Universidad de Cantabria (UC)

Facultad de Medicina

Departamento de Biología Molecular

Instituto de Biomedicina y Biotecnología de Cantabria (IBBTEC)

Departamento de Microbiología y Genómica

Desarrollo de un sistema de expresión de proteínas heterólogas basado en *Streptomyces rimosus*

Tesis Doctoral presentada por

Andrés Felipe Carrillo Rincón

Para optar por el grado de Doctor por la Universidad de Cantabria

Santander, 2016

University of Cantabria

School of Medicine

Molecular Biology department

Institute of Biomedicine & Biotechnology of Cantabria (IBBTEC)

Department of Microbiology and Genomics

Development of *Streptomyces rimosus* expression system for production of heterologous proteins

DOCTORAL DISERTATION

Andrés Felipe Carrillo Rincón

Santander, 2016

Don Gabriel Moncalian Montes, Profesor titular de Genética, perteneciente al departamento de Biología Molecular de la Universidad de Cantabria, y **Don Hrvoje Petković**, profesor titular de Microbiología, perteneciente al departamento de Biotecnología de la Universidad de Liubliana.

CERTIFICAN: Que **Don Andrés Felipe Carrillo Rincón** ha realizado bajo su dirección el trabajo que lleva por título "Desarrollo de un sistema de expresión de proteínas heterólogas basado *Streptomyces rimosus*".

Consideramos que dicho trabajo se encuentra terminado y reúne los requisitos necesarios para su presentación como Memoria de Doctorado, al objeto de poder optar al grado de Doctor.

Fdo. **Gabriel Moncalian Montes** Fdo. **Hrvoje Petković**

El presente trabajo se ha realizado entre el departamento de Biología Molecular de la Universidad de Cantabria, el departamento de Microbiología y Genómica del Instituto de Biomedicina y Biotecnología de Cantabria, la empresa Acies Bio Ltd. (Eslovenia) y el departamento de Biotecnología de la Universidad de Liubliana, bajo la dirección del Dr. Hrvoje Petković y el Dr. Gabriel Moncalian, gracias a un contrato de la Universidad de Cantabria asociado al proyecto "Augusto González Linares" de la Universidad de Cantabria.

Durante este periodo se han realizado tres estancias, la primera por un periodo de 18 meses en la compañía Acies Bio Ltd. localizada en Liubliana, Eslovenia. La segunda por un periodo de 6 meses en el laboratorio del Dr. Petković localizado en el departamento de Biotecnología de la Universidad de Liubliana, Eslovenia; y la tercera por un periodo de 12 meses en el laboratorio del Dr. Müller ubicado en Helmholtz Institute for pharmaceutical research, Saarland, Alemania.

Supervisor: **Prof. Dr. Hrvoje Petković**

Co-advisor: **Dr. Gabriel Moncalian**

Acknowledgments

I would like to thank:

Dr. Hrvoje Petković, his view of science inspired me and I appreciate the time and effort he invested in my education. Petko shared with me all his wisdom; sometimes I did not hear him. Petko trained me as a scientist and as a businessman as well, but most important, he gave me life lessons which made me a better person.

Dr. Moncalian for his support over the last four years.

University of Cantabria, despite all the drawbacks with my nationality, Unican administration always found a solution to get my documents in order.

The company Acies Bio, they gave me the specific training and resources which allowed me to work in the Streptomyces field.

Dr. Rolf Müller trusted in my ideas and allowed me to use his research facilities.

Dr. César Poza, I arrived to Spain with a lot of theory in my head but my lab skills were clearly absent. With patience and dedication Dr. Poza built my technical skills and now I can proudly say that I am really good in the lab.

Dr. Vasilka Magdevska, she transferred me all her knowledge of S. rimosus. Even though Streptomyces are not the easiest microorganism to work with, Vasilka made them look a piece of cake.

Thank you very much,

Muchas gracias,

Hvala lepa,

Danke

Contents

Chapter 1. Expression of the acyltransferase of module 4 of tacrolimus

Chapter 2. Development of protein expression system based on *Streptomyces rimosus* as host

Abbreviations and symbols

ACP	Acyl carrier protein
Amp	Ampicillin
amp^r	Ampicillin resistance gene
AT	Acyltransferase
bp	Base pair
dH_2O	Distillated water
DNA	Deoxyribonucleic acid
dNTP	Nucleotides
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
FDA	Food and Drug Association
GRAS	Generally Regarded As Safe
H (His)	Histidine
HCl	Hydrogen chloride
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kb	Kilobase
kDa	Kilodalton
KS	Ketosynthase
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
Mb	Megabase
MeCN	Acetonitrile
NRP	Nonribosomal peptides
NRPS	Nonribosomal peptides synthases
NaCl	Sodium chloride
OD	Optical density
OTC	Oxytetracycline
PCR	Polymerase Chain Reaction
PEG	Polyethylene glicol
*PermE**	Erythromycin promoter
PKS	Polyketide synthases
PnitA-PnitR	Nitirilase promoter
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
TCA	Trichloroacetic acid
Tcp830	Tetracycline promoter adapted for *Streptomyces*
TE	Buffer from Tris-EDTA
TES	Buffer from Tris, EDTA and sucrose
tetRiS	tetR adapted for expression in *Streptomyces*
Tris	2-Amino-2-hydroxymethyl-propane-1,3-diol
Ts	Thiostrepton
ts^r	Thiostrepton resistance gene
U	Unit

Chapter 1

EXPRESSION OF THE ACYLTRANSFERASE OF MODULE 4 OF TACROLIMUS

1 Introduction

1.1 Type I Polyketides

Polyketides are a large group of secondary metabolites, mainly produced by the soil-borne and marine Gram-positive actinomycetes, which display astonishing structural diversity [1]. This structural variety is reflected in a large spectrum of pharmacological activities such as antibacterial (erythromycin, oxytetracycline), antifungal (nystatin, amphotericin), anti-cancer (epothilone, rapamycin), anti-helminthic (avermectin), immunosuppressive (tacrolimus) and other biological activities [2]. These natural products are synthesized by a diverse group of biosynthetic enzymes termed polyketide synthases (PKS), which are classified into three groups according to the biosynthetic mechanisms. Type I PKS are organized in modules where each module typically contains at least three domains, beta-ketoacyl synthase (KS), acyltransferase (AT) and acyl carrier protein (ACP). Besides typical type I PKS, there are two types of atypical type I PKS: the AT-less type I PKS, which lack AT domains in each module and instead contain at least one discrete acyltransferase acting in trans, [3] and the iterative type I PKS, whose domains act iteratively [4].

Each module of the typical type I PKS catalyzes one step of the polyketide chain extension [5]. Thus, the AT domain determines the choice of the extender unit to be incorporated and transfers it to the corresponding phosphopantetheine arm of the ACP, while the KS domain catalyzes the decarboxylative condensation between the previous acyl-KS intermediate and the selected extender unit to form the β-ketoacyl-ACP intermediate, therefore continuing the elongation of the polyketide chain [5]. In addition of the three main domains, each module may contain a combination of additional reductive domains including the ketoreductase (KR), enoylreductase (ER) and dehydratase (DH) domains, which catalyze the reduction of the beta-keto unit, thus significantly expanding chemical diversity of these compounds [2]. The module that initiates the biosynthesis of the polyketide chain selecting the “starter unit” is termed the “loading module” and subsequent elongation steps are carried out by “extender modules” [5]. Many modules involved in the final elongation step also contain a thioesterase (TE) domain, which releases the nascent polyketide chain from the PKS enzyme, generating the cyclic macrolide/macrolactone structure. Erythromycin, rapamycin and avermectin are examples of type I PKS [5].

Great structural diversity of polyketides arises from the different choices that these enzymatic systems can make during carbon chain assembly. For example, starter units can be selected from acetate, propionate, butyrate and other unusual CoA-activated carboxylic acids [6]. Most often, AT domains of extender modules recognize and incorporate the “standard” extender units, malonyl-CoA, methylmalonyl-CoA and ethylmalonyl-CoA. However, in some polyketides, other unusual extender units are also incorporated [6]. Further diversity is caused by different degrees of reduction of the beta-keto unit after each elongation step as well as subsequent “post-PKS” modifications of the polyketide structures [2].

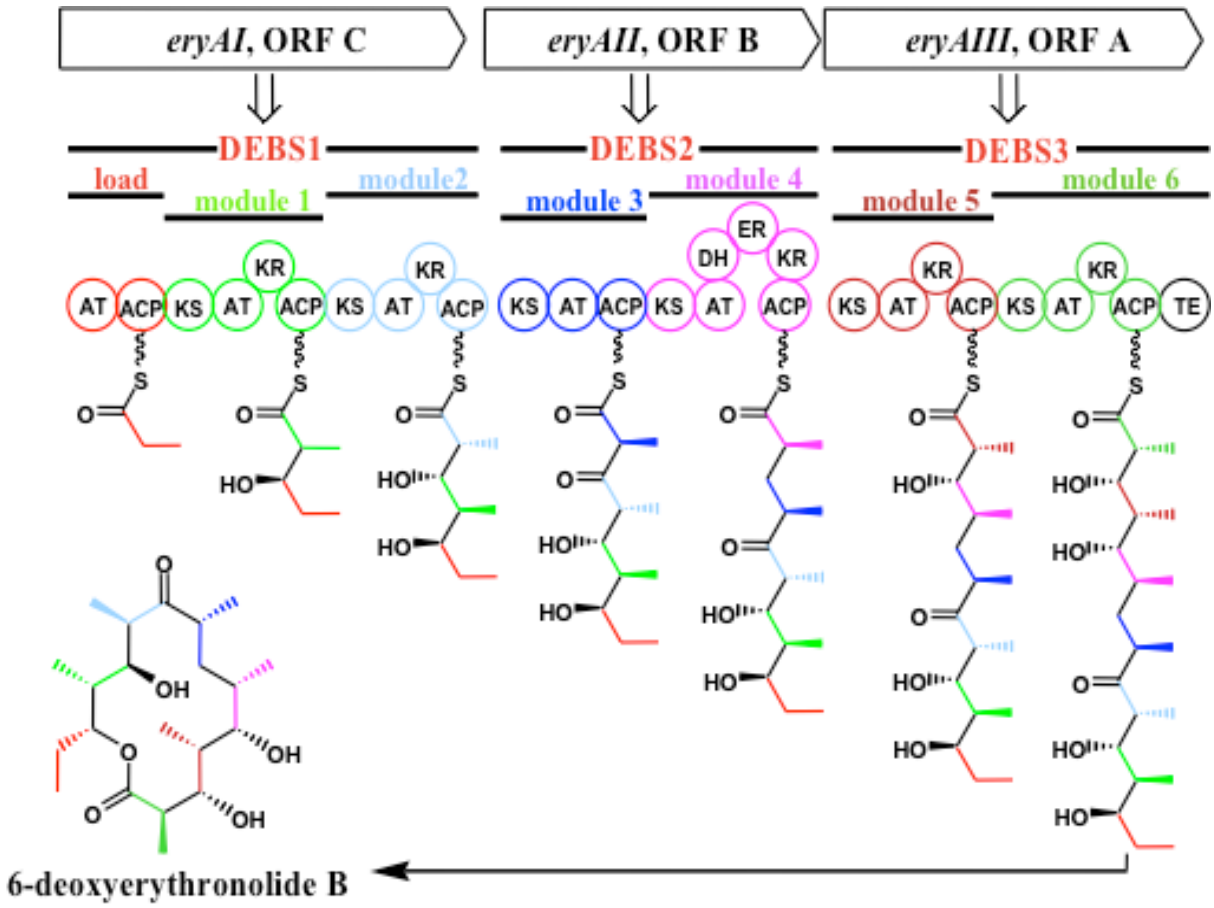


Figure 1. Schematic presentation of 6-deoxyerythromycin synthase (6-DEBS). The model system to study type I PKSs.

1.2 Specificity of acyltransferases in PKS type I

The specificity of AT domains towards the extender units has been studied intensively, and certain conserved motifs clearly influence substrate specificity; however, this information is limited to the most common extender units. Common amino acid sequence motifs were sought in sequence alignments of different regions of either malonyl-CoA or methylmalonyl-CoA specific AT domains. Initially, a motif located about 25-30 residues upstream of the active site serine of AT domains was identified to differentiate between malonyl-CoA and methylmalonyl-CoA selection [7]. With increasing number of sequences available, an additional region towards the C-terminus of AT domains, sometimes termed the “hypervariable region” was suggested to determine their specificity [8]. However, this region was later found not to be a key determinant of domain specificity. One key motif, which is usually located 90–100 amino acid residues downstream from the conserved active site serine (GHS-motif) was then found to particularly influence the specificity of AT domains [9]. In general, AT domains recognizing malonyl-CoA extender unit possess a HAFH sequence motif in this location, while AT domains specific for methylmalonyl-CoA contain the YASH motif, and AT domains responsible for incorporation of ethylmalonyl-CoA the sequence XAGH can be identified (where X is often F, T, V or H) [9].

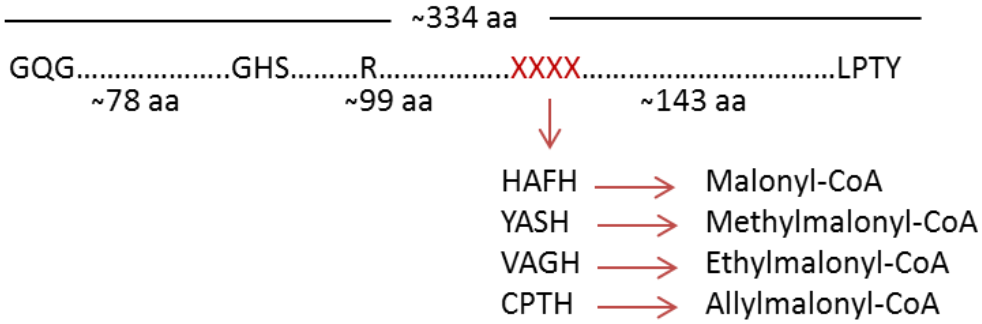


Figure 2. The amino acid sequence of typical AT domains starting with GQG amino acid sequence and finishing with LPTY amino acid sequence. Active site GHS and catalytically important arginine residue (R) are marked on the sequence. The location of the motif believed to be responsible of selecting the extender unit is labelled in red.

While these sequence motifs are still believed to be the most influential determinant of AT domain selectivity, subsequent characterization and/or detailed studies of numerous AT domains in native or engineered context revealed additional factors affecting selectivity and catalytic efficiency of AT domains. On one hand, AT domains with non-consensus signature motifs were identified, often selecting for a mixture of substrates [9].

In recent years, a limited number of rare AT domains have been identified, which have the capacity to incorporate non-standard, structurally complex extender units into the polyketide chains [10]. In addition, based on the mixtures of analogous polyketides produced, some AT domains show a certain degree of relaxed specificity (promiscuity) in substrate selection, while still excluding the "standard" and most abundant extender units, such as malonyl-CoA and methylmalonyl-CoA [11]. This feature makes them promising candidates for biosynthetic engineering. Interesting examples of such acyltransferase domains are the AT1 domain of the unusual *salA* PKS gene, involved in the biosynthesis of salinosporamide, which incorporates up to 5 different extender units including the unusual chloroethylmalonyl-CoA [12], the AT domain of the 13th module of the sanglifehrin PKS which incorporates 2-(2-oxo-butyl)malonyl-CoA [13], and the AT domain of the 4th module, designated as designed as fkAT4 domain, involved in the biosynthesis of the important immunosuppressant FK506, also known as tacrolimus [10]. The fkAT4 domain incorporates unusual allylmalonyl-CoA extender units while also showing certain degree of promiscuity towards other natural and unnatural extender units [11].

1.3 Acyltransferase of module 4 (fkAT4) of tacrolimus (FK506)

Tacrolimus is a medically important immunosuppressant drug, which belongs to the class of combined PKS/NRPS-derived metabolites of great medical importance produced by *Streptomyces tsukubaensis* [14]. Interestingly, the AT domain of the fourth module of FK506 PKS incorporates an unusual allylmalonyl-CoA extender unit, resulting in a side chain with a chemically reactive terminal double bond (Fig. 3). The incorporation of allylmalonyl-CoA results in the presence of the allyl group at the

carbon 21 of the FK506 structure, a feature which distinguishes FK506 from the structurally and biosynthetically related compound FK520, which possesses an ethyl group at equivalent position (Fig 3) [15]. Furthermore, FK520 is also co-produced by FK506-producing strains if ethylmalonyl-CoA is incorporated into the polyketide chain instead of allylmalonyl-CoA by the fkAT4 domain [11]. The fkAT4 domain of FK506 PKS displays relaxed specificity and can accept, not only naturally occurring allylmalonyl-CoA, ethylmalonyl-CoA and propylmalonyl-CoA, but also unnatural 2-methylallylmalonyl-CoA and fluoroethylmalonyl-CoA extender units, while "standard" extenders malonyl-CoA and methylmalonyl-CoA are not selected by the fkAT4 domain [11]. Interestingly, the recognition motif in the fkAT4 domain is a unique an unusual CPTH sequence, differing substantially from the usual HAFH and YASH motifs (Fig 2) [9]. Incorporation of unusual and unnatural extender units by the fkAT4 domain, but at the same time strong selectivity against the standard extender units, is very unique property of fkAT4 domain, thus presenting a very promising opportunity to study the selectivity of unusual extender units.

1.4 Acyltransferases of type I PKS studied to date

Acyltransferase domains are the gatekeepers in the polyketide assembly line, its function is to recognize particular extender units, transferring the acyl units from acyl-CoA or acyl-S-ACP donors into holo–ACP via a ping pong mechanism [16]. The reaction is divided in two parts, self-acylation and trans-acylation. First, in the self-acylation reaction, the AT selects the extender unit from acyl-CoA or acyl-ACP donors to form the acyl-AT intermediates, thus releasing CoA and/or ACP respectively. Second, in the trans-acylation reaction, the acyl-AT intermediates transfer the acyl unit into the holo-ACPs producing the acyl-S-ACP [5]. Together, the AT, KS and ACP domains comprise the minimal modular components required for polyketide chain elongation, the AT selects the building block to be incorporated and transfers it to the ACP domain present in the same module, and then the KS domain catalyzes chain elongation between the growing polyketide chain and the ACP-bound extender unit.

To date, the structural information of five AT domains related to type I PKS have been published, these AT domains include: the AT3 and AT5 domains of 6-deoxyerythronolide B synthase (DEBS), the type I PKS involved in the biosynthesis of erythromycin (PDB IDs: 2QO3 & 2HG4) [17, 18]. The AT10 domain of enedyne PKS, an iterative type I PKS involved in the biosynthesis of dynemicin [19]. The trans-AT from the "AT-less" disorazole synthase (PDB ID: 3RGI) [20], and the AT2 domain (ZMA-AT) of the hybrid polyketide-nonribosomal system involved in the biosynthesis of Zwittermicin A (PDB ID: 4QBU) [21]. These AT domains characterized, with the exception of ZMA-AT, are selecting "standard" units; therefore structural information related to AT domains selecting "rare" extender units linked to polyketide biosynthesis could bring a better understanding of AT selection.

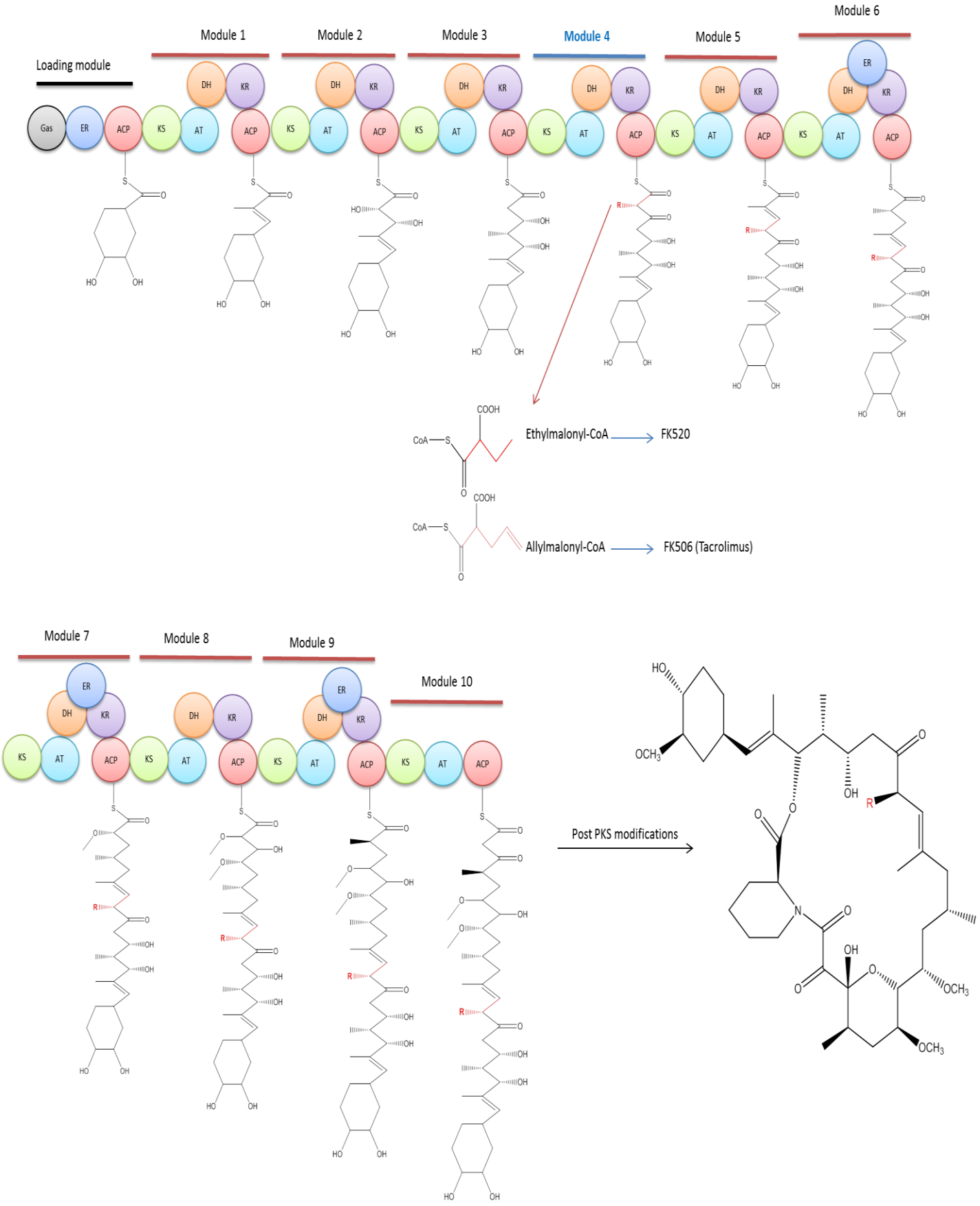


Figure 3. Schematic presentation of FK506 polyketide synthase and the range of extender units that can be selected by the AT4 domain.

Studies on the AT4 and AT10 domains of FK506 PKS have been published recently [22, 23], however not at structural level and only confirming the selectivity of the fkAT4 and fkAT10 domains for allylmalonyl or ethylmalonyl and malonyl respectively [22, 23]. The biochemical characterization of fkAT4 and ZMA-AT revealed that the unusual extender units allylmalonyl and hydroxymalonyl selected by fkAT4 and ZMA-AT respectively, are linked to a donor ACP instead of CoA [21, 22]. These results suggest that during the first self-acylation step, depending on the AT, the ones selecting acyl-CoA units interact with CoA by a protein-small molecule interaction, whereas the ATs specific for extender units linked to ACP carriers, the recognition occurs via protein-protein interaction [21, 22]. However, the specific mechanism to select these rare extender units and deselect the common ones is still unknown.

2 Aim of this study

It is widely recognized that acyltransferase domains (ATs) of modular PKS are responsible for selection of extender units, which are used in the biosynthesis of polyketide backbones of many secondary metabolites. Almost exclusively, extender units used in polyketide biosynthesis are limited to malonyl-, methylmalonyl- and less often to ethylmalonyl-CoA activated extender units, thus significantly limiting structural diversity of polyketide compounds, compared to e.g. peptide antibiotics, produced by non-ribosomal peptide synthetases. The specificity of AT domains has been studied intensively, and certain conserved motifs clearly influence substrate specificity; however, this information is still limited to the most common extender units. Therefore, there is a need to extend the understanding of specificity of AT domains to more unusual extender units.

We believe that the unique substrate specificity of the acyltransferase domain of module 4 of FK506 PKS (fkAT4) is of particular interest. It is our intention to take advantage of this unique property of fkAT4, that makes fkAT4 domain capable of accepting diverse substrates, and most importantly, how fkAT4 deselects standard extenders such as malonyl-CoA and methylmalonyl-CoA.

3 Hypothesis

We hypothesize that the understanding of the fkAT4 domain will reveal key information on protein-substrate interaction. This knowledge will enable us to modify AT domains in order to influence selectivity towards target substrate, expanding the specificity of AT domains to more diverse extender units, this way introducing much greater structural variability, and chemically amenable moieties to type I polyketides.

4 Material and Methods

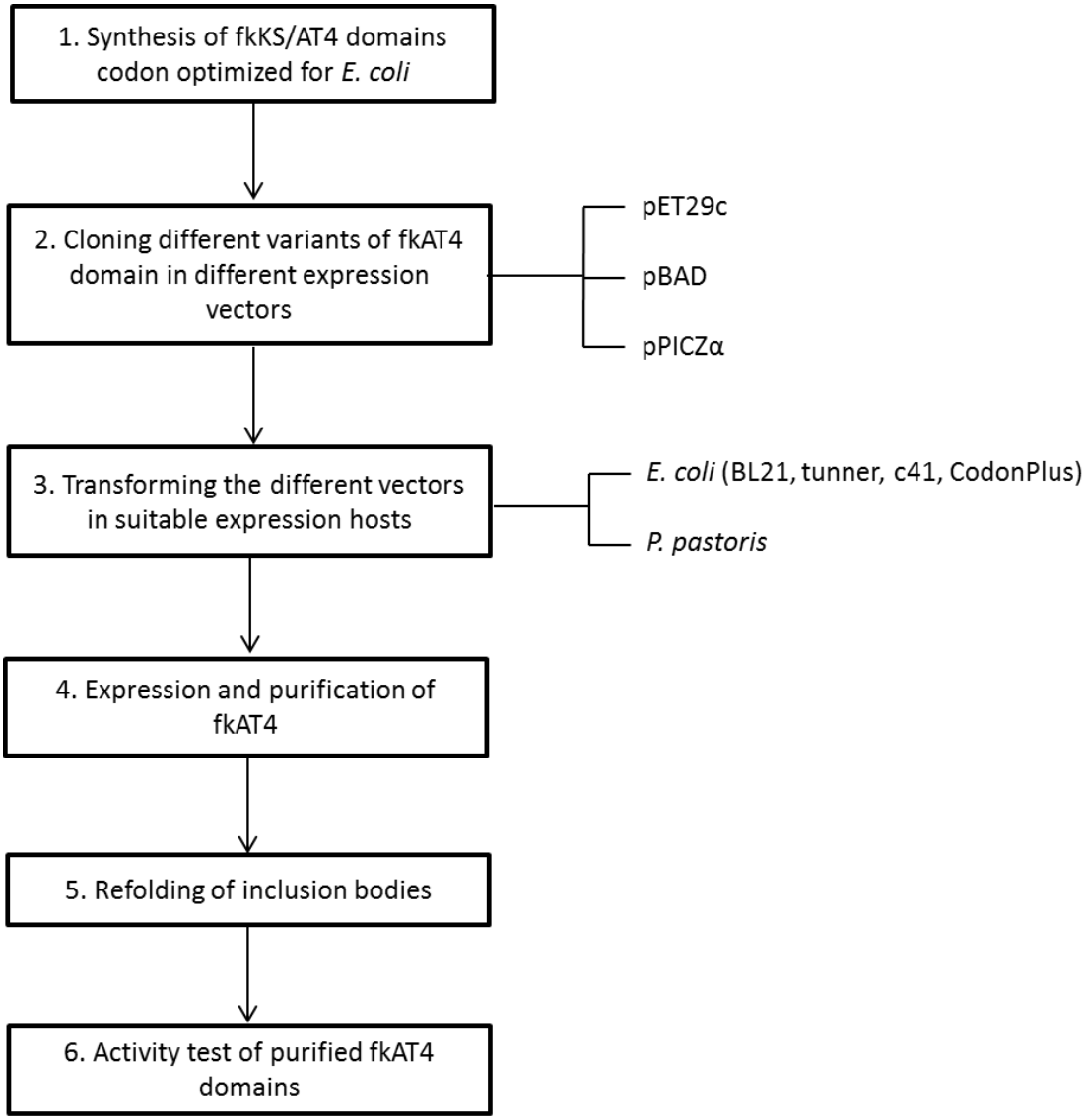


Figure 4. Experiment flow chart.

4.1 Cultivation media

Medium LB

1% tryptone, 0.5% yeast extract, 0.5% NaCl.
2% agar (*Optional: If making LB plates*)

1. Dissolve 10 g yeast extract, 5 g tryptone and 5 g NaCl in 900 ml water
2. Adjust pH to 7.5 and complete to 1 L with dH_2O
3. *Optional*: Add 20 g of agar if making LB plates.
4. Autoclave for 20 minutes.

Medium 2TY

1.6% tryptone, 1% yeast extract, 0.5% Sodium chloride.
2% agar (*Optional: If making 2TY plates*)

1. Dissolve 16 g yeast extract, 10 g tryptone and 5 g NaCl in 900 ml water
2. Adjust pH to 7.5 and complete to 1 L with dH_2O
3. *Optional*: Add 20 g of agar if making 2TY plates.
4. Autoclave for 20 minutes.

Medium YPD

1% yeast extract, 2% peptone.
2% agar (*Optional: If making YPD plates*)
2% glucose

1. Dissolve 10 g yeast extract and 20 g peptone in 900 ml water
2. *Optional*: Add 20 g of agar if making YPD plates.
3. Autoclave for 20 minutes.
4. Add 100 ml of 2% glucose (filter sterilize).

Medium YPDS

1% yeast extract, 2% peptone, 1 M sorbitol
2% agar
2% glucose

1. Dissolve 10 g yeast extract, 182.2 g sorbitol, 20 g peptone in 900 ml water
2. Add 20 g of agar to the solution and dissolve.
3. Autoclave for 20 minutes.
4. Add 100 ml of 2% glucose (filter sterilize).

Buffered Glycerol-complex Medium BMGY and Buffered Methanol-complex Medium BMMY

1% yeast extract, 2% peptone, 100 mM potassium phosphate pH 6.0, 1.34% Yeast Nitrogen Base (YNB), 4×10^{-5}% biotin and 1% glycerol or 0.5% methanol respectively.

1. Dissolve 10 g of yeast extract, 20 g peptone in 700 ml water.
2. Autoclave 20 minutes.
3. Cool to room temperature, then add the following and mix well:
    - 100 ml 1 M potassium phosphate buffer, pH 6.0
    - 100 ml 10X YNB (13.4% yeast nitrogen base with ammonium sulfate and without amino acids, filter sterilize)
    - 2 ml 500X B (0.2 mg/ml biotin, filter sterilize. Store at 4°C)
    - 100 ml 10X GY (10% Glycerol)
4. For BMMY, add 100 ml 10X M (5% Methanol) instead of glycerol.

5. Store media at 4°C.

4.2 Strains used during this PhD project

The strains used in this work are listed below.

4.2.1 *Escherichia coli* strains

DH10β: strain used for cloning and propagation of plasmids. Genotype: F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara, leu*) 7697 *gal*U *gal*K λ- *rps*L *nup*G.

BL21 (DE3): T7 Expression Strain. Genotype: F- *huA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*

C41 (DE3): BL21(DE3) derivative effective in expressing toxic and membrane proteins. Genotype: F- ompT gal dcm hsdS_B(r_B^- m_B^-)(DE3).

Tuner: BL21(DE3) derivative with lac permease mutation (*lacZY*), this mutation allows uniform entry of IPTG into all cells of the population enabling liner titration of IPTG to control the level of induction. Genotype: Genotype: F- *ompT hsdS_B(r_B^- m_B^-) gal dcm lacY1*(DE3).

BL21-CodonPlus(DE3): BL21(DE3) derivative with extra copies of the *argU* and *proL* genes (encoding tRNAs that recognize the arginine codons AGA and AGG and the proline codon CCC, respectively). Genotype: F- ompT hsdS (r_B^- m_B^-) *dcm+ Tet^r gal λ* (DE3) *endA Hte* [*argU proL Cam^r*].

4.2.2 *Pichia pastoris* strains

X-33: wild-type strain useful for selection on Zeocin.

4.3 Antibiotics and inducers

Table 1. Antibiotics and indicators used in this study, their final concentrations as selection markers in selected media and stock solution concentration.

Antibiotic or inducer	Selective concentration	Stock solution
Ampicillin	100 µg/ml	100 mg/ml in dH_2O
Kanamycin	50 µg/ml	100 mg/ml in dH_2O
Zeocin	25 µg/ml *E. coli* 100 µg/ml *P. pastoris*	100 mg/ml in dH_2O
Chloramphenicol	10 µg/ml	20 mg/ml in 100% ethanol
IPTG	From 0.1mM to 1 mM	100mM in dH_2O
Arabinose	From 0.001% - 1%	10% w/w in dH_2O

The stock solutions were filter sterilized with a 0.22μm filter and stored at -20°C. Stock solutions were added to molten agar, previously cooled to 55°C.

4.4 Cultivation conditions of *E. coli* and *P. pastoris* strains

4.4.1 Cultivation of *E. coli* strains

E. coli strains were inoculated from glycerol stocks on 2TY plates supplemented with the appropriate antibiotic, when needed, and cultivated overnight at 37°C. Liquid cultures were started from single colonies and incubated overnight at 37°C and 220 rpm. *E. coli* strains on 2TY plates were kept at 4°C for up to 1 month.

4.4.2 Cultivation of *P. pastoris* X-33 strains

P. pastoris X-33 strains were inoculated from glycerol stocks on YPD plates supplemented with zeocin, when needed, for 48 hours at 30°C. Liquid cultures were started from single colonies incubated overnight at 30°C and 220 rpm. X-33 strains on YPD plates were kept at 4°C for up to 1 month.

4.5 Preservation of *E. coli* and *P. pastoris* strains

E. coli and *P. pastoris* strains were maintained as glycerol stocks. A single colony of each strain was cultivated overnight in 5 mL 2TY media at 37°C for *E. coli*, and 5 mL YPD at 30°C for *P. pastoris*, supplemented with the respective selective marker, when needed. 1 mL was spin down at 14000 rpm and resuspended in 1 mL 20% (w/v) glycerol and frozen at -20°C or at -80°C. The strains were revived by scraping the surface of the frozen suspension with a sterile loop and either inoculating in liquid broth or streaking on agar plates supplemented with the proper antibiotic.

4.6 Transformation of *E. coli* and *P. pastoris*

4.6.1 Preparation of *E. coli* electrocompetent cells

2 liter flasks containing 400 mL of 2TY broth medium were inoculated with 4 mL overnight culture of the required *E. coli* strain. The culture was grown until OD_{600} 0.9 (approximately 2-3 hours at 37°C in shaker at 220 rpm). The cells were transferred to 4 sterile centrifuge bottles and place on ice for 20 minutes. After stopping the cell growth, the cells were centrifuged at 4000 rpm for 10 minutes at 4°C and the supernatant was discarded. The cell pellet was washed twice with 200 mL and 100 mL of 1mM Hepes pH 7.0, and once more with 100 mL 10% glycerol/1mM Hepes pH 7.0. Every wash was followed by centrifugation at 4000 rpm for 10 minutes at 4°C. Finally,

the cell pellet was resuspended in 3 mL 10% glycerol/1mM Hepes pH 7.0 and distributed in 40 µL aliquots which were frozen at -80°C. The electrocompetent cells retained their competence over several months without losing their transformation efficiency.

4.6.2 Transformation of *E. coli* with plasmid DNA

An aliquot of electrocompetent cells was thawed on ice and mixed gently with 2 µL plasmid DNA (1-100 ng). The mixture was transferred to a 0.1 cm electroporation cuvette which was subject to an electric pulse of 1.3 kV. The pulsed cells were resuspended in 950 µL 2TY medium and incubated at 37°C for 45 minutes before being spread on 2TY plates supplemented with the required selective marker.

4.6.3 Preparation of *P. pastoris* electrocompetent cells

2 liter flasks containing 500 mL of YPD broth were inoculated with 500 µL overnight culture of *P. pastoris*. The culture was cultivated overnight to an OD_{600} 1.5 at 30°C. The cells were centrifuged at 1500 g for 5 minutes at 4°C and resuspended with 500 mL of ice cold dH_2O. The cells were centrifuged again and the pellet was washed once again with 250 mL ice cold dH_2O. The suspension was centrifuged again and the pellet was resuspended in 20 mL ice cold 1 M sorbitol. After one extra round of centrifugation, the pellet was resuspended in 1 mL ice cold 1 M sorbitol and kept on ice to be transformed the same day, the electrocompetent cells were not stored.

4.6.4 Transformation of *P. pastoris* with plasmid DNA

80 µL of electrocompetent cells were gently mixed with plasmid DNA (50-100 µg) and transferred to 0.2 cm ice cold electroporation cuvette. The mixture was incubated on ice for 5 minutes and pulsed at 2.5 kV. The pulsed cells were resuspended in 1 mL ice cold 1 M sorbitol and transferred to 15 mL tube to be incubated at 30°C without shaking for 2 hours before being spread on YPDS plates containing zeocin. The plates were incubated for 3-10 days at 30°C and the colonies were streak on fresh YPD containing zeocin for further work.

4.7 DNA manipulation

4.7.1 Synthesis of KS/AT4 domains

Due the codon usage difference between *E. coli* (50% GC content) and *S. tsukubaensis* (71% GC content), it was necessary to synthesize the open reading frame encoding the KS/AT4 domains of FK506 codon optimized for efficient expression in *E. coli*, thus avoiding possible transcriptional issues. The DNA fragment was synthesized by Eurofins (Luxemburg).

4.7.2 Enzymes

Restriction and DNA modifying enzymes were purchased from Fermentas (United States). Enzymes were used according to the manufacturer's instructions.

4.7.3 Isolation of plasmid DNA from *E. coli*

Plasmid DNA was isolated according to the manufacturer's protocol using GeneJET Plasmid Miniprep Kit from Thermo Scientific (United States). Overnight cultures of *E. coli* were obtained from liquid 2TY medium with appropriate antibiotics.

4.7.4 DNA Gel Electrophoresis

DNA fragments were separated and visualized in 0.7-1% agarose gels using 1X TAE (50X TAE: 242 g Tris-base, 57,1 ml glacial acetic acid, 100 ml 0,5 M EDTA (pH 8.0), distillated water to 1 liter) as running buffer. SybrR Safe solution (1 µL each 100 mL agarose/1x TAE) was added on the gel before the electrophoresis run to stain the DNA. Samples were prepared by adding 6X loading buffer to the DNA solution. Electrophoresis was performed in Sub-Cell systems from Bio-Rad (Unites States) and DNA was detected under UV light (312 nm). The following DNA ladders were used to estimate the size of the DNA fragments: Fermentas Lambda DNA/*EcoRI*, Fermentas Lambda DNA/*EcoRI+HindIII* and Fermentas GeneRulerTM 1kb DNA Ladder.

4.7.5 DNA extractions from agarose Gel

DNA fragments were cut from gel with a scalpel under UV light. DNA fragments were recovered from TAE agarose gels using the Wizard Promega (United States) kit according to the manufacturer's protocol.

4.7.6 Polymerase chain reaction (PCR)

Phusion High-Fidelity DNA Polymerase was used for *in-vitro* DNA amplification according to manufacturer's instruction. Reaction mixture of 25 or 50 µL contained 1X GC buffer, 200 µM of each dNTP, 0.5 µM of forward and reverse primers, between 1 to 20 ng of DNA template, 0.02 U/mL of DNA polymerase and dH_2O to fix the mentioned concentrations. In addition, DMSO (1-5%) was added when required. Cycling conditions usually included 3 steps: Initial denaturation for 40 seconds at 98^{o}C followed by 35 cycles (denaturation for 10 sec at 98^{o}C, primer annealing at temperature +3^{o}C of the lower Tm primer for 10 seconds and extension of DNA fragment at 72^{o}C for 15 seconds/kb) and a final extension at 72^{o}C for 10 min. The samples were stored at 4^{o}C for further analysis once the amplification program was completed.

4.7.7 Primers

Table 2. Primers used in this study are listed in the following table.

.

Primer name	Sequence (5´ to 3´)	Amplicon
1S-AT4 Forward	AAAAACATATGCTGGCGATCGTAGGT	KS/AT4 for pET29c
1E-AT4 Reverse	CCGACGTATGCGTTCCTCGAGTTTTT	KS/AT4; AT4 for pET29c
2S-AT4 Forward	AAAAACATATGCCGCTGCCTGTGTCCGCG	LAT4 for pET29c
1E-AT4 Reverse	CCGACGTATGCGTTCCTCGAGTTTTT	LAT4; AT4 for pET29c
3S-AT4 Reverse	AAAAACATATGACCGTGTTCGTGTTCCCG	AT4; AT4- for pET29c
2E-AT4 Reverse	GCTAAAGCGTATACTGCCCTCGAGTTTTT	AT4; LAT4- for pET29c
Fw pBAD-AT4syn	AAAAACCATGGTGTTCGTGTTCCCGGGGCAG	AT4 for pBAD
Rv pBAD-AT4syn	CATGATCTGCCGACGTATGCGTTCCACCACCACCAC CACCACTGAAGCTTTTTTT	KS/AT4; AT4 for pBAD
Fw KS.AT4syn	AAAAACCATGGTAGGTATGGCATGTCGCCT	KS/AT4; AT4 for pBAD
Fw: P.p1	AAAAACTCGAGAAAAGACTGGCGATCGTAGGTATG GC	KS/AT4 for pPICZαA
Rw P.p2	ACGTATGCGTTCCATCATCATCATCATCATTGATCTA GATTTTT	KS/AT4; AT4 for pPICZαA
Fw P.p	AAAAACTCGAGAAAAGAACCGTGTTCGTGTTCCCG	AT4; AT4- for pPICZαA
Rv P.p4	AAAGCGTATACTGCCCATCATCATCATCATCATTGAT CTAGATTTTT	AT4- for pPICZαA
FW AT4 screening	CCCCGTGTGCTGGGC	Insert screening
RV AT4 screening	CCGCAACTGCGGCC	Insert screening

4.7.8 Digestion of DNA

Fast digest enzymes were used to cut both, plasmids and PCR products. Usually the reaction mixture contained up to 1 µg of DNA, 1x restriction buffer, 1 U of required restriction enzyme and completed to 10 µl dH_2O. For digestion of larger amounts of DNA, the volumes were scaled up accordingly. The reactions were incubated at 37°C in a heat block or water thermostat for up to 2 hours according to the DNA concentration. Reactions were stopped at –20^{o}C for further analysis.

4.7.9 Release of 5'- and 3'-phosphate groups from DNA

Dephosphorylation of cloning vectors to prevent recircularization during ligation was performed using FastAP Thermosensitive Alkaline Phosphatase. After plasmid digestion, 1 U of FastAP was added to the sample and incubated at 37^{o}C for 15 minutes to avoid false positives after ligation.

4.7.10 Ligation of DNA fragments

After digestion by restriction enzyme(s) and vector dephosphorylation, both fragments, vector and insert, were purified by gel electrophoresis and gel extraction prior to ligation. T4 DNA Ligase was used to join DNA fragments with either cohesive or blunt termini catalyzing the formation of a phosphodiester bond between juxtaposed 5'-phosphate and 3'-hydroxyl termini. Ligation of DNA fragments was carried out usually at a molar ratio 3:1 (insert: vector) in 10 µl volume using 5 Weiss U DNA T4 ligase and 1x T4 ligase buffer. Ligation mixtures were incubated at room temperature for up to 3 hours or at 4^{o}C overnight and subsequently transformed into *E. coli* DH10β.

4.7.11 DNA sequencing

DNA Sequencing was performed by Macrogen (South Korea) using Standard-seq system. Sequence analysis was performed with the software codon code aligner and BLAST (Basic Local Alignment Search Tool) via NCBI web site. Primers used to sequence are listed in table 3.

Table 3. Sequencing primers used in this work.

Primer	Sequence (5'-3')	MCS of vector
T7 promoter	TAATACGACTCACTATAGGG	pET29c
T7terminator	GCTAGTTATTGCTCAGCGG	pET29c
pBAD Forward	ATGCCATAGCATTTTTATCC	pBAD
pBAD Reverse	GATTTAATCTGTATCAGG	pBAD
5'AOX1	GACTGGTTCCAATTGACAAGC	pPICZαA
3'AOX1	GCAAATGGCATTCTGACATCC	pPICZαA

4.7.12 Vectors

Table 4. Vectors used in this study.

Plasmid name	Resistance	Organism	Source
pBluescript II SK(+) KS/AT4 synthetic	Ampicillin	*E. coli*	Eurofins
pET29c	Kanamycin	*E. coli*	Invitrogen
pET29c KS/AT4	Kanamycin	*E. coli*	This study
pET29c LAT4	Kanamycin	*E. coli*	This study
pET29c AT4	Kanamycin	*E. coli*	This study
pET29c AT4-	Kanamycin	*E. coli*	This study
pET29c LAT4-	Kanamycin	*E. coli*	This study
pBAD	Ampicillin	*E. coli*	Invitrogen
pBAD AT4	Ampicillin	*E. coli*	This study
pBAD KS/AT4	Ampicillin	*E. coli*	This study
L1SL2/pETcoco-2	Ampicillin	*E. coli*	Prof. P.F Leadlay

pPICZαA	Zeocin	*E. coli/ P. pastoris*	Invitrogen
pPICZαA KS/AT4	Zeocin	*E. coli/ P. pastoris*	This study
pPICZαA AT4	Zeocin	*E. coli/ P. pastoris*	This study
pPICZαA AT4-	Zeocin	*E. coli/ P. pastoris*	This study

The 2.5 kb synthetic DNA region encoding the KS/AT4 domains and other 4 constructs of different configurations of the fkAT4 domain were cloned in three different expression vectors, pET29c, pBAD and pPICZα. The 5 different DNA constructions correspond to: KS/AT4 domain (KS/AT4), the linker joining the KS domain with AT4 domain (LAT4), the AT4 domain (AT4), a truncated version of the AT4 domain (AT4-) and the linker to the KS4 domain with the truncated AT4 domain (LAT4-).

These 5 constructions were PCR amplified from pBluescript II SK(+) KS/AT4 synthetic vector adding *NdeI* and *XhoI* restriction places at 5´ and 3´ positions respectively, further digested with the mentioned enzymes and ligated into pET29c vector previously digested with *Nde*I and *XhoI* in order to accommodate the five constructs in frame with the native His tag of the pET29c vector. Once cloned, the five vectors (pET29c KS/AT4, pET29c LAT4, pET29c AT4, pET29c AT4- and pET29c LAT4-) were transformed into the following strains: *E. coli* BL21 (DE3*), E. coli* C41, *E. coli* tuner and *E. coli* BL21-CodonPlus (DE3).

Besides, the constructs KS/AT4 and AT4 were ligated into the pBAD expression vector. These DNA fragments were PCR amplified from pBluescript II SK(+) KS/AT4 synthetic vector incorporating *NcoI* at 5´ position and *HindIII* with a His tag at 3´ position for further ligation in pBAD vector previously digested with the same enzymes. The resulting plasmids pBAD KS/AT4 and pBAD AT4 were subsequently transformed into *E. coli* BL21 (DE3) and *E. coli* tuner.

The DNA constructs KS/AT4, AT4 and AT4- were transformed into the eukaryotic expression system pPICZαA of *P. pastoris*. The three DNA constructs were PCR amplified from pBluescript II SK (+) KS/AT4 synthetic vector adding *XhoI* at 5´ position and *XbaI* with a His tag at 3´ position. The constructs were cloned in frame with the α factor of pPICZαA vector, previously digested with same set of enzymes generating the plasmids pPICZαA KS/AT4, pPICZαA AT4 and pPICZαA AT4- which were further transformed into *P. pastoris* X-33 strain.

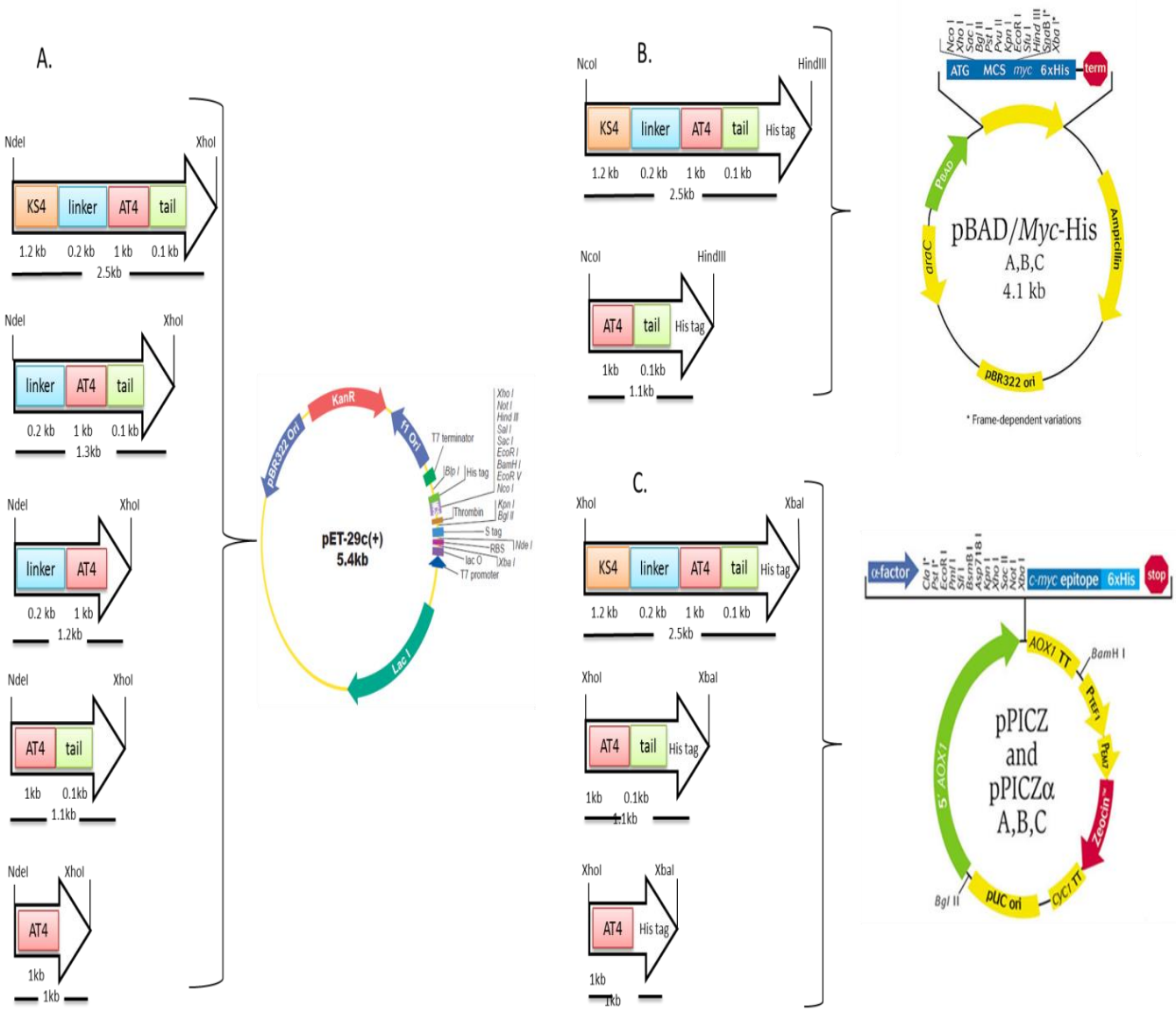


Figure 5. General cloning scheme of the different variants of the fkAT4 domain cloned in **A.** pET29c **B.** pBAD and **C.** pPICZαA.

4.8 Screening of *E. coli* Transformants

The screening of *E. coli* transformants was done using AmpliTaq Gold polymerase mixture from Thermo Fischer Scientific (United States) according to manufacturer's instructions. Bacterial lysates (1 µL) were used as DNA templates. Lysates were prepared by boiling a colony in water (200 µL, 95°C for 5 minutes).

4.9 Screening of *P. pastoris* Transformants

The screening of *P. pastoris* transformants was done using AmpliTaq Gold polymerase mixture according to manufacturer's instructions. Yeast lysates (1 µL) were used as DNA templates. Lysates were prepared resuspending a colony in 20 µL of NaOH 20

mM, the suspension was boiled for 10 minutes at 99°C and left on ice until the cell debris settled down.

4.10 Expression of the fkAT4 domains in *E. coli* BL21 (DE3), C41 and tuner strains

Plasmids pET29c and pBAD containing the different variants of fkKS/AT4 domains fused to a N-terminal His-tag were transformed into *E. coli* BL21 (DE3), C41 and tuner strains. The different strains were grown overnight in 5 mL LB medium at 37°C and 220 rpm, further 500 µL of the overnight culture were transferred to fresh 10 mL LB and the cultures were cultivated at 37°C to an OD_{600} 0.6-0.8. Once the OD_{600} was reached the cultures were cooled down to 18°C and induced with IPTG to a final concentration between 0.1 and 1 mM for pET29c series plasmids or 0.001% and 1% arabinose for pBAD series plasmids. Protein expression continued for an additional time (3, 6 or 20 hours) at different temperatures (20°C, 30°C and 37°C). Once the expression time was completed, the cells were harvested by centrifugation and frozen at -20°C. Cell pellets were analyzed by SDS-PAGE to confirm heterologous expression of fkAT4 domain. Cultures were supplemented with 100 µg/ml ampicillin or 50 µg/ml kanamycin for pBAD or pET29c respectively.

Cell pellets showing expression of the fkAT4 domain were thawed to 4°C and resuspended in lysis buffer (100 mM Tris-HCl, pH 7.5, 300 mM NaCl and 20 mM imidazole) and the cell weres disrupted via sonication. After ten rounds of sonication (10 burst of 10 seconds followed by intervals of 10 seconds for cooling down) the cell lysate was centrifuged and both, cell debris and supernatant, were analyzed through SDS-PAGE to confirm the solubility of the target recombinant proteins.

4.11 Expression of the fkAT4 domains in *E. coli* BL21 (DE3) codon plus RP strain containing the plasmid L1SL2/pETcoco-2

E. coli BL21 (DE3) codon plus RP strain containing the plasmid L1SL2/pETcoco-2 was independently transformed with pET29c KS/AT4, pET29c AT4, pET29c LAT4, pET29c LAT4- and pET29c AT4-. Protein expression was performed as mentioned in the section 4.10 with slightly modifications. The cells were cultivated in LB medium supplemented with 50 µg/mL ampicillin, 30 µg/mL kanamycin, 34 µg/mL chloramphenicol and 0.2% w/v glucose.

