Universidade de Aveiro Departamento de Química
2021

Margarida de Sousa Gomes

Peptone from casein: its influence in the transcription of lanthipeptides and the proteome of *Pedobacter lusitanus* NL19

Peptona de caseína: a sua influência na transcrição de lantipéptidos e no proteoma de *Pedobacter lusitanus* NL19

Universidade de Aveiro Departamento de Química
2021

Margarida de Sousa Gomes

Peptone from casein: its influence in the transcription of lanthipeptides and the proteome of *Pedobacter lusitanus* NL19

Peptona de caseína: a sua influência na transcrição de lantipéptidos e no proteoma de *Pedobacter lusitanus* NL19

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Biotecnologia Molecular, realizada sob a orientação científica da Doutora Tânia Caetano, Investigadora Auxiliar do Departamento de Biologia e da Professora Doutora Sónia Mendo, Professora Auxiliar com Agregação do Departamento de Biologia da Universidade de Aveiro.

Dedico aos meus pais Arminda e António, e à minha irmã Raquel.

o júri

presidente

Doutora Elisabete Verde Martins Coelho

investigadora auxiliar (Nivel 1) do Departamento de Quimica da Universidade de Aveiro

Doutora Joana Cristina Pacheco Barbosa

investigadora doutorada da Escola Superior de Biotecnologia da Universidade Católica Portuguesa

Doutora Tânia Isabel Sousa Caetano

investigadora equiparada a auxiliar do Departamento de Biologia da Universidade de Aveiro

agradecimentos

Em primeiro lugar quero agradecer à professora Tânia pela paciência, disponibilidade, auxílio e confiança que me deu. Um profundo obrigado por ter aceitado ser minha orientadora. Sem a sua ajuda para resolver os problemas que foram surgindo não teria conseguido realizar este trabalho.

Não posso deixar de agradecer à professora Sónia por me ter recebido no seu laboratório e por me ter integrado na sua "família LBM". Obrigada pela simpatia, interesse e apoio demonstrados.

Um obrigado ao professor Hugo Osório do i3S da Universidade do Porto, pela ajuda, explicações e disponibilidade na realização da análise proteómica.

Agradeço também às ettes (Susana e Letícia) pela amizade e por todos os momentos de descompressão. À Joana e à Diana por me ajudarem sempre que precisei e pela alegria no laboratório. Ao Gonçalo e ao Rúben pela ajuda e amizade. A todos os elementos do LBM um muito sincero obrigado.

Um grande obrigado à Família Martim (Drika, Joana, João Luís, Ana, Dri, Governo, Xana, Carol, Victor, Daniela e Bubas). Pelos momentos que partilhados juntos, pelo apoio, diversão e amizade. O que Coimbra uniu ninguém separa.

Por último quero agradecer à minha família. Aos meus pais, irmã e sobrinhas. Pelo apoio incondicional e pelo interesse no meu percurso académico. Obrigado por estarem sempre presentes.

palavras-chave

Lantipéptidos, RT-qPCR, transcriptómica, proteómica, sequenciação, mineração de genomas, *Sphingobacteriaceae*

resumo

Pedobacter lusitanus NL19 é uma bactéria de Gram-negativo, da família *Sphingobacteriaceae*, que foi isolada de uma mina de urânio desativada em Portugal. Esta estirpe produz péptidos não ribossomais (NRPs), designados por pedopeptinas. A produção destes péptidos é reprimida em meio com elevadas concentrações de peptona de caseína (PC). Para além de clusters biosintéticos (BGCs) que codificam NRPs, o genoma da estirpe NL19 também possui BGCs de outros metabolitos secundários (SMs), incluindo lantipéptidos (4 BGCs: *ped*8, *ped*14, *ped*15 e *ped*17). Lantipéptidos são péptidos de síntese ribossomal com modificações pós-traducionais (RiPPs), que exibem uma ampla variedade de atividades biológicas, incluindo antimicrobiana e antialodínica. Estes péptidos são caracterizados pela presença de resíduos de lantionina (Lan) e metillantionina (MeLan) e dividem-se em quatro classes, definidas pelas enzimas que catalisam as reações que originam esses resíduos. Este estudo focou-se na estirpe NL19 e os principais objetivos foram determinar o efeito de elevadas concentrações de PC: i) na transcrição de lantipéptidos e ii) no proteoma da estirpe, em particular proteínas envolvidas na síntese de SMs. Devido à pandemia COVID-19 e consequente confinamento, foi definido um terceiro objetivo, que consistiu na identificação e análise *in silico* de BGCs de lantipéptidos presentes nos genomas de outros géneros da família *Sphingobacteriaceae*. O objetivo i) envolveu a análise transcricional do BGC *ped*15, que codifica dois péptidos percursores e outras proteínas biosintéticas. Para tal, procedeu-se à sequenciação da região a montante deste cluster, por *primer walking*, o que permitiu identificar outros seis genes de péptidos percursores. A análise transcricional foi realizada por RT-qPCR e revelou que elevadas concentrações de PC não alteram a expressão do BGC *ped*15, ao contrário do que acontece com o BGC das pedopeptinas. De uma forma geral, os resultados de RT-qPCR validaram os resultados de RNA-seq disponíveis, e mostram que o efeito repressor da PC não é transversal à produção de todos os SMs. O objetivo ii) incluiu a análise do proteoma da estirpe NL19 cultivada com elevadas concentrações de PC (e respetivo controlo) por nano LC-ESI-MS/MS e permitiu detetar a expressão diferencial de várias proteínas relacionadas com a biosíntese de SMs, incluindo as péptido sintetases não ribossomais das pedopeptinas e de um precursor de lantipéptido codificado no BGC *ped*8. O objetivo iii) envolveu a análise de 446 BGCs da família *Sphingobacteriaceae* com as ferramentas bioinformáticas antiSMASH e BiG-SCAPE. Foram identificados BGCs de lantipéptidos de classe I e classe III nos géneros *Mucilaginibacter* e *Sphingobacterium*. Os BGCs de classe III codificam enzimas LanKC com domínios liase um pouco distintos, o que pode indicar que estas enzimas utilizam um mecanismo de formação de Lan/MeLan relativamente diferente daquele já conhecido. Este estudo contribui para o conhecimento da resposta bacteriana à manipulação de meios de cultura, em particular na produção de SMs com potencial biotecnológico como os lantipéptidos. Para além disso, permitiu identificar o potencial de géneros bacterianos já conhecidos, mas até agora inexplorados, para produzir novos lantipétidos, cuja caracterização biosintética, estrutural e funcional é ainda desconhecida.

