UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA VEGETAL

Heteromeric endolysins encoded in a single gene: occurrence in Gram-positive targeting phages and relevance for lytic activity

Raquel Bastos Gonçalo

Mestrado em Biologia Molecular e Genética

Dissertação orientada por:
Prof. Doutor Carlos São-José
Prof. Doutora Rita Zilhão

2021

Acknowledgements

I would like to express my gratitude to all the people directly or indirectly involved in this thesis.

Firstly, I would like to thank Professor Carlos São-José for the opportunity to be a part of his team and to develop this project. Also, for all the support, guidance and discussions that led to the conclusion of this work. I would also like to thank Professor Rita Zilhão for the support and relevant feedback in this work.

A very special thanks to Daniela Pinto, for all the support and patience to teach me, for the motivation that eventually led to the conclusion of this project, and for the revisions and comments during the writing process of the thesis.

I would also like to thank one more special member during my time in the lab for all the patience and guidance, Ana Gouveia.

A huge thank you for the support of my family and friends which allowed the conclusion of this work and for making life so much better.

Resumo

As infeções bacterianas foram durante muitos séculos uma das principais causas de morte a nível mundial, não só pelas más condições de vida existentes, mas também pela inabilidade de combater e tratar estas infeções. Portanto, a descoberta de uma ferramenta para combater as infeções bacterianas, os antibióticos, foi de uma importância sem precedentes, uma vez que permitiu aumentar o tempo médio de vida e melhorar a qualidade de vida do Homem. Os antibióticos desde que foram descobertos têm sido amplamente utilizados para tratar infeções bacterianas muito devido à sua elevada eficácia, baixo custo de produção e reduzida toxicidade. Funcionam como uma arma que na maior parte dos casos mata de forma indiscriminada as bactérias (patogénicas ou não), geralmente sem ser necessário identificar a bactéria patogénica que está a causar a infeção, o que por um lado acelera e facilita o processo de tratamento, mas por outro pode ter efeitos colaterais indesejados ao atacar a microbiota humana. Para além da sua utilização na vertente clínica, os antibióticos passaram depois também a ser utilizados em diversas outras áreas como a indústria e a agricultura, nomeadamente na produção animal para consumo humano, de forma a controlar a produtividade e o surgimento de bactérias nestes ambientes, contribuindo assim a segurança alimentar.

No entanto, com o passar do tempo, muito devido à facilidade de acesso aos antibióticos a baixo custo e ao desconhecimento do perigo associado à recorrente utilização destes, o uso de antibióticos foi-se tornando cada vez mais indiscriminado, generalizado e excessivo o que criou uma das grandes ameaças que a nossa sociedade enfrenta atualmente, o aumento da resistência aos antibióticos por parte de algumas bactérias. Este problema nos dias de hoje é já considerado como responsável pela morte de milhares de pessoas em todo o mundo, e se nada for feito para reverter esse cenário, este número continuará a aumentar para muitos milhões, o que tem estimulado e apressado a pesquisa de alternativas. Uma das possíveis alternativas para combater esta ameaça é a utilização de endolisinas.

As endolisinas são enzimas produzidas pelos bacteriófagos (ou fagos), os vírus que infetam bactérias. Os fagos são parasitas intracelulares obrigatórios que, durante a infecção da bactéria hospedeira, expressam o seu genoma e formam novas partículas virais usando a maquinaria celular da célula hospedeira. No fim do seu ciclo replicativo, os fagos utilizam as endolisinas para promover a lise da célula hospedeira e assim permitir a libertação da progenia viral. Isto acontece porque as endolisinas degradam o principal constituinte da parede celular bacteriana, o peptidoglicano. Esta degradação, especialmente em bactérias Gram-positivas, onde a parede celular é o principal componente estrutural do invólucro celular, resulta na lise osmótica e leva à morte da célula bacteriana. A possibilidade de utilizar as endolisinas como possível alternativa ao uso de antibióticos reside neste seu potencial de degradar a parede celular bacteriana, o qual também se verifica quando estas enzimas (enzibióticos) são adicionadas exogenamente às bactérias, particularmente quando adicionadas a espécies Gram-positivas. Para além de conseguirem atuar sobre bactérias resistentes aos antibióticos, a grande maioria das endolisinas estão evolutivamente desenhadas para aturarem especificamente no peptidoglicano de um género ou espécie bacteriana, de acordo com o fago de origem, protegendo assim as bactérias benéficas da microbiota humana. Por outro lado, os estudos disponíveis indicam que as bactérias apresentam uma capacidade muito reduzida de desenvolver resistência às endolisinas quando comparadas com os antibióticos, o que é outra das vantagem do uso das endolisinas.

As endolisinas que têm como alvo bactérias Gram-positivas têm normalmente uma arquitetura modular, sendo tipicamente compostas por uma região N-terminal com um ou dois domínios catalíticos (que degradam o peptidoglicano), e por um módulo C-terminal com um ou mais motivos de ligação à parede celular bacteriana. Embora a maior parte das enzimas descritas sejam monoméricas, também já foram descritas endolisinas heteroméricas compostas por duas subunidades diferentes. Entre as enzimas heteroméricas que foram descobertas mais recentemente, verificou-se que as duas subunidades eram produzidas a partir de um único gene, isto graças à presença de um local interno de iniciação da tradução (composto por um local de ligação do ribossoma seguido de um codão de iniciação). Para estas endolisinas, o gene contém duas grelhas de leitura que estão *in frame*, têm o mesmo codão de terminação e levam à produção de dois polipéptidos distintos. Nestes casos ainda com poucos exemplos e, portanto, considerados até agora pouco representativos, as endolisinas são constituídas normalmente por uma subunidade correspondente ao polipéptido produzido a partir da leitura total do gene, e três subunidades produzidas a partir da grelha de leitura iniciada no local interno de iniciação da tradução (porção 3' da grelha de leitura total do gene).

A presença deste local interno de iniciação da tradução parece, nos exemplos já descritos, fornecer um mecanismo para otimizar a afinidade da enzima à parede celular bacteriana, pois há um aumento dos motivos de ligação à parede bacteriana. Esta otimização, parece refletir-se depois na melhoria da atividade lítica e por isso a compreensão deste fenómeno parece tão importante e interessante principalmente no contexto atual. Após uma análise bioinformática, que foi capaz de identificar diversos genes de endolisinas com um putativo local interno de iniciação da tradução, produzidas por fagos que infetam bactérias Gram-positivas, escolhemos algumas destas para confirmar experimentalmente esta previsão. Nesta escolha tivemos em conta a composição das endolisinas quanto ao número e tipo de domínios conservados, de modo a aumentar a diversidade da nossa amostra. Para além disso, tivemos também em conta se as enzimas eram produzidas por fagos que infetam bactérias de géneros diferentes dos exemplos já descritos com enzimas heteroméricas produzidas através deste mecanismo, e se as bactérias infetadas apresentavam relevância para o Homem (quer em ambiente clínico quer em ambiente industrial).

Após a escolha das endolisinas, investigou-se então se efetivamente ocorria a produção de dois polipéptidos codificados a partir do mesmo gene, sendo o polipéptido menor produzido a partir da grelha de leitura iniciada no local interno de iniciação da tradução. Investigou-se também se estes dois polipéptidos interagiam formando endolisinas heteroméricas e que impacto isso teria na sua atividade lítica contra as bactérias alvo. Verificámos que nas endolisinas escolhidas e testadas dois polipéptidos são, de facto, produzidos a partir de um mesmo gene, sendo o polipéptido menor produzido a partir da grelha de leitura iniciada no local interno de iniciação da tradução. Estes resultados sugerem que os fagos que levam à produção de endolisinas potencialmente heteroméricas a partir deste mecanismo podem ser mais frequentes do que inicialmente se pensava.

Contudo, numa das endolisinas testadas (LysJavan488), ao contrário do esperado, demonstramos que, pelo menos nas condições testadas, os dois polipéptidos produzidos não parecem interagir de forma a gerar uma estrutura heteromérica. Foi possível demonstrar também, no caso de outra das endolisinas testadas (LysPhi7951), que é necessária a associação entre os dois polipéptidos codificados no mesmo gene para que a enzima exiba a sua máxima atividade lítica. A análise bioquímica indicou ainda que este heteromultímero deverá ser composto por uma subunidade do polipéptido produzido a partir da leitura total do gene associada a entre quatro a seis subunidades correspondentes ao polipéptido produzido a partir da grelha de leitura iniciada

no local interno de iniciação da tradução, o que configura uma estrutura heterométrica nunca antes descrita para uma endolisina.

Palavras chave: Bacteriófagos; Endolisinas; Heteromérica; Local interno de iniciação da tradução; Enzimas líticas

Abstract

Antibiotic resistance is one of the major threats of our time, which has been driving research on antibacterial alternatives. Among these are endolysins, enzymes that digest the peptidoglycan, the major and essential constituent of the bacterial cell wall. Endolysins are produced by bacteriophages, the viruses that infect bacteria, which use them to lyse host cells at the end of the viral replicative cycle for virion progeny release. Those that target Gram-positive bacteria have a typical modular architecture, with one or more catalytic domains responsible for peptidoglycan cleavage in the N-terminal region, and a C-terminal module carrying one or more motifs involved in cell wall binding. Although generally monomeric, two-subunit heteromeric endolysins have already been described. For the recently discovered examples, it was found that the two subunits are produced from a single gene, thanks to the presence of an internal translation start site (ITSS) that allows translation of a smaller reading frame into a C-terminal product of the endolysin. In these cases, the heteromeric endolysins are composed typically by one subunit of the full-length polypeptide, and three copies of the smaller polypeptide.

Based on a handful of examples of endolysins with diverse domain architectures that carried putative ITSS, and which are produced by phages that infect clinically or industrially relevant bacteria, we have demonstrated that two different polypeptides are indeed produced from a single gene. These results suggest that this mechanism may be more frequent than anticipated. For one of the studied endolysins, we have also shown that the association of the two polypeptides, with one as a multimer, is necessary for full enzymatic activity. The biochemical analysis also showed that this heteromultimeric endolysin should be composed by one subunit of the full-length polypeptide and between four to six of the smaller polypeptides, indicating a structure never described before.

Key words: Bacteriophages; Endolysins; Heteromeric; Internal translation start site; Lytic enzymes

Content

Figure index

Table index

Abbreviations

AFC – affinity chromatography

AGE – agarose gel electrophoresis

BCE – bacterial cell envelope

CAP– covalently attached protein

CDs – catalytic domains

CM – cytoplasmic membrane

CTP – C-terminal product

CW – cell wall

CWBD – cell wall binding domain

EPSs – exopolysaccharides

FLP – full-length polypeptide

His_6 – hexahistidine

His_6 -Lys – endolysins with N-terminal fused to a hexahistidine tag

I – T_1 extract insoluble fraction

IMP – integral membrane protein

IPTG – isopropylβ-D-thiogalactopyranoside

ITSS – internal translation start site

K_{av} – protein distribution coefficient

LB – Luria Bertani

LP – lipoprotein

LPS – lipopolysaccharide

LTA – lipoteichoic acid

Lys-His_6 – endolysins with C-terminal fused to a hexahistidine tag

m-DAP – meso-diaminopimelic acid

NAG – N-acetylglucosamine

NAM – N-acetylmuramic acid

OD600 – optical density at 600nm

OM – outer membrane

OMP – outer membrane protein

ORF – open reading frame

PBS – phosphate-buffered saline

PCR – polymerase chain reaction

PMF – proton-motive force

Rs – stokes radii

S – T_1 extract soluble fraction

SDS-PAGE – sodium dodecyl sulfate–polyacrylamide gel electrophoresis

SEC – size exclusion chromatography

T_0 – total extract before induction of protein production

T_1 – total extract after induction of protein production

TCEP – tris(2-carboxyethyl)phosphine hydrochloride

V_e – elution volume

V_0 – column void volume

V_t – column total volume

WHO – World Health Organization

WTA – wall teichoic acid

X-Gal – 5-bromo-4-chloro-3-indolyl-β-d-galactopyranoside

1. Introduction

1.1 Background

Antibiotics are defined as organic or synthetic compounds that inhibit or kill bacteria by interfering with essential cellular processes (Mohr, 2016). Since the landmark discovery of penicillin in 1928, antibiotics have been used to treat bacterial infection diseases (Davies and Davies, 2010; Mohr, 2016; Singh *et al.*, 2017). This has increased the quality and expectancy of life in about 30 years in developed countries (Laxminarayan *et al*., 2016).

Over the years there has been an increase in bacterial resistance to antibiotics especially due to overuse in clinical, agriculture and industrial environments, leading to accelerated evolution and expanded repertoire of antibiotic resistance genes (Aminov, 2009; Ventola, 2015). Bacterial resistance to antibiotics can occur through different mechanisms that can include: (1) reduction of drug entry; (2) inactivation of the drug; (3) expression of efflux pumps to eliminate drugs from inside bacterial cells; and (4) modification of the target, altering the drug-binding affinity. Sometimes bacteria are even resistant to more than one class of antibiotics simultaneously, diminishing treatment options (Singh *et al.*, 2017). The World Health Organization (WHO) and other organizations recognize antibiotic resistance as a major threat to human health and global economy. Nowadays, at a global scale, it is estimated that at least 700 000 patients die per year due to antibiotic resistance and that can increase to 10 million by 2050 if development of antimicrobial resistance is not controlled (Lesho and Laguio-Vila, 2019). This problem has been motivating research on alternative antimicrobials.

Phage therapy is a possible alternative that started to be applied in humans in 1919 by Felix d'Hérelle. It involves the targeted application of bacteriophages (phages), the viruses that infect bacteria, to lyse the bacterial pathogen that is causing the infection (Abedon *et al*., 2011; Gordillo Altamirano and Barr, 2019). Phages are highly diverse intracellular parasites that infect bacteria through the recognition and binding to the host cell surface. Without compromising cell viability, they then hijack the metabolism of the host cell, expressing their genome (that can go from 17 kbp to 0.5 Mbp) and multiplying the number of virus particles (Carter and Saunders, 2013; Oliveira *et al*., 2013; Poranen *et al*., 2002). At the end of the replication cycle, phages disrupt the bacterial cell envelope (BCE) causing cell lysis and the consequent release of progeny virions (Carter and Saunders, 2013; Fernandes and São-José, 2018; Young *et al*., 2014).

The poor knowledge on phage biology, and the logistical and technical obstacles in developing phage therapy led to its widespread abandonment after the discovery of antibiotics (Lin *et al*., 2017). Only a few countries (e.g., Georgia and Poland) kept researching and practicing (Górski *et al*., 2018). Nowadays, because it only recently regained attention of the western medicine, it still exhibits safety concerns because: (1) the success of phage therapy is distrusted and considered poorly documented (Górski *et al*., 2018); and (2) is a live-biological agent that can potentially interact with the immune system, can replicate, and evolve during manufacture or use (Loc-Carrillo and Abedon, 2011).

Endolysins, however, which are key enzymes in phage-mediated bacteriolysis, may at least partially overcome these problems, being their use currently explored as alternative antimicrobials.

1.2 Endolysins

Endolysins are enzymes produced by phages to disrupt the major constituent of the bacterial cell wall (CW), the peptidoglycan, at the end of the replication cycle. With this essential cell structure compromised, bacteria tend to quickly undergo osmotic cell lysis (Fernandes and São-José, 2018; Young, 2014). These enzymes act specifically on peptidoglycan of certain bacterial species or genus (Fernandes and São-José, 2018), which makes them perfectly suited to destroy a target pathogen without affecting commensal bacteria (Schmelcher, Donovan *et al.*, 2012). Besides endolysin's clinical applications, they can also be used in diagnostics and in food products from agriculture to packaging (de Melo *et al.*, 2018; Schmelcher, Donovan *et al.*, 2012). Together, all these characteristics have been the basis for the intense exploration of endolysins as enzybiotics (São-José, 2018). However, the endolysins, especially their structure and regulation, are still poorly understood. More knowledge on endolysins could lead to the development of versions with improved stability, activity, or host range and thus improved applicability (Dunne *et al*., 2016).

1.2.1 The bacterial cell wall

The peptidoglycan CW, structure targeted by the endolysins, is a conserved component of the multilayered BCE. The BCE protects cell structure from environment insults, determines its shape, ensures its homeostasis and allows influx and efflux of products to sustain growth (Fernandes and São-José, 2018; Silhavy *et al*., 2010). The BCE, in Gram-positive bacteria, contains the cytoplasmic membrane (CM), mainly composed of phospholipids, and the CW. Gram-negative bacteria and mycobacteria have a more complex BCE with a lipidic membrane surrounding the CW, designated outer membrane (OM). In Gram-negative bacteria the OM functions as an additional barrier controlling influx and efflux of compounds (Hoffmann *et al*., 2008; Silhavy *et al*., 2010) (Fig. 1.1).

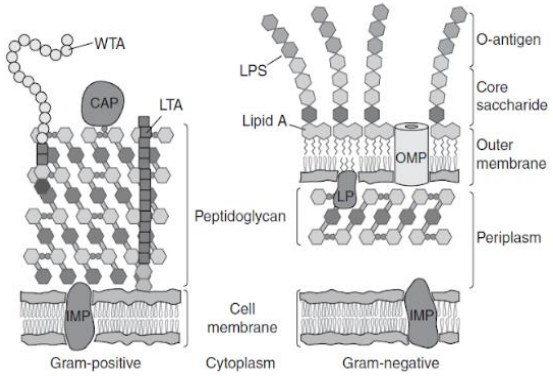


Figure 1.1 - Schematic representation of the bacterial cell envelope (BCE) of Gram-positive and Gram-negative bacteria. CAP, covalently attached protein; IMP, integral membrane protein; LP, lipoprotein; LPS, lipopolysaccharide; LTA, lipoteichoic acid; OMP, outer membrane protein; WTA, wall teichoic acid. (Fig. from Silhavy *et al*., 2010).

The peptidoglycan macromolecule forms a sacculus that surrounds the bacterial CM and confers the necessary mechanical resistance to avoid cell lysis as a result of turgor pressure (Vollmer *et al.*, 2008). Therefore, uncontrolled breakdown of the structure, as done by the endolysins, can result in osmotic cell lysis (São-José, 2018). Each peptidoglycan layer is a linear glycan chain, composed, as shown in Figure 1.2, of repeated units of a disaccharide made of N-acetylglucosamine (NAG) and N-acetylmuramic acid (NAM), linked by β-1,4 glycosidic bonds between the C1 of NAM and the C4 of NAG. Adjacent glycan strands are cross-linked by short peptides that are attached to NAM via amide

bonds. For most bacteria, these cross-links involve the amino acid residues at positions 3 (often L-Lys or meso-diaminopimelic acid (m-DAP)) and 4 (D-Ala) of complementary peptide chains. The peptide stems may be linked by a direct interpeptide bond (in a few Gram-positive species) or via an interpeptide bridge (in most Gram-positive bacteria). The different types of bacterial peptidoglycan are mainly defined by variations within the amino acidic composition of these interpeptide bridges (White *et al.*, 2012; Schleifer and Kandler, 1972; Vollmer *et al.*, 2008). In Gram-positive bacteria the multiple layers of peptidoglycan can make the CW much thicker than that of Gram-negative bacteria (Poranen *et al*., 2002).