4.12 Expression of the fkAT4 domains in *P. pastoris* X-33

Single colonies of the recombinants strains of *P. pastoris* X-33 harboring the three different variants of KS/AT4 domains were inoculated in 100 mL BMGY medium and cultivated at 30°C and 200 rpm to an OD_{600} 2-6. The cells were then harvested at 3000 g for 5 minutes at room temperature, the supernatant was discarded and the cell pellet was resuspended in 20 mL BMMY to reach a final OD_{600} 1. The cultures were induced every day during 3 days with 100% methanol to a final concentration of 0.5%. 1 mL samples were collected every 12 hours and analyzed by SDS-PAGE, both pellet and supernatant. After three days of culture, the cells were harvested and the

supernatant was loaded onto a HisTrap HP nickel column to concentrate the recombinant proteins, fractions containing the target protein were analyzed by SDS-PAGE.

4.13 Purification and refolding of fkAT4 inclusion bodies.

Inclusion bodies containing the different variants of the fkAT4 domain were obtained from the *E.coli* tuner strain carrying the pET29c fkAT4 based vectors. Protein expression was performed as described in section 4.10. Specifically, the recombinant *E. coli* tuners strains were induced with 0.1 mM IPTG and cultivated for 3 hours at 37^{o}C. The volumes were scaled up to 1 L and after the 3 hours of induction, the cells were harvested and resuspended with 10 mL buffer A (50 mM Tris-HCl pH 7.5, 10% sucrose, 5 mM $MgCl_2$, 1 mM EDTA and 0.1 mM PMSF) and incubated for 30 minutes at 37^{o}C with 0.7 mg/mL lysozyme. After the incubation period, 10 mL of buffer B (50 mM Tris-HCl pH 7.5, 1 M NaCl, 2 mM $MgCl_2$, 0.1 mM EDTA and 0.6% w/v triton X-100) were added and the mixture was rock on ice for 10 minutes followed by sonication. To obtain the inclusion bodies the cell lysate was centrifuged at 27,000g for 10 min at 4^{o}C, and the pellet was resuspended and washed overnight with 20 mL buffer C (6 M guanidine hydrochloride, 50 mM Tris-HCl pH 7.5, 0.05 mM EDTA, 200 mM NaCl and 50 mM Imidazole). Unfolded proteins were purified and refolded using Ni-Sepharose affinity chromatography on an Akta FPLC apparatus. Briefly, after filtration of the mixture containing the unfolded proteins through 0.45 µm membrane, the solution was loaded onto a 5-mL HisTrap HP nickel column previously equilibrated with 2 column volumes of buffer C. The proteins were refolded inside the column by washing the 6 M guanidine using a linear gradient with buffer D (50 mM Tris-HCl pH 7.5, 0.05 mM EDTA, 200 mM NaCl and 50 mM Imidazole). The refolded proteins were eluted with a linear gradient of buffer E (50 mM Tris-HCl pH 7.5, 0.05 mM EDTA, 200 mM NaCl and 500 mM Imidazole). Fractions containing the recombinant fkAT4 domains were dialyzed overnight with buffer E (50 mM Tris-HCl pH 7.5, 0.05 mM EDTA and 200 mM NaCl) to remove the excess of imidazole. Samples containing the recombinant proteins were frozen for further analysis.

4.14 Enzymatic assay of the refolded fkAT4 domains produced by *E. coli* tuner

The acyltransferase activity of the refolded AT4 domain was determined spectrophotometrically by monitoring the increase in the A412 absorbance using Ellman's reagent (DTNB). Briefly, 20 µl reaction mixture containing the refolded proteins and reaction buffer (75 mM Tris-HCl pH 7.5, 3 mM DNTB, 10% DMSO and 5 mM allylmalonyl-SNAC) was incubated for up to 20 hours at 37^{o}C. Three negative controls were included, without DNTB, without allylmalonyl-SNAC and without the refolded proteins. After the incubation time the absorbance was determined at the wavelength mentioned before.

5 Results

5.1 Cloning the fkAT4 domains in the expression plasmids pET29c and pBAD

It is known that the acyltransferase domains are in charge of selecting the extender units. However, the specific residues involved in this selection are still unknown; moreover inter-domain regions could play a significant role in the selection process. In order to stablish whether or not the inter-domain regions are also influencing the selection of the extender units, we decided to evaluate 5 different variants of the fkAT4 domain. The KS/AT4 domains, the AT4 domain, the linker that joins the KS4 and AT4 domains with the AT4 domain (LAT4), a truncated version of the AT4 domain (AT4-) and the linker that joins the KS and AT4 domains with the truncated AT4 domain (LAT4-) were cloned in the different expression vectors (Fig. 6). After sequence confirmation, the different plasmids were transformed in the expression *E. coli* strains BL21, C41, tuner and codon plus as described in materials and methods.

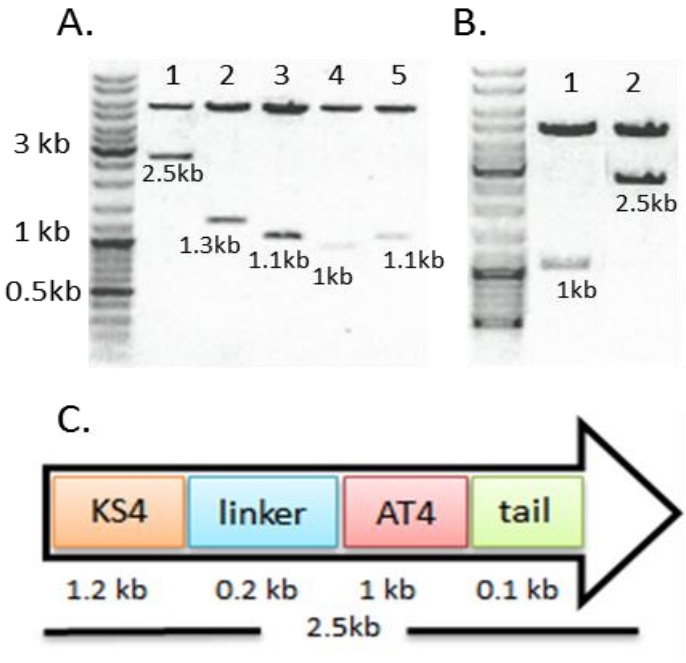


Figure 6. **A.** pET29c with 1) KS/AT4 2) LAT4 3) AT4 4) AT4- 5) LAT4- digested with *NdeI/XhoI*. **B.** pBAD with 1) AT4 2) KS/AT4 digested with *NcoI/HindIII*. **C.** Synthetic ORF encoding the KS/AT4 domains synthesized with codon usage for *E. coli*.

5.2 Expression of fkAT4 domains in *E. coli* strains BL21 and C41.

The BL 21 and C41 *E. coli* strains transformed with pET29c and pBAD vectors containing the different variants of the fkAT4 domain probed to be incapable to express any of the different versions of the fkAT4 domain. Different cultivation conditions were modified, including variations in temperature, inducer (IPTG or arabinose) concentrations and time of expression (Fig. 7). However, BL21 and C41 transformants were not even able to produce inclusion bodies of the target proteins (Data not shown).

5.3 Expression of fkAT4 domain in *E. coli* tuner strain.

E. coli tuner strain efficiently overexpressed the 5 different variants of the fkAT4 domain but only with the pET29c expression system. The pBAD tuner transformants were incapable to express the fkAT4 domain. Nonetheless, the tuner transformants containing the pET29c vector with the different variants of the fkAT4 domain were only capable to produce insoluble aggregates of the target proteins. Several attempts were performed trying to improve solubility of the recombinant proteins, mainly focused on cultivation conditions such as modifying the IPTG concentration, induction time and temperature (Fig. 7). However, none of the strategies tested improved the solubility of the target proteins (Fig. 8).

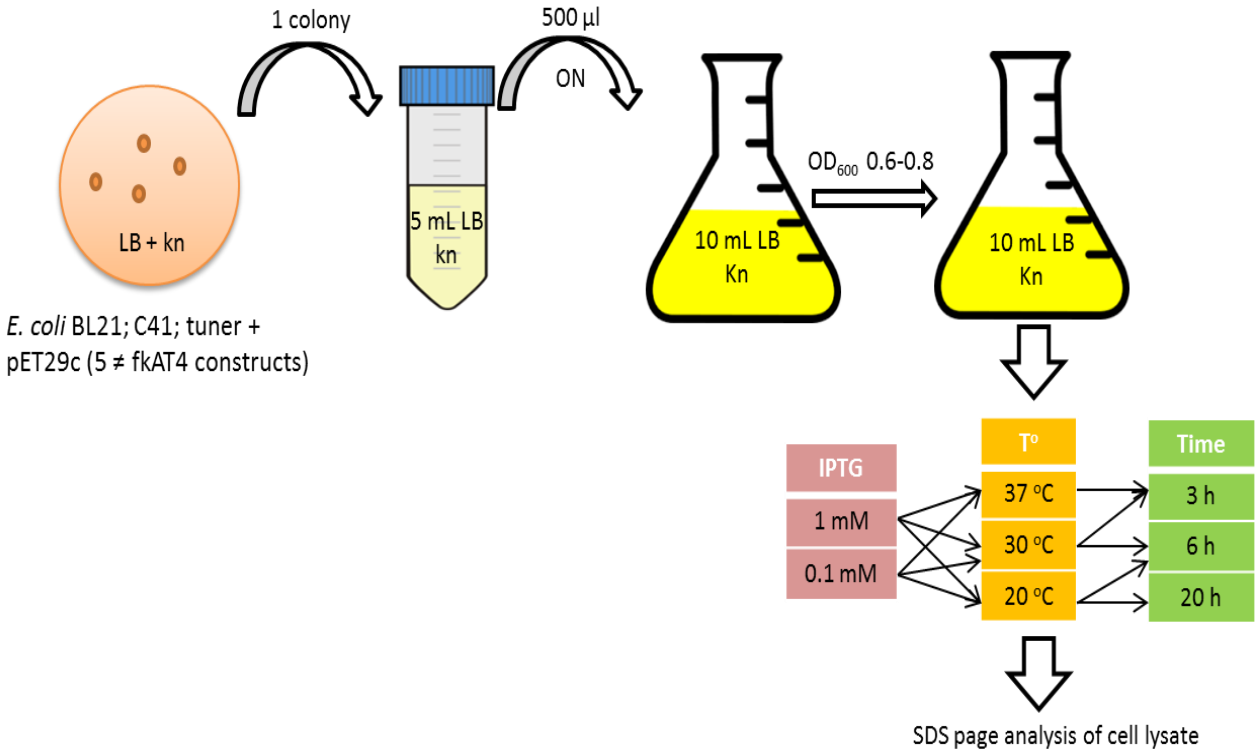


Figure 7. General scheme displaying the different cultivation conditions tested to overexpress the recombinant fkAT4 domain in *E. coli*.

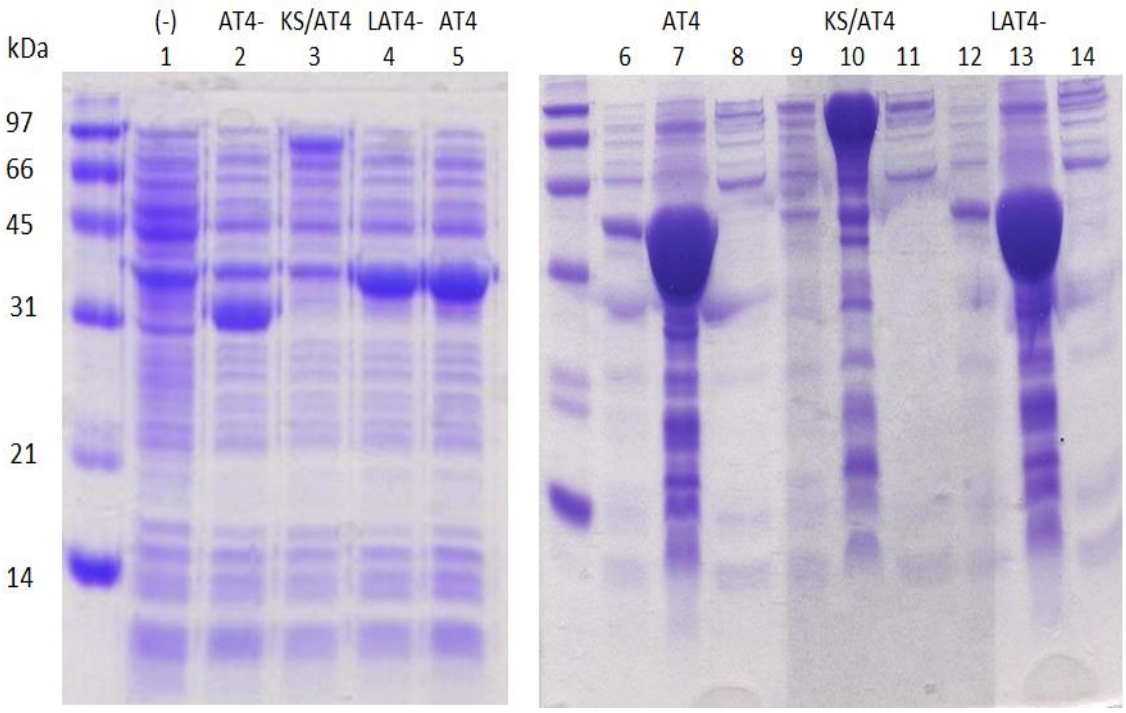


Figure 8. Induction of fkAT4 variants in *E. coli* tuner strain containing pET29c expression vector. Induction conditions: 0.1 mM IPTG for 3 hours at 37°C. **Crude extract of tuner strains containing** 1. Empty plasmid 2 and 6. pET29c AT4-. 3 and 9. pET29c KS/AT4. 4 and 12. pET29c LAT4-. 5. pET29c AT4. **Pellet of tuner strains containing** 7. pET29c AT4. 10. pET29c KS/AT4. 13. pET29c LAT4. **Soluble fraction containing** 8. pET29c AT4. 11. pET29c KS/AT4. 14. pET29c LAT4.

5.4 Expression of fkAT4 domains in *E. coli* BL21 (DE3) codon plus RP strain containing the plasmid L1SL2/pETcoco-2.

Heterologous expression of proteins from *Streptomyces* origin using *E. coli* as a host has proven to be a difficult task. In order to increase solubility of fkAT4 domain we evaluated the *E. coli* BL21 (DE3) codon plus RP strain containing the plasmid L1SL2/pETcoco-2 kindly provided by professor Dr. Leadlay (U. Cambridge). This strain contains extra copies of the tRNA gene *argU*, which recognizes the AGA and AGG arginine codons, and the tRNA gene *proL*, which recognizes the proline codon CCC. Although our fkAT4 domain was codon optimized for *E. coli* expression, we were interested in the plasmid L1SL2/pETcoco-2 which encodes the chaperonins *GroEL1*, *GroES* and *GroEL2* from *Streptomyces coelicolor*. Leadlay and collaborators demonstrated that co-expression of target proteins from *Streptomyces* using these chaperonins in *E. coli* enhanced solubility of the target protein, thus avoiding the formation of inclusion bodies [24]. *E. coli* BL21 (DE3) codon plus RP strain containing the plasmid L1SL2/pETcoco-2 were independently transformed with pET29c KS/AT4, pET29c AT4, pET29c LAT4, pET29c LAT4- and pET29c AT4- plasmids. Different cultivation settings (temperature, IPTG concentration, time of expression) were evaluated in order to overexpress the recombinant fkAT4 domain, however no soluble protein was detected (data not shown) using these strains.

5.5 Cloning and expressing the fkAT4 domains in the expression host *P. pastoris* X-33

We tried to express the fkAT4 domain in *P. pastoris* due its advantage to secret proteins to the extracellular space, thus increasing the chances to produce soluble functional fkAT4 domain. The KS/AT4, AT4 and AT4- constructs were cloned in frame with the α-factor under the regulation of AOX1 promoter present in the expression vector pPICZαA. *P. pastoris* X-33 transformants were analyzed to confirm integration of the plasmids in the chromosome through colony PCR amplifying a region of 574 bp of the AT4 domain (Fig. 9). Positive colonies displaying correct integration of fkAT4 domain in the chromosome were selected for protein expression experiments.

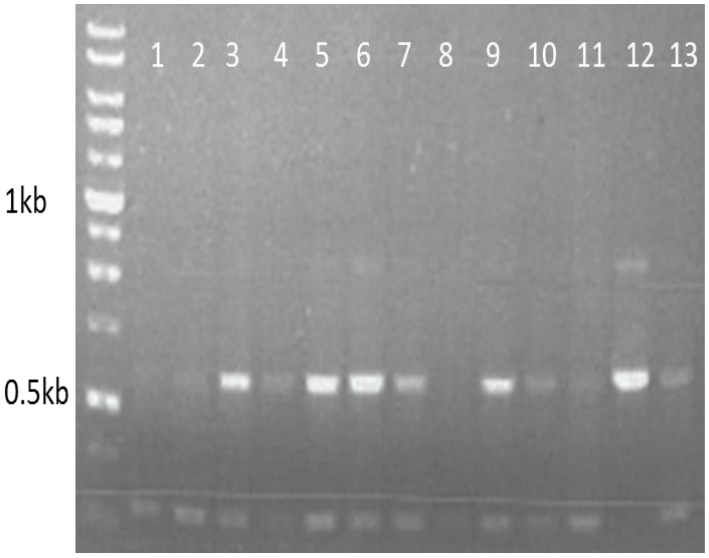


Figure 9. Confirmation of pPICZαA KS/AT4 (lines 1-4), pPICZαA AT4 (lines 5-8) and pPICZαA AT4- (lines 10-13) integration into *P. pastoris* X-33 chromosome. A band of 574 bp in lines 3, 4, 5, 6, 7, 9, 10, 12 and 13 indicated the correct integration of the plasmids into *P. pastoris* chromosome.

Protein expression in *P. pastoris* X-33 was carried out as described in material and methods by induction with methanol and collecting samples every 12 hours. After 72 hours of expression, the supernatant was loaded onto a HisTrap HP nickel column in order to concentrate the target protein. SDS-page analysis revealed no protein expression, neither extra- nor intra-cellular, using *P. pastoris* expression system.

5.6 Refolding the fkAT4 inclusion bodies

Despite the different strains and cultivation conditions tested to get soluble fkAT domain, only *E. coli* tuner strain was able to produce inclusion bodies of the 5 different versions of the fkAT4. The protein aggregates produced by the tuner strain were purified and solubilized using a HisTrap HP nickel column, as described in materials and methods. The 5 constructs were successfully refolded and purified, showing the corresponding bands of 32, 36, 39, 43 and 88 kDa for AT4-, AT4, LAT4-, LAT4 and KS/AT4 respectively. Nonetheless the efficacy of this strategy to refold the 5 variants of

fkAT4 domain, the KS/AT4 construct showed an extra band of around 40 kDa most probably because the protein was cleaved during the refolding process (Fig. 10).

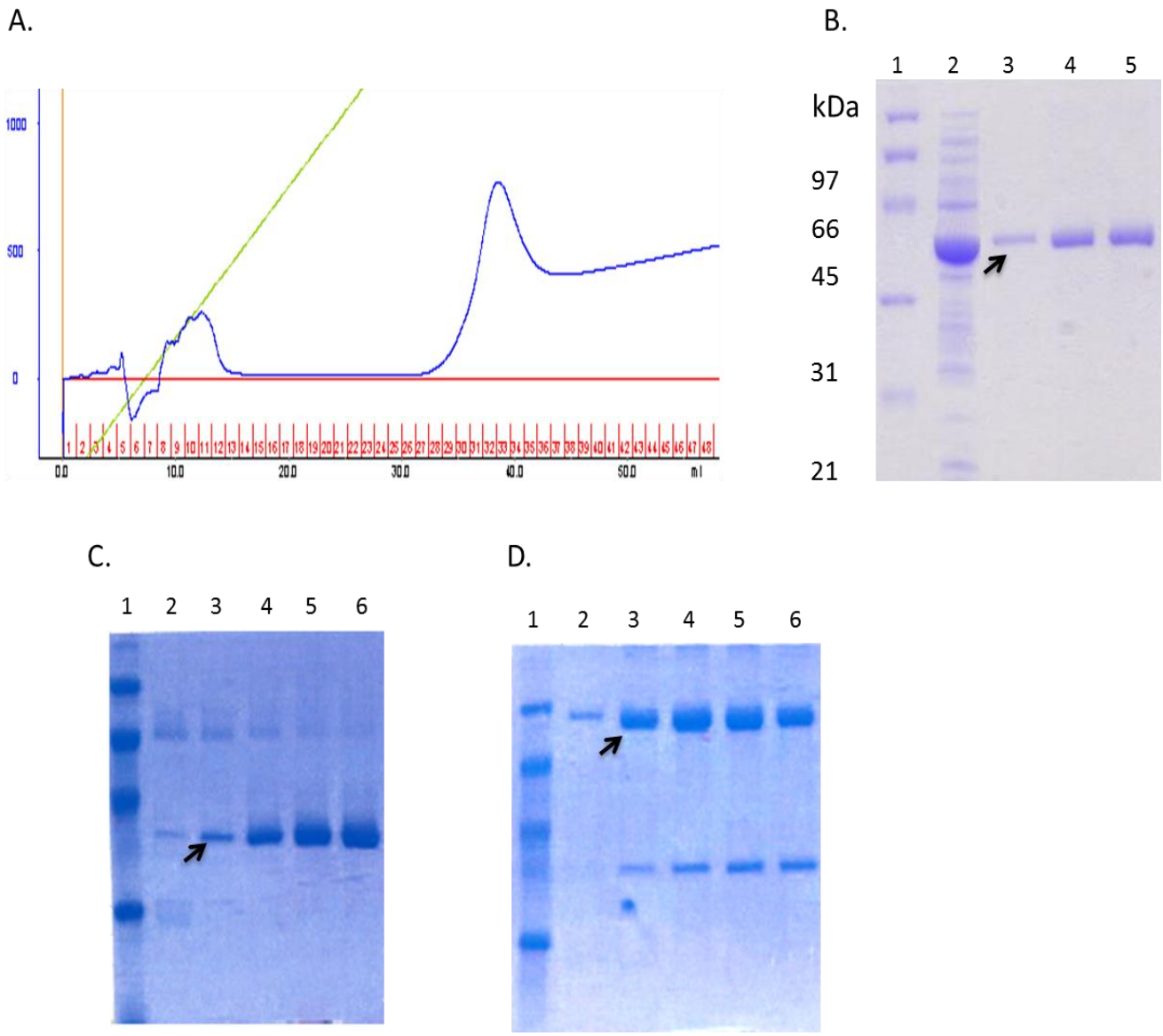


Figure 10. A. Typical chromatogram observed for refolded fkAT4 domains after removing the 6M guanidine hydrochloride and eluting with imidazole. **B.** SDS-page of refolded fkAT4 domain (line 1: Protein marker, line 2: pellet, lines 3, 4 and 5: solubilized fractions of fkAT4). **C.** SDS-page of refolded LAT4- (Line 1: protein marker, following lines: solubilized fractions of LAT4-) **D.** SDS-page of refolded KS/AT4 (Line 1: protein marker, following lines: solubilized fractions of KS/AT4)

5.7 Enzymatic assay to evaluate the refolded fkAT4 domains

The acyltransferase activity of the refolded fkAT4 domains was evaluated using the Ellman's reagent (DNTB), which is used to quantity free sulfhydryl groups in solution. Before doing the reaction, the purified fkAT domains were dialyzed to remove the imidazole because it's yellowish appearance and also because it reacts with DNTB. The

coupled spectrophotometric assay was designed to detect the acyltransferase activity of the fkAT4 domain by the release of free CoA from the extender unit allylmalonyl-CoA, thus DTNB would react with the thiol group of the free CoA generating an increase in the A412 absorbance (Fig. 11).

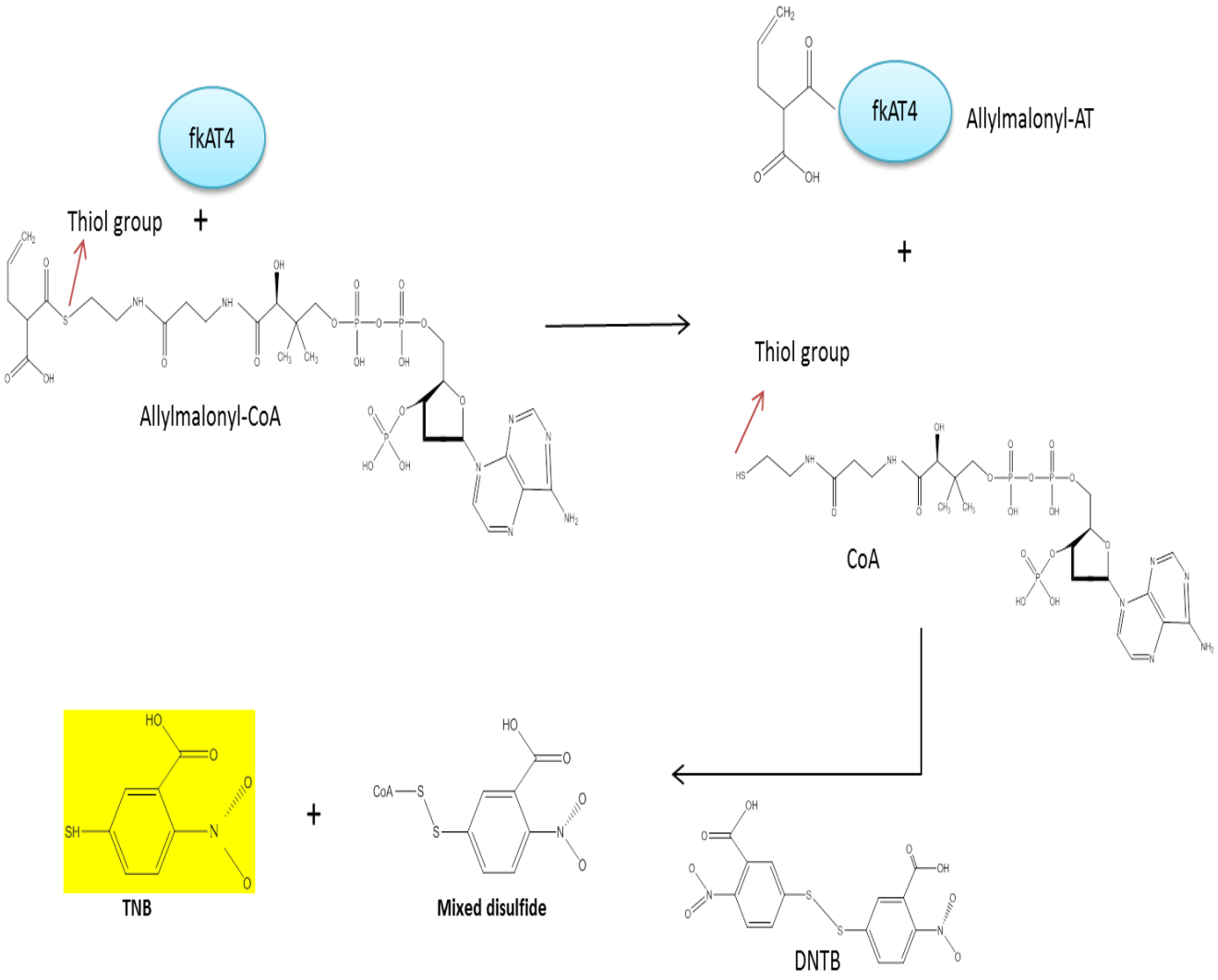


Figure 11. Expected reaction between purified fkAT4 domain and the extender unit allylmalonyl-CoA. In the first step of the reaction the fkAT4 domain selects the allylmalonyl extender unit releasing CoA. In the second part of the reaction the ellman´s reagent interacts with the free thiol group of CoA generating TNB which is detected at A412.

Unfortunately, the fkAT4 domains, previously denatured and refolded, started to precipitate once they were thawed, thus suggesting they were not properly refolded. The enzymatic assay was carried out and confirmed our suspects. The 5 different variants of the fkAT4 domain could not restore their natural three-dimensional arrangement, thus making impossible the assessment of their activity *in vitro*. Different psychochemical parameters of the reaction were modified (pH, temperature and incubation time); however no change in the absorbance was observed when the proteins were exposed to DNTB.

6 Discussion

Incorporation of unusual extenders units into the backbone of type I polyketides could introduce much greater structural variability to this natural products. Better understanding of substrate specificity of AT domains would provide knowledge about how to modulate their specificity. The fkAT4 acyltransferase domain presents a great opportunity as a model AT system for further studies of substrate specificity, thus improving the overall understanding of extender unit selection in PKS enzymes. We directed our efforts in the understanding of the unusual selectivity of the fkAT4 domain through *in vitro* studies. To achieve this goal, we first synthesized the 2.5 Kb open reading frame encoding the fkKS/AT4 domains optimized for codon usage in *E. coli,* and we expressed 5 different variants this domain in two expression systems, one prokaryotic and one eukaryotic.

E. coli is the most convenient and frequently used overexpression system to produce recombinant proteins [25], not in vain we synthesized the fkAT4 domain with codon bias for this expression host. First results obtained with the *E. coli* strains BL21 and C41 showed no recombinant protein production, thus indicating that the fkAT4 domain is toxic when pET expression system is used in to express this domain. To reduce toxicity of the recombinant protein in *E. coli* it is necessary to control the level of induction. Two strategies have been reported to do so, tuning the promoters and/or using strains that allow control of induction [25]. We tried with arabinose *araPBAD* promoter included in the pBAD vector. Arabinose promoter offers a tighter regulation reducing the strength of the induction, in comparison with pET expression vectors regulated by the T7 promoter [26]. However, no recombinant protein was detected using pBAD containing transformants. On the other hand, the *E. coli* tuner strain is a BL21 derivative with a lac permease (*lacY*) mutation that allows uniform entry of IPTG into all *LacY*− cells in the population, which produces a concentration-dependent, homogeneous level of induction [27]. The *E. coli* tuner strain enabled us to purify inclusion bodies, at least.

Inclusion bodies are protein aggregates with non-native conformations. Its formation is the result of unbalanced equilibrium between protein aggregation and solubilization [28]. *E. coli* offers a specific microenvironment which may differ from the original host, in our case *S. tsukubaensis*, in terms of pH, osmolarity, redox potential, cofactors, and folding mechanisms. Even slightly modifications of these parameters could lead to protein instability and aggregation [29]. Disulfide bonds are also a matter of concern regarding the generation of inclusion bodies because erroneous disulfide bonds formation can lead to protein misfolding and protein aggregation [25]. However, the fkAT4 domain is not predicted to form disulfide bridges, according DISULFIND [30], discarding this potential problem as responsible for the formation of inclusion bodies. Several strategies have been published to ameliorate the formation of inclusion bodies, mainly focused in the co-expression of chaperons and/or slowing down *E. coli* metabolism [25]. We tried both, using chaperons from *S. coelicolor* [24] and lowering production rate by decreasing incubation temperature [31]. However, *E. coli* expression system proved to be incapable to produce the native conformation of the fkAT4 domain.

The yeast *Pichia pastoris* has been widely used for the expression of recombinant proteins exploiting its capacity to secrete the target protein [32]. The KS/AT4, AT4 and AT4- were cloned in frame with the native signal sequence of the commercial vector pPICZαa to secrete the different variants of the fkAT4 domain, thus trying to avoid the formation of inclusion bodies obtained with *E. coli*. Although the advantage of *P. pastoris* to export the target protein, this host couldn't produce the recombinant fkAT4 domain. The failure of *P. pastoris* to produce the fkAT4 domain could be due the presence of extracellular proteases hydrolyzing the target protein, the difference in GC content (table 5) [32] or simply the incapacity of this microorganism to produce this kind of proteins.

Table 5. GC content of the microorganisms used in this study and the synthesized fkAT4 domain.

DNA	GC (%)
S. tsukubaensis	72
E. coli	51
P. pastoris	Genome: 41; coding genes: 79
Codon optimized fkAT4	61

Despite it was not possible to obtain the fkAT4 domain properly folded, we decided to continue our work with the inclusion bodies produced by *E. coli* tuner strain. The aggressive strategy for re-folding inclusion bodies implies the denaturation of the protein aggregates with guanidine hydrochloride, a protein destabilizer that at high concentrations denatures proteins by a chaotropic effect [28]. Once the protein aggregates are denatured, the guanidine hydrochloride is removed using a gradient to refold the target recombinant protein, the fkAT4 domain in our case. After the removal of the guanidine hydrochloride is expected to recuperate the protein in its native tertiary structure. Following the re-folding procedure, it was possible to recover the 5 variants of the fkAT4 domain as soluble proteins. However, the recombinant proteins purified were aggregated and/or misfolded during refolding. Nonetheless, we evaluated the activity of the refolded fkAT domain using the Ellman reagent. Unfortunately, the refolded fkAT4 domain was not re-folded correctly, preventing the assessment of its *in vitro* activity.

Recently Li and collaborator reported the purification of the fkAT4 domain using *E. coli* BL21 strain as a heterologous host [22]. The fkAT4 domain was produced using the pET28 expression vector, inducing with 0.1 mM IPTG for 6 hours at 30^{o}C, conditions also tested in this work with negative results. Only one difference was observed regarding our experimental procedure, the open reading frame selected by Jiang et al. corresponds to the LAT4- construct (truncated version of the fkAT4 domain with the linker to the KS domain). However, the linker region that we evaluated has 94 amino acids, while the linker used by Li and colleagues has 42 amino acids [22]. Li and coworkers confirmed that the fkAT4 domain is active for allylmalonyl and ethylmalonyl

but not for methylmalonyl and malonyl units in transacylation reactions. Moreover, the LAT4- transfers allylmalonyl and ethylmalonyl from the acyl donors ACP or CoA, thus suggesting the promiscuity of fkAT4 for ACP or CoA [22]. Additionally, they replaced the specific motif CPTH found in fkAT4 domain for TAGH, specific motif for ethylmalonyl, this modification abolished the transacylation reaction but not the self-acylation reaction [22]. Their findings suggest that fkAT4 transfers various acyl units from various acyl donors to various acyl acceptors, emphasizing that CoA is not necessary for the incorporation of unusual extender units and that the acyl donor ACP domain may be the common pathway [22]. Although the impressive amount of data collected by Jiang et al. there are still questions to answer regarding the fkAT4 domain such as i) how is the fkAT4 domain interacting with the extender units? Ii) What are the key residues involved in selection of the unusual extender units? iii) Why is the fkAT4 domain deselecting standard extender units? All these questions are still to be solved making the fkAT4 domain a challenging acyltransferase for *in vitro* studies. The answers for these questions will enable us to specifically modify the AT domains for the incorporation of diverse extender units, thus expanding the chemical diversity of type I polyketides.

Since 1982 the biopharmaceutical market has launched more than 200 products [33]. The production of recombinant human insulin by *E. coli* proved that microorganisms are efficient biofactories for the production of industrial relevant enzymes. Different prokaryotic and eukaryotic expression systems have been developed for the efficient production of industrially important proteins. In this work we tried to understand the unusual selectivity of the acyltransferase of the module 4 of the immunosuppressive drug tacrolimus using standard approaches. *E. coli* and *P. pastoris* are the usual microorganisms selected for production of target recombinant proteins [25, 32]. However, the expression of the recombinant fkAT4 was an unattainable challenge for both microbial cell factories. *E. coli* could produce inclusion bodies in the best scenario, while *P. pastoris* failed producing any form of fkAT4.

The development of different expression systems capable to succeed where *E. coli* and *P. pastoris* have failed is essential in order to produce novel recombinant enzymes that current expression hosts are unable to produce. To accomplish this goal, new expression hosts are required that fulfil all the drawbacks inherent of the current expression hosts. Streptomycetes are the most economically valuable group of bacteria producing over 60% of all known antibiotics and a plethora of compounds with biological activity [34]. A number of *Streptomyces* species have been used for the heterologous expression of gene clusters related to the biosynthesis of secondary metabolites [35-39], but their use as heterologous host for the production of recombinant proteins has been limited to the use *S. lividans* host [33]. In the next chapter we describe the development of *Streptomyces rimosus* host for expression of heterologous proteins.

Chapter 2

Development of protein expression system based on *Streptomyces rimosus* as host

7 Introduction

7.1 Available heterologous protein expression systems

There is a growing demand for heterologous protein production in all areas of biotechnology. Recombinant proteins play a crucial role in both, academia and industrial applications. The possibility to isolate and purify specific proteins allows researchers to study the biochemical and structural properties of these proteins *in vitro*, thus leading to much deeper understanding of complex biological processes [40]. Gained knowledge can then be applied by the industry to produce proteins (enzymes) with improved activity through directed evolution or rational design [41]. However, usually the bottleneck is the production of the target protein. Several expression systems have been developed for the production of recombinant proteins which are relatively widely used including bacteria, yeast, filamentous fungi, insect cells, mammalian cells and transgenic animals and plants (Table 6.).

Table 6. The most commonly used expression systems (only major features are listed).

	Bacteria			**Yeast**	**Filamentous fungi**
Characteristics	***E. coli***	***Bacillus***	***S. lividans***	***P. pastoris/ S. cerevisiae***	***Aspergillus sp.***
Generation time	30 min	55 min	95 min	90 min	2 hours
Complexity of cultivation medium	minimum	minimum	minimum	minimum	minimum
Cost of cultivation medium	low	low	low	low	low
Strength of expression	high	low	low	high	high
Export out of the cell	No	Yes	Yes	Yes	Yes
Disulfide bonds formation	poorly adapted	Yes	Yes	Yes	Yes
Inclusion bodies formation	frequently	rarely	rarely	rarely	rarely
Posttranslational modifications	No	No	No	Yes, but not correctly	Yes, but not correctly

Characteristics	Insect cells	Mammalian cells	Transgenic	In vitro
Generation time	18-24 h	24 h	up to years	4-16h
Complexity of cultivation medium	complex	Complex	complex	None
Cost of cultivation medium	high	high	high	high
Strength of expression	variable	Low	high	Low
Export out of the cell	Yes	Yes	Yes	None
Disulfide bonds formation	Yes	Yes	Yes	yes
Inclusion bodies formation	rarely	rarely	no data	None
Posttranslational modifications	Yes	Yes	Yes	None

No expression system can accommodate all the needs for heterologous protein expression. Among them *E. coli* is the most commonly used expression platform; it has a diverse number of vectors and promoters, a number of engineered strains are available and many cultivation strategies exist, most often allowing the production of high yield of target recombinant protein by this bacterium in a short period of time [25]. However, it frequently fails producing active proteins and instead produces insoluble aggregates, considering that the target protein cannot be exported outside of the cell [28]. The yeast *Pichia pastoris* is another established expression host [32] which does not have that diverse genetic tools but, like *E. coli*, it has short generation time and it is easy to genetically manipulate, often ensuring sufficient yields of recombinant protein, which are most often produced intracellularly at high yields [32]. Besides, *P. pastoris* can also export the target protein [25]. The major drawback of *P. pastoris* is its tendency to hyper-glycosylate expressed proteins, an undesirable feature when the recombinant protein is of prokaryotic origin [32]. *Saccharomyces cerevisiae* host also offers many advantages to produce heterologous proteins including simple and efficient methodologies, correct processing of eukaryotic proteins, and industrial scalability. Dozens of pharmaceutical proteins such as insulin, vaccines and blood factors, produced by *S. cerevisiae* have entered the market [42], however this host also has tendency to over-glycosylate target proteins. These three microorganisms are the most commonly used hosts to produce recombinant proteins. Other organisms have also been used to produce recombinant proteins, which however, have not been developed as well as *E. coli, S. cerevisiae and P. pastoris.* They don't offer sufficiently high yields or often lack simple and rapid methods for genetic manipulation compared to *E. coli*, *P. pastoris* and *S. cerevisiae*. For instance *B. subtilis* produces many proteases that can degrade the target recombinant proteins, and is also quite difficult to genetically manipulate [43]. On the other hand, *Streptomyces lividans* is not producing extracellular proteases and, it is capable of secreting recombinant proteins directly into the culture medium. The use of *S. lividans* as a host facilitates extracellular protein folding, subsequently simplifying purification procedure. Unfortunately, the yields of the target heterologous protein are often low, therefore not providing sufficient yields of the target protein for industrial applications [33]. Although *S. lividans* is undoubtedly the most versatile *Streptomyces* strain for production of heterologous proteins, relatively slow and filamentous growth also presents disadvantage compared to *E. coli.* It has been demonstrated, that filamentous fungi such as *Aspergillus* sp. or

Trichoderma sp. can produce heterologous proteins in very high yields [44]. However, there are a number of drawbacks, when using these heterologous hosts such as slow growth and relatively difficult genetic manipulation. In addition, a considerable number of recombinant proteins are lost or stuck in the secretory pathways because of incorrect processing, modifications or misfolding, resulting in the degradation of recombinant proteins by native proteases [45]. Additionally, similarly to *P. pastoris*, filamentous fungi hyper-glycosylate target heterologous proteins [45]. Other expression systems, such as insect or mammalian cells, are most often not economical for producing bulk enzymes. In addition, they are impractical for routinely use in the laboratory, and they carry out more complex posttranslational modifications [46, 47]. Thus, considering available expression systems, only a few strains qualify as "industrially useful".

The filamentous bacterium *Streptomyces rimosus* is the industrial producer of oxytetracycline (OTC) and it holds a "GRAS" status (Generally Regarded As Safe), since it is used in commercial bioprocesses for over 50 years. *S. rimosus* is registered as a producer of OTC by the FDA and EMA (Food and Drug Administration and European Medicine Agency respectively) and other regulatory institutions. *S. rimosus* has long been used in industrial settings for OTC production and due to its medical and commercial importance, an extensive knowledge on the general biology, physiology, genetics and molecular biology techniques of this microorganism have been accumulated since 1950s, when industrial bioprocesses for oxytetracycline production was initially developed [48]. Thus, considering all circumstances, *S. rimosus* may present potentially valuable host for production of heterologous proteins. There are a number of advantageous properties of *S. rimosus*, which make this industrial strain potentially attractive expression system for heterologous protein production, such as i) rapid growth of the culture reaching high cell density in short period of time, similarly to *E. coli*, ii) it is relatively easily transformable and iii) its capacity to secrete proteins extracellularly. Good understanding of the genetics and the physiology, and the availability of well-developed industrial media and bioprocesses of *S. rimosus* at industrial scale constitute a solid basis for the re-development of this host as a potentially useful heterologous host for heterologous protein production. *S. rimosus* has a number of properties that can be compared to *E. coli*, but it also has some disadvantages such as i) relatively high production of extracellular proteolytic enzymes that can potentially degrade the target recombinant protein and ii) codon bias of streptomycetes has a high GC-content (> 70%) [49], thus the heterologous gene to be expressed in *S. rimosus* has to be synthesized with the codon usage for this bacterium in order to avoid transcriptional issues. Similarly, as most prokaryotes such as *E. coli*, *S. rimosus* does not glycosylate proteins, which also represents additional disadvantage but also an advantage, depending on the target protein. Notwithstanding, it is the aim of this thesis to evaluate the expression system based on *S. rimosus* host. When further improved and bioprocess developed, *S. rimosus* could become useful heterologous protein expression system to fulfill academic and, in some cases, even industrial requirements, making of this actinomycete an alternative host for recombinant protein production.

7.2 Genetics of *Streptomyces rimosus*

The interest for *S. rimosus* started in 1950 with the discovery of oxytetracycline (OTC) by Pfizer. OTC was the second tetracycline antibiotic discovered [50], since then four *S. rimosus* strains have been used intensively in clinic. These *S. rimosus* strains are known to produce OTC and have been studied over the years. They include the so-called Russian strain (LS-T118), the wild type ATCC 10970 strain (NRRL 2234) from the American Type Culture Collection also designated as R7, the Zagreb strain (R6), and the Pfizer strain (M4018) derived from R7 strain [48]. The R6 and M4018 strains have been used for production of oxytetracycline [48].

After initial studies of genetic linkage map of *S. rimosus* strains, carried out during the 60s and 70s [51, 52], studies of the *S. rimosus* chromosome by pulsed field gel electrophoresis (PFGE) which were conducted in the 90s allowed the construction of the physical map of *S. rimosus* [53]. These studies revealed the presence of a linear chromosome of around 8 Mb and a giant linear plasmid of 387 Kb in the R6 strain [53, 54]. Recently, relatively low quality data on the genome sequences of four *S. rimosus* strains were published [55, 56]. The draft genome is estimated to have a total of 8,416 protein-coding genes and the overall size is consistent with the physical map derived by PFGE, with 9.5 Mb and 72% G+C content, including one giant linear plasmid, which is present in all *S. rimosus* strains [55]. The four genome sequences of *S. rimosus* available in public databases include the R7 strain [55] and three R6 derivatives (the industrial OTC producer R6-500 strain and two non OTC producers R6-500 derivatives) [56]. Genome sequence analysis showed > 99% identity between R6 and R7 strains indicating that R6 was derived from the R7 soil isolate ancestor. The only obvious difference found between the four sequences analyzed was the reduced size of R6 strains genomes compared to R7 [55, 56]. R6 strains were subject of intense strain improvement programs to improve OTC production which probably resulted in extensive deletions (> 600 kb) [56].

Programs of strain improvement not only gave rise to OTC high-producing strains through random mutagenesis, essential part of strain development programs, but it also caused some undesired morphological and physiological properties of industrial *S. rimosus* strains [57], thus reducing interest for this bacterium in the scientific community. Industrial strains of *S. rimosus* suffered a phenomenon known as “strain degeneration” due to intense selection programs, thus leading to physiological, morphological, and sometimes also genetic instabilities [57]. Genetic instability in *S. rimosus* R6 is well documented [53, 56, 57]. Characteristics that are commonly affected include reduction of antibiotic production and resistance, loss of sporulation, loss of pigmentation, and changes in colony morphology [57]. Over the years, R6 strain (also called Pliva strain) was subject of extensive strain improvement through intensive mutagenesis to increase OTC production. Interestingly, over the strain improvement process, R6 strain lost around 600 kb of genome, which could also be the reason for its genetic instability [56]. However, unlike R6 strain, the wild type strain R7 and its derivative M4018 do not seem to display such profound instability.

7.3 Morphological and physiological properties of *Streptomyces rimosus*

Streptomyces rimosus is a gram positive, aerobic, filamentous bacterium belonging to the order *Actinomycetales*, family *Streptomycetaceae* and genus *Streptomyces*. The complex life cycle of *Streptomyces* spp. can be divided into six stages: 1) A single spore germinates to form a germ tube under favorable conditions; 2) the germ tube develops forming a hyphae that branches into substrate mycelium growing across and deep down into the substrate to reach nutrients; 3) aerial hyphae emerge into the air from substrate mycelium in response to nutrient depletion and other signals; 4) the aerial hyphae differentiate into a long chain of pre-spore compartments; 5) spores are dispersed in the environment; 6) the spores germinate again when they find favorable conditions starting the life cycle once more [58]. The end of the second and beginning of the third stage (late growth phase or stationary phase), after the main period of rapid vegetative growth and assimilative metabolism, marks the beginning of the so-called “idiophase” coinciding with production of secondary metabolites (antibiotics) and hydrolytic enzymes [59]. This phase is under the control of various nutritional factors such as the type and amount of carbon, nitrogen, and phosphorous sources, minerals and trace elements [59].

As expected, *S. rimosus* meets all the properties described before, which are typical for *Streptomyces* spp. However it differs in another aspect; *S. rimosus* is one of the rare, so called, rapidly growing species of *Streptomyces*. Considering relatively rapid growth rate, and other physiological parameters, often relevant to industrial bioprocesses, *S. rimosus* is somehow comparable to the classical industrial heterologous protein producers, such as *E. coli* or *P. pastoris*. It grows in the form of short fragments and reaches exponential phase in the rich medium such as TSB in 20 hours of incubation, similarly to medium LB used for *E. coli* [60]. This characteristic of *S. rimosus* makes this strain a potentially attractive host for research and industrial development, when considering expression of heterologous proteins.

7.4 Secondary metabolites produced by *Streptomyces rimosus*

Streptomyces rimosus is important industrial microorganism due its ability to produce the medically important antibiotic oxytetracycline (OTC) [48]. Structurally, OTC belongs to the group of aromatic polyketides. OTC displays broad spectrum activity against both, Gram-positive and Gram-negative pathogens [61]. The biosynthesis of OTC was intensively studied over decades [48, 50]. In addition to OTC, *S. rimosus* also produces rimocidin, a polyene macrolide polyketide with antifungal activity [62]. Although rimocidin was first isolated from *S. rimosus* [63], the molecular genetics of rimocidin biosynthesis has been evaluated using other microorganisms (*S. mauvecolor* and *S. diastaticus*) [62, 64] .

Oxytetracycline is an aromatic polyketide synthesized by a type II polyketide synthase (type II PKS) [50]. The gene cluster responsible for OTC biosynthesis contains at least 24 open reading frames (ORFs) clustered in a 34 kb DNA segment [65]. The genes include the minimal PKS (*oxyABC*), an amidotransferase (*oxyD*) and putative

acyltransferase/thiosterase (*oxyP*), all involved in the priming reactions of the nascent polyketide backbone. Further, the gene cluster contains two putative oxygensases (*oxyE* and *oxyL*), one C-6 methyltransferase (*oxyF*), three putative cyclases/aromatases *oxyIKN*, one ketoreductase (*oxyJ*), one aminotransferase (*oxyQ*), a transcriptional regulator (*otcG*), a SARP regulator (*otcR*), two resistance genes (*otrA* and *otrB*), one reductase (*oxyR*), one hydroxylase (*oxyS*) and five proteins with unknown function (*oxyGHMOT*) (Fig. 12) [50, 65, 66].

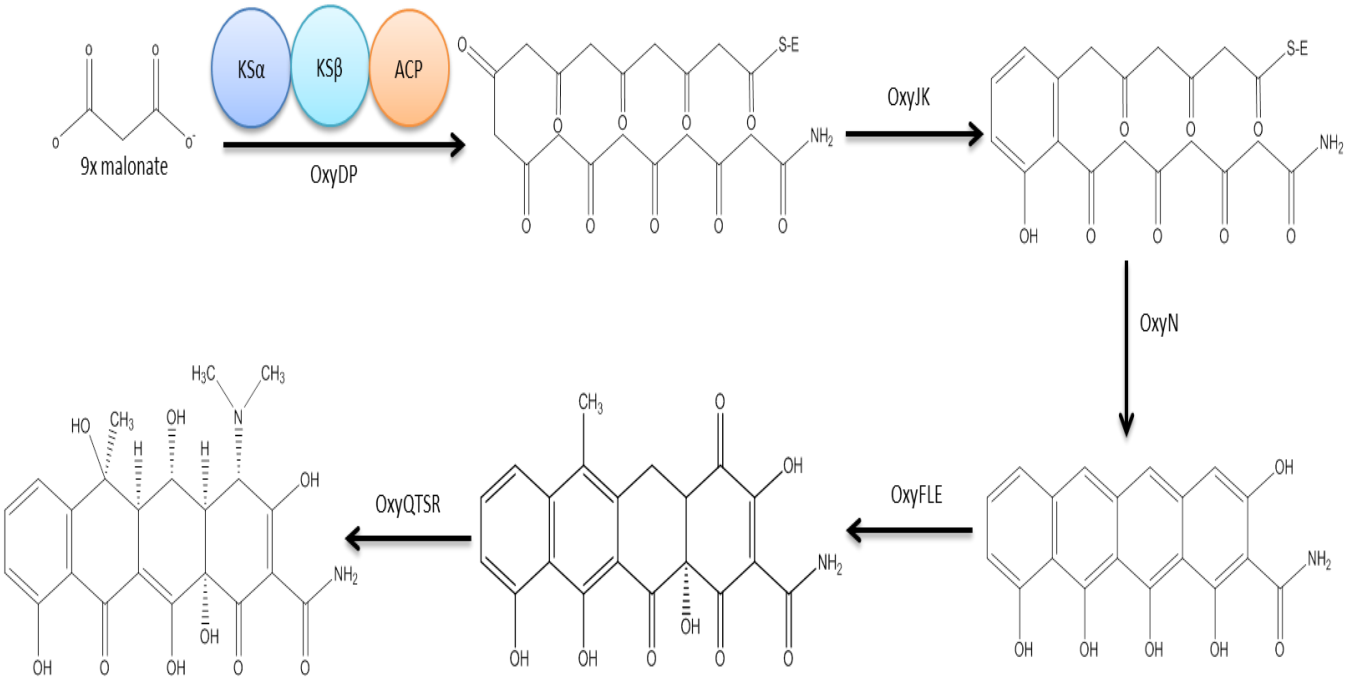


Figure 12. Proposed biosynthetic pathway of oxytetracycline in *Streptomyces rimosus*.

During the last three decades, numerous genes involved in OTC biosynthesis have been studied. Thus, the catalytic activity of majority of enzymes involved in OTC biosynthesis has been elucidated [50, 66, 67]. However, although intensive efforts have been invested in the study of the priming steps of OTC biosynthesis, the first step involved in the priming reaction of OTC is unknown. Attempts to constitute the active minimal PKS and the putative acyltransferase/thiosterase *oxyP* in *in-vitro* conditions have failed mostly due the impossibility to obtain soluble proteins [67].

Unlike OTC, rimocidin is a polyene macrolactone produced through the action of modular type I PKS [68]. The biosynthetic pathway catalyzing biosynthesis of rimocidin is encoded by 11 putative genes clustered in a 19 kb DNA segment [69]. The genes were designated *rimABCDEFGHJK*, where *rimAB* encode the type I PKSs carrying the loading module and 13 elongation modules responsible for the correct formation of the polyketide chain (Fig. 13) [64]. The other 9 open reading frames (ORFs) encode for: putative tyrosine phosphatase *rimC*, putative cholesterol oxidase *rimD*, putative glycosyl-transferase *rimE*, putative aminotransferase *rimF*, putative monooxygenase *rimG*, putative ferredoxin *rimH*, putative crotonyl-CoA reductase homologue *rimJ* and putative acetyltransferase (*rimK*) [69].