keywords

Lanthipeptides, RT-qPCR, transcriptomics, proteomics, sequencing, genome mining, *Sphingobacteriaceae*

abstract

Pedobacter lusitanus NL19 is a Gram-negative bacterium from the family *Sphingobacteriaceae*, which was isolated from a deactivated uranium mine in Portugal. This strain produces nonribosomal peptides (NRPs), called pedopeptins. The production of these peptides is repressed by high concentrations of peptone from casein (PC). In addition to biosynthetic gene clusters (BGCs) encoding NRPs, the NL19 strain genome also has BGCs of other secondary metabolites (SMs), including lanthipeptides (4 BGCs: *ped*8, *ped*14, *ped*15 and *ped*17). Lanthipeptides are ribosomally synthesized and post-translationally modified peptides (RiPPs), which exhibit a wide variety of biological activities, including antimicrobial and antiallodynic. Lanthipeptides are characterized by the presence of lanthionine (Lan) and methyllanthionine (MeLan) residues and are divided into four classes, defined by the enzymes that catalyze the reactions needed for the installation of these residues. This study focused on the NL19 strain and the main objectives were to determine the effect of high concentrations of PC: i) on the transcription of lanthipeptides and ii) on the proteome of the strain, with special interest in proteins involved in the biosynthesis of SMs. Due to the COVID-19 pandemics and confinement, a third objective was defined, which aimed to identify and analyze, *in silico*, lanthipeptide BGCs from the genomes of other genera of the family *Sphingobacteriaceae*. Objective i) involved the transcriptional analysis of BGC *ped*15 that encodes two precursor peptides and other biosynthetic proteins. It was necessary to sequence the upstream region of this cluster, through primer walking, which allowed the identification of six other structural peptide genes. The transcriptional analysis was performed by RT-qPCR and revealed that high concentrations of PC do not affect the expression of the BGC *ped*15, contrary to what was found for the pedopeptins BGC. In general, the RT-qPCR results validated the available RNA-seq results and showed that the transcriptional repression caused by high concentrations of PC is not transversal to the production of other SMs. Objective ii) involved the analysis of the proteome of the NL19 strain grown in high concentrations of PC (and its control) by nano LC-ESI-MS/MS and allowed the detection of the differential expression of some proteins related to the biosynthesis of SMs, including the pedopeptins nonribosomal peptide synthetases and a lanthipeptide precursor encoded in the *ped*8 BGC. Objective iii) involved the analysis of 446 BGCs of the family *Sphingobacteriaceae* with the antiSMASH and BiG-SCAPE tools. Class I and class III lanthipeptide BGCs were identified in the genera *Mucilaginibacter* and *Sphingobacterium*. Class III BGCs encode LanKC enzymes with slightly different lyase domains, which may indicate that these enzymes use a different mechanism for the installation of Lan/MeLan. This study contributes to the body of knowledge of the bacterial response to the manipulation of culture media, in particular in the production of SMs with biotechnological potential as lanthipeptides. In addition, this study identified the potential of bacterial genera already known, but still underexplored, for the production of new lanthipeptides, whose biosynthetic, structural and functional characterization is unknown.