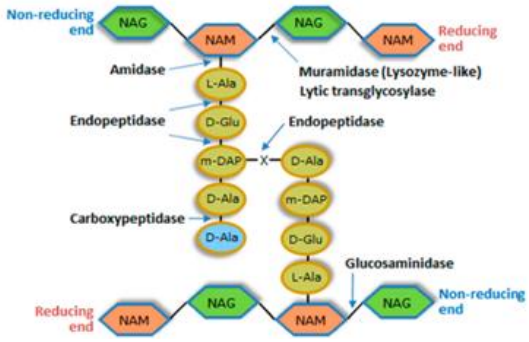


Figure 1.2 - Basic structure of the bacterial CW peptidoglycan with NAG and NAM linked by glycosidic bonds. The (X) indicates the cross-linking between m-DAP and D-Ala of a neighbour peptide chain. The D-Ala residue in blue may be lost after peptidoglycan maturation. The arrows indicate where the different classes of endolysins cleave the peptidoglycan. (Fig. from Fernandes and São-José, 2018).

1.2.2 Endolysin classification

Endolysins can be classified according to the peptidoglycan bond they target, with the three major classes being glycosidases, amidases and endopeptidases. The glycosidases cleave one of the two glycosidic bonds in the glycan chain and can be subdivided into glucosaminidases (N-acetyl-β-D-glucosaminidases), lytic transglycosylases, muramidases or lysozymes (N-acetyl-β-D-muramidases). The amidases (N-acetylmuramoyl-L-alanine amidases) cleave the amide bond connecting NAM to the first amino acid residue of the peptide stem. Finally, the endopeptidases cleave within or between the peptide strands (Fig. 1.2) (Nelson *et al*., 2012; São-José, 2018).

Endolysins can also be distinguished regarding their dependence on the holin to reach the CW. The holin is another phage-encoded protein that has a key role in lysis. At the appropriate time, holins form lethal holes in the CM that cause the collapse of the CM proton-motive force (PMF) (Young *et al*., 2014). The so-called canonical endolysins require the holes formed by the holins to cross the CM and access the CW, while the exported endolysins use host cell transport machinery, like de Sec system, to access the CW. However, both canonical and exported endolysins seem to have their activity enhanced by the PMF-collapsing action of the holin (Fernandes and São-José, 2016; Fernandes and São-José, 2018) (Fig. 1.3).

In the case of Gram-positive bacteria, endolysins when added exogenously can have the capacity to induce bacterial cell lysis due to the lack of an OM, which in Gram-negative and Mycobacteria hinders access of the lytic enzymes to the CW (São-José, 2018; Schmelcher, Donovan *et al.*, 2012) (Fig. 1.3).

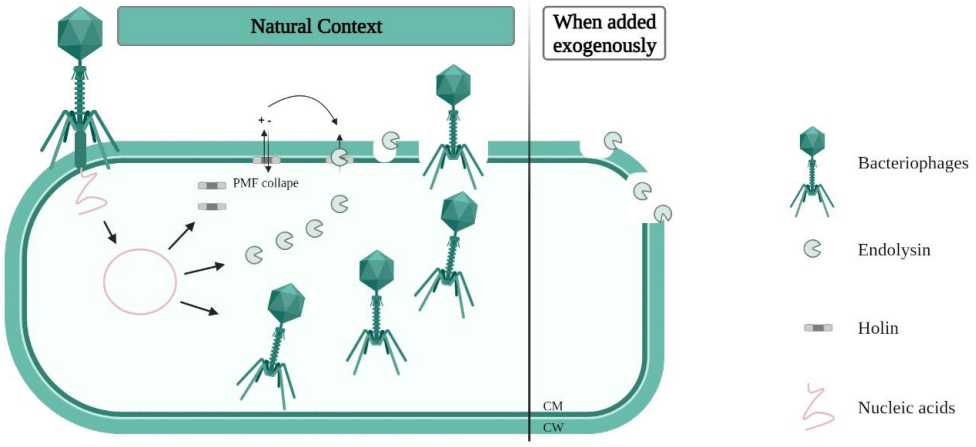


Figure 1.3 - Mode of action of endolysins in Gram-positive bacteria in their natural context (requiring the action of the holin) and when added exogenously.

1.2.3. Endolysin domain architecture

Endolysins targeting Gram-positive bacteria have a typical modular architecture, in which one or more N-terminal catalytic domains (CDs) responsible for peptidoglycan cleavage are connected by a flexible linker to a C-terminal region involved in cell wall binding (CWBD). There are some examples that deviate from this most common architecture, namely regarding the number and relative position of CDs and CWBDs, as shown in Figure 1.4 (São-José, 2018).

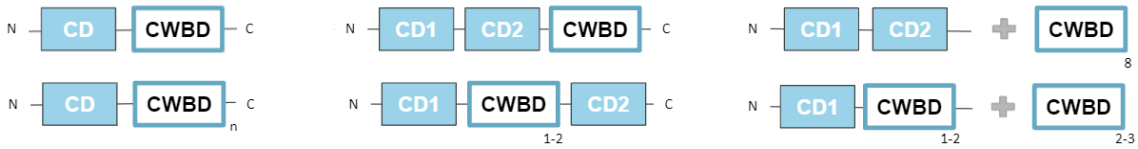


Figure 1.4 - Domain architectures of endolysins from Gram-positive systems. The "n" letter indicates that a variable number of CW binding motifs may compose the CWBD (2 to 7 copies) connected to the CD. The "+" sign denotes few cases of known heteromeric enzymes, which result from the association of the indicated subunits. (Fig. adapted from São-José, 2018).

Although endolysins are generally regarded as monomeric proteins, a few cases of multimeric endolysins have been described (rightmost architectures in Fig. 1.4):

(1) In one case (PlyC), two different polypeptides were shown to associate and compose the endolysin. PlyC is specific for streptococcal species and is composed of two subunits coded by separate genes. The *plyCA* gene codes a dual-CD polypeptide, and the *plyCB* gene codes the enzyme's CWBD. The active protein is a 1:8 heteromultimer of PlyCA:PlyCB (McGowan *et al*., 2012).

(2) In other cases, the different polypeptides composing the heteromultimer are encoded by the same gene. The endolysin gene has two in frame overlapping reading frames (ORF) that share a stop codon. This arrangement requires an internal ribosome binding site and an internal translation start site (ITSS) to allow the translation of the smaller reading frame, which corresponds to the 3' portion of the full-length reading frame.

a. Lys170, of the enterococcal phage F170/08, was the first heteromeric endolysin described to be encoded in a single gene. Based on biochemical studies, the active endolysin was proposed to be

composed by one subunit corresponding to the full-length polypeptide (FLP), as deduced from the *lys170* gene, and three copies of a C-terminal product (CTP) corresponding to the enzyme's CWBD (Proença *et al*., 2015), a structure that meanwhile gained further experimental support (Xu *et al*., 2021). The CWBD was found to be independently produced from an *in frame*, alternative ITSS located inside the *lys170* gene (Proença *et al*., 2015).

b. The discovery of ITSSs with similar role in heteromerization were in the meantime described for four other endolysins targeting *Clostridioides difficile* (Dunne *et al*., 2014), *Clostridium tyrobutyricum* (Dunne *et al*., 2016), *Clostridium sporogenes* (Dunne *et al*., 2016) and *Enterococcus faecalis* (Zhou *et al*., 2020). In the last case, 3D structure determination revealed an enterococcal endolysin sharing the 1:3 oligomeric state as Lys170.

In the described cases of heteromeric enzymes coded by single genes, the presence of the ITSS seems to provide a mechanism to increase the number of CWBDs of the endolysin and thus contribute to the enzymes' affinity, efficiency and activity (Proença *et al*., 2015).

1.3 Aims

A bioinformatic analysis led to the identification of several endolysin genes with putative ITSS. These endolysins were identified in a wide variety of phages that infect Gram-positive bacterial species. Hence, a few examples of clinical and industrial relevance were selected with the aim of testing these predictions. To achieve that, the following objectives were set: (1) provide experimental evidence that two different polypeptides are produced from a single selected endolysin gene with putative ITSS; (2) determine if, as in the majority of the described cases, the two polypeptides interact to form a heteromeric architecture; (3) investigate the impact of this interaction in the endolysins' activity.

2. Materials and Methods

2.1. Bacteria, plasmids and growth conditions

Escherichia coli strain XL1-Blue MRF' (Stratagene, USA), used for plasmid isolation and propagation, was grown at 37°C in Luria Bertani (LB) medium with aeration. The *E. coli* expression strain CG61 (São-José *et al.*, 2000), corresponding to strain BL21 (Stratagene, USA) harbouring the plasmid pGP1-2, its derivatives were grown in LB at 28°C with aeration and used for protein production. *Streptococcus pyogenes* strains M1 GAS and M6 (ST382) were grown at 37°C with 5% of carbon dioxide in brain heart infusion medium, whereas *Streptococcus thermophilus* strain 4078 was grown at 42°C in M17 medium supplemented with 1.5% lactose under anaerobiosis (anaerobic jar with AnaeroGen sachets, Thermo Scientific, USA). *S. pyogenes* and *S. thermophilus* strains were provided by the labs of Professors Mário Ramirez (IMM, FMULisboa, Portugal) and Douwe van Sinderen (University College Cork, Cork, Ireland), respectively.

Plasmid pGP1-2 carries genes coding: (1) resistance to kanamycin, (2) the phage T7 RNA polymerase under the control of a phage lambda promoter, and (3) the cI857 thermosensitive variant of the lambda CI repressor under the control of a pLac promoter. When the temperature is lower than 30°C, the repressor cI857 is functional and inhibits the lambda promoter, shutting down the expression of the T7 RNA polymerase. When the temperature is higher than 42°C, the protein cI857 is inactivated, so there is production of the T7 RNA polymerase. This will drive the transcription of the gene of interest carried in the expression plasmid under the control of a T7 promoter.

The expression vectors pIVEX2.3d and pIVEX2.4d (Roche Applied Science, Germany), which are selected in the presence of ampicillin, were used for endolysin overproduction in *E. coli*. In these vectors, the expression of cloned genes is under the control of the phage T7 φ10 promoter. Using pIVEX2.3d allowed the production of the endolysins with their C-terminus fused to a hexahistidine tag (Lys-His_6), whereas pIVEX2.4d was used for production of the endolysins N-terminally fused to the His_6 tag (His_6-Lys).

In the case of the endolysin LysbIL67, its gene was also used for cloning in the expression vector pET21a$^+$ (EMD/Novagen, USA) and in the cloning/expression vectors pBluescript II KS$^+$ and pBluescript II SK$^+$ (both from Stratagene, USA). As for pIVEX2.3d, vector pET21a$^+$ allows the production of the endolysins C-terminally fused to His_6 under the control of the phage T7 promoter. These three plasmids confer resistance to ampicillin.

When required, ampicillin (100 μg/mL) and/or kanamycin (40 μg/mL) were added to the culture medium for plasmid selection.

2.2 General DNA techniques

A survey and bioinformatics analysis of more than 3000 endolysin genes recovered from the Phage Lytic Proteins database (PhaLP, https://www.phalp.org/) allowed the identification of putative in frame ITSS in many endolysins targeting Gram-positive bacteria (D. Pinto, unpublished results). From these, several candidates were chosen to be used in this work based on following criteria: (1) diverse endolysin domain architecture, (2) preference for endolysins targeting bacteria different from those

targeted by the heteromeric endolysins previously described, (3) availability of the corresponding phage and host bacteria, and (4) potential relevance of target bacteria to human health and industry (Fig. S5.1). The selected endolysins genes were amplified with the DNA polymerase NZYProof 2XColourless Master Mix (NZYTech, Portugal) by polymerase chain reaction (PCR), using phage DNA as template and specific primers (Table S5.1).

Purification of the PCR amplification products was performed with the commercial kit GeneJET PCR Purification Kit (Thermo Scientific™, USA) following the manufacturer's instructions. The phage DNA used in the amplification of LysJavan488 was present in *S. pyogenes* strain M1. The endolysin LW32 gene was not amplified but instead commercially synthetized (NZYTech, Portugal). The endolysin genes were cloned into the expression vectors pIVEX2.3d or pIVEX2.4d, using the restriction endonucleases *Nco*I and *Cfr*9I (*Xma*I), and T4 DNA ligase from Invitrogen ™ (Thermo Scientific, USA). The *lysbIL67* was also cloned into the vector pET21a$^+$, pBluescript II KS$^+$ and pBluescript II SK$^+$ using T4 DNA ligase and the restriction endonucleases *Spe*I (for the vectors pBluescript) or *Nhe*I (for the case of the vector pET21a$^+$) and *Xho*I from Invitrogen ™ (Thermo Scientific, USA).

Development of *E. coli* competence and transformation was done according to the method of Chung *et al*., 1989. Briefly, the *E. coli* strain XL1-Blue MRF' was grown until early exponential phase and resuspended in a transformation and storage solution (which consists of LB with 10% polyethylene glycol, 5% dimethyl sulfoxide, 50 mM $MgCl_2$, pH 6.5). Plasmid DNA was added to the competent cells and incubated on ice for 30 minutes. Afterwards 0.9 mL of LB with 20 mM glucose were added and the cells were incubated at the *E. coli* strain growth conditions for one hour.

Selection of *E. coli* transformants was done in LB agar plates supplemented with ampicillin (100 μg/mL). The presence of the desired cloning products was confirmed by colony PCR using NZYTaq II 2x Green Master Mix (NZYTech, Portugal), with specific primers (NcoI-*lysX* and pIVEXRev or *lysX*-XmaI and T7promoter, with "*X*" denoting the different endolysin designations; Tables S5.1 and S5.2), according to the polymerase manufacturer's instructions. AGE for visualization of the amplification products was done as described above. Transformants carrying pBluescript II KS$^+$::*lysbIL67* and pBluescript II SK$^+$::*lysbIL67,* on the other hand, were selected on LB plates with X-Gal (5-bromo-4-chloro-3-indolyl-β-d-galactopyranoside, 40 μg/mL), IPTG (isopropylβ-D-thiogalactopyranoside, 20 mM), and ampicillin (100 μg/mL). Upon induction of *lac* operon by IPTG, X-Gal will be hydrolysed by the produced β-galactosidase, resulting in the appearance of blue colonies. If cloning was successful, the β-galactosidase gene will be disrupted by the insertion of *lysbIL67* and white colonies will appear. In the specific case of pIVEX2.4d with the gene *lysJavan488,* to analyse the transformants a restriction analysis was performed with *Pvu*II from Invitrogen ™ (Thermo Scientific, USA) following the manufacturer's instructions.

Transformants confirmed by PCR or restriction profile were then grown in LB medium with ampicillin (100 μg/mL) and incubated overnight at 37°C with aeration. Plasmid extraction and purification was performed with the commercial kit GeneJET Plasmid Miniprep Kit (Thermo Scientific™, USA) following the manufacturer's instructions. Sequence correctness of endolysin genes carried in the recombinant plasmids was confirmed by DNA sequencing (Eurofins Scientific, Luxemburg). The confirmed plasmids were then used to transform the *E. coli* strain CG61 as described above. CG61 transformants were selected in LB agar plates supplemented with ampicillin (100 μg/mL) and kanamycin (40 μg/mL) at 28°C.

2.3 General techniques for protein production and analysis

In order to induce protein production, the expression strains were grown in LB or in LB buffered with 0.2 M sodium phosphate at pH 7.2 and supplemented with 0.5 M of sorbitol to maximize protein solubility. The cultures were incubated at 28°C until an optical density at 600 nm (OD600) of 0.8, after which protein production was induced by moving the cultures to a shaking water bath set to 42°C for 30 minutes, in order to inactivate the cI857 repressor and induce recombinant protein production. After that, the cultures were incubated either for 3 hours at 37°C or overnight at 16°C, with agitation in both cases.

The cells from induced cultures were recovered by centrifugation (4900 *g*, 15 min, 4°C) and concentrated 20 to 40 times in lysis buffer (50 mM Hepes, 500 mM NaCl, 50 mM imidazole, 1% glycerol pH 7), supplemented with 20 µg/mL of DNase, 10 mM of $MgCl_2$ and 1X Complete Mini EDTA-free Protease Inhibitor Cocktail (Roche Applied Science, Germany). In the specific case of the endolysin LysJavan488, initial tests were also performed with lysis buffers containing 0.1% Triton X-100, 1 mM TCEP [Tris(2-carboxyethyl)phosphine hydrochloride] and/or 30% glycerol. In the case of LysLW32, initial tests were performed in the same conditions described for LysJavan488 and additionally: (1) with the pH of the lysis buffer adjusted to 5 or 8 (for pH 5 was used Tris instead of Hepes), (2) the LB buffered with 0.2 M sodium phosphate at pH 8, and (3) the culture incubated until an OD600 of 0.4 before induction and incubated at 16ºC for 30 minutes to 2.5 hours after induction.

The cells recovered in lysis buffer were kept on ice and disrupted by sonication with 10 cycles of 15 seconds (amplitude 50, pulse 0.5) intercalated with pauses of 45 seconds. The lysate was centrifuged (17400 *g*, 30 min, 4°C) and the supernatant (soluble fraction) collected.

Total extract before induction of protein production (T_0) and after induction (T_1) as well as the T_1 extract soluble fraction (S) and T_1 extract insoluble fraction (I) after resuspension in 7 M of urea and 2 M of thiourea, were mixed with reducing Laemmli buffer, heated for 10 minutes at 95°C, and subjected to sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) in a 15% polyacrylamide resolving gel. In the case of LysLW32, the samples were analysed in a 17.5% gel. The protein bands were visualized after staining gels with Coomassie Blue. PageRuler Prestained Protein Ladder (Thermo Scientific™, USA) was used as a protein marker in SDS-PAGE.

After SDS-PAGE, proteins were transferred onto 0.20 µm nitrocellulose membranes (Bio-Rad, USA) using the Trans-blot Cell Tank Transfer system (Bio-Rad, USA) for 90 minutes at 100 volts. Immuno-detection of the His_6 tag was performed with the Anti-His_6-Peroxidase mouse monoclonal serum (Roche Applied Science) diluted 1:5000. Antigen/antibody complexes were detected with the BM Chemiluminescence Western Blotting Kit (Roche Applied Science, Germany) according to the manufacturer's instructions. This detection is possible since the transformants carry a translation gene fusion vector which will form Lys-His_6.

2.4 Purification of endolysins

The soluble fraction collected from the supernatant, obtained as described above, was filtered through a 0.45 µm filter. Histidine tagged endolysins were purified by affinity chromatography (AFC) using HisTrap HP columns (GE Healthcare, USA) coupled to an ÄKTA-Prime system (GE Healthcare, USA). The binding and elution buffers had the same composition of the lysis buffer, except for the imidazole concentration in the elution buffer being 500 mM. The eluted fractions were analysed by

SDS-PAGE. Endolysins from AFC fractions were pooled, and the buffer exchanged to an imidazole-free buffer (50 mM Hepes, 500 Mm NaCl, 30% glycerol with or without 0.1% Triton X-100, pH 7) using HiTrap Desalting columns (GE Healthcare, USA). When deemed necessary, the AFC fractions were subjected to size exclusion chromatography (SEC) using a HiLoad 16/60 Superdex 75 pg column (GE Healthcare, USA) equilibrated in SEC buffer (50 mM Hepes, 500 mM NaCl and 1% glycerol, pH 7).