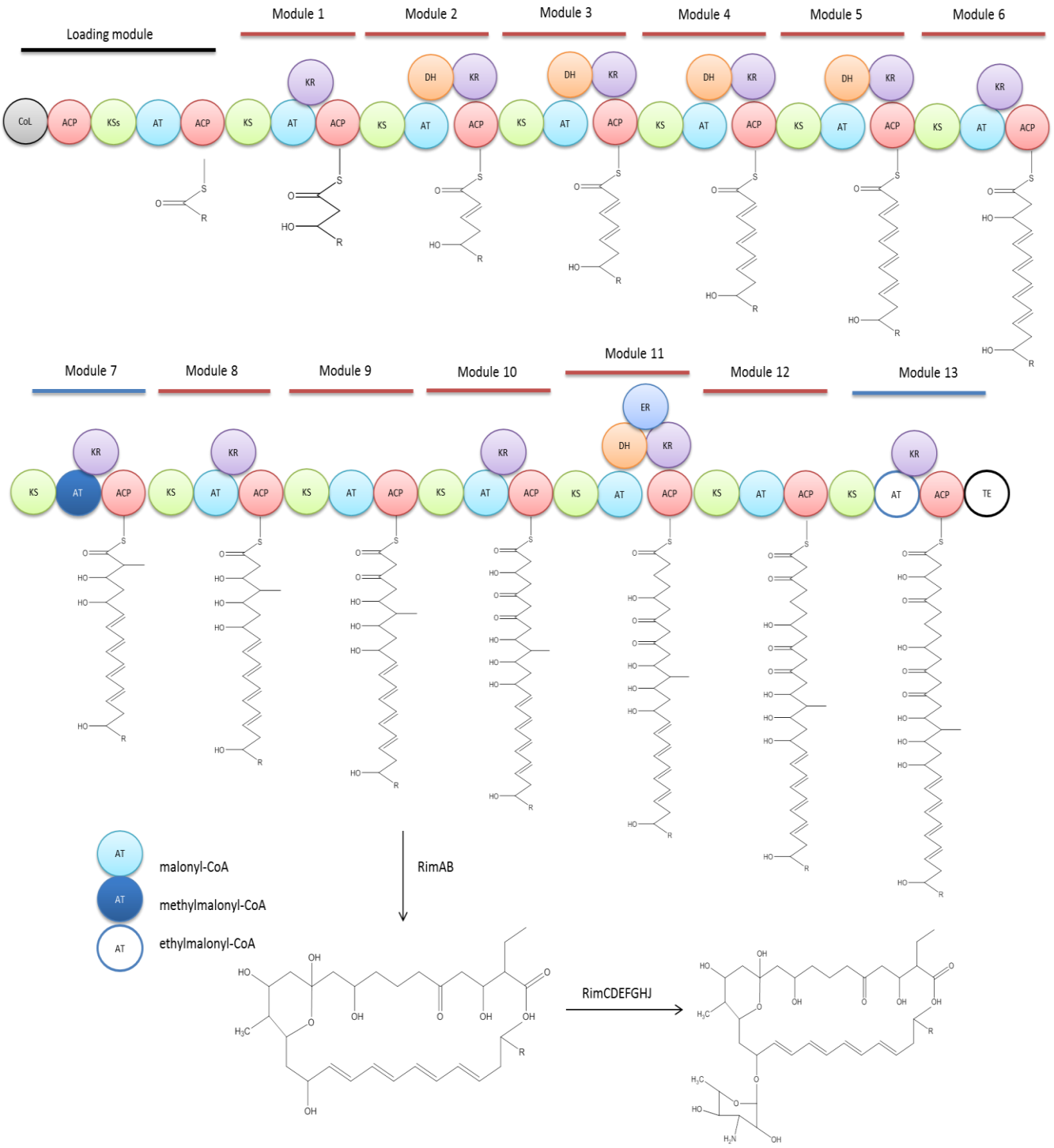


Figure 13. Schematic presentation of rimocidin biosynthesis.

Extensive knowledge on physiology of *S. rimosus*, coupled with solid understanding of OTC biosynthesis, generated over decades, presents excellent source of information, which can be applied in *S. rimosus* heterologous protein production efforts. Well-developed microbiological and molecular-biology methodologies for manipulation of *S. rimosus*, as well as available gene tools have been used in the scope of this PhD thesis. Better understanding of OTC biosynthesis, the expression of *otc* gene cluster, the molecular biology methods and gene tools for *otc* cluster manipulation, which were

developed to study OTC biosynthesis in *S. rimosus* presented good starting point for work carried out in this PhD. When considering the development of the platform for heterologous protein expression based on *S. rimosus* host, it was necessary to carry out construction of the expression vectors, select and test suitable resistance genes (markers), promoters, and to evaluate the selection of replicative and integrative vectors, in order to demonstrate that *S. rimosus* could potentially be developed as a platform for heterologous protein production.

7.5 Hydrolytic enzymes produced by *Streptomyces rimosus*

Streptomyces inhabit a wide range of niches. Generally they are found in soil, although numerous strains have adapted to aquatic environments, while some streptomycetes are known as plant and animal pathogens [58]. Generally recognized as soil bacteria, streptomycetes are saprophytes that employ a catalogue of extracellular hydrolytic enzymes to degrade organic material in their natural environment in order to obtain nutrients required for their metabolism [70]. The soil is a harsh and hostile environment with many competitors and in order to succeed in this niche, it is believed that *Streptomyces* spp. developed diverse set of secondary metabolites used to compete against other microorganism. Additionally, these bacteria also produce extracellular enzymes to degrade and exploit the complex organic materials present in the soil. As a consequence, *Streptomyces* spp. not only secretes antimicrobials, but also a large variety of hydrolytic enzymes [70].

It has been reported that *Streptomyces* spp. produce numerous hydrolases, proteases/peptidases, chitinases/chitosanases, cellulases/endoglucanases, amylases, pectinases, lignin degrading enzymes and many other hydrolases [70, 71]. Interestingly, a number of extracellular enzymes have been isolated from *S. rimosus*, such as an endodeoxyribonuclease [72], an alpha-amylase [73], a lipase [74] and a metalloendopeptidase [75]. Based on the bioinformatics analysis, *S. rimosus* is very likely producing many other extracellular enzymes. The capability of *S. rimosus* to produce numerous extracellular hydrolytic enzymes can present advantage in the development of the host for production of extracellular heterologous proteins. When considering protein export systems and choice of the secretion pathway, hydrolytic enzymes produced by *S. rimosus* offer very attractive source of signal sequences necessary for the export of target heterologous proteins. However, the presence of endogenous proteases in *S. rimosus*, on the other hand, could interfere or degrade the target recombinant proteins, aspect which has to be considered, when developing *S. rimosus* as heterologous hosts for heterologous protein production.

7.6 Recombinant proteins produced by *Streptomyces rimosus*

To our knowledge, to date, two publications are describing the potential use of *S. rimosus* as heterologous host [76, 77]. Magdevska et al. demonstrated that *S. rimosus* was capable to produce the reporter system chalcone synthase when the *rppA* gene, under the regulation of *PermE* and *PermE** promoters, was integrated in the chromosome of this bacterium using the pSAM2 containing plasmid pTS55 [77]. On the

other hand, Champreda et al. have reported the heterologous expression of triacyl glycerol lipase (BTA hydrolase), an enzyme used for degradation of bioplastics made out of adipic acid (BTA) [76]. BTA hydrolase was isolated from *Thermobifida sp.,* a thermophilic actinomycete. According to Champreda and collaborators, the heterologous production of this enzyme failed in both, *E. coli* and *P. pastoris* [76]. The gene encoding the BTA hydrolase, together with its signal sequence and a His-tag located at the C-terminal, was cloned in the conjugative vector pIJ8600 containing the ϕC31 integrase, thus allowing stably integration of the vector in the chromosome of *S. rimosus*. The new plasmid was designated pIJ-BTA and the expression of BTA hydrolase was regulated by *tipA* promoter, which is induced with thiostrepton (also incorporated in pIJ8600 vector) [76]. Champreda and colleagues reported the production of BTA-hydrolase using *S. rimosus* host. Besides, the authors were able to purify BTA-hydrolase from *S. rimosus* supernatant. However, the yield of the recombinant protein was too low to satisfy academic and/or commercial applications. Even though the poor yields of recombinant BTA hydrolase achieved from *S. rimosus* R7 host, authors have demonstrated that *S. rimosus* is capable of producing recombinant proteins, which *E. coli* and *P. pastoris* fail to produce. Therefore, further optimization of *S. rimosus* as heterologous host and development of necessary tools such as vectors, promoters and signal sequences will certainly be needed. In addition, media and process optimization, which were originally developed for the production of OTC, will likely have to be re-developed, in order to ensure higher yields of target proteins, potentially making *S. rimosus* industrially applicable heterologous host for production of selected recombinant proteins.

7.7 Available gene tools for manipulation of *Streptomyces rimosus* as heterologous host

As described by Magdevska (PhD thesis, 2011) [78] some initial efforts have been carried out to develop different gene tools based mostly on the plasmids used for studying OTC biosynthesis in *S. rimosus* and other actinomycetes. Thus, plasmid vectors for production of heterologous proteins in *S. rimosus* containing different promoters, signal sequences and other components of expression vectors were evaluated as summarized below:

1. Integrative plasmids used for heterologous expression of target enzymes: two *E. coli–S. rimosus* shuttle vectors were developed to integrate foreign DNA stably in *S. rimosus* chromosome. The integrative plasmids series designated pVM and pAB04 differ in their integration attachment site based on pSAM2 and ϕC31 recombinase respectively [79, 80]. pVM vector integrates in *S. rimosus* genome in a single location [81] and despite being very useful its integration efficiency is relatively low compared to pAB04 with much higher frequency of integration. pAB04 vector contains the ϕC31 *attP/int* site-specific recombination system from phage ϕC31. Multiple integration *attB* sites specific for ϕC31 recombinase in different *Streptomyces* spp. have been reported in literature [82]. However, for *S. rimosus* the presence of secondary *attB* sites of ϕC31 integrative vectors have never been confirmed.

2. Evaluation of common actinomycete promoters in *S. rimosus*: The erythromycin resistance gene promoter (*PermE*) and its derivative (*PermE**) from *Saccharopolyspora erythraea* have been widely use to express homologous and heterologous genes in *Streptomyces* spp. [83]. The strength of these constitutive promoters was studied intensively in *S. rimosus* using the reporter gene *xylE* [78] from *Pseudomonas putida* encoding a catechol 2,3-dioxygenase [84]. It was demonstrated that *PermE** promoter is 10-fold stronger compared to *PermE* promoter in *S. rimosus*, achieving 265 mU/mg compared to 20 mU/mg of XylE activity respectively [78]. On the other hand the ORF4/*PactI* activator/promoter system from the actinorhodin biosynthetic gene cluster of *S. coelicolor* [85], another widely used promoter in *Streptomyces* spp., did not show any activity in *S. rimosus* [78].

3. The choice of signal sequences for export of target heterologous proteins: Important advantage of using *S. rimosus* as a host to produce heterologous proteins is its capacity to secrete proteins to the medium. 8 different signal sequences were cloned/synthesized and their efficiency was tested using *PermE** promoter and the reporter system XylE, this way assessing secretion capacity of heterologous proteins in *S. rimosus* [78]. XylE proved to be very suitable reporter system, considering this protein is not exported in the native host *Pseudomonas putida* [84]. Different degrees of secretion efficiency of XylE were observed by the selected signal sequences. Analysis was carried out by comparing intra and extracellular XylE enzyme activity in cultures of *S. rimosus* M4018 transformants harboring different signal sequences fused to *xylE* gene [78]. *S. halstedii* xylanase (*xysA*) [86] and *S. rimosus* lipase (*lip*) [87] demonstrated the highest secretion capacity with 83.7% and 73.3% of extracellular XylE [78] (Fig. 14). Subtilisin inhibitor, α-amylase, and mutated α-amylase signal sequences from *S. venezuelae* [88] exported around 50% of the enzyme, while α-amylase and protease B signal sequences from *S. griseus* and trypsin-like proteinase signal sequence from *S. rimosus* were able to secrete less than 30% of the recombinant XylE protein [78] (Fig. 14). When considering the total production of the target heterologous protein and secretion efficiency together, maximal production of secreted XylE was achieved using the signal sequences of lipase and trypsin-like proteinase signal peptides, both from *S. rimosus*, thus indicating that selection of the signal peptide has significant effect on total production of the recombinant protein [78]. Nevertheless, there is still need to further evaluate the secretion capacity of *S. rimosus* using selected heterologous proteins of potential commercial importance. Obviously, there is still room for further selection and testing of more efficient signal sequences from *S. rimosus*, considering *S. rimosus* produces numerous extracellular proteins. The additional issue is also the construction of signal-sequence splice-sites, considering XylE is not exported in its natural host *Pseudomonas sp.* In addition to the activity of any target protein, it is also important to evaluate the quality of the protein, this means correct folding and correct post-translational modifications, such as disulfide bond formation.

7.8 Gene tools developed in diverse *Streptomyces* strains for expression of heterologous proteins

The establishment of a robust heterologous host requires the development of a set of different tools including suitable replicative and integrative vectors, promoters, signal sequences and different culture media to achieve quality and sufficient yield of target recombinant protein. *E. coli* is most often the first choice to produce recombinant proteins due to the vast knowledge and versatile gene tools generated in this host, considering it is easy to manipulate, and it reaches high cell density in short period of time [89]. It can be considered, that *S. rimosus* has comparable characteristics when compared to *E. coli* in terms of specific growth rate. Importantly, *S. rimosus* can export proteins to the culture media, thus facilitating further purification processes. However, the genetic tool kit of *S. rimosus* is not sufficiently developed for heterologous protein expression. Petković and collaborators only carried out initial evaluation of *S. rimosus* as heterologous host for protein production by using integrative plasmids and selected signal sequences [78]. They have also carried out initial evaluation of the strength of selected common *Streptomyces* promoters using the reporter gene *xylE* [78]. The current gene tools for *S. rimosus* manipulation and procedures used for heterologous expression of target proteins did not prove to be sufficient and industrially applicable. To improve the efficiency of *S. rimosus* as heterologous hosts, there is a critical need to improve the gene tools of *S. rimosus*. Importantly, all the strains available today, have been manipulated by random mutagenesis and selection in the scope of strain improvement activities, and do not suite ideally for production of heterologous proteins. To date, two integrative plasmids (pVM vector congaing pSAM2 recombinase and pAB04 containing ϕC31 recombinase), several constitutive promoters (*PermE, PermE*,* ORF4/PactI) and eight signal sequences have been used and evaluated for the heterologous expression of recombinant proteins in *S. rimosus* with limited degree of success [78]. Additional gene tools that could be used in *S. rimosus*, this way improving its properties as heterologous host are described below.

7.8.1 Integration of foreign DNA into *Streptomyces* spp. by mediated site-specific recombination

The phages recombinases originated from ϕC31, ϕBT1, SV1 and TG1 phages have been adapted to a number of plasmids in order to stably integrate exogenous DNA into the chromosome of several *Streptomyces* species. Another recombinase from R4 bacteriophage, isolated from *S. albus*, has not been adapted as a tool for insertion of genes in actinomycetes, however it has been shown to be a useful tool for engineering human cells [90]. The recombinase from *S. venezuelae* phage SV1 was the last one to be adapted into integrative vectors and it successfully integrated heterologous DNA in several streptomycetes, including *S. coelicolor* and *S. lividans* [91]. TG1 bacteriophage was isolated from *Streptomyces cattleya* and its recombinase was adapted to integrate target DNA in different *Streptomyces* strains with high rate of success [92]. ϕBT1 and ϕC31 are related phages, however the recombinases of these phages target different *attB* sites [93]. ϕBT1 and ϕC31 *int/attP* have been widely used by the *Streptomyces*

community to integrated exogenous DNA in several streptomycetes [91]. ϕC31 recombinase was the first to be adapted to integrative vectors by the researchers at Eli Lilley group back in 1991 [94]. It is not therefore surprisingly, that ϕC31-based vectors have been used widely to integrate DNA in streptomycetes [90]. Recently Fayed et al. demonstrated the versatility of these recombinases inserting the erythromycin gene cluster in *S. coelicolor* and *S. lividans*, thus resulting in production of erythromycin in these hosts [95]. Considering that TG1, ϕBT1, SV1 and ϕC31 integrases have different *attP* sites and target different *attB* sites, Smith and colleagues divided the erythromycin gene cluster in four sections, thus inserting *eryAI*, *eryAII*, and *eryAIII* genes into TG1 *int/attP*, SV1 *int/attP*, and ϕBT1 *int/attP*-containing plasmids, respectively, while tailoring genes were inserted into the ϕC31 *int/attP*-containing plasmid. These vectors were introduced into *S. coelicolor* and *S. lividans* and the recombinant strains were able to produce erythromycin, confirming efficiency of phage recombinase to stable introduce DNA into streptomycetes [95].

7.8.2 Inducible promoters developed for streptomycetes

The use of stronger or inducible promoters could improve protein production in *S. rimosus* host. The constitutive expression of proteins or genes may cause a reduction of the specific growth rate, resulting in reduced overall yield of the desired product. This potential issue can be avoided by the development of inducible promoters, thus suppressing the expression of target gene, until the culture reaches high cell density. This way, at the stage when culture reaches sufficient cell density, the promoter is induced, thus initiating the expression of the target protein. To date, based on the literature information, at least five regulatory systems have been successfully adapted in actinomycetes: the thiostrepton *PtipA* promoter from *S. lividans* [96], the nitrilase promoter *PnitA-NitR* from *Rhodococcus rhodochrous* [97], the tetracycline repressor/operator (*TetR/tetO*) system from transposon Tn10 [98] and most recently the resorcinol- and cumate-inducible gene switches from *Corynebacterium glutamicum* and *Pseudomonas putida* respectively [99]. The thiostrepton *PtipA* promoter displayed low levels of induction when induced with thiostrepton antibiotic in *S. rimosus* [76]. In addition, thiostrepton is the selective markers used in *S. rimosus* plasmids, thus *PtipA* promoter is not most suitable regulatory system for *S. rimosus* host. The resorcinol and the cumate regulatory systems were evaluated in *S. albus* and despite these promoters proved to be dose dependent by the addition of cumate and resorcinol, only cumate promoter was active in *S. lividans* and *S. erythraea* [99], therefore their versatility in other actinomycetes is always questionable. The nitrilase promoter *PnitA-NitR* (Fig. 14) from *Rhodococcus rhodochrous* is induced by ε-caprolactam, an inexpensive and environmentally safe inducer [97]. *PnitA-NitR* promoter/regulator system includes the *PnitA* promoter and the transcriptional positive regulator *NitR*. In the presence of ε-caprolactam, the transcriptional factor NitR binds to the *PnitA* promoter activating the transcription of the genes downstream of *PnitA* promoter [97]. The efficiency of *PnitA-NitR* promoter/regulator system was evaluated with the heterologous expression of two reporter genes (isonitrile hydratase *inhA* and catechol 2,3-dioxygenase *xylE*) and nitrilase gene (*nitA*) in *S. lividans*, *S. coelicolor*, *S. avermitilis* and *S. griseus* using a multicopy plasmid based on pIJ101 replicon (up to 300 per cell) [97]. Herai et al.

reported that *PnitA-NitR* promoter/regulator system displayed suppressed basal expression and addition of ε-caprolactam induced expression of the target protein, yielding levels as high as ~40% of all soluble protein in *S. lividans* TK24 [97], therefore this promoter could also be functional in *S. rimosus* considering it reported activity in several streptomycetes.

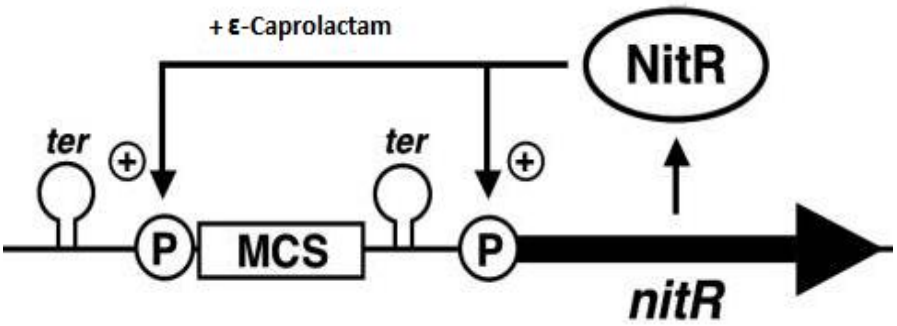


Figure 14. Schematic presentation of *PnitA-NitR* promoter/regulator system [97]. P: *PnitA* promoter; *ter*: terminator; MCS: multi-cloning site.

The tetracycline (TC) inducible repressor (*tetR*) operator (*tetO*) system from the *E. coli* transposon Tn*10* was also successfully adapted for expression of single genes encoding proteins and complete gene clusters in *S. coelicolor* [98, 100]. In *E. coli* the Tet repressor protein (*TetR*) inhibits transcription of the resistance protein in the absence of tetracycline by binding to *tetO* sequences present in the promoter region of the resistance gene. When TetR senses the presence of tetracycline they bind together, the complex TC-TetR diminishes the affinity of the repressor protein to the *tetO* sequences by 9 orders of magnitude resulting in expression of the resistance gene [101]. The tetracycline promoter was adapted to *S. coelicolor* using tetracycline as inducer, thus when TC is added, the affinity of the repressor TetR for *tetO* is reduced resulting in the induced state of tetracycline promoter [98]. To adapt the tetracycline promoter for protein expression in *S. coelicolor* three modification were carried out to the original (TC) repressor/operator (*TetR/tetO*) system. First, the *tetR* gene was codon optimized for efficient expression in *Streptomyces* spp. and it was renamed '*tetRiS*', which stands for *tetR* adapted for expression in *Streptomyces* [98]. Second, *tetRiS* was designed to be under the regulation of the constitutive promoter *SF14* [102] and third, the tetracycline promoters (*tcp*) were constructed containing promoter elements from *PermE** promoter and two or three operators using both, *tetO1* and *tetO2*, sequences [98]. Following this architecture, 6 tetracycline promoters were designed and named *tcp* promoters [98]. The strength of *tcp* promoters was evaluated using *luxAB* reporter system expressing luciferase, and the *neo* gene conferring kanamycin resistance [98]. The synthetic promoters displayed different levels of expression in the repressed and derepressed states. The synthetic tetracycline promoter *tcp830* (Fig. 15) displayed the highest difference in the levels of kanamycin resistance between induced and un-induced states, and it was among the strongest promoters of this group [98]. Interestingly, *tcp830* promoter alone without *tetRiS* showed to be an inducible system in *S. coelicolor, S. avermitilis*, *S. ambofaciens* and *S. lividans* indicating that these strains contain a functional *tetRiS* homologue and an innate TC-controllable promoter, while in *S. roseosporus*, *S. griseus* and *S. venezuelae* strains, the *tcp830* promoter was

constitutively active, even in the presence of *tetRiS*. [98]. The *tcp830* promoter was also used for the transcription of entire genetic clusters in *S. coelicolor* [100]. It successfully transcribed the biosynthetic gene cluster of novobiocin (a single 18-kb polycistronic mRNA) demonstrating the robustness of this genetic tool. The result obtained of *tcp830* promoter in different streptomycetes strains indicates its potential use in *S. rimosus* host to express heterologous proteins.

Figure 15. Schematic presentation of the DNA sequence of *tcp830* promoter. Blue boxes: −10 and −35 elements from *PermE** promoter; Green box: *tet01*; Orange box: *tet02*; Red box: Ribosome binding site. Tetracycline operator (*tetO*) sequences are recognized by the tetracycline repressor (TetR). When tetracycline is present, TetR binds to TC, inducing transcription of *tcp830* promoter.

7.8.3 Replicative plasmids used in streptomycetes

In addition to stronger promoters, the use of multicopy vectors could potentially increase the yield of target recombinant protein in *S. rimosus*. Multiple copies of the gene encoding the recombinant protein can potentially lead to a subsequent increase in protein expression. Currently there are a limited number of replicative plasmids available for use in *Streptomyces* spp., those include the pIJ101 based vectors reported to have between 50 to 300 copies per chromosome in *S. lividans* [103-105], the pJV1 derivative vectors with 150 copies per chromosome reported in *S. phaeochromogenes* [106], the pSG5 based vectors with 20 to 50 copies per chromosome reported in *S. ghanaensis* [107] and SCP2 plasmid from *S. coelicolor*, a conjugative low copy number plasmid (up to 4 copies per cell) of 31 kb [108]. These plasmids and their derivatives replicate by rolling-circle mechanism and have all the required machinery to be transferred through conjugation from *E. coli* to *Streptomyces* spp. with the exception of SCP2 plasmid, which is not replicating by a rolling-circle mechanism [104, 106-108]. pJV1 plasmid shares the replication proteins and organization with pIJ101 vector and it has not been used for heterologous protein expression in *Streptomyces* spp. [109]. The pSG5 plasmid differs in the genetic organization and transfer functions from pIJ101 and pJV1 plasmids, and due to its naturally temperature-sensitive replication, pSG5 derivatives have been used as suicide vectors for transposon delivery, gene disruption and replacement experiments [107]. Only the pIJ101 based vectors have been used to produce recombinant proteins in *Streptomyces* spp. displaying different degrees of stability in different host backgrounds. For example, nitrilase promoter strength was evaluated in *S. lividans*, *S. coelicolor*, *S. avermitilis* and *S. griseus* using the pSH19 vector, a derivative version of pIJ101 [97]. Thiostrepton promoter strength was evaluated in *S. coelicolor* and *S. lividans* using the high copy vector pIJ6021, a pIJ101 based vector [96]. Despite the observed stability of pIJ101-based vectors found in several streptomycetes, in *S. rimosus* these plasmids often proved to be unstable. *S. rimosus* transconjugants harboring pIJ101 based vectors displayed a pock phenotype

similar to lethal zygosis due its conjugative elements [103]. Introduction of pIJ303 plasmid, a pIJ101 derivative, into *S. rimosus* didn't produce the pock phenotype, however, after plasmid isolation, pIJ303 plasmid couldn't be recovered and instead, derivatives of deleted plasmid were obtained [103]. The pPZ12 vector, a derivative of pIJ303 formed by *in-vivo* deletion, was selected as the progenitor vector for plasmid development by Pfizer group based on its relatively small size (6.2 kb) and its non-conjugative nature [110]. The pPZ74 plasmid was constructed based on pPZ12 plasmid, which is a bifunctional cosmid vector used to construct a genomic library of *S. rimosus* [111]. Despite the structural stability of pPZ74 plasmid in *S. rimosus*, this vector was only used to clone streptomycetes genetic fragments into *E. coli* for sequencing and mapping purposes [111], therefore the plasmid was never tested to express heterologous proteins in *S. rimosus*.

The incorporation of inducible promoters and stable replicative plasmids to *S. rimosus* genetic tool kit could improve the abilities of this microorganism as a heterologous host for protein production, thus significantly increasing applicability and versatility of this expression system. Based on the literature, the inducible promoters *PnitA-NitR* and *tcp830* seem to be efficient promoters in a wide range of *Streptomyces* hosts, and the pPZ12 plasmid proved to be stable in *S. rimosus*. Therefore, the incorporation of strong inducible promoters potentially would enhance gene expression levels in *S. rimosus* host. Besides, a stable replicative plasmid would increase the copy number of a target gene, thus leading to increase in protein expression.

7.9 Phytase AppA as industrially important model system for evaluation of *Streptomyces rimosus* host for heterologous protein production

Phytase is important enzyme used by the agricultural industry. Phytase is a phosphatase that catalyzes the hydrolysis of phytic acid releasing inorganic phosphate and myoinositol [112]. Phytic acid or phytate is the primary storage form of phosphorous in seeds of cereals and legumes, however phosphorous in this from can't be assimilated by monogastric animals [113]. Phytase besides allowing better phosphorus uptake by pigs and poultry from phytate, it also contributes to reduce phosphorus pollution in areas of intensive animal production [114]. The phytase (AppA) isolated from *E. coli* has shown to have the highest activity among the different recombinant phytases evaluated [115]. It is a periplasmic enzyme encoded by the gene *appA* with a molar mass of 47 kDa [116]. Of particular interest is the presence of four nonconsecutive disulfide bonds essential for the correct folding of AppA [117]. Despite the fact that *E. coli* is the natural hots of AppA, its overexpression in this host resulted in the production of inclusion bodies [116]. *E. coli*, the best studied heterologous host, accumulates the recombinant proteins in the cytoplasm and the reduced state of this solution prevents formation of disulfide bridges, therefore proteins with disulfide bonds often are not folded correctly, leading to the production of inclusion bodies by *E. coli* [118]. Production of soluble disulfide-bonded proteins in *E. coli* is compartmentalized to the periplasm, a region poorly adapted to maintain proteins in high yields [119]. Consequently, AppA is purified from the periplasm of *E. coli*, a tedious and labor intensive task [114]. Briefly, the *E. coli* cells are cultured for 24 hours

supplemented with glucose and trace elements, when the OD_{675} reaches 36, phytase gene expression is induced by changing the carbon source from glucose to lactose [116]. Once the biosynthetic process is completed, AppA is isolated and concentrated from the periplasm and purified further by multiple chromatographic steps, thus increasing the production costs [116]. To facilitate phytase purification in *E. coli*, Miksch et al. managed to produce AppA as an extracellular enzyme and purify it in a single chromatographic step, the His-tagged phytase was produced by fed-batch fermentation but this time the cultivation time had to be extended for 72 hours, reaching 33.2 U/mL of extracellular phytase activity produced by *E. coli* [114]. Alternative hosts have been evaluated to produce AppA: *Pichia pastoris* and *S. lividans* [120]. These microorganisms are well documented to secrete proteins, thus avoiding the reduced environment of the cytoplasm increasing the chances to produce correctly folded phytase. Although both microorganisms effectively produce the recombinant phytase, both presented some limitations. *P. pastoris* produced 237 U/mL of AppA, however the protein was hyper-glycosylated [115]. In contrast, *S. lividans* produced the recombinant protein of desired quality, but at relatively low yields (0.95 U/mL) [120].

Considering that phytase AppA has been heterologously produced in the past by different research groups, phytase offers an interesting model system to evaluate the capacity of *S. rimosus* to produce recombinant proteins. In addition, besides its industrial importance, AppA has been produced by several industrial hosts. During these attempts, the post-translational modifications (disulfide bonds) often presented an obstacle to produce this enzyme intracellularly. Therefore, secretion of AppA out of the cells is required to produce this enzyme in its correctly folded and active form. Consequently, phytase will allow us to evaluate the capacity of *S. rimosus* to export and produce AppA and compare yields with other hosts.

8 Aim of this study

Many applications in biotechnology and basic research require the purification of highly purified soluble proteins and/or high yields of active and correctly folded proteins, to either carry out biochemical analysis, or for their direct use in therapeutic, diagnostic and industrial applications. In general, ensuring high yields of active recombinant proteins is often difficult and unpredictable with currently available expression systems [119]. The necessity of sufficient quantity and quality of a target product of acceptable price demands continuous optimization of the existing and/or the development of new expression systems and bioprocesses. Although, a variety of expression systems for the production and biosynthesis of heterologous proteins have been developed, to date, there is no universal expression system, which would fulfill requirements for production of every heterologous enzyme at industrial settings.

S. rimosus is a potentially useful heterologous host for production of secondary metabolites and heterologous proteins, considering that it offers several attractive advantages such as:

1. Rapid growth rate.
2. The culture can reach high cell density.
3. Rapid genetic manipulation available.
4. It is generally regarded as safe (GRAS).
5. Abundant supply of common biosynthetic building blocks involved in the biosynthesis of secondary metabolites.
6. Numerous economically affordable and industrially useful carbon and nitrogen sources can be used.
7. Protein secretion capacity, thus facilitating protein folding and downstream procedures (protein purification).
8. Availability of industrial media and bioprocess parameters at industrial level.

The development of an efficient and diverse set of genetic tools to ensure production of heterologous proteins in *S. rimosus* is of high importance, when considering the development of *S. rimosus* based heterologous protein production platform. In addition to the existing integrative vectors, we intend to develop a stable replicative vector for this host, and to build up and evaluate the most interesting promoters described in scientific literature. The potential use of *S. rimosus* as heterologous host was already explored with the production of BTA hydrolase [76], catechol 2,3-dioxygenase and partially also with phytase [78]. However, the yields achieved were relatively low, compared to the commercial expression systems such as *E. coli* or *P. pastoris*. To improve significantly protein production system based on *S. rimosus* host, it is important to improve all components of *S. rimosus* expression system, and to improve our understanding of certain biological processes such as gene regulation and export systems of this streptomycete. This includes evaluation of all component of the expression system including vectors, promoters, signal sequences and media optimization, to achieve the most optimal results.

Additionally, the transcriptome of *S. rimosus* in two substantially different cultivation media will be studied. The transcriptome will likely also bring valuable information about the genetic potential of this host, hopefully making it possible to identify the genes required for the efficient expression and secretion of recombinant proteins (e.g. Tat and Sec-pathways), thus improving the efficiency of *S. rimosus* as heterologous host for protein production. To conduct all these studies, catechol 1,2-dioxygenase XylE enzyme and commercial phytase AppA from *E. coli* will be used as model proteins.

9 Hypothesis

Based on the listed arguments, described in the introduction section, our hypothesis states that the bacterium *S. rimosus* is a convenient host for the development of a novel expression system for production of recombinant proteins. We expect that the incorporation of a replicative vector and more efficient and robust promoters, coupled with suitable signal sequences into the heterologous expression platform of *S. rimosus* will improve the production of target heterologous protein.

10 Materials and methods

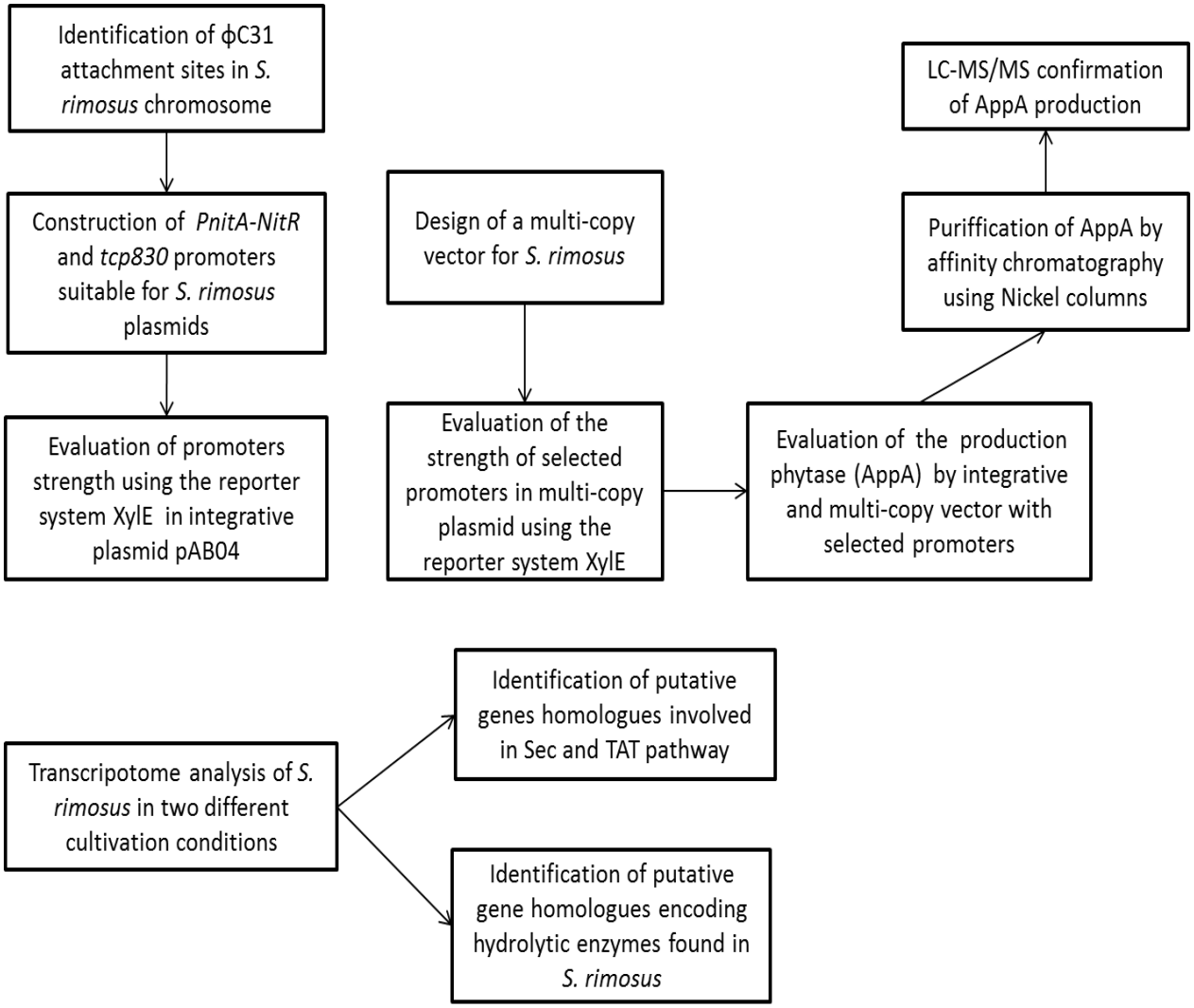


Figure 16. Experiment flow chart.

10.1 Cultivation media

Medium 2TY

1.6% tryptone, 1% yeast extract, 0.5% Sodium chloride.
2% agar (*Optional: If making 2TY plates*)

1. Dissolve 16 g tryptone, 10 g yeast extract and 5 g NaCl in 900 ml dH_2O.
2. Adjust pH to 7.5 and complete to 1 L with dH_2O.
3. *Optional*: Add 20 g of agar if making 2TY plates.
4. Autoclave for 20 minutes at 121^{o}C.

Medium TSB (TRYPTIC SOY BROTH)

1.7% tryptone, 0.3% soy bean, 0.5% Sodium chloride, 0.25% Dipotassium Phosphate, 0.25% glucose. 2% agar (*Optional: If making TSB plates*)

1. Dissolve 17 g tryptone, 3 g soy bean, 5 g NaCl, 2.5 g K_2HPO_4 and 2.5 g glucose in 900 dH_2O
2. Adjust pH to 7.3 and complete to 1 L with dH_2O
3. *Optional*: Add 20 g of agar if making TSB plates.
4. Autoclave for 20 minutes at 121^oC.

Medium SM (soya-mannitol agar)

2% soy meal, 2% D-mannitol and 2% agar.

1. Dissolve 20 g soy meal and 20 g D-mannitol in 900 mL dH_2O.
2. Adjust pH to 7 and complete to 1 L with dH_2O.
3. In 500 mL flask add 5 g of agar and pour 250 mL of the pH 7 solution.
4. Autoclave for 30 minutes at 121^oC.
5. Wait to 90^oC to open flasks.

Seed medium for OTC production

1.5% tryptone, 1% glucose, 0.1% calcium carbonate and 0.5% yeast extract.

1. Dissolve 15 g tryptone, 10 g glucose, 1 g $CaCO_3$ and 5 g yeast extract in 1 L dH_2O.
2. Autoclave for 20 minutes at 121^oC.

OTC production medium (complex medium)

0.7% MOPS, 4.2% soy meal, 0.6% ammonium sulfate, 0.2% magnesium chloride, 0.15% sodium chloride, 2.8% corn starch, 0.01% Zinc sulfate, 0.00375% manganese sulfate and 2% soya oil.

1. Dissolve 7 g MOPS, 42 g soy meal, 6 g $(NH_4)_2SO_4$, 3.77 g $MgCl_2 \bullet 6H_2O$, 1.5 g NaCl, 7.3 g $CaCO_3$ and 28 g corn starch in 900 mL dH_2O.
2. Add 10 mL $ZnSO_4$ (1% solution) and 3.75 mL $MnSO_4$ (1% solution).
3. Heat the medium to 80^oC stirring all the time.
4. Cool it down to room temperature, fix pH 6.25 and complete to 1 L with dH_2O.
5. Add 0.1 mL soya oil in 5 mL (falcon) or 1 mL soya oil in 50 mL (Erlenmeyer flask)
6. Autoclave for 30 minutes at 121^oC.

10.2 Strains used during this PhD project

The strains used in this work are listed below.

10.2.1 *Escherichia coli* strains

DH10β: strain used for cloning and propagation of plasmids. F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara, leu*)7697 *gal*U *gal*K λ- *rps*L *nup*G.

ET12567: methylation-deficient *E. coli* strain (without *dam* and *dcm* methylases). Genotype: F-dam-13::Tn9, *dcm-6*, *hsdM*, *hsdR*, recF143::Tn1I, *galK2*, *galT22, ara14, lacY1, xyl5, leuB6, thi-1, tonA31, rpsL136, hisG4, tsx-78, mtl-1 glnV44*

10.2.2 *Streptomyces rimosus* strains

R7: Wild type strain [121].

M4018: Pfizer strain [122].

"white" strain: OTC non-producing strain (I. S. Hunter, University of Strathclyde, Glasgow).

10.3 Antibiotics and inducers used in this study

Table 7. Antibiotics and indicators used in this study, their final concentrations as selection markers in selected media and stock solution concentration.

Antibiotic or inducer	Selective concentration	Stock solution
Ampicillin	100 µg/ml	100 mg/ml in dH_2O
Apramacyn	75 µg/ml	100 mg/ml in dH_2O
Thiostrepton	30 µg/ml solid media 5 µg/ml liquid media	100 mg/ml in DMSO
Chloramphenicol	10 µg/ml	20 mg/ml in 100% ethanol
ε-caprolactam	0.1% wt/vol	10 wt/vol in dH_2O
Tetracycline	1ug/mL	5 mg/ml in 70% Ethanol

The stock solutions were filter sterilized with a 0.22µm filter and stored at -20^{o}C. Stock solutions were added to molten agar, previously cooled to 55^{o}C.

10.4 Growth conditions for cultivation of *Streptomyces rimosus* strains

Liquid cultures were started in TSB from 1% v/v spore suspensions supplemented with thiostrepton, when required, and cultivated overnight at 30^{o}C and 220 rpm.

10.5 Preservation of *Streptomyces rimosus* strains

Spores were propagated on SM medium supplemented with thiostrepton, when required, at 30^{0}C. After incubation for 8 to 10 days, the spores were collected and frozen at -80^{0}C in 20% glycerol.

10.6 Preparation of *Streptomyces rimosus* electrocompetent cells

Electrocompetent cells of *S. rimosus* were prepared according Pigac and Schrempf [60] with small modifications. Briefly, 1% v/v spore suspension was inoculated in 5 mL TSB and cultivated for 20-24 hours at 30^{o}C and 220 rpm. 100 mL TSB were inoculated with 1 mL of the overnight culture and after 24 hours of incubation the culture was transferred into two falcon tubes and left on ice for 20 minutes. Once the cell growth was stopped, the cells were centrifuged at 10000 rpm for 15 minutes at 4^{o}C and the supernatant was discarded. The mycelium was washed twice with ice-cold 10.3% sucrose (100 mL and 50 ml respectively) and once with ice-cold 15% glycerol. Washing steps were followed by centrifugation at 10000 rpm for 15 minutes at 4^{o}C. After the last washing step, the mycelium was re-suspended in 10 mL 15% glycerol supplemented with lysozyme (400 µg/mL) and incubated for 40 minutes at 37^{o}C. The cells were centrifuged and washed twice with ice-cold 15% glycerol and resuspended once again in 3 mL 30% (w/v) PEG 1000, 10% glycerol and 6,5% sucrose solution. The suspension was distributed in 50 µL aliquots and stored at -80^{o}C. The electrocompetent cells retained their competence over several months without losing their transformation efficiency.

10.7 Transformation of *Streptomyces rimosus* with plasmid DNA

An aliquot of electrocompetent cells was thawed on ice and mixed gently with 10 µl plasmid DNA (1-4 µg). The mixture was transferred to a 0.2 cm electroporation cuvette, which was subject to an electric pulse of 2 kV. The pulsed cells were re-suspended in 950 µl TSB medium and incubated at 30^{o}C for 3 hours before being spread on TSB plates supplemented with thiostrepton. Transformants started to appear after 48 hours.

10.8 DNA manipulation

The general molecular biology techniques were carried out as described in section 1.4.6.

10.8.1 Construction of *PnitA-NitR* and *tcp830* promoters

Nitrilase *PnitA-NitR* and tetracycline *tcp830* promoters were synthesized by Eurofins (Luxembourg) and delivered in pEX based vectors. Both promoters were synthesized according to published data [97, 98]. The multi-cloning site (MCS) was designed specifically to accommodate different signal sequences and the gene of interest, and additional restriction sites were located flanking the promoters to facilitate further cloning to shuttle vectors (Fig. 20).

10.8.2 Genomic DNA isolation from *Streptomyces rimosus*

Genomic DNA was isolated according to the manufacturer's instruction from gram-positive bacteria using GenElute Bacterial Genomic DNA Kit Protocol from Sigma-Aldrich (United States). Biomass was obtained from overnight cultures of *S. rimosus* in liquid TSB medium supplemented with thiostrepton when required.

10.8.3 Plasmid DNA Isolation from *Streptomyces rimosus*

Plasmid DNA was isolated according to the manufacturer's instruction from gram-positive bacteria using GeneJET Plasmid Miniprep Kit from Thermo Fischer Scientific (United States). Biomass was obtained from overnight cultures of *S. rimosus* in liquid TSB medium supplemented with thiostrepton.

10.8.4 Primers

Table 8. Primers used in this study were purchased at Sigma-Aldrich (United States) and are listed in the following table.

Primer name	Sequence (5´ to 3´)	Amplicon
Fw-srT-appA-NdeI	AAAAACATATGTTGCAGAGCTACCTGAAGCACCT	appA His-tag
Rv-appA-XbaI	AAAAATCTAGAGTCAGTGGTGGTGGTGGTGGTGGAGG GAGCACGCCGGGATGC	appA His-tag
Fw P1 P.res	CACCGCGGCTTCGAGACC	Plasmid rescue
Rv P1 P.res	CTCCGGCCTCAGTGGCCGT	Plasmid rescue
Fw P2 P.res	CACGACGAAGGCGCTGATGC	Plasmid rescue
Rv P2 P.res	TCACGTATCCCGCGTGTCCTG	Plasmid rescue

10.8.5 DNA sequencing

DNA Sequencing was carried out by Macrogen (South Korea). Sequence analysis was performed with the software codon code aligner and BLAST (Basic Local Alignment Search Tool) via NCBI web site.

Table 9. Sequencing primers used in this study.

Primer	Sequence (5' to 3')	MCS of vector
pEX-For	GGAGCAGACAAGCCCGTCAGG	pEX series vectors
pEX-Rev	CAGGCT TTACACTTTATGCTTCCGGC	pEX series vectors
M13F-pUC	GTTTTCCCAGTCACGAC	pUC19 and pVF
M13R-pUC	CAGGAAACAGCTATGAC	pUC19 and pVF
pVF-EcoRI-RV	AGCGGATCAAGGGGTTTG	pVF
pAB04 Fw	CTATTACGCCAGCTGGCGAA	pAB04
pAB04 Rv	CTCACTCATTAGGCACCCCA	pAB04
ϕC31 fish Rv	CGTTGGCGCTACGCTGTGT	Plasmid rescue

Table 10. Primers used to sequence pVF vector.

Primer name	Sequence (5´ To 3´)
pVF_Fw_fw_15	CACCGAGTCCCACGCCA
pVF_Fw_fw_13	GACGCCTTCCGCGAT
pVF_pUC4	GTAACGCCAGGGTTTTCCCAGT
pVF_pUC2	GCTTATCAGTGAGGCACCT
pVF_pUC1	CGACACGGAAATGTTGAATACTCATAC
pVF_pUC2	GTATATATGAGTAAACTTGGTCTGACAG
pVF10	ACCCCCGACAGCGGATC
pVF_pUC3	CTGTAGGTATCTCAGTTCGGTGTA
pVF_14	CTAAATACATTCAAATATGTATCCGCTCAT
pVF_15	GCAGTGCTGCCATAACCATGAGT
pVF_Fw_fw_19	GCCGCGGGAGTAATCCT
pVF13	GCCCGGCTCGCAAGT
pVF_16	ACTCAAGACGATAGTTACCGGATAAG
pVF11	GATCGTCGCCACCTTCGA
pVF_Fw_fw_17	ACCGCAATTGCCCACACAC
pVF12	TACTCGTGCCAGCGCGAG

10.8.6 Vectors

Table 11. List vectors used in this study.

Plasmid name	Selective marker	Promoter	ORF	Microorganism	Source
pUC19	Amp	----	----	*E. coli*	Thermo fisher
pYAC-OTC	Amp/Ts	*OTC*	*OTC*	*E. coli/S. rimosus*	Acies Bio
pYAC	Amp/Ts	----	----	*E. coli/S. rimosus*	Acies Bio
pEX-*PnitA*	Amp	*PnitA-NitR*	----	*E. coli*	Eurofins
pEX-*PnitA*-1	Amp	*PnitA-NitR*	*xylE*	*E. coli*	This study
pEX-*PnitA*-2	Amp	*PnitA-NitR*	*srT/appA*	*E. coli*	This study
pEX-tcp830	Amp	*tcp830*	----	*E. coli*	Eurofins
pEX-tcp830-1	Amp	*tcp830*	*xylE*	*E. coli*	This study

pEX-tcp830-1/*tetRiS*	Amp	*tcp830*	*xylE*	*E. coli*	This study
pEX-tcp830-2	Amp	*tcp830*	*srT/appA*	*E. coli*	This study
pAB04-1	Apr/Ts	*PermE**	*xylE*	*E. coli/S. rimosus*	Dr. Magdveska
pAB04-2	Apr/Ts	*PnitA-NitR*	*xylE*	*E. coli/S. rimosus*	This study
pAB04-3	Apr/Ts	*tcp830*	*xylE*	*E. coli/S. rimosus*	This study
pAB04-4	Apr/Ts	*PermE**	*srT/appA*	*E. coli/S. rimosus*	Dr. Magdevska
pPZ12	Ts	----	----	*S. rimosus*	Dr. Magdevska
pVF	Amp/Ts	----	----	*E. coli/S. rimosus*	This study
pVF-1	Amp/Ts	*PermE**	*xylE*	*E. coli/S. rimosus*	This study
pVF-2	Amp/Ts	*PnitA-NitR*	*xylE*	*E. coli/S. rimosus*	This study
pVF-3	Amp/Ts	*tcp830*	*xylE*	*E. coli/S. rimosus*	This study
pVF-3/*tetRiS*	Amp/Ts	*tcp830*	*xylE*	*E. coli/S. rimosus*	This study
pVF-4	Amp/Ts	*PnitA-NitR*	*srT/appA*	*E. coli/S. rimosus*	This study
pVF-5	Amp/Ts	*tcp830*	*srT/appA*	*E. coli/S. rimosus*	This study

Construction of pVF vector:

pVF vector was constructed using pPZ12 [123] and pUC19 (Invitrogen) vectors. Both plasmids were digested with *PstI*, gel purified and ligated to produce pVF vector. The plasmid pVF was transformed and isolated from *S. rimosus* and sequenced through primer walking (Primers are listed in table. 10).

Vectors containing *xylE* reporter gene:

Nitrilase *PnitA-NitR* and tetracycline *tcp830* promoters were delivered in pEX-based vector. Both vectors (pEX-*PnitA* and pEX-*tcp830*) were digested with *NdeI/XbaI* to accommodate *xylE* gene, previously isolated from pAB04-1 using the same set of enzymes, yielding pEX-*PnitA*-1 and pEX-*tcp830*-1. pAB04-1 vector was digested with *EcoRI/XbaI* to purify the cassette containing *PermE*/xylE* which was ligated in pVF vector previously treated with the same set of enzymes generating pVF-1. To produce pAB04-2, pEX-*PnitA*-1 was digested with *HindIII* and the cassette *PnitA/xylE/nitR* was gel purified, end-filled with klenow and cloned in pAB04, previously digested via *EcorI/ XbaI*, end filled with klenow and dephosphorylated. Further, pEX-*PnitA*-1 and pVF were both digested with *HindIII* and the cassette *PnitA/xylE/nitR* was ligated in pVF, yielding pVF-2. For *tcp830* promoter, pEX-*tcp830*-1 was digested with *EcoRI* and the cassette *tcp830*/*xylE* was ligated in pAB04 and pVF vectors previously treated with *EcoRI*, yielding pAB04-3 and pVF-3. The transcriptional regulator *tetRiS* was synthesized under the regulation of *SF14* promoter and delivered in pEX based vector. The cassette *SF14-tetRiS* was isolated using *SacI/BamHI*, treated with klenow enzyme and ligated in the vector pEX-*tcp830*-1, previously digested with *SacI*, end filled and dephosphorylated, generating pEX-*tcp830*-1/*tetRiS* vector. The cassette *tcp830/xylE/SF14/tetRiS* was liberated from pEX-*tcp830*-1/*tetRiS* using *KpnI/XbaI*, end filled and ligated in pVF, previously digested with *EcoRI* and end filled and dephosphorylated, to generate pVF-3/*tetRiS*.

Vectors containing *appA* gene:

Phytase *(appA)* gene was kindly provided by Dr. Magdevska. *appA* gene was previously subject of codon optimization for efficient expression in *S. rimosus* in order to avoid

transcriptional issues considering the difference in GC content between *E. coli* K12, natural host of phytase *appA* gene, and *S. rimosus*. We fused the signal sequence of a putative serine like protease (*srT*) encoded by *S. rimosus* genome to *appA* [78]. The cassette *srT*/*appA* was PCR amplified from pAB04-4 using the primers Fw-srT-*appA*-*NdeI* and Rv-*appA*-*XbaI*, which incorporated *NdeI* restriction site at 5' position and a His-tag and *XbaI* at 3' position. After gel purification and restriction with *NdeI/XbaI*, the *srT/appA* cassette was ligated in pEX-*PnitA*-*NitR* and pEX-*tcp830*, both digested with *NdeI/XbaI* yielding the plasmids pEX-*PnitA*-2 and pEX-*tcp830*-2. Further, pEX-*PnitA*-2 was treated with *HindIII* and the cassette PnitA/*srt*/*appA*/*nitR* was ligated in pVF vector digested also with *HindIII* producing pVF-4 vector. For pVF-5 an identical ligation approach was followed but *EcoRI* was used instead of *HindIII* in the final cloning step. After sequence confirmation, plasmids were transformed into *S. rimosus*.

10.8.7 Plasmid rescue strategy to identify the *attB* sites in *Streptomyces rimosus* chromosome targeted by the ϕC31 recombinase

With the aim of identifying the ϕC31 *attB* sites present in *S. rimosus* chromosome, the plasmid pYAC-OTC containing the ϕC31 recombinase and the *attP* site was transformed into *S. rimosus* M4018. These transformants were kindly provided by the company Acies Bio Ltd. (Slovenia). The plasmid was rescued from *S. rimosus* chromosome to identify the DNA sequence targeted by the ϕC31 recombinase. Briefly, 1-2 µg gDNA of pYAC-OTC *S. rimosus* transformants and gDNA of R7, M4018 and "white" strains were digested independently overnight with 2 µL *SacI* (2 FDA) in 20 µL reaction. Digested gDNA was precipitated with ethanol as follows, the digested gDNA volume was completed to 200 µL with dH_2O and an equal volume of neutral phenol was added to remove the proteins, the mixture was centrifuged 1 minute at 14000 rpm and the aqueous phase was transferred to a new Eppendorf. 1/10 volume of 3 M sodium acetate pH 5.2 was added to the sample followed by the addition of 2.5 volumes of 100% ethanol. The mixture was incubated at -20°C overnight. Afterwards the samples were centrifuged for 10 minutes and 14000 rpm at 4°C and the supernatant was discarded, the pellet was washed with 70% ethanol, centrifuged again for 5 minutes and 14000 rpm at 4°C. Finally, the ethanol was removed and the digested DNA was resuspended in 15 µL TE buffer. Ligation of digested DNA was carried out overnight at 16°C after addition of 2 µL 10X ligase buffer, 1 µL T4 DNA ligase and completed with dH_2O to 20 µL. 2 µL of ligation mix were transformed into *E. coli* DH10β and colonies were selected against ampicillin. Plasmid DNA was isolated from the colonies obtained and its restriction pattern was analysed with *NcoI/BamHI*. The plasmids displaying different restriction patterns were sequenced to analyse the integration *attB* sites in *S. rimosus* chromosome.