Table of contents

List of Figures

List of Tables

List of abbreviations

ABC	Adenosine triphosphate binding cassette
AMS	ABC-transporter maturation and secretion
ATP	Adenosine triphosphate
BGC	Biosynthetic gene cluster
bp	Base pairs
cDNA	Complementary deoxyribonucleic acid
Ct	Threshold cycle
ddNTPs	Dideoxynucleotide triphosphates
dNTPs	Deoxynucleotides triphosphates
DEG	Differentially expressed gene
DEP	Differentially expressed protein
Dha	2,3-didehydroalanine
Dhb	2,3-didehydrobutyrine
HMMs	Hidden Markov Models
ITS	Intergenic transcribed spacers
LA	Luria-Bertani agar
Lan	Lanthionine
LB	Luria-Bertani broth
LC-ESI	Liquid chromatography electrospray ionization
MDR	Multidrug-resistant
MeLan	Methyllanthionine
MRSA	Methicillin-resistant *Staphylococcus aureus*
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
NGS	Next-generation sequencing
NRPs	Nonribosomal peptides
NRPS	Nonribosomal peptide synthetase
ORF	Open reading frame
OMPs	Outer membrane proteins
PC	Peptone from casein
PCR	Polymerase chain reaction
pSer	Phosphoserine

pThr	Phosphothreonine
PTMs	Posttranslational modifications
RiPPs	Ribosomally synthesized and post-translationally modified peptides
rRNA	Ribosomal ribonucleic acid
RT-qPCR	Reverse transcription quantitative polymerase chain reaction
SAGE	Serial analysis of gene expression
SMs	Secondary metabolites
TSA	Tryptic soy agar
TSB	Tryptic soy broth
VRE	Vancomycin resistant enterococci

Chapter I -Introduction

1.1 Secondary metabolites

Secondary metabolites (SMs), also called natural products, are not essential for the growth, development or reproduction of an organism [1–3]. Bacteria, fungi, plants, and other organisms produce them as a result of adaptation to the surrounding environment or as a defense mechanism against predators. Therefore, the SMs confer the producing organisms an evolutionary advantage [2,4]. SMs exhibit highly selective and specific biological activities, resulting in a variety of industrial and pharmaceutical applications [2]. These properties helped advancing the knowledge of biology and the development of medicine making them the largest contributors to drugs in the history of medicine [5]. For instance, acetylsalicylic acid, commonly known as aspirin, is perhaps the most famous and most used anti-inflammatory agent and derives from a natural compound, salicin, isolated from *Salix alba* [1].

The use of SMs has been known since 2900 BCE in Egyptian medicine. In 2600 BCE the use of approximately 1000 plant-derived substances was documented in Mesopotamia, some of which are still used nowadays to treat parasitic infections and inflammation [6]. Also, microorganisms have synthesized SMs with antibiotic capacity for billions of years [3]. More than 5000 antibiotics were identified among members of the order Actinomycetales, with the genus *Streptomyces* having a huge contribution for this prominent number. In the 1940s, the discovery of streptomycin and streptothricin, produced by *Streptomyces* genus, triggered large-scale screening to identify new antibiotics from this genus [3,5]. Also, the genus *Bacillus* [7] and *Myxobacteria* [8] have played a key role in antibiotic discovery contributing with important SMs such as gramicidin and ambruticin, respectively [7].

SMs can be classified based on their biosynthetic origin. Throughout the 20th and 21st centuries, research for natural products identified five prevalent groups: alkaloids, nonribosomal peptides (NRPs), polyketides, terpenoids and ribosomally synthesized and post-translationally modified peptides (RiPPs) [9]. The latter is a class of peptides with diversified structures characterized by extensive posttranslational modifications (PTMs) such as hetero- or macrocyclization, dehydration, acylation, glycosylation, halogenation, prenylation and epimerization [10,11]. Through *in silico* analysis of genomic data, the discovery and understanding of RiPPs is rapidly increasing [10]. Among the most well-studied subfamilies of RiPPs are the lanthipeptides characterized by lanthionine (Lan) and/or methyllanthionine (MeLan) thioether crosslinks [9].

1.2 Lanthipeptides

Lanthipeptides are RiPPs characterized by the presence of Lan or MeLan bridges [12]. The name of the peptide group is a short-hand nomenclature for lanthionine-containing peptides. A Lan consists of two alanine residues linked by a thioether group that connects their β-carbons; in the case of MeLan, it contains one additional methyl group [9]. What unifies this family of compounds is the mechanism by which the thioether crosslinks are formed. The PTMs reactions leading to the installation of Lan residues include a dehydration and a cyclization step. Firstly, Ser and Thr residues are dehydrated to 2,3-didehydroalanine (Dha) and 2,3-didehydrobutyrine (Dhb), respectively. Afterwards, Cys residues attack the dehydrated residues, via Michael type addition, generating the Lan and MeLan thioether cross-linked amino acids (**Figure 1**) [13].

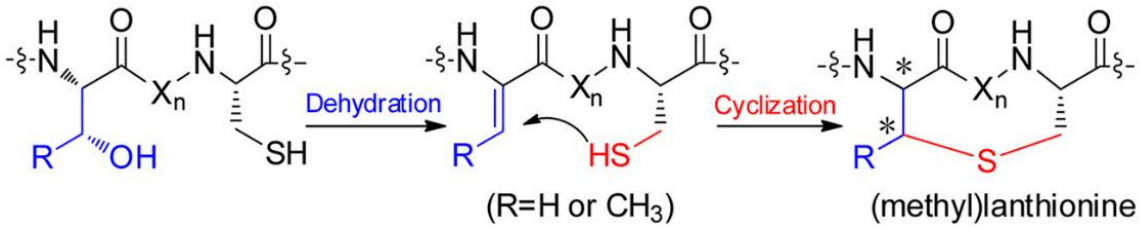

Figure 1: Post-translation modifications reactions in (methyl)lanthionine biosynthesis [14].