Protein concentrations were determined by the Bradford method. The Bradford reagent (Bio-Rad, USA) in contact with a protein changes its colour and that change is measured spectroscopically. A calibration curve using bovine serum albumin or bovine gamma globulin was used to infer protein concentration in the samples. The enzymes were divided in small aliquots and kept at -80°C.

2.5 Protein purification under denaturing conditions

Protein purification under denaturing conditions was tested as described in Schmelcher, Korobova *et al*., 2012. The cells from induced cultures were resuspended in 16 mL of 100 mM NaH_2PO_4, 10 mM Tris, 8 M urea, pH 8 per 800 mL culture and incubated at room temperature for 1 hour under agitation (in order to disrupt the cells and dissolve putative inclusion bodies). Residual insoluble material was eliminated by centrifugation for 25 minutes at 10000 g. The supernatant was applied to an ÄKTA-Prime system (GE Healthcare, USA) with a coupled HiTrap HP columns (GE Healthcare, USA). The column was washed with 100 mM NaH_2PO_4, 10 mM Tris, 8 M urea, pH 6 and target proteins were eluted with 100 mM NaH_2PO_4, 10 mM Tris, 8 M urea, pH 4.5. The fractions were analysed by SDS-PAGE, concentrated using VIVASPIN 20 5000MWCO PES (Sartorius Stedium, Germany) and dialyzed twice for 2 hours against 20 mM NaH_2PO_4, 150 mM NaCl, 15% glycerol, pH 7.5 using Slide-A-Lyzer™ Dialysis Cassettes, 3.5K MWCO (Thermo Scientific™, USA). The protein preparations were then centrifuged for 25 minutes at 10000 *g* and analysed again by SDS-PAGE.

2.6 Analytical size exclusion chromatography

Samples of 100 µL of purified Lys-His_6 at 1.25 mg/mL of FLP, 1.25 mg/mL of native FLP/CTP complex and 2.5 mg/mL of CTP, were applied to a Superose 6 10/300 GL column (GE Healthcare, USA) equilibrated in protein buffer (same as the SEC buffer, see above) and run at a flow rate of 0.4 mL/min. It was also applied a combination of 1.25 mg/mL FLP and 2.5 mg/mL of CTP after being incubated for 1 hour at 37°C. The column void (V_0) and total (V_t) volumes were determined using the elution volumes (V_e) of blue dextran 2000 (GE Healthcare, USA) and 4% acetone, respectively. The protein standard was composed of thyroglobulin (670 kDa and stokes radius [Rs] of 8.6 nm), γ-globulin (158 kDa and Rs of 5.1 nm), ovalbumin (44 kDa and Rs of 2.73 nm), myoglobin (17 kDa and Rs of 1.91nm) and vitamin B12 (1.35 kDa and Rs of 0.85 nm) from Gel Filtration Standard (Bio-Rad, USA), and conalbumin (with 75 kDa and Rs 3.72 nm) (GE Healthcare, USA) (Cabré *et al.*, 1989; Momand *et al.*, 2017; Stetefeld *et al.*, 2016).

The molecular mass of purified Lys-His_6 was estimated through two different extrapolations (Cabré *et al.*, 1989). One from a plot of the standard proteins' distribution coefficient (K_{av}) *versus* the logarithm of their molecular masses, where,

$$(2.1)\ K_{av} = \frac{V_e - V_0}{V_t - V_0},$$

The second one by the determination of the purified Lys-His_6 Rs from a plot of the standard proteins' Rs and their K_{av},

(2.2) Rs = - $\sqrt[2]{logK_{av}}$,

after which a regression of the Rs *versus* the standard proteins' logarithmic molecular masses that allowed to estimate the molecular mass of the purified Lys-His_6.

2.7 Evaluation of endolysin lytic action

a. LysJavan488

Purified endolysin polypeptides were tested against their bacterial target by spotting different protein quantities in dense lawns of viable cells. *S. pyogenes* M6 was grown until an OD600 of 0.8, centrifuged, and the cells were concentrated 100 times in fresh growth medium. A sample of the cell suspension was 100-fold diluted in LysJavan488 reaction buffer (35 mM ammonium acetate, 150 mM NaCl and 10 mM of $CaCl_2$, pH 6.25, based on Celia *et al*., 2008 and Pritchard *et al*., 2004), supplemented with 0.7% agar and poured into a Petri dish. This semi-solid reaction buffer was prepared by mixing equal parts of 2-fold concentrated reaction buffer with a melted 1.4% agar solution. Different protein quantities of the endolysin polypeptides (0 to 16 μM, in 10 μL final volume) were spotted and, after an overnight incubation at 37°C under an atmosphere of 5% carbon dioxide, observed for the presence of transparent lysis halos.

b. LysPhi7951

S. thermophilus strain 4078 was grown overnight as described above. Then, cells were collected and resuspended in 1X Phosphate-buffered saline (PBS) supplemented with 16% glycerol to an OD600 of 80 and stored at -80°C. A sample of these cell suspensions was 100-fold diluted in LysPhi7951 reaction buffer (25 mM Hepes and 5 mM $CaCl_2$ at pH 7), supplemented with 0.7% agar and poured into a Petri dish (as described above for LysJavan488). Defined amounts of the endolysin (FLP, CTP, native FLP/CTP complex and reconstituted complex) were spotted in 10 μL final volume and, after an overnight incubation at 42°C, observed for the presence of transparent lysis halos.

In another assay, *S. thermophilus* was grown until an OD600 of 0.4, centrifuged, washed and the cells resuspended in LysPhi7951 reaction buffer to an OD600 of 0.8. The activity of the different LysPhi7951 polypeptides (or combinations) was monitored based on the reduction of OD600 of cell suspensions. This cell suspension was composed by 200 μL of final volume prepared by mixing equal parts of the cells resuspended and of LysPhi7951 polypeptides. The OD600 was measured in a 96-well plate every hour, for a total of 12 hours, using a Varioskan LUX Multimode Microplate Reader (Thermo Scientific, USA). Plates were incubated at 37°C, with cell suspensions agitated before each OD600 measurement.

Optimization of the reaction buffers was performed. The following buffers were tested: (1) 25 mM Hepes, 5 mM $CaCl_2$, pH 8; (2) 25m M Hepes, 5 mM $CaCl_2$, 250 mM NaCl, pH 7; (3) 25 mM Hepes, 5 mM $CaCl_2$, pH 7; (4) 25 mM Hepes, 5 mM $CaCl_2$, 500 mM NaCl, pH 7; (5) 25 mM Hepes, 5 mM $CaCl_2$, 500 mM NaCl, 1 mM of TCEP, pH 7; (6) 25mM Pipes, 5 mM $CaCl_2$, pH 6; (7) 50 mM Sodium acetate / acetic acid buffer, 5 mM $CaCl_2$, pH 5; (8) 50 mM Sodium acetate /acetic acid buffer, 5 mM $CaCl_2$, pH 4.

Negative controls were equally prepared, except that the same volume of the protein buffer was added instead of the endolysin.

2.8 Mutagenesis of the ITSS of *lysPhi7951*

The mutagenesis of the *lysPhi7951* ITSS was performed by overlap extension PCR. First, two PCR products covering the 5' and 3' end of the endolysin gene, and overlapping in the ITSS region, were produced using the primer pairs NcoI-*lysPhi7951*/ LysP7951_ITSSminus_rev and LysP7951_ITSSminus_frw/ *lysPhi7951*-XmaI (Tables S5.1 and S5.3). Then, the two PCR products were used as template to amplify and assemble the entire endolysin gene with changed ITSS, using the primer pair NcoI-*lysPhi7951/ lysPhi7951*-XmaI (Tables S5.1). After cloning, the mutagenesis was confirmed by DNA sequencing (Eurofins Scientific, Luxemburg). The changes introduced at the gene and protein level are indicated in Table 2.1.

Table 2.1 – Mutagenesis of the ITSS

LysPhi7951	With the ITSS	Without the ITSS
DNA sequence (5' → 3')	...GTAGATAAGAAAGA**G**GAAGAAG AAAAT**A**TGGATTATGTAGTTCGA…	…GTAGATAAGAAAGA**A**GAAGAAGA AAAT**C**TGGATTATGTAGTTCGA…
Protein sequence (N→ C)	…VDKKEEEENMDYVVR...	…VDKKEEEEN**L**DYVVR ...

Underline is the ITSS (in the DNA and protein sequence) and in bold is represented the changes made.

3. Results and Discussion

The gene and proteins sequences of the selected endolysins (Fig. S5.1) can be accessed through the Phage Lytic Proteins database (PhaLP, https://www.phalp.org/), using the corresponding UniProt accession number. These were:

- Two endolysins from phages which specifically infect *Lactococcus lactis* - LysbIL67 (UniProt Q38241), from phage bIL67, and LysLW32 (UniProt A0A1W6JHS3) from phage LW32. *L. lactis* is the most widely utilized bacteria in dairy lactic starter cultures (Lavelle *et al*., 2020), because of their rapid lactic acid production from lactose that leads to the preservation of the otherwise quickly spoiled milk. They also contribute to both flavour and texture of the final product (Mierau and Kleerebezem, 2005). *L. lactis* is generally regarded as safe and due to its economic importance, these bacteria have been intensely investigated and characterized. More recently they have been seen as a factory for antigen production, a vehicle to deliver therapeutics and a probiotic due to their ability to survive through the gastrointestinal tract, not invading or colonizing the mucosal surfaces of the host. Moreover, since they do not have lipopolysaccharides they do not stimulate host immune responses (Cook *et al.*, 2018; Mierau and Kleerebezem, 2005; Song *et al.*, 2017).

- LysPhi13 (UniProt Q8SDJ9), the endolysin of *Staphylococcus aureus* phage Phi13. *Staphylococcus aureus* causes a variety of tissues infections, bloodstream infections, toxic shock syndrome, and food poisoning (Gutiérrez *et al.*, 2018). They have become a serious threat in clinical environments because of their reported increased resistance to antibiotics, leading the WHO to indicate methicillin-resistant *Staphylococcus aureus* as a high priority pathogen against which novel treatment options are urgently needed (WHO, 2017).

- LysJavan488 (UniProt A0A4D6BBK1), the endolysin of phage Javan488, which specifically infects *S. pyogenes*. *S. pyogenes* is the leading cause of pharyngitis in children and adolescents (Kanwal and Vaitla, 2020) and can also cause life-threatening invasive illnesses, such as, necrotizing fasciitis and pneumonia. For those reasons, it has been considered by WHO, the ninth leading infectious etiology of human mortality (Ferretti *et al.*, 2016; Wilkening and Federle, 2017).

- LysPhi7951 (UniProt A0A286QQJ8), the endolysin of phage Phi7951. This phage infects *S. thermophilus*, a valuable microorganism to the dairy industry due to its sugar metabolism, proteolysis and production of important metabolites. Amongst such important metabolites are exopolysaccharides (EPSs), which affect the organoleptic properties of fermented foods and protocooperation with *Lactobacillus delbrueckii* and *L. lactis* (Markakiou *et al*., 2020; McDonnell *et al*., 2020). Besides that, their EPSs can also be used as thickeners, stabilizers, emulsifiers, and water-binding agents. Recently, it was found that these EPSs can be antioxidant, anticancer, anti-inflammatory, and immunomodulatory agents increasing their use by the pharmaceutical industry, and as probiotics (Cui *et al*., 2017; Martinović *et al*., 2020). In addition, the bacterium's ability to survive and be metabolically active within the gastrointestinal tract allows it to be considered a tool for targeted delivery of various biological molecules (Lecomte *et al*., 2016).

3.1 LysbIL67

LysbIL67 coding gene was amplified by PCR, digested with *Nco*I and *Xma*I and ligated into the expression vector pIVEX2.3d cut with the same restriction enzymes. No transformants containing pIVEX2.3d::*lysbIL67* could be retrieved, as assessed by PCR. In order to overcome this issue, cloning into a different expression vector (pET21a$^+$) was attempted. In this case the vector was digested with *Nhe*I and *Xho*I, while the insert was cut with *Spe*I and *Xho*I, generating compatible ends. Also in this case, no transformants containing pET21a$^+$::*lysbIL67* could be obtained. At this point, suspicions that LysbIL67 could be toxic to *E. coli* arose, so cloning of *lysbIL67* into a typical cloning vector, not carrying a ribosome binding site in this translation gene fusion vector, was attempted. Both the vector pBluescript II KS$^+$ and insert (*lysbIL67*) were digested with *Spe*I and *Xho*I, ligated and transformed into *E. coli*. Once more, no pBluescript II KS$^+$::*lysbIL67* transformants were obtained. A final attempt was made to clone *lysbIL67* into pBluescript II SK$^+$, also via *Spe*I and *Xho*I, to explore a different orientation of the insert. However, not even this allowed the obtention of transformants carrying *lysbIL67*.

We have then hypothesized that *lysbIL67* cloning is problematic most likely due to toxicity issues. In fact, in a previous article a similar problem was described with the endolysin of phi-vML3, which shares 94% identity with the LysbIL67. There, the authors show that expression of full endolysin gene in *E. coli* leads to the cell death (Shearman *et al*., 1994), supporting the toxicity hypothesis.

3.2 LysPhi13

After the amplification and digestion of *lysPhi13*, the endolysin gene (Table S5.4) was ligated to the expression vector pIVEX2.3d (pIVEX2.3d::*lysPhi13*) and used to transform *E. coli* XL1-Blue MRF'. Transformants were screened through colony PCR and the amplification products were analysed by AGE (Fig. 3.1). Some of the colonies tested showed the amplification product with the expected size (around 893 bp, including vector borne sequences). One clone was selected, its plasmid was extracted and confirmed by DNA sequencing.

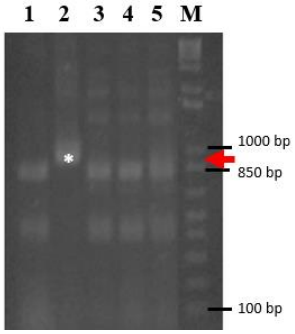


Figure 3.1 - Analysis of *E. coli* strain XL1-Blue MRF' derivatives transformed with pIVEX2.3d::*lysPhi13*.
M – 1 Kb Plus DNA Ladder Invitrogen™ (Thermo Fisher Scientific, USA); 1 – transformant #6; 2 – transformant #7; 3 – transformant #8; 4 – transformant #9; 5 – transformant #10. The arrow indicates the position of the expected amplification product and the asterisk the clone chosen for sequencing.

pIVEX2.3d::*lysPhi13* extracted from clone #7 was used to transform *E. coli* CG61. After induction of gene expression, production of the expected polypeptides was not observed (Fig. 3.2A) – FLP had an expected mass of 30.2 kDa and the CTP an expected mass of 13.7 kDa. To check if the two

polypeptides were present at low amounts not detected by SDS-PAGE, a Western blot analysis was made with an Anti-His_6-Peroxidase mouse monoclonal antibody, since the two expected polypeptides contained a C-terminal His_6 tag. However, the results did not indicate the presence of any polypeptide containing the His_6 tag (Fig. 3.2B), suggesting that they were not being produced from the pIVEX2.3d::*lysPhi13*.

These results indicated a problem in LysPhi13-His_6 synthesis, and therefore we tested other culture conditions with the aim of maximizing protein production. Previous work with other endolysins showed that growing the expression *E. coli* strain in LB buffered with 0.2 M sodium phosphate and supplemented with 0.5 M sorbitol could result in improvement of protein production and/or solubility. In addition, cultures were incubated overnight at 16°C with aeration following induction of protein synthesis. However, despite the new conditions the LysPhi13-His_6 polypeptides were not detected (Fig. 3.2C).

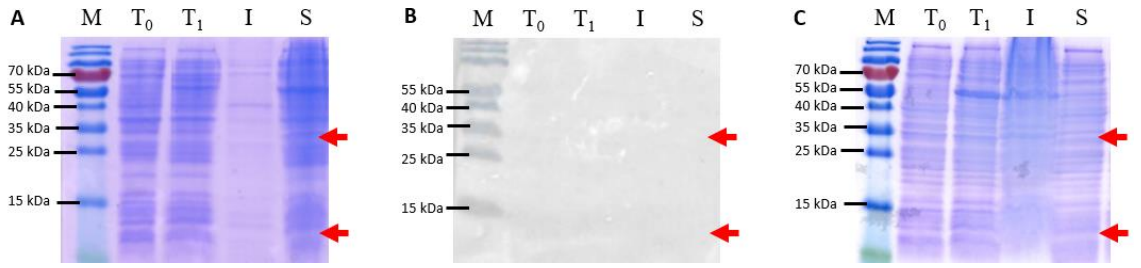


Figure 3.2 - Analysis of LysPhi13-His6 production from *E. coli* CG61 carrying the recombinant plasmid pIVEX2.3d::*lysPhi13*: (A) SDS-PAGE analysis of protein extracts before and after induction of protein production (in LB medium, 3 hours at 37°C after induction); (B) Anti-His_6 Western blot analysis of the protein extracts shown in A; (C) SDS-PAGE analysis of protein extracts before and after induction of protein production (cells grown in LB buffered with 0.2 M sodium phosphate and supplemented 0.5 M sorbitol, and protein production occurring during overnight incubation at 16°C, with aeration).
M – PageRuler Prestained Protein Ladder (Thermo Scientific™, USA); T_0 – Total extract before induction of protein production; T_1 – Total extract after induction of protein production; I – T_1 extract insoluble fraction; S – T_1 extract soluble fraction. The arrows indicate the expected molecular mass of the FLP (30.2 kDa) and the CTP (13.7 kDa).

Lys87, an endolysin similar to LysPhi13, could be produced in a previous study, although with poor solubility (Fernandes *et al*., 2012). However, for reasons yet unclear, we could not detect LysPhi13 production in the conditions tested.

3.3 LysLW32

The endolysin LW32 gene (Table S5.4) was cloned into the vector pIVEX2.3d. The desired plasmid, pIVEX2.3d::*lysLW32*, was obtained in *E. coli* XL1-Blue MRF' after colony PCR screening (as in Fig. 3.1). The clone #3 was selected, its plasmid extracted and confirmed by DNA sequencing.

E. coli CG61 was then transformed with the confirmed pIVEX2.3d::*lysLW32* plasmid and protein production tested as for LysPhi13. Protein induction resulted in the accumulation of two polypeptides, one with an apparent molecular mass compatible with that expected for the FLP (29.3 kDa), and another with an apparent molecular mass compatible with that expected for the CTP (8.4 kDa, Fig. 3.3A). A Western blot analysis with the Anti-His_6 antibody confirmed that both polypeptides contained the His_6 tag (Fig. 3.3B). Given that pIVEX2.3d generates C-terminal His_6 tag fusions, this

result supports the hypothesis that the two polypeptides were products of the same gene, being the smaller one likely initiated at the *in frame* ITSS.