10.8.8 Design of the probes for Southern blot studies to confirm the location of *attB* sites in *Streptomyces rimosus* chromosome

DIG-labeled DNA probes were generating with the PCR DIG Probe Synthesis Kit from Roche (Switzerland). Primers to generate the probes (table 8) were designed according

to the DNA sequence obtained from the rescued plasmids, and rescued plasmids were also used as PCR templates to generate the hybridization probes. Probes were labeled with DIG-dUTP (alkali-labile), by the method of polymerase chain reaction according Roche´s protocol.

10.8.9 Pulsed field gel electrophoresis (PFGE) to establish the restriction map of *Streptomyces rimosus* transformants containing integrated vector pYAC-OTC

PFGE method was adapted from Kieser et al. [49]. The blocks containing genomic DNA of pYAC-OTC transformants and R7, M4018 and "white" *S. rimosus* strains were prepared from overnight cultures grown in 5 mL TSB medium. 2 mL of the 24 hours old culture were centrifuged at 10000 rpm for 5 minutes, the supernatant was discarded and the mycelium was washed twice with 2 mL SucTE buffer (0.3 M sucrose, 25 mM tris pH 8 and 25 mM EDTA pH 8). After the second wash the cells were resuspended in SucTE buffer and diluted (if necessary) to get a final OD_{600}= 1.8-2. 500 µL of mycelial suspension were mixed with an equal volume of 2% low melting point agarose (LMP) previously melted and kept at 50^{o}C. The mixture was poured into the block mould adapter, which was incubated at 4^{o}C for 10 minutes. Once the blocks solidified, they were removed from the mould and incubated overnight at 37^{o}C in lysis solution 1 (SucTE buffer and 2 mg/mL lysozyme) (1 mL/block). The next day the lysis solution 1 was replaced by proteinase K solution (0.5 M EDTA pH 8, 1% N-laurylsarcosine and 2 mg/mL proteinase K) and the blocks were incubated for 24 hours at 50^{o}C. Afterwards, the proteinase K solution was removed and the blocks were washed 4 times 1 hour each with TE buffer (10 mM tris pH 8 and 1 mM EDTA pH) and washed once more 2 times 1 hour each with PMSF solution (1 mM PMSF in TE buffer). After the washing steps, the blocks were ready for treatment with restriction endonucleases or stored at 4^{o}C in 0.5 M EDTA. The blocks were treated with *AseI* restriction endonuclease (1 block in 400 µL 1X fast digest buffer and 3 µL of enzyme) and incubated overnight to avoid partial digestion. PFGE was performed in a CHEF Mapper apparatus from Bio-Rad (United States) in 100 mL 1% LMP agarose gel run in 0.5X TBE buffer (10X TBE buffer: 108 g Tris base, 55 g boric acid and 40 mL EDTA pH 8 dissolved in 1 L dH_2O). The gels were stained with ethidium bromide (10 µL of 10 mg/mL ethidium bromide solution in 200 mL dH_2O) for 20 minutes, washed twice with 200 mL dH_2O and revealed in GEL Doc instrument.

10.8.10 Southern blot using PFGE gels to determine the location of the *attB* sites targeted by the ϕC31 recombinase in the restriction map of *Streptomyces rimosus* chromosome

PFGE gels were analysed by Southern blot with probes previously designed from rescued plasmids in order to identify the location of the *attB* sites specific for ϕC31 recombinase in *S. rimosus* chromosome. The DNA was first transferred from the PFGE gel to a positively charged nylon membrane through vacuum blotting. Briefly, each gel was incubated for 15 minutes in 0.25 M HCl (depurination), and then washed twice with dH_2O; 3 minutes each. The DNA was denatured with 0.5 M NaOH/1 M NH4Ac for

15 minutes and transferred to the nylon membrane previously wet with 0.4 M NaOH. The DNA was transferred using the vacuum blotter 785 from Bio-Rad (United States). After the transfer, the membrane was washed for 15 minutes in 1 M NH4Ac and dried for 20 minutes at 100^{o}C. Membranes were stored at room temperature and ready for hybridization. The DIG Nonradioactive System method from Roche (Switzerland) was used for nucleic acid labeling and detection. Shortly, the membrane was washed for 5 minutes with 25 mL 5X SCC buffer (20X SSC buffer: 3 M NaCl in 0.3M sodium citrate pH 7.0), then the 5X SCC buffer was replaced by pre-heated 60^{o}C Dig easy hyb buffer (20 mL/100 cm^{2} membrane) and the membrane was incubated for 30 minutes with gentle agitation. Meanwhile, the probe and Lambda *HindIII* ladder were denatured at 99^{o}C for 5 minutes and subsequently cooled down on ice for 1 minute. The denatured probe and ladder were mixed with preheated Dig easy hyb buffer (3.5 mL/100 cm^{2} membrane) and after 30 minutes of incubation the dig easy hyb buffer was replaced by the one containing the probe and ladder and left overnight with gentle agitation at 60^{o}C. Next day the buffer was removed and the membrane was washed twice 5 minutes each with 2X SCC buffer and 0.1% SDS (20 mL/100 cm^{2} membrane) at 20^{o}C and washed again twice 15 minutes each with 0.5X SCC buffer and 0.1% SDS (20 mL/100 cm^{2} membrane) at 68^{o}C. The immunological detection was carried out at 20^{o}C, washing the membrane for 5 minutes with 20 mL washing buffer (0.1 M maleic acid, 0.15 M NaCl pH 7.5 and 0.3% v/v tween 20) and incubating it for 30 minutes in blocking solution (10X blocking solution: 10% blocking reagent in 0.1 M maleic acid, 0.15 M NaCl pH 7.5) (100 mL/100 cm^{2} membrane). Afterwards, the membrane was incubated for 20 minutes in antibody solution (Anti-Digoxigenin-AP dissolved blocking solution 150 mU/mL) (20 mL/100 cm^{2} membrane) and washed twice 15 minutes each with 100 mL washing buffer. The membrane was equilibrated for 5 minutes with detection buffer (0.1 M Tris-HCl and 0.1 M NaCl pH 9.5) (20 mL/100 cm^{2} membrane) and revealed with 10 mL color-substrate solution (1 NBT/BCIP tablet in 10 mL detection buffer) in the dark without agitation until the appearance of the bands, the reaction was stopped with dH_2O.

10.9 Heterologous expression of catechol 2,3-dioxygenase XylE and phytase AppA in *Streptomyces rimosus* host

Recombinant strains of *S. rimosus* were cultivated in 5 mL of TSB medium or complex medium to evaluate XylE and AppA expression. Bioprocess was initiated by inoculating 1% v/v of spore suspensions in TSB medium, after 20 hours 10% v/v of the overnight culture was transferred to fresh soluble medium TSB and samples were taken after 20 hours of growth for further analysis. For seed and production media, 1% v/v of spore suspension was inoculated in seed medium, after 20 hours 10% v/v of the overnight culture was transferred to complex medium and samples were taken every 24 hours during 7 days. The culture media were supplemented with thiostrepton (5 µg/mL) when required and cultures were grown at 30^{0}C and 220 rpm. Inducers, ε-caprolactam (0.1% w/v) or tetracycline (1 µg/mL), were added after 4 or 60 hours of the bioprocess in TSB and complex medium respectively. 1 mL samples were taken after bioprocess was completed. Cells were harvested at 4^{o}C and 14000 rpm for 5 minutes. Pellet and supernatant were frozen at -20^{o}C for further enzymatic analysis.

10.10 Enzymatic assays to evaluate XylE activity in *Streptomyces rimosus* culture

Catechol 2,3-dioxygenase XylE activity was assayed as described previously [49]. Briefly, the cell pellet was washed with 1 mL 2 mM potassium phosphate buffer pH 7.5, and after centrifugation for 10 minutes at 14000 rpm, the cells were resuspended in 0.5 mL 100 mM potassium phosphate buffer pH 7.5, 20 mM EDTA and 10% acetone and disrupted by sonication on ice 4 x 15 sec, with 30 sec intervals to avoid foaming. After sonication, 5 µl 10% triton X-100 was added, and the extracts were incubated on ice for 15 min. Cellular debris was removed by centrifugation for 5 min at 14000 rpm at 4^{o}C. To measure XylE activity, 10 µL of the supernatant of the cell extract were mixed with 190 µL assay buffer (100 mM potassium phosphate buffer pH 7.2, 0.2 mM catechol) and incubated for 10 minutes at 37^{o}C. Catechol 2,3-dioxygenase activity was evaluated spectrophotometrically at 375 nm. One mU of XylE activity was calculated as the rate of change in absorbance at 375 nm per minute [49]:

$$\text{mU catechol dioxygenase } [\text{nmol min}^{-1}] = \frac{30.03 \times \Delta A375}{time\ [\text{min}]}$$

Specific activity (mU/mg of total proteins) was calculated as the rate of catechol 2,3-dioxygenase activity per milligram of total proteins [49]:

$$\text{Specific activity } [\text{mU mg}^{-1}] = \frac{\text{mU}}{\text{Protein concentration}}$$

Protein concentration was determined by Bradford method [124]. Briefly, 10 µl of supernatant used to measure XylE activity, was mixed with 200 µl Bradford buffer (4:1 ratio, 4 volumes of distilled water and 1 volume Bradford). After incubation for 5 minutes at room temperature the absorbance was measured at 595 nm. Blanks were made using 10 µL distilled water. The concentration of total proteins, manifested like mg of protein per milliliter supernatant, was determined by calibration curve made following the same procedure using different concentrations of bovine serum albumin as standards. In case the sample's absorbance exceeded the range of calibration curve, dilutions with distilled water were made.

10.11 Enzymatic assays to evaluate phytase AppA activity in *Streptomyces rimosus* culture

Phytase activity was determined using ammonium molybdate method described previously [125] with minor modifications. Briefly, 1 mL of cell culture were spin-down at 14000 rpm for 5 minutes and the cell pellet was sonicated as described before, but using 50 mM Tris pH 7 buffer instead in order to measure intracellular AppA activity. Extracellular AppA activity was deduced directly from supernatant. 30 µl of cells extracts or supernatant (dilutions were made if necessary) were mixed with 120 µl sodium phytate reagent (0.1 M acetic acid buffer pH 5, 2 mM phytic acid sodium salt)

and incubated for 15 minutes at 37^{o}C. The reaction was stopped by adding 150 µL 15% trichloroacetic acid (TCA) and samples were mixed on vortex. Blanks were made by adding TCA to the samples and incubated at 37^{o}C for 15 minutes before adding the sodium phytate reagent. Afterwards, 300 µL of color reagent solution (3:1:1 mixture of 1 M sulfuric acid, 2,5% ammonium molybdate (w/v), and 10% ascorbic acid (w/v) respectively) were added, followed by incubation at 50^{o}C for 15 min. The phosphorous released from phytic acid by AppA was measured spectrophotometrically at 820 nm. One phytase unit (FTU) was defined as the amount of enzyme required to release 1 µmol of inorganic phosphorus (Pi) per minute from sodium phytate at 37^{o}C.

10.12 Purification of secreted phytase AppA from *Streptomyces rimosus* culture using His-Tag/Ni-affinity chromatography

Phytase (AppA) production was scaled up to 200 mL culture. The production process started inoculating 1% v/v spore suspension of the recombinant *S. rimosus* "white" strain harboring pVF-5 (pVF-*tcp830/srT/appA*) replicative plasmid in OTC seed medium supplemented with thiostrepton. The culture was cultivated at 30^{0}C and 220 rpm for 20 hours. 10% v/v of the overnight culture was transferred to complex medium supplemented with thiostrepton and cultivated for 4 days. Once the biosynthetic process was completed, the culture was centrifuged at 10000 rpm. The pellet was discarded and the supernatant was filtered through filter paper to remove cellular debris and medium components, which could block the HisTrap column. The clear supernatant (around 150 mL) was mixed with 200 mL equilibration buffer (20 mM imidazole pH 8, 300 mM NaCl and 50 mM tris pH 7) and loaded into 5 mL HisTrap column previously equilibrated with 5 column volumes of washing buffer (25 mM imidazole pH 8, 500 mM NaCl and 50 mM tris pH 7). The proteins were washed with 25 column volumes of washing buffer and the recombinant protein was eluted isocratically with elution buffer (300 mM NaCL, 250 mM imidazole pH 8 and 50 mM tris pH 7). Fractions containing the recombinant protein were collected and run on SDS-PAGE to confirm the presence of the 45 kDa band belonging to AppA, tested for phytase activity and frozen for further analysis.

10.13 In-gel digestion and mass spectrometric peptide analysis (LC-MS/MS) to confirm the identity of the purified protein AppA from *Streptomyces rimosus* supernatant

In-gel digestion and LC-MS/MS was performed to confirm the identity of the purified phytase (AppA). The protocol was adapted from Shevchenko et al [126]. After SDS-PAGE, the gel was washed twice with dH_2O 10 minutes each time and the band was cut out from the polyacrylamide gel into small cubes and transferred to an Eppendorf tube. The cubes were washed once with dH_2O for 15 minutes and washed again with 1:1 dH_2O:acetonitrile (MeCN) for additional 15 minutes. The liquid phase was replaced with MeCN and the sample was incubated at room temperature until the cubes shrunk. Then the MeCN was removed and the cubes were rehydrated with 0.1 M NH_4HCO_3 for 5 minutes and an equal volume of MeCN was added, and the sample was

incubated for additional 15 minutes. The liquid was removed and the cubes were dried for 1 hour, using a vacuum centrifuge, to later swell them with 10 mM dithiothreitol/0.1 M NH_4HCO_3. The swollen particles were incubated for 45 minutes at 56^{o}C and cooled down at room temperature. Further, the liquid was removed and replaced with the same volume of 55 mM iodoacetamide in 0.1 M NH_4HCO_3 and the sample was left it in the dark for 30 minutes. After the dark incubation period, the iodoacetamide solution was removed and the cubes were washed twice with 0.1 M NH_4HCO_3/ MeCN (0.1 M NH_4HCO_3 for 5 minutes, and then addition of MeCN and incubation for 15 minutes). After the two washes, the liquid was removed and the cubes were shrunk with MeCN, followed by liquid removal and vacuum centrifugation for 1 hour to dry the cubes. To digest the protein, the cubes were rehydrated by adding the chilled digestion buffer (50 mM NH_4HCO_3, 5 mM CaCl2 and 12.5 ng/µL of trypsin) and incubated at 37^{o}C for 45 minutes. Enough 50 mM NH_4HCO_3 was added to cover the cubes if necessary; otherwise the samples were incubated overnight at 37^{o}C. To recover/extract the peptides, one volume of 25 mM NH_4HCO_3 was added to the overnight sample and incubated for 10 minutes at room temperature. Then the same volume of MeCN was added and the sample was incubated for additional 10 minutes. The supernatant was recovered and the extraction was repeated twice but this time using 5% formic acid instead of NH_4HCO_3. Finally, the extracts were dried in a vacuum centrifuge and the peptides were analyzed by LC-MS/MS method.

10.14 *Streptomyces rimosus* transcriptome

The transcriptome of *S. rimosus* M4018 was kindly provided by the biotechnology company Acies Bio Ltd. (Slovenia). The mRNA was isolated at 6 different time points from *S. rimosus* culture cultivated in TSB and complex medium. For TSB medium, the mRNA was analyzed after 9, 13, 16, 20, 24 and 36 hours of cultivation, while for complex medium the mRNA was analyzed after 12, 20, 36, 48, 72 and 96 hours of the bioprocess.

11 Results

11.1 ϕC31 phage site-specific recombinase-mediated integration of heterologous DNA into *Streptomyces rimosus* chromosome

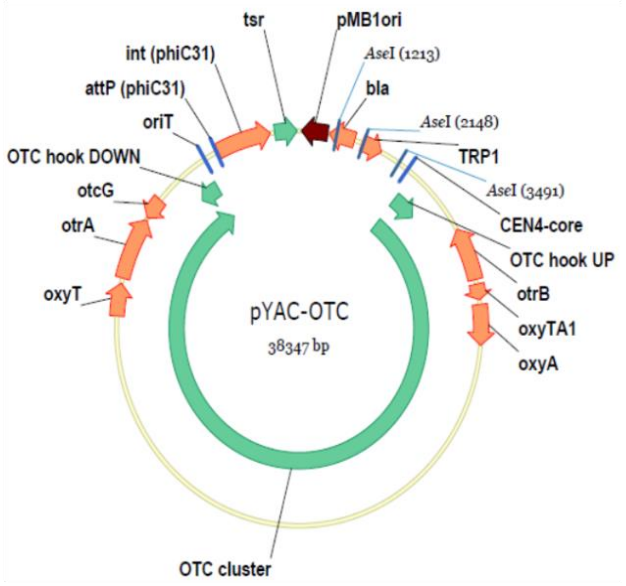
Integrative vectors are most often integrated into the chromosome of the host organism stably, thus avoiding the use of selective markers (e.g. antibiotics) to maintain the plasmid, feature which is highly desired for industrial purposes. The ϕC31 site-specific integrase gene from *Streptomyces* phage ϕC31 has been incorporated into a number of streptomycetes vectors, thus facilitating integration of heterologous DNA into the chromosome of numerous actinomycetes [127]. The integration of plasmid DNA is accomplished by ϕC31 recombinase that catalyzes the site-specific recombination between the phage attachment site *attP* (present in the vector) and the bacterial attachment site *attB* (located in the host chromosome) [92]. ϕC31 recombinase recognizes the short DNA fragment consisting of 51 bp corresponding to the *attB* site and induces recombination between *attP* and *attB* sites, resulting in the hybrid sites *attL* (left) and *attR* (right) [127]. Analysis of the integration mechanism of ϕC31 based vectors into the chromosome of several *Streptomyces* species revealed the presence of multiple *attB* sites, thus allowing ϕC31 based plasmids to integrate in the target chromosome in multiple locations [82]. Thamchaipenet and collaborators identified the DNA sequence of the perfect *attB* site targeted by ϕC31 recombinase in *S. rimosus* R7 strain. However, they did not identify the location and the presence of additional or pseudo *attB* sites in *S. rimosus* genome [128]. At Acies Bio Ltd. (Slovenia), a ϕC31 based plasmid containing the entire OTC cluster was constructed and designated as pYAC-OTC (Fig. 17A), and transformed into *S. rimosus* M4018. Southern blot analysis, where selected OTC genes were DIG-labeled and used as probes, surprisingly revealed integration of pYAC-OTC into multiple sites of *S. rimosus* chromosome. Thus, in addition to the native OTC cluster, often one or even two additional copies of *otc* gene cluster were identified in the chromosome of *S. rimosus* pYAC-OTC containing transformants (unpublished data). The study of expression of the *otc* gene cluster in *S. rimosus* is not in the focus of this thesis. However, considering that we do not really have data on alternative integration sites of ϕC31-based integrative vectors and, considering that we need to evaluate the effect of an alternative integrative site on expression of introduced heterologous genes in *S. rimosus* genome, we decided to identify the location of the *attB* sites using transformants containing multiple copies of the pYAC-OTC plasmid in *S. rimosus* chromosome. *S. rimosus* transformants containing additional copies of pYAC-OTC were kindly provided by Acies Bio Ltd (Slovenia).

Plasmid rescue approach was carried out to determine the DNA sequences of the *attB* sites present in *S. rimosus* genome. Plasmid rescue is a technique used to recover bacterial plasmids integrated in the chromosome of a host organism [129]. In addition to the plasmid itself, this approach allows the recovery of flanking genomic sequences [129], thus allowing identification of the DNA sequences specific for the *attB* sites present in *S. rimosus* pYAC-OTC transformants. In order to complete the plasmid rescue experiment, the DNA integrated plasmid sequence has to contain the origin of replication for *E. coli* plasmid, an antibiotic resistance gene expressed in *E. coli* and a

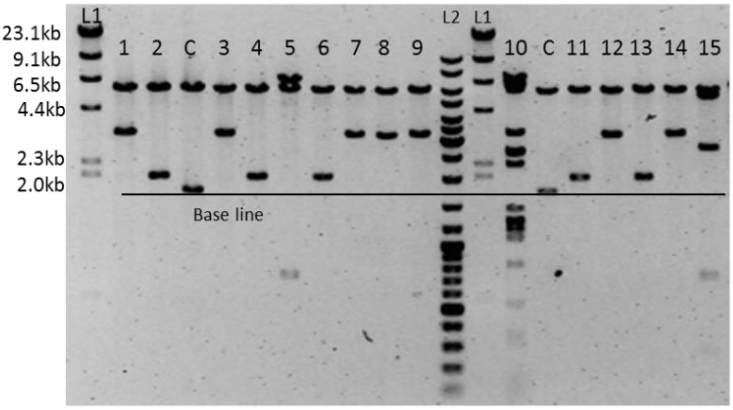
suitable restriction site that cuts in the DNA of the integrated vector, however, located downstream and only on one side of the *E. coli* plasmid replication origin and resistance gene. In addition, the selected "restriction site" has to be relatively common in the genome of the host (in our case *S. rimosus*); so that the DNA fragment that will be re-ligated and re-transformed into *E. coli* does not reach size which is too large and thus, difficult to transform and sequence. Therefore, the rescued fragment with the plasmid contains at least, the origin of replication and resistance gene from the integrated vector and the adjacent upstream DNA sequence of the *attB* sites targeted in *S. rimosus* genome.

Our strategy (Fig. 17C) was designed to rescue the genomic sequence upstream *SacI* restriction site introduced with pYAC-OTC vector, thus the rescued plasmids contained the pYAC plasmid without the *otc* gene cluster, but includes the ampicillin resistance gene, the origin of replication pMB1, the thiostrepton resistance gene, the ϕC31 integrase gene, the *attL* sites and the flanking genomic region upstream the *attB* sites specific for ϕC31 recombinase, thus the DNA sequence continuing until the first *SacI* restriction site located in the genome of *S. rimosus*. This DNA fragment, adjacent to the flanking DNA sequence of ϕC31 recombinase, is actually the DNA sequence that we were interested in, because it allowed us to identify the *attB* sequences, and also let us design hybridization probes which enabled us to locate the *attB* sites in *S. rimosus* chromosome. Restriction pattern of the rescued plasmids displayed two different plasmid architectures, compared to the empty pYAC vector, thus indicating that ϕC31 recombinase integrated in two different loci in *S. rimosus* chromosome (Fig. 17B). The identification of the perfect *attB* and a pseudo *attB* sites was possible through sequencing the rescued plasmids using the ϕC31 fish Rv primer (Table. 9), this oligonucleotide was designed to sequence the DNA region upstream the *int(ϕC31)* gene, thus allowing the identification of the rescued region (Fig. 17C). The DNA sequences obtained with ϕC31 fish Rv primer, using the two different rescued plasmids as templates, were located in *S. rimosus* genome, and Frame plot software enabled us to identify the correct open reading frames (ORFs) targeted by ϕC31 *int/attP* containing plasmid. The perfect *attB* site, containing correct *attB* DNA sequence lies within a putative gene encoding for a pirin-like protein, which was already identified in *S. rimosus* R7 by Phornphisutthimas et al. [128], and among other species of *Streptomyces* including *S. ambofaciens*, *S. aureofaciens*, *S. cinnamonesis*, *S. clavuligerus*, *S. coelicolor*, *S. fradiae*, *S. griseus*, *S. hygroscopicus* and *S. lividans* [82, 94, 130]. The secondary or so called pseudo *attB* site lies within a putative acyl-CoA dehydrogenase gene. Both genes containing the *attB* sites are disrupted when DNA is integrated in these locations. Further analysis of the two rescued DNA sequences allowed the identification of the complete sequence of the perfect *attB* and pseudo *attB* sites. As expected, both *attB* sites contained the core signature sequence 5´TT, region at which cross-over occurs (Fig. 17D) [82, 94].

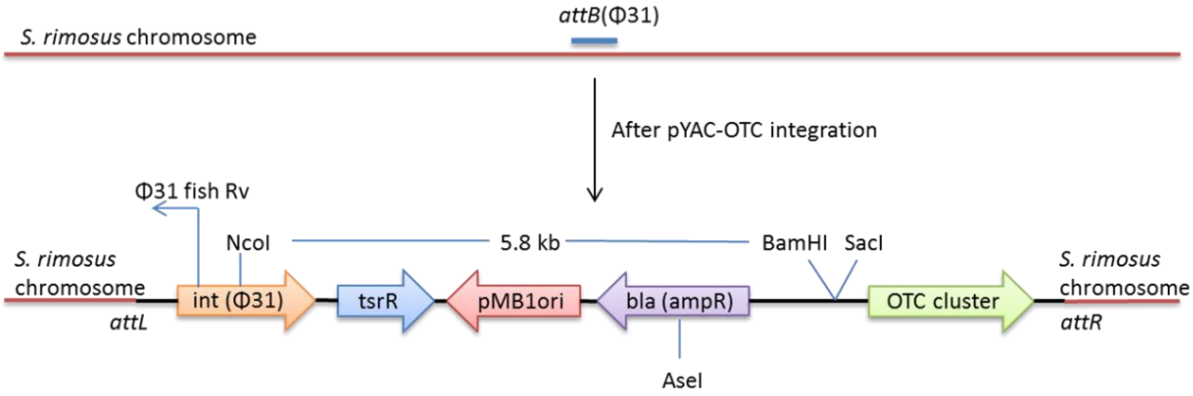
A.



B.



C.



D.

attP ΦC31: GTAGTGCCCCAACTGGGGTAACCTTTGAGTTCTCTCAGTTGGGGGCGTAGG
attB: CGGTGCGGGTGCCAGGGGGTGCCCTTGGGCTCGCCCGGCGCGTACTCCACC
Pseudo *attB*: GAGAAGTCCGACCCGGGCGTCTCCTTCGGCGCCCCGGAGAAGAAGCTCGGC

E.

attL (perfect): CGGTGCGGGTGCCAGGGGGTGCCCTTGAGTTCTCTCAGTTGGGGGCGTAGG
attR (perfect): GTAGTGCCCCAACTGGGGTAACCTTTGGGCTCGCCCGGCGCGTACTCCACC
attL (pseudo): GAGAAGTCCGACCCGGGCGTCTCCTTGAGTTCTCTCAGTTGGGGGCGTAGG
attR (pseudo): GTAGTGCCCCAACTGGGGTAACCTTTCGGCGCCCCGGAGAAGAAGCTCGGC

Figure 17. A. pYAC-OTC plasmid map. **B.** Restriction pattern of pYAC plasmid without *otc* gene cluster and rescued plasmids generated with *BamHI/NcoI*. pYAC vector (7,5kb) digested with *BamHI/NcoI* results in two bands of 1.7kb and 5.8 kb respectively. The 5.8kb band includes *bla*, pMB1 and *tsr* genes and 0.7 kb of *int* (ϕC31) gene, while the 1.7 kb band (base lane) contains 1.2 kb of *int* (ϕC31) gene, 53 bp of the *attP* site and 0.4 kb of structural nucleotides. Gel order: L1: *Lambda* DNA/*HindIII* ladder; L2*:* 1kb gene ruler; C: pYAC *BamHI/NcoI*; 1-14: rescued plasmids treated with *BamHI/NcoI*. All plasmids showed the corresponding 5.8 kb band (core structure of pYAC) and the rescued plasmids displayed two different patterns compared to the base line of pYAC vector. **C.** Methodology to rescue pYAC vector from *S. rimosus* chromosome.

Chromosomal DNA of pYAC-OTC *S. rimosus* transformants was isolated, digested with *SacI*, ligated and transformed into *E. coli* selected against ampicillin. **D & E.** Sequences of the attachment sites recognized by ϕC31 *int/attP* recombinase in *S. rimosus* genome. **D.** Comparison of *attP*, *attB* and pseudo *attB* sites before integration **E.** modification of *attB* sites after integration. *attP* sequence is underlined, conserved nucleotides of the *attB* sites are highlighted and the core region 5´TT where cross over occurs is indicated inside boxes.

Plasmid rescue strategy allowed us to identify the DNA sequence of two *attB* sites targeted by the ϕC31 recombinase in *S. rimosus* genome. However, this information was not sufficient to identify the exact location of the corresponding *attB* sites in *S. rimosus* chromosome considering the poor quality of the *S. rimosus* genome sequences available [55]. The current available genome sequence of *S. rimosus* is presented in 453 contigs and the entire chromosome has not yet been assembled. Therefore, in order to identify the location of the *attB* sites specific for ϕC31 recombinase on the chromosome of *S. rimosus*, we conducted pulsed field gel electrophoresis (PFGE) technique, which is able to separate large DNA fragments of up to 2Mb in size, after cutting the total DNA with rare cutting restriction enzymes. In the case of G+C rich content genomes, isolated from an actinomycete such as *S. rimosus,* the enzymes recognizing DNA sequence containing A+T more than 6 bp in size would cut entire genome very few times. We have restricted chromosomal DNA using *AseI* enzyme containing the recognition sequence 5´ ATTAAT 3´, to separate the chromosome DNA into large fragments. Importantly, the restriction site for *AseI* enzyme is also present in the pYAC-OTC plasmid. Thus, *S. rimosus* transformants containing integrated pYAC-OTC showed different restriction pattern of the entire chromosome due the additional *AseI* restriction sites. *S. rimosus* genome sequence indicates the presence of 11 *AseI* restriction places, thus leading to 13 fragments after PFGE. pYAC-OTC itself contains 3 *AseI* restriction sites. However, when analyzing with PFGE, these 3 *AseI* sites will be considered as a unique *AseI* site considering that 1 Kb fragments cannot be identified/separated by PFGE, and thus only contribute to the restriction pattern of the chromosome cut with *AseI* enzyme once. By comparing PFGE and Southern blot analysis of R7, “white”, M4108 *S. rimosus* strains and pYAC-OTC containing *S. rimosus* transformants, we have revealed that both *attB* sites are located in the largest DNA fragment A (~2.3Mb, Fig. 18A). The perfect *attB* site is located in the middle of the ~2.3 Mb fragment A (Fig. 18D). Therefore, a single integration event of pYAC-OTC in this location introduces an extra *AseI* into the ~2.3 Mb fragment, hence the *AseI* restriction pattern is altered giving rise to two bands of 1.3 and 1.0 Mb respectively, instead of the ~2.3 Mb fragment (transformant in lane 7 PFGE Fig. 18A). According to our analysis using PFGE, the pseudo *attB* site is located around ~250 Kb away from the first *AseI* restriction site of the ~2.3 Mb fragment A. Double integration event of pYAC-OTC incorporates 2 *AseI* restriction sites cleaving the ~2.3 Mb fragment three times in total, resulting in three bands of 250 Kb, 970 kb and 1.1 Mb (transformant lane 8 PFGE Fig. 18A). In addition to PFGE, we complemented our study with Southern blot analysis using the rescued DNA sequences as hybridization probes, thus allowing us to confirm the chromosomal location of the *attB* sites. Two hybridization patterns were observed in pYAC-OTC *S. rimosus* transformants confirming that ϕC31 recombinase can integrate once or twice in *S. rimosus* chromosome. The Southern blot analysis confirmed that both *attB* sites are located in the ~2.3 Mb fragment A given that both

probes targeted this DNA fragment in the R7, “white” and M4018 strains, while the *S. rimosus* pYAC-OTC containing transformants displayed different hybridization patterns (Fig. 18 B and C). Transformant in lane 7 underwent a single integration even, considering it only contains one extra OTC copy, therefore a single integration even occurred in the perfect *attB* site (targeting the pirin-like protein homologue), thus both probes hybridized with the 1.3 Mb fragment (lane 7 Fig. 18 B and C). Transformant in lane 8 underwent a double integration event, thus both probes targeted the 970 Kb fragment (lane 8 Fig. 18 B and C). Two DNA fragments have been identified by Southern blot analysis in transformants in lanes 5 and 6, this transformants have 2 extra OTC clusters, therefore the result obtained suggested partial digestion or, a subpopulation of each transformant has only one extra OTC copy integrated in the perfect *attB* site (lane 5 and 6 Fig. 18 B and C). A diagram explaining all the possible integration events mediated ϕC31 recombinase in *S. rimosus* chromosome is presented in Figure 18D.

Hranueli and colleagues constructed the restriction map of *S. rimosus* chromosome by applying PFGE and restriction analysis using *AseI* enzyme in the late 90s [53]. Their analysis showed that *AseI* restriction enzyme digests *S. rimosus* chromosome in 13 DNA fragments (Fig. 19), and that the otc cluster is located in the C1 fragment of 795 kb in size, while the *attB* site targeted by the pSAM2 integrase lies in the A fragment (Fig. 19) [53]. Due the poor quality of the entire genome sequence of *S. rimosus* available, we have used the map proposed by Hranueli et al. [53] as reference for analysis of *S. rimosus* chromosome. Besides confirming the presence of two *attB* sites specific for the ϕC31 recombinase, our analysis also allowed us to locate the *attB* sites in *S. rimosus* chromosome. Both *attB* sites targeted by the ϕC31 recombinase are located in the A fragment (~2.3 Mb) (Fig. 19), and when *S. rimosus* is transformed with ϕC31 recombinase containing plasmids, two different integration events can potentially occur i) a single integration event in the perfect *attB* site, thus disrupting the putative pirin like protein gene or ii) a double integration event, in the perfect and in the pseudo *attB* sites, consequently disrupting the putative pirin like protein and the putative acyl-CoA dehydrogenase genes respectively.

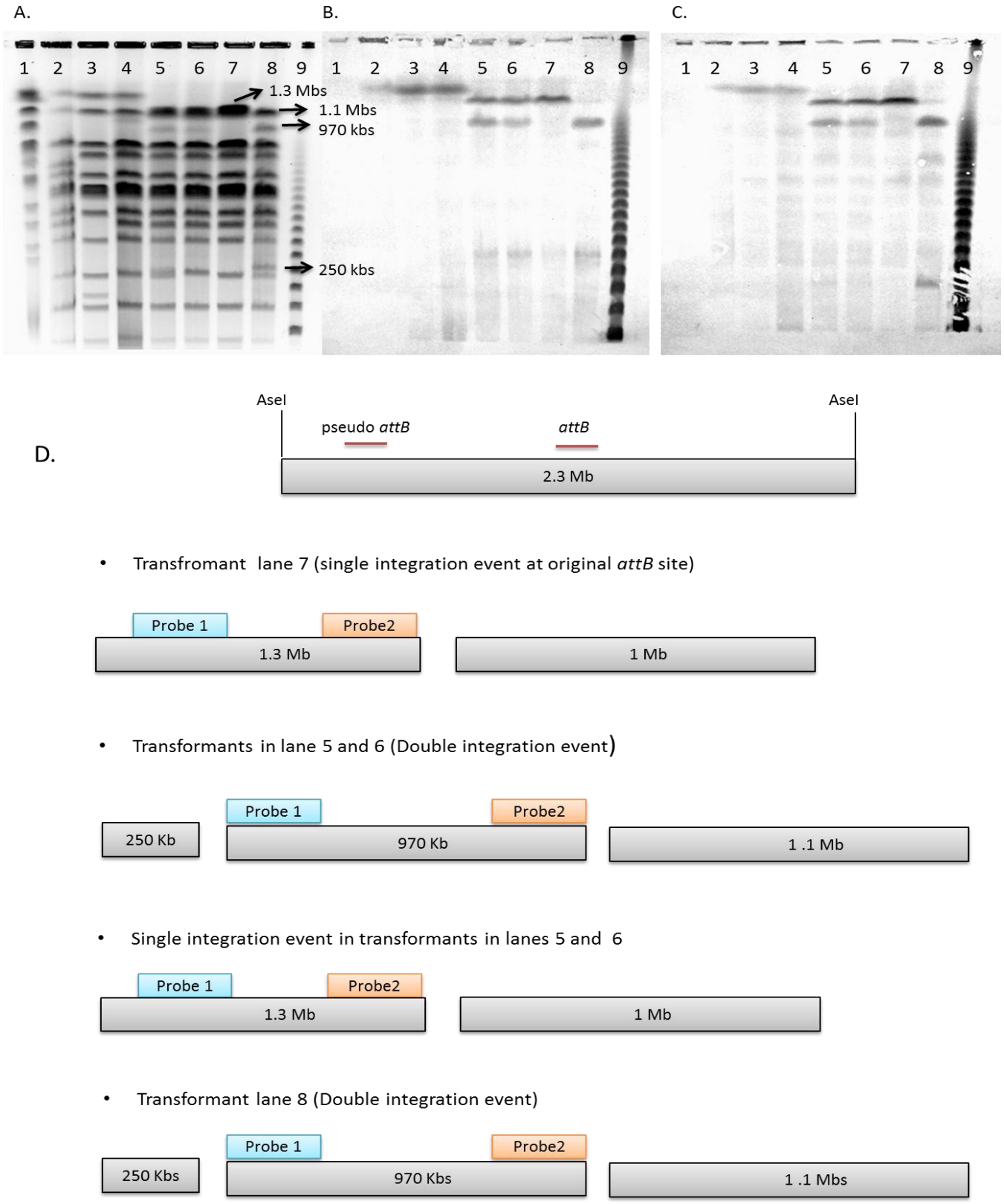


Figure 18. A. PFGE analysis of chromosomal DNA digested with *AseI*. Pulse program: 25 Kb to 1100 Kb. Lanes: 1. *Saccharomyces cerevisiae* chromosome marker; 2. R7; 3. "white"; 4. M4018; 5.6.7.8 pYAC-OTC *S. rimosus* transformants 9. Lambda PFGE ladder. **B.C.** Southern blot of PFGE hybridized with probes generated from rescued plasmids; each membrane was treated with an independent probe **D.** Restriction pattern of the ~2.3 Mb fragment A after single or double integration event.

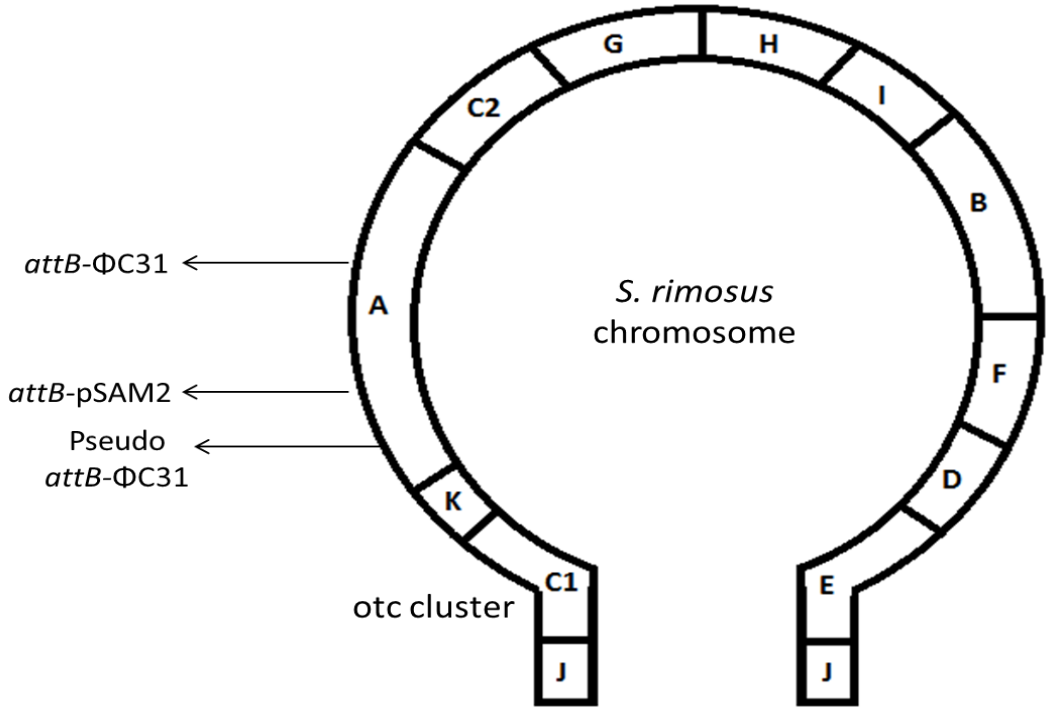


Figure 19. Proposed locations of the ϕC31 and pSAM2 integration sites on the restriction map of *S. rimosus* chromosome generated with *AseI* restriction endonuclease.

11.2 Evaluation of the expression profiles of *PnitA-NitR* and *tcp830* promoters in *Streptomyces rimosus*

Nitrilase *PnitA-NitR* and tetracycline *tcp830* are inducible promoters activated in the presence of ε-caprolactam and tetracycline, respectively (Fig. 20). These two promoters have been reported to be inducible in several *Streptomyces* species, as described in section 7.8.2 [97, 98, 100], thus indicating their potential use in *S. rimosus*. In the scope of this thesis, the expression profiles of both promoters were evaluated through the expression of the reporter gene *xylE* in *S. rimosus* M4018 strain. Both promoters carrying *xylE* as reporter gene were cloned in pAB04 vector containing ϕC31 recombinase and transformed into *S. rimosus*. Transformants were cultivated in TSB medium as described in material and methods and the inducers ε-caprolactam and tetracycline were added at different time points. Surprisingly, catechol 2,3-dioxygenase (XylE) activity was barely detected in *S. rimosus* transformants harboring up to two copies per cell of *xylE* regulated by *PnitA-NitR* or *tcp830* promoters, thus indicating that both promoters are not as efficient compared to the constitutive promoter *PermE** in *S. rimosus*, even in the presence of the corresponding inducers (Fig. 21). It has been reported that some analogs of ε-caprolactam and tetracycline improve the induction levels of each promoter [131, 132]. In the case of *PnitA-NitR* promoter, the addition of isovaleronitrile instead of ε-caprolactam was reported to activate its expression in *M. smegmatis* and *M. tuberculosis* [131], and for tetracycline promoter, anhydrotetracycline binds to the Tet repressor 35-fold more strongly than Tetracycline [132], however in *S. rimosus* these analogs did not seem to activate the corresponding promoters.

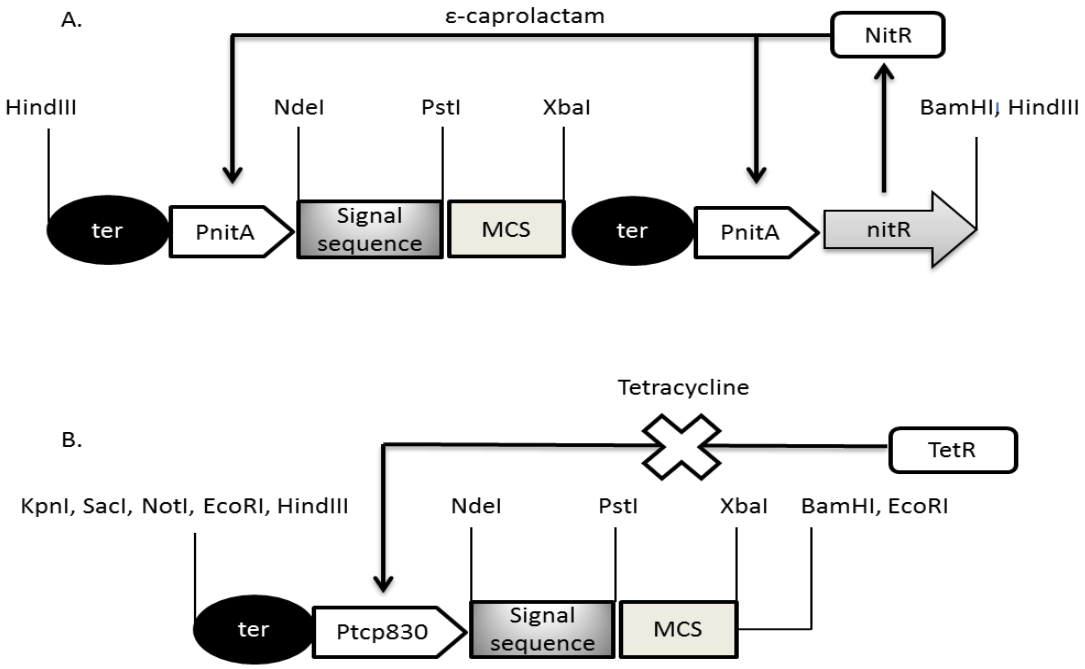


Figure 20. Inducible promoters evaluated in *S. rimosus*. **A.** Schematic presentation of *PnitA-NitR* promoter. Addition of ε-caprolactam induces the binding of the transcriptional factor *NitR* to *PnitA* and thus activating transcription of *PnitA* promoter [97]. **B.** Schematic presentation of *tcp830* promoter. Addition of tetracycline removes TetR repressor from *tcp830* promoter initiating transcription [98]. Both promoters were designed to facilitate cloning of genes with the same set of enzymes.

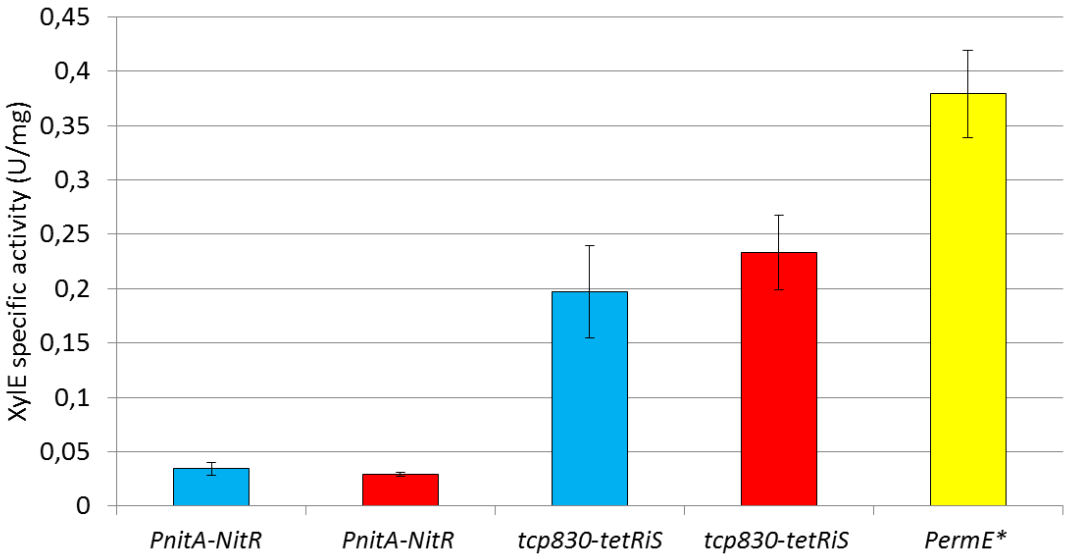


Figure 21. Evaluation of selected promoters in *S. rimosus* using the reporter gene *xylE*. The XylE activity was assayed after cultivating *S. rimosus* transformants for 24 hours in TSB medium. Transformants contained the integrative plasmid pAB04. Blue bars: cultures non-induced; Red bars: cultures induced with ε-caprolactam and tetracycline for *PnitA-NitR* and *tcp830* promoters respectively; yellow bar: transformants containing *PermE** promoter. Each construct was analyzed with 3 different transformants and results are presented as the means ± standard error. pVF-*tcp830/xylE* transformants evaluated contained *tetRiS* gene regulated by *SF14* promoter.

11.3 Construction of the *E. coli- S. rimosus* replicative shuttle vector pVF

S. rimosus has been characterized for its repellency to carry multi-copy vectors [103]. However, it was demonstrated that pPZ12 and pPZ74 plasmids, derived from pIJ101, were relatively stable plasmids in this strain [111], thus indicating the potential use of these vectors to introduce and express heterologous genes in *S. rimosus*. It was important to attempt construction of functional *S. rimosus* replicative vector, considering increased copy-number of the target gene potentially could lead to higher yields of the recombinant protein, and pPZ-based plasmids seemed to be good candidates for work with *S. rimosus* host. We constructed the pVF plasmid (Fig. 22A), a bi-functional multi-copy shuttle vector generated by merging pUC19 *E. coli* commercial vector and pPZ12 *S. rimosus* vector. This new plasmid has the benefits of both parental vectors, from one side pUC19 introduces the genetic elements to maintain the vector in *E. coli* allowing rapid DNA manipulation. On the other side, pPZ12 vector potentially ensures ability to introduce and maintain multiple copies of the plasmid in *S. rimosus*. To confirm the structural stability of the new chimeric vector, pVF was transformed into *S. rimosus* and the plasmid was isolated from a number of independent transformants to verify the restriction pattern. *PstI* restriction endonuclease, which divides/releases the DNA fragment containing pUC19 (2.6 Kb) from pPZ12 (6.2 Kb) was used for this purpose. The vector displayed the expected restriction pattern and *S. rimosus* transformants didn't show any morphological or physiological instability or any other physiological changes, compared to the wild type strain.

Once the structural stability of pVF vector was confirmed, we sequenced the entire plasmid pVF trough primer walking technique (primers are listed in table 10). The analysis of the entire DNA sequence of pVF plasmid confirmed that the DNA sequence corresponding to the pUC19 vector remained unaltered. The DNA sequence corresponding to the pPZ12 vector sequence revealed genetic elements from pIJ101-series vector, as expected. The sequencing of pVF plasmid revealed 8787 bp nucleotides with 64.5% G+C average content. By applying frame plot software [133] we identified 6 open reading frames (ORFs), two from pUC19 vector (*amp*^{r} and *rep*) and four corresponding to pPZ12 vector. As expected, one ORF belonging to pVF vector encodes for thiostrepton resistance gene, while the remaining ORFs (*rep, korB* and orf56) have homologous found in the pIJ101 plasmid [104]. When comparing pVF with pIJ101, they both share the origin of replication responsible for autonomous replication of the plasmid in *Streptomyces* spp. [104]. However, all the genetic elements involved in genetic transfer of pIJ101 (*tra, spdAB, korA, kilA* and *kilB*) are absent in pVF vector. pVF plasmid is containing only the orf56 and *korB* genes, which function is unknown [134].

To complete the characterization of pVF plasmid, we evaluated its potential segregational instability. The segregation instability information was needed to establish the rate at which pVF transformants lose all the copies of the plasmid in the absence of antibiotic selection, thiostrepton in this case. This frequency was calculated by sub-culturing *S. rimosus* pVF transformants in liquid TSB medium with or without selective marker every 24 hours. Each day serial dilutions of cultures were plated on the selective medium containing antibiotic and non-selective TSB plates, and CFU were

counted. According to the data gathered, around 30% of the cells lose the plasmid at every step of the sub-cultivation, or every 24 hours (Fig. 21B).

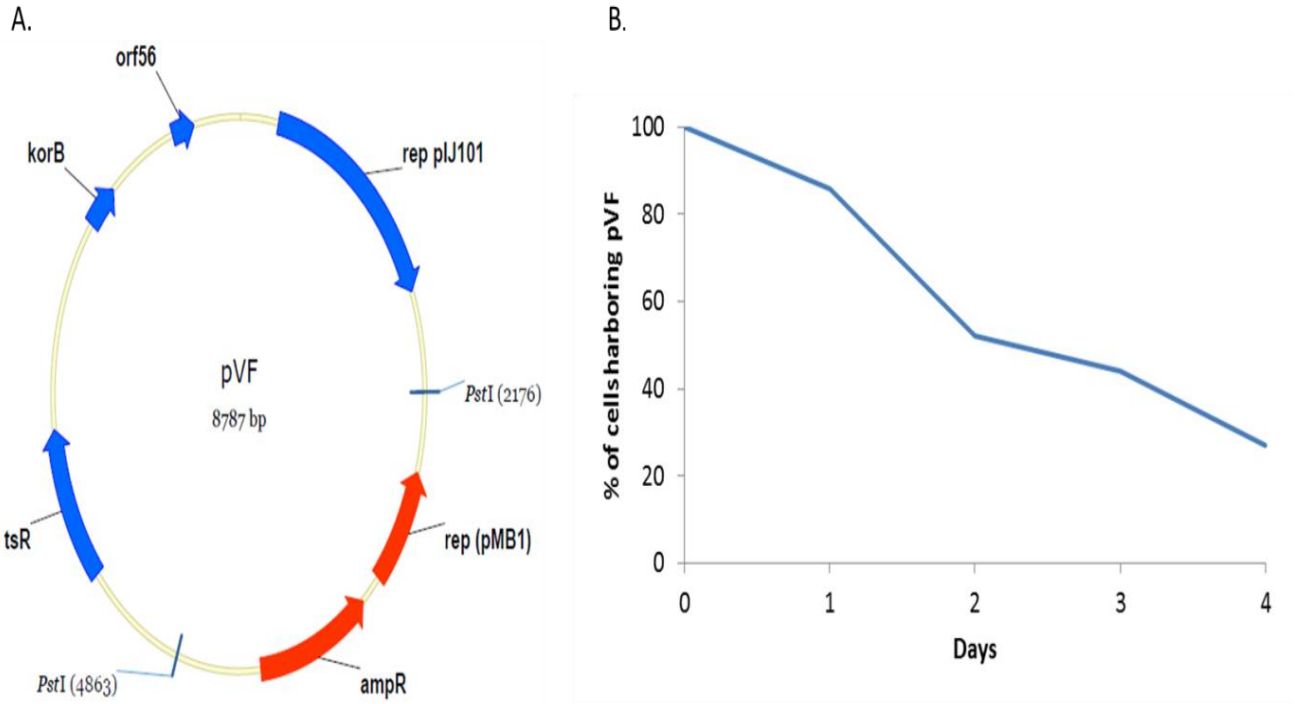


Figure 22. A. Map of pVF vector and **B.** Segregation instability of pVF.

11.4 Expression of the reporter gene *xylE* under *PermE** promoter using the replicative vector pVF in *Streptomyces rimosus*

Magdevska et al. [78] expressed the reporter system XylE in *S. rimosus* under the regulation of *PermE** promoter and introduced with the integrative plasmid pAB04, which contains the ϕC31 recombinase. These transformants were able to produce around 0.3 U/mg of XylE specific activity [78]. It was reasonable to expect, that increased number of copies of *xylE* gene should lead to higher activity of the recombinant protein XylE, and pVF vector offered the possibility to test this hypothesis. By applying pVF vector it was expected to increase the number of copies from 2 (maximum number of integration events obtained with ϕC31 recombinase in *S. rimosus*), to up to a few hundred copies of *xylE* per cell [105]. When transformed, *S. rimosus* transformants carrying pVF-*PermE**/*xylE* displayed an abnormal morphology. These transformants formed soft and bold colonies without visible sporulation and were not attached to the TSB agar plates, as is the case with the parent strain *S. rimosus* M4018. Interestingly, there wasn't any noticeable difference in the sporulation capabilities of these transformants when cultivated on SM medium. 30 independent pVF-*PermE**/*xylE S. rimosus* transformants were evaluated for catechol 2,3-dioxygenase activity and only 1 was positive for XylE activity reaching around 7 U/mg (Fig. 23). These transformants were further analyzed, and the plasmid DNA from each transformant was isolated to verify the restriction pattern. All 29 transformants, which did not show XylE activity, displayed a diverse restriction pattern, thus indicating

in-vivo rearrangement of the plasmid, while the positive transformant exhibited the expected restriction pattern. Despite the efficiency showed by the pVF-*PermE*/xylE S. rimosus* transformant with correct restriction pattern, with 23-fold higher XylE activity compared to the integrative version (from 0.3 U/mg to 7 U/mg), this expression system based on the replicative vector pVF and *PermE** promoter cannot be considered as a robust and useful genetic tool to produce target proteins in *S. rimosus* due its genetic instability.