Depending on their constrained conformations, the resulting polycyclic peptides, display a wide variety of biological activities, from antimicrobial to antiallodynic [13]. Currently, lanthipeptides are divided into four classes depending on the biosynthetic enzymes responsible for installing Lan and MeLan thioether bridges (**Figure 2**) [9,13]. Regardless of the class or final conformation, the genes that encode the lanthipeptide machinery have been designated with the generic locus symbol *lan* with a more specific designation regarding each lanthipeptide (e.g., *nis* for nisin, *gdm* for gallidermin, *spa* for subtilin) [15]. Typically, the lanthipeptide BGCs comprise genes encoding the precursor peptide (*lanA*), the dehydration and cross-linkage enzymes (*lanB* and *lanC*, *lanM*, *lanKC* or *lanL*) and the protease/exporter (*lanP*/*lanT*), often clustered in the same operon. In addition, genes encoding enzymes that install other PTMs on some lanthipeptides can be identified in these gene clusters. For instance, *aplG*, encodes a glycosyltransferase responsible for the installation of a deoxyhexose onto a Trp residue in NAI-112, a class III lanthipeptide (see section **1.2.3**; **Figure 5**) [16]. Moreover, regulatory (*lanKR*) and immunity (*lanIFEG*) genes may be included in these BGCs, usually in different transcription units [17,18].

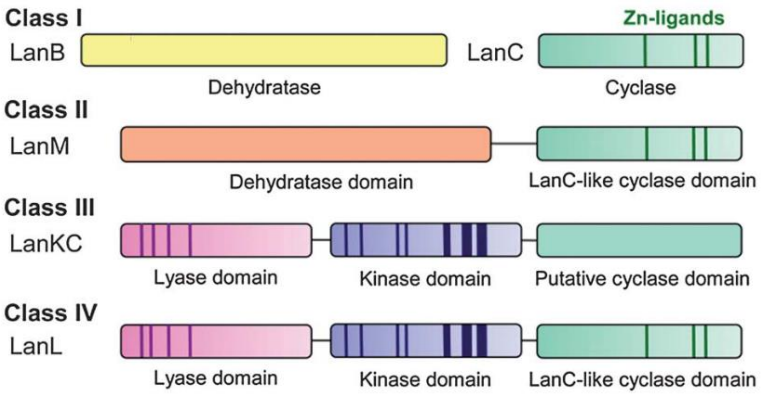

Figure 2: Representation of the four different classes of lanthionine-generating enzymes. The dark areas shown conserved regions important for catalytic activity [11].

1.2.1 Class I lanthipeptides

In the case of class I lanthipeptides, a dehydratase LanB is responsible for the dehydration step and the cyclization is catalyzed by a LanC cyclase (**Figure 2**). Comparative studies of class I lanthipeptide BGCs revealed that they are normally composed by, at least, 4 genes (*lanA, lanB, lanC* and *lanT*) [13]. Nisin was the first lanthipeptide ever reported and it is the model molecule of class I lanthipeptides. In 1928, Rogers discovered that nisin, which was produced by *Lactococcus lactis* could inhibit the growth of *Lactobacillus bulgaricus* [19]. The interest for this substance increased since it had an antibiotic capacity. Hence, new investigations were made in order to elucidate nisin's structure and biosynthesis (**Figure 3A**) and its BGC was found to be composed by 11 genes (**Figure 3B**). These include genes encoding: i) the precursor peptide (*nisA*), ii) three proteins involved in PTMs (*nisB*, *nisC* and *nisP*), iii) an ATP-binding cassette (ABC) transporter (*nisT*), iv) immunity proteins (*nisIFEG*) and v) the regulatory two-component system (*nisRK*) [20].

