched the qualitative interpretation of the infra-red spectra, but also provided a basis for the quantitative estimation of protein secondary structure (Kauppinen *et al.*, 1981).

Despite a well-recognised conformational sensitivity of protein infrared bands, the analysis of the spectra in terms of protein secondary structure was not straightforward. Particularly controversial was the "quantitative" aspect of the infrared method. The FT-IR studies of peptides and proteins have made efforts in identifying characteristic frequencies and determining their relations to the structures of protein molecules. Nonetheless, FT-IR spectra are complicated in that different ambient circumstances, including side-chain absorbance, make it difficult to precisely assign the secondary structure and frequency. The estimation of secondary structure elements in Lss (and truncated forms) with the FT-IR method should be considered only as a good approximation (Venyaminov and Kalnin, 1990).

4. Lysostaphin tertiary structure prediction

4.1. Tertiary structure

Proteins are complex organic compounds that consist of amino acids joined by peptide bonds. The composition of amino acids in a protein defines the three dimensional form that the protein folds to. These structures are unique in the sense that a given sequence of amino acids always folds into almost the same structure under the same environmental conditions (pressure, temperature, pH, etc.). The tertiary structure is the native state, or folded form, of a single protein chain. This form is also called the functional form. The tertiary structure of a protein includes the coordinates of its residues in three dimensional space.

There are a number of methods with varying resolution to determine the structure of proteins. X-ray crystallography and NMR are the most commonly used methods to determine protein structure (Branden and Tooze, 1998). There are also less frequently used techniques like electron microscopy (Rhodes, 2006). These methods require time and effort to obtain good resolution. There are certain drawbacks for each method. For example, during X-ray crystallography, the protein is crystallised, and on rare occasions this distorts portions of a structure (Rhodes, 2006). Additionally, different laboratories using different methods may publish different structures for the same protein.

Despite their drawbacks there is still no acceptable alternative to the laboratory based methods. The next best methods for protein structure determination are theory based methods such as homology modelling. This method is referred to as a structure prediction method. It can briefly be

described as fitting a known sequence to the experimentally determined three dimensional structure of a protein that is similar in sequence (Rhodes, 2006).

4.2. Further purification steps for X-ray crystallography and NMR purposes

In order to obtain a pure sample of LssEdom protein, at high concentration and without His-tag (sometimes known to prevent crystallisation), pSGAT2-based plasmids were constructed. The LssEdom sequence was sub-cloned in a pSGAT2 vector downstream of sequences encoding for a His-tag and a thrombin cleavage site. Likewise, Lss sequence was sub-cloned in pSGAT2 vector.

4.2.1. Sub-cloning in pSGAT2

Lss and LssEdom sequences sub-cloning in the pSGAT2 vector was achieved using a PCR method. pET21a(+)-$Lss_{248-493}$ (pEA3) was used as a template along with LssForw (forward) and LssRev (reverse) primers to amplify the Lss sequence, and with LssForw (forward) and LssEdomRev (reverse) primers to amplify the LssEdom sequence. The products of amplification, named Lss (762bp) and LssEdom (438bp) PCR products are shown in Figure 54.

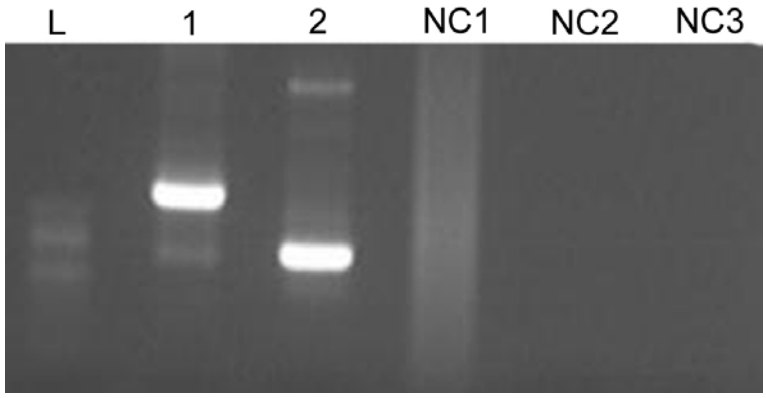


Figure 54: PCR amplification of Lss and LssEdom DNA template from pET21a(+)-$Lss_{248-493}$ (pEA3). L: Ladder, **1:** Lss PCR product (762 bp), **2:** LssEdom PCR product (438 bp). **NC: Negative Control . NC1:** presence of LssForw and LssRev primers, absence of template, **NC2:** presence of LssForw and LssEdomRev primers, absence of template, **NC3:** presence of pEA3 template, absence of primers.

Restrictions of pSGAT2 with *Bam*HI and *Hind*III resulted in the excision of a 1.1 kb insert observed by agarose gel electrophoresis, shown in Figure 55. Lss and LssEdom PCR products digested with *Bam*HI and *Hind*III were purified and cloned into pSGAT2 and used to transform DH5α cells. The clones were confirmed by PCR on recombinant plasmid (pSGAT2-Lss and pSGAT2-LssEdom) and by DNA sequencing, using T7PP (forward) and T7TP (reverse) primers. DNA sequencing data were compared to Lss and LssEdom sequences with the ClustalW program. The results confirmed the presence of Lss and LssEdom inserts in *Bam*HI-*Hind*III fragment.

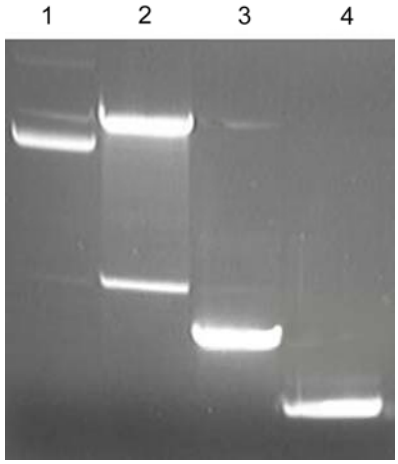


Figure 55: Restriction of pSGAT2, Lss and LssEdom PCR products with *Bam*HI and *Hind*III. 1: Supercoiled pSGAT2, **2:** Digested pSGAT2 (4429 bp), **3:** Digested Lss PCR product (748 bp), **4:** Digested LssEdom PCR product (427 bp).

Subsequent to pSGAT2-Lss and pSGAT2-LssEdom constructions, two modifications of the DNA sequences were made. First, the codon corresponding to the third glycine of the pentaglycine sequence, which is downstream of the starting codon in pSGAT2, was mutated by PCR to code for a serine residue, in order to obtain a translated sequence that could not be cleaved by the lysostaphin activity domain. Secondly, a stop codon was inserted at the end of LssEdom DNA sequence.

For the first round of PCR, site-directed mutants were constructed from pSGAT2-Lss and pSGAT2-LssEdom vectors by PCR, using the sense/antisense primer pair: T7PP (forward)/Gly3mutpSGAT2 (reverse) primers. The purified PCR-amplified DNA fragment was digested with *Bam*HI and *Xba*I, gel-purified, and sub-cloned into pSGAT2-Lss and pSGAT2-LssEdom plasmids digested with the appropriate restriction endonucleases. For the second round of PCR, a site-directed mutant was constructed from pSGAT2-LssEdom vector by PCR, using the sense/antisense primer pair: pSGAT2edomFOR (forward)/pSGAT2edomREV (reverse). The purified

PCR-amplified DNA fragment was digested with *Bam*HI and *Hind*III, gel-purified, and sub-cloned into pSGAT2-LssEdom plasmid backbone digested with the appropriate restriction endonucleases. A stop codon was inserted at the end of LssEdom DNA sequence using the sense/antisense primer pair: pSGAT2EdomFor (forward)/pSGAT2EdomRev (reverse). The PCR product was digested with *Bam*HI and *Hind*III and sub-cloned into pSGAT2-LssEdom plasmids. In each case, the fidelity of the PCR-amplified DNA fragments was established by nucleotide sequencing after sub-cloning into the expression vector (Figure 56).

Translation of pSGAT2-Ser3-LssEdom sequencing data compared with LssEdom sequence

```
E1T7PP      EGDISMNTIHHHHHHNTSGSGGSGGRLVPRGSAATHEHSAQWLNNYKKGYGYGPYPLGIN 60
Edom        ------------------------------MAATHEHSAQWLNNYKKGYGYGPYPLGIN 29
                                           ****************************

E1T7PP     GGMHYGVDFFMNIGTPVKAISSGKIVEAGWSNYGGGNQIGLIENDGVHRQWYMHLSKYNV 120
Edom       GGMHYGVDFFMNIGTPVKAISSGKIVEAGWSNYGGGNQIGLIENDGVHRQWYMHLSKYNV 89
           ************************************************************

E1T7PP     KVGDYVKAGQIIGWSGSTGYSTAPHLHFQRMVNSFSNSTAQDPMPFLKSA 179
Edom       KVGDYVKAGQIIGWSGSTGYSTAPHLHFQRMVNSFSNSTAQDPMPFLKSA 138
           ***************************************************
```

Translation of pSGAT2-Ser3-Lss sequencing data compared with Lss sequence

```
L1T7PP      EGDISMNTIHHHHHHNTSGSGGSGGRLVPRGSAATHEHSAQWLNNYKKGYGYGPYPLGIN 60
Lss         ------------------------------MAATHEHSAQWLNNYKKGYGYGPYPLGIN 29
                                           ****************************

L1T7PP      GGMHYGVDFFMNIGTPVKAISSGKIVEAGWSNYGGGNQIGLIENDGVHRQWYMHLSKYNV120
Lss         GGMHYGVDFFMNIGTPVKAISSGKIVEAGWSNYGGGNQIGLIENDGVHRQWYMHLSKYNV 89
            ************************************************************

L1T7PP      KVGDYVKAGQIIGWSGSTGYSTAPHLHFQRMVNSFSNSTAQDPMPFLKSAGYGKAGGTVT180
Lss         KVGDYVKAGQIIGWSGSTGYSTAPHLHFQRMVNSFSNSTAQDPMPFLKSAGYGKAGGTVT149
            ************************************************************

L1T7PP      PTPNTGWKTNKYGTLYKSESASFTPNTDIITRTTGPFRSMPQSGVLKAGQTIHYDEVMKQ240
Lss         PTPNTGWKTNKYGTLYKSESASFTPNTDIITRTTGPFRSMPQSGVLKAGQTIHYDEVMKQ209
            ************************************************************

L1T7PP      DGHVWVGYTGNSGQRIYLPVRTWNKSTNTLGVLWGTIK 277
Lss         DGHVWVGYTGNSGQRIYLPVRTWNKSTNTLGVLWGTIK 246
            ******************************************
```

Figure 56: ClustalW homology analysis showing sequence alignment of E1T7PP and L1T7PP sequences, identified by nucleotide sequencing, and LssEdom or Lss sequence respectively. Stars indicate a perfect match in between the theoretical sequence and the sequencing data.

To summarise, for this study two vectors were built: pSGAT2-Ser-Lss and pSGAT2-Ser-LssEdom. The difference between those plasmids occurs in the nucleotide sequence they carry: Lss and LssEdom, respectively. The expression of Lss and LssEdom is under the control of the T7 promoter inducible by IPTG. After expression, the His-tag sequence is fused to the *N*-terminal region of the protein and can be removed by thrombin cleavage. For future studies, a $pSGAT2_{Gly \rightarrow Ser}$ vector with a substitution of a glycine in a serine residue in position 3 of the pentaglycine sequence was also built.

4.2.2. His-tag removal

The recombinant plasmids: pSGAT2-Ser-Lss or pSGAT2-Ser-LssEdom, were transformed into the host, *E.coli* BL21 (DE3), and addition of isopropyl β-thiogalactoside (IPTG) induced the overexpression of a band of approximately 30 and 18 kDa molecular weight, corresponding respectively to H_6-Lss and H_6-LssEdom fusion proteins, when compared to the uninduced control in SDS-PAGE. The proteins were overexpressed in soluble form in the cytoplasm.

The H_6-Lss and H_6-LssEdom were purified to near homogeneity, observed on SDS-PAGE, using a single affinity chromatographic step on zinc agarose. In order to obtain a pure sample of LssEdom, without His-tag, known to help obtain highly ordered crystals, the His-tag was removed by enzymatic cleavage and separated by chromatography. Thrombin recognised the consensus sequence Leu-Val-Pro-Arg-Gly-Ser and cleaved the peptide bond between arginine and glycine. Such a protease cleavage site allowed removal of an upstream domain “His-Tag”. Kinetic experiments were run with different

concentrations of thrombin (1U, 3U, 5U of thrombin per mg of protein), for different incubation times (1h, 4h, 16h) and at different temperatures (19ºC, 4ºC). As shown by SDS-PAGE (Figure 57), cleavage was completed following a 16 h treatment, at 19ºC, with 3 units per mg of protein, in iexA buffer (10 mM Tris pH 7.0, 50 mM NaCl, 5% (v/v) glycerol).

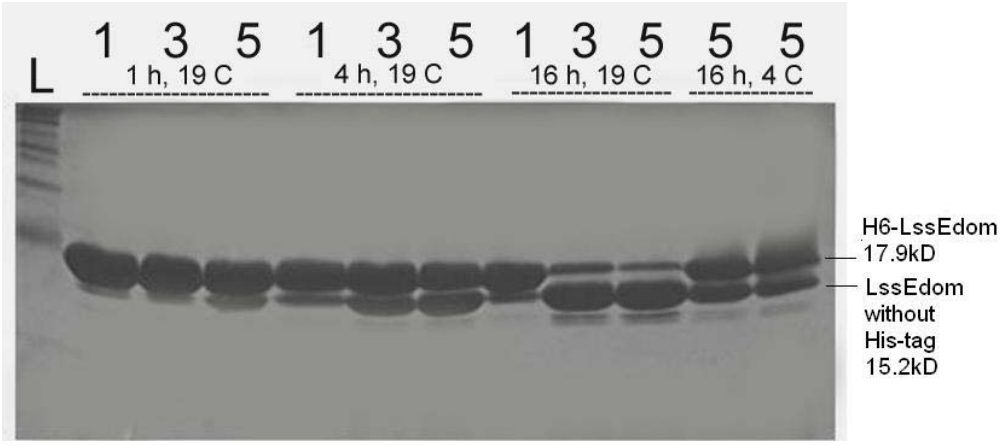


Figure 57: Thrombin His-tag cleavage optimisation. 1: Edom + 1U thrombin, **3:** Edom + 3U thrombin, **5:** Edom + 5U thrombin.

4.2.3. Further purification of LssEdom

Thrombin removal was performed by ion-exchange chromatography. The selected program permitted the specific adsorption of LssEdom on a cation-exchange column under low salt concentration conditions and the elution of LssEdom under high salt concentration conditions. Thrombin, negatively charged under the pH condition selected, was not retained on the column. Samples collected during the elution step were loaded on a SDS-PAGE (Figure 58).

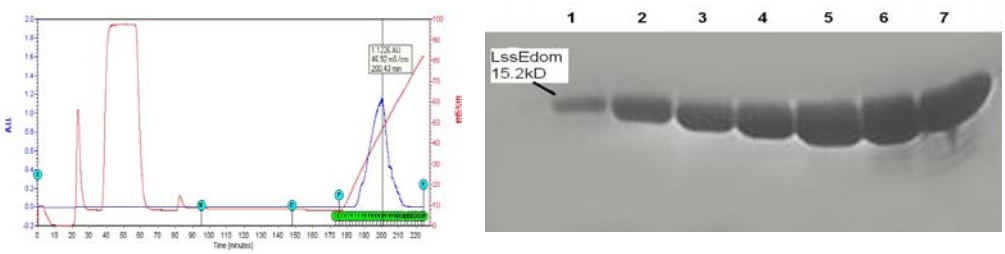


Figure 58: LssEdom purification by ion-exchange chromatography: chromatogram and SDS-PAGE showing the fractions containing LssEdom (1-5).

To prove that thrombin was removed from solution by the ion-exchange chromatography described above, a test was done on the unstable LssEdom expressed from pSGAT2-Ser-LssEdom vector without a stop codon at the end of the DNA sequence. H_6-LssEdom was dialysed in buffer A (20 mM sodium phosphate (pH 7.4), 5 mM imidazole, 0.5 M sodium chloride, 5% (v/v) glycerol), then 5U of thrombin per mg of protein were added. Ion-exchange chromatography was run directly and the supernatant left for 3 days at 19ºC. The negative control, where H_6-LssEdom in bufferA was not exposed to thrombin, did not show any degradation after 3 days at 19ºC. On the contrary, for the positive control, in which was added 5U of thrombin per mg of protein after chromatography in accordance with the test, H_6-LssEdom was completely degraded after 3 days at 19ºC. Proteins were loaded on SDS-PAGE as shown Figure 59.

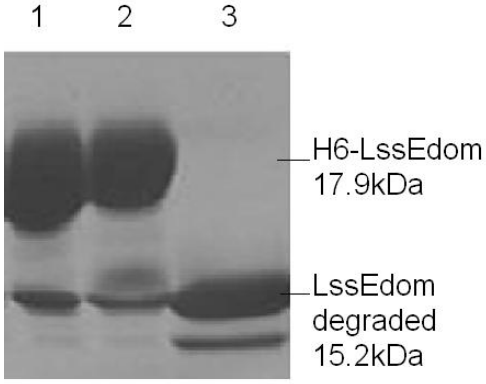


Figure 59: Verification of thrombin removal by ion-exchange chromatography by cleavage assay on unstable LssEdom expressed from pSGAT2-Ser-LssEdom vector without a stop codon at the end of the DNA sequence. 1: Thrombin cleavage prior ion-exchange chromatography, **2:** Negative control, **3:** Positive control.

The H_6-LssEdom and removed His–tag were physically separated from LssEdom using a single affinity chromatographic step on zinc agarose. LssEdom was homogeneous as observed on SDS-PAGE, shown in Figure 60.

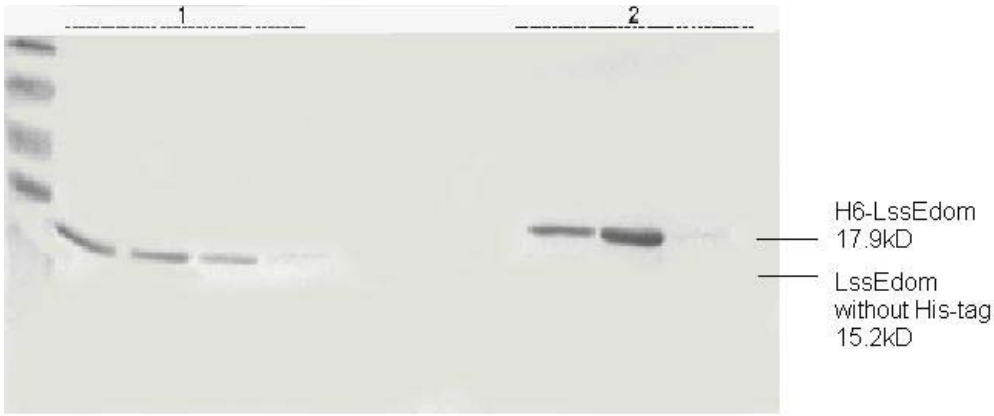


Figure 60: LssEdom purification by zinc-affinity chromatography. 1: LssEdom without His-tag eluted with the flowthrough on zinc-NTA agarose, **2:** H_6-LssEdom retained on zinc-NTA agarose.

LssEdom was further purified by gel filtration chromatography as detailed in Materials and Methods (cf. section 2.2.1.). Protein concentration was determined by UV spectroscopy as detailed in Materials and Methods (cf. section 2.2.5.). The protein solution was concentrated to obtain suitable crystals for X-ray diffraction. For this purpose a vertical filter tube was used and the protein concentration was determined after ultra filtration. The results showed that the maximal solubility of the lysostaphin endopeptidase domain, when expressed in 2YT medium, was reached and the concentration value remained as: $C_{LssEdom} = 220\ \mu M$.

4.3. Stability studies

4.3.1. Differential scanning calorimetry (DSC)

We considered using differential scanning calorimetry (DSC) to define the thermal stability of untagged LssEdom. The DSC thermal denaturation profile of untagged LssEdom, purified on zinc, in 10 mM HEPES pH 7.0 and 150 mM NaCl (Figure 61) displayed a single peak centered at 62.8ºC (12.7 μM LssEdom, v = 20ºC.h^{-1}) and denaturation proceeded with an enthalpy range, ΔH, of 21.8 ± 0.3 kcal.mol^{-1}. Therefore, LssEdom is stable at room temperature, at which structural studies were run.

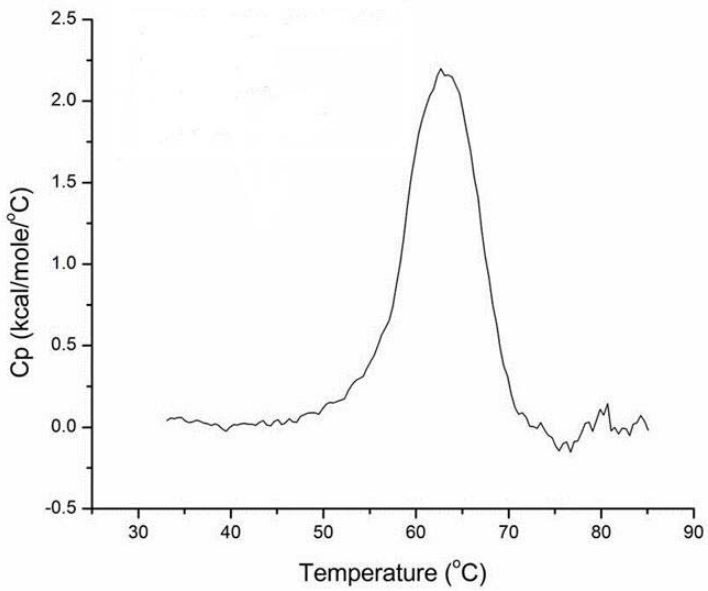


Figure 61: Exprimental heat capacity profile of untagged LssEdom.
The trace shows the thermal denaturation of LssEdom (12.7 μM) dialysed in 10 mM HEPES pH 7.0, 150 mM NaCl, at a scan rate of 20ºC.h^{-1}.

4.3.2. Circular Dichroism (CD)

Circular dichroism (CD) is commonly used in denaturation experiments in which the CD signal of a protein is monitored while the protein is perturbed in some fashion (e.g., increasing temperature or chemical denaturant). Changes in CD signal reflect changes in the protein structure. Information about protein stability or folding intermediates can be obtained (Woody, 1996). We employed the circular dichroism (CD) technique to define the thermal stability of LssEdom-H_6. Shown in Figure 62 are the spectra of LssEdom-H_6 subjected to a scale up (UP) and scale down (DN) of temperature. Spectra were recorded at 224 nm using 26 µM LssEdom in 1 mM sodium phosphate buffer, pH 7.0. These studies indicated that LssEdom-H_6 denaturation occurred above 60ºC, as shown by previous DSC experiments run with untagged LssEdom, and that LssEdom does not refold spontaneously.

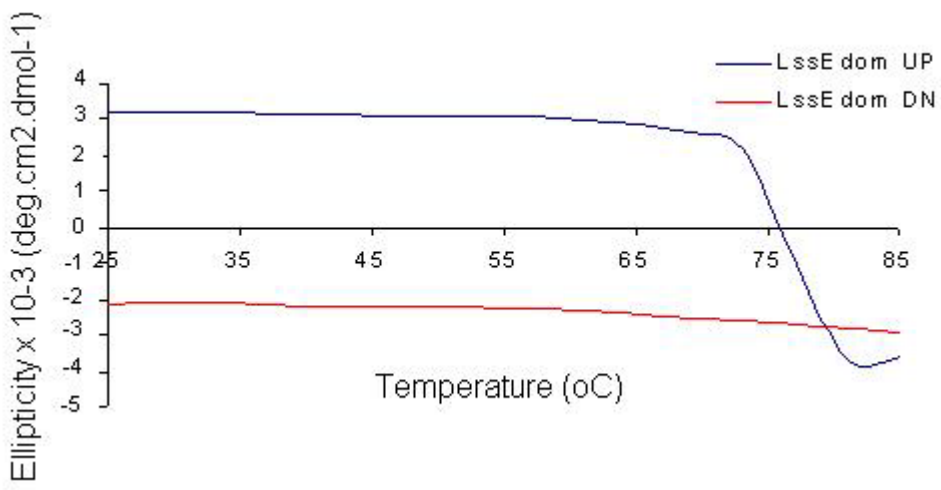


Figure 62: LssEdom-H_6 CD denaturation experiment monitored while increasing the temperature (scale UP) or deacreasing the temperature (scale down DN). Thermal denaturation recorded at 224 nm. The protein concentration was 0.30 mg.ml^{-1}. The buffer employed was 1 mM sodium phosphate, pH 7.0**.**

Those informations are corroborated by Zygmunt and Tavormina (1972) that demonstrated inactivation of lysostaphin by heating at 75ºC for 15 minutes.

4.4. LssEdom crystallisation

4.4.1. Crystal screening

In initial high throughput crystallisation trials, 2016 different conditions were tested on freshly purified untagged LssEdom. About ten small, individual diamond shaped crystals could be observed with the condition described in Figure 63, which was considered as the "Hit condition". This condition came from the NextalBiotech™ $AmSO_4$™ screen and was made of 0.1 M MES sodium salt pH 6.5, 2.0 M ammonium sulphate and 5% (w/v) PEG 400. No significant crystalline material could be observed with any other condition.

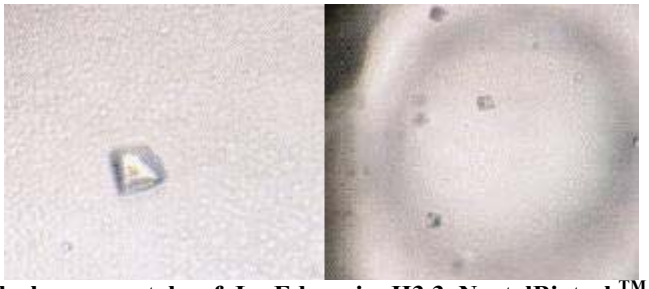


Figure 63: Diamond shape crystals of LssEdom in H3.3 NextalBiotech™ $AmSO_4$™ screen. LssEdom dialysed in 10 mM Tris pH 7.0 and 150 mM sodium chloride buffer (1 µl in the drop): 66 µM. Reservoir solution (1 µl in the drop, 60 µl in the reservoir): 0.1 M MES sodium salt pH 6.5, 2.0 M ammonium sulphate and 5% (w/v) PEG 400.

Screens were performed manually around the "Hit condition" of the high throughput screens, using the sitting-drop vapour-diffusion technique. The effect of various concentrations of ammonium sulphate and PEG on the protein crystallisation, were tested. A control was made with student-prepared

solutions and with commercial solutions. No crystals were observed for the different conditions.

The experiments made previously with tagged LssEdom-H_6 indicated that salts (ammonium sulphate, magnesium sulphate and sodium acetate) helped the protein to crystallise. As a consequence, in addition to 0.1 M MES sodium salt pH 6.5, 2.0 M ammonium sulphate and 5% (w/v) PEG 400, 100 mM and 200 mM sodium chloride was also added. However, no significant crystalline material was observed.

Therefore, screens were performed with the streak seeding technique. This technique is utilised for epitaxial jumps to obtain crystals under conditions which may be significantly different from those in which the original seed was obtained. A whisker cat was used to gently touch a rod and dislodge seeds. The seeds were transferred to a new protein-precipitant drop composed as the seed source drop: 1.70 M ammonium sulphate, 3% PEG 400, 0.1 M MES sodium salt pH 6.5. The only difference consisted in the concentration of protein used: 219, 100 and 37 □M. As no significant crystalline material was observed, another streak-seeding experiment was performed with the solution that contained residual crystals after cryo-mounting operation. The seeds were placed into pre-equilibrated drops that were prepared with commercial solutions and 54 μM of untagged LssEdom (ratio 4μl:4μl). Only phase separation of the solution and precipitation of the protein were observed.

4.4.2. X-ray diffraction

Data collection was carried out at the EMBC Grenoble outstation. One crystal mounted on a 0.05 mm loop was found. With the highest magnification

a very weak and disordered diffraction was observed, limited to 3.0-3.2 Å. No work was carried on with these data.

4.5. Nuclear Magnetic Resonance (NMR) of LssEdom in solution

Considering the difficulties encountered in crystallising LssEdom, another strategy to obtain its structure was investigated. Nuclear magnetic resonance (NMR) spectroscopy was explored as a replacement for X-ray crystallography, for the determination of protein structure in solution. To be suitable for NMR spectroscopy, the protein must be small (below 16 kDa), highly pure, stable at room temperature and very soluble. In preliminary studies, differential scanning calorimetry (DSC) and circular dichroism (CD) techniques were used to determine LssEdom stability at room temperature. Thus, optimisation of LssEdom expression in a minimal medium, purification and solubility were required prior to NMR spectroscopy studies.

4.5.1. Effects of temperature and additives on LssEdom expression level

To perform NMR spectroscopy studies, proteins must be expressed in a minimal medium (MM) to be able to ^{15}N and ^{13}C label them. The level of expression of LssEdom is very much dependent on the medium used. As shown Figure 64, LssEdom expressed in MM at 30°C gave mostly insoluble proteins.

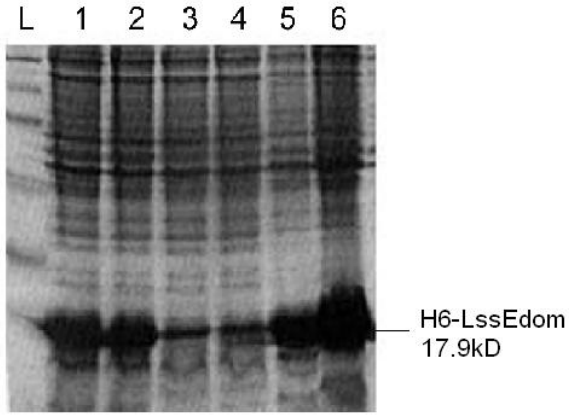


Figure 64: LssEdom expression in MM at 30˚C.
Out of all protein expressed in MM (**1,2**), LssEdom is mostly detected in the insoluble fraction (**5,6**) rather than in the soluble one (**3,4**).

$ZnCl_2$ was also added to MM at a final concentration of 1 mM, which is lower than the concentration that inhibits lysostaphin activity (Szweda *et al.*, 2005). LssEdom contains a Zn^{2+} ion in its activity site and it was believed that addition of zinc cation in solution would help the protein to fold, and thus, go into the soluble fraction. No changes were observed in protein solubility. Thus, the effects of salt and temperature were tested on the expression of LssEdom in the soluble fraction. By increasing the amount of salt in MM to 5 $g.l^{-1}$ and 10 $g.l^{-1}$, and decreasing the temperature to 27˚C and 20˚C, we observed that the expression of LssEdom in the soluble fraction was largely improved as shown in Figure 65.

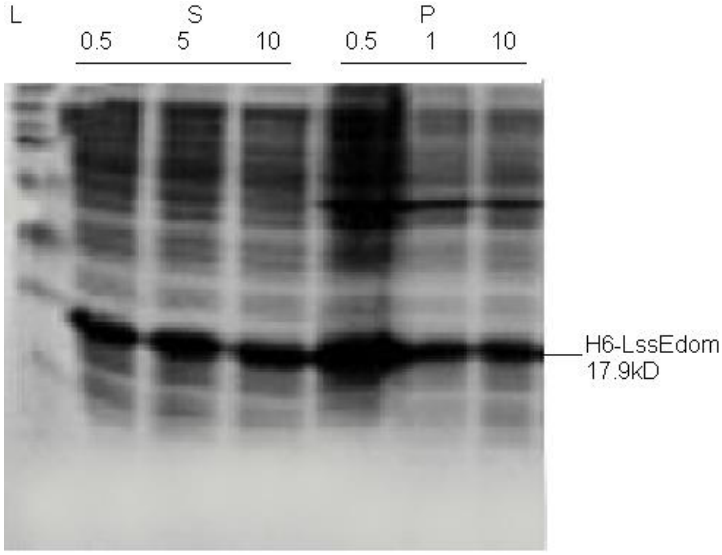


Figure 65: LssEdom expression in MM at 20ºC with added in solution different amounts of salt (0.5, 5, 10 g.l^{-1}). Under those conditions, LssEdom is mostly detected in the soluble fraction **(S)** rather than in the insoluble one **(P).**

In consequence, for the next experiments, LssEdom was expressed in MM containing 0.5 g.l^{-1} NaCl as no significant changes were observed when adding higher amounts of salt, and was incubated at 20°C rather than 30°C, as this had been shown to improve folding of the protein.

4.5.2. Effects of buffers, pH and additives on LssEdom solubility

To perform NMR spectroscopy studies, proteins must be dialysed in a solution containing no element that could give a distinct signal. A good choice of buffer is reported to be sodium phosphate buffer and the amount of NaCl should be kept relatively low.

At first, LssEdom expressed in MM containing 0.5 g.l^{-1} NaCl and incubated at 20°C, was purified by affinity chromatography on zinc agarose and dialysed in a buffer containing 10 mM sodium phosphate, pH 7.4 and 150 mM NaCl. The highest concentration of LssEdom obtained was 55 µM which was too low for NMR studies. Thus, different dialysis buffers were tested,

HEPES and MES, at the same concentration and with the same amount of salt in solution. No improvement was observed. Therefore, additives were tested to improve protein solubility. According to Golovanov and coworkers, simultaneous addition of charged amino acids, L-arginine (L-Arg) and L-glutamic acid (L-Glu) at 50 mM to the buffer could dramatically increase maximum achievable concentration of soluble protein (Golovanov *et al.*, 2004). LssEdom that had been expressed and purified was dialysed in a buffer containing 10 mM sodium phosphate, pH 7.4, 100 mM NaCl, 50 mM L-Arg and 50 mM L-Glu. A slight improvement in protein concentration was measured: 75 μM, compared to the initial 55 μM , which was far from the 8.7 times increase reported in the literature. Also, different pHs were tested to test their effect on protein solubility. Sodium phosphate buffer at pH 6.5 and pH 8.5 with 100 mM NaCl, 50 mM L-Arg and 50 mM L-Glu presented no differences. Addition in the dialysis buffer of an uncleavable LssEdom substrate (Gly_3 is believed to be the product of pentaglycine cleavage by lysostaphin) at a ratio 1:1, enzyme:substrate, did not help the protein to fold and go into solution. Finally, transfer of transformed bacteria from LB medium for growth culture at 30°C to MM for LssEdom expression at 20°C (to enhance chaperone production and help folding of the protein) did not help to obtain a better concentration of the protein after purification and dialysis in 10 mM sodium phosphate, pH 7.4, 100 mM NaCl, 50 mM L-Arg and 50 mM L-Glu.

Therefore, the maximal concentration obtained for LssEdom expressed in minimum medium at 20°C, purified to near homogeneity as described for crystallisation studies (cf. section 4.2.3.) and dialysed in a buffer containing 10

mM sodium phosphate, pH 7.4, 100 mM NaCl, 50 mM L-Arg and 50 mM L-Glu, was 75 μM. It is well below the 1mM working concentration used in NMR spectroscopy studies on a Brüker AV-600 spectrometer operating at 600 MHz.

It was suggested that a chaperone was missing in the minimum medium used for protein expression as higher concentration of protein (220 μM) could be obtained when the protein was expressed in rich 2YT medium. The pro-region of lysostaphin was reported by Baba and Schneewind to potentially function as an intramolecular chaperone for folding (Baba and Schneewind, 1996), as was also reported for refolding of subtilisin by Zhu and coworkers (Zhu *et al.*, 1989).

4.5.3. Heteronuclear Single Quantum Coherence (HSQC)

The NMR preliminary studies, carried out in collaboration with Dr J. Long (Centre for Biomolecular Sciences, University of Nottingham, Nottingham) showed a lower molecular weight impurity, and so, LssEdom was further purified by size-exclusion chromatography. To determine the nature of the impurity, mass spectrometry (MS) studies were achieved in collaboration with G. Coxhill (Chemistry, University of Nottingham, Nottingham) on LssEdom samples before and after purification by size-exclusion chromatography. Examination by MS (ESI positive ion) on the MicroToF did show some peaks additional to the multiply charged ions of the protein, in particular a relatively intense singly charged ion at m/z=719 (cf. Appendix 2). On the contrary, in the spectra produced on the Apex FTMS system, this peak was not present (cf. Appendix 3). The spectra indicated that

the intensity of the ion at m/z=719 is dependent on the voltage difference between the Capillary Exit and the Skimmer 1. This dependency suggested that this peak was probably a single charged fragment from the protein. Thus, no differences were found between the "before size-exclusion chromatography" and the "after size-exclusion chromatography" spectra and no impurity was detected by MS (ESI positive and negative ions).

The NMR preliminary HSQC study, carried out in collaboration with Dr J. Long (Centre for Biomolecular Sciences, University of Nottingham, Nottingham) was run with ^{15}N labelled LssEdom at a concentration of 0.44 mg in 600 µl of 10 mM sodium phosphate, pH 7.4, 150 mM NaCl, 0.04% azide, 10% D_2O in ddH_2O. The experiment was run over-night. The spectrum was dispersed and no impurity was detected (Figure 66). Nonetheless, the maximal concentration of LssEdom obtained, 75 µM, is well below the 1 mM working concentration necessary for structure generation on a Bruker AV-600 spectrometer operating at 600 MHz.

It was later found that the presence of zinc cations in solution, in the different steps of purification and dialysis, were responsible for the precipitation of the LssEdom and Lss proteins. By adding Chelex© resin, and thus chelating divalent ions in the solutions used for the dialysis of the protein solutions, the concentration of both proteins was greatly increased achieving a maximum level of 500 µM. These improvements in the maximal concentrations found for LssEdom may have been factors enabling the performance of NMR studies into the structure of LssEdom in interaction with the Gly-Gly-L-Ser-Gly-L-Ser peptide that is not cleavable by the glycinase

action of LssEdom; this in collaboration with Dr F. Guerlesquin at the IMR02, IBSM, Marseille.

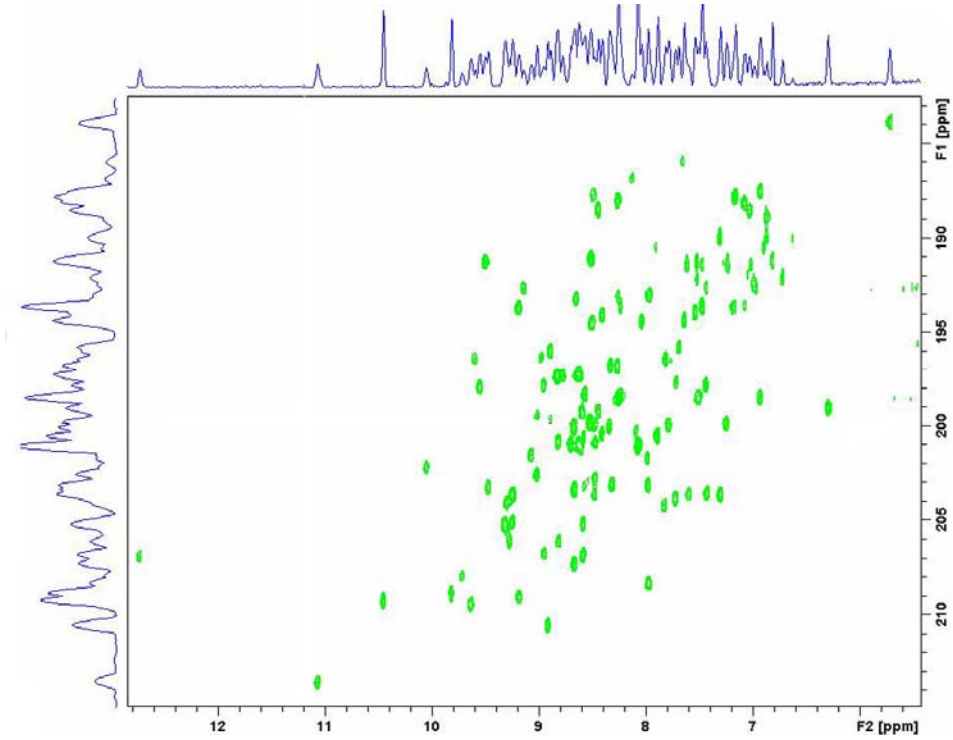


Figure 66: The NMR preliminary HSQC spectrum was obtained with a 600MHz Bruker Spectrometer, on a N^{15} labelled LssEdom peptide at a concentration of 75 µM in 10 mM sodium phosphate, 150 mM NaCl, pH 7.4, 0.04% azide, 10% D_2O in ddH_2O.

4.6. Lysostaphin tertiary structure modelling

4.6.1. Sequence to structure

The primary structure (i.e., amino acid sequence) of a protein is much easier to determine than its tertiary structure. The gap between the number of proteins of known sequence and number of proteins of known tertiary structure is increasing in an accelerating manner. This stems from the limitations of laboratory based methods exploited in this chapter. In order to close this gap, there have been many searches on how to determine the tertiary structure of a

protein from its sequence. Tertiary structure prediction methods come into picture at this point.

The CASP (Critical Assessment of techniques for protein Structure Prediction) experiment is an international platform where researchers try to predict tertiary structures of proteins whose structure are not available (PSPC, 2008). Automated structure prediction provides models of reasonable quality to researchers who are not themselves structure prediction experts (Marti-Renom *et al.*, 2000).

The method used for the construction of an atomic-resolution model of Lss was homology modelling, also known as comparative modelling. Protein structures ("templates") that likely resemble the structure of the query sequence ("target"), from the basis of the resemblance with its primary sequence, were selected and used to produce an alignment that maps residues in the target sequence to residues in the template sequence. The sequence alignment and the template structure were then used to produce a structural model of the target.

4.6.2. Template selection

There are a few exceptions to the general rule that proteins sharing significant sequence identity will share a fold. For example, a judicious set of mutations of less than 50% of a protein can cause the protein to adopt a completely different fold (Dalal *et al.*, 1997). However, such a massive structural rearrangement is unlikely to occur in evolution, especially since the protein is usually under the constraint that it must fold properly and carry out its function in the cell.

In silico analysis of lysostaphin gene from *S. simulans* biovar *staphylolyticus* showed that LytM (GI: 49482513) and lysostaphin (GI: 3287967) were sharing the same traits (Sharma *et al.*, 2006).

In a subsection of the MEROPS peptidase database (Rawlings *et al.*, 2008) which searches for protein homologues in a genome, ALE-1 from *S. capitis* (Uniprot code O05156), LytM from *S. aureus* (UniProt code Q8NYG1) and Millericin B from *S. milleri* (MEROPS code MERO13517) were very close in evolution to lysostaphin from *S. simulans* and *S. staphylolyticus* (Uniprot code P10547 and P10548, respectively) (Figure 67).

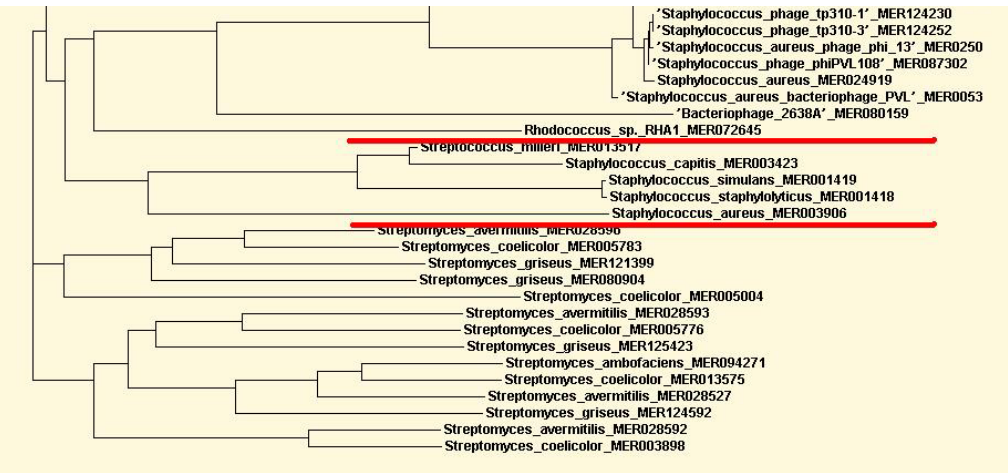


Figure 67: MEROPS phylogenetic tree with entry lysostaphin.

The outcome of a BLAST search with entry of the subfamily M23B is a list of proteins that are homologous to lysostaphin from *S. staphylolyticus* (UniProt code P10548):

- ALE-1 from *S. capitis* (UniProt code O05156),
- ORF013 from *S. aureus* phage3A (UniProt code Q4ZCS7),
- NMB1333 from *Neisseria meningitidis* (UniProt code Q9JZ20),
- LytM from *S. aureus* Mu50 (UniProt code Q8NYG1),
- IscS from *Buchnera aphidicola* (UniProt code Q89A19),

- ZooA from *Haemophylus influenzae* (UniProt code P44693),
- YebA from *Shigella flexneri* (UniProt code P0AFT1),
- LytH from *Bacillus subtilis* (UniProt code O32130),
- two other putative peptidases whose Uniprot code were Q4A0H0 and Q4L983.

The sequences of these proteins were used as input to construct a phylogenetic tree using MULTALIN software. The outcome is shown Figure 68.

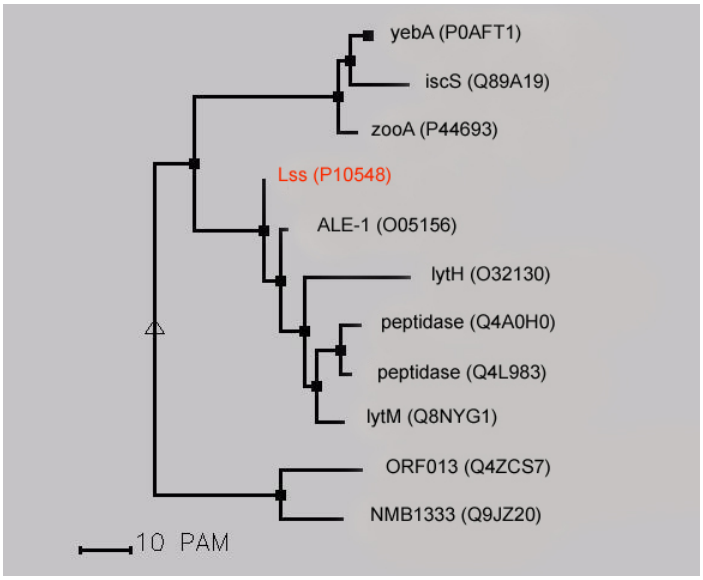


Figure 68: Phylogenetic tree constructed with MULTALIN software with M23B subfamily peptidase proteins as input sequences. (PAM, rule line to count the position in the alignment).

From these data it can be observed that ALE-1 and LytM are very close to Lss on the evolutionary tree, showing similarities in sequence.

4.6.3. Structure prediction with Modeller 8v2

Lysostaphin ($Lss_{253-493}$) protein sequence was submitted to the Protein Homology/analogY Recognition Engine (PHYRE) (Bennett-Lovsey *et al.*, 2008), which searched for three-dimensional protein structures available in the Brookhaven protein databank (Bernstein *et al.*, 1977). Three-dimensional structural homologues became fortuitously available for the homology model

of Lss during the course of my PhD. Two reference proteins, LytM (PDB code 2B44) and ALE-1 (PDB code 1R77), were used to model the structure of Lss. The identity scores compared to the target protein were 48% and 83% for $LytM_{186\text{-}314}$ and for $ALE\text{-}1_{272\text{-}365}$ respectively (Figure 69).

(A)

```
LytM          ---HAKDASWLTSRKQLQPYGQYHG---GGAHYGVDYAMPENSPVYSLTDGTVVQAGWSN
LssEdom       AATHEHSAQWLNNYKKGYGYGPYPLGINGGMHYGVDFFMNIGTPVKAISSGKIVEAGWSN
                 * :.*.**.. *:   ** *     ** *****: *  .:** ::::.*.:*:*****

LytM          YGGGNQVTIKEANSNNYQWYMHNNRLTVSAGDKVKAGDQIAYSGSTGNSTAPHVHFQRMS
LssEdom       YGGGNQIGLIENDGVHRQWYMHLSKYNVKVGDYVKAGQIIGWSGSTGYSTAPHLHFQRMV
              ******: : * :. : ***** .: .*..** ****: *.:***** *****:*****

LytM          GGIGNQYAVDPTSYLQSR
LssEdom       NSFSNSTAQDPMPFLKSA
              ..:.*. * ** .:*:*
```

(B)

```
ALE-1         -YKTNKYGTLYKSESASFTANTDIITRLTGPFRSMPQSGVLRKGLTIKYDEVMKQDGHVW
LssTdom       GWKTNKYGTLYKSESASFTPNTDIITRTTGPFRSMPQSGVLKAGQTIHYDEVMKQDGHVW
               :*****************.******* *************: * **:************

ALE-1         VGYNTNSGKRVYLPVRTWNESTGELGPLWGTIK
LssTdom       VGYTGNSGQRIYLPVRTWNKSTNTLGVLWGTIK
              ***. ***:*:********:**. ** ******
```

Figure 69: Amino acid sequence alignment of target with template proteins using ClustalW. (A) LssEdom aligned with LytM. **(B)** LssTdom aligned with ALE-1.

It should be noted that the amino acid sequence determines the tertiary structure of a protein (Anfinsen *et al.*, 1954). On one hand, accurate alignment between the target sequence and template structure with a high level of sequence homology can guarantee better homology modelling. On the other hand, the tertiary structures of proteins are tightly coupled with their activities and functions and might not only be determined by their primary sequences. It was reported in sections 1.5.1 and 1.5.2 that LytM and ALE-1 possessed a similar function to Lss.

The linker region between the *N*-terminal and *C*-terminal domains of Lss was also submitted to a subsection of the Brookhaven protein databank (Bernstein *et al.*, 1977) which searched sequences that had published three-

dimensional structures. The search result indicated that no PDB sequence was identical to the one submitted, nevertheless, the best match for this linker region appeared to be one of the loops in the carboxypeptidase A structure starting from amino acid 148 and finishing with amino acid 165 (Greenblatt *et al.*, 1998). The identity scores compared to the target protein was 45% (Figure 70).

```
1ARL            DAGFGKAGASSSPCSETY
Lsslinker       SAGYGKAGGTVTPTPNTG
                .**:****.: :* .:*
```

Figure 70: Amino acid sequence alignment of target (Lss linker A.A. 383-400) with template protein (1ARL A.A. 148-165) using ClustalW.

Given the above cited templates, an alignment was generated with the Lss target (Figure 71).

```
>P1;2B44ZN
structureX:2B44ZN:253:::::::
AKDASWLTSRKQLQPYGQYHGGGAHYGVDYAMPENSPVYSLTDGTVVQAGWSNYGGGNQV
TIKEANSNNYQWYMHNNRLTVSAGDKVKAGDQIAYSGSTGNSTAPHVHFQRMSGGIGNQY
AVDPTSYLQ*

>P1;1ARLlinker
structureX:1ARLlinker:383:::::::
------------------------------------------------------------
------------------------------------------------------------
---------DAGFGKAGASSSPCSETY*

>P1;1R77
structureX:1R77:401:::::::
------------------------------------------------------------
------------------------------------------------------------
---------------------------YKTNKYGTLYKSESASFTANTDIITRLTGPFRS
MPQSGVLRKGLTIKYDEVMKQDGHVWVGYNTNSGKRVYLPVRTWNESTGELGPLWGTIK*

>P1;Lss
sequence:Lss:::::::
AQWLNNYKKGYGYGPYPLGINGGMHYGVDFFMNIGTPVKAISSGKIVEAGWSNYGGGNQI
GLIENDGVHRQWYMHLSKYNVKVGDYVKAGQIIGWSGSTGYSTAPHLHFQRMVNSFSNST
AQDPMPFLKSAGYGKAGGTVTPTPNTGWKTNKYGTLYKSESASFTPNTDIITRTTGPFRS
MPQSGVLKAGQTIHYDEVMKQDGHVWVGYTGNSGQRIYLPVRTWNKSTNTLGVLWGTIK*
```

Figure 71: Alignment of the complete Lss sequence target with "core" structural fragments identified in closely related solved structure: LytM (PDB code 2B44) and ALE-1 (PDB code 1R77), and with variable region: carboxypeptidase A (PDB code 1ARL).

Given the above alignment of the cited templates with Lss target, this information was used to generate a three-dimensional structural model of the target, represented as a set of Cartesian coordinates for each atom in the

protein (Baker and Sali, 2001). The unsolved protein structure of Lss could be modelled using MODELLER 8v2. The input file is shown in Figure 72.

```
              #Homology modelling by automodel class

from modeller.automodel import *  #Load the automodel class

log.verbose()    #request verbose output
env = environ()  #create a new MODELLER environment to build this model in

#directories for input atom files
#env.io.atom_files_directory='../atom_files'

a=loopmodel(env,
alnfile ='lss.ali',    #alignment filename
knowns =('2B44','1ARLlinker','1R77'), #codes of the templates
sequence = 'Lss')       #code of the target
a.starting_model= 1                    # index of the first model
a.ending_model  = 1                    # index of the last model
                 #(determines how many models to calculate)
      a.md_level=None       #No refinement of model

      a.loop.starting_model = 1      # First loop model
      a.loop.ending_model   = 1      # Last loop model
      a.loop.md_level = refine.fast

a.make()                      # do modeling and loop refinement
```

Figure 72: Script input file containing commands for MODELLER 8v2 for automated comparative modelling.

All the files were transferred to the C:\Program Files\Modeller8v2\examples\automodel directory for the program to run. MODELLER 8v2 treated the aligned residues as structurally equivalent. Distances and dihedral angles restraints on the target sequence were calculated from the alignment with template three-dimensional structures. The spatial restraints were obtained empirically from a database of protein structure alignments. The model was obtained by optimising the spatial restraints. The output file was a 3D model for the target sequence containing all mainchain and sidechain non-hydrogen atoms. Representations of the model are shown in Figure 73.

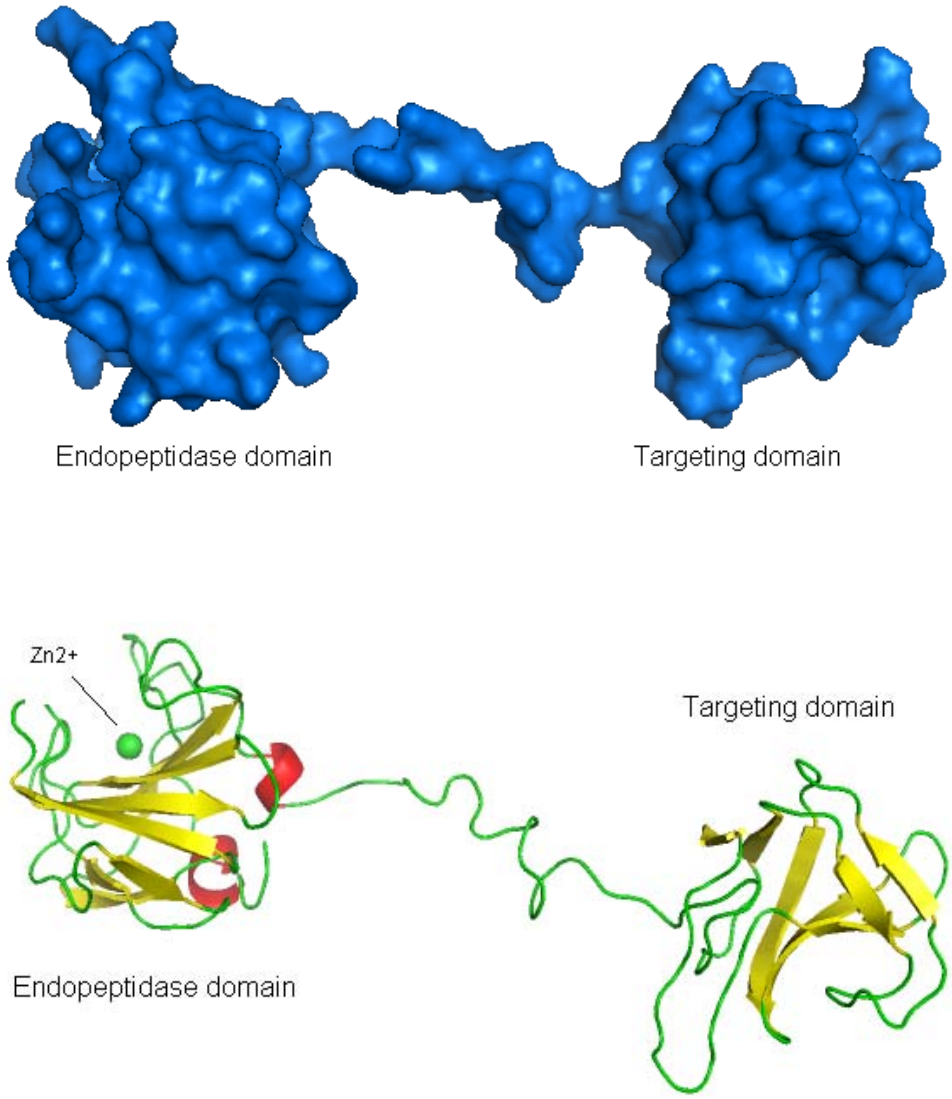


Figure 73: The predicted structure of Lss based on automated comparative modelling using MODELLER 8v2, shown as surface (top) and cartoon (bottom).

Similar to the crystal structure of $LytM_{186-314}$ (PDB code 2B44), we can see from Figure 74 that LssEdom has 8 β-sheets and 2 α-helices that will superimpose. The Zn^{2+} chelating residues (i.e., H33, D37 and H116) in LssEdom active site show strong spatial resemblances with those in LytM.

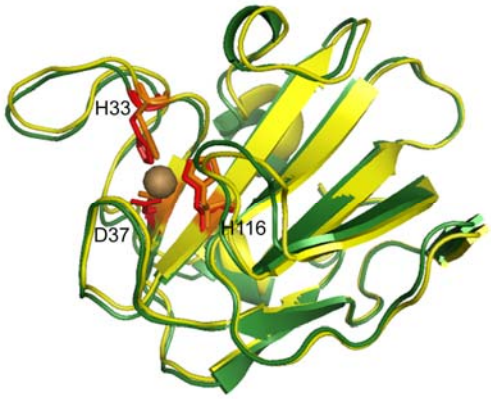


Figure 74: The three-dimensional structure of $LytM_{186-314}$ obtained by X-ray crystallography (green) superimposed with the predicted structure of LssEdom based on automated comparative modelling (yellow). The calculated RMSD comparing backbone structures is 2.618. Also indicated are the Zn^{2+} chelating residues in LssEDom active site.

Similar to the crystal structure of ALE-$1_{272-365}$ (PDB code 1R77), we can see from Figure 75 that LssTdom has 8 β-sheets that will superimpose.

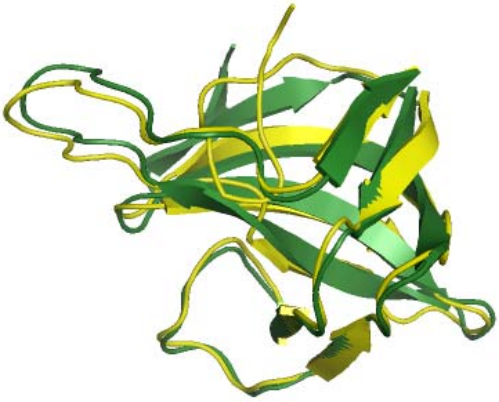


Figure 75: The three-dimensional structure of ALE-$1_{272-365}$ obtained by X-ray crystallography (green) superimposed with the predicted structure of LssTdom based on automated comparative modelling (yellow). The calculated RMSD comparing backbone structures is 5.170.

4.6.4. Evaluation of prediction accuracy

The aim of the evaluation is to predict whether or not a model is acceptable. If the model violates restraints, it might be because the optimiser used in the modelling program did not find a good optimum of the objective function. The integrity of the model was checked using MolProbity software, which analyses the bond angles of each residue in the protein. Ramachandran has shown that all combinations of phi and psi angles cannot exist in nature, simply because of steric collisions of residue atoms (Ramachandran and Sasisekharan, 1968). Ramachandran has taken into account only phi and psi angles of alanine dipeptides and has shown that there are certain conformations an alanine dipeptide cannot conform to (Karplus, 1996). These restrictions stem from the collision of two atoms in the disallowed conformations. Ramachandran (Ramachandran *et al.*, 1963) has chosen to represent the distribution of angle pairs in a single chain or multiple chains with a scatter plot, which is today called a Ramachandran plot. Some of the regions are more populated than others and some of the regions are observed in secondary structure elements like α-helices or β-sheets. The accuracy of the Lss three-dimensional structure model has been assessed by measuring the percentage of the number of correct predictions over all predictions for each atom. The Ramachandran plot (Figure 76) showed 84.8% (201/237) of all residues in the favoured regions and 92.8% (220/237) of all residues in the allowed regions.

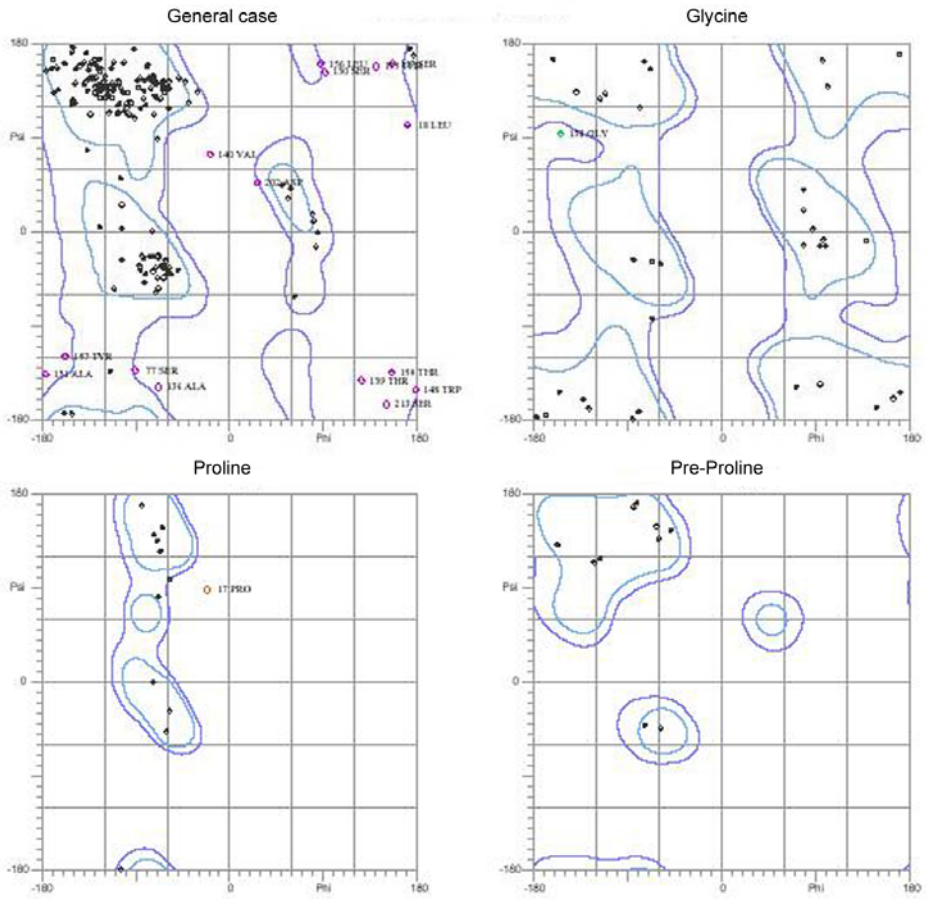


Figure 76: MolProbity Ramachandran plot of Lss model. The X-axis shows the phi angles and the y-axis shows the psi angles. Grid lines are placed every 60 degrees, showing the phi and psi angle combination (-180,180) degrees. The light blue areas indicate favoured regions; the dark blue areas indicate allowed regions.

The outcome of this search is a list of identified outliers that are listed in Table 9 and localised on the Lss model structure as shown in Figure 77.

Table 9: List of outliers in the MolProbity Ramachandran analysis of Lss model.

Outliers (Residue as numbered on the model)	phi, psi angles
Pro 17	-22.0, 89.3
Leu 18	-170.6, 103.4
Ser 77	-91.4, -132.2
Ser 117	155.2, 162.8
Ser 130	91.7, 154.4
Ala 131	-177.8, -136.4
Ala 136	-69.6, -148.1
Gly 138	-155.2, 95.3
Thr 139	126.2, -142.9
Val 140	-19.9, 75.2
Trp 148	178.5, -151.8
Thr 150	155.3, -134.4
Thr 155	140.7, 159.1
Leu 156	87.9, 162.4
Tyr 157	-158.5, -119.6
Asp 202	26.1, 48.9
Ser 213	150.6, -165.4

Figure 77: Localisation on the Lss model of outlier residues (residues highlighted in purple).

Outliers were mainly localised in the linker connecting LssEdom to LssTdom. This region is flexible and is likely to take different conformations in solution. It is of interest to understand how the linker folds in the presence of substrate accommodated by LssEdom and LssTdom in full length Lss.

Models of lysostaphin were later found in the SwissModel repository of the Expasy server (Kopp and Schwede, 2004, Kopp and Schwede, 2006). Two models were displayed, one predicting the structure for Lss *N*-terminal domain (AA 238-371) based on LytM structure (PDB code 2B13), the other predicting the structure for Lss *C*-terminal domain (AA 389-480) based on ALE-1 structure (PDB code 1R77). These models were superimposed with the Lss model predicted from MODELLER 8v2 automodel and with the LytM (PDB code 2B13) and ALE-1 (1R77) structures.

SYBYL 7.3 software was used to calculate the RMSD comparing the backbone structures of the SwissModel for LssEdom and LytM (PDB code 2B13). The RMSD for the SwissModel (0.97) is better than for the automodel MODELLER 8v2 (2.62) predicting LssEdom structure when compared to LytM structure (PDB code 2B13). In addition, SYBYL 7.3 software was used to calculate the RMSD comparing the backbone structures of the SwissModel for LssTdom and ALE-1 (PDB code 1R77). The RMSD for the SwissModel (0.08) is better than for the automodel MODELLER 8v2 (5.17) predicting LssTdom structure when compared to ALE-1 structure (PDB code 1R77).

In Figure 78 and in Figure 79 the similarities in the secondary structures (β-sheets) and the variabilities in the structures of the loops can be seen. These regions are flexible and likely to adopt different conformations in solution and in the presence of ligand so as to accommodate it.

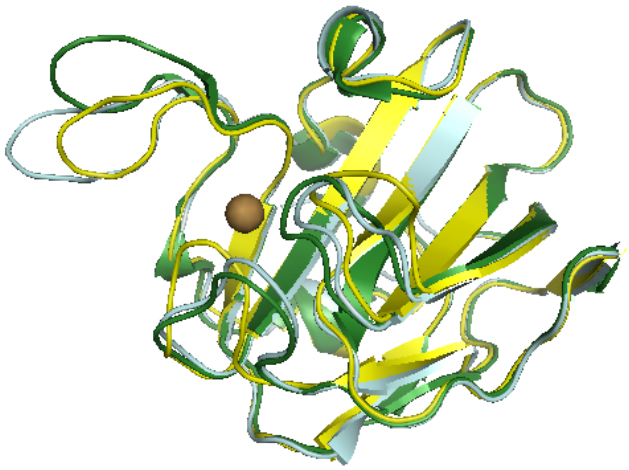


Figure 78: Swissmodel (light blue), LytM (PDB code 2B13) (green) and LssEdom model predicted from MODELLER 8v2 automodel (yellow) were superimposed to show variable regions.

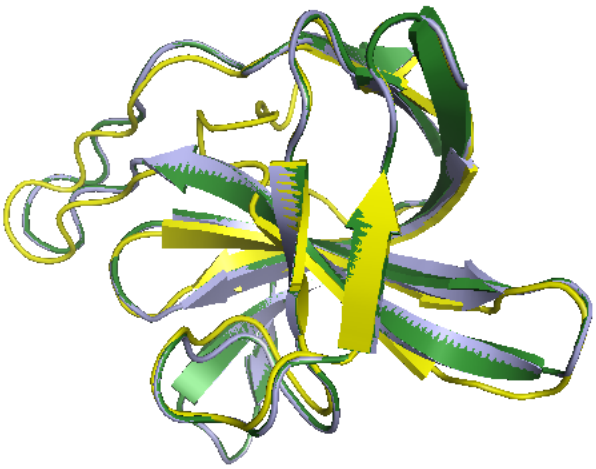


Figure 79: Swissmodel (light blue), ALE-1 (PDB code 1R77) (green) and LssTdom model predicted from MODELLER 8v2 automodel (yellow) were superimposed to show variable regions.

4.7. Conclusions

Even with highly pure and correctly folded protein, good structural data for LssEdom has not been obtained, this is in spite of the fact that the homologous $LytM_{186-314}$ protein structure has been solved (Firczuk *et al.*, 2005). This could be due to weakening of the β-sheet structure in the protein worked on, as shown in the FT-IR spectra (cf. section 3.5.3.). A model of the lysostaphin structure was built to circumvent the problems encountered; the reliability of which was tested using MOLProbity software.

The compact structure of full length Lss compared with LssTdom was demonstrated by size exclusion chromatography, where Lss was retained longer in the gel-filtration column than LssTdom (data not shown). This could be indicative of an interaction between LssEdom and LssTdom in the full length structure of Lss, bringing both domains together to accommodate the substrate. We look forward for the crystal structure of Lss in complex with its substrate (i.e., the Gly-Gly-$_L$-Ser-Gly-$_L$-Ser peptide that is not cleaved by glycinase activity of lysostaphin) to give more insight in this matter.

5. Interaction studies

The target specificity of lysostaphin is dependent on its *C*-terminal targeting domain (Baba and Schneewind, 1996). Studies on the lysostaphin and ALE-1 cell wall targeting domains, which are very similar, have shown that these targeting domains bind purified peptidoglycan from staphylococci in a manner that required the presence of pentaglycine cross-bridges (Grundling and Schneewind, 2006). It was also said that increased linearised peptidoglycan structure greatly enhanced binding of the targeting domain (Baba and Schneewind, 1996). A question that arose from those studies was to identify the receptors for the targeting domain of lysostaphin in the cell wall of *S. aureus*. Our studies aimed at identifying the receptors of lysostaphin to uncover the molecular interactions of the lysostaphin targeting domain (LssTdom) to components of the cell wall of staphylococci (Gly_5, *N*-acetylmuramyl-Ala-D-isoglutaminyl-N^{ε}-stearoyl-Lys, acetyl-Lys-D-Ala-D-Ala). Indeed, knowing the selectivity of the lysostaphin targeting domain for certain components of the cell wall of different staphylococci, would aid the evolution of the ligand-binding properties of lysostaphin targeting domain using the phage display technique (Smith and Petrenko, 1997).

The conventional techniques used for direct detection of non-covalent complexes [e.g. gel permeation chromatography (GCP), centrifugation, sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and infrared spectroscopy (Liu and Barth, 2003)] provide limited or no information about the mass and binding stoichiometry of the complex. Advanced spectroscopic techniques such as X-ray and NMR methods have been used for molecular interaction studies, but these methods require good quality crystals

and considerable time and effort to obtain good resolution structures. Methods that do not require labelling of any of the interacting molecules, such as surface plasmon resonance (SPR), isothermal titration calorimetry (ITC), and electrospray ionisation mass spectrometry (ESI-MS), are mostly preferred. Mass spectrometry (MS), with recent advances in ionisation methods, can directly provide the molecular mass and stoichiometry of a complex and allow the quantification of the interaction strength (Ganem *et al.*, 1991, Veros and Oldham, 2007, Peschke *et al.*, 2004). This technique is relatively simple, robust and easily automated (Tjernberg *et al.*, 2004).

5.1. Attenuated Total Reflection (ATR) spectrometry

LssH114A is a lysostaphin inactive mutant and its stability was investigated employing FT-IR. The activity test run with LssWT on MV11 substrate signified that LssWT was in its native conformation. Thus LssWT served as a control protein to measure any change in conformation in the mutant LssH114A. The proteins, LssH114A and LssWT, were analysed at 30 μM concentration in 1 mM sodium phosphate (pH 7.0) buffer. Infrared spectra of the amide I and II bands were obtained for both proteins. The most sensitive spectral region of protein secondary structure in the infrared is the amide I band (1700-1600 cm^{-1}), which is due almost entirely to the C=O stretch vibrations of the peptide linkages (approximately 80%). The amide II band, in contrast, derives mainly from in-plane NH bending (40-60% of the potential energy) and from the CN stretching vibration (18-40%) (Krimm and Bandekar, 1986), showing much less protein conformational sensitivity than its amide I counterpart. As with many other proteins, the amide I contour is a composite

of overlapping component bands that represent different elements of the protein secondary structure. Some of these components are usually resolvable by the computational procedure of band narrowing by Fourier self-deconvolution. The deconvolved spectra of LssH114A and LssWT in solution are shown Figure 79.

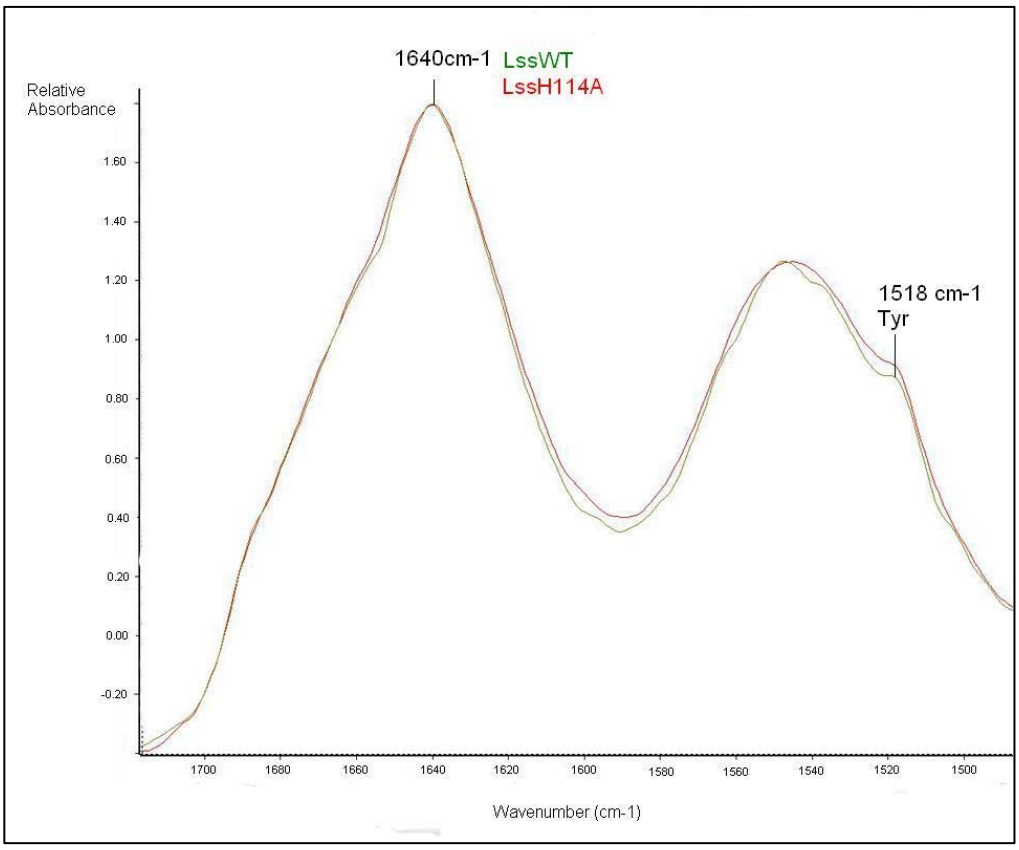


Figure 79: Infrared deconvolved spectra of LssWT (green) and LssH114A (red) in H_2O medium. The spectra are overlaid to show the similarity.

Both traces represent a broad band contour with a maximum at 1640 cm^{-1}. A significant absorbance can also be observed at 1518 cm^{-1}. It is indicative of the presence of 16 tyrosines in both protein sequences.

The presence of certain peptide secondary structure elements was determined based on the positions of the component bands. In order to

quantify the approximate amount of each protein secondary structure, a curve fitting of deconvolved amide I band contour was performed with Gaussian function. The assignments of the main infrared absorption bands are given in Table 10, where the percentage of the area of a component band assigned to a certain conformation in the total area of all bands was considered to reflect the approximate amount of each structure.

Table 10: Infrared absorption band positions (in wavenumbers, cm^{-1}) determined with the aid of Fourier self deconvolution procedure for amide I for LssWT and LssH114A. Assignments are given based on literature data (Byler and Susi, 1986, Goormaghtigh *et al.*, 1994). Italic numbers in brackets show the tentative percentage of each structure obtained from curve fitting.

Protein	Turns	β-sheets	Unordered
LssWT	1662 (*56*) 1689 (*4*)	1630 (*35*)	1641 (*5*)
LssH114A	1665 (*34*) 1688 (*2*)	1621 (*2*) 1624 (*47*)	1639 (*15*)

For LssWT spectra, the band at 1630 cm^{-1} is representative of β-sheet backbone conformation, while the ones at 1662 and 1689 cm^{-1}, can be assigned to ordered features like turns (Byler and Susi, 1986, Goormaghtigh *et al.*, 1994). As shown on this fitting, there is a weak band at 1641 cm^{-1} that may originate from unordered features, although the wavenumber at which the band occurs is a bit low. Its contribution could be explained by the presence of the flexible linker in between the activity domain (LssEdom) and the targeting domain (LssTdom) of LssWT. Similarly for LssH114A spectra, the bands at 1620 and 1624 cm^{-1} are representative of a β-sheet backbone conformation; those at 1664 and 1688 cm^{-1} are representative of ordered features such as turns. The lower spectral position of the sheet's main band and the contribution of 15% of weak β-sheet backbone conformation in LssH114A compared to LssWT could be explained by the mutation of histidine (His) into

alanine (Ala) in LssH114A that could cause a fast conformational equilibrium between β-sheet and unordered conformation.

Curve fitting analysis shows limitations for the quantitative assessment of the "absolute" content of protein secondary structure. However it provides a sensitive tool for monitoring, in relative terms, the nature of changes in the conformation of protein backbone (Surewicz *et al.*, 1993). Overall, both spectra show strong similarities and this suggested that LssH114A had a native conformation, comparable with that of LssWT.

By concentrating LssH114A by evaporation of the 45 µM solution on the attenuated total reflection (ATR) diamond, it was possible to see changes in conformation with the increase in concentration of the protein. In Figure 80 shows the spectra from 29 multiple scans that were run with 25 accumulations each. The concentration of LssH114A could be estimated by integration of the amide I band area.

#scan	Integral Amide I	[LssH1144A] µM	
0	0,25	45	Solution
1	0,24	43	Solution
2	0,24	44	Solution
3	0,2	36	Solution
4	0,24	43	Solution
5	0,21	37	Solution
6	0,25	46	Solution
7	0,31	57	Solution
8	0,35	63	Solution
9	0,45	82	Solution
10	0,7	128	Solution
11	0,78	142	Solution
12	0,63	114	Solution
13	1	181	Solution
14	1,05	191	Solution
15	1,1	200	Solution
16	1,34	244	Solution
17	1,44	261	Solution
18	1,78	325	Solution
19	2,3	419	Solution
20	2,93	533	Solution
21	4,55	829	½ film

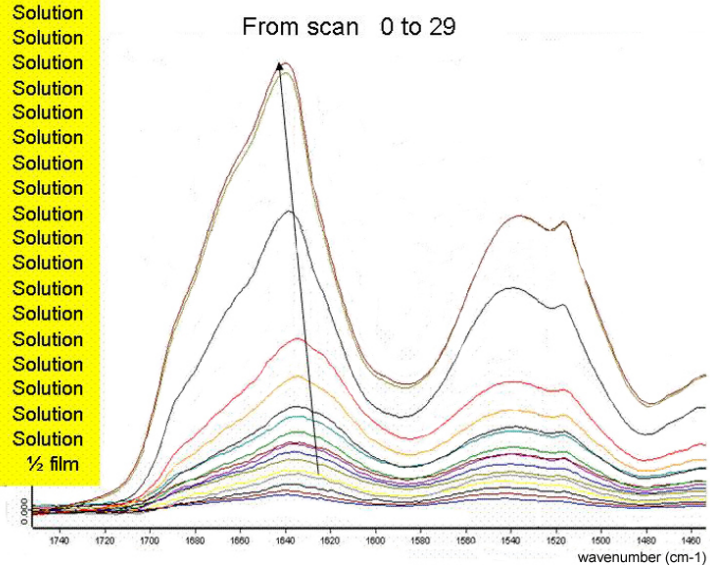


Figure 80: Infrared spectra of LssH114A at different concentrations. The table is showing an estimation of LssH114A concentration for the different spectra (i.e., 0-21).

Changes in conformation could be observed between 60 µM and 80 µM and above. When the protein was concentrated above 500 µM, forming a film, the sample thickness allowed greater interaction of the protein with the evanescent wave. However, the variations observed were compatible with a transformation of β-sheets into a compact form (amyloids), which were suggestive of an aggregation state (Halverson *et al.*, 1991, Ashburn *et al.*, 1992).

Figure 81 shows the deconvolved spectra for LssH114A in the 1580-1760 cm^{-1} region, where identical spectra have been averaged for clarity. Changes in LssH114A conformation could be observed above 63 µM.

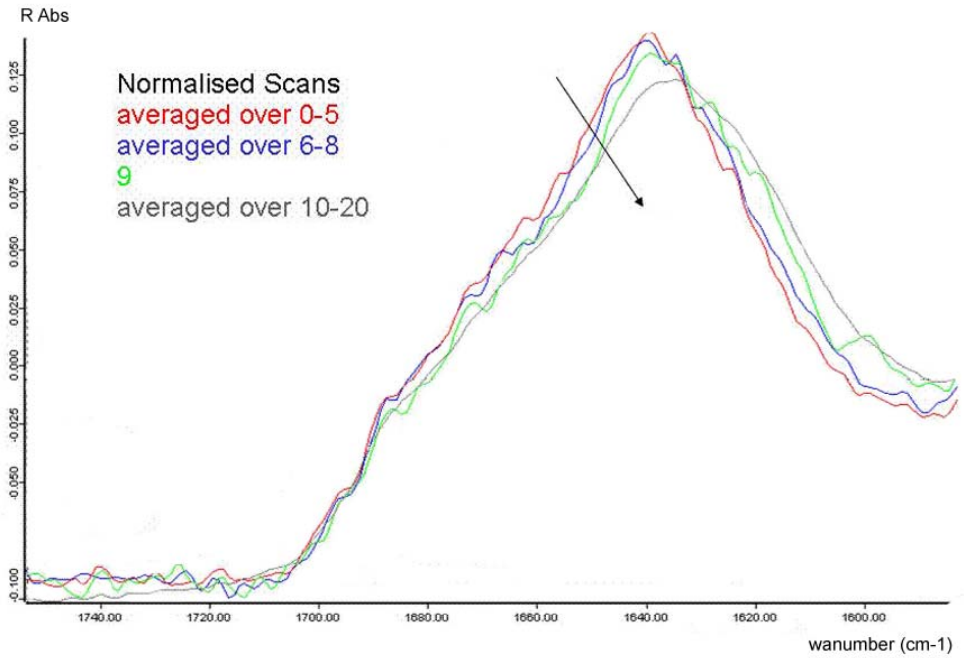


Figure 81: Infrared deconvolved spectra of LssH114A at different concentrations. Identical spectra were averaged: spectra 0-5 (red), spectra 6-8 (blue), spectra 9 (green), spectra 10-20 (grey).

From those observations, it seems that LssH114A is unstable at concentrations above 70 µM.

The infrared method was used to probe conformational transitions induced by ligand binding. To investigate if the peptidoglycan from *Staphylococcus aureus* was interacting with LssH114A and inducing conformational changes in the protein structure, 30 µM of protein was mixed with 0.14 mg.ml^{-1} peptidoglycan and compared with the protein alone. The spectra collected (Figure 82) were the result of 500 accumulations repeated twice, averaged and deconvolved.

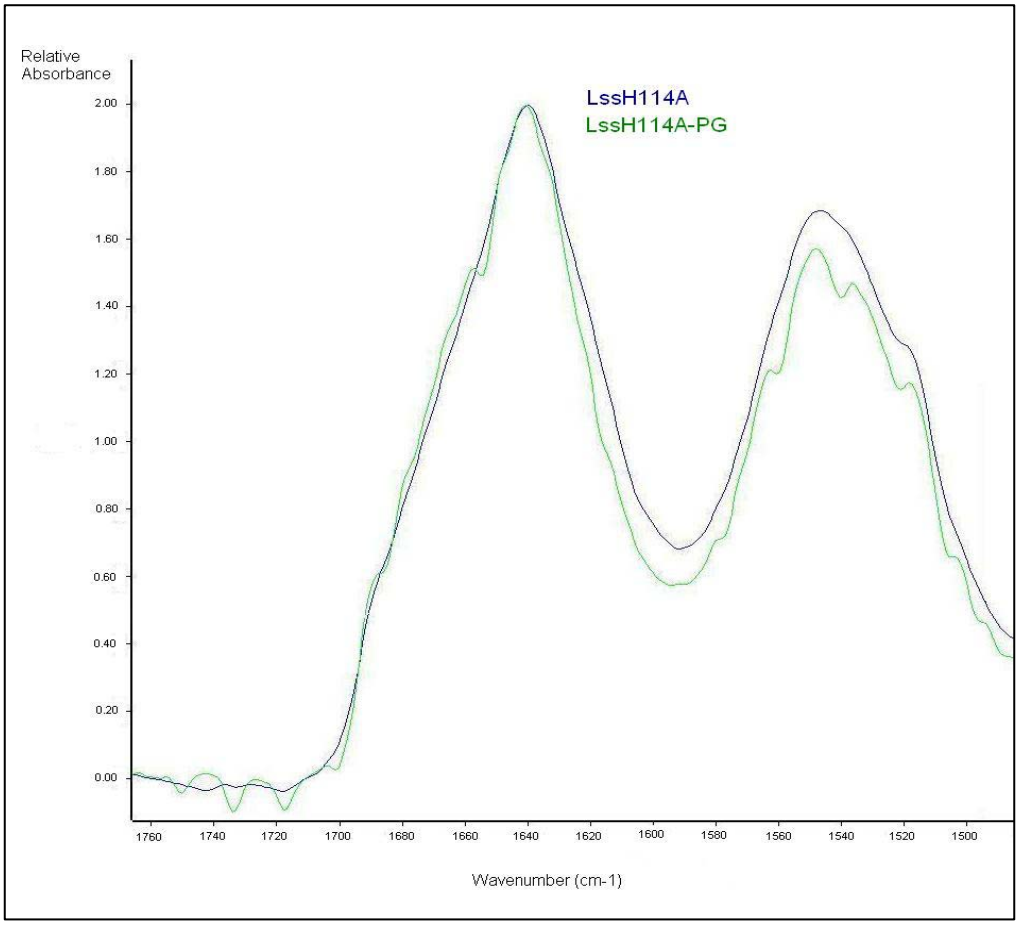


Figure 82: Overlaid infrared deconvolved spectra of LssH114A (blue) and LssH114A in solution with peptidoglycan from *S. aureus* (green).

Decomposed amide I band positions, percentage areas and assignment of the spectral components for LssH114A and LssH114A in solution with peptidoglycan from *S. aureus* are shown in Table 11.

Table 11: Infrared absorption band positions (in wavenumbers, cm^{-1}) determined with the aid of Fourier self deconvolution procedure for amide I for LssH114A and LssH114A in solution with peptidoglycan from *S. aureus*. Assignments are given based on the literature data (Byler and Susi, 1986, Goormaghtigh *et al.*, 1994). Italic numbers in brackets show tentative percentage of each structure obtained from curve fitting.

Protein	Turns	β-sheets	Unordered
LssH114A	1665 (*34*) 1688 (*2*)	1621 (*2*) 1624 (*47*)	1639 (*15*)
LssH114A + PG *S.aureus*	1665 (*41*) 1688 (*3*)	**_1618 (*5*)** 1637 (*51*)	

For LssH114A with added peptidoglycan from *S. aureus*, the curve at 1637 cm^{-1} is representative of β-sheet backbone conformation, the ones at 1665 and 1688 cm^{-1}, are representative of ordered features like turns. The weak band at 1618 cm^{-1} represents a tyrosine side chain vibration that may be enhanced by changes in conformation.

The shift in the spectral position of the β-sheets to higher wavenumber could indicate that there is more order for LssH114A mixed with peptidoglycan from *S. aureus* than for LssH114A alone, which fitting showed a contribution of 15% of weak β-sheet backbone conformation. The difference in the spectral shape from LssH114A alone and LssH114A in presence of peptidoglycan may represent an excess of unsubtracted substrate. It could also be that insoluble peptidoglycan is interacting with the protein and orientating it on the surface of the diamond, giving more order to the protein orientation, as seen for membrane proteins (Rothschild *et al.*, 1980).

To investigate if the pentaglycine (Gly_5) ligand was interacting with LssH114A, 45 μM of protein was mixed with 0.14 $mg.ml^{-1}$ of the ligand and compared with the protein alone using infrared. Concentration of LssH114A was performed by evaporation of the 45 μM solution on the diamond. Seven multiple scans were run with 25 accumulations each. The spectra collected are

shown in Figure 83. Concentrations were estimated by integration of the amide I band.

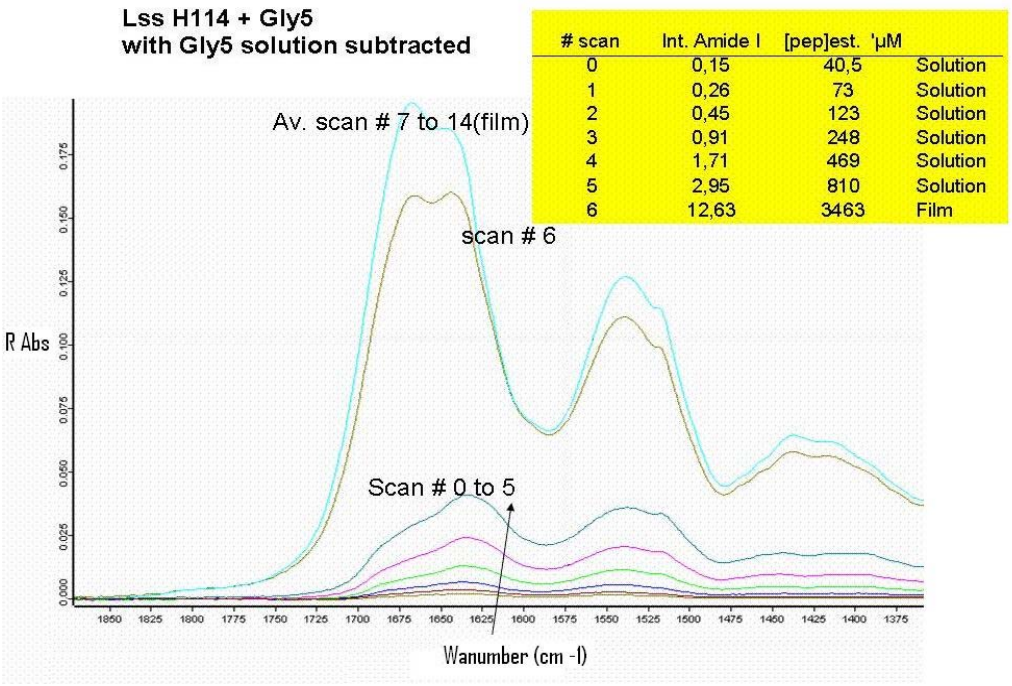


Figure 83: Infrared spectra of LssH114A in the presence of Gly_5 at different concentrations. The table shows an estimation of LssH114A concentration for the different spectra (i.e., 1-6).

Changes in conformation could be observed between 123 µM and 248 µM and above. When the protein was concentrated above 810 µM, forming a film, the sample thickness allowed greater interaction of the protein with the evanescent wave. However, the variations observed were compatible with a transformation of β-sheets into a compact form (amyloids), which were suggestive of an aggregation state (Halverson *et al.*, 1991, Ashburn *et al.*, 1992)

Figure 84 and Figure 85 show, in more detail, the infrared deconvolved spectra of LssH114A and LssH114A in presence of Gly_5; identical spectra were averaged for clarity. The spectra for LssH114A and of LssH114A in the presence of Gly_5 were superimposed to look for evidence of conformational changes.

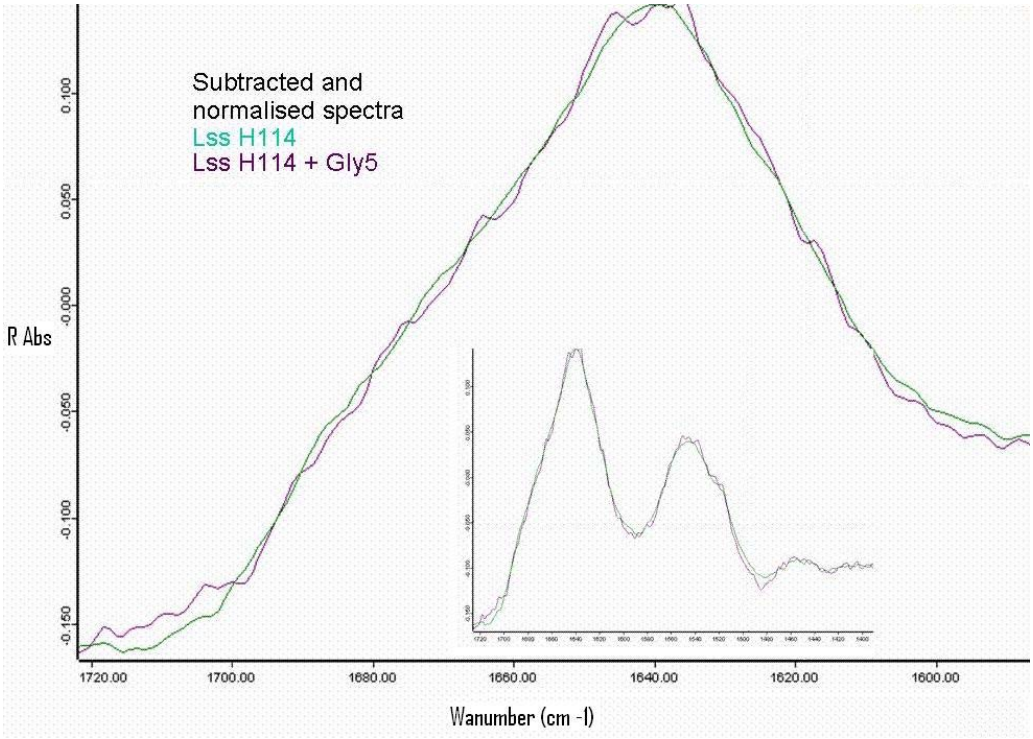


Figure 84: Overlaid infrared deconvolved spectra of LssH114A (blue) and LssH114A in solution with Gly_5 (purple). Identical spectra were averaged: spectra 0-5 for LssH114A alone and spectra 0-2 for LssH114A with Gly_5.

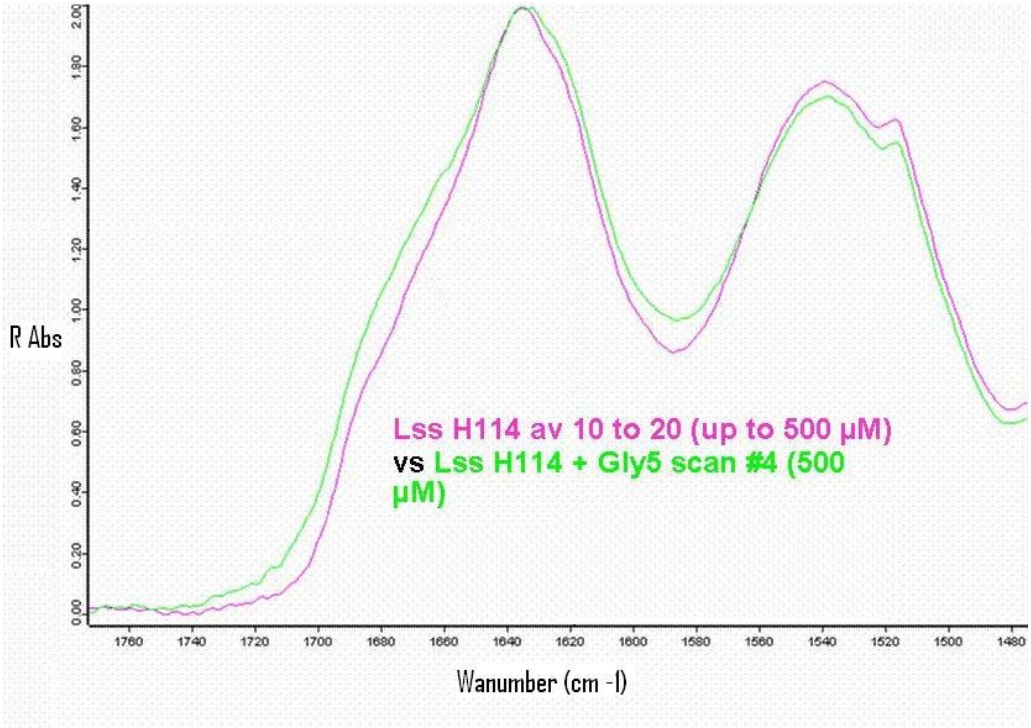
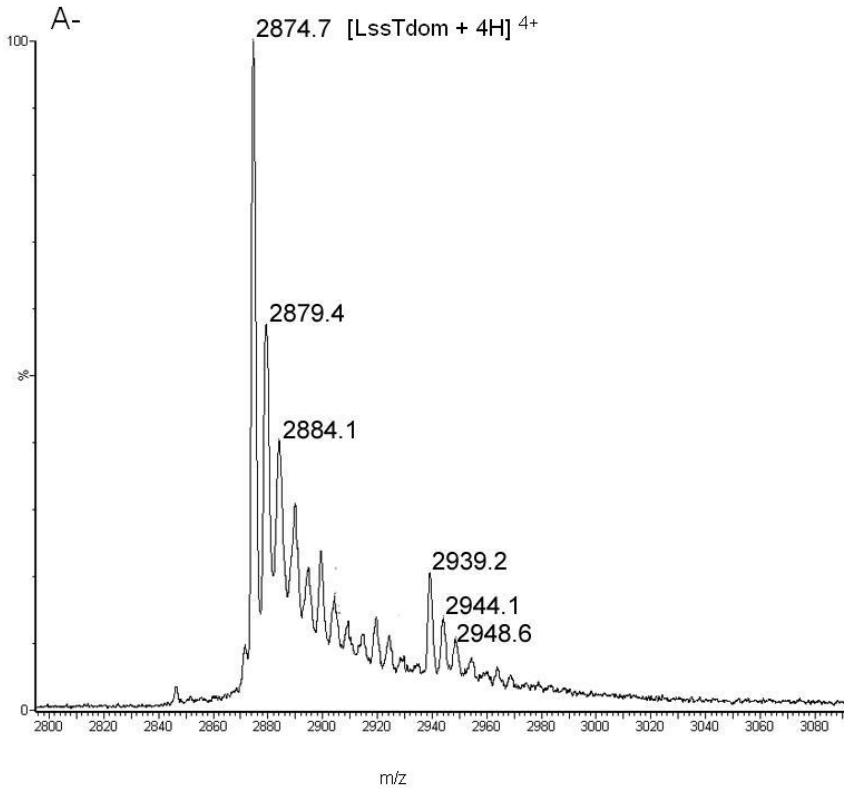


Figure 85: Overlaid infrared deconvolved spectra of LssH114A (pink) and LssH114A in solution with Gly_5 (green). Identical spectra were averaged: spectra 10-20 for LssH114A alone and spectra 4 for LssH114A with Gly_5.

The difference in the shape of the spectra from LssH114A alone and LssH114A in the presence of Gly_5 may represent an excess of unsubtracted substrate. It could also mean that Gly_5 is stabilising LssH114A, forming a complex.

5.2. Mass spectrometry

Analysis of LssTdom (5 μM) by ESI-MS, in triethylammonium acetate (25 mM, pH 7.4), gave a simple spectrum dominated by two peaks at *m/z* 2874.7 (77%) (Figure 86A) and 2300.0 (23%) (Figure 86B). These peaks correspond to the $[M+4H]^{4+}$ and $[M+5H]^{5+}$ ions of LssTdom respectively, where methionine has been cleaved from the *N*-terminal of the protein. (measured mass of LssTdom without Met +1 11495.0 Da; calculated mass of LssTdom without Met +1 11491.9 Da).



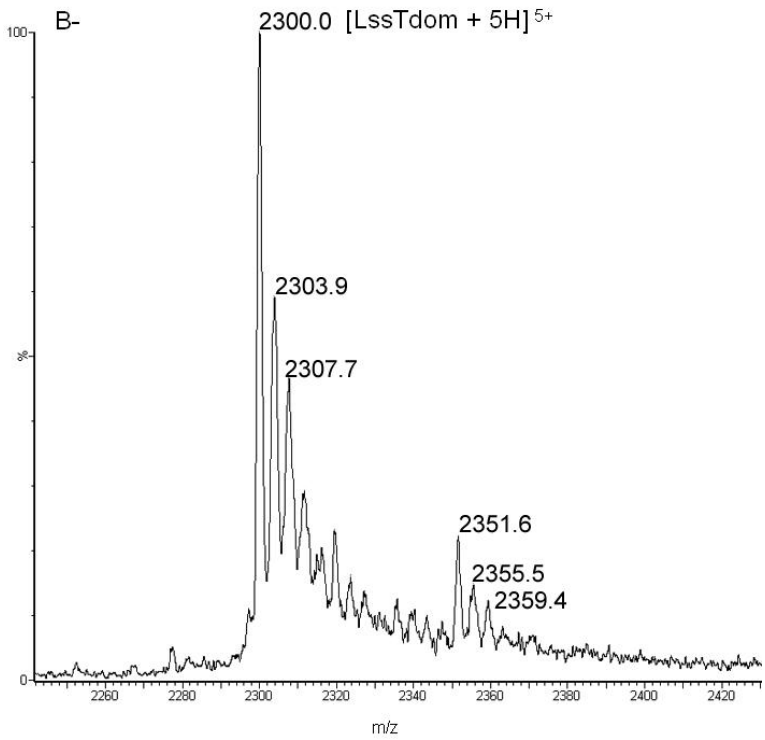


Figure 86: (A) ESI mass spectra of the $[M+4H]^{4+}$ charge state of the LssTdom showing the apoprotein signal at *m/z* 2874.7. **(B)** ESI mass spectra of the $[M+5H]^{5+}$ charge state of the LssTdom showing the apoprotein signal at *m/z* 2300.0.

This spectrum is complicated by adducts. The peaks at *m/z* 2879.4 and 2303.9 may be originated from addition of one ammonium ion (MW 18.04) and the ones at *m/z* 2884.1 and 2307.7 may be originated from addition of two ammonium ions (MW 36.08), on the apoprotein (LssTdom without Met +1). Distinctive peaks at *m/z* 2939.2 and 2351.6 could correspond to LssTdom still containing methionine on the *N*-terminus (Met +1) (calculated mass of LssTdom 11623.0 Da) with added triethylammonium ion (MW 102.2) and most likely a molecule of methanol (MW 32.0).

The peptide Gly_5 is mimicking the cross-bridge found in *S. aureus* peptidoglycan, a target of lysostaphin. Following the addition of 40 µM Gly_5 (G_5) (8 molar equivalents), an additional set of signals could be observed at *m/z* 2950.5 (Figure 87A) and 2360.6 (Figure 87B), corresponding to the

$[M+4H]^{4+}$ and $[M+5H]^{5+}$ charge states of LssTdom-G_5 non-covalent complex (measured mass of LssTdom-G_5 11796.1 Da; calculated mass of LssTdom-G_5 11795.2 Da).

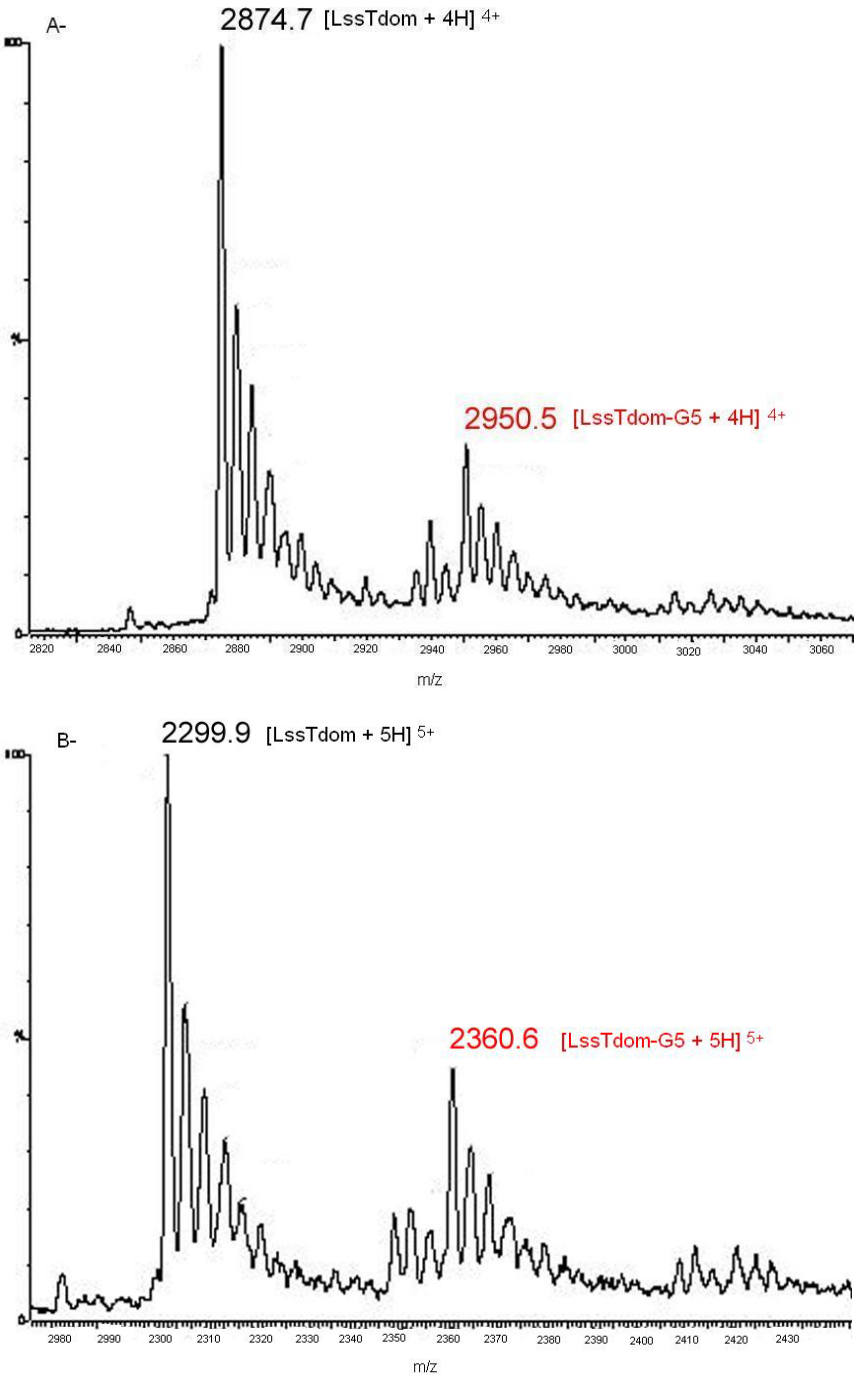


Figure 87: (A) ESI mass spectra of the $[M+4H]^{4+}$ charge state of the LssTdom showing the LssTdom-G_5 non-covalent complex signal at *m/z* 2950.5. **(B)** ESI mass spectra of the $[M+5H]^{5+}$ charge state of LssTdom showing the LssTdom-G_5 non-covalent complex signal at *m/z* 2360.6.

In order to observe the apparent dissociation constant K_D for LssTdom-G_5 complexes, a series of ESI-MS measurements were made at different concentrations of Gly_5 (ranging from 2 to 8 molar equivalents). Figure 88 A-D shows results for the LssTdom-G_5 complex.

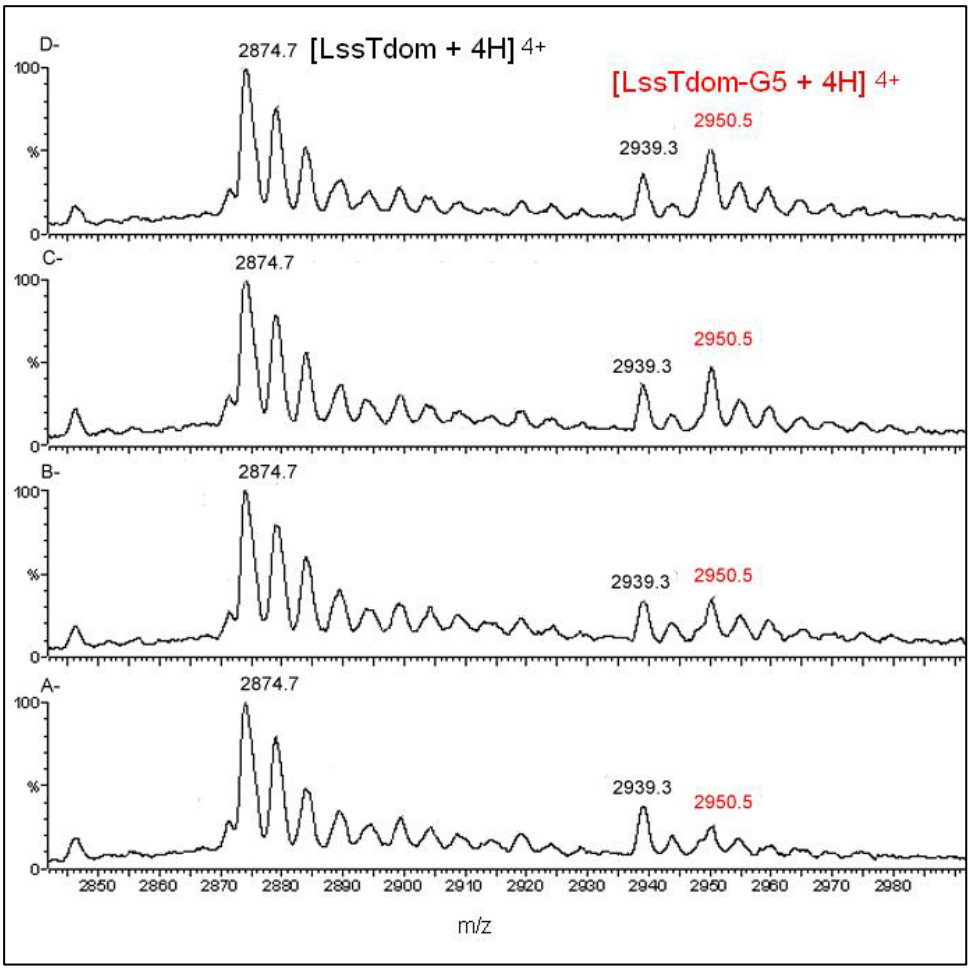


Figure 88: ESI mass spectra of the $[M+4H]^{4+}$ charge state of the LssTdom showing the LssTdom-G_5 non-covalent complex signal at *m/z* 2950.5 at different concentrations: **(A)** 2 molar equivalents, **(B)** 4 molar equivalents, **(C)** 6 molar equivalents, **(D)** 8 molar equivalents.

A smooth increase in the intensity of the ion at *m/z* 2950.5 and 2360.6, relative to *m/z* 2874.7 and 2299.9 respectively, is observed as the total ligand concentration is increased. Using the relative intensities of the signals in the mass spectrum (combined data for $[M+4H]^{4+}$ and $[M+5H]^{5+}$ charge states), together with the total protein and ligand concentrations, the values of protein

concentration free in solution [P], of protein-ligand complex concentration [PL] and of ligand concentration free in solution [L] were calculated (Table 12). The dissociation constant K_D was determined from plots of [PL]/[L] vs. [L] (Figure 89). Gly_5 bound to LssTdom with a dissociation constant of 9 (± 1) μM.

Table 12: Values for [P] (protein concentration), [L] (ligand concentration) and [PL] (protein-ligand complex).

[L] (μM)	[P] (μM)	[PL] (μM)
8.93	3.93	1.07
18.63	3.63	1.37
28.39	3.31	1.69
38.23	3.23	1.77

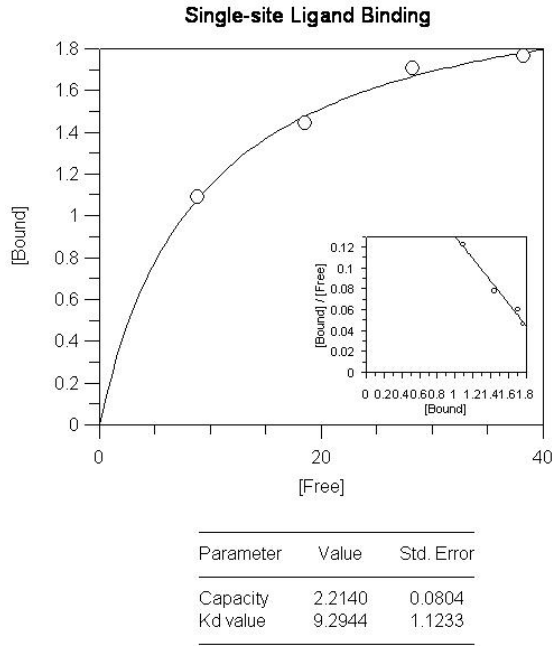


Figure 89: GraFit4 plot of the LssTdom-G_5/free ligand concentrations vs. free ligand concentrations giving a K_D value of 9 (± 1) μM (Chi^2 = 0.0014).

The peptide Gly-Gly-L-Ser-Gly-L-Ser is mimicking the cross-bridge found in lysostaphin host bacteria peptidoglycan – *S. simulans*. Addition of 40 μM Gly-Gly-L-Ser-Gly-L-Ser (G_2SGS) (8 molar equivalents) to the LssTdom solution could be observed at *m/z* 2965.5 (Figure 90A) and 2372.5 (Figure 90B), corresponding to the $[M+4H]^{4+}$ and $[M+5H]^{5+}$ charge states of

LssTdom-G_2SGS non-covalent complex (measured mass of LssTdom-G_2SGS 11857.7 Da; calculated mass of LssTdom-G_2SGS 11855.3 Da).

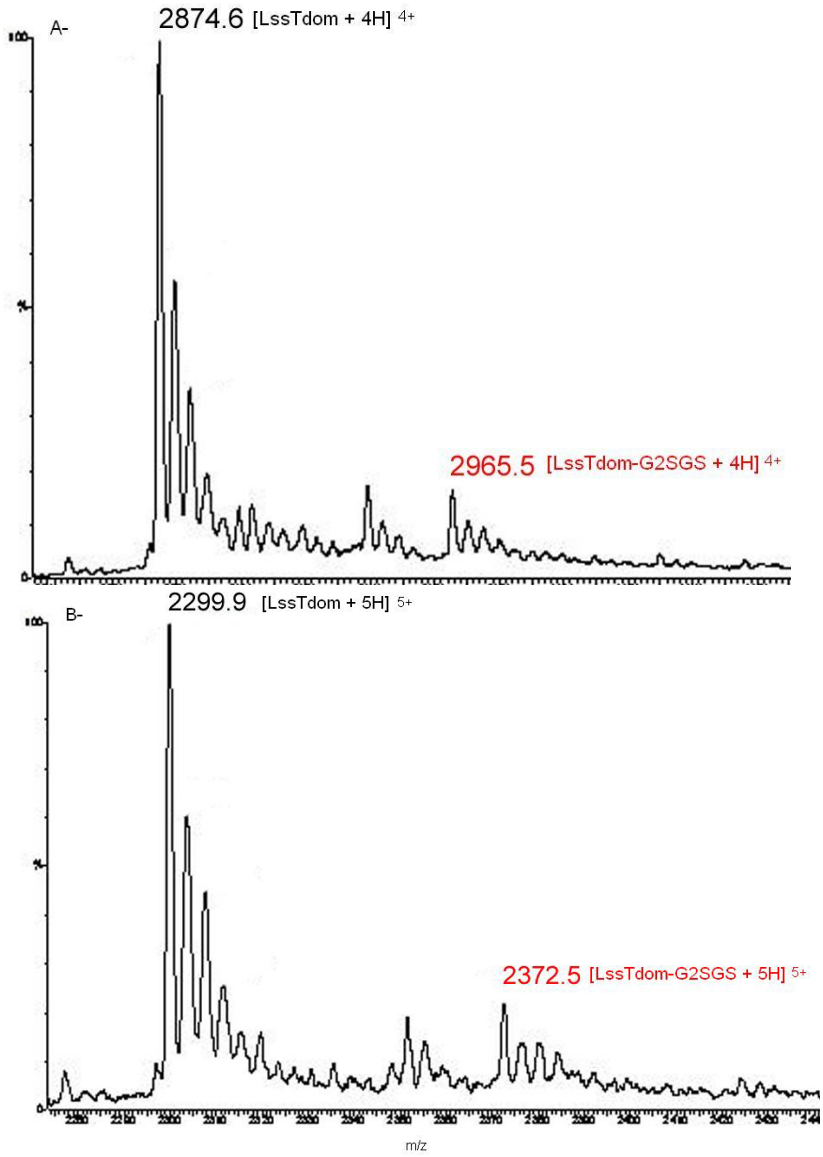


Figure 90: (A) ESI mass spectra of the $[M+4H]^{4+}$ charge state of LssTdom showing the LssTdom-G_2SGS non-covalent complex signal at *m/z* 2965.5. **(B)** ESI mass spectra of the $[M+5H]^{5+}$ charge state of the LssTdom showing the LssTdom- G_2SGS non-covalent complex signal at *m/z* 2372.5.

In order to observe the apparent dissociation constant K_D for LssTdom-G_2SGS complexes, a series of ESI-MS measurements were made at different

concentrations of Gly-Gly-L-Ser-Gly-L-Ser (ranging from 2 to 8 molar equivalents). Figure 91 A-D shows results for the LssTdom-G_2SGS complex.

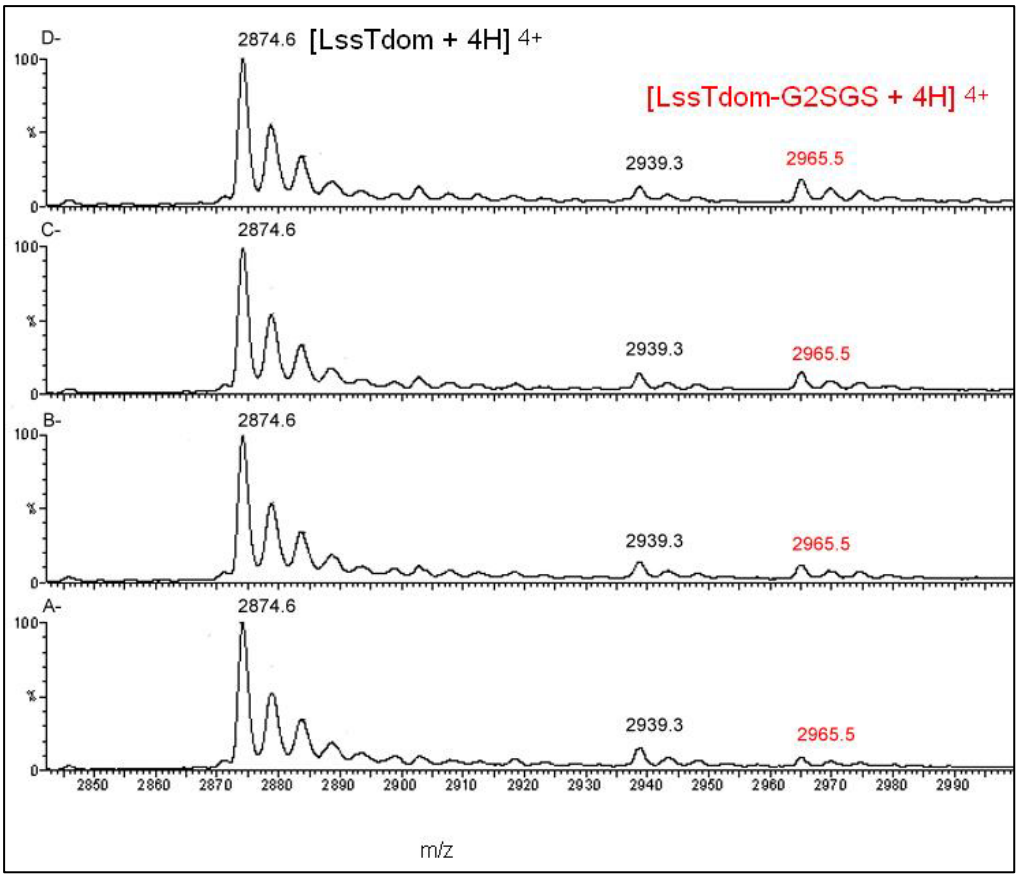


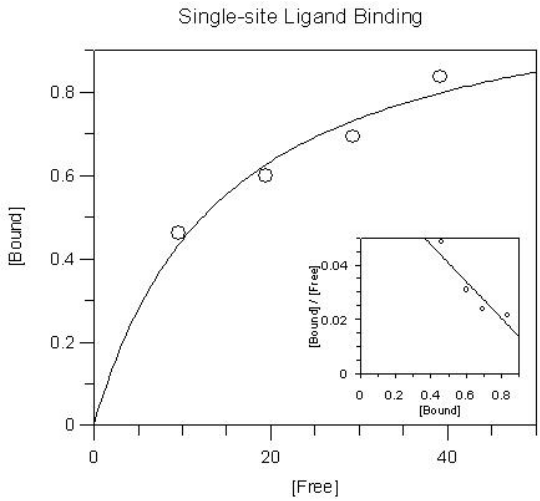
Figure 91: ESI mass spectra of the $[M+4H]^{4+}$ charge state of the LssTdom showing the LssTdom-G_2SGS non-covalent complex signal at m/z 2965.5 at different concentrations: **(A)** 2 molar equivalents, **(B)** 4 molar equivalents, **(C)** 6 molar equivalents, **(D)** 8 molar equivalents.

A smooth increase in the intensity of the ion at *m/z* 2965.5 and 2372.5, relative to *m/z* 2874.6 and 2299.9 respectively, is observed as the total ligand concentration is increased. Using the relative intensities of the signals in the mass spectrum (combined data for $[M+4H]^{4+}$ and $[M+5H]^{5+}$ charge states), together with the total protein and ligand concentrations, the values of protein concentration free in solution [P], of protein-ligand complex concentration [PL] and of ligand concentration free in solution [L] were calculated (Table 13). The dissociation constant K_D was determined from plots of [PL]/[L] vs.

[L] (Figure 92). Gly-Gly-L-Ser-Gly-L-Ser bound to LssTdom with a dissociation constant of 15 (± 5) μM.

Table 13: Values for [P] (protein concentration), [L] (ligand concentration) and [PL] (protein-ligand complex).

[L] (μM)	[P] (μM)	[PL] (μM)
9.56	4.56	0.44
19.42	4.41	0.58
29.32	4.32	0.68
39.18	4.18	0.82



Parameter	Value	Std. Error
Capacity	1.0970	0.1419
Kd value	14.6688	5.0379

Figure 92: Plot of the LssTdom-G_2SGS/free ligand concentrations vs. free ligand concentrations giving a K_D value of 15 (± 5) μM (Chi^2 = 0.0022).

A competition experiment between LssTdom and 8 molar equivalents of Gly_5 and Gly-Gly-L-Ser-Gly-L-Ser showed both receptor-ligand complexes at *m/z* 2950.3 and 2965.5 respectively for the $[M+4H]^{4+}$ charge state (Figure 93D). Peak intensities revealed that the LssTdom-G_5 complex was 1.3 times more abundant than LssTdom-G_2SGS, in agreement with the above mentioned K_D values (i.e., 9.3 μM for the LssTdom-G_5 complex and 14.7 μM for the LssTdom-G_2SGS complex). The competitive binding experiment measured

the binding of a single concentration of Gly_5 ligand (40 μM) in the presence of various concentrations of G_2SGS ligand (10 to 80 μM) (Figure 93 A-H).

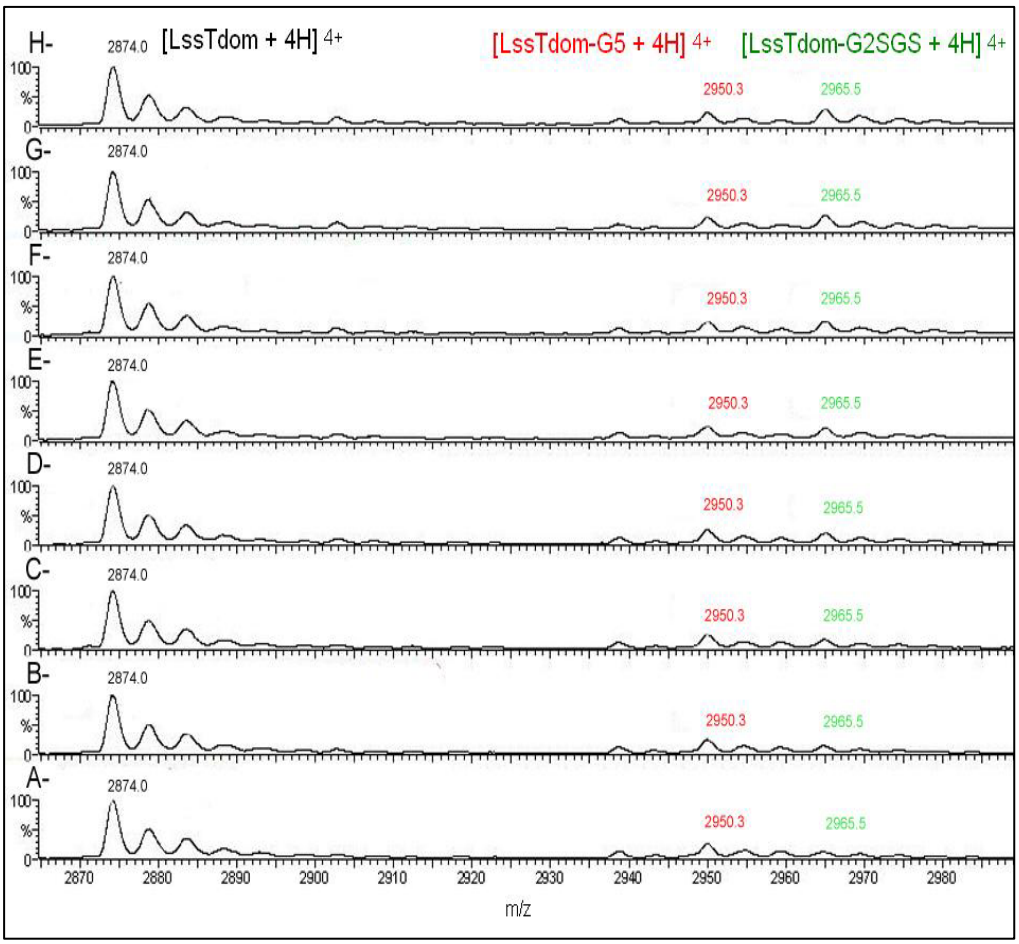


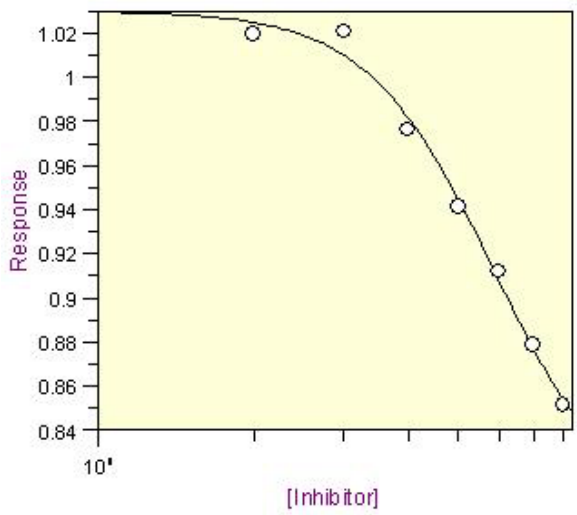
Figure 93: ESI mass spectra of the $[M+4H]^{4+}$ charge state of the LssTdom showing the LssTdom-Gly_5 non-covalent complex at m/z 2950.3 at fixed concentration (8 molar equivalents) and the competitive LssTdom-G_2SGS non-covalent complex signal at *m/z* 2965.5 at different concentrations: **(A)** 2 molar equivalents, **(B)** 4 molar equivalents, **(C)** 6 molar equivalents, **(D)** 8 molar equivalents, **(E)** 10 molar equivalents, **(F)** 12 molar equivalents, **(G)** 14 molar equivalents, **(H)** 16 molar equivalents.

A smooth increase in the intensity of the ion at *m/z* 2965.5 and 2372.5, and a smooth decrease in the intensity of the ion at *m/z* 2950.3 and 2360.3, relative to *m/z* 2874.0 and 2299.8 respectively, is observed as G_2SGS ligand concentration is increased. Using the relative intensities of the signals in the mass spectrum (combined data for $[M+4H]^{4+}$ and $[M+5H]^{5+}$ charge states), together with the total protein and ligand concentrations, the value of

[LssTdom-G_5] was calculated (Table 14). The median inhibition concentration (IC_{50}) which is the concentration of G_2SGS that reduces the LssTdom-G_5 non-covalent complex formation by 50%, was determined from plots of log[G_2SGS] concentrations vs [LssTdom-G_5] concentrations (Figure 94). The concentration of G_2SGS that reduces Gly_5 binding by half was 59 μM (± 9 μM), which is very close to [Gly_5] + K_D (i.e., 40 μM + 9 μM (± 1 μM)).

Table 14: Values for [inhibitor] ([G_2SGS] concentration) and response ([LssTdom-G_5]).

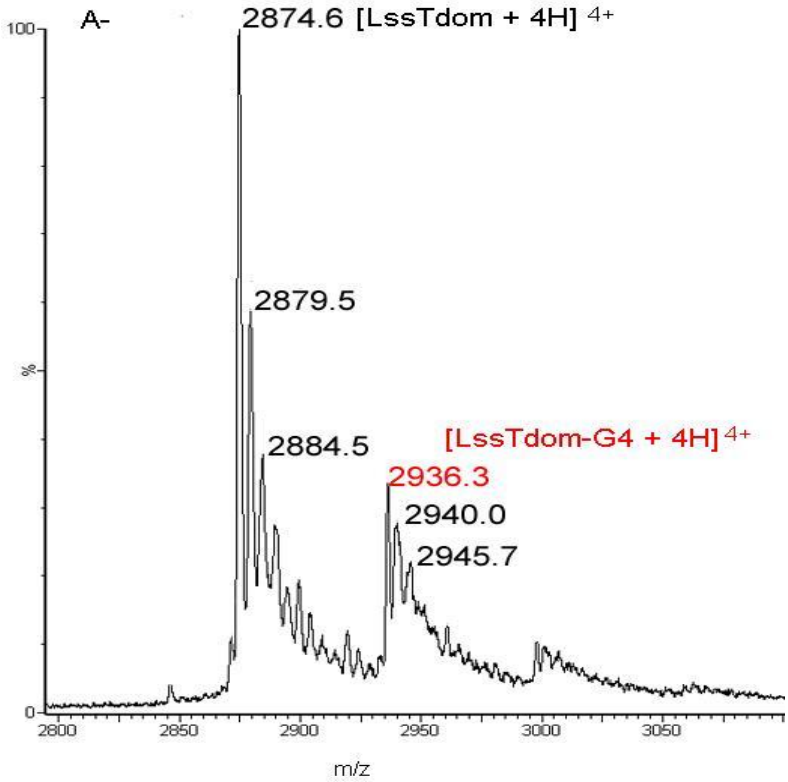
[inhibitor] (=[G_2SGS]) (μM)	Response (=[LssTdom-Gly_5]) (μM))
20	1.0195
30	1.0210
40	0.9763
50	0.9412
60	0.9120
70	0.8788
80	0.8516



Parameter	Value	Std. Error
Y Range	0.2352	0.0684
IC 50	58.7021	8.6433
Slope factor	3.5336	1.2497
Background	0.7948	0.0596

Figure 94: Plot of log[G_2SGS] concentrations vs [LssTdom-G_5] concentrations giving an IC_{50} value of 59 (± 9) μM.

Following the addition of 40 μM Gly_4 (G_4) (8 molar equivalents) to the LssTdom solution, two additional peaks could be observed at *m/z* 2936.3 (Figure 95A) and 2349.1 (Figure 95B), corresponding to the $[M+4H]^{4+}$ and $[M+5H]^{5+}$ charge states of LssTdom-G_4 non-covalent complex (measured mass of LssTdom-G_4 11740.9 Da; calculated mass of LssTdom-G_4 11738.1 Da).



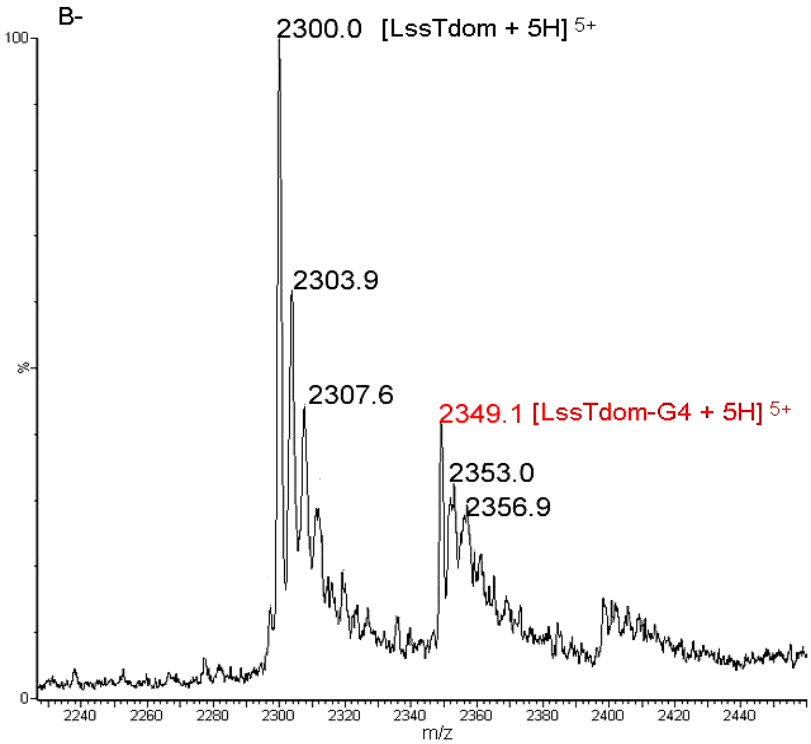


Figure 95: (A) ESI mass spectra of the $[M+4H]^{4+}$ charge state of the LssTdom showing the LssTdom-G_4 non-covalent complex signal at *m/z* 2936.3. **(B)** ESI mass spectra of the $[M+5H]^{5+}$ charge state of the LssTdom showing the LssTdom-G_4 non-covalent complex signal at *m/z* 2349.1.

The additional peaks at *m/z* 2939.2 and 2351.6 that are present in LssTdom spectrum when not associated with Gly_4 may be hidden and complicate the accurate measurement of the intensities of the peaks at *m/z* 2936.3 and 2349.1, corresponding to the $[M+4H]^{4+}$ and $[M+5H]^{5+}$ charge states of LssTdom-G_4 complex. Nonetheless, in order to observe the apparent dissociation constant K_D for LssTdom-G_4 complexes, a series of ESI-MS measurements were made at different concentrations of Gly_4 (ranging from 2 to 16 molar equivalents). Figure 96 A-E shows results for LssTdom-G_4 complex.

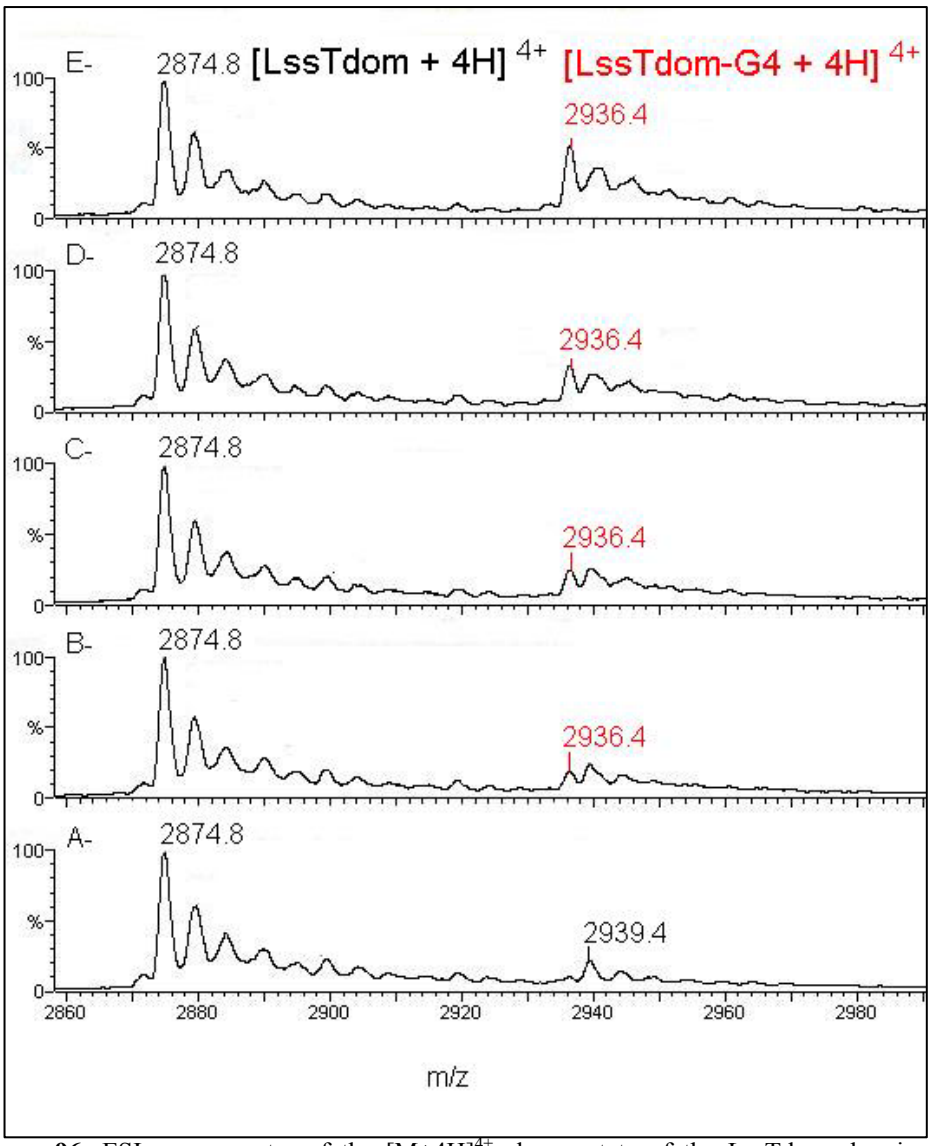


Figure 96: ESI mass spectra of the $[M+4H]^{4+}$ charge state of the LssTdom showing the LssTdom-G_4 non-covalent complex signal at *m/z* 2936.3 at different concentrations: **(A)** 2 molar equivalents, **(B)** 4 molar equivalents, **(C)** 6 molar equivalents, **(D)** 8 molar equivalents, **(E)** 16 molar equivalents.

A smooth increase in the intensity of the ion at *m/z* 2936.3 and 2349.1, relative to *m/z* 2874.7 and 2300.0 respectively, is observed as the total ligand concentration is increased. Using the relative intensities of the signals in the mass spectrum (combined data for $[M+4H]^{4+}$ and $[M+5H]^{5+}$ charge states), together with the total protein and ligand concentrations, the values of protein concentration free in solution [P], of protein-ligand complex concentration [PL] and of ligand concentration free in solution [L] were calculated (Table

15). The dissociation constant K_D was determined from plots of [PL]/[L] vs. [L] (Figure 97). Gly_4 bound to LssTdom with a dissociation constant of 38 (± 5) μM.

Table 15: Values for [P] (protein concentration), [L] (ligand concentration) and [PL] (protein-ligand complex).

[L] (μM)	[P] (μM)	[PL] (μM)
9.49	4.49	0.51
19.17	4.17	0.83
28.96	3.96	1.04
38.74	3.24	1.26

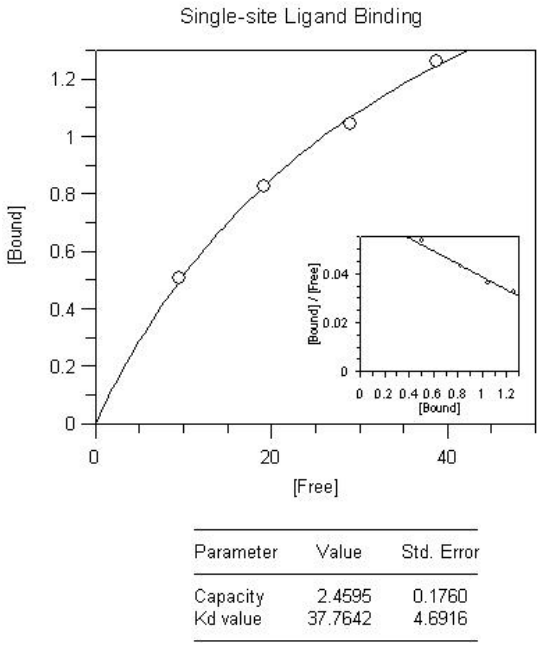


Figure 97: GraFit4 plot of the LssTdom-G_4/free ligand concentrations vs. free ligand concentrations giving a K_D value of 38 (± 5) μM (Chi^2 = 0.0004).

Following the addition of 40 μM Gly_2 (G_2) (8 molar equivalents) to the LssTdom solution, two additional peaks could be observed at *m/z* 2906.6 (Figure 98A) and 2325.6 (Figure 98B), corresponding to the $[M+4H]^{4+}$ and $[M+5H]^{5+}$ charge states of LssTdom-G_2 non-covalent complex (measured mass of LssTdom-G_2 11622.4 Da; calculated mass of LssTdom-G_2 11624.0 Da).

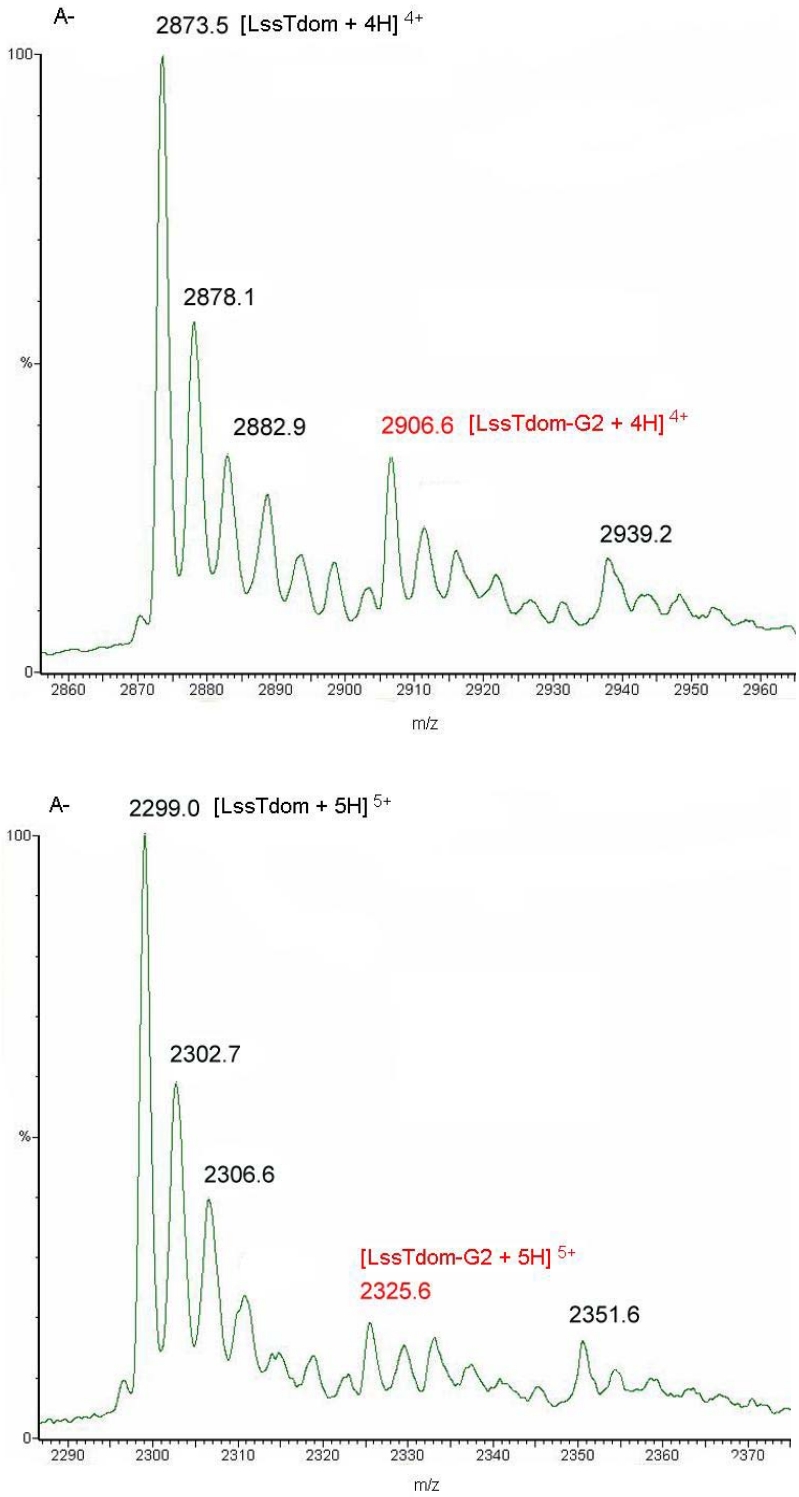


Figure 98: (A) ESI mass spectra of the $[M+4H]^{4+}$ charge state of the LssTdom showing the LssTdom-G_2 non-covalent complex signal at *m/z* 2906.6. **(B)** ESI mass spectra of the $[M+5H]^{5+}$ charge state of the LssTdom showing the LssTdom-G_2 non-covalent complex signal at *m/z* 2325.6.

In order to observe the apparent dissociation constant K_D for LssTdom-G_2 complexes, a series of ESI-MS measurements were made at different concentrations of Gly_2 (ranging from 2 to 12 molar equivalents). Figure 99 A-G shows results for LssTdom-G_2 complex.

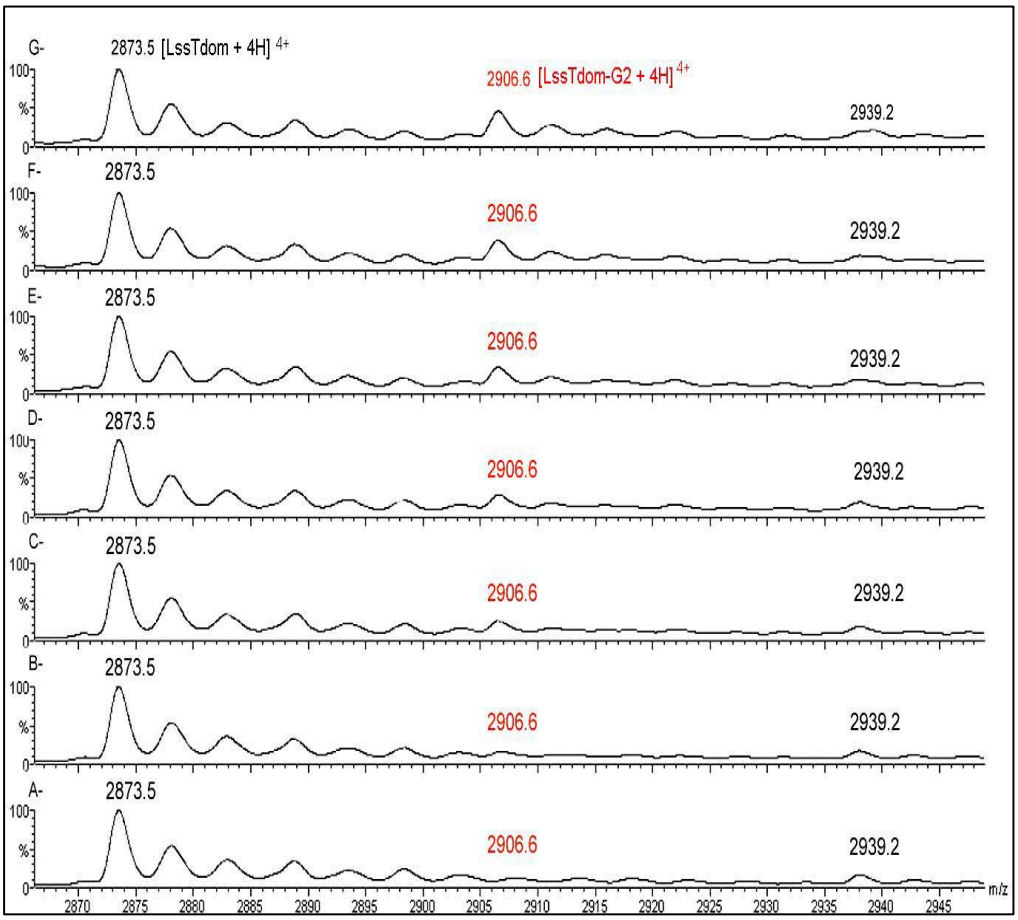


Figure 99: ESI mass spectra of the $[M+4H]^{4+}$ charge state of the LssTdom showing the LssTdom-G_2 non-covalent complex signal at *m/z* 2906.6 at different concentrations: **(A)** no ligand, **(B)** 2 molar equivalents, **(C)** 4 molar equivalents, **(D)** 6 molar equivalents, **(E)** 8 molar equivalents, **(F)** 10 molar equivalents, **(G)** 12 molar equivalents.

A smooth increase in the intensity of the ion at *m/z* 2906.6 and 2325.6, relative to *m/z* 2873.5 and 2299.0 respectively, is observed as the total ligand concentration is increased. Using the relative intensities of the signals in the mass spectrum (combined data for $[M+4H]^{4+}$ and $[M+5H]^{5+}$ charge states), together with the total protein and ligand concentrations, the values of protein concentration free in solution [P], of protein-ligand complex concentration

[PL] and of ligand concentration free in solution [L] were calculated (Table 16). The dissociation constant K_D was determined from plots of [PL]/[L] vs. [L] (Figure 100). Gly_2 bound to LssTdom with a dissociation constant of 14 (± 2) µM.

Table 16: Values for [P] (protein concentration), [L] (ligand concentration) and [PL] (protein-ligand complex).

[L] (µM)	[P] (µM)	[PL] (µM)
9.35	4.35	0.65
19.08	4.08	0.92
28.97	3.97	1.03
38.85	3.85	1.15
48.73	3.73	1.27

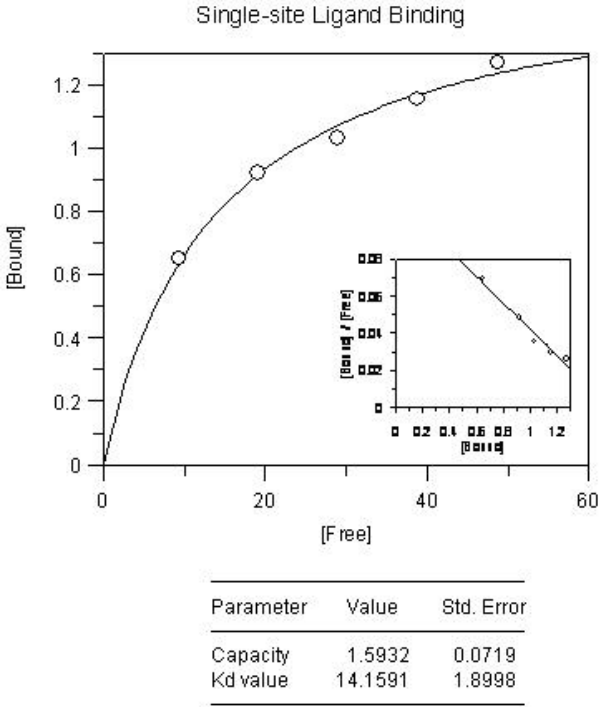


Figure 100: GraFit4 plot of the LssTdom-G_2/free ligand concentrations vs. free ligand concentrations giving a K_D value of 14 (± 2) µM (Chi^2 = 0.0011).

The peptide acetyl-Lys-D-Ala-D-Ala is mimicking part of the stem peptide found in *S. aureus* peptidoglycan. Following addition of 40 µM acetyl-Lys-D-Ala-D-Ala (KA_2) (8 molar equivalents) to the LssTdom solution, two additional peaks could be observed at *m/z* 2957.4 (Figure 101A) and 2365.9 (Figure 101B), corresponding to the $[M+4H]^{4+}$ and $[M+5H]^{5+}$ charge states of

LssTdom-KA_2 non-covalent complex (measured mass of LssTdom-KA_2 11825.0 Da; calculated mass of LssTdom-KA_2 11822.2 Da).

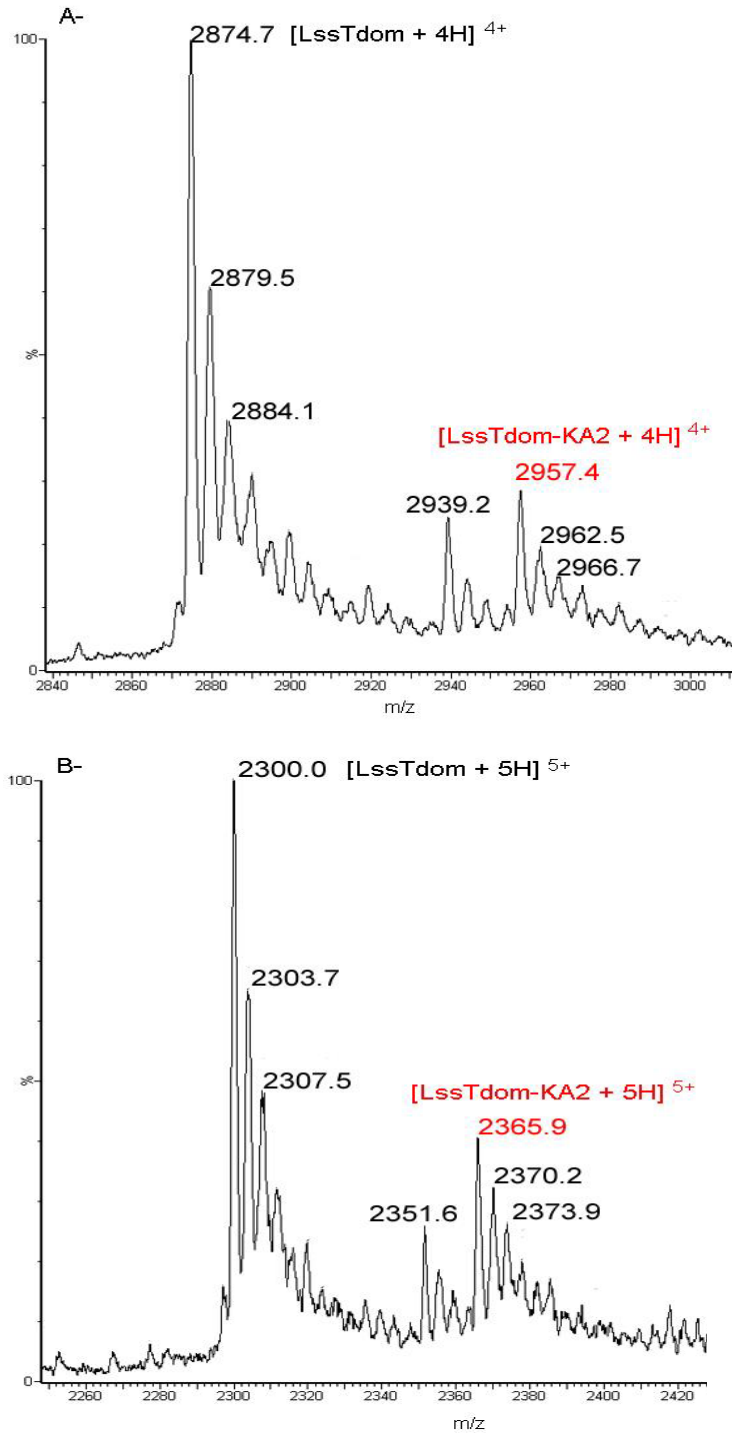


Figure 101: (A) ESI mass spectra of the $[M+4H]^{4+}$ charge state of LssTdom showing the LssTdom-KA_2 non-covalent complex signal at *m/z* 2957.4. **(B)** ESI mass spectra of the $[M+5H]^{5+}$ charge state of the LssTdom showing the LssTdom-KA_2 non-covalent complex signal at *m/z* 2365.9.

In order to observe the apparent dissociation constant K_D for LssTdom-KA_2 complexes, a series of ESI-MS measurements were made at different concentrations of acetyl-Lys-D-Ala-D-Ala (ranging from 2 to 8 molar equivalents). Figure 102 A-D shows results for LssTdom-KA_2 complex.

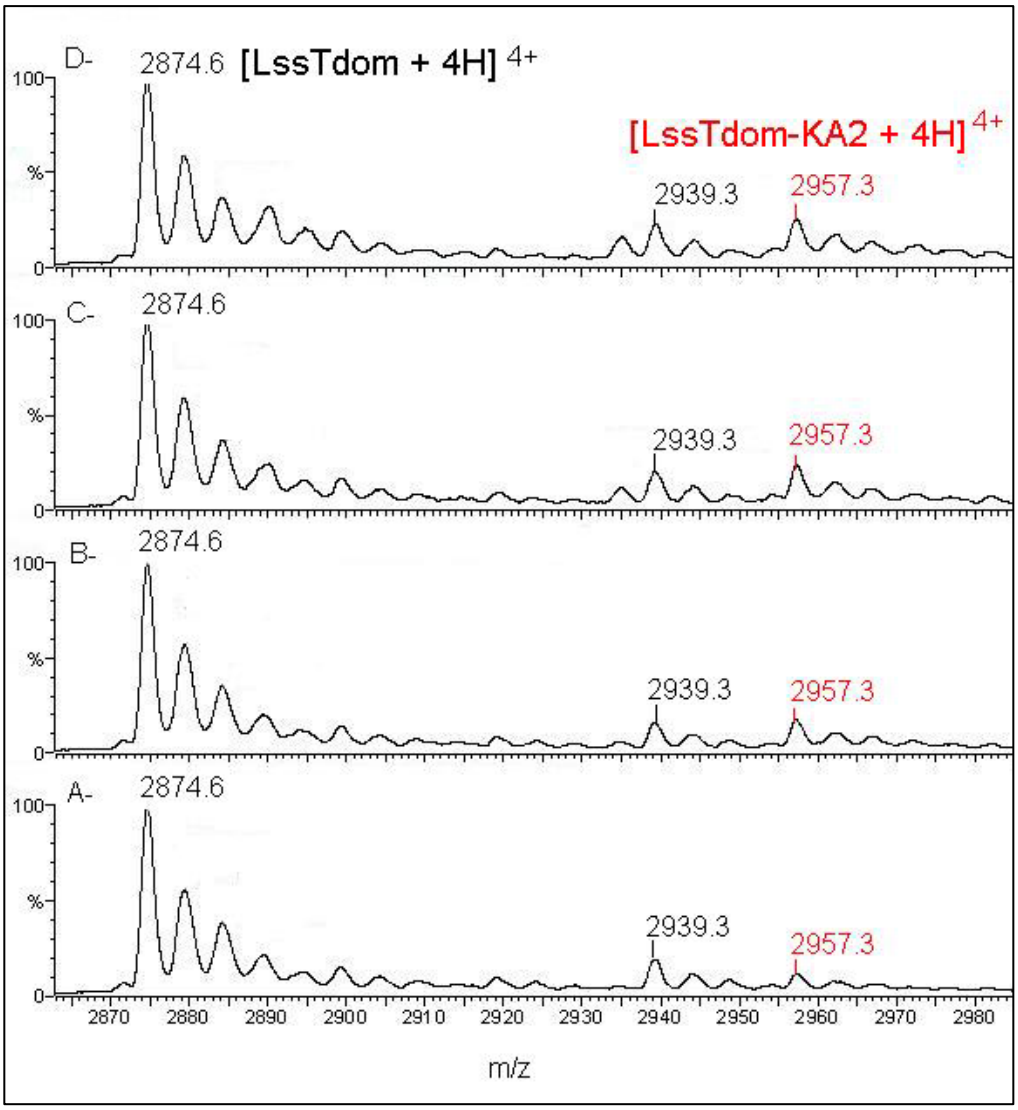


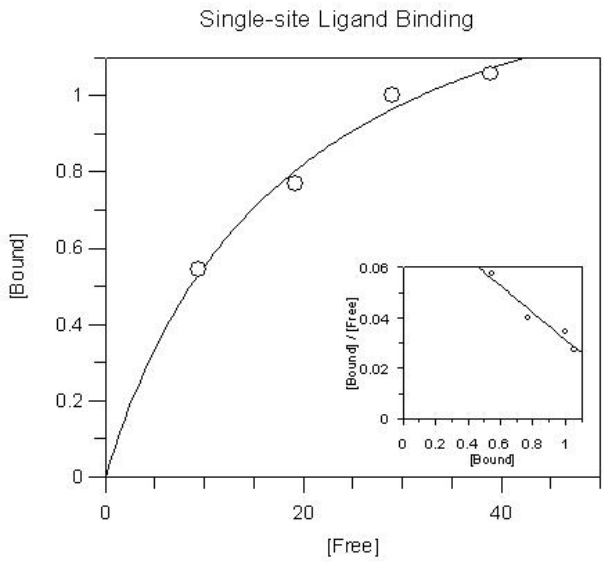
Figure 102: ESI mass spectra $[M+4H]^{4+}$ charge state of the LssTdom showing the LssTdom-KA_2 non-covalent complex signal at m/z 2957.0 at different concentrations: **(A)** 2 molar equivalents, **(B)** 4 molar equivalents, **(C)** 6 molar equivalents, **(D)** 8 molar equivalents.

A smooth increase in the intensity of the ion at *m/z* 2957.0 and 2365.7, relative to *m/z* 2874.3 and 2299.6 respectively, is observed as the total ligand concentration is increased. Using the relative intensities of the signals in the

mass spectrum (combined data for $[M+4H]^{4+}$ and $[M+5H]^{5+}$ charge states), together with the total protein and ligand concentrations, the values of protein concentration free in solution [P], of protein-ligand complex concentration [PL] and of ligand concentration free in solution [L] were calculated (Table 17). The dissociation constant K_D was determined from plots of [PL]/[L] vs. [L] (Figure 103). acetyl-Lys-D-Ala-D-Ala bound to LssTdom with a dissociation constant of 19 (± 4) µM.

Table 17: Values for [P] (protein concentration), [L] (ligand concentration) and [PL] (protein-ligand complex).

[L] (µM)	[P] (µM)	[PL] (µM)
9.46	4.46	0.54
19.23	4.23	0.77
29.00	4.00	1.00
38.94	3.94	1.06



Parameter	Value	Std. Error
Capacity	1.5850	0.1479
Kd value	18.6939	4.0941

Figure 103: Plot of the LssTdom-KA_2/free ligand concentrations vs. free ligand concentrations giving a K_D value of 19 (± 4) µM (Chi^2 = 0.0015).

A competition experiment between LssTdom and 8 molar equivalents of Gly_2 and acetyl-Lys-D-Ala-D-Ala showed both receptor-ligand complexes at *m/z* 2906.6 and 2956.1 respectively for the $[M+4H]^{4+}$ charge state (Figure 104 D). Peak intensities revealed that LssTdom-G_2 complex was equivalently abundant to LssTdom-KA_2, in agreement with the above mentioned K_D values (i.e., 14.2 μM for LssTdom-G_2 complex and 18.7 μM for LssTdom-KA_2 complex). The competitive binding experiment measured the binding of a single concentration of Gly_2 ligand (40 μM) in the presence of various concentrations of KA_2 ligand (10 to 100 μM) (Figure 104 A-I).

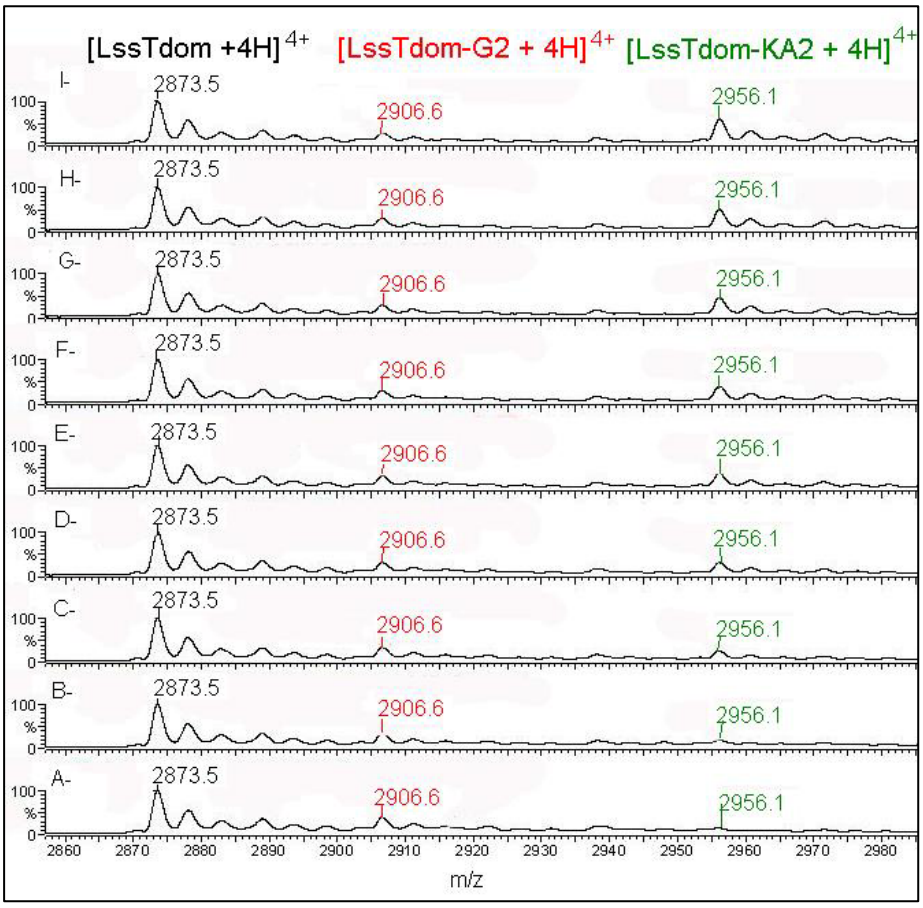
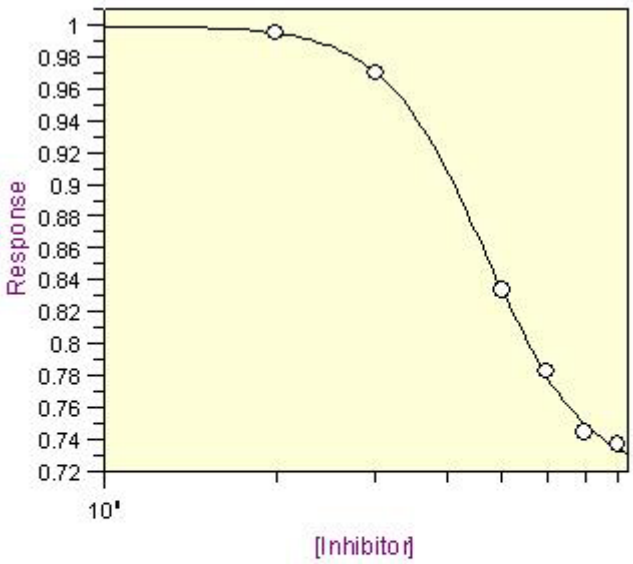


Figure 104: ESI mass spectra of the $[M+4H]^{4+}$ charge state of the LssTdom showing the LssTdom-Gly_2 non-covalent complex at m/z 2906.6 at fixed concentration (8 molar equivalent) and the competitive LssTdom-KA_2 non-covalent complex signal at *m/z* 2956.1 at different concentrations: **(A)** 2 molar equivalents, **(B)** 4 molar equivalents, **(C)** 6 molar equivalents, **(D)** 8 molar equivalents, **(E)** 10 molar equivalents, **(F)** 12 molar equivalents, **(G)** 14 molar equivalents, **(H)** 16 molar equivalents, **(I)** 20 molar equivalents.

A smooth increase in the intensity of the ion at *m/z* 2956.1 and 2365.9, and a smooth decrease in the intensity of the ion at *m/z* 2906.6 and 2325.6, relative to *m/z* 2873.5 and 2299.8 respectively, is observed as the amount of KA_2 ligand concentration is increased. Using the relative intensities of the signals in the mass spectrum (combined data for $[M+4H]^{4+}$ and $[M+5H]^{5+}$ charge states), together with the total protein and ligand concentrations, the value of [LssTdom-G_2] was calculated (Table 18). The median inhibition concentration (IC_{50}) which is the concentration that reduces the LssTdom-G_2 non-covalent complex formation by 50%, was determined from plots of log[KA_2] concentrations vs [LssTdom-G_2] concentrations (Figure 105). The concentration of KA_2 that reduces Gly_2 binding by half was 47 μM (± 1 μM), which is very close to [Gly_2] + K_D (i.e., 40 μM + 14 μM (± 2 μM)).

Table 18: Values for [inhibitor] ([KA_2] concentration) and response ([LssTdom-G_2]).

[inhibitor] (=[G_2SGS]) (μM)	Response (=[LssTdom-Gly_5]) (μM))
20	0.9949
30	0.9696
50	0.8338
60	0.7826
70	0.7444
80	0.7367

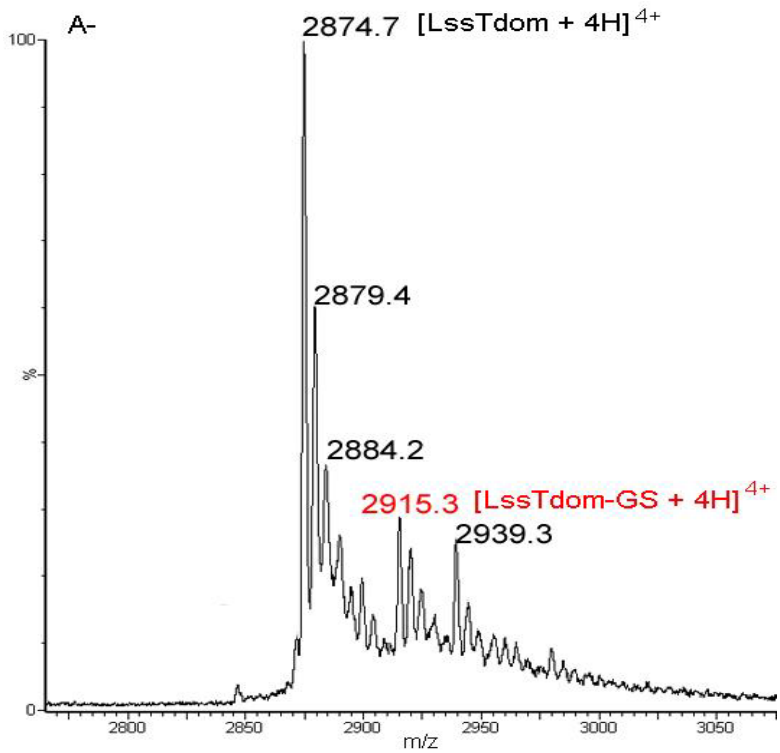


Parameter	Value	Std. Error
Y Range	0.2845	0.0134
IC 50	46.8302	0.9041
Slope factor	4.8827	0.5422
Background	0.7145	0.0091

Figure 105: Plot of log[KA_2] concentrations vs [LssTdom-G_2] concentrations giving an IC_{50} value of 47 (± 1) µM.

This very same set of experiment was run with LssTdom in complex with 8 molar equivalent Gly_5, competing for the same binding site with different concentrations of acetyl-Lys-D-Ala-D-Ala (from 2 to 20 molar equivalents). The resolution on the peaks for LssTdom-G_5 (2946.6 and 2360.6 respectively for $[M+4H]^{4+}$ and $[M+5H]^{5+}$ charge states) and LssTdom-KA_2 (2955.6 and 2365.9 respectively for $[M+4H]^{4+}$ and $[M+5H]^{5+}$ charge states) was poor and did not allow any consistent measure on the intensity of the peaks. Nonetheless, a smooth increase in the intensity of the ion at *m/z* 2946.6 and 2360.6, and a smooth decrease in the intensity of the ion at *m/z* 2955.6 and 2365.9, relative to *m/z* 2873.5 and 2299.8 respectively, is observed as the amount of KA_2 ligand concentration is increased.

Following addition of 40 µM Gly-L-Ser (GS) (8 molar equivalents) to the LssTdom solution, two additional peaks could be observed at *m/z* 2915.3 (Figure 106A) and 2331.6 (Figure 106B), corresponding to the $[M+4H]^{4+}$ and $[M+5H]^{5+}$ charge states of LssTdom-GS non-covalent complex (measured mass of LssTdom-GS 11655.1 Da; calculated mass of LssTdom-GS 11654.0 Da).



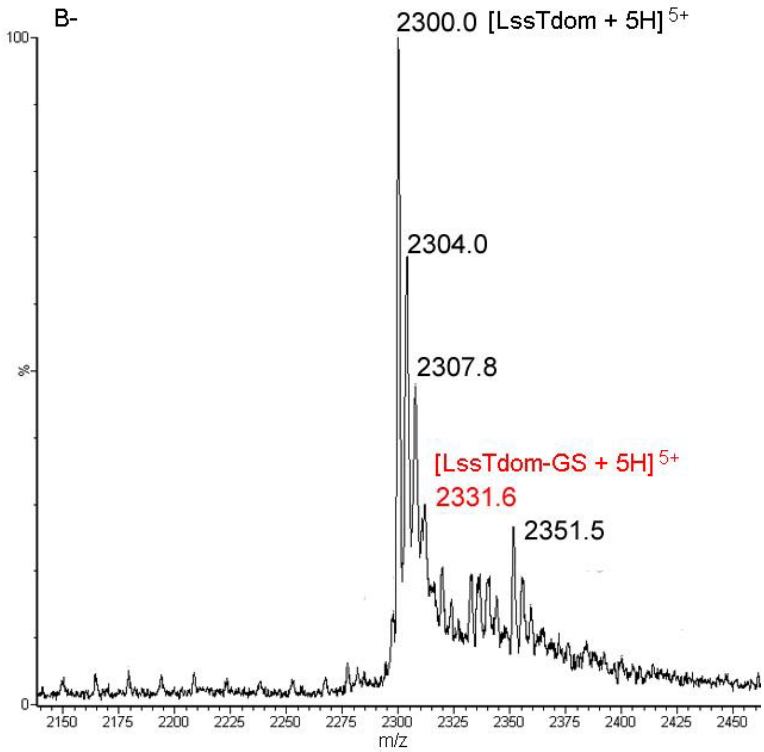


Figure 106: (A) ESI mass spectra of the $[M+4H]^{4+}$ Charge state of the LssTdom showing the LssTdom-GS non-covalent complex signal at *m/z* 2915.3. **(B)** ESI mass spectra of the $[M+5H]^{5+}$ charge state of the LssTdom showing the LssTdom-GS non-covalent complex signal at *m/z* 2331.6.

Thorough dialysis of LssTdom in milliQ water led to significant changes in protein ESI mass spectra. As shown in Figure 107, LssTdom unfolded in milliQ water resulted in ions having higher charge states (6^+-14^+). This may be compared with the same experiment performed in 25 mM triethylammonium acetate, pH 7.4, which promoted a folded conformation with only two charge states (i.e., 4^+-5^+). Moreover the width of the charge state distributions was also larger for unfolded LssTdom due to structural heterogeneity of the corresponding states in solution. The denatured state of a protein exposes more sites for protonation, resulting in the formation of higher charge states (Pramanik *et al.*, 1998).

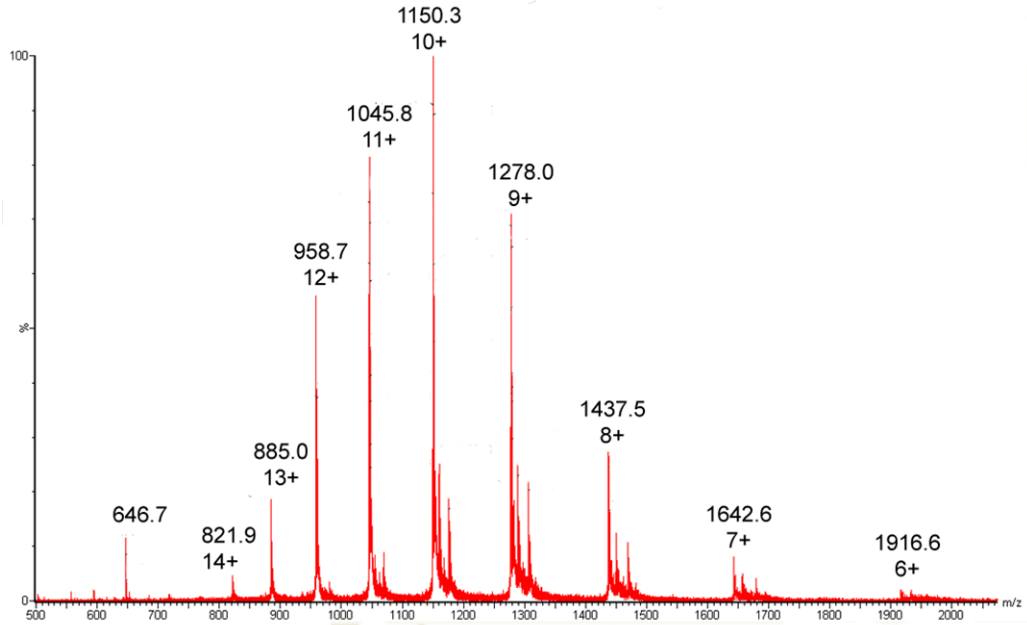


Figure 107: ESI mass spectra of LssTdom thoroughly dialysed in milliQ water showing charge state distribution characteristic of an unfolded protein.

As a negative control LssTdom denatured in milliQ water showed no binding to any of the ligands by ESI-MS (Figure 108). This spectrum is complicated by adducts. The peak at *m/z* 1278.0 correspond to the $[M+9H]^{9+}$ ion of LssTdom where methionine has been cleaved from the *N*-terminal of the protein. The peak at *m/z* 1306.7 could correspond to LssTdom still containing methionine on the *N*-terminus (Met +1), with adducts coming from the different commercial ligand solutions.

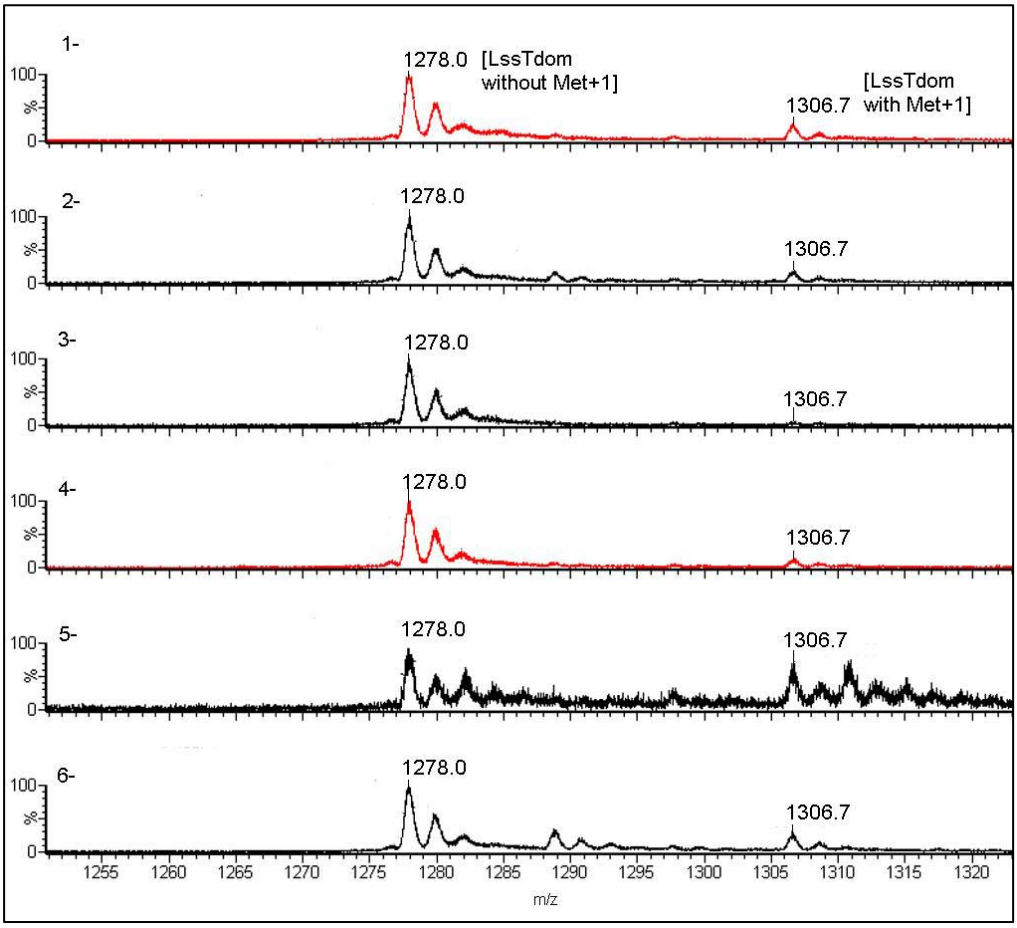


Figure 108: ESI mass spectra of the $[M+9H]^{9+}$ charge state of denatured LssTdom showing the apoprotein signal at *m/z* 1277.9 (**6**.) [**1.** in presence of Gly_5 (160 µM); **2.** In presence of Gly_4 (160 µM); **3.** In presence of Gly_3 (160 µM); **4.** In presence of Lys-D-Ala-D-Ala (160 µM); **5.** In presence of NAM-Ala-iGlu-Lys (160 µM)].

As a negative control no complexes could be observed between LssTdom and Gly_3 (Figure 109A) or L-serine (L-Ser) (Figure 109B) or NAM-Ala-iGlu-Lys (Figure 109C) ligands, following the addition of 8 molar equivalents of the ligands. The peptide Gly3 is believed to be the product of pentaglycine cleavage by lysostaphin and the peptide NAM-Ala-iGlu-Lys is mimicking part of the stem peptide found in *S. aureus* peptidoglycan.

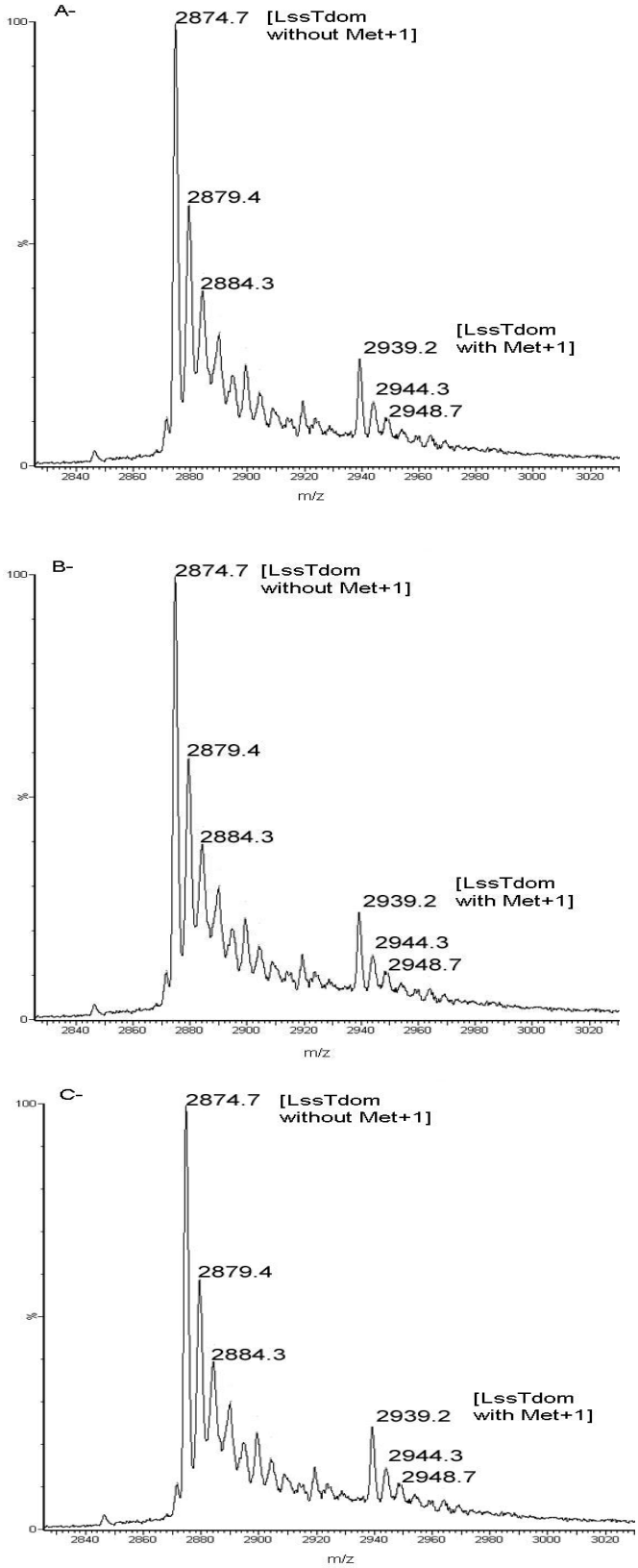


Figure 109: ESI mass spectra of the $[M+4H]^{4+}$ charge state of LssTdom showing the apoprotein signal at *m/z* 2874.7 **(A)** in presence of Gly_3 (40 µM); **(B)** in presence of L-Ser (40 µM); **(C)** in presence of NAM-Ala-iGlu-Lys (40 µM).

Table 19 summarises the results obtained by MS on the complexes formed by LssTdom with ligands mimicking parts of the *S. aureus* peptidoglycan.

Table 19: Table summarising the results on the complexes obtained by MS. Abbreviations used are: K_D, constant of dissociation**, IC50,** median inhibition concentration**, G,** glycine**, S,** serine**, K,** lysine**, A,** alanine**, Q,** glutamine**, NAM,** *N*-acetylmuramic acid.

Complex	**K_D (µM)**	**IC_{50} (µM)**
LssTdom.G_5	9 (±1)	59 (±9)
LssTdom.G_2SGS	15 (±5)	
LssTdom.G_4	38 (±5)	No study
LssTdom.G_3	No binding	No study
LssTdom.G_2	14 (±2)	47 (±1)
LssTdom.KA_2	19 (±4)	
LssTdom.GS	Binding	No study
LssTdom.S	No binding	No study
LssTdom.NAM-AQK	No binding	No study

5.3. Conclusions

Gründling and coworkers showed that the lysostaphin targeting domain binds to purified peptidoglycan and that intact pentaglycine cross bridges are essential for binding (Gründling and Schneewind, 2006). However, in those studies, binding of the lysostaphin targeting domain was not inhibited by addition of excess amounts of pentaglycine. Our results showed a concentration dependent binding of the lysostaphin targeting domain to pentaglycine. A reasonable dissociation constant, with a K_D value of 9×10^{-6} ($\pm 1 \times 10^{-6}$) M for the LssTdom-G_5 complex, was measured. Although ESI-MS is specific and sensitive, caution is needed when interpreting results as the avidity of the gas-phase complex may not be similar to that found in solution, depending on the nature of non-covalent interactions. Complexes where the bond in solution involves a significant hydrophobic component have high K_A values in solution, but low bond strength in the gas-phase. Such complexes will readily decompose in the clean up stages and lead to values of K_A that are too low.

An increasing association with increasing n (Gly_3, Gly_4, Gly_5) was observed. These results suggested the occurrence of a weak, probably specific, interaction between LssTdom and Gly_n (n>3). Lysostaphin targeting domain was shown not to bind Gly_3, which is speculated to be very similar to the product of pentaglycine cleavage by lysostaphin. Endeed, enzymes have evolved to have low affinity to prevent product inhibition. Those results are consistent with the results presented by Warfield and coworkers, where a triglycine FRET substrate was not cleaved at all and a tetraglycine FRET substrate was hydrolysed approximately five times more slowly than a

pentaglycine FRET substrate by full length lysostaphin (Warfield *et al.*, 2006). Firczuk and coworkers briefly reported the testing of tetraglycine-[*N*-(2,4-dinitrophenyl)ethylene diamine] as a fluorescent substrate for LytM (Firczuk *et al.*, 2005). They found this compound to be a moderate substrate (K_M = 2.3 mM). The K_D value found for the LssTdom-G_4 complex by ESI-MS was 38 x 10^{-6} (± 5 x 10^{-6}) M, which is tighter than that found previously. An explanation would be that the substrate used for the FRET assay is not as well accommodated as tetraglycine alone owing to the steric effect of *N*-(2,4-dinitrophenyl)ethylene diamine. It could also be that as LytM and lysostaphin do not share the same targeting domain, they behave differently regarding the substrate.

Additionally, a non-covalent LssTdom-G_2 complex was observed with a K_D value of 14 x 10^{-6} (± 2 x 10^{-6}) M. Gründling and coworkers examined the binding of the lysostaphin targeting domain to the *S. aureus* mutant *femAB* with monoglycine bridges and its isogenic parent with pentaglycine cross-bridges (Gründling and Schneewind, 2006). A drastic decrease in the binding of the lysostaphin targeting domain to the cell wall envelope of *femAB* mutant staphylococci was observed compared with the isogenic wild-type strain. Thus, according to these studies, the binding was dependent on the length of the interpeptide bridge. The unsuspected interaction of the lysostaphin targeting domain to Gly_2 ligand observed by ESI-MS might be explained by a change in conformation of the lysostaphin targeting domain, accommodating a small ligand like Gly-L-Ser and Gly-Gly rather than peptidoglycan from staphylococci.

Data on the K_M (200 x 10^{-6} M) of lysostaphin (full length) using the FRET assay against the Abz-Gly_5-EDDnp FRET substrate (cf. section 3.3) gave a 20 times lower association value compared to that observed on ESI-MS for the LssTdom-Gly_5 complex. An explanation would be that the substrate used for the FRET assay is not as well accommodated as pentaglycine alone, due to the steric effect of the *N*-(2-aminobenzoyl) and *N*-(2,4-dinitrophenyl) ethylene diamine groups, which are fluorophore and quencher respectively in the FRET substrate.

In our study, the dissociation constant values for the LssTdom-G_5 and LssTdom-GGSGS non-covalent complexes were of the same order: 9 x 10^{-6} (± 1 x 10^{-6}) M for the LssTdom-G_5 complex and 15 x 10^{-6} (± 5 x 10^{-6}) M for the LssTdom-GGSGS complex. Data from the competition assay showed that the concentration of G_2SGS that reduced Gly_5 binding by half was 59 μM (± 9 μM). Lu and coworkers compared the binding of the ALE-1 targeting domain to various staphylococcal peptidoglycans with genetically modified interpeptide bridges and their isogenic parents (Lu *et al.*, 2006). Once Gly_5 became Gly_2-Ser-Gly_2 within the peptidoglycan, binding was reduced considerably but not entirely abolished. The targeting domain showed a strong preference for Gly_5 in this study. A question that arose was whether the LssTdom-GGSGS non-covalent complex signal observed by ESI-MS was significant of a specific interaction. Indeed, the lysostaphin targeting domain is positively charged at pH 7.4 and is likely to favour Ser association (polar interaction) over Gly in the gas-phase.

Gründling and coworkers suggested that the lysostaphin targeting domain binding to peptidoglycan required not only pentaglycine cross bridges

but also other structural features of cross-linked peptidoglycan. His studies showed that cross-linked peptidoglycan, harbouring murein disaccharides, wall peptides, and cross bridges, appeared to serve as a receptor for the lysostaphin targeting domain. Our results showed that the lysostaphin targeting domain bound to acetyl-Lys-D-Ala-D-Ala but did not bind to *N*-acetylmuramyl-Ala-D-isoglutaminyl-N^{ε}-stearoyl-Lys; both are fragments of the stem peptides in staphylococci. Those observations are corroborated by the studies of Gründling and coworkers studies (Gründling and Schneewind, 2006) where a *S. aureus dltB* mutant with a higher amount of exposed D-Ala, not linked to polyglycerol-phosphate (LTA) or polyribitol-phosphate (WTA), showed increased binding to the lysostaphin targeting domain. However, Guan and coworkers determined the crystal structure of the *C*-terminal peptidoglycan-binding domain of human peptidoglycan recognition protein (PGRP-Iα) in a complex with a *N*-acetylmuramyl-Ala-D-isoglutaminyl-N^{ε}-stearoyl-Lys representing the conserved core of lysine-type peptidoglycans (Guan *et al.*, 2005). Human peptidoglycan recognition protein hydrolyses the amide bond between the *N*-acetylmuramyl and L-alanine moieties of peptidoglycan and lysostaphin hydrolyses the pentaglycine cross-bridge; this difference of function can explain the difference observed in the binding substrates.

Data from the competition assay showed that the concentration of acetyl-Lys-D-Ala-D-Ala that reduced Gly_2 binding on that lysostaphin targeting domain by half was 47 μM (± 1 μM). These data suggested that acetyl-Lys-D-Ala-D-Ala and Gly_2 are competing for the same binding site in the lysostaphin targeting domain. No quantitative measure was made on acetyl-Lys-D-Ala-D-Ala competition with Gly_5, but the same tendency was

observed: acetyl-Lys-D-Ala-D-Ala and Gly_5 seemed to compete for the same binding site in the lysostaphin targeting domain. This contradicted our first hypothesis which was that lysostaphin targeting domain was accommodating both of those compounds in the same binding site rather than accommodating one at a time. This would make LssTdom very non specific and able to bind to bacteria that do not have the pentaglycine peptidoglycan cross-bridge (i.e, *Streptococcus mutans* which peptidoglycan structure differs from that of *S. aureus* in its interpeptide bridge and contains L-Ala-L-Ala). It is in disagreement with what was reported by Lu and coworkers where ALE-1 targeting domain binding to peptidoglycan of *S. mutans* was significantly lower than to peptidoglycan of *S. aureus* 209P (Lu et al., 2006).

6. Discussion

Lysostaphin is a potential therapeutic antibiotic for Methicillin-resistant *Staphylococcus aureus* (MRSA) for which new antibacterials are required owing to the widespread occurrence of multi-drug resistant strains. *lif*-encoded resistance to lysostaphin is associated with the site-specific incorporation of serine residues into the pentaglycine cross-bridge (Ehlert *et al.*, 2000; Bardelang *et al.*, 2008). Thus a structure is desirable to enable rationally guided design of lysostaphin variants able to bypass the *lif* mechanism of resistance. The investigations in this thesis raised interrogations on the structural requirements for enzymatic activity and on the potential targets in the peptidoglycan for the binding of lysostaphin.

Fourier transform infrared (FT-IR) spectroscopy allowed the detection of conformational changes in lysostaphin and its endopeptidase domain. The proteins infrared bands showed a more ordered conformation for the full length lysostaphin than its endopeptidase domain alone (cf. section 3.5.3.). The compact structure of the full length lysostaphin compared with its cell wall targeting domain was demonstrated by size exclusion chromatography, where lysostaphin was retained longer in the gel filtration column than its targeting domain alone (cf. section 4.7.). It was then proposed that the endopeptidase domain and the cell wall targeting domain interacted in the full length structure of lysostaphin, bringing both domains together to accommodate the same substrate (cf. section 4.7.).

However, in the homology model presented in this thesis (cf. section 4.6.3.), the linker between LssEdom and LssTdom is modelled in mostly extended conformation, bringing the two domains 40 Å apart. Firczuk (2005),

Lu (2006) and their coworkers suggested that LytM activity domain and ALE-1 cell wall binding domain respectively, accommodated the same peptide motif, a polyglycine. This would imply that lysostaphin binding to peptidoglycan would require cross-linked peptidoglycan with two pentaglycine regions separated by at least 40 Å.

The initial prediction of lysostaphin structure from its sequence information is not very accurate considering the error on the protein linker structure prediction that may cause the model not to resemble the native state of the protein (cf. section 4.6.4.). Nevertheless, structure prediction is only a step in protein structure determination. To obtain good crystals of Lss in complex with its substrate (i.e., the non-hydolysable Gly-Gly-L-Ser-Gly-L-Ser peptide or a peptide that possess two non-hydrolysable Gly-Gly-L-Ser-Gly-L-Ser sequences, provided the distance between these motifs is appropriate for the enzyme) that would diffract between 2 and 3 Å, would allow the resolution of its tertiary structure using the technique of molecular replacement where LytM (Firczuk *et al.*, 2005) and ALE-1 (Lu *et al.*, 2006) structures would serve as models. Once the structure of lysostaphin solved, it would be interesting to compare it with LytM structure to understand better the substrate specificity of those two glycyl-glycine endopeptidases.

The interaction studies performed in this thesis (cf. section 5.2.) aimed at defining the potential targets in the peptidoglycan for the binding of lysostaphin. Our results showed a concentration dependent binding of the lysostaphin targeting domain to pentaglycine, Gly-Gly-L-Ser-Gly-L-Ser, tetraglycine, diglycine, acetyl-Lys-D-Ala-D-Ala. Interactions of LssTdom to diglycine was unsuspected as it was not in agreement with what had

previously been reported in the literature (Lu *et al.*, 2006; Gründling and Schneewind, 2006). A question that arose from the binding studies performed was whether the interactions observed in the gas phase resembled those formed in solution. Indeed, the method makes a number of assumptions (Peschke *et al.*, 2004), most notably that complexes are successfully transferred from solution to the gas phase without significant perturbation of the intermolecular non-covalent forces responsible for their association. A limitation with this type of approach is that it is only valid for complexes dominated by polar or ionic interactions, where the dissociation of the complex in the gaseous phase can be neglected (a compound forming a complex that has significant contributions from non-polar interactions is thermodynamically favoured as a dissociation molecule in the gaseous phase).

To define the type of system involved in the association of the lysostaphin targeting domain to the peptidoglycan components of staphylococci, we looked at the residues likely to be involved in the binding. A non-exhaustive list of the residues likely involved in the binding of the ALE-1 targeting domain to pentaglycine is: tyrosine 276 and 280, isoleucine 293, arginine 296 and tryptophan 358 (Lu *et al.*, 2006). As the cell wall targeting domain of lysostaphin and ALE-1 share a high degree of homology, it seems safe to assume that both enzymes behave similarly. Tryptophan and isoleucine are hydrophobic residues, tyrosine is partially hydrophobic and arginine is hydrophilic and charged. As those residues are mainly hydrophobic, we can predict that the complexes of lysostaphin with peptidoglycan components formed were dominated by non-polar interactions and appeared relatively weaker in the gas-phase.

The data obtained for K_D and IC_{50} for the different complexes were under scrutiny, with the question as to whether it was possible to quantify the interactions observed. Indeed, the dissociation constants (K_D) obtained for the binding of the lysostaphin targeting domain are in the μM range which is very low compared to the nM range expected for high affinity binders. Thus it is difficult to assess whether the complexes formed were due to non-specific binding. Nonetheless, the value of K_M obtained for full length lysostaphin by FRET assay measurement showed a very poor binding affinity (K_M = 200 μM), in agreement with that which was obtained by ESI-MS. This gives more confidence in the reliability of the measures obtained by mass spectrometry.

To confirm the data obtained by ESI-MS, a potential future approach could be to use isothermal titration calorimetry (ITC) to measure non-covalent LssTdom-ligand complexes. This would confirm whether or not resistance to lysostaphin action due to incorporation of serine residues in the crossbridge is a result of the endopeptidase domain being unable to cleave this sequence and not due to the targeting domain being unable to bind it. By demonstrating that lysostaphin targeting domain presents a difference of affinity between pentaglycine and Gly-Gly-L-Ser-Gly-L-Ser in solution using ITC technique and thus contradicting the ESI-MS results obtained in the gas-phase, might enable the discovery of lysostaphin targeting domain variants that could acquire altered binding properties using the phage display technique. Indeed, the key aspect of this technique is to find an affinity selection test adapted for the discovery of protein variants, which can be resumed by this quote “you get what you select for”.

-A-

ABERGEL, C., COUTARD, B., BYRNE, D., CHENIVESSE, S., CLAUDE, J. B., DEREGNAUCOURT, C., FRICAUX, T., GIANESINI-BOUTREUX, C., JEUDY, S., LEBRUN, R., MAZA, C., NOTREDAME, C., POIROT, O., SUHRE, K., VARAGNOL, M. & CLAVERIE, J. M. (2003) Structural genomics of highly conserved microbial genes of unknown function in search of new antibacterial targets. *J Struct Funct Genomics*, 4, 141-57.

ALTSCHUL, S. F., MADDEN, T. L., SCHAFFER, A. A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, 25, 3389-402.

ANDRUSHCHENKO, V. V., VOGEL, H. J. & PRENNER, E. J. (2007) Optimization of the hydrochloric acid concentration used for trifluoroacetate removal from synthetic peptides. *J Pept Sci*, 13, 37-43.

ANFINSEN, C. B., REDFIELD, R. R., CHOATE, W. L., PAGE, J. & CARROLL, W. R. (1954) Studies on the gross structure, cross-linkages, and terminal sequences in ribonuclease. *J Biol Chem*, 207, 201-10.

ASHBURN, T. T., AUGER, M. & LANSBURY, P. T. (1992) The structural basis of pancreatic amyloid formation: isotope-edited spectroscopy in the solid state. *J. Am. Chem. Soc.*, 114, 790-1.

-B-

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, 4789-97.

BAGUI, T. K., GHOSH, M. & DATTA, A. K. (1996) Two conformationally vicinal thiols at the active site of Leishmania donovani adenosine kinase. *Biochem J*, 316 (Pt 2), 439-45.

BAKER, D. & SALI, A. (2001) Protein structure prediction and structural genomics. *Science*, 294, 93-6.

BARDELANG, P., VANKEMMELBEKE, M., ZHANG, Y., JARVIS, H., ANTONIADOU, E., ROCHETTE, S., THOMAS, N.R., PENFOLD, C.N., JAMES, R. (2008) Design of a polypeptide FRET substrate that facilitates study of the antimicrobial protease lysostaphin. *Biochem J*. [Epub ahead of print].

BARTH, A. (2007) Infrared spectroscopy of proteins. *Biochim Biophys Acta*, 1767, 1073-101.

BEATSON, S. A., SLOAN, G. L. & SIMMONDS, R. S. (1998) Zoocin A immunity factor: a *fem*A-like gene found in a group C streptococcus. *FEMS Microbiol Lett*, 163, 73-7.

BECKER, S. C., FOSTER-FREY, J. & DONOVAN, D. M. (2008) The phage K lytic enzyme LysK and lysostaphin act synergistically to kill MRSA. *FEMS Microbiol Lett*, 287, 185-91.

BENNETT-LOVSEY, R. M., HERBERT, A. D., STERNBERG, M. J. & KELLEY, L. A. (2008) Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. *Proteins*, 70, 611-25.

BERGER-BACHI, B., BARBERIS-MAINO, L., STRASSLE, A. & KAYSER, F. H. (1989) FemA, a host-mediated factor essential for methicillin resistance in *Staphylococcus aureus*: molecular cloning and characterization. *Mol Gen Genet*, 219, 263-9.

BERGFORS, T. M. (1999) Protein crystallization: techniques, strategies and tips, International University Line, La Jolla, CA.

BERNSTEIN, F. C., KOETZLE, T. F., WILLIAMS, G. J., MEYER, E. F., JR., BRICE, M. D., RODGERS, J. R., KENNARD, O., SHIMANOUCHI, T. & TASUMI, M. (1977) The Protein Data Bank: a computer-based archival file for macromolecular structures. *J Mol Biol*, 112, 535-42.

BETTS, M. J. & RUSSELL, R. B. (2003) Amino acid properties and consequences of substitutions, in Bioinformatics for Geneticist. Barnes, M.R. and Gray, I.C.

BLOOM, J. D., MEYER, M. M., MEINHOLD, P., OTEY, C. R., MACMILLAN, D. & ARNOLD, F. H. (2005) Evolving strategies for enzyme engineering. *Curr Opin Struct Biol*, 15, 447-52.

BOCHTLER, M., ODINTSOV, S. G., MARCYJANIAK, M. & SABALA, I. (2004) Similar active sites in lysostaphins and D-Ala-D-Ala metallopeptidases. *Protein Sci*, 13, 854-61.

BOYLE-VAVRA, S., LABISCHINSKI, H., EBERT, C. C., EHLERT, K. & DAUM, R. S. (2001) A spectrum of changes occurs in peptidoglycan composition of glycopeptide-intermediate clinical *Staphylococcus aureus* isolates. *Antimicrob Agents Chemother*, 45, 280-7.

BRANDEN, C.I., TOOZE, J. (1998) Introduction to protein structure, Garland Publishing Inc., Chap 18, p373-392

BROWDER, H. P., ZYGMUNT, W. A., YOUNG, J. R. & TAVORMINA, P. A. (1965) Lysostaphin: Enzymatic Mode of Action. *Biochem Biophys Res Commun*, 19, 383-9.

BUNN, W. J., HEATH, H. E., LEBLANC, P. A. & SLOAN, G. L. (1998) Wall-associated processing of extracellular enzymes of *Staphylococcus simulans* biovar *staphylolyticus*. *FEMS Microbiol Lett*, 165, 123-7.

BYLER, D. M. & SUSI, H. (1986) Examination of the secondary structure of proteins by deconvolved FTIR spectra. *Biopolymers*, 25, 469-87.

-C-

CEDRONE, F., MENEZ, A. & QUEMENEUR, E. (2000) Tailoring new enzyme functions by rational redesign. *Curr Opin Struct Biol*, 10, 405-10.

CHAMBERS, H.F., SACHDEVA, M. (1990) Binding of β-lactam antibiotics to penicillin-binding proteins in methicillin-resistant *Staphylococcus aureus*. *J Infect Dis*, 161, 1170-6.

CHAYEN, N. E. (2004) Turning protein crystallisation from an art into a science. *Curr Opin Struct Biol*, 14, 577-83.

CLIMO, M. W., EHLERT, K. & ARCHER, G. L. (2001) Mechanism and suppression of lysostaphin resistance in oxacillin-resistant *Staphylococcus aureus*. *Antimicrob Agents Chemother*, 45, 1431-7.

CLIMO, M. W., PATRON, R. L., GOLDSTEIN, B. P. & ARCHER, G. L. (1998) Lysostaphin treatment of experimental methicillin-resistant *Staphylococcus aureus* aortic valve endocarditis. *Antimicrob Agents Chemother*, 42, 1355-60.

COLMAN, P. M., JANSONIUS, J. N. & MATTHEWS, B. W. (1972) The structure of thermolysin: an electron density map at 2-3 Å resolution. *J Mol Biol*, 70, 701-24.

COMPTON, L. A. & JOHNSON, W. C., JR. (1986) Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. *Anal Biochem*, 155, 155-67.

CORPET, F. (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res*, 16, 10881-90.

-D-

DALAL, S., BALASUBRAMANIAN, S. & REGAN, L. (1997) Protein alchemy: changing beta-sheet into alpha-helix. *Nat Struct Biol*, 4, 548-52.

DAVIS, I. W., LEAVER-FAY, A., CHEN, V. B., BLOCK, J. N., KAPRAL, G. J., WANG, X., MURRAY, L. W., ARENDALL, W. B., 3RD, SNOEYINK, J., RICHARDSON, J. S. & RICHARDSON, D. C. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res*, 35, W375-83.

DAVYDOV, A. S. (1971) Theory of molecular excitons, New York, Plenum Press.

DEHART, H. P., HEATH, H. E., HEATH, L. S., LEBLANC, P. A. & SLOAN, G. L. (1995) The lysostaphin endopeptidase resistance gene (epr) specifies modification of peptidoglycan cross bridges in *Staphylococcus simulans* and *Staphylococcus aureus*. *Appl Environ Microbiol*, 61, 1475-9.

-E-

EHLERT, K., TSCHIERSKE, M., MORI, C., SCHRODER, W. & BERGER-BACHI, B. (2000) Site-specific serine incorporation by Lif and Epr into positions 3 and 5 of the staphylococcal peptidoglycan interpeptide bridge. *J Bacteriol*, 182, 2635-8.

ENDO, S. (1962) Study on protease by thermophilic bacteria. *J Ferment Technol*, 40, 346-53.

-F-

FASMAN, G. D. (1996) Circular dichroism and the conformational analysis of biomolecules, New York, Plenum Press.

FENN, J. B., MANN, M., MENG, C. K., WONG, S. F. & WHITEHOUSE, C. M. (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science*, 246, 64-71.

FIRCZUK, M., MUCHA, A. & BOCHTLER, M. (2005) Crystal structures of active LytM. *J Mol Biol*, 354, 578-90.

FÖRSTER, T. (1948) Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann Physik* 437, 55-75.

-G-

GANEM, B., LI, Y. T. & HENION, J. D. (1991) Detection of noncovalent receptor-ligand complexes by mass spectrometry. *J Am Chem Soc*, 113, 6294-6.

GHUYSEN, J. M., TIPPER, D. J. & STROMINGER, J. L. (1965) Structure of the cell wall of *Staphylococcus aureus*, strain Copenhagen. IV. The teichoic acid-glycopeptide complex. *Biochemistry*, 4, 474-85.

GHUYSEN, J. M., TIPPER, D. J. & STROMINGER, J. L. (1966) Enzymes that degrade bacterial cell wall. *Methods Enzymol*, 8, 885-99.

GILL, S. C. & VONHIPPEL, P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem*, 182, 319-26.

GOLDBERG, L. M., DEFRANCO, J. M., WATANAKUNAKORN, C. & HAMBURGER, M. (1967) Studies in experimental staphylococcal endocarditis in dogs. VI. Treatment with lysostaphin. *Antimicrob Agents Chemother*, 7, 45-53.

GOLOVANOV, A. P., HAUTBERGUE, G. M., WILSON, S. A. & LIAN, L. Y. (2004) A simple method for improving protein solubility and long-term stability. *J Am Chem Soc*, 126, 8933-9.

GOORMAGHTIGH, E., CABIAUX, V. & RUYSSCHAERT, J. M. (1990) Secondary structure and dosage of soluble and membrane proteins by attenuated total reflection Fourier-transform infrared spectroscopy on hydrated films. *Eur J Biochem*, 193, 409-20.

GOORMAGHTIGH, E., CABIAUX, V. & RUYSSCHAERT, J. M. (1994) Determination of soluble and membrane protein structure by Fourier transform infrared spectroscopy. III. Secondary structures. *Subcell Biochem*, 23, 405-50.

GREENBLATT, H. M., FEINBERG, H., TUCKER, P. A. & SHOHAM, G. (1998) Carboxypeptidase A: native, zinc-removed and mercury-replaced forms. *Acta Crystallogr D Biol Crystallogr*, 54, 289-305.

GRISHINA, I. B. & WOODY, R. W. (1994) Contributions of tryptophan side chains to the circular dichroism of globular proteins: exciton couplets and coupled oscillators. *Faraday Discuss*, 245-62.

GRUNDLING, A. & SCHNEEWIND, O. (2006) Cross-linked peptidoglycan mediates lysostaphin binding to the cell wall envelope of *Staphylococcus aureus*. *J Bacteriol*, 188, 2463-72.

GUAN, R., BROWN, P. H., SWAMINATHAN, C. P., ROYCHOWDHURY, A., BOONS, G. J. & MARIUZZA, R. A. (2006) Crystal structure of human peptidoglycan recognition protein I alpha bound to a muramyl pentapeptide from Gram-positive bacteria. *Protein Sci*, 15, 1199-206.

GUAN, R., ROYCHOWDURY, A., EMBER, B., KUMAR, S., BOONS, G. J. & MARIUZZA, R. A. (2005) Crystal structure of a peptidoglycan recognition protein (PGRP) in complex with a muramyl tripeptide from Gram-positive bacteria. *J Endotoxin Res*, 11, 41-6.

-H-

HALL, T. M., PORTER, J. A., BEACHY, P. A. & LEAHY, D. J. (1995) A potential catalytic site revealed by the 1.7-Å crystal structure of the amino-terminal signalling domain of Sonic hedgehog. *Nature*, 378, 212-6.

HALVERSON, K. J., SUCHOLEIKI, I., ASHBURN, T. T. & LANSBURY, P. T. (1991) Location of beta-sheet-forming sequences in amyloid proteins by FTIR. *J Am Chem Soc*, 113, 6701-2.

HAMBLY, E. & SUTTLE, C. A. (2005) The viriosphere, diversity, and genetic exchange within phage communities. *Curr Opin Microbiol*, 8, 444-50.

HANGAUER, D. G., MONZINGO, A. F. & MATTHEWS, B. W. (1984) An interactive computer graphics study of thermolysin-catalyzed peptide cleavage and inhibition by *N*-carboxymethyl dipeptides. *Biochemistry*, 23, 5730-41.

HARRICK, N. J. (1967) Internal Reflection Spectroscopy, John Wiley & Sons Inc., New York: Interscience.

HARRISON, E. F. & ZYGMUNT, W. A. (1967) Lysostaphin in experimental renal infections. *J Bacteriol*, 93, 520-4.

HEATH, H. E., HEATH, L. S., ROSE, K. E. & SLOAN, G. L. (1991) Beta-lactamase is encoded on plasmid pACK3 in *Staphylococcus simulans* biovar *staphylolyticus*. *FEMS Microbiol Lett*, 61, 113-6.

HEATH, L. S., HEATH, H. E., LEBLANC, P. A., SMITHBERG, S. R., DUFOUR, M., SIMMONDS, R. S. & SLOAN, G. L. (2004) The streptococcolytic enzyme zoocin A is a penicillin-binding protein. *FEMS Microbiol Lett*, 236, 205-11.

HEGDE, S. S. & SHRADER, T. E. (2001) FemABX family members are novel nonribosomal peptidyltransferases and important pathogen-specific drug targets. *J Biol Chem*, 276, 6998-7003.

HEINRICH, P., ROSENSTEIN, R., BOHMER, M., SONNER, P. & GOTZ, F. (1987) The molecular organization of the lysostaphin gene and its sequences repeated in tandem. *Mol Gen Genet*, 209, 563-9.

HENZE, U., SIDOW, T., WECKE, J., LABISCHINSKI, H. & BERGER-BACHI, B. (1993) Influence of *fem*B on methicillin resistance and peptidoglycan metabolism in *Staphylococcus aureus*. *J Bacteriol*, 175, 1612-20.

HOLDEN, H. M., TRONRUD, D. E., MONZINGO, A. F., WEAVER, L. H. & MATTHEWS, B. M. (1987) Slow- and fast-binding inhibitors of thermolysin display different modes of binding: crystallographic analysis of extended phosphonamidate transition-state analogues. *Biochemistry*, 26, 8542-53.

HUBER, M. M. & HUBER, T. W. (1989) Susceptibility of methicillin-resistant *Staphylococcus aureus* to lysostaphin. *J Clin Microbiol*, 27, 1122-4.

-I-

IVERSEN, O. J. & GROV, A. (1973) Studies on lysostaphin. Separation and characterization of three enzymes. *Eur J Biochem*, 38, 293-300.

-J-

JACK, R. W., TAGG, J. R. & RAY, B. (1995) Bacteriocins of gram-positive bacteria. *Microbiol Rev*, 59, 171-200.

JOERGER, R. D. (2003) Alternatives to antibiotics: bacteriocins, antimicrobial peptides and bacteriophages. *Poult Sci*, 640-7.

JOHNSON, W. C. (1999) Analyzing protein circular dichroism spectra for accurate secondary structures. *Proteins*, 35, 307-12.

JOHNSON, W. C., JR. & TINOCO, I., JR. (1972) Circular dichroism of polypeptide solutions in the vacuum ultraviolet. *J Am Chem Soc*, 94, 4389-92.

-K-

KABSCH, W. & SANDER, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22, 2577-637.

KANG, D., LIU, G., LUNDSTROM, J., GELIUS, E. & STEINER, H. (1998) A peptidoglycan recognition protein in innate immunity conserved from insects to mammals. *Proc Natl Acad Sci U S A*, 1998, 10078-82.

KARPLUS, P. A. (1996) Experimentally observed conformation-dependent geometry and hidden strain in proteins. *Protein Sci*, 5, 1406-20.

KAUPPINEN, J. K., MOFFAT, D. J., MANTSCH, H. H. & CAMERON, D. G. (1981) Fourier Self-deconvolution: a method for resolving intrinsically overlapped bands. *Applied Spectroscopy*, 35, 255-334.

KEBARLE, P. (2000) A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. *J Mass Spectrom*, 35, 804-17.

KESTER, W. R. & MATTHEWS, B. M. (1977) Crystallographic study of the binding of dipeptide inhibitors to thermolysin: implications for the mechanism of catalysis. *Biochemistry*, 16, 2506-16.

KIM, M. S., BYUN, M. & OH, B. H. (2003) Crystal structure of peptidoglycan recognition protein LB from *Drosophila melanogaster*. *Nat Immunol*, 4, 787-93.

KIRI, N., ARCHER, G. & CLIMO, M. W. (2002) Combinations of lysostaphin with beta-lactams are synergistic against oxacillin-resistant *Staphylococcus epidermidis*. *Antimicrob Agents Chemother*, 46, 2017-20.

KOCH, C.A., ANDERSON, D., MORAN, M.F., ELLIS, C. & PAWSON, T. (1991) SH2 and SH3 domains: elements that control interactions of cytoplasmic signalling proteins. *Science*, 252, 668-74

KOEHL, J. L., MUTHAIYAN, A., JAYASWAL, R. K., EHLERT, K., LABISCHINSKI, H. & WILKINSON, B. J. (2004) Cell wall composition and decreased autolytic activity and lysostaphin susceptibility of glycopeptide-intermediate *Staphylococcus aureus*. *Antimicrob Agents Chemother*, 48, 3749-57.

KOKAI-KUN, J. F., WALSH, S. M., CHANTURIYA, T. & MOND, J. J. (2003) Lysostaphin cream eradicates *Staphylococcus aureus* nasal colonization in a cotton rat model. *Antimicrob Agents Chemother*, 47, 1589-97.

KOPP, J. & SCHWEDE, T. (2004) The SWISS-MODEL repository of annotated three-dimensional protein structure homology models. *Nucleic Acids Res*, 32, 230-4.

KOPP, J. & SCHWEDE, T. (2006) The SWISS-MODEL repository: new features and functionalities. *Nucleic Acids Res*, 34, 315-8.

KRIMM, S. & BANDEKAR, J. (1986) Vibrational spectroscopy and conformation of peptides, polypeptides, and proteins. *Adv Protein Chem*, 38, 181-364.

KUMAR, S., ROYCHOWDHURY, A., EMBER, B., WANG, Q., GUAN, R., MARIUZZA, R. A. & BOONS, G. J. (2005) Selective recognition of synthetic lysine and meso-diaminopimelic acid-type peptidoglycan fragments by human peptidoglycan recognition proteins Iα and S. *J Biol Chem*, 280, 37005-12.

-L-

LAKOWICZ, J. R. (1999) Principles of Fluorescence spectroscopy New York, Kluwer Academic/Plenum.

LEMAIRE, D., MARIE, G., SERANI, L. & LAPREVOTE, O. (2001) Stabilization of gas-phase noncovalent macromolecular complexes in electrospray mass spectrometry using aqueous triethylammonium bicarbonate buffer. *Anal Chem*, 73, 1699-706.

LIANG, Q., SIMMONDS, R. S. & TIMKOVICH, R. (2004) NMR evidence for independent domain structures in zoocin A, an antibacterial exoenzyme. *Biochem Biophys Res Commun*, 317, 527-30.

LING, B. & BERGER-BACHI, B. (1998) Increased overall antibiotic susceptibility in *Staphylococcus aureus fem*AB null mutants. *Antimicrob Agents Chemother*, 42, 936-8.

LIPSCOMB, W. N. & STRATER, N. (1996) Recent Advances in Zinc Enzymology. *Chem Rev*, 96, 2375-2434.

LIU, C., XU, Z., GUPTA, D. & DZIARSKI, R. (2001) Peptidoglycan recognition proteins: a novel family of four human innate immunity pattern recognition molecules. *J Biol Chem*, 2001, 34686-94.

LIU, M. & BARTH, A. (2003) Mapping interactions between the Ca^{2+}-ATPase and its substrate ATP with infrared spectroscopy. *J Biol Chem*, 278, 10112-8.

LIVINGSTONE, C. D. & BARTON, G. J. (1993) Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Comput Appl Biosci*, 9, 745-56.

LOESSNER, M. J. (2005) Bacteriophage endolysins-current state of research and applications. *Curr Opin Microbiol*, 2005, 8, 480-7.

LOLL, P. J. & AXELSEN, P. H. (2000) The structural biology of molecular recognition by vancomycin. *Annu Rev Biophys Biomol Struct*, 29, 265-89.

LOPEZ, R. & GARCIA, E. (2004) Recent trends on the molecular biology of pneumococcal capsules, lytic enzymes, and bacteriophage. *FEMS Microbiol Lett*, 28, 553-80.

LOW, L. Y., YANG, C., PEREGO, M., OSTERMAN, A. & LIDDINGTON, R. C. (2005) Structure and lytic activity of a *Bacillus anthracis* prophage endolysin. *J Biol Chem*, 280, 35433-9.

LU, J. Z., FUJIWARA, T., KOMATSUZAWA, H., SUGAI, M. & SAKON, J. (2006) Cell wall-targeting domain of glycylglycine endopeptidase distinguishes among peptidoglycan cross-bridges. *J Biol Chem*, 281, 549-58.

-M-

MAIDHOF, H., REINICKE, B., BLUMEL, P., BERGER-BACHI, B. & LABISCHINSKI, H. (1991) *fem*A, which encodes a factor essential for expression of methicillin resistance, affects glycine content of peptidoglycan in methicillin-resistant and methicillin-susceptible *Staphylococcus aureus* strains. *J Bacteriol*, 173, 3507-13.

MANI, N., TOBIN, P. & JAYASWAL, R. K. (1993) Isolation and characterization of autolysis-defective mutants of *Staphylococcus aureus* created by Tn917-lacZ mutagenesis. *J Bacteriol*, 175, 1493-9.

MANNING, M. C. & WOODY, R. W. (1987) Theoretical determination of the CD of proteins containing closely packed antiparallel beta-sheets. *Biopolymers*, 26, 1731-52.

MARTI-RENOM, M. A., STUART, A. C., FISER, A., SANCHEZ, R., MELO, F. & SALI, A. (2000) Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct*, 29, 291-325.

McCOY, M. (2004) Killing Staph together, Start-up Biosynexus places fermentation project in newly strated Avecia facility. *Chem Engin News*.

MENGIN-LECREULX, D. & LEMAITRE, B. (2005) Structure and metabolism of peptidoglycan and molecular requirements allowing its detection by the Drosophila innate immune system. *J Endotoxin Res*, 11, 105-11.

MEROUEH, S. O., BENCZE, K. Z., HESEK, D., LEE, M., FISHER, J. F., STEMMLER, T. L. & MOBASHERY, S. (2006) Three-dimensional structure of the bacterial cell wall peptidoglycan. *Proc Natl Acad Sci* U S A, 103, 4404-9.

MORA, J. F., VAN BERKEL, G. J., ENKE, C. G., COLE, R. B., MARTINEZ-SANCHEZ, M. & FENN, J. B. (2000) Electrochemical processes in electrospray ionization mass spectrometry. *J Mass Spectrom*, 35, 939-52.

MOREIRA, B., BOYLE-VAVRA, S., DeJONGE, B.L.M., DAUM, R.S. (1997) Increased production of penicillin-binding protein 2, increased detection of other penicillin-binding proteins, and decreased coagulase activity associated with glycopeptide resistance in *Staphylococcus aureus*. *Antimicrob Agents Chemother*, 41, 1788-93

-N-

NAVARRE, W. 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.

NEUMANN, V. C., HEATH, H. E., LEBLANC, P. A. & SLOAN, G. L. (1993) Extracellular proteolytic activation of bacteriolytic peptidoglycan hydrolases of *Staphylococcus simulans* biovar *staphylolyticus*. *FEMS Microbiol Lett*, 110, 205-11.

NIGUTOVA, K., SERENCOVA, L., PIKNOVA, M., JAVORSKY, P. & PRISTAS, P. (2008) Heterologous expression of functionally active enterolysin A, class III bacteriocin from *Enterococcus faecalis*, in *Escherichia coli*. *Prot Expr Purif*, 60, 20-4.

NILSEN, T., NES, I. F. & HOLO, H. (2003) Enterolysin A, a cell wall-degrading bacteriocin from *Enterococcus faecalis* LMG 2333. *Appl Environ Microbiol*, 69, 2975-84.

NOGUCHI, T., INOUE, Y. & TANG, X. S. (1999) Structure of a histidine ligand in the photosynthetic oxygen-evolving complex as studied by light-induced Fourier transform infrared difference spectroscopy. *Biochemistry*, 38, 10187-95.

-O-

ODINTSOV, S. G., SABALA, I., MARCYJANIAK, M. & BOCHTLER, M. (2004) Latent LytM at 1.3Å resolution. *J Mol Biol*, 335, 775-85.

PACE, C. (1995) How to measure and predict the molar absorption coefficient of a protein. *Prot Sci*, 4, 2411-23.

PANGBURN, M. K. & WALSH, K. A. (1975) Thermolysin and neutral protease: mechanistic considerations. *Biochemistry*, 14, 4050-4.

PARK, P. W., SENIOR, R. M., GRIFFIN, G. L., BROEKELMANN, T. J., MUDD, M. S. & MECHAM, R. P. (1995) Binding and degradation of elastin by the staphylolytic enzyme lysostaphin. *Int J Biochem Cell Biol*, 27, 139-46.

PATRON, R. L., CLIMO, M. W., GOLDSTEIN, B. P. & ARCHER, G. L. (1999) Lysostaphin treatment of experimental aortic valve endocarditis caused by a *Staphylococcus aureus* isolate with reduced susceptibility to vancomycin. *Antimicrob Agents Chemother*, 43, 1754-5.

PELMENSCHIKOV, V., BLOMBERG, M. R. & SIEGBAHN, P. E. (2002) A theoretical study of the mechanism for peptide hydrolysis by thermolysin. *J Biol Inorg Chem*, 7, 284-98.

PERANEN, J., RIKKONEN, M., HYVONEN, M. & KAARIAINEN, L. (1996) T7 vectors with modified T7lac promoter for expression of proteins in *Escherichia coli*. *Anal Biochem*, 236, 371-3.

PERRY, A. M., TON-THAT, H., MAZMANIAN, S. K. & SCHNEEWIND, O. (2002) Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. III. Lipid II is an in vivo peptidoglycan substrate for sortase-catalyzed surface protein anchoring. *J Biol Chem*, 277, 16241-8.

PESCHKE, M., BLADES, A. & KEBARLE, P. (2002) Charged states of proteins. Reactions of doubly protonated alkyldiamines with NH(3): solvation or deprotonation. Extension of two proton cases to multiply protonated globular proteins observed in the gas phase. *J Am Chem Soc*, 124, 11519-30.

PESCHKE, M., VERKERK, U. H. & KEBARLE, P. (2004) Features of the ESI mechanism that affect the observation of multiply charged noncovalent protein complexes and the determination of the association constant by the titration method. *J Am Soc Mass Spectrom*, 15, 1424-34.

PORTER, C. J., SCHUCH, R., PELZEK, A. J., BUCKLE, A. M., MCGOWAN, S., WILCE, M. C. J., ROSSJOHN, J., RUSSELL, R., NELSON, D., FISCHETTI, V. A. & WHISSTOCK, J. C. (2007) The 1.6 Å crystal structure of the catalytic domain of PlyB, a bacteriophage lysin active against *Bacillus anthracis*. *J Mol Biol*, 366, 540-50.

PRAMANIK, B. N., BARTNER, P. L., MIRZA, U. A., LIU, Y. H. & GANGULY, A. K. (1998) Electrospray ionization mass spectrometry for the study of non-covalent complexes: an emerging technology. *J Mass Spectrom*, 33, 911-20.

PSPC (2008) http://www.predictioncenter.org. Protein Structure Prediction Center, University of California.

-Q-

QUICKEL, K. E., JR., SELDEN, R., CALDWELL, J. R., NORA, N. F. & SCHAFFNER, W. (1971) Efficacy and safety of topical lysostaphin treatment of persistent nasal carriage of *Staphylococcus aureus*. *Appl Microbiol*, 22, 446-50.

-R-

RAMACHANDRAN, G. N., RAMAKRISHNAN, C. & SASISEKHARAN, V. (1963) Stereochemistry of polypeptide chain configurations. *J Mol Biol*, 7, 95-9.

RAMACHANDRAN, G. N. & SASISEKHARAN, V. (1968) Conformation of polypeptides and proteins. *Adv Protein Chem*, 23, 283-438.

RAMADURAI, L. & JAYASWAL, R. K. (1997) Molecular cloning, sequencing, and expression of *lyt*M, a unique autolytic gene of *Staphylococcus aureus*. *J Bacteriol*, 179, 3625-31.

RAMADURAI, L., LOCKWOOD, K. J., NADAKAVUKAREN, M. J. & JAYASWAL, R. K. (1999) Characterization of a chromosomally encoded glycylglycine endopeptidase of *Staphylococcus aureus*. *Microbiology*, 145, 801-8.

RAO, M. B., TANKSALE, A. M., GHATGE, M. S. & DESHPANDE, V. V. (1998) Molecular and biotechnological aspects of microbial proteases. *Microbiol Mol Biol Rev*, 62, 597-635.

RAWLINGS, N. D., MORTON, F. R., KOK, C. Y., KONG, J. & BARRETT, A. J. (2008) MEROPS: the peptidase database. *Nucleic Acids Res*, 36, D320-5.

RAWLINGS, N. D., O'BRIEN, E. & BARRETT, A. J. (2002) MEROPS: The protease database. *Nucleic Acids Res*, 30, 343-6.

RAYLEIGH, C. (1882) On the equilibrium of liquid conducting. Masses charged with electricity. *Phil Mag*, 14, 184-6.

RECSEI, P. A., GRUSS, A. D. & NOVICK, R. P. (1987) Cloning, sequence, and expression of the lysostaphin gene from *Staphylococcus simulans*. *Proc Natl Acad Sci* U S A, 84, 1127-31.

RHODES, G. (2006) Crystallography made crystal clear: a guide for users of macromolecular models, Academic Press, Portland, USA, p34, p227-30, p269-91.

ROCHE, F. M., MASSEY, R., PEACOCK, S. J., DAY, N. P., VISAI, L., SPEZIALE, P. & AL., E. (2003) Characetrization of novel LPXTG-containing proteins of *Staphylococcus aureus* identified from genome sequences. *Microbiology*, 149, 643-54.

ROHRER, S. & BERGER-BACHI, B. (2003a) FemABX peptidyl transferases: a link between branched-chain cell wall peptide formation and beta-lactam resistance in Gram-positive cocci. *Antimicrob Agents Chemother*, 47, 837-46.

ROHRER, S., EHLERT, K., TSCHIERSKE, M., LABISCHINSKI, H. & BERGER-BACHI, B. (1999) The essential *Staphylococcus aureus* gene *fmh*B is involved in the first step of peptidoglycan pentaglycine interpeptide formation. *Proc Natl Acad Sci* U S A, 96, 9351-6.

ROTHSCHILD, K. J., SANCHES, R., HSIAO, T. L. & CLARK, N. A. (1980) A spectroscopic study of rhodopsin alpha-helix orientation. *Biophys J*, 31, 53-64.

RUPP, B. & WANG, J. (2004) Predictive models for protein crystallization. *Methods*, 34, 390-407.

-S-

SABATH, L. D., LEAF, C. D., GERSTEIN, D. A. & FINLAND, M. (1969) Cell walls of methicillin-resistant *Staphylococcus aureus*. *Antimicrob Agents Chemother*, 73, 73-7.

SALI, A. & BLUNDELL, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*, 234, 779-815.

SAMBROOK, J., FRITSH, E. F. & MANIATIS, T. (1989) Molecular cloning: a laboratory manual, Cold Spring Harbor Laboratory Press.

SARKAR, G. & SOMMER, S. S. (1990) The "megaprimer" method of site-directed mutagenesis. *Biotechniques*, 8, 404-7.

SCHAFFNER, M. A., MELLY, M. A. & KOENIG, M. G. (1967) Lysostaphin: an enzymatic approach to staphylococcal desease. II. *In vivo* studies. Yale *J Biol Med*, 39, 230-44.

SCHELLMAN, J. A. (1968) Symmetry rules for optical rotation. *Acc Chem Res*, 1, 144-51.

SCHINDLER, C. A. & SCHUHARDT, V. T. (1964) Lysostaphin: a new bacteriolytic agent for the staphylococcus. *Proc Natl Acad Sci* U S A, 51, 414-21.

SCHINDLER, C. A. & SCHUHARDT, V. T. (1965) Purification and properties of lysostaphin-a lytic agent for *Staphylococcus aureus*. *Biochim Biophys Acta*, 97, 242-50.

SCHLEIFER, K. H. & KANDLER, O. (1972) Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriol Rev*, 36, 407-77.

SCHNEEWIND, O., FOWLER, A. & FAULL, K.F. (1995) Structure of the cell wall anchor of surface proteins in *Staphylococcus aureus*. *Science*, 268, 103-6

SCHNEIDER, T., SENN, M. M., BERGER-BACHI, 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, 675-85.

SHAH, A., MOND, J. & WALSH, S. (2004) Lysostaphin-coated catheters eradicate *Staphylococccus aureus* challenge and block surface colonization. *Antimicrob Agents Chemother*, 48, 2704-7.

SHARMA, R., SHARMA, P. R., CHOUDHARY, M. L., PANDE, A. & KHATRI, G. S. (2006) Cytoplasmic expression of mature glycylglycine endopeptidase lysostaphin with an amino terminal hexa-histidine in a soluble and catalytically active form in *Escherichia coli*. *Protein Expr Purif*, 45, 206-15.

SHEEHAN, M. M., GARCIA, J. L., LOPEZ, R. & GARCIA, P. (1996) Analysis of the lysin of the lactococcal bacteriophage Tuc2009 by chimeric gene assembling. *FEMS Microbiol Lett*, 140, 23-8.

SIMMONDS, R. S., NAIDOO, J., JONES, C. L. & TAGG, J. R. (1995) The Streptococcal bacteriocin-like inhibitory substance, zoocin A, reduces the proportion of *Streptococcus mutans* in an artificial plaque. *Microb Ecol Health Dis*, 8, 281-92.

SIMMONDS, R. S., PEARSON, L., KENNEDY, R. C. & TAGG, J. R. (1996) Mode of action of a lysostaphin-like bacteriolytic agent produced by *Streptococcus zooepidemicus* 4881. *Appl Environ Microbiol*, 62, 4536-41.

SIMMONDS, R. S., SIMPSON, W. J. & TAGG, J. R. (1997) Cloning and sequence analysis of zooA, a *Streptococcus zooepidemicus* gene encoding a bacteriocin-like inhibitory substance having a domain structure similar to that of lysostaphin. *Gene*, 189, 255-61.

SINGH, J. & THORNTON, J. M. (1992) Atlas of protein side-chain interactions. Oxford University Press, Oxford.

SLOAN, G. L., SMITH, E. C. & LANCASTER, J. H. (1977) Lysostaphin endopeptidase-catalysed transpeptidation reactions of the imino-transfer type. *Biochem J*, 167, 293-6.

SMITH, G. P. & PETRENKO, V. A. (1997) Phage Display. *Chem Rev*, 97, 391-410.

SPRATT, B. G. (1994) Resistance to antibiotics mediated by target alterations. *Science*, 264, 388-93.

STRANDEN, A. M., EHLERT, K., LABISCHINSKI, H. & BERGER-BACHI, B. (1997) Cell wall monoglycine cross-bridges and methicillin hypersusceptibility in a *fem*AB null mutant of methicillin-resistant *Staphylococcus aureus*. *J Bacteriol*, 179, 9-16.

SUGAI, M., FUJIWARA, T., AKIYAMA, T., OHARA, M., KOMATSUZAWA, H., INOUE, S. & SUGINAKA, H. (1997a) Purification and molecular characterization of glycylglycine endopeptidase produced by *Staphylococcus capitis* EPK1. *J Bacteriol*, 179, 1193-202.

SUGAI, M., FUJIWARA, T., OHTA, K., KOMATSUZAWA, H., OHARA, M. & SUGINAKA, H. (1997b) epr, which encodes glycylglycine endopeptidase resistance, is homologous to *fem*AB and affects serine content of peptidoglycan cross bridges in *Staphylococcus capitis* and *Staphylococcus aureus*. *J Bacteriol*, 179, 4311-8.

SUREWICZ, W. K. & MANTSCH, H. H. (1988) New insight into protein secondary structure from resolution-enhanced infrared spectra. *Biochim Biophys Acta*, 952, 115-30.

SUREWICZ, W. K., LEDDY, J. J. & MANTSCH, H. H. (1990) Structure, stability, and receptor interaction of cholera toxin as studied by Fourier-transform infrared spectroscopy. *Biochemistry*, 29, 8106-11.

SUREWICZ, W. K., MANTSCH, H. H. & CHAPMAN, D. (1993) Determination of protein secondary structure by Fourier transform infrared spectroscopy: a critical assessment. *Biochemistry*, 32, 389-94.

SZWEDA, P., KOTLOWSKI, R. & KUR, J. (2005) New effective sources of the *Staphylococcus simulans* lysostaphin. *J Biotechnol*, 117, 203-13.

-T-

TABAC, M., WITTE, A. & BLASI, U. (2005) Functional analysis of the lysis genes of *Staphylococcus aureus* phage P68 in *Escherichia coli*. *Microbiology*, 151, 2331-42.

TANG, X. J., BREWER, C. F., SAHA, S., CHERNUSHEVICH, I., ENS, W. & STANDING, K. G. (1994) Investigation of protein-protein noncovalent interactions in soybean agglutinin by electrospray ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom*, 8, 750-4.

THOMPSON, J. D., HIGGINS, D. G. & GIBSON, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through

sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, 22, 4673-80.

THUMM, G. & GOTZ, F. (1997) Studies on prolysostaphin processing and characterization of the lysostaphin immunity factor (Lif) of *Staphylococcus simulans* biovar *staphylolyticus*. *Mol Microbiol*, 23, 1251-65.

TJERNBERG, A., CARNO, S., OLIV, F., BENKESTOCK, K., EDLUND, P. O., GRIFFITHS, W. J. & HALLEN, D. (2004) Determination of dissociation constants for protein-ligand complexes by electrospray ionization mass spectrometry. *Anal Chem*, 76, 4325-31.

TON-THAT, H., MARRAFFINI, L. A. & SCHNEEWIND, O. (2004) Protein sorting to the cell wall envelope of Gram-positive bacteria. *Biochim Biophys Acta*, 1694, 269-78.

TRAYER, H. R. & BUCKLEY, C. E., 3RD (1970) Molecular properties of lysostaphin, a bacteriolytic agent specific for *Staphylococcus aureus*. *J Biol Chem*, 245, 4842-6.

TSCHIERSKE, M., EHLERT, K., STRANDEN, A. M. & BERGER-BACHI, B. (1997) Lif, the lysostaphin immunity factor, complements FemB in staphylococcal peptidoglycan interpeptide bridge formation. *FEMS Microbiol Lett*, 153, 261-4.

TYDELL, C. C., YUAN, J., TRAN, P. & SELSTED, M. E. (2006) Bovine peptidoglycan recognition protein-S: antimicrobial activity, localization, secretion, and binding properties. *J Immunol*, 2006, 1154-62.

-V-

VENYAMINOV, S. & KALNIN, N. N. (1990) Quantitative IR spectrophotometry of peptide compounds in water (H_2O) solutions. I. Spectral parameters of amino acid residue absorption bands. *Biopolymers*, 30, 1243-57.

VEROS, C. T. & OLDHAM, N. J. (2007) Quantitative determination of lysozyme-ligand binding in the solution and gas phases by electrospray ionisation mass spectrometry. *Rapid Commun Mass Spectrom*, 21, 3505-10.

VON EIFF, C., KOKAI-KUN, J. F., BECKER, K. & PETERS, G. (2003) *In vitro* activity of recombinant lysostaphin against *Staphylococcus aureus* isolates from anterior nares and blood. *Antimicrob Agents Chemother*, 47, 3613-5.

-W-

WALSH, S., SHAH, A. & MOND, J. (2003) Improved pharmacokinetics and reduced antibody reactivity of lysostaphin conjugated to polyethylene glycol. *Antimicrob Agents Chemother*, 47, 554-8.

WANG, Z. M., LI, X., COCKLIN, R. R., WANG, M., FUKASE, K., INAMURA, S., KUSUMOTO, S., GUPTA, D. & DZIARSKI, R. (2003) Human peptidoglycan recognition protein-L is an *N*-acetylmuramoyl-L-alanine amidase. *J Biol Chem*, 278, 49044-52.

WARFIELD, R., BARDELANG, P., SAUNDERS, H., CHAN, W. C., PENFOLD, C., JAMES, R. & THOMAS, N. R. (2006) Internally quenched peptides for the study of lysostaphin: an antimicrobial protease that kills *Staphylococcus aureus*. *Org Biomol Chem*, 4, 3626-38.

WERNER, T., LIU, G., KANG, D., EKENGREN, S., STEINER, H. & HULTMARK, D. (2000) A family of peptidoglycan recognition proteins in the fruit fly *Drosophila melanogaster*. *Proc Natl Acad Sci* U S A, 2000, 13772-7.

WILKINSON, B. J., DORIAN, K. J. & SABATH, L. D. (1978) Cell wall composition and associated properties of methicillin-resistant *Staphylococcus aureus* strains. *J Bacteriol*, 136, 976-82.

WILM, M. & MANN, M. (1996) Analytical properties of the nanoelectrospray ion source. *Anal Chem*, 68, 1-8.

WOODY, R. W. (1996) Theory of circular dichroism of proteins. In Circular Dichroism and the conformational analysis of biomolecules, New York, Plenum.

WU, J. A., KUSUMA, C., MOND, J. J. & KOKAI-KUN, J. F. (2003) Lysostaphin disrupts *Staphylococcus aureus* and *Staphylococcus epidermidis* biofilms on artificial surfaces. *Antimicrob Agents Chemother*, 47, 3407-14.

-X-

XU, N., HUANG, Z-H., DeJONGE B.L. & GAGE, D.A. (1997) Structural characterization of peptidoglycan muropeptides by matrix-assisted laser desorption ionization mass spectrometry and postsource decay analysis. *Anal Biochem*, 248, 7-14.

-Y-

YOUNG, R. Y., WANG, I.-N. & ROOF, W. D. (2000) Phages will out: strategies of host cell lysis. *Trends Microbiol*, 8, 120-8.

-Z-

ZHU, X.L., OHTA, Y., JORDAN, F. & INOUYE, M. (1989) Pro-sequence of subtilisin can guide the refolding of denatured subtilisin in an intermolecular process. *Nature*, 339, 483.

ZYGMUNT, W. A., BROWDER, H. P. & TAVORMINA, P. A. (1966) Influence of blood and serum on the antistaphylococcal activity of lysostaphin. *J Bacteriol*, 91, 725-8.

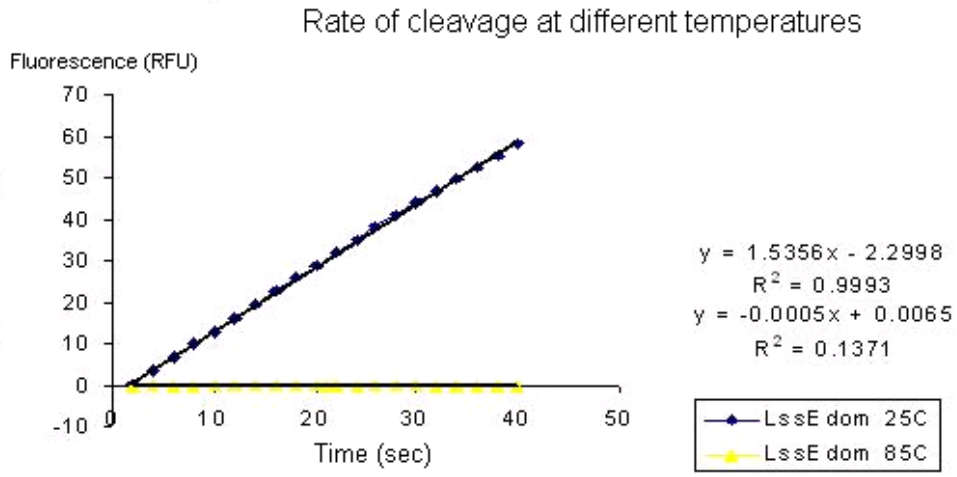
ZYGMUNT, W. A., BROWDER, H. P. & TAVORMINA, P. A. (1967) Lytic action of lysostaphin on susceptible and resistant strains of Staphylococcus aureus. *Can J Microbiol*, 13, 845-53.

ZYGMUNT, W. A. & TAVORMINA, P. A. (1972) Lysostaphin: model for a specific enzymatic approach to infectious disease. *Prog Drug Res*, 16, 309-33.

Appendix 1

Additional graphic for Chapter 3 (Section 3.4.1.).

Activities of LssEdom before and after denaturation by treatment at high temperature (85ºC) were assessed by FRET assay at 37ºC. The MV11 substrate concentration used was 2.0 μM taken from a 0.19 mM stock solution, in 1 mM sodium phosphate (pH 7.0) buffer. The LssEdom concentration used in each case was 0.100 μM.



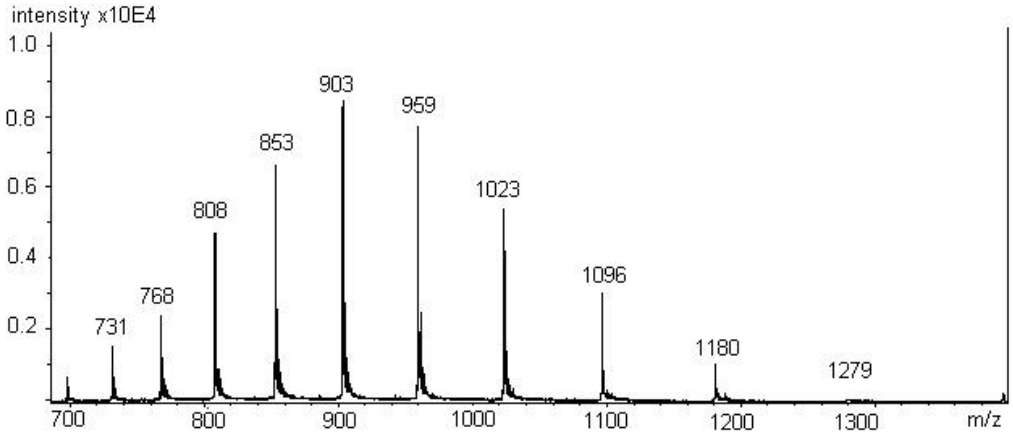
Cleavage of 2.0 μM MV11 by 0.100 μM LssEdom, before (25ºC) and after denaturation at 85ºC.

The rate of cleavage is approaching zero when LssEdom is denatured by treatment at high temperature (85ºC).

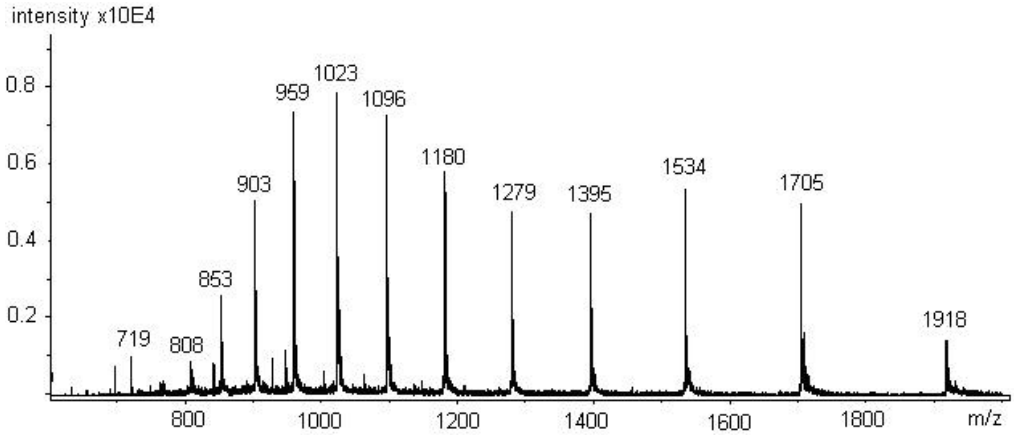
Appendix 2

Additional spectrum for Chapter 4 (Section 4.5.3.).

Examination of LssEdom protein sample by MS (ESI positive ion) on the MicroToF did show some peaks additional to the multiply charged ions of the protein, in particular a relatively intense singly charged ion at m/z=719. Intensity of the ion at m/z=719 is dependent on the voltage difference between the Capillary Exit and the Skimmer 1.



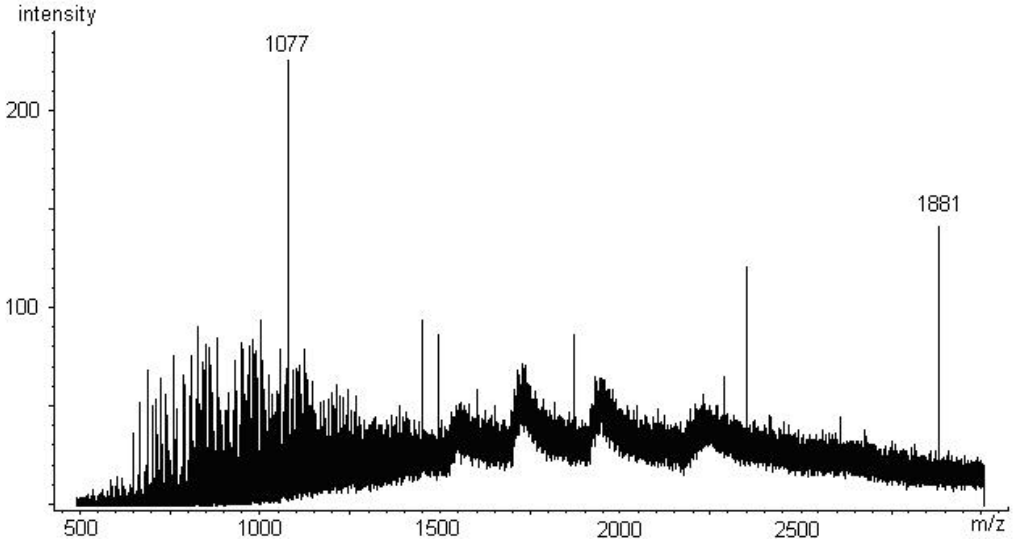
Mass spectrum (ESI positive ion) at 80 V of LssEdom sample used for NMR studies.



Appendix 3

Additional spectrum for Chapter 4 (Section 4.5.3.).

Examination of LssEdom protein sample by MS (negative ion) on the Apex FTMS system, did not show the ion at m/z=719.



Appendix 4

Poster presentation

Lysostaphin: an enzybiotic active against Staphylococci

Sophie Rochette[*,#], Philip Bardelang[#], Richard James[#] and Neil R. Thomas[*]

Presented at the conferences organised by the Royal Society of Chemistry: "Antibiotics, where now?" (2008), London, U.K.

Appendix 5

Publication

Bardelang, P., Vankemmelbeke, M., Zhang, Y., Jarvis, H., Antoniadou, E., Rochette, S., Thomas, N.R., Penfold, C.N., James, R.

Design of a polypeptide FRET substrate that facilitates study of the antimicrobial protease lysostaphin.
Biochem J. (2009), 418, 615-24.

Appendix 6

Publication

Rochette, S., Hopper, J., James, R., Oldham, N., Thomas, N.R.
Binding interactions of lysostaphin with peptidoglycan fragments,
In progress