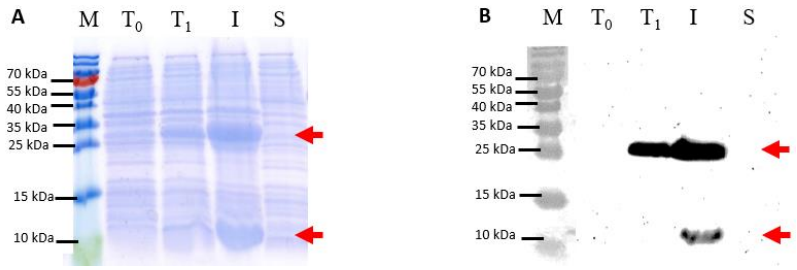


Figure 3.3 - Analysis of LysLW32-His$_6$ production from *E. coli* CG61 carrying the recombinant plasmid pIVEX2.3d::*lysLW32*. (A) SDS-PAGE analysis of protein extracts before and after induction of protein production (in LB medium, 3 hours at 37°C after induction); (B) Anti-His$_6$ Western blot analysis of the protein extracts shown in A.
M – PageRuler Prestained Protein Ladder (Thermo Scientific™, USA); T_0 – Total extract before induction of protein production; T_1 – Total extract after induction of protein production; I – T_1 extract insoluble fraction; S – T_1 extract soluble fraction. The arrows indicate the expected size of the FLP (29.3 kDa) and of the CTP (8.4 kDa).

Unfortunately, the two polypeptides could only be detected in the insoluble fraction of protein extracts, which hindered their purification. In order to increase the solubility of the recombinant proteins, other production conditions were tested, including those tested for LysPhi13 (see above). Nonetheless, neither of them increased solubility, at least not significantly enough to allow detection of the polypeptides in the soluble fractions (Fig. S5.2).

Purification under denaturing conditions was then pursued. A small amount of the two LysLW32 polypeptides (Fig. 3.4) was obtained by AFC (AFC elution profile in Fig. S5.3). Fractions AF3 to AF6 and AF9 to AF12 were concentrated and then subjected to dialysis to remove the denaturation agent and promote protein renaturation, followed by centrifugation to eliminate possible precipitates. Protein precipitation was observed during the process, which explains the weak protein bands seen in the SDS-PAGE (Fig. 3.4 in the 2ndC). Hence, our attempt to obtain soluble LysLW32 after protein purification under denaturing conditions was not successful.

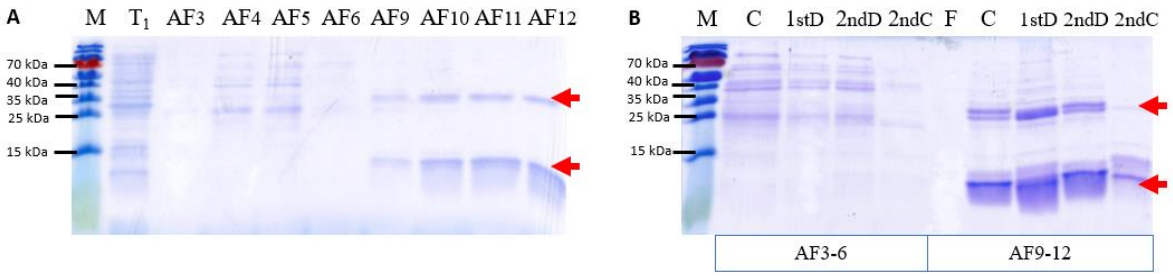


Figure 3.4 – SDS-PAGE analysis of the production and purification of LysLW32-His$_6$ under denaturing conditions. (A) Analysis of the AFC elution fractions; (B) Analysis of samples from the concentration and dialysis process.
M – PageRuler Prestained Protein Ladder (Thermo Scientific™, USA); T_1 – total protein extracts after induction of protein production; AF3 to AF6 and AF9 to AF12 – samples of the AFC elution fractions 3 to 6 and 9 to 12; C – concentrated AFC fractions; 1stD – first dialysis; 2ndD – second dialysis; 2ndC – second dialysis after centrifugation; F – flow-through from the concentration. The arrows indicate the expected size of the FLP (29.3 kDa) and of the CTP (8.4 kDa).

In conclusion, in the case of LysLW32 we demonstrated the production of two polypeptides from its gene, compatible with initiation of translation at the ITSS. However, soluble LysLW32 could not be obtained for further studies.

3.4 LysJavan488

3.4.1 Cloning and protein production

The *lysJavan488* gene (Table S5.4) ligated to the expression vector pIVEX2.3d (pIVEX2.3d::*lysJavan488*) was used to transform the *E. coli* strain XL1-Blue MRF'. Transformants were screened by colony PCR and AGE analysis, as before. Clone #6 was selected, and its plasmid extracted and confirmed by DNA sequencing.

E. coli CG61 was transformed with pIVEX2.3d::*lysJavan488* extracted from clone #6. Induction of protein production resulted in the accumulation of two polypeptides, one corresponding to the FLP (expected molecular mass of 45.3 kDa) and the other to the CTP (expected molecular mass of 27.8 kDa, Fig. 3.5A). Through a Western blot analysis with the Anti-His_6 antibody we confirmed that both polypeptides contained a His_6 tag. This indicated that the two polypeptides were products of the same gene, as pIVEX2.3d produced a His_6 C-terminal fusion. Together these results support the hypothesis that the two polypeptides are being produced from the same gene, being the smaller one likely initiated at the *in frame* ITSS (Fig. 3.5B).

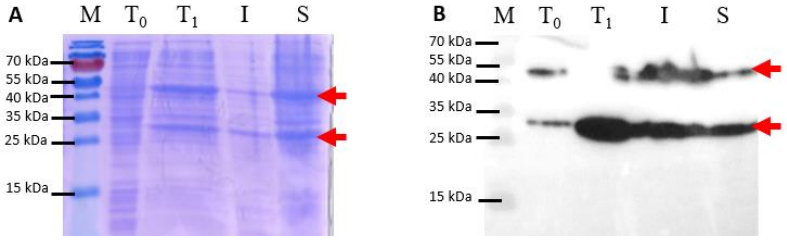


Figure 3.5 - Analysis of LysJavan488-His_6 production from *E. coli* CG61 carrying the recombinant plasmid pIVEX2.3d::*lysJavan488*. (A) SDS-PAGE analysis of protein extracts before and after induction of protein production (in LB medium, 3 hours at 37°C after induction); (B) Anti-His_6 Western blot analysis of samples shown in A.
M – PageRuler Prestained Protein Ladder (Thermo Scientific™, USA); T_0 – Total extract before induction of protein production; T_1 – Total extract after induction of protein production; I – T_1 extract insoluble fraction; S – T_1 extract soluble fraction. The arrows indicate the expected size of the FLP (45.3 kDa) and of the CTP (27.8 kDa).

Under the assumption that the two polypeptides would interact, as in the previously studied examples (Dunne *et al*., 2014; Dunne *et al*., 2016; Proença *et al*., 2015; Zhou *et al*., 2020), a purification strategy was developed that consisted in generating a N-terminal His_6-tagged LysJavan488. This would allow the capture and purification of the protein complex containing the His_6-LysJavan488FLP and the CTP when associated. For that purpose, *E. coli* XL1-Blue MRF' was transformed with a ligation of pIVEX2.4d and *lysJavan488* via *Nco*I and *Xma*I restriction sites. Transformants were screened by colony PCR (Fig. 3.6A).

The obtained results included some non-specific amplifications and since the negative control also showed a band close to the relevant size of 1340 bp, we attempted a different approach. Plasmids

were extracted from transformants #23 and #25 and were subjected to restriction with the enzyme *Pvu*II. The restriction profile was analysed after AGE. Three bands with 821, 1566 and 2366 bp were expected for the correct plasmid structure. The plasmid extracted from clone #23 showed the expected restriction profile (Fig. 3.6B). DNA sequencing provided further confirmation. Then, pIVEX2.4d::*lysJavan488* was transferred to *E. coli* CG61 for protein production, which confirmed the presence of the two expected polypeptides.

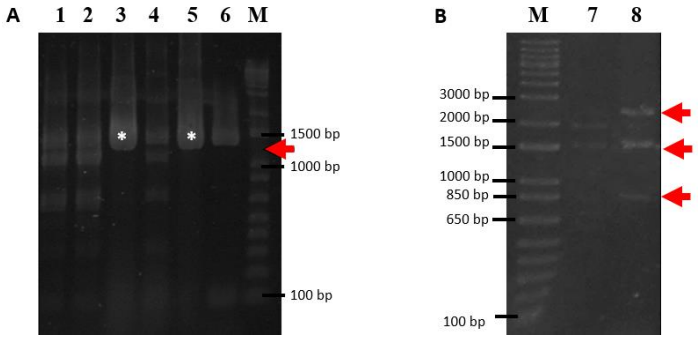


Figure 3.6 - Analysis of *E. coli* strain XL1-Blue MRF' derivatives transformed with pIVEX2.4d::*lysJavan488*: (A) screening of the transformants; (B) restriction profile with the enzyme *PvuII.*
M – 1 Kb Plus DNA Ladder Invitrogen™ (Thermo Fisher Scientific, USA); 1 – transformant #21; 2 – transformant #22; 3 – transformant #23; 4 – transformant #24; 5 – transformant #25; 6 – negative control; 7 – plasmid isolated from transformant #25; 8 – plasmid isolated from transformant #23. The arrow indicates the expected product and the asterisk the clones with the amplification product close to the expected.

3.4.2 Protein purification

His_6-LysJavan488 was purified by AFC (Fig. S5.4A). Fractions of the AFC peak were analysed by SDS-PAGE and showed that the two polypeptides, one corresponding to the expected FLP and the other to the CTP, were co-purified. However, the LysJavan488 FLP was purified in a much higher quantity than the CTP (Fig. 3.7A), which was suggestive that the interaction between the two polypeptides was not occurring in these conditions.

This inability to form a complex could be related to the protein conformation, which may be altered by some of the buffer compounds (the TCEP, because it eliminates the disulphide bonds, or the detergent Triton X-100). These possibilities were tested by generating protein extracts and performing protein purification, first without TCEP (Fig. 3.7B and Fig. S5.4C) and then without TCEP and Triton X-100 (Fig. 3.7C and Fig. S5.4E). These conditions did not change the previous observations. The AFC peak fractions of the three purifications were pooled and subjected to buffer exchange to remove the imidazole, using HiTrap Desalting columns, and the fractions were then analysed by SDS-PAGE (Fig. S5.4B, D and F).

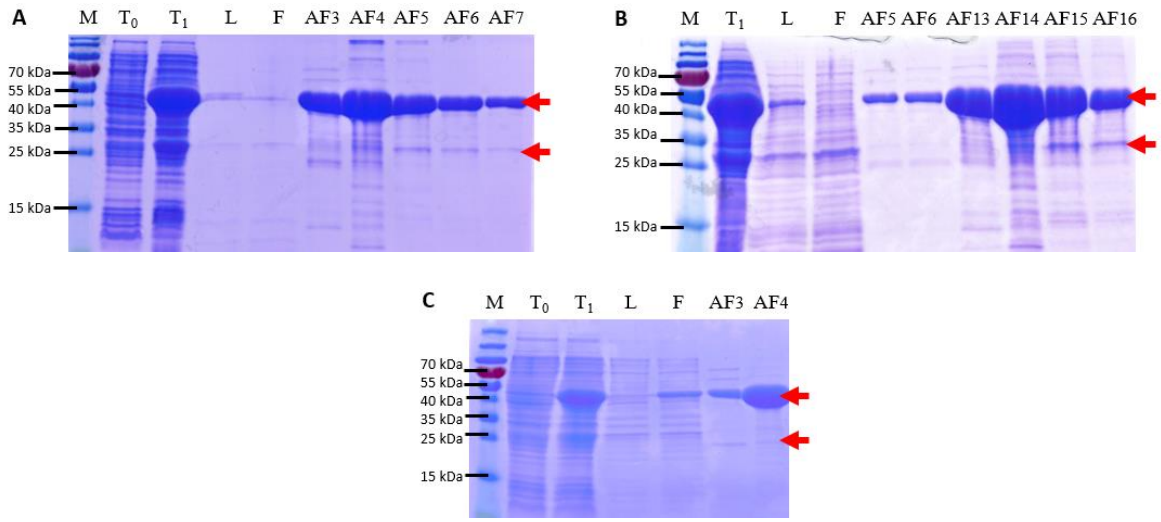


Figure 3.7 - Analysis of the production and purification of His_6-LysJavan488 (cells grown in LB buffered with 0.2 M sodium phosphate and supplemented 0.5 M sorbitol, with protein production occurring during overnight incubation at 16°C, with aeration) by AFC: (A) under reducing conditions (1 mM TCEP) and in the presence of Triton X-100, (B) under oxidizing conditions and (C) under oxidizing conditions and in the absence of detergent.
M – PageRuler Prestained Protein Ladder (Thermo Scientific™, USA); T_0 – Total extract before induction of protein production; T_1 – Total extract after induction of protein production; L – lysate's soluble fraction before loading into the column; F – flow-through of the AFC; AF3 to AF7 and AF13 to AF16 – eluted fractions 3 to 7 and 13 to 16 of the AFC. The arrows indicate the expected size of the FLP (47.4 kDa) and of the CTP (27.1 kDa).

Subsequently, the purification by AFC of the LysJavan488-His_6 (from pIVEX2.3d::*lysJavan488*) was also performed (Fig. 3.8A-B) and followed by a SEC. If the two polypeptides were interacting, we should have a peak in the SEC corresponding to a form of at least 73.1 kDa (the sum of 45.3 kDa of one FLP and 27.8 kDa of one CTP). The peaks seen in a SEC, the fractions of which were analysed by SDS-PAGE (Fig. 3.8C-F), corresponded to the individual, non-associated polypeptides. Therefore, in these conditions no evidences of FLP/CTP interaction were obtained.

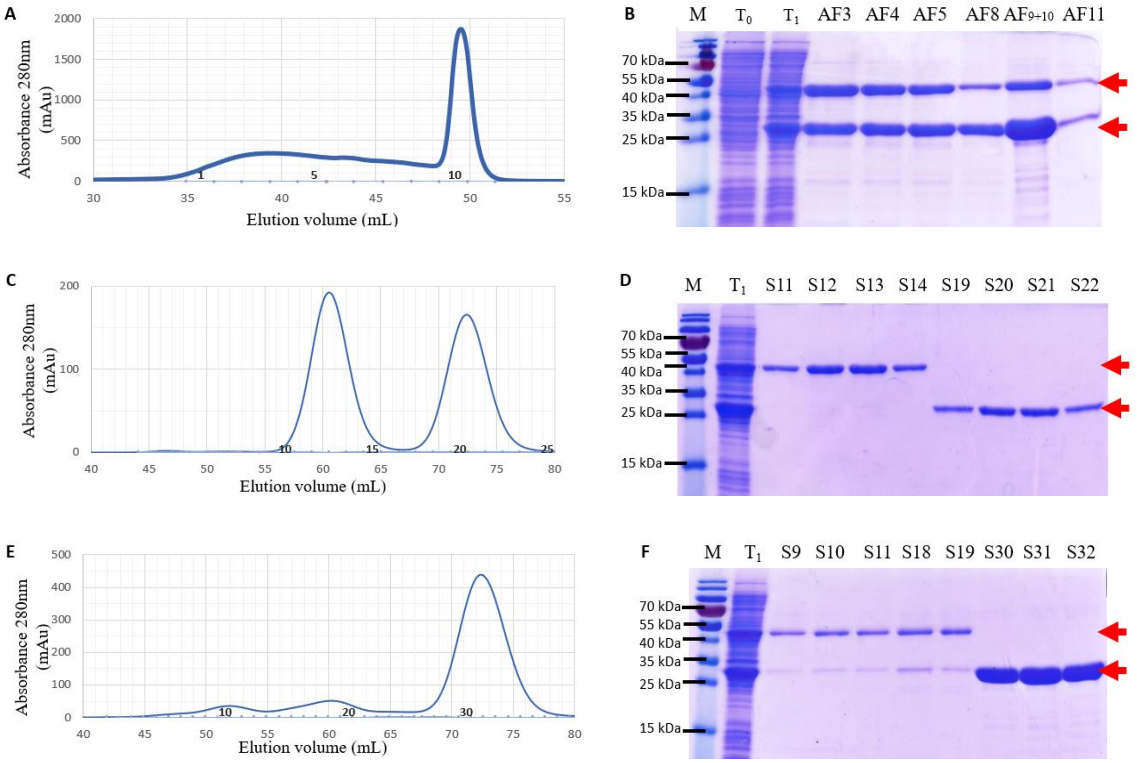


Figure 3.8 - Purification of the proteins produced from pIVEX2.3d::*lysJavan488* (with the same conditions as in Fig. 3.7). (A) and (B) Elution profile and SDS-PAGE analysis of the AFC peaks, respectively; (C) and (D) Elution profile of the SEC performed with fractions 1 to 5 of the AFC and SDS-PAGE analysis of the corresponding SEC fractions, respectively; (E) and (F) Elution profile of the SEC performed with fractions 9 and 10 of the AFC and SDS-PAGE analysis of the corresponding SEC fractions, respectively. The numbers shown along the horizontal axis identify the fraction number.
M – PageRuler Prestained Protein Ladder (Thermo Scientific™, USA); T_0 – Total extract before induction of protein production; T_1 – Total extract after induction of protein production; AF3 to AF11 – fractions 3 to 11 of the AFC; S9 to S14, S18 to S22 and S30 to S32 – fractions 9 to 14, 18 to 22 and 30 to 32 of the SEC. The arrows indicate the expected size of the FLP (45.3 kDa) and of the CTP (27.8 kDa).

3.4.3 LysJavan488 lytic activity

The lytic activity of LysJavan488 was analysed, FLP and the CTP were used, alone or in combination, against *S. pyogenes* cells. Before going into the activity results, we should note that: (1) when we refer to the concentration of the mixture, each protein was added at half the target concentration; (2) we have used FLP with a His_6 tag in the N-terminus (fractions DS2-4 in Fig. S5.4D) and the CTP with the His_6 tag in the C-terminal (fractions S30-31 in Fig 3.8F); and (3) due to the quality of the purification (Fig. 3.7) the FLP preparation (His_6-LysJavan488) contains a small amount of CTP.

The endolysin polypeptides were spotted in dense lawns of viable cells and the formation of transparent halos resulting from cell lysis checked after overnight incubation. Although not possible to capture in photographs, halos were visible at concentrations between 8 and 16 µM of FLP, CTP and the mixture of both, although with the FLP produced smaller and less translucid halos. At 6 µM, halos were only observed with the CTP, the FLP/CTP mixture and in some assays also with the FLP but again less transparent and smaller. Bellow this concentration no more halos were observed.

These qualitative results give the first hints concerning the role of the two species produced from the *lysJavan488* (Table 3.1):

(1) The CTP is more potent than the FLP (since it is active at lower concentrations).

(2) A synergistic effect is observed between the species. At 6 μM of the mixture of both species we have 3 μM of CTP and 3 μM of FLP, concentrations at which each species individually is no longer active against *S. pyogenes* cells. However, when combined they become active. One could argue that, given that the FLP contains the CD and the CWBD which make up the CTP (Table S5.4), we are just observing the effect of having 6 μM of a CTP composed by us (with 3 μM of CTP and 3 μM of FLP). Still, that does not seem to be the case as 6 μM of FLP is not always active, although carrying the CTP, and when it produced halos these are smaller and less transparent than those observed with 6 μM of the mixture.

Table 3.1 – Lytic activity of the LysJavan488 against *S. pyogenes*

	Concentration (μM)								
	0	2	4	6	8	10	12	14	16
FLP	-	-	-	?	+	+	++	++	+++
CTP	-	-	-	+++	+++	+++	+++	+++	+++
Mixture of both	-	-	-	+++	+++	+++	+++	+++	+++

The sensitivity of S. pyogenes to the lytic action of LysJavan488 polypeptides (FLP, CTP and a mixture of both) was evaluated based on the observation of lysis halos on a relative scale ranging from opaque/small (+) to translucent/large (+++). Resistance to lytic action is indicated with (-) and with (?) is indicated the inconsistent results with the indicated concentration (some assays showed a very small and poorly translucid halo, while in others no obvious lysis halo was observed).

3.4.4 Conclusions

We have demonstrated that two polypeptides are produced from *lysJavan488* using an *in frame* ITSS. However, contrary to other examples in the literature (Dunne *et al*., 2014; Dunne *et al*., 2016; Proença *et al*., 2015; Zhou *et al*., 2020), for LysJavan488 there is no evidence that the two produced polypeptides associate, in the conditions tested. Why the two isoforms are produced and what is the biological relevance of it is under investigation.

3.5 LysPhi7951

3.5.1 Cloning and construction of pIVEX2.3d::*lysPhi7951*

Transformants of the *E. coli* strain XL1-Blue MRF' with the endolysin gene (Table S5.4) cloned in the expression vector pIVEX2.3d (pIVEX2.3d::*lysPhi7951*) were screened by colony PCR and AGE analysis of the products (as in Fig 3.1). All colonies tested showed the expected amplification product of 1037 bp. The clone #25 was selected, its plasmid extracted and confirmed by DNA sequencing.

E. coli CG61 was transformed with plasmid pIVEX2.3d::*lysPhi7951* isolated from clone #25. A transformant was chosen and induction of protein production resulted in the accumulation of two polypeptides, the FLP with a molecular mass of 34.4 kDa and the CTP with a molecular mass of 10.8 kDa (Fig. 3.9A). A Western blot analysis with the Anti-His_6 antibody confirmed that both polypeptides contained a His_6 tag, supporting the idea that they are produced from the same gene, and the smaller one is initiated at the *in frame* ITSS (Fig. 3.9B).

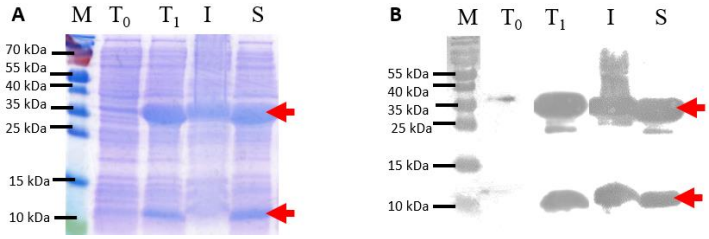


Figure 3.9 - Analysis of LysPhi7951-His$_6$ production from *E. coli* CG61 carrying the recombinant plasmid pIVEX2.3d::*lysPhi7951*. (A) SDS-PAGE analysis of protein extracts before and after induction of protein production (in LB medium, 3 hours at 37°C after induction); (B) Anti-His$_6$ Western blot analysis of samples shown in A.
M – PageRuler Prestained Protein Ladder (Thermo Scientific™, USA); T_0 – Total extract before induction of protein production; T_1 – Total extract after induction of protein production; I – T_1 extract insoluble fraction; S – T_1 extract soluble fraction. The arrows indicate the expected molecular mass of the FLP (34.4 kDa) and the CTP (10.8 kDa).

Purification of LysPhi7951-His$_6$ was performed by AFC, followed by a SEC. The peak fractions from both chromatographies were analysed by SDS-PAGE (Fig. 3.10). AFC generated two different peaks of hight and shape (Fig. 3.10A) and for that reason two SEC were performed: one with the AFC fractions of the first peak (AF3-4) and another with the fractions of the second (AF5-6).

SEC from the first AFC peak (Fig. 3.10C) reveals a single peak with an elution volume compatible with a species of about 35 kDa. SDS-PAGE analysis of the SEC fractions (Fig. 3.10D) is compatible with this peak corresponding to the FLP of LysPhi7951 with only a minor presence of the corresponding CTP.

SEC from the second AFC peak (Fig. 3.10E) showed something different: two peaks are observed, being the highest one compatible with a species with a much higher molecular mass than that of the FLP (and of the CTP). SDS-PAGE analysis (Fig. 3.10F) shows both species (FLP and CTP) at comparable amounts, peaking at the same elution volume. These observations are compatible with the second AFC peak corresponding to a FLP-CTP multimeric complex.

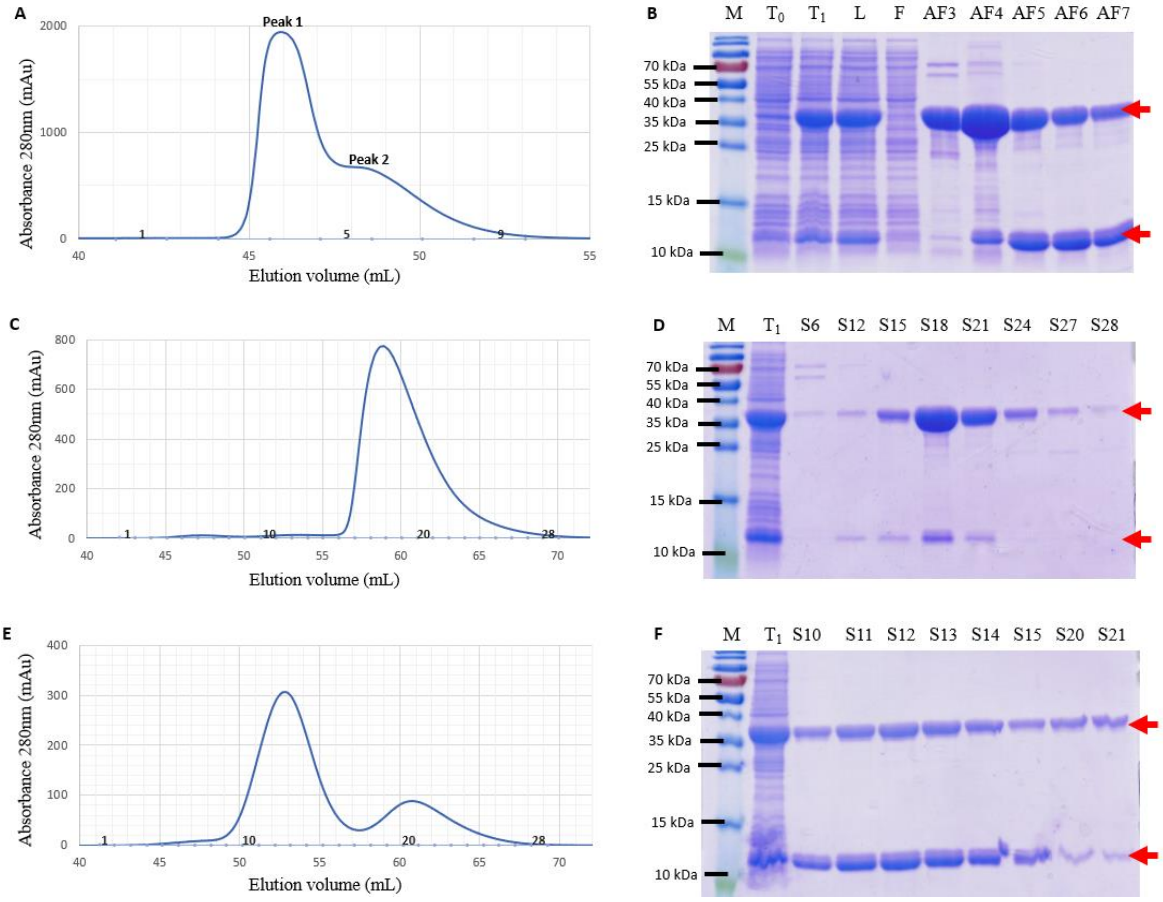


Figure 3.10 - Purification of the proteins produced from pIVEX2.3d::*lysPhi7951* (with protein production as in Fig. 3.9). (A) and (B) Elution profile and SDS-PAGE analysis of the AFC peaks, respectively; (C) and (D) Elution profile and SDS-PAGE analysis of the SEC performed with the fractions 3 and 4 of the AFC (Peak 1); (E) and (F) Elution profile and SDS-PAGE analysis of the SEC performed with the fraction 5 and 6 of the AFC (Peak 2). The numbers shown close to the horizontal axis identify the fraction number.
M – PageRuler Prestained Protein Ladder (Thermo Scientific™, USA); T_0 – Total extract before induction of protein production; T_1 – Total extract after induction of protein production; L – lysate's soluble fraction before loading into the column; F – flow-through of the AFC; AF3 to AF7 – fractions 3 to 7 of the AFC; S6, S10 to S15, S18, S20, S21, S24, S27 and S28 – fractions of 6, 10 to 15, 18, 20, 21, 24, 27 and 28 of the SEC. The arrows indicate the expected molecular mass of the FLP (34.4 kDa) and the CTP (10.8 kDa).

3.5.2 Cloning and production of the CTP

In order to obtain only the CWBD of LysPhi7951, which in the case of this endolysin corresponds to the CTP, *E. coli* XL1-Blue MRF' was transformed with the smaller ORF starting at the ITSS inserted in the expression vector pIVEX2.3d (pIVEX2.3d::*lysPhi7951CTP*). Transformants were screened through colony PCR and AGE analysis of the amplification products, as before. Clone #4 was selected since it showed the amplification product with the expected molecular weight of 413 bp. Its plasmid was extracted and confirmed by DNA sequencing. *E. coli* CG61 was transformed with pIVEX2.3d::*lysPhi7951CTP* #4, a transformant was chosen and protein production tested. The results confirmed the production and accumulation of the CTP (10.8 kDa, Fig. 3.11A). The C-terminally His_6-tagged LysPhi7951CTP (from pIVEX2.3d::*lysPhi7951CTP*) was purified by AFC (Fig. 3.11B and S5.5A), followed by a SEC of the AFC single peak fractions (Fig. 3.11B and S5.5B). Purified CTP of LysPhi7591 was successfully obtained.

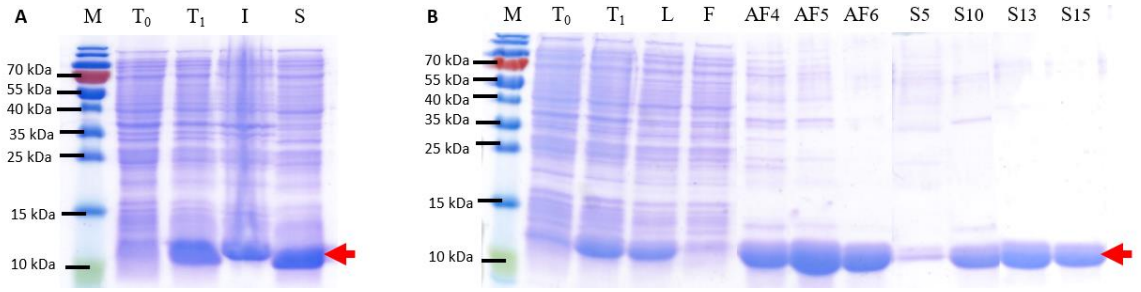


Figure 3.11 - Analysis of the production and purification of LysPhi7951CTP-His$_6$: (A) SDS-PAGE analysis of protein extracts before and after induction of protein production (as in Fig. 3.9); (B) SDS-PAGE analysis of the AFC and SEC purification steps (SEC performed with the fractions 4 to 6 of the AFC).
M – PageRuler Prestained Protein Ladder (Thermo Scientific™, USA); T_0 – Total extract before induction of protein production; T_1 – Total extract after induction of protein production; I – T_1 extract insoluble fraction; S – T_1 extract soluble fraction; L – lysate's soluble fraction before loading into the column; F – flow-through of the AFC; AF4 to AF6 – fractions 4 to 6 of the AFC; S5, S10, S13 and S15 – fractions 5, 10, 13 and 15 of the SEC. The arrow indicates the expected molecular mass of the CTP (10.8 kDa).

3.5.3 Cloning and production of the FLP

Aiming at a higher quality purification of the FLP of LysPhi7951 and to confirm that the CTP is being, in fact, produced from the putative ITSS, *E. coli* XL1-Blue MRF' was transformed with the endolysin gene containing a mutated ITSS (see Table 2.1 for sequence details) cloned in the expression vector pIVEX2.3d (pIVEX2.3d::*lysPhi7951FLP*). The introduced mutation disrupted the ITSS so no translation should be initiated at that site and hence, no CTP should be generated if it was, in fact, being initiated at the ITSS.

Transformants were screened by colony PCR and AGE analysis of the amplification products (as in Fig. 3.1). One clone was selected that showed the amplification product with the expected molecular weight of 423 bp. Its plasmid was extracted and confirmed by DNA sequencing. *E. coli* CG61 was transformed with pIVEX2.3d::*lysPhi7951FLP* #16 and protein production tested. The results showed the production of only one polypeptide with a molecular mass compatible with the FLP (34.4 kDa, Fig. 3.12A), confirming that no CTP was being synthesized and consequently that the bioinformatically predicted ITSS was driving its production. LysPhi7951FLP was purified by AFC, followed by a SEC (Fig. 3.12B and S5.6).

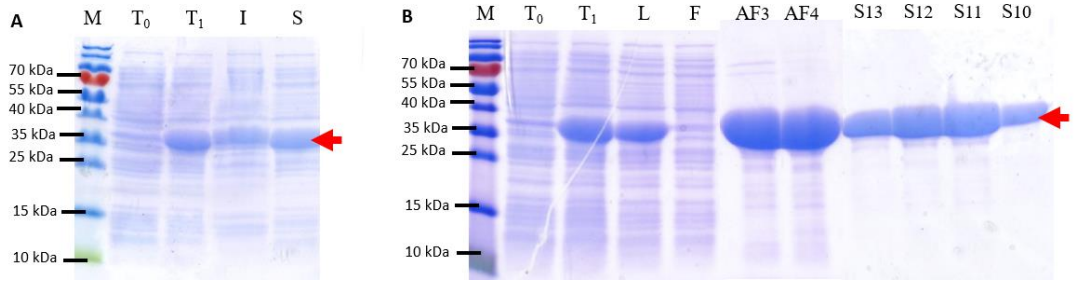


Figure 3.12 - Analysis of the production and purification of LysPhi7951FLP-His$_6$: (A) SDS-PAGE analysis of protein extracts before and after induction of protein production (as in Fig. 3.9); (B) SDS-PAGE analysis of the AFC and SEC purification steps (SEC performed with fractions AF3 and AF4).
M – PageRuler Prestained Protein Ladder (Thermo Scientific™, USA); T_0 – Total extract before induction of protein production; T_1 – Total extract after induction of protein production; I – T_1 extract insoluble fraction; S – T_1 extract soluble fraction; L – lysate's soluble fraction before loading into the column; F – flow-through of the AFC; AF3 to AF4 – fractions 3

to 4 of the AFC; S10 to S13 – fractions 10 to 13 of the SEC. The arrow indicates the expected molecular mass of the FLP (34.4 kDa).

3.5.4 Cloning and protein production from pIVEX2.4d::*lysPhi7951*

The data shown in Figure 3.10E-F suggested that the two polypeptides produced from LysPhi7951 were interacting to form a multimeric complex. If so, we most likely could purify the multimer by AFC of the His_6-LysPhi7951. This would allow the capture and purification of the protein multimer, the His_6-LysPhi7951FLP and the untagged CTP only if associated with the His_6-LysPhi7951FLP. Therefore, *E. coli* XL1-Blue MRF' was transformed with a ligation reaction of pIVEX2.4d and *lysPhi7951* via with *Nco*I and *Xma*I. A colony PCR screening was made and clone #18 showed the expected amplification size of 993 bps, as before. Its plasmid was extracted and confirmed by DNA sequencing. The plasmid pIVEX2.4d::*lysPhi7951* #18 was used to transform *E. coli* CG61 and the induction of protein production resulted in the accumulation of the two expected polypeptides - FLP (36.2 kDa) and CTP (10.0 kDa) - as seen in Figure 3.13A.

Purification by AFC was performed as before and SDS-PAGE analysis of the AFC peak fractions showed the expected FLP and CTP polypeptides. While co-purification was observed, FLP of LysPhi7951 was obtained in much larger quantity than the CTP (Fig. 3.13B and S5.7A). This uneven production of both polypeptides, already visible during production of LysPhi7951-His_6 from pIVEX2.3d::*lysPhi7951* (Fig. 3.10A), seems to be more accentuated when the endolysin is expressed from pIVEX2.4d::*lysPhi7951* (T_1 in Fig. 3.13A). SEC was performed with AF10-11 fractions (Fig. 3.1B) but in these conditions we could not confirm if CTP was interacting with the FLP.

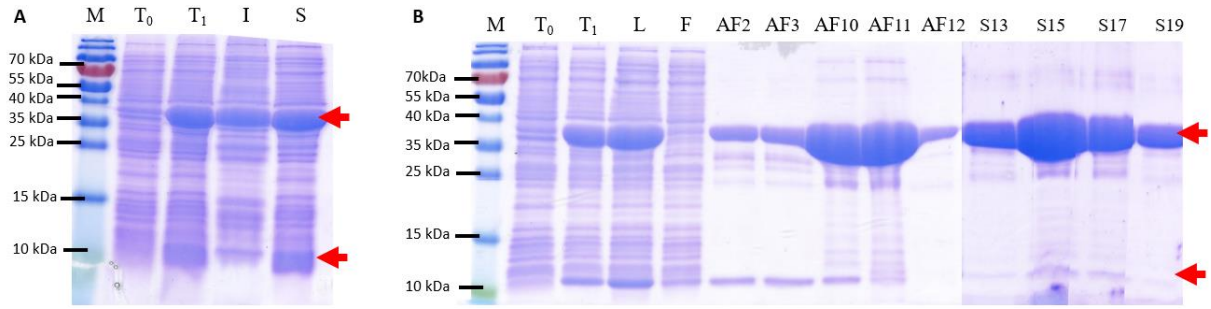


Figure 3.13 - Analysis of the production and purification of His_6-LysPhi7951: (A) SDS-PAGE analysis of protein extracts before and after induction of protein production (as in Fig. 3.9); (B) SDS-PAGE analysis of the AFC and SEC purification steps, respectively (SEC performed with fractions AF10-11).
M – PageRuler Prestained Protein Ladder (Thermo Scientific™, USA); T_0 – Total extract before induction of protein production; T_1 – Total extract after induction of protein production; I – T_1 extract insoluble fraction; S – T_1 extract soluble fraction; L – lysate's soluble fraction before loading into the column; F – flow-through of the AFC; AF2 to AF3 and AF10 to AF12 – fractions 2 to 3 and 10 to 12 of the AFC; S13, S15, S17 and S19 – fractions 13, 15, 17and 19 of the SEC. The arrows indicate the expected molecular mass of the FLP (36.2 kDa) and the CTP (10.0 kDa).

3.5.5 Lytic activity of the endolysin

The lytic activity of LysPhi7951 was tested using the FLP, the CTP, a combination of both and the multimeric complex spotted in dense lawns of *S. thermophilus* cells. We started by optimising the assay using different reaction buffer compositions (Fig. S5.8) and concluded that the best one was 25

mM Hepes with 5 mM $CaCl_2$ at pH 7. Using this reaction buffer, the spotted endolysins produced transparent lysis halos after overnight incubation.

Lysis halos were not observed for the FLP and the CTP, but were clearly visible for the LysPhi7951 complex. For that reason, the purified FLP and CTP were used to try reconstituting the active enzymatic complex – they were combined at different ratios keeping the FLP amount fixed and changing that of the CTP. Halos were observed and, as the proportion of CTP increased, the halos become bigger, more visible, and more similar to the ones observed with the multimeric complex (Fig. 3.14). These results support the hypothesis that the two polypeptides produced from *lysPhi7951* form a multimeric complex to generate the lytic enzyme, with the amount of active multimer increasing as the proportion of CTP augmented.

When we describe ratios here, we are considering:

(1) the difference of molecular mass between the FLP and the CTP of approximately ⅓ (34.4 kDa / 10.8 kDa), so for example a mixture of 1 μg of FLP and 0.33 μg of CTP will produce a molar ratio of approximately 1:1.
(2) the mixtures tested will allow the formation of more multimers than the ones present in the same amount of purified multimer (i.e., a 1:6 mixture having 0.5 μg FLP and 1 μg CTP will have in theory more multimers than 0.5 μg of the purified multimer, because in this the amount 0.5 μg is the sum of α μg FLP with β μg CTP).

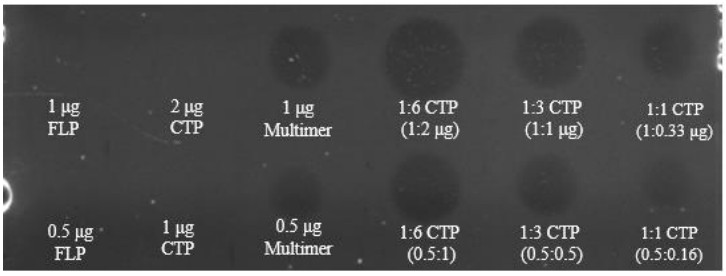


Figure 3.14 - Lytic activity assay of LysPhi7951 against *S. thermophilus* 4078 using the FLP, CTP, FLP/CTP multimer and the combination of the CTP and FLP at different ratios. The FLP/CTP molar ratios are indicated below the halos and reflect the proportion of FLP and CTP subunits in the mixture (mass proportion is shown in parenthesis).

In another assay, the activity of the LysPhi7951 multimer and of the FLP/CTP combination at different ratios was investigated in liquid medium. As in the previous assay, different mixtures of the purified subunits were performed. In accordance with the data of the previous assay, the purified complex shows lytic activity against *S. thermophilus* while the individual polypeptides do not. Additionally, the mixtures of CTP and FLP show activity (in a dose dependent way) suggesting that the multimer is reconstituted. These results also showed that the ratios 1FLP:3CTP and 1:6 behave more similarly to the complex (Fig. 3.15).

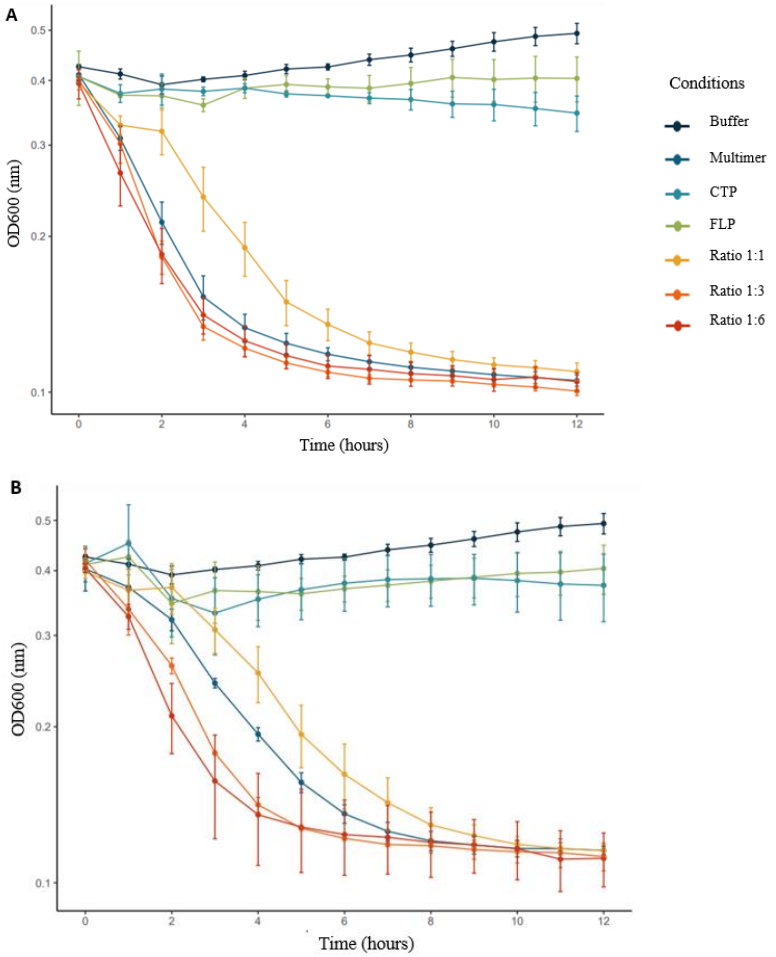


Figure 3.15 - Lytic activity assay of LysPhi7951 in liquid medium using the FLP, CTP, multimer and the combination of different ratios of the CTP and FLP against *S. thermophilus* 4078. OD600 measurements were taken once every hour after addition of the proteins. The vertical lines indicate the standard deviation of three independent experiments. Different protein concentrations were used in (A) and (B), with the ratios constructed based on a fixed amount of FLP and changing the amount of CTP. In (A) 0.625 µg/µL FLP fixed amount, 1.25 µg/µL the CTP, 0.625 µg/µL of the complex, ratio 1:1 with the fixed amount of FLP and 0.208 µg/µL of CTP, ratio 1:3 with 0.625 µg/µL of CTP and ratio 1:6 with 1.25 µg/µL of CTP. In (B) 0.3125 µg/µL FLP fixed amount, 0.625 µg/µL the CTP, 0.3125 µg/µL of the complex, ratio 1:1 with 0.104 µg/µL of CTP, ratio 1:3 with 0.3125 µg/µL of CTP and ratio 1:6 with 0.625 µg/µL of CTP.

3.5.6 Analytical SEC

To gain insight on the composition of LysPhi7951 multimeric complex, the characterization of the recombinant proteins was made by analytical SEC (Fig. 3.16A top). The first striking observation is that the absorbance at 280 nm of FLP, CTP and multimer is very different. This occurs because the CTP has a very low extinction coefficient (2980 $M^{-1}cm^{-1}$ according to Protparam) due to the aminoacidic composition leading to the observed low absorbance. On the contrary, the FLP has a higher absorbance due to the higher extinction coefficient between 52830 (assuming all cysteine residues are reduced) and 53080 $M^{-1}cm^{-1}$ (according to Protparam). The multimer showed an absorbance similar to the CTP, suggesting that in the multimer the ratio of subunits is skewed towards the CTP.

An additional analytic SEC analysis was performed using the 1:6 ratio of FLP:CTP. As seen in Figure 3.16A (bottom), most of this mixture eluted as a single peak (although with a right "shoulder"), with an elution volume (V_e = 16.87 mL) very close to that of the multimer (shown on top), and

consequently having an apparent molecular mass similar to the purified complex. Additionally, SDS-PAGE analysis of the samples composing the peak (Fig. 3.16B) showed that both polypeptides (FLP and CTP) and that the peak "shoulder" probably corresponds to free FLP. Overall, this analysis supported that the multimer is reconstituted after the mixture of the two species, which agreed with the activity data described above. When we tested the lytic activity of the fractions, we observed that the highest activity was in the peak fractions, also supporting the idea that the mixture of the two polypeptides reconstituted the enzymatic complex, since neither of the controls (FLP or CTP) produced lysis halos (Fig. 3.16C).

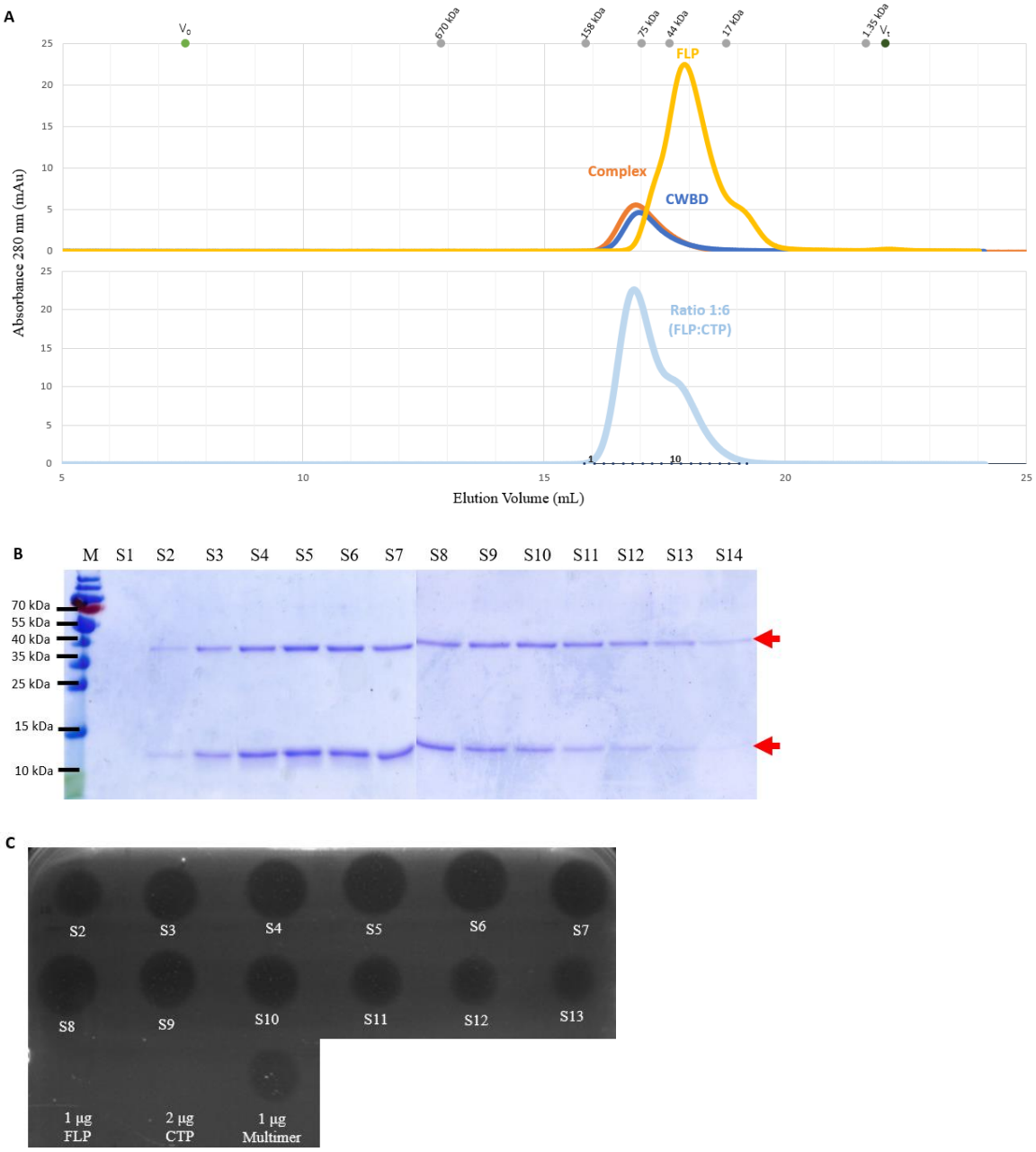


Figure 3.16 - Characterization of the recombinant proteins by analytical SEC. (A) Samples of purified LysPhi7951-His$_6$ were applied to an analytical size exclusion column and run with a flow rate of 0.4 mL/min. The elution profile of the proteins was monitored by absorbance measurements at 280nm. The grey dots indicate the V_e (elution volume) for the protein standard (thyroglobulin at 12.85mL with 670 kDa, γ-globulin at 15.86 mL with 158 kDa, conalbumin at 17.02 mL with 75 kDa, ovalbumin at 17.59 mL with 44 kDa, myoglobin at 18.76 mL with 17 kDa, and vitamin B12 at 21.66 mL with 1.35 kDa). The multimer eluted in a single peak at 16.89 mL (in orange), the FLP eluted at 17.91 mL (in yellow) while the CTP eluted at 16.93 mL (in dark blue). The column void volume (V_0) was 7.55 mL and the column total volume (V_t) was 22.08 mL. In light blue we can see the elution profile of a FLP/CTP mixture that had been incubated for 1 hour at 37°C before the run (peak at 16.87 mL). The numbers shown close to the horizontal axis identify the fraction number and the relevant ones (S1 to S14) were analysed by SDS-PAGE (B). (C) To analyse the lytic activity of the fractions from S2 to S13 were spotted in a dense lawn of *S. thermophilus* 4078 cells.

The K_{av} is related to the size of the protein and the Rs with the apparent radius of a molecule sedimenting under centrifugal force. Both can be correlated with the molecular mass of the proteins, and for that reason the K_{av} and the Rs were estimated for the LysPhi7951 protein species. Through the plot

of K_{av} or Rs for each protein of the standard against the corresponding logarithmic molecular mass (Fig. 3.17 using the equations 2.1 and 2.2), the molecular mass of the proteins was inferred. FLP molecular mass was estimated to be 35.0 kDa, not far from the theorical value of 34.4 kDa. On the contrary, the CTP had a molecular mass of 75.7 kDa, which is significantly larger than the theorical one (10.8 kDa), and suggestive of interaction between seven CTP monomers. The molecular mass estimated through the SEC data for the complex was 78.1 kDa (Table 3.2).

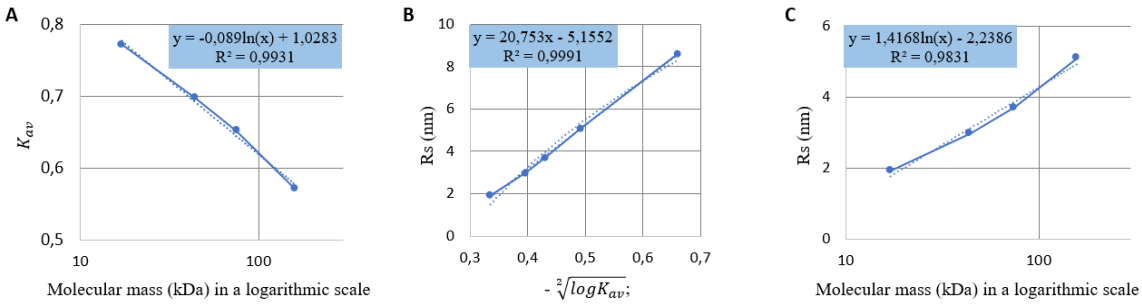


Figure 3.17 - Plots that allowed estimation of the apparent molecular masses of the different LysPhi7951 species based on their elution volumes. Protein standard (V_e, molecular mass, Rs): thyroglobulin (12.85 mL, 670 kDa, 8.6 nm), γ-globulin (15.86 mL, 158 kDa, 5.1 nm), conalbumin (17.02 mL , 75 kDa, 2.7 nm), ovalbumin (17.59 mL, 44 kDa, 1.91 nm) and myoglobin (18.76 mL, 17 kDa). Using this information, the following plots were constructed: (A) K_{av} vs. molecular mass; (B) Rs vs. $-\sqrt[2]{logK_{av}}$; (C) Rs vs. molecular mass. These plots and their fitting equations were used to infer in the molecular mass calculations.

Table 3.2 - V_e, K_{av}, Rs and the apparent molecular masses of the LysPhi7951 proteins.

	Theoretical molecular mass (kDa)	Ve (mL)	K_{av}	Molecular mass (kDa) estimated from K_{av} vs. molecular mass plot (Fig. 3.25A)	Rs calculated (nm)	Molecular mass (kDa) estimated from Rs vs. molecular mass plot (Fig.3.25C)
FLP	34.4	17.91	0.713	34.6	2.80	35.0
Multimer	?	16.89	0.643	76.0	3.94	78.1
CTP	10.8	16.93	0.646	73.7	3.89	75.7

Indicated with (?) is the unknown molecular mass.

This analysis allowed the generation of hypothesis concerning the composition of the LysPhi7951 multimer. We should note that: (1) the CTP seems to oligomerize; (2) the molecular mass inferred for the FLP is not far from the expect for its monomer; (3) the low absorbance at 280 nm of the LysPhi7951 complex strongly suggests a multimer with a higher number of CTP subunits compared to FLP; and (4) the multimeric endolysins reported to date contain typically one FLP associated to n CTP (Dunne *et al*., 2014; Dunne *et al*., 2016; Proença *et al*., 2015; Zhou *et al*., 2020).

Given our data, we propose that LysPhi7951 multimer consists of one FLP associated with four to six CTP. Considering the theoretical and determined (by the analytic SEC) molecular masses:

(1) The complex has a determined molecular mass of 78.1 kDa. This could correspond to one FLP with four CTP (34.4 kDa + 4 ×10.8 kDa = 77.6 kDa) or one FLP with five CTP (34.4 kDa + 5 ×10.8 kDa = 88.4 kDa);
(2) If the complex has only one FLP then, 78.1 kDa - 35.0 kDa = 43.1 kDa would correspond to the CTP subunits. Considering that the molecular mass of the CTP monomer is 10.8 kDa, it indicated approximately four CTP (= 43.1/10.8);
(3) The determined molecular mass of the CTP was 75.7 kDa indicating the association of seven CTP (75.7 kDa : 10.8 kDa ≈ 7). Considering that the FLP is composed also by a CTP then we would have one FLP (composed of one CD and one CTP) with six CTP;

3.5.7 Conclusions

We have demonstrated that the *lysPhi7951* gene produces two polypeptides using an *in frame* ITSS and that the two polypeptides associate. As reported previously for Lys170 (Proença *et al*., 2015), the presence of the ITSS in LysPhi7951 provides a mechanism to increase the number of CWBDs (which in this case corresponds to the CTP) present in the active endolysin, most likely optimizing the enzyme's affinity to its substrate. In fact, oligomerization of the CWBD subunits in association to the FLP is necessary to generate the active endolysin against *S. thermophilus*. Based on our biochemical and functional analysis, we propose that the LysPhi7951 complex is composed of one FLP associated with 4 to 6 CTP subunits.