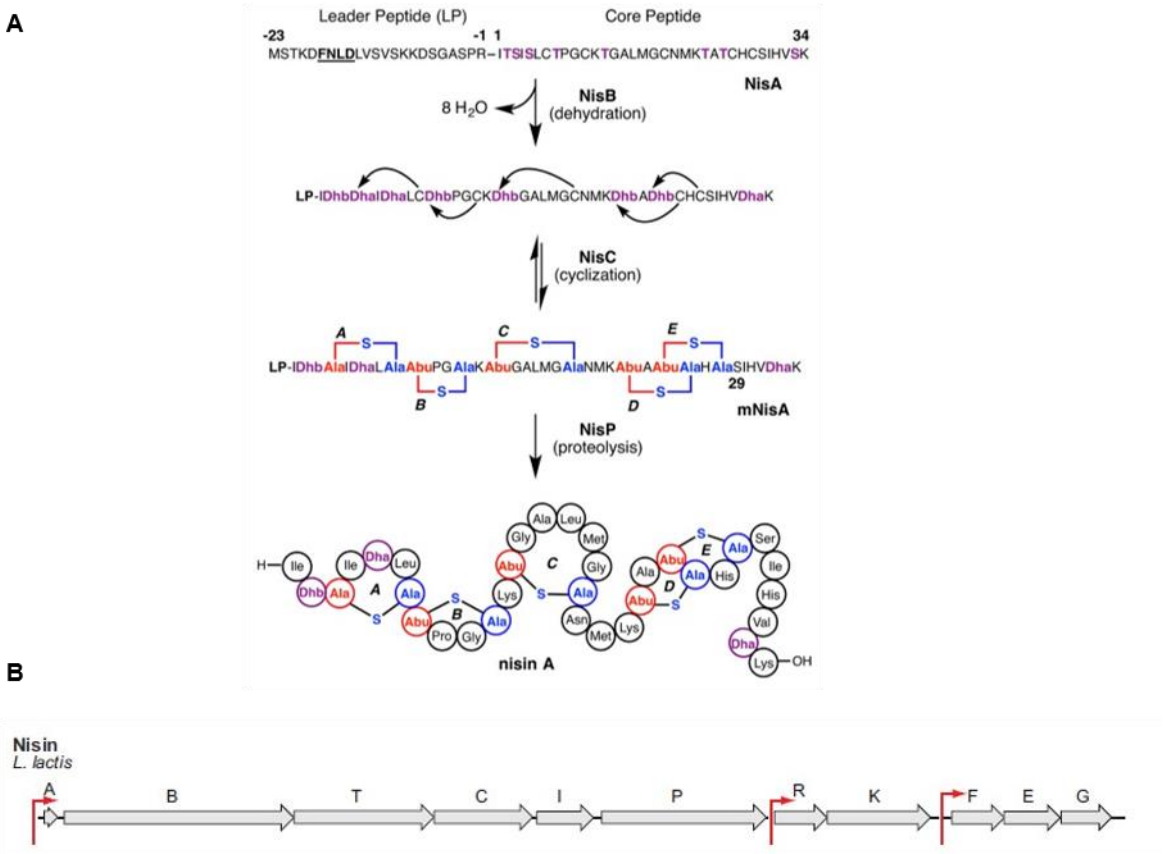

Figure 3: Schematic representation of nisin biosynthesis, including the necessary enzymes to install Lan residues (A) [11]. Representative BGC of nisin; transcriptional units are indicated by red arrows (B) [18].

The mechanisms involved in the dehydration step in class I lanthipeptides were elucidated for the nisin's lanthipeptide dehydratase NisB [21,22]. It was found that, firstly, LanB dehydratases require glutamate, presumably derived from glutamyl-tRNA^{Glu} [21], to form an ester linkage with the side chain hydroxyl groups of Ser/Thr within the precursor peptide. Subsequently, the peptide-glutamyl adducts are eliminated, giving rise to the dehydro amino acid [22]. Therefore, LanB proteins, such as NisB, depend on a glutamylation (N-terminal) and an elimination (C-terminal) domain to function properly [21,23]. This unique mechanism is relevant due to the tRNA specificity of LanBs. In other words, the nucleotide sequences present in the acceptor stem of tRNA^{Glu} are likely responsible for helping to achieve a suitable recognition and catalysis by LanB enzymes [24].

The thioether structure that characterizes the lanthionine rings, is generated by the intramolecular attack of a Cys thiolate onto the dehydro amino acid. In class I lanthipeptides, this reaction is accomplished by LanC. Stoichiometric amounts of zinc were found in two LanC enzymes (NisC and SpaC) [25], highlighting the role of this ion. Zinc was proposed to be critical in the activation of thiols for nucleophilic attack, functioning as a Lewis acid to lower the pKa and enhancing the reactivity of the thiolate [13]. Moreover, LanC enzymes feature a cyclase domain with

a conserved Cys–Cys–His zinc-binding motif also present in other class I, class II and class IV cyclases (**Figure 2**) [26]. Mutations of these residues abolish the cyclase activity [13].

1.2.2 Class II lanthipeptides

In class II, the dehydration and cyclization steps are carried out by a two-domain LanM synthetase (**Figure 2**). This synthetase is composed of an N-terminal dehydration domain, without sequence homology to LanB and a C-terminal cyclization LanC-like domain [27]. Owing to the lack of homology between LanB and LanM N-terminal, the dehydration mechanism in class I and II is fundamentally different. Studies proved that LanMs require ATP/Mg^{2+} to phosphorylate Ser/Thr and subsequently ADP/Mg^{2+} to eliminate the phosphate esters to yield Dha/Dhb [13,15].

The C-terminal cyclase domain of LanM shares ~25% sequence identity to LanC, including conservation of the zinc-binding residues [11]. The cyclization events catalyzed by LanMs are thought to resemble that of the LanCs, meaning that cyclization involves activation of a Cys thiol by coordination to the Zn^{2+} ion [13].

A LanM enzyme was first found during the biosynthesis of lacticin 481 (**Figure 4A**), whose gene cluster was sequenced in the mid-1990s (**Figure 4B**). The cluster includes a structural gene (*lctA*), a gene encoding the modification enzyme (*lctM*), a gene encoding a bifunctional transporter/protease enzyme ($lctT_p$) and genes involved in immunity (*lctFEG*) [11,13].