11.5 Expression of the reporter gene *xylE* using the inducible promoters *PnitA-NitR* and *tcp830* with replicative vector pVF in *Streptomyces rimosus*

Initial results indicated that the inducible promoters *PnitA-NitR* and *tcp830* are not strong enough to produce XylE when they are integrated in the chromosome of *S. rimosus* using the ϕC31 integrase, when compared with *PermE** promoter (see 11.2). Surprisingly, on the other hand, the replicative plasmid pVF containing *xylE* reporter gene, under *PermE** promoter displayed high degree of instability. Presumably, this is the case due to high strength of *PermE** promoter (see 11.4). Therefore, to avoid this pronounced genetic and morphological instability observed earlier using the replicative vector pVF equipped with *PermE** promoter, we decided to test different promoters. Initially we tested pVF vector with *xylE* under the control of *PnitA-NitR* promoter designated as pVF-2 plasmid. This plasmid construct was transformed into *S. rimosus* and unlike pVF-*PermE*/xylE* containing transformants, all *S. rimosus* pVF-2 transformants displayed no difference in morphological or physiological properties compared to the parent strain *S. rimosus* M4018. The plasmid was isolated from pVF-2 *S. rimosus* transformants and the restriction pattern revealed that the plasmid was not being modified. XylE activity of the transformants was detectable, in contrast to the transformants with integrative version of *PnitA-NitR* promoter, where almost no activity was detected. By using the multi-copy vector pVF-2, XylE activity reached around 1 U/mg. However, the *PnitA-NitR* promoter didn't behave as an inducible promoter (Fig. 23).

To evaluate *tcp830* promoter activity in the replicative vector, we fused *xylE* gene with *tcp830* promoter creating pVF-3 vector and transformed *S. rimosus*. pVF-3 containing transformants did not display any morphological anomalies. Plasmid isolated from pVF-3 *S. rimosus* transformants showed the expected restriction pattern proving stable maintenance of the plasmid. Surprisingly, XylE activity reached up to 42 U/mg, a 7-fold increase when compared to *S. rimosus* transformants containing pVF- *PermE*/xylE*. Similarly to *PnitA-NitR*, *tcp830* promoter behaved rather as constitutive promoter in *S. rimosus* background.

We designed the *tcp830* promoter-containing plasmid without the transcriptional regulator *tetR* (*tetRiS*) because *S. rimosus* has more than 100 TetR family transcriptional regulator homologues present in the chromosome. Moreover, in the original paper describing *tcp830* promoter, it was demonstrated that *S. coelicolor* has an innate Tc-responsive repressor [98]. Therefore, we assumed that the promoter could also be potentially repressed by any of the *TetRiS* homologues present in *S.*

rimosus chromosome. However, initial results proved the TetR homologues present in *S. rimosus* don't interact with *tcp830* promoter. Therefore, in order to turn the promoter activity off, we incorporated the transcriptional repressor *tetRiS* under the control of *SF14* constitutive promoter to pVF-3 plasmid construct. The addition of the transcriptional regulator didn't influence the constitutive activity of *tcp830* promoter in *S. rimosus*. This way indicating that *tetRiS* either was not transcribed, because *SF14* promoter is not active in *S. rimosus,* or it cross-reacts with other genetic elements of the chromosome. To confirm *tetRiS* expression, we carried out a reverse transcription polymerase chain reaction (RT-PCR) analysis, and we have confirmed the activity of the *SF14* to transcribe the transcriptional regulator *tetRiS* (Fig. 24). Interestingly, Yin et al. also confirmed recently the capacity of *SF14* promoter to express genes in *S. rimosus* [65]. These data suggested that *tetRiS* has higher affinity for other *tetO* sequences present in *S. rimosus* chromosome.

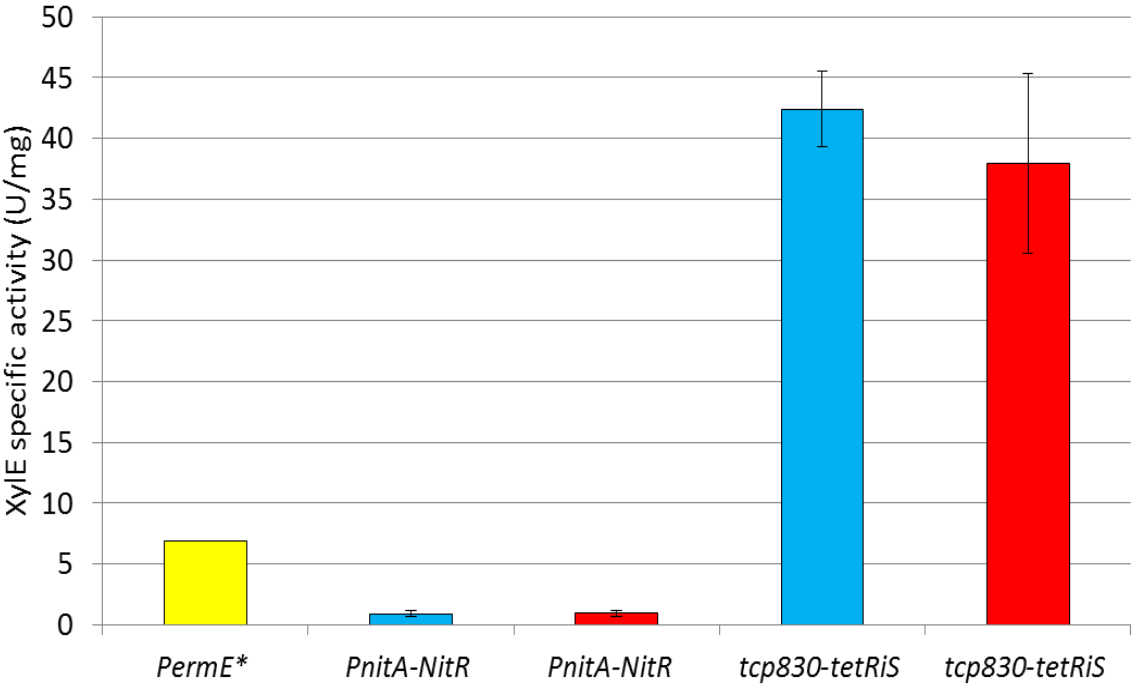


Figure 23. Evaluation of the strength of selected promoters in *S. rimosus* using the reporter system XylE and the replicative plasmid pVF. The XylE activity was evaluated after 24 hours of cultivation in TSB medium. Yellow bar: unique pVF-*PermE*/xylE* transformants positive for XylE activity; Blue bars: non-induced cultures; and red bars: induced cultures with ε-caprolactam or tetracycline for *PnitA-NitR* and *tcp830* promoters respectively. pVF-*tcp830/xylE* transformants evaluated contained *tetRiS* gene regulated by *SF14* promoter. 3 different transformants of each construct were analyzed and results are presented as the means ± standard error.

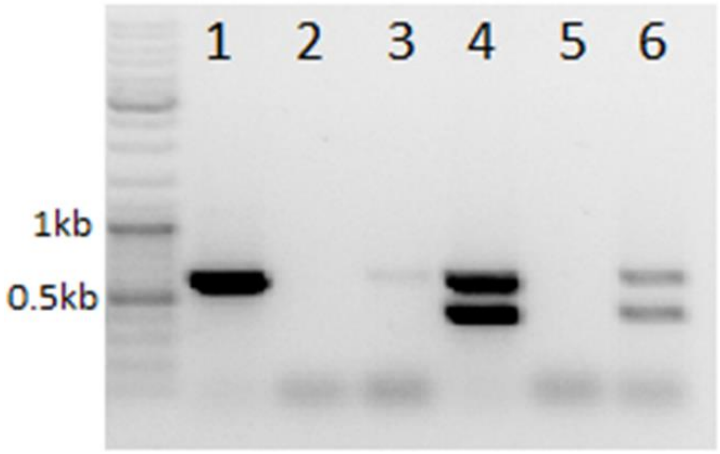


Figure 24. Evaluation of RNA expression of *tetRiS* gene under the control of *SF14* promoter. RNA from *S. rimosus* was isolated after 24 hours of cultivation in TSB medium. RNA polymerase principal sigma factor *hrdD* was used as positive control. A set of primers was designed to PCR amplify a DNA region of 600 bp of *hrdD* and 400 bp of *tetRiS*. Lines 1, 2 and 3 correspond to *S. rimosus* M4018. Line: 1 DNA, line 2: RNA, line 3: cDNA. Lines 4, 5 and 6 correspond to *S. rimosus* transformant containing pVF*/tcp830-xylE/SF14-tetRiS* vector. Line 4: DNA, line 5: RNA, line 6: cDNA. The PCR product corresponding to *hrdD* is present in the DNA a cDNA of both strains, while the PCR product specific for *tetRiS* is only present in the DNA and cDNA of pVF*/tcp830-xylE/SF14-tetRiS S. rimosus* transformant.

11.6 Production of recombinant phytase (AppA) using the replicative plasmids pVF-*PnitA-NitR* and pVF-*tcp830* in *Streptomyces rimosus* host

The specific activity of XylE was significantly improved using nitrilase and tetracycline promoters when introduced into *S. rimosus* on the replicative plasmid pVF. These data demonstrated the efficiency of the new expression vectors, thus indicating the potential of *S. rimosus* as a host for production of heterologous proteins. To corroborate the usefulness of the new gene tools, we tested them with the production of phytase (AppA), an industrial enzyme used as animal feed supplement [114]. Recently, it was demonstrated by our group that *S. rimosus* is able to produce and secrete the phytase from *E. coli* when the synthetic *appA* gene, codon optimized for *S. rimosus*, under the regulation of *PermE** promoter is fused to a trypsin-like proteinase signal sequence (*srT*) (unpublished data). The construct was integrated in *S. rimosus* chromosome using the ϕC31 recombinase, reaching AppA yields of around 3 U/mL [78]. Phytase production in *S. rimosus* is only achieved in the complex medium used for oxytetracycline (OTC) production. However, the new genetic tools were only tested in TSB medium; therefore their behavior in industrial medium was still not evaluated. We sub-cloned the *appA* gene *in frame* with trypsin-like proteinase signal sequence (*srT*) under the control of nitrilase and tetracycline promoters in pVF vector and transformed into *S. rimosus*. Transformants displayed normal morphology comparable to the wild type strain, confirming the genetic stability of the constructed expression vectors designed as pVF-4 (*PnitA-NitR/srT/appA*) and pVF-5 (*tcp830/srT/appA*). *S. rimosus* transformants were inoculated in the seed medium for OTC production and

after 20 hours, 1% of the seed culture was transferred to the complex medium. Samples were collected every 24 hours for 7 days to evaluate phytase activity during the cultivation. pVF-*PnitA-NitR* plasmid construct resulted in relatively low activity in complex medium reaching maximum around 1 U/mL of extracellular AppA activity. On the other hand, pVF-*tcp830* displayed much higher AppA activity reaching up to 6.5 U/mL (Fig. 25). The addition of the inducers ε-caprolactam or tetracycline didn't influence significantly the final activity of AppA, thus confirming the constitutive nature of these promoters in *S. rimosus*. Our results also confirmed the efficiency of *S. rimosus* to secrete the recombinant protein. Importantly, intracellular AppA activity never exceeded 15% of the total phytase activity, which is important advantage of this expression system (Fig. 25).

In the industrial production of oxytetracycline, maximum yield of antibiotic is obtained after seven days of cultivation. However, protein production in *S. rimosus* follows a different pattern. Time course experiment demonstrated that the best time to collect the recombinant protein is during the 4th day of cultivation, after this time the AppA activity starts decreasing (Fig. 25). When using pVF series vectors in *S. rimosus*, we observed significant segregation of this replicative vector, when cultivated on the medium without antibiotic. Our data demonstrated that every 24 hours around 30% of the cells lose the plasmid. When calculated, after around 96 hours of incubation only 25% of the cells still contained the vector, and this is the time when the highest yield of recombinant protein was reached in complex medium. Nevertheless, the segregation instability was tested in TSB, a rich medium where *S. rimosus* has much shorter generation time, reaching stationary phase already after 20 hours. To confirm whether or not pVF vector behaves similarly in complex medium, we analyzed AppA production in the absence of the selective marker thiostrepton. Our results demonstrated that the addition of thiostrepton to complex medium has a profound impact on the yields of AppA. For instance, after 4 days of cultivation *S. rimosus* transformants containing replicative expression vector pVF-*tcp830*, a 3-fold difference between medium containing thiostrepton (6.5 U/mL) and medium without thiostrepton (2 U/mL) was observed, thus confirming the plasmid segregation described in TSB medium also applies for complex medium. On the other hand, the use of the selective marker had much less influence to the final yield of phytase, when integrated in *S. rimosus* chromosome by ϕC31 recombinase and regulated with *PermE** promoter. Actually, there was no difference in AppA yields, with or without thiostrepton; reaching activity of around 3 U/mL in both cases. Nevertheless, the addition thiostrepton retarded AppA production, reaching maximum yields after 4 days, while in the absence of thiostrepton similar AppA yields were reached after only 48 hours (Fig. 25).

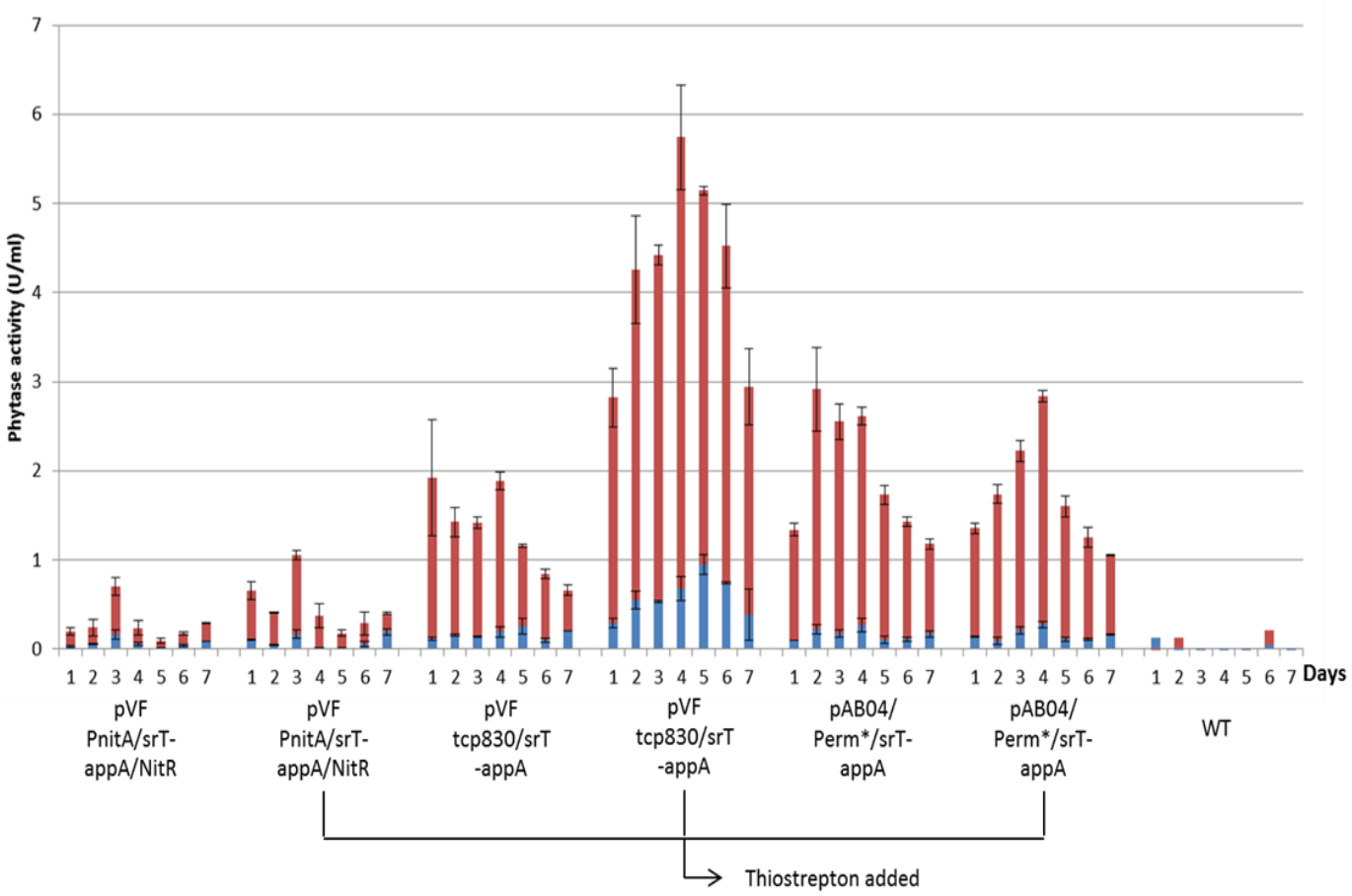


Figure 25. Phytase AppA activity measured in complex medium using *S. rimosus* M4018 as heterologous host. Blue bars indicate intracellular phytase activity and red bars extracellular phytase activity. Each construct was analyzed with 7 different transformants and results are presented as the means ± standard error. Recombinants strains cultivated with the selective marker thiostrepton are indicated.

Once we established the optimal conditions to produce phytase using *S. rimosus*, after 4 days of cultivation in complex medium using the pVF-*tcp830* expression vector, we carried out limited colony selection effort in order to select the highest-producing strains, thus estimating what would be the highest production of phytase, without the need of intensive mutagenesis or further medium and process development at shaker level. It is well known to the expert in the field of industrial microbiology, that intensive mutagenesis and strain improvement efforts combined with further medium and process development in fermenters can significantly improve final yield of the target product. The shake flask experiment with limited strain selection (without mutagenesis) will thus give us initial information on the variability and early estimate what is the highest yield of phytase that we can achieve by a simple and limited selection procedure. This information is therefore very useful, because it will aid further assessment of the potential of the current expression system with available promoters at industrial settings. We therefore selected the best *S. rimosus* recombinant strains carrying pVF-*tcp830/srT/appA* and pAB04-*PermE*/srT/appA* and spread these two strains on TSB plates to generate single colonies. 18 independent fast-growing colonies of each construct were patched on SM medium and after 7 days of incubation plugs were inoculated in seed medium for OTC production. The

cultivation media were supplemented with thiostrepton during all stages of cultivation to gain maximal activity of phytase. The 36 daughter colonies were evaluated for extracellular phytase activity at the fourth day of the bioprocess. Interestingly, the limited number of tested pAB04-*PermE** containing displayed minimal variability, thus indicating that once the construct is integrated into the chromosome, the expression is stable over time and generations; demonstrating robustness and reproducibility of the selected transformants. On the other hand, we were able to isolate easily significantly improved strains with higher AppA activity containing replicative plasmid pVF-*tcp830*. In some cases reaching up to 20 U/mL of extracellular AppA activity (3-fold increase from the parental strain), thus suggesting rapid adaptation of the replicative plasmid in *S. rimosus* background (Fig. 26). However, this also means limited robustness of these recombinant strains, considering selective pressure with thiostrepton has to be used; not preferred strategy when considering industrial applications.

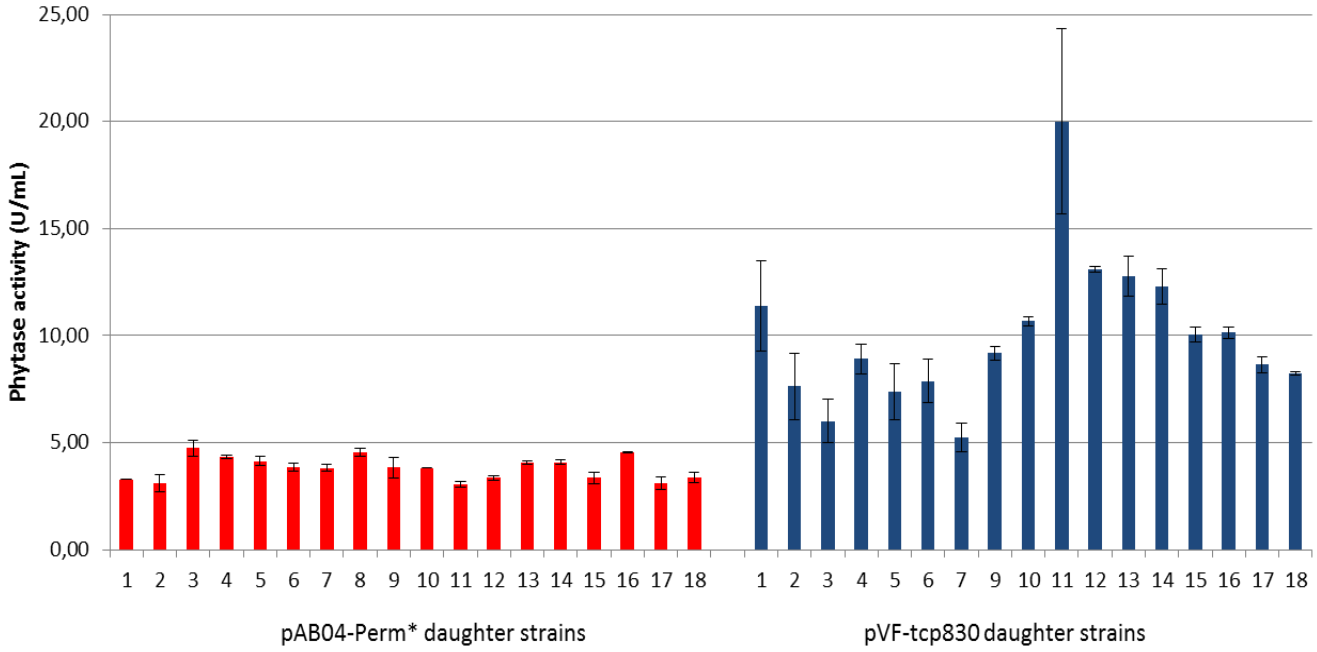


Figure 26. Extracellular phytase AppA activity measured in complex medium after 4 days of cultivation. Red bars: different *S. rimosus* daughter colonies from the initially selected parental strain carrying integrative vector pAB04-*PermE*/srT/appA*. Blue bars: different *S. rimosus* daughter colonies from the initially selected parental strain carrying replicative vector pVF-*tcp830/srT/appA*. Each isolate was evaluated in two separate experiments with two replicates and results are presented as the means ± standard error.

11.7 Purification and identification of the recombinant extracellular phytase (AppA) produced by pVF-*tcp830/srT/appA Streptomyces rimosus* transformant using His-Tag/Ni affinity chromatography

During the cloning procedure, we designed the version of the *appA* gene containing C-terminal polyhistidine-tag to facilitate its purification. pVF-*tcp830/srT/appA* expression vector was transformed into *S. rimosus* "white" strain. With the increased yields of

AppA that we gained using the replicative vector pVF-*tcp830/srT/appA*, we attempted to test whether the expression of the recombinant AppA was high enough to facilitate simple isolation procedure of extracellular AppA from the complex medium supernatant. The highest producing *S. rimosus* "white" isolate containing replicative vector pVF-*tcp830/srT/appA* was cultivated in 200 mL of complex medium, following the optimal cultivation conditions for production of phytase described earlier in the section 10.12. After 4 days of cultivation the *S. rimosus* culture broth was centrifuged, and the cell pellet was discarded. Supernatant was cleared by vacuum filtration through filter paper to remove residual debris from the culture. Supernatant was supplemented with TRIS buffer and loaded into a HisTrap column as described in materials and methods. SDS-page of the eluted protein fractions identified a band of 45 kDa presumable belonging to AppA (Fig. 27). Further LC-MS/MS analysis of the isolated band from the Coomassie stained gel confirmed the identity of the purified protein band corresponds to phytase (Supplementary information), thus proving for the first time that *S. rimosus* is a robust heterologous host. Protein concentration of phytase was estimated to reach 1.8 mg/mL in a final volume of 2 mL, therefore from 200 mL culture of *S. rimosus* it was possible to isolate 2.6 mg of phytase.

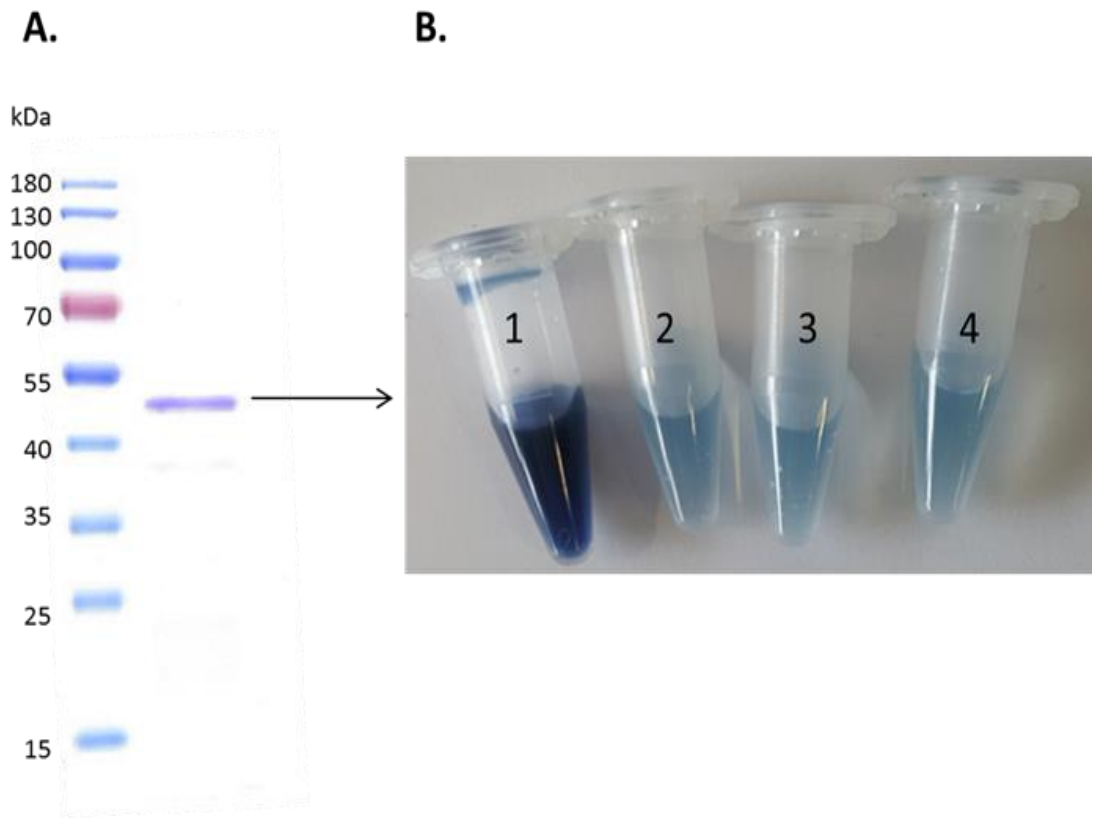


Figure 27. A. Recombinant AppA purified with HisTrap affinity column from the supernatant of *S. rimosus* "white" strain harboring the replicative vector pVF-*tcp830/srT/appA*. The band was in-gel digested and analyzed through LC-MS/MS confirming the identity of AppA (supplementary information). **B.** Enzymatic assay to measure AppA activity using 1. Purified phytase, 2. Purified phytase treated with TCA, 3. Elution buffer, and 4. H_2O.

11.8 Transcriptome analysis of *Streptomyces rimosus* under different conditions

Notwithstanding the evidence gathered about the feasibility of using *S. rimosus* to produce heterologous proteins, the physiology of this microorganism related to heterologous protein expression is still not well understood. To date, rather low quality *S. rimosus* genome sequence of different strains are available [55, 56] and only the pathway of oxytetracycline (OTC) biosynthesis is well characterized [50]. However, key information related to gene expression and protein production, such as the Sec and Tat pathways, or presence of molecular chaperones, has not been studied in *S. rimosus*. Like many streptomycetes, *S. rimosus* encodes numerous putative hydrolytic enzymes [135], and it also encodes several putative gene clusters involved in the biosynthesis of secondary metabolite [136], whose products have not been identified to date. In addition to *S. rimosus* genome annotation, valuable information on gene expression and production of secondary metabolites or enzymes produced by *S. rimosus* can be gathered by a comparative transcriptome study. We have thus carried out comparative transcriptome analysis of *S. rimosus* cultivated on two media, laboratory medium TSB, which ensures rapid growth of *S. rimosus*, but does not significantly induce secondary metabolism, and complex medium, which was originally developed for OTC production at a shake flask level.

S. rimosus is a fast-growing streptomycete and this feature is readily demonstrated in TSB medium, where *S. rimosus* culture reaches stationary phase in 20 hours, depleting all nutrients after 24 hours of cultivation. On the other hand, when cultivated in the complex medium, the biosynthetic process can last up to seven days. Therefore, analysis of *S. rimosus* transcriptome cultivated on these two entirely different media should bring very useful information on gene expression and gene regulation. The transcriptome was analyzed at 6 different time points (Fig. 28) to have a broad picture of what is being transcribed at different times. While soluble TSB medium ensures rapid growth, complex medium, on the other hand, induces secondary metabolism, and thus stimulating production of secondary metabolites and hydrolytic enzymes.

S. rimosus transcriptome was kindly provided by the company Acies Bio Ltd. (Slovenia). Whole transcriptome shotgun sequencing (WTSS) or RNA-seq is a technique used to quantify gene expression levels at certain conditions and there are several ways to interpret the data extracted from this analysis. We used the Reads per kilobase per million mapped reads (RPKM) method, the most widely used approach in next-generation sequencing research for length normalization [137]. Briefly, RNA-seq protocols use mRNA fragmentation approach prior sequencing. Thus, the total number of reads for a given transcript is proportional to the expression level of the transcript multiplied by the length of the transcript. Explained in a simplified manner, a long transcript will have more reads compared to a short gene of similar expression [138]. Therefore RPKM is a length normalization method to compensate this variability and it was the method of choice to present the gene expression levels of the different genes under investigation in this study.

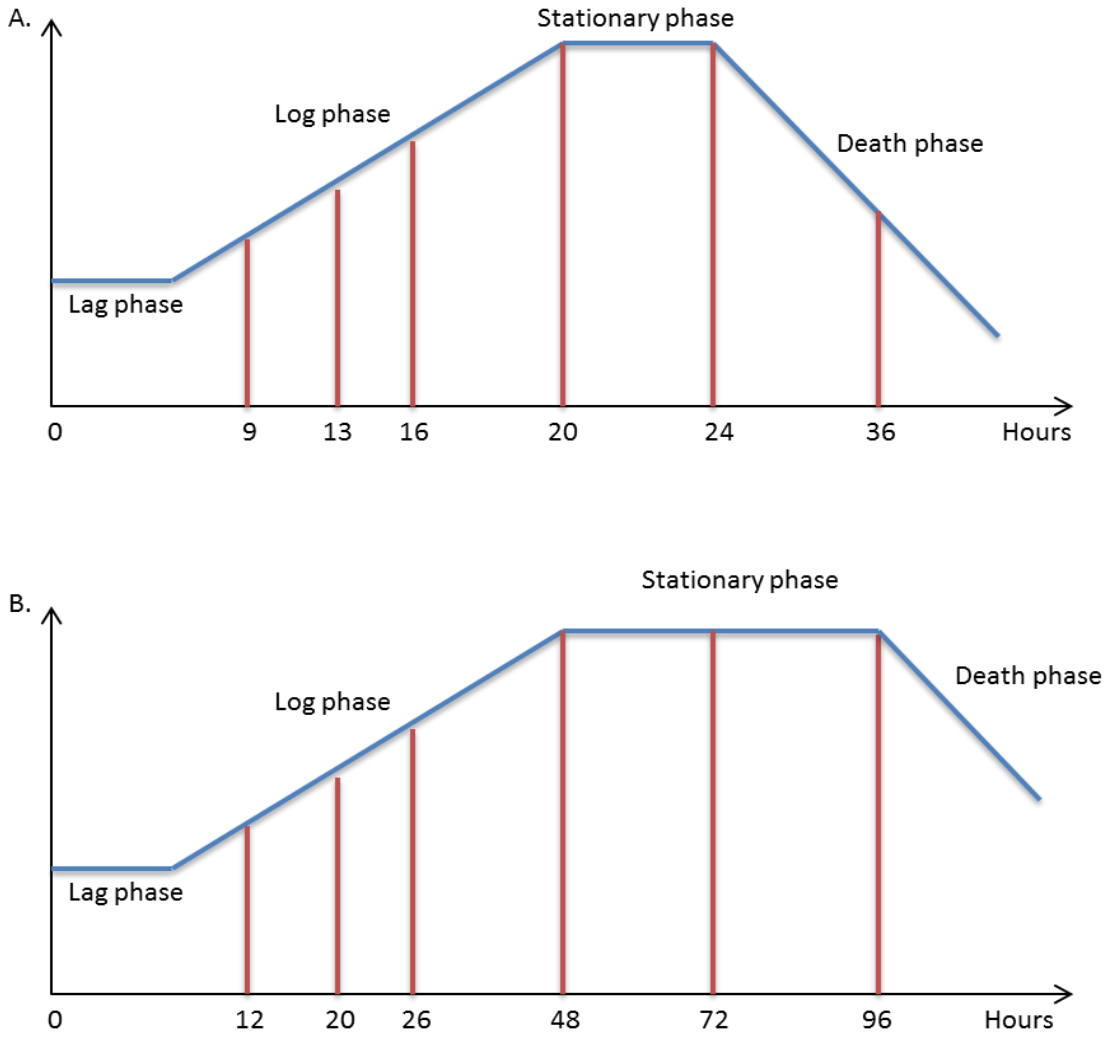


Figure 28. *S. rimosus* growth curve in **A.** TSB medium and **B.** complex medium. The time when the mRNA was isolated is marked with red lines.

11.8.1 Identification and analysis of putative gene homologues involved in the formation of the Sec pathway in *Streptomyces rimosus*

Streptomyces species are soil-dwelling bacteria specialized to survive in harsh and hostile environment. These microorganisms developed a diverse metabolism, which enables them to produce a large number of secondary metabolites and hydrolytic enzymes. According to one of the existing hypothesis, secondary metabolites have evolved to suppress growth of other microorganisms [70]. It is known that *S. rimosus* produces at least two antibiotics, oxytetracycline (OTC) and the polyene antifungal rimocidin. Like in other organisms, a supply of nutritional resources and synthesis of diverse building blocks is also required for *S. rimosus* cells, which is achieved by exploiting complex organic materials in the environment, and this is why streptomycetes generally produce numerous hydrolytic enzymes [135]. *S. rimosus* genome contains the Sec and Tat pathways to secrete these enzymes. In the scope of

this thesis, we have decided to employ the Sec pathway to produce target heterologous enzymes extracellularly (e.g. phytase), thus facilitating enzyme purification in the scope of downstream processing.

The Sec-dependent pathway exports proteins in an unfolded conformation guided by a signal peptide to a translocon where an ATPase pump pushes the proteins out of the cell [139]. Firstly, the protein targeted to the Sec-dependent pathway is biosynthesized as pre-protein with an N-terminal signal sequence, which contains the signal peptidase recognition site at the C-terminal of the signal peptide. The unfolded protein is captured by the chaperone SecB which prevents its folding and delivers it to SecA protein. The ATPase encoded by SecA interacts also with the translocon formed by SecYEG. This translocon serves as a channel where the pre-protein is pushed out by the action of SecA in an ATP-dependent mechanism assisted by SecDF and YajC. Finally, once the protein is completely translocated, the signal peptide is cleaved off by a type I signal peptidase releasing the mature protein into the culture supernatant [70, 140, 141]. We have identified in *S. rimosus* genome gene homologues presumably involved in the formation of the Sec pathway. Gene homologues of *secAYEGDF*, *yajC* and five type I signal peptidases were identified in *S. rimosus* genome with the exception of the chaperone SecB. However, based on the literature data, this chaperone is often missing in Gram-positive bacteria [141], and its role is substituted by the chaperones DnaKJE [142-144]. We have identified in *S. rimosus* genome two *dnaK* and three *dnaJ* gene homologues.

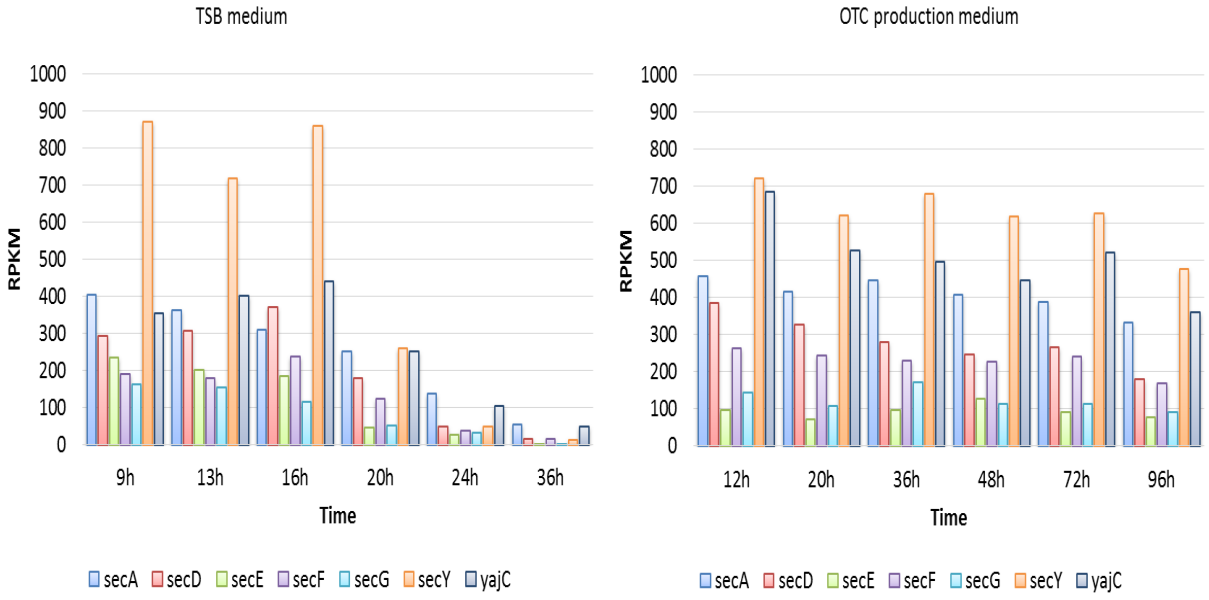
Transcriptome data revealed that the identified putative genes involved in the formation of the translocon are expressed in both media, in TSB and in complex medium (Fig. 29A). However, it is reasonable to expect that after 24 hours of incubation the secretory capacity in TSB medium is exhausted considering all nutrients are already depleted in accordance with the stationary stage of the culture (Fig. 28 and Fig. 29A). In the complex medium the expression of the putative translocon is stable throughout the different time points suggesting good capacity of *S. rimosus* to export proteins in this industrial-like medium.

Examination of the putative chaperones required to drive the unfolded protein to the translocon in the absence of SecB allowed us to identify five potential candidates, possibly “replacing” the role of SecB; two homologues of *dnaK*, which according to our transcriptome analysis are not transcribed in both media, and three putative gene homologues of *dnaJ* which we designed as *dnaJ* I, *dnaJ* II and *dnaJ* III, displaying different expression levels. Putative *dnaJ* II homologue is not expressed in any condition while putative *dnaJ* I and *dnaJ* III gene homologues are active in both media, thus indicating their potential role assisting the secretion of proteins (Fig. 29B). Interestingly, based on transcriptome data, in TSB medium the expression of *dnaJ* I and *dnaJ* III does not seem coordinated with the expression of the genes of Sec pathway, considering that *secAYEGDF* expression is decreasing over time, while the expression of *dnaJ* I and *dnaJ* III is increasing during the same time, even in the late stationary phase. DnaJ chaperone, also designated as Hsp40 (for heat shock protein 40 kDa), is a co-chaperone of DnaK (Hsp70). DnaJ unaccompanied by DnaK can bind unfolded proteins as well as native proteins exerting a strong unfolding activity [145]. DnaK and

DnaJ cooperate to assist the folding of newly synthesized or unfolded polypeptides. However, if it is to believe transcriptome data, their role in *S. rimosus* might be questioned, due the lack of expression of *dnaK* gene homologues. On the contrary, high concentrations of DnaJ inhibit DnaK-DnaJ-mediated refolding [146], phenomena that is likely occurring in *S. rimosus*, where the maintenance of the global protein homeostasis is influenced by the two DnaJ chaperones. The increased expression levels of *dnaJ* I in TSB medium at the late stationary phase is likely due to nutrient limitation, cells try to reduce starvation stress, as a consequence heath shock proteins try to repair damaged proteins otherwise degrade them restoring protein homeostasis promoting cell survival [147]. Complex medium on the other hand provides sufficient amount of nutrients to avoid starvation during the first 96 hours, therefore it is possible to conclude that *dnaJ I* and *dnaJ* III were active at stable levels in this industrial medium, functioning as helpers in protein folding and driving pre-proteins to the Sec pathway. Although in the last time point (96 hours), a slightly increase in *dnaJ* I expression can be observed, thus indicating the end of the stationary phase (Fig. 29B).

Five type I signal peptidases (SPases) homologues were identified in *S. rimosus* genome, these enzymes are responsible for the proteolytic cleavage of the signal peptide of secreted proteins and most bacterial membranes contain one or two SPases [148]. *S. lividans* encodes 4 SPases designated SipWXYZ suggesting the important role of extracellular enzymes in the life cycle of streptomycetes [149]. Transcriptome data showed that the five SPases are upregulated in TSB medium during the first 20 hours. During the starvation period, at late stationary phase, the SPases expression declines, like the case with the translocon (Fig. 29C). In complex medium the SPases are active during different time points suggesting their cooperation with the translocon secreting proteins. Interestingly, the type I peptidase II gene homologue is particularly strongly expressed in both media in accordance with reports in *S. lividans*, where SipY is the major SPase and its absence causes deficient extracellular protease production, while the remaining three exert a secondary role helping SipY [149].

A.



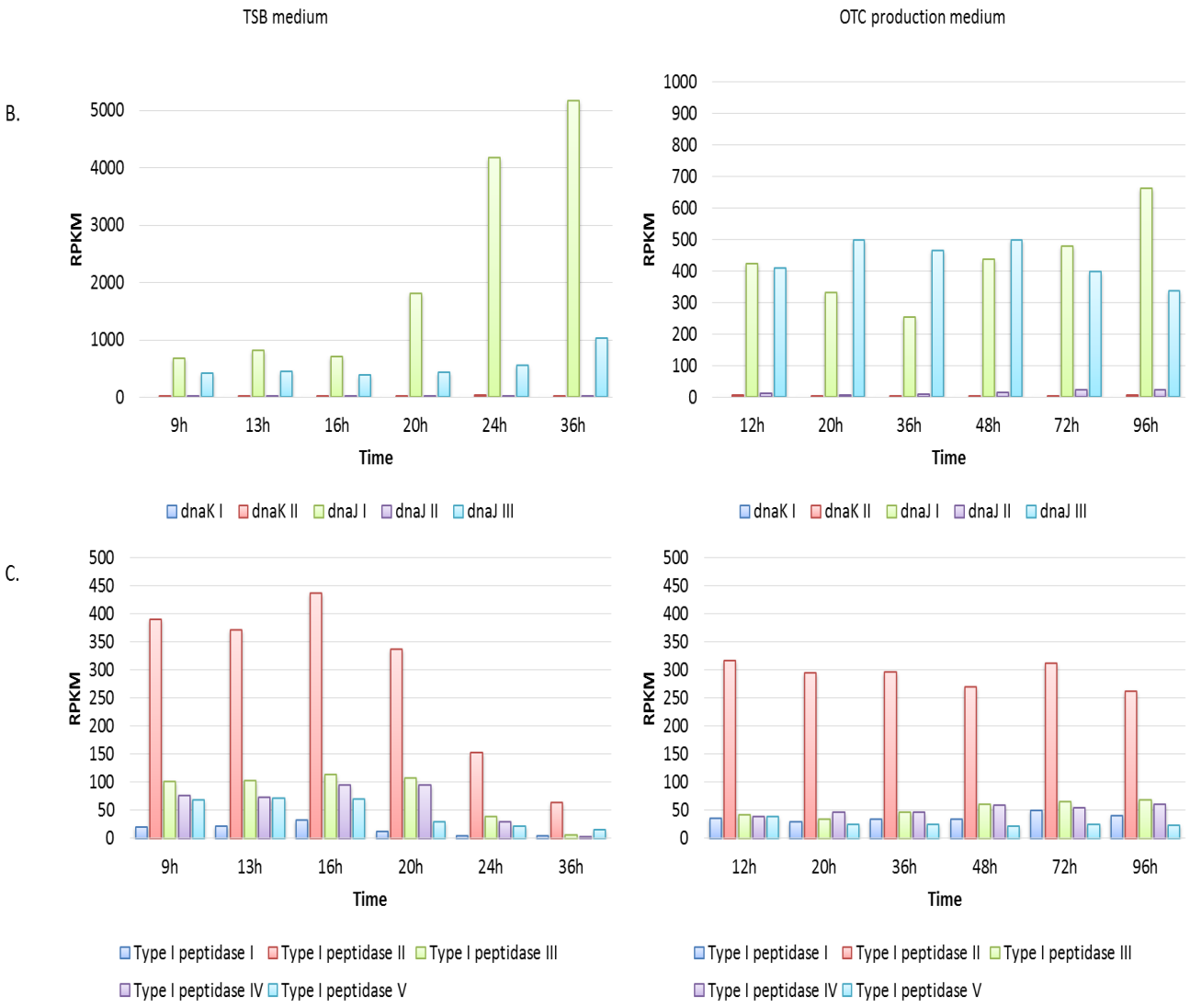


Figure 29. Transcriptome analysis of the gene homologues involved in the formation of the Sec-dependent pathway in *S. rimosus* cultivated in two different culture media. **A.** *secADEFGY* and *yajC* gene homologues putatively involved in the formation of the translocon. **B.** *danK* and *dnaJ* gene homologues encoding for putative molecular chaperones potentially replacing the role of the chaperone SecB. **C.** Putative type I signal peptidases gene homologues.

11.8.2 The twin-arginine translocation pathway (Tat pathway) in *Streptomyces rimosus*

Although we only used a signal sequence of the Sec-dependent pathway for the construction of expression plasmids in our heterologous protein expression experiments in *S. rimosus*, this strain also contains gene homologues belonging to the Tat pathway. Unlike the Sec pathway, which transports proteins in an unfolded manner, the Tat pathway exports proteins in their final or folded conformation; therefore its mechanism of export differs from the Sec pathway [150]. The Tat system in Gram-positive bacteria is composed of two proteins designated as TatA and TatC, unlike Gram-negatives which carry three membrane proteins (TatABC) [151]. In *S. rimosus* we identified two putative *tatA* homologues and one putative *tatC* homologue, as expected for Gram-positive bacteria, but differing from other

streptomycetes like *S. lividans,* where TatABC complex was identified [70]. In addition to the *tat* gene homologues, the target protein destined for transport via Tat pathway contains a signal peptide with the highly conserved twin-arginine motif, defined as SRRXFLK [152]. The only common feature that Sec and Tat pathway share is the cleavage of the signal peptide by a SPase I to release the translocated proteins [153]. TatA and TatC proteins form the translocon, which transports the proteins by mediated proton motive force [150]. Although its mechanism is still not well understood, it is known that TatA has a dual role, carrying out both TatA-like and TatB-like roles [154]. In Gram-negatives TatB and TatC form an integral membrane complex which binds the signal peptide, then TatA is recruited to form the TatABC complex which drives the protein out of the membrane to be later cleaved by SPase I [152]. *S rimosus* transcriptome suggests that *tatA* II and *tatC* homologues are active in both media that were tested. Analogously to the Sec pathway, the transcription of the gene homologues of Tat system is dependent on the culture growth stage. The transcription of the Tat pathway gene homologues starts decreasing during the stationary phase in TSB medium. In contrast, transcription of Tat gene homologues is active during entire 96 hours of incubation in complex medium (Fig. 30). According to the transcriptome analysis, *tatA* I gene homologue is not transcribed. Interestingly, based on the high degree of homology, the *tatA* I is probably a direct gene duplication of the *tatA* II, which is often found elsewhere on the genomes of several species of gram-positive bacteria [150].

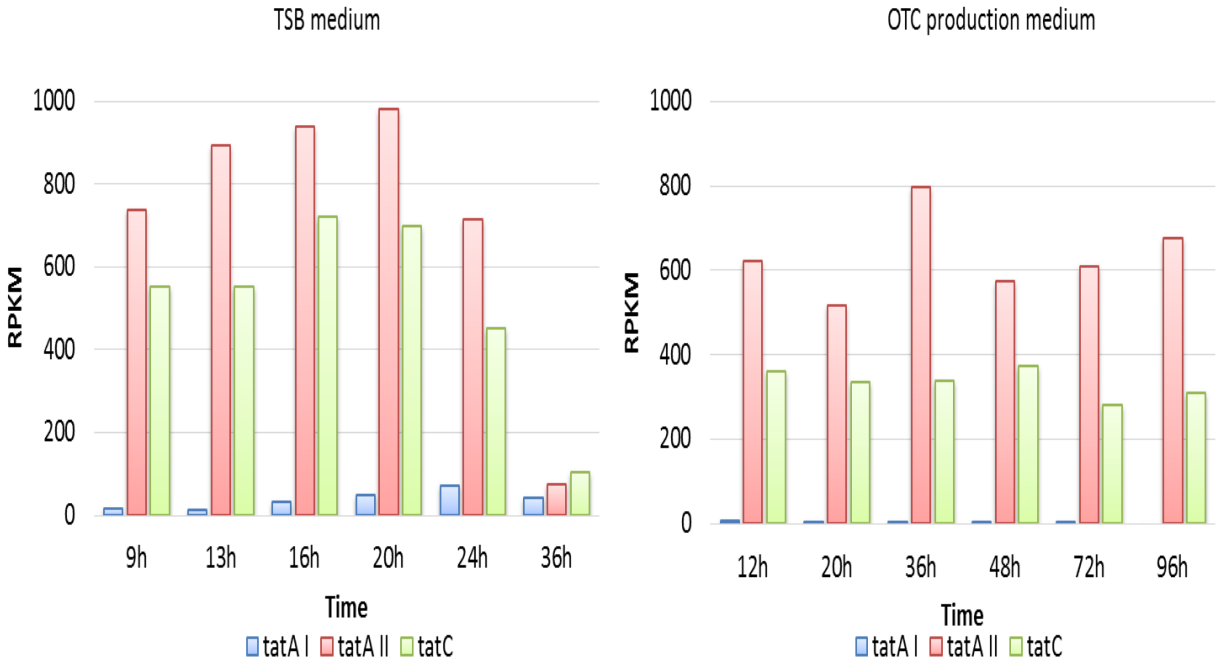


Figure 30. Transcriptome analysis of the gene homologues involved in the formation of the Tat pathway in *S. rimosus* cultivated in different culture media.

11.8.3 Endogenous proteases of *Streptomyces rimosus*

One potential drawback, when producing recombinant proteins extracellularly is the activity of endogenous extracellular proteases, which can degrade the target heterologous protein [155]. Historically, *S. lividans* has been recognized as the most useful heterologous host within the *Actinomyces* genus because it displays a low extracellular protease activity [156]. *S. lividans* also lacks the restriction systems common in streptomycetes that prevents the entry of *E. coli* derived DNA (vectors), thus avoiding the use of non-methylated DNA [156]. *Bacillus subtilis* is another widely used expression host exploited for its capacity to export proteins. However, this bacterium produces at least eight extracellular proteases potentially involved in the degradation of heterologous secreted proteins leading to commercial losses [157]. Another microorganism recognized for high capacity to produce extracellular recombinant proteins is *Pichia pastoris* and despite several efforts to alleviate undesired proteolysis, including media development and engineering of protease-deficient strains, proteases can still be an issue when using this yeast strain for production of extracellular recombinant proteins [32]. Gene tools developed in *S. rimosus* to produce and secrete recombinant proteins, specifically phytase in our case, demonstrate relatively high robustness of this expression system. However, it is difficult to access the negative effect of extracellular proteases on the final yield of the target heterologous enzyme produced by this host.

Proteases (EC 3.4.21-24) are enzymes that break down peptide bonds of proteins via hydrolysis [158]. These enzymes can be divided into six categories: serine proteases, cysteine proteases, aspartic proteases, glutamate proteases, threonine proteases and metalloproteases, based on the functional group present at the active site and their catalytic mechanism [158, 159]. Transcriptome analysis identified 8 putative subtilisin-like serine proteases, 10 putative trypsin-like serine proteases and 18 putative metalloproteases. Thus suggesting, *S. rimosus* is very rich source of proteases.

Subtilisin-like serine proteases or subtilases belong to the serine protease family [160]. These enzymes are found extracellularly as well as intracellularly and these group of proteases show different degrees of specificity [160]. Extracellularly, these proteases are unspecific, most likely playing a role in defense and nutrition. In contrast, intracellularly they are highly specific, acting upon pro-proteins and pro-peptides [161]. Transcriptome data reveals that in TSB medium, three subtilases gene homologues are transcribed (subtilase III, IV & V). While in the complex medium five out of eight subtilase homologues identified are transcribed (subtilase II, III, IV, V & VI). Interestingly, the subtilase V is highly transcribed in both tested culture media (Fig. 31).

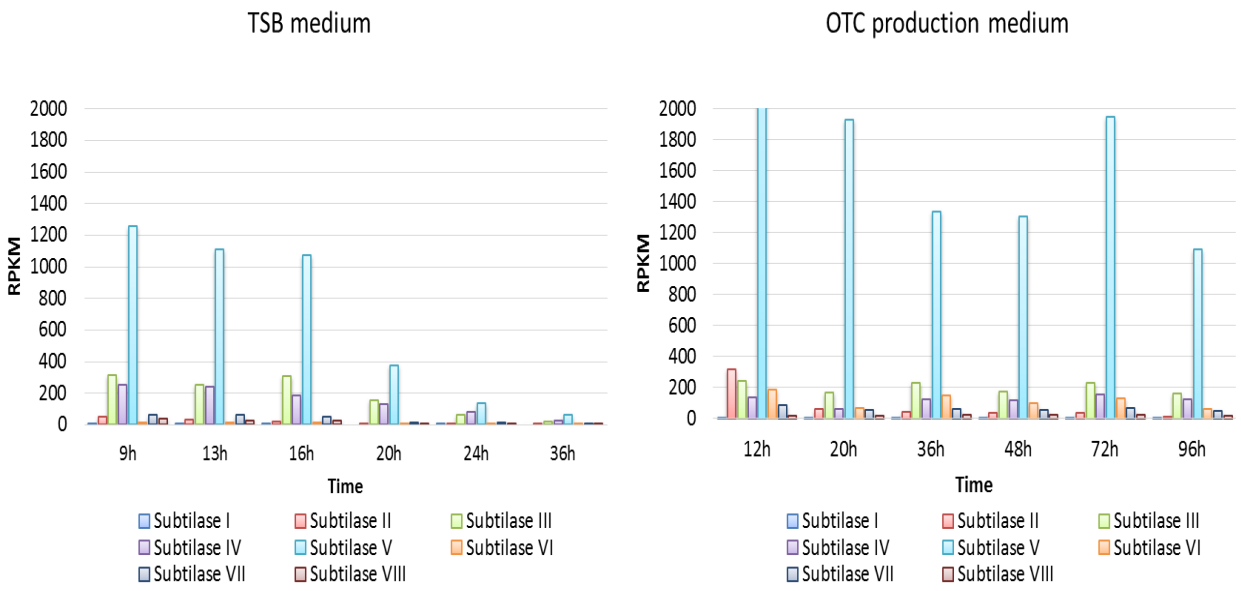


Figure 31. Transcriptome analysis of putative subtilisin-like serine protease genes identified in *S. rimosus* cultivated in two different culture media.

Trypsine-like serine proteases also belong to the serine protease family [162], and are usually extracellular proteases biosynthesized as pre-proenzymes found mainly in the digestive tract of animals and microbial sources [163, 164]. *S. rimosus* encodes ten trypsine-like serine protease homologues, nonetheless only one homologue is strongly expressed in the complex medium (trypsine-like serine protease V). Based on the transcriptome data, the remaining nine gene homologues are not expressed significantly, similarly as found in the TSB medium, where their transcription is barely detected (Fig. 32).

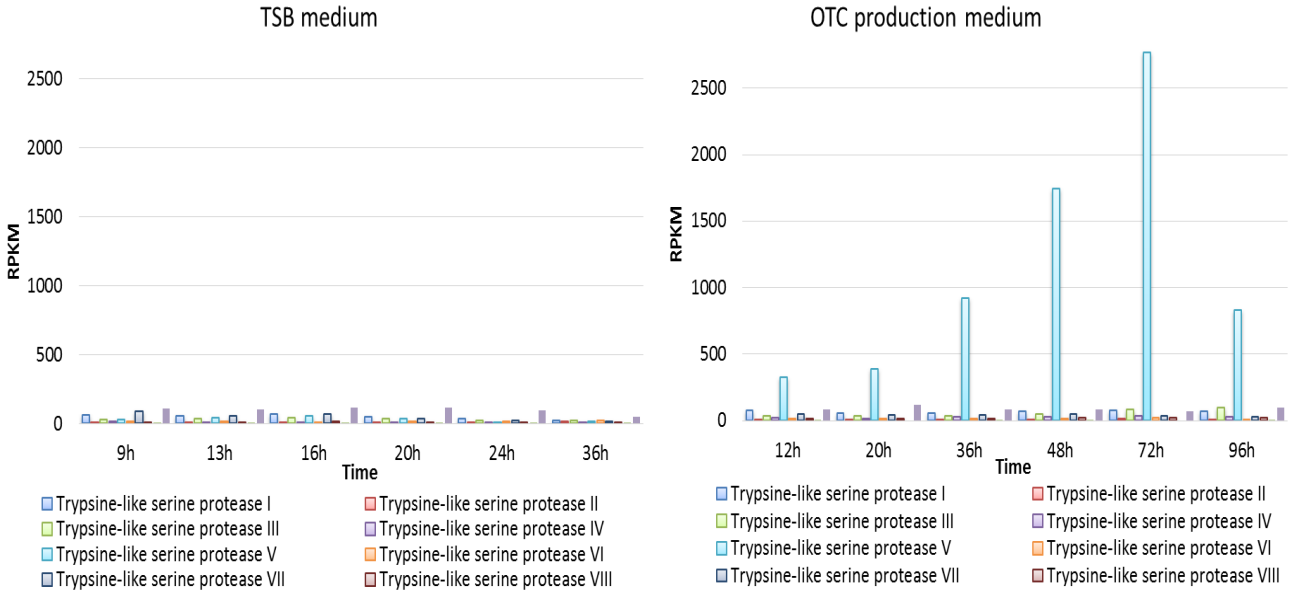


Figure 32. Transcriptome analysis of putative trypsine-like serine proteases identified in *S. rimosus* cultivated in two different culture media.

Metalloproteases belong to the group of proteases that contain one or two metal ions (Zn^{2+}, Mg^{2+}, Ni^{2+}, or Cu^{2+}) in their active centers and are produced by all species of plants, animals, and microorganisms [165]. Bacteria synthesize these enzymes as zymogens secreted into the environment to degrade proteins and peptides [165]. *S. rimosus* is equipped with 10 Zn^{2+} dependent metalloprotease homologues, including 4 thermolysin-like proteases, thermostable neutral zinc metalloproteinases used to produce aspartame [166]. The metal ion present in the active site of the other 4 metalloprotease homologues is unknown. Transcriptome analysis revealed that the metalloproteases with unknown metal ion and the thermolysin-like homologues are not transcribed significantly in TSB medium (Fig. 33 B and C). We identified only three actively transcribed Zn^{2+} metalloproteases (Zn^{2+} metalloprotease IV, VII & VIII) in this culture medium (Fig. 33A). Different situation is observed when *S. rimosus* is cultivated in complex medium. Besides the observed transcription of few putative Zn^{2+} metalloprotease genes, 2 putative metalloprotease (metalloprotease III & IV) and 2 putative thermolysin like proteases genes (thermolysin I and IV) are significantly more transcribed on this industrial medium (Fig. 33). Transcription of one of the putative metalloprotease gene designated as metalloprotease IV, transcribed at the beginning of the bioprocess, is decreasing its expression over time (Fig. 33C). Two thermolysin-like protease homologues (thermolysin I and IV) are also transcribed. The thermolysin IV is expressed during the entire cultivation process, while thermolysin I is transcribed only during the initial stages of OTC production (Fig. 33B). In addition to the identified proteases genes already described, a few additional putative protease homologues were identified, but with a clear role of maintaining proteostasis or preventing proteotoxicity during environmental and cellular stress, like the intracellular Clp ATP-dependent protease or Lon ATP-dependent protease genes, which were therefore not analyzed in our study [167, 168].

Importantly, independently on the transcriptome analysis data described in this thesis, no protease activity was detected experimentally in supernatant of the *S. rimosus* in TSB culture. In contrast, *S. rimosus* displays relatively high extracellular proteolytic activity in the complex medium. Proteolytic activity in the complex medium is particularly increasing towards the second part of the bioprocess, simultaneously with the production of OTC; suggesting that production of proteases and other hydrolytic enzymes belongs to the secondary metabolism.

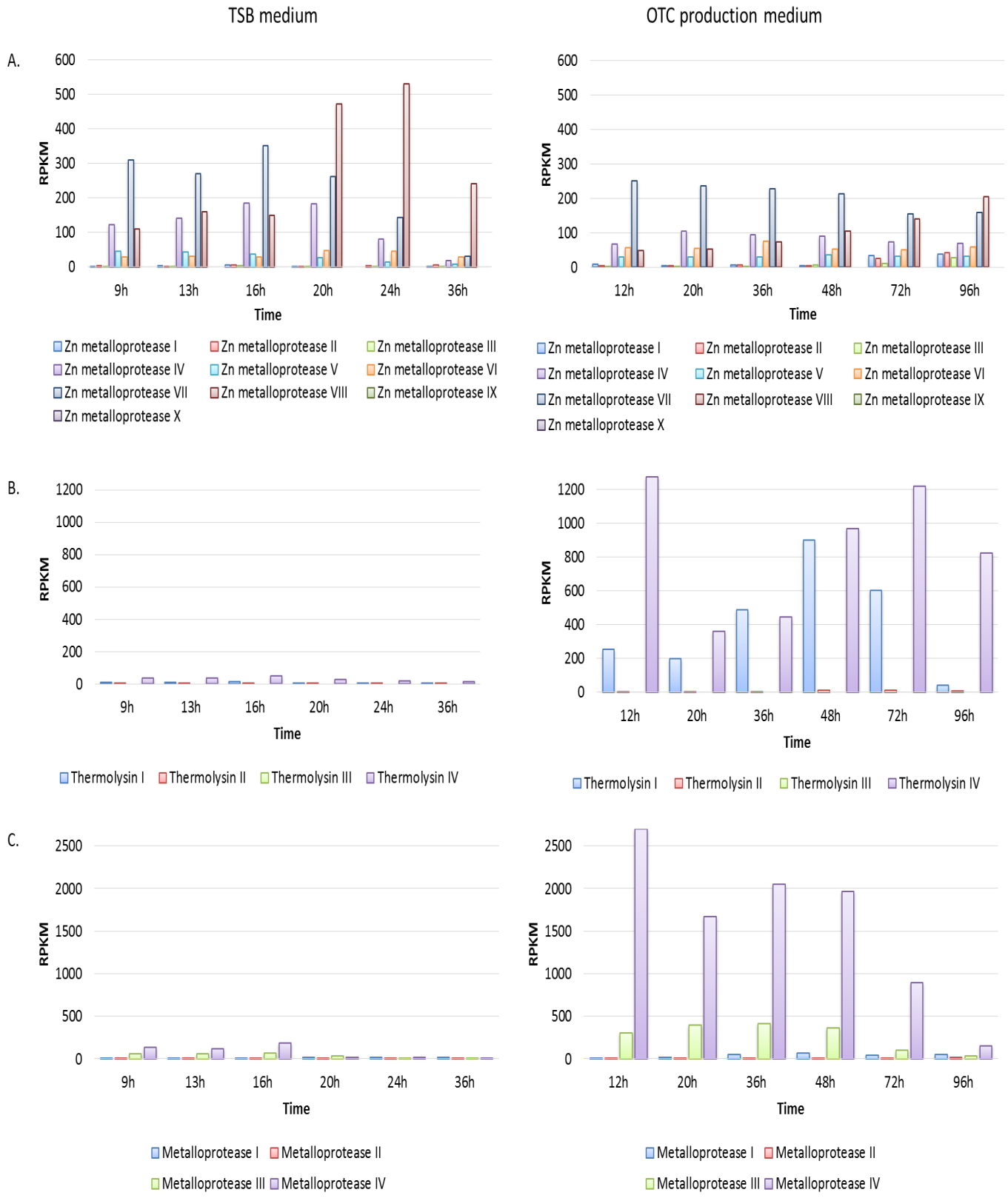


Figure 33. Transcriptome analysis of **A.** Putative Zn^{2+} dependent-like metalloprotease. **B.** Putative thermolysin proteases and **C.** unknown metalloproteases identified in *S. rimosus* cultivated in two different culture media.

11.8.4 *Streptomyces rimosus* as chassis for production of heterologous proteins

In the scope of the work on phytase AppA production and purification using *S. rimosus* host, we have observed the presence of many other extracellular proteins that attached to the HisTrap column. Interestingly, particularly three proteins of 40, 45 and 75 kDa in size displayed relatively high affinity to the Ni^{2+} column, thus interfering with the purification of target protein AppA (Fig. 34). At that time, we even considered the possibility that the different proteins displaying such high affinity to Ni^{2+} column might actually be different conformations or differently processed version of phytase AppA. We therefore aimed to establish identity of these proteins by applying LC-MS/MS. The peptide mass fingerprint revealed that the 40, 45 and 70 kDa proteins correspond to a glycerophosphoryl diester phosphodiesterase, a secreted tripeptidyl aminopeptidase and a phospholipase C, respectively (supplementary information). Fortunately the net charge of these three proteins is positive while the net charge of phytase AppA is negative at pH 7, therefore by increasing the stringency of the washes before elution with a higher concentration of NaCl (from 300 mM to 500 mM) allowed the purification of phytase, as it was demonstrated before (Fig. 27). The glycerophosphoryl diester phosphodiesterase and the phospholipase C might be involved in the glycerolphosphodiester metabolism [169]. The glycerophosphoryl diester phosphodiesterase degrades periplasmic glycerophosphodiesters to produce sn-glycerol-3-phosphate (G3P) and the corresponding alcohols [169]. The phospholipase C hydrolyzes sn-glycerol-3-phosphate (a glycerol based phospholipid) to yield diacylglycerol and a phosphomonoester [170]. The secreted tripeptidyl aminopeptidase cleaves tripeptides from the N-termini of proteins [171], thus this peptidase may be cleaving the secreted proteins reducing their stability [172]. The non-specific binding of several native proteins from *S. rimosus* to the Ni^{2+} column due to unspecific electrostatic interactions made us increase the stringency of the washes steps in the purification process of AppA, thus affecting the overall yield of secreted His-tagged recombinant AppA from *S. rimosus*. However, the presence of these native exported proteins of *S. rimosus* in the culture supernatant demonstrated the potential of *S. rimosus* as a chassis for the production of commercial enzymes and other heterologous proteins. Extracellular proteins identified in the supernatant present valuable source of signal sequences and promoter regions. Therefore, we decided to extend genome data analysis also to other hydrolytic enzymes present in *S. rimosus*. A summary of the enzymes identified in *S. rimosus* genome is presented in table 12 with the corresponding transcriptional status.

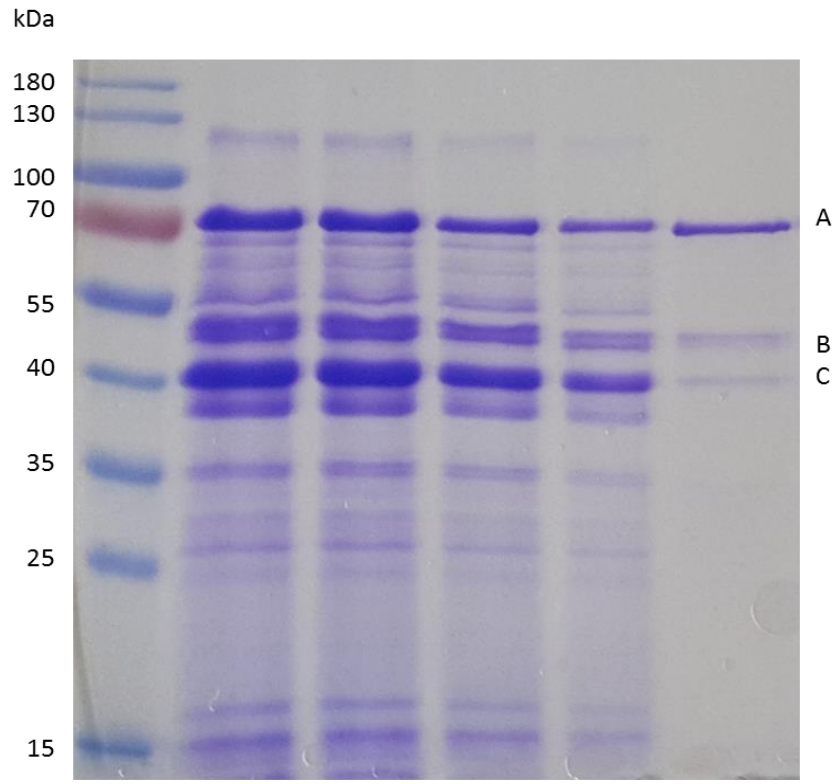


Figure 34. SDS-page electrophoresis of eluted fractions after his-tag purification of supernatant of *S. rimosus* "white" strain carrying pVF-3 plasmid cultivated in complex medium. The elution process was carried out by a linear gradient of imidazole. Every lane represents an eluted fraction with increasing concentration of imidazole. **A.** phospholipase C. **B.** Two bands are observed, one belongs to phytase and the other to secreted tripeptidyl aminopeptidase. **C.** Glycerophosphoryl diester phosphodiesterase. In-gel digestion and LC-MS/MS analysis allowed identification of the secreted proteins (supplementary data).

Table 13. List of putative gene homologues encoding hydrolytic enzymes with potential biotechnological applications identified in *S. rimosus* genome.

Gene coding for putative	# of putative genes	Transcriptional status	Function	Biotechnological application
Chitinases	8	off	Hydrolysis of chitin.	Bio-pesticides, production of chito-oligosaccharides used for medicine, cosmetics and dietary supplements [173].
Lipases	10	9 off 1 on	Hydrolysis of lipids.	Food, dairy, detergent and pharmaceutical industries [174].
Collagenase	1	off	Hydrolysis of collagen.	Medical, cosmetic and food production applications [175].
Nitrilases	2	off	Hydrolysis of nitriles.	Biosynthesis of acrylamide, food, chemical, pharmaceutical, wastewater treatment and textile industries [176].
α-Amylases	6	off	Hydrolysis of starch.	Food, fermentation, textile, paper, detergent, and pharmaceutical industries [177]
Cellulase	1	off	Hydrolysis of cellulose	Pulp and paper, textile, laundry, biofuel production, food and feed industry, brewing, and agriculture [178]

Chitosanase	1	on	Hydrolysis of chitosan.	Bioremediation of fishery wastes and production of chitosan oligosaccharides used by the pharmaceutical industry [179, 180].
Xylanases	5	off	Hydrolysis of xylan.	Management of waste, biofuels production, food, agro-fiber, and the paper and pulp industries [181].
Pectate lyases	2	off	Hydrolysis of pectin.	Plant fiber processing, especially in textile, paper, fruit juice, coffee, and tea fermentation industries [182]
Nitroreductases	2	off	Metabolize nitroaromatic and nitroheterocyclic derivatives	Bioremediation of nitroaromatic and nitroheterocyclic explosives, prodrug activation in gene directed anticancer therapies, activation of hypoxic anticancer nitro drugs in solid tumors, and enzyme-based biosensors for nitro-sensitive compounds [183].
Carbonic anhydrasen	3	1 on 2 off	Hydration and dehydration of carbon dioxide and bicarbonate	Carbon sequestration and biofuel production [184]
Transglutaminase	1	off	Formation of isopeptide bonds between proteins.	Manufacture of cheese and other dairy products, meat processing, production of edible films and bakery products [185].
Cholesterol oxidases	2	off	Catalyzes the oxidation of cholesterol into 4-cholesten-3-one.	Clinical laboratories (determination of serum cholesterol), Biocatalysis for the production of a number of steroids and insecticidal activity [186].
Glycoside hydrolase	1	off	Hydrolysis of glycosidic bonds.	Food industry [187]
Glycosyltransferase	1	off	Catalyzes the attachment of a sugar to an aglycon	Pharmaceutical industry (biosynthesis of glycosylated natural products) [188].
α-galactosidase	1	off	Hydrolysis of α-1,6-linked galactose	Food and feed, sugar and paper and pulp industries [189].
xylan 1,4-beta-xylosidase	1	off	Hydrolysis of non-reducing end xylose residues from xylooligosaccharides	Bakery and feed industry, production of D-xylose for xylitol manufacture, and deinking of recycled paper [190].
Pullulanases	4	off	Hydrolysis of starch and pullulan.	Starch-based industries especially those aimed for glucose production [191].

12 Discussion

Many research applications currently require the purification of relatively high amounts of active proteins for either its study (biochemical analysis, X-ray crystallography, etc.), or for its direct use (industrial enzymes, therapeutic and diagnostic applications, etc.). To date, no such a host from which all the heterologous proteins can be successfully produced in the correctly folded form in sufficient quantities is available. Therefore, it is of high importance to provide a variety of expression systems in order to increase the possibilities to get functional proteins either, for academic or industrial purposes. *E. coli* is often the first choice for expression of the target protein, and despite the progress experienced in the last decades, the development of new genetically modified strains and plasmid expression systems, *E. coli* often fails to produce target functional proteins; often producing incorrectly folded proteins or even inclusion bodies [28]. Alternative microbial hosts like *Bacillus subtilis*, *Streptomyces lividans* or *Pichia pastoris* offer other advantages for protein expression; including the capacity to export the recombinant proteins extracellularly, thus avoiding the formation of insoluble aggregates [32, 70, 192]. However, the yields of recombinant protein obtained by these hosts are often too low or hyper-glycosylated, in the case of *P. pastoris*. Considering a number of advanced properties of *S. rimosus*, as discussed in this thesis, we aimed to develop suitable gene tools and test the capacity of *S. rimosus* as heterologous host for protein production. The gene tools developed for *S. rimosus* allowed us to compare this expression system to other commonly used hosts. For this purposes, we used model reporter system XylE and commercial phytase AppA, which is a native protein of *E. coli*, produced industrially also by *E. coli*. During the last decade, a number of research groups have attempted to produce this enzyme in different hosts [120, 125] such as *Saccharomyces cerevisiae*, *Pichia pastoris* and *Streptomyces lividans*, thus representing ideal model enzyme for comparative evaluation of a new expression system.

12.1 Development of gene tools for production of heterologous proteins in *Streptomyces rimosus*

12.1.1 Stable integration of target genes into *Streptomyces rimosus* genome

Heterologous DNA can be introduced into microorganisms using replicative plasmids, however plasmids offer limited control of copy number, and segregational instability can be a significant issue even in medium containing the corresponding selective antibiotic [193, 194]. In addition, the use of antibiotics to maintain the replicative plasmids increases the overall bioprocess cost and generates environmental concerns [195], thus limiting their industrial utility. On the other hand, irreversibly integration of exogenous DNA directly into the host chromosome offers the advantage of stable maintenance of target DNA without the need of antibiotic selection [196]. Generally, stable integration of DNA into streptomycetes is methodologically very tedious work when carried out through homologous recombination. A single homologous recombination-mediated event can reversibly integrate the entire plasmid into the

chromosome, resulting in unstable single-crossover event, which can revert to the wild-type genotype via homologous recombination [194]. The desired stable double-crossover event is much more difficult to achieve, as they result from the two-step procedure of two homologous recombination events in a single cell [194]. In streptomycetes, both single-crossover and double-crossover mutants are obtained after introducing the plasmid with the homologous regions, thereafter, colonies have to be propagated (sub-cultured) one or two generations to increase the ratio of double-crossover mutants. Double-crossover mutants are subsequently screened by replica plating approach, which is often extremely tedious and time-consuming work [197]. Additionally, the filamentous nature of these microorganisms, which contains more than one cell in a single filament, complicates even more the double cross-over selection process due the difficulty to isolate single cells with desired genotype. This is the reason why homologous recombination in streptomycetes is a complex procedure, which can take a number of months to isolate a single mutant with desired genotype [49].

To overcome the drawbacks of homologous recombination, phage-directed site-specific integrases appeared like the best option to stably integrate target DNA into the chromosome of a target streptomycete. Bacteriophages typically insert their genomes into the host chromosomes *via* integrase-mediated site-specific integration between two sites, the *attB* present on the bacterial genome and *attP* from the phage, thus forming the *attL* and *attR* sites stablishing the lysogenic state [198]. At least five temperate phages (ϕC31, ϕBT1, R4, TG1 and SV1) are known for *Streptomyces* species [90, 95]. In *S. rimosus* only ϕBT1 and ϕC31-based vectors have been used successfully to integrate DNA into the chromosome of this bacterium. pSAM2, a self-transmissible plasmid [199], can also integrate DNA in *S. rimosus* chromosome [48]. Hranueli et al. established that the *attB* site targeted by the integrase of pSAM2 plasmid is located in the A fragment of *S. rimosus* chromosome restriction map (Fig. 19) [53]. However, pSAM2 derived vectors are not often preferred choice to integrate DNA in streptomycetes because the *attB* site overlaps the gene coding for the $tRNA^{pro}$ [200]. Besides, disadvantage of pSAM2 integrase in *S. rimosus* is its relatively low integration efficacy (Petković H, personal communication). In this work we intended to elucidate the DNA sequence of the *attB* sites, where ϕC31 recombinase-containing vectors integrate in *S rimosus* chromosome, and for this purpose we analyzed *S. rimosus* transformants containing chromosomally-integrated pYAC-OTC vector.

The pYAC-OTC plasmid containing the entire OTC gene cluster, which carries the ϕC31 integrase**,** was constructed by Acies Bio Ltd. (Slovenia, unpublished data). It is known, that the ϕC31 recombinase has broader spectrum of hosts, and has been widely used to introduce DNA in *Streptomyces* spp.. It has been demonstrated by Combes et al [82], that ϕC31 integrase often mediates integration into multiple sites in *S. coelicolor* and *S. lividans*, both carrying in their chromosome the perfect *attB* site and the so called secondary or pseudo *attB* sites [82]. The ϕC31 *int/attP* integration system was tested in *S. rimosus* and it was concluded that *S. rimosus* has a single integration site (the perfect *attB* site), when the DNA was introduced by conjugation [128]. However, in the present study, the introduction of DNA into *S. rimosus* was carried out by electroporation instead of conjugation, because the electroporation procedure with *S.*

rimosus is rapid and ensures very high efficiency [60]. The capacity of ϕC31 recombinase to integrate the entire OTC cluster (32 kb) in *S. rimosus* M4018 was already demonstrated (Unpublished data, Acies Bio Ltd.; Slovenia). We have demonstrated that plasmid vector containing ϕC31 recombinase can integrate at two loci via independent recombination (see 11.1). Interestingly, Thamchaipenet and colleagues identified a single integration site in *S. rimosus* [128]. In their study they have used the R7 and M4018 *S. rimosus* strains and the conjugative plasmid pIJ8600 (pSET152 derivative with ϕC31 recombinase). Thamchaipenet and colleagues demonstrated that transconjugants of both strains gave identical southern hybridization pattern, implying that there was a single integration site at which ϕC31 integration system integrated in *S. rimosus* chromosome [128]. In contrast, our results indeed confirmed the presence of the perfect *attB* site, as demonstrated by Thamchaipenet and colleagues, and an additional secondary or pseudo *attB* was found in the R7, M4018 and "white" strains (see 11.1). The only obvious difference between Thamchaipenet and colleagues and work carried out in this study was the methodology used to introduce the plasmid into *S. rimosus*. While we introduced the DNA by electroporation, Phornphisutthimas et al. used conjugation instead [128]. *S. coelicolor* and *S. lividans* also contain more than one *attB* site in their chromosomes. These *Streptomyces* species have multiple *attB* sites (the original *attB* and three additional pseudo *attB* sites) [82]. Therefore the integration of ϕC31 *int/attP* in several loci via independent recombination is a common feature in several *Streptomyces* species, thus confirming that ϕC31 integrase has relaxed specificity. The investigation of the *attB* sites present in *S. coelicolor* and *S. lividans* was carried out with the plasmid pSET152. Similarly to the study carried out in *S. rimosus* by Phornphisutthimas et al.[128], the method used to introduce the pSET152 plasmid into these hosts was also by conjugation. We did not carry out comparative study, where conjugation and electroporation procedures would be directly compared, but it is possible that these two methods influence the efficacy of integrase mediate-chromosomal integration of the plasmid DNA. However, independently of the methods used, our plasmid rescue experiments, combined with PFGE and Southern hybridization analysis clearly demonstrated the presence of at least two potential integration sites specific for ϕC31 recombinase in *S. rimosus* genome.

XylE and AppA activities were much lower when these recombinant genes were integrated in *S. rimosus* chromosome using ϕC31 recombinase compared to the activities obtained using the multi-copy plasmid pVF. However the capacity to integrate foreign DNA stably at single copy is a valuable approach, and there are at least 3 recombinases (TG1, ΦBT1 and SV1) that have not been evaluated in *S. rimosus* to express heterologous proteins. In eukaryotes, the location of an integrated heterologous gene on the chromosome is known to have effect on its expression [201]. Besides, Lee and colleagues [201] demonstrated that the expression of the reporter gene *gfp* inserted in *E. coli* at different locations of the genome, under the same promoter varies by ~300-fold, depending on its precise position on the chromosome [201]. Therefore, identification of the integration locations of the three remaining recombinases in *S. rimosus* is of interest. This kind of information would be very valuable when considering the engineering efforts in *S. rimosus* when using different recombinases.

In addition to heterologous protein expression, the heterologous expression of any biosynthetic pathway is an indispensable tool for drug discovery and development of novel natural products [202]. Therefore, it is not always the main aim to produce high yields of the target recombinant protein, but rather functional enzymes catalyzing the biosynthesis or biotransformation of the target compound. Actinomycetes are among the richest sources producing numerous structurally diverse compounds. Therefore, the bacteriophage integrase containing vectors represent very powerful tool in genetic engineering of antibiotics-producing actinomycetes.

12.1.2 Construction and properties of the replicative plasmid pVF in *S. rimosus*

To date, the pET vectors are the most useful system for production of recombinant proteins in *E. coli*. These vectors offer numerous possibilities to produce recombinant proteins in *E. coli*, thus being one of the principal features of the genetic tool kit of this microorganism to overexpress proteins. Leaving aside the promoter and fusion tags included in these plasmid constructs, the major feature of pET vectors is the pBR322 origin of replication which allows up to 20 molecules of vector per cell [203], highlighting the importance of a multi-copy vector in the final yields of recombinant proteins produced by *E. coli*. The value of extra copies, when using a replicative vector is also observed in *S. lividans* and *S. coelicolor*, where overexpression of recombinant proteins in these hosts is boosted with the use of pJ101-based vectors with up to 300 copies per cell [204]. To improve the abilities of *S. rimosus* as heterologous host, it was important to develop stably maintained replicative plasmid, thus significantly increasing the yield of the target enzyme, as demonstrated in this work. Several attempts had been carried out by Pfizer, to identify a stable replicative vector for *S. rimosus*, all of them based in the conjugative plasmid pJ101. However, *S. rimosus* showed the incapability to harbor a replicative plasmid with conjugative elements [111, 123]. *In-vivo* rearrangements of pJ101 in *S. rimosus* generated the pPZ12 vector, the first stable multi-copy vector in *S. rimosus*, from which the bi-functional cosmid vector pPZ74 was developed [111]. This was extremely valuable achievement by Pfizer´s scientist. It is worth mentioning that in 1984, when the plasmid was described [111] and in 1985 when the patent was filled [123], only few promoters were available in the scientific community. We recovered pPZ12 vector and fused it with pUC19 plasmid creating the shuttle vector pVF, a plasmid which is easily manageable in both, *E. coli* and *S. rimosus*.

The sequencing of pVF vector revealed that all the conjugative machinery of pIJ101 is absent, only retaining the origin of replication, the thiostrepton resistance gene and the putative genes orf56 and *korB*, whose role is believed to control the copy number of the replicative vector [205]. Future work could include removing *korB* and orf56 genes, to confirm whether or not they are essential for the stability and copy number of pVF. In the case these putative genes do not have essential function in pVF vector, the size of the plasmid could be reduced for additional 2.5 kb. This way, the vector would be more versatile for cloning purposes, and potentially the vector could stably accommodate larger DNA inserts.

In addition to the genetic stability displayed by pVF vector in *S. rimosus*, we also evaluated its segregation stability. When introducing pVF vector, besides increasing recombinant protein production in *S. rimosus*, we have also demonstrated the relatively high stability over long time in the absence of the selective marker thiostrepton. We have demonstrated that around 30% of the cells lose the pVF plasmid in TSB medium in 24 hours, which is still relatively acceptable in the short bioprocess. Therefore, it is possible to envisage the production of recombinant protein in a rich medium, where the biosynthetic process is completed in less than 24 hours, such as TSB, without supplementation with antibiotic (thiostrepton). In contrast, in the complex medium, production process takes more than two days, thus the addition of thiostrepton is of critical importance to maintain plasmid copy number high and thus achieving maximum yields of target recombinant protein. Phytase activity obtained with *S. rimosus* pVF-*tcp830/srT/appA* transformants was much higher when antibiotic marker (thiostrepton) was added into the complex medium. In the absence of thiostrepton, AppA activity was drastically reduced around 40%, compared to the maximum activity in the complex medium with thiostrepton. The use of antibiotics as selective markers, in this particular case thiostrepton, would increase the production cost at industrial settings significantly. On the other hand, this would also potentially cause issues regarding the registration of produced products, considering antibiotics cannot be present in the final product. Recombinant proteins, intended to be used for industrial and/or commercial applications have to be obtained from microorganism classified as GRAS (Generally Recognized As Safe) and, with strong preference to use media without antibiotics. *S. rimosus* is already classified as GRAS and the use auxotrophic markers could remove the need for antibiotic selection.

The so called metabolic-dependent markers were successfully adapted in yeasts, after generation of mutants for genes encoding enzymes in pathways for the biosynthesis of important metabolic building blocks, such as amino acids or nucleotides [206]. Auxotrophic markers can then compensate the nutritional deficiency caused by the mutation, therefore the mutant strain is cultivated in a medium with the specific nutrient and after transformation, only cells carrying the plasmid with the auxotrophic marker will be able to grow in the medium without the required nutrient [206]. Auxotrophic markers were also successfully adapted in *Mycobacterium bovis* by knocking out the amino acid leucine *leuD* gene. Expression of *leuD* on a plasmid not only allowed complementation, but also acted as a selectable marker [207]. This strategy could be applied to pVF vector in order to remove the thiostrepton selective marker, thus allowing production of heterologous proteins by *S. rimosus* in antibiotic-free medium.

A stable replicative vector was one of the most important gene tools needed for genetic manipulation of *S. rimosus*. Thus, pVF *E. coli-Streptomyces* shuttle vector, stably replicating in *S. rimosus* now represents extremely valuable tool, which can now be used for expression of recombinant proteins and other metabolites in *S. rimosus*. This shuttle vector is easy to handle in both, *E. coli* and *S. rimosus*, and its multi-copy state facilitates its isolation directly from both microorganisms. The multi-cloning site of pUC19 allows multiple options to clone diverse DNA fragments, and the good genetic stability observed implies the potential use of pVF vector to produce

recombinant proteins, as demonstrated by significant increase of phytase activity (see 11.6). Although not yet tested, this plasmid may be also useful in other *Actinomyces* spp. host systems.

12.1.3 Construction and evaluation of *PnitA-NitR* and *tcp830* promoters in *Streptomyces rimosus*

The *PnitA-NitR* and *tcp830* promoters probed to be useful promoters to produce heterologous proteins in *S. rimosus* when used with the replicative plasmid pVF. However, the chromosomally integrated versions of these promoters by using integrative plasmid pAB04 displayed weak expression. Thus suggesting that with copy number significantly increased, the efficiency of *PnitA-NitR* and *tcp830* promoters was increased as well. Therefore, when expressing recombinant proteins in *S. rimosus* using the multi-copy vector pVF, weak promoters are essential to maintain the genetic stability of pVF vector. In contrasts, the strong promoter *PermE** incorporated to pVF vector induces genetic and morphological instability in *S. rimosus*. All *S. rimosus* transformants harboring either *PnitA-NitR* or *tcp830* in the multi-copy vector pVF maintained the wild type phenotype, and the plasmids isolated from these transformants preserved the original architecture. Unfortunately, both promoters behaved constitutive rather than inducible.

Herai et al. [97] reported that transformants of *S. lividans* TK 24 carrying pHS19 + *PnitA-NitR* was able to produce 45 U/mg of specific activity of XylE when the recombinant strain was cultivated for 120 hours in YEME medium [97]. The nitrilase promoter did not display such strength in *S. rimosus* when expressing XylE, only reaching 1 U/mg. However, XylE activity reached 42 U/mg when the reporter gene was expressed with *tcp830* promoter and cloned into pVF vector, thus suggesting the potential of *S. rimosus* which reached similar XylE activity as compared with *S. lividans* but in only 24 hours of cultivation, 4 days before *S. lividans* reached maximal activity. In addition, *S. rimosus* media and process optimization has not yet been optimized, therefore there is a plenty of room for improvement.

The constitutive activity of *tcp830* promoter was somehow unexpected considering that none of the around 100 *tetR* homologues identified in *S. rimosus* genome repressed *tcp830* promoter. Therefore, our expectation, that at least one *tetR* homologue could turn off the activity of *tcp830* promoter was not correct. It seemed that no TetR homologue is specific enough to repress the activity of the *tcp830* promoter. The *tetRiS* gene, the original *tcp* repressor, was therefore incorporated into pVF-*tcp830/xylE* vector under the regulation of the constitutive promoter *SF14* from *S. ghanaensis* phage I19 [102], however the regulator didn't shut down the *tcp830* promoter. By applying RT-PCR analysis, we confirmed that *tetRiS* was actively transcribed, thus the constitutive activity of *tcp830* promoter was difficult to explain. *S. rimosus* is the natural producer of tetracycline antibiotic oxytetracycline, therefore, one potential hypothesis was that *tcp830* promoter was auto-induced by *S. rimosus* native OTC production. However, XylE expression was evaluated in TSB medium, where OTC is not produced at high yields. Transcriptome data on *otc* gene cluster

expression demonstrated very low transcription levels, if at all, when applying these culture media. However, even small amounts of OTC produced, could trigger the activity of promoter, considering the promoter under *tetR* regulation is already activated at concentrations below 1 µg/mL. To confirm whether or not OTC was inducing the *tcp830* promoter, we transformed the pVF-t*cp830/xylE/tetRiS* plasmid into *S. rimosus* "white" strain. In this strain OTC production was abolished, considering that OTC gene cluster was deleted. Thus *tcp830* promoter should be off in the case OTC was the responsible for *tcp830* activation. In the "white" strain (ΔOTC) the promoter *tcp830* also displayed constitutive phenotype. Thus, based on this information, we can conclude that OTC is not activating the promoter. Rodriguez-Garcia et al. who designed the *tcp830* promoter for *Streptomyces* concluded that for some species of *Streptomyces* this promoter doesn't have an off state [98]. Therefore, it can be concluded that *S. rimosus* also belongs to this group of species.

In conclusion, for now, we did not manage to develop an inducible promoter for *S. rimosus* host. However, two valuable promoters are now available to produce heterologous proteins in this host. The combination of the replicative vector pVF with *tcp830* promoter increased the activity of the reporter system XylE for almost 140 times, compared to the integrative-based vector containing the ϕC31 recombinase combined with *PermE** promoter. Importantly, the new expression plasmids (pVF-*PnitA-NitR* and pVF-*tcp830*) also demonstrated high genetic stability, considering all *S. rimosus* transformants evaluated harboring pVF replicative vector with *PnitA-NitR/xylE* or *tcp830/xylE* displayed reproducible XylE activity, thus clearly demonstrating the robustness of this heterologous protein production platform.

12.2 Heterologous expression of recombinant phytase (AppA) in complex medium using *Streptomyces rimosus* host

To further evaluate the potential of *S. rimosus* for heterologous protein production and also to test the strength of the genetic tools developed in this work, we selected phytase AppA as model system. Phytase (AppA) was selected because i) Phytase is particularly difficult to express as a functional protein in other microorganisms (even in *E. coli*, the natural host) [114, 120, 125, 208], ii) The expression of phytase AppA was attempted also in other hosts of industrial importance such as *E. coli* [114, 116], the yeasts *P. pastoris* and *S. cerevisiae* [125, 208] as well and importantly, also in *Streptomyces lividans*, so far the best *Streptomyces* heterologous host [120], thus allowing direct comparison when evaluating *S. rimosus* as expression host, iii) The enzymatic assay to evaluate AppA production is relatively easy to perform, thus simplifying evaluation procedure and iv) this enzyme is industrially relevant product.

When evaluating *S. rimosus* pAB04-*PermE*/srT/appA* strains at shaker scale, we could reach up to 3 U/mL of extracellular AppA activity in complex medium. Therefore, it was our immediate aim to evaluate maximal AppA yields which could be achieved based on the replicative vector pVF in *S. rimosus* (section 11.6). Production of recombinant AppA by *S. rimosus* was only observed in complex medium. Surprisingly, although it was possible to detect XylE activity in TSB medium by the same plasmid constructs,

attempts to produce AppA in TSB medium failed, considering no AppA enzymatic activity was observed [78]. It is difficult to explain why phytase AppA is only produced in complex medium. AppA was designed to be exported out of the cell by fusing a trypsin-like proteinase signal sequences (*srT*), specific for the Sec-pathway. Transcriptome analysis suggests that the genes involved in the formation of the Sec-pathway are weakly expressed after 24 hours on the TSB medium, thus *S. rimosus* likely cannot secrete the recombinant protein in this condition. Besides, the inappropriate redox environment of the cytoplasm could also influence the production of the enzyme. Both eukaryotic and prokaryotic organisms maintain their cytoplasm reduced and, consequently, disulfide bond formation is not favored in the cytoplasm [118]. Phytase has four nonconsecutive disulfide bonds, thus it is quite possible that the intracellularly recombinant protein, which is not exported out of the cell, undergoes incorrect folding. Despite this potential drawback, recombinant phytase was efficiently produced as an extracellular protein by *S. rimosus* in complex medium. This is not surprising considering that the transcriptome analysis indicates the upregulation of all the genes involved in the formation of the Sec-pathway in this industrial-like medium.

We proceeded to evaluate the efficiency of the new expression plasmids in complex medium with *S. rimosus* transformants carrying the pVF multi-copy based vectors containing *PnitA-NitR* or *tpc830* promoters, and equipped with the serine like protease signal sequence (*srT*) fused to the phytase (*appA*) gene. We have regularly observed significant activity of AppA, when using this plasmid constructs, thus indicating the high stability of pVF vector. Analogously, as observed in the experiments with *xylE* reporter gene, the constitutive nature of both promoters was also detected in complex medium. Recombinant phytase production was increased significantly when pVF-based vectors were used. 3 U/mL of phytase activity were obtained with pAB04-*Perm*/srT/appA S. rimosus* transformants compared to 20 U/mL obtained with pVF-*tcp830/srT/appA* transformants, thus achieving almost 7-fold increase in AppA activity. The expression of phytase AppA fused to *srT* signal peptide under *tcp830* promoter, when incorporated to the replicative vector pVF, ensured relatively high extracellular activity of AppA. This *S. rimosus* recombinant strain allowed simple and straight forward purification of AppA through affinity chromatography by using supernatant of *S. rimosus* culture. This way avoiding the need to disrupt the cells, thus significantly facilitating the purification process (see 11.7).

12.3 Overall assessment of the *Streptomyces rimosus* host and expression plasmids

It is important to stress that the gene tools developed in this study for *S. rimosus* as host system represent only initial efforts towards the industrially-applicable heterologous protein production platform. When comparing to the available expression systems, we believed that the genetic tool kit developed in the scope of this work for *S. rimosus* host could in the future present a potential alternative system; particularly, when other expression systems fail to produce the target protein in the correctly folded form. In addition to catechol 2,3-dioxygenase (XylE), which is relatively reliable reporter gene working in actinomycetes [49], we have carried out our work

with commercial phytase gene *appA* as a model system. The phytase AppA from *E. coli* has received special attention because among the phytase enzymes studied, this one has been reported to demonstrate the greatest specific activity degrading phytic acid compared to those from other sources [209]. Independently on the target protein, it is generally very difficult to predict optimal expression host/platform. The production of the target protein becomes even more difficult when the target protein contains post-translational modifications, such as disulfide bonds, which are of critical importance for correct protein folding, stability, and activity of the target protein [210]. Therefore, we suspected that the efficient production of correctly folded and active protein AppA by *S. rimosus* would demonstrate the potential robustness and versatility of this host. A large proportion of proteins contain disulfide bonds. For example, analysis of the human genome revealed that 30% of the human proteins are predicted to target the endoplasmic reticulum (ER) and half of those are predicted to form disulfide bonds [211]. Therefore, it is of critical importance to have available expression systems, which can efficiently formulate disulfide-bonded correctly folded proteins.

AppA overexpression in *E. coli* resulted in the production of inclusion bodies [116]. However, *E. coli*, the native host of AppA, still is the best candidate to produce this enzyme industrially. *E. coli* has often a limited capacity for formation of proteins containing disulfide-bonds, considering that *E. coli* processes disulfide bonded proteins in the periplasmic space. The oxidative environment and the presence of periplasmic chaperons facilitate the formation of disulfide bonds *in-vivo* in the periplasmic space [210, 212]. As a consequence, *E. coli* secretes and anchors AppA to the plasma membrane where it is exposed to the environment. However, when the protein is overexpressed intracellularly in *E. coli*, this host is not supporting efficient processing of disulfide bonds, therefore incorrect post-translational modifications often result in the production of inclusion bodies of AppA by *E. coli* [114].

P. pastoris, on the other has already proven as very useful industrial host for protein production, and attempts to produce AppA in this host have already been explored [115, 120, 125, 208]. *P. pastoris* has a potent secretory apparatus [32] facilitating secretion of phytase. The phytase yields obtained with this methylotrophic yeast (117 U/mL) were 3.5 fold higher in comparison to *E. coli* [120]. However, the enzyme was hyper-glycosylated making necessary an extra step of enzymatic deglycosylation, hence significantly increasing the production costs [120], which probably complicates the development of the industrial process. The yields of AppA were later increased to 204 U/mL by medium optimization and process improvement, nonetheless the recombinant enzyme was extensively glycosylated [208]. Attempts to produce phytase (AppA) using *S. cerevisiae* were also carried out. However, the phytase activity yields in this system were much lower than those produced by *P. pastoris* [125]. *Streptomyces lividans* was also evaluated to produce phytase AppA [120] because, like *P. pastoris, S. lividans* has functional secretory machinery [70], which facilitates the production of extracellular AppA. However, as described in the literature, the AppA yields generated by *S. lividans* host were not very high (0.95 U/mL), although the phytase was not glycosylated [120].

The highest AppA activity achieved by *S. rimosus*, in the scope of this work reached around 20 U/mL. The results published by other research groups using the most popular expression hosts, such as *E. coli* and *P. pastoris*, demonstrate that *S. rimosus* as host system still needs significant improvements, to be qualified as industrially useful. However, it is very difficult to give an estimate on economic and industrial issues of all tested heterologous hosts, considering commercial aspects do not depend only on the final yield of the target product, but also on the upstream and downstream process costs, which are related to the medium composition, issues related to the consumption of energy and other economic parameters. The quality of the product, in addition to the quantitative parameters also play very important role, when estimating the process parameters. Unfortunately, information related to the properties of heterologously produced enzyme AppA from different heterologous hosts are not available in the literature. For example, phytase AppA produced by *P. pastoris* is glycosylated. Thus, although the activity of *P. pastoris* phytase is very high, based on the literature estimate, ten times higher than the yield of AppA achieved in *S. rimosus*, information related to other properties of the recombinant enzyme is not available. When developing new process for AppA production, for example, it would be of great advantage to achieve similar of even improved properties of the enzyme; which might be influenced by glycosylation and other conditions that could influence different folding of AppA. Equally, the same issues apply to any other heterologously produced proteins of potential industrial value. On the other hand, when considering *S. lividans* as heterologous host, relatively slowly growth of this strain, in comparison to *S. rimosus*, represent disadvantage, considering that industrial bioprocess might take much longer, this way significantly increasing costs. However, the processes for AppA production with *S. lividans* or *S. rimosus* can most likely be improved by industrial medium and process development approaches, considering that all experiments with *S. lividans*, as reported in the publication by Stahl et al. [120] have been carried out on the standard laboratory media used primarily in the scope of academic research. Equally applies for *S. rimosus* host, which we have tested in this work. Media used in our experiments have originally been developed for the production of oxytetracycline. In addition, *S. rimosus* host strains used in this study very likely are not the most suitable strains for industrial production of phytase or any other heterologous enzymes, considering they were optimized during strain improvement regimes over 6 decades for the production of oxytetracycline. Therefore, independently on the *Streptomyces* species host used, in both cases, intensive strain and process improvement, accomplished with selected metabolic engineering strategies, would have to be carried out in order to reach desired economic targets.

Important morphological property of *S. rimosus* is also its rapid growth rate. When using the complex medium, maximal yield of phytase is reached after 4 days of incubation, thus significantly longer bioprocess, when compared to bioprocess with *E. coli* host, where bioprocess takes 2-3 day [114]. However, *S. ri*mosus can grow much faster in rich media such as TSB, and can reach maximal biomass in time comparable to the processes based on *E. coli* host. Unfortunately, phytase production by *S. rimosus* was not observed in TSB medium. It is difficult to explain rationally the reason for not being produced in TSB, considering that XylE is produced when using identical strains. On the other hand, complex medium used for production of AppA by *S. rimosus* is

much cheaper compared to TSB or media used for *E. coli*, thus significantly reducing the bioprocess costs. *S. rimosus* is an organism with GRAS status, and it has all stages of industrial bioprocess already developed to the industrial production scale, which is also significant advantage of *S. rimosus* as heterologous protein expression system for production of target proteins. However, for different reasons, the use of entirely soluble media, such as TSB would have numerous advantages compared to complex medium. Therefore, it would be of great importance to adapt production process when using *S. rimosus* on media containing entirely soluble media components.

12.4 Challenges and future perspectives of *Streptomyces rimosus* platform to produce heterologous proteins

As discussed in the literature, four factors are essential to achieve production of extracellular heterologous proteins [70]: i) a robust host, ii) well established bioprocess parameters (media composition and cultivation strategies) iii) efficient gene tools including plasmids and promoters to introduce and express the target gene, and iv) effective signal sequences which allow secretion of target protein. In this work we have mostly used *S. rimosus* M4018 to produce phytase, and this strain was subject of intense selection over decades to produce oxytetracycline (OTC) [48]. Besides, phytase was produced in a medium originally developed to produce OTC. Moreover, the gene tools (e.g. promoters) were selected from different hosts based on the available data from literature and adapted in *S. rimosus*. Therefore, it is reasonable to expect that *S. rimosus* platform can still undergone substantial further improvements to reach much higher yields of target heterologous protein, when developing the four factors mentioned before.

S. rimosus M4018 produces relatively high yield of oxytetracycline (OTC), which is in the scope of heterologous protein production actually undesired property. However, the capacity to produces OTC can easily be removed by deletion of the entire gene cluster encoding OTC production. In our work, in addition to “4018” strain, we have also routinely used the so-called “white” strain, which has *otc* gene cluster deleted Even though the “white” strain lacks the *otc* gene cluster, this strain is still similar to M4018, therefore both strains were developed originally for OTC production and not for the production of heterologous proteins. Alternatively, the *S. rimosus* wild type strain R7 does not produce such high yields of OTC compared to M4018 strain. Therefore, the production of phytase by R7 strain could bring valuable information regarding the most optimal *S. rimosus* strain to produce heterologous proteins, and to continue further development of the most optimal *S. rimosus* host.

Independently of the *S. rimosus* strain selected to produce heterologous proteins, the chassis of this microorganism needs further adaption for protein production. The genome sequence analysis of *S. rimosus* R7 strain revealed the presence of numerous putative secondary metabolite biosynthetic gene clusters, in addition to OTC [55]. In the complex medium, and based on transcriptome analysis, only two biosynthetic gene clusters are transcribed by *S. rimosus* M4018, OTC and the siderophore enterobactin. The biosynthetic gene clusters in *S. rimosus* sum up to around 0.5 Mb of

"uncharacterized genetic information" which could be removed, thus simplifying the genotypic background of *S. rimosus*. This way likely, enhancing *S. rimosus* capacity of heterologous protein production, similarly as in the case of *E. coli* K12 with reduced genome [213]. *E. coli* strains with reduced genome displayed higher electroporation efficiency and accurate propagation of recombinant genes and plasmids compared to the wild type strain [213]. Another example is *S. avermitilis* with reduced genome, denominated SUKA 5 and SUKA 17, which were capable to produce higher yields of target secondary metabolites compared to the wild type strain [214]. Consequently, removing unnecessary genetic information already proved to be a useful strategy in the scope of developing microbial chassis. In addition, *S. rimosus* transcriptome data also support our efforts to identify essential pathways facilitating production of heterologous proteins as well as deleting undesired metabolic pathways.

Ideally, a chassis specialized to produce heterologous proteins should not produce endogenous proteases, which can degrade the target protein. Transcriptome analysis of S*. rimosus* indicated that a few proteolytic enzymes are being transcribed, thus suggesting a potential risk for degradation of target heterologous proteins by native proteases. The optimal *S. rimosus* chassis specialized for protein production should therefore have reduced content of endogenous proteinases, thus avoiding undesired proteolysis of heterologous proteins produced by *S. rimosus*. This strategy was already applied in *E. coli*, *B. subtilis, S. cerevisiae* and *P. pastoris* [32, 215-217], showing that these protease-deficient strains display increased productivity of the target heterologous proteins. Thus knocking out highly transcribed proteases in *S. rimosus* could likely improve the chassis of this host for enhanced heterologous protein production.

In addition to the proteases produced by *S. rimosus*, other proteins could also be interfering with the production of heterologous proteins in this host. The first attempts to purify recombinant phytase from *S. rimosus* supernatant demonstrated that a number of extracellular enzymes were binding to the HisTrap column. The most abundant proteins attached to the HisTrap column were identified by applying LC/MS-MS. Using PRED-TAT method [218] we identified the extracellular phospholipase C and the tripeptidyl aminopeptidase containing Sec signal sequences, while the glycerophosphoryl diester phosphodiesterase has a Tat signal peptide. These extracellular proteins, besides interfering with the purification process, forcing us to increase the stringency of the washing steps, and most probably reducing significantly the final yield of the phytase produced by *S. rimosus*. Naturally, these extracellular proteins are clearly "competing" for the Sec and Tat secretion systems. Therefore, in the scope of improvement of production of extracellular heterologous proteins in *S. rimosus*, its chassis should contain reduced number of genes encoding unnecessary secreted proteins, thus relieving the capacity Sec and Tat pathways for export of the target protein, consequently likely increasing secretion of target protein. On the other hand, reduction of the number of extracellular enzymes in the supernatant in the liquid culture would also make downstream processing steps much more efficient.

Currently the phospholipase C, the tripeptidyl aminopeptidase and the glycerophosphoryl diester phosphodiesterase isolated from *S. rimosus* supernatant are

lowering the purification efficiency of the recombinant phytase. However, these enzymes present an attractive source for signal sequences and promoters. The high expression of these three extracellular proteins (section 11.8.4), their identification through LC/MS-MS together with available *S. rimosus* genome sequence allowed us to identify their promoter region, and PRED-TAT predictor software enabled us to establish their signal sequences. Therefore, these gene tools could be adapted to existing plasmids, thus enhancing production of heterologous proteins in *S. rimosus*. Results presented in the Figure 34 indicated that these extracellular proteins are abundant in *S. rimosus* supernatant under tested cultivation conditions and they also display affinity for the HisTrap column. Consequently, their signal sequences are likely very efficiently exporting corresponding proteins, and most importantly, the promoters regulating these genes are likely strong enough to produce high amounts of the target. Hopefully, this kind of promoters, present as a single copy per chromosome, could be incorporated into plasmids containing phage recombinase as described in section 7.8.1, thus producing recombinant proteins in the absence of the selective marker thiostrepton. In addition, these promoters could also be tested in the multicopy vector pVF to produce recombinant proteins. However, as described in section 11.4, the strong constitutive promoter *PermE** together with pVF vector displayed significantly genetic instability, likely due to extensive burden of replicative machinery of the plasmid to the host cell.

S. rimosus chassis for production of heterologous proteins can also be optimized by improving the secretory machinery. In *S. rimosus* both secretion pathways (Sec and Tat) are active under the tested cultivation conditions. In the scope of this thesis, we used the export machinery of the Sec pathway to secrete phytase. However, although most of the phytase was exported out of *S. rimosus* cells, the Sec pathway could present a bottleneck for production of extracellular heterologous proteins in *S. rimosus* with significantly improved strains. Wen and collaborators [219, 220] demonstrated that in *B. subtilis*, the overexpression of the chaperone DnaK and the lipoprotein PrsA enhanced secretion capabilities of this host, thus improving production of extracellular proteins. In addition, as described in several patents and scientific reports, the overexpression and/or rational modification of chaperones and Sec proteins resulted in better secretory capacities in *B. subtilis* [221-226]. Therefore, although the genes required for the formation of the Sec pathway are expressed in *S. rimosus*, there might be room for upgrading the secretory machinery in this host by overexpressing the genes encoding each of the components of the Sec pathway, thus improving secretion capacity and subsequently increasing production of target extracellular heterologous proteins in *S. rimosus*.

It has been proposed that the Sec-dependent pathway is the major secretion system in bacteria with the exception of streptomycetes [227]. Unlike other bacterial systems, according to the literature data, the *Streptomyces* Tat system is the major secretion pathway in terms of the numbers of exported proteins [227]. Independently of the selection of the secretory system, and depending on the protein selected to be heterologously expressed extracellularly, Sec or Tat pathways have significant impact on the final yield of the target protein. Some proteins require specific conditions to be properly folded, therefore intracellular or extracellular environment play significant

role in the correct processing of different proteins [228], consequently selection of secretion system is not trivial. Mellado and collaborators [139] found that proteins exported through the Sec pathway cannot be secreted via Tat pathway, when the Sec signal peptide was exchanged for a Tat specific signal sequence. In contrast, proteins secreted by the Tat pathway can also be exported via Sec pathway, when interchanging the signal peptides. There are a number of unknown factors, which influence the predisposition of certain protein to be prone for the Sec or the Tat pathway. However, structure information of Tat and Sec dependent proteins suggest that the Tat route is more suitable for less structured proteins [139]. On the other hand, Gauthier et al. [71] used simultaneously the Sec and Tat secretion systems in *S. lividans* to secrete and produce xylanase B. Interestingly, the simultaneous activity of both pathways enhanced secretion capabilities in *S. lividans* [71]. Based on *S. rimosus* transcriptome, we have confirmed the importance of both pathways in the life cycle of this bacterium, and despite using the Sec pathway in this work, based on the literature data, we could also consider the Tat pathway in *S. rimosus* to secrete heterologous proteins. LC/MS-MS analysis and PRED-TAT predictor allowed us to identify the glycerophosphoryl diester phosphodiesterase Tat signal sequence, which was purified from *S. rimosus* supernatant. Therefore, this Tat-specific signal sequence could be adapted to expression plasmids in order to evaluate the Tat route of *S. rimosus*.

The *S. rimosus* chassis should be optimized for the production of heterologous proteins; however the gene tools evaluated in this work also need further optimization. The three promoters (*PermE**, *PnitA-NitR* and *tcp830*) tested to express the reporter system XylE and the phytase AppA were designed without terminator regions in the plasmid DNA. Only *PnitA-NitR* promoted included a terminator, however the transcription of the transcriptional regulator *nitR* was not controlled by any terminator. Terminators play an important role both, in completing the transcription process, and impacting mRNA half-life [229]. Incorporation of terminator sequences to expression constructs revealed that transcription levels were enhanced by 3-fold and protein output 4-fold [229]. Consequently, the addition of terminator sequences to current gene tools of *S. rimosus* could boost protein production, therefore its addition should be considered for future development of *S. rimosus* gene tools.

There is no doubt that *E. coli* is still the first choice in recombinant protein production at laboratory and industrial scale. However, for complex proteins with post-translational modifications, this host often tends to produce inclusion bodies. Therefore alternative expression systems are necessary to circumvent this kind of issue. We believe that we have demonstrated in this work that *S. rimosus* arises as an interesting alternative expression system, which meets number of requirements as potential industrial expression host system. The capacity to secrete the target protein represents likely one of the most important benefits of *S. rimosus* compared to *E. coli*, and this is probably the main reason why *S. rimosus* host system still can be considered as potentially industrially useful hosts.

13 Conclusion

Based on our results, we can conclude that *S. rimosus* is a potential alternative host for production of heterologous proteins. Similarly, as in the *E. coli* or *P. pastoris* expression systems, we have developed a functional genetic tool kit that accommodates to *S. rimosus* host background, which includes:

- Two model heterologous proteins; catechol 2,3-dioxygenase XylE from *Pseudomonas* sp. [84] and phytase AppA from *E. coli* [116], which can be used routinely for further development of *S. rimosus* protein expression system.
- The integrative vector pAB04, containing ϕC31 phage recombinase.
- The *E. coli-S. rimosus* shuttle replicative vector pVF, based on pUC19 (*E. coli*) and pPZ12 (*S. rimosus*) plasmids.
- Three functional, predominantly constitutive promoters (*PermE**, *PnitA-NitR* and *tcp830*).
- A signal sequence present in the trypsin like proteinase enzyme from *S. rimosus* which efficiently exports target heterologous proteins.
- Microbiological procedures to overexpress and purify target heterologous proteins, intracellularly or extracellularly, using *S. rimosus* host.

We have adapted periplasmic phytase AppA from *E. coli* as model system in order to access the potential of *S. rimosus* host. AppA, with four nonconsecutive disulfide bonds, is an enzyme relatively demanding to produce in the correctly folded form. Even in *E. coli*, the preferred heterologous host for the production of AppA, it generates undesired insoluble aggregates of phytase, while *P. pastoris*, another industrially valuable expression host, hyper-glycosylates AppA. We have demonstrated that *S. rimosus* efficiently secretes the recombinant phytase to the extracellular medium, where it was fully active and most likely correctly folded. The AppA yields obtained with *S. rimosus* were comparable to the yields obtained with *E. coli*, 20 U/mL and 33 U/mL [114], respectively.

Probably the best feature of *S. rimosus* as heterologous host is its capacity to export the target protein. Genome and transcriptome analysis suggested the potent secretory apparatus of *S. rimosus*. We used phytase AppA to evaluate the Sec-dependent pathway of *S. rimosus*, and we have demonstrated that a protein fused to a Sec signal peptide is efficiently exported to the extracellular medium. Potentially, any protein could be secreted by *S. rimosus* facilitating downstream processing, thus avoiding the formation of inclusion bodies (a common feature of *E. coli*), and importantly, avoiding disruption of *S. rimosus* cells, which is a great advantage when considering downstream processing. Thus, considering the efficient secretion of the target protein, we have demonstrated that the biosynthetic process for production of extracellular recombinant proteins by *S. rimosus* can be carried out at relatively low volume. However, we do expect that the procedure can be transferred to large industrial scale fermenters, considering all the main process manipulations for *S. rimosus* at the industrial fermenters already exist.

S. rimosus has been used over 70 years for the industrial production of oxytetracycline. The genome sequence and transcriptome analysis of *S. rimosus* revealed the potential of this industrial host to produce heterologous proteins. Additionally, a number of enzymes with biotechnological applications were identified. The genetic tool kit developed in this work presents very solid base for further development of *S. rimosus* platform for expression of heterologous proteins. Therefore, we believe that we have demonstrated that *S. rimosus* is a useful alternative of the current expression systems for production of heterologous proteins.

14 Summary in Spanish

15 Capítulo 1: Selectividad del dominio aciltransferasa fkAT4 involucrado en la biosíntesis de tacrolimus

15.1 Introducción

Los policétidos tipo I son una clase de metabolitos secundarios sintetizados por un grupo de enzimas denominadas policétido sintasas (PKS tipo I). Estos productos naturales muestran una impresionante diversidad estructural lo que se refleja en un gran espectro de actividades farmacológicas [5]. Las policétido sintasas (PKS) se organizan en módulos de modo que cada módulo cataliza un paso de la extensión de la cadena del policétido [5]. Cada módulo consiste en al menos tres dominios enzimáticos: betacetoacil sintasa (KS), aciltransferasa (AT) y proteína portadora de acilo (ACP), donde el dominio AT determina la elección de la unidad extensora (o iniciadora) para cada paso de extensión [5]. Además cada módulo puede contener una combinación de dominios adicionales como cetoreductasa (KR), enoilreductasa (ER) y deshidratasa (DH), que catalizan la reducción de la unidad beta-ceto [5]. Las unidades extensoras activadas usadas en la biosíntesis de policétidos están limitadas casi exclusivamente a malonil-, metilmalonil- y etilmalonil-CoA [230], como consecuencia la diversidad estructural de los compuestos policétidos está limitada a estas tres unidades extensoras. Así todo, existen algunos dominios AT capaces de seleccionar diferentes unidades extensoras, como por ejemplo el dominio fkAT4 implicado en la biosíntesis de tacrolimus (FK506), fkAT4 es capaz de seleccionar etilmalonil-, alilmalonil- o propilmalonil-CoA, pero no selecciona las unidades extensoras más comunes, como lo son malonil- o metilmalonil-CoA [10]. La especificidad de los dominios AT ha sido estudiada en detalle habiéndose identificado ciertos motivos conservados que intervienen en la especificidad de sustrato. Sin embargo, estos estudios se han limitado, de momento, al estudio de las unidades extensoras más comunes [9] siendo necesario mejorar el entendimiento de la especificidad del dominio AT a estas otras unidades extensoras menos usuales. Una mejor comprensión de la especificidad de sustrato de los dominios AT involucrados en la biosíntesis de policétidos nos indicará cómo alterar esa especificidad. De este modo se podrá introducir una mayor variabilidad estructural a partir de diferentes unidades extensoras que permitan posteriores esfuerzos semisintéticos, los cuales facilitaran el desarrollo de nuevos fármacos. Además de las diversas unidades extensoras naturales, el dominio fkAT4 puede aceptar diversas unidades extensoras no naturales usando aproximaciones combinadas de ingeniería (químico)biosintética y metabólica [11]. Es por esto que queremos estudiar el dominio fkAT4 dado su tendencia por seleccionar específicamente unidades extensoras atípicas haciendo de esta aciltransferasa el modelo perfecto para estudiar la selectividad de los dominios AT.

15.2 Hipótesis

La producción, purificación y estudio del dominio fkAT4 mejorara el entendimiento de la selectividad de los dominios AT involucrados en la biosíntesis de policétidos. La interacción proteína-substrato del dominio fkAT4 purificado con los diferentes substratos permitirá entender el mecanismo por el cual este dominio selecciona determinadas unidades extensoras, lo cual facilitara modificar diferentes módulos AT incrementando la diversidad estructural de los policétidos tipo I.

15.3 Materiales y métodos

15.3.1 Plásmidos y microorganismos

El grupo de genes que codifican las policétido sintasas involucradas en la biosíntesis de tacrolimus se encuentran en el microorganismo *Streptomyces tsukubaensis*. Para su expresión se diseñó un fragmento de ADN sintético codificando la aciltransferasa fkAT4 con el uso de codones optimizado para *E. coli* dado que se quería purificar el dominio fkAT4 usando este huésped (tabla. 5).

Cinco diferentes variables del dominio fkAT4 se clonaron en el vector de expresión pET29 el cual se induce con IPTG en *E. coli* e incorpora una cola de histidinas en el C-terminal a la proteína recombinante para facilitar su purificación. Las cinco versiones del módulo fkAT4 incluyen: (i) los dominios fkKS/AT4 (ii) el dominio fkAT4 (iii) versión truncada del dominio fkAT4 (iv) el dominio fkAT con el enlace al dominio fkKS4 y (v) el dominio fkAT4 truncado con el enlace al dominio fkKS4.

Los cinco plásmidos generados a partir del vector pET29 y las versiones del dominio fkAT4 se trasformaron en las cepas de *E. coli* BL 21, tuner, C41 y codón plus.

Alternativamente se clonaron dos variantes del dominio fkAT4 (los dominios fkKS/AT4 y el dominio fkAT4) en el vector de expresión pBAD el cual se induce con arabinosa en *E. coli*. Estos dos plásmidos de transformaron a *E. coli* BL21.

Adicionalmente se clonaron tres variantes del dominio fkAT4 (los dominios fkKS/AT4, el dominio fkAT4 y la versión truncada del dominio fkAT4) en el vector de expresión pPICZαA. Los tres nuevos vectores se transformaron a la levadura *Pichia pastoris*.

15.3.2 Expresión del dominio fkAT4 en *E. coli*

Las cepas de *E. coli* portadoras de los diferentes plásmidos de expresión (pET29c y pBAD) con los respectivos dominios fkAT4 se cultivaron en medio de cultivo LB a 37^{o}C y 220 rpm. Una vez los cultivos alcanzaron OD_{600} entre 0.6 y 0.8 se adiciono el inductor, IPTG o arabinosa según el plásmido de expresión, y se continuo con el crecimiento celular por 3, 6 o 24 horas según la temperatura de crecimiento (37^{o}C, 30^{o}C o 20^{o}C) (Grafico. 6). Al finalizar la inducción las células fueron recolectadas y se analizó el

contenido celular mediante electroforesis en geles de poliacrilamida/SDS en busca de las diferentes versiones del dominio fkAT4 recombinantes.

15.3.3 Expresión del dominio fkAT4 en *P. pastoris*

Las cepas de *P. pastoris* portadoras de los plásmidos de expresión con las variantes del dominio fkAT4 fueron inoculadas en medio BMGY y crecidas a 30^{o}C hasta alcanzar OD_{600} 2-6, las células se recolectaron y se resuspendieron en medio BMMY hasta alcanzar OD_{600} 1. Los cultivos fueron inducidos con metanol cada 24 horas durante 3 días. Cada 12 horas se recolectaron muestras que se analizaron mediante electroforesis en geles de poliacrilamida/SDS.

15.3.4 Repliegue de los cuerpos de inclusión contenido los dominios fkAT4 formados por *E. coli*

Para replegar los cuerpos de inclusión contenido los dominios fkAT4, las células fueron recolectadas, lisadas y posteriormente el contenido celular, el cual incluye los cuerpos de inclusión, fue tratado con una solución 6M de hidrocloruro de guanidinio (agente desnaturalizante). Los cuerpos de inclusión desnaturalizados fueron cargados en una columna de níquel Histrap (GE) para capturar los dominios fkAT4 recombinantes. En dicha columna se eliminó gradualmente el hidrocloruro de guanidinio para replegar los dominios fkAT4 recombinantes y se eluyeron usando una solución con imidazol, el cual fue posteriormente removido mediante diálisis.

15.3.5 Ensayo químico para evaluar la actividad de los dominios fkAT4 replegados

Para establecer la actividad de las variables del dominio fkAT4 replegadas se empelo la reacción de Ellman. Al mezclar los dominios fkAT4 con la unidad extensora alilmalonil-SNAC el dominio aciltransferasa captura únicamente alilmalonil liberando SNAC (compuesto análogo de Coenzima A). El grupo tiol presente en el SNAC liberado reaccionara con el reactivo de Ellman produciendo un color amarillo detectable a una absorbancia A412.

15.4 Resultados

De las diferentes variables del dominio fkAT4 clonadas en el plásmido pET29c y expresadas en las cepas de *E. coli* BL21, tuner, codón plus y C41, solo la cepa tuner fue capaz de sobre-expresar las variantes recombinantes del dominio fkAT4. Sin embargo, estas variables recombinantes del dominio fkAT4 solo fueron producidas como cuerpos de inclusión. Tampoco el plásmido pBAD controlando la expresión de los dominios fkKS/AT4 o el dominio fkAT4 en la cepa de *E. coli* BL21, fue capaz de producir la proteína recombinante de interés. Las variables del dominio fkAT4 diseñadas para ser expresadas en *P. pastoris* no fueron producidas por esta levadura.

Dado que solo fue posible la obtención de agregados insolubles del dominio fkAT4, estos cuerpos de inclusión se desnaturalizaron y replegaron en una columna de níquel usando un gradiente de hidrocloruro de guanidinio. La desnaturalización y posterior renaturalización de los cuerpos de inclusión de las variables del dominio fkAT4 permitió obtener dichos dominios en una forma soluble (Figura 35).

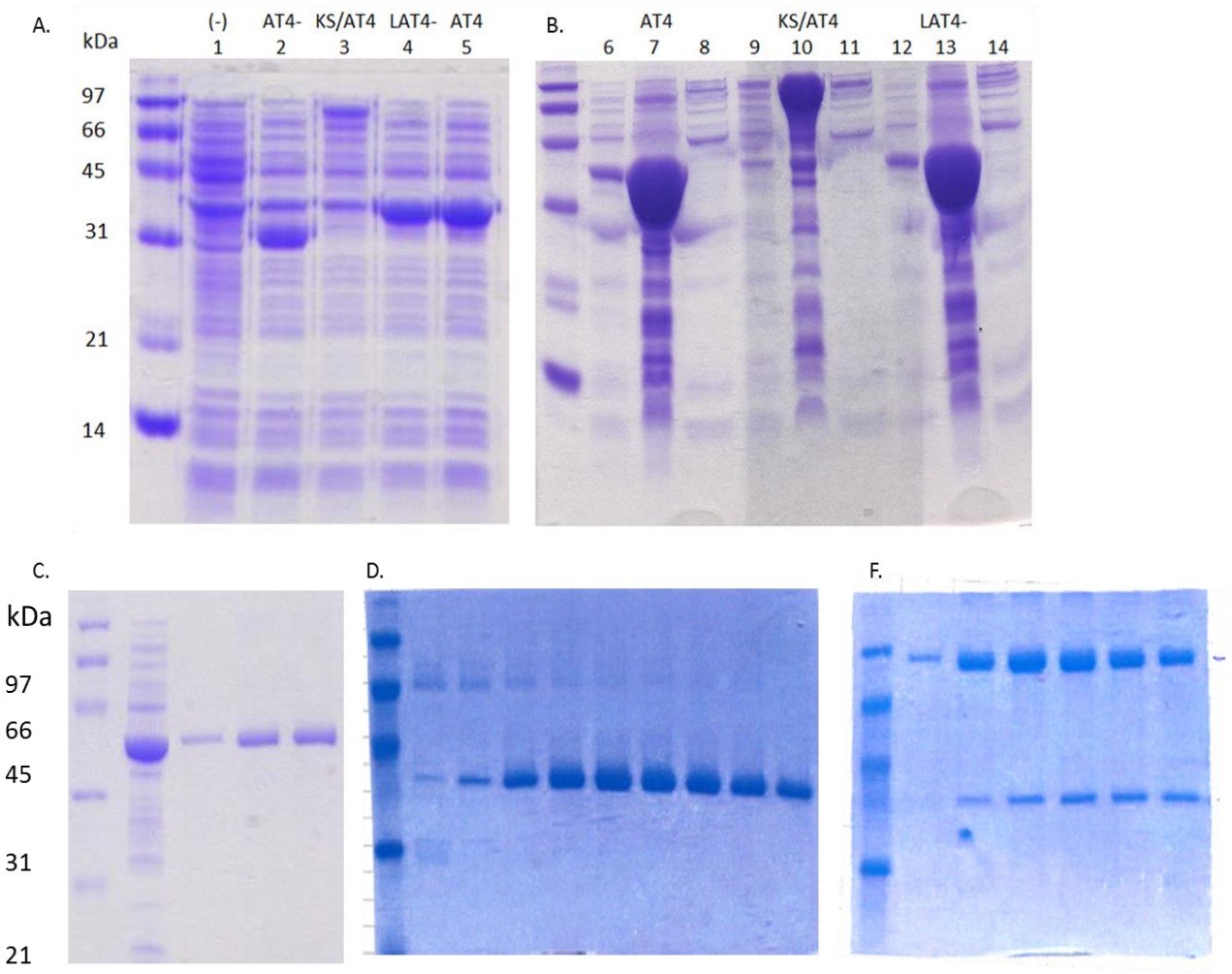


Figura 35. Purificación de fkAT4. **A.** extracto celular de la diferentes variables del dominio fkAT4 sobre-expresadas en la cepa de *E. coli* tuner. (1: Cepa salvaje de *E. coli* tuner; 2: AT4-: versión truncada del dominio fkAT4; 3: KS/AT4: los dominios fkKS/AT4; 4:LAT4-: dominio fkAT4 truncado con el enlace al dominio fkKS4; 5: AT4: dominio fkAT4). **B.** Análisis del extracto crudo, pellet y proteínas solubles después de lisar las cepas tuner recombinantes (6,7 y 8: dominio fkAT4; 9, 10 y 11: dominios fkKS/AT4; 12, 13 y 14: dominio fkAT4 truncado con el enlace al dominio fkKS4). **C. D. y F.** dominios fkAT4, dominio fkAT4 truncado con el enlace al dominio fkKS4 y los dominios fkKS/AT4 respectivamente luego de ser replegados.

Una vez se replegaron las diferentes variables del dominio fkAT4 se observó que estas proteínas recombinantes re-naturalizadas tendían a agregarse y precipitar. Aun así procedimos a efectuar los ensayos *in-vitro* para verificar la actividad aciltransferasa de los dominios fkAT4. Los ensayos realizados con los dominios fkAT4 re-solubilizados, la unidad extensora alilmalonil-SNAC y el reactivo de Ellman indicaron que los dominios fkAT4 no se replegaron correctamente dado que no se evidencio cambió de color en la

reacción enzimática. Así, el proyecto se detuvo hasta encontrar un huésped capaz de expresar el dominio fkAT4 plegado correctamente.

15.5 Discusión

Se ha evidenciado que los huéspedes *E. coli* y *P. pastoris* no son capaces de producir proteínas provenientes de streptomycetes. Posibles razones para tal resultado incluyen: la diferencia en el uso de codones o porque estos microrganismos no tienen la maquinaria (chaperonas) para poder plegar correctamente las proteínas de *Streptomyces* spp. [24]. Se hace entonces necesario el desarrollo de un huésped, idealmente streptomycete, que pueda sobre-expresar dichas proteínas y de este modo facilitar la obtención de proteínas recombinantes que *E. coli* o *P. pastoris* no pueden producir.

16 Capítulo 2: Desarrollo del kit genético para producir proteínas recombinantes en *Streptomyces rimosus*

16.1 Introducción

Streptomyces spp. son un grupo de bacterias gram-positivas con un contenido elevado de G+C en sus genomas [135]. Estos microorganismos se encuentran en diversos ambientes, pero se caracterizan por vivir en el suelo [58]. Dado que en este ecosistema habitan una gran diversidad de organismos, los streptomycetes han desarrollado una serie de metabolitos secundarios y enzimas extracelulares para poder colonizar este ambiente hostil [135]. Dentro de los metabolitos secundarios producidos por streptomycetes se encuentran gran diversidad de compuestos antibacterianos (tetraciclina), antifúngicos (nyastin), anticancerígenos (rapamicina), antihelmínticos (avermectin) e inmunosupresores (tacrolimus), por nombrar algunos [5]. *S. rimosus* es conocido ampliamente por la industria farmacéutica ya que es usado para la producción de oxitetracilina, antibiótico de amplio espectro [48]. Tanto la genética de *S. rimosus* como los medios de cultivo para la producción de oxitetracilina se ha estudiado y desarrollado desde 1950, cuando se descubrió la capacidad de este streptomycete para producir este importante antibiótico [48].

S. rimosus es reconocido como un agente GRAS (Generalmente Reconocido como Seguro) por la FDA y la EMA (Agencia de Alimentos y Medicamentos americana y la Agencia Europea de Medicamentos, respectivamente). Además, este microrganismo se caracteriza por ser un streptomycete de rápido crecimiento, alcanzando la fase exponencial en menos de 24 horas. Aparte de ser conocido por producir oxitetracilina, también se han purificado 4 enzimas extracelulares (una endodeoxyribonucleasa [72], una alfa-amilasa [73], una lipasa [74] y una metaloendopeptidasa [77], lo que indican la gran capacidad secretora de esta bacteria. Todas estas características hacen de *S. rimosus* un potencial huésped para la expresión de proteínas heterólogas que *E. coli* no puede producir correctamente. Actualmente existen pocas herramientas genéticas para usar *S. rimosus* como huésped para la expresión de proteínas recombinantes. Las herramientas existentes incluyen un plásmido integrativo basado en la recombinasa ϕC31, un promotor constitutivo (*PermE**) y un péptido señal para exportar las proteínas recombinantes (*srT*) [78].

La fitasa es una enzima usada como aditivo alimentario, especialmente por la industria porcina y aviar, mejorando la ingesta de fosforo a partir de ácido fitico [114]. La producción de fitasa codificada naturalmente en *E. coli* es particularmente difícil dado que esta enzima presenta cuatro enlaces disulfuro en la secuencia proteica [117]. Se han empleado diferentes microorganismos para producir la fitasa de *E. coli*, entre ellos *Pichia pastoris*, *Saccharomyces cerevisiae, Streptomyces lividans* y el mismo *E. coli*, sin embargo los rendimientos no son los mejores [114, 120]. La sobre-expresión de fitasa en su huésped natural, *E. coli*, genera la producción cuerpos de inclusión dado la presencia de los enlaces disulfuro [116], en *P. pastoris* y *S. cerevisiae* la fitasa recombinante es super-glicosilada [208], mientras que en *S. lividans* la producción de fitasa es apenas detectable [120].

16.2 Objetivo de este estudio

La incorporación de herramientas genéticas al kit de *S. rimosus* mejoraran las capacidades de este huésped para producir proteínas recombinantes. Recientemente se han desarrollado dos promotores inducibles los cuales presentaron actividad en gran variedad de streptomycetes. Dichos promotores, denominados *PnitA-NitR* y *tcp830*, probaron ser lo suficientemente eficientes para expresar proteínas recombinantes en diferentes streptomycetes [97, 98]. Además se hace indispensable la incorporación de un plásmido replicativo al kit genético de *S. rimosus*. Por tanto, este estudio tiene como objetivo diseñar un plásmido replicativo para *S. rimosus* y evaluar diferentes promotores los cuales permitan a *S. rimosus* producir proteínas heterólogas. Dicho análisis se llevará a cabo en *S. rimosus* con la producción del gen reportero que codifica para catecol 2,3-dioxigenasa (XylE) proveniente de *Pseudomonas* sp. y la fitasa (AppA) de *E. coli*. Adicionalmente se estudiará el transcriptoma de *S. rimosus* para identificar los diferentes genes involucrados tanto en la producción de proteínas (sistema de secreción Sec y Tat) como en posibles genes involucrados en la producción enzimas con potencial biotecnológico.

16.3 Hipótesis

La bacteria *S. rimosus* es un huésped conveniente para el desarrollo de un nuevo sistema de expresión para la producción de proteínas heterólogas. Predecimos que la incorporación de un vector replicativo y mejores promotores al kit de herramientas genéticas de *S. rimosus* permitirá la producción de proteínas recombinantes en cantidades similares a las obtenidas por los actuales sistemas de expresión, como *E. coli* o *P. pastoris*.

16.4 Materiales y métodos

16.4.1 Rescate del plásmido pYAC-OTC del cromosoma de *S. rimosus* para determinar los sitios *attB* específicos para la recombinasa ϕC31

El ADN cromosómico de transformantes de *S. rimosus* portadores del plásmido pYAC-OTC fue aislado, digerido con *SacI*, ligado y transformado en *E. coli* con el fin de rescatar el plásmido pYAC junto con la región adyacente al sitio de integración *attB* específico para la recombinasa ϕC31. Se secuenciaron los plásmidos rescatados y a partir de la secuencia obtenida se generaron sondas de hibridación para localizar los sitios de integración *attB* en el cromosoma de *S. rimosus*. El ADN cromosómico de *S. rimosus* se digirió nuevamente con *AseI* y se separó mediante electroforesis de campo pulsado. El ADN separado se transfirió a una membrana de nylon y se hibridó con las sondas generadas previamente.

16.4.2 Síntesis de los promotores inducibles *PnitA-NitR* y *tcp830*

Los promotores *PnitA-NitR* y *tcp830* se sintetizaron de acuerdo a las publicaciones originales describiendo dichos promotores [97, 98]. Se incorporó un sitio de clonaje para facilitar la manipulación y posterior clonaje de los promotores a plásmidos de *S. rimosus*.

16.4.3 Generación del plásmido replicativo para *E. coli* y *S. rimosus*

El plásmido pPZ12 [111] se digirió con *PstI*. A su vez el plásmido pUC19 se digirió con la misma enzima y ambos plásmidos fueron ligados y transformados a *E. coli* generando el plásmido bifuncional pVF. El plásmido pVF se transformó a *S. rimosus*, se aisló y se secuencio.

16.4.4 Evaluación de los promotores inducibles *PnitA-NitR* y *tcp830 en S. rimosus*

El gen reportero *xylE* se acomodó para ser regulado por los promotores *PermE**, *PnitA-NitR* y *tcp830*. Las construcciones *PermE*/xylE*, *PnitA-NitR/xylE* y *tcp830/xylE* se clonaron en los plásmidos pAB04 y pVF los cuales se transformaron a *S. rimosus*. Posteriormente se clonó el gen que codifica para la fitasa (*appA*) de *E. coli* para ser regulada con los promotores *PnitA-NitR* y *tcp830*. El gen fitasa fue amplificado del plásmido pAB04-4 proporcionado por la Dra. Magdevska (Universidad de Liubliana, Eslovenia). En pAB04-4 la fitasa se encuentra fusionada al péptido señal *srT* [78]. El gen *appA* fue amplificado junto con el péptido señal y se le incorporo una cola de histidinas en el C-terminal para facilitar la purificación de la fitasa recombinante. Las construcciones *PnitA-NitR/srT/appA* y *tcp830/srT/appA* fueron clonadas en el plásmido replicativo pVF y transformadas a *S. rimosus*.

16.4.5 Ensayos enzimáticos para establecer la actividad del gen reportero *xylE* y fitasa

La actividad enzimática de la catecol-2,3-dioxigenasa (XylE) y fitasa (AppA) fueron determinadas según métodos descritos anteriormente [49, 125].

16.4.6 Producción y purificación de fitasa recombinante producida por *S. rimosus*

La fitasa producida por la cepa recombinante de *S. rimosus* portadora del plásmido pVF-*tcp830/srT/appA* fue purificada del medio extracelular mediante cromatografía de afinidad usando columnas de níquel. La proteína purificada se analizó media cromatografía liquida acoplada a espectrometría de masas (LC-MS/MS).

16.4.7 Transcriptoma de *S rimosus*

El transcriptoma de *S. rimosus* fue realizado por la compañía Acies Bio Ltd. (Eslovenia). El ARN fue obtenido de los medios de cultivo TSB y medio de cultivo para la producción de oxitetraciclina. Se tomaron muestras de ARN a las 9, 13, 16, 20, 24 y 36 horas en TSB mientras que en el medio para la producción de oxitetraciclina las muestras se tomaron a las 12, 20, 36, 48, 72 y 96 horas.

16.5 Resultados

Dado la importancia de entender el número de veces y localización de los sitios de integración específicos de la recombinasa ϕC31 en el cromosoma de *S. rimosus*, se emplearon cepas recombinantes de *S. rimosus* portadoras del plásmido pYAC-OTC, el incluye la recombinasa ϕC31, y a partir del ADN genómico de estas cepas se rescató el plásmido pYAC para identificar los sitios de integración (*attB*) presentes en el cromosoma de *S. rimosus*. Los plásmidos rescatados permitieron establecer que en el cromosoma de *S. rimosus* hay dos sitios de integración (*attB*) específicos para la recombinasa ϕC31. La hibridación southern usando las sondas generadas de los plásmidos rescatados demostró que ambos sitios de integración se localizan en el fragmento de 2.3 Mb cuando el cromosoma de *S. rimosus* es digerido con la enzima de restricción *AseI*, enzima con la cual se estableció el mapa genético de *S. rimosus* [53]. Además se estableció que los plásmidos que contienen la recombinasa ϕC31 una vez son transformados en *S. rimosus* se pueden integrar en el sitio de integración perfecto (*attB*) y además en un sitio secundario de integración (o falso *attB*).

Cuando se emplearon los promotores *PnitA-NitR* y *tcp830* regulando el gen reportero *xylE* y se integraron al cromosoma de *S. rimosus* usando la recombinasa ϕC31 se detectó muy leve actividad XylE, inferior a la obtenida cuando el gen reportero es expresado con el promotor constitutivo *PermE**. Este resultado indica que ambos promotores no son lo suficientemente fuertes para expresar proteínas, al menos cuando estos son integrados en el cromosoma de *S. rimosus*.

El plásmido replicativo pVF, generado a partir de los plásmidos pPZ12 y pUC19, demostró ser genéticamente estable en *S. rimosus*. La secuenciación del plásmido pVF permitió identificar la maquinaria del plásmido pUC19 (origen de replicación pMB1 y marcador de selección para ampicilina), y además el origen de replicación del plásmido pIJ101, el marcador de selección para thiostrepton y 2 marcos abiertos de lectura que codifican los genes *korB* y orf56 (Figura 36). Estudios de segregación del plásmido pVF en *S. rimosus* permitieron establecer que alrededor de 30% del plásmido se pierde cada 24 horas cuando la cepa *S. rimosus* pVF se cultiva en la ausencia de thiostrepton (Figura 36)

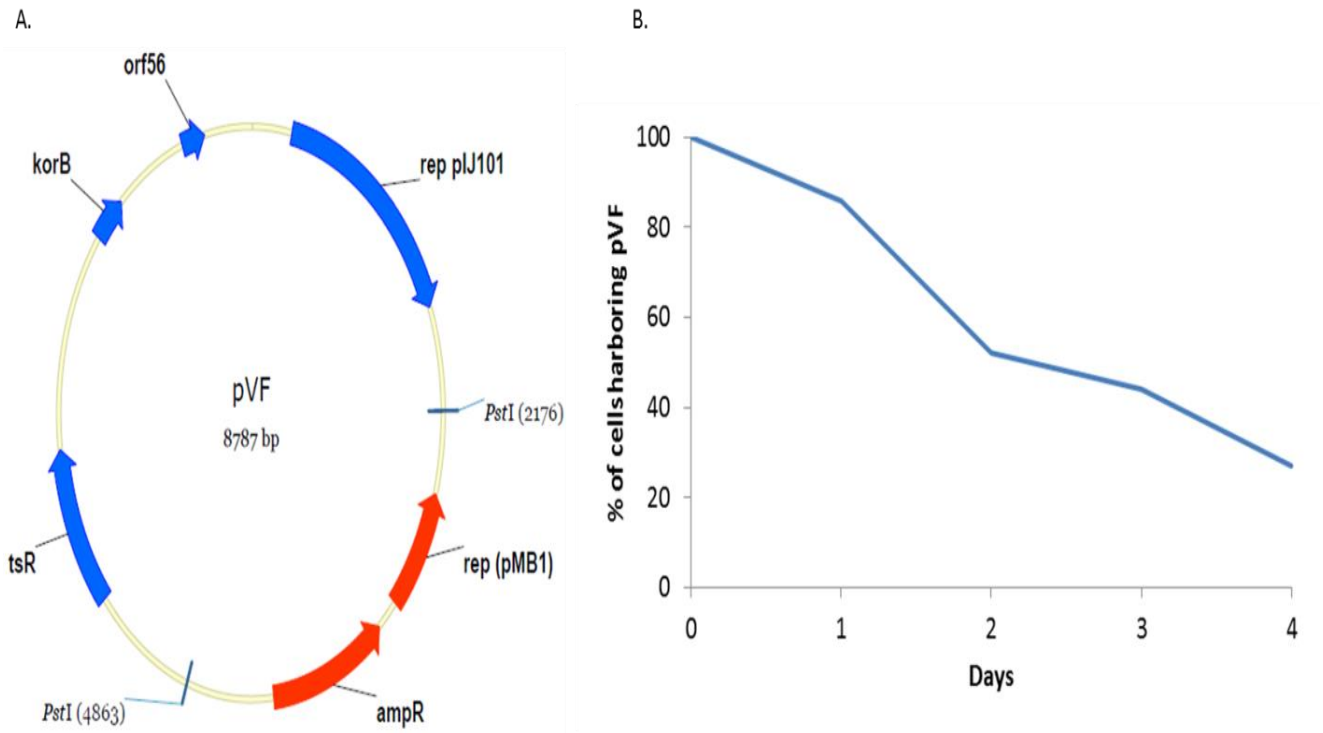


Figura 36. A. mapa del plásmido pVF. **B.** Estabilidad segregacional del plásmido pVF.

Inicialmente se probó el plásmido pVF con el promotor constitutivo *PermE** para producir XylE. Las cepas de *S. rimosus* portadoras del plásmido pVF-*PermE*/xylE* exhibieron un fenotipo distinto al de la cepa salvaje. El patrón de restricción mostrado por el plásmido pVF-*PermE*/xylE* tras ser aislado de las cepas "enfermas" de *S. rimosus* mostro que el vector sufrió deleciones *in-vivo*. Así, el sistema de expresión compuesto del plásmido pVF y el promotor *PermE** no es estable en *S. rimosus*. Cuando se evaluaron los plásmidos pVF-*PnitA-NitR/xylE* y pVF-*tcp830/xylE* en *S. rimosus* se evidenció que las cepas recombinantes portadoras de dichos plásmidos no exhibieron diferencias morfológicas con respecto a la cepa salvaje. Los plásmidos pVF-*PnitA-NitR/xylE* y pVF-*tcp830/xylE* se aislaron de *S. rimosus* confirmando que ambos plásmidos conservaron su integridad estructural. Al cultivar las cepas recombinantes de *S. rimosus* portadoras de los plásmidos pVF-*PnitA-NitR/xylE* y pVF-*tcp830/xylE* en medio TSB se detectó actividad XylE, aunque en ambos casos la actividad fue detectada sin la adición del inductor (ϵ-caprolactamo y tetraciclina para *PnitA-NitR* y *tcp830* respectivamente) indicando la actividad constitutiva de ambos promotores en *S. rimosus* (Figura 37).

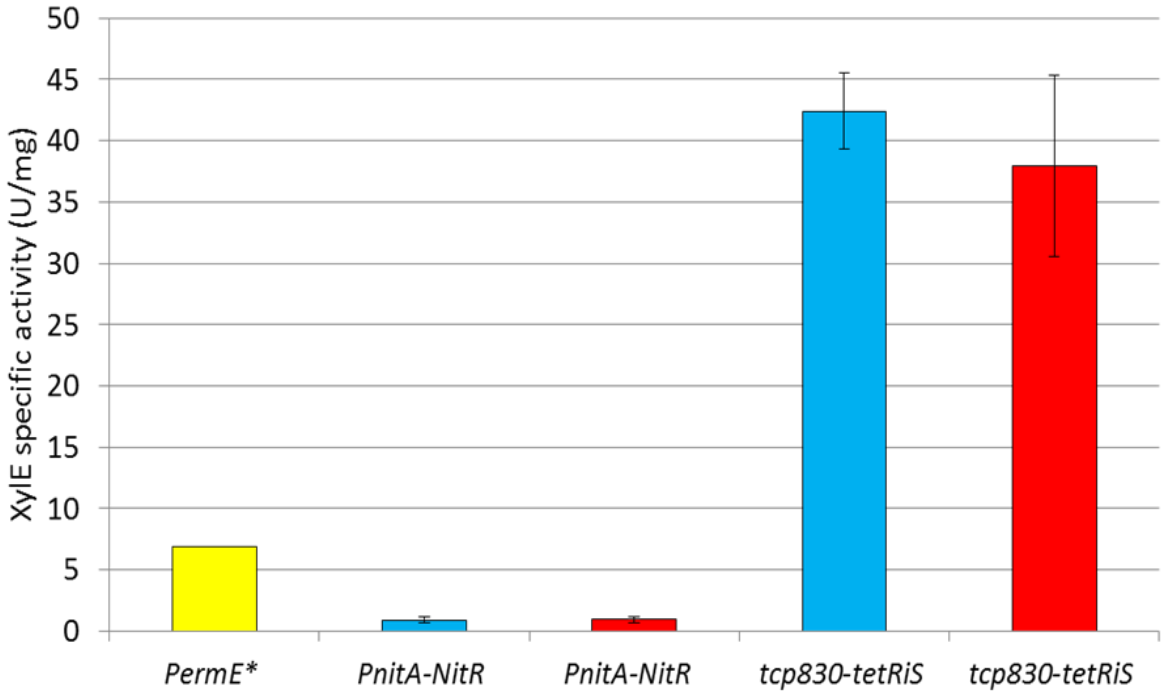


Figura 37. Actividad XylE detectada en diferentes cepas recombinantes de *S. rimosus* portadoras del plásmido pVF y los promotores *PnitA-NitR*, *tcp830* y *Perm**. Las barras azules indican los cultivos no inducidos mientras que las barras rojas indican los cultivos inducidos con ϵ-caprolactamo y tetraciclina para *PnitA-NitR* y *tcp830* respectivamente. La barra amarilla indica la cepa de *S. rimosus* portadora del plásmido pVF/*PermE*-xylE*. Los resultados se expresan como el promedio ± la desviación estándar calculada a partir de 3 colonias diferentes. Para el promotor *PermE** no se pudo calcular promedio dado que de más de 30 colonias analizadas solo 1 mostro actividad XylE.

Una vez se confirmó la estabilidad genética del plásmido pVF con los promotores *PnitA-NitR* y *tcp830* se procedió a evaluar la producción de fitasa en *S. rimosus* con los nuevos sistemas de expresión desarrollados. Las cepas recombinantes de *S. rimosus* portadoras de los plásmidos pVF-*PnitA-NitR/srT/appA* y pVF-*tcp830/srT/appA* fueron cultivadas en el medio de cultivo empleado en la producción de oxitetraciclina y se recogieron muestras cada día durante 7 días para determinar la actividad fitasa tanto intra como extracelularmente (Figura 38).

Los resultados obtenidos indican que el mejor sistema de expresión de *S. rimosus* está compuesto por el vector pVF y el promotor *tcp830*. Además se evidenció que la fitasa es secretada al medio extracelular y que la máxima producción se alcanza después de 4 días de cultivo cuando este incluye el marcador de selección thiostrepton. Las cepas recombinantes de *S. rimosus* portadoras del plásmido pVF cultivadas en la ausencia del marcador de selección mostraron rendimientos más bajos de actividad fitasa, lo cual confirmo la estabilidad segregacional del plásmidos pVF descrita anteriormente. Por su parte, las cepas de *S. rimosus* portadoras del plásmido pAB04-*PermE*/srT/appA* mostraron la misma actividad fitasa cuando estas son cultivadas con o sin thiostrepton, lo que indica que una vez el plásmido se ha integrado en el cromosoma de *S. rimosus*

mediante la recombinasa ϕC31, el mantenimiento del ADN integrado en el cromosoma de *S. rimosus* no requiere la presencia del marcador de selección.

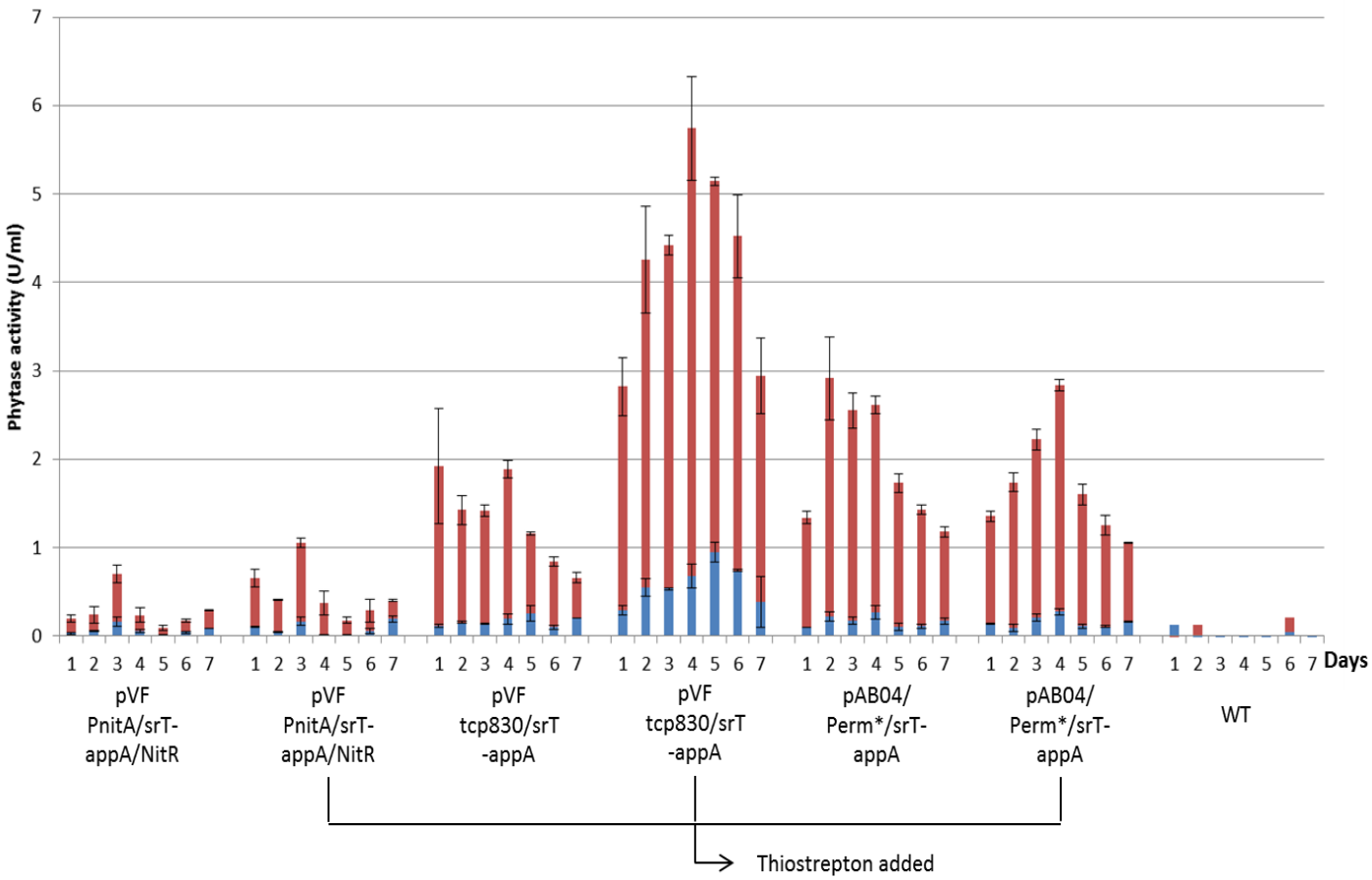


Figura 38. Actividad fitasa en diferentes cepas recombinantes de *S. rimosus*. Las barras rojas indican la actividad detectada extracelularmente mientras que las barras azules indican la actividad intracelular. Se indica los cultivos a los cuales se les adiciono thiostrepton. Cada construcción se analizó con 7 transformantes diferentes y los resultados se expresan como el promedio ± la desviación estándar.

Con las condiciones optimizadas para la producción de fitasa por parte de *S. rimosus* en 5 mL de cultivo, se decidió a aumentar la producción de esta enzima. En este caso la producción se escaló a 200 mL de medio de cultivo y se inoculo *S. rimosus* pVF-*tcp830/srT/appA*. Al cuarto día se recogieron y removieron las células y se cargó el sobrenadante en una columna de níquel para aislar la fitasa usando la cola de histidinas incorporada en el C-terminal de la fitasa recombinante. Se logró purificar la fitasa recombinante usando cromatografía de afinidad con columna de níquel, y se confirmó la identidad de la fitasa purificada después de correr está en un gel de poliacrilamida/SDS (Figura 27). Posteriormente se digirió la banda con tripsina y se confirmó su identidad mediante LC-MS/MS (material suplementario).

Además de comprobar la capacidad de *S. rimosus* para producir proteínas heterólogas, se estudió el transcriptoma de esta bacteria en 2 medios de cultivos diferentes, en TSB y en el medio de cultivo empleado para la producción de oxitetraciclina. La información suministrada por el transcriptoma permitió determinar el estado transcripcional de los genes putativos involucrados en los sistemas de secreción de

proteínas, las posibles proteasas que pueden degradar las proteínas recombinantes, además de una serie de genes putativos que codifican enzimas de potencial valor industrial. Los datos obtenidos del transcriptoma de *S. rimosus* indican que los sistemas de secreción de proteínas Sec y Tat están activos en ambos medios de cultivo. Sin embargo, después de 20 horas de crecimiento en TSB, el sistema Sec empieza a desactivarse mientras que el sistema Tat empieza a perder actividad a partir de las 24 horas de cultivo. En medio de producción ambos sistemas de secreción de proteínas están activos las primeras 96 horas de crecimiento. El transcriptoma también permitió identificar 36 proteasas putativas de las cuales 4 están activas en TSB y 7 están activas en medio de cultivo de producción. También se identificaron 52 enzimas putativas con potencial uso biotecnológico de las cuales solo dos están activas. Entre estas enzimas se encuentran diversas lipasas, amilasas, nitrilasas, quitinasas, colagenasas, celulasas y xilanasas, entre otras (tabla. 13).

16.6 Discusión

El presente estudio permitió incorporar nuevas herramientas al kit genético de *S. rimosus*. Además, se analizaron herramientas desarrolladas previamente que no habían sido analizadas en *S. rimosus*. Es conocido que los plásmidos integrativos basados en la recombinasa ϕC31 pueden integrarse en el cromosoma de streptomycetes en múltiples ocasiones [82]. Sin embargo, en *S. rimosus* solo se había caracterizado el sitio perfecto de integración (*attB*) [128]. El estudio de transformantes de *S. rimosus* portando el plásmido pYAC-OTC, el cual se integra en el cromosoma mediante la recombinasa ϕC31, permitió establecer que *S. rimosus* tiene dos sitios de integración para esta recombinasa, *attB* y falso *attB*. El plásmido pYAC rescatado del cromosoma de *S. rimosus* permitió establecer las secuencias genéticas de los sitios *attB* y estudios de southern indicaron que ambos sitios de integración se encuentran en la banda de 2.3 Mb cuando el cromosoma de *S. rimosus* es mapeado con la enzima de restricción *AseI*. El uso de plásmidos integrativos mediante la recombinasa ϕC31 en el cromosoma de *S. rimosus* deja claro que estos plásmidos se pueden integrar en el cromosoma de esta bacteria en dos loci diferentes.

Los promotores inducibles *PnitA-NitR* y *tcp830* mostraron ser débiles para expresar XylE cuando estos fueron integrados en el cromosoma de *S. rimosus* usando la recombinasa ϕC31. Dado este resultado, se hizo necesario el desarrollo de un plásmido replicativo para *S. rimosus*. Estudios anteriores demostraron que el promotor *PnitA-NitR* solo era funcional en *S. lividans*, *S. avermitilis*, *S. griseus* y *S. coelicolor* cuando este fue usado en un plásmido replicativo basado en el vector pIJ101 [97]. El plásmido pPZ12, identificado por el grupo Pfizer, mostro ser replicativo y estable en *S. rimosus* [111]. En el presente estudio usamos el plásmido pPZ12 y le integramos el plásmido replicativo de *E. coli* pUC19, generando de este modo el vector bifuncional pVF para *E. coli* y para *S. rimosus*. Cuando se probó este vector replicativo con el promotor *PermE**los resultaros mostraron la inestabilidad genética de este sistema de expresión, probablemente dado que *PermE** es un fuerte promotor constitutivo en *S. rimosus*, y las alteraciones genéticas en el plásmido pudieron haber surgido debido a la presencia de las más de 300 copias de este promotor en cada célula lo cual mostro ser

toxico para *S. rimosus*. Los promotores *PnitA-NitR* y *tcp830* mostraron ser débiles cuando se integraron en el cromosoma de *S. rimosus*. Por tanto, fueron funcionales en el plásmido replicativo pVF. Al evaluar transformantes de *S. rimosus* con pVF-*PnitA-NitR/xylE* y pVF-*tcp830/xylE,* estos no mostraron diferencias morfológicas con respecto a la cepa salvaje. Los plásmidos aislados de estas cepas recombinantes mostraron estabilidad estructural y la actividad de XylE observada supera significativamente a la obtenida previamente con el sistema de expresión pAB04-*PermE*/xylE*.

Dado que los nuevos sistemas de expresión de *S. rimosus* conformados por el vector replicativo pVF y los promotores *PnitA-NitR* y *tcp830* mostraron su eficiencia para producir XylE, se decido entonces a producir fitasa, enzima usada por la industria porcina y avícola [116]. La fitasa (AppA) se produce por *S. rimosus* usando el medio de cultivo desarrollado para la producción de oxitetraciclina y la fitasa esta fusionada al péptido señal *srT* el cual es exportado fuera de la célula mediante el mecanismo de secreción Sec. Los resultados obtenidos indicaron que el promotor *tcp830* junto al plásmido pVF es el mejor sistema de sobreexpresión de proteínas recombinantes en *S. rimosus*, alcanzando hasta 20 U/mL de actividad fitasa extracelular, rendimientos comparables a los producidos por *E. coli* (33.2 U/mL de actividad fitasa [114]).

El análisis del transcriptoma de *S. rimosus* mostro que los sistemas de secreción Sec y Tat son ampliamente usados por este microorganismo para exportar proteínas. En el presente estudio usamos el sistema Sec para exportar fitasa, aunque queda abierta la posibilidad de integrar el sistema Tat al kit de herramientas genéticas de *S. rimosus*. Con respecto a las proteasas putativas identificadas, no es posible afirmar que estas proteasas tengan actividad sobre las proteínas recombinantes producidas por *S. rimosus*, en este caso concreto sobre la fitasa. Solo la generación de mutantes de *S. rimosus* incapaces de producir estas proteasas endógenas podrá establecer si estas enzimas están degradando las proteínas recombinantes. El transcriptoma también permitió identificar 52 enzimas putativas con potencial biotecnológico. El kit genético desarrollado para *S. rimosus* en este estudio no solo tiene el potencial de producir proteínas heterólogas en esta bacteria, sino que también puede activar grupos de genes presentes en *S. rimosus* involucrados en la biosíntesis de metabolitos secundarios o producir proteínas nativas que no se han purificado de este microorganismo dado que se encuentran inactivas. Dichas enzimas pueden tener potencial uso biotecnológico. Actualmente *S. rimosus* es conocido por la comunidad científica por su capacidad para producir oxitetraciclina, pero según nuestros resultados, de este microorganismo, potencialmente, se podrían comercializar más productos de origen biológico.

16.7 Conclusión

En la actualidad *E. coli* y *P. pastoris* son los huéspedes más usados para producir proteínas heterólogas. Sin embargo, estos huéspedes no siempre pueden producir la proteína de interés. Este estudio permitió comprobar la capacidad de *S. rimosus* para producir proteínas heterólogas, por tanto *S. rimosus* ofrece una alternativa cuando *E. coli* o *P. pastoris* no pueden producir determinadas proteínas. Las herramientas

genéticas generadas en este estudio permitieron comprobar la capacidad de *S. rimosus* para producir proteínas heterólogas y exportarlas al medio de cultivo facilitando su purificación. Además, se estableció un protocolo para purificar las proteínas exportadas al medio de cultivo usando cromatografía de afinidad. Más aun, el estudio del transcriptoma de *S. rimosus* deja claro el potencial biotecnológico de esta bacteria no se limita a la producción de oxitetraciclina ya que aún quedan metabolitos secundarios y enzimas endógenas presentes en el genoma de *S. rimosus* que pueden ser activadas con las herramientas genéticas desarrolladas en el presente estudio. Dichos productos naturales pueden ser relevantes en la industria farmacéutica, alimentaria, agrícola, de detergentes, entre otras, con lo cual se ampliaría el espectro de productos que podrían ser comercializados a partir de *S. rimosus*.

17 References

1. Bilyk, O. and A. Luzhetskyy, *Metabolic engineering of natural product biosynthesis in actinobacteria.* Curr Opin Biotechnol, 2016. **42**: p. 98-107.
2. Hertweck, C., *The biosynthetic logic of polyketide diversity.* Angew Chem Int Ed Engl, 2009. **48**(26): p. 4688-716.
3. Lim, S.K., et al., *iso-Migrastatin, migrastatin, and dorrigocin production in Streptomyces platensis NRRL 18993 is governed by a single biosynthetic machinery featuring an acyltransferase-less type I polyketide synthase.* J Biol Chem, 2009. **284**(43): p. 29746-56.
4. Chen, H. and L. Du, *Iterative polyketide biosynthesis by modular polyketide synthases in bacteria.* Appl Microbiol Biotechnol, 2016. **100**(2): p. 541-57.
5. Staunton, J. and K.J. Weissman, *Polyketide biosynthesis: a millennium review.* Nat Prod Rep, 2001. **18**(4): p. 380-416.
6. Ray, L. and B.S. Moore, *Recent advances in the biosynthesis of unusual polyketide synthase substrates.* Nat Prod Rep, 2016. **33**(2): p. 150-61.
7. Haydock, S.F., et al., *Divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA:acyl carrier protein transacylase domains in modular polyketide synthases.* FEBS Lett, 1995. **374**(2): p. 246-8.
8. Lau, J., et al., *Dissecting the role of acyltransferase domains of modular polyketide synthases in the choice and stereochemical fate of extender units.* Biochemistry, 1999. **38**(5): p. 1643-51.
9. Petkovic, H., et al., *Substrate specificity of the acyl transferase domains of EpoC from the epothilone polyketide synthase.* Org Biomol Chem, 2008. **6**(3): p. 500-6.
10. Goranovic, D., et al., *Origin of the allyl group in FK506 biosynthesis.* J Biol Chem, 2010. **285**(19): p. 14292-300.
11. Kosec, G., et al., *Novel chemobiosynthetic approach for exclusive production of FK506.* Metab Eng, 2012. **14**(1): p. 39-46.
12. Eustaquio, A.S., et al., *Biosynthesis of the salinosporamide A polyketide synthase substrate chloroethylmalonyl-coenzyme A from S-adenosyl-L-methionine.* Proc Natl Acad Sci U S A, 2009. **106**(30): p. 12295-300.
13. Qu, X., et al., *Cloning, sequencing and characterization of the biosynthetic gene cluster of sanglifehrin A, a potent cyclophilin inhibitor.* Mol Biosyst, 2011. **7**(3): p. 852-61.
14. Husain, S. and N. Singh, *The impact of novel immunosuppressive agents on infections in organ transplant recipients and the interactions of these agents with antimicrobials.* Clin Infect Dis, 2002. **35**(1): p. 53-61.
15. Wu, K., et al., *The FK520 gene cluster of Streptomyces hygroscopicus var. ascomyceticus (ATCC 14891) contains genes for biosynthesis of unusual polyketide extender units.* Gene, 2000. **251**(1): p. 81-90.
16. Dunn, B.J., D.E. Cane, and C. Khosla, *Mechanism and specificity of an acyltransferase domain from a modular polyketide synthase.* Biochemistry, 2013. **52**(11): p. 1839-41.
17. Tang, Y., et al., *Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B synthase.* Chem Biol, 2007. **14**(8): p. 931-43.

18. Tang, Y., et al., *The 2.7-Angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase.* Proc Natl Acad Sci U S A, 2006. **103**(30): p. 11124-9.
19. Liew, C.W., et al., *Crystal structure of the acyltransferase domain of the iterative polyketide synthase in enediyne biosynthesis.* J Biol Chem, 2012. **287**(27): p. 23203-15.
20. Wong, F.T., et al., *Structure and mechanism of the trans-acting acyltransferase from the disorazole synthase.* Biochemistry, 2011. **50**(30): p. 6539-48.
21. Park, H., et al., *A polyketide synthase acyltransferase domain structure suggests a recognition mechanism for its hydroxymalonyl-acyl carrier protein substrate.* PLoS One, 2014. **9**(10): p. e110965.
22. Jiang, H., et al., *An acyltransferase domain of FK506 polyketide synthase recognizing both an acyl carrier protein and coenzyme A as acyl donors to transfer allylmalonyl and ethylmalonyl units.* FEBS J, 2015. **282**(13): p. 2527-39.
23. Wang, Y.Y., et al., *Biochemical characterization of a malonyl-specific acyltransferase domain of FK506 biosynthetic polyketide synthase.* Protein Pept Lett, 2015. **22**(1): p. 2-7.
24. Moncrieffe, M.C., et al., *Structure of the glycosyltransferase EryCIII in complex with its activating P450 homologue EryCII.* J Mol Biol, 2012. **415**(1): p. 92-101.
25. Rosano, G.L. and E.A. Ceccarelli, *Recombinant protein expression in Escherichia coli: advances and challenges.* Front Microbiol, 2014. **5**: p. 172.
26. Schleif, R., *AraC protein, regulation of the l-arabinose operon in Escherichia coli, and the light switch mechanism of AraC action.* FEMS Microbiol Rev, 2010. **34**(5): p. 779-96.
27. Khlebnikov, A. and J.D. Keasling, *Effect of lacY expression on homogeneity of induction from the P(tac) and P(trc) promoters by natural and synthetic inducers.* Biotechnol Prog, 2002. **18**(3): p. 672-4.
28. Yamaguchi, H. and M. Miyazaki, *Refolding techniques for recovering biologically active recombinant proteins from inclusion bodies.* Biomolecules, 2014. **4**(1): p. 235-51.
29. Carrio, M.M. and A. Villaverde, *Construction and deconstruction of bacterial inclusion bodies.* J Biotechnol, 2002. **96**(1): p. 3-12.
30. Ceroni, A., et al., *DISULFIND: a disulfide bonding state and cysteine connectivity prediction server.* Nucleic Acids Res, 2006. **34**(Web Server issue): p. W177-81.
31. Vera, A., et al., *The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures.* Biotechnol Bioeng, 2007. **96**(6): p. 1101-6.
32. Ahmad, M., et al., *Protein expression in Pichia pastoris: recent achievements and perspectives for heterologous protein production.* Appl Microbiol Biotechnol, 2014. **98**(12): p. 5301-17.
33. Anne, J., et al., *Recombinant protein production and streptomycetes.* J Biotechnol, 2012. **158**(4): p. 159-67.
34. Klementz, D., et al., *StreptomeDB 2.0-an extended resource of natural products produced by streptomycetes.* Nucleic Acids Res, 2016. **44**(D1): p. D509-14.
35. Gomez-Escribano, J.P. and M.J. Bibb, *Engineering Streptomyces coelicolor for heterologous expression of secondary metabolite gene clusters.* Microb Biotechnol, 2011. **4**(2): p. 207-15.

36. Gomez-Escribano, J.P. and M.J. Bibb, *Streptomyces coelicolor as an expression host for heterologous gene clusters.* Methods Enzymol, 2012. **517**: p. 279-300.
37. Komatsu, M., et al., *Engineered Streptomyces avermitilis host for heterologous expression of biosynthetic gene cluster for secondary metabolites.* ACS Synth Biol, 2013. **2**(7): p. 384-96.
38. Ikeda, H., S.Y. Kazuo, and S. Omura, *Genome mining of the Streptomyces avermitilis genome and development of genome-minimized hosts for heterologous expression of biosynthetic gene clusters.* J Ind Microbiol Biotechnol, 2014. **41**(2): p. 233-50.
39. Thanapipatsiri, A., et al., *A Streptomyces coelicolor host for the heterologous expression of Type III polyketide synthase genes.* Microb Cell Fact, 2015. **14**: p. 145.
40. Delic, M., et al., *Engineering of protein folding and secretion-strategies to overcome bottlenecks for efficient production of recombinant proteins.* Antioxid Redox Signal, 2014. **21**(3): p. 414-37.
41. Kumar, A. and S. Singh, *Directed evolution: tailoring biocatalysts for industrial applications.* Crit Rev Biotechnol, 2013. **33**(4): p. 365-78.
42. Martinez, J.L., et al., *Pharmaceutical protein production by yeast: towards production of human blood proteins by microbial fermentation.* Curr Opin Biotechnol, 2012. **23**(6): p. 965-71.
43. Demain, A.L. and P. Vaishnav, *Production of recombinant proteins by microbes and higher organisms.* Biotechnol Adv, 2009. **27**(3): p. 297-306.
44. Ward, O.P., *Production of recombinant proteins by filamentous fungi.* Biotechnol Adv, 2012. **30**(5): p. 1119-39.
45. Nevalainen, H. and R. Peterson, *Making recombinant proteins in filamentous fungi- are we expecting too much?* Front Microbiol, 2014. **5**: p. 75.
46. Khan, K.H., *Gene expression in Mammalian cells and its applications.* Adv Pharm Bull, 2013. **3**(2): p. 257-63.
47. Unger, T. and Y. Peleg, *Recombinant protein expression in the baculovirus-infected insect cell system.* Methods Mol Biol, 2012. **800**: p. 187-99.
48. Petkovic, H., et al., *Genetics of Streptomyces rimosus, the oxytetracycline producer.* Microbiol Mol Biol Rev, 2006. **70**(3): p. 704-28.
49. Kieser, T., et al., *Practical Streptomyces genetics*. 2000, Norwich: The John Innes Foundation.
50. Pickens, L.B. and Y. Tang, *Oxytetracycline biosynthesis.* J Biol Chem, 2010. **285**(36): p. 27509-15.
51. Friend, E.J. and D.A. Hopwood, *The linkage map of Streptomyces rimosus.* J Gen Microbiol, 1971. **68**(2): p. 187-97.
52. Nyiri, L., *Variability of Streptomyces rimosus.* Acta Microbiol Acad Sci Hung, 1961. **8**: p. 257-73.
53. Pandza, K., et al., *Physical mapping shows that the unstable oxytetracycline gene cluster of Streptomyces rimosus lies close to one end of the linear chromosome.* Microbiology, 1997. **143 (Pt 5)**: p. 1493-501.
54. Pandza, S., et al., *Recombination between the linear plasmid pPZG101 and the linear chromosome of Streptomyces rimosus can lead to exchange of ends.* Mol Microbiol, 1998. **28**(6): p. 1165-76.

55. Pethick, F.E., et al., *Draft Genome Sequence of the Oxytetracycline-Producing Bacterium Streptomyces rimosus ATCC 10970.* Genome Announc, 2013. **1**(2): p. e0006313.
56. Baranasic, D., et al., *Genome Sequences of the Oxytetracycline Production Strain Streptomyces rimosus R6-500 and Two Mutants with Chromosomal Rearrangements.* Genome Announc, 2014. **2**(4).
57. Gravius, B., et al., *Genetic instability and strain degeneration in Streptomyces rimosus.* Appl Environ Microbiol, 1993. **59**(7): p. 2220-8.
58. Flardh, K. and M.J. Buttner, *Streptomyces morphogenetics: dissecting differentiation in a filamentous bacterium.* Nat Rev Microbiol, 2009. **7**(1): p. 36-49.
59. Horinouchi, S., *Mining and polishing of the treasure trove in the bacterial genus streptomyces.* Biosci Biotechnol Biochem, 2007. **71**(2): p. 283-99.
60. Pigac, J. and H. Schrempf, *A Simple and Rapid Method of Transformation of Streptomyces rimosus R6 and Other Streptomycetes by Electroporation.* Appl Environ Microbiol, 1995. **61**(1): p. 352-6.
61. Tang, Z., et al., *Improved oxytetracycline production in Streptomyces rimosus M4018 by metabolic engineering of the G6PDH gene in the pentose phosphate pathway.* Enzyme Microb Technol, 2011. **49**(1): p. 17-24.
62. Jeon, B.J., et al., *Antifungal activity of rimocidin and a new rimocidin derivative BU16 produced by Streptomyces mauvecolor BU16 and their effects on pepper anthracnose.* J Appl Microbiol, 2016.
63. Davisson, J.W., et al., *Rimocidin, a new antibiotic.* Antibiot Chemother (Northfield), 1951. **1**(5): p. 289-90.
64. Escudero, L., et al., *New Rimocidin/CE-108 Derivatives Obtained by a Crotonyl-CoA Carboxylase/Reductase Gene Disruption in Streptomyces diastaticus var. 108: Substrates for the Polyene Carboxamide Synthase PcsA.* PLoS One, 2015. **10**(8): p. e0135891.
65. Yin, S., et al., *Identification of a cluster-situated activator of oxytetracycline biosynthesis and manipulation of its expression for improved oxytetracycline production in Streptomyces rimosus.* Microb Cell Fact, 2015. **14**: p. 46.
66. Wang, P., et al., *Uncovering the enzymes that catalyze the final steps in oxytetracycline biosynthesis.* J Am Chem Soc, 2013. **135**(19): p. 7138-41.
67. Wang, P., et al., *Genetic characterization of enzymes involved in the priming steps of oxytetracycline biosynthesis in Streptomyces rimosus.* Microbiology, 2011. **157**(Pt 8): p. 2401-9.
68. Smith, A.B., 3rd, et al., *(+)-Rimocidin synthetic studies: construction of the C(1-27) aglycone skeleton.* J Org Chem, 2009. **74**(16): p. 5987-6001.
69. Seco, E.M., et al., *Starter unit choice determines the production of two tetraene macrolides, rimocidin and CE-108, in Streptomyces diastaticus var. 108.* Chem Biol, 2004. **11**(3): p. 357-66.
70. Anne, J., et al., *Protein secretion biotechnology in Gram-positive bacteria with special emphasis on Streptomyces lividans.* Biochim Biophys Acta, 2014. **1843**(8): p. 1750-61.
71. Gauthier, C., H. Li, and R. Morosoli, *Increase in xylanase production by Streptomyces lividans through simultaneous use of the Sec- and Tat-dependent protein export systems.* Appl Environ Microbiol, 2005. **71**(6): p. 3085-92.

72. Vukelic, B., A. Ritonja, and L. Vitale, *Endo-deoxyribonuclease from Streptomyces rimosus.* Appl Microbiol Biotechnol, 1995. **43**(6): p. 1056-60.
73. Vukelic, B., et al., *Extracellular alpha-amylase from Streptomyces rimosus.* Appl Microbiol Biotechnol, 1992. **37**(2): p. 202-4.
74. Lescic, I., et al., *Structural characterization of extracellular lipase from Streptomyces rimosus: assignment of disulfide bridge pattern by mass spectrometry.* Biol Chem, 2004. **385**(12): p. 1147-56.
75. Vitale, L., B. Vukelic, and I. Krizaj, *Extracellular metalloendopeptidase of Streptomyces rimosus.* Arch Microbiol, 2006. **185**(3): p. 183-91.
76. Sinsereekul, N., et al., *Recombinant expression of BTA hydrolase in Streptomyces rimosus and catalytic analysis on polyesters by surface plasmon resonance.* Appl Microbiol Biotechnol, 2010. **86**(6): p. 1775-84.
77. Magdevska, V., et al., *Robust reporter system based on chalcone synthase rppA gene from Saccharopolyspora erythraea.* J Microbiol Methods, 2010. **83**(2): p. 111-9.
78. MAGDEVSKA, V., *Streptomyces rimosus AS A POTENTIAL HOST FOR HETEROLOGOUS PROTEIN EXPRESSION*, in *BIOTECHNICAL FACULTY*. 2011, UNIVERSITY OF LJUBLJANA.
79. Sezonov, G., et al., *Replicase, excisionase, and integrase genes of the Streptomyces element pSAM2 constitute an operon positively regulated by the pra gene.* J Bacteriol, 1998. **180**(12): p. 3056-61.
80. Thorpe, H.M., S.E. Wilson, and M.C. Smith, *Control of directionality in the site-specific recombination system of the Streptomyces phage phiC31.* Mol Microbiol, 2000. **38**(2): p. 232-41.
81. Smokvina, T., F. Boccard, and M. Guerineau, *Cloning and expression vectors in an actinomycetes strain, process for transformation of this strain, actinomycetes strain obtained and preparation of proteins*. 1997, Google Patents.
82. Combes, P., et al., *The streptomyces genome contains multiple pseudo-attB sites for the (phi)C31-encoded site-specific recombination system.* J Bacteriol, 2002. **184**(20): p. 5746-52.
83. Bibb, M.J., G.R. Janssen, and J.M. Ward, *Cloning and analysis of the promoter region of the erythromycin resistance gene (ermE) of Streptomyces erythraeus.* Gene, 1985. **38**(1-3): p. 215-26.
84. Ingram, C., et al., *xylE functions as an efficient reporter gene in Streptomyces spp.: use for the study of galP1, a catabolite-controlled promoter.* J Bacteriol, 1989. **171**(12): p. 6617-24.
85. Wilkinson, C.J., et al., *Increasing the efficiency of heterologous promoters in actinomycetes.* J Mol Microbiol Biotechnol, 2002. **4**(4): p. 417-26.
86. Ruiz-Arribas, A., et al., *Analysis of xysA, a gene from Streptomyces halstedii JM8 that encodes a 45-kilodalton modular xylanase, Xys1.* Appl Environ Microbiol, 1997. **63**(8): p. 2983-8.
87. Vujaklija, D., et al., *A novel streptomycete lipase: cloning, sequencing and high-level expression of the Streptomyces rimosus GDS(L)-lipase gene.* Arch Microbiol, 2002. **178**(2): p. 124-30.

88. Lammertyn, E. and J. Anne, *Modifications of Streptomyces signal peptides and their effects on protein production and secretion.* FEMS Microbiol Lett, 1998. **160**(1): p. 1-10.
89. Gao, X., P. Wang, and Y. Tang, *Engineered polyketide biosynthesis and biocatalysis in Escherichia coli.* Appl Microbiol Biotechnol, 2010. **88**(6): p. 1233-42.
90. Baltz, R.H., *Streptomyces temperate bacteriophage integration systems for stable genetic engineering of actinomycetes (and other organisms).* J Ind Microbiol Biotechnol, 2012. **39**(5): p. 661-72.
91. Fayed, B., et al., *A novel Streptomyces spp. integration vector derived from the S. venezuelae phage, SV1.* BMC Biotechnol, 2014. **14**: p. 51.
92. Morita, K., et al., *The site-specific recombination system of actinophage TG1.* FEMS Microbiol Lett, 2009. **297**(2): p. 234-40.
93. Gregory, M.A., R. Till, and M.C. Smith, *Integration site for Streptomyces phage phiBT1 and development of site-specific integrating vectors.* J Bacteriol, 2003. **185**(17): p. 5320-3.
94. Kuhstoss, S. and R.N. Rao, *Analysis of the integration function of the streptomycete bacteriophage phi C31.* J Mol Biol, 1991. **222**(4): p. 897-908.
95. Fayed, B., et al., *Multiplexed integrating plasmids for engineering of the erythromycin gene cluster for expression in Streptomyces spp. and combinatorial biosynthesis.* Appl Environ Microbiol, 2015. **81**(24): p. 8402-13.
96. Takano, E., et al., *Construction of thiostrepton-inducible, high-copy-number expression vectors for use in Streptomyces spp.* Gene, 1995. **166**(1): p. 133-7.
97. Herai, S., et al., *Hyper-inducible expression system for streptomycetes.* Proc Natl Acad Sci U S A, 2004. **101**(39): p. 14031-5.
98. Rodriguez-Garcia, A., et al., *Natural and synthetic tetracycline-inducible promoters for use in the antibiotic-producing bacteria Streptomyces.* Nucleic Acids Res, 2005. **33**(9): p. e87.
99. Horbal, L., V. Fedorenko, and A. Luzhetskyy, *Novel and tightly regulated resorcinol and cumate-inducible expression systems for Streptomyces and other actinobacteria.* Appl Microbiol Biotechnol, 2014. **98**(20): p. 8641-55.
100. Dangel, V., et al., *Use of an inducible promoter for antibiotic production in a heterologous host.* Appl Microbiol Biotechnol, 2010. **87**(1): p. 261-9.
101. Bertram, R. and W. Hillen, *The application of Tet repressor in prokaryotic gene regulation and expression.* Microb Biotechnol, 2008. **1**(1): p. 2-16.
102. Labes, G., M. Bibb, and W. Wohlleben, *Isolation and characterization of a strong promoter element from the Streptomyces ghanaensis phage I19 using the gentamicin resistance gene (aacC1) of Tn 1696 as reporter.* Microbiology, 1997. **143 (Pt 5)**: p. 1503-12.
103. Kieser, T., et al., *pIJ101, a multi-copy broad host-range Streptomyces plasmid: functional analysis and development of DNA cloning vectors.* Mol Gen Genet, 1982. **185**(2): p. 223-8.
104. Kendall, K.J. and S.N. Cohen, *Complete nucleotide sequence of the Streptomyces lividans plasmid pIJ101 and correlation of the sequence with genetic properties.* J Bacteriol, 1988. **170**(10): p. 4634-51.
105. Gusek, T.W. and J.E. Kinsella, *Review of the Streptomyces lividans/vector pIJ702 system for gene cloning.* Crit Rev Microbiol, 1992. **18**(4): p. 247-60.

106. Servin-Gonzalez, L., et al., *Sequence and functional analysis of the Streptomyces phaeochromogenes plasmid pJV1 reveals a modular organization of Streptomyces plasmids that replicate by rolling circle.* Microbiology, 1995. **141 (Pt 10)**: p. 2499-510.
107. Maas, R.M., et al., *The conjugative plasmid pSG5 from Streptomyces ghanaensis DSM 2932 differs in its transfer functions from other Streptomyces rolling-circle-type plasmids.* Microbiology, 1998. **144 (Pt 10)**: p. 2809-17.
108. Haug, I., et al., *Streptomyces coelicolor A3(2) plasmid SCP2*: deductions from the complete sequence.* Microbiology, 2003. **149**(Pt 2): p. 505-13.
109. Servin-Gonzalez, L., *Relationship between the replication functions of Streptomyces plasmids pJV1 and pIJ101.* Plasmid, 1993. **30**(2): p. 131-40.
110. Butler, M.J., et al., *Molecular cloning of resistance genes and architecture of a linked gene cluster involved in biosynthesis of oxytetracycline by Streptomyces rimosus.* Mol Gen Genet, 1989. **215**(2): p. 231-8.
111. CHAMBERS, A.E. and I.S. HUNTER, *Construction and use of a bifunctional streptomycete cosmid.* Biochemical Society Transactions, 1984. **12**(4): p. 644-645.
112. Jia, Z., et al., *Purification, crystallization and preliminary X-ray analysis of the Escherichia coli phytase.* Acta Crystallogr D Biol Crystallogr, 1998. **54**(Pt 4): p. 647-9.
113. Lim, D., et al., *Crystal structures of Escherichia coli phytase and its complex with phytate.* Nat Struct Biol, 2000. **7**(2): p. 108-13.
114. Miksch, G., et al., *Overexpression of the phytase from Escherichia coli and its extracellular production in bioreactors.* Appl Microbiol Biotechnol, 2002. **59**(6): p. 685-94.
115. Akbarzadeh, A., et al., *Optimization of Recombinant Expression of Synthetic Bacterial Phytase in Pichia pastoris Using Response Surface Methodology.* Jundishapur J Microbiol, 2015. **8**(12): p. e27553.
116. Golovan, S., et al., *Characterization and overproduction of the Escherichia coli appA encoded bifunctional enzyme that exhibits both phytase and acid phosphatase activities.* Can J Microbiol, 2000. **46**(1): p. 59-71.
117. Berkmen, M., D. Boyd, and J. Beckwith, *The nonconsecutive disulfide bond of Escherichia coli phytase (AppA) renders it dependent on the protein-disulfide isomerase, DsbC.* J Biol Chem, 2005. **280**(12): p. 11387-94.
118. de Marco, A., *Strategies for successful recombinant expression of disulfide bond-dependent proteins in Escherichia coli.* Microb Cell Fact, 2009. **8**: p. 26.
119. Lobstein, J., et al., *SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm.* Microb Cell Fact, 2012. **11**: p. 56.
120. Stahl, C.H., D.B. Wilson, and X.G. Lei, *Comparison of extracellular Escherichia coli AppA phytases expressed in Streptomyces lividans and Pichia pastoris.* Biotechnol Lett, 2003. **25**(10): p. 827-31.
121. Hranueli, D., J. Pigac, and M. Vesligaj, *Characterization and persistence of actinophage RP2 isolated from Streptomyces rimosus ATCC 10970.* J Gen Microbiol, 1979. **114**(2): p. 295-303.

122. RHODES, P.M., et al., *Biochemical and Genetic Characterization of Streptomyces rimosus Mutants Impaired in Oxytetracycline Biosynthesis.* Microbiology, 1981. **124**(2): p. 329-338.
123. Hunter, I.S.D., *Cosmid cloning vectors*. 1985, Google Patents.
124. Bradford, M.M., *A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding.* Anal Biochem, 1976. **72**: p. 248-54.
125. Lee, S., et al., *Expression of Escherichia coli AppA2 phytase in four yeast systems.* Biotechnol Lett, 2005. **27**(5): p. 327-34.
126. Shevchenko, A., et al., *In-gel digestion for mass spectrometric characterization of proteins and proteomes.* Nat. Protocols, 2007. **1**(6): p. 2856-2860.
127. Ginsburg, D.S. and M.P. Calos, *Site-specific integration with phiC31 integrase for prolonged expression of therapeutic genes.* Adv Genet, 2005. **54**: p. 179-87.
128. Phornphisutthimas, S., et al., *Development of an intergeneric conjugal transfer system for rimocidin-producing Streptomyces rimosus.* Lett Appl Microbiol, 2010. **50**(5): p. 530-6.
129. Nan, G.L. and V. Walbot, *Plasmid rescue: recovery of flanking genomic sequences from transgenic transposon insertion sites.* Methods Mol Biol, 2009. **526**: p. 101-9.
130. Kim, M.K., H.S. Ha, and S.U. Choi, *Conjugal transfer using the bacteriophage phiC31 att/int system and properties of the attB site in Streptomyces ambofaciens.* Biotechnol Lett, 2008. **30**(4): p. 695-9.
131. Pandey, A.K., et al., *Nitrile-inducible gene expression in mycobacteria.* Tuberculosis (Edinb), 2009. **89**(1): p. 12-6.
132. Degenkolb, J., et al., *Structural requirements of tetracycline-Tet repressor interaction: determination of equilibrium binding constants for tetracycline analogs with the Tet repressor.* Antimicrob Agents Chemother, 1991. **35**(8): p. 1591-5.
133. Ishikawa, J. and K. Hotta, *FramePlot: a new implementation of the frame analysis for predicting protein-coding regions in bacterial DNA with a high G + C content.* FEMS Microbiol Lett, 1999. **174**(2): p. 251-3.
134. Kendall, K.J. and S.N. Cohen, *Plasmid transfer in Streptomyces lividans: identification of a kil-kor system associated with the transfer region of pIJ101.* J Bacteriol, 1987. **169**(9): p. 4177-83.
135. Chater, K.F., et al., *The complex extracellular biology of Streptomyces.* FEMS Microbiol Rev, 2010. **34**(2): p. 171-98.
136. Baltz, R.H., *Genetic manipulation of secondary metabolite biosynthesis for improved production in Streptomyces and other actinomycetes.* J Ind Microbiol Biotechnol, 2016. **43**(2-3): p. 343-70.
137. Li, P., et al., *Comparing the normalization methods for the differential analysis of Illumina high-throughput RNA-Seq data.* BMC Bioinformatics, 2015. **16**: p. 347.
138. Oshlack, A. and M.J. Wakefield, *Transcript length bias in RNA-seq data confounds systems biology.* Biol Direct, 2009. **4**: p. 14.
139. Gullon, S., et al., *Exploring the Feasibility of the Sec Route to Secrete Proteins Using the Tat Route in Streptomyces lividans.* Mol Biotechnol, 2015. **57**(10): p. 931-8.

140. Beckwith, J., *The Sec-dependent pathway.* Res Microbiol, 2013. **164**(6): p. 497-504.
141. Schneewind, O. and D. Missiakas, *Sec-secretion and sortase-mediated anchoring of proteins in Gram-positive bacteria.* Biochim Biophys Acta, 2014. **1843**(8): p. 1687-97.
142. Qi, H.Y., J.B. Hyndman, and H.D. Bernstein, *DnaK promotes the selective export of outer membrane protein precursors in SecA-deficient Escherichia coli.* J Biol Chem, 2002. **277**(52): p. 51077-83.
143. Moran-Barrio, J., A.S. Limansky, and A.M. Viale, *Secretion of GOB metallo-beta-lactamase in Escherichia coli depends strictly on the cooperation between the cytoplasmic DnaK chaperone system and the Sec machinery: completion of folding and Zn(II) ion acquisition occur in the bacterial periplasm.* Antimicrob Agents Chemother, 2009. **53**(7): p. 2908-17.
144. Wild, J., et al., *DnaK and DnaJ heat shock proteins participate in protein export in Escherichia coli.* Genes Dev, 1992. **6**(7): p. 1165-72.
145. Tiwari, S., et al., *Unique structural modulation of a non-native substrate by cochaperone DnaJ.* Biochemistry, 2013. **52**(6): p. 1011-8.
146. Srinivasan, S.R., et al., *Molecular chaperones DnaK and DnaJ share predicted binding sites on most proteins in the E. coli proteome.* Mol Biosyst, 2012. **8**(9): p. 2323-33.
147. Jolly, C. and R.I. Morimoto, *Role of the heat shock response and molecular chaperones in oncogenesis and cell death.* J Natl Cancer Inst, 2000. **92**(19): p. 1564-72.
148. Geukens, N., et al., *Membrane topology of the Streptomyces lividans type I signal peptidases.* J Bacteriol, 2001. **183**(16): p. 4752-60.
149. Palacin, A., et al., *SipY Is the Streptomyces lividans type I signal peptidase exerting a major effect on protein secretion.* J Bacteriol, 2002. **184**(17): p. 4875-80.
150. Goosens, V.J., C.G. Monteferrante, and J.M. van Dijl, *The Tat system of Gram-positive bacteria.* Biochim Biophys Acta, 2014. **1843**(8): p. 1698-706.
151. Kudva, R., et al., *Protein translocation across the inner membrane of Gram-negative bacteria: the Sec and Tat dependent protein transport pathways.* Res Microbiol, 2013. **164**(6): p. 505-34.
152. Palmer, T. and B.C. Berks, *The twin-arginine translocation (Tat) protein export pathway.* Nat Rev Micro, 2012. **10**(7): p. 483-496.
153. Auclair, S.M., M.K. Bhanu, and D.A. Kendall, *Signal peptidase I: cleaving the way to mature proteins.* Protein Sci, 2012. **21**(1): p. 13-25.
154. Barnett, J.P., et al., *A minimal Tat system from a gram-positive organism: a bifunctional TatA subunit participates in discrete TatAC and TatA complexes.* J Biol Chem, 2008. **283**(5): p. 2534-42.
155. Sinha, J., et al., *Causes of proteolytic degradation of secreted recombinant proteins produced in methylotrophic yeast Pichia pastoris: case study with recombinant ovine interferon-tau.* Biotechnol Bioeng, 2005. **89**(1): p. 102-12.
156. Binnie, C., J.D. Cossar, and D.I. Stewart, *Heterologous biopharmaceutical protein expression in Streptomyces.* Trends Biotechnol, 1997. **15**(8): p. 315-20.
157. Krishnappa, L., et al., *Degradation of extracytoplasmic catalysts for protein folding in Bacillus subtilis.* Appl Environ Microbiol, 2014. **80**(4): p. 1463-8.

158. Niyonzima, F.N. and S. More, *Detergent-compatible proteases: microbial production, properties, and stain removal analysis.* Prep Biochem Biotechnol, 2015. **45**(3): p. 233-58.
159. Li, Q., et al., *Commercial proteases: present and future.* FEBS Lett, 2013. **587**(8): p. 1155-63.
160. Siezen, R.J. and J.A. Leunissen, *Subtilases: the superfamily of subtilisin-like serine proteases.* Protein Sci, 1997. **6**(3): p. 501-23.
161. Nonaka, T., et al., *The crystal structure of an oxidatively stable subtilisin-like alkaline serine protease, KP-43, with a C-terminal beta-barrel domain.* J Biol Chem, 2004. **279**(45): p. 47344-51.
162. Page, M.J. and E. Di Cera, *Serine peptidases: classification, structure and function.* Cell Mol Life Sci, 2008. **65**(7-8): p. 1220-36.
163. Screen, S.E. and R.J. St Leger, *Cloning, expression, and substrate specificity of a fungal chymotrypsin. Evidence for lateral gene transfer from an actinomycete bacterium.* J Biol Chem, 2000. **275**(9): p. 6689-94.
164. Volonte, F., et al., *Overexpression of a bacterial chymotrypsin: application for L-amino acid ester hydrolysis.* Enzyme Microb Technol, 2011. **49**(6-7): p. 560-6.
165. Wu, J.W. and X.L. Chen, *Extracellular metalloproteases from bacteria.* Appl Microbiol Biotechnol, 2011. **92**(2): p. 253-62.
166. Inouye, K., et al., *Engineering, expression, purification, and production of recombinant thermolysin.* Biotechnol Annu Rev, 2007. **13**: p. 43-64.
167. Venkatesh, S., et al., *Multitasking in the mitochondrion by the ATP-dependent Lon protease.* Biochim Biophys Acta, 2012. **1823**(1): p. 56-66.
168. Frees, D., U. Gerth, and H. Ingmer, *Clp chaperones and proteases are central in stress survival, virulence and antibiotic resistance of Staphylococcus aureus.* Int J Med Microbiol, 2014. **304**(2): p. 142-9.
169. Santos-Beneit, F., et al., *Phosphate and carbon source regulation of two PhoP-dependent glycerophosphodiester phosphodiesterase genes of Streptomyces coelicolor.* Microbiology, 2009. **155**(Pt 6): p. 1800-11.
170. Aloulou, A., et al., *Phospholipases: an overview.* Methods Mol Biol, 2012. **861**: p. 63-85.
171. Butler, M.J., et al., *Cloning and characterization of a gene encoding a secreted tripeptidyl aminopeptidase from Streptomyces lividans 66.* Appl Environ Microbiol, 1995. **61**(8): p. 3145-50.
172. Holm, K.A., et al., *Tripeptidyl aminopeptidase*. 1998, Google Patents.
173. Hartl, L., S. Zach, and V. Seidl-Seiboth, *Fungal chitinases: diversity, mechanistic properties and biotechnological potential.* Appl Microbiol Biotechnol, 2012. **93**(2): p. 533-43.
174. Gupta, R., N. Gupta, and P. Rathi, *Bacterial lipases: an overview of production, purification and biochemical properties.* Appl Microbiol Biotechnol, 2004. **64**(6): p. 763-81.
175. Suphatharaprateep, W., B. Cheirsilp, and A. Jongjareonrak, *Production and properties of two collagenases from bacteria and their application for collagen extraction.* N Biotechnol, 2011. **28**(6): p. 649-55.
176. Gong, J.S., et al., *Nitrilases in nitrile biocatalysis: recent progress and forthcoming research.* Microb Cell Fact, 2012. **11**: p. 142.

177. de Souza, P.M. and P. de Oliveira Magalhaes, *Application of microbial alpha-amylase in industry - A review.* Braz J Microbiol, 2010. **41**(4): p. 850-61.
178. Kuhad, R.C., R. Gupta, and A. Singh, *Microbial cellulases and their industrial applications.* Enzyme Res, 2011. **2011**: p. 280696.
179. Zhou, Z., et al., *Extracellular overexpression of chitosanase from Bacillus sp. TS in Escherichia coli.* Appl Biochem Biotechnol, 2015. **175**(7): p. 3271-86.
180. Liang, T.W., et al., *Purification of chitinase/chitosanase from Bacillus cereus and discovery of an enzyme inhibitor.* Int J Biol Macromol, 2014. **63**: p. 8-14.
181. Beg, Q.K., et al., *Microbial xylanases and their industrial applications: a review.* Appl Microbiol Biotechnol, 2001. **56**(3-4): p. 326-38.
182. Li, X., et al., *Cloning, expression and characterization of a pectate lyase from Paenibacillus sp. 0602 in recombinant Escherichia coli.* BMC Biotechnol, 2014. **14**: p. 18.
183. Cortial, S., et al., *NADH oxidase activity of Bacillus subtilis nitroreductase NfrA1: insight into its biological role.* FEBS Lett, 2010. **584**(18): p. 3916-22.
184. Gonzalez, J.M. and S.Z. Fisher, *Carbonic anhydrases in industrial applications.* Subcell Biochem, 2014. **75**: p. 405-26.
185. Kieliszek, M. and A. Misiewicz, *Microbial transglutaminase and its application in the food industry. A review.* Folia Microbiol (Praha), 2014. **59**(3): p. 241-50.
186. Pollegioni, L., L. Piubelli, and G. Molla, *Cholesterol oxidase: biotechnological applications.* FEBS J, 2009. **276**(23): p. 6857-70.
187. Sathya, T.A. and M. Khan, *Diversity of glycosyl hydrolase enzymes from metagenome and their application in food industry.* J Food Sci, 2014. **79**(11): p. R2149-56.
188. Luo, S.L., et al., *Cloning and heterologous expression of UDP-glycosyltransferase genes from Bacillus subtilis and its application in the glycosylation of ginsenoside Rh1.* Lett Appl Microbiol, 2015. **60**(1): p. 72-8.
189. Katrolia, P., et al., *Biotechnological potential of microbial alpha-galactosidases.* Crit Rev Biotechnol, 2014. **34**(4): p. 307-17.
190. Jordan, D.B. and K. Wagschal, *Properties and applications of microbial beta-D-xylosidases featuring the catalytically efficient enzyme from Selenomonas ruminantium.* Appl Microbiol Biotechnol, 2010. **86**(6): p. 1647-58.
191. Hii, S.L., et al., *Pullulanase: role in starch hydrolysis and potential industrial applications.* Enzyme Res, 2012. **2012**: p. 921362.
192. Vavrova, L., K. Muchova, and I. Barak, *Comparison of different Bacillus subtilis expression systems.* Res Microbiol, 2010. **161**(9): p. 791-7.
193. Da Silva, N.A. and S. Srikrishnan, *Introduction and expression of genes for metabolic engineering applications in Saccharomyces cerevisiae.* FEMS Yeast Res, 2012. **12**(2): p. 197-214.
194. Heap, J.T., et al., *Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker.* Nucleic Acids Res, 2012. **40**(8): p. e59.
195. Gu, P., et al., *A rapid and reliable strategy for chromosomal integration of gene(s) with multiple copies.* Sci Rep, 2015. **5**: p. 9684.
196. Kuhlman, T.E. and E.C. Cox, *Site-specific chromosomal integration of large synthetic constructs.* Nucleic Acids Res, 2010. **38**(6): p. e92.
197. Li, P., et al., *An efficient blue-white screening based gene inactivation system for Streptomyces.* Appl Microbiol Biotechnol, 2015. **99**(4): p. 1923-33.

198. Zhang, L., et al., *Control of directionality in Streptomyces phage phiBT1 integrase-mediated site-specific recombination.* PLoS One, 2013. **8**(11): p. e80434.
199. Boccard, F., et al., *Structural analysis of loci involved in pSAM2 site-specific integration in Streptomyces.* Plasmid, 1989. **21**(1): p. 59-70.
200. Mazodier, P., C. Thompson, and F. Boccard, *The chromosomal integration site of the Streptomyces element pSAM2 overlaps a putative tRNA gene conserved among actinomycetes.* Mol Gen Genet, 1990. **222**(2-3): p. 431-4.
201. Bryant, J.A., et al., *Chromosome position effects on gene expression in Escherichia coli K-12.* Nucleic Acids Res, 2014. **42**(18): p. 11383-92.
202. Kim, E.J., I. Yang, and Y.J. Yoon, *Developing Streptomyces venezuelae as a cell factory for the production of small molecules used in drug discovery.* Arch Pharm Res, 2015. **38**(9): p. 1606-16.
203. Balbas, P. and F. Bolivar, *Back to basics: pBR322 and protein expression systems in E. coli.* Methods Mol Biol, 2004. **267**: p. 77-90.
204. Sun, N., et al., *Construction of over-expression shuttle vectors in Streptomyces.* Annals of Microbiology, 2012. **62**(4): p. 1541-1546.
205. Ducote, M.J. and G.S. Pettis, *Common and distinguishing regulatory and expression characteristics of the highly related KorB proteins of streptomycete plasmids pIJ101 and pSB24.2.* J Bacteriol, 2003. **185**(13): p. 3780-7.
206. Mulleder, M., et al., *A prototrophic deletion mutant collection for yeast metabolomics and systems biology.* Nat Biotechnol, 2012. **30**(12): p. 1176-8.
207. Borsuk, S., et al., *Auxotrophic complementation as a selectable marker for stable expression of foreign antigens in Mycobacterium bovis BCG.* Tuberculosis (Edinb), 2007. **87**(6): p. 474-80.
208. Chen, C.-C., et al., *A Pichia pastoris fermentation strategy for enhancing the heterologous expression of an Escherichia coli phytase.* Enzyme and Microbial Technology, 2004. **35**(4): p. 315-320.
209. Wyss, M., et al., *Biochemical characterization of fungal phytases (myo-inositol hexakisphosphate phosphohydrolases): catalytic properties.* Appl Environ Microbiol, 1999. **65**(2): p. 367-73.
210. Zhang, L., C.P. Chou, and M. Moo-Young, *Disulfide bond formation and its impact on the biological activity and stability of recombinant therapeutic proteins produced by Escherichia coli expression system.* Biotechnol Adv, 2011. **29**(6): p. 923-9.
211. Wong, J.W., S.Y. Ho, and P.J. Hogg, *Disulfide bond acquisition through eukaryotic protein evolution.* Mol Biol Evol, 2011. **28**(1): p. 327-34.
212. Kamionka, M., *Engineering of therapeutic proteins production in Escherichia coli.* Curr Pharm Biotechnol, 2011. **12**(2): p. 268-74.
213. Posfai, G., et al., *Emergent properties of reduced-genome Escherichia coli.* Science, 2006. **312**(5776): p. 1044-6.
214. Beites, T. and M.V. Mendes, *Chassis optimization as a cornerstone for the application of synthetic biology based strategies in microbial secondary metabolism.* Front Microbiol, 2015. **6**: p. 906.
215. Ellis, M., et al., *Development of a high yielding E. coli periplasmic expression system for the production of humanized Fab' fragments.* Biotechnol Prog, 2016.

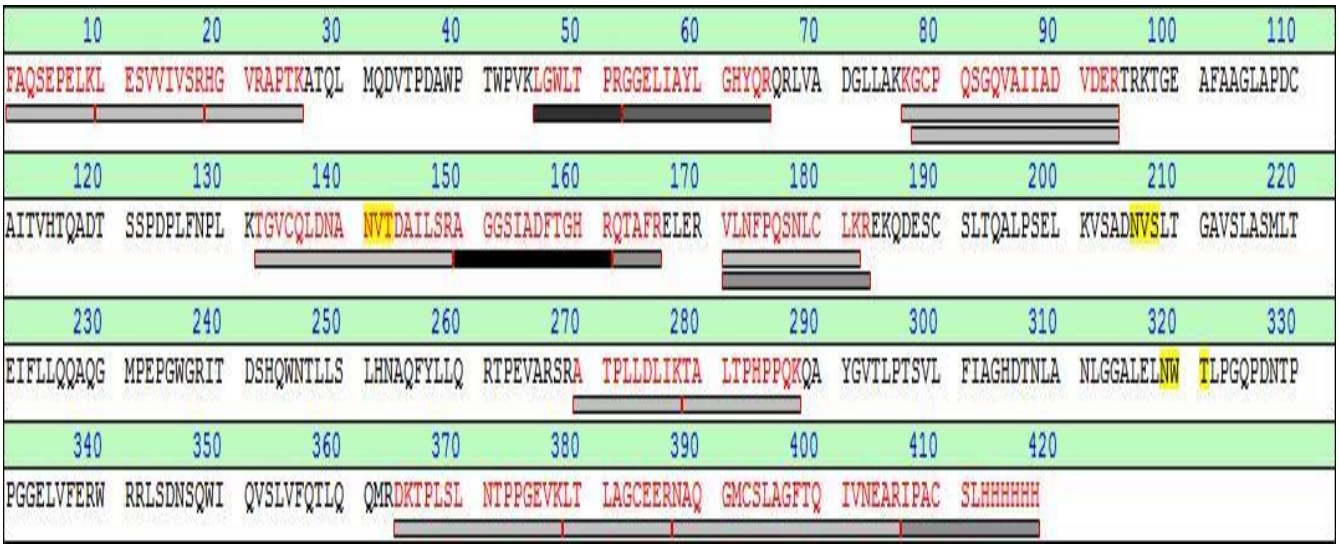
216. Nguyen, T.T., T.D. Quyen, and H.T. Le, *Cloning and enhancing production of a detergent- and organic-solvent-resistant nattokinase from Bacillus subtilis VTCC-DVN-12-01 by using an eight-protease-gene-deficient Bacillus subtilis WB800.* Microb Cell Fact, 2013. **12**: p. 79.
217. Tomimoto, K., et al., *Protease-deficient Saccharomyces cerevisiae strains for the synthesis of human-compatible glycoproteins.* Biosci Biotechnol Biochem, 2013. **77**(12): p. 2461-6.
218. Bagos, P.G., et al., *Combined prediction of Tat and Sec signal peptides with hidden Markov models.* Bioinformatics, 2010. **26**(22): p. 2811-7.
219. Chen, J., et al., *Combinatorial Sec pathway analysis for improved heterologous protein secretion in Bacillus subtilis: identification of bottlenecks by systematic gene overexpression.* Microbial Cell Factories, 2015. **14**(1): p. 92.
220. Chen, J., et al., *Enhanced extracellular production of alpha-amylase in Bacillus subtilis by optimization of regulatory elements and over-expression of PrsA lipoprotein.* Biotechnol Lett, 2015. **37**.
221. Bos, J.W., et al., *Production of active Pseudomonas glumae lipase in homologous or heterologous hosts*. 1998, Google Patents.
222. Quax, W.J., R. Kerkman, and C.P. Broekhuizen, *Secretion factors for gram-positive microorganisms genes encoding them and methods of using it*. 2002, Google Patents.
223. Diao, L., et al., *Functional implementation of the posttranslational SecB-SecA protein-targeting pathway in Bacillus subtilis.* Appl Environ Microbiol, 2012. **78**(3): p. 651-9.
224. Kakeshita, H., et al., *Enhanced extracellular production of heterologous proteins in Bacillus subtilis by deleting the C-terminal region of the SecA secretory machinery.* Mol Biotechnol, 2010. **46**(3): p. 250-7.
225. Mulder, K.C., J. Bandola, and W. Schumann, *Construction of an artificial secYEG operon allowing high level secretion of alpha-amylase.* Protein Expr Purif, 2013. **89**(1): p. 92-6.
226. Malten, M., et al., *Coexpression of the type I signal peptidase gene sipM increases recombinant protein production and export in Bacillus megaterium MS941.* Biotechnol Bioeng, 2005. **91**(5): p. 616-21.
227. Widdick, D.A., et al., *The twin-arginine translocation pathway is a major route of protein export in Streptomyces coelicolor.* Proc Natl Acad Sci U S A, 2006. **103**(47): p. 17927-32.
228. Miyazaki, T., et al., *Hyper secretion of Thermobifida fusca beta-glucosidase via a Tat-dependent signal peptide using Streptomyces lividans.* Microb Cell Fact, 2013. **12**: p. 88.
229. Curran, K.A., et al., *Short Synthetic Terminators for Improved Heterologous Gene Expression in Yeast.* ACS Synth Biol, 2015. **4**(7): p. 824-32.
230. Chan, Y.A., et al., *Biosynthesis of polyketide synthase extender units.* Nat Prod Rep, 2009. **26**(1): p. 90-114.

18 Supplementary information

18.1 LC-MS/MS analysis of phytase purified from *S. rimosus* supernatant

Sequence Data:

Intensity Coverage: 27.3%
Sequence coverage MS: 44.5%
pI (Isoelectric point): 6.4



Phytase amino acid sequence

MW:46158.581

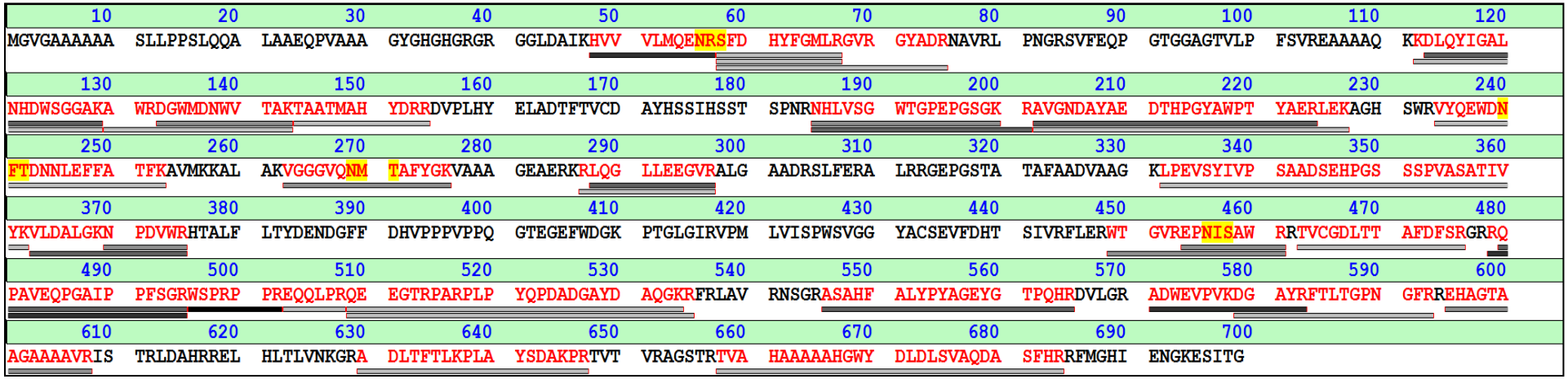
FAQSEPELKLESVVIVSRHGVRAPTKATQLMQDVTPDAWPTWPVKLGWLTPRGGELIAY
LGHYQRQRLVADGLLAKKGCPQSGQVAIIADVDERTRKTGEAFAAGLAPDCAITVHTQA
DTSSPDPLFNPLKTGVCQLDNANVTDAILSRAGGSIADFTGHRQTAFRELERVLNFPQS
NLCLKREKQDESCSLTQALPSELKVSADNVSLTGAVSLASMLTEIFLLQQAQGMPEPGW
GRITDSHQWNTLLSLHNAQFYLLQRTPEVARSRATPLLDLIKTALTPHPPQKQAYGVTL
PTSVLFIAGHDTNLANLGGALELNWTLPGQPDNTPPGGELVFERWRRLSDNSQWIQVSL
VFQTLQQMRDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSLH
HHHHH

18.2 LC-MS/MS identification of phospholipase C

Protein 1: phosphocholine-specific phospholipase C [Streptomyces rimosus subsp. rimosus ATCC 10970]

Accession:	gi|440615177	**Score:**	166.00
Database:	NCBInr	**MW [kDa]:**	75.90
Seq. Coverage [%]:	55.70 %	**pI:**	9.02
		No. of Peptides:	35

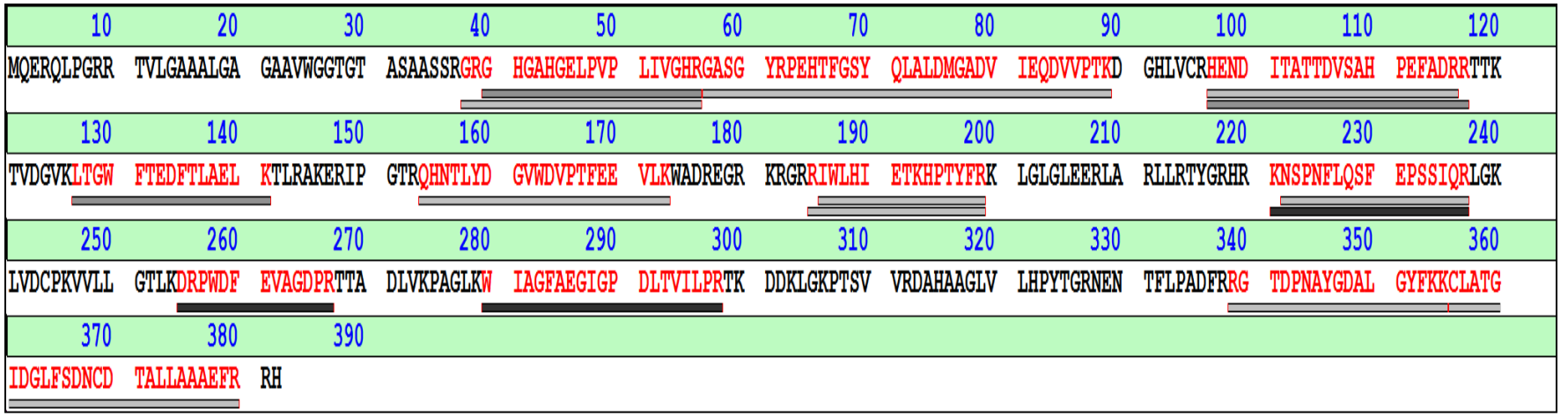
Modification(s): Carbamidomethyl, Oxidation



18.3 LC-MS/MS identification of glycerophosphoryl diester phosphodiesterase

Protein 1:	MULTISPECIES: glycerophosphoryl diester phosphodiesterase [Streptomyces]		
Accession:	gi|490080452	**Score:**	86.70
Database:	NCBInr	**MW [kDa]:**	42.00
Seq. Coverage [%]:	56.00 %	**pI:**	9.17
		No. of Peptides:	15

Modification(s): Carbamidomethyl



18.4 LC-MS/MS identification of secreted tripeptidyl aminopeptidase

Protein 1:	MULTISPECIES: secreted tripeptidyl aminopeptidase [Streptomyces]		
Accession:	gi|490078006	**Score:**	80.60
Database:	NCBInr	**MW [kDa]:**	53.80
Seq. Coverage [%]:	37.20 %	**pI:**	9.24
		No. of Peptides:	16

Modification(s): Oxidation