3.6. Main conclusions and future perspectives

In this work, we have demonstrated that from the five endolysin genes tested we were able to clone and saw protein production only in three (*lysLW32*, *lysJavan488* and *lysPhi7951*). Those three use ITSSs to generate two different polypeptides from one single gene using two *in frame* ORFs that share a stop codon. We also showed that heteromerization generates an active endolysin, as seen for LysPhi7951. However, not in all cases is heteromerization occurring between the two produced polypeptides, as seen with LysJavan488 at least in the conditions tested.

New questions are raised: (1) since not in all cases is multimerization occurring between the endolysin polypeptides, what is the biological relevance of the production of FLP and CTP in LysJavan488?; (2) can LysJavan488 be used in a clinical setting to kill *S. pyogenes*?; and (3) what is the exact composition of the LysPhi7951 multimer? Regarding the last question, we have recently initiated analysis by native mass spectrometry of the LysPhi7951 complex and of its FLP and CTP subunits to disclose the stoichiometry of the endolysin multimer.

4. References

Abedon, S., Kuhl, S., Blasdel, B. and Kutter, E., 2011. Phage treatment of human infections. *Bacteriophage*, 1(2), pp.66-85.

Aminov, R., 2009. The role of antibiotics and antibiotic resistance in nature. *Environmental Microbiology*, 11(12), pp.2970-2988.

Cabré, F., Canela, E. and Canela, M., 1989. Accuracy and precision in the determination of stokes radii and molecular masses of proteins by gel filtration chromatography. *Journal of Chromatography A*, 472, pp.347-356.

Carter J., Saunders V. *Virology: Principles and Applications.* 2nd ed. Wiley & Sons Inc.; New York, NY, USA: 2013.

Celia, L., Nelson, D. and Kerr, D., 2008. Characterization of a bacteriophage lysin (Ply700) from *Streptococcus uberis. Veterinary Microbiology*, 130(1-2), pp.107-117.

Chung, C., Niemela, S. and Miller, R., 1989. One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *Proceedings of the National Academy of Sciences*, 86(7), pp.2172-2175.

Cook, D., Gysemans, C. and Mathieu, C., 2018. *Lactococcus lactis* As a Versatile Vehicle for Tolerogenic Immunotherapy. *Frontiers in Immunology*, 8.

Cui, Y., Xu, T., Qu, X., Hu, T., Jiang, X. and Zhao, C., 2016. New Insights into Various Production Characteristics of *Streptococcus thermophilus* Strains. *International Journal of Molecular Sciences*, 17(10), p.1701.

Davies, J. and Davies, D., 2010. Origins and Evolution of Antibiotic Resistance. *Microbiology and Molecular Biology Reviews*, 74(3), pp.417-433.

de Melo, A., Levesque, S. and Moineau, S., 2018. Phages as friends and enemies in food processing. *Current Opinion in Biotechnology*, 49, pp.185-190.

Dunne, M., Leicht, S., Krichel, B., Mertens, H., Thompson, A., Krijgsveld, J., Svergun, D., Gómez-Torres, N., Garde, S., Uetrecht, C., Narbad, A., Mayer, M. and Meijers, R., 2016. Crystal Structure of the CTP1L Endolysin Reveals How Its Activity Is Regulated by a Secondary Translation Product. *Journal of Biological Chemistry*, 291(10), pp.4882-4893.

Dunne, M., Mertens, H., Garefalaki, V., Jeffries, C., Thompson, A., Lemke, E., Svergun, D., Mayer, M., Narbad, A. and Meijers, R., 2014. The CD27L and CTP1L Endolysins Targeting Clostridia Contain a Built-in Trigger and Release Factor. *PLoS Pathogens*, 10(7), p.e1004228.

Fernandes, S. and São-José, C., 2016. More than a hole: the holin lethal function may be required to fully sensitize bacteria to the lytic action of canonical endolysins. *Molecular Microbiology*, 102(1), pp.92-106.

Fernandes, S. and São-José, C., 2018. Enzymes and Mechanisms Employed by Tailed Bacteriophages to Breach the Bacterial Cell Barriers. *Viruses*, 10(8), p.396.

Fernandes, S., Proença, D., Cantante, C., Silva, F., Leandro, C., Lourenço, S., Milheiriço, C., de Lencastre, H., Cavaco-Silva, P., Pimentel, M. and São-José, C., 2012. Novel Chimerical Endolysins with Broad Antimicrobial Activity Against Methicillin-Resistant *Staphylococcus aureus*. *Microbial Drug Resistance*, 18(3), pp.333-343.

Ferretti, J., Stevens, D. and Fischetti, V., 2016. *Streptococcus pyogenes*. University of Oklahoma Health Sciences Center.

Gordillo Altamirano, F. and Barr, J., 2019. Phage Therapy in the Postantibiotic Era. *Clinical Microbiology Reviews*, 32(2).

Górski, A., Międzybrodzki, R., Łobocka, M., Głowacka-Rutkowska, A., Bednarek, A., Borysowski, J., Jończyk-Matysiak, E., Łusiak-Szelachowska, M., Weber-Dąbrowska, B., Bagińska, N., Letkiewicz, S., Dąbrowska, K. and Scheres, J., 2018. Phage Therapy: What Have We Learned?. *Viruses*, 10(6), p.288.

Gutiérrez, D., Fernández, L., Rodríguez, A. and García, P., 2018. Are Phage Lytic Proteins the Secret Weapon To Kill *Staphylococcus aureus*?. *mBio*, 9(1).

Hoffmann, C., Leis, A., Niederweis, M., Plitzko, J. and Engelhardt, H., 2008. Disclosure of the mycobacterial outer membrane: Cryo-electron tomography and vitreous sections reveal the lipid bilayer structure. *Proceedings of the National Academy of Sciences*, 105(10), pp.3963-3967.

Kanwal, S. and Vaitla, P., 2020. *Streptococcus Pyogenes*. [online] Ncbi.nlm.nih.gov. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK554528/> [Accessed 7 December 2020].

Lavelle, K., Goulet, A., McDonnell, B., Spinelli, S., Sinderen, D., Mahony, J. and Cambillau, C., 2020. Revisiting the host adhesion determinants of *Streptococcus thermophilus* siphophages. *Microbial Biotechnology*, 13(6), pp.1765-1779.

Laxminarayan, R., Matsoso, P., Pant, S., Brower, C., Røttingen, J.-A., Klugman, K., & Davies, S., 2016. Access to effective antimicrobials: a worldwide challenge. *The Lancet*, 387(10014), pp. 168–175.

Lecomte, X., Gagnaire, V., Lortal, S., Dary, A. and Genay, M., 2016. *Streptococcus thermophilus*, an emerging and promising tool for heterologous expression: Advantages and future trends. *Food Microbiology*, 53, pp.2-9.

Lesho, E. and Laguio-Vila, M., 2019. The Slow-Motion Catastrophe of Antimicrobial Resistance and Practical Interventions for All Prescribers. *Mayo Clinic Proceedings*, 94(6), pp.1040-1047.

Lin, D., Koskella, B. and Lin, H., 2017. Phage therapy: An alternative to antibiotics in the age of multi-drug resistance. *World Journal of Gastrointestinal Pharmacology and Therapeutics*, 8(3), p.162.

Loc-Carrillo, C. and Abedon, S., 2011. Pros and cons of phage therapy. *Bacteriophage*, 1(2), pp.111-114.

Markakiou, S., Gaspar, P., Johansen, E., Zeidan, A. and Neves, A., 2020. Harnessing the metabolic potential of *Streptococcus thermophilus* for new biotechnological applications. *Current Opinion in Biotechnology*, 61, pp.142-152.

Martinović, A., Cocuzzi, R., Arioli, S. and Mora, D., 2020. *Streptococcus thermophilus*: To Survive, or Not to Survive the Gastrointestinal Tract, That Is the Question!. *Nutrients*, 12(8), p.2175.

McDonnell, B., Hanemaaijer, L., Bottacini, F., Kelleher, P., Lavelle, K., Sadovskaya, I., Vinogradov, E., Ver Loren van Themaat, E., Kouwen, T., Mahony, J. and Sinderen, D., 2020. A cell wall-associated polysaccharide is required for bacteriophage adsorption to the *Streptococcus thermophilus* cell surface. *Molecular Microbiology*, 114(1), pp.31-45.

McGowan, S., Buckle, A., Mitchell, M., Hoopes, J., Gallagher, D., Heselpoth, R., Shen, Y., Reboul, C., Law, R., Fischetti, V., Whisstock, J. and Nelson, D., 2012. X-ray crystal structure of the streptococcal specific phage lysin PlyC. *Proceedings of the National Academy of Sciences*, 109(31), pp.12752-12757.

Mierau, I. and Kleerebezem, M., 2005. 10 years of the nisin-controlled gene expression system (NICE) in *Lactococcus lactis*. *Applied Microbiology and Biotechnology*, 68(6), pp.705-717.

Mohr, K., 2016. History of Antibiotics Research. *Current Topics in Microbiology and Immunology*, pp.237-272.

Momand, J., Magdziarz, P., Feng, Y., Jiang, D., Parga, E., Celis, A., Denny, E., Wang, X., Phillips, M., Monterroso, E., Kane, S. and Zhou, F., 2017. t-Darpp is an elongated monomer that binds calcium and is phosphorylated by cyclin-dependent kinases 1 and 5. *FEBS Open Bio*, 7(9), pp.1328-1337.

Nelson, D., Schmelcher, M., Rodriguez-Rubio, L., Klumpp, J., Pritchard, D., Dong, S. and Donovan, D., 2012. Endolysins as Antimicrobials. *Advances in Virus Research*, pp.299-365.

Oliveira, H., Melo, L., Santos, S., Nobrega, F., Ferreira, E., Cerca, N., Azeredo, J. and Kluskens, L., 2013. Molecular Aspects and Comparative Genomics of Bacteriophage Endolysins. *Journal of Virology*, 87(8), pp.4558-4570.

Poranen, M., Daugelavičius, R. and Bamford, D., 2002. Common Principles in Viral Entry. *Annual Review of Microbiology*, 56(1), pp.521-538.

Pritchard, D., Dong, S., Baker, J. and Engler, J., 2004. The bifunctional peptidoglycan lysin of *Streptococcus agalactiae* bacteriophage B30. *Microbiology*, 150(7), pp.2079-2087.

Proença, D., Velours, C., Leandro, C., Garcia, M., Pimentel, M. and São-José, C., 2015. A two-component, multimeric endolysin encoded by a single gene. *Molecular Microbiology*, 95(5), pp.739-753.

São-José, C., 2018. Engineering of Phage-Derived Lytic Enzymes: Improving Their Potential as Antimicrobials. *Antibiotics*, 7(2), p.29.

São-José, C., Parreira, R., Vieira, G. and Santos, M., 2000. The N-Terminal Region of the Oenococcus oeniBacteriophage fOg44 Lysin Behaves as a Bona Fide Signal Peptide in *Escherichia coli* and as a cis-Inhibitory Element, Preventing Lytic Activity on Oenococcal Cells. *Journal of Bacteriology*, 182(20), pp.5823-5831.

Schleifer, K. and Kandler, O., 1972. Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriological Reviews*, 36(4), pp.407-477.

Schmelcher, M., Donovan, D. and Loessner, M., 2012. Bacteriophage endolysins as novel antimicrobials. *Future Microbiology*, 7(10), pp.1147-1171.

Schmelcher, M., Korobova, O., Schischkova, N., Kiseleva, N., Kopylov, P., Pryamchuk, S., Donovan, D. and Abaev, I., 2012. *Staphylococcus haemolyticus* prophage ΦSH2 endolysin relies on cysteine, histidine-dependent amidohydrolases/peptidases activity for lysis 'from without'. Journal of Biotechnology, 162(2-3), pp.289-298.

Shearman, C., Jury, K. and Gasson, M., 1994. Controlled expression and structural organization of a *Lactococcus lactis* bacteriophage lysin encoded by two overlapping genes. *Applied and Environmental Microbiology*, 60(9), pp.3063-3073.

Silhavy, T., Kahne, D. and Walker, S., 2010. The Bacterial Cell Envelope. *Cold Spring Harbor Perspectives in Biology*, 2(5), pp.a000414-a000414.

Singh, S., Young, K. and Silver, L., 2017. What is an "ideal" antibiotic? Discovery challenges and path forward. *Biochemical Pharmacology*, 133, pp.63-73.

Song, A., In, L., Lim, S. and Rahim, R., 2017. A review on *Lactococcus lactis*: from food to factory. *Microbial Cell Factories*, 16(1).

Stetefeld, J., McKenna, S. and Patel, T., 2016. Dynamic light scattering: a practical guide and applications in biomedical sciences. *Biophysical Reviews*, 8(4), pp.409-427.

Ventola, C., 2015. The Antibiotic Resistance Crisis Part 1: Causes and Threats. *Pharmacy and Therapeutics*, 40(4), pp.278-283.

Vollmer, W., Blanot, D. and De Pedro, M., 2008. Peptidoglycan structure and architecture. *FEMS Microbiology Reviews*, 32(2), pp.149-167.

White, D., Drummond, J. and Fuqua, C., 2012. *The Physiology And Biochemistry Of Prokaryotes*. 4th ed. Oxford University Press.

WHO, 2017. [online] Available at: < https://www.who.int/medicines/publications/WHO-PPL-Short_Summary_25Feb-ET_NM_WHO.pdf> [Accessed 20 December 2020]

Wilkening, R. and Federle, M., 2017. Evolutionary Constraints Shaping *Streptococcus pyogenes* –Host Interactions. *Trends in Microbiology*, 25(7), pp.562-572.

Xu, X., Zhang, D., Zhou, B., Zhen, X. and Ouyang, S., 2021. Structural and biochemical analyses of the tetrameric cell binding domain of Lys170 from enterococcal phage F170/08. *European Biophysics Journal*, 50(5), pp.721-729.

Young, R. Phage lysis: Three steps, three choices, one outcome. *J. Microbiol*. 2014, 52, 243–258.

Zhou, B., Zhen, X., Zhou, H., Zhao, F., Fan, C., Perčulija, V., Tong, Y., Mi, Z. and Ouyang, S., 2020. Structural and functional insights into a novel two-component endolysin encoded by a single gene in *Enterococcus faecalis* phage. *PLOS Pathogens*, 16(3), p.e1008394.

5. Supplementary Data

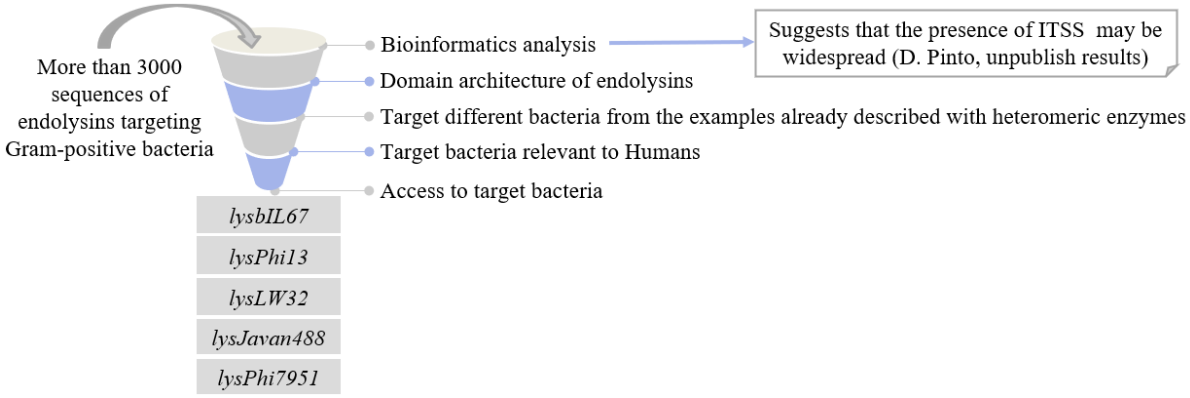


Figure S5.1 - Schematic representation of the methodology underlying the survey, bioinformatics analysis and selection of endolysins targeting Gram-positive bacteria encoded by genes carrying putative ITSS (D. Pinto, unpublished results). The five indicated endolysin genes were selected for heterologous expression in this work, with the goal of confirming experimentally the ITSS predictions. For this selection we took into account: (1) the composition of endolysins regarding the number and type of conserved domains, in order to increase the diversity of our sample; (2) if the enzymes targeted bacterial genera different from those targeted by the previously described heteromeric enzymes produced through this mechanism; (3) if the targeted bacteria were of relevance to humans (either in a clinical setting or in industrial environment); and (4) if we had access to both the endolysin sequence and the bacteria they target.

Table S5.1- Primers used for the endolysin gene

	Name	Sequence (5'→3')	Melting temperature (°C)	GC-content (%)
lysbIL67	NcoI-*lysBIL67*	ATGCTCCCATGGAAATATCTCAA AACGGTTTGAACT	60	38.9
	lysBIL67-XmaI	ATGCTACCCGGGTTTTACTTCTCC TGAAGCCATG	65	50.0
	SpeI-NcoI-*lysBIL67*	ATGCACTAGTCCTCCCATGGAAA TATCTCAAAACGGTTTGAACT	55	41.0
	lysBIL67-XmaI-XhoI	ATGCTACTCGAGCCCGGGTTTTAC TTCTCCTGAAGCCATG	54	32.0
lysJavan488	NcoI-*lysJavan488*	ATGCTCCCATGGCCTTTTTAGATA AAATTAAACAAGGC	50	36.8
	lysJavan488-XmaI	ATGCTACCCGGGATTTAATTTACC CCAAAGGCTGATG	65	45.9
lysPhi13	NcoI-*lysPhi13*	ATGCTCCCATGGAAACATACAGT GAAGCAAGAGC	61	47.1
	lysPhi13-XmaI	ATGCTACCCGGGAAACACTTCTTT CACAATCAATCTC	63	43.2
lysLW32	NcoI-*lysLW32CWBD*	ATCGCTCCATGGATTTAATTTTCA CTGTAGACACTAAACG	56	38.0
lysPhi7951	NcoI-*lysPhi7951*	ATCGATCCATGGAAAAAGGTGAT TACTTTATCG	56	36.3
	lysPhi7951-XmaI	ATCGATCCCGGGTTTTCCAGTTTG ACCTTGTGTAGC	66	50.0
	NcoI-*lysPhi7951CWBD*	ATCGCTCCATGGATTATGTAGTTC GAAGCGAAAG	60	44.0

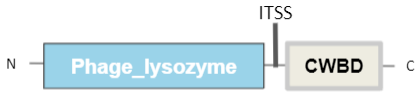
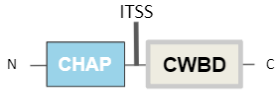
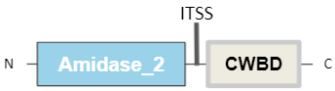
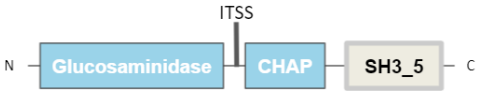
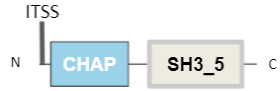
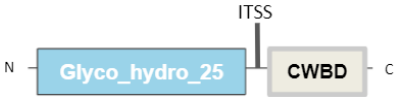
Table S5.2 - Primers complementary to pIVEX2.3d, pIVEX2.4d , pET21a$^+$, pBluescript II KS$^+$ and pBluescript II SK$^+$

	Name	Sequence (5'→3')	Melting temperature (°C)	GC-content (%)
pIVEX2.3d pIVEX2.4d pET21a$^+$	T7 promoter	TAATACGACTCACTATAGGG	46	40.0
	pIVEXRev	ATTGCTCAGCGGTGGCAG	61	61.1
pBluescript II KS$^+$ pBluescript II SK$^+$	pUC-down	AGGCGATTAAGTTGGGTAAC	54	45
	pUC-up	TCACACAGGAAACAGCTATG	54	45

Table S5.3 - Primers used for the mutagenesis of *lysPhi7951*

	Name	Sequence (5'→3')	Melting temperature (°C)	GC-content (%)
lysPhi7951	LysP7951_ITSSminus_frw	GAAGAAGAAGAAAATCTGG ATTATGTAGTTCG	63.1	34.4
	LysP7951_ITSSminus_rev	CTACATAATCCAGATTTTCT TCTTCTTCTTTCTTATCTAC	65.4	30.0

Table S5.4 - Length of the endolysin gene and the produced proteins (retrieved from the PhaLP database).

Endolysin designation		Composition	Gene length (bp)	Protein (kDa / pI)	Results
LysbIL67	Full-length		678	25.6 / 8.63	Cloning failed
	CTP		213	8.3 / 9.35	
LysPhi13	Full-length		753	29.1 / 9.10	Production failed
	CTP		330	12.6 / 9.71	
LysLW32	Full-length		741	28.0 / 6.12	Production insoluble
	CTP		180	7.1 / 9.10	
LysJavan488	Full-length		1200	44.2 / 4.67	Purified and analysed
	CTP		732	26.7 / 4.53	
LysPhi9751	Full-length		903	32.9 / 4.50	Purified and analysed
	CTP		279	9.8 / 4.91	

The filled blue boxes represent the N-terminal CDs and the filled grey boxes represent the CWBDs. Inside each CD box are indicated the Pfam domains.

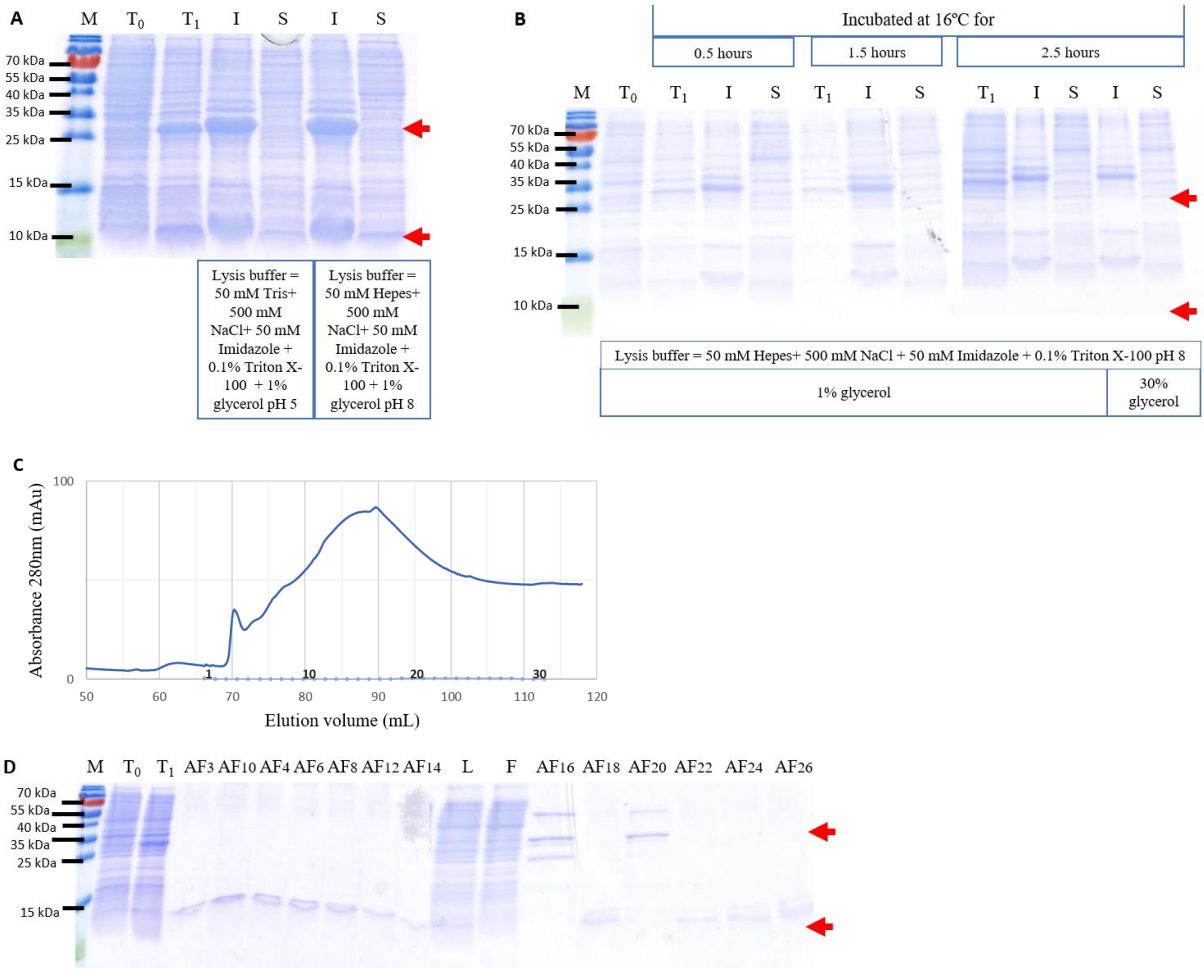


Figure S5.2 - Different conditions tested to improve solubility in the induction of protein production at 16ºC from pIVEX2.3d::*lysLW32* in LB buffered with 0.2 M sodium phosphate at pH 8 and supplemented with 0.5 M of sorbitol: SDS-PAGE analysis of after overnight induction (A) or 1 hour, 2 hours or 3 hours of induction (B). The different lysis buffers are indicated in the figures. In (C) we have the analysis of the eluting profile of an AFC from the lysate's soluble fraction and in (D) the SDS-PAGE analysis of the relevant fractions of the AFC.
M - PageRuler Prestained Protein Ladder (Thermo Scientific™, USA); T_0 – total protein extract before induction; T_1 – total protein extract after induction; I – T_1 extract insoluble fraction; S – T_1 extract soluble fraction; L - lysate's soluble fraction before loading into the column; F - flow through; AF3 to AF24 - fractions 3 to 24 of the AFC. The arrows indicate the expected molecular mass of the FLP (29.3 kDa) and the C-terminal product (8.4 kDa) and the numbers above the horizontal axis in (C) identify the fractions number.

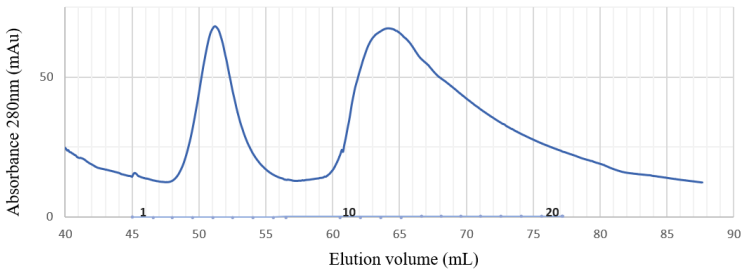


Figure S5.3 – AFC eluting profile of the proteins produced from pIVEX2.3d::*lysLW32* under denaturing conditions. The numbers present above the horizontal axis identify the fraction number and the relevant ones were analysed by SDS-PAGE (Fig. 3.4).

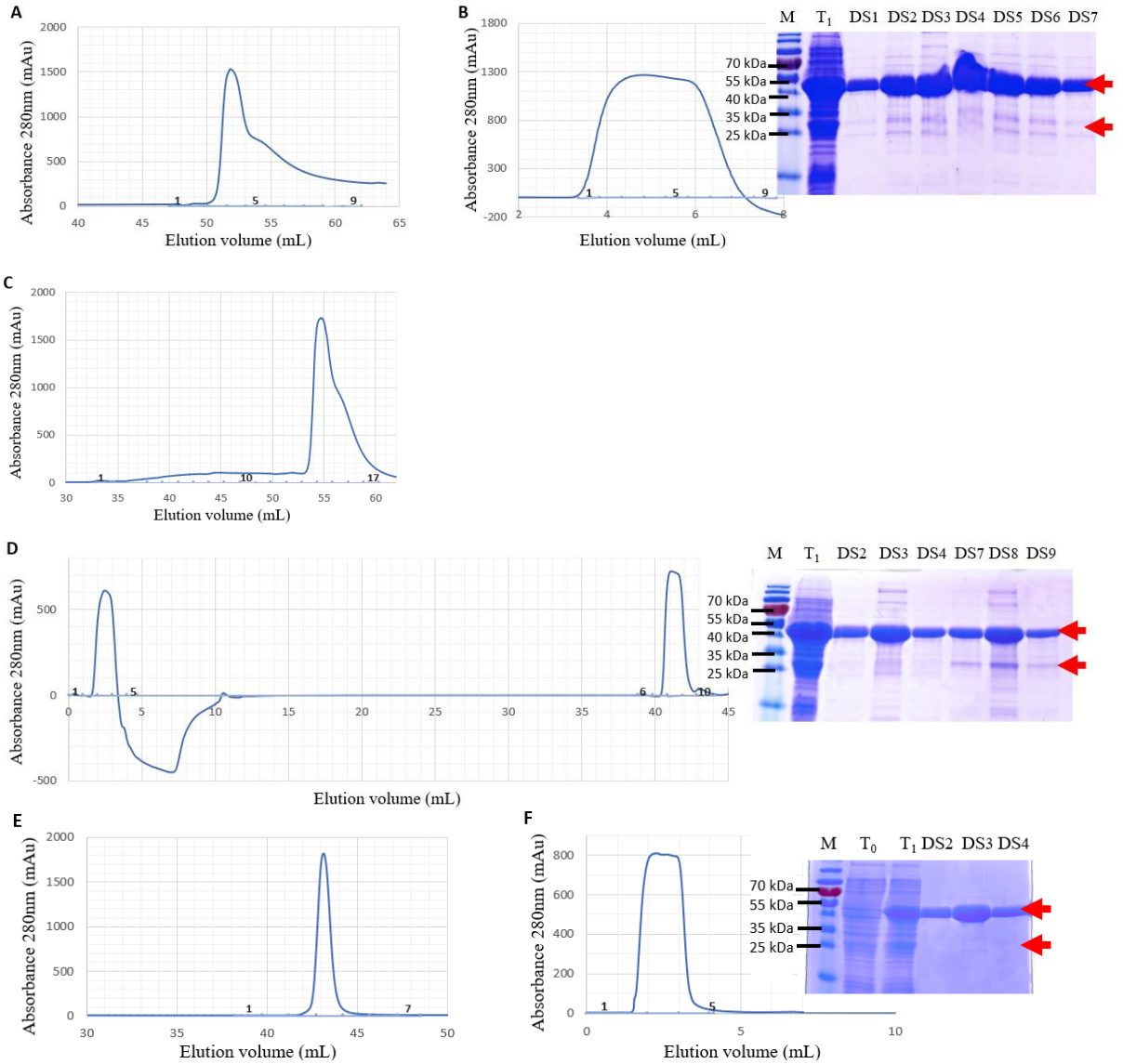


Figure S5.4 - Eluting profile of the proteins produced from pIVEX2.4d::*lysJavan488*: (A), (C) and (E) correspond to the AFC profiles; (B), (D) and (F) to the desalting with the analysis of the relevant fractions by SDS-PAGE; (A) and (B) the binding buffer with 50 mM Hepes, 500 mM NaCl, 50 mM imidazole, 1 mM TCEP, 1% glycerol, 0.1% Triton X-100 at pH 7; (C) and (D) the binding buffer did not have TCEP; (E) and (F) the binding buffer did not have TCEP and Triton X-100. The numbers shown above the horizontal axis identify the fraction number and the relevant ones were analysed by SDS-PAGE. M - PageRuler Prestained Protein Ladder (Thermo Scientific™, USA); T_0 - total protein extract before induction; T_1 - total protein extract after induction; DS1 to DS9 - fractions 1 to 9 of the desalting; The arrows indicate the expected size of the FLP (47.4 kDa) and of the CTP (27.1 kDa).

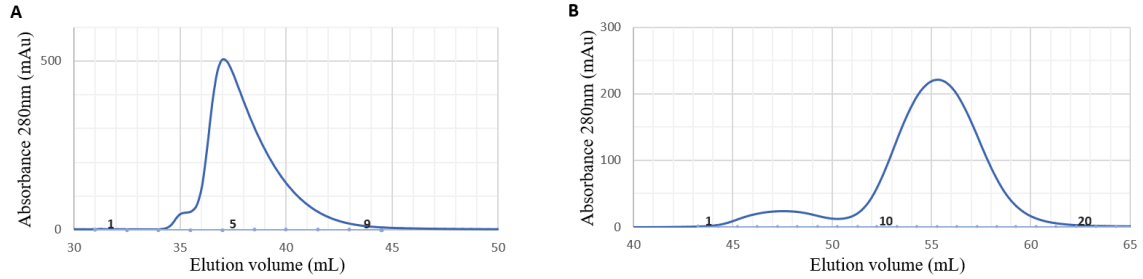


Figure S5.5 - Eluting profiles of the proteins produced from pIVEX2.3d::*lysPhi7951CTP*: (A) profile from AFC and (B) from SEC of the fraction 4 to 6 of the AFC. The numbers shown above the horizontal axis identify the fraction number and the relevant ones were analysed by SDS-PAGE (Fig. 3.11).

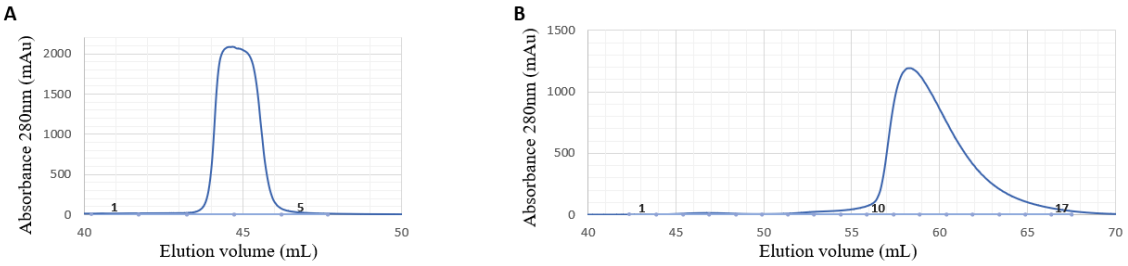


Figure S5.6 - Eluting profile of the proteins produced from pIVEX2.3d::*lysPhi7951FLP*: (A) elution profile from AFC and (B) from SEC of AF3-4. The numbers shown above the horizontal axis identify the fraction number and the relevant ones were analysed by SDS-PAGE (Fig. 3.12).

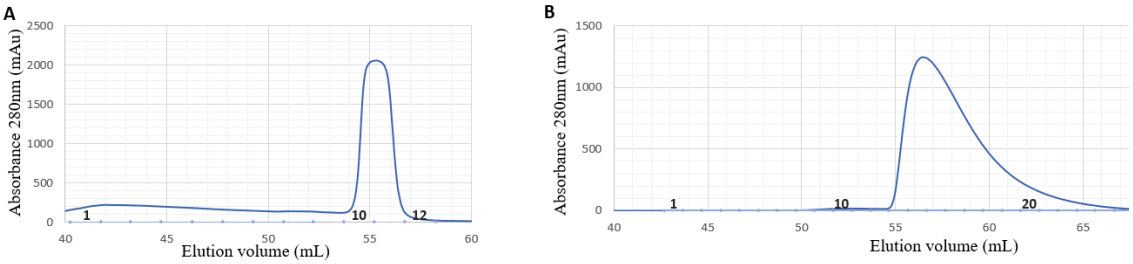


Figure S5.7 - Eluting profile of the proteins produced from pIVEX2.4d::*lysPhi7951*: (A) elution profile from AFC and (B) from SEC of the fractions AF10-11. The numbers shown above the x axis identify the fraction number and the relevant ones were analysed by SDS-PAGE (Fig. 3.13).

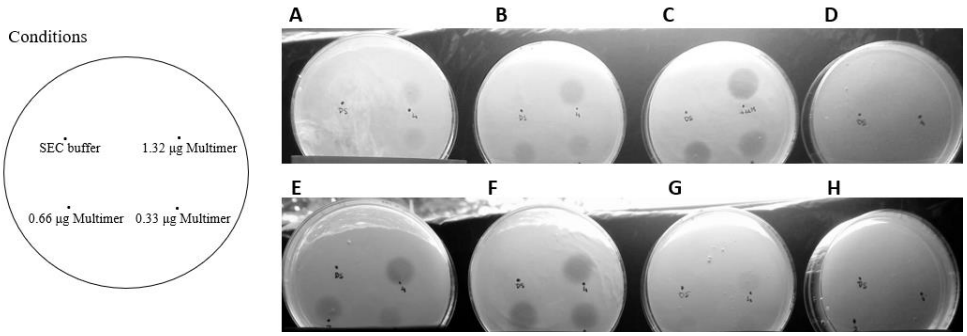


Figure S5.8 – Improvement of the lysis conditions for LysPhi7951. The buffers tested were: (A) 25 mM Hepes, 5 mM $CaCl_2$, pH 8; (B) 25m M Hepes, 5 mM $CaCl_2$, 250 mM NaCl, pH 7; (C) 25 mM Hepes, 5 mM $CaCl_2$, pH 7; (D) 25 mM Hepes, 5 mM $CaCl_2$, 500 mM NaCl, pH 7; (E) 25 mM Hepes, 5 mM $CaCl_2$, 500 mM NaCl, pH 7, 1 mM TCEP; (F) 25mM Pipes, 5 mM $CaCl_2$, pH 6; (G) 50 mM Sodium acetate/acetic acid buffer, 5 mM $CaCl_2$, pH 5; and (H) 50 mM Sodium acetate/acetic acid buffer, 5 mM $CaCl_2$, pH 4.