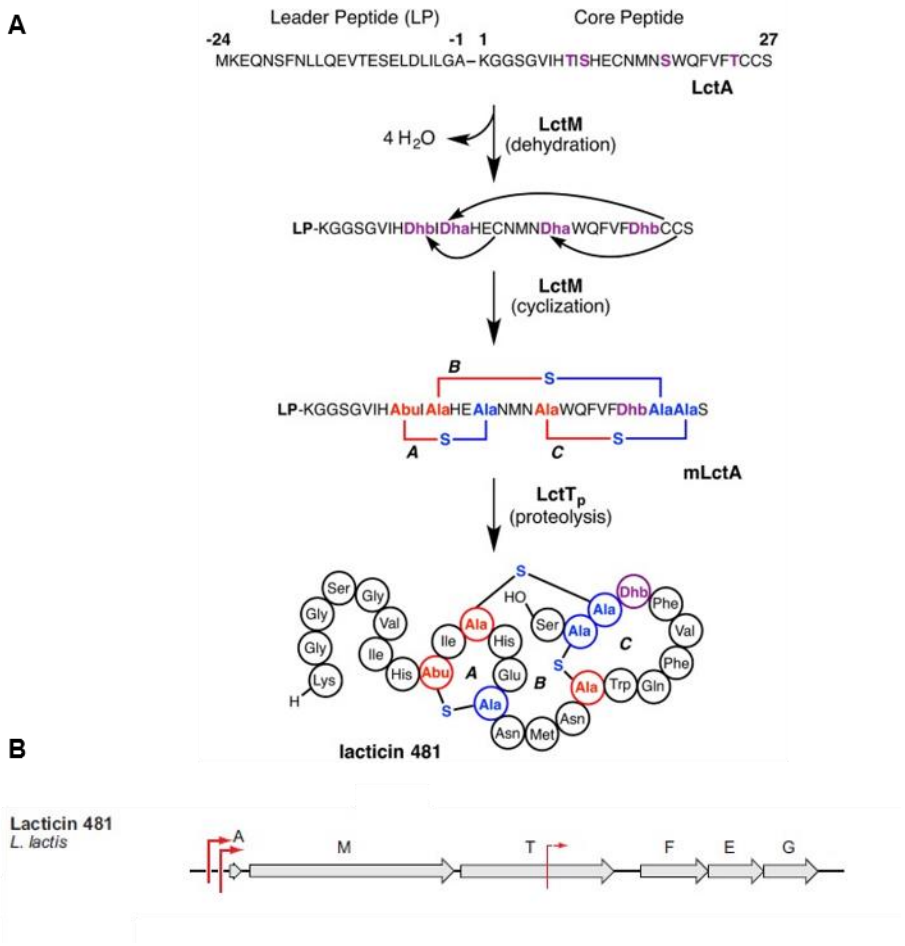

Figure 4: Schematic representation of lacticin 481 biosynthesis, including the necessary enzymes to install Lan residues (A) [13]. Representative BGC of lacticin 481; transcriptional units are indicated by red arrows (B) [18].

1.2.3 Class III lanthipeptides

Class III lanthipeptides are modified by a trifunctional synthetase called LanKC bearing an N-terminal phosphoserine(pSer)/phosphothreonine(pThr) lyase domain, a central Ser/Thr kinase domain, and a putative C-terminal cyclase domain (**Figure 2**) [9,11,13]. The dehydration process, via phosphorylation, involves two separate active sites. First, the phosphorylation of the precursor peptide LanA occurs in the central kinase domain, with the help of Mg^{2+} and ATP or in some cases GTP [11]. Then, the phosphate elimination in the N-terminal pSer/pThr lyase domain. These two domains share high homology with those of class IV (see section **1.2.4**) [26].

The cyclization of cysteine residues is performed by the C-terminal cyclase domain [13,28] that, despite having homology with other cyclase domains/enzymes, lacks the three zinc ligands [9]. Since the zinc ligands are missing in this class it remains unclear how thiol activation of Cys residues is achieved [26]. Interestingly, this class is the only characterized with labionin-containing

lanthipeptides. These entwined, bicyclic structures are formed when an enolate intermediate undergoes a second Michael addition with another Dha moiety [29].

An example of a class III lanthipeptide is NAI-112 (**Figure 5A**) [16,29,30]. It is produced by *Actinoplanes* sp. and its BGC is composed by seven genes (**Figure 5B**). Along with the structural gene (*aplA*), the other genes encode: i) the synthetase enzyme (*aplKC*), ii) an glycosyltransferase tailoring enzyme (*aplG*), iii) ABC transporters with putative immunity and secretion functions (*aplI*, *aplT1*, and *aplT2*) and iv) a protease (*aplP*) [16,30].

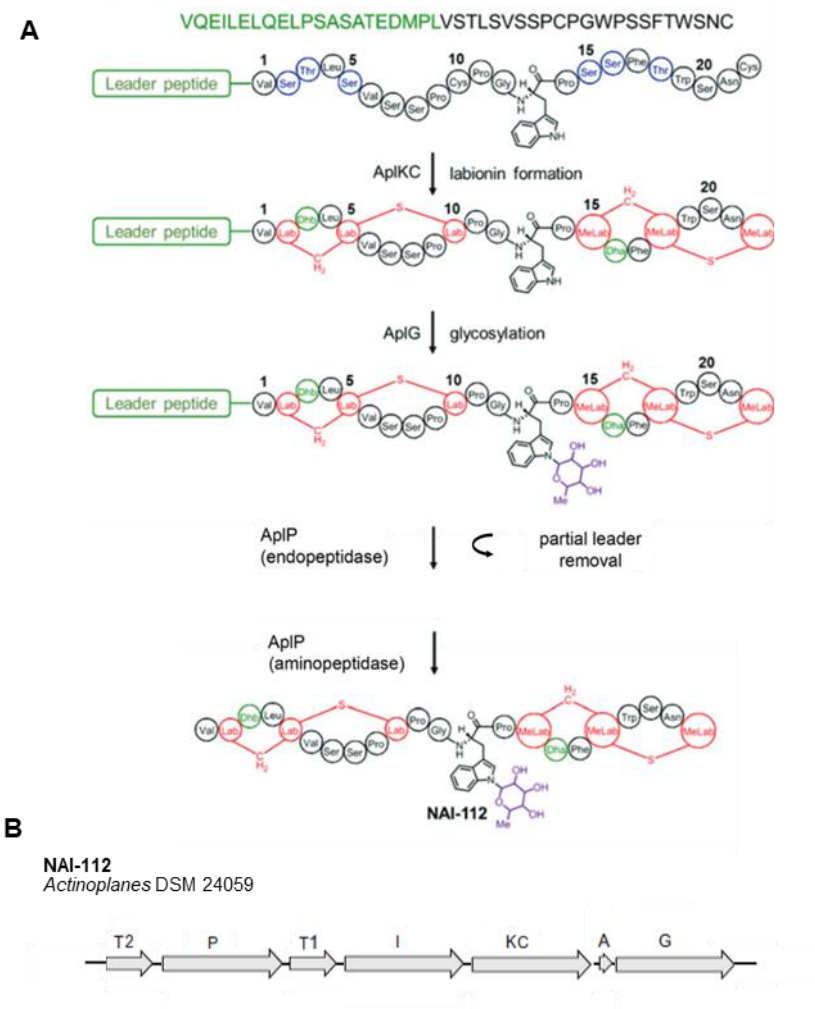

Figure 5: Schematic representation of NAI-112 biosynthesis, including the necessary enzymes to install Lan residues (A). Representative BGC of NAI-112 (B) [16].

1.2.4 Class IV lanthipeptides

As in class III, class IV lanthipeptides are formed through trifunctional enzymes. Class IV lanthipeptides are modified by a LanL enzyme (**Figure 2**). LanL contains a lyase N-terminal domain and a central kinase domain, both of which similar to LanKC and thus the catalytic activity is assumed to work in the same way [11]. Moreover, the kinase domains in class III/IV present conserved catalytic residues and structural features with certain structurally characterized human Ser/Thr kinases and with the mycobacterial protein kinases B (PknB) and G (PknG). Likewise, the lyase domains, in class III/IV, share many of the conserved catalytic residues with the type III effector proteins of the OspF family, which are pThr lyases produced by pathogens [26]. Additionally, LanL features a C-terminal cyclase with conversed zinc binding motif homologous to LanC and LanM [11].

The prototype for class IV lanthipeptides is venezuelin produced by *Streptomyces venezuelae* (**Figure 6A**). Its BGC, represented in **Figure 6B**, consists of the structural gene (*venA*), the synthetase gene (*venL*), and two components of a putative ABC transporter (*venT*) and (*venH*), encoding the ATP-binding and permease subunits, respectively [13,31].

