NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTAMENT OF CHEMISTRY

MARLENE PIMENTEL LEITE

BSc in Biochemistry

STRUCTURAL CHARACTERIZATION OF THE C-TERMINAL DOMAIN OF ENDOLYSIN PLYMW2 AND INTERACTION WITH THE PEPTIDOGLYCAN OF *STAPHYLOCOCCUS AUREUS*

MASTER IN BIOCHEMISTRY

NOVA University Lisbon

October, 2022

NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTAMENT OF CHEMISTRY

MARLENE PIMENTEL LEITE

BSc in Biochemistry

STRUCTURAL CHARACTERIZATION OF THE C-TERMINAL DOMAIN OF ENDOLYSIN PLYMW2 AND INTERACTION WITH THE PEPTIDOGLYCAN OF *STAPHYLOCOCCUS AUREUS*

MASTER IN BIOCHEMISTRY

NOVA University Lisbon

October, 2022

NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTAMENT OF CHEMISTRY

STRUCTURAL CHARACTERIZATION OF THE C-TERMINAL DOMAIN OF ENDOLYSIN PLYMW2 AND INTERACTION WITH THE PEPTIDOGLYCAN OF *STAPHYLOCOCCUS AUREUS*

MARLENE PIMENTEL LEITE

BSc in Biochemistry

Adviser: Doctor Jorge da Silva Dias
Investigator, NOVA University Lisbon

Co-advisers: Professor Doctor Rita Gonçalves Sobral
Assistant professor, NOVA University Lisbon

Examination Committee:

Chair: Professor Doctor Pedro António de Brito Tavares,
Assistant professor, NOVA University Lisbon

Rapporteurs: Doctor Luís Pedro Gafeira Gonçalves,
Investigator, ITQB

Adviser: Doctor Jorge da Silva Dias,
Investigator, NOVA University Lisbon

MASTER IN BIOCHEMISTRY

NOVA University Lisbon
October, 2022

Structural characterization of the C-terminal domain of endolysin PlyMW2 and interaction with the peptidoglycan of *Staphylococcus aureus*

ACKNOWLEDGMENTS

Em primeiro lugar quero agradecer ao meu orientador Dr. Jorge Dias, por me ter aceite e ter-me dado esta oportunidade de aprender e evoluir. Obrigada por toda a disponibilidade, paciência e ajuda!

Também quero agradecer a minha coorientadora Dra. Rita Sobral por toda a ajuda que me deu durante este trabalho, e por estar sempre disponível.

Obrigada à UCIBIO, ao Departamento de Química e à NOVA School of Science and Technology da Universidade Nova de Lisboa por providenciarem todos os recursos e espaços necessários ao desenvolvimento deste trabalho.

Quero também agradecer ao Prof. Eurico Cabrita por uma vez mais me receber no seu laboratório e de o disponibilizar para que este trabalho fosse realizado. Obrigada a todos os membros de laboratório 302 por me receberem tão bem, por estarem sempre disponíveis para me ajudarem e pelas boas partidas de cartas.

Obrigada aos membros do laboratório 333, por me receberem sempre muito bem, por estarem sempre prontos a ajudar e bem dispostos.

Obrigada à minha família, por todo o apoio que sempre me deram independentemente de estar a mais de 1000 km de distância e as saudades serem imensas.

Obrigada "As do Piso do Meio" por tudo, em especial à Alícia Candeias, à Joana Galhano e à Marta Vieira por todas as gargalhadas e bons momentos partilhados. De certo que tornaram esse percurso mais leve.

Também quero deixar um agradecimento especial ao Vicente Almeida, por ser um amigo incansável, por ter sempre uma palavra motivadora para me dar e pela longas conversas.

Por fim, e não menos importante, obrigada Jorge Fernandes. Obrigada por todo o apoio e motivação. Obrigada pela motivação extra nos dias em que estive mais em baixo.

Obrigada de coração!

ABSTRACT

Staphylococcus aureus (*S. aureus*) is a major pathogen with the ability to acquire resistance to most antibiotics and a threat for global health. *S. aureus* is remarkably capable of developing resistance to commonly used antibiotics and novel approaches to treat these infections are urgent. The phage lytic enzymes or endolysins are alternatives with high potential for antibacterial therapy.

PlyMW2 lysin found in a prophage in *S. aureus* subspecies MW2 has two structural/functional domains: a CHAP domain and a divergent cell wall-binding domain (CBD) with no known structural homologs. The main goal is to determine how this unique CBD interacts with the *S. aureus* cell wall, specifically with its peptidoglycan. The gene for CBDMW2 was amplified from genomic DNA and cloned in modified pET vectors. The overexpressed proteins were purified and used for the structural and interactions studies. Peptidoglycan was extracted, purified from *staphylococcal* cultures and digested with muralytic enzymes. The resulting muropeptides were purified by reverse phase-HPLC and analyzed and characterized by NMR and ESI-MS.

The interaction of CBDMW2 with *S. aureus* cells was analyzed using Fluorescence microscopy and showed that the GFP-CBDMW2 binds to localized regions of the cell wall, likely where the peptidoglycan is more accessible. MicroScale Thermophoresis was used to measure the interaction between the CBDMW2 and a mixture of muropeptide digested with mutanolysin and amidase, yielding fragments containing peptides with different lengths. An apparent K_D of 211±110 nM was obtained showing that CBDMW2 binds to the peptide component with high affinity. Backbone resonance assignments were achieved from NMR 3D-experiments and used to analyze the ^{1}H-^{15}N-HSQC titration of the CBDMW2 with a monomeric muropeptide. CBDMW2 NMR structure revealed SH3b-type fold although low/none sequence homology, with striking differences on the loop regions. The binding region identified for the monomeric muropeptide reveals unique features involving non-conserved amino acids with structural and functional homologs.

Keywords: *Staphylococcus aureus;* antibiotic resistance; endolysin; CBDMW2 domain

RESUMO

Staphylococcus aureus (*S. aureus*) é um agente patogénico com a capacidade de adquirir resistência à maioria dos antibióticos, sendo uma grande ameaça para a saúde global. *S. aureus* é capaz de desenvolver resistência aos antibióticos comummente utilizados, sendo urgentemente necessário novas abordagens contra as infeções. As enzimas fagolíticas (endolisinas) são alternativas com elevado potencial na terapia antibacteriana.

A lisina PlyMW2 encontrada num profago da subespécie *S. aureus* MW2 tem dois domínios estruturais/funcionais: um domínio CHAP e um domínio de ligação à parede celular (CBD), sem homólogos estruturais conhecidos. O principal objetivo é determinar como o CDB interage com a parede celular de *S. aureus*, especialmente com o peptidoglicano. O gene do CBDMW2 foi amplificado do ADN genómico e clonado em vetores pET modificados. As proteínas sobre-expressas foram purificadas e utilizadas para estudos estruturais e interação. Além disso, peptidoglicano de *S. aureus* foi extraído, purificado, e digerido com hidrólases de peptidoglicano. Os muropéptidos foram purificados por cromatografia de fase reversa e analisados e caracterizados por RMN e ESI-MS.

A interação do CBDMW2 com células de *S. aureus* foi analisada utilizando Microscopia de Fluorescência, e mostrou que a GFP-CBDMW2 liga-se a regiões específicas na parede celular, possivelmente onde o peptidoglicano é mais acessível. *MicroScale Termophoresis* foi utilizada para medir a interação entre o CBDMW2 e uma mistura de muropéptidos, obtendo-se um K_D aparente de 211±110 nM, demonstrando uma alta afinidade entre as duas moléculas. As atribuições das ressonâncias da cadeia principal foram obtidas através de experiências 3D RMN e utilizadas para analisar o ^{1}H-^{15}N-HSQC da titulação do CBDMW2 com um muropéptido monomérico. A estrutura do CBDMW2 revelou ser do tipo SH3b, embora tenha baixa/nenhuma homologia de sequência, com diferenças notáveis nas regiões do *loop*. A região de ligação identificada para o muropéptido demonstra características únicas envolvendo aminoácidos não conservados nos homólogos estruturais/ funcionais.

Palavras chave: *Staphylococcus aureus*; resistência a antibióticos; endolisina; domínio CBDMW2

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

^{1}H	Proton
^{13}C	Carbon 13 isotope
^{15}N	Nitrogen 15 isotope
aa	Amino Acids
Abs	Absorbance
Amp	Ampicillin
APS	Ammonium Persulfate
BME	2-Mercaptoethanol
CARA	Computer-Aided Resonance Assignment
CBD	Cell Wall-binding Domain
CHAP	Cysteine Histidine-Dependent Amidohydrolase/Peptidase Domain
COSY	Correlation Spectroscopy
CSP	Chemical Shift Perturbation
Da	Dalton
DNA	Deoxyribonucleic Acid
dNTPs	Deoxynucleotide triphosphates
EAD	Enzymatically Active Domain
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid

ESI-MS	Electrospray Ionization Mass Spectrometry
GFP	Green Fluorescent Protein
GlcNAc	N-acetylglucosamine
HSQC	Heteronuclear Single Quantum Correlation NMR experiment
IPTG	Isopropyl-β-D-thiogalactopyranoside
IMAC	Immobilized Metal ion Affinity Chromatography
Kan	Kanamycin
LB	Lysogeny Broth
MRSA	Methicilin-resistant *Staphylococcus aureus*
MST	MicroScale Thermophoresis
MurNAc	N-Acetyl-Muramic Acid
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
OD	Optical Density
PBP	Penicillin-Binding-Protein
O/N	Overnight
PBS	Phosphate-Buffered Saline
PCR	Polymerase Chain Reaction
PG	Peptidoglycan
RPM	Rotations Per Minute
RT	Room Temperature
RP-HPLC	Reversed-Phase-High-Performance Liquid Chromatography
SDS-PAGE	Sodium Dodecyl-Sulfate Polyacrylamide Gel Electrophoresis
S. aureus	*Staphylococcus aureus*
TEMED	N,N,N′,N′-Tetramethylethylenediamine
TEV	Tobacco Etch Virus nuclear-inclusion-a endopeptidase (EC 3.4.22.44.)

TFA Trifluoracetic Acid

TOCSY Total Correlated Spectroscopy

TSB Tryptic Soy Broth

TSP 3-(trimethylsilyl) propionic-2,2,3,3-d4 acid

UV-Vis Ultraviolet-visible spectroscopy

Amino acids		
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartate	Asp	D
Cysteine	Cys	C
Glutamate	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysin	Lys	K
Methionine	Met	M
phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	T	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

1

INTRODUCTION

1.1 *Staphylococcus aureus*: an overview

Staphylococcus aureus (*S. aureus*) is a gram-positive bacterium that belongs to the *Staphylococcaceae* family, has the shape of a coccus (**Figure 1.1**), and is a facultative anaerobe, that can grow through aerobic respiration or fermentation and reproduces asexually via binary division [2].

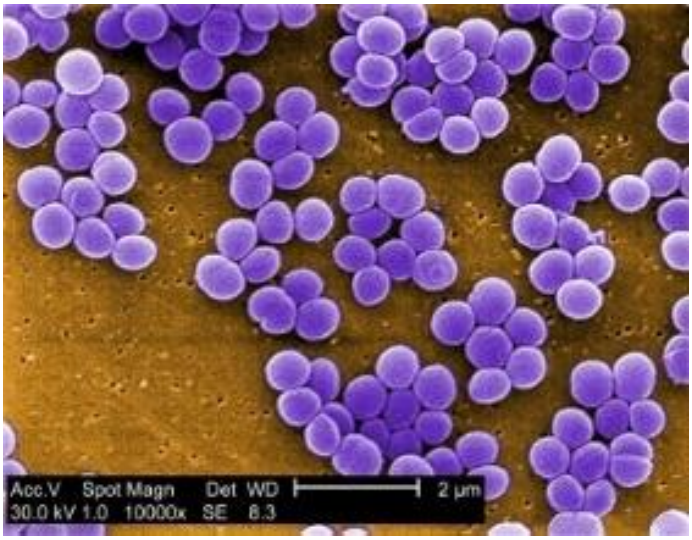


Figure 1.1. *Staphylococcus aureus* by a digitally-colorized scanning electron microscopic image. Photo Credit: Janice Carr Content Providers(s): CDC/ Matthew J. Arduino, DRPH; Janice Carr - This media comes from the Centers for Disease Control and Prevention's Public Health Image Library (PHIL), with identification number #6486.

S. aureus can act as a commensal or a pathogenic agent, inhabiting the upper respiratory tract and the skin of healthy individuals, without developing an infection, however it can become an opportunistic pathogen in the bloodstream. Thus, causing various illnesses, such as infections of the skin, soft tissue, bone (osteomyelitis), blood (bacteremia), lung (pneumonia) and toxic shock syndrome. An estimated 20% of people are persistent nasal *S. aureus* carriers, and about 30% are intermittent carriers [3, 4].

The major concern associated with *S. aureus* is its remarkable ability to acquire resistance to antibacterial agents, such as the methicillin-resistant strains, known as MRSA, which complicates treatments.

1.1.1 Methicillin-Resistant *S. aureus*

S. aureus is recognized for its ability to easily acquire and maintain resistance to antibiotics, due to its flexible genome, which allows this organism to easily adapt to external pressures caused by antibiotics, thus becoming a constant challenge to public health care.

In 1940, the mortality rate due to *S. aureus* infections was almost 80%, to overcome this problem, penicillin was used against *S. aureus*. Penicillin is an antibiotic with a β-lactam nucleus, which resembles the terminus D-alanyl-D-alanine bond of the pentapeptide. This antibiotic was used to inhibit transpeptidation activity by penicillin-binding-proteins (PBPs) thus interfering with the cross-linking reaction. Approximately ten years after penicillin was first introduced, the first cases of antibiotic resistance were recorded. The mechanism of resistance was associated with the production of the enzyme β-lactamase, which hydrolyzes the β-lactam bonds, rendering the antibiotic inactive [3].

To overcome the difficulties imposed by penicillin-resistant strains, methicillin, a synthetic penicillin, was introduced in 1959. Methicillin is resistant to the action of the enzyme β-lactamase, due to the methoxy groups that produce a steric hindrance around the amide bond, reducing its affinity for β-lactamases. However, within a year, Methicilin-resistant *S. aureus* (MRSA) began to appear. These strains increased quickly becoming one of the main causes of infections worldwide, acquiring resistance through the incorporation of a *mecA* gene. A penicillin-binding protein 2a (PBP2a) homolog is encoded by this gene. The PBP2a has a low affinity for β-lactams, is similar to PBPa, and can take over transpeptidation reactions, overcoming peptidoglycan biosynthesis inhibition and allowing resistance to β-lactam antibiotics [3, 4].

In the first years, MRSA infections were mostly associated with hospitalized patients, the health care-associated MRSA (HA-MRSA), however, shortly after the multidrug-resistant strains of MRSA emerged in the community, the community-associated MRSA (CA-MRSA). Lower rates of clindamycin resistance at first served as a defining characteristic of CA-MRSA isolates, but nowadays, the genotypic differences are almost homogenized. The emergence of additional resistance mechanisms has occurred in both MRSA subtypes [4, 5, 6].

High rates of morbidity, mortality, and increasing treatment costs were the results of the advent of MRSA strains. One approach to combat resistance is bacteriophage derived enzymes, also known by endolysins, and other peptidoglycan (PG) hydrolases with the capability to hydrolyze bonds in cell wall components.

1.1.2 Peptidoglycan

S. aureus is a gram-positive bacterium, and thus possesses a thicker layer of peptidoglycan than a gram-negative bacterium in its cell wall. The cell wall is crucial for the survival of the bacterium, as it has very important functions in preserving cell integrity, for example resisting turgor pressure, but also facilitating the host-pathogen interactions. It is composed mostly of peptidoglycan, a polysaccharide constituted glycan chains (N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc) linked together by β-1,4 glycosidic bonds) cross-linked by stem peptides (L-Ala-D-iso-Gln-L-Lys-D-Ala-D-Ala) and pentaglycine bridges [7, 8]. The glycine peptide binds the D-Alanine at position 4 (after hydrolysis of the D-Ala residue at position 5) to the L-lysine side chain in an adjacent peptide. This layer has anchored to it, the teichoic acids that are composed of glycerol-phosphate and ribitol-phosphate, with a negative charge establishing covalent bonds with the peptidoglycan (**Figure 1.2**).

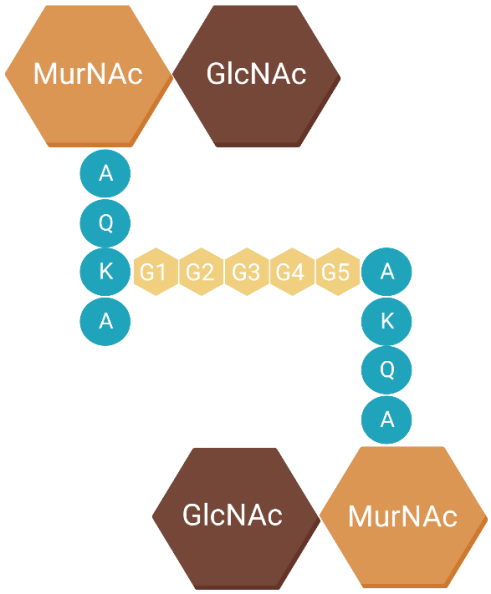


Figure 1.2. *S. aureus* peptidoglycan structure: alternate residues of N-Acetylmuramic acid and N-Acetylglucosamine with β-1,4 linkages, the stem peptide (L-Ala-D-iso-Gln-L-Lys-D-Ala) and the pentaglycine bridge that makes the cross link between different stem peptides.

Peptidoglycan synthesis occurs in three compartments: the first one is the cytoplasm, where the synthesis is carried out by the Mur ligase family of enzymes (MurA-MurF), the second compartment is the cytoplasmic membrane, and the last one is the cell wall (**Figure 1.3**).

During the synthesis, first occurs the transformation of fructose-6-phosphate (F6P) into glucosamine-6-phosphate (GlcN6P) by the aminotransferase GlmS. Then, phosphoglucosamine mutase GlmM and uridyltransferase GlmU60, transforms GlcN6P into N-acetylglucosamine (UDP-GlcNAc). Next, the UDP-GlcNAc is converted into N-acetylmuramic acid (UDP-MurNAc), by the enzymes

MurA and MurB. After the five amino acids of the stem peptide are added to UDP-MurNAc, forming UDP-MurNAc- L-Ala-D-Glu-L-Lys-D-Ala-D-Ala, by the Mur ligases (MurC-F). The enzyme MurC adds L-Ala, MurD adds D-Glu, MurE adds L-Lys, and MurF adds the final two amino acids D-Ala [9].

Afterward, the UDP-MurNAc-stem peptide binds to the interior of the cytoplasmic membrane, forming lipid I. Lipid II is synthesized by the addition of UDP-GlcNAc, which is catalyzed by the MurG enzyme from lipid I. Then, the five glycines are added to L-Lys. This addition is catalyzed by the enzymes FemX, FemA and FemB. Next follows the amidation of the D-Glu residue in lipid II producing a D-iso-Gln residue, catalyzed by the MurT enzyme [10]. Afterwards the lipid II is transferred to the outer part of the cytoplasmic membrane, where the final step occurs [9].

In the cell wall occurs the incorporation of lipid II-Gly_5 into the peptidoglycan layer through transglycosylation and transpeptidation reactions. These reactions are catalyzed by the penicillin binding protein (PBP). As these proteins are crucial for the final stage of peptidoglycan synthesis, they became an important target of β-lactam antibiotics [9, 11]. Therefore, any inhibition in peptidoglycan biosynthesis or any specific degradation will result in bacterial cell lysis [12].

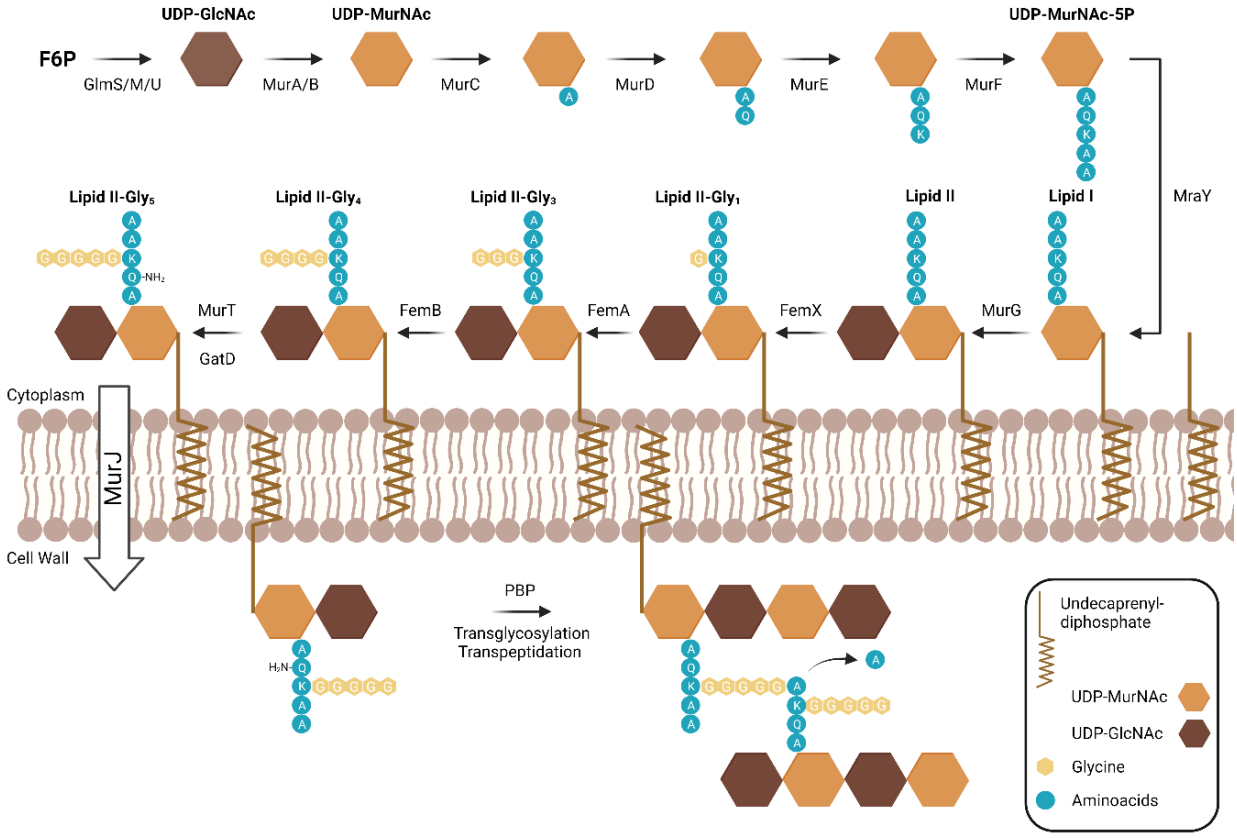


Figure 1.3. Schematic representation of *S. aureus* peptidoglycan biosynthesis. Adapted from [9].

Endolysins, which are bacteriophage derived enzymes with the potential to hydrolyze bonds in cell wall components, are another strategy for fighting *S. aureus*. These hydrolytic enzymes can be used against antibiotic-resistant bacterial pathogens such as MRSA.

1.2 Bacteriophages

Bacteriophages (or phages) are viruses that only infect bacterial cells. All phages are constituted of genetic material enclosed inside of a protein capsid. Phages are highly species-specific towards their host cell and their ability to infect and kill infectious bacterial cells supports their potential as a supplement or replacement of antibiotic agents. Additionally, phages are the most abundant biological entities on earth, and can be found in almost every environment [13].

Their mechanism of infection consists in the attachment of the phage to the host cell and the entry of their genome into the host cell, which will control the host cellular machinery in order to replicate. Phages can replicate through two types of lifecycles that differ in the mechanism of DNA replication: the lytic cycle, where, DNA replicates separately from the host DNA, and the lysogenic cycle, where, the viral DNA is incorporated into the host DNA (**Figure 1.4**). For a successful infection to occur, an essential class of phage-encoded enzymes with antibacterial capability are needed, known as the phage lytic enzymes.

In order to infect the host cell, the phage degrades the cell wall locally with virion-associated peptidoglycan hydrolase, a phage lytic enzyme or with lytic structural proteins. Following replication, two important phage lytic enzymes are expressed, the holins and endolysins. The holins are pore-forming oligomers, located at the cytoplasmic membrane, that help to permeabilize the membrane, and release the endolysin to the periplasm. Thus, endolysins have the capacity to extensively degrade the peptidoglycan layer from the inside out, leading to a sudden osmotic lysis of the host cell and release of phage progeny (lytic cycle) [14, 15]. Meanwhile, the lysogenic cycle is a temperate and non-virulent infection, considering that the host cell can proceed with normal reproduction, transmitting the phage DNA to its daughter cells during bacterial cell division (prophage), until the lytic cycle is triggered by an environmental stress [16].

Bacteriophages have been a great approach to use in therapeutics, ever since their discovery, with the phage lytic enzymes being an alternative with a high potential for antibacterial therapy.

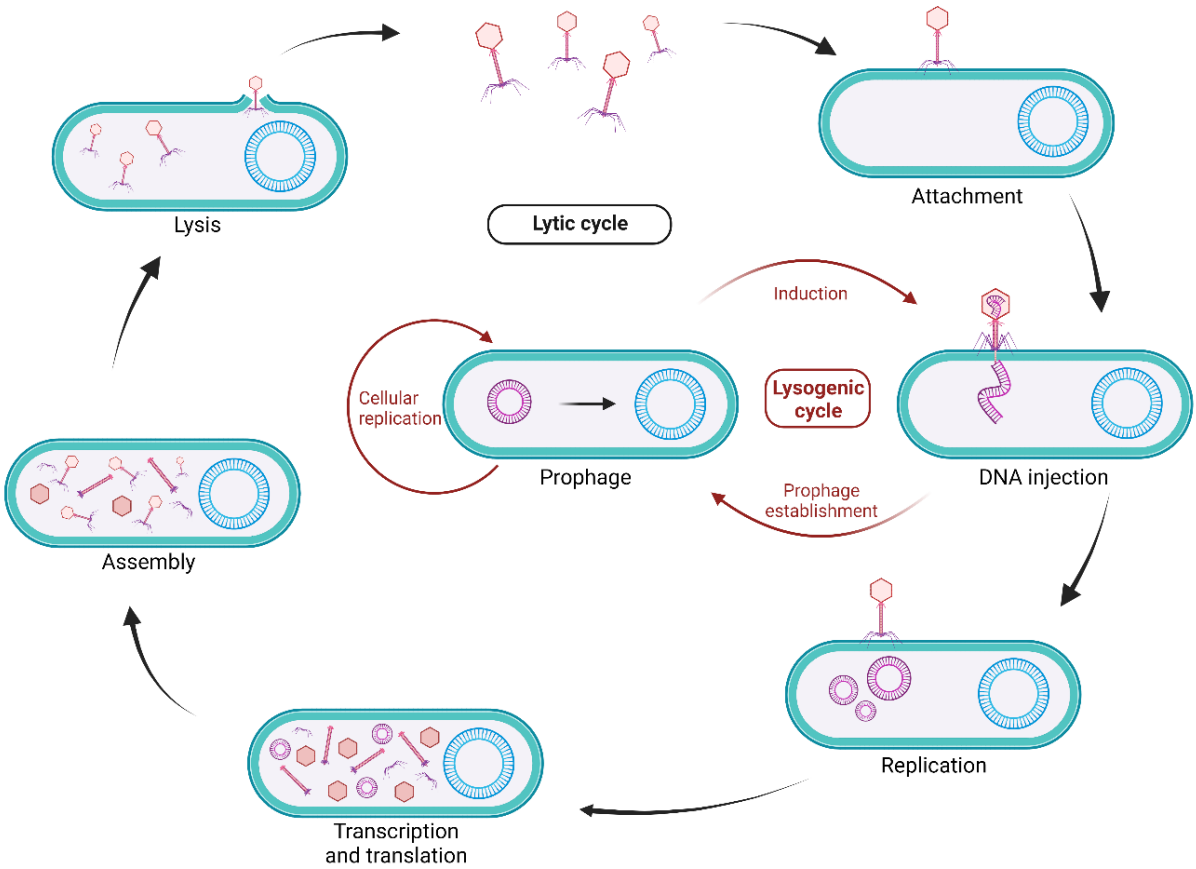


Figure 1.4. Phage life cycles: lytic and lysogenic cycle. In the lytic cycle, the phage will initially interact with the receptors on the host cell, attach to it and inject its genome. Then the phage replicates, and produces new viral offspring. Then they are released from the infected cell using holins and endolysins to lyse the host cell. In the lysogenic cycle, the viral genome is termed a prophage and replicates in tandem with the host DNA to spread to succeeding generations. Prophages can leave the lysogenic state when under stress and produce more virions that are expelled from the bacterium. Adapted from [17].

1.2.1 Lytic enzymes

The phage lysins or endolysins are PG hydrolases expressed by bacteriophages, which are related with the cell lysis during the phage replication.

Endolysins that act against Gram-positive bacteria present a modular structure: a C-terminus region with a cell wall-binding domain (CBD) and a N-terminus region with one or two enzymatically active domains (EADs) (**Figure 1.5**). The domains are linked by a flexible linker with variable size [18]. The EAD specifically cleaves certain bonds in the peptidoglycan of the host, acting as an enzyme, and the CBD function is to recognize and bind tightly to different epitopes in the cell wall, helping the catalytic effect of the EAD.

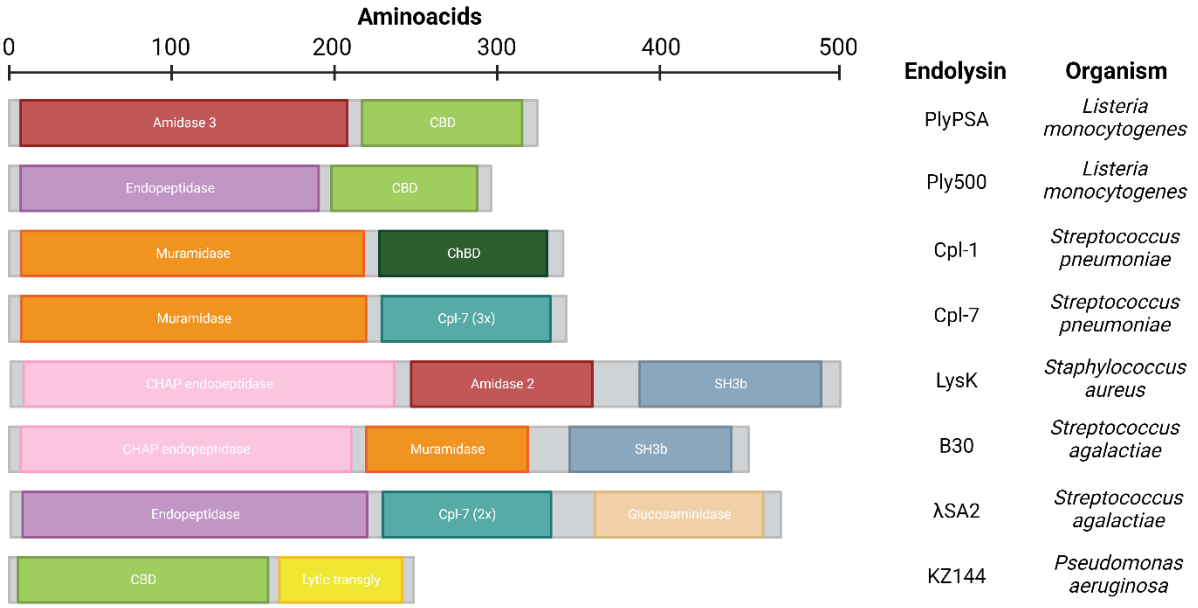


Figure 1.5. Representation of the modular architecture of a selected number of phage endolysins. Scale bar indicates aminoacid number. Cpl-7: Cpl-7-like cell wall-binding domain; ChBD: Choline binding domain; SH3b: Bacterial Src homology 3 domain. Adapted from [19].

Depending on the specific peptidoglycan bond that the EAD is targeting, the endolysins are divided into various groups: (i) N-acetyl-D-muramidases, transglycosylases, (ii) N-acetyl-D-glucosaminidases, (iii) N-acetylmuramoyl-L-alanine amidases, (iv) interpeptide bridge endopeptidases and (v) L-alanoyl-D-glutamate endopeptidases (**Figure 1.6**). (i) N-acetyl-D-muramidase and the transglycosylase cleaves the N-acetylmuramyl-β-1,4-N-acetylglucosamine bond. (ii) The glycosidic bond, N-acetylglucosaminyl-β-1,4-N-acetylmuramine is hydrolized by the N-acetyl-D-glucosaminidase. (iii) The bond between the L-Ala of the stem peptide and MurNAc is hydrolyzed by N-acetylmuramoyl-L-alanine amidase. (iv) The interpeptide bridge endopeptidase cleaves in between the bond of two glycines in the pentaglycine peptide. (v) L-alanoyl-D-glutamate endopeptidase cuts in the stem peptide, between the L-Lys and L-Ala. All endolysins are hydrolases, except for the lytic transglycosylases, and the glucosaminidases and muramidases are classified as glycosidases [20].

Metal-dependent PG amidases and endopeptidases are examples of other enzymes classified by their cleavage mechanism and not by the specific bond they cleave in the peptidoglycan. The Cysteine, Histidine-dependent Amidohydrolases/Peptidases (CHAP) domain is one of them. It depends on two invariant residues, a cysteine and a histidine, which compose the active site of these proteins. Some CHAP domains have amidase and peptidase activity and others are endopeptidases. The active site of this domain is located in a cleft between a small half β-barrel packed and a cluster of two helices (linked by a long loop), and comprises a conserved cysteine and histidine residues that are responsible for the catalytic activity [19, 21, 22].

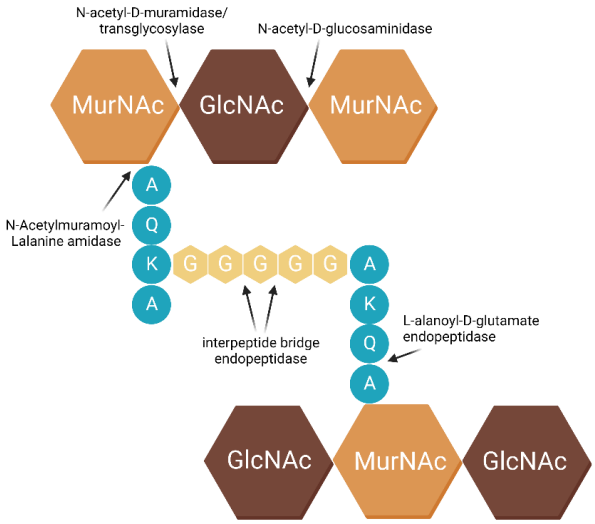


Figure 1.6. Structure of *S. aureus* peptidoglycan and identification of some of the muralytic activities. The muralytic enzymes can be divided in: endopeptidases, amidases, muramidases and glucosaminidases. Adapted from [20].

Besides the characteristics domains of the N-terminus, the C-terminus domains, the CBD, also have specific features. As previously mentioned, the CBD recognizes and binds specific ligands in the cell wall and targets the catalytic domain to its substrate, influencing the activity range.

The endolysins CBDs that recognize conserved components of the cell wall from various hosts, usually present a similar sequence and structure. Hence, the folds of the CBDs could be grouped into four main classes: (i) Choline-binding modules; (ii) Bacterial SH3 (SH3b) domains; (iii) Three-helix bundles; (iv) α/β multimers.

The (i) choline-binding modules ChBDs are versatile structures that attach to the choline-containing teichoic acid of the pneumococcal cell wall. These domains contain two sub-domains: superhelical moiety in the the N-terminus, formed with four ChBr repeats to create the superhelix, and between these repeats are the choline-binding positions with conserved tryptophan residues, and the C-terminus with β-sheet. (ii) Bacterial SH3 (SH3b) domains tend to vary, presenting different number of residues, β-strands and linker lengths, but overall, they are formed by a SH3 β-barrel fold, with five to seven β-strands, normally positioned as two antiparallel β-sheets, connected by loops. The SH3b domains can be involved in the binding of specific ligands with high affinity, mostly known to recognize and bind specifically to the pentaglycine bridge (transpeptidation peptide), but others showed to bind to another ligands, for example to carbohydrates. Most of the staphylococcal phages have a SH3b-type CBD. (iii) Endolysins with a three-helix bundles domain contain three repeats of three helix bundle to bind to the *N*-acetyl-D-glucosaminyl-(β-1,4)-*N*-acetylmuramyl-L-alanyl-D-isoglutamine, aligning the EAD with the cleavage sites in the peptide and the sugar strand. Finally, (iv) α/β multimers domains involve a central parallel or four-stranded antiparallel β-sheet surrounded by two helices (**Figure 1.7**) [15, 19, 22]

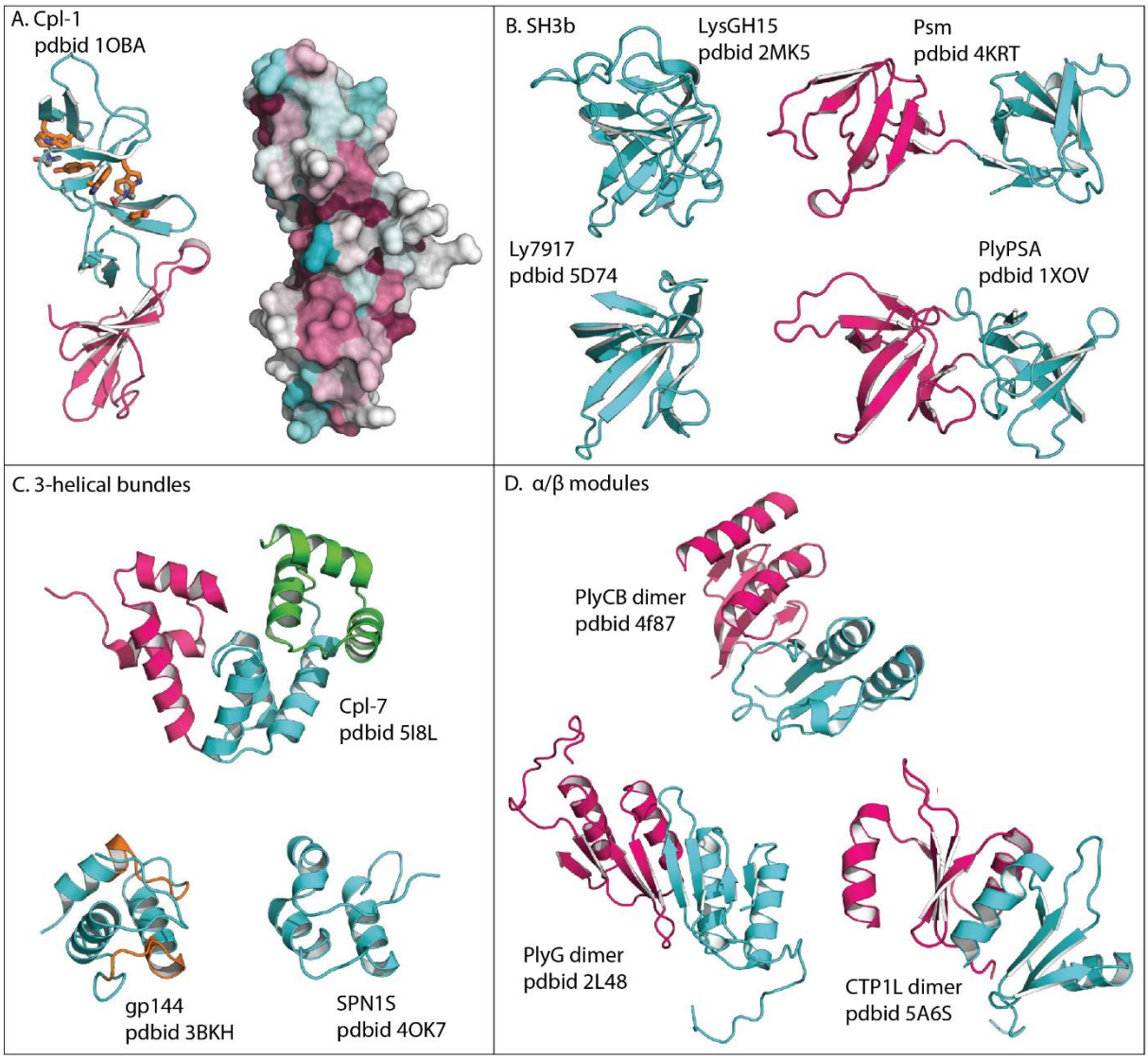


Figure 1.7. Main classes of CBD domains folds. (**A**) The choline-binding module example: Cpl-1. **(B)** Four bacterial SH3 domains examples: LysGH15, Psm, Ly7917 and PlyPSA. **(C)** Different three-helical bundles examples: Cpl-7, gp144 and SPN1S. **(D)** Three different examples of α/β modules: PlyCB dimer, PlyG dimer and CTP1L dimer. Adapted from [22].

In terms of applicability, endolysins were shown to be effective in a variety of areas, such as in the medicine, veterinary, agriculture, food and biotechnology fields (**Figure 1.8**).

Despite the identification of endolysins with broad-range lytic activity, most phage endolysins exhibit specificity to certain bacterial species, which is thought to be one of their most advantageous traits in the current era of widespread antibiotic resistance, by preventing selective pressure on commensal populations. Endolysins also bind and cleave highly immutable and conserved targets in host cell wall structures, and any alteration to these structures may be harmful to the host. Furthermore, a number of endolysins have two catalytic domains that hydrolyze several peptidoglycan linkages, which is thought to significantly lessen the possibility of resistance forming. Additionally, the modular design of endolysins presents a unique opportunity for protein engineering, enabling the

alteration of the bacteriolytic selectivity, solubility, activity and other physicochemical properties. The use of endolysins in combination with antibiotics is another strategy to increase their therapeutic activity while preventing the emergence of resistance [15, 18].

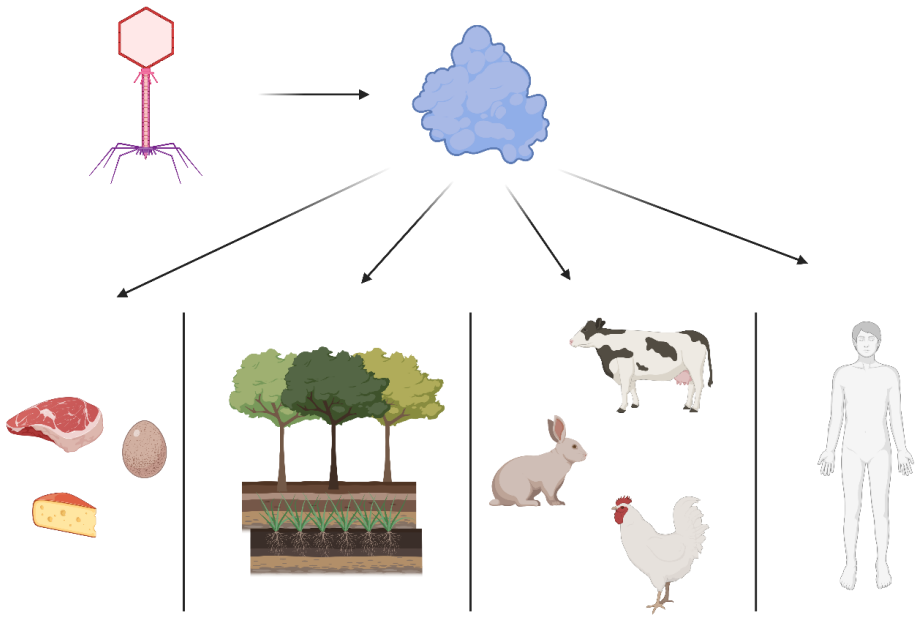


Figure 1.8. Endolysins from bacteriophages are used as antibiotics in a variety of industries. Endolysin can be engineered in the treatment of people, animals, and plants infections that are highly resistant to antibiotics. It also prevents the spread of antibiotic resistance through other food chains. Adapted from [18].

Even though endolysin therapy is a relatively new concept, some businesses in both Europe and America have made significant progress toward bringing their products to market. Endolysins like Staphefekt SA.100 by Micreos Human Health BV are already on the market and have been used to treat patients with chronic skin conditions provoked by *S. aureus* infections [18, 19]. These endolysins are applied in a topical cream which was shown to have a very targeted mechanism of action and did not harm commensal skin bacteria [23].

1.2.2 PlyMw2 Endolysin

S. aureus MRSA strains, a superbug that can infect a variety of sites on the human body, have become much more common over the past decade. The nosocomial infections are frequently challenging to treat and are commonly contracted by patients who are already ill [19]. As such, the study and identification of endolysins is an interest for public health improvement.

The PlyMW2 endolysin from the prophage found in the *S. aureus* subspecies aureus MW2 DNA, a CA-MRSA strain. This endolysin has two functional and structural domains, 251 amino acids (aa) long, a molecular weight of 29 kDa and a theoretical pI of 9.25. Also, contains a cysteine, histidine-dependent amidohydrolases/peptidase domain (CHAP endopeptidase, residues 1-150 aa), and a cell wall-binding (CBD) domain (residues 153-251 aa) (**Figure 1.9**).

Figure 1.9. Modular structure of the endolysin PlyMW2: a catalytic domain (CHAP endopeptidase) and a cell-wall-binding domain (CBD), connected by a linker (L).

This endolysin was found by searching for homologues of the N-terminus CHAP domain in the *S. aureus* genome, a unique domain for the phages that have *Staphylococcus* as a target, recognizing the pentaglycine bridge. On the other hand, the C-terminus domain does not have any homologues of sequence known, information given by a bioinformatic analysis using the Swiss-Model program [24]. Thus, this work will focus on the cell wall-binding domain, and how this domain interacts with the peptidoglycan.

1.3 Nuclear Magnetic Resonance

Nuclear magnetic resonance (NMR) was used to study the CBDMW2 domain: calculate its 3D structure and understand how the interaction between the CBDMW2 domain and the peptidoglycan of *S. aureus* is accomplished.

NMR is a non-destructive analytical method that depends on the magnetic properties of the atoms to provide information about molecular structure, chemical reaction rates, and diffusion processes. It is one of the few techniques available that can be used on all three states of matter, and enables to mimic physiological fluids or extreme conditions that leads to protein denaturation. Furthermore, it allows for he structure determination with atomic resolution, once the range of the NMR interactions is between 0.1 Angstrom and 10 Angstrom (1 nm).

The basis for NMR is the interaction of an externally applied radiofrequency radiation with the atomic nuclei. There is a net exchange of energy caused by this interaction, which leads to a change in an intrinsic property of the atomic nuclei called nuclear spin (I).

Depending on the isotope, the nuclear spin changes, but only atomic nuclei with I≠0 are detectable by this technique, and are called NMR-active nuclei, they can be ^{1}H, ^{2}H, ^{13}C and ^{15}N. When aligned to an external magnetic field (B_0), the nuclear spins can adopt two different orientations. The orientation parallel to the external magnetic field, corresponds to the lowest energy level of the

nucleus, and the antiparallel orientation, corresponds to the highest energy level of the nucleus. The difference between energy levels (ΔE) depends on the magnetic field (B_0) and the magnetogyric ratio (γ) and affects the sensitivity of the technique. To produce an NMR signal, it is necessary to provide energy corresponding to the transition between two spin states. The energy difference between the two states can be calculated, by equation (1). Where h is the Plank's constant and v is the resonant frequency (Hz) [25, 26].

$$\Delta E = hv = \frac{\gamma h B_0}{2\pi} \quad (1)$$

1.4 Objectives

In summary, this work aims to characterize the structure of the C-terminus domain of the endolysin PlyMW2, understand the molecular mechanisms involved in the interaction and specificity of CBDMW2 with the peptidoglycan of *S. aureus* using different techniques and identify the molecular determinants of the interaction between the C-terminus domain and the cell wall of *S. aureus*.

To this end, the main tasks defined for this work are:

1. Construction of recombinant plasmids for the endolysin, and its respective domains;
2. Expression and purification of isotopically labeled CBDMW2 with ^{15}N or ^{15}N and ^{13}C;
3. Expression and purification of CBDMW2 with a fluorescent tag for Fluorescence Microscopy and MicroScale Thermophoresis;
4. Determination of the 3D structure of the cell wall-binding domain of CBDMW2 by NMR;
5. Extraction and purification of peptidoglycan from *S. aureus* COL and enzymatic fragmentation;
6. Identification of molecular determinants for ligand or substrate recognition by the domain and activity using various biophysical and biochemical methods, with a particular emphasis on NMR:
   6.1. Titration of CBDMW2 with peptidoglycan fragments monitored by ^{1}H-^{15}N-HSQC;
   6.2. Identification of essential residues involved in peptidoglycan binding by 3D structure mapping of the most pronounced chemical shift perturbations (CSP).

2

METHODS

2.1 Strains, plasmids, and growth conditions

Table 2.1 lists the strains and plasmids used in this study. *E. coli* strains were propagated in Luria Bertani medium (LB, Liofilchem, Teramo, Italy) and *S. aureus* strains were propagated in Tryptic Soy Broth (TSB, Becton, Dickinson and Company, Franklin Lakes, USA) at 37 °C overnight (O/N) with agitation. The antibiotic ampicillin (100 µg/mL, Sigma-Aldrich) was used when working with the pET21c plasmids, and the antibiotic kanamycin (30 µg/mL, Sigma-Aldrich, St. Louis, USA) was used when working with the pET28 plasmids.

Table 2.1. Strains and plasmids used in this study.

Strain	Description	Source
Escherichia coli		
DH5α	*recA endA*1 *gyrA*96 *thi*-1 *hsdR*17 *supE*44 *relA*1 *F*80 Δ*lacZ*Δ*M15*	Invitrogen
BL21 (DE3)	F- *ompT hsdS*$_B$ (r_B-m_B-) *gal dcm* (DE3)	Invitrogen
Staphylococcus aureus		
COL	MRSA strain	Rockefeller University Collection
MW2	MRSA strain	Rockefeller University Collection

Plasmids	Description	Source
pET21c	Expression vector with T7 promoter and lac operator; C-terminus 6-histidine tail; Amp^r	Novagen
pET28-8HT-GB1-TEV	Expression vector with T7 promoter and lac operator; pET28a vector changed with an 8-histidine tail, GB1 protein, and a cutting site for TEV protease on N-terminus; Kan^r.	From the lab
pET21c-PlyMW2	pET21c cloned with the PlyMW2 protein by the traditional cloning method without tags; Amp^r.	This study
pET21c-CHAP	pET21c cloned with the CHAP domain, from PlyMW2, by the traditional cloning method with a His-tag in the C-terminus; Amp^r.	This study
pET28-8HT-GB1-TEV-CBDMW2	pET28-8HT-GB1-TEV cloned with the CBDMW2 domain, from PlyMW2, by the traditional cloning method with a His-tag in the N-terminus; Kan^r.	This study
pET28-8HT-GFP-TEV-CBDMW2	pET28-8HT-GB1-TEV-CBDMW2 cloned with the GFP, by the restriction-free cloning method with a His-tag in the N-terminus; Kan^r.	This study

* Kan^r: Kanamycin resistance; Amp^r: ampicillin resistance;

2.2 Genomic DNA extraction

The genomic DNA was extracted using the Wizard® Genomic DNA Purification Kit from Promega, readapting the Isolating Genomic DNA from Gram-Positive and Gram-Negative Bacteria protocol, with minor changes in the duration of some steps or the buffers used.

Genomic DNA extraction was conducted for *S. aureus* subspecies aureus MW2 (CA-MRSA). The culture was grown in TSB (Becton, Dickinson and Company, Franklin Lakes, EUA), and incubated O/N at 37 °C and 180 rpm.

Afterward, 1 mL of the culture was centrifuged at 15000 × g for 2 minutes to pellet the cells, and the supernatant was removed. Then, the pellet was resuspended in 100 µL of 100 mM Tris-HCl pH 8.0 with 1 µL of lysostaphin and 3 µL RNAse (10 mg/mL) and incubated at 37 °C for 120 minutes. Posteriorly, 600 µL of Nuclei Lysis Solution was added to the sample and incubated at 80 °C for 5 minutes to lyse the cells, then cooled to room temperature. In the next step, 200 µL of Protein Precipitation Solution was added to the cell lysate, mixed vigorously for 20 seconds, and incubated on ice for 10 minutes. After incubation, the sample was centrifuged at 15000 × g for 30 minutes at 4 °C. The resulting supernatant (with the DNA) was transferred to a clean Eppendorf containing 600 µL of isopropanol and incubated at room temperature for 15 minutes. Then, the sample was centrifuged at 15000 × g for 30 minutes at 4 °C. The supernatant was discarded, 600 µL of 70% ethanol was added to the pellet, and the tube was gently inverted to wash the DNA pellet. The sample was centrifuged at 15000 × g for 15 minutes at 4 °C, and the ethanol was carefully aspirated. The pellet was allowed to air-dry for 15 minutes, over a clean absorbent paper. The DNA extracted was rehydrated in 100 µL of TE buffer in incubation at 65 °C for 1 hour. Resulting DNA was stored at 4 °C for later use.

DNA samples were quantified by UV spectrophotometry using the NanoDrop ND-1000 (Thermo Fisher Scientific, Waltham, USA), obtaining for each sample the respective concentrations of 103.5 ng/µL and 107.3 ng/µL.

2.3 Construction of recombinant plasmids

Different plasmid constructions were produced to obtain plasmids to synthesize the full protein, the two domains individually, and the CBDMW2 domain with Green fluorescent protein (GFP). Various fragments from the genomic DNA of *S. aureus* subspecies aureus MW2 were amplified using different primers (**Table 2.2**) and cloned into the chosen plasmids (**Table 2.3**), using the traditional cloning method. For the construction of the CBDMW2 domain with GFP, restriction-free cloning (RF cloning) was used.

In the traditional cloning method (**Figure 2.1. A**), restriction enzymes are used to produce DNA fragments with end sequences complementary to the endings of the plasmid digested with the same enzymes, allowing them to be connected by a DNA ligase. In the digestion step, one or two restriction enzymes can be used. The most common method is performed with two different enzymes for both insert and vector, for directional cloning. To prevent self-ligation of the vector, dephosphorylation may be needed. The alkaline phosphatase removes the phosphate groups at the 5' ends in this reaction. The last step is the ligation of the insert with the vector.

RF cloning is independent of restriction enzymes (**Figure 2.1. B**), this way the technique allows to insert any sequence into any position in the plasmid, independent of restriction enzyme recognition sites. Thus, a pair of hybrid primers are used. These primers are meant to include complementary sequences for the insert and also for the target plasmid. In this technique, two polymerase chain reactions (PCR) are performed, the first one is to amplify the insert, and the resulting product is used as a "mega-primer" for the second PCR, using the target plasmid as a template. Lastly, DpnI is used to degrade any parental plasmid.

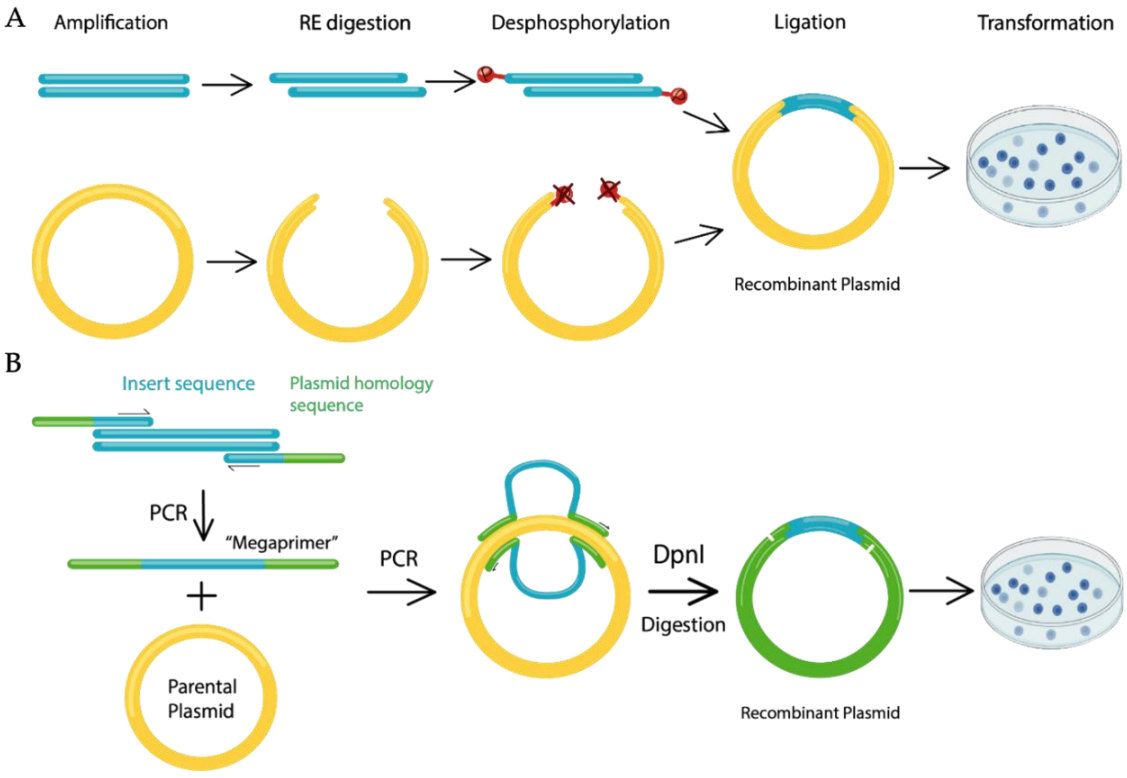


Figure 2.1. (A) Traditional cloning method scheme. **(B)** Restriction Free cloning scheme.

PCR amplifications were performed using the Phusion™ High-Fidelity DNA Polymerase (Thermo Fisher Scientific). FastAP Thermosensitive Alkaline Phosphatase (AP) (Thermo Fisher Scientific) was used for the dephosphorylation of the vector. The T4 DNA ligase (Thermo Fisher

Scientific) was used for the ligation reaction. The DpnI (Thermo Fisher Scientific) was used to cleave the plasmids with methylated recognition sites. The amplification and digestion products were analyzed by an 0.8% agarose gel, and purified using the NZYgelpure kit (NZYTech).

For the construction of the plasmid pET21c-PlyMW2, the full gene of the PlyMW2 protein was amplified, using 5-Nde-PhiMW2 and 3-Xho-PhiMW2 primers, and cloned into the pET21c plasmid (**Figure 5.1**, annexes). These primers used inserted restriction sites on the amplified fragment for NdeI and XhoI restriction enzymes. The plasmid pET21c-CHAP was built using the 5-Nde-PhiMW2 and 3-Xho-CHAPMW2 primers, to amplify the CHAP domain, and restriction sites for NdeI and XhoI were also inserted, while the resultant fragment was cloned into the pET21c. To produce the plasmid pET28-8HT-GB1-TEV-CBDMW2, 5-BmH1-Sh3bMW2 and 3-Xho-PhiMW2 primers were used to amplify the CBDMW2 domain gene and to insert restriction sites for BamHI and XhoI enzymes. This fragment was cloned into the pET28-8HT-GB1-TEV (**Figure 5.2**, annexes).

The NdeI, XhoI, and BamHI fast digest enzymes (Thermo Fisher Scientific) were used following the manufacturer's instructions for the restriction enzyme digestion. The mix for the plasmids contained 2 µL of 10x buffer for fast digest enzymes (Thermo Fisher Scientific), 1 µL of each enzyme (according to the plasmid), 1 µg of plasmid DNA and nuclease-free water to make 20 µL. For the fragment, the mix contained 3 µL of 10x buffer for fast digest enzymes (Thermo Fisher Scientific), 1 µL of each enzyme (according to the fragment), 0.2 µg of the fragment, and nuclease-free water to make 30 µL. Then, the samples were incubated at 37 °C for 5 minutes for plasmids and 60 minutes for fragments when using NdeI.

For the dephosphorylation of the vectors, the mix was composed of 1 µg of linear DNA, 2 µL of the 10x buffer for the AP, 1 µL of the AP, and nuclease-free water to make 20 µL. Then, the mix was incubated for 10 min at 37 °C, and after that, it went to heat for 5 minutes at 75 °C to inactive the enzymes.

At last, the ligation reaction was achieved by mixing 2 µL of 10x buffer for T4 DNA ligase, 1 µL of T4 DNA ligase, 8 µL of the plasmid, and 9 µL of the insert, and left at 4 °C O/N.

To construct plasmid pET28-8HT-GFP-TEV-CBDMW2, 5-pET28-HIS-GFP_Fw, and, 3-pET28-GFP-TEV_Rev primers were used, first to amplify the GFP gene, producing the "mega-primer" and used to insert the resultant fragment into the pET28-8HT-GB1-TEV-CBDMW2 plasmid. The primers were designed using the online tool RF-Cloning.org [27].

To the PCR product was added 1 µL of DpnI (10 U/µL) and it was left to incubate for 2 hours at 37 °C. Afterwards, the mix was stored at 4 °C O/N.

The recombinant plasmids were used to transform *E. coli* DH5α cells to perform a colony screening using PCR to verify its cloning products. This step is needed since, in the transformation reaction, the DNA used is a cloning product that can contain a mix of different things: a plasmid

without any fragment, a plasmid with another fragment than the desired, and a plasmid with the wanted fragment. Cells that incorporate the plasmid alone or the plasmid with the incorrect fragment can also grow because they contain the antibiotic resistance gene, and this step doesn't guarantee that all the colonies that grow have only the desired construct. Thus, the colony screening identifies transformed colonies that contain the right insert. The colony screening method analyzes some cells from the colony and requires the primers used to specify the fragment's amplification.

The NZYTaq II DNA polymerase (NZYTech) was used in the colony screening. First, some random colonies were chosen and boiled in 10 µL of nuclease-free water for 10 minutes. Afterward, the PCR mixture was added to the boiled colonies. All the colony's screening products were analyzed with an 0.8% agarose gel (**Table 5.1**, annexes), and the ones that showed a band, were sent to sequencing done by STAB Vida (STAB Vida, Almada, Portugal). To prepare the samples for the sequencing, each colony was selected through an agarose gel inoculated in 5 mL of LB medium (**Table 5.2**, annexes) and incubated at 37 °C with agitation O/N. The DNA from each culture was extracted with the kit ZR Plasmid mini prep-classic (Zymo Research, Orange, CA, EUA). The positive results from sequencing were used to transform *E. coli* BL21(DE3) cells.

Table 2.2. Primers used in this study.

Primer	Nucleotide Sequence (5'-3')	Annealing Temperature	Source
5-Nde-PhiMW2	gcgcatatgaaaacatacagtgaagc	56.92 °C	This study
3-Xho-PhiMW2	gtgctcgagttaaaacacttctttcac	56.50 °C	This study
3-Xho-CHAPMW2	gtgctcgagtttaactttatttgtgatc	54.78 °C	This study
5-BmH1-Sh3bMW2	ataggatccgctcaaaaagcagtcgg	60.78 °C	This study
5-pET28-HIS-GFP_Fw	accatcaccatcaccatcacggatctatgagcaaaggagaagaacttttcac	67.43 °C	This study
3-pET28-GFP-TEV_Rev	gatccctggaagtacaggttttcagatcctttgtagagctcatccatgcc	67.60 °C	This study

Table 2.3. Plasmids used in this study.

Fragment to amplify	Primers used	Plasmid base	Final result
PlyMW2	5-Nde-PhiMW2 3-Xho-PhiMW2	pET21c	pET21c-PlyMW2
CHAP domain	5-Nde-PhiMW2 3-Xho-CHAPMW2	pET21c	pET21c-CHAP
CBDMW2 domain	5-BmH1-Sh3bMW2 3-Xho-PhiMW2	pET28-8HT-GB1-TEV	pET28-8HT-GB1-TEV-CBDMW2
GFP	5-pET28-HIS-GFP_Fw 3- pET28-GFP-TEV_Rev	pET28-8HT-GB1-TEV-CBDMW2	pET28-8HT-GFP-TEV-CBDMW2

2.4 Expression and purification of CBDMW2 domain

After confirming sequencing results, the plasmids pET28-8HT-GB1-TEV-CBDMW2 and pET28-8HT-GFP-TEV-CBDMW2 were used to transform competent *E. coli* BL21(DE3) cells, via the heat shock method (chemical transformation). The first step for the transformation was to mix competent cells with 50 ng of plasmid, and then incubate the mix for 15 minutes in ice, followed by a thermic shock of 60 seconds at 42 °C, and returned to the ice for 5 minutes. Afterwards, the cells were cultured in antibiotic-free LB for 45 minutes at 37 °C and 180 rpm, to allow their recovery from the thermic shock. Later, cells were inoculated in an LB agar plate with kanamycin (Kan) 50 µg/mL and left at 37 °C O/N.

In order to prepare for protein expression, a pre-inoculum was prepared with some selected colonies in LB medium with Kan 50 µg/mL and left O/N at 37 °C and 180 rpm. For plasmid pET28-8HT-GB1-TEV-CBDMW2, the pre-inoculum was inoculated in an M9 minimal medium (**Table 5.3**, annexes) isotopically enriched with ^{15}N, using $^{15}NH_4Cl$ (Cambridge Isotope Laboratories CIL – 99%), 50 µg/mL Kan, and different concentrations of glucose, one culture with 2 g/L and another with 4 g/L (for expression tests), at 37 °C and 180 rpm. Therefore, with the results of expression tests, the CBDMW2 domain was expressed again in an M9 medium with 2 g/L of labeled glucose ^{13}C (U-$^{13}C_6$-glucose >99%). For plasmid pET28-8HT-GFP-TEV-CBDMW2, the pre-inoculum was inoculated in an LB medium with Kan 50 µg/mL. Once cultures reached an OD_{600nm} of approximately 0.6 AU, protein expression was induced with 1 mM of IPTG, and cultures were further incubated O/N at 25°C.

The expression tests, for plasmid pET28-8HT-GB1-TEV-CBDMW2, were performed to evaluate the yield of protein obtained in the two conditions. This way, samples were collected at different expression times: before the induction with IPTG, after 2 hours of induction, and after 16 hours of induction. The collected samples were later analyzed in an SDS-PAGE gel (**Table 5.4**, **Table 5.5** and **Table 5.6**, annexes).

Extraction and purification steps have remained almost the same for both proteins.

Cells were centrifuged for 10 minutes at 8000 rpm and 8 °C (Beckman Coulter Avanti® J-HC High Capacity Centrifuge + JA-14 Fixed-Angle Rotor), and the pellet was resuspended in a phosphate buffer (PBS) (10 mM PO_4 pH 6.0, 200 mM NaCl, 1 mM BME) (**Table 5.7**) and stored at 4 °C O/N.

Next, the cell lysis was induced by sonication, using 8 cycles of 1 minute, 80% amplitude, and 0.5 output (Hielscher UP50H ultrasonicator). The cell extract was centrifuged for 40 minutes, 10000 rpm, and 8 °C (Eppendorf 5804 R centrifuge + F-34-6-38 Rotor), and the supernatant was recovered to purify the protein from the extract, via an immobilized metal affinity chromatography (IMAC). The supernatant was injected on a HisTrapTM HP column (GE HealthCare), that has resin immobilized with Ni^{2+}, in which proteins containing a histidine tag, are adsorbed to the column.

The purification was done by applying an imidazole concentration gradient, starting with a washing step using PBS (10 mM PO_4 pH 6.0, 200 mM NaCl, and 1 mM BME). The first part of the gradient was between 0-60 mM imidazole for 12 mL, the second part between 60-600 mM imidazole for 20 mL, and the last part 1000 mM imidazole for 0 mL, using a PBS with 10 mM PO_4 pH 6.0 with 200 mM NaCl, 1 mM BME and 1 M imidazole. The elution of the protein was monitored by UV-spectroscopy at 280 nm. The collected fractions were analyzed by SDS-PAGE.

The fraction containing the protein was dialyzed for PBS (10 mM PO_4 pH 6.0, 100 mM NaCl, 1 mM BME) (**Table 5.7**, annexes). For dialysis of the protein 8HT-GB1-CBDMW2, TEV protease and 50 µM EDTA was added to enable the isolation of the CBDMW2 from the GB1 fusion protein with the 8 histidine sequence (8HT-GB1). Thus, 1 mg of TEV protease was used. Dialysis was achieved by SnakeSkin™ Dialysis Tubing, 3.5K MWCO, 22 mm (Thermo Scientific), and was performed O/N with slow agitation at RT.

A second IMAC was performed for the protein 8HT-GB1-CBDMW2, to allow the separation of the 8HT-GB1 from the CDB, since the 8HT-GB1 absorbs into the Ni^{2+} column, and the CBDMW2 is eluted in the flow-through. The elution was monitored by UV-spectroscopy at 280 nm, and collected fractions were analyzed by SDS-PAGE.

The purified CBDMW2 and GFP-CBDMW2 were concentrated (EMD Millipore UFSC05001 5122 Amicon Stirred Cell Model 8050; Filter with cutoff 3 kDa), and their final concentrations were calculated through UV-spectroscopy at 280 nm (Thermo Scientific NanoDropTM 100 spectrophotometer) with the molar extinction coefficient of 37390 $M^{-1}.cm^{-1}$ for CBDMW2, and 58330

$M^{-1}.cm^{-1}$ for GFP-CBDMW2, both estimated by the software ProtParam ExPASy. The final samples were stored at 4 °C until used to prepare samples for NMR, fluorescence microscopy, and MST.

2.5 Nuclear Magnetic Resonance

NMR experiments were acquired on the Bruker Avance III NMR spectrometers (Bruker BioSpin), located in the UCIBIO-FCT NOVA NMR facility, operating at 600 MHz for ^{1}H and the other at 500 MHz for ^{1}H, equipped with a 5 mm triple resonance cryogenic probe (CP TCI) with -z gradients and the N_2-cooled cryogenic probe (Prodigy) optimized for 5 mm ^{1}H with -z gradients, respectively.

TOPSPIN 3.4 software was used for the data acquisition, TOPSPIN 4.1.1 software for processing and CARA (Computer Aided Resonance Assignment) [28] to analyze and create figures of all spectra. The samples were prepared in 3 mm (NORELL, NC, USA) and 5 mm NMR tubes (Shigemi). For chemical shift reference, 3-(trimethylsilyl) propionic-2,2,3,3-d4 acid (TSP, Sigma-Aldrich, MO, USA) was used (δ TSP = 0 ppm).

2.5.1 Preparation of NMR samples

For acquisition of triple resonance experiments and 3D NOESY experiments, a sample of CBDMW2 labeled with ^{15}N and ^{13}C was prepared in 10 mM PO_4 buffer pH 6.0 with 1 mM BME and 10% D_2O, with a final concentration of 600 µM, and transferred to a 5 mm NMR tube.

2.5.2 Assignment of the main chain resonances of CBDMW2 domain

2.5.2.1 Triple Resonance Method

In order to do the assignment of the resonances of CBDMW2, the triple resonance method was used. This method needs an isotopic enrichment of proteins with ^{13}C and ^{15}N and uses the scalar couplings to correlate the different nuclei along the protein backbone, allowing to identify neighboring residues. This method became a standard approach to make resonance assignments, since the triple resonance spectra are simple, containing only a few signals per residue, and have high sensitivity since the transfer of magnetization is made via the covalent chemical bonds.

Therefore, the following experiments were acquired for the assignment of the main chain resonances: HNCO, HN(CA)CO, HNCA, HN(CO)CA, HN(CO)CACB, and HNCACB, at 293 K. For the assignment of the side chains and structure calculation, these experiences have been acquired: CC(CO)NH, HBHA(CO)NH, H(CCO)NH, hCCH-TOCSY, HcCH-TOCSY, ^{15}N TOCSY, ^{1}H-^{13}C-

HSQC, ^{15}N-NOESY, ^{13}C-NOESY aliphatic, and ^{13}C-NOESY aromatic, at 293 K. The assignment of the side chains and structure calculation was made by another member of the lab group. The residue atoms were assigned using the nomenclature by CYANA.

2.5.2.2 Assignment of the main chain resonances

To begin with the assignment, all resonances in the ^{1}H-^{15}N-HSQC spectrum were picked, all these resonances correspond to NH correlations either from amide groups in the backbone, or the NH groups of the side chains (NH – tryptophan, and arginine; NH_2 – asparagine and glutamine). The next step was the analysis of triple resonance spectra.

The first experiment analyzed was the HNCO, where the magnetization is transferred from the H_N of the residue itself (i), via the N (i), to the CO of the precedent residue (i-1). In the HN(CA)CO experiment, the magnetization is transferred from the H_N (i), via the N (i) and Cα (i and i-1), to the CO of the residue itself (i), and the previous residue (i-1). The Cα only transfers the magnetization, it is not detected (**Figure 2.2**).

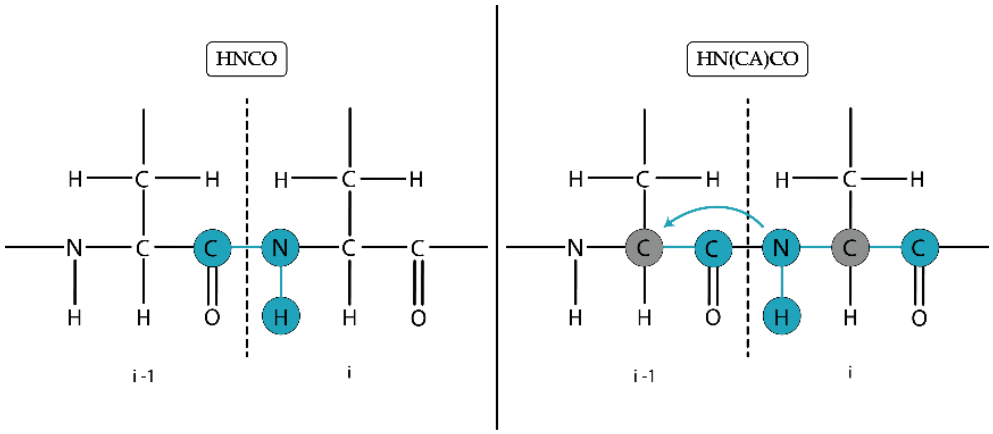


Figure 2.2. Schematic representation of the correlations observed in the 3D triple resonance experiments HNCO and HN(CA)CO. The nuclei whose frequencies are detected in one of the three dimensions are marked in blue. Nuclei only involved in the transfers of the magnetization which are not detected are marked in grey.

Next, the HN(CO)CA experiment correlates the H_N (i), via N (i) and CO (i-1), with the Cα of the previous residue (i-1). CO only transfers the magnetization and is not detected. The HNCA correlates the H_N (i), via N (i), with the Cα of residue itself (i) and the preceding residue (i-1) (**Figure 2.3**).

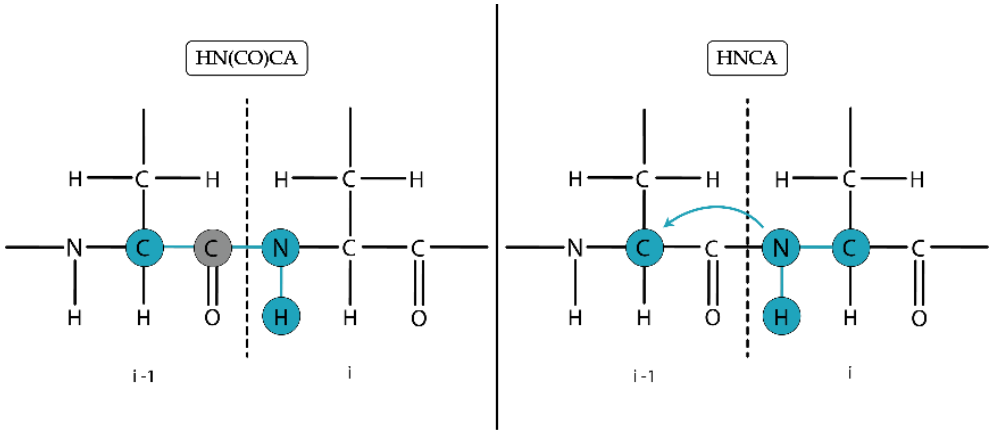


Figure 2.3. Schematic representation of the correlations observed in the 3D triple resonance experiments HN(CO)CA and HNCA. The nuclei whose frequencies are detected in one of the three dimensions are marked in blue. Nuclei only involved in the transfers of the magnetization which are not detected are marked in grey.

Then, the HN(CO)CACB experiment, where the magnetization is transferred from the H_N (i), via N (i), to the Cβ and Cα of residue i-1. The HNCACB experiment correlates the H_N (i), via N (i), to the Cβ and Cα of residue i and i-1 (**Figure 2.4**).

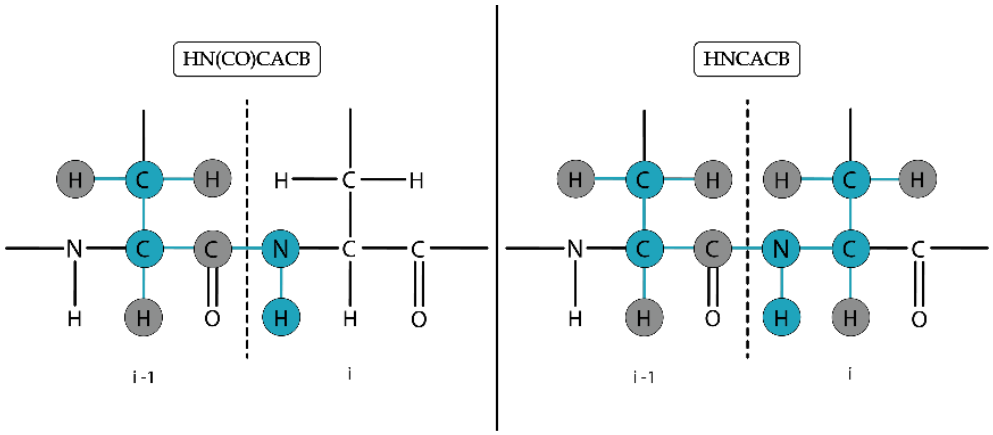


Figure 2.4. Schematic representation of the correlations observed in the 3D triple resonance experiments HN(CO)CACB and HNCACB. The nuclei whose frequencies are detected in one of the three dimensions are marked in blue. Nuclei only involved in the transfers of the magnetization which are not detected are marked in grey.

The analysis of all these spectra, as a whole, made it possible to make the sequential assignment of the residues of the CBDMW2 domain.

2.6 Fluorescence microscopy

Fluorescence microscopy was used to analyze if the CBDMW2 domain would bind to the peptidoglycan of *S. aureus*. Two different strains were used, *S. aureus* COL and MW2, both MRSA

strains, and the fluorescent label used was GFP, which was merged with the CBDMW2 domain (GFP-CBDMW2).

The *S. aureus* strains, MW2 and COL, were grown in TSB at 37 °C and 180 rpm, to an OD_{600} of 0.6 AU [29]. Then, cells were harvested, washed with 10 mM PO_4 pH 6.0, and incubated with 27 µM of GFP-CBDMW2 and 16 µM of GFP, for control, for 20 minutes at 4 °C. Afterwards, samples were washed two times with the same buffer (10 mM PO_4 pH 6.0). To attach the cells to the glass slides, PB (10 mM PO_4 pH 6.0) with 1% agarose was used. The images were acquired using the fluorescence microscope Axio Imager D2 (Zeiss) and processed with Fiji software [30]. To excite the GFP, the wavelength of 488 nm was used.

2.7 Extraction and purification of peptidoglycan from *Staphylococcus aureus*

The strain of *Staphylococcus* used to extract peptidoglycan was COL. The protocol used was from Bárbara Gonçalves et al., 2019 [10]. Using the same protocol, two different media were used, a TSB medium and a minimal medium to label the peptidoglycan with ^{15}N, using $^{15}NH_4Cl$. To prepare the minimal medium, the following solutions must be prepared: metal solution and the mother solution (**Table 5.8**, annexes). The metals solution was prepared, then filtered. Next the mother solution was prepared, and the pH was adjusted to 7.7. Then the casamino was added. The final step was the addition of 1 mL of the metal solution (1000x), 3.33 mL of the mother solution (300x), and glucose to a final concentration of 7.5 mM, and the medium was filled with water to a final volume of 1 L.

Initially, cells were propagated in 2 L of TSB and 2 L of the minimal medium at 37 °C and 220 rpm O/N, with an initial OD_{620} of 0.02 AU. The growth was stopped when cultures reached an OD_{620} of 0.7 AU for the TSB medium and an OD_{620} of 0.9 AU for the minimal medium and immediately chilled in an ice and ethanol bath until the temperature dropped below 10 °C. Then the culture was centrifuged at 10000 rpm at 4 °C for 10 minutes. The pellet was resuspended in ddH_2O, and boiled in 4% SDS for 30 minutes to isolate cell walls. The SDS was removed by washing cells several times with ddH_2O, and centrifuges of 15 minutes and 12000 rpm at room temperature (RT). The Hayashi method was used to check if SDS was completely removed. The SDS-free pellets were resuspended in ddH_2O, and the cell walls were broken using glass beads of 160 µm (Sigma-Aldrich) and a Digital Disruptor Genie device (Scientific Industries, New York, EUA), making 20 cycles of 1 minute in the device followed by 1 minute in ice. Glass beads were removed by filtration, and then the filtered solution was centrifuged for 15 minutes at 15000 rpm at RT. Pellets were resuspended in 100 mM Tris pH 7.5.

To purify the cell walls it was added to the cells, 20 mM $MgSO_4$, 10 µg/mL of DNAse (Sigma-Aldrich), and 50 µg/mL of RNAse (Sigma-Aldrich) and incubated at 37 °C for 2 hours. Later, 10 mM of $CaCl_2$ (Merck) and 100 µg/mL of trypsin (Worthington Industries, Columbus, EUA) were added and incubated O/N at 37 °C with agitation. The next step involved adding 1% of SDS and boiling the sample for 15 min. Then, to remove SDS, the sample was centrifuged for 15 minutes at 15000 rpm at RT, and ddH_2O was used to wash the pellet. After centrifugation, the pellet was resuspended in 8M $LiCl_2$ and incubated for 30 minutes at 37 °C and the sample was centrifuged for 15 minutes at 15000 rpm at RT. Afterward, the pellet was resuspended in 0.1 M EDTA pH 7.0 incubated for 30 minutes at 37 °C, and centrifuged for 15 minutes at 15000 rpm at RT. Next, the pellet was resuspended in 20 mL of ddH_2O and centrifuged for 15 minutes at 15000 rpm at RT, for 4 times, and the resulting pellet was resuspended in a small volume of ddH_2O and lyophilized.

To purify the peptidoglycan, 5 mg of lyophilized cells were incubated with 2 mL of 46% hydrofluoric acid (Labkem, Dublin, República da Irlanda) for 48 hours at 4 °C with agitation. ddH_2O was added to the sample after incubation to make a final volume of 30 mL and centrifuged for 50 minutes at 20.000 rpm at 4 °C. The pellet was resuspended in ddH_2O, centrifugation was repeated, and the resultant pellet was resuspended in 10 mM Tris pH 7.0 and centrifuged again in the same conditions. Then, the pellet was resuspended in ddH_2O and centrifuged in the same conditions, and this step was repeated once more. Finally, ddH_2O was added to the sample and then lyophilized.

The lyophilized pellet was resuspended in 0.05% NaN_3 to a concentration of 10 mg/mL and added 25 mM sodium phosphate buffer pH 5.5. To the sample, was added 150 µL of mutanolysin (N-acetylmuramidase) at 5 mg/mL (Sigma-Aldritch, M9901), which was then incubated for a week at 37 °C. Afterward, the sample was centrifuged for 5 minutes at 14000 rpm at RT. Then, 0.5 M borate buffer pH 9.0 and sodium borohydride powder was added to the muropeptides and incubated for 30 minutes at RT. The reduction of MurNAc originate a non sugar derivative, for simplicity we will maintain the MurNAc terminology. To prepare the sample for RP-HPLC, pH was lowered below 2 with trifluoroacetic acid (TFA), and centrifuged for 5 minutes at 13500 rpm at RT. Lastly, the supernatant was removed and used in the RP-HPLC.

For the separation and analysis in the RP-HPLC, a C8 reversed-phase column (LiChroCART 250-4, LiChrospher 100 - RP-8, 10 um) was used. To perform the elution of the muropeptides, a buffer with Milli Q water, 0.05% TFA, and 20% acetonitrile was used, in a linear gradient between 0 and 10%. Molecules with a polar nature were eluted faster than molecules with a nonpolar nature. The elution was monitored by UV-spectroscopy at 206 nm. The fractions were separated by the size of the fragments (I – smaller fragments; VI – longer fragments) and were lyophilized.

The collected fractions with longer fragments (V and VI) were resuspended in 25 mM Tris-HCl pH 7.5 and 50 mM NaCl and digested with lysostaphin. Thus 50 µL of lysostaphin at 3.63 mg/mL, was

added to the sample and incubated O/N at 37 °C. The resultant fragments were separated and analyzed by RP-HPLC. To prepare the samples for the RP-HPLC, the pH was lowered below 2 with TFA then centrifuged for 5 minutes at 13500 rpm at RT and the supernatant was removed and injected into the RP-HPLC. The previously stated definitions were used, a C8 reversed-phase column and acetonitrile linear gradient between 0 and 10% to elute the muropeptides. Also the elution was monitored by UV-spectroscopy at 206 nm. The collected fractions were lyophilized.

2.7.1 Preparation of NMR samples and quantification

To evaluate and characterize the muropeptides, some samples with different fragments were prepared for NMR. Fractions I, II, and double digested were resuspended in Milli-Q water with 10% D_2O, and the sample was transferred to 3 mm NMR tubes (Shigemi).

1D-^{1}H spectra were acquired to calculate the concentration of the fragments, using 50 µM TSP as a reference, and ^{1}H-^{15}N-HSQC, ^{1}H-^{13}C-HSQC, TOCSY, and NOESY experiments were acquired to evaluate purity and characterize the muropeptides. Two TOCSY experiments were acquired with a mixing time of 60 and 30 ms, respectively. The NOESY experiment was acquired with a mixing time of 250 ms.

The concentration of the double digested samples were calculated through the 1D-^{1}H spectra, using TSP, that has in its constitution 9 protons and appears in the 1D-^{1}H spectrum as a singlet. Integrating the TSP singlet and calibrating it as a function of concentration, the concentration value that corresponds to the integral is obtained. The concentration of TSP in the samples was 0.05 mM, and TSP has nine protons, thus the concentration value that corresponds to the integral of TSP is 0.45 mM. Then the signal of the sample is integrated. For that, the signals used corresponded to the methyl groups of L and D-Alanine, the D-Lac group of MurNAc, and the proton from the HG2 from L-Lysine, making a total of ten protons (**Figure 2.5**). The resulting integral value is divided by ten, giving a final concentration of 2.07 mM for the double digested muropeptide sample.

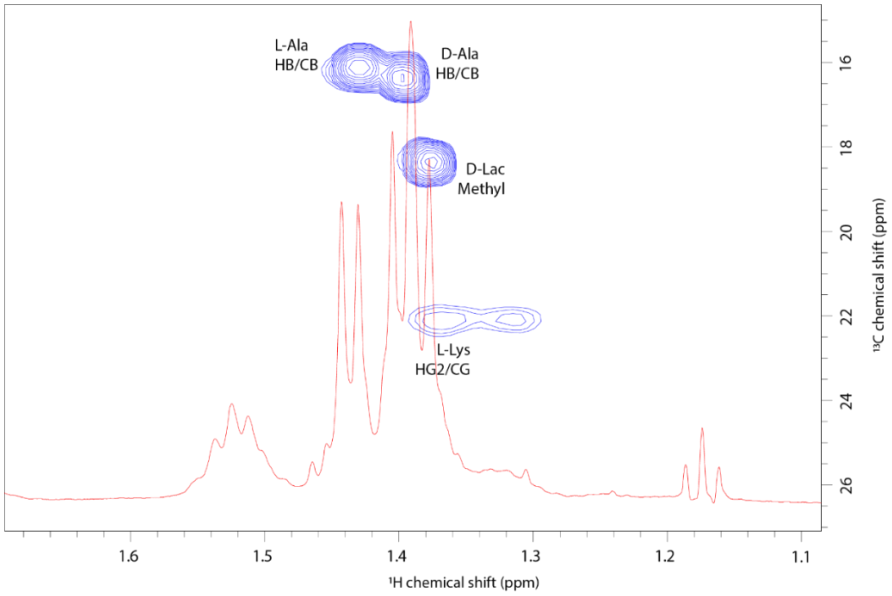


Figure 2.5. Peaks used to quantify the muropeptide. In red the 1D-^{1}H spectrum and blue the ^{1}H-^{13}C-HSQC spectrum.

2.7.2 Electrospray Ionization Mass Spectrometry

The samples analyzed by Electrospray Ionization Mass Spectrometry (ESI-MS) were fraction II and the fraction that was double digested. Both were analyzed in the same buffer used in the NMR analysis, Milli-Q water, and 10% D_2O, with a concentration of 240 µM for fraction II and 2.07 mM for the fraction double digested. The analyses were performed by "Laboratório de Análises" LAQV Requimte (LAQV Requimte, Almada, Portugal), using an Agilent 1200 G1365B MWD, an Agilent 1200 Series G1316A TCC (Thermostatted Column Compartment), an Agilent 1100 HPLC G1313A Autosampler, and an Agilent 1260 Infinity II Binary Pump G7112B.

2.8 MicroScale Thermophoresis

MicroScale Thermophoresis (MST) is a technique used to determine binding affinities in buffers and biological liquids, observing the motion of fluorescent molecules along a microscopic temperature gradient in a capillary. MST is usually used to analyze the binding between a protein and small molecules, in this case, the interaction between GFP-CBDMW2 and PG fragments of *S. aureus*. The temperature gradient is induced by an IR laser, that heats the capillary at the center of its length. Heat leads to the movement of molecules to the edges of the capillary and once the heat application stops, molecules leave the capillary borders (**Figure 2.6**).

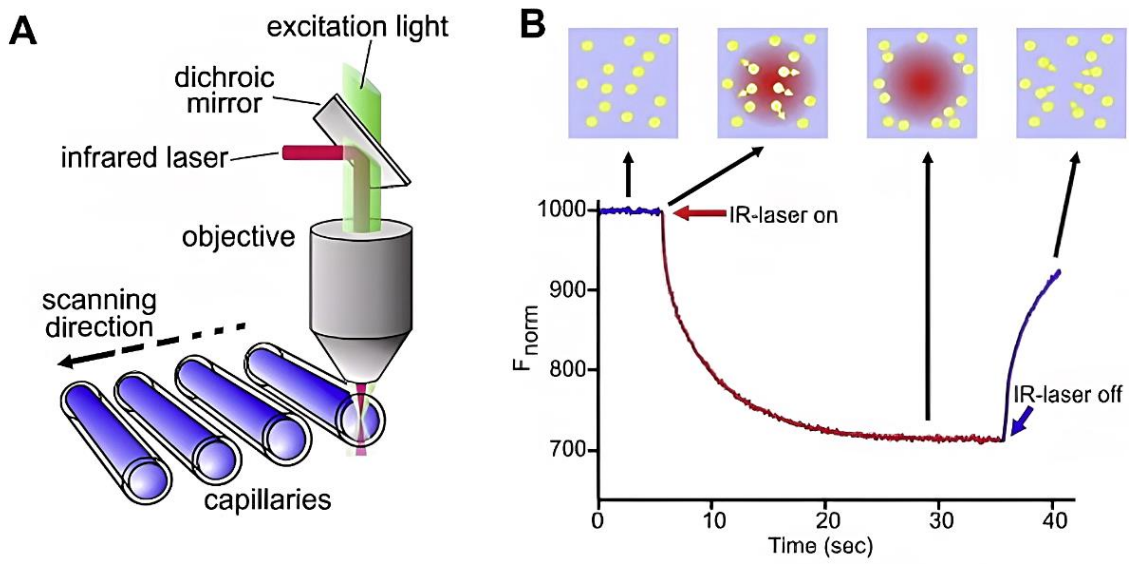


Figure 2.6. MicroScale Thermophoresis setup and experiments. (A) Schematic of technical setup of the MST. Through the objective, the center of the capillary is heated by an IR laser, and the thermophoretic movement of the molecules through the temperature gradient is detected. (B) Typical signal of a MST experiment. An initial fluorescence is detected in the homogeneous sample. Then, the temperature gradient is induced, and a T-Jump (temperature jump) occurs, where the fluorescence decreases due to the fast temperature change, and the molecules move to the edges of the capillary. Once the heat application stops, the molecules return to the initial state. Adapted from [26].

The MST assay was performed using the Nano Temper technologies Monolith NT.115 (BLUE/RED) (NanoTemper Technologies, GmbH, Munich, Germany) at 25 °C.

For the assay, GFP-CBDMW2 was prepared in 10 mM PO_4, 200 mM NaCl, 1 mM BME, and 0.1% Tween20, with a final concentration of 100 nM. The ligand used was PG fragmented with mutanolysin and amidase. For the samples, serial dilutions were prepared from the stock ligand, obtaining 16 samples with concentrations between 37.5 µM and 1.14×10^{-3} µM of ligand, and triplicates were prepared. 16 Monolith NT.115 capillaries were filled with samples, and to obtain measurements an excitation power of 20% NANO-BLUE and a medium MST-Power were used. All data were analyzed using the MO. Affinity Analysis 2.3 software.

2.9 Titration of CBDMW2

The CBDMW2 was titrated with two different PG fragments: a peptide with 5 glycines and muropeptides (**Figure 2.7**). For the assay with the peptide of glycines, final concentration of CBDMW2 was 50 µM in 50 mM PO_4 pH 6.0, 10% D_2O, with different ratios of peptide (1:0, 1:17, 1:31.5, 1:63.5, 1:95 and 1:190). In the case of muropeptides, final CBDMW2 concentration was 25 µM in 50 mM PO_4 pH 6.0, 10% D_2O, and muropeptide concentration ratios were: 1:0, 1:0.8, 1:1.6, 1:3, 1:6.6, 1:11, 1:16.6 and 1:33.

The monitorization of the resonances was accomplished through the acquired ^{1}H-^{15}N-HSQC spectra to follow the interaction between the CBDMW2 and the fragments by observing the chemical shifts of peaks in the spectrum. All spectra were acquired at 293 K, and data was analyzed using Bruker TopSpin 4.1.1 software.

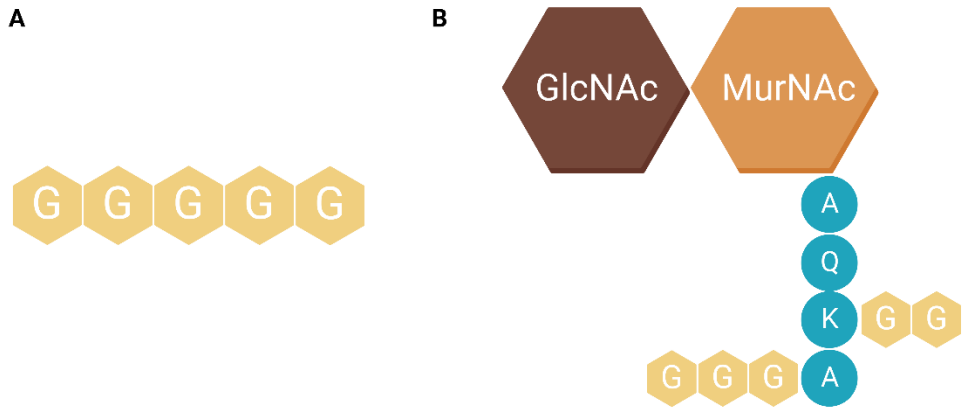


Figure 2.7. (A) Peptide with 5 glycines. **(B)** Muropeptide. Since lysostaphin does not have a defined cut site, the resulting peptide may have a variable number of glycines in the cross-linking bridge, which are marked in gray.

3

RESULTS & DISCUSSION

3.1 Construction of recombinant plasmids

Recombinant plasmids were built for the full-length protein, for the CHAP domain, for the CBDMW2 domain, and for the CBDMW2 domain with the GFP. **Figure 3.1** shows the amplification products for the full-length protein and the two domains in the agarose gel and the negative controls for the different amplifications.

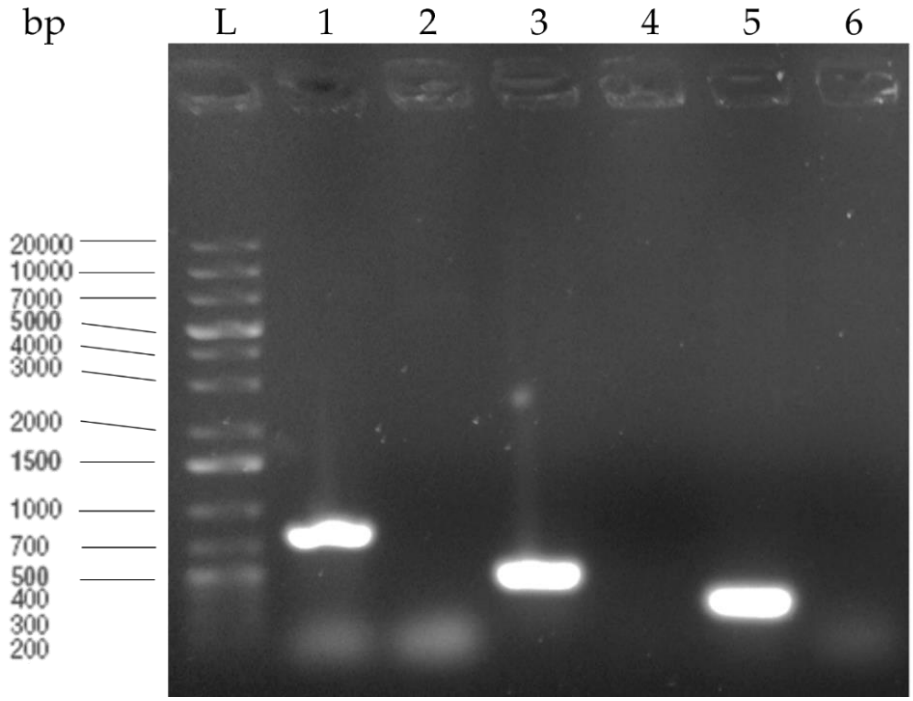


Figure 3.1. Amplification products analyzed by a 0.8% agarose gel (90V). **L-** GeneRuler 1 kb Plus DNA Ladder, ready-to-use (Thermo Scientific); **1-** Amplification product of the full-length protein; **2-** Negative control for full-length protein; **3-** Amplification product of the N-terminus domain (CHAP); **4-** Negative control for the N-terminus domain; **5-** Amplification product of the C-terminus domain (CBDMW2); **6-** Negative control for the C-terminus domain.

Agarose electrophoresis confirmed that the amplification had been successfully executed (**Figure 3.1**) since the gene for the full-length protein has a total of 753 base pairs, the N-terminus domain has 450 base pairs and the C-terminus domain has 303 base pairs. Plus, the negative controls did not show any bands.

The next step was the digestion of the fragments with their respective restriction enzymes, and the products were analyzed by an 0.8% agarose gel (**Figure 3.2**).

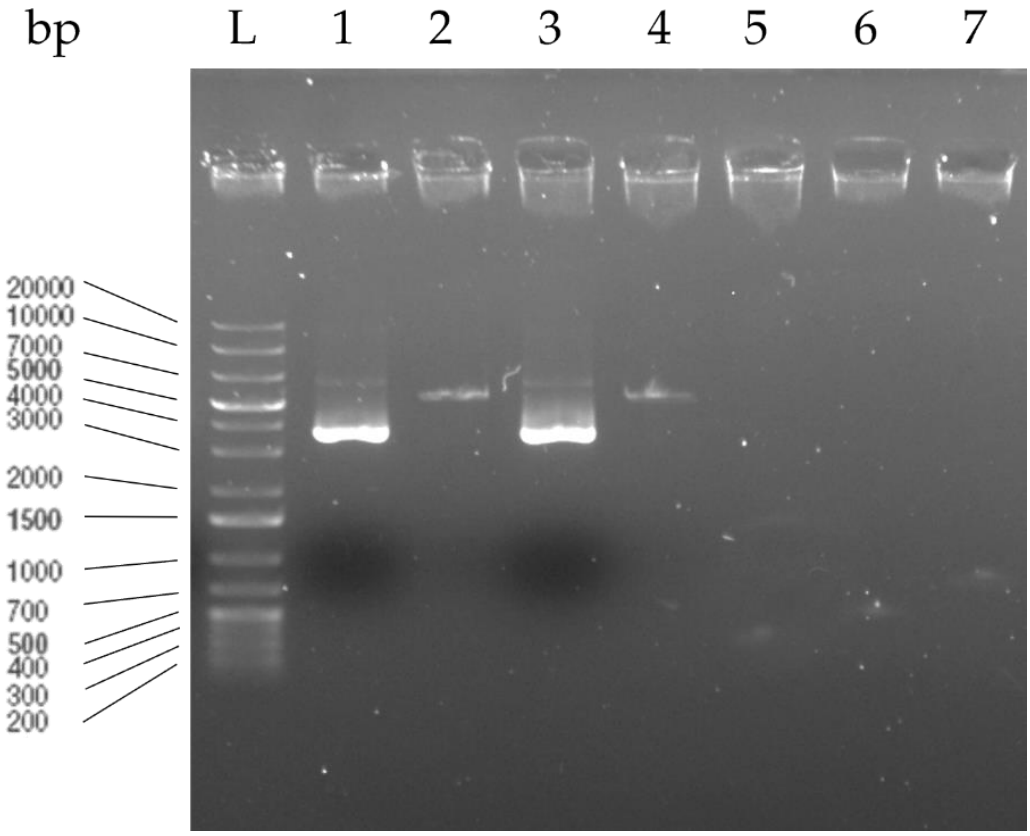


Figure 3.2. Digestion products analyzed by a 0.8% agarose gel (90V). **L-** GeneRuler 1 kb Plus DNA Ladder, ready-to-use (Thermo Scientific); **1-** pET28-8HT-GB1-TEV undigested; **2-** pET28-8HT-GB1-TEV digest with NdeI and XhoI; **3-** pET21c undigested; **4-** pET21c digest with NdeI and XhoI; **5-** Digested product of the C-terminus domain (CBDMW2) with BamHI and XhoI; **6-** Digested product of the N-terminus domain (CHAP) with NdeI and XhoI; **7-** Digested product of the full-length protein with NdeI and XhoI.

Posteriorly, the dephosphorylation of the vectors was completed, and then the ligation reaction was performed O/N to promote the connection of the digested fragment with the respective plasmid, giving rise to the recombinant plasmid.

To produce the plasmid for the CBDMW2 domain with the GFP, RF-cloning was used. In this technique, two consecutive PCR are performed. In the first PCR is produced a megaprimer, which is used for the second PCR.

Figure 3.3 shows the products of the first PCR, the megaprimer, which in this case is the *GFP* gene. This gene has 730 bp [32].

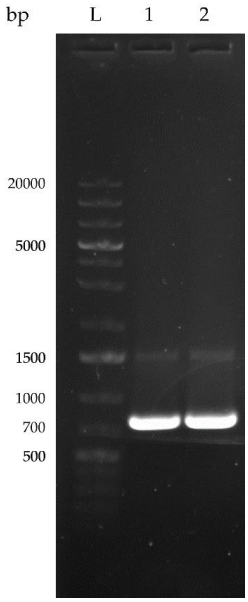


Figure 3.3. Amplification products of the first PCR analyzed by a 0.8% agarose gel (90V). **L**- GeneRuler 1 kb Plus DNA Ladder, ready-to-use (Thermo Scientific); **1**- Amplification product for GFP; **2**- Duplicate of the amplification product for GFP.

In the second PCR, is used 150 ng of the megaprimer produced in the earlier PCR, to amplify the plasmid. **Figure 3.4** shows the products of the second PCR. In this step, the recombinant plasmid is assembled.

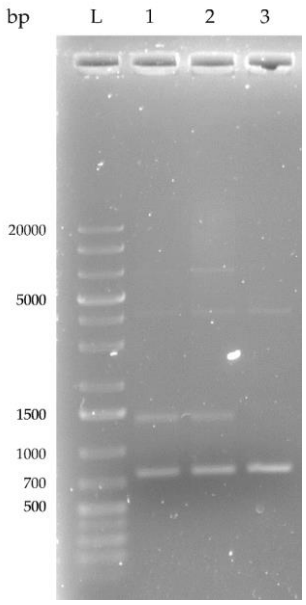


Figure 3.4. Amplification products of the second PCR analyzed by a 0.8% agarose gel (90V). **L**- GeneRuler 1 kb Plus DNA Ladder, ready-to-use (Thermo Scientific); **1**- Amplification product for pET28-8HT-GFP-TEV-CBDMW2; **2**- Duplicate of the amplification product for pET28-8HT-GFP-TEV-CBDMW2; **3**- Negative control for pET28-8HT-GFP-TEV-CBDMW2.

In order to eliminate the methylated parental plasmids, digestion was performed with DpnI. Later, DH5α cells were transformed with the products for the full-length protein, for the CHAP domain, for the CBDMW2 domain, and for the CBDMW2 domain with the GFP, and some random colonies were screened by PCR and analyzed by an 0.8% agarose gel (**Figure 3.5**). The ones that showed bands, were picked and sent for sequencing.

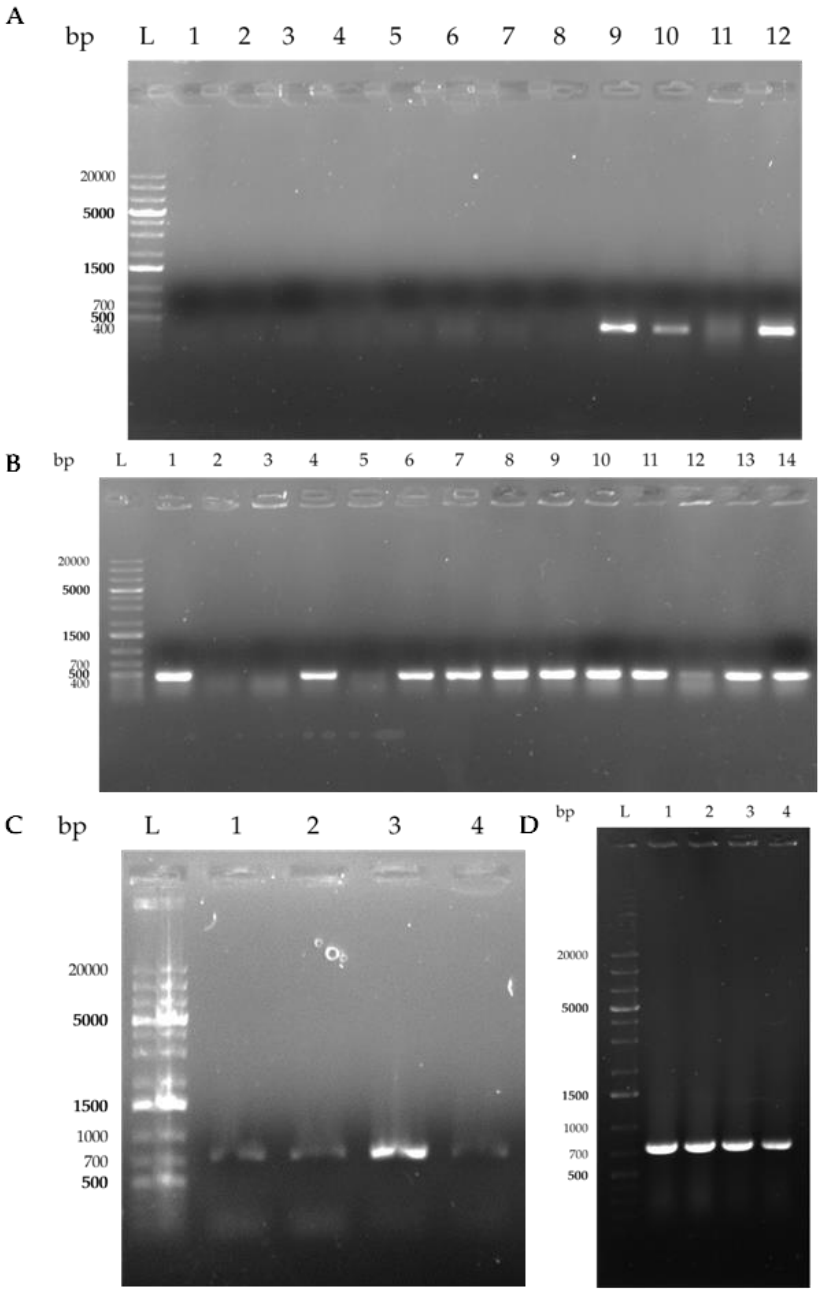


Figure 3.5. Screening products analyzed by a 0.8% agarose gel (90V). **L-** GeneRuler 1 kb Plus DNA Ladder, ready-to-use (Thermo Scientific). **(A) 1, 2, 3, 4-** PlyMW2 screening products; **5, 6, 7, 8-** CHAP domain screening products; **9, 10, 11, 12-** CBDMW2 domain screening products. **(B) 1 – 14-** CHAP domain screening products. **(C) 1 – 4-** PlyMW2 screening products. **(D) 1-4** GFP-CBDMW2 domain screening products.

All the constructs were built successfully. The plasmids pET28-8HT-GB1-TEV-CBDMW2 and pET28-8HT-GFP-TEV-CBDMW2 were used for the expression of CBDMW2 and GFP-CBDMW2 protein, respectively.

3.2 Expression and purification of CBDMW2 domain

Competent *E. coli* BL21(DE3) cells were transformed with pET28-8HT-GB1-TEV-CBDMW2 and pET28-8HT-GFP-TEV-CBDMW2 through the heat shock method, and the cells were incubated O/N in LB agar plates. Then, some colonies were used to prepare pre-inoculums.

For plasmid pET28-8HT-GB1-TEV-CBDMW2, an expression test was performed, to evaluate the yield. Two M9 minimal mediums were prepared, one of them with 2 g/L and another with 4 g/L of glucose. **Figure 3.6** shows the results of the analysis of the samples collected at different expression times.

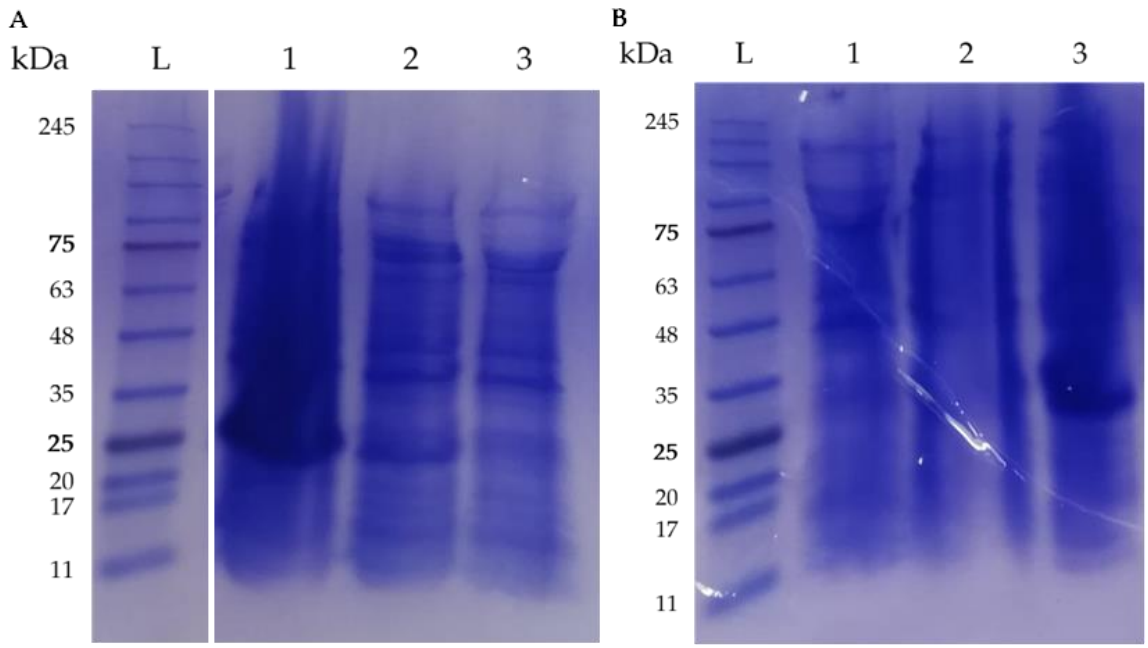


Figure 3.6. SDS-PAGE of the expression test, with a 10% Tris-Tricine gel. **L**-NZYColour Protein Marker II (NZYTech). **(A)** 2 g/L of glucose. **1-** After 16 hours of induction; **2-** After 2 hours of induction; **3-** Before induction with IPTG. **(B)** 4 g/L of glucose. **1-** Before induction with IPTG; **2-** After 2 hours of induction; **3-** After 16 hours of induction.

The yield for 2 and 4 g/L of glucose was 41.6 mg/L and 59.8 mg/L, respectively. Thus, a new expression of the CBDMW2 domain was performed using 2 g/L of labeled glucose.

The extraction and purification steps have remained almost the same for both proteins, but for the protein, CBDMW2 has been done with two affinity chromatography steps whereas the protein GFP-CBDMW2 was purified with only one affinity chromatography step.

Figure 3.7 shows the purity of the proteins after the purification through IMAC.

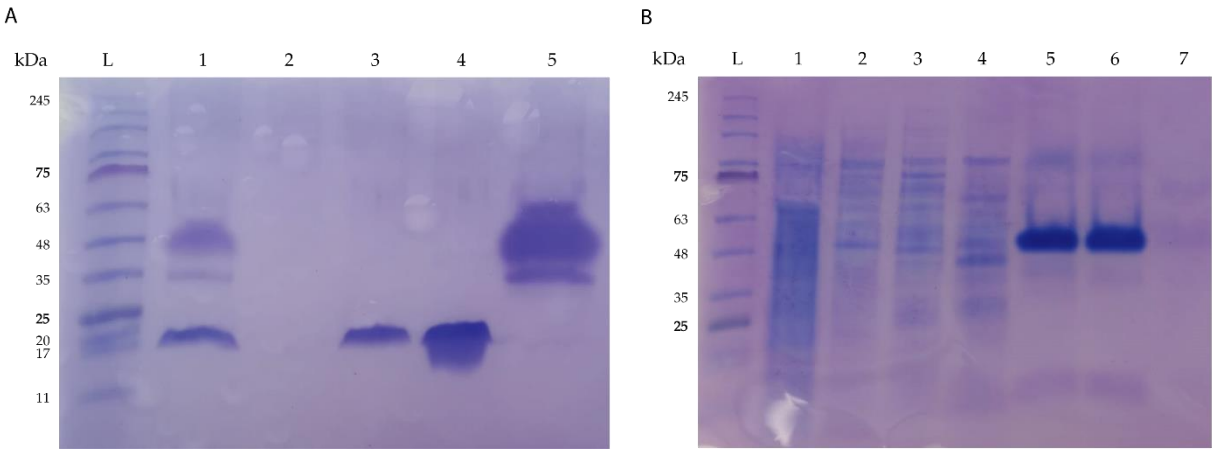


Figure 3.7. Protein purification analysis by an SDS-PAGE, with a 10% Tris-Tricine gel. **L**-NZYColour Protein Marker II (NZYTech). **(A)** CBDMW2 protein. **1-** After dialysis; **2-** Empty; **3, 4-** 0 mM imidazole; **5-** 1000 mM imidazole. **(B)** GFP-CBDMW2 protein. **1, 2-** Flow-through; **3, 4-** 60 mM imidazole; **5, 6-** 600 mM imidazole; **7-** 1000 mM imidazole.

The yield for CBDMW2 was 17.1 mg/L. After concentrating the proteins, the final concentration for CBDMW2 was 600 µM and was used in NMR, and 21.4 µM for GFP-CBDMW2 and was used in fluorescence microscopy and MST.

3.3 Assignment of the main chain resonances of CBDMW2 domain

The triple resonance method was used to do the assignment of the main chain resonances of CBDMW2 domain and the experiments used were as follow: HNCO, HN(CA)CO, HNCA, HN(CO)CA, HN(CO)CACB, and HNCACB. The CARA program was used to do the assignment [28].

To do the assignment, the first experiments analyzed were the HNCO and HN(CA)CO. A single slice from the spectrum was taken at a ^{15}N selected frequency, resulting in a two dimensional slice. Next, strips were extracted, centering in a H_N shift that make it possible to cover the entire ^{13}CA or ^{13}CO window (**Figure 3.8**) [33].

In the HNCO, only one signal is detected. This signal observed in the strip for H_N (i) corresponds to the CO of the precedent residue (i-1), that is, the information is from the previous amino acid. Meanwhile, in the HN(CA)CO two signals are observed in the strip for H_N (i): one from the CO of the precedent residue (i-1), and the other from to CO of the residue itself (i) [34].

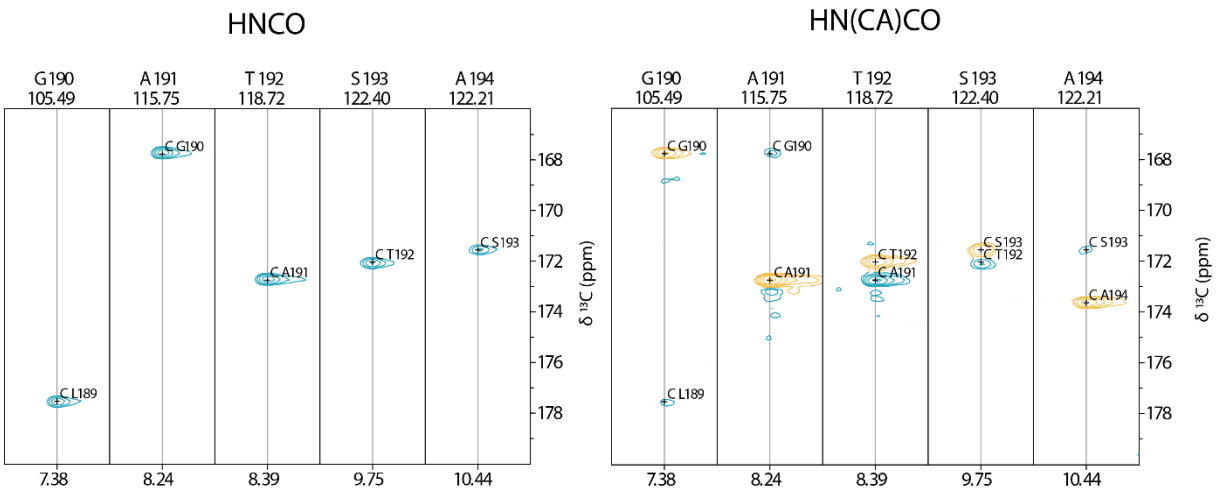


Figure 3.8. Section of the HNCO and HN(CA)CO spectrum of the CBDMW2 domain using the Stripscope of the CARA program. Strip plots are shown for five residues (G190-A191-T192-S193-A194). Each strip is extracted from a ^{15}N plane and centered around the H_N chemical shift. For each H_N (i), the HNCO strip contains one peak and the HN(CA)CO strip contains two peaks. The CO peak observed in both strips corresponds to CO of the precedent residue (i-1) (blue). The CO peak observed only in the HN(CA)CO strip corresponds to CO of the residue itself (i) (yellow).

Next, the HN(CO)CA and the HNCA were analyzed, obtaining information about the Cα (**Figure 3.9**). In the HN(CO)CA, only one signal is detected. The signal observed in the strip for H_N (i) corresponds to the Cα of the precedent residue (i-1). In the HN(CA)CO two signals are observed in the strip for H_N (i), one corresponding to the same Cα of the precedent residue (i-1) observed in the HN(CO)CA, and the other from to Cα of the residue itself (i). In case of only one Cα peak is observed in the HNCA strip, then both Cα are overlapped [34].

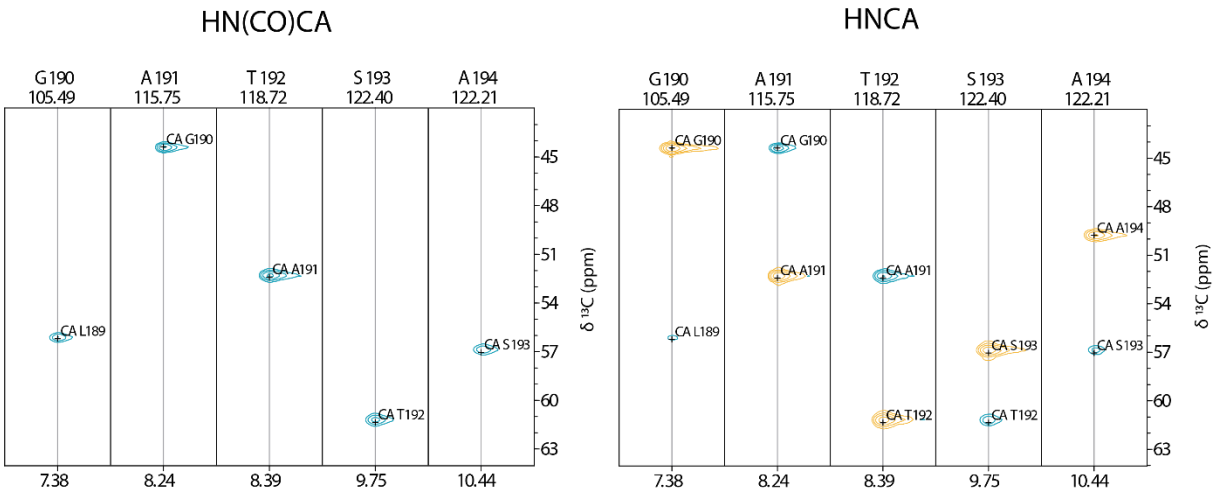


Figure 3.9. Section of the HN(CO)CA and HNCA spectrum of the CBDMW2 domain using the Stripscope of the CARA program. Strip plots are shown for five residues (G190-A191-T192-S193-A194). Each strip is extracted from a ^{15}N plane and centered around the H_N chemical shift. For each H_N (i), the HN(CO)CA strip contains a single peak and the HNCA strip contains two peaks. The Cα peak observed in both strips belongs to the previous residue (i-1) (blue). The Cα peak observed only in strip HNCA belongs to residue itself (i) (yellow).

Next, HN(CO)CACB and HNCACB experiments were analyzed (**Figure 3.10**). These experiments gives information about the chemical shifts of Cα and Cβ. Being possible to easily differentiate them, since the phase for Cα and Cβ are different in the HNCACB.

In the HN(CO)CACB, the signals observed in the strip for H_N (i) corresponds to the Cα and Cβ of the precedent residue (i-1). In the HNCACB the signals observed in the strip for H_N (i) correspond to the Cα and Cβ from the precedent residue (i-1), and the Cα and Cβ of the residue itself (i). The peaks of the Cβ, both for i-1 and i, are in a different phase from the peaks of the Cα [34].

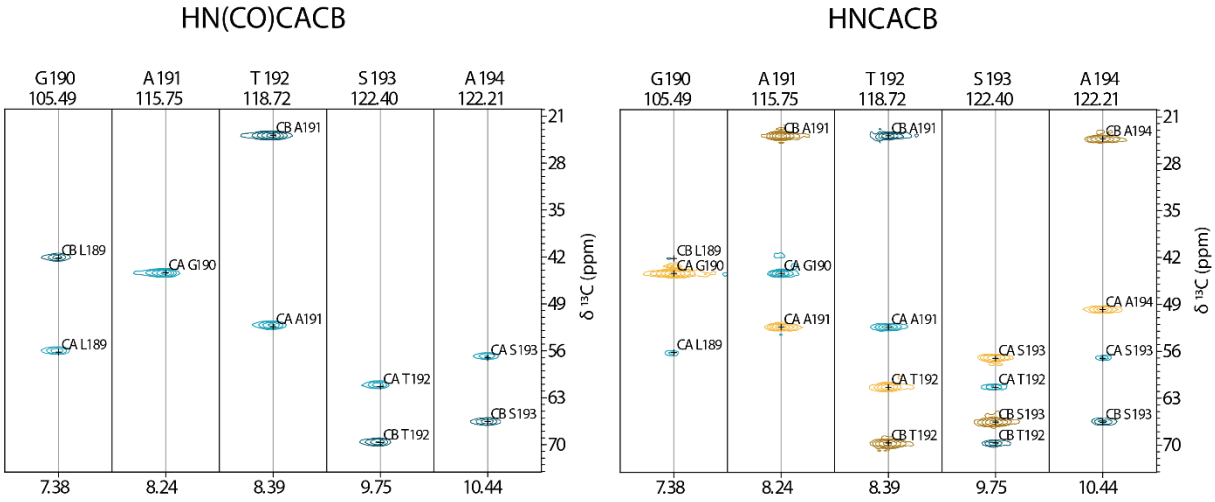


Figure 3.10. Section of the HN(CO)CACB and HNCACB spectrum of the CBDMW2 domain using the Stripscope of the CARA program. Strip plots are shown for five residues (G190-A191-T192-S193-A194). Each strip is extracted from a ^{15}N plane and centered around the H_N chemical shift. For each H_N (i), the HN(CO)CACB strip contains one or two peaks and the HNCACB strip contains two to four peaks. In HN(CO)CACB, all Cα peaks are positive, while all Cβ peaks are negative. The Cα and Cβ peaks observed in both strips belong to the preceding residue (i-1) (blue and dark blue respectively). The Cα and CB peaks observed in the HNCACB strip belong to the residue itself (i) (yellow and dark yellow respectively).

With the information obtained from the six experiments for each spin system, Cβ (i-1), Cα (i-1), CO (i-1), N (i), HN (i), Cβ (i), Cα (i) and CO (i), the sequential assignment could be made, correlating the spin systems with the same CO, Cα and Cβ chemical shifts. For instance, a spin system (1st) precedes another (2nd) when the shift for the CO (i) of the first spin system is equal to the shift of CO (i-1) observed in the second spin system, and the same for the shifts of Cα and Cβ.

After that, the sequential assignment of the spin systems could be made, comparing the shifts for the residue itself (i) with the reference chemical shifts for each amino acid, allowing to identify the amino acid that the spin system corresponded to. Some amino acids, such as alanine, serine, threonine and glycine have characteristic chemical shifts for Cα and Cβ, making them easy to identify, being the best residues to start the assignment. For example, glycine does not present a Cβ. Nonetheless, it was not possible to connect all the spin systems, resulting in different fragments, due to the existence of proline residues that interrupts the sequential assignment [34].

With all this information it was possible to do the main chain sequence assignment and a assigned ^{1}H-^{15}N-HSQC spectrum could be obtained (**Figure 3.11**).

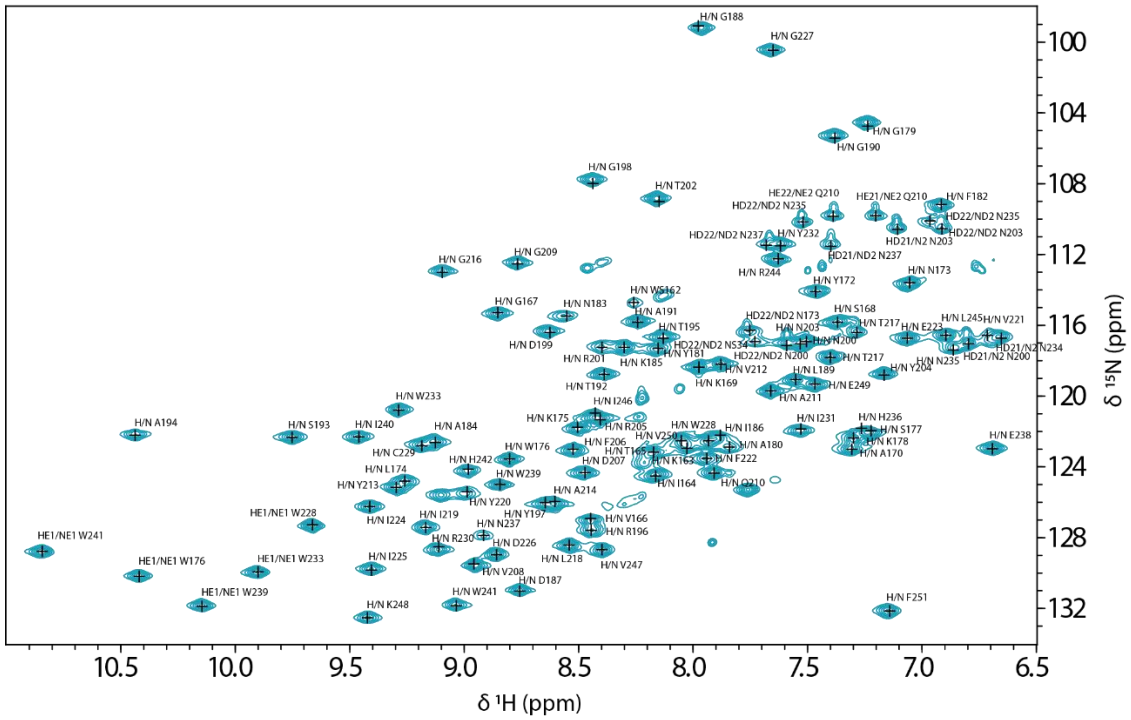


Figure 3.11. Assigned ^{1}H-^{15}N-HSQC spectrum of the CBDMW2 domain. ^{1}H-^{15}N-HSQC spectrum with 600 µM of CBDMW2 prepared with 10 mM PO_4 pH 6.0 buffer, 1 mM BME and 90% H_2O/10% D_2O.

The assignment of the side chains and structure calculation (in progress) was made by another members of the lab group, using the following experiments: CC(CO)NH, HBHA(CO)NH, H(CCO)NH, hCCH-TOCSY, HcCH-TOCSY, ^{15}N TOCSY, ^{1}H-^{13}C-HSQC, ^{15}N-NOESY, ^{13}C-NOESY aliphatic, and ^{13}C-NOESY aromatic. The preliminary 3D structure of the CBDMW2 domain, shown in **Figure 3.12.A** was calculated using the noeassign module of the CYANA program. This preliminary structure from noeassign presented a target function of 0.72, RMSD values of 0.58 Å/1.06 Å; 1392 NOEs (597 intra; 303 sequential; 80 medium range; 286 long distance; 126 dihedral angles) and 15.47 restrictions/residue. Another structure for this domain was predicted using the AlphaFold program [35, 36]. The sequence of the CBDMW2 was introduced in the program, and by comparison with others structures in the database, the program predicted the 3D structure showed in **Figure 3.12.B**. In the two models the CBDMW2 domain present a SH3b fold, as the majority of the cell wall binding domains from staphylococcal phages (LysF1, endolysin from phage K1/420, pdbid: 5O1Q) [37] or from staphylococcus targeting proteins, like lysostaphin, a bacteriocin from *Staphylococcus simulans* (pdbid: 6RK4) [38] and ALE-1, an autolysin from *Staphylococcus capitis* (pdbid: 1R77) [39]. The preliminary NMR and AlphaFold structures superimpose with a RMSD of 2.01 Å (Pymol from residue 167 to 251), indicating that both structures are similar.

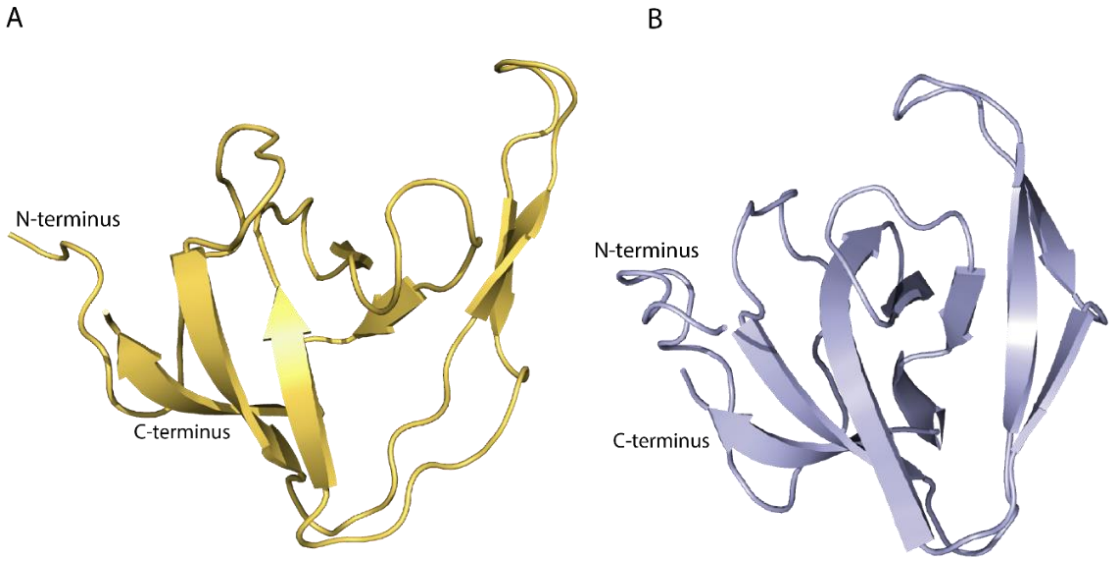


Figure 3.12. (A) CBMW2 preliminary 3D structure calculated using the CYANA program. **(B)** AlphaFold 3D structure predicted for the CBDMW2. Both structures align with a RMSD 2.01 Å (Pymol).

Although being a Sh3b fold, like the CBD of lysostaphin, some striking differences are notorious: i) Lysostaphin has an extended loop between β1 and β2 in comparison to CDBMW2. Also, the N-terminus region between the two domains is very divergent; ii) CBDMW2 has a more extended RT loop than the loop of lysostaphin. iii) The extra loop between sheets β7 and β8 is more extended in lysostaphin than the loop of CBDMW2 (**Figure 3.13** and **Figure 5.4**, annexes).

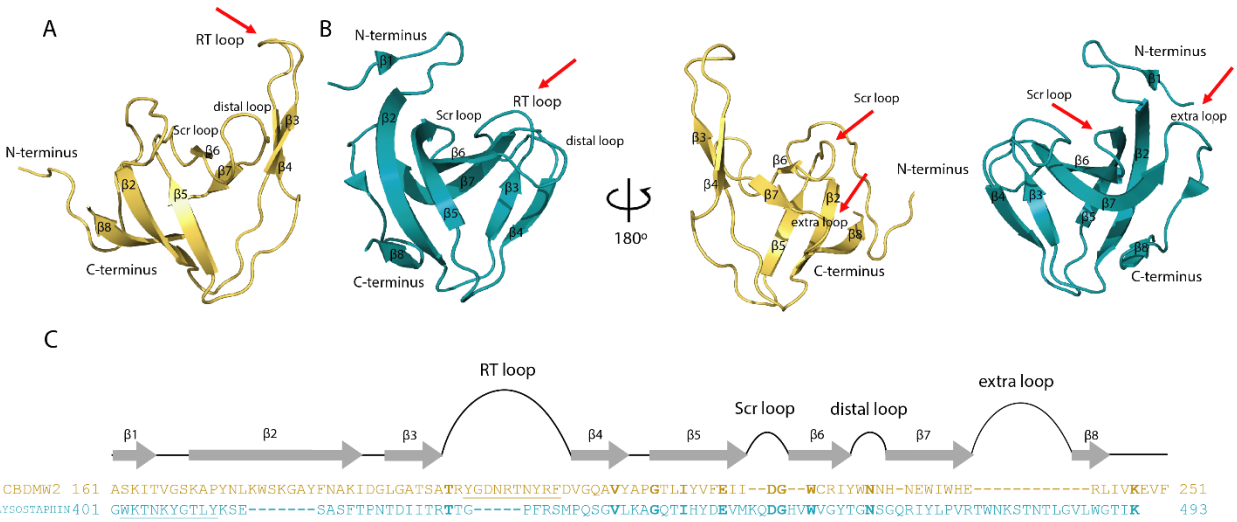


Figure 3.13. 3D structure of the CBDMW2 domain and lysostaphin (pdbid: 6RK4), and the proteins alignment. **(A)** 3D structure of CBDMW2 with numbered beta sheets. The red arrows point to the loops with the most noticeable differences. **(B)** 3D structure of lysostaphin with numbered beta sheets. **(C)** Sequence alignment of the CBDMW2 with the lysostaphin, according to their structures (conserved residues are in **bold**). The two 3D structures are showed in the same orientation [38].

Lysostaphin has a unique mode of binding to *S. aureus* peptidoglycan. It has two binding sites on opposite sides of the SH3 domain. On one side binds to the stem peptide, and on the other binds to

the pentaglycine bridge. These binding sites are not cooperative [38]. Due to the difference between the CBDMW2 domain and lysostaphin structures, mainly in the loops, the CBDMW2 binding mode to PG might be different.

3.4 Fluorescence Microscopy

Fluorescence microscopy was used to analyze if the CBDMW2 domain would bind to the peptidoglycan of *S. aureus* (**Figure 3.14**). Two different strains were used, *S. aureus* COL and MW2, both MRSA strains, and the fluorescent label used was GFP, which was merged with the CBDMW2 domain (GFP-CBDMW2).

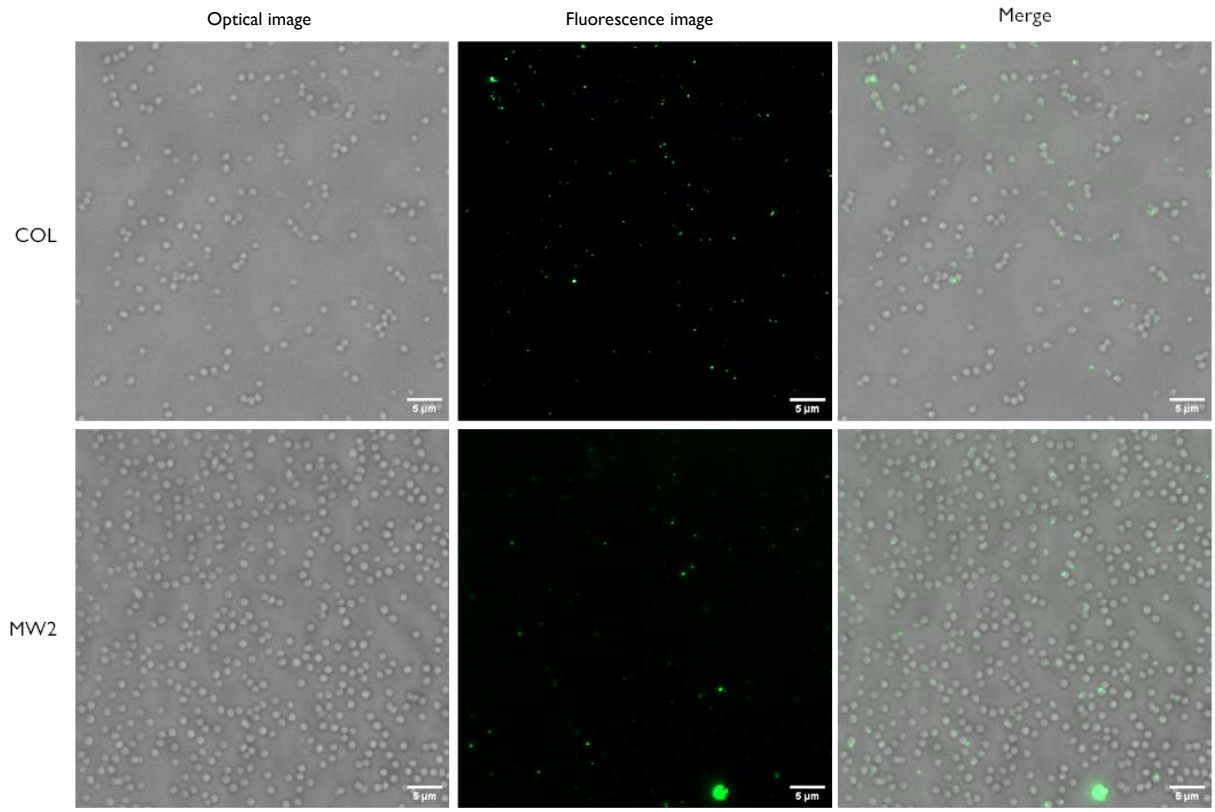


Figure 3.14. CBDMW2 binds to *S. aureus* strains at specific locations of the cell wall. On the left are the microcopy images, in the middle the fluorescence images and in the right is the superposition of the two imagens. The images on top and bottom are from COL and MW2 strains, respectively.

The protein was expected to bind randomly to the cell, covering almost all of it surface. Nonetheless, we observed that the protein binds to cells at specific locations, hence, only a small green dot on several cells was visible in the analysis in **Figure 3.14** (merge image). The reason for the CBDMW2 not binding to the entire cell surface may be because of the PG is not as exposed, and the accessibility is low because of the teichoic acids or others cell wall components. Those regions where

the protein binds could be cell division regions, and have less cell wall components that block the way to the PG.

3.5 Extraction of Peptidoglycan

For the peptidoglycan extraction, the strain COL of *Staphylococcus* was used, grown in two different media: TSB medium and minimal medium labeled with ^{15}N. The cells were propagated in 2 L for each medium with an initial OD_{620} of 0.02 AU. The final OD_{620} was 0.7 AU for TSB, and 0.9 AU for minimal medium. The yield for unlabeled peptidoglycan was 90.7 mg/L and 140.1 mg/L for ^{15}N peptidoglycan.

3.5.1 Enzymatic digestion and Muropeptide separation by RP-HPLC

After extraction and purification, the peptidoglycan was fragmented with mutanolysin, an enzyme that hydrolyses the β-1,4 linkages between N-Acetylmuramic acid and N-Acetylglucosamine. The peptidoglycan fragments (muropeptides) produced were analyzed and separated by RP-HPLC, with a C8 reversed-phase column.

The samples were collected per peak at characteristic retention times, that correspond to different peptidoglycan fragments, with the first fragments being the smallest, and over time, there is the elution of larger fragments, due to the overall charge distribution, the more non-polar and consequently the longer the retention time. Thus, the order of elution is, first monomers, then a repetition of two monomers (dimers), next trimers, tetramers, and oligomers with more repetitions [39]. These fragments are represented in **Figure 3.15**, by I – monomers, II – dimers, III- trimers, IV – tetramers, V and VI - bigger fragments, with more repetitions of the monomer.

Similar fractions from different injections were collected and lyophilized.

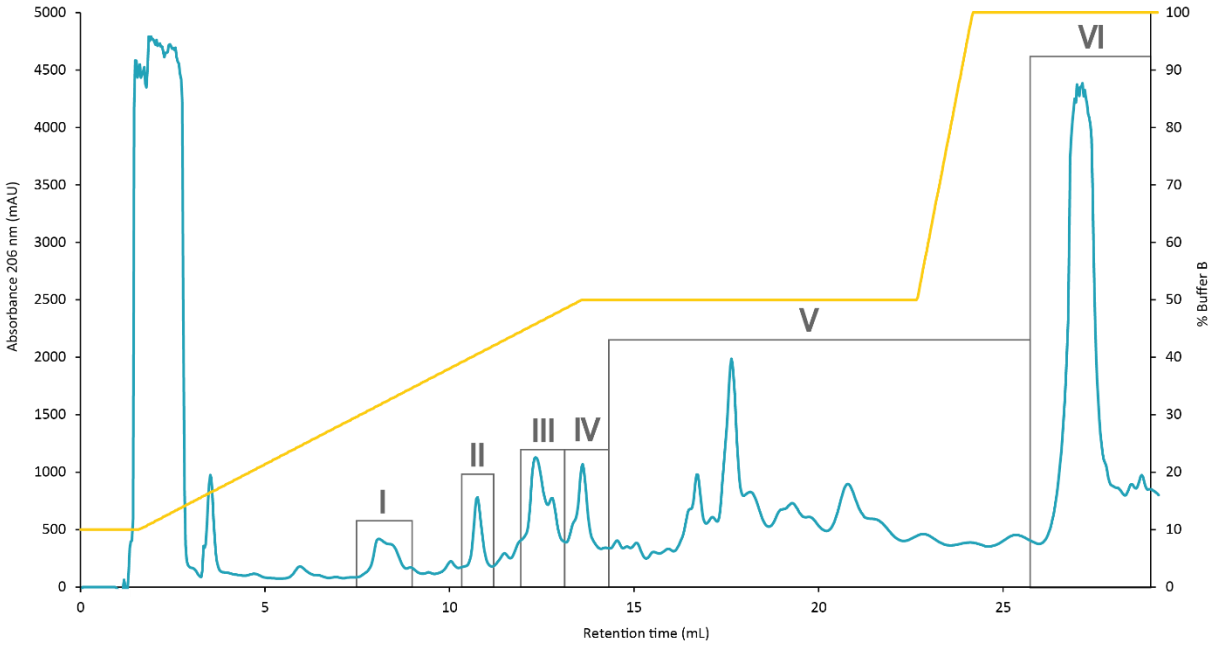


Figure 3.15. RP-HPLC profile of peptidoglycan from *S. aureus* strain COL. I – monomers, II – dimers, III- trimers, IV – tetramers, V and VI - fragments with more repetitions of the monomer. Gradient elution: Buffer A: Milli Q water and 0.05% TFA; Buffer B: Milli Q water + 0.05% TFA + 20% acetonitrile [40].

The samples containing fractions V and VI were resuspended and further digested with lysostaphin, while the remaining fractions were stored. This enzyme cleaves between the second and the third glycine residue of the pentaglycine cross-bridge [41]. The new digestion sample was separated and analyzed by RP-HPLC (**Figure 3.16**). In this separation, the only sample collected was the peak highlighted in **Figure 3.16** in red that showed a shift in retention time from the original fraction, indicating that lysostaphin digestion was successful. This sample was lyophilized and resuspended in appropriate solvent and characterized by NMR and ESI-MS, and used to titrate the CBDMW2 domain, monitored by ^{1}H-^{15}N-HSQC.

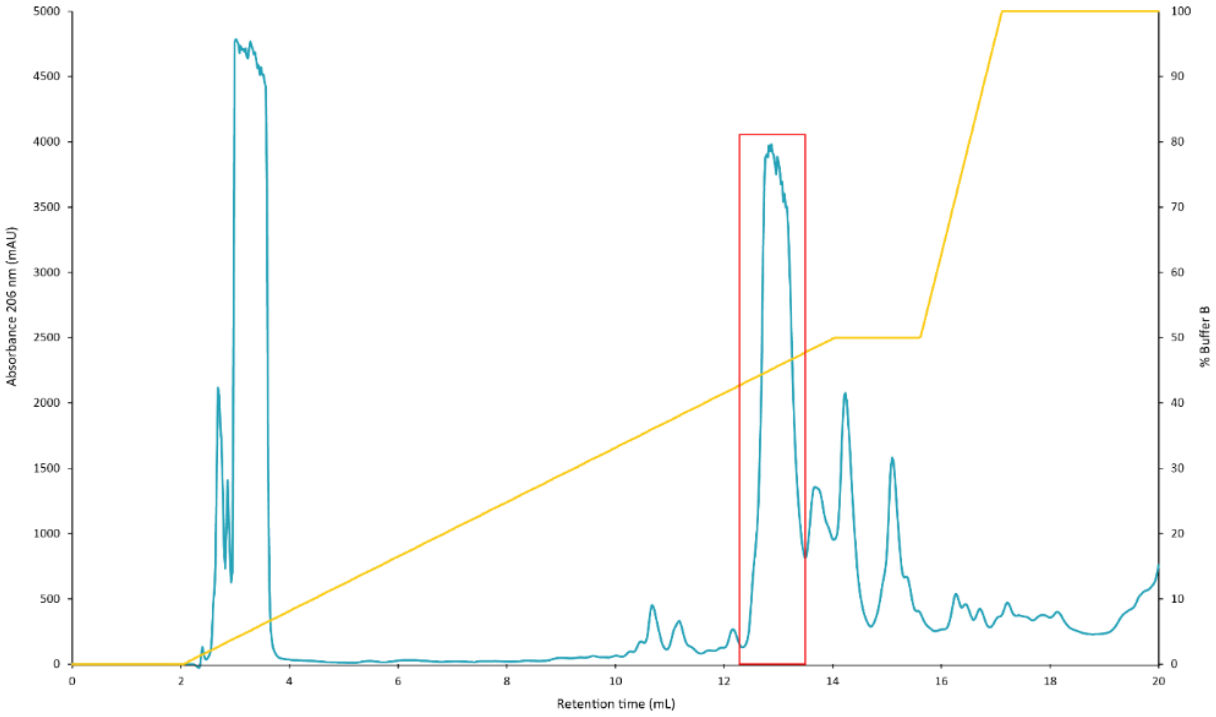


Figure 3.16. RP-HPLC profile (C8 column) of peptidoglycan from *S. aureus* strain COL. Peak of the muropeptide in. Gradient elution: Buffer A: Milli Q water and 0.05% TFA; Buffer B: Milli Q water + 0.05% TFA + 20% acetonitrile.

3.5.2 Muropeptides characterization

To evaluate the purity and to characterize the double digested muropeptide, the fraction was resuspended in 90% H_2O/10% D_2O, and 1D-^{1}H spectra, TOCSY, NOESY, ^{1}H-^{13}C-HSQC and ^{1}H-^{15}N-HSQC experiments were acquired.

TOCSY and NOESY spectra were used to assign the protons, once TOCSY correlates all protons within a given spin system, the magnetization transfer is based on scalar coupling, and NOESY correlates protons through space dipolar interaction, and is used to determine which protons are closer to each other in space, even though they are not connected.

All residues from the stem peptide and the glycines from the cross bridge were assigned through the typical ^{1}H-^{1}H TOCSY patterns of the amino acids (**Figure 3.17**, **Table 5.9** and **Table 5.10**, annexes). Four glycines were found, and to distinguish them, the NOESY spectrum was used. However, only two glycines could be correctly assigned, glycine 1 and glycine 5. Gly 1 presented a NOE between the Hα and the H_N of the side chain of L-Lys, the adjacent residue. Gly 5 presented a NOE between the H_N of the Gly with D-Ala Hα. The other two glycines were not possible to distinguish since Gly resonances are degenerated.

For the assignment of GlcNAc and MurNAc residues (**Figure 5.5**), TOCSY spectra was used. One anomeric proton was identified in the ^{1}H-^{13}C-HSQC spectra (^{13}C chemical shift around 100 ppm). This proton belongs to the GlcNAc residue. The remaining GlcNAc resonances were identified based

on the TOCSY patterns and expected chemical shift for this sugar. The MurNAc residue in this muropeptide is a reduced derivative of MurNAc (non sugar) and no anomeric proton is expected. The NHAc resonance and TOCSY patterns were used to assign the MurNAc resonances (**Figure 3.18** and **Table 5.11**, annexes).

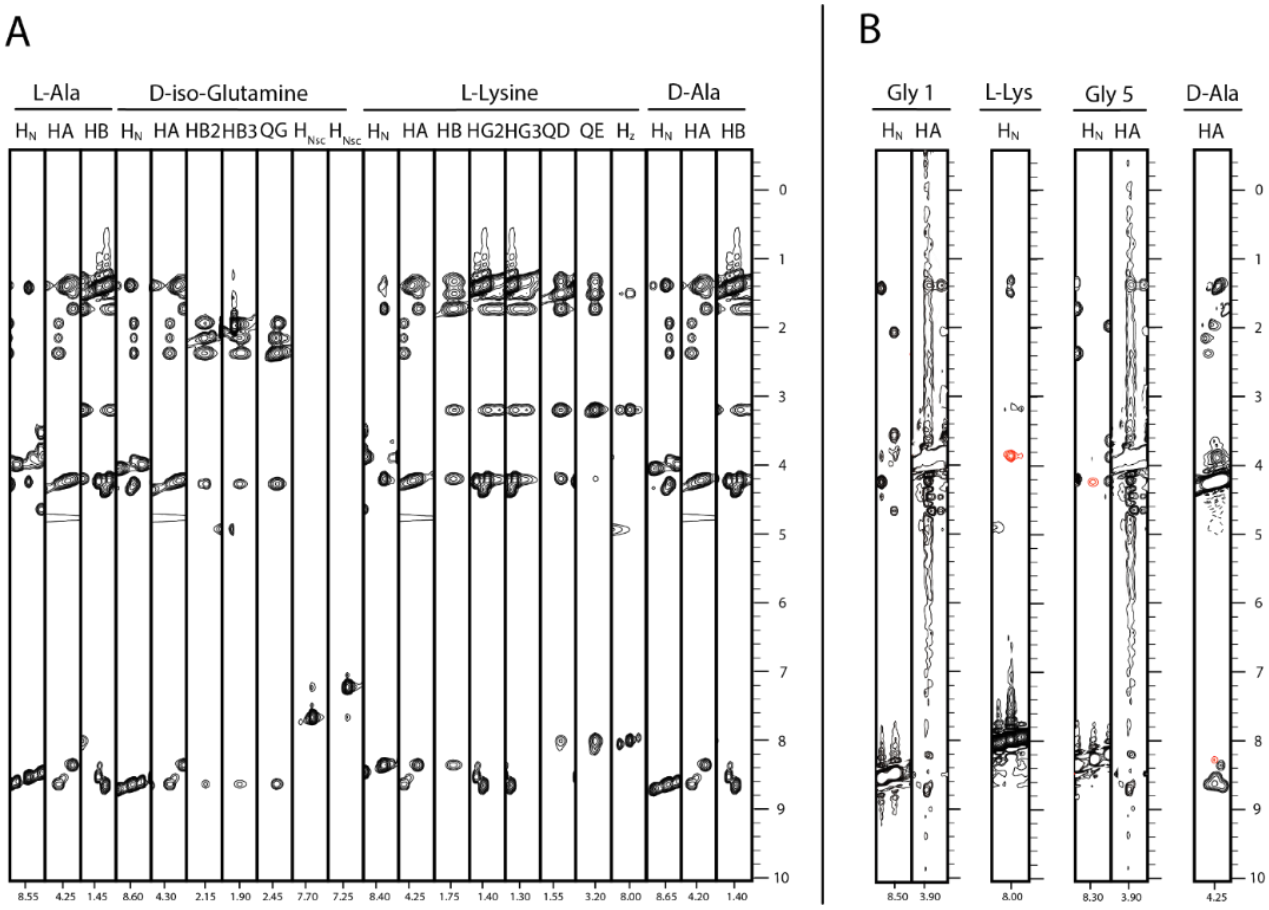


Figure 3.17. Strips from TOCSY and NOESY spectra of the muropeptide fragment. **(A)** Strips from TOCSY spectrum of the residues of the stem peptide. **(B)** Strips from NOESY spectrum of glycine 1 and the H_N of L-Lys. In red is represented the NOE between these two signals. Strips from NOESY spectrum of glycine 5 and the Hα D-Ala. In red is represented the NOE between these two signals.

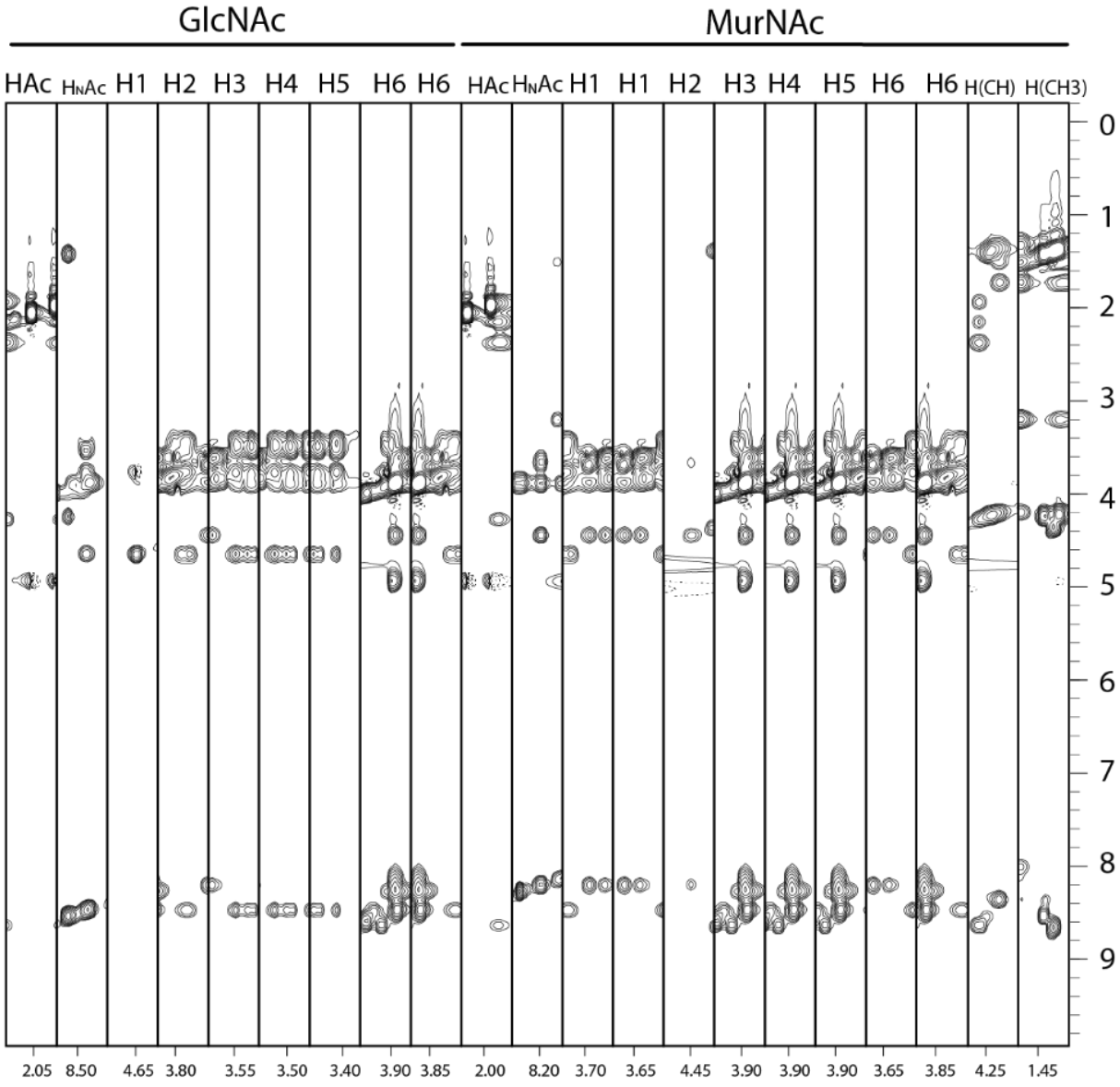


Figure 3.18. Strips from TOCSY spectrum of the sugars of the double digested muropeptide. On the left are the strips from TOCSY spectrum of GlcNAc, and in the right are the strips from TOCSY spectrum of MurNAc.

^{13}C and ^{15}N resonances were assigned from ^{1}H-^{13}C-HSQC and ^{1}H-^{15}N-HSQC spectra from proton resonances (**Figure 3.19** and **Figure 3.20**) [42, 43].

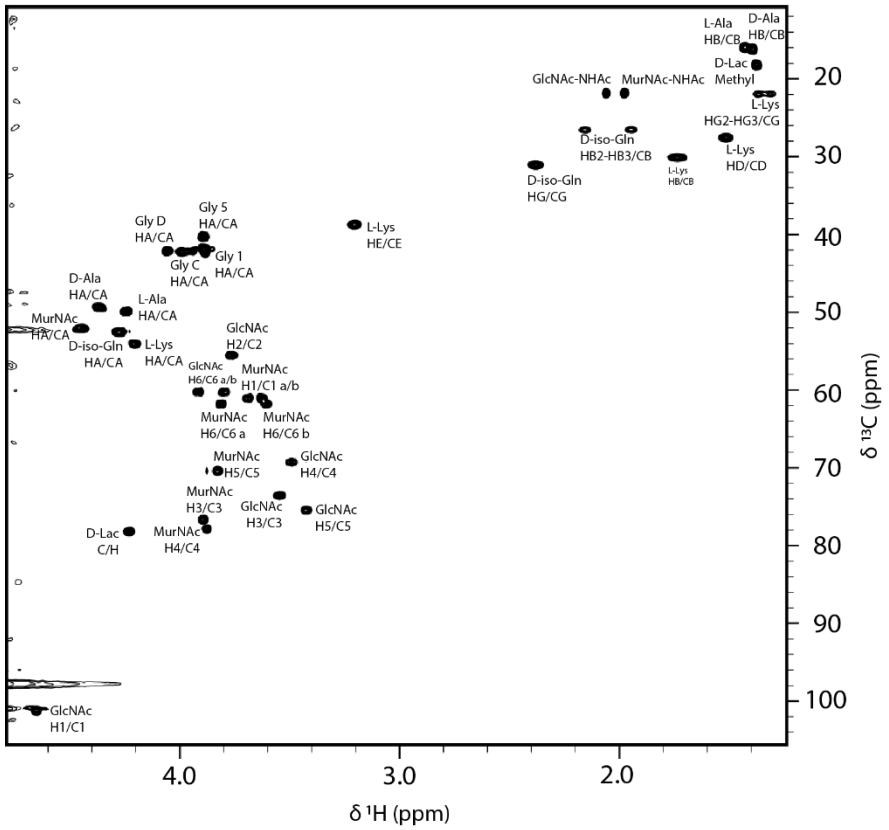


Figure 3.19. Fully assigned ^{1}H-^{13}C-HSQC spectrum of the muropeptide fragment at 2.07 mM of Muropeptide fragments in 90% H_2O/10% D_2O.

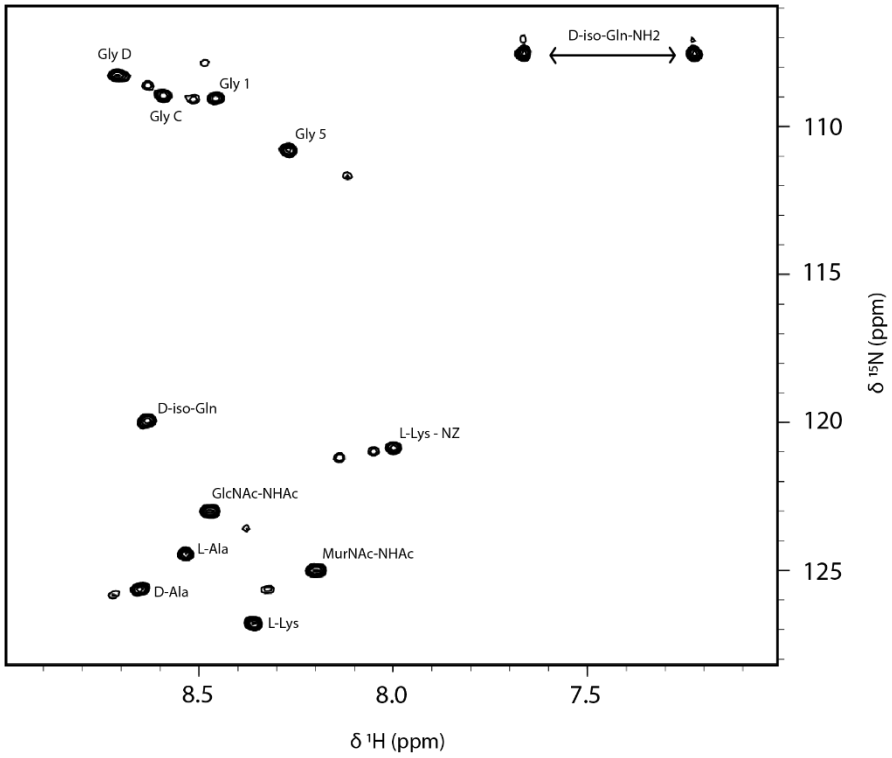


Figure 3.20. Fully assigned ^{1}H-^{15}N-HSQC spectrum of the muropeptide fragment at 2.07 mM of Muropeptide fragments in 90% H_2O/10% D_2O.

3.5.3 Electrospray Ionization Mass Spectrometry

The double digested sample was analyzed by Electrospray Ionization Mass Spectrometry (**Figure 3.21**). The PG was digested with mutanolysin, that hydrolyses the β-1,4 linkages between N-Acetylmuramic acid and N-Acetylglucosamine, and lysostaphin, that cleaves the bond between the glycine 2 and 3 in the pentaglycine peptide. This way, it is expected that the fragments contains the two sugars, GlcNAc and MurNAc, the stem peptide and (L-Ala-D-iso-Gln-L-Lys-D-Ala), and five glycines, two bonded to the L-Lysine, and three bonded to the D-Alanine.

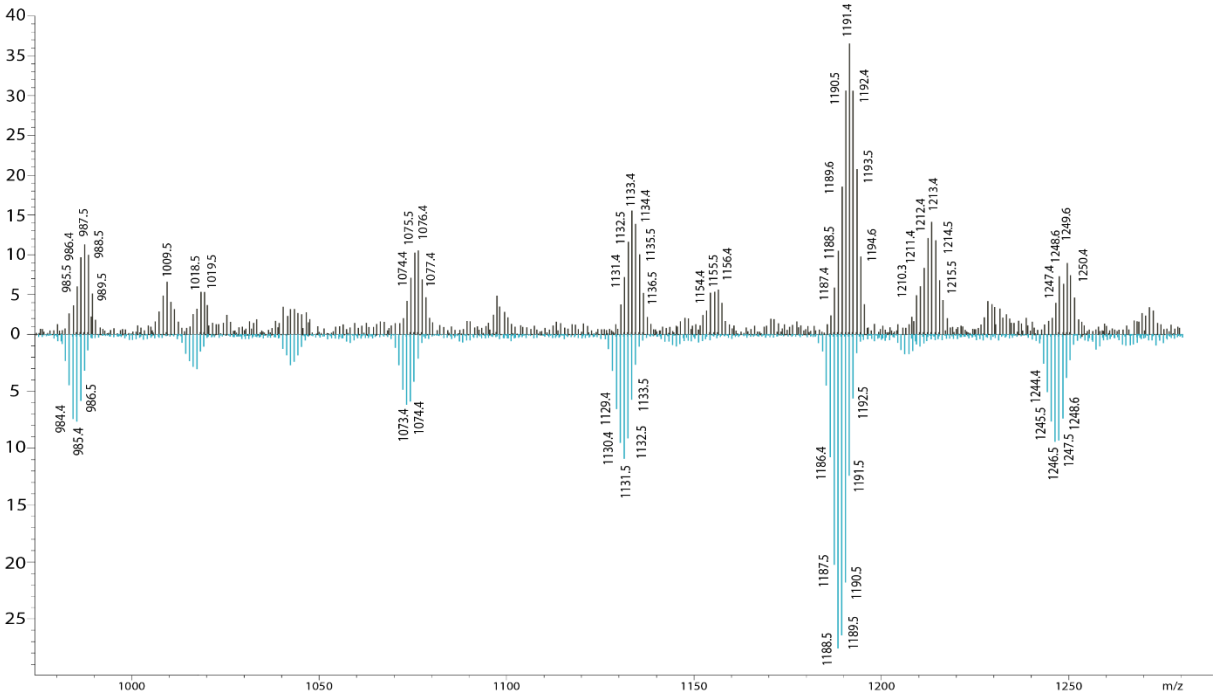


Figure 3.21. Plot of experimental ESI-MS/MS spectra of the double digested sample. In black the measure was performed in positive ion mode. Blue is the plot measured in negative ion mode.

The theoretical mass intensity for the fragment is 1182 Da, assuming a neutral charge [44]. Through the analysis of mass chromatogram it is possible to observe a major mass peak around 1191.4 for positive mode (+1 charge), and around 1189.5 for the negative mode (-1 charge). The 8 Da difference between the theoretical mass and the actual mass, can be attributed to an isotopic effect/enrichment (natural abundance). Based on this m/z values, the molecule with most abundance is the muropeptide with two sugars: GlcNAc and MurNAc; the stem peptide (L-Ala-D-iso-Gln-L-Lys-D-Ala), and five glycines. The other less abundant peaks differ on a value of 57 Da, that corresponds to the molecular weight of the glycine residue. This way, the most abundant fragment present 5 glycines.

With all the information given by the NMR and the ESI-MS the structure obtained for the major species and is represented in **Figure 3.22**. This molecule present a molecular weight of 1182 Da.

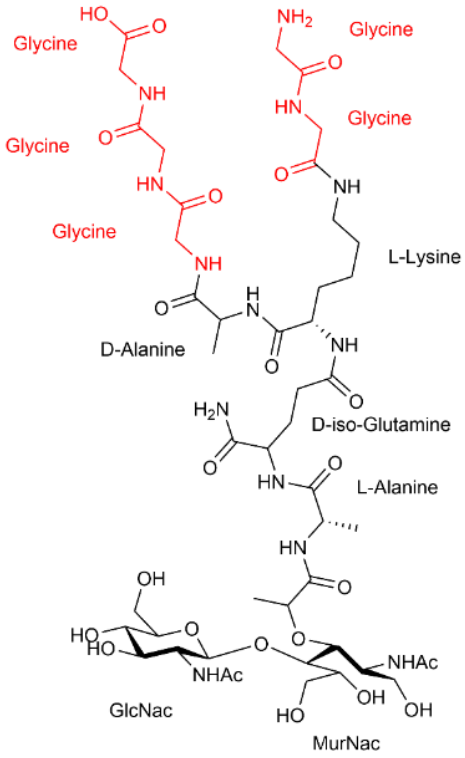


Figure 3.22. Structure of the double digested muropeptide. The muropeptide contains two sugars: GlcNAc and MurNAc, the stem peptide (L-Ala-D-iso-Gln-L-Lys-D-Ala), three glycines linked to the D-Alanine, and two glycines linked to the L-lysine, represented in red.

3.6 MicroScale Thermophoresis

MST is based on thermophoresis, the movement of molecules in a temperature gradient. Likewise, this technique has the ability to derive dissociation constants and accurately track molecular interactions. This way, MST was used to analyze the binding between the CBDMW2 domain and small PG fragments of *S. aureus*. The MST assay was performed using a final concentration of 100 nM of GFP-CBDMW2, and serial dilutions of the ligand, with concentrations between 37.5 µM and 1.14 x 10^{-3} µM (37.5 µM, 18.8 µM, 9.38 µM, 4.69 µM, 2.34 µM, 1.17 µM, 5.86 x 10^{-1} µM, 2.93 x 10^{-1} µM, 1.46 x 10^{-1} µM, 7.32 x 10^{-2} µM, 3.66 x 10^{-2} µM, 1.83 x 10^{-2} µM, 9.16 x 10^{-3} µM, 4.58 x 10^{-3} µM, 2.29 x 10^{-3} µM, 1.14 x 10^{-3} µM). The ligand used was PG fragmented first with amidase and then with mutanolysin. This digestion originated peptides with different sizes. The muropeptide mix was quantified by 1D-^{1}H NMR spectra, using TSP as reference. This estimated concentration is the concentration of "monomeric forms".

In MST experiment a K_D is determined, through the evaluation of thermophoresis of the labeled molecule over the 16 different samples, and can then be evaluated as a function of unlabeled-partner concentration, with K_D being a fitted parameter in the analysis of this "binding curve" [45]. The fit yielded an apparent K_D of 211.39 ± 110.16 nM from three independent measurements (**Figure 3.23**).

This value indicates the formation of a high affinity complex. Complementary experiments with different muropeptides and controls were not able to be done because the equipment broke.

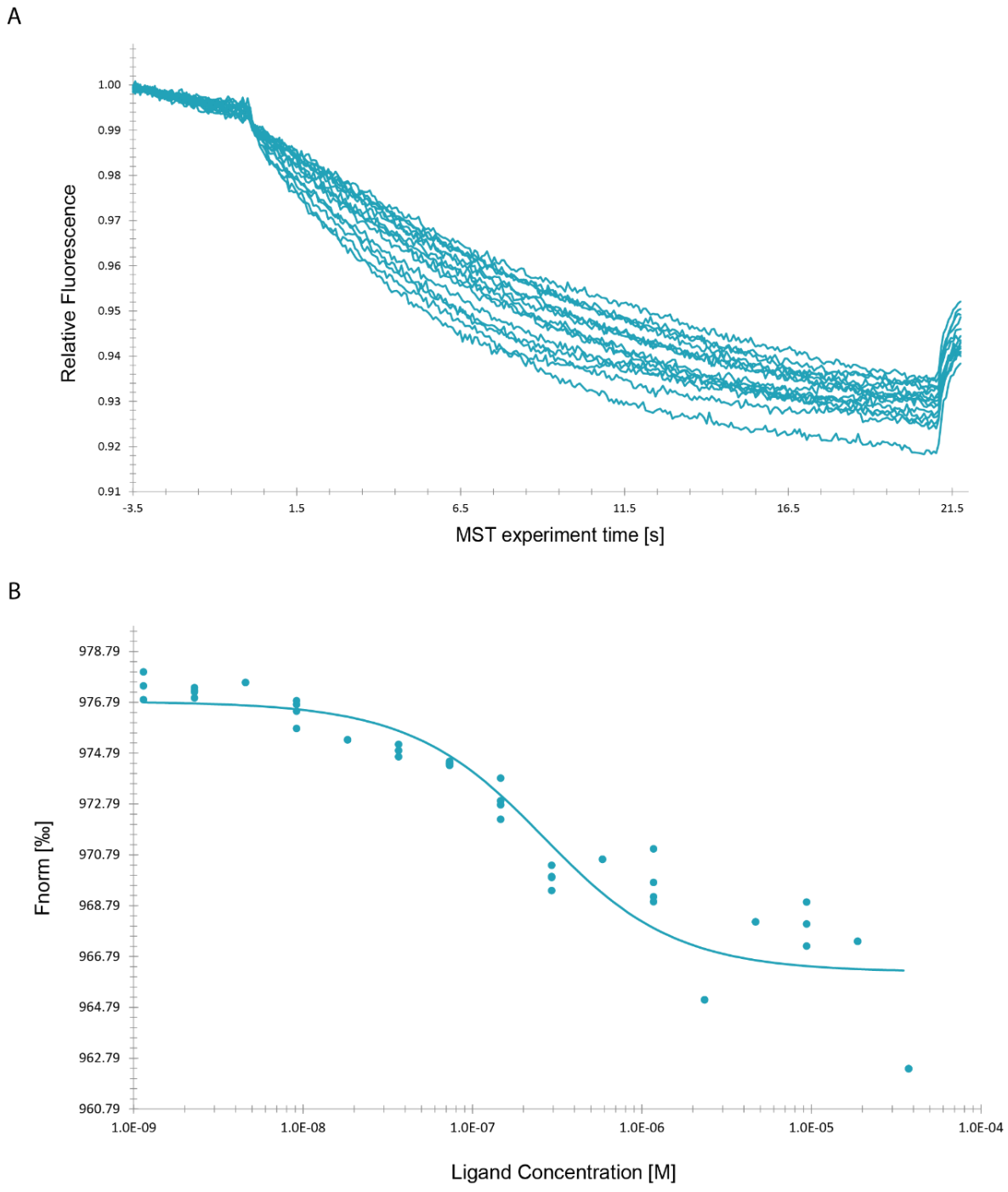


Figure 3.23. Thermophoretic analysis of the interaction between GFP-CBDMW2 and PG fragments. (**A**) MST trace of titrations of PG fragments against 100 nM GFP-CBDMW2. (**B**) Changes in thermophoresis of a titration with small PG fragments with concentrations from 37.5 µM to 1.14 x 10^{-3} µM against 100 nM GFP-CBDMW2 are shown (concentrations of the PG fragments: 37.5 µM, 18.8 µM, 9.38 µM, 4.69 µM, 2.34 µM, 1.17 µM, 5.86 x 10^{-1} µM, 2.93 x 10^{-1} µM, 1.46 x 10^{-1} µM, 7.32 x 10^{-2} µM, 3.66 x 10^{-2} µM, 1.83 x 10^{-2} µM, 9.16 x 10^{-3} µM, 4.58 x 10^{-3} µM, 2.29 x 10^{-3} µM, 1.14 x 10^{-3} µM). Fit of the changes in thermophoresis signal yielded a K_D of 211.39 ± 110.16 nM.

3.7 Titration of CBDMW2 domain

3.7.1 5 Glycine Bridge

The CBDMW2 was titrated with a peptide with five glycines with different ratios of peptide: 1:0, 1:17, 1:31.5, 1:63.5, 1:95, and 1:190. The interaction studies between the domain and the peptide were monitored by NMR using the ^{1}H-^{15}N-HSQC spectrum, acquiring spectra for the different concentrations of peptide.

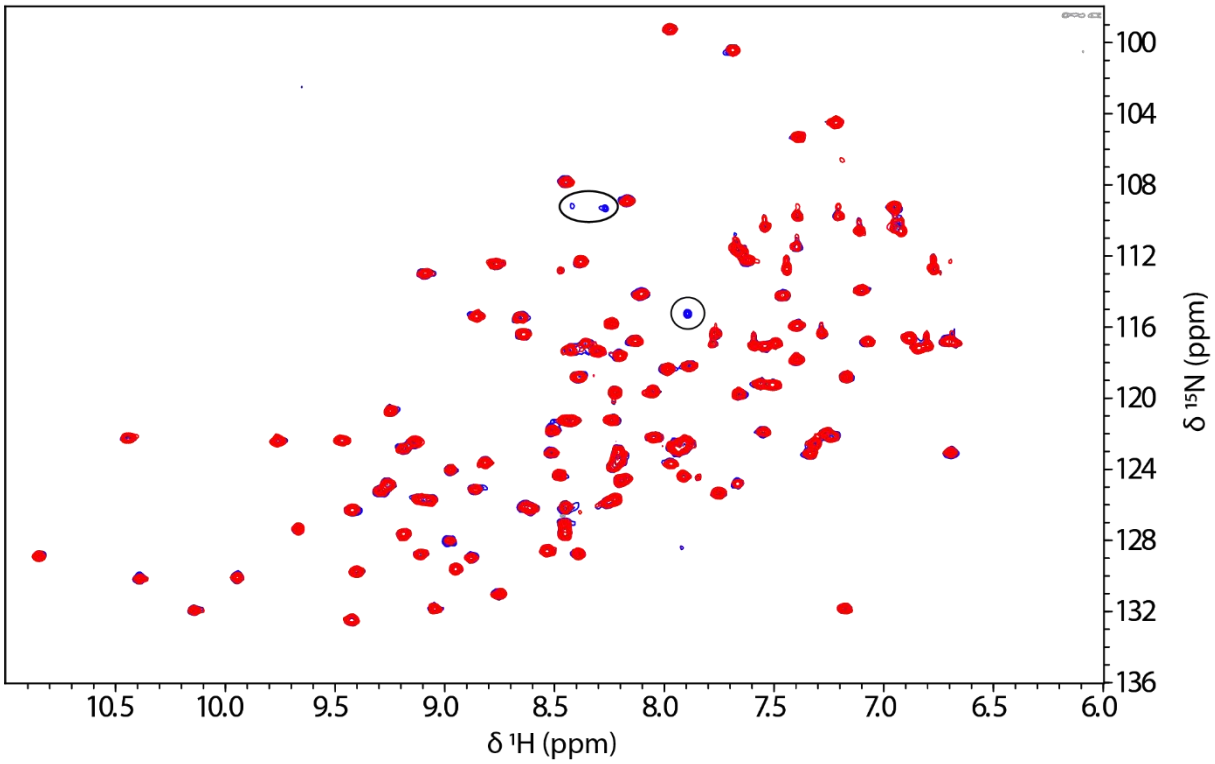


Figure 3.24. Monitoring the interaction between CBDMW2 protein and a peptide with 5 glycines using ^{1}H-^{15}N-HSQC spectrum. Overlay of ^{1}H-^{15}N-HSQC spectra acquired with the ratios 1:0 (red) and 1:190 (blue). In black is marked the signals corresponding to the glycines of the peptide. CBDMW2 samples were prepared with 10 mM PO_4 pH 6.0, 1 mM BME and 90% H_2O/10% D_2O.

In the analysis of the titration of the CBDMW2 domain with the peptide of glycines, no chemical shift or intensity variation could be observed in the ^{1}H-^{15}N-HSQC spectrum. The observed change is the appearance of signals corresponding to the glycines of the peptide (marked in black, **Figure 3.24**). Under the tested conditions there is no interaction between the CBDMW2 and the peptide with five glycines.

3.7.2 *S. aureus* muropeptide

For the titration of the CBDMW2 domain with the muropeptide, the concentration ratios used were: 1:0, 1:0.8, 1:1.58, 1:3, 1:6.6, 1:11, 1:16.6, and 1:33. To follow the interaction between the CBDMW2 domain and the muropeptide, ^{1}H-^{15}N-HSQC were acquired (**Figure 3.25**).

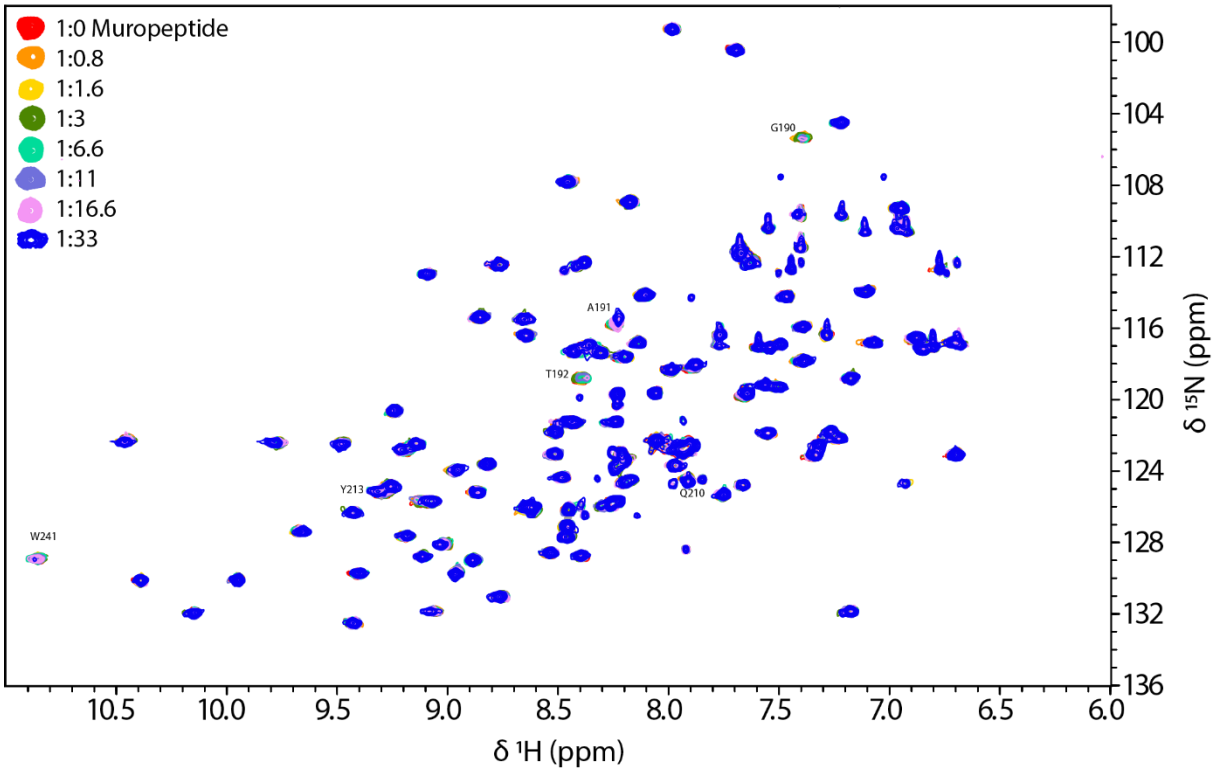


Figure 3.25. Monitoring the interaction between CBDMW2 protein and muropeptide using ^{1}H-^{15}N-HSQC spectrum. Overlay of ^{1}H-^{15}N-HSQC spectra acquired with the ratios 1:0 (red), 1:0.8 (orange), 1:1.6 (yellow), 1:3 (green), 1:6.6 (emerald green), 1:11 (purple), 1:16.6 (pink) and 1:33 (blue). CBDMW2 samples were prepared with 10 mM PO_4 pH 6.0, 1 mM BME and 90% H_2O/10% D_2O.

Figure 3.25 shows some chemical shift variations and/or line broadening, in specific the signals G190, A191, T192, Q210, Y213, and the side chain signal for W241. As an illustration, **Figure 3.26** displays the chemical shift of two residues, G190 and A191. However, for residue G190 and others, it was not possible to monitor their resonances at all titration points as they showed a line broadening. This information demonstrates that the CBDMW2 domain interacts with the double digested muropeptide of *S. aureus*.

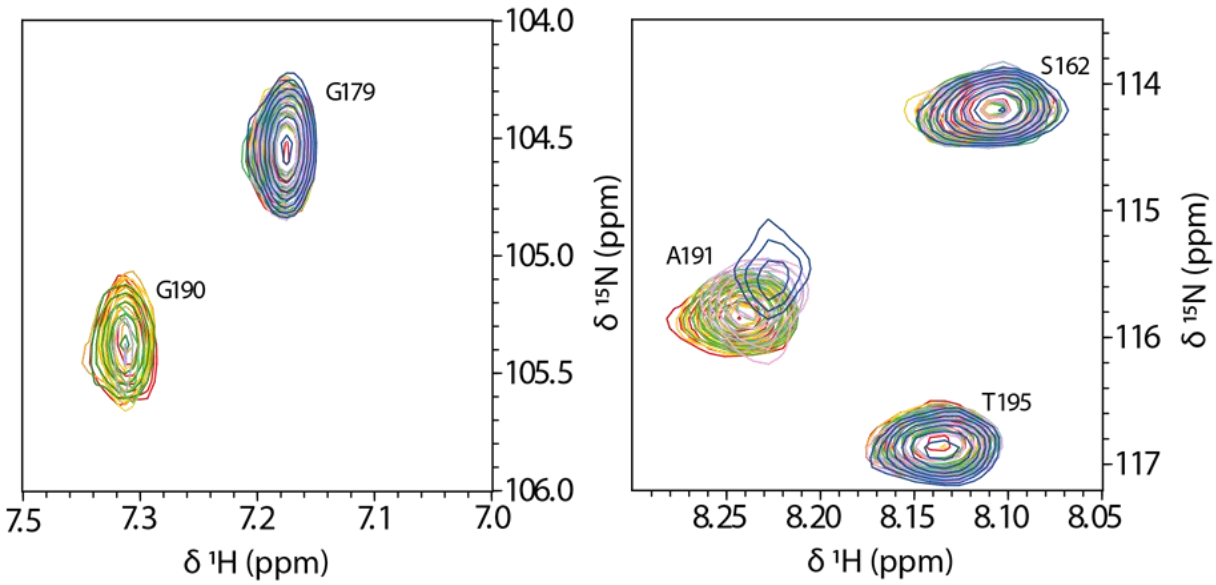


Figure 3.26. Monitoring chemical shift of the residue G190 and A191 in the interaction between CBDMW2 protein and the PG fragment Overlay of ^{1}H-^{15}N-HSQC spectra acquired with the ratios 1:0 (red), 1:0.8 (orange), 1:1.6 (yellow), 1:3 (green), 1:6.6 (emerald green), 1:11 (purple), 1:16.6 (pink) and 1:33 (blue). CBDMW2 samples were prepared with 10 mM PO_4 pH 6.0, 1 mM BME and 90% H_2O/10% D_2O.

In order to understand the binding site on CBDMW2 the chemical shift perturbation (CSP) was used. This technique mostly use the changes of chemical shift in the ^{1}H-^{15}N-HSQC spectrum. Thus, for all the titration spectra, the CSP for the backbone and side chains amides, were calculated using the following equation (2) [46]:

$$\Delta\delta total = \sqrt{[(\Delta\delta_H)^2 + (0.2 \times \Delta\delta_N)^2]} \quad (2)$$

The CSP was calculated for all the titration points (1:0, 1:0.8, 1:1.6, 1:3, 1:6.6, 1:11, 1:16.6 and 1:33). For the last point of the titration, two cut-off values were selected to identify the residues with the most pronounced values of CSP, 0.02 for the backbone and for the side chain, this way it is possible to identify the residues that are most involved in the interaction. The residues with asterisks, G190, T192 and R244, are the ones that could not be monitored for their deviations, and were assigned a maximum CSP value (0.045 for backbone, 0.040 for side chain) (**Figure 3.27**).

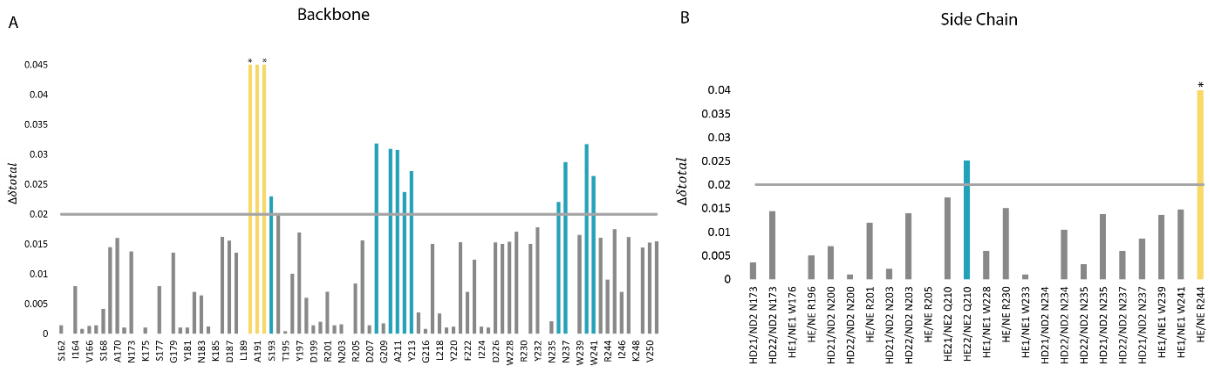


Figure 3.27. Chemical shift perturbation analysis (CSP) of the interaction between the CBDMW2 domain and the fragments of peptidoglycan from *S. aureus* (1:33). **(A)** CSP values for the backbone. Blue (cut-off of 0.02) and yellow indicate the residues whose signals in the ^{1}H-^{15}N-HSQC spectrum have disappeared due to extensive line broadening, and were assigned a maximum CSP value. **(B)** CSP values for the side chain. Blue are the residue with CSP value higher than the cut-off of 0.02, and yellow indicate the residues whose signals in the ^{1}H-^{15}N-HSQC spectrum have disappeared due to extensive line broadening, and were assigned a maximum CSP value.

In the 3D structure of the CBDMW2 domain, the residues with higher CSP values were mapped. (**Figure 3.28**). The residues involved in the interaction with the PG fragments that presented higher CSP values are located in a single region.

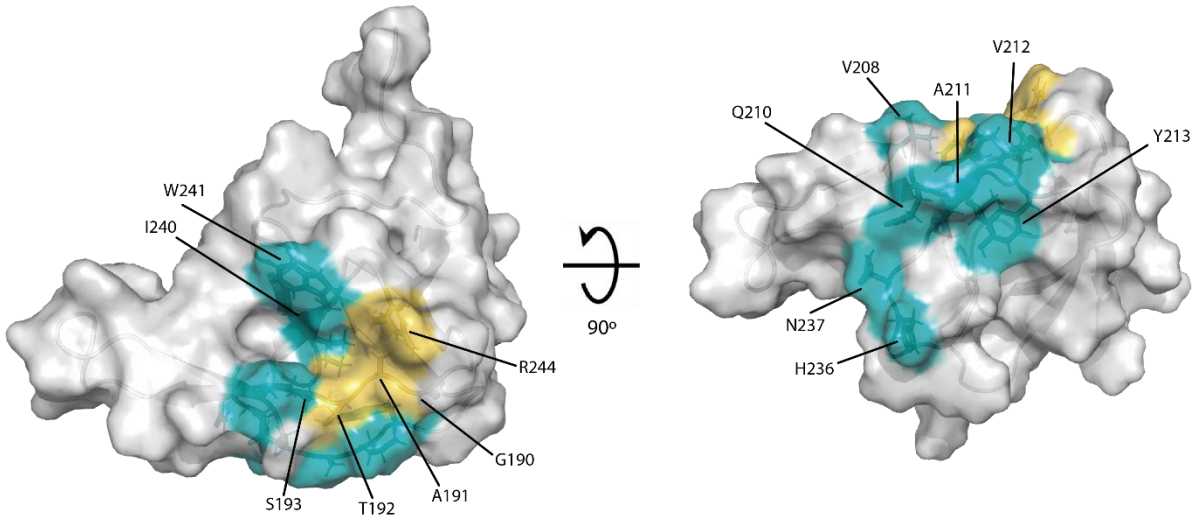


Figure 3.28. Mapping of the most pronounced values of CSP in the 3D structure. Blue is for the residues higher than the cut-off 0.02 and yellow are the residues whose signals in the ^{1}H-^{15}N-HSQC spectrum have disappeared due to extensive line broadening, and were assigned a maximum CSP value.

The results of the interaction were compared with the data of the SH3b domain of lysostaphin (pdbid: 6RK4) [38]. Since, the lysostaphin also have a region that interacts with stem peptide of the peptidoglycan (**Figure 3.29**).

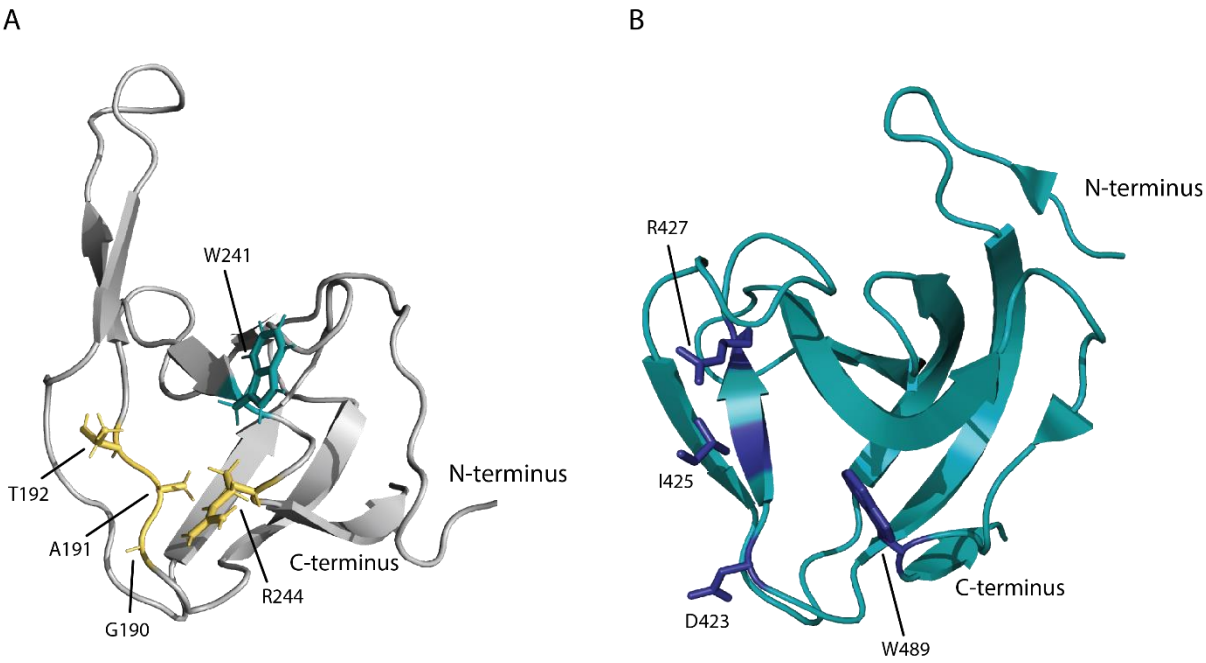


Figure 3.29. **(A)** 3D structure of CBDMW2 with the main residues that interacts with the stem peptide **(B)** 3D structure of the SH3b domain of Lysostaphin (pdbid: 6RK4) with the main residues that interacts with the stem peptide [38]. The two 3D structures have the same orientation.

The region of interaction from CBDMW2 domain overlap with region in the lysostaphin SH3b domain involved in binding to the stem peptide of the peptidoglycan. The two proteins present a common region of interaction but with different amino acids. No amino acid homology exists between both regions, indicating a different binding mode for the stem peptide. Yet both regions present a Trp and Arg, but their role in CBDMW2 interaction with the stem peptide remains unclear. In addition, the role of the glycine bridge is not clear. More complex muropeptides with the stem peptide and the complete glycine bridge (five glycines) and with different oligomerization might help to answer this question.

4

CONCLUSIONS AND FUTURE PERSPECTIVES

Antimicrobial resistance (AMR) is a global threat to human, animal, and plant health, food security and the achievement of many of the UN Sustainable Development Goals. In addition, it has a negative impact on food production, the environment and the global economy. *S. aureus* are among the most common human fast-growing bacteria causing acute human infections; and MRSA bacteremia is associated with significant healthcare costs, morbidity, and mortality. MRSA resistant strains to 3rd-generation cephalosporin are largely disseminated and found in high frequency in human infections observed in hospitals all over the world and increasingly more frequent in the community. These types of infections lead to the increase in use of the last resort drugs (e.g., vancomycin) against which new strains are emerging. A more effective management of MRSA strains by early detection and consequent elimination of these strains will reduce the risk of new infections and help control MRSA spread and ultimately diminish the number of new infections. In addition, prompted efforts to develop novel antimicrobials are current underway, with a preference for the more sustainable and innovative techniques brought forth by the advances in recent years in the fields of synthetic biology, biotechnology and biochemistry.

A promising contribution lies on repurposing lytic enzymes produced either by bacteriophages known to infect the *Staphylococcus* genus or by staphylococcal species themselves as part of their inter-specific competition mechanism. For instance, lysostaphin, an endopeptidase secreted by *S. simulans*, has been expansively studied as a therapeutic agent to treat MRSA infections, being considered a valid treatment, either alone or in combination with antibiotics.

The PlyMW2 endolysin found in a prophage in the *S. aureus* subspecies aureus MW2 widens the field of available modules due to its divergent cell wall binding domain. PlyMW2 has two structural and functional domains: a cysteine, histidine-dependent amidohydrolases/peptidase (CHAP) domain, and the cell wall-binding domain (CBD) with no known structural homologs.

The main objective or goal of this work was to understand the interaction between the CBDMW2 domain with the peptidoglycan of *S. aureus* using different biochemical techniques such as NMR.

To achieve this, the gene for the CBDMW2 was amplified from *S. aureus* MW2 strain genomic DNA with compatible primers and cloned into modified pET vectors containing N-terminus His-tag, GB1 fusion or GFP fusion. All tags could be removed with the TEV protease. The expression and purification of CBDMW2 was achieved by IMAC with high yields, approximately 17 mg of pure protein per liter of culture. Isotopically enriched CBDMW2 samples (^{15}N and $^{15}N/^{13}C$) were purified from cells grown in enriched M9 minimal media ($^{15}NH_4Cl$ and ^{13}C-glucose). Double labeled sample was used to acquire 3D-triple resonance experiments for resonance assignment and for noe based-structure calculation.

Additionally, unlabeled and ^{15}N-enriched peptidoglycan was extracted and purified from cultures of *S. aureus* COL strain grown in TSB or in a defined minimal media with $^{15}NH_4Cl$ as the nitrogen source. The yield for unlabeled and ^{15}N-enriched peptidoglycan was 90.7 mg/L and 140.1 mg/L, respectively. The main difference in yields in pure peptidoglycan can be due to the fact that the culture in TSB was centrifuged at a lower OD_{600} of 0.7 while the COL culture in the defined minimal media was centrifuged at an OD_{600} of 0.9 with a consequent higher cell mass.

The unlabeled peptidoglycan was digested with mutanolysin (N-acetylmuramidase) and the resulting muropeptides were separated by reverse phase HPLC. The muropeptides were collected according to their oligomerization/size. The purified muropeptides corresponding to monomers and to "dimers" (containing 2 repeats) were analyzed by 1D NMR and were not pure. In addition, the yields for these two muropeptides were low and not enough to pursue to the interactions studies. An alternative was to use the fractions of bigger muropeptides and hydrolyze these with lysostaphin (endopeptidase with Glycyl-Glycyl activity). The double digested muropeptide was purified by reverse phase HPLC and analyzed by ESI-MS and NMR. Resonances for double digested muropeptide was achieved using TOCSY, NOESY, ^{1}H-^{13}C-HSQC and ^{1}H-^{15}N-HSQC experiments. The muropeptide analyzed from the lysostaphin digestion contains the two sugars, GlcNAc and MurNAc (or a reduced derivative due to the treatment with sodium borohydride), the stem peptide (L-Ala-D-iso-Gln-L-Lys-D-Ala), three glycines linked to the D-Alanine, and two glycines linked to the L-lysine. This sample was used for NMR titrations with CBDMW2 and allow us to identify a number of residues involved in this interaction, including residues G190, A191, T182, W241 and R244 that suffer extensive broadening and cluster in the same region of the CBDMW2 structure.

The solution structure of CBDMW2 is still in progress and is being done by other members of the lab. Yet, preliminary structures obtained from noeassign module from CYANA (RMSD 20

structures 0.58 Å/1.06 Å; TF 0.72 with 15.47 restrictions per residues) allowed us to say that CBDMW2 has a SH3b-type fold common for CBDs of other lysins and involved in peptidoglycan binding. This preliminary NMR structure was validated by the one obtained from the AlphaFold program (RMSD 2 Å). Despite this structural homology there is still low sequence homology. The CBDMW2 structure shows remarkably differences, in length and sequence, in the loops characteristic for these domains (namely RT loop, src, loop, extra loop).

These differences are notorious in the muropeptide binding. The CBDMW2 region of interaction overlaps with the region in the lysostaphin SH3b domain that is involved in binding to the stem peptide of the peptidoglycan although both domains do not share conserved amino acid residues. The absence of amino acid homology between the two regions suggests that they bind differently to the stem peptide.

In order to understand the role of these amino acids in CBDMW2 interaction with the stem peptide and to understand the role of the glycine pentapeptide in peptidoglycan binding, more complex muropeptides with the stem peptide and the complete pentaglycine bridge and with different oligomerization will be produced (including ^{15}N labeled). In addition, the use of mutagenesis in the amino acids involved in the interaction should help understand their importance in the interaction. The effects of these modifications can be tested by co-precipitation assays. New MST experiments should be done, to better understand the interaction of the CBDMW2 with the peptidoglycan. STD-NMR experiments with different ligands can help better understand about the binding mode of the CBDMW2 domain with the peptidoglycan. Crystallography will also be attempted to obtain a high-resolution structure of the CBDMW2 with the muropeptide fragments.

BIBLIOGRAPHY

[1] J. M. Lourenço, "The NOVAthesis LATEX Template User's Manual," [Online]. Available: https://github.com/joaomlourenco/novathesis_word/raw/master/novathesis_word-FINAL-EN.pdf. [Acedido em 13 October 2022].

[2] T. J. Foster e J. A. Geoghegan, "Staphylococcus aureus," em *Molecular Medical Microbiology*, Dublin, Academic Press, 2015, pp. 655-674.

[3] S. Lakhundi e K. Zhang, "Methicillin-Resistant Staphylococcus aureus: Molecular Characterization, Evolution, and Epidemiology," *Clinical Microbiology Reviews,* vol. 31, nº 4, 18 september 2018.

[4] N. A. Turner, B. K. S.-. Kuinkel, S. A. Maskarinec, E. M. Eichenberger, P. P. Shah, M. Carugati, T. L. Holland e V. G. F. Jr, "Methicillin- resistant Staphylococcus aureus: an overview of basic and clinical research," *Nature Reviews Microbiology,* vol. 17, nº 4, pp. 203-218, 8 february 2019.

[5] M. Patel, "Community-Associated Meticillin-Resistant Staphylococcus aureus Infections: Epidemiology, Recognition and Management," *Drugs,* vol. 69, nº 6, pp. 693-716, 2009.

[6] M. Gajdács, "The Continuing Threat of Methicillin-Resistant Staphylococcus aureus," *antibiotics,* vol. 8, nº 2, p. 52, 2 may 2019.

[7] W. A. McGuinness, N. Malachowa e F. R. DeLeo, "Vancomycin Resistance in Staphylococcus aureus," *Yale Journal of Biology and Medicine,* vol. 90, nº 2, pp. 269-281, 23 June 2017.

[8] S. Porfírio e R. W. C. P. Azadi, "Elucidating Peptidoglycan Structure: An Analytical Toolset," *Trends in Microbiology,* vol. 27, nº 7, pp. 607-622, July 2019.

[9] M. Jarick, U. Bertsche, M. Stahl, D. Schultz, K. Methling, M. Lalk, C. Stigloher, M. S. A. Schlosser e K. Ohlsen, "The serine/threonine kinase Stk and the phosphatase Stp regulate cell wall synthesis in Staphylococcus aureus," *Scientific Reports,* vol. 8, 12 September 2018.

[10] B. V. Gonçalves, R. Portela, R. Lobo, T. A. Figueiredo, I. R. Grilo, A. M. Ludovice, H. d. Lencastre, J. S. Dias e R. G. Sobral, "Role of MurT C-Terminal Domain in the Amidation of Staphylococcus aureus Peptidoglycan," *Antimicrobial Agents and Chemotherapy,* vol. 63, nº 10, October 2019.

[11] J. A. F. Sutton, O. T. Carnell, L. Lafage, J. Gray, J. Biboy, J. F. Gibson, E. J. G. Pollitt, S. C. Tazoll, W. Turnbull, N. H. Hajdamowicz, B. Salamaga, G. R. Pidwill, A. M. Condliffe e S. J. Foster, "Staphylococcus aureus cell wall structure and dynamics during host-pathogen interaction," *PLoS Pathogens,* vol. 17, nº 3, p. e1009468, 31 March 2021.

[12] J. M. Monteiro, G. Covas, D. Rausch, S. R. Filipe, T. Schneider, H.-G. Sahl e M. G. Pinho, "The pentaglycine bridges of Staphylococcus aureus peptidoglycan are essential for cell integrity," *Scientific Reports,* vol. 9, nº 1, 21 March 2019.

[13] T. Olszak, A. Latka, B. Roszniowski, M. A. Valvano e Z. Drulis-Kawa, "Phage life cycles behind bacterial biodiversity," *Current medicinal chemistry,* vol. 24, nº 36, pp. 3987-4001, 2017.

[14] K. Abdelkader, H. Gerstmans, A. Saafan, T. Dishisha e Y. Briers, "The Preclinical and Clinical Progress of Bacteriophages and Their Lytic Enzymes: The Parts are Easier than the Whole," *viruses,* vol. 11, nº 2, p. 96, 24 January 2019.

[15] H. H. Kashani, M. Schmelcher, H. Sabzalipoor, E. S. Hosseini e R. Moniri, "Recombinant Endolysins as Potential Therapeutics against Antibiotic-Resistant Staphylococcus aureus: Current Status of Research and Novel Delivery Strategies," *Clinical Microbiology Reviews,* vol. 31, nº 1, pp. e00071-17, January 2018.

[16] M. R. Clokie, A. D. Millard, A. V. Letarov e S. Heaphy, "Phages in nature," *Bacteriophage,* vol. 1, nº 1, pp. 31-45, 1 January 2011.

[17] G. P. C. Salmond e P. C. Fineran, "A century of the phage: past, present and future," *Nature Reviews Microbiology,* vol. 13, nº 12, pp. 777-786, 9 November 2015.

[18] F. Abdelrahman, M. Easwaran, O. I. Daramola, S. Ragab, S. Lynch, T. J. Oduselu, F. M. Khan, A. Ayobami, F. Adnan, E. Torrents, S. Sanmukh e A. El-Shibiny, "Phage-Encoded Endolysins," *Antibiotics,* vol. 10, nº 2, p. 124, 28 January 2021.

[19] M. Schmelcher, D. M. Donovan e M. J. Loessner, "Bacteriophage endolysins as novel antimicrobials," *Future Microbiology,* vol. 7, nº 10, pp. 1147-1171, October 2012.

[20] D. Gutiérrez, L. Fernández, A. Rodríguez e P. Garcíaa, "Are Phage Lytic Proteins the Secret Weapon To Kill Staphylococcus aureus?," *mBio,* vol. 9, nº 1, pp. 9:e01923-17, 23 January 2018.

[21] A. Bateman e N. D. Rawlings, "The CHAP domain: a large family of amidases including GSP amidase and peptidoglycan hydrolases," *Trends in Biochemal Sciences,* vol. 28, nº 5, pp. 234-237, May 2003.

[22] S. S. Broendum, A. M. Buckle e S. McGowan, "Catalytic diversity and cell wall binding repeats in the phage-encoded endolysins," *Molecular Microbiology,* vol. 110, nº 6, pp. 879-896, 13 November 2018.

[23] E. Murray, L. A. Draper, R. P. Ross e C. Hill, "The Advantages and Challenges of Using Endolysins in a Clinical Setting," *Viruses,* vol. 13, nº 4, p. 680, April 2021.

[24] A. Waterhous, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. Heer, T. Beer, C. Rempfer, L. Bordoli, R. Lepore e T. Schwede, "SWISS-MODEL: homology modelling of protein structures and complexes.," *Nucleic Acids Research,* vol. 46, nº W1, pp. W296-W303, 2 July 2018.

[25] J. Lambert, E. Mazzola e C. Ridge, Nuclear Magnetic Resonance Spectroscopy: An Introduction to Principles, Applications, and Experimental Methods, 2nd Edition, Chennai, India: John Wiley & Sons Ltd, 2018.

[26] D. Marion, "An Introduction to Biological NMR Spectroscopy," *Molecular & Cellular Proteomics,* vol. 12, nº 11, pp. 3006-3025, November 2013.

[27] S. R. Bond and C. C. Naus, "RF-Cloning.org: an online tool for the design of restriction-free cloning projects," *Nucleic Acids Research,* vol. 40, no. W1, p. W209–W213, 8 May 2012.

[28] R. L. J. Keller, The Computer Aided Resonance Assignment Tutorial, Zürich: CANTINA Verlag, 2004.

[29] L. Y. Low, C. Yang, M. Perego, A. Osterman e R. C. Liddington, "Structure and lytic activity of a Bacillus anthracis prophage endolysin," *The Journal of Biological Chemistry,* vol. 280, nº 42, pp. 35433-35439, 21 October 2005.

[30] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak e A. Cardona, "Fiji: an open-source platform for biological-image analysis," *Nature Methods,* vol. 9, nº 7, p. 676–682, 28 June 2012.

[31] M. Jerabek-Willemsen, T. André, R. Wanner, H. M. Roth, S. Duhr, P. Baaske e D. Breitsprecher, "MicroScale Thermophoresis: Interaction analysis and beyond," *Journal of Molecular Structure,* vol. 1077, pp. 101-113, 5 December 2014.

[32] W. Uckert, L. Pedersen e W. Günzburg, "Green Fluorescent Protein Retroviral Vector: Generation of High-Titer Producer Cells and Virus Supernatant," em *Gene Therapy of Cancer*, Totowa, NJ, Human Press, 2000, pp. 275-285.

[33] G. Rule e K. Hitchens, "Resonance Assignments: Heteronuclear Methods," em *Fundamentals of Protein NMR Spectroscopy*, vol. 5, Dordrecht, Springer, 2006, pp. 277-312.

[34] C. Redfield, "Assignment of Protein NMR Spectra Using Heteronuclear NMR—A Tutorial," em *Protein NMR*, Boston, MA, Springer, 2015, pp. 1-42.

[35] J. Jumper, R. Evans e A. Pritzel, "Highly accurate protein structure prediction with AlphaFold," *Nature,* vol. 596, nº 7873, p. 583–589, July 2021.

[36] M. Varadi, S. Anyango e M. Deshpande, "AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models," *Nucleic Acids Research,* vol. 50, nº D1, p. D439–D444, November 2021.

[37] M. Benešík, J. Nováček, L. Janda, R. Dopitová, M. Pernisová, K. Melková, L. Tišáková, J. Doškař, L. Žídek, J. Hejátko e R. Pantůček, "Role of SH3b binding domain in a natural deletion mutant of Kayvirus endolysin LysF1 with a broad range of lytic activity," *Virus Genes,* vol. 54, nº 1, pp. 130-139, 2018.

[38] L. S. Gonzalez-Delgado, H. Walters-Morgan, B. Salamaga, A. J. Robertson, A. M. Hounslow, E. Jagielska, I. Sabała, M. P. Williamson, A. L. Lovering e S. Mesnage, "Two-site recognition of Staphylococcus aureus peptidoglycan by lysostaphin SH3b," *nature chemical biology,* vol. 16, nº 1, pp. 24-30, November 2019.

[39] J. Z. Lu, T. Fujiwara e H. Komatsuzawa, "Cell Wall-targeting Domain of Glycylglycine Endopeptidase Distinguishes among Peptidoglycan Cross-bridges," *Protein Structure and Folding,* vol. 281, nº 1, pp. 549-558, 2006.

[40] B. L. M. d. Jonge, Y.-S. Chang, D. Gage e A. Tomasz, "Peptidoglycan composition of a highly methicillin-resistant Staphylococcus aureus strain. The role of penicillin binding protein 2A," *The Journal of Biological Chemistry,* vol. 267, nº 16, pp. 11248-11254, 5 June 1992.

[41] M. d. C. d. F. Bastos e M. L. V. Coelho, "Staphylococcal Antimicrobial Peptides: Relevant Properties and Potential Biotechnological Applications," *Current Pharmaceutical Biotechnology,* vol. 10, nº 1, pp. 38-61, 2009.

[42] K. O. Lee, M. Kong, I. Kim, K.-S. Ryu, S. Ryu e J.-Y. Suh, "Structural Basis for Cell-Wall Recognition by Bacteriophage PBC5 Endolysin," *Structure,* vol. 27, nº 9, pp. 1355-1365, 2019.

[43] R. Maya-Martinez, J. A. N. Alexander, C. F. Otten, I. Ayala, D. Vollmer, J. Gray, C. Bougault, A. Burt, C. Laguri, M. Fonvielle, M. Arthur, N. Strynadka, W. Vollmer e J.-P. Simorre, "Recognition

of Peptidoglycan Fragments by the Transpeptidase PBP4 From Staphylococcus aureus," *Frontiers in Microbiology,* vol. 9, 2019.

[44] "Isotope Distribution Calculator and Mass Spec Plotter," Adaptas Solutions , [Online]. Available: https://www.sisweb.com/mstools/isotope.htm. [Acedido em 15 September 2022].

[45] T. H. Scheuermann, S. B. Padrick, K. H. Gardner e C. A. Brautigam, "On the acquisition and analysis of microscale thermophoresis data," *Analytical Biochemistry,* vol. 496, pp. 79-93, 1 March 2016.

[46] M. P. Williamson, "Chemical Shift Perturbation," em *Modern Magnetic Resonance*, Springer, Cham, 2018, pp. 995-1012.

[47] J. H. Carr, "Public Health Image Library (PHIL)," Centers for Disease Control and Prevention, 2001. [Online]. Available: https://phil.cdc.gov/details.aspx?pid=11153. [Acedido em 25 august 2022].

5

APPENDIX

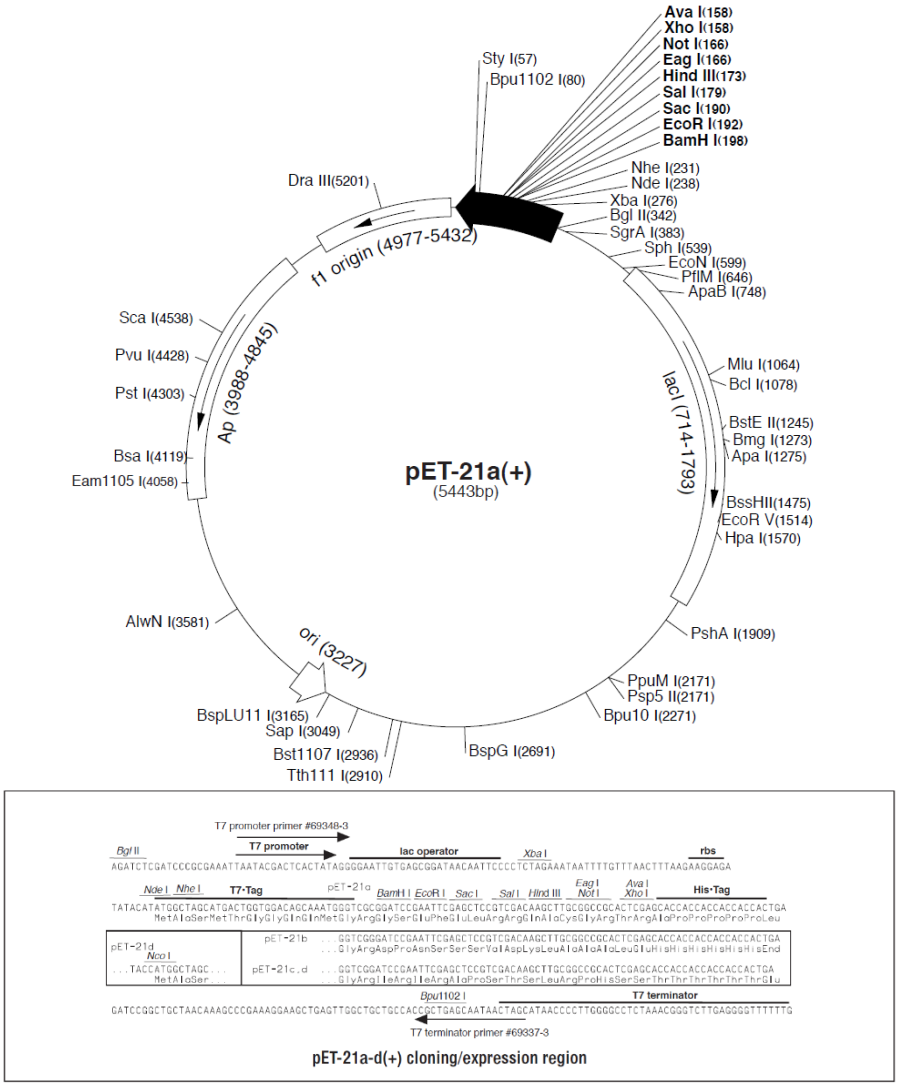
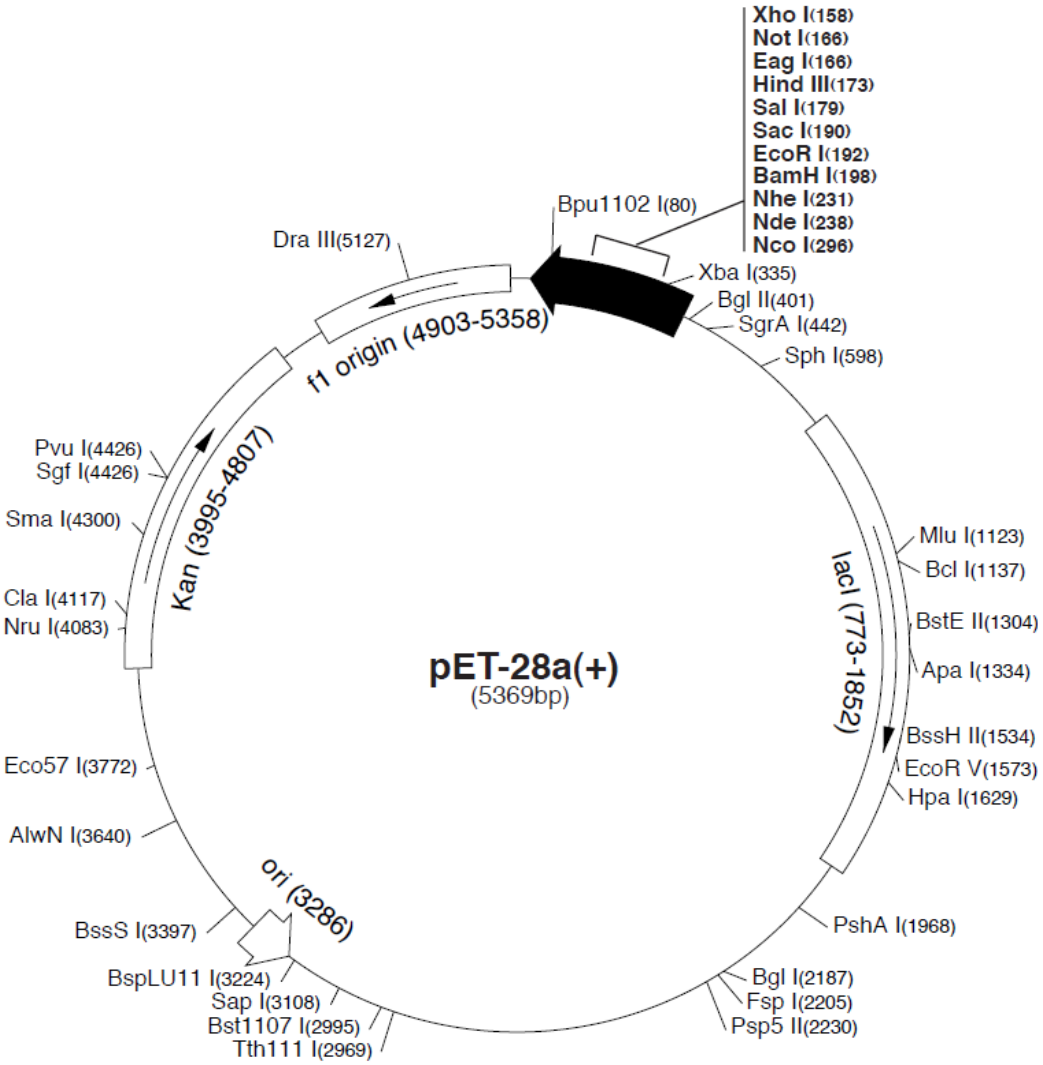


Figure 5.1. Map of the pET21 expression vector (Novagen) and its restriction sites for cloning and expression.



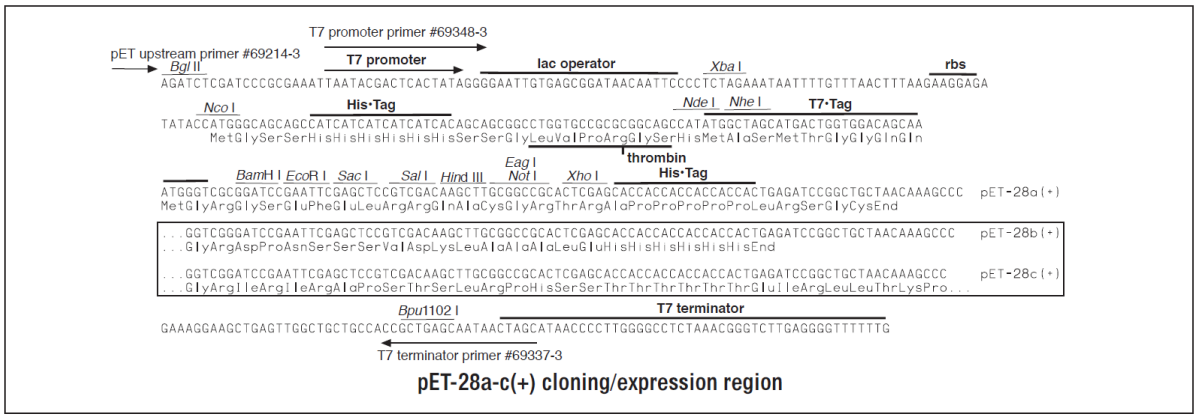


Figure 5.2. Map of the pET28 expression vector (Novagen) and its restriction sites for cloning and expression.

Table 5.1. Compounds to prepare an agarose gel 0.8%.

Compounds	**For 80 mL**
TAE buffer (mL)	80
Agarose (g)	0.64
GreenSafe Premium (NZYTech) (µL)	4

Table 5.2. Composition of LB and LA medium (1 L).

Compounds	**For 1 L**
Tryptone (g)	10
Yeast Extract (g)	5
NaCl (g)	10
Milli-Q water (mL)	Up to 1 L
Agar (g) – For LA medium	15

Table 5.3. Compounds to prepare M9 medium.

Compounds (M9 medium)	**For 250 mL**
H_2O (mL)	230
M9 salts (mL)	25
$^{15}NH_4Cl$ (g)	0.25
$MgSO_4$ 2 M (mL)	0.25
$CaCl_2$ (mL)	0.25
Glucose (g)	1
Thiamine-HCl 0.1% (mL)	0.125
$FeSO_4$ 0.1 M (mL)	0.25
Kanamycin 100 mg/mL (µL)	125

Table 5.4. Compounds to prepare a SDS-PAGE gel.

Compounds	**Separating gel (12%)**	**Stacking gel (4%)**
dH_2O (mL)	2.1	1.95
3 M Tris-HCl/SDS pH 8.45 (mL)	2.5	0.775
30% acrylamide (mL)	3	0.4
Glycerol (mL)	0.75	-
10% APS (µL)	21	21

TEMED (µL)	7	7

Table 5.5. Composition of the buffers for SDS-PAGE analysis (1 L).

Compounds	**Race buffer 10x (pH 8.3)**	**Anode buffer 10x (pH 8.8)**
Tris-base (g)	121.10	242
Tricine (g)	179.20	-
20% SDS (mL)	50	-
Milli-Q water (mL)	Up to 1 L	Up to 1 L

Table 5.6. Composition of sample buffer for SDS-PAGE analysis (50 mL).

Compounds	**Sample Buffer (4x)**
1 M Tris-HCl pH 6.8 (mL)	10
Glycerol (mL)	24
10% SDS (g)	8
DTT (g)	3.10
Comassie blue R250 (mg)	20
Milli-Q water (mL)	Up to 50 mL

Table 5.7. Compounds to buffers. The volume of water is added until the final volume is reached.

Compounds	**10 mM PO_4 pH=7.0** **100 mM NaCl** **For 500 mL**	**10 mM PO_4 pH=7.0** **100 mM NaCl** **6M Urea** **For 100 mL**	**10 mM PO_4 pH=7.0** **100 mM NaCl** **1M Imidazole** **For 200 mL**
PO_4 (mL)	5	1	2
NaCl (g)	2.92	0.58	1.17
Urea (mL)	-	36.04	-
Imidazole (mL)	-	-	13.62

Table 5.8. Compounds to prepare the minimal medium for PG extraction.

Metal Solution		
Compounds	**[Final Concentration]**	**For 1 L**
$FeCl_2$	7.5µM	1.4926 g (First $FeCl_2$ is dissolved in 1 M HCl in 5 mL. Then is added to the final solution, leaving the HCl at 5 mM.)
$ZnCl_2$	0.51 µM	0.0695g
$MnCl_2$	0.5 µM	0.0989g
Boric acid	0.097 µM	0.0060
$CoCl_2$	1.46 µM	0.348
$CuCl_2$	0.015 µM	0.5 M-30 µL ($CuSO_4$)
$NiCl_2$	0.1 µM	0.0236
Na_2MoO_4	0.148 µM	0.036
Mother Solution		
Compounds	**[Final Concentration]**	**For 1 L**
Na_2HPO_4	10 mM	1.779 g
KH_2PO_4	10 mM	1.36 g
NH_4Cl	9.34 mM	0.499 g
NaCl	8.55 mM	0.499 g
$MgSO_4$	0.81 mM	0.111 g
Citric acid	0.142 mM	0.029 g
Casamin + tryptophan (100mg/L)	1% Casamin + 100 mg/L tryptophan	10 g + 100 mg tryptophan
MOPS	40 mM	40 mL/g

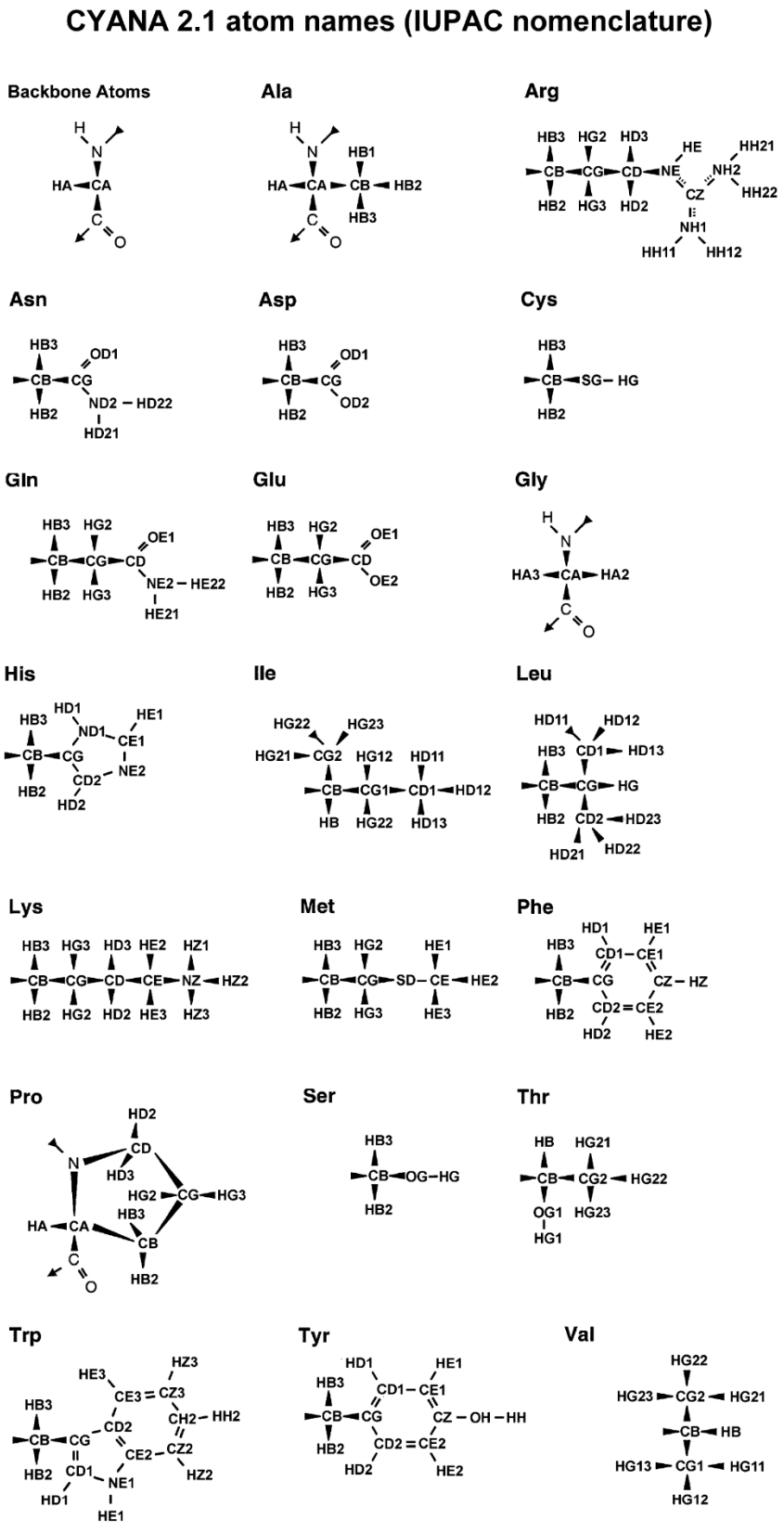


Figure 5.3. CYANA nomenclature of amino acid residues.

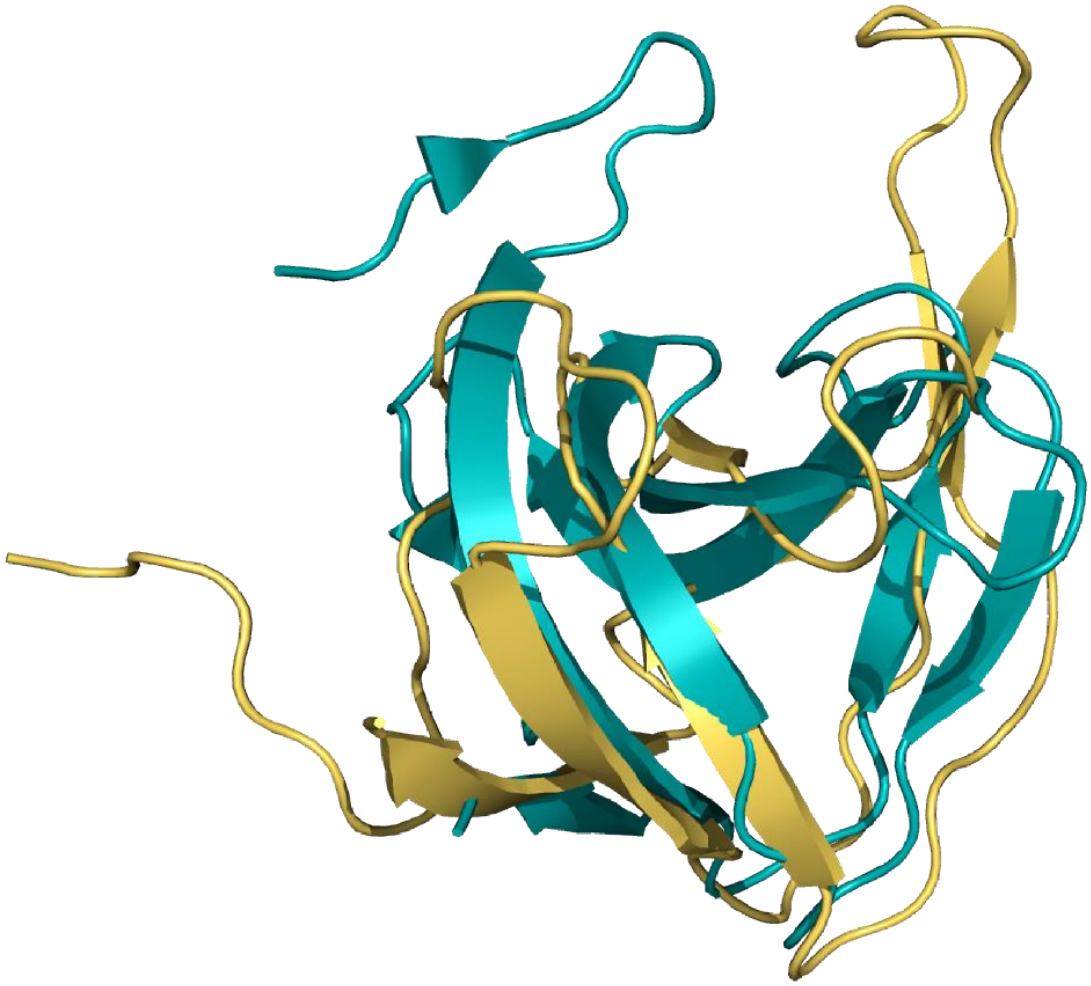


Figure 5.4. The superimpose of the preliminary 3D structure of CBDMW2 and the Lysostaphin structure, show similar SH3 fold with different loop lengths.

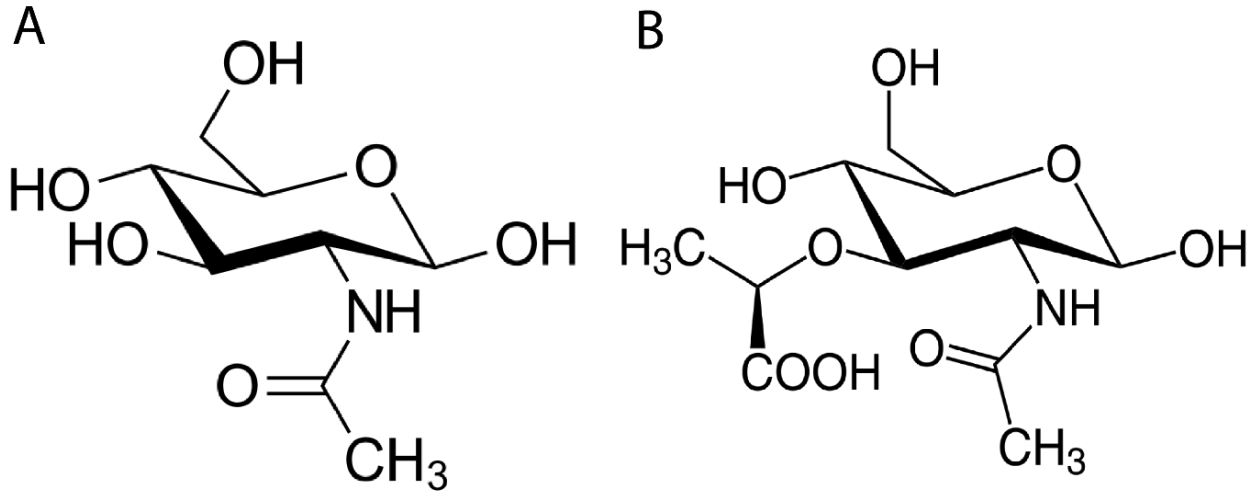


Figure 5.5. Structure of the two sugars that constitutes muropeptide. (**A**) GlcNAc. (**B**) MurNAc.

Table 5.9. Chemical shifts of the stem peptide and D-Lac from the muropeptide.

Tetrapéptido+D-Lac			
Amino acid /Protons	**ppm**	**Amino acid /Protons**	**ppm**
L-Alanine x		**L-Lysine**	
H_N	8.531	**H_N**	8.353
HA	4.248	**HA**	4.195
HB	1.434	**HB**	1.723
N_H	124.50	**QD**	1.5105
CA	50	**HG3**	1.3189
CB	16.14	**HG2**	1.3643
D-iso-Glutamine		**QE**	3.206
H_N	8.627	**H_z**	8.005/8.05/8.14
HA	4.265	**N_H**	126.85
QG	2.371	**N_z**	120
HB2	2.146	**CA**	54.21
HB3	1.934	**CB**	30.31
N_H	119.92	**CG**	22.06
N_{H2}	107.64	**CD**	27.76
CA	54.22	**CE**	39.02
CB	26.65	**D-Alanine**	
CG	31.25	**H_N**	8.653
D-Lac		**HA**	4.354
H(CH)	4.23	**HB**	1.4
C(CH)	78.28	**N_H**	125.66
H(CH3)	1.38	**CA**	49.54
C(CH3)	18.36	**CB**	16.33

Table 5.10. Chemical shifts of the glycines from the muropeptide.

Glycines			
Amino acid /Protons	**ppm**	**Amino acid /Protons**	**ppm**
Glycine A/1		**Glycine B/5**	

H_N	8.463	**H_N**	8.27
HA	3.88	**HA**	3.895
CA	42.177	**CA**	40.478
N_H	109.151	**N_H**	110.883
Glycine C x		**Glycine D x**	
H_N	8.6	**H_N**	8.7
HA	4	**HA**	4.056
CA	42.46	**CA**	42.26
N_H	109	**N_H**	108.39

Table 5.11. Chemical shifts of sugars: GlcNAc and MurNAc

Sugar /Protons	ppm	Sugar /Protons	ppm
GlcNAc		**MurNAc**	
H_NAc	8.4733	**H_NAc**	8.2018
HAc	2.06	**HAc**	1.9833
H1	4.65	**H1**	3.634/3.686
H2	3.786	**H2**	4.45
H3	3.552	**H3**	3.895
H4	3.499	**H4**	3.88
H5	3.426	**H5**	3.83
H6	3.79/3.917	**H6**	3.61/3.81
N_HAc	123.03	**N_HAc**	125.04
CAc	22	**CAc**	21.994
C1	101	**C1**	61.18
C2	55.61	**C2**	52.17
C3	73.8	**C3**	76.76
C4	69.34	**C4**	77.96
C5	75.50	**C5**	70.47
C6	60.35	**C6**	61.78