

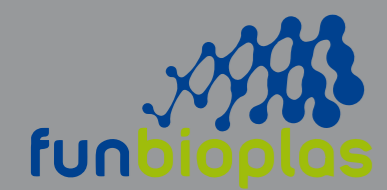


João Fernando Machado Alves

Development of a molecular biology platform for the production of antimicrobial peptides

**Universidade do Minho**  
Escola de Ciências





**FCT** Fundação  
para a Ciência  
e a Tecnologia





**Universidade do Minho**  
Escola de Ciências

João Fernando Machado Alves

**Development of a molecular biology  
platform for the production of  
antimicrobial peptides**

Dissertação de Mestrado  
Mestrado em Genética Molecular

Trabalho efetuado sob a orientação do  
**Doutor Raul Machado**  
**Doutor André Costa**

Outubro 2019

## DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos. Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada. Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.



**Atribuição-NãoComercial-SemDerivações**  
**CC BY-NC-ND**

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

## ACKNOWLEDGEMENTS

I cannot express enough thanks to my supervisors for their continued support and encouragement: Dr. Raul Machado and Dr. André Costa. I offer my sincere appreciation for the learning opportunities provided.

The completion of this project could not have been accomplished without the support of my classmates and lab colleagues, Catarina, Mafalda, Rita, Erica, Diana, Margarida, Maria and Marisol.

Thanks to my parents as well, Mr. António Alves and Mrs. Julia Machado. In addition, special thanks to the rest of my friends and loved ones, for the solace and/or game nights.

Este trabalho é financiado por fundos nacionais através da FCT – Fundação para a Ciência e a Tecnologia, I.P., no âmbito do projeto "FunBioPlas - Novel synthetic biocomposites for biomedical devices" com referência ERA-IB-2-6/0004/2014.



## **STATEMENT OF INTEGRITY**

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

## RESUMO - Desenvolvimento de uma plataforma de biologia molecular para a produção de péptidos antimicrobianos

Os peptídeos antimicrobianos (AMPs) são importantes candidatos a antibióticos devido à baixa ou até mesmo ausente capacidade de induzir o desenvolvimento de resistências pelos microrganismos. Este facto é importante devido ao aparecimento de bactérias multirresistentes. O presente trabalho tem por objetivo a criação de uma plataforma para a produção e purificação de AMPs explorando *elastin-like polypeptides* (ELPs) como método de purificação não cromatográfica e inteínas como mecanismo de auto-clivagem para o isolamento de AMPs.

Como estratégia geral de clonagem, foram desenhados diferentes adaptadores (total de nove) que foram utilizados para construir quatro vetores de expressão. Dois vetores, denominados de pA1C e pA3N, foram desenhados para inserir o ELP (A60) e diferentes AMPs (ABP-CM4; HAMP; BMAP-28) em cada um dos terminais da inteína. O vetor pA1C explora a utilização de uma inteína  $\Delta$ I-SM mutada num aminoácido para bloquear a clivagem no N-terminal enquanto o vetor pA3N usa a inteína  $\Delta$ I-SM mutada num aminoácido para bloquear a clivagem no C-terminal. Ambas inteínas são mutantes da inteína *Mtu* reca. Como prova de conceito, o péptido ABP-CM4 foi inserido no C-terminal do vetor pA1C e no N-terminal do vetor pA3N. Após obtenção das construções genéticas, as proteínas de fusão foram expressas em *E. coli* BL21 (DE3) e purificadas com recurso a *inverse temperature cycling* (ITC). Depois da purificação, a clivagem no N-terminal da inteína  $\Delta$ I-SM (mutada) foi induzida com a adição do agente redutor DTT. A clivagem no C-terminal da inteína  $\Delta$ I-SM (mutada) foi provocada com uma solução de clivagem a pH 6. Apesar de ambas as construções mostrarem a ocorrência de auto-clivagem do AMP, os resultados obtidos com a construção pA3N foram mais promissores que os obtidos com a construção pA1C, já que este último demonstrou a ocorrência de *splicing* residual.

Adicionalmente, foram obtidas construções adicionais, com e sem His-tag, envolvendo a utilização da inteína *split Npu* DnaE e com o objetivo de obter a concatenação/multimerização de biopolímeros como ELPs e *silk-elastin-like* proteins (SELPs). Para tal, as construções genéticas exploram a utilização dos domínios de dimerização FRB e FKBP fundidos com as inteínas *split*, as quais, após adição de rapamicina, conduzem ao *splicing* e concatenação. Os vetores de expressão com e sem His-tag foram denominados por pA2H (com His-tag) e pA2 $\emptyset$  (sem His-tag). A análise dos resultados mostrou haver concatenação não controlada do biopolímero A60, conduzindo à formação de moléculas de maior peso molecular.

Palavras-chave: Concatenação; Inteína; Péptido Antimicrobiano; *Splicing*

## **ABSTRACT** - Development of a molecular biology platform for the production of antimicrobial peptides

Antimicrobial peptides (AMPs) are important antibiotic candidates due to their reduced or even absent propensity to induce development of resistance from microbes. This fact is important due to the rise of multi-resistant bacteria. This work aims at developing a platform for the production and purification of AMPs using elastin-like polypeptides (ELPs) for the non-chromatographic purification and inteins as self-cleaving tags for the isolation of the AMPs.

As general molecular cloning strategy, different adapters (total of nine) were designed and used for the genetic construction of four different expression vectors. Two vectors, named pA1C and pA3N, were designed to insert both ELP (A60) and different AMPs (ABP-CM4; HAMP; BMAP-28) on different sides of the intein. The vector pA1C uses a mutant intein  $\Delta$ I-CM with the N-terminal cleavage blocked by a point mutation. The pA3N vector, on the other hand, uses a mutant intein  $\Delta$ I-SM with the C-terminal cleavage blocked by a point mutation. Both inteins are mutants of the *Mtu* recA intein. As proof of concept for the molecular design, the ABP-CM4 peptide was inserted at the C-terminal in the pA1C vector and at the N-terminal in the pA3N vector. Following genetic constructions, the fusion proteins were expressed in *E. coli* BL21(DE3) and purified by inverse temperature cycling. After purification, the N-terminal cleaving of the  $\Delta$ I-SM (mutated) intein was triggered by addition of the reducing agent DTT. The C-terminal cleaving of the  $\Delta$ I-CM (mutated) intein was triggered with a cleaving solution at pH 6. Although both constructions demonstrated to be able to self-cleave the AMP, the pA3N vector showed to be more promising than pA1C vector, with the latter displaying residual splicing.

Additional constructions, with and without His-tags, involved the use of the split intein *Npu* DnaE aiming at the concatenation/multimerization of biopolymers such as ELPs and silk-elastin-like proteins (SELPs). The genetic constructions explore the use of FRB and FKBP dimerizing domains fused to split inteins to trigger splicing and concatenation after addition of rapamycin. The expression vectors were termed as pA2H (with His-tag) and pA2 $\emptyset$  (without His-tag). Results with A60 demonstrated the concatenation of the biopolymer leading to the formation of molecules with higher molecular weight, although at an uncontrolled rate.

Key words: Concatenation; Inteins; Antimicrobial Peptide; Splicing



# TABLE OF CONTENTS

LIST OF ABBREVIATIONS.....	x
LIST OF FIGURES.....	xiii
LIST OF TABLES.....	xvi
<b>1. Introduction.....</b>	<b>1</b>
<b>1.1. The Antibiotic Crisis.....</b>	<b>1</b>
<b>1.2. Antimicrobial peptides.....</b>	<b>1</b>
1.2.1. Structure of AMPs.....	2
1.2.2. Mode of action of Antimicrobial Peptides.....	2
1.2.3. Advantages of Antimicrobial Peptides.....	5
1.2.4. Antimicrobial peptide: ABP-CM4.....	5
1.2.5. Antimicrobial peptide: HAMP.....	6
1.2.6. Antimicrobial peptide: BMAP-28.....	6
1.2.7. Heterologous synthesis of AMPs.....	6
<b>1.3. Protein-Based Polymers (PBPs).....</b>	<b>7</b>
1.3.1. Protein-Based Polymers as fusion partners for protein purification.....	7
<b>1.4. Inteins.....</b>	<b>8</b>
1.4.1. Contiguous inteins and split inteins.....	9
1.4.2. Splicing Mechanism.....	9
1.4.3. Self-cleaving tags.....	11
1.4.4. Concatenation tags.....	12
1.4.5. Conditional protein splicing.....	12
<b>1.5. Objectives.....</b>	<b>13</b>
<b>2. Materials and Methods.....</b>	<b>15</b>
<b>2.1. Biological Materials.....</b>	<b>15</b>
2.1.1. Bacterial strains.....	15
2.1.2. Culture media.....	15
2.1.3. Plasmids.....	15
2.1.4. Protein-based polymers (PBP) sequences.....	15
2.1.5. Antimicrobial peptides.....	16
2.1.6. Oligonucleotides.....	16
2.1.7. Inteins.....	17
2.1.8. FKBP and FRB.....	18

2.1.9. Adapters design .....	19
2.1.10. Oligonucleotides for Site Directed Mutagenesis (SDM) .....	19
<b>2.2. Methods</b> .....	<b>20</b>
2.2.1. Plasmid extraction .....	20
2.2.2. Competent cells.....	20
2.2.3. Bacterial Transformation via Heat Shock Method.....	20
2.2.4. Plasmid preparation.....	20
2.2.5. Site Directed Mutagenesis.....	21
2.2.6. Insert preparation .....	21
2.2.7. Insertion of adapters into pET25 .....	23
2.2.8. Preparation of the adapters for ELP/SELP insertion.....	23
2.2.9. Insertion of ELP/SELP .....	24
2.2.10. Preparation of adapters for AMP insertion .....	24
2.2.11. Insertion of ABP-CM4.....	25
2.2.13. Production screenings .....	25
2.2.14. Protein Purification by Inverse Transition Cycling (ITC).....	26
2.2.15. Cleavage test.....	26
2.2.16. Concatenation test.....	27
2.2.17. Western Blot.....	27
<b>3. Results and Discussion</b> .....	<b>28</b>
<b>3.1. Adapter design</b> .....	<b>28</b>
3.1.1. N-terminal cleavage adapter.....	29
3.1.2. C-terminal cleavage adapter.....	29
3.1.3. Concatenation adapters .....	30
3.1.4. Variants of the concatenation adapter .....	31
3.1.5. Alternative adapter for the production of AMPs .....	33
<b>3.2. Construction of the expression vector</b> .....	<b>34</b>
3.2.1. Site directed mutagenesis .....	34
3.2.2. Insertion in vector pET25 .....	35
3.2.3. Insertion of the ELPs.....	37
3.2.5. Insertion of ABP-CM4.....	41
<b>3.3. Production screenings</b> .....	<b>41</b>
3.3.1. Production screening of hisA60-int <sub>pH</sub> -CM4 <sub>cys</sub> , hisA60-int <sub>pH</sub> -CM4, hisA60-CM4 <sub>cys</sub> and hisA60-CM4 proteins.....	42

3.3.2. Production screenings of <i>cysCM4-A60<sub>his</sub></i> , <i>CM4-A60<sub>his</sub></i> , <i>cysCM4-intRed-A60<sub>his</sub></i> and <i>CM4-intRed-A60<sub>his</sub></i> proteins .....	43
3.3.3. Production screenings of <i>intcon-C-A60<sub>his</sub>-intcon-N</i> , <i>intcon-C-A60-intcon-N</i> , <i>intcon-C-SELP<sub>his</sub>-intcon-N</i> and <i>intcon-C-SELP-intcon-N</i> proteins .....	45
<b>3.4. Protein Purification</b> .....	46
3.4.1. <i>hisA60-intpH-CM4</i> .....	47
3.4.2. <i>hisA60-CM4<sub>cys</sub></i> .....	49
3.4.3. <i>CM4-intRed-A60<sub>his</sub></i> .....	50
3.4.4. <i>cysCM4-A60<sub>his</sub></i> .....	51
3.4.5. <i>intcon-C-A60<sub>his</sub>-intcon-N</i> .....	52
<b>3.5. Cleavage test</b> .....	53
3.5.1. <i>CM4-intRed-A60<sub>his</sub></i> .....	53
3.5.2. <i>hisA60-intpH-CM4</i> .....	57
<b>3.6. Concatenation test</b> .....	60
<b>4. Conclusions</b> .....	66
<b>4. References</b> .....	67
<b>5. Annexes and Appendixes</b> .....	78

## LIST OF ABBREVIATIONS

aa: Amino acid

ABP-CM4: Antibacterial Peptide CM4

AMP: Antimicrobial peptide

BMAP-28: Bovine myeloid antimicrobial peptide 28

CFU: Colony forming unit

CPP: Cell Penetrating Peptide

DNA: Deoxyribonucleic acid

DOD: Dodecapeptide

DTT: Dithiothreitol

ELP: Elastin-like polypeptide

ELR: elastin-like recombinamer

FKBP: FK506 binding protein

FRB: FKPB-rapamycin binding domain

GRAS: Generally Regarded As Safe

HAMP: Hepcidin antimicrobial peptide

His: Histidine

*Has. Halobacterium salinarum*

*Hut. Halorhabdus utahensis*

ITC: Inverse Transition Cycling

LB: Lysogeny Broth

*Mtu. Mycobacterium tuberculosis*

MW: Molecular weight

*Npu. Nostoc punctiforme*

PBP: Protein-based polymers

PBS: Phosphate-Buffered Saline

PCR: Polymerase chain reaction

RNA: Ribonucleic acid

SELP: Silk-elastin-like protein

SDM: Site Directed Mutagenesis

*Ssp. Synechocystis species*

TB: Terrific Broth

TB<sub>lac</sub>: Terrific Broth (induction with lactose)

TE: Tris-EDTA

ΔI: Mini intein

ΔI-CM: Mini intein-Cleaving Mutant

ΔI-SM: Mini intein-Splicing Mutant

## LIST OF FIGURES

<b>Figure 1.</b> Modes of action of membrane-active antimicrobial peptides. ....	<b>Erro! Marcador não definido.</b>
<b>Figure 2.</b> Models of membrane permeability.....	<b>Erro! Marcador não definido.</b> 5
<b>Figure 3.</b> Representation of the difference between a typical large intein and a minimal or mini intein. .	9
<b>Figure 4.</b> Representation of the <i>trans</i> -splicing of a split intein. ....	99
<b>Figure 5.</b> Display of the splicing mechanism performed by the different intein class. ....	11
<b>Figure 6.</b> A- Representation of the insertion confirmation method using <i>Ppu21</i> restriction enzyme. B- Simulation of a 1% agarose gel in the SnapGene® program .....	2424
<b>Figure 7.</b> Overview of the designed adapters. ....	28
<b>Figure 8.</b> Schematics for the designed universal adapter for N-terminal self-cleaving tag.....	29
<b>Figure 9.</b> Schematics for the designed universal adapter for C-terminal self-cleaving tag.....	30
<b>Figure 10.</b> Schematics for the designed adapters. A- Self-assembly adapter without his-tag. B- Self-assembly adapter with his-tag. ....	31
<b>Figure 11.</b> Design schematics of thenon-utilized self-assembly adapters. A- Self-assembly adapter with no form of conditional splicing. B- Self-assembly adapter bordered by FRB domains. C- Self-assembly adapter bordered by FKBP. The constructions of the adapters B and C are part of the same dual system for protein multimerization. D- Self-assembly adapter conditioned by chemical caging with disulphide bridges...	333
<b>Figure 12.</b> Schematics for the designed adapter for production of AMPs that split the AMP with an intein halfway.....	34
<b>Figure 13.</b> Electrophoretic analysis of the PCR product of the site directed mutagenesis in a 1% agarose gel.....	35
<b>Figure 14.</b> A- Simulation of a 1% agarose gel in the SnapGene® program. B- Electrophoretic analysis in a 1% agarose gel of the digestion of vectors with <i>Eco72I</i> and <i>Kpr2I</i> . ....	36
<b>Figure 15.</b> A- Simulation of a 1% agarose gel in the SnapGene® program. B- Electrophoretic analysis in a 1% agarose gel of the digestion of vectors with <i>KprI</i> and <i>Kpr2I</i> . ....	36
<b>Figure 16.</b> A- Simulation of a 1% agarose gel in the SnapGene® program. B- Electrophoretic analysis in a 1% agarose gel of the digestion of vectors with <i>Ppu2I</i> . ....	37
<b>Figure 17.</b> A- Simulation of a 1% agarose gel in the SnapGene® program. B- Electrophoretic analysis in a 1% agarose gel of the digestion of vectors with <i>KprI</i> .....	38
<b>Figure 18.</b> A- Simulation of a 1% agarose gel in the SnapGene® program. B- Electrophoretic analysis in a 1% agarose gel of the digestion of vectors with <i>XhoI</i> and <i>NdeI</i> . ....	39

<b>Figure 19.</b> A- Simulation of a 1% agarose gel in the SnapGene® program. B- Electrophoretic analysis in a 1% agarose gel of the digestion of vectors with <i>Xba</i> I and <i>Ava</i> I.....	40
<b>Figure 20.</b> A- Simulation of a 1% agarose gel in the SnapGene® program. B- Electrophoretic analysis in a 1% agarose gel of the digestion of vectors with <i>Ppu</i> II.....	40
<b>Figure 21.</b> SDS-PAGE analysis of <i>his</i> A60-int <sub>pH</sub> -CM4 <sub>cys</sub> (A), <i>his</i> A60-int <sub>pH</sub> -CM4 (B), <i>his</i> A60-CM4 <sub>cys</sub> (C) and <i>his</i> A60-CM4 (D) production screenings in a 10% polyacrylamide gel.....	43
<b>Figure 22.</b> SDS-PAGE analysis of <i>cys</i> CM4-int <sub>Red</sub> -A60 <sub>his</sub> (A), CM4-int <sub>Red</sub> -A60 <sub>his</sub> (B), <i>cys</i> CM4-A60 <sub>his</sub> (C) and CM4-A60 <sub>his</sub> (D) production screenings in a 10% polyacrylamide gel.....	44
<b>Figure 23.</b> SDS-PAGE analysis of int <sub>con-C</sub> -A60 <sub>his</sub> -int <sub>con-N</sub> (A) and int <sub>con-C</sub> -A60-int <sub>con-N</sub> (B) production screenings in a 10% polyacrylamide gel.....	45
<b>Figure 24.</b> SDS-PAGE analysis of int <sub>con-C</sub> -SELP <sub>his</sub> -int <sub>con-N</sub> (A) and int <sub>con-C</sub> -SELP-int <sub>con-N</sub> (B) production screenings in a 10% polyacrylamide gel.....	46
<b>Figure 25.</b> SDS-PAGE analysis of <i>his</i> A60-int <sub>pH</sub> -CM4 purification in a 10% polyacrylamide gel.....	49
<b>Figure 26.</b> SDS-PAGE analysis of <i>his</i> A60-CM4 <sub>cys</sub> purification in a 10% polyacrylamide gel.....	50
<b>Figure 27.</b> SDS-PAGE analysis of CM4-int <sub>Red</sub> -A60 <sub>his</sub> purification in a 10% polyacrylamide gel.....	51
<b>Figure 28.</b> SDS-PAGE analysis of <i>cys</i> CM4-A60 <sub>his</sub> purification in a 10% polyacrylamide gel.....	52
<b>Figure 29.</b> SDS-PAGE analysis of int <sub>con-C</sub> -A60 <sub>his</sub> -int <sub>con-N</sub> purification in a 10% polyacrylamide gel.....	53
<b>Figure 30.</b> A- SDS-PAGE analysis of CM4-int <sub>Red</sub> -A60 <sub>his</sub> cleavage test in a 10% polyacrylamide gel. B- Western Blot analysis of a gel equal to the one in A.....	55
<b>Figure 31.</b> Theorized mechanism for intra-intein cleavage that would explain the appearance of the intermediary bands during the cleavage.....	56
<b>Figure 32.</b> SDS-PAGE analysis of CM4-int <sub>Red</sub> -A60 <sub>his</sub> cleavage test in a 10% polyacrylamide gel. Samples were taken after cleavage induction of 20 h and after a hot cycle centrifugation. The same gel was stained with Coomassie Blue (A) and AgNO <sub>3</sub> (B).....	57
<b>Figure 33.</b> A- SDS-PAGE analysis of <i>his</i> A60-int <sub>pH</sub> -CM4 cleavage test in a 10% polyacrylamide gel.. B- Western Blot analysis of a gel similar to the one in A.....	59
<b>Figure 34.</b> SDS-PAGE analysis of <i>his</i> A60-int <sub>pH</sub> -CM4 cleavage test in a 10% polyacrylamide gel. Samples were taken after cleavage induction of 24 h at room temperature and other 24 h at 30 °C. The same gel was stained with Coomassie Blue (A) and AgNO <sub>3</sub> (B). .....	60
<b>Figure 35.</b> SDS-PAGE analysis of int <sub>con-C</sub> -A60 <sub>his</sub> -int <sub>con-N</sub> concatenation test after incubation for 4 h at room temperature.. .....	62

**Figure 36.** SDS-PAGE analysis of int<sub>con-C</sub>-A60<sub>his</sub>-int<sub>con-N</sub> concatenation test after incubation for 24 h at room temperature. .... 62

**Figure 37.** A- SDS-PAGE analysis of int<sub>con-C</sub>-A60<sub>his</sub>-int<sub>con-N</sub> concatenation after incubation for 24 h at room temperature. B- Western Blot analysis with anti-His and anti-mouse as primary and secondary antibodies, respectively..... 63



## LIST OF TABLES

<b>Table 1.</b> Sequence and length of ELPs and SELP. ....	16
<b>Table 2.</b> Amino acid sequences and length of the antimicrobial peptides used in this work.....	16
<b>Table 3.</b> Primer sequences used to amplify ABP-CM4. Underlined nucleotides represent where the annealing with the ABP-CM4 sequence occurs. ....	17
<b>Table 4.</b> Amino acid sequences of the inteins used for the design of adapters.....	17
<b>Table 5.</b> Cleavage strategies used for each of the selected intein. ....	18
<b>Table 6.</b> Amino acid sequence and length of the FKBP and FRB domains used for the design of concatenation adapters.....	19
<b>Table 7.</b> Mutagenesis primers to mutate nucleotides of the concatenation adapters.....	19
<b>Table 8.</b> Oligonucleotides used for each amplification and restriction enzymes used for digestion as well as the target adapter for insertion. ....	22
<b>Table 9.</b> Schematic representation of all the final constructions. ....	41
<b>Table 10.</b> Protein Modules and corresponding molecular weights .....	46
<b>Table 11.</b> Attribution of proteins to the bands considering the size and percentage of ELP.....	<b>Erro!</b>

**Marcador não definido.**

# 1. Introduction

## 1.1. The Antibiotic Crisis

The world faces, nowadays, a health problem with the emergence of multi-resistant strains of bacteria compromising the treatment of bacterial infections. Every year almost 99,000 deaths occur associated with antibacterial resistant pathogens in the United States of America (Ventola, 2015), while in Europe around 33000 deaths per have been attributed to infections by antibiotic-resistant bacteria (Cassini *et al.*, 2019). Therefore, there is a need to bypass antimicrobial resistance to antibiotics, by developing alternatives to the current antibiotics.

The multi-resistant strains are mostly the result of antibiotic overuse. Reasons for the abuse of antibiotics go from unsuited antibiotic prescriptions to overprescription of antibiotics, as well as due to the agricultural use of antibiotics as animal growth supplements – a practice now banned by the European Union. All these factors lead to a selective pressure for microbes to adapt and become resistant, ultimately leading to the rise of multi-resistance strains (Ventola, 2015). Another important factor influencing the antibiotic crisis is the lack of development of new antibiotics by pharmaceutical companies. Antimicrobial peptides may provide ways to bypass microbial resistance to antibiotics.

## 1.2. Antimicrobial peptides

Antimicrobial peptides (AMPs) are produced by a variety of different organisms and serve as a first line of defence against pathogens and, in the case of microorganisms, to acquire a competitive advantage against other organisms (Marcos & Manzanares, 2012). From a biotechnological point of view, these peptides have the potential to enter the field of antibiotic design and production. Especially, due to a specific characteristic: AMPs targets seem to become resistant to these peptides at a slower rate than conventional antibiotics, and in this way, they constitute an alternative to standard antibiotics. Indeed, as AMPs target highly conserved structures and pathways in microbial cells, it is harder to cells to evolve and create resistance mechanisms (Spohn *et al.*, 2019).

Many unicellular - such as bacteria and protozoa - and multicellular organisms - like plants and mammals - produce AMPs. The first evidence of AMPs existence was reported in 1939 by René Dubos in a bacillus found in the soil (Dubos, 1939). Since then, multiple AMPs have been discovered, characterized and synthesized. In the case of plants, the first AMP isolated and described was purotionin in 1942, isolated from *Triticum aestivum*; as for animals, a defensin AMP was the first peptide isolated and

described to have antibacterial activity, back in 1956 (Balls, 1942; Hirsch, 1956). Currently, one of the best curated collection of antimicrobial peptides (CAMP<sub>R3</sub>, <http://www.camp3.bicnirrh.res.in>) contain over 10000 sequences of AMPs (Waghu *et al.*, 2015).

### 1.2.1. Structure of AMPs

Generally, AMPs consist of about five to a hundred amino acids and are mainly cationic, even though anionic AMPs exist. AMPs affect different targets from very different organisms like bacteria, virus, fungi, or even cancer cells (Marcos & Manzanares, 2012). Charge, conformational structure, amphipacity and hydrophobicity are key factors that influence the specificity of an AMP, determining what type of targets, and more specifically, what type of membrane it can interact. Factors like conformation and structure are prone to change depending on the conditions they are exposed to, for example when an AMP is exposed to a hydrophobic environment like that of the inside of a membrane (Park & Hahm, 2005). The cationic charge of AMPs enables their interaction with the negatively charged phosphate groups of phospholipids, in bacterial membranes. This interaction is the most discussed mode of action of AMPs, as in the case of gram-negative bacteria, where the exterior membrane generally displays an anionic charge where AMPs can interact (Sierra *et al.*, 2017). These mechanisms will be further discussed below.

While most of described AMPs display a helical structure, four different types of secondary structures are found for these peptides:  $\alpha$ -helix,  $\beta$ -sheet, extended, or loop (Huang *et al.*, 2010). Due to their size, AMPs have simple secondary structures, but in rare occasions, it is possible to find an AMP with more than one secondary structure (Huang *et al.*, 2010). Due to their high number and diversity, AMPs can be divided into several categories or families. One can divide AMPs into five classes, according to their structure and amino acid composition: linear cationic  $\alpha$ -helix peptides; cationic peptides rich in specific amino acids; anionic and cationic peptides that contain cysteines and form disulphide bonds; anionic peptides; and anionic and cationic peptides that are fragments of larger proteins (Park & Hahm, 2005). Several other divisions by family (Khamis *et al.*, 2014), origin (Li *et al.*, 2012), and other structural characteristics (Wang, Li & Wang, 2015) can be used but will not be further discussed in this text.

### 1.2.2. Mode of action of Antimicrobial Peptides

AMPs activity is defined by the interaction with different conserved targets, depending on the organism being studied. In the case of peptides that are active against bacteria, they can be divided in two groups: the membrane-active AMPs, which disrupt the integrity of the cell membrane to kill the cell

(Yeaman & Yount, 2003), and the intracellularly active AMPs, that act on targets inside the cell (Nicolas, 2009).

Most of the membrane-active AMPs are amphipathic, giving them the ability to interact and enter periodically in the bacterial membrane to cause disruption and leading to cell death. Grage *et al.* and Yeaman & Yount described different models to explain the membrane disruption activity of AMPs (Figure 1) namely, the barrel-stave, toroidal pore, aggregate and carpet models (Yeaman & Yount, 2003) whereas, the membrane thinning model was later added (Grage *et al.*, 2016). In the barrel-stave model, the cell membrane permeability increases when the AMPs aggregate themselves in a perpendicular way to the membrane forming porous structures that allow for exchanges between intra and extracellular fractions, resulting in membrane disruption and cell death (Yeaman & Yount, 2003). The toroidal pore model is very similar to the barrel-stave model, but the arrangement of the AMPs forces the phospholipids heads to be always oriented towards the structured AMPs (Shai, 2002). In the aggregate model, the AMPs combine with the phospholipids to form complexes that result in the formation of heterogeneous non-specifically oriented pores increasing the membrane permeability (Sierra *et al.*, 2017). The membrane thinning model works by increasing the distance between phospholipids in the membrane causing the membrane to thin and be disrupted. Commonly, this is achieved by parallel insertion of AMPs on one side of the membrane or, less commonly, perpendicular insertion of AMPs on one side of the membrane (Grage *et al.*, 2016). The carpet model proposes that the AMPs acts as a detergent, with the creation of micelles after AMPs interaction with the membrane, destabilizing it and creating holes that results in cell death (Shai, 2002).

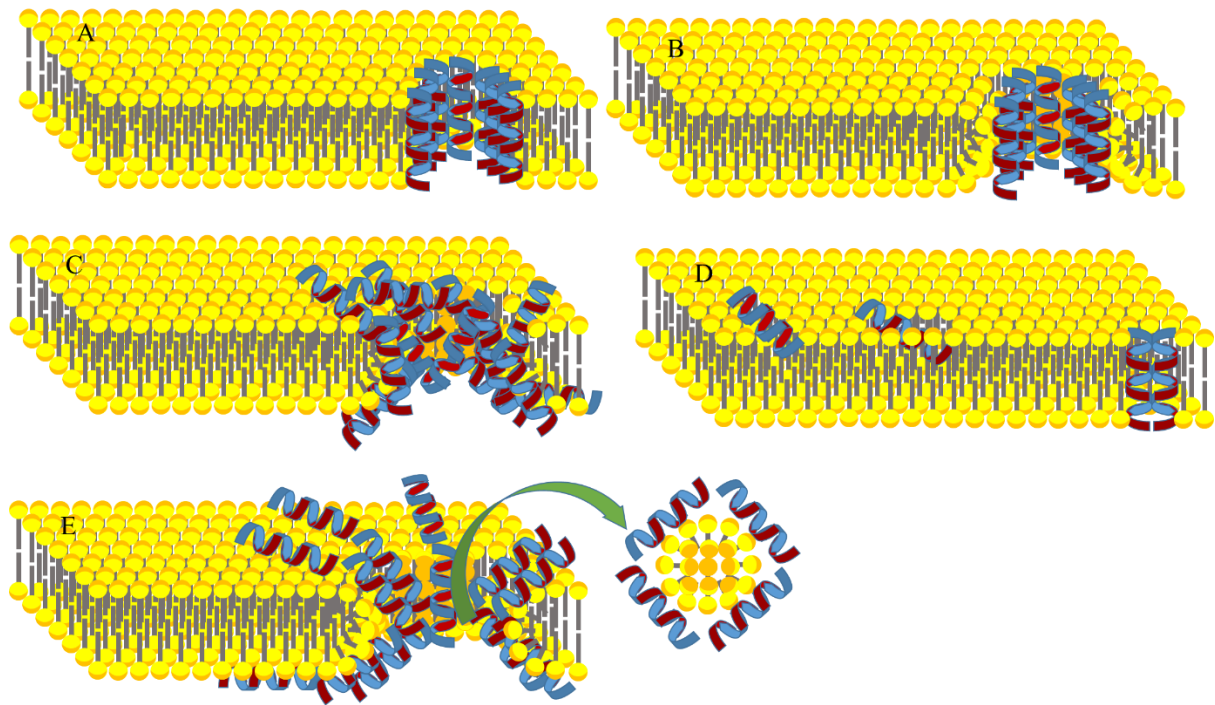


Figure 1. Modes of action of membrane-active antimicrobial peptides via the (A) barrel-stave, (B) toroidal pore, (C) aggregate, (D) membrane thinning and (E) carpet models. The membrane thinning model is shown in three different variants of that type. Based on Laverty, Gorman & Gilmore, 2011 and Grage *et al.*, 2016.

Intracellular-active AMPs have the ability to cross the plasmatic membrane in order to act against different targets. The way peptides cross this lipidic barrier is still not completely understood and two main models have been proposed (Marcos & Manzanares, 2012). Specifically, it was suggested that intracellularly active AMPs could cross the membrane either through spontaneous lipid-assisted translocation or stereospecific receptor-mediated membrane translocation (Nicolas, 2009). A comparison with Cell Penetrating Peptides (CPPs) can be made as they hold many similar properties to AMPs, such as being small, cationic and amphipathic peptides. CPPs are notorious for serving as transduction domains that lead proteins of interest to the cell interior, in a parallel way to how intracellular-active AMPs cross the cell membrane (Derakhshankhah & Jafari, 2018). Figure 2A illustrates the model for spontaneous lipid-assisted translocation in which, the cationic AMP starts by assuming a parallel orientation to the membrane due to electrostatic interactions. This interaction is followed by a temporary permeability of the membrane, repositioning the AMP perpendicularly in the membrane, leading to the translocation of the peptide to the cell interior (Nicolas, 2009). The stereospecific receptor-mediated model (Figure 2B) for membrane translocation happens at lower concentrations than the first described model. In this model, the interaction of the AMP with a membrane protein or peptidoglycan gives the AMP the potential to translocate through the membrane (Marcos & Manzanares, 2012). AMPs intracellular

activity varies from inhibition of DNA, RNA or protein synthesis to the interruption of enzymatic activity or protein folding (Nicolas, 2009).

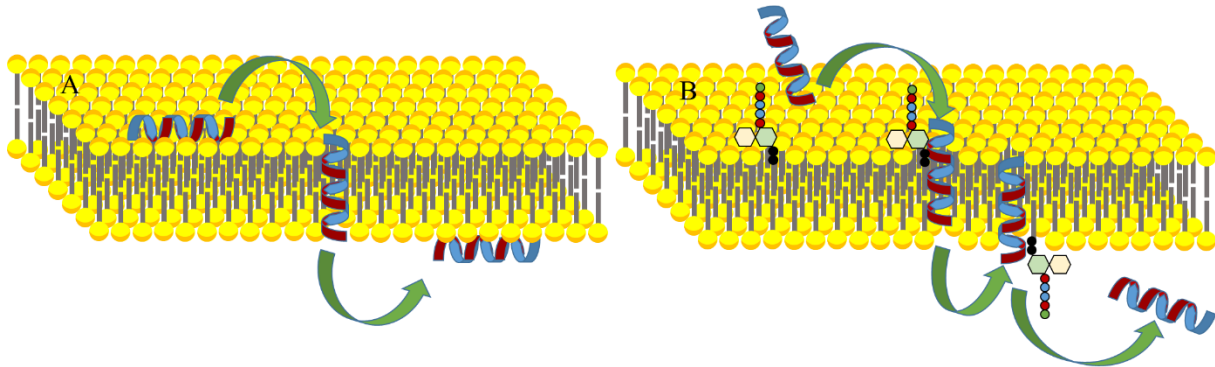


Figure 2. Models of membrane permeability: (A) Spontaneous lipid-assisted translocation, and (B) stereospecific receptor-mediated. Based on Münch & Sahl, 2015 and Marcos & Manzanares, 2012

### 1.2.3. Advantages of Antimicrobial Peptides

The targets of AMPs are more intrinsic than normal antibiotics and may involve more than a simple horizontal transfer of genes to acquire resistance (Spohn *et al.*, 2019). The mechanisms through which microbes can become resistant to AMPs usually involve either changing the extracellular acidity or reducing their net negative charge, although this is far less likely to occur than the resistance methods found for chemical antibiotics (Zasloff, 2002; Sierra *et al.*, 2017). Furthermore, the reason for AMPs specificity to microbial cells (or even cancer cells) relates to membrane potential and their constituents. The presence of cholesterol in mammalian cell membranes makes these cells more resistant to AMPs than prokaryotic cells. This occurs mostly due to the control of stability and fluidity of the plasmatic membrane by cholesterol molecules and to the inhibition of membrane disruption by AMPs (Zasloff, 2002).

In this dissertation, three AMPs were considered out from the plethora of AMPs available: ABP-CM4, BMAP-28 and HAMP. These were chosen based on previous works from the research group and are briefly described.

### 1.2.4. Antimicrobial peptide: ABP-CM4

The antibacterial peptide CM4, usually termed ABP-CM4, is an AMP extracted from the hemolymph of the silkworm *Bombyx mori* (Tu, Qu & Xu, 1989). ABP-CM4 activities include antibacterial, antifungal and even anti-tumoral, while being non-toxic to normal mammalian cells (Cheng, Lu, Zhang &

Cao, 2010). This 35 amino acids-long cationic peptide presents a  $\alpha$ -helix secondary structure and net charge of +5 at neutral pH. The helix structure holds on one of the faces the charged residues while the other include mainly hydrophobic amino acids. On interaction with the membrane, ABP-CM4 forms ion channels that depolarize the cell and lead to cell death. (Li *et al.*, 2011).

#### 1.2.5. Antimicrobial peptide: HAMP

Hepcidin antimicrobial peptide (HAMP) is a hormone primarily found in the liver and considered to be the main responsible for regulation of iron homeostasis in mammals (Pietrangelo & Trautwein, 2004). Its richness in cysteines allows  $\beta$ -sheet secondary structures to form, critical for its iron regulation activity (Clark *et al.*, 2011). Besides its regulatory activity, hepcidin is a known AMP that acts against bacterial, fungal and viral infections (Liu *et al.*, 2012; Michels *et al.*, 2015). One of the reasons for its antimicrobial properties is thought to be the reduction of the iron available that is necessary for pathogens to survive (Pietrangelo & Trautwein, 2004; Pigeon *et al.*, 2000), although *in vitro* studies have shown other modes of action (Houamel *et al.*, 2015).

#### 1.2.6. Antimicrobial peptide: BMAP-28

The bovine myeloid antimicrobial peptide 28 (BMAP-28), as the name implies, was isolated from the neutrophils of bovines (Huang, Ross & Blecha, 1997). The AMP has an  $\alpha$ -helix structure with a hydrophobic tail. BMAP-28 induces the death of bacteria and fungi by permeabilizing either the mitochondrial or the plasmatic membrane of pathogens by forming pores in the membrane (Guo, Xun & Han, 2018; Risso *et al.*, 2002).

#### 1.2.7. Heterologous synthesis of AMPs

Chemical synthesis is the major alternative used to for the obtention of AMPs, having several disadvantages such as the several steps of production, expensive production in large-scale and limitations in purification. The heterologous expression of AMPs can be optimized to obtain a more rentable method, with fewer steps and the ability for large scale productions (Li *et al.*, 2011; Pachon-Ibáñez *et al.*, 2017).

*E. coli* is a very versatile host to be used in heterologous production with several commercially vectors available (Rosano & Ceccarelli, 2014). *E. coli* is a well-studied Gram-negative facultative anaerobic bacterium, generally regarded as safe (GRAS) and commonly found in the intestine of mammals (Madigan *et al.*, 2015). Due to its versatility, this host has been widely used for the heterologous expression of AMPs with the strain *E. coli* BL21 (DE3) being the most widely used (Sinha & Shukla, 2019). Nevertheless, the heterologous expression of AMPs faces a number of problems. The main problem refers to AMP's toxicity that might affect the producing host, while their small size and nature makes AMPs prone to

proteases (Zorko & Jerala, 2009). Purification is also hindered in many cases due to AMP's small size and amphipathic nature. For these reasons, a purification tag is required in most of the cases, increasing solubility and facilitating the purification while masking the toxicity of the AMP.

### **1.3. Protein-Based Polymers (PBPs)**

Protein-based polymers such as elastin-like recombinamers/polypeptides (ELR/ELPs) and silk-elastin-like proteins (SELPs) can serve as fusion partners for the production of recombinant AMPs. These proteins exhibit interesting properties for biomedical applications, are non-cytotoxic, and can be purified using non-chromatographic techniques (Machado *et al.*, 2013a; Machado *et al.*, 2013b; da Costa *et al.*, 2015; da Costa *et al.*, 2018). This is useful for providing cost-efficient purification of AMPs. Furthermore, these polymers have the potential to be used to produce scaffolds for the regeneration of tissues due to their biocompatibility (Tejeda-Montes *et al.*, 2014; Machado *et al.*, 2013b).

In the work of this dissertation, adapters were designed with two objectives in mind: to allow the fusion of different ELPs/SELPs with different AMPs in order to obtain biopolymers with antimicrobial properties and to produce those peptides using these polymers as fusion tags. Recently, ABP-CM4 peptide fused with A200 (ELP) was produced at approximately 120 mg/L in batch cultures, while using the pET25 vector in *E. coli*. This shows the versatility of ELPs to be used as production and purification tags as well as base material for the production of antimicrobial films (da Costa *et al.*, 2015).

#### 1.3.1. Protein-Based Polymers as fusion partners for protein purification

Fusion tags provide better protein solubilisation, detection, purification and other properties to proteins otherwise would not have them. ELPs used as fusion tags improve the solubility of the AMPs, while it also prevent the produced AMPs to exert toxicity towards the production host (Hu *et al.*, 2009). In the case of ELPs, cycles of hot and cold incubation and centrifugation steps can be used for purification due to their thermoresponsive behaviour in a process called Inverse Transition Cycling (ITC) (Meyer & Chilkoti, 1999; Machado *et al.*, 2009; da Costa *et al.*, 2018). Salting-out is another non-chromatographic technique that is used for the purification of SELPs (Machado *et al.*, 2013a). Both ITC and salting out methods are cost efficient types of purification. Moreover, AMPs can be used to confer antimicrobial properties to these biopolymers. After purification and if the end goal is to have the protein of interest without the fusion tag, that fusion tag must be removed. For that purpose, the use of inteins as fusion partners is advantageous, as these can be used as to form self-cleaving tags with no extra amino acids (aa) being added to the peptide (Casali & Preston, 2003; Burgess-Brown, 2017). This is an important



issue to be considered upon the design of AMPs, especially due to their small size. The introduction of additional amino acids could affect the structure and mode of action of the AMP, leading to a lack of function.

#### 1.4. Inteins

Protein splicing occurs when a protein has one of its center domains removed and the bordering sequences are connected (Perler *et al.*, 1994). A common way to explain this process is to parallel it to RNA splicing. Intein domains are spliced from their native protein in a similar way to how introns are removed from the precursor RNA. The name intein derives from internal protein sequence (Perler *et al.*, 1994). The first intein was discovered almost 30 years ago, as a part of the gene *VMA1* from *Saccharomyces cerevisiae*, when the researchers noticed that a gene was expressing two proteins and one of those protein's sequence was inside the sequence of the other. A precursor protein was shown to be able form two separate proteins removing the protein inside of the other one (Hirata *et al.*, 1990; Yamashiro *et al.*, 1990). Since then, inteins have been discovered in all three domains of life (Perler, 2002).

The parts of the precursor protein that are not intein are called exteins and, logically, an extein is a C-extein if it is situated in the C-terminal of the protein or a N-extein if it is situated in the N-terminal of the protein (Figure 3). After splicing, the exteins are combined and compose the mature protein (Perler *et al.*, 1994). Sometimes an intein might have a linker sequence in the middle that is itself another protein domain which is usually a dodecapeptide (DOD) homing endonuclease. Such inteins are called large inteins while inteins that lack that endonuclease are called mini (minimal) inteins (Figure 3). It has been demonstrated that removal of the DOD homing endonuclease are not important for the protein splicing and, in this way, it is thus possible to produce artificial mini inteins (Chong & Xu, 1997; Shingledecker, Jiang & Paulus, 1998).

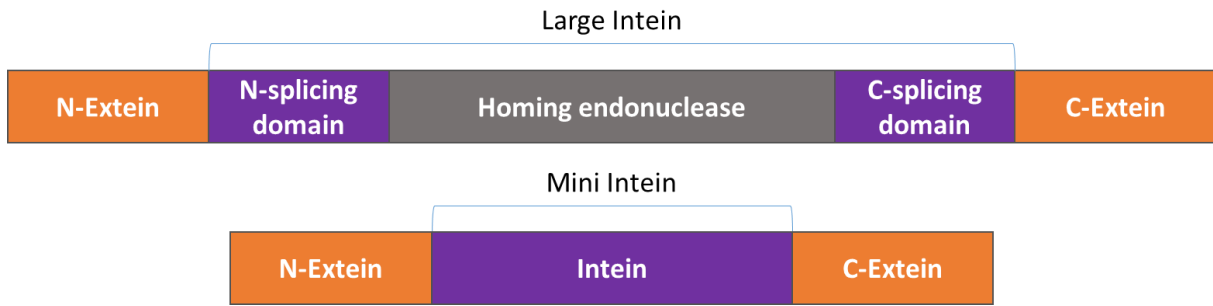


Figure 1. Representation of the difference between a typical large intein and a minimal or mini intein.

#### 1.4.1. Contiguous inteins and split inteins

Inteins can be divided into contiguous inteins and split inteins. While contiguous inteins are expressed in the same protein as a whole domain, split inteins are located apart in the DNA sequence and expressed scattered in separate proteins. Then, the N- and the C-splicing domains interact with each other for the splicing process to occur, using the same process as contiguous inteins, uniting the two exteins. This process called *trans*-splicing (Figure 4) was discovered in 1998 and it stands in contrast to the *cis*-splicing that occurs in contiguous inteins (Wu, Hu & Liu, 1998). Split intein domains are usually termed C-intein or N-intein depending if they are the portion of the C-splicing domain or the N-splicing domain, respectively (Figure 4). While these two types of inteins exist in Nature, *in vitro* experiments have shown that split inteins can be engineered from contiguous inteins and *vice versa* (Elleuche & Pöggeler, 2007; Mootz & Muir, 2002; Mills *et al.*, 1998; Southworth *et al.*, 1998).

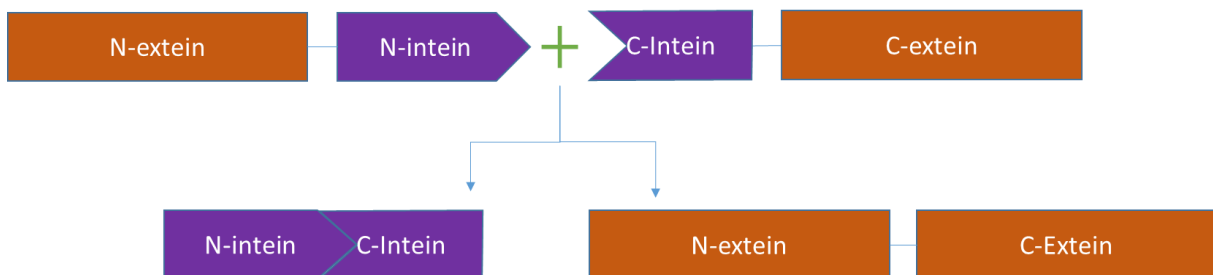


Figure 2. Representation of the *trans*-splicing of a split intein.

#### 1.4.2. Splicing Mechanism

Based on the beginning of their splicing mechanism, inteins are divided into three classes: class 1, class 2 and class 3 inteins. Class 1 inteins are by far the most common and are the ones used

throughout this dissertation. In class 1 inteins, the first step of the splicing mechanism includes an N-S/N-O acyl shift that requires a Cysteine, a Serine or a Threonine as the first residue of the inteins that causes an intermediate thioester bond important for the second step. The thioester bond is transferred from the first aa of the intein to the aa +1 of the C-extein. This aa is, once again, either a Cysteine, a Serine or a Threonine. This step finishes with the N-terminal cleavage, connecting both N- and C-extein (Cooper & Stevens, 1995). The N-terminal cleavage is followed by the cyclization of the carboxamide group on the last aa of the intein, typically a Glutamine or an Asparagine, leading to the cleavage of the C-terminal (Petrokovski, 1998). Then, the cyclization on the intein's C-terminal is hydrolysed and undone. The now connected exteins go through a final N-S/N-O acyl shift that restores the peptide bond, resulting in the mature protein (Amitai, Dassa & Petrokovski, 2003).

The difference between the Class 1 of inteins and the other two classes (Class 2 and 3) is mostly observed in the first step of the splicing. The N-S/N-O acyl shift of Class 1 inteins is replaced by a nucleophilic displacement to the first aa of the C-extein. In Class 2 inteins, instead of a Cysteine, Serine or Threonine, an Alanine is present as the first aa of the intein (Southworth, Benner & Perler, 2000). Class 3 inteins also start with an Alanine in the +1 position. In a different way, these inteins induce a N-S acyl shift with a Cysteine present in the middle of the intein, instead of doing it with the first aa (Brace *et al.*, 2010). The rest of the splicing occurs similar to the Class 1 inteins. These differences are visually represented in Figure 5.

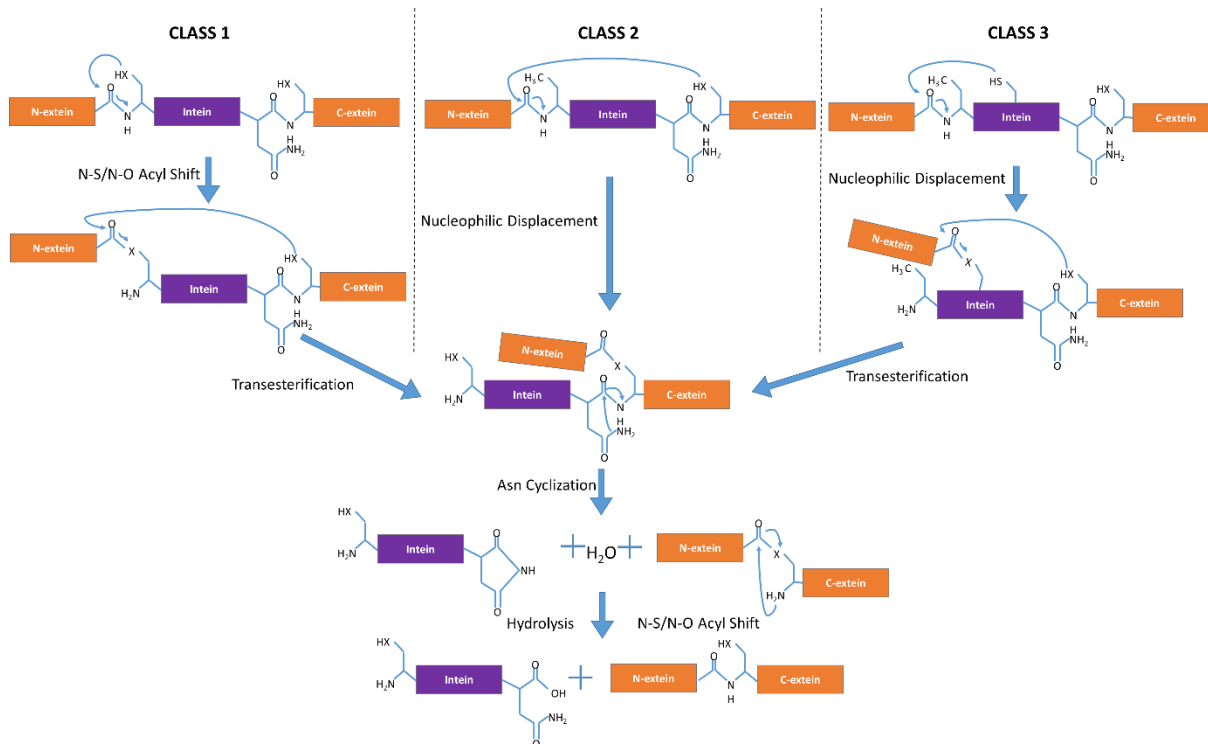


Figure 3. Display of the splicing mechanism performed by the different intein class. From left to right: Class 1 inteins; Class 2 inteins; Class 3 inteins. Based on Brace *et al.*, 2010.

### 1.4.3. Self-cleaving tags

The splicing ability of inteins can be used to create self-cleaving tags that can be used for protein expression and purification. To separate the purification tag from the protein of interest, specific inteins that only cleave on one side have been designed (Batjargal, Walters & Petersson, 2015). The main advantage of these self-cleaving tags is the lack of using cleavage systems like proteases (Carrington *et al.*, 1989) or harsh chemicals (Rais-Beghdadi *et al.*, 1998), making it easier for large-scale productions. The mutations needed to create a one-side cleavable intein are the first residues of either the N- or C-terminus. If an intein is mutated on the first aa, the N-cleavage is blocked because the N-S/N-O acyl shift no longer occur, leaving an intein that can cleave only in the C-terminal. In a similar way, when an intein is mutated in the last aa, the C-cleavage process does not occur leading to an intein that can only cleave in the N-terminal. Using these mutants, the protein of interest can be cleaved from the purification tag (Zhao *et al.*, 2008).

#### 1.4.3.1. The case of *Mycobacterium tuberculosis recA* intein and its mutants

The *Mtu recA* is a contiguous large intein with quite a number of mini-intein ( $\Delta I$ ) mutants described. One of those mutants ( $\Delta I$ -Cleaving Mutant) has proven to be useful as a self-cleaving tag in

conjunction with ELPs (Shi, Meng & Wood, 2012). The  $\Delta$ I-Splicing Mutant ( $\Delta$ I-SM) was constructed by removing the homing endonuclease between residues 111-382 and with the mutation V67L (Valine 67 to Leucine) to restore splicing activity (Hiraga *et al.*, 2005). The  $\Delta$ I-CM was obtained by performing the mutations D24G and D422G in  $\Delta$ I-SM, ensuring enhanced activity as well as a pH-dependent cleavage of the C-terminal (Van Roey *et al.*, 2007). These two mutant inteins were considered for this dissertation.

#### *1.4.3.2. The case of Synechocystis DnaX and its mutants*

The *Ssp* DnaX intein is a contiguous intein that has not received much attention as a purification tag although it has been used for mutagenic studies (Qi, Meng & Liu, 2011). It has a mutant devised to incorporate a His-tag in the middle of the intein as a linker sequence instead of a homing endonuclease. It was reported that the splicing of the *Ssp* DnaX mini-intein was efficient, indicating the potential of this intein for the production and purification of AMPs (Qi, Meng & Liu, 2011).

#### 1.4.4. Concatenation tags

Working with inteins provides many tools to accomplish different post-translational modifications. One of these modifications is the ability to multimerise polypeptides using inteins as monomers. Split inteins are able to do splicing even though they are in separate proteins. Bowen *et al.* (2018) have used this *trans*-splicing to concatenate molecules to achieve larger sizes. For instance, dragline spideroin molecules of 556 kDa were produced using this method (Bowen *et al.*, 2018). In this work, two constructions - 96mer-Intein<sup>N</sup> and Intein<sup>C</sup>-96mer - were produced and allowed to interact with each other, resulting in the dimerization of a larger molecule composed of 192 monomers (Bowen *et al.*, 2018). These constructions used an artificial intein obtained from the consensus sequence of different split inteins that had fast splicing activity in order to obtain higher stability and speed (Stevens *et al.*, 2016).

#### *1.4.4.1. The case of Nostoc punctiforme DnaE*

*Npu* DnaE is a split intein that was considered in the design of some adapters possessing a high rate of splicing. The intein has been used for production of specific antibodies through the platform named Bispecific Antibody by Protein *Trans*-Splicing and it was used for protein labelling with fluorescent tags (Han *et al.*, 2019; Lee *et al.*, 2018).

#### 1.4.5. Conditional protein splicing

Conditional protein splicing refers to the ways splicing can be blocked or activated. One way to trigger splicing activity is to fuse split inteins to heterodimerizing domains (Schwartz *et al.*, 2006). The

most prevalent example of heterodimerization is the fusion with the FKPB-rapamycin binding domain (FRB) and FK506 binding protein (FKBP). This system have been used to increase the splicing activity of the split intein *Scd* VMA. These domains do not interact when alone, but when rapamycin is added, they are able to form a complex that increases the splicing activity of the intein (Brenzel, Kurpiers & Mootz, 2006; Schwartz *et al.*, 2006).

Caging the intein chemically is a way to block splicing activity. An intein can be blocked using disulphide bridges by placing a Cysteine in the -3 position of the N-extein. Commonly, the first residue of an intein is a Cysteine and a disulphide bond will form between the first Cys with the introduced one, preventing the N-S acyl shift and inhibiting the splicing process. This bond can be undone by introducing a reducing agent that leads to activation of splicing (Callahan, Stanger & Belfort, 2013; Cui *et al.*, 2006). Other form of intein caging is to connect the molecule 6-nitroveratryl (3,4-dimethoxy-6-nitro-benzyl, Nvl) to the C-terminal fragment of the intein that blocks splicing. This molecule can be removed by irradiation at 365 nm allowing splicing to occur (Berrade, Kwon & Camarero, 2010).

Several other examples of conditional splicing exist with different regulatory processes: through addition of a ligand, as in the *Mtu recA* mutant  $\Delta$ I-SM<sup>TR</sup> intein (Skretas & Wood, 2005), using halophilic inteins such as *Hsa* PolII and the *Hut* MCM2 inteins (Ciragan *et al.*, 2016; Reitter *et al.*, 2016), or even as in the case of the  $\Delta$ I-CM intein with pH-dependent C-terminal cleavage (Van Roey *et al.*, 2007).

## 1.5. Objectives

The main goal of this thesis was to develop a molecular biology platform, consisting of a set of “universal” cloning adapters, in which different AMP sequences could be seamless inserted into the pET25 expression system for further bioproduction using *E. coli*. To achieve the proposed goal, the cloning adapters have to combine several characteristics:

- Be compatible with integration of PBPs such as ELPs and SELPs, to be used as solubilisation tags, purification tags, to reduce AMPs toxicity and to be used as base for the creation of antimicrobial materials.
- Contain inteins as self-cleaving tags, for the purification and isolation of AMPs.
- Be compatible with incorporation of AMPs either at the N- or C- terminus.
- Inclusion of an His-tag for chromatographic purification, if needed.

In addition to the main objective, and since split inteins can provide a way to concatenate proteins, a secondary goal of this project was the design of an adapter for the concatenation/multimerization of ELP and SELP biopolymers.

## 2. Materials and Methods

### 2.1. Biological Materials

#### 2.1.1. Bacterial strains

*Escherichia coli* XL1-Blue (Agilent) was used for cloning and maintenance of plasmids while *E. coli* BL21(DE3) was used for protein expression. These strains were maintained in both liquid and solid culture media and preserved at -80 °C in 2 ml cryotubes with a glycerol solution at 30% (w/v) concentration.

#### 2.1.2. Culture media

For the growth of *E. coli*, the culture media used were Lysogeny Broth (LB, 1 L: 10 g tryptone, 5 g yeast extract, 5 g NaCl), and Terrific Broth (TB, 1L: 12 g tryptone, 24 g yeast extract, 0.017 M KH<sub>2</sub>PO<sub>4</sub>, 0.072 M K<sub>2</sub>HPO<sub>4</sub>, 5 g glycerol) supplemented with 2 g/L of lactose for auto-induction purposes (TB<sub>lac</sub>). Maintenance and growth of *E. coli* XL1-Blue and *E. coli* BL21 (DE3) was performed using LB in solid medium (2% agar-agar) and in liquid LB medium. Protein expression was achieved using *E. coli* BL21 (DE3) grown in TB<sub>lac</sub> at 37 °C, for 22 h and 200 rpm. For all media, either ampicillin (100 µg/mL) or kanamycin (50 µg/mL) were added, depending of the respective resistance-marker.

#### 2.1.3. Plasmids

The plasmids pUC57 and pDrive were used as cloning vectors in which the adapters and ELP/SELPs were inserted, respectively. The vector pET25b (+) (see Annex A) was used throughout this work as expression plasmid and as the basis for the adapters' design. All these vectors confer resistance to ampicillin giving the possibility to use ampicillin as a selection marker.

#### 2.1.4. Protein-based polymers (PBP) sequences

Table 1 displays the monomeric sequence of the protein-based polymers (PBPs) used in the experimental work. To make the constructions, ELPs and SELPs were used from previous works of the research group (Machado *et al.*, 2009; Machado *et al.*, 2013a; da Costa *et al.*, 2018). These sequences had been inserted in a pDrive and cryopreserved in *E. coli* strains at -80 °C. Only A60 and SELP-59-A



were used experimentally during this work, but other constructions from the group were also considered in the design of the adapters.

Table 1. Sequence and length of ELPs and SELP.

Protein	Sequence	Length/Size
<b>SELP-59-A</b>	S <sub>5</sub> E <sub>9</sub> -[(GAGAGS) <sub>5</sub> (VPAVG) <sub>9</sub> ] <sub>9</sub>	751 aa; 58.3 kDa
<b>A200</b>	(VPAVG) <sub>200</sub>	1001 aa; 84.8 kDa
<b>A140</b>	(VPAVG) <sub>140</sub>	701 aa; 59.4 kDa
<b>A60</b>	(VPAVG) <sub>60</sub>	301 aa; 25.5 kDa

#### 2.1.5. Antimicrobial peptides

Three different AMP sequences were considered as potential peptides for its use in the cloning steps. Table 2 depicts the sequence and length. The AMPs were obtained from pDrive plasmids from previous works of the research group but only ABP-CM4 was used experimentally during the work, although the other peptides were also considered in the design of the adapters.

Table 2. Amino acid sequences and length of the antimicrobial peptides used in this work.

Antimicrobial peptide	Aminoacid sequence	Length/Size
<b>ABP-CM4</b>	RWKIFKKIEKVGQNIRDGIVKAGPAVAVVGQAATI	35 aa; 3.8 kDa
<b>HAMP</b>	TSDTHFPICIFCCGCCHRKCGMCCKT	27 aa; 3.0 kDa
<b>BMAP-28</b>	GGLRSLGRKILRAWKKYGPIIVPIIRIG	28 aa; 3.1 kDa

#### 2.1.6. Oligonucleotides

The oligonucleotides used to amplify ABP-CM4 for the different adapters are displayed in Table 3. Among these, some – number 4 and 8 - were designed specifically to add a terminal Cysteine to the ABP-CM4 sequence for future functionalization of surfaces by grafting. As mentioned above, although only ABP-CM4 was considered for experimental work and as a proof-of-concept, the strategy for the other AMPs (HAMP and BMAP-28) was also devised and is presented in appendix (See Appendixes A and B).

Table 3. Primer sequences used to amplify ABP-CM4. Underlined nucleotides represent where the annealing with the ABP-CM4 sequence occurs.

N°	Primer	Nucleotide sequence	Length
1	FWD_A1_CM4_Wintein_Blunt	GCGTGGTGGTGCATAAT <u>CGTTGGAAAATCT</u> <u>TCAAG</u>	25 bp
2	FWD_A1_CM4_Wout_int_KpnI	AGAGGGTACCC <u>CGTTGGAAAATCTTCAAG</u>	28 bp
3	rev_A1_CM4_HindIII	ACAAAAGCTTAAATGGTTGCGGCCTG	26 bp
4	rev_A1_CM4_Hind_wcys_term	AAAAAAGCTTAGCAAATGGTTGCGGCCTG	29 bp
5	FWD_A2_CM4_Wint-c_NheI	CGCGGCTAGCAATTGCC <u>CGTTGGAAAATCTT</u> <u>CAAG</u>	34 bp
6	FWD_A2_CM4_Woutint-cNdeI	CGCGC <u>CATATGCGTTGGAAAATC</u>	22 bp
7	rev_A2_CM4_Sall	AAAAGTCGACAAATGGTTGCGGCCTG	26 bp
8	FWD_A3_CM4_Wint_Cystrm	CGCGCATATGTGCC <u>CGTTGGAAAATCTTCAA</u> <u>G</u>	31 bp
9	rev_A3_CM4_Wintein_KpnI	CGCGGGTACCTTCCGCCAGGCAAATGGTTG <u>CGGCCTG</u>	37 bp

### 2.1.7. Inteins

Table 4 displays the intein sequences used in adapter design, while Table 5 displays the conditions that trigger the desired activity for each of the chosen inteins.

Table 4. Amino acid sequences of the inteins used for the design of adapters.

Inteins	Aminoacid sequence	Length/Size
<i>Mtu recA</i>	KCLAEGTRIFDPVTGTTTHRIEDVVDGRKPIHVAAAKDGLHARPVVSWFQDQTRDVLRIAGGAIVW ATPDHKVLTEYGWRAAGELRKGDRVAQPRRFDGFGDSAPIPADHARLLCYLIGDGRDGVWGGKTPIN FINVQRALIDDVTRIAATLGCAAHPQGRISLAIHRPGERNGVADLCQQAGIYGKLAWEKTI PNWFFEP DIAADIVGNLLFGLFESDGWVSREQTGALRVGYTTTSEQLAHQIHWLLLRFVGVSTVRDYDPTQKRPSI VNGRRIQSKRQVFEVRISGMDNVTAFESVPMWVGPRGAALIQAPEATQGRRRGSQATYLAEMTDAV LNYLDERGVTAQEAAAMIGVASGDPRGGMKQVLGASRLRRDRVQALADALDDKFLHDMLAEELRYSV IREVLPTRRARTFDLEVEELHTLVAEGVVHNC	44 aa; 48.5 kDa

Table 4 (continuation).

<b><math>\Delta I</math> (<i>Mtu recA</i>)</b>	KCLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVAAAKDGLHARPVSWFDQGTRDVIGLRIAGGAIWV ATPDHKVLTEYGWRAAGELRKGDRAQPRRFDGFGDSAPIPARVQALADALDDKFLHDMMLAEELRYS VIREVLPTRRARTFDLEVEELHTLVAEGVWHNC	168 aa; 18.8 kDa
<b><math>\Delta I</math>-SM (<i>Mtu recA</i>)</b>	KCLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVAAAKDGLHARPVSWFDQGTRDVIGLRIAGGAILW ATPDHKVLTEYGWRAAGELRKGDRAQPRRFDGFGDSAPIPARVQALADALDDKFLHDMMLAEELRYS VIREVLPTRRARTFDLEVEELHTLVAEGVWHNC	168 aa; 18.8 kDa
<b><math>\Delta I</math>-CM (<i>Mtu recA</i>)</b>	KCLAEGTRIFDPVTGTTHRIEDVVGGRKPIHVAAAKDGLHARPVSWFDQGTRDVIGLRIAGGAILW ATPDHKVLTEYGWRAAGELRKGDRAQPRRFDGFGDSAPIPARVQALADALDDKFLHDMMLAEELRYS VIREVLPTRRARTFGLEVEELHTLVAEGVWHNC	168 aa; 18.8 kDa
<b><i>Ssp DnaX</i></b>	ECLTGDSQVLTRNGLMSIDNPQIKGREVLSYNETLQQWEYKQVLRWLDRGEKQTLSEIKTKNSTVRCTA NHLIRTEQGWTRAENITPGMKILSPASVDVDNLSQSTALTASLGGLSGAINYEAINDDKNTTSLSLK KQKQPDPFVNADVAKNLIQHFCSAKKEELKVSNPAGEDIPTKKATDFGISEQKKLHQGNRWEQKF SVLSTEPCLGMEVLTIPTHIADSPACDGPTAPSSQNGWNIKRQDWDVCHPKYDSQPIKAMGKVP KPVVPQTLLMFSASQSNLEVKENKFLRNGSRISLKEWLGTTWTVPSLFPNLGVHGFYSYQRAFSRK KINLLNGLPIEDIPPVQNPVQPIAEALTAKPITTKWEQWPPASGYRTWKSIPSPQWHTNFEEVESVTKGQ VEKVYDLEVEDNHNHFNGLLVHNC	430 aa; 48.4 kDa
<b><i>Ssp DnaX</i> (mini-intein)</b>	ECLTGDSQVLTRNGLMSIDNPQIKGREVLSYNETLQQWEYKQVLRWLDRGEKQTLSEIKTKNSTVRCTA NHLIRTEQGWTRAENITPGMKILSPASASGHHHHHHGGSGSPQWHTNFEEVESVTKGQVEKVYDL EVEDNHNHFNGLLVHNC	150 aa; 17.2 kDa
<b><i>Npu DnaE-n</i></b>	YCLSYETEILTVEYGLLPIGKIVEKRIECTVYSDNNGNIYQVPAQWHRGEQEVFEYCLEGDSLIRATK DHFMTVDGQMLPIDEIFERELDLMRVDNLPN	102 aa; 12 kDa
<b><i>Npu DnaE-c</i></b>	MIKIATRKYLGKQNVYDIGVERDHNFKNGFIASNC	36 aa; 4.2 kDa

Table 5. Cleavage strategies used for each of the selected intein.

Intein	Activity	Triggering conditions
<b><math>\Delta I</math>-SM (<i>Mtu recA</i>)</b>	N-terminal cleavage	pH 6.0 (Wu, Mee, Califano, Banki & Wood, 2006)
<b><math>\Delta I</math>-CM (<i>Mtu recA</i>)</b>	C-terminal cleavage	Addition of DTT (Hiraga, Derbyshire, Dansereau, Van Roey & Belfort, 2005)
<b><i>Npu DnaE</i></b>	Splicing	Addition of DTT (Ramirez, Valdes, Guan & Chen, 2012)

### 2.1.8. FKBP and FRB

FK506 binding protein and the FKBP-rapamycin binding domain were included in the sequence of the concatenation adapter. The amino acid sequences for these domains are presented in Table 6.

Table 6. Amino acid sequence and length of the FKBP and FRB domains used for the design of concatenation adapters.

Domain	Aminoacid sequence	Length/Size
FKBP	GVQVETISPGDGRTFFPKRGQTCVWHYTGMLEDGKKFDSSRDRNKPFKFMLGK	107 aa; 11.8 kDa
	QEVIRGWEEGVAQM SVGQRAKLTISP DYAYGATGHPGIIPPHATLVFDVELLKL	
	E	
FRB	ILWHEMWHEGLEEASRLYFGERNVKGMFEVLEPLHAMMERGPQTLKETSFN	93 aa; 11.3 kDa
	QAYGRDLMEAQEWCRKYMKSGNVKDLLQAWDLYYHVFRRISK	

### 2.1.9. Adapters design

The adapters were designed with the aid of SnapGene® software, chemically synthesized (NZYtech), and delivered in a pUC57 vector. Three main types of adapters were designed with different strategies in mind: i) the post-translational concatenation of polymers using split inteins; ii) creation of a self-cleaving tag that cleaves in the C-terminal of the introduced PBP; and, iii) creation of a self-cleaving tag that cleaves on the N-terminal of the PBP. Experimentally, and to test the feasibility of the devised system, four adapters were chemically synthesized (NZYtech): C-cleavage adapter, N-cleavage adapter, and two concatenation adapters without and with His-tag, considering possible purification through chromatographic processes (further represented in Figure 7). Other adapters for concatenation, with or without conditional splicing, and adapters for splicing were designed but not used experimentally. These are also represented in Figure 7 and discussed in the results section.

### 2.1.10. Oligonucleotides for Site Directed Mutagenesis (SDM)

As the adapters for concatenation presented an extra restriction site for the *AvaI* enzyme, hindering subsequent insertion of the adapter into the pET25 plasmid, site directed mutagenesis was employed to remove the extra restriction site. The sequences of the mutagenesis primers used for SDM are described in Table 7.

Table 7. Mutagenesis primers to mutate nucleotides of the concatenation adapters.

Primer	Nucleotide sequence	Length
Fwd_mut_ad2a	GTCGAAACGATTTCCGCCGGGTGATGGCC	28 bp
Rev_mut_ad2a	GGCCATCACCCGCGAAATCGTTTCGAC	28 bp

## 2.2. Methods

### 2.2.1. Plasmid extraction

All plasmids used during this work were extracted from cells using the GenElute™ HP Plasmid Miniprep kit (Sigma-Aldrich), following manufacturers' instructions. Elution of the DNA was performed in ultrapure water at 60 °C for work aliquots or in Elution Buffer at 60 °C for DNA storage. The plasmid-containing cells were previously harvested from an overnight inoculum made in LB supplemented with ampicillin grown at 37 °C and 200 rpm. The DNA was stored at -20 °C prior to use.

### 2.2.2. Competent cells

Cells were made competent by two different protocols: a RbCl protocol was used for the XL1-Blue strain while a CaCl<sub>2</sub> protocol was used for the BL21 (DE3) strain (see Appendixes C and D). The efficiency of the competent cells was measured by transforming 1 ng of pUC18 plasmid and counting the number of colony forming units (CFUs).

### 2.2.3. Bacterial Transformation via Heat Shock Method

Plasmid DNA was added to an aliquot of XL1-Blue or BL21(DE3) competent cells (200 µl) and gently mixed. The cells were incubated on ice for 30 min and heat shocked at 42 °C for 1 min. The tubes were then placed on ice for 10 min, followed by addition of LB medium up to 1 ml, and incubated at 37 °C and 200 rpm for 1 h. The cells were then centrifuged for 1 min at 12100 xg and resuspended in 100 µl of the supernatant. The suspension was spread in LB solid medium with 100 µg/ml of ampicillin and incubated overnight at 37 °C.

### 2.2.4. Plasmid preparation

The pET25 (+) plasmid was digested in two different ways: a double digestion with *Ava*I and *Nde*I; or a double digestion with *Xho*I and *Nde*I (Fast Digest™, Thermo Scientific). Digestion was carried out at 37 °C for four hours to ensure complete digestion. The vector was then dephosphorylated using FastAP Thermosensitive Alkaline Phosphatase (Thermo Scientific) for an hour, followed by enzyme inactivation at 75 °C for 10 min. The digestion products were separated on a 1% (w/v) agarose gel at 90 mV and

stained with Midori Green or GreenSafe for visualisation. The band corresponding to the linearized vectors was excised from the gel and purified using NucleoSpin Gel and PCR Clean-up kit (Machery-Nagel), following the manufacturers' instructions. The linearized plasmids were then stored at -20 °C until further use.

#### 2.2.5. Site Directed Mutagenesis

Site directed mutagenesis was achieved using the mutagenesis primers described in Table 7. The amplification reaction was performed with 25 µl of ACCUZYME™ mix (Bioline), 2 µl of each primer (10 µM), 60 ng of DNA and ultrapure water up to 50 µl, using the following conditions:

1. Initial Denaturation: 95 °C – 30 s
  2. Denaturation: 98 °C – 30 s
  3. Annealing: 41,6 °C – 60 s
  4. Extension: 68 °C – 8 min
  5. Final Extension: 68 °C – 7 min
- } X 18 cycles

After reaction, *DpnI* restriction enzyme (Thermo Scientific) was used to digest the original methylated template DNA, leaving the amplified DNA intact. *DpnI* was then inactivated at 80 °C for 20 min. XL1-Blue was transformed with the resulting plasmid via heat shock methodology. From the colonies of the transformation, an inoculum was made with LB liquid medium and pDNA was extracted with GenElute™ HP Plasmid Miniprep kit. The mutation was then confirmed by DNA sequencing.

#### 2.2.6. Insert preparation

Similarly to the double digestions done with the pET25 plasmid, the adapters that were previously inserted in the pUC57 plasmid were also doubly digested. The concatenation adapters were digested using the *AvaI* and *NdeI* enzymes, while the self-cleaving tag adapters were digested with *XhoI* and *NdeI*. The adapter was then separated from the remaining pDNA by electrophoresis in a 1% agarose gel and the corresponding band was extracted from the gel using the NucleoSpin Gel and PCR Clean-up kit (Machery-Nagel) following the manufacturer's instructions. The eluted DNA was preserved at -20 °C until further use. DNA sequences of SELP-59-A and A60 were previously inserted in pDrive between two restriction sites for the endonuclease *Eam1104I*. After digestion with *Eam1104I* (Thermo Scientific) a linear DNA sequence encoding the desired polymer with sticky ends was obtained. The digestion was

separated by electrophoresis in a 1% agarose gel and the desired band was extracted and purified with the use of the NucleoSpin Gel and PCR Clean-up kit (Machery-Nagel). The sticky ends were extended at 72 °C for 30 min with ACCUZYME™ mix (Bioline) (1:1), in order to obtain blunt ends. The blunt-ended DNA fragment was then purified with NucleoSpin Gel and PCR Clean-up kit (Machery-Nagel), and stored at -20 °C.

Before preparing the genetic sequence of ABP-CM4 for insertion, the ABP-CM4 in a pUC57 plasmid was firstly amplified via Polymerase Chain Reaction (PCR) with ACCUZYME™ mix (Bioline) and the primers described in Table 3. The amplicon was then prepared for insertion by amplification with the primers described in Table 8. The PCR conditions used were as follows:

1. Initial Denaturation: 98 °C – 3 min
  2. Denaturation: 98 °C – 15s
  3. Annealing: 48 °C – 15s
  4. Extension: 72 °C – 1,5 min
  5. Final Extension: 72 °C – 1 min
- } X 30 cycles

The PCR product was then purified using NucleoSpin Gel and PCR Clean-up kit (Machery-Nagel). The amplicons were digested using restriction enzymes from Thermo Scientific according to Table 8, and inactivated at 80 °C for 20 min.

Table 8. Oligonucleotides used for each amplification and restriction enzymes used for digestion as well as the target adapter for insertion.

Amplification	Primers used	Restriction enzymes used	Target Adapter
1	FWD_A1_CM4_Wintein_Blunt rev_A1_CM4_HindIII	<i>HindIII</i>	N-cleavage adapter
2	FWD_A1_CM4_Wout_int_KpnI rev_A1_CM4_HindIII	<i>KpnI</i> and <i>HindIII</i>	N-cleavage adapter
3	FWD_A1_CM4_Wintein_Blunt rev_A1_CM4_Hind_wcys_term	<i>HindIII</i>	N-cleavage adapter
4	FWD_A1_CM4_Wout_int_KpnI rev_A1_CM4_Hind_wcys_term	<i>KpnI</i> and <i>HindIII</i>	N-cleavage adapter
5	FWD_A2_CM4_Woutint-cNdeI	<i>NdeI</i> and <i>SalI</i>	C-cleavage adapter

	rev_A2_CM4_Sall		
6	<u>FWD_A2_CM4_Woutint-cNdel</u> rev_A3_CM4_Wintein_KpnI	<i>NdeI</i> and <i>KpnI</i>	C-cleavage adapter
7	<u>FWD_A3_CM4_Wint_Cystrm</u> rev_A2_CM4_Sall	<i>NdeI</i> and <i>SalI</i>	C-cleavage adapter
8	<u>FWD_A3_CM4_Wint_Cystrm</u> rev_A3_CM4_Wintein_KpnI	<i>NdeI</i> and <i>KpnI</i>	C-cleavage adapter
9*	<u>FWD_A2_CM4_Wint-c_NheI</u> rev_A2_CM4_Sall	<i>NheI</i> and <i>SalI</i>	Concatenation adapter
10*	<u>FWD_A2_CM4_Woutint-cNdel</u> rev_A2_CM4_Sall	<i>NdeI</i> and <i>SalI</i>	Concatenation adapter

\*this combination of primers was not used experimentally but the strategy was devised.

### 2.2.7. Insertion of adapters into pET25

The digested adapters and digested pET25 plasmid were ligated using the T4 DNA ligase (NZYtech) with 1:5 molar ratio of vector:insert. The ligation was performed overnight at room temperature and further used to transform *E. coli* XL1-Blue. From the transformation plates, colonies were selected and used to inoculate sterile LB for further plasmid extraction, as described previously. The extracted DNA was then digested with either *NdeI* and *XhoI* or *NdeI* and *AvaI* and separated in a 1% agarose gel to confirm the insertion. The adapters' sequence was confirmed by DNA sequencing and cells were cryopreserved at -80 °C.

### 2.2.8. Preparation of the adapters for ELP/SELP insertion

All adapters have an *Eco72I* restriction site to allow the insertion of SELP or ELP through blunt ligation. The pUC57 and pET25 plasmids containing the adapters were digested with *Eco72I* Fast Digest (Thermo scientific) for 4 hours, followed by heat inactivation at 80 °C for 10 min. The linearized plasmid was dephosphorylated using FastAP Thermosensitive Alkaline Phosphatase (Thermo Scientific) for 1 hour, and inactivated at 75 °C for 10 min. The DNA was preserved at -20 °C until further use.



### 2.2.9. Insertion of ELP/SELP

The blunt-end ligation between adapters and ELP/SELP was performed following instructions for the T4 DNA Ligase (Thermo Scientific) while using a 50% PEG4000 solution (Thermo Scientific) to increase the ligation efficiency. After 2 hours of ligation, XL1-Blue was transformed with the ligation product and grown colonies were used to inoculate LB medium for DNA extraction. The insertion was confirmed by digestion with *Ppu211* (Figure 6) and by DNA sequencing. The DNA was stored at -20 °C and cells cryopreserved at -80 °C.

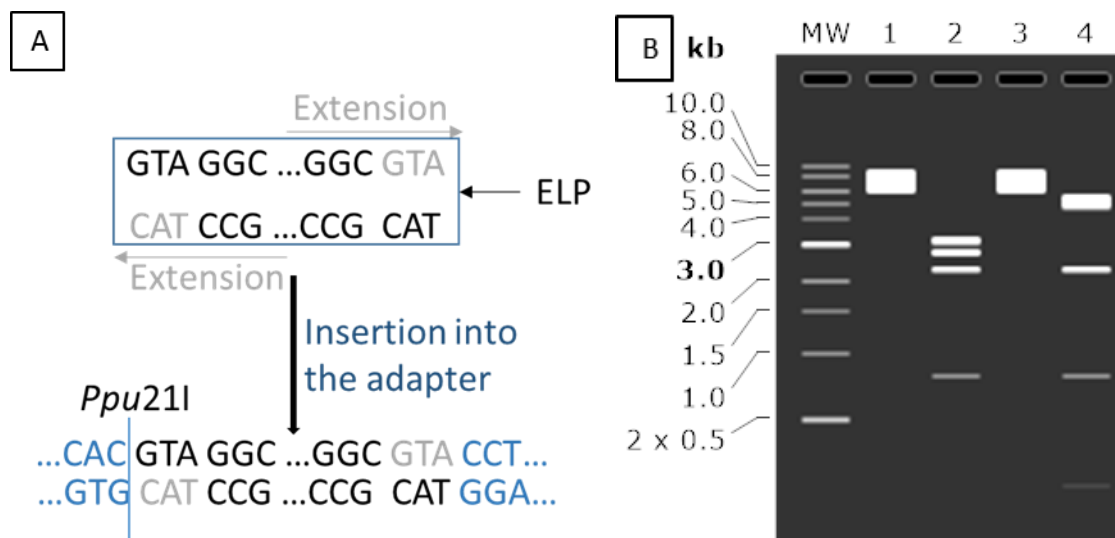


Figure 4. A- Representation of insertion confirmation with *Ppu211* restriction enzyme. B- Simulation of a 1% agarose gel with SnapGene®: Lane 1 (digestion with *NdeI*) and lane 2 (digestion with *Ppu211*) represent the electrophoretic pattern of correctly inserted A60. Lane 3 (digestion with *NdeI*) and lane 4 (digestion with *Ppu211*) represent the digestion pattern of incorrectly inserted A60.

### 2.2.10. Preparation of adapters for AMP insertion

The N-cleavage adapter was double digested in two different ways: with *HindIII* and *Eco147I* to insert ABP-CM4 while keeping the intein, and with *HindIII* and *KpnI* to insert the ABP-CM4 and remove the intein (for the creation of an ELP with antimicrobial properties). The C-cleavage adapter was also double digested in two different ways: with *KpnI* and *NdeI* to insert ABP-CM4 while keeping the intein, and with *NdeI* and *SaI* to insert the ABP-CM4 and remove the intein. The digested adapters were separated in a 1% agarose gel and extracted using NucleoSpin Gel and PCR Clean-up kit (Machery-Nagel). After extraction, the DNA was dephosphorylated using FastAP Thermosensitive Alkaline Phosphatase (Thermo Scientific) for 1 h and inactivated at 75 °C for 10 min. The DNA was stored at -20 °C.

### 2.2.11. Insertion of ABP-CM4

The insertion of ABP-CM4 into the adapters was performed with T4 DNA ligase (Thermo Scientific) supplemented with 50% PEG4000 for the blunt-end ligations, and T4 DNA ligase (NZYtech) for the sticky-end ligations, followed by incubation at room temperature for 2 h or overnight, respectively. *E. coli* XL1-Blue was transformed with the ligation product, and positive colonies were used to inoculate LB and grown overnight at 37 °C. After the incubation period, the plasmid DNA was extracted and the constructions were confirmed by DNA sequencing. The plasmids with the correct construction were conserved at -20 °C and cells stored at -80°C.

### 2.2.13. Production screenings

Due to the large number of constructions, the feasibility of the devised genetic platform was assessed only with representative constructions. For this, selected plasmid constructions were transformed into *E. coli* BL21 (DE3) and used for protein expression studies. Initial production screening studies were carried out with transformed colonies grown in TB<sub>lac</sub> for 22 h at 37 °C and 200 rpm. Non-transformed (empty) *E. coli* BL21 (DE3) was used as negative control and cells transformed with previously obtained CM4-A60 or SELP-59-A were used as positive controls. After incubation, the OD<sub>600</sub> of each sample was measured and 1 ml of each culture was collected by centrifugation and frozen at -20 °C. The pH of the cell culture was also measured to ensure it was not affecting the pH-sensitive inteins. The cell pellets were then resuspended in 100 µl TE buffer (10 mM Tris pH 8; 1 mM EDTA, pH 8) and 25 µl of loading buffer (see Appendix E), followed by centrifugation at 13300 xg for 10 min. The soluble fraction of each sample was analysed by SDS-PAGE, in a 10% polyacrylamide gel (see Appendix E) and separated at 15 mA for 2 h. For each sample, the volumes to be loaded were normalized to an OD<sub>600</sub> of 0.1 according to the following the formula:

$$\text{OD}_{600} \times X = 0.1 \times 125 \mu\text{l}$$

where, OD<sub>600</sub> represents the optical density of the cell culture measured at 600 nm, X is the volume of sample to be loaded in the gel, 0.1 is the normalization factor, and 125 µl is the final volume used to resuspend the cell pellets (100 µl of TE + 25 µl of loading buffer).

After electrophoresis, the gels were negatively stained with 0.3 M of copper chloride, recorded and visually compared in order to determine the best protein producing colony. Additionally, Coomassie

Blue staining (See Appendix E) was also used in the negatively stained gels, preceded by a destaining step with TE buffer (10X). The best producing colonies were cryopreserved at -80°C for further use.

#### 2.2.14. Protein Purification by Inverse Transition Cycling (ITC)

For purification of the constructions with the ELP A60, the best producing colonies were inoculated in 2 L of TB<sub>lac</sub> medium and incubated for 22 h at 37 °C and 200 rpm. After the incubation period, cells were collected by centrifugation at 10400 xg at 4 °C for 10 min and resuspended in 60 ml of TE buffer. While maintained on ice, the resuspended cells were lysed by ultrasonic disruption using a Vibra-Cell sonicator (750 watts, Biolock Scientific) at 60% amplitude with 3 s of pulse on, 9 s of pulse off and a total sonication time of 10 min. For samples without inteins sensitive to pH, the pH of the cell crude extract was set up at ~3.5 and incubated overnight at 4 °C, followed by centrifugation at 10700 xg for 20 min. For the samples with inteins, the cell crude extract was immediately centrifuged for three times at 15000 xg for 30 min. Purification of the recombinant proteins was achieved by employing three ITC rounds consisting of hot-cold cycles with hot and cold incubation and centrifugation steps (See Annex R for a schematic representation of the purification protocol). For each hot cycle, NaCl 0.5 M was added to the protein solution and incubated at 37 °C for 2 h, followed by centrifugation at 40 °C for 20 min at 10700 xg. The supernatant is discarded and the pellet is resuspended in ice-cold ddH<sub>2</sub>O. The resuspended pellet is settled overnight at 4°C with agitation (cold cycle), followed by a cold centrifugation at 4 °C for 20 min at 10700 xg. After centrifugation, the supernatant is transferred to a new tube (the pellet is discarded) and the hot cycle is re-initiated. After the ITC, the protein was dialysed to remove salts and lyophilized.

#### 2.2.15. Cleavage test

The cleaving of the  $\Delta$ I-CM intein was tested by incubating the purified ELP-int-AMP (1.1 mg/ml) protein at room temperature overnight in a cleaving solution (Wu *et al.*, 2006). The cleaving solution consists of PBS buffer (See appendix F) with 20 mM of MES, pH 6.0, and 2 mM of EDTA. An additional cleaving test was also performed using a cleaving solution with 10 mM of DTT. TE buffer, solutions at pH 8.0, and ddH<sub>2</sub>O were used as negative controls. Since the SDS-PAGE loading buffer contains  $\beta$ -mercaptoethanol, which is a reducing agent, experiments were performed with and without  $\beta$ -mercaptoethanol.

The cleaving test of  $\Delta$ I-SM intein was performed by incubating the AMP-int-ELP protein (1.5 mg/ml) overnight at room temperature with different concentrations of DTT (Hiraga *et al.*, 2005): 0, 50 and 200 mM. For each concentration of DTT, the assay was repeated in ddH<sub>2</sub>O and in TE buffer, pH 8.0, to assess the influence of the pH in cleavage activity. After incubation, the samples were analysed by SDS-PAGE with Coomassie Blue staining (See Appendix E), followed by AgNO<sub>3</sub> staining (See Appendix F).

#### 2.2.16. Concatenation test

The concatenation constructions (1.0 mg/ml) were tested with five different combinations of solutions incubated overnight at room temperature: ddH<sub>2</sub>O, 50 mM of DTT, 100  $\mu$ M rapamycin (0.1% of DMSO), 0.1% of DMSO and 50 mM of DTT with 100  $\mu$ M rapamycin (0.1% of DMSO) (Brenzel, Kurpiers & Mootz, 2006; Ramirez, Valdes, Guan & Chen, 2012; Schwartz, Saez, Young & Muir, 2006). As mentioned above, since the loading buffer used for SDS-PAGE analysis contains  $\beta$ -mercaptoethanol, which is a reducing agent, experiments were performed with and without  $\beta$ -mercaptoethanol.

#### 2.2.17. Western Blot

The insertion of a His-tag in the constructions allows its detection by Western Blot, using an anti-His6 and anti-mouse as primary and secondary antibodies, respectively. The protocol used for Western Blotting is described in Appendix G.

### 3. Results and Discussion

#### 3.1. Adapter design

As mentioned earlier, different types of adapters were designed (Figure 7) using SnapGene®. Adapters intended for the production of AMPs were designed using self-cleaving tags, either at the C-terminal region, at the N-terminal region or between the AMP sequence. Adapters designed for the post-translational concatenation of biopolymers were designed considering the use of split inteins. Due to the high number of devised strategies as well as due to time constraints, the experimental work considered a total of four adapters: one for each type of self-cleaving tag (N- and C-terminus cleavage) and two regarding the post-translational concatenation (with and without His-tag). The remaining adapters will be discussed throughout the design section of the results but were not used experimentally. The adapters to be inserted into the pET25 vector are flanked by sequence recognition sites for *NdeI* and *XhoI* or *NdeI* and *AvaI*. The *Eco72I* restriction site was used for the blunt-end insertion of biopolymers, a strategy that had been successfully used in previous works from the group. Moreover, it is a strategic restriction site because it allows the verification of gene orientation through digestion with *Ppu21I* enzyme.

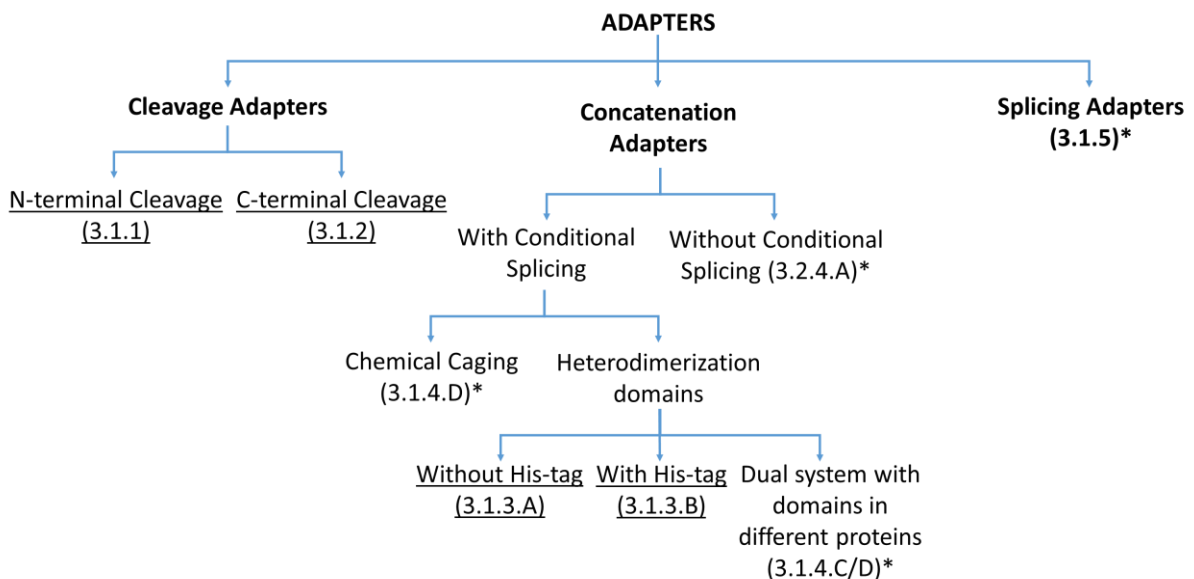


Figure 5. Overview of all the designed adapters. The experimentally used adapters are underlined while the others are identified with an asterisk (\*).

### 3.1.1. N-terminal cleavage adapter

This adapter was designed for AMP purification and contains a N-terminal cleavage intein (Figure 8). The intein used for this adapter was the *Mtu recA* mutant  $\Delta$ I-SM, which is a mini intein version of the original intein, the *Mtu recA*, with similar splicing rates (Hiraga *et al.*, 2005). For simplicity, this intein also was given the name of int<sub>Red</sub>, where 'Red' refers to the reducing agent that triggers the splicing. An Alanine replaced an Asparagine as the last aa of the intein in order to block C-cleavage. For AMP insertion, the *NdeI* and *KpnI* enzymes are used or alternatively, *NdeI* and *SalI*, if the intention is to remove the intein between the AMP and the inserted biopolymer. The *KpnI* restriction site is present in the beginning of the intein for a seamless integration of the AMP thus avoiding adding extra aa to the AMP sequence. For intein removal with the insertion of the AMP, the *SalI* restriction enzyme was utilized together with *NdeI*. The sequence recognized by *SalI* digestion is partially shared with the recognition sequence of *Eco72I* (utilized for the insertion of the SELP/ELP) to avoid the insertion of extra aa. Furthermore, *SalI* sequence also shares bases with the codon coding for the Cysteine introduced at the end of the intein sequence. The introduction of a His-tag into the adapter was also considered for Western blotting and chromatographic purification purposes. This adapter was designed to for two different constructions: AMP-int<sub>Red</sub>-SELP/ELP or AMP-SELP/ELP.

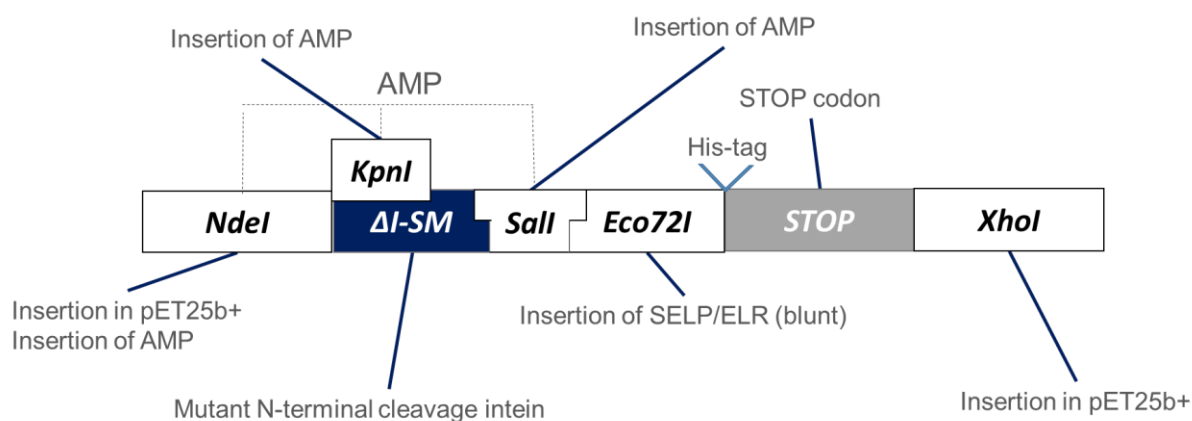


Figure 6. Schematic representation of the adapter designed for N-terminal self-cleaving.

### 3.1.2. C-terminal cleavage adapter

Similarly to the N-terminal cleavage adapter, the C-terminal cleavage adapter (Figure 9) uses a mini-intein mutant of the *Mtu recA* intein (Van Roey *et al.*, 2007). The  $\Delta$ I-CM was chosen as it shows enhanced C-cleavage activity with its first aa (Cysteine) replaced by an Alanine to block N-terminal

cleavage. Furthermore, this intein shows a pH-dependent activity, triggered when pH is lowered to 6 or below (Wu *et al.*, 2006). As such, this intein was named  $\text{int}_{\text{pH}}$ . The AMP sequence can be inserted using the restriction sites *HindIII* and *StuI* or *HindIII* and *KpnI* if the aim is to remove the intein. Even though the adapter already contains a stop codon, it leaves two extra aa resulting from the restriction enzyme (*HindIII*) recognition sequence. To avoid it, another stop codon can be inserted by PCR at the end of the AMP sequence for a seamless insertion of the peptide. The *StuI* restriction site is placed near the end of the intein to avoid the addition of extra aa to the AMP. The *KpnI* was added between the *Eco72I* and the intein to make possible the removal of the intein. This adapter also has an His-tag incorporated in the design and was designed to produce mirrored constructions to the first described adapter - either SELP/ELP- $\text{int}_{\text{pH}}$ -AMP or SELP/ELP-AMP.

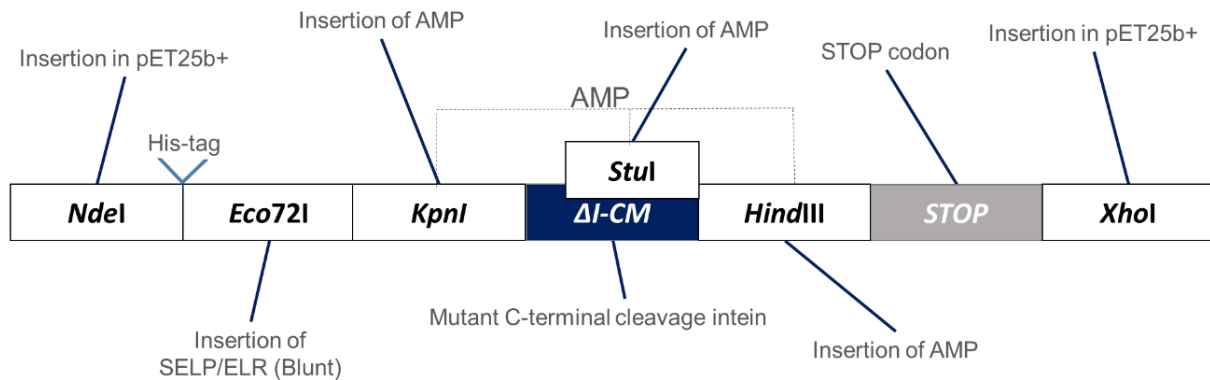


Figure 7. Schematic representation of the adapter designed for C-terminal self-cleaving.

### 3.1.3. Concatenation adapters

The main objective of these adapters (Figure 10) is to create multimerized forms of the insert polymers through concatenation. The adapters use the *Npu* DnaE split intein, with no mutations, and divided into N-intein ( $\text{int}_{\text{N}}$ ) and C-intein ( $\text{int}_{\text{C}}$ ) (Ramirez *et al.*, 2012). The FRB and FKPB domains were added to condition the splicing and dimerization in order to occur only after addition of rapamycin (Brenzel, Kurpiers & Mootz, 2006; Schwartz *et al.*, 2006). For simplicity, the fused domains FRB- $\text{int}_{\text{C}}$  and  $\text{int}_{\text{N}}$ -FKBP were named, respectively,  $\text{int}_{\text{con-C}}$  and  $\text{int}_{\text{con-N}}$ , where 'con' refers to the concatenation purpose of the adapter. Although initially considered, this adapter also allows the insertion of AMPs in two different ways: using both *NdeI* and *SaI*, as in previous adapters, to produce an AMP-SELP/ELP- $\text{int}_{\text{N}}$ -FKBP sequence that can only be fused to other sequences at its C-terminus; or *SaI* and *NheI*, to produce

a FRB-int<sub>C</sub>-AMP-SELP/ELP-int<sub>N</sub>-FKBP that can be concatenated at both terminus. Two adapters were chemically synthesized with this objective: one without (Figure 10A) and another with (Figure 10B) an His-tag for affinity chromatography purification or Western-blot detection purposes. However, one disadvantage of these adapters relies on the potential circularization of the protein if the split inteins of the same protein splice with each other. While it represents a possible disadvantage for the specific objective of the adapters, it can be useful for the production of circular AMPs (Garcia *et al.*, 2011). The primary constructions obtained with this adapter were termed as int<sub>con</sub>-C-SELP/ELP-int<sub>con</sub>-N.

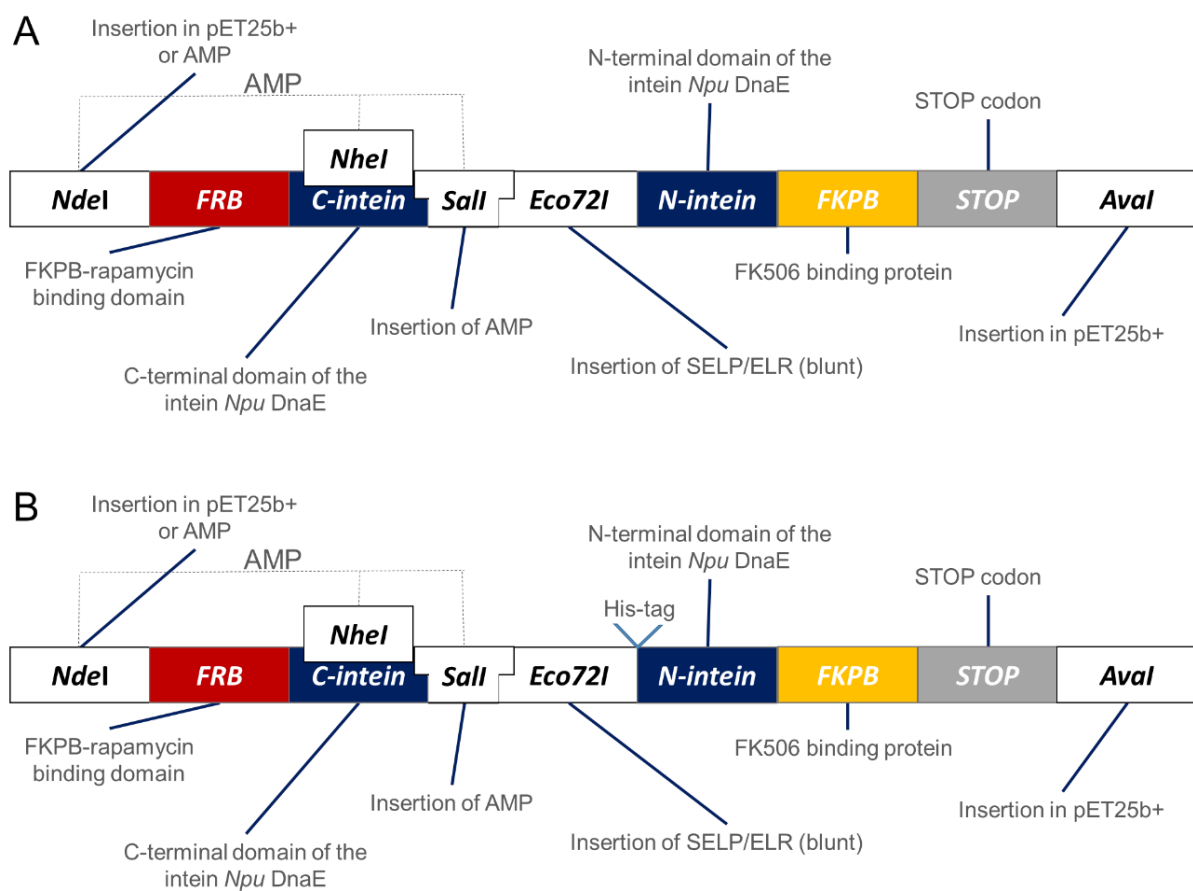


Figure 8. Schematic representation of the designed adapters. A- Concatenation adapter without His-tag. B- Concatenation adapter with His-tag.

#### 3.1.4. Variants of the concatenation adapter

The chemically synthesized concatenation adapters were preceded by different designs that were not experimentally used (Figure 11). The first design had a simple objective: allow the multimerization of SELPs/ELPs (Figure 11A). It presents restriction sites to insert the biopolymer flanked by both C-intein



and N-intein. The main disadvantage of this adapter is the lack of conditioned splicing, meaning there is less control of the splicing process. This design was the basis for most of the designed adapters for concatenation, either by heterodimerizing domains (Figure 10; Figure 11B; Figure 11C) or by having the intein chemically caged using disulphide bonds (Figure 11D).

Adapters from Figures 11B and 11C represent a dual system very similar to the adapter from Figure 10. The differing point is the heterodimerization domain present in the construction. While the adapters from Figure 10 have both FRB and FKBP domains in the same construction, the constructions represented in Figures 11B and 11C only have either FRB (Figure 11B) or FKBP (Figure 11C) at the both terminals of the construction. The goal behind this arrangement is to reduce the potential circularization of the biopolymer, since FRB only dimerize with FKBP after rapamycin addition, and do not dimerize with domains of the same nature. As main disadvantage, it was identified the need to produce both proteins for subsequent concatenation.

The adapter from Figure 11D uses a chemical caging method to produce a conditional splicing for concatenation (Callahan, Stanger & Belfort, 2013). A Cysteine was added in the -3 position of the extein relative to the N-intein. This Cysteine forms a disulphide bond with the Cysteine present in the first position of the N-intein preventing the characteristic N-O/S-O acyl shift that starts the protein splicing process. This bond can then be removed by the addition of a reducing agent allowing splicing to occur. The reducing conditions inside a cell can, however, result in the breakage of the disulphide bridge leading to a loss of effectiveness of the caging method.

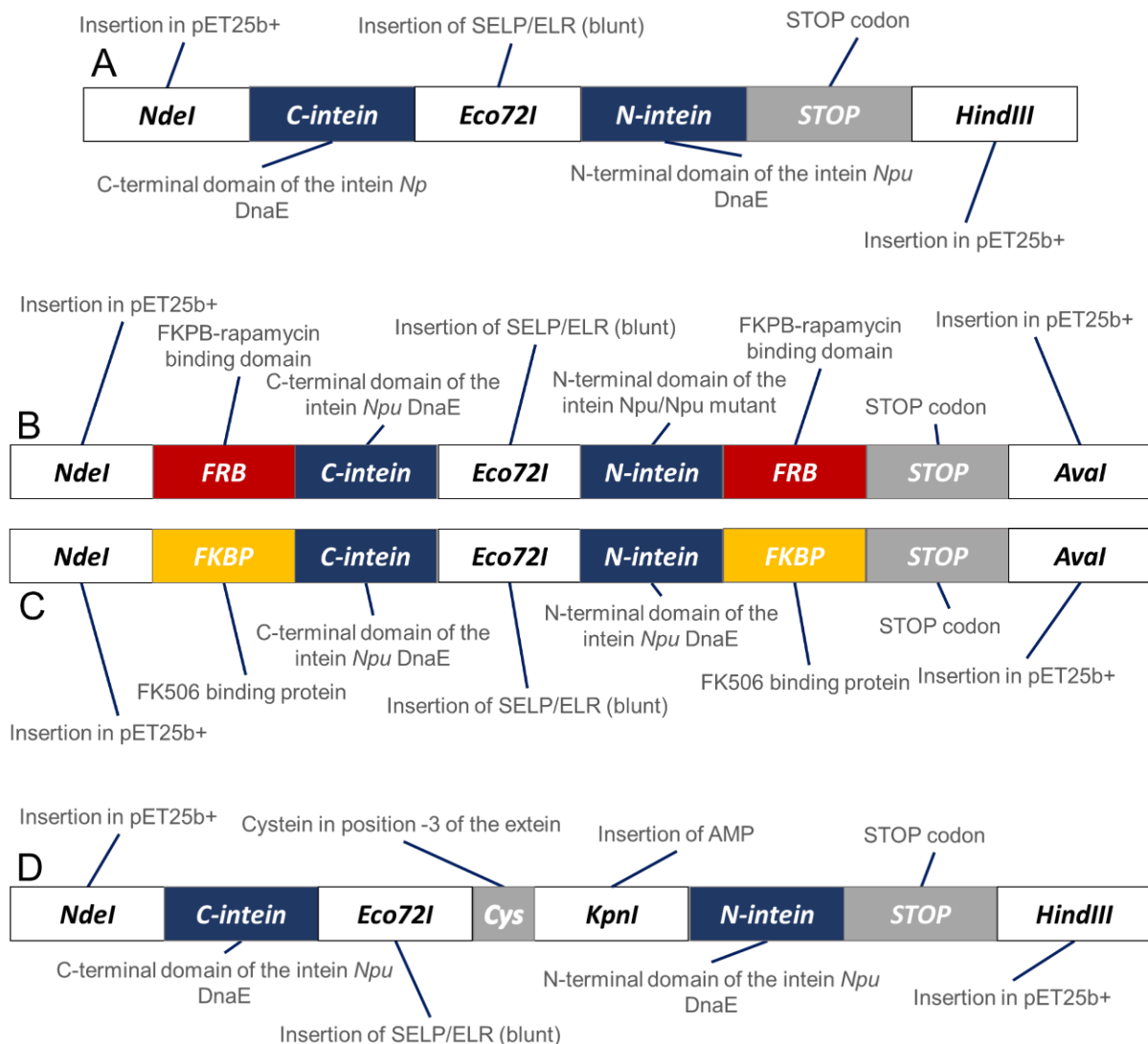


Figure 9. Schematic presentation of the designed concatenation adapters. A- Concatenation adapter with no conditional splicing. B- Concatenation adapter flanked by FRB domains. C- Concatenation adapter flanked by FKBP. The constructions of the adapters B and C are part of the same dual system for protein multimerization. D- Concatenation adapter conditioned by chemical caging with disulphide bridges. To note that these designs were not used in experimental work.

### 3.1.5. Alternative adapter for the production of AMPs

An alternative adapter was designed for the production of AMPs that induce toxicity to the host which can result in higher production yields. In this strategy, the AMP is divided in two parts that are later joined together by splicing (Qi, Meng & Liu, 2011). The intein is a mini-intein mutant of the *Ssp* DnaX intein with an His-tag (Qi, Meng & Liu, 2011). This adapter does not rely on the use of the ELP/SELP biopolymers as purification tags and, instead, relies on the use of an His-tag for affinity chromatography purification. A Cysteine, Serine or Threonine in the position +1 of C-extein is essential for the splicing to occur (Perler, 2002); meaning that, if the AMP sequence does not hold any of these residues with OH- or

SH-groups, the system is not functional. This fact is not a limiting issue for AMPs such as HAMP or BMAP-28, as they both include Cysteines or Serines; however, it is not the case of ABP-CM4. To work with other AMPs such as ABP-CM4, there is the need to add one of the required amino acids to the middle of the sequence. Nevertheless, this could result in differences of structure and activity. Thus, this system is specific of the AMP sequence. As a result, oligonucleotides were designed to amplify HAMP and BMAP-28 to prepare its insertion into this adapter (See Appendix A and B for information on the primers). The *NdeI* and *Eco72I* restriction sites are used to insert the N-terminal half of the AMP while the *BcuI* and *HindIII* are used to insert the C-terminal half of the AMP. The *Eco72I* site shares some of its sequence with the *Ssp DnaX* mini-intein allowing for the seamless insertion of an AMP half, meaning that no extra aa are added. Similarly, the *BcuI* site is situated near the end of the intein so that the other half can be inserted with no extra aa.

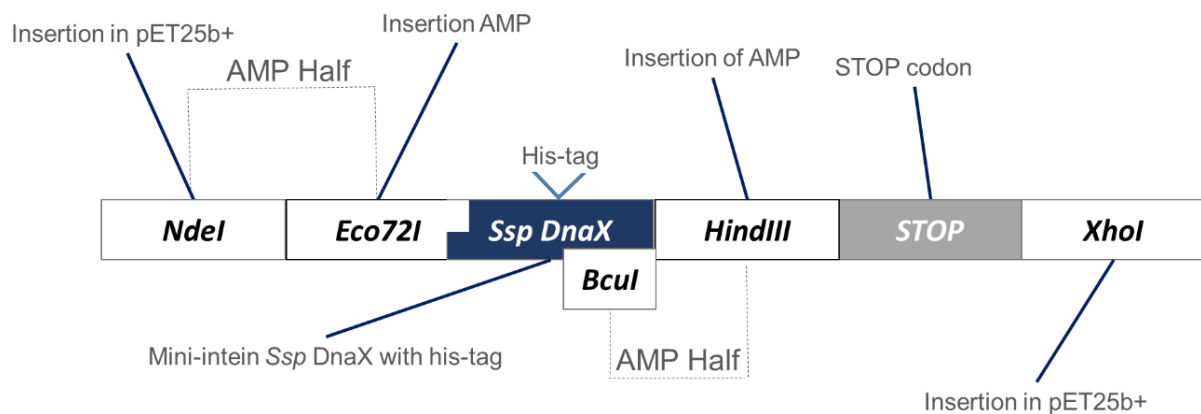


Figure 10. Schematic representation of the adapter designed for the production of parted AMPs relying on the splicing mechanism to join each part/half.

### 3.2. Construction of the expression vector

The selected adapters for the experimental work were chemically synthesized and delivered in pUC57 vectors. The following results reflect the cloning steps used for inserting the adapters into the pET25b (+) expression vector.

#### 3.2.1. Site directed mutagenesis

The pUC57 vectors containing the adapters for concatenation were submitted to site directed mutagenesis using the primers of Table 7, in order to remove an *AvaI* restriction site. Two PCR reactions were performed, one for each adapter. The amplification products were separated in a 1% agarose gel

(Figure 13) with results showing a band near 4000 bp (between 3 kb and 4 kb), corresponding to the size of the pUC57 with adapter (~3750 bp).

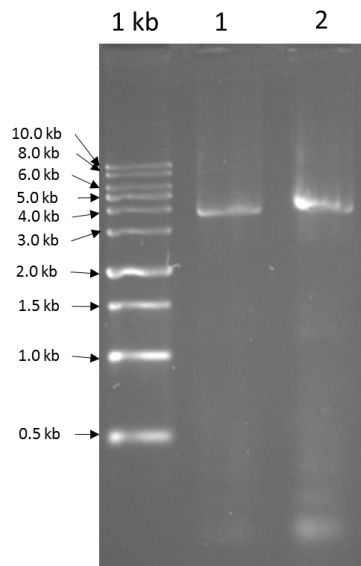


Figure 11. Electrophoretic analysis of the site directed mutagenesis amplification product, in a 1% agarose gel. Lane 1 corresponds to the PCR reaction of the concatenation adapter without his-tag. Lane 2 corresponds to the PCR reaction of the concatenation adapter with His-tag.

### 3.2.2. Insertion in vector pET25

After SDM, the pUC57 plasmids were digested with *AvaI* and *NdeI* and ligated to linearized and dephosphorylated pET25, previously digested with the same enzymes. The ligation was confirmed by digestion with *Eco72I* and *Kpn2I* (Figure 14), showing the expected molecular weight bands. The plasmids with the concatenation adapters with and without His-tag were named, pA2H and pA2 $\emptyset$ , respectively. The final construction was then confirmed through DNA sequencing (see Annexes B, C, D and E for the sequencing chromatograms).

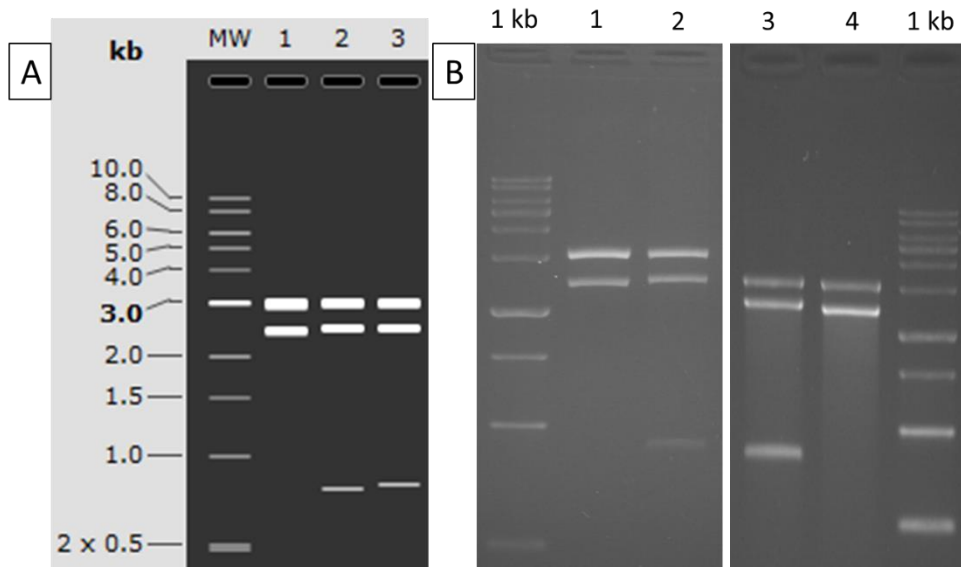


Figure 12. A- *In silico* simulation (SnapGene®) of a 1% agarose gel showing the digestion products after restriction digestion with *Kpn*I and *Eco*II. B- Electrophoretic analysis in a 1% agarose gel of the digestion with *Eco*II and *Kpn*I. Lanes 1 and 4 correspond to the digestion of the pET25 vector (used as control); Lane 2 is the digestion of pA2Ø; Lane 3 is the digestion of pA2H.

The ligation of the adapters with a self-cleaving tag was also confirmed by restriction digestion and is depicted in Figure 15. Both adapters with N-terminal cleavage and C-terminal cleavage were digested with *Kpn*I and *Kpn*II, showing the successful insertion of the adapters. The resulting plasmids for the C-terminal cleavage tag and the N-terminal cleavage tag were named, respectively, pA1C and pA3N. These were further confirmed by DNA sequencing (see Annexes F and G for the sequencing chromatograms).

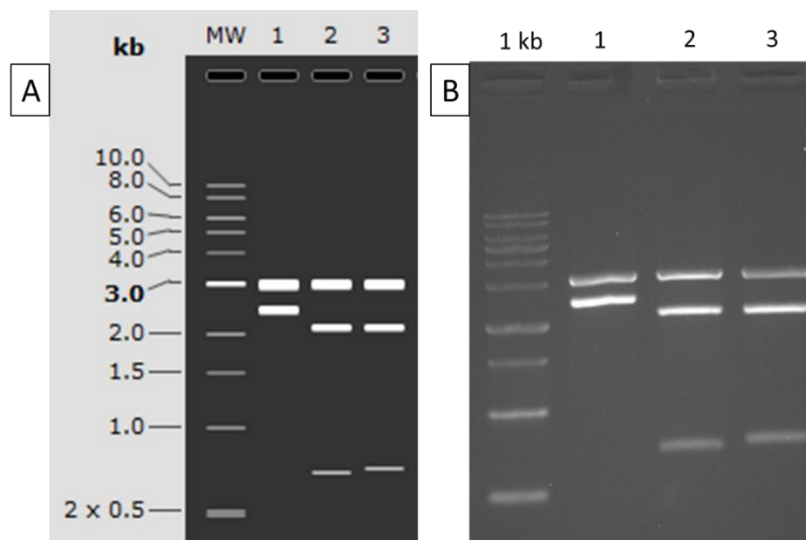


Figure 13. A- *In silico* simulation (SnapGene®) of a 1% agarose gel showing the digestion products after restriction digestion with *Kpn*II and *Kpn*I. B- Electrophoretic analysis in a 1% agarose gel of the digestion with *Kpn*I and *Kpn*II. Lane 1 correspond to the digestion of the pET25 vector (used as control); Lane 2 is the digestion of pA1C construction; Lane 3 is the digestion of the pA3N vector.

### 3.2.3. Insertion of the ELPs

Due to the high number of devised constructions, work was focused in using only the ELP A60. Nevertheless, the designed strategy is compatible with A140 and A200. The decision to use A60 in detriment of A140 or A200 was based on the size of the DNA sequences as the larger molecular weight of A140 and A200 can hamper the sub-cloning steps. The blunt-end ligation of A60 DNA sequence into pA2H and pA2 $\emptyset$  plasmids (concatenation) was confirmed by digestion with *Ppu211* (Figure 16). Results show positive clones for each of the constructions (lanes 3, 6 and 10) but also, as a consequence of the blunt-end ligation, an incorrectly oriented insert (lane 8). The samples were further confirmed by DNA sequencing (see Annexes H and I for the sequencing chromatograms).

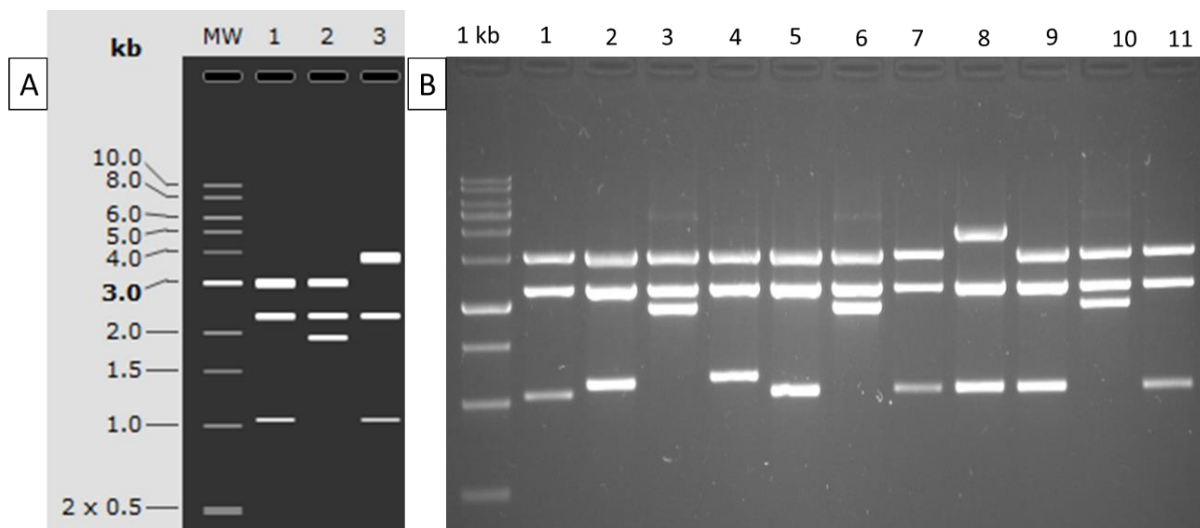


Figure 14. A- *In silico* simulation (SnapGene®) of a 1% agarose gel showing the digestion products after restriction digestion with *Ppu211*: Lane 1 corresponds to the pET25 with the concatenation adapter (used as control); Lane 2 represents the correct insertion of A60 into the adapter; Lane 3 represents the insertion of incorrectly oriented A60. B- Electrophoretic analysis in a 1% agarose gel of the digestion with *Ppu211*: Lanes 1 and 11 correspond to the pA2H and pA2 $\emptyset$ , respectively; Lanes 2-6 show the results of digestion, derived from different colonies transformed with the ligation product of A60 with pA2H; Lanes 7-10 show the results of digestion, derived from different colonies transformed with the ligation product of A60 with pA2 $\emptyset$ . Lanes 3, 6, 8 and 10 show the insertion of A60 into the vector, although showing an incorrect orientation in Lane 8.

The insertion of A60 gene into the pA3N vector (N-terminal cleavage) was confirmed by restriction digestion with *KpnI*. However, this does not allow to infer on the orientation of the insert which was only possible after DNA sequencing. In Figure 17, it is possible to see the difference between the negative control and the DNA of positive colony indicating the insertion of the A60 (~900 bp). This construction was then confirmed by sequencing (see Annex J for the sequencing chromatogram).

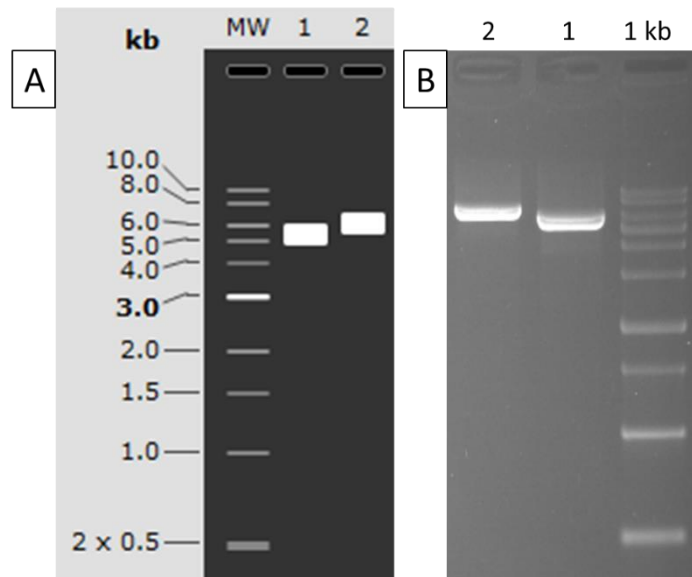


Figure 15. A- *In silico* simulation (SnapGene®) of a 1% agarose gel showing the digestion products after restriction digestion with *KpnI*. B- Electrophoretic analysis in a 1% agarose gel of the digestion with *KpnI*. Lane 1 corresponds to the pA3N vector; Lane 2 shows the result obtained for a representative colony derived from the transformation of the ligation of A60 with pA3N.

Ligation of A60 to the pA1C vector (C-terminal cleavage) was confirmed using the *NdeI* and *XhoI* enzymes that flank the adapter (Figure 18). Results show one positive insertion with the expected molecular weight (Lane 9), but also two insertions with lower size than the expected for the A60 sequence (Lanes 2 and 4). The smaller size is likely related to DNA contamination occurring upon digestion of A60 and purification from the pDrive source plasmid. The construction further confirmed by DNA sequencing (see Annex K for the sequencing chromatogram). For convenience, the A60 protein was named either A60 or A60<sub>his</sub> depending if it holds or not a His-tag in the C-terminal of A60. Alternatively, if the His-tag sequence was present in N-terminal of the A60 it was termed <sub>his</sub>A60.

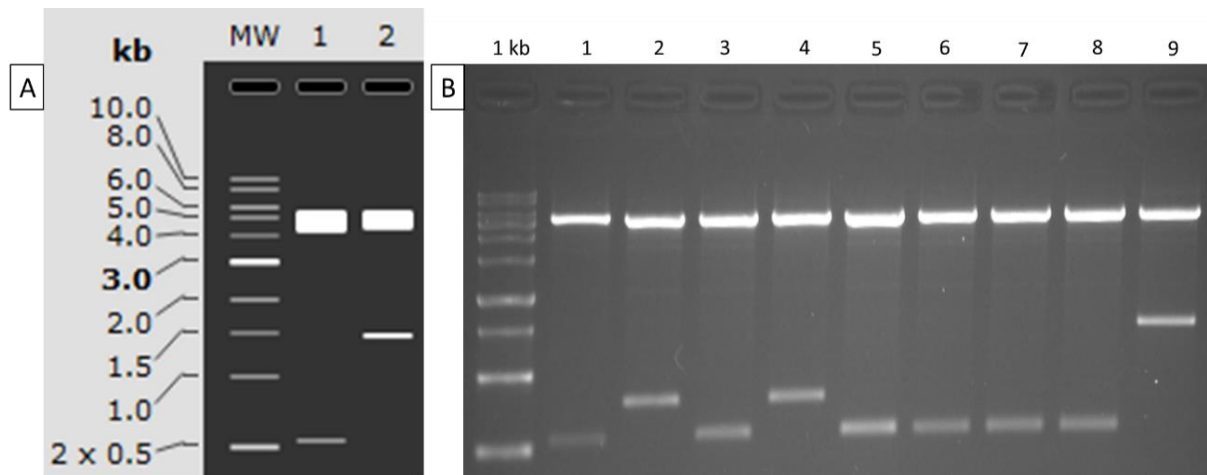


Figure 16. A- *In silico* simulation (SnapGene®) of a 1% agarose gel showing the digestion products after restriction digestion with *XhoI* and *NdeI*. B- Electrophoretic analysis in a 1% agarose gel of the digestion with *XhoI* and *NdeI*. Lane 1 corresponds to the pA1C vector; Lanes 2-9 correspond to colonies resulting from the ligation of A60 and the pA1C vector.

### 3.2.4. Insertion of SELP

Due to its biotechnological interest, the SELP SELP-59-A was also inserted in some of the adapters namely, pA1C, PA2H and pA2Ø. The ligation of the SELP-59-A sequence to the pA1C vector was confirmed through digestion with *XbaI* and *AvaI* (Figure 19), and the correct orientation of the insert was further confirmed by DNA sequencing (see Annex L for the sequencing chromatogram). Lanes 3 and 5 of the Figure 19B showed positive clones for SELP insertion. For the purpose of developing high molecular weight protein polymers, SELP-59-A was also inserted into the concatenation adapters. The SELP sequence was ligated to the pA2H vector and confirmed with *XbaI* and *AvaI* restriction digestion (Figure 19). Three positive clones were obtained (Lanes 6, 8 and 9) with one of the results (Lane 7) not matching any of the expected electrophoretic pattern, probably due to contamination. Again, the orientation of the insert was verified by DNA sequencing. These constructions were named either SELP or SELP<sub>his</sub> depending if they hold a His-tag or not in the C-terminal of the SELP. As for the ligation of SELP with pA2Ø, the DNA was digested with *Ppu21I* (Figure 20) to assess the insertion of SELP and analyse the insert orientation. Analysis of results revealed two positive transformants with the correct orientation (Lanes 3 and 5). To notice that the two higher molecular weight bands of the positive and correct insertion (lane 2 in Figure 20A) partially overlap, making difficult its visualization in 1% agarose gel electrophoresis, appearing as a thick band (lanes 3 and 5 in Figure 20). Nevertheless, two transformants presented positive results for both insertion and orientation (lanes 3 and 5) and were further confirmed by DNA sequencing (see Annexes M and N for the sequencing chromatograms).



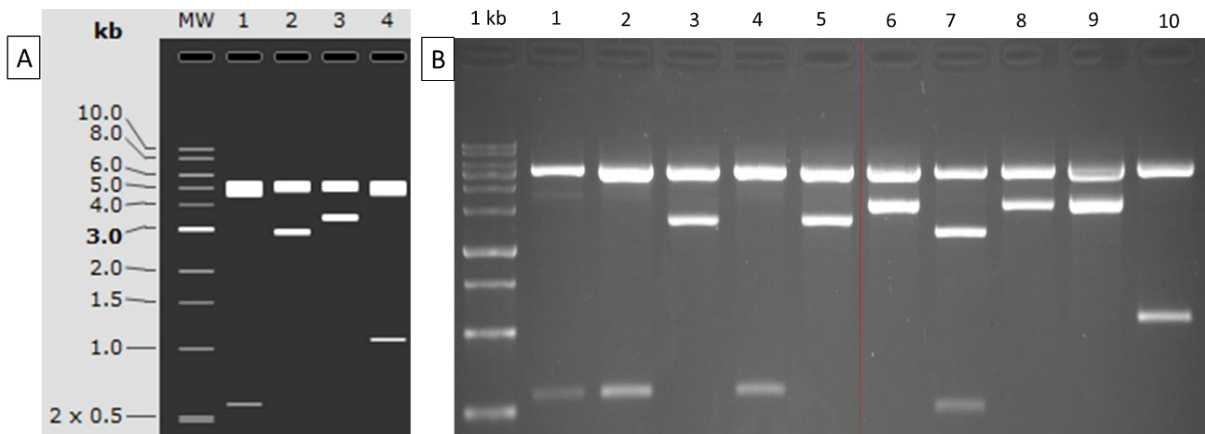


Figure 17. A- *In silico* simulation (SnapGene®) of a 1% agarose gel showing the digestion products after restriction digestion with *Xba*I and *Ava*I: Lane 1 corresponds to pA1C (used as control); Lane 2 represents the insertion of SELP-59-A into pA1C; Lane 3 represents the insertion of SELP-59-A into pA2H; lane 4 corresponds to pA2H (used as control). B- Electrophoretic analysis in a 1% agarose gel of the digestion with *Xba*I and *Ava*I: Lane 1 and 10 correspond to pA1C and pA2H, respectively; Lanes 2-5 show the result of digestion derived from different colonies transformed with the ligation product of SELP-59-A with pA1C. Lanes 6-9 show the result of digestion derived from different colonies transformed with the ligation product of SELP-59-A with pA2H.

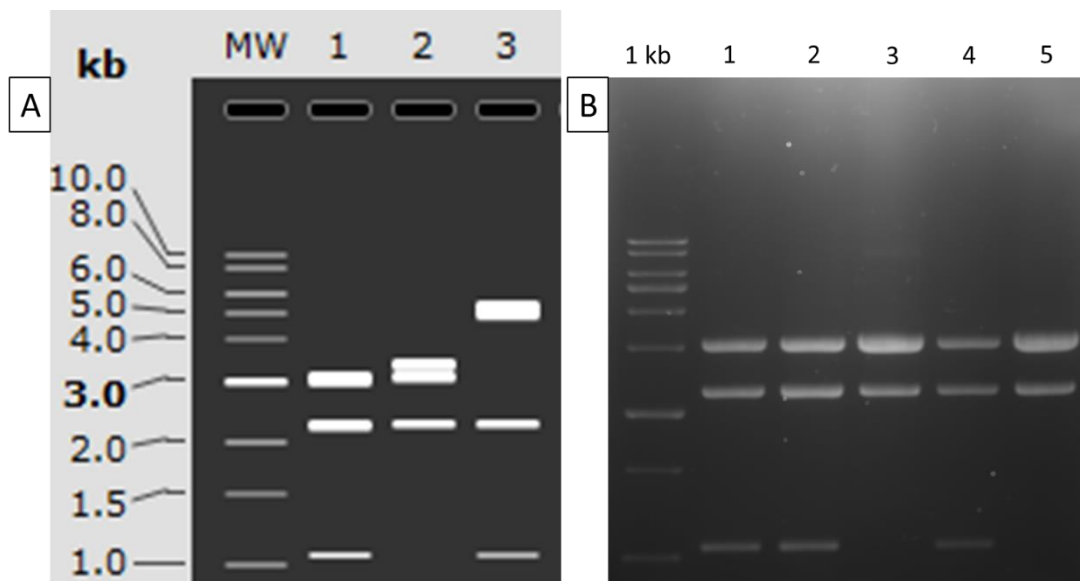


Figure 18. A- *In silico* simulation (SnapGene®) of a 1% agarose gel showing the digestion products after restriction digestion with *Ppu*21I: Lane 1 corresponds to pA2 $\emptyset$  (used as control); Lane 2 represents the insertion of SELP-59-A into pA2 $\emptyset$  in the correct orientation; Lane 3 represents the incorrectly oriented insertion of SELP-59-A into pA2 $\emptyset$ . B- Electrophoretic analysis in a 1% agarose gel of the digestion with *Ppu*21I: Lane 1 corresponds to pA2 $\emptyset$ ; Lanes 2-5 show the results of digestion derived from different colonies transformed with the ligation product of SELP-59-A with pA2 $\emptyset$ . There are two positive results in Lanes 3 and 5.

### 3.2.5. Insertion of ABP-CM4

As stated above, due to the high number of constructions, work focused only the insertion of ABP-CM4 as a proof-of-concept to test the feasibility of the molecular biology platform. The AMP was amplified using the oligonucleotides described in Table 8 and used for insertion into the adapters using the enzymes described in the same Table 8. Several variations of the constructions were made: the ABP-CM4 was inserted into the pA1C:A60 and pA3N:A60 vectors with or without a terminal cysteine while considering the removal or maintaining the intein. The ABP-CM4 was named either CM4<sub>cys</sub>, cysCM4 or CM4 depending if it holds a C-terminal Cysteine, a N-terminal Cysteine or none, respectively. The insertions were all confirmed by sequencing (see Annexes O, P and Q for the sequencing chromatograms) before proceeding to next experiments.

### 3.3. Production screenings

The twelve final constructions are resumed in Table 9. These final constructions were transformed into BL21 (DE3) and screened to detect protein expression levels. To note that the intermediate constructions of the self-cleaving adaptors with no AMP are not represented in the table.

Table 9. Schematic representation of all the final constructions.





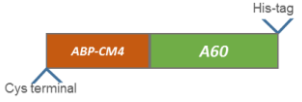







Construction	Construction representation	Construction name
1.		hisA60-int <sub>pH</sub> -CM4 <sub>cys</sub>
2.		hisA60-int <sub>pH</sub> -CM4
3.		hisA60-CM4 <sub>cys</sub>
4.		hisA60-CM4
5.		cysCM4-A60 <sub>his</sub>
6.		CM4-A60 <sub>his</sub>

Table 9 (continuation).

7.		cysCM4-intRed-A60 <sub>his</sub>
8.		CM4-intRed-A60 <sub>his</sub>
9.		intcon-C-A60 <sub>his</sub> -intcon-N
10.		intcon-C-A60-intcon-N
11.		intcon-C-SELP <sub>his</sub> -intcon-N
12.		intcon-C-SELP-intcon-N

### 3.3.1. Production screening of hisA60-int<sub>pH</sub>-CM4<sub>cys</sub>, hisA60-int<sub>pH</sub>-CM4, hisA60-CM4<sub>cys</sub> and hisA60-CM4 proteins

The four constructions resulting from the insertion of A60 and ABP-CM4 into the pA1C vector are described and represented on Table 9 by the numbers 1-4. These four constructions vary in two aspects: the presence or absence of the int<sub>pH</sub> and the presence or absence of a terminal cysteine. The insertion of a terminal Cysteine serves to introduce the possibility of immobilising the AMP onto a surface by grafting with Cysteine for the functionalization of surfaces (Mishra *et al.*, 2014). Production screening studies for all the constructions were carried out with samples normalised to an OD<sub>600</sub>=0.1, in order to choose the colony with the best production per cell (Figure 21). After SDS-PAGE and subsequent staining, the bands were compared and the colony showing the highest production, *i.e.* the band with larger area, was chosen as the best producing colony (pointed by an arrow). This analysis was done with the aid of ImageJ software (Schneider, Rasband & Eliceiri, 2012), using a densitometry analysis. The hisA60-int<sub>pH</sub>-CM4<sub>cys</sub> protein showed the lower expression values when compared to the control (CM4-A60).

The estimated size of the proteins with intein is ~49 kDa whereas, those without intein display a molecular weight of ~30 kDa. As observed in Figure 21, the overexpressed bands corresponding to the fusion protein appear at a higher MW than expected. This disparity between the estimated and displayed molecular weight has been reported several times in works with ELPs and attributed to the hydrophobic nature of the protein polymer (McPherson, Xu & Urry, 1996; Meyer & Chilkoti, 2002; Machado *et al.*, 2009; da Costa *et al.*, 2018). In fact, the displayed MW of ELP-based proteins is approximately 20%

higher than the theoretical MW, meaning that the proteins with and without intein would roughly display molecular weights around 59 kDa and the 36 kDa, respectively. The following equation was used to predict the displayed molecular weight:  $MW_{estimated} \times 1.2 = MW_{displayed}$ . Taking into account the mathematical expression and the fact that more than 50% of the fusion proteins is composed by the ELP sequence, it is possible to observe that the adjusted MW matches the MW shown in the SDS-PAGE.

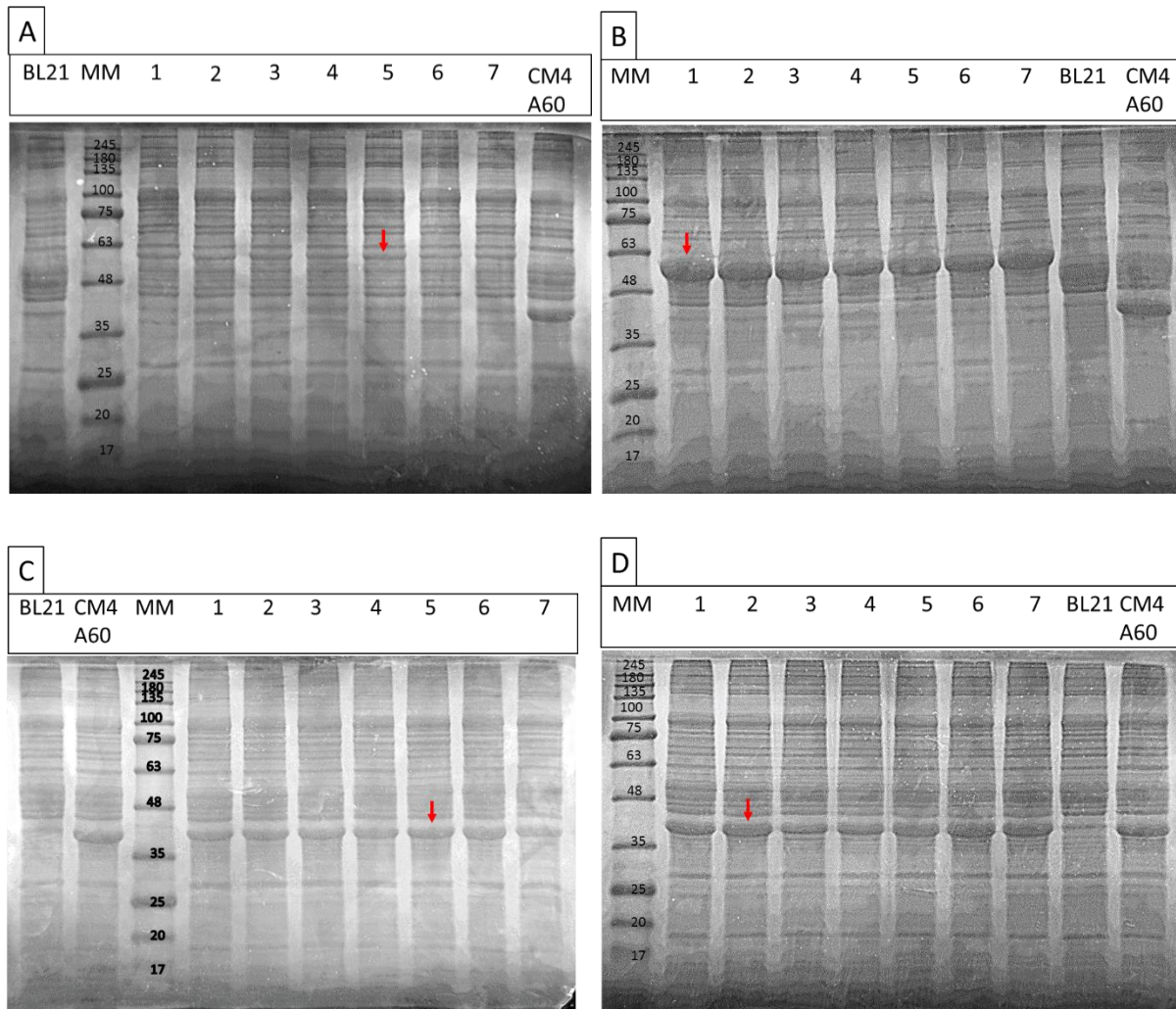


Figure 19. SDS-PAGE analysis of (A)  $hisA60-int_{pH}-CM4_{cys}$ , (B)  $hisA60-int_{pH}-CM4$ , (C)  $hisA60-CM4_{cys}$ , and (D)  $hisA60-CM4$  production screenings in a 10% polyacrylamide gel. Cells were grown in  $TB_{lac}$  for 22 h at 37 °C and 200 rpm. MM - Molecular weight marker (NZYcolour Protein Marker II, NZYTech); Lanes 1-7 correspond to production screenings of seven different colonies. Empty BL21(DE3), *i.e.* without plasmid, was used as negative control for production. CM4-A60 (from previous work of the group) was used as positive control for overexpression. The arrows point to the best producing colony.

### 3.3.2. Production screenings of $cysCM4-A60_{his}$ , $CM4-A60_{his}$ , $cysCM4-int_{Red}-A60_{his}$ and $CM4-int_{Red}-A60_{his}$ proteins

The four constructions resulting from the insertion of A60 and ABP-CM4 into the pA3N vector are described and represented on Table 9 by the numbers 5-8. Once again, these proteins vary between each

other from the presence or absence of the  $\Delta$ I-SM intein and the presence or absence of a terminal cysteine. Production screening studies were carried out with samples normalised to an  $OD_{600}=0.1$ , analysed by SDS-PAGE, and the bands with highest expression of protein were chosen with the help of ImageJ program (marked in the gel by an arrow). Similarly to the previous production screening, the fusion proteins' theoretical molecular weight differs from the displayed molecular weight. Considering the mathematical expression discussed in section 3.4.1, it was possible to observe that the calculated MW relates with those found in the SDS-PAGE analysis.

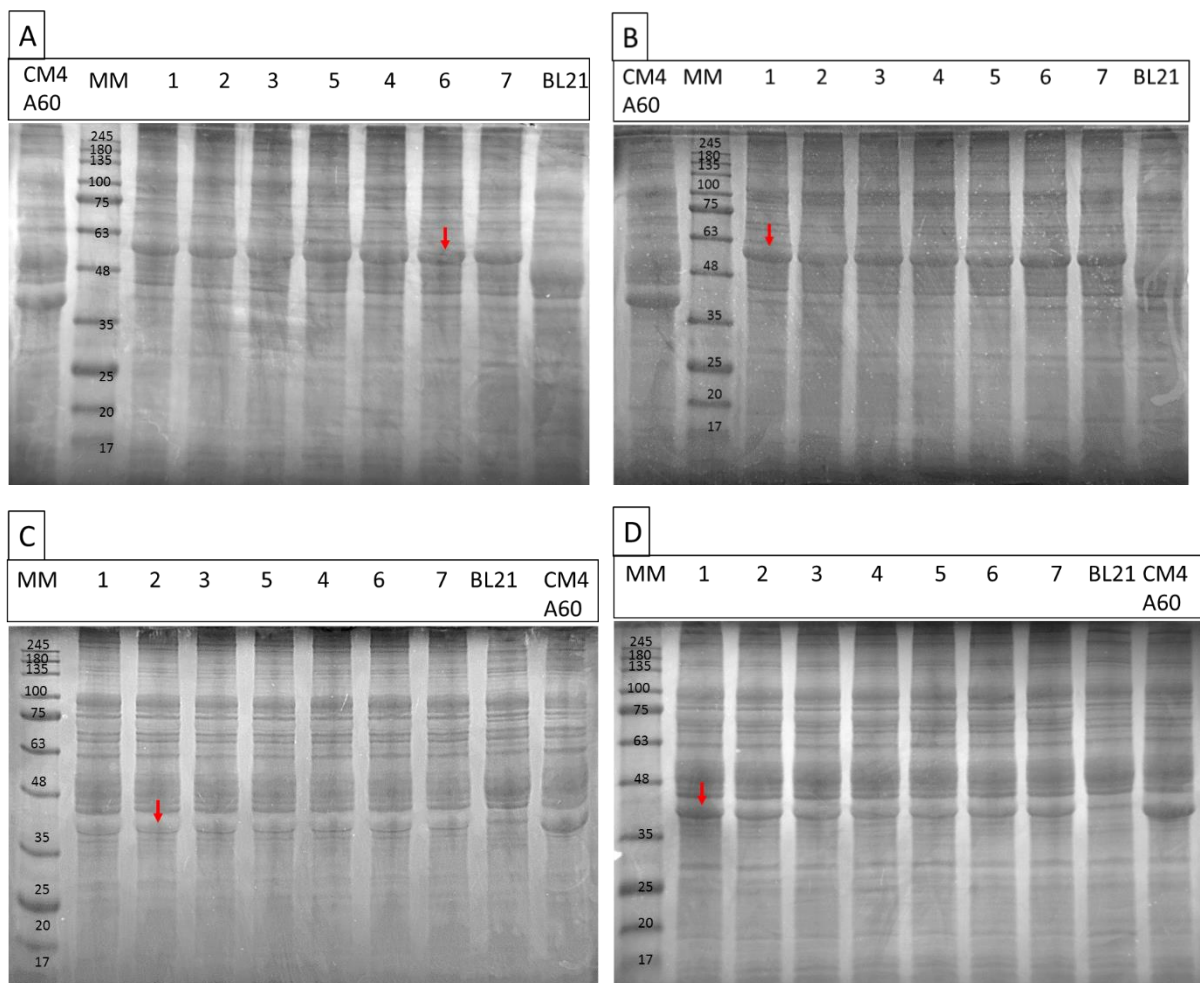


Figure 20. SDS-PAGE analysis of (A)  $cysCM4-int_{Red}-A60_{his}$ , (B)  $CM4-int_{Red}-A60_{his}$ , (C)  $cysCM4-A60_{his}$ , and (D)  $CM4-A60_{his}$  production screenings in a 10% polyacrylamide gel. Cells were grown in  $TB_{LAC}$  for 22 h at 37 °C and 200 rpm. MM - Molecular weight marker (NZYcolour Protein Marker II, NZYTech); Lanes 1-7 correspond to production screenings of seven different colonies. Empty BL21(DE3), *i.e.* without plasmid, was used as negative control for production. CM4-A60 (from previous work of the group) was used as positive control for overexpression. The arrows point to the best producing colony.

### 3.3.3. Production screenings of $\text{int}_{\text{con-C-A60his-int}_{\text{con-N}}$ , $\text{int}_{\text{con-C-A60-int}_{\text{con-N}}$ , $\text{int}_{\text{con-C-SELP}_{\text{his-int}_{\text{con-N}}$ and $\text{int}_{\text{con-C-SELP-int}_{\text{con-N}}$ proteins

As in the previous sections 3.3.1. and 3.3.2., the samples of the concatenation constructions were normalized for an  $\text{OD}_{600}=0.1$ , to allow a direct comparison of the production of the protein of interest by SDS-PAGE. Figures 22 and 23 show the results of the production screening studies for the constructions 9-12 represented in Table 9. The expression of all concatenation proteins tested displayed a less accentuated band when compared to the control (CM4-A60), probably due to the occurrence of partial protein splicing. The colonies chosen with highest expression are marked with an arrow for each screening.

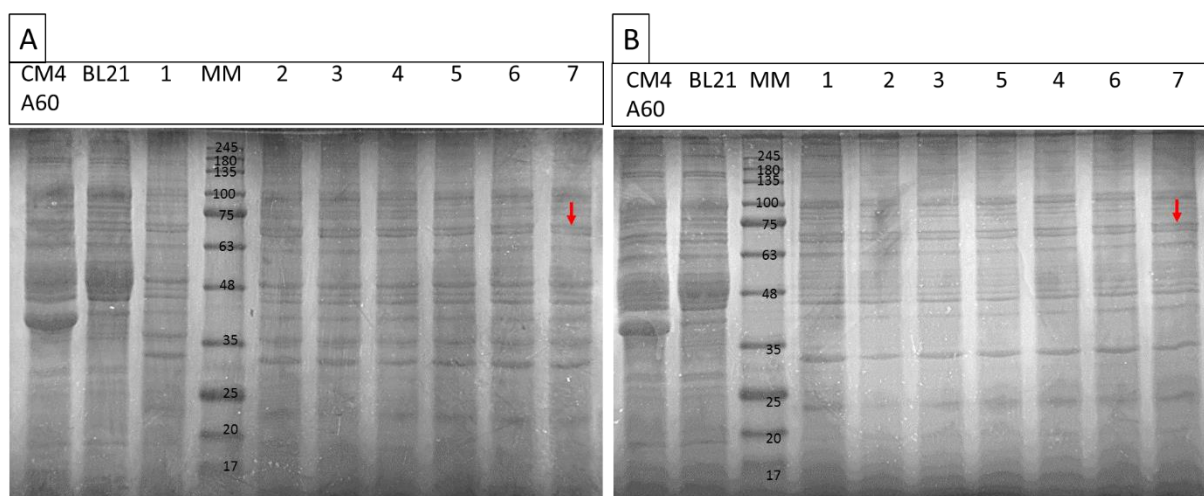


Figure 21. SDS-PAGE analysis of (A)  $\text{int}_{\text{con-C-A60his-int}_{\text{con-N}}$ , and (B)  $\text{int}_{\text{con-C-A60-int}_{\text{con-N}}$  production screenings in a 10% polyacrylamide gel. Cells were grown in  $\text{TB}_{\text{LAC}}$  for 22 h at 37 °C and 200 rpm. MM - Molecular weight marker (NZYcolour Protein Marker II, NZYTech); Lanes 1-7 correspond to production screenings of seven different colonies. Empty BL21(DE3), *i.e.* without plasmid, was used as negative control for production. CM4-A60 (from previous work of the group) was used as positive control for overexpression. The arrows point to the best producing colony.

Unlike the previous constructions, the fusion proteins for concatenation demonstrated to have a lower deviation between the theoretical and the displayed molecular weight. Most likely, this is a consequence of the higher amount of polar aa in these constructions, resulting in a decrease of the overall hydrophobicity of the protein. As such, the estimated molecular weight value ( $\sim 65$  kDa) matches the molecular weight displayed in the SDS-PAGE. The bands of the SELP-based constructions (Figure 24) also seem to match the predicted molecular weight of  $\sim 98$  kDa. Again, this can be explained by the contribution of polar aa that decrease the overall protein hydrophobicity. Interestingly, the band just below 35 kDa that appears in all samples, but not in the controls, could be related with a premature cleavage

of int<sub>con</sub>. This suggests that the actual expression of the protein is higher than the one perceptible by this analysis.

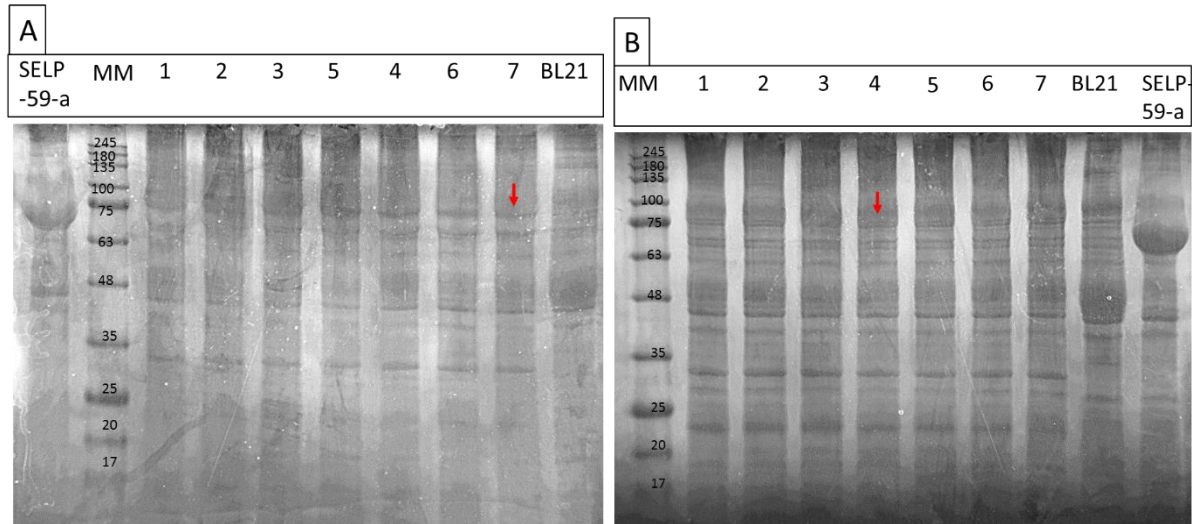


Figure 22. SDS-PAGE analysis of (A) int<sub>con</sub>-C-SELP<sub>his</sub>-int<sub>con</sub>-N, and (B) int<sub>con</sub>-C-SELP-int<sub>con</sub>-N production screenings in a 10% polyacrylamide gel. Cells were grown in TB<sub>LAC</sub> for 22 h at 37 °C and 200 rpm. MM - Molecular weight marker (NZYcolour Protein Marker II, NZYTech); Lanes 1-7 correspond to production screenings of seven different colonies. Empty BL21(DE3), *i.e.* without plasmid, was used as negative control for production. CM4-A60 (from previous work of the group) was used as positive control for overexpression. The arrows point to the best producing colony.

### 3.4. Protein Purification

To assess the feasibility of the devised strategies, five representative constructions (constructions 2, 3, 5, 8 and 9; Table 8) were selected for protein production and purification, and used to test the different cleavage/splicing methodologies. For production, all proteins were expressed by means of auto-induction with TB<sub>lac</sub>, using a protocol established in the LBM laboratory (Machado *et al.*, 2013a).

For ease of result interpretation, Table 10 summarises the molecular weights of each full protein, as well as of all the individual components, for the selected constructions discussed in the following sections.

Table 10. Protein Modules and corresponding molecular weights

module	Molecular Weight (kDa)
hisA60-int <sub>pH</sub> -CM4	48.8
hisA60-CM4 <sub>cys</sub>	30.6

Table 10(continuation).

CM4-int <sub>Red</sub> -A60 <sub>his</sub>	49.2
cysCM4-A60 <sub>his</sub>	30.6
int <sub>con-C</sub> -A60 <sub>his</sub> -int <sub>con-N</sub>	66.1
A60 <sub>his</sub>	26.7
hisA60	26.6
int <sub>pH</sub> /int <sub>Red</sub>	18.5
Int <sub>con-N</sub>	23.8
int <sub>con-C</sub>	15.9
ABP-CM4	3.8

#### 3.4.1. hisA60-int<sub>pH</sub>-CM4

The optimized purification procedure normally used in the lab group for ELPs and SELPs, includes an initial step of acidification at pH 3.5 after cell lysis, in order to precipitate *E. coli* endogenous proteins and lead to a clear supernatant extract (Machado *et al.*, 2013a). Then, the proteins of interest are purified from the clear supernatant through the use of ITC (ELPs) or by ammonium sulphate precipitation (SELPs). However, since the intein from hisA60-int<sub>pH</sub>-CM4 is sensitive to pH, the acidification step was excluded from the purification protocol, as it could trigger the cleavage of the AMP. Instead, separation of the soluble and insoluble fractions from the cell lysate was performed by increasing the number and velocity of the centrifugation steps. Still, even with the additional centrifugation cycles, a large amount of *E. coli* contaminants was still present, as it can be seen in Figure 25 (S1TE). These contaminants were further removed by employing three rounds of ITC (hot and cold incubation and centrifugation cycles). ELPs have a transition temperature above which they start aggregating and segregate from solution, allowing to be recovered by centrifugation. On the other hand, as the temperature is cooled down below a specific threshold, the ELP aggregates solubilize completely whereas, most of the contaminating proteins remain precipitated. The ITC take advantage of this thermoresponsive behaviour to purify ELP-based polymers using hot (ELP precipitates) and cold (ELP solubilizes) cycles, allowing its separation from other contaminating proteins (Meyer & Chilkoti, 1999). Once the pellet during the hot cycle is formed and the



supernatant discarded, the ELP-enriched pellet can be solubilised at lower temperatures (usually in ice-cold water) followed by a cold centrifugation, to achieve a pure ELP fraction. Due to water sensitivity to pH changes, purification of hisA60-int<sub>pH</sub>-CM4 was performed in parallel with TE buffer and ddH<sub>2</sub>O to check if there was cleavage or splicing. Furthermore, the presence of salts in the TE buffer aids the purification of ELPs by lowering the transition temperature, thus making it easier to pellet the polymer after heating (Reguera *et al.*, 2007; Machado *et al.*, 2009).

For purification, and as mentioned in Materials and Methods, each ITC round involves the following cycles: i) a hot cycle in which the solution is incubated at 37 °C followed by a hot centrifugation and; ii) a cold cycle in which the pellet from the hot cycle is resuspended in ice-cold ddH<sub>2</sub>O or TE (See Annex R for a schematic representation of the purification protocol). SDS-PAGE analysis of samples purified after three ITC cycles (Figure 25) reveals two bands with similar intensities at approximately 45 kDa and another at approximately 38 kDa. While this remains to be elucidated, the two bands with the larger molecular weight correlate with the expected sizes for the complete construction and the construction without the ABP-CM4 peptide, respectively. As for the band with smaller molecular size, we can speculate that is the result of the residual splicing of the intein domain, causing the appearance of the protein A60-CM4.

Interestingly, a comparison between P2<sub>TE</sub> and P2 allows to establish some differences: i) the sample P2<sub>TE</sub> presents more protein than P2 and, ii) while P2<sub>TE</sub> shows a single intense band at approximately 45 kDa, the sample P2 is characterized by two bands near 45 kDa with similar intensities. The different band intensities observed between P2<sub>TE</sub> and P2 are most likely a consequence of pipetting errors and/or due to different resuspension volumes between cycles. The presence of two bands in the P2 sample when compared to the single band of the P2<sub>TE</sub> suggests that this phenomenon only occurs in pure protein fractions.

To note that ELPs do not stain or stain very poorly with Coomassie Blue staining (Machado *et al.*, 2009; Hassouneh, Christensen & Chilkoti, 2010). Although His-tags and other fused proteins (as in fusion proteins in which ELPs are used as purification tags) are stained with Coomassie Blue, the size of the ELP exerts great influence, and can result in faded or “ghost” bands. To overcome this limitation, negative staining with copper chloride is often used (Machado *et al.*, 2009). In this construction, the int<sub>pH</sub>-CM4 was 22.3 kDa (resulting from the A60 cleavage), which is likely to be the band we see in the gel in lane S2 and S2TE near the 21.5 kDa MM band.

We can also notice in Figure 25 that the protein is having some residual splicing that resulted in a band corresponding to the  $_{\text{hisA60}}$  (at around 32 kDa marked by an arrow) that is barely visible due to the fact that Coomassie Blue does not stain the ELP, staining only the His-tag domain (Hassouneh, Christensen & Chilkoti, 2010). Through GelAnalyzer® software, we calculated that the size of the observed band is in fact the size expected for the A60 sequence (around 32 kDa, due to the hydrophobicity issues already described). This residual splicing is most probably occurring during the production step where the intracellular pH decrease with the acids produced during cell metabolism, although the final extracellular pH value was of 6.96 at 26.4 °C, due to the medium buffer.

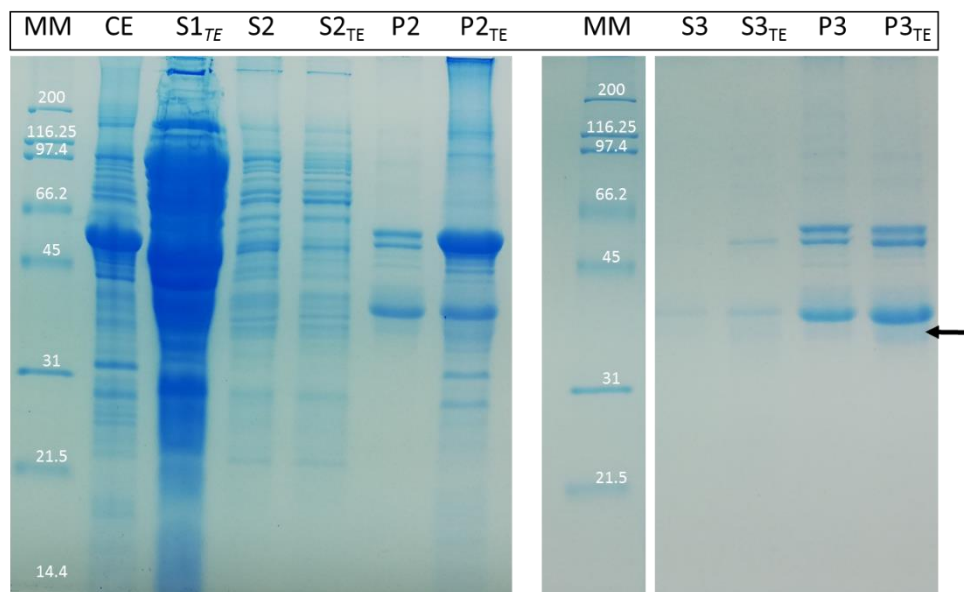


Figure 23. SDS-PAGE analysis of the different purification steps of  $_{\text{hisA60-int}_{\text{pH-CM4}}$  by ITC. MM - Molecular weight marker (Broad Range SDS-PAGE marker, Bio-Rad); Lane CE – cell crude extract; Lane S1<sub>TE</sub> - supernatant after the first hot cycle (no ELP expected). Lane S2 and S2<sub>TE</sub> - samples of the supernatant after the second hot cycle in ddH<sub>2</sub>O or in TE, respectively (no ELP expected); Lanes P2 and P2<sub>TE</sub> - samples of the pellet from the second hot cycle, solubilized in ddH<sub>2</sub>O or TE, respectively (presence of ELP expected); Lanes S3 and S3<sub>TE</sub> - samples of the supernatant after the third hot cycle in ddH<sub>2</sub>O or TE, respectively (no ELP expected); Lane P3 and P3<sub>TE</sub> - samples of pellet from the third hot cycle, solubilized in ddH<sub>2</sub>O or in TE, respectively (presence of ELP expected). SDS-PAGE was conducted in 10% polyacrylamide gels. No modifications other than cropping and resizing were applied to the images. The arrow points to the protein band with 32 kDa.

### 3.4.2. $_{\text{hisA60-CM4}_{\text{cys}}}$

For this construction, purification was achieved using ddH<sub>2</sub>O and addition of NaCl 0.5 M was here introduced, in order to reduce the ELP transition temperature (Machado *et al.*, 2009) and improve the recovery of the recombinant protein during purification by ITC. Due to the absence of the intein domain, purification was achieved by employing routine protocols without the need of buffered solutions. As observed by SDS-PAGE (Figure 26), the supernatants of the purification cycles (lanes S1, S2 and S3) do not present any detectable protein, thus indicating little or no loss of  $_{\text{hisA60-CM4}_{\text{cys}}}$  during the ITC

rounds. Some contaminants were still present after three rounds of purification, suggesting that more cycles are required to obtain pure protein. Due to the higher size influence of the ELP in this construct (in opposition to the 3.8 kDa of CM4), gel staining with Coomassie Blue is not efficient and thus, copper chloride negative staining was used instead.

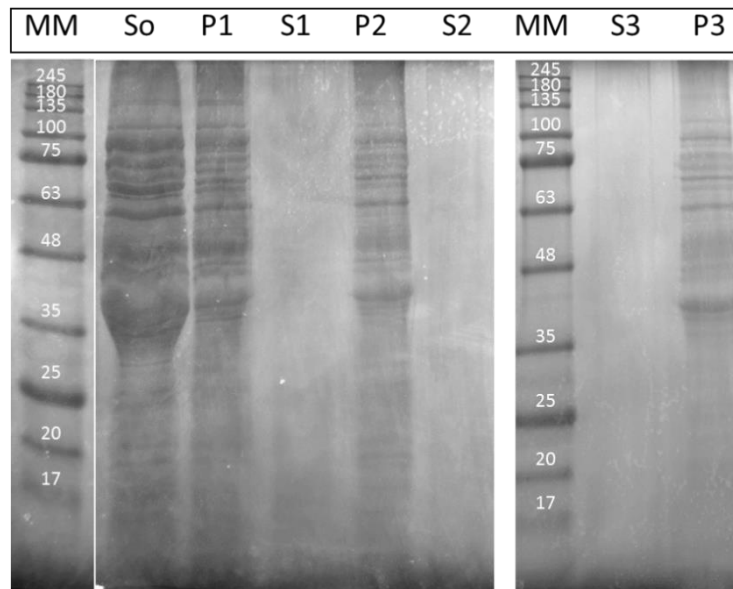


Figure 24. SDS-PAGE analysis of the different purification steps of  $_{hisA60-CM4_{cys}}$  by ITC purification in a 10% polyacrylamide gel. MM - Molecular weight marker (NZYcolour Protein Marker II, NZYTech); Lane So – cell crude extract (lysed cells); Lane P1 - pellet after the first hot cycle; Lane S1 - supernatant after the first hot cycle; Lane S2 - supernatant after the second hot cycle; Lane P2 - pellet after the second hot cycle; Lane S3 - supernatant after the third hot cycle; Lane P3 - pellet after the third hot cycle. Pellets from the hot cycles were resuspended in ice-cold ddH<sub>2</sub>O. SDS-PAGE was conducted in 10% polyacrylamide gels. No modifications other than cropping and resizing were applied to the images.

### 3.4.3. CM4-int<sub>Red</sub>-A60<sub>his</sub>

Purification of CM4-int<sub>Red</sub>-A60<sub>his</sub> was achieved by following the same methodology as in section 3.4.1. for  $_{hisA60-int_{pH}-CM4}$ . However, in this case, only ddH<sub>2</sub>O was used to resolubilize the ELP-enriched pellets in the ITC rounds. Even though the intein present in this protein is not responsive to pH, the acidification step was bypassed, as the low pH could induce changes in intein conformation and decrease the cleavage efficiency. SDS-PAGE analysis of ITC purification (Figure 27) revealed the presence of protein in the supernatant of the second cycle of purification (lane S2), which was not expected. This can be explained by the contribution of intein to reduce the overall hydrophobicity of the system, resulting in an increase of the ELP transition temperature. As such, the conditions tested (37 °C) are most likely not sufficient to induce an effective phase transition behaviour of the ELP molecules; consequently, the

recombinant protein or part of remain in the supernatant of the hot cycle. Possible ways to improve the efficiency of the ITC rely on the use of higher temperatures for the hot cycles or through addition of NaCl in order to lower the transition temperature and thus achieve a complete recovery of the recombinant protein.

Considering the results, by the end of the purification it was possible to observe the presence of extra bands (Figure 27), but it is unclear if these are due to splicing or to the presence of contaminants.

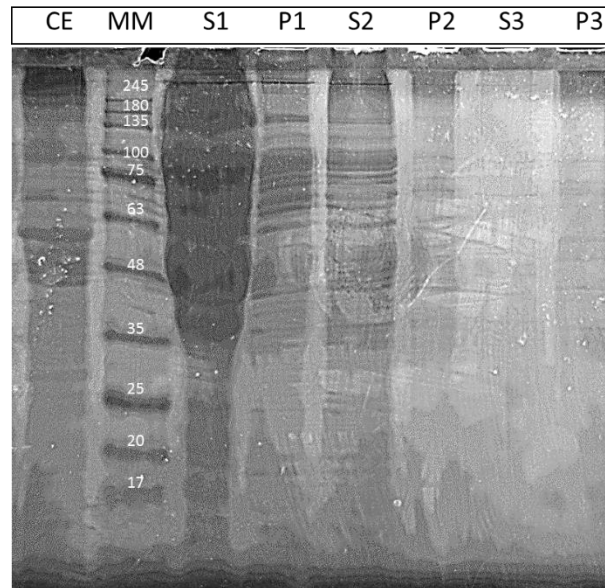


Figure 25. SDS-PAGE analysis of the different purification steps of CM4-int<sub>Red</sub>-A60<sub>his</sub> by ITC. MM - Molecular weight marker (NZYcolour Protein Marker II, NZYTech); Lane CE – cell crude extract; Lane P1 - pellet after the first hot cycle; Lane S1 - supernatant after the first hot cycle; Lane S2 - supernatant after the second hot cycle; Lane P2 - pellet after the second hot cycle; Lane S3 - supernatant after the third hot cycle; Lane P3 - pellet after the third hot cycle. Pellets from the hot cycles were resuspended in ice-cold ddH<sub>2</sub>O. SDS-PAGE was conducted in 10% polyacrylamide gels.

#### 3.4.4. cysCM4-A60<sub>his</sub>

Purification of cysCM4-A60<sub>his</sub> followed the same procedure as in the purification of hisA60-CM4<sub>cys</sub> (section 3.4.2.). The pH of the cell crude lysate was adjusted to 3.5 and 0.5 M of NaCl was used. SDS-PAGE analysis (Figure 28) revealed complete or almost complete recovery of the recombinant protein during purification by ITC (pointed by an arrow); strengthening the suggestion that addition of NaCl 0.5 M could be used to improve the purification of CM4-int<sub>Red</sub>-A60<sub>his</sub> (section 3.4.3.). By the end of purification, it was possible to observe the presence of extra bands that could be related with the presence of contaminants, thus indicating that extra cycles should be performed to increase purity.

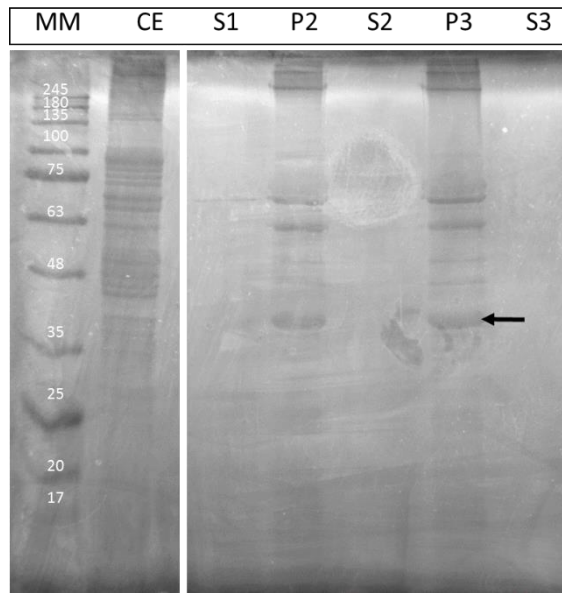


Figure 26. SDS-PAGE analysis of the different purification steps of  $cysCM4-A60_{his}$  by ITC. MM - Molecular weight marker (NZYcolour Protein Marker II, NZYTech); Lane CE – cell crude extract; Lane S1 - supernatant after the first hot cycle; Lane S2 - supernatant after the second hot cycle; Lane P2 - pellet after the second hot cycle; Lane S3 - supernatant after the third hot cycle; Lane P3 - pellet after the third hot cycle. Pellets from the hot cycles were resuspended in ice-cold ddH<sub>2</sub>O. SDS-PAGE was conducted in 10% polyacrylamide gels. No modifications other than cropping and resizing were applied to the images. The arrow point to the protein band that is  $cysCM4-A60_{his}$ .

#### 3.4.5. $int_{con-C-A60_{his}}-int_{con-N}$

Purification of  $int_{con-C-A60_{his}}-int_{con-N}$  was achieved by employing the initial acidification step, followed by ITC purification with addition of 0.5 M NaCl. As observed in the SDS-PAGE gel (Figure 29), acidification of the cell crude extract to pH 3.5 allowed to remove most of *E. coli* endogenous proteins. As previously observed for  $hisA60-CM4_{cys}$  (section 3.4.2.) and  $cysCM4-A60_{his}$  (section 3.4.4.), addition of NaCl allowed the full recovery of the recombinant protein through ITC purification. Interestingly, the use of ITC for the purification of  $int_{con-C-A60_{his}}-int_{con-N}$  was not so effective as the recombinant protein is present in the supernatant of the first heating cycle (Lane S1). Nevertheless, the recombinant protein is fully recovered in the following cycles with no detectable protein in the supernatants of the second and third hot cycles (Lanes S2 and S3). As in the purification of the other proteins, after completing the three rounds of ITC, there are still some extra proteins in the recombinant protein-enriched pellet (Lane P3). These could be a result of the presence of contaminants or possible residues of splicing or cleaving. Still, as described earlier, one of the extra bands can be a result of the circularization of A60.

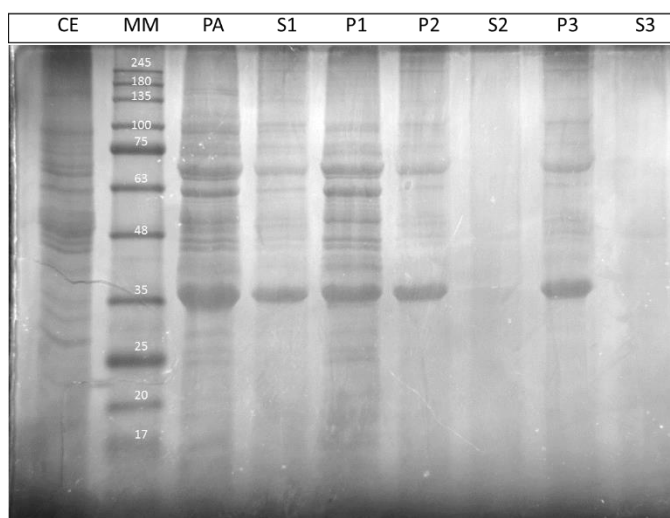


Figure 27. SDS-PAGE analysis of the different purification steps of  $\text{int}_{\text{con-C-A60}_{\text{his}}-\text{int}_{\text{con-N}}}$  by ITC. MM - Molecular weight marker (NZYcolour Protein Marker II, NZYTech); Lane CE – cell crude extract; Lane PA – supernatant of the lysate after acidification at pH 3.5; Lane P1 - pellet after the first hot cycle; Lane S1 - supernatant after the first hot cycle; Lane S2 - supernatant after the second hot cycle; Lane P2 - pellet after the second hot cycle; Lane S3 - supernatant after the third hot cycle; Lane P3 - pellet after the third hot cycle. Pellets from the hot cycles were resuspended in ice-cold ddH<sub>2</sub>O. SDS-PAGE was conducted in 10% polyacrylamide gels.

### 3.5. Cleavage test

This section describes the tests performed to evaluate the cleaving ability of the constructs with self-cleaving tags.

#### 3.5.1. CM4-int<sub>Red</sub>-A60<sub>his</sub>

The int<sub>Red</sub> intein has its N-terminal cleavage triggered by the addition of a reducing agent, such as DTT, resulting in the release of the antimicrobial peptide. The recombinant CM4-int<sub>Red</sub>-A60<sub>his</sub>, purified either with ddH<sub>2</sub>O or with TE, was incubated overnight at room temperature with two different concentrations of DTT (50 and 200 mM). The cleavage product was analyzed by SDS-PAGE without  $\beta$ -mercaptoethanol as it functions as a reducing agent and is often interchangeable with DTT (Batjargal, Walters & Petersson, 2015). As observed in Figure 30A, the samples with no reducing agent (lanes 0 and 0<sub>TE</sub>, used as control) display three bands at molecular weights of 74, 69 and 66 kDa. In opposition, these bands are absent in the samples with DTT (lanes 50, 50<sub>TE</sub>, 200 and 200<sub>TE</sub>) and the electrophoretic pattern is characterized by the presence of six bands with molecular weights of 58, 51, 43, 41, 36 and 34 kDa. Moreover, it is possible to observe the presence of a small molecular weight band well below 6.5 kDa which is more evident in the samples with 200 mM of DTT. This low molecular weight band corresponds to the molecular weight of ABP-CM4 (3.8 kDa) suggesting that the antimicrobial peptide is cleaved from CM4-int<sub>Red</sub>-A60<sub>his</sub>. The intensity of this band is stronger at the highest concentration of

DTT, indicating that 50 mM of the reducing agent might not be sufficient to induce total cleavage. Comparing the samples in ddH<sub>2</sub>O or in TE, there are no clear differences between them, indicating that pH is not a significant factor for the N-terminal cleavage but rather, the concentration of reducing agent.

The protein CM4-int<sub>Red</sub>-A60<sub>His</sub> contains a His-tag that can be used for detection by Western blot using an anti-His antibody. Comparing with SDS-PAGE, analysis of the Western blot (Figure 30B) reveals a positive detection for the His-tag in all of the three bands from the non-reducing conditions (0 and 0<sub>TE</sub>). At reducing conditions (50, 50<sub>TE</sub>, 200 and 200<sub>TE</sub>) these bands are absent and the six protein bands with molecular weights of 58, 51, 43, 41, 36 and 34 kDa are positively detected (Figure 30A). Considering that all of these bands have the His-tag fused with A60, and that ELP proteins have a displayed molecular weight 20% higher than the theoretical value, one can infer the following assumptions: i) the protein band at 58 kDa corresponds to the CM4-int<sub>Red</sub>-A60<sub>His</sub> protein ( $49.2 \text{ kDa} \times 1.2 = 59.04 \text{ kDa}$ ); ii) the protein band at 51 kDa band is most likely the int<sub>Red</sub>-A60<sub>His</sub> protein after cleavage of the antimicrobial peptide ( $45.3 \text{ kDa} \times 1.2 = 54.36 \text{ kDa}$ ); iii) the 36 kDa band is probably attribute to CM4-A60<sub>His</sub> ( $30.6 \text{ kDa} \times 1.2 = 36.72 \text{ kDa}$ ) which indicates that residual splicing might be occurring; iv) the 34 kDa band is probably the A60<sub>His</sub> protein cleaved from the intein, which explains the reason for the faded staining with Coomassie Blue. As for the remaining protein bands with 41 and 43 kDa, these do not match any of the expected results. A possible explanation could be attributed to incomplete translation however, this hypothesis was discarded as the His-tag domain includes the last residues from the fusion protein. Assuming that A60 is present in the two proteins detected at 41 and 43 kDa, the equation  $MW_{estimated} \times 1.2 = MW_{displayed}$  was reversible employed to determine the estimated molecular weight. Accordingly, the estimated molecular weights are  $\sim 36$  and  $\sim 34$  kDa, for the 43 and 41 kDa bands, respectively. On a view of these data, it was theorized that, after the N-cleavage of ABP-CM4, the intein suffers another cleavage by using residues downstream in the sequence that are able to do the N-S/N-O acyl shift (Cysteine, Serine or Threonine). A (thio)ester intermediate might be forming with those aa, similarly to what occurs in class 3 inteins (Brace *et al.*, 2010). Alternatively, these protein bands could be the result of protein degradation but the bands seem too defined for that to be the case. A detailed analysis of the amino acidic sequence of CM4-int<sub>Red</sub>-A60<sub>His</sub> revealed a Serine in position 141, that if cleaved would give a protein with an estimated molecular weight of  $\sim 34$  kDa which explains the 41 kDa band. In addition, it was also found a Threonine at +113 position of the protein which, if cleaved, would generate a protein with an estimated MW of  $\sim 37$  kDa which is just a bit higher than the estimated MW of  $\sim 36$  kDa of the 43 kDa band. This difference could be related with the error associated with the estimation of protein sizes by GelAnalyzer software.

Other Serine and Threonine aa exist in the sequence of the intein but they do not match the molecular weights required to explain the bands. A schematic representation of the theorized mechanism to explain these bands is depicted in Figure 31. Nevertheless, these are all assumptions that need a more detailed validation.

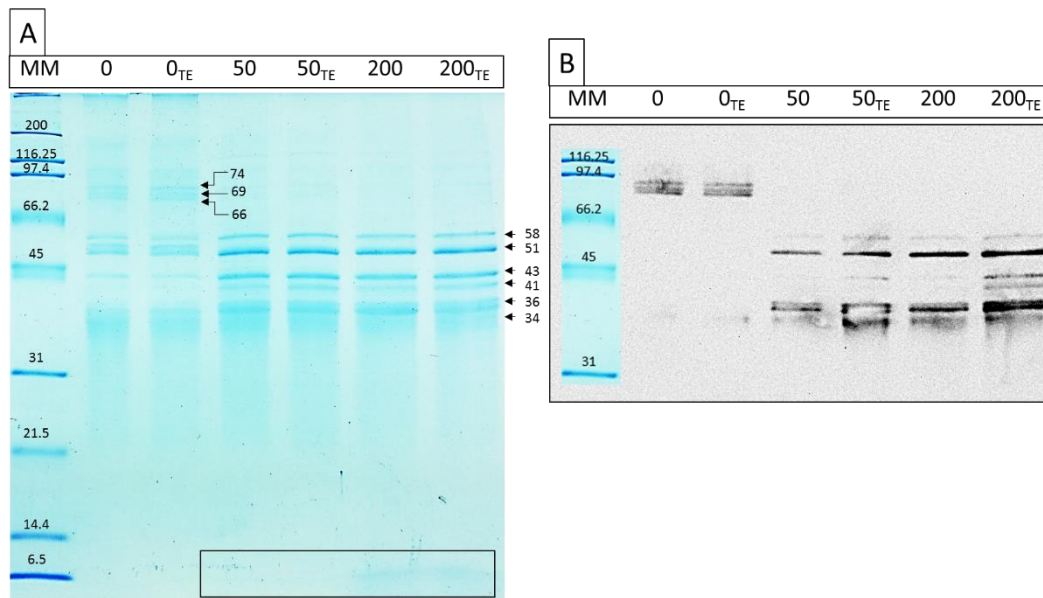


Figure 28. A- SDS-PAGE analysis of CM4-int<sub>Red</sub>-A60<sub>His</sub> cleavage test after incubation for 20 h. MM - Molecular weight marker (Broad Range SDS-PAGE marker, Bio-Rad); Lanes 0 and 0<sub>TE</sub> - non-reducing conditions (without DTT) for samples in ddH<sub>2</sub>O or TE, respectively; Lanes 50 and 50<sub>TE</sub> – reducing conditions with 50 mM of DTT for samples in ddH<sub>2</sub>O and in TE; Lanes 200 and 200<sub>TE</sub> - samples in reducing conditions with 200 mM of DTT for samples in ddH<sub>2</sub>O or TE, respectively. The cleaved ABP-CM4 band is surrounded by the box in the lower end of the gel. The molecular weight of the bands pointed by the arrows was determined with the use of GelAnalyzer software. B- Western Blot analysis of the same samples with anti-His and anti-mouse as primary and secondary antibodies, respectively.



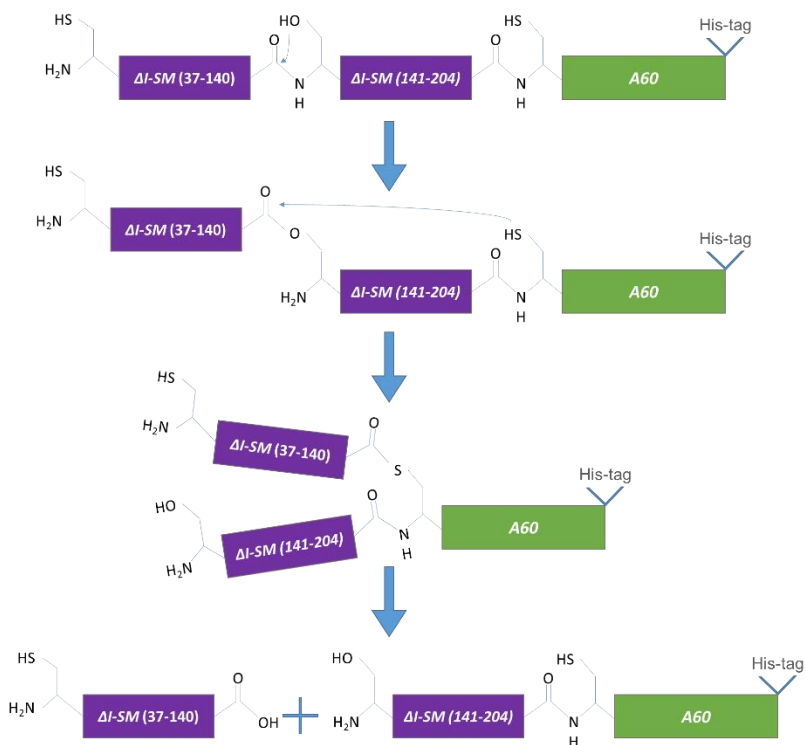


Figure 29. Theorized mechanism for the intra-intein cleavage.

The main objective of this thesis is the creation of adapters in order to establish a molecular biology platform for the production and purification of AMPs. Having developed the molecular tools and tested the feasibility of the cleavage system, the next step was to test if the AMP could be easily purified with this strategy. To do so, the cleaved proteins of CM4-int<sub>Red</sub>-A60<sub>his</sub> in TE were incubated at 50 °C for 1 h followed by centrifugation. Due to the thermoresponsive properties of the ELP, it is expected that all proteins containing A60 would precipitate after centrifugation, leaving the pure ABP-CM4 in the supernatant. Figure 32 shows the SDS-PAGE analysis of this temperature purification step, revealing the successful separation of ABP-CM4 from the remaining protein bands. The gels were stained with Coomassie Blue (Figure 32A) followed by silver nitrate (Figure 32B) to allow a more clear visualization of the less intense bands; silver nitrate is ~10-30 times more sensitive than Colloidal Coomassie Blue (Chevallet, Luche & Rabilloud, 2006). Remarkably, one single hot incubation step was sufficient to separate all the cleaved peptide from the rest of the proteins, clearly demonstrating the high efficiency of the system. As previously observed in Figure 30, cleavage of the antimicrobial peptide is conditioned by the concentration of the reducing agent. This observation is clearly evident by comparing the samples S0, S50 and S100 in Figure 32B, in which the highest amount of cleaved AMP is observed at the highest concentration of DTT.

The strong silver stained band observed at near 6.5 kDa in the MM lane is an artefact of the molecular weight marker and appeared in all SDS-PAGE gels stained with silver nitrate.

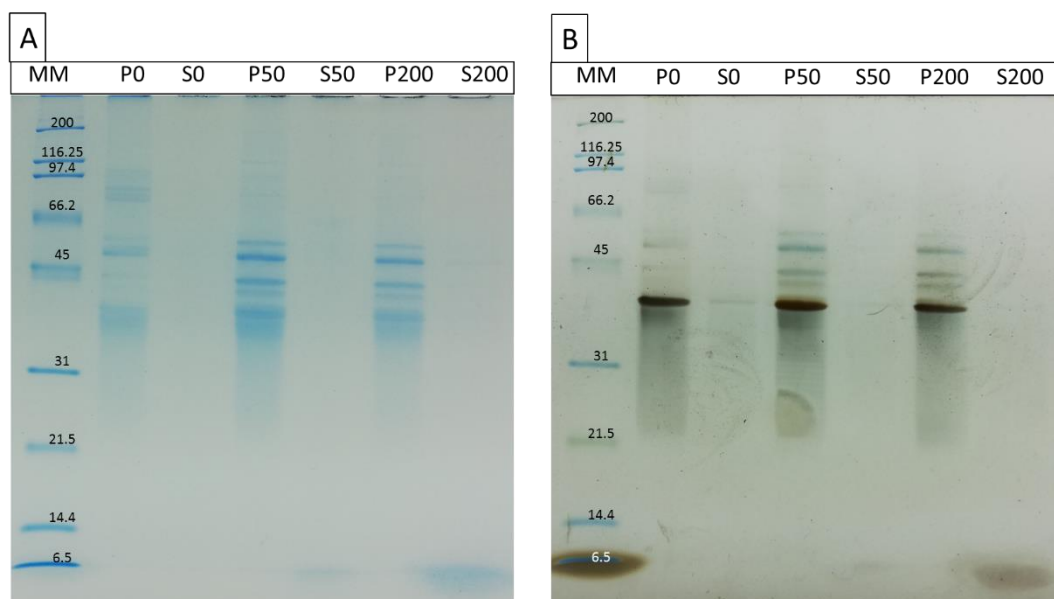


Figure 30. SDS-PAGE analysis with (A) Coomassie Blue staining and (B) silver nitrate staining of the temperature purification test used to isolate the cleaved antimicrobial peptide ABP-CM4 from CM4-int<sub>Red</sub>-A60<sub>his</sub>. The cleaved protein solution in TE (after incubation for 20 h with 0, 50 or 200 mM of DTT) was incubated at 50 °C for 1 h, followed by a centrifugation at 40 °C. MM - Molecular weight marker (Broad Range SDS-PAGE marker, Bio-Rad); Lanes P0 and S0 - samples of the pellet and supernatant, respectively, in non-reducing conditions (without DTT); Lanes P50 and S50 - samples of the pellet and supernatant, respectively, in reducing conditions with 50 mM of DTT; Lanes P200 and S200 - samples of the pellet and supernatant, respectively, in reducing conditions with 200 mM of DTT.

### 3.5.2. *his*A60-int<sub>pH</sub>-CM4

Despite the successful demonstration of the cleavage system while using reducing agents, as well as the successful isolation of the antimicrobial peptide through the use of a simple hot cycle, there is still the need to remove DTT from the final purified product fraction. This can be easily circumvented by, for instance, dialysis; nevertheless, it requires an additional step.

In a try to simplify the overall process of isolation and purification, the *his*A60-int<sub>pH</sub>-CM4 construction was devised to promote cleavage at pH 6. The purified *his*A60-int<sub>pH</sub>-CM4 was thus submitted to the cleaving conditions described in Materials and Methods (cleavage solution at pH 6, at room temperature for 24 h), and analysed by SDS-PAGE (Figure 33A). The negative controls were done with ddH<sub>2</sub>O (pH 6.45 at room temperature for 24 h) and with TE buffer (pH 8.0 at room temperature for 24 h). Results demonstrate that the system was ineffective, with no detectable bands corresponding to free ABP-CM4, thus indicating the absence of cleavage. The presence of  $\beta$ -mercaptoethanol in the loading

buffer also demonstrated to not influence cleavage, showing similar results independently of its addition. In fact, all conditions presented a similar electrophoretic pattern. Further analysis by Western Blotting (Figure 33B) revealed the presence of four protein bands at 33, 35, 52 and 56 kDa. Considering that the His-tag is fused to the N-terminal of A60, it is reasonable to assume that all the detected bands have at least 26.6 kDa, corresponding to the pair His-Tag/A60. Following the rationale described in section 3.5.1., and considering the deviation in the electrophoretic mobility of ELP-based proteins of  $\sim 20\%$ , it is possible to deduce that the four detected bands probably correspond to different portions of  $_{his}A60-int_{pH}-CM4$ , deriving from unspecific and/or spontaneous cleavage. Accordingly, with the help of the mathematical expression  $MW_{estimated} \times 1.2 = MW_{displayed}$ , the 56 kDa band finds a probable correspondence with the full protein  $_{his}A60-int_{pH}-CM4$  ( $48.9 \text{ kDa} \times 1.2 = 58.7 \text{ kDa}$ ). Following the same reasoning, the 52 kDa band can be attributed to the  $_{his}A60-int_{pH}$  portion ( $45.1 \text{ kDa} \times 1.2 = 54.1 \text{ kDa}$ ) whereas, the 35 kDa band is probably the  $_{his}A60-CM4$  portion ( $30.4 \text{ kDa} \times 1.2 = 36.5 \text{ kDa}$ ). The 33 kDa is likely the  $_{his}A60$  portion ( $26.6 \text{ kDa} \times 1.2 = 31.9 \text{ kDa}$ ), cleaved from the construction by the intein's N-terminal activity, and would explain the diffused staining by Coomassie Blue; especially considering that these bands are highly marked in the Western Blot (Figure 33B). Considering the aforementioned assumptions, the lack of protein bands corresponding to  $int_{pH}-CM4$  and  $int_{pH}$  suggests that splicing/cleavage occurred somewhere during production or purification, which can explain the lower molecular weight protein band observed during purification (Lanes P2 and P3 in Figure 25). Nevertheless, one should consider that these are all assumptions that need further experimental validation.

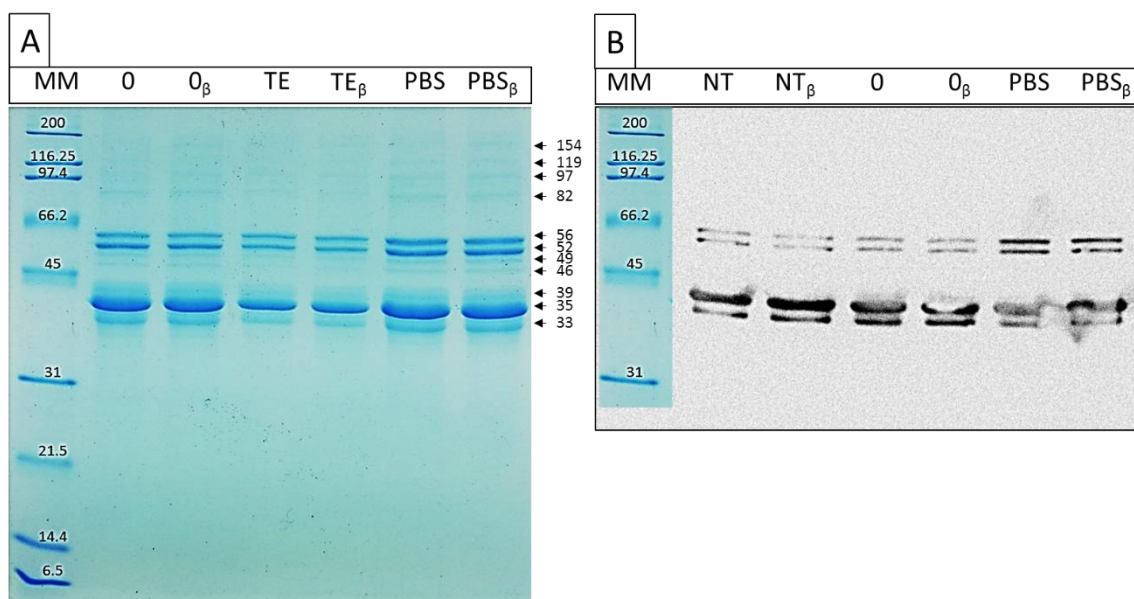


Figure 31. A- SDS-PAGE analysis of *hisA60-int<sub>pH</sub>-CM4* cleavage test after incubation for 24 h in cleavage solution at pH 6 and room temperature. Samples with  $\beta$ -mercaptoethanol in the loading buffer are identified with a  $\beta$ . MM - Molecular weight marker (Broad Range SDS-PAGE marker, Bio-Rad); Lanes 0 and 0 <sub>$\beta$</sub>  - samples of the negative control in ddH<sub>2</sub>O; Lanes TE and TE <sub>$\beta$</sub>  - samples of the negative control in TE buffer at pH 8; Lanes PBS and PBS <sub>$\beta$</sub>  - samples in cleavage solution. The molecular weight of the bands pointed by the arrows was determined with the use of GelAnalyzer software. B- Western Blot analysis with anti-His and anti-mouse as primary and secondary antibodies, respectively. Lanes NT and NT <sub>$\beta$</sub>  - samples of the negative control that received no treatment nor incubation.

The protein bands at 154, 119, 97, 82, 49, 46 and 39 kDa were not detected by Western Blot, indicating the lack of the His-tag and thus, are most likely contaminants. Again, a more detailed analysis by, for example, amino acid analysis/sequencing or purification by affinity chromatography, could provide insights into the origin of these bands or to improve purity. However, as mentioned above, the main objective of this thesis was the development of the molecular biology platform and thus, optimization of purification is beyond the scope of this work.

Due to the absence of cleavage and release of ABP-CM4, further experiments involved performing the cleavage reaction at a higher temperature of 30 °C, for 24 h (Wu *et al.*, 2006). In addition, the presence of DTT (10 mM) in the cleavage solution was also considered. Figure 34A and 34B show the results obtained by SDS-PAGE stained with Coomassie Blue and silver nitrate, respectively, while using these conditions. Despite the increase in temperature, the electrophoretic pattern was similar for all the tested samples. However, after silver staining, which is much more sensitive than Coomassie Blue, there is a clear band at approximately 6.5 kDa in the PBS<sub>DTT</sub> sample, indicating the presence of ABP-CM4 (Figure 34B). Surprisingly, the presence of free ABP-CM4 was only detected in the PBS<sub>DTT</sub> sample which contains 10 mM of DTT, meaning that the presence of the reducing agent is critical for the ABP-CM4 C-

terminal cleavage to occur. This was unexpected as cleavage should be promoted by changing the pH to 6, independently of using a reducing agent.

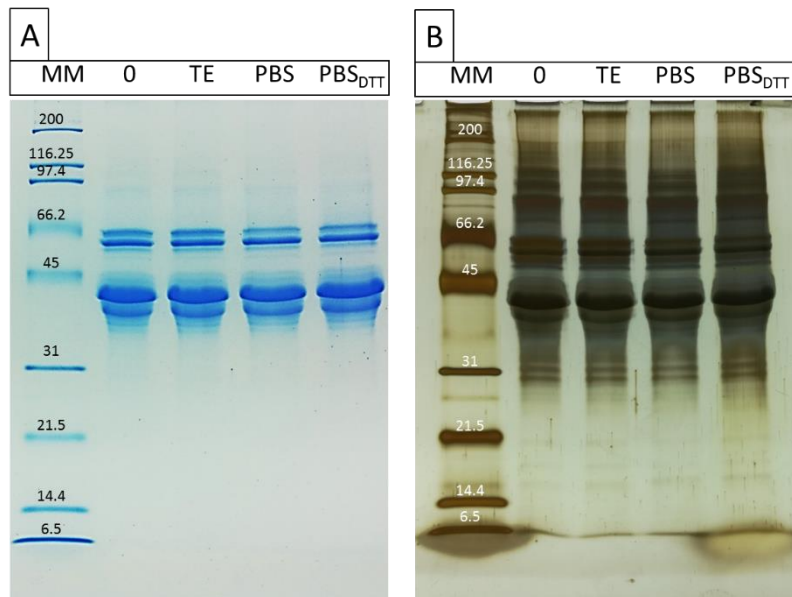


Figure 32. SDS-PAGE analysis of *hisA60-int<sub>pH</sub>-CM4* cleavage test in a 10% polyacrylamide gel. Samples were taken after cleavage induction of 24 h at room temperature and other 24 h at 30 °C. The same gel was stained with Coomassie Blue (A) and AgNO<sub>3</sub> (B). MM - Molecular weight marker (Broad Range SDS-PAGE marker, Bio-Rad); Lane 0 - sample of the negative control solubilized in ddH<sub>2</sub>O; Lane TE - samples of the negative control solubilized in TE buffer, pH 8; Lane PBS - samples of the pellet and supernatant of the induction with cleaving solution, pH 6; Lanes PBS<sub>DTT</sub> - samples of the pellet and supernatant of the induction with Cleaving solution, pH 6, 10 mM of DTT. In all samples, β-mercaptoethanol was added in the loading buffer.

Interestingly, the protein band at 35 kDa (Figures 33A and 34B) is the most intense, indicating an higher amount of protein. This band was attributed to the *hisA60-CM4* portion, and corresponds to a high amount of residual splicing arising from a high rate of N-terminal activity. Even though the initial Cysteine is mutated, splicing can occur due to the amount of Threonines near the beginning of the intein (at positions -1 and +6) that can initiate the splicing.

The purpose of this system was to produce and cleave ABP-CM4 by means of pH changing. This objective was partially accomplished; however, the system still needs a reducing agent for the reaction to occur.

### 3.6. Concatenation test

The biggest issue with the *int<sub>con</sub>-C-A60<sub>his</sub>-int<sub>con</sub>-N* protein is that throughout purification and even in the production screening studies, splicing was apparently already occurring. The *Npu* DnaE intein displays improved splicing when a reducing agent like DTT or β-mercaptoethanol is added. In the

construction  $\text{int}_{\text{con-C-A60his-int}_{\text{con-N}}$ , the FRB and FKBP domains flank the protein polymer sequence to trigger  $\text{int}_{\text{con-C}}$  and  $\text{int}_{\text{con-N}}$  dimerization by addition of rapamycin. In this way, it is possible to concatenate the protein polymer to obtain high molecular weight variants. However, this strategy presents a disadvantage, as there is the possibility of A60 circularization.

To evaluate the feasibility of the devised strategy,  $\text{int}_{\text{con-C-A60his-int}_{\text{con-N}}$  purified as previously described was incubated for 4 h and 24 h at room temperature with DTT and rapamycin. The splicing products were then analysed by SDS-PAGE using loading buffer with and without  $\beta$ -mercaptoethanol (Figure 35). Analysis of results after 4 h, revealed the same electrophoretic pattern for all the samples independently of the conditions tested (Figure 35A). The gel is characterized by strong bands at  $\sim 66$  kDa and  $\sim 35$  kDa which correspond to the unspliced protein and to the circularized A60, respectively. After 24 h of incubation (Figure 35B), the samples containing a reducing agent seem to display a less intense band corresponding to the unspliced protein (66 kDa) whereas, the band attributed to the circularization of A60 is maintained for all samples. Also, it is possible to observe a series of high molecular weight protein bands at 106, 138, 181 and 214 kDa that can be attributed to ELP concatenation products. The bands at molecular weights  $<66$  kDa are likely a result of cleavage/splicing and will be discussed later.

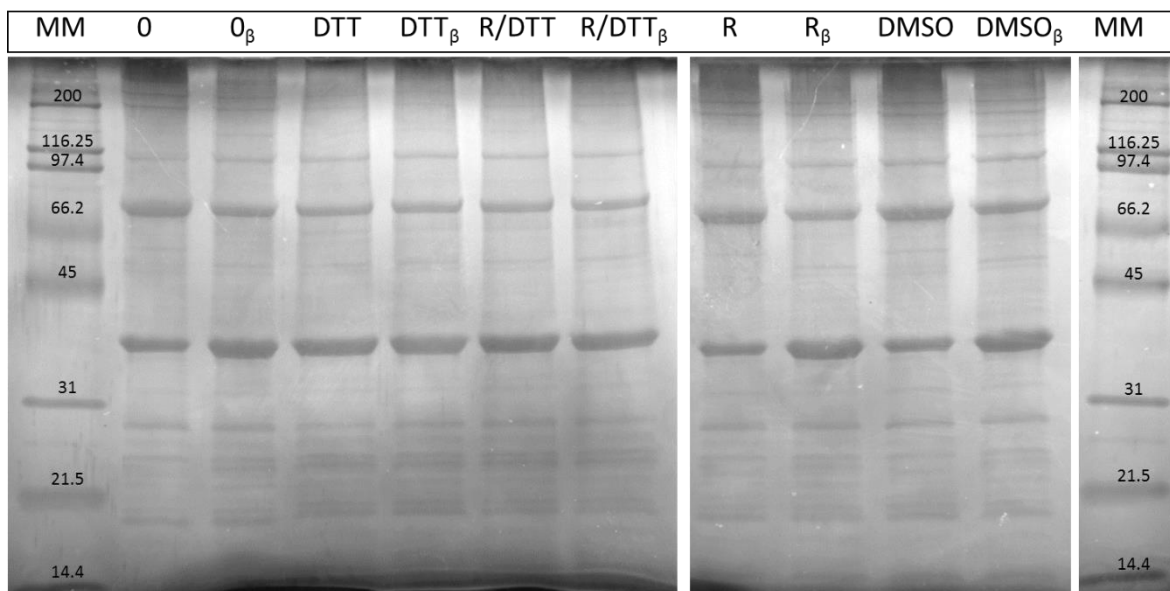


Figure 35. SDS-PAGE analysis of  $\text{int}_{\text{con-C-A60his-int}_{\text{con-N}}$  concatenation test after incubation for 4 h at room temperature. Samples with  $\beta$ -mercaptoethanol in the loading buffer are identified with a  $\beta$ . MM - Molecular weight marker (Broad Range SDS-PAGE marker, Bio-Rad); Lanes 0 and  $0_{\beta}$  - samples solubilized in ddH<sub>2</sub>O (negative control); Lanes DTT and  $DTT_{\beta}$  - samples with 50 mM of DTT; Lanes R and  $R_{\beta}$  - samples with 100  $\mu$ M rapamycin with 0.1% DMSO; Lanes R/DTT and  $R/DTT_{\beta}$  - samples with 100  $\mu$ M rapamycin with 0.1% DMSO and 50 mM of DTT; Lanes DMSO and  $DMSO_{\beta}$  - samples of the negative control with 0.1% DMSO.

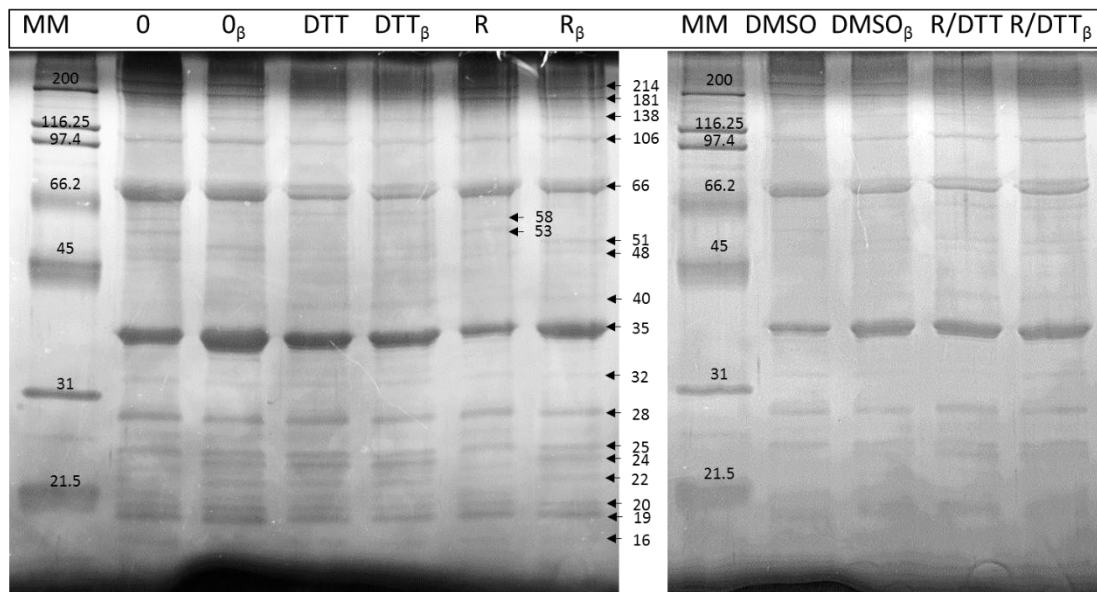


Figure 36. SDS-PAGE analysis of  $\text{int}_{\text{con-C-A60}_{\text{his}}-\text{int}_{\text{con-N}}}$  concatenation test after incubation for 24 h at room temperature. Samples with  $\beta$ -mercaptoethanol in the loading buffer are identified with a  $\beta$ . MM - Molecular weight marker (Broad Range SDS-PAGE marker, Bio-Rad); Lanes 0 and  $0_{\beta}$  - samples solubilized in ddH<sub>2</sub>O (negative control); Lanes DTT and DTT $_{\beta}$  - samples with 50 mM of DTT; Lanes R and R $_{\beta}$  - samples with 100  $\mu$ M rapamycin with 0.1% DMSO; Lanes R/DTT and R/DTT $_{\beta}$  - samples with 100  $\mu$ M rapamycin with 0.1% DMSO and 50 mM of DTT; Lanes DMSO and DMSO $_{\beta}$  - samples of the negative control with 0.1% DMSO. The molecular weight of the bands pointed by the arrows was determined with the use of GelAnalyzer software.

According to the results obtained after 24 h, nineteen bands were identified and the molecular weights were determined through the use of the software GelAnalyzer. From highest to lowest, the protein bands were estimated to present molecular weight values of 214, 181, 138, 106, 66, 58, 53, 51, 48, 40, 35, 32, 28, 25, 24, 22, 20, 19 and 16 kDa (marked by arrows in Figure 36). For a more detailed analysis on the origin of these bands, the samples after 24 h of incubation were further stained with Coomassie Blue (Figure 37A) and compared with Western Blotting (Figure 37B). As observed in the Western Blot, nine bands with sizes  $\geq 35$  kDa were positively detected for the presence of the His-tag whereas, the remaining protein bands below 35 kDa were absent. A tenth band marked by an asterisk (\*) does not match any of the bands previously seen in the SDS-PAGE gel, meaning that an ELP-based protein is present but at concentration below the detection limit of the stains. As mentioned, the strong protein bands at 35 kDa and 66 kDa correspond to the circularization of A60 and to the unspliced protein, respectively. Interestingly, the control samples present several very high molecular weight bands ( $> 245$  kDa), which are more evident in the sample with  $\beta$ -mercaptoethanol, but are absent in the remaining samples containing DTT. These very high MW protein bands are present in the Western Blot (Figure 37B), suggesting that are a product of the concatenation of A60.

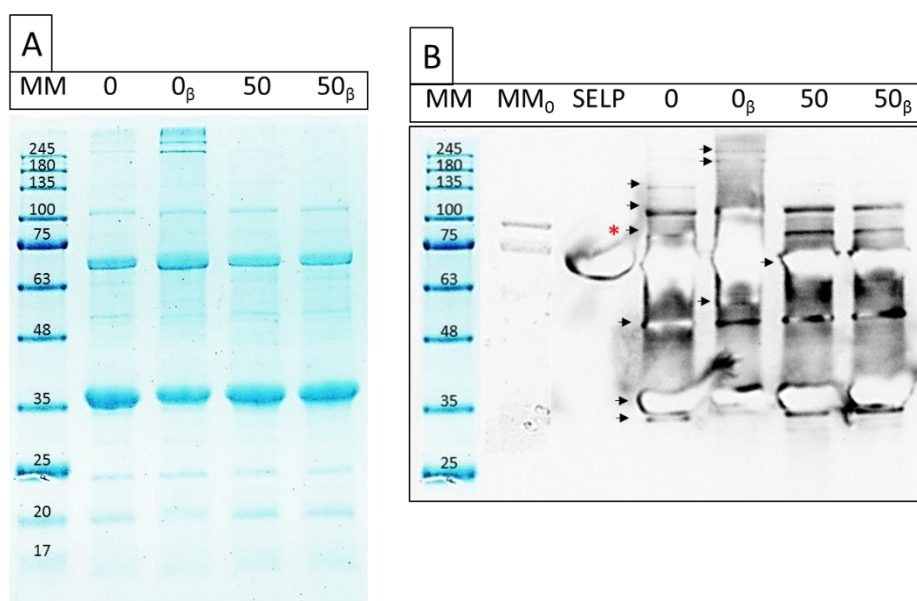


Figure 37. A- SDS-PAGE analysis of int<sub>con</sub>-C-A60<sub>his</sub>-int<sub>con</sub>-N concatenation after incubation for 24 h at room temperature. Samples with β-mercaptoethanol in the loading buffer are identified with a β. MM - Molecular weight marker (Broad Range SDS-PAGE marker, Bio-Rad; Lanes 0 and 0<sub>β</sub> - samples solubilized in ddH<sub>2</sub>O (negative control); Lanes 50 and 50<sub>β</sub> - samples with 50 mM DTT. B- Western Blot analysis with anti-His and anti-mouse as primary and secondary antibodies, respectively. Lane MM<sub>0</sub> corresponds to the molecular weight marker in the Western Blot membrane; SELP-59-A, a silk-elastin-like protein with a His-tag was used as positive control (SELP). Protein bands detected by Western Blot are pointed by arrows; the band identified with an asterisk (\*) points out to a protein band not detected in SDS-PAGE.

In an attempt to relate some of the detected protein bands in Figure 36 with potential cleavage/splicing products, the estimated molecular weight was calculated according to the expression  $MW_{estimated} \times 1.2 = MW_{displayed}$ . As mentioned earlier, ELP-based proteins display an abnormal SDS-PAGE gel mobility with an increase of 20% in molecular weight, in comparison to the theoretical/estimated value. Considering this, calculations were performed by following the same methodology as described earlier: proteins that were thought to contain more than 50% ELP were adjusted by a factor of 1.2 (corresponding to an increase of 20%). For simplicity, Table 11 presents the theorized results obtained for 12 of the 19 detected bands and the corresponding protein or protein portion. These attributions were performed not only by corresponding the molecular weights, but also while comparing the SDS-PAGE gels with the Western Blot results. Nevertheless, although these estimations are based on scientific assumptions, it must be noted that these values are theoretical and need further scientific validation.



Table 11. Attribution of proteins to the bands considering the size and percentage of ELP.

Size (kDa)	Corresponding protein(s)	Estimated MW (kDa)	>50% ELP	Displayed MW (kDa)
214*	int <sub>con</sub> -C-(A60 <sub>his</sub> ) <sub>5</sub> -int <sub>con</sub> -N	168	✓	201.6
181*	int <sub>con</sub> -C-(A60 <sub>his</sub> ) <sub>4</sub> -int <sub>con</sub> -N	142	✓	170.4
138	int <sub>con</sub> -C-(A60 <sub>his</sub> ) <sub>3</sub> -int <sub>con</sub> -N	117	✓	140.4
106	int <sub>con</sub> -C-(A60 <sub>his</sub> ) <sub>2</sub> -int <sub>con</sub> -N	91	✓	109.2
66	int <sub>con</sub> -C-A60 <sub>his</sub> -int <sub>con</sub> -N	66	✗	-
58	A60 <sub>his</sub> -int <sub>con</sub> -N	49	✓	58.8
48	int <sub>con</sub> -C-A60 <sub>his</sub>	41	✓	49.2
40	int <sub>con</sub> -C + int <sub>con</sub> -N **	39	✗	-
35	A60 <sub>his</sub> (circular)	-	-	-
32	A60 <sub>his</sub> (linear)	26	✓	31.2
25	int <sub>con</sub> -N	24	✗	-
16	int <sub>con</sub> -C	16	✗	-

\*Proteins might be too big to exactly determine its size; \*\*Dimers

The bands of 214, 181, 138, 106 and 66 kDa from Figure 37B match the marked bands in the Western Blot, which gives reasons to think that are the result of a concatenation of A60. The 58 kDa and 48 kDa bands also display an equivalent in the Western Blot, validating the hypothesis that they are the result of cleavage of either int<sub>con</sub>-C or int<sub>con</sub>-N. The 35 kDa band seems to be the circularization of A60<sub>his</sub> whereas, the 32 kDa band corresponds to A60<sub>his</sub> in which both int<sub>con</sub>-C and int<sub>con</sub>-N domains were cleaved, resulting in a linearized A60<sub>his</sub>. The circular form of A60<sub>his</sub> runs slowly in the gel due to its shape and thus, displays a higher MW. The band marked with an asterisk in the Western Blot was determined to have a MW of 86 kDa with GelAnalyzer software. This protein band contains an A60<sub>his</sub> (as it was positively detected in the Western Blot) and, considering its size, it is possible that could be the result of cleavage of either the int<sub>con</sub>-C or the int<sub>con</sub>-N domain of the dimer int<sub>con</sub>-C-(A60<sub>his</sub>)<sub>2</sub>-int<sub>con</sub>-N protein formed after concatenation. In theory, this would be the result of N-cleavage or C-cleavage of the intein and would result in a (A60<sub>his</sub>)<sub>2</sub>-int<sub>con</sub>-N (76 kDa) or a int<sub>con</sub>-C-(A60<sub>his</sub>)<sub>2</sub> (68 kDa) protein. Adjusting the molecular weight by 20%, a possible (A60<sub>his</sub>)<sub>2</sub>-int<sub>con</sub>-N or int<sub>con</sub>-C-(A60<sub>his</sub>)<sub>2</sub> protein would, respectively, display 91.2 kDa or 81.6 kDa. Since the protein band was determined as displaying a MW of 86 kDa, it

is reasonable to assume that it could be either  $(A60_{\text{his}})_2\text{-int}_{\text{con-N}}$  or a  $\text{int}_{\text{con-C}}(A60_{\text{his}})_2$ . As for the band with 40 kDa, this protein is absent in the Western Blot indicating the lack of a His-tag. This suggests that this band corresponds to a combination of  $\text{int}_{\text{con-N}}$  with  $\text{int}_{\text{con-C}}$  either by the dimerization of FKBP with FRB and rapamycin, or to the interaction of the split halves from the *Npu* DnaE intein. The bands at 53, 51, 28, 24, 22, 20 and 19 kDa protein bands are most likely attributed to the cleavage of the inteins through internal Serine or Threonine aa, as previously explained in Figure 31 (section 3.5.1.). A pair of Serines in the positions +580 and +581 could explain the 53 and 51 kDa protein bands. The 28, 24, 22, 20 and 19 kDa protein bands are, by the same logic, either a combination of cleaved inteins or contaminants.

The *Npu* DnaE intein is apparently, not perfectly suited for this kind of concatenation since it is very active. Instead, the use of dual systems of self-assembly such as those described in Figures 11A and 11B, could be more efficient and prevent circularization. The intein best suited for a system like that system would probably be an artificially split intein, as it would not interact easily and splice freely. In turn, that would make the dimerizing domains more effective at regulating the splicing. However, the system  $\text{int}_{\text{con-C}}A60_{\text{his}}\text{-int}_{\text{con-N}}$  have potential use for the circularization of proteins as this phenomenon occurs at some extent. In the overall, although not perfect, this strategy allowed to obtain concatenated variants of A60, as demonstrated in Figure 35 and 36.

## 4. Conclusions

In the overall, the goals devised in this project were successfully achieved giving origin to a molecular biology platform for the purification of antimicrobial peptides. This provides the basis to set up a library of constructs with different AMPs that can be easily purified by employing simple processes.

Regarding the overall results, from the twelve different constructions, three were selected for subsequent experimental validation: two designed for the purification of AMPs and one for the concatenation of protein polymers. The constructions designed for the production of AMPs were able to successfully self-cleave the AMP from a protein-based polymer (PBP) tag which was isolated using a non-chromatographic purification. This encouraging result means that other AMPs could potentially be purified by the same way or, alternatively, fused with PBPs to produce polymers with antimicrobial properties. Nevertheless, there are still some issues that must be considered such as the removal of DTT from the purified AMP fraction which can be challenging (even by dialysis) due to the small size of the AMP. While promising results were obtained with CM4-int<sub>Red</sub>-A60<sub>his</sub>, the construction <sub>his</sub>A60-int<sub>pH</sub>-CM4 presents additional issues such as residual splicing. The adapters designed for the constructions can be also used for the functionalization of surfaces: the introduction of a terminal Cysteine allows the immobilization of AMPs onto surfaces through grafting. This strategy will be validated in the future and can have high potential for the functionalization of materials.

The other construction (int<sub>con</sub>-C-A60<sub>his</sub>-int<sub>con</sub>-N) was designed for the concatenation of PBPs for the purpose of obtaining protein polymers with very high molecular weights (above the limit of *E. coli*) that otherwise, are not possible to be produced by bacteria. In this strategy, the ineffective conditional splicing represented the major problem, as splicing occurs independently of the presence of rapamycin. This could be solved by using a split intein with less effective splicing. Other issue to be solved is the circularization of the protein as it demonstrated to promote circularization at some extent. However, circularization can be beneficial for instance, for the circularization of some AMPs which is an interesting subject to be explored.

#### 4. References

- Amitai, G., Dassa, B., & Pietrokovski, S. (2004). Protein Splicing of Inteins with Atypical Glutamine and Aspartate C-terminal Residues. *Journal of Biological Chemistry*, *279*(5), 3121-3131. <https://doi.org/10.1074/jbc.m311343200>
- Balls, A. (1942). A crystalline protein obtained from a lipoprotein of wheat flour. *Cereal Chemistry*, *19*, 279–288.
- Batjargal, S., Walters, C., & Petersson, E. (2015). Inteins as Traceless Purification Tags for Unnatural Amino Acid Proteins. *Journal Of The American Chemical Society*, *137*(5), 1734-1737. <https://doi.org/10.1021/ja5103019>
- Berrade, L., Kwon, Y., & Camarero, J. (2010). Photomodulation of Protein *Trans*-Splicing Through Backbone Photocaging of the DnaE Split Intein. *Chembiochem*, *11*(10), 1368-1372. <https://doi.org/10.1002/cbic.201000157>
- Bowen, C., Dai, B., Sargent, C., Bai, W., Ladiwala, P., & Feng, H. *et al.* (2018). Recombinant Spidroins Fully Replicate Primary Mechanical Properties of Natural Spider Silk. *Biomacromolecules*, *19*(9), 3853-3860. <https://doi.org/10.1021/acs.biomac.8b00980>
- Brace, L., Southworth, M., Tori, K., Cushing, M., & Perler, F. (2010). The *Deinococcus radiodurans* Snf2 intein caught in the act: Detection of the Class 3 intein signature Block F branched intermediate. *Protein Science*, *19*(8), 1525-1533. <https://doi.org/10.1002/pro.431>
- Brenzel, S., Kurpiers, T., & Mootz, H. (2006). Engineering Artificially Split Inteins for Applications in Protein Chemistry: Biochemical Characterization of the Split *Ssp* DnaB Intein and Comparison to the Split *Sce* VMA Intein. *Biochemistry*, *45*(6), 1571-1578. <https://doi.org/10.1021/bi051697+>
- Burgess-Brown, N. (2017). *Heterologous gene expression in E.coli*. New York: Humana Press.
- Callahan, B., Stanger, M., & Belfort, M. (2013). A redox trap to augment the intein toolbox. *Biotechnology And Bioengineering*, *110*(6), 1565-1573. <https://doi.org/10.1002/bit.24821>
- Carrington, J., Cary, S., Parks, T., & Dougherty, W. (1989). A second proteinase encoded by a plant potyvirus genome. *The EMBO Journal*, *8*(2), 365-370. <https://doi.org/10.1002/j.1460-2075.1989.tb03386.x>

- Casali, N., & Preston, A. (2003). *E. coli plasmid vectors*. Totowa, N.J.: Humana Press.
- Cassini, A., Högberg, L., Plachouras, D., Quattrocchi, A., Hoxha, A., & Simonsen, G. *et al.* (2019). Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. *The Lancet Infectious Diseases*, *19*(1), 56-66. [https://doi.org/10.1016/s1473-3099\(18\)30605-4](https://doi.org/10.1016/s1473-3099(18)30605-4)
- Cheng, X., Lu, W., Zhang, S., & Cao, P. (2010). Expression and purification of antimicrobial peptide CM4 by Npro fusion technology in *E. coli*. *Amino Acids*, *39*(5), 1545-1552. <https://doi.org/10.1007/s00726-010-0625-0>
- Chevallet, M., Luche, S., & Rabilloud, T. (2006). Silver staining of proteins in polyacrylamide gels. *Nature Protocols*, *1*(4), 1852-1858. <https://doi.org/10.1007/10.1038/nprot.2006.288>
- Chong, S., & Xu, M. (1997). Protein Splicing of the *Saccharomyces cerevisiae* VMA Intein without the Endonuclease Motifs. *Journal of Biological Chemistry*, *272*(25), 15587-15590. <https://doi.org/10.1074/jbc.272.25.15587>
- Ciragan, A., Aranko, A., Tascon, I., & Iwai, H. (2016). Salt-inducible Protein Splicing in *cis* and *trans* by Inteins from Extremely Halophilic Archaea as a Novel Protein-Engineering Tool. *Journal Of Molecular Biology*, *428*(23), 4573-4588. <https://doi.org/10.1016/j.jmb.2016.10.006>
- Clark, R., Tan, C., Preza, G., Nemeth, E., Ganz, T., & Craik, D. (2011). Understanding the Structure/Activity Relationships of the Iron Regulatory Peptide Hepcidin. *Chemistry & Biology*, *18*(3), 336-343. <https://doi.org/10.1016/j.chembiol.2010.12.009>
- Cooper, A., & Stevens, T. (1995). Protein splicing: self-splicing of genetically mobile elements at the protein level. *Trends In Biochemical Sciences*, *20*(9), 351-356. [https://doi.org/10.1016/s0968-0004\(00\)89075-1](https://doi.org/10.1016/s0968-0004(00)89075-1)
- Cui, C., Zhao, W., Chen, J., Wang, J., & Li, Q. (2006). Elimination of *in vivo* cleavage between target protein and intein in the intein-mediated protein purification systems. *Protein Expression And Purification*, *50*(1), 74-81. <https://doi.org/10.1016/j.pep.2006.05.019>
- Da Costa, A., Machado, R., Ribeiro, A., Collins, T., Thiagarajan, V., Neves-Petersen, M. T., Rodríguez-Cabello, J.C., Gomes A. C. & Casal, M. (2015). Development of elastin-like recombinamer films

- with antimicrobial activity. *Biomacromolecules*, *16*(2), 625-635.  
<https://doi.org/10.1021/bm5016706>.
- Da Costa, A., Pereira, A. M., Gomes, A. C., Rodríguez-Cabello, J. C., Casal M. & Machado, R. (2018). Production of bioactive hepcidin by recombinant DNA tagging with an elastin-like recombinamer. *New Biotechnology*, *46*, 45-53. <https://doi.org/10.1016/j.nbt.2018.07.001>
- Derakhshankhah, H., & Jafari, S. (2018). Cell penetrating peptides: A concise review with emphasis on biomedical applications. *Biomedicine & Pharmacotherapy*, *108*, 1090-1096.  
<https://doi.org/10.1016/j.biopha.2018.09.097>
- Dubos, R. (1939). Studies on a bactericidal agent extracted from a soil Bacillus: I. Preparation of the agent. Its activity *in vitro*. *Journal of Experimental Medicine*, *70*(1), 1-10.  
<https://doi.org/10.1084/jem.70.1.1>
- Dubos, R. (1939). Studies on a bactericidal agent extracted from a soil Bacillus: II. Protective effect of the bactericidal agent against experimental Pneumococcus infections in mice. *Journal of Experimental Medicine*, *70*(1), 11-17. <https://doi.org/10.1084/jem.70.1.11>
- Elleuche, S., & Pöggeler, S. (2007). *Trans*-splicing of an artificially split fungal mini-intein. *Biochemical And Biophysical Research Communications*, *355*(3), 830-834.  
<https://doi.org/10.1016/j.bbrc.2007.02.035>
- Garcia, A., Tai, K., Puttamadappa, S., Shekhtman, A., Ouellette, A., & Camarero, J. (2011). Biosynthesis and Antimicrobial Evaluation of Backbone-Cyclized  $\alpha$ -Defensins. *Biochemistry*, *50*(48), 10508-10519. <https://doi.org/10.1021/bi201430f>
- Grage, S., Afonin, S., Kara, S., Buth, G., & Ulrich, A. (2016). Membrane thinning and thickening induced by membrane-active amphipathic peptides. *Frontiers In Cell And Developmental Biology*, *4*.  
<https://doi.org/10.3389/fcell.2016.00065>
- Guo, Y., Xun, M., & Han, J. (2018). A bovine myeloid antimicrobial peptide (BMAP-28) and its analogs kill pan-drug-resistant *Acinetobacter baumannii* by interacting with outer membrane protein A (OmpA). *Medicine*, *97*(42), e12832. <https://doi.org/10.1097/md.00000000000012832>
- Han, L., Zong, H., Zhou, Y., Pan, Z., Chen, J., & Ding, K. *et al.* (2019). Naturally split intein Npu DnaE mediated rapid generation of bispecific IgG antibodies. *Methods*, *154*, 32-37.  
<https://doi.org/10.1016/j.ymeth.2018.10.001>

- Hassouneh, W., Christensen, T., & Chilkoti, A. (2010). Elastin-like polypeptides as a purification tag for recombinant proteins. *Current Protocols in Protein Science*. <https://doi.org/10.1002/0471140864.ps0611s61>
- Hiraga, K., Derbyshire, V., Dansereau, J., Van Roey, P., & Belfort, M. (2005). Minimization and stabilization of the *Mycobacterium tuberculosis* recA intein. *Journal Of Molecular Biology*, *354*(4), 916-926. <https://doi.org/10.1016/j.jmb.2005.09.088>
- Hirata, R., Ohsumi, Y., Nakano, A., Kawasaki, H., Suzuki, K., & Anraku, Y. (1990). Molecular structure of a gene, VMA1, encoding the catalytic subunit of H<sup>+</sup>-translocating adenosine triphosphatase from vacuolar membranes of *Saccharomyces cerevisiae*. *Journal of Biological Chemistry*, *265*(12), 6726-6733.
- Hirsch, J. (1956). Phagocytin: A bactericidal substance from polymorphonuclear leucocytes. *Journal Of Experimental Medicine*, *103*(5), 589-611. <https://doi.org/10.1084/jem.103.5.589>
- Houamel, D., Ducrot, N., Lefebvre, T., Daher, R., Moulouel, B., & Sari, M. *et al.* (2015). Heparin as a Major Component of Renal Antibacterial Defenses against Uropathogenic *Escherichia coli*. *Journal Of The American Society Of Nephrology*, *27*(3), 835-846. <https://doi.org/10.1681/asn.2014101035>
- Hu, F., Ke, T., Li, X., Mao, P., Jin, X., & Hui, F. *et al.* (2009). Expression and Purification of an Antimicrobial Peptide by Fusion with Elastin-like Polypeptides in *Escherichia coli*. *Applied Biochemistry And Biotechnology*, *160*(8), 2377-2387. <https://doi.org/10.1007/s12010-009-8850-2>
- Huang, H.-J., Ross, C. R., & Blecha, F. (1997). Chemoattractant properties of PR-39, a neutrophil antibacterial peptide. *Journal of Leukocyte Biology*. <https://doi.org/10.1002/jlb.61.5.624>
- Huang, Y., Huang, J., & Chen, Y. (2010). Alpha-helical cationic antimicrobial peptides: relationships of structure and function. *Protein & Cell*, *1*(2), 143-152. <https://doi.org/10.1007/s13238-010-0004-3>
- Khamis, A., Essack, M., Gao, X., & Bajic, V. (2014). Distinct profiling of antimicrobial peptide families. *Bioinformatics*, *31*(6), 849-856. <https://doi.org/10.1093/bioinformatics/btu738>

- Lavery, G., Gorman, S., & Gilmore, B. (2011). The Potential of Antimicrobial Peptides as Biocides. *International Journal Of Molecular Sciences*, *12*(10), 6566-6596. <https://doi.org/10.3390/ijms12106566>
- Lee, E., Min, K., Chang, Y., & Kwon, Y. (2018). Efficient and wash-free labeling of membrane proteins using engineered *Npu* DnaE split-inteins. *Protein Science*, *27*(9), 1568-1574. <https://doi.org/10.1002/pro.3455>
- Li, J., Zhang, J., Xu, X., Han, Y., Cui, X., Chen, Y., & Zhang, S. (2011). The antibacterial peptide ABP-CM4: the current state of its production and applications. *Amino Acids*, *42*(6), 2393-2402. <https://doi.org/10.1007/s00726-011-0982-3>
- Li, Y., Xiang, Q., Zhang, Q., Huang, Y., & Su, Z. (2012). Overview on the recent study of antimicrobial peptides: Origins, functions, relative mechanisms and application. *Peptides*, *37*(2), 207-215. <https://doi.org/10.1016/j.peptides.2012.07.001>
- Liu, H., Trinh, T., Dong, H., Keith, R., Nelson, D., & Liu, C. (2012). Iron Regulator Heparin Exhibits Antiviral Activity against Hepatitis C Virus. *Plos ONE*, *7*(10), e46631. <https://doi.org/10.1371/journal.pone.0046631>
- Machado, R., Azevedo-Silva, J., Correia, C., Collins, T., Arias, F. J., Rodríguez-Cabello, J. C. & Casal, M. (2013a). High level expression and facile purification of recombinant silk-elastin-like polymers in auto induction shake flask cultures. *AMB Express*, *3*(1), 1-15. <https://doi.org/10.1186/2191-0855-3-11>
- Machado, R., da Costa, A., Sencadas, V., Garcia-Arévalo, C., Costa, C. M, Padrão, J., Gomes, A. C., Lanceros-Méndez, S., Rodríguez-Cabello, J. C. & Casal, M. (2013b). Electrospun silk-elastin-like fibre mats for tissue engineering applications. *Biomedical Materials*, *8*(6), 065009. <https://doi.org/10.1088/1748-6041/8/6/065009>
- Machado, R., Ribeiro, A., Padrão, J., Silva, D., Nobre, A., & Teixeira, J. *et al.* (2009). Exploiting the Sequence of Naturally Occurring Elastin: Construction, Production and Characterization of a Recombinant Thermoplastic Protein-Based Polymer. *Journal Of Nano Research*, *6*, 133-145. <https://doi.org/10.4028/www.scientific.net/jnanor.6.133>
- Madigan, M., Martinko, J., Bender, K., Buckley, D., & Stahl, D. (2015). *Brock biology of microorganisms* (14<sup>th</sup> ed.). Boston: Pearson.



- Marcos, J., & Manzanares, P. (2012). Chapter 8: Antimicrobial Peptides. In J. Lagaron, M. Ocio & A. Lopez-Rubio, *Antimicrobial Polymers* (1st ed., pp. 195-225). Hoboken: Wiley.
- McPherson, D., Xu, J., & Urry, D. (1996). Product Purification by Reversible Phase Transition Following *Escherichia coli* Expression of Genes Encoding up to 251 Repeats of the Elastomeric Pentapeptide GVGVP. *Protein Expression And Purification*, 7(1), 51-57. <https://doi.org/10.1006/prep.1996.0008>
- Meyer, D., & Chilkoti, A. (1999). Purification of recombinant proteins by fusion with thermally-responsive polypeptides. *Nature Biotechnology*, 17(11), 1112-1115. <https://doi.org/10.1038/15100>
- Meyer, D., & Chilkoti, A. (2002). Genetically Encoded Synthesis of Protein-Based Polymers with Precisely Specified Molecular Weight and Sequence by Recursive Directional Ligation: Examples from the Elastin-like Polypeptide System. *Biomacromolecules*, 3(2), 357-367. <https://doi.org/10.1021/bm015630n>
- Michels, K., Nemeth, E., Ganz, T., & Mehrad, B. (2015). Hepcidin and Host Defense against Infectious Diseases. *PLOS Pathogens*, 11(8), e1004998. <https://doi.org/10.1371/journal.ppat.1004998>
- Mills, K., Lew, B., Jiang, S., & Paulus, H. (1998). Protein splicing in *trans* by purified N- and C-terminal fragments of the *Mycobacterium tuberculosis* RecA intein. *Proceedings Of The National Academy Of Sciences*, 95(7), 3543-3548. <https://doi.org/10.1073/pnas.95.7.3543>
- Mishra, B., Basu, A., Chua, R., Saravanan, R., Tambyah, P., & Ho, B. *et al.* (2014). Site specific immobilization of a potent antimicrobial peptide onto silicone catheters: evaluation against urinary tract infection pathogens. *Journal Of Materials Chemistry B*, 2(12), 1706. <https://doi.org/10.1039/c3tb21300e>
- Mootz, H., & Muir, T. (2002). Protein Splicing Triggered by a Small Molecule. *Journal Of The American Chemical Society*, 124(31), 9044-9045. <https://doi.org/10.1021/ja026769o>
- Münch, D., & Sahl, H. (2015). Structural variations of the cell wall precursor lipid II in Gram-positive bacteria – Impact on binding and efficacy of antimicrobial peptides. *Biochimica Et Biophysica Acta (BBA) - Biomembranes*, 1848(11), 3062-3071. <https://doi.org/10.1016/j.bbamem.2015.04.014>
- Nicolas, P. (2009). Multifunctional host defense peptides: intracellular-targeting antimicrobial peptides. *FEBS Journal*, 276(22), 6483-6496. <https://doi.org/10.1111/j.1742-4658.2009.07359.x>

- Pachón-Ibáñez, M., Smani, Y., Pachón, J., & Sánchez-Céspedes, J. (2017). Perspectives for clinical use of engineered human host defense antimicrobial peptides. *FEMS Microbiology Reviews*, *41*(3), 323-342. <https://doi.org/10.1093/femsre/fux012>
- Park, Y., & Hahm, K. (2005). Antimicrobial peptides (AMPs): peptide structure and mode of action. *BMB Reports*, *38*(5), 507-516. <https://doi.org/10.5483/bmbrep.2005.38.5.507>
- Perler, F. B., Davis, E. O., Dean, G. E., Gimble, F. S., Jack, W. E., Neff, N., Noren, C., Thorner, J., & Belfort, M. (1994). Protein splicing elements : inteins and exteins a definition of terms and recommended nomenclature. *Nucleic Acids Research*, *22*(7), 1125–1127. <https://doi.org/10.1093/nar/22.7.1125>
- Perler, F. (2002). InBase: the Intein Database. *Nucleic Acids Research*, *30*(1), 383-384. <https://doi.org/10.1093/nar/30.1.383>
- Pietrangelo, A., & Trautwein, C. (2004). Mechanisms of Disease: the role of hepcidin in iron homeostasis - implications for hemochromatosis and other disorders. *Nature Clinical Practice Gastroenterology & Hepatology*, *1*(1), 39-45. <https://doi.org/10.1038/ncpgasthep0019>
- Petrokovski, S. (1998). Identification of a virus intein and a possible variation in the protein-splicing reaction. *Current Biology*, *8*(18), R634-R638. [https://doi.org/10.1016/s0960-9822\(07\)00409-5](https://doi.org/10.1016/s0960-9822(07)00409-5)
- Pigeon, C., Ilyin, G., Courselaud, B., Leroyer, P., Turlin, B., Brissot, P., & Loréal, O. (2000). A New Mouse Liver-specific Gene, Encoding a Protein Homologous to Human Antimicrobial Peptide Hepcidin, Is Overexpressed during Iron Overload. *Journal Of Biological Chemistry*, *276*(11), 7811-7819. <https://doi.org/10.1074/jbc.m008923200>
- Qi, X., Meng, Q., & Liu, X. (2011). Spontaneous C-cleavage of a mini-intein without its conserved N-terminal motif A. *FEBS Letters*, *585*(15), 2513-2518. <https://doi.org/10.1016/j.febslet.2011.06.035>
- Rais-Beghdadi, C., Roggero, M., Fasel, N., & Reymond, C. (1998). Purification of recombinant proteins by chemical removal of the affinity tag. *Applied Biochemistry And Biotechnology*, *74*(2), 95-103. <https://doi.org/10.1007/bf02787176>

- Ramirez, M., Valdes, N., Guan, D., & Chen, Z. (2012). Engineering split intein DnaE from *Nostoc punctiforme* for rapid protein purification. *Protein Engineering Design And Selection*, 26(3), 215-223. <https://doi.org/10.1093/protein/gzs097>
- Risso, A., Braidot, E., Sordano, M., Vianello, A., Macri, F., & Skerlavaj, B. *et al.* (2002). BMAP-28, an Antibiotic Peptide of Innate Immunity, Induces Cell Death through Opening of the Mitochondrial Permeability Transition Pore. *Molecular And Cellular Biology*, 22(6), 1926-1935. <https://doi.org/10.1128/mcb.22.6.1926-1935.2002>
- Reguera, J., Urry, D., Parker, T., McPherson, D., & Rodriguez-Cabello, J. (2007). Effect of NaCl on the Exothermic and Endothermic Components of the Inverse Temperature Transition of a Model Elastin-like Polymer. *Biomacromolecules*, 8(2), 354-358. <https://doi.org/10.1021/bm060936l>
- Reitter, J., Cousin, C., Nicastrì, M., Jaramillo, M., & Mills, K. (2016). Salt-Dependent Conditional Protein Splicing of an Intein from *Halobacterium salinarum*. *Biochemistry*, 55(9), 1279-1282. <https://doi.org/10.1021/acs.biochem.6b00128>
- Rosano, G. L. & Ceccarelli, E. A. (2014). Recombinant protein expression in *Escherichia coli*: advances and challenges. *Frontiers in Microbiology*, 5, 172. <https://doi.org/10.3389/fmicb.2014.00172>
- Schneider, C., Rasband, W., & Eliceiri, K. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 9(7), 671-675. <https://doi.org/10.1038/nmeth.2089>
- Schwartz, E., Saez, L., Young, M., & Muir, T. (2006). Post-translational enzyme activation in an animal via optimized conditional protein splicing. *Nature Chemical Biology*, 3(1), 50-54. <https://doi.org/10.1038/nchembio832>
- Shai, Y. (2002). Mode of action of membrane active antimicrobial peptides. *Biopolymers*, 66(4), 236-248. <https://doi.org/10.1002/bip.10260>
- Shi, C., Meng, Q., & Wood, D. (2012). A dual ELP-tagged split intein system for non-chromatographic recombinant protein purification. *Applied Microbiology And Biotechnology*, 97(2), 829-835. <https://doi.org/10.1007/s00253-012-4601-3>
- Shingledecker, K., Jiang, S., & Paulus, H. (1998). Molecular dissection of the *Mycobacterium tuberculosis* RecA intein: design of a minimal intein and of a *trans*-splicing system involving two intein fragments. *Gene*, 207(2), 187-195. [https://doi.org/10.1016/s0378-1119\(97\)00624-0](https://doi.org/10.1016/s0378-1119(97)00624-0)

- Sierra, J., Fusté, E., Rabanal, F., Vinuesa, T., & Viñas, M. (2017). An overview of antimicrobial peptides and the latest advances in their development. *Expert Opinion On Biological Therapy*, *17*(6), 663-676. <https://doi.org/10.1080/14712598.2017.1315402>
- Sinha, R., & Shukla, P. (2019). Antimicrobial Peptides: Recent Insights on Biotechnological Interventions and Future Perspectives. *Protein & Peptide Letters*, *26*(2), 79-87. <https://doi.org/10.1080/10.2174/0929866525666181026160852>
- Skretas, G., & Wood, D. (2005). Regulation of protein activity with small-molecule-controlled inteins. *Protein Science*, *14*(2), 523-532. <https://doi.org/10.1110/ps.04996905>
- Southworth, M., Adam, E., Panne, D., Byer, R., Kautz, R., & Perler, F. (1998). Control of protein splicing by intein fragment reassembly. *The EMBO Journal*, *17*(4), 918-926. <https://doi.org/10.1093/emboj/17.4.918>
- Southworth, M., Benner, J., & Perler, F. (2000). An alternative protein splicing mechanism for inteins lacking an N-terminal nucleophile. *The EMBO Journal*, *19*(18), 5019-5026. <https://doi.org/10.1093/emboj/19.18.5019>
- Spohn, R., Daruka, L., Lázár, V., Martins, A., Vidovics, F., & Grézal, G. *et al.* (2019). Integrated evolutionary analysis reveals antimicrobial peptides with limited resistance. *Nature Communications*, *10*(1). <https://doi.org/10.1038/s41467-019-12364-6>
- Stevens, A., Brown, Z., Shah, N., Sekar, G., Cowburn, D., & Muir, T. (2016). Design of a Split Intein with Exceptional Protein Splicing Activity. *Journal Of The American Chemical Society*, *138*(7), 2162-2165. <https://doi.org/10.1021/jacs.5b13528>
- Tejeda-Montes, E., Klymov, A., Nejadnik, M., Alonso, M., Rodriguez-Cabello, J., Walboomers, X., & Mata, A. (2014). Mineralization and bone regeneration using a bioactive elastin-like recombinamer membrane. *Biomaterials*, *35*(29), 8339-8347. <https://doi.org/10.1016/j.biomaterials.2014.05.095>
- Tu, Y-Z., Qu, X-M., & Xu, T-S. (1989). Purification, characterization and structure of CM2Ph1, an antibacterial peptide from *Bombyx mori*. *Science China B*, *32*(9), 1072-1081.
- Van Roey, P., Pereira, B., Li, Z., Hiraga, K., Belfort, M., & Derbyshire, V. (2007). Crystallographic and Mutational Studies of Mycobacterium tuberculosis recA Mini-inteins Suggest a Pivotal Role for a

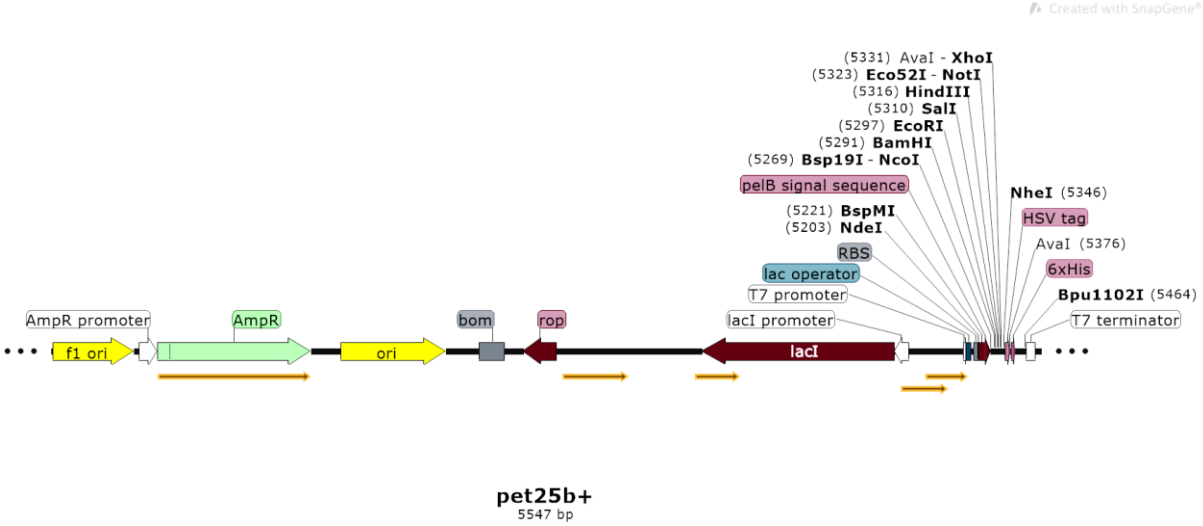
- Highly Conserved Aspartate Residue. *Journal Of Molecular Biology*, 367(1), 162-173.  
<https://doi.org/10.1016/j.jmb.2006.12.050>
- Ventola, C. (2015). The antibiotic resistance crisis: part 1: causes and threats. *Pharmacy and Therapeutics*, 40(4), 277–283.
- Waghu, F., Barai, R., Gurung, P., & Idicula-Thomas, S. (2015). CAMP<sub>RS</sub>: a database on sequences, structures and signatures of antimicrobial peptides: Table 1. *Nucleic Acids Research*, 44(D1), 1094-1097. <https://doi.org/10.1093/nar/gkv1051>
- Wang, G., Li, X., & Wang, Z. (2015). APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Research*, 44(D1), D1087-D1093.  
<https://doi.org/10.1093/nar/gkv1278>
- Wu, H., Hu, Z., & Liu, X. (1998). Protein *trans*-splicing by a split intein encoded in a split DnaE gene of *Synechocystis sp.* PCC6803. *Proceedings Of The National Academy Of Sciences*, 95(16), 9226-9231. <https://doi.org/10.1073/pnas.95.16.9226>
- Wu, W., Mee, C., Califano, F., Banki, R., & Wood, D. (2006). Recombinant protein purification by self-cleaving aggregation tag. *Nature Protocols*, 1(5), 2257-2262.  
<https://doi.org/10.1038/nprot.2006.314>
- Yamashiro, C., Stevens, T., Neff, N., Kane, P., Goebel, M., & Wolczyk, D. (2006). Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. *Science*, 250(4981), 651–657. <https://doi.org/10.1126/science.2146742>
- Yeaman, M., & Yount, N. (2003). Mechanisms of Antimicrobial Peptide Action and Resistance. *Pharmacological Reviews*, 55(1), 27-55. <https://doi.org/10.1124/pr.55.1.2>
- Zasloff, M. (2002). Antimicrobial peptides of multicellular organisms. *Nature*, 415(6870), 389-395.  
<https://doi.org/10.1038/41538>
- Zhao, Z., Lu, W., Dun, B., Jin, D., Ping, S., & Zhang, W. *et al.* (2008). Purification of green fluorescent protein using a two-intein system. *Applied Microbiology And Biotechnology*, 77(5), 1175-1180.  
<https://doi.org/10.1007/s00253-007-1233-0>
- Zorko, M., & Jerala, R. (2009). Production of Recombinant Antimicrobial Peptides in Bacteria. *Methods In Molecular Biology*, 61-76. [https://doi.org/10.1007/978-1-60761-594-1\\_5](https://doi.org/10.1007/978-1-60761-594-1_5)



5. Annexes and Appendixes

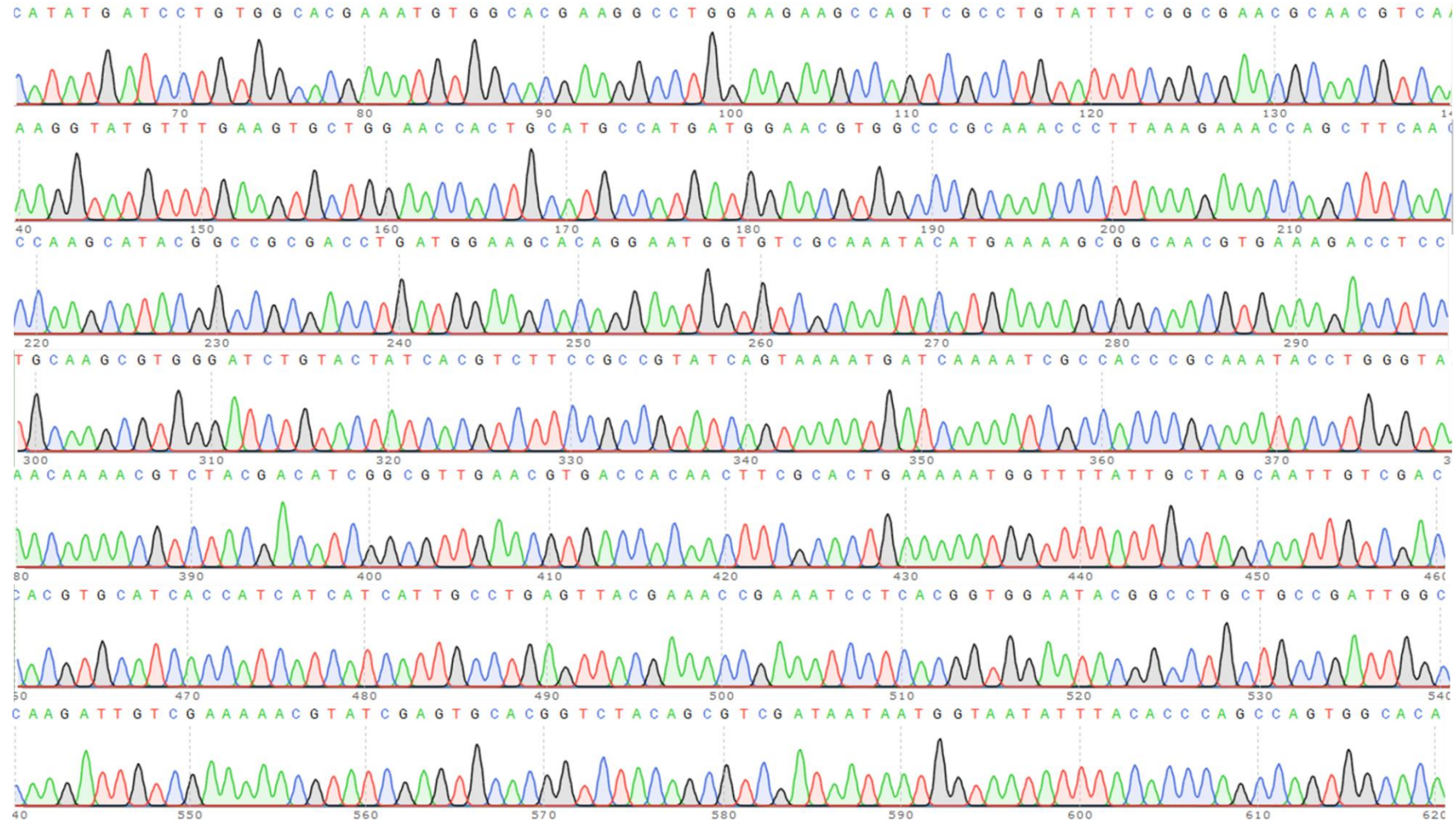
Annex A

pet25b+ sequence scheme



Annex B

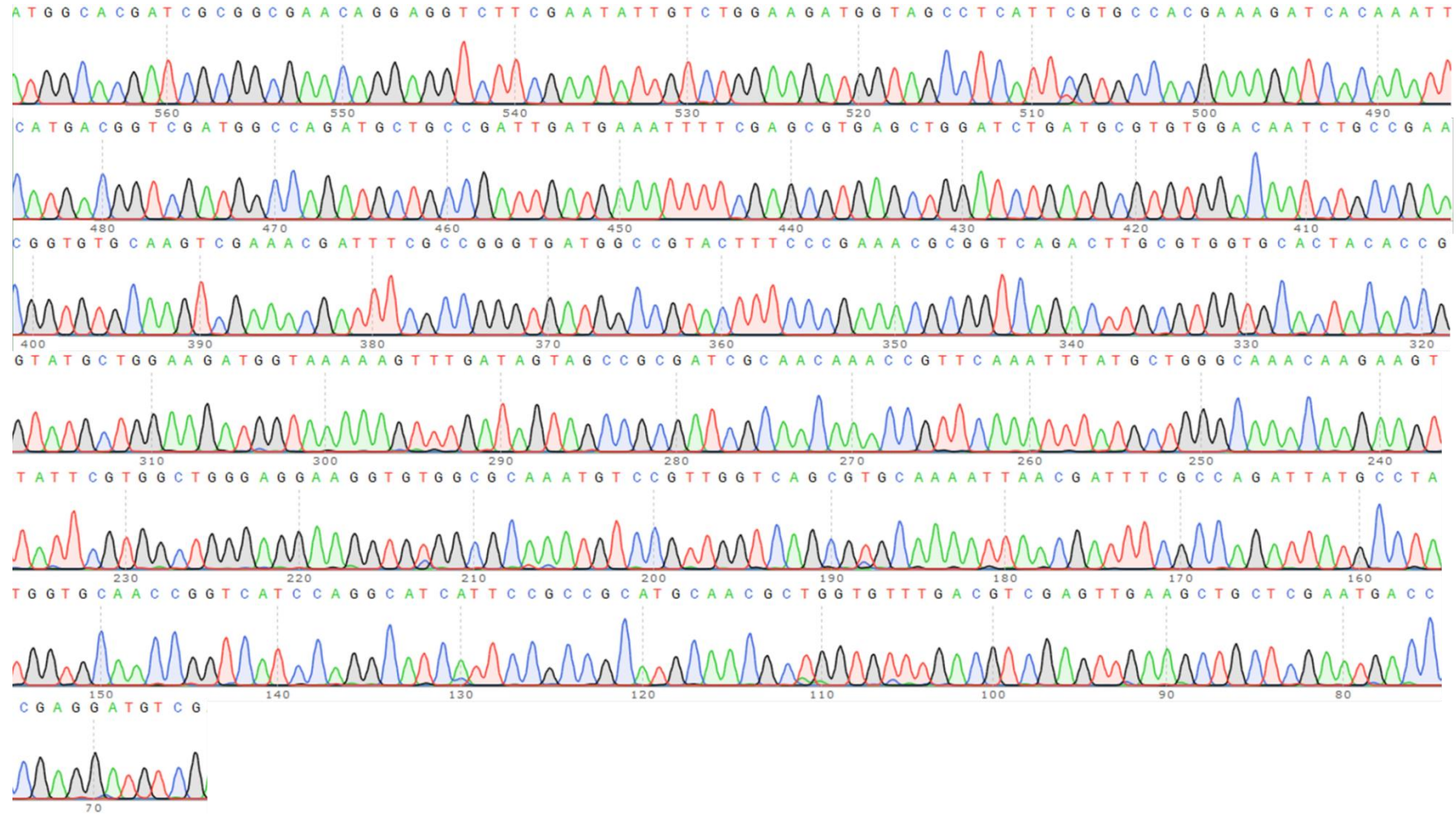
pA2H sequencing chromatogram using the primer T7 Fwd





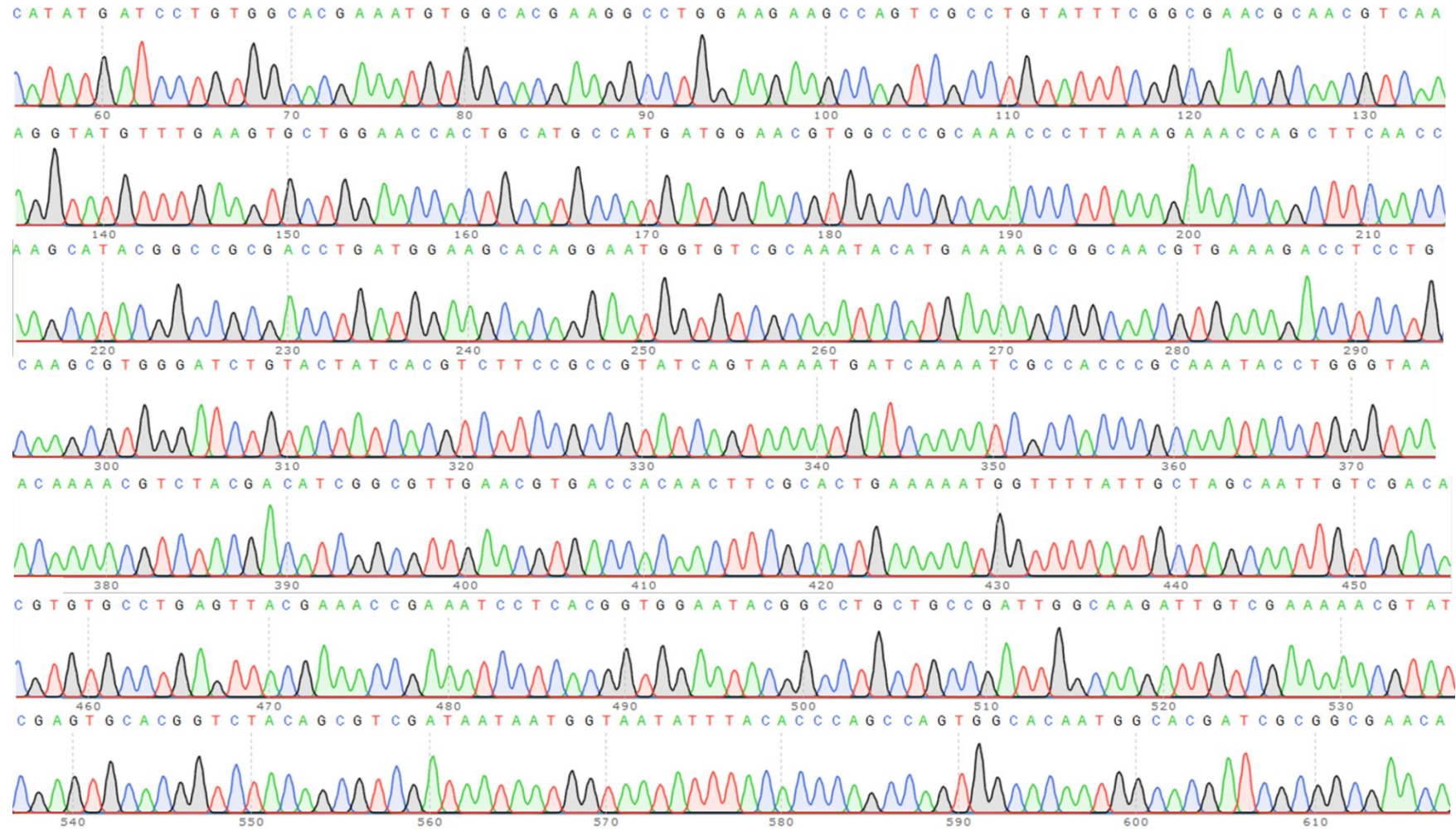
Annex C

Continuation of the pA2H sequencing chromatogram using the primer T7 terminal



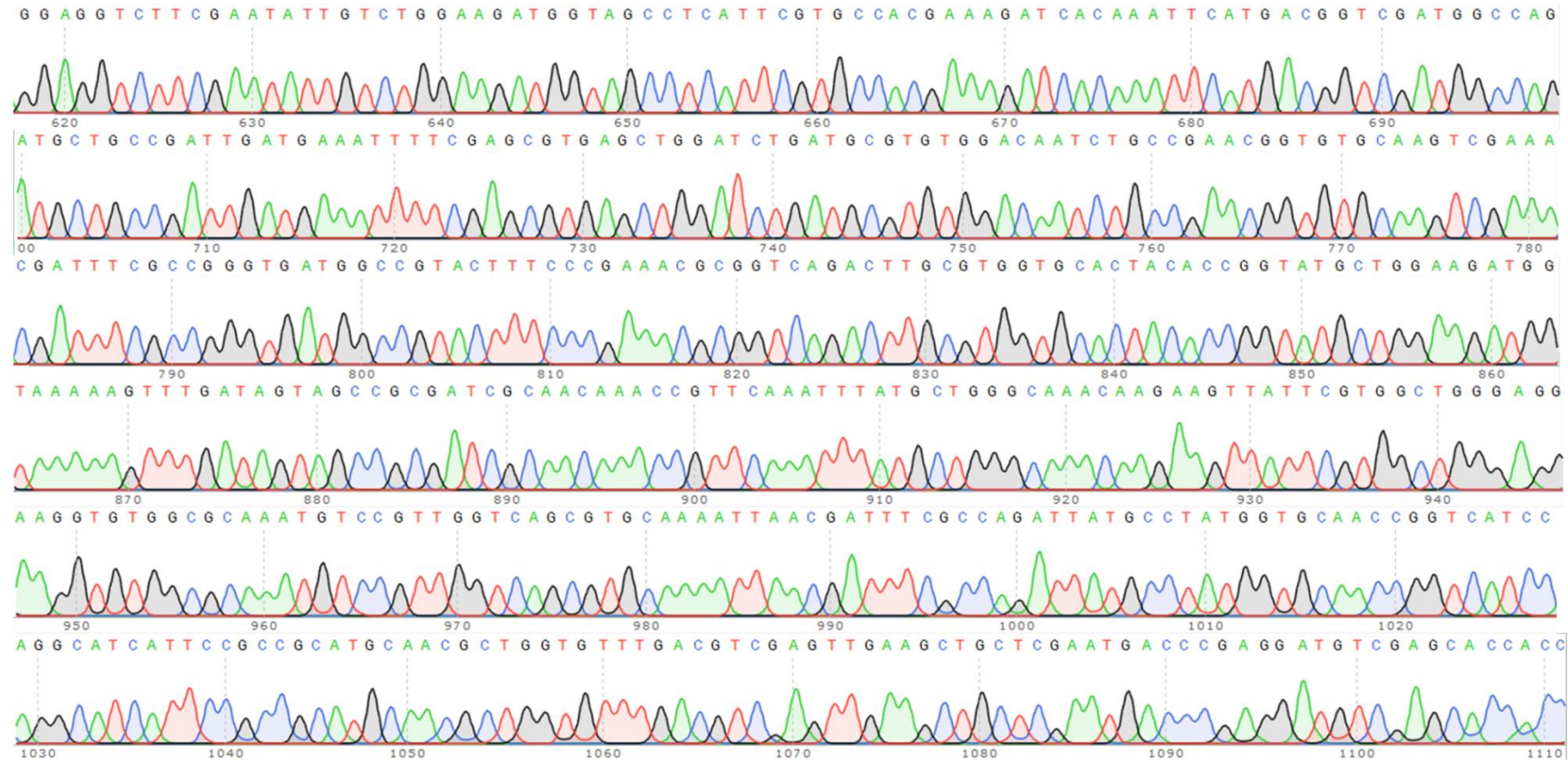
Annex D

pA2Ø sequencing chromatogram using the primer T7 Fwd



Annex E

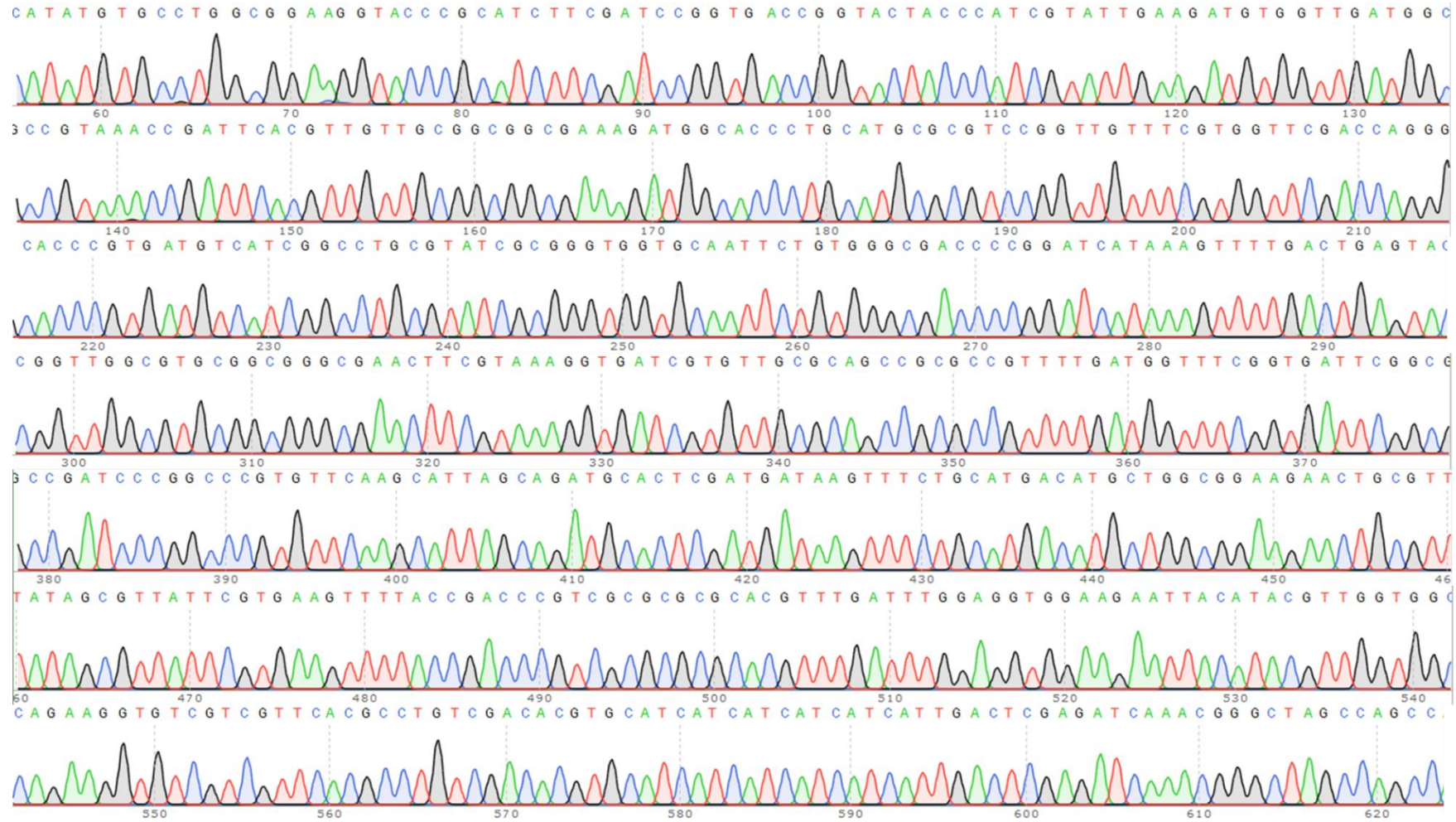
Continuation of the pA2 $\phi$  sequencing chromatogram using the primer T7 Fwd





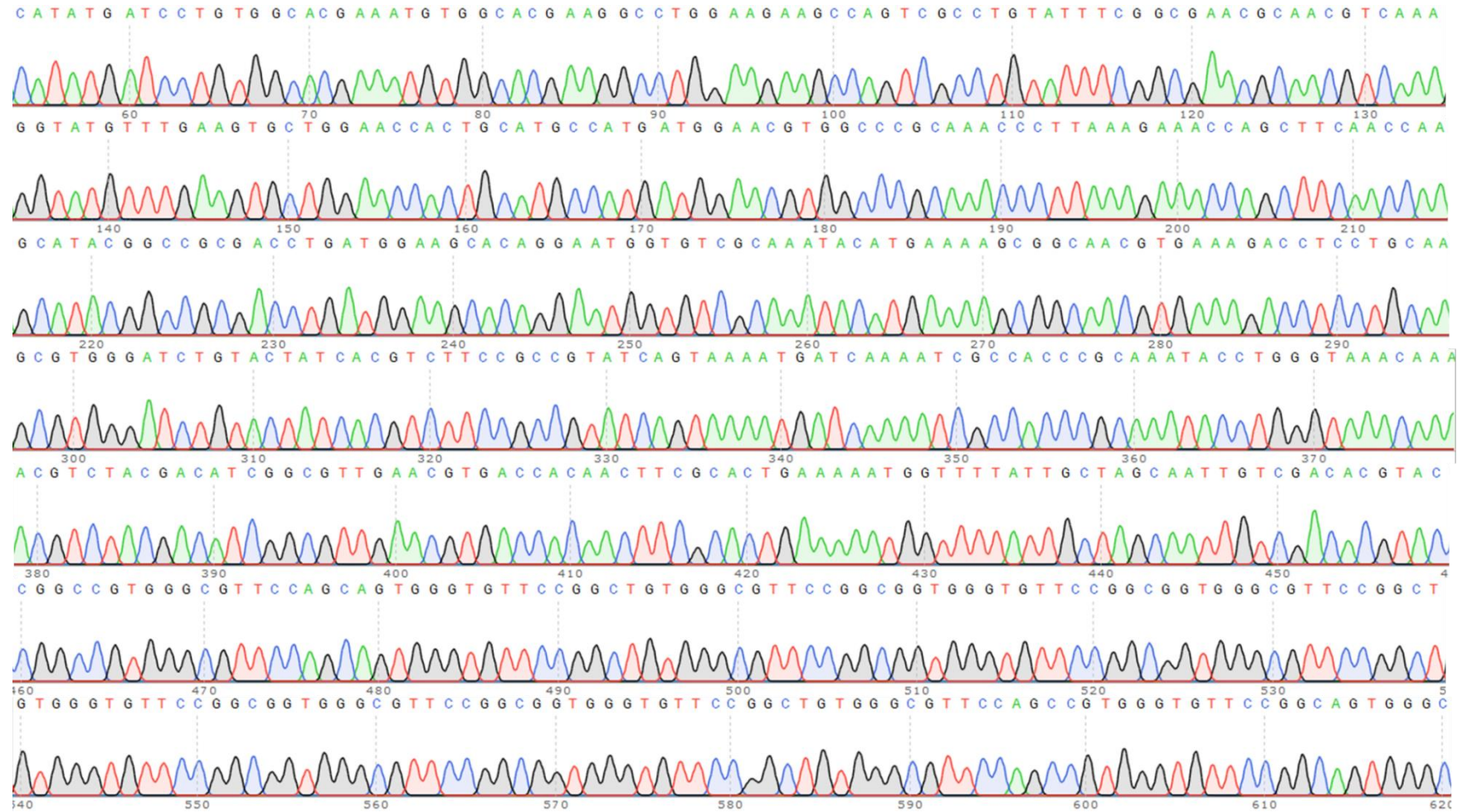
Annex G

pA3N sequencing chromatogram using the primer T7 Fwd



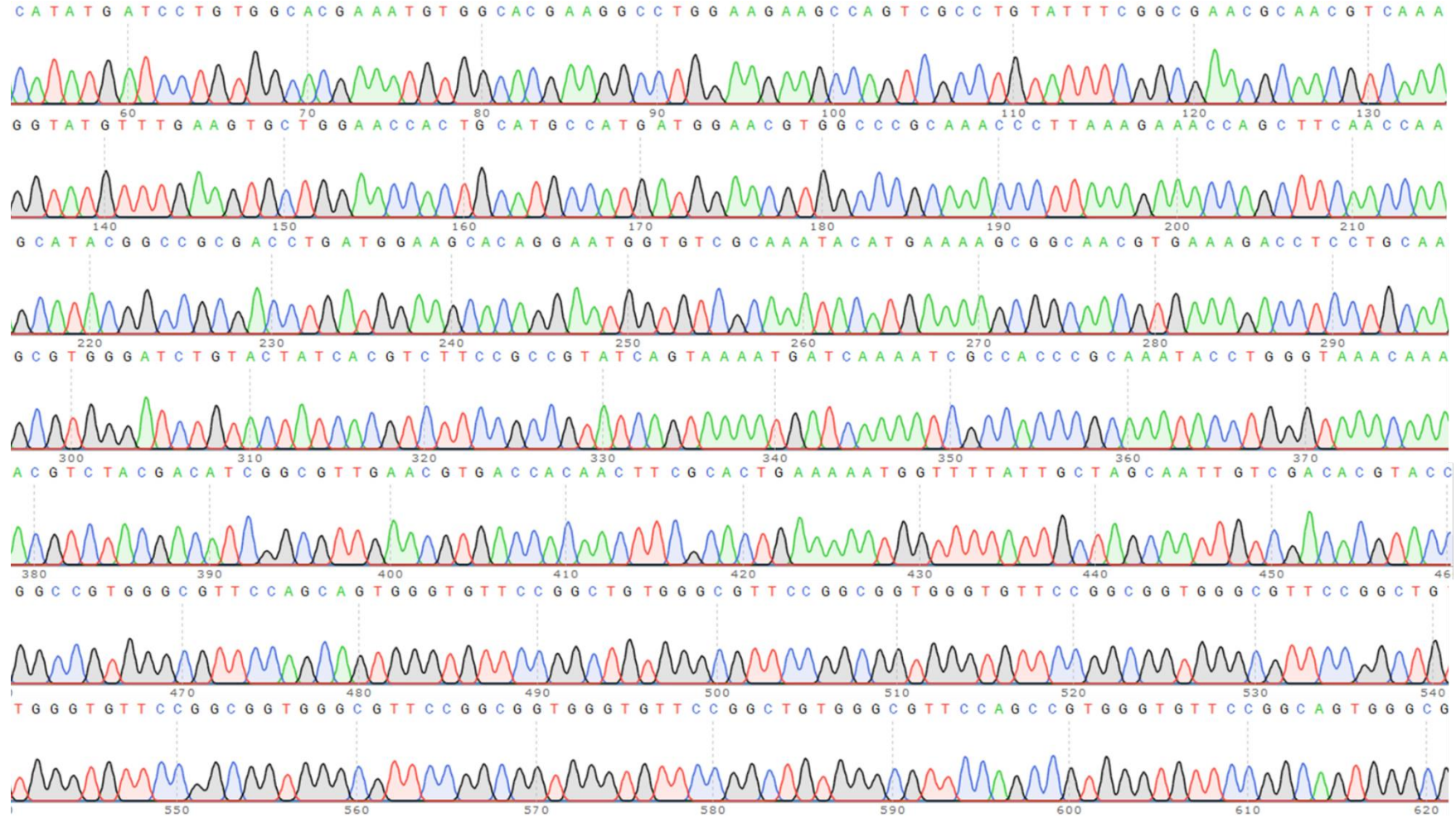
Annex H

pA2H::A60 sequencing chromatogram using the primer T7 Fwd



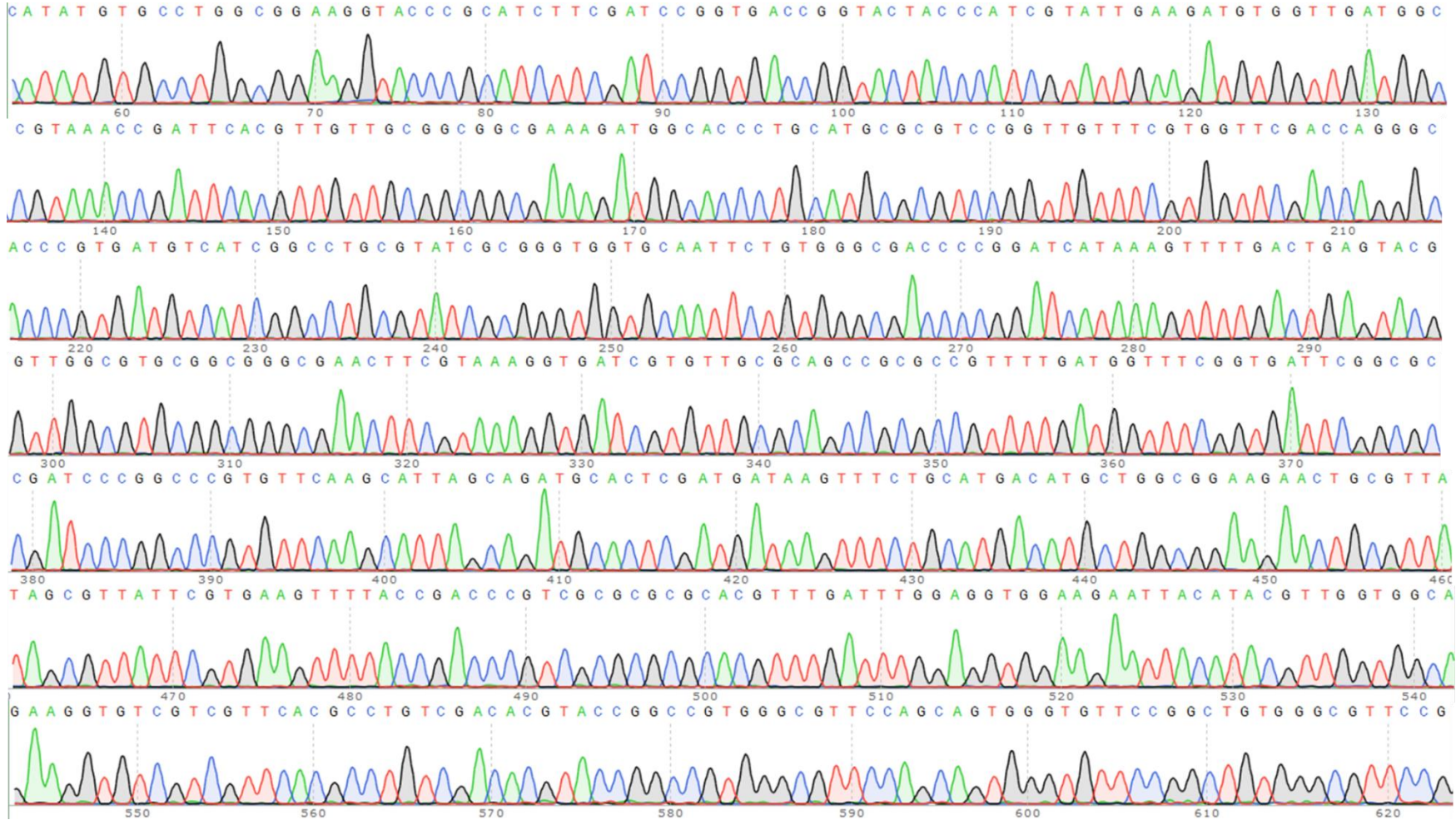
Annex I

pA2 $\emptyset$ ::A60 sequencing chromatogram using the primer T7 Fwd



Annex J

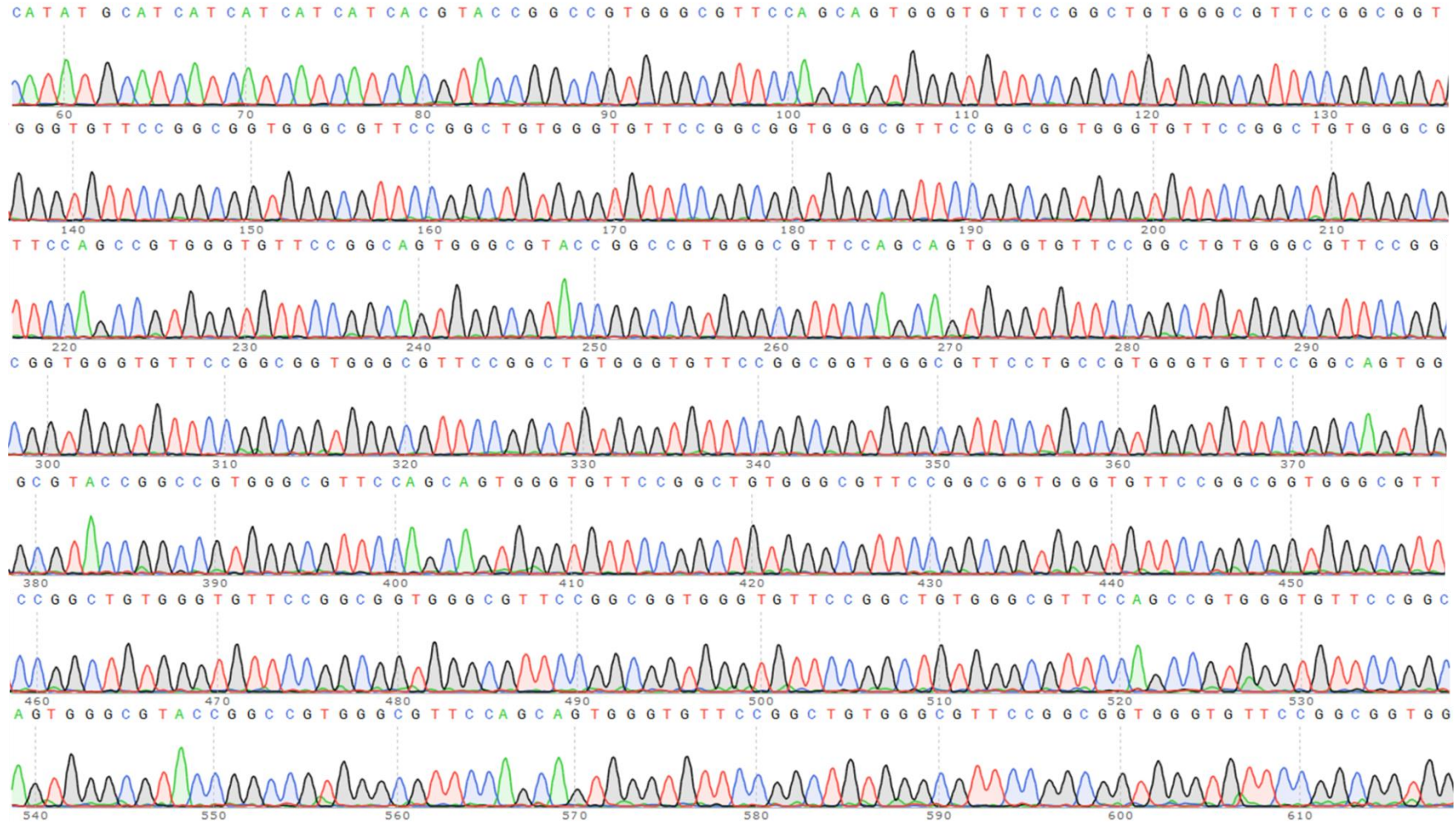
pA3N::A60 sequencing chromatogram using the primer T7 Fwd





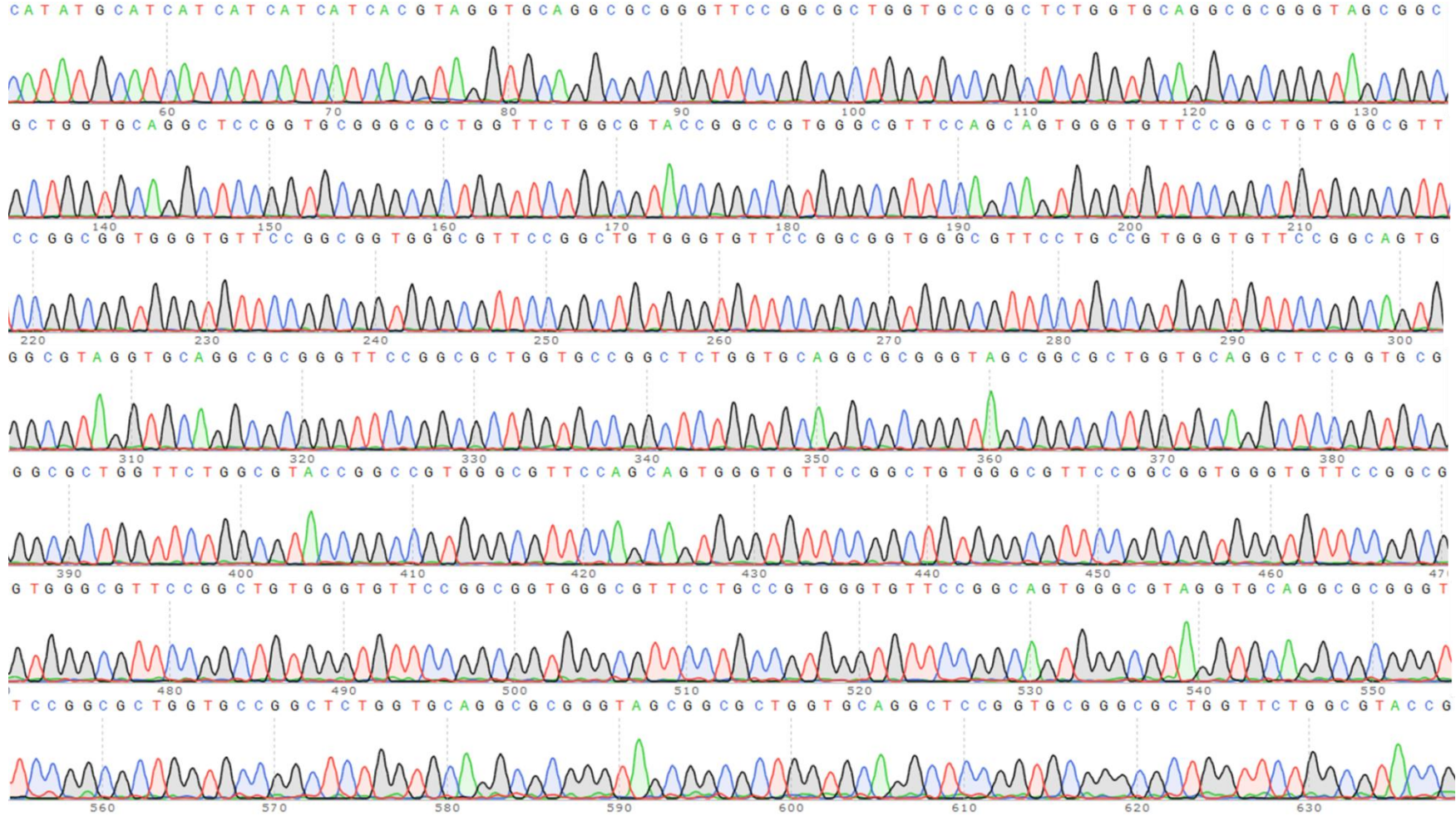
Annex K

pA1C::A60 sequencing chromatogram using the primer T7 Fwd



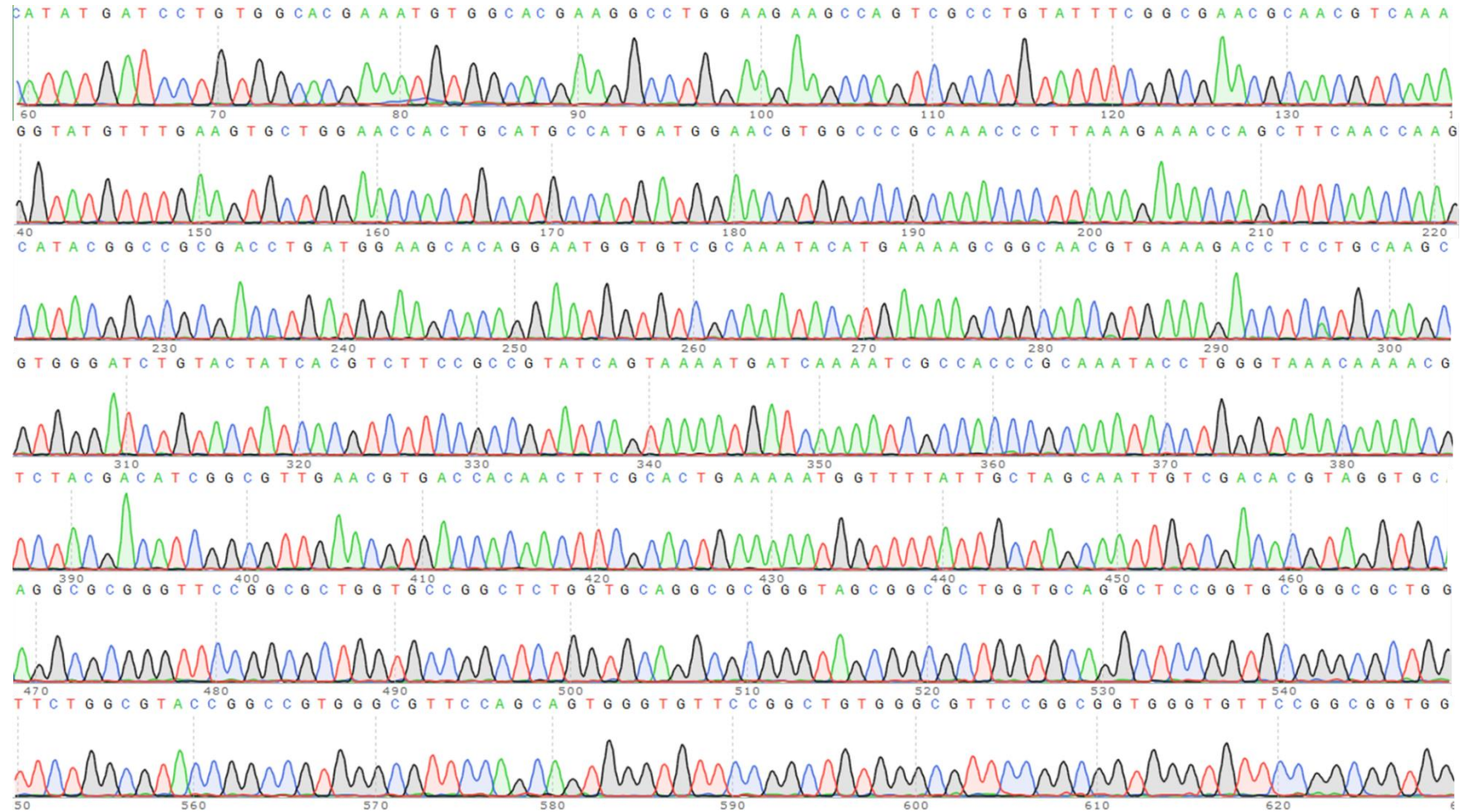
Annex L

pA1C::SELP59 sequencing chromatogram using the primer T7 Fwd



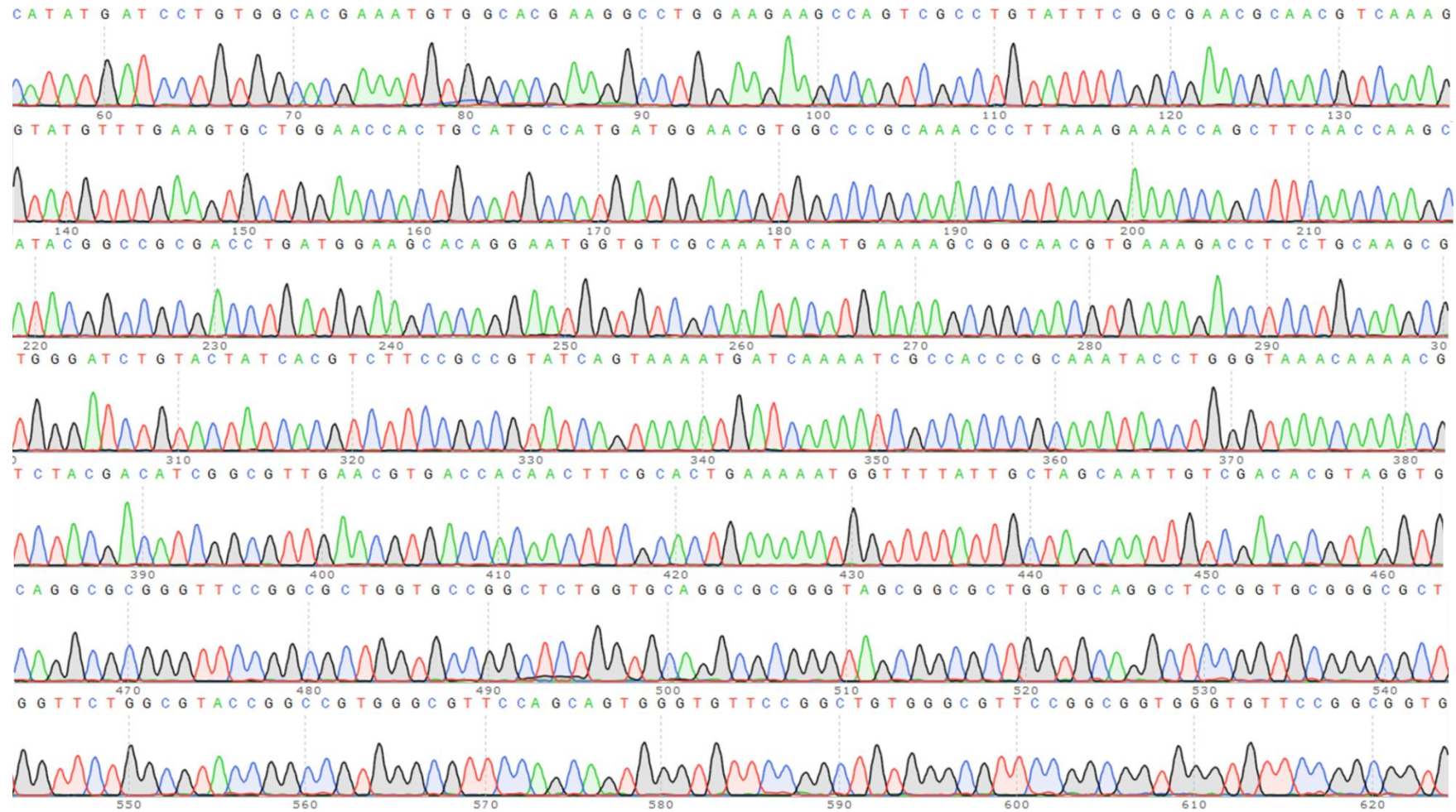
Annex M

pA2H::SELP59 sequencing chromatogram using the primer T7 Fwd



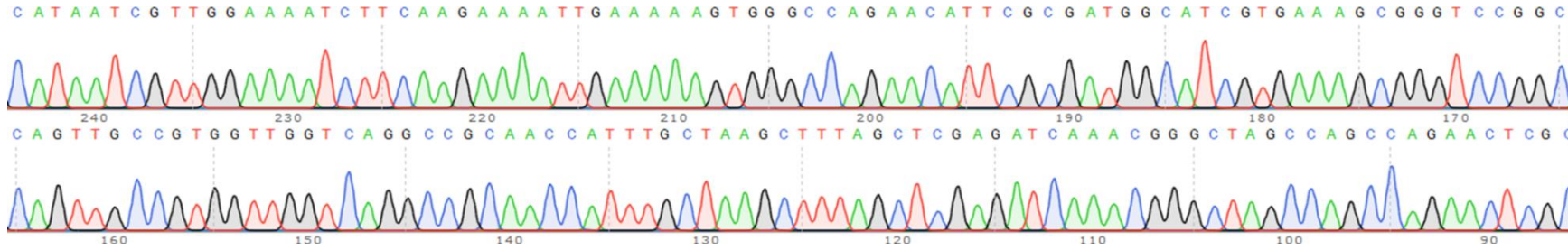
Annex N

pA2∅::SELP59 sequencing chromatogram using the primer T7 Fwd

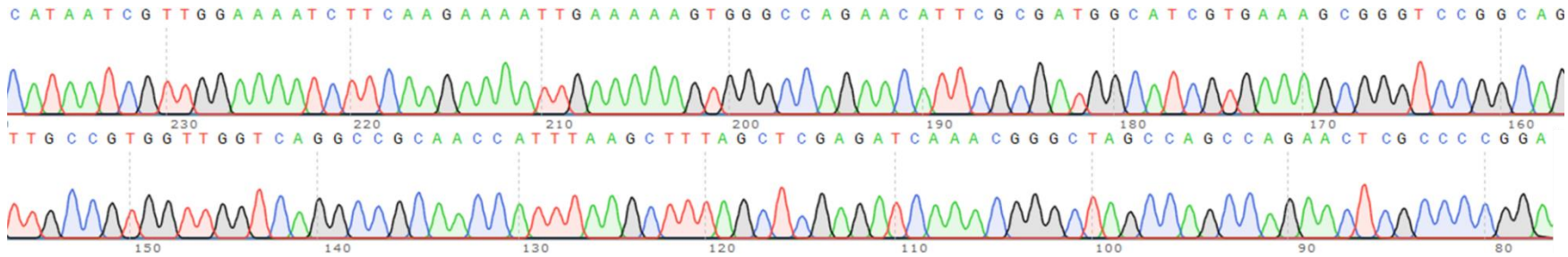


Annex O

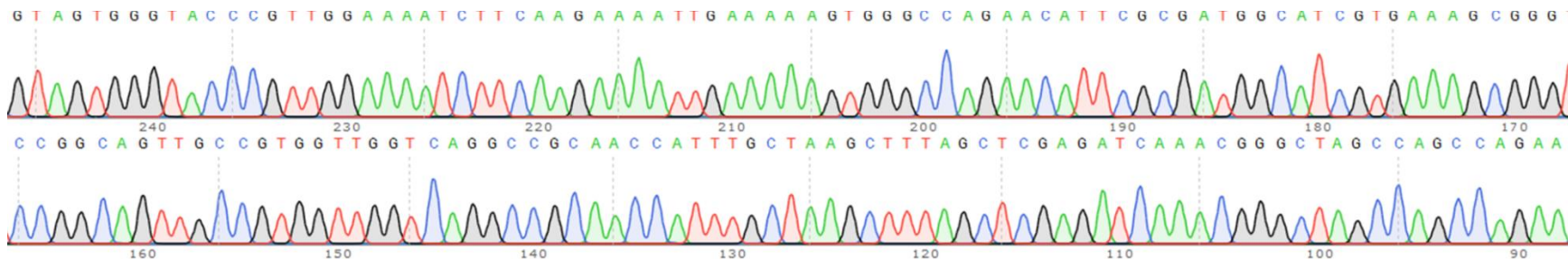
pA1C::A60:CM4<sub>cys</sub> sequencing chromatogram using the primer T7 terminal



pA1C::A60:CM4 sequencing chromatogram using the primer T7 terminal

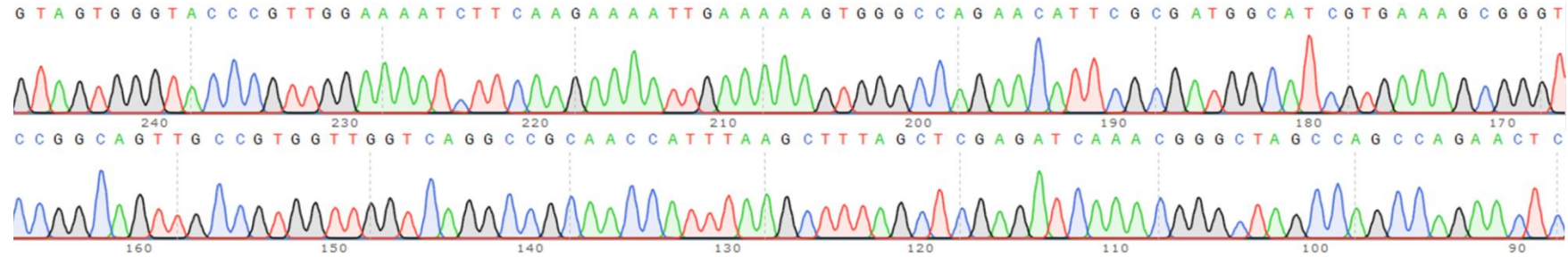


pA1C::A60:CM4<sub>cys</sub> (intein removed) sequencing chromatogram using the primer T7 terminal

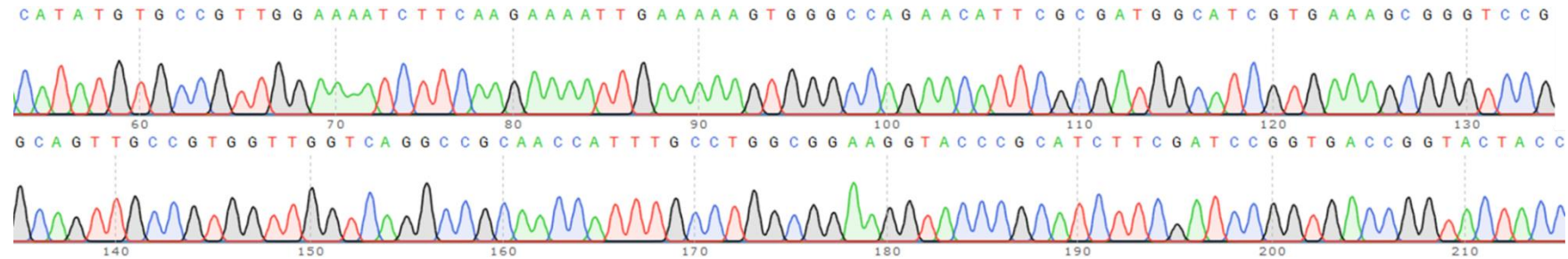


Annex P

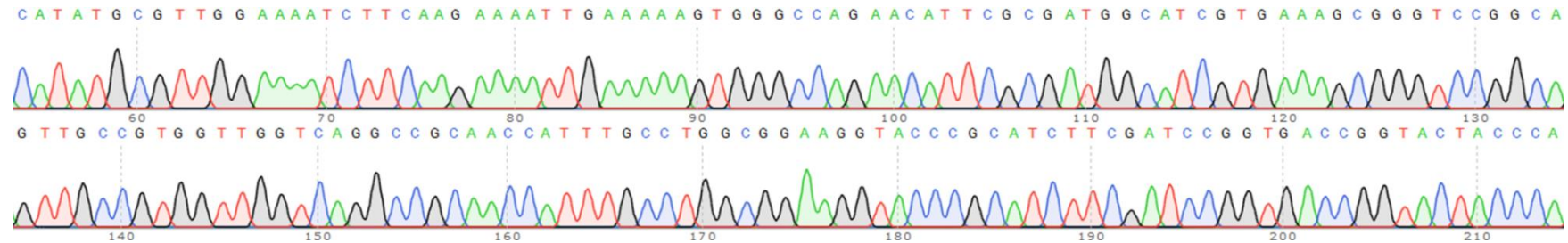
pA1C::A60:CM4 (intein removed)sequencing chromatogram using the primer T7 terminal



pA3N::A60:cysCM4 sequencing chromatogram using the primer T7 Fwd

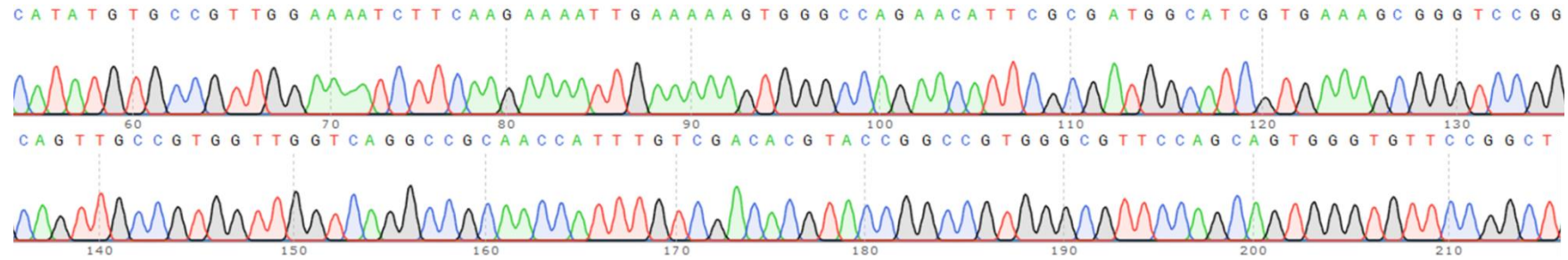


pA3N::A60:CM4 sequencing chromatogram using the primer T7 Fwd

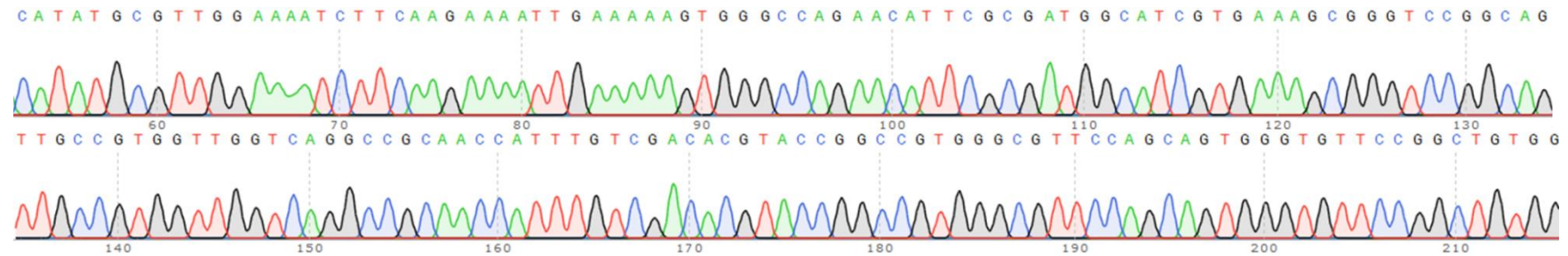


Annex Q

pA3N::A60:cysCM4 (intein removed) sequencing chromatogram using the primer T7 Fwd

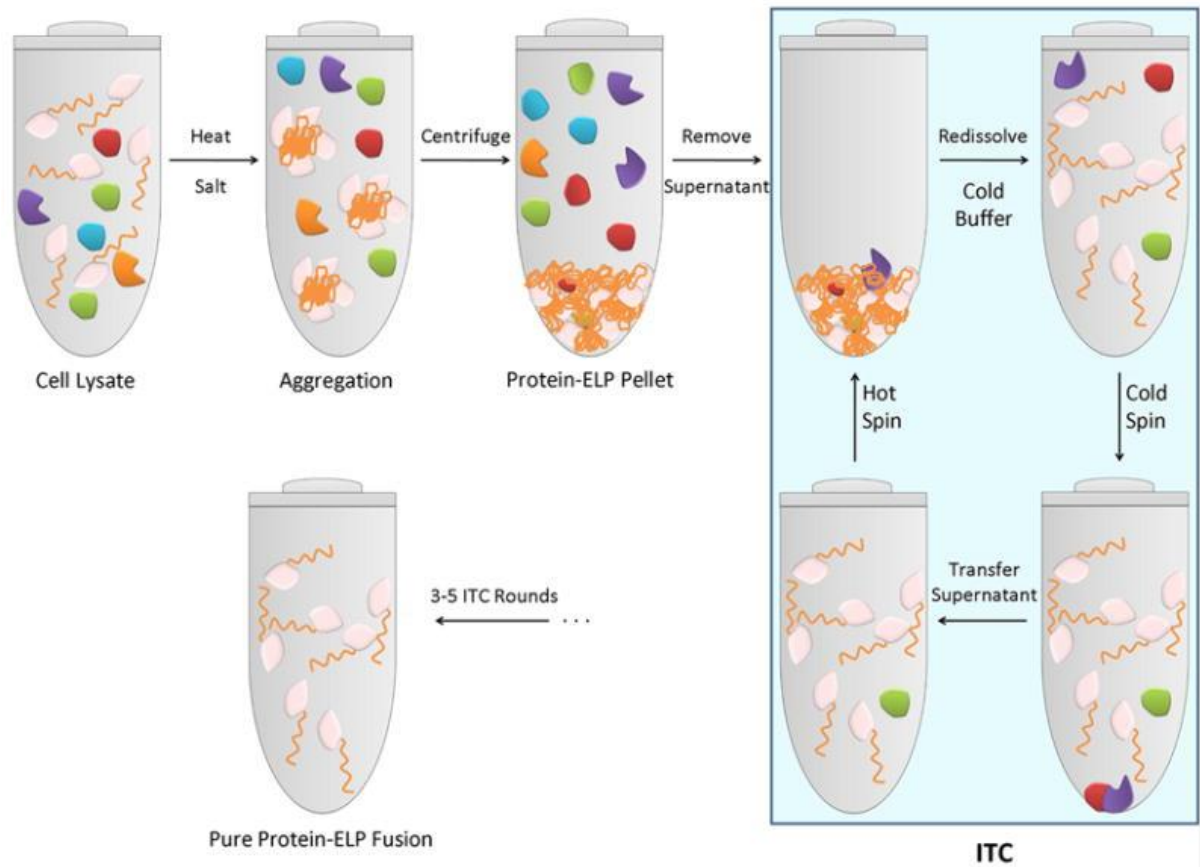


pA3N::A60:CM4 (intein removed) sequencing chromatogram using the primer T7 Fwd



Annex R

Schematic representation of the purification of recombinant ELP fusion proteins (reproduced from Hassouneh, Christensen & Chilkoti, 2010)





Appendix A

Primer sequences to amplify HAMP in order to insert in the pA1C, pA3N, pA2H and pA2Ø vectors.

Primer	Nucleotide sequence	Length
FWD_A1_Hep_W/_intein_Blunt	GCGTGGTGGTGCATAAATACTAGCGATA CTCATTTTC	35 bp
FWD_A1_Hep_W/out_intein_KpnI	AGAGGGTACCACTAGCGATACTCATTTC C	28 bp
reverse_A1_Hep_HindIII	CGCGAAGCTTAAGTTTTGCAACACATA CC	29 bp
reverse_A1_Hep_HindIII_w/_Cys_terminal	CGCGAAGCTTAGCAAGTTTTGCAACAC ATACC	32 bp
FWD_A2_Hep_W/_Int-c_NheI	AGCGGCTAGCAATTGCACTAGCGATAC TCATTTTC	34 bp
FWD_A2_Hep_W/out_int-c_NdeI	CGCGCATATGACTAGCGATACTCATTTC C	28 bp
reverse_A2_Hep_SalI	AGCGGTCGACAAGTTTTGCAACACATA CC	29 bp
FWD_A3_Hep_W/_int_NdeI_Cys_terminal	CGCGCATATGTGCACTAGCGATACTCA TTTC	31 bp
reverse_A3_Hep_W/_intein_KpnI	AAAAGGTACCTTCCGCCAGGCAAGTTT TGCAACACATACC	40 bp

Primer sequences to amplify HAMP to insert in the adapter described in 3.2.5.

Primer	Nucleotide Sequence	Length
FWD_A2_Hep_W/out_int-c_NdeI	CGCGCATATGACTAGCGATACTCATTTC	28 bp
Reverse_A4a_hep	CCACAGCAAAAAATACAT	18 bp
FWD_A4b_hep	CGCGACTAGTGCATAATTGTTGCCATCGTAGTAAA	35 bp
reverse_A1_Hep_HindIII	CGCGAAGCTTAAGTTTTGCAACACATACC	29 bp

Appendix B

Primer sequences to amplify BMAP-28 peptide to insert it into the pA1C, pA3N, pA2H and pA2Ø vectors.

Primer	Nucleotide Sequence	Length
FWD_A1_28_W/_intein_Blunt	GCGTGGTGGTGCATAATGGCGGTCTGCG TAGC	32 bp
FWD_A1_28_W/out_intein_KpnI	AGAAGGTACCGGCGGTCTGCGTAGC	25 bp
reverse_A1_28_HindIII	CGCGAAGCTTAACCGATACGGATAATCG GC	30 bp
reverse_A1_28_HindIII_w/_Cys	CGCGAAGCTTAGCAACCGATACGGATAAT CGGC	33 bp
FWD_A2_28_W/_Int-c_NheI	AAGAGCTAGCAATTGCGGCGGTCTGCGT AGC	31 bp
FWD_A2_28_W/out_Int-c_NdeI	AAAGCATATGGGCGGTCTGCG	21 bp
reverse_A2_28_Sall	CGCGGTGACAAACCGATACGGATAATCG GC	30 bp
FWD_A3 _28_W/_int_NdeI_Cys_terminal	AAACCATATGTGCGGCGGTCTGCGTAGC	28 bp
reverse_A3_28_W/_intein_KpnI	ACACGGTACCTTCCGCCAGGCAACCGAT ACGGATAATCGGC	41 bp

Primer sequences to amplify BMAP-28 to insert in the adapter described in 3.2.5.

Primer	Nucleotide Sequence	Length
FWD_A2_28_W/out_Int- c_NdeI	AAAGCATATGGGCGGTCTGCG	21 bp
Reverse_A4a_Bmap28	CGCAGACCGCCCATATG	17 bp
FWD_A4B_Bmap28	CGCGACTAGTGCATAATAGCCTGGGCCGAAAATT	35 bp
reverse_A1_28_HindIII	CGCGAAGCTTAACCGATACGGATAATCGGC	30 bp

## Appendix C

### **RbCl competent cells protocol:**

- Grow a 10 ml inoculum of *E. coli* cells (desired strain) in LB liquid medium overnight at 37 °C and 200 rpm.
- From that inoculum, take 400 µl and use it to inoculate 10 ml of LB liquid medium and incubate it at 37 °C and 200 rpm.
- When the OD<sub>550</sub> of the culture reaches 0.3, use 4 ml of the inoculum to inoculate 100 ml of LB liquid medium and incubate it at 37 °C and 200 rpm.
- When the OD<sub>550</sub> reaches 0.3, incubate the culture for 5 min in ice and centrifuge it 5 min at 750 xg at 4 °C.
- Gently resuspend the pellet in 20 ml of TFBII solution.
- Centrifuge the resuspended pellet 5 min at 750 xg at 4 °C.
- Resuspend the pellet gently in 2.5 ml of TFBII and incubate it at 4 °C for 5 min.
- Divide the cells into 100 µl aliquots on sterile microtubes before being flash frozen in liquid nitrogen and preserved at -80 °C.

#### **TFBI solution**

- Weigh components to obtain a final concentration of 30 mM CH<sub>3</sub>COOK, 50 mM MnCl<sub>2</sub>, 10 mM CaCl<sub>2</sub>, 100 mM RbCl and 1.67 M of glycerol.
- Dissolve it in ddH<sub>2</sub>O and adjust the pH to 5.8 using 0.2 M CH<sub>3</sub>COOH solution.
- Filter the solution using a 0.22 µm filter and stored at 4 °C.

#### **TFBII solution**

- Weigh components to obtain a final concentration of 10 mM MOPS sodium salt, 75 mM CaCl<sub>2</sub>, 10 mM RbCl and 1.67 M of glycerol.
- Dissolve it in ddH<sub>2</sub>O and adjust the pH to 6.5 using 0.5 M NaOH solution.
- Filter the solution using a 0.22 µm filter and stored at 4 °C.

## Appendix D

### CaCl<sub>2</sub> competent cells protocol

- Grow a 10 ml inoculum of *E. coli* cells in LB liquid medium overnight at 37 °C and 200 rpm.
- From that inoculum, take 1 ml and use it to inoculate 100 ml of LB liquid medium and incubate it at 37 °C and 200 rpm.
- When the OD<sub>600</sub> of the culture reaches 0.6, incubate the culture for 10 min in ice and centrifuge it 10 min at 750 xg at 4 °C.
- Resuspend gently the pellet in 8 ml of TB Buffer and add DMSO to a final concentration of 7 %, as cryopreservative.
- Incubate the suspension for 10 min in ice.
- Divide the cells into 200 µl aliquots on sterile microtubes before being flash frozen in liquid nitrogen and preserved at -80 °C.

### TB Buffer

- Weigh the components for a final concentration of 10 mM HEPES, 15 mM CaCl<sub>2</sub> and 250 mM KCl.
- Dissolve in ddH<sub>2</sub>O and adjust the pH to 6.7 using KOH.
- Add MnCl<sub>2</sub> to a final concentration of 55mM.
- Filter the solution using a 0.22 µm filter and stored at 4 °C.

## Appendix E

### Recipe for 10 % acrylamide gel to be used in SDS-PAGE

Reagents	Running gel	Stacking gel
Acrylamide/Bis-acrylamide, 30% solution	1.565 ml	0.415 ml
1.875 M Tris, pH 8.8	1.875 ml	-
0.5 M Tris, pH 6.8	-	0.63 ml
ddH <sub>2</sub> O	1.105 ml	1.355 ml
10 % SDS	47 $\mu$ l	25 $\mu$ l
10 % APS	37.5 $\mu$ l	12.5 $\mu$ l
TEMED	3.5 $\mu$ l	2.5 $\mu$ l

### Loading buffer (SDS-PAGE, 5x concentrated)

10 % SDS

40 % Glycerol

0.2 M Tris-HCl

50 mM  $\beta$ -mercaptoethanol

0.1 % Bromophenol Blue

### Coomassie Blue staining solution

0.20 % Coomassie Blue G-250

10 % Acetic Acid

40 % Ethanol

## Appendix F

### PBS recipe:

8 g/L NaCl

0.2 g/L KCl

1.44 g/L Na<sub>2</sub>PO<sub>4</sub>

0.24 g/L KH<sub>2</sub>PO<sub>4</sub>

### Protocol for AgNO<sub>3</sub> gel staining

- 1- Stir 20 min. in solution 1 (50% methanol and 5% glacial acetic acid).
- 2- Stir 10 min. in solution 2 de (50% methanol).
- 3- Stir 10 min. in upH<sub>2</sub>O.
- 4- Stir 1 min. in solution 3 (0.02% sodium thiosulphate).
- 5- Wash gel 2 times with upH<sub>2</sub>O, 1 min. each wash.
- 6- Stir 20 min. in solution 4 (0.1% silver nitrate) at 4°C.
- 7- Put a little of solution 5 (3% sodium carbonate with 50 µl formaldehyde) and dispose of the liquid.  
Drop the rest of the solution and shake it until the bands are properly revealed.
- 8- To stop the revelation use solution 6 (5% glacial acetic acid).
- 9- Leave the gel in ddH<sub>2</sub>O.

## Appendix G

### Western Blot

-Put the nitrocellulose membrane in contact with the polyacrylamide gel in a transfer cassette in order sponge-filter-gel-membrane-filter sponge. Ensure all pieces are fully in contact with no bubbles. In the transfer system, fill it with icepack and running buffer (30g/L Tris, 144 g/L Glycine and 10/g SDS) with 10% methanol.

-The transference lasts 1.5 h at 4 °C at 54 mA per membrane.

-After the transference, wash the membrane with TBST (50 mM Tris-HCl, pH 7.5, 150 mM NaCl and 0.1% Tween 20). Stir the membrane using Ponceau solution for 10 min to reveal the bands and verify if the transference was successful. Wash with ddH<sub>2</sub>O to remove the coloration.

-Block the membrane with TBST with 5% milk protein for 1 h in agitation. Wash with TBST after blocking.

-Incubate at 4 °C overnight with the primary antibody in TBST with 5% milk protein (1:1000).

-Wash with TBST two times for 5 min. Wash the membrane with TBST for 15 min.

-Incubate at 4 °C 1 h with the secondary antibody in TBST with 5% milk protein (1:2000).

-Wash membrane five times for 5 min with TBST.

-Reveal the membrane with horseradish peroxidase and alkaline phosphatase. Test different revelation times using appropriate equipment.