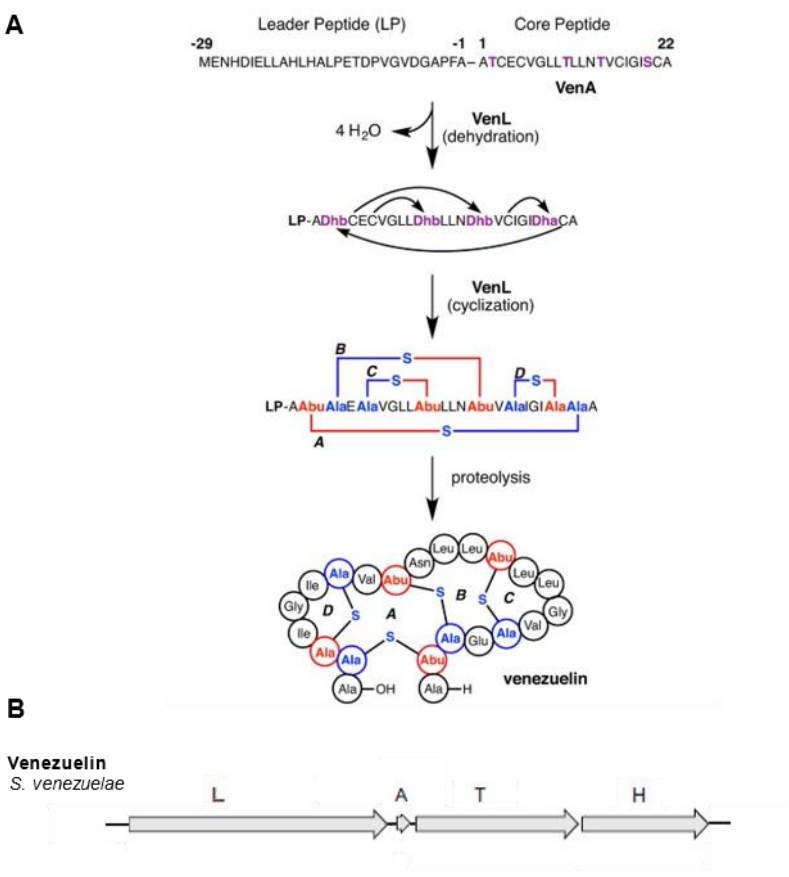

Figure 6: Schematic representation of venezuelin biosynthesis, including the necessary enzymes to install Lan residues (A) [13]. Representative BGC of venezuelin (B) [31].

1.2.5 Leader peptide removal and export

The starting point of lanthipeptide biosynthesis is the linear precursor peptide, LanA, that is structurally divided into a leader and a core peptide (**Figure 7**). The peptide is the substrate to the lanthipeptide modification enzymes that install the Lan and MeLan rings on the core peptide, as well as additional PTMs, forming the modified mLanA peptide (**Figure 7**). However, the complete biosynthesis of a lanthipeptide normally also involves its export, which is usually directed and accompanied by leader peptide proteolysis [9,11,13].

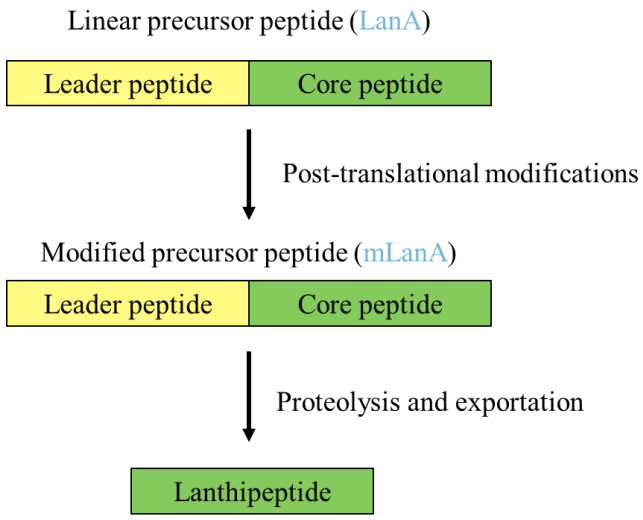

Figure 7: General biosynthetic pathway for lanthipeptides [13].

The leader peptide removal and transport through the membrane can be catalyzed either by independent or bifunctional enzymes. In the first case, a LanP is responsible for the removal of the leader peptide in the proteolytic step. It is a dedicated subtilisin-like serine protease, that may be membrane anchored or cytoplasmic [11]. The exportation is carried out by a LanT that has sequence similarities to integral membrane transporters. Analysis of its amino acid sequence showed two Walker-motif sequences involved in ATP-binding, demonstrating that LanT is a transmembrane ABC transporter [11]. This mechanism is mostly found in class I but also in some class II lanthipeptides and examples include epidermin, lactocin S and nisin (**Figure 3**) [13]. The bifunctional transporter/protease enzymes are named $LanT_p$ and are composed by: i) a N-terminal catalytic activity, responsible for cleaving the leader peptide at a conserved Gly motif (GG/GA/GS) [13] and ii) an ATP-binding and a transmembrane domain that make them members of the ABC-transporter maturation and secretion (AMS) protein family [13]. This mechanism was previously described as class II exclusive but was most recently found within class I lanthipeptide BGCs, especially in the

phyla Proteobacteria [13] and Bacteroidetes [32]. Some examples include pinensins, bovicin HJ50 and lacticin 481 (**Figure 4**) [13].

Class III and IV lanthipeptides still have the mechanism of leader peptide removal poorly elucidated since most clusters lack a potential LanP that can be responsible for this step. A study was conducted for the class III lanthipeptide NAI-112 since its BGC encodes the peptidase AplP (**Figure 5**). It was found that AplP exhibits both endopeptidase and aminopeptidase activities not only with the NAI-112 precursor peptide but also with other class III precursor peptides [26]. AplP homologs were found in class III and IV lanthipeptide producing strains and it is proposed that, at least some of these leader peptides can undergo a stepwise trimming by endopeptidase and/or aminopeptidase activity [26]. The transport of class III and IV lanthipeptides is also poorly known. Nonetheless, analysis of class III BGCs from phylum Actinobacteria and different *Bacillus* and *Staphylococcus* species (phylum Firmicutes) revealed the presence of putative ABC-type transporters [28,33].

1.2.6 The regulation of the production of lanthipeptides

Regulation of lanthipeptides biosynthesis can be coordinated by a two-component sensory system composed of a membrane-bound histidine kinase (LanK) and a transcriptional response regulator (LanR). The regulation process begins with an external environment signal triggering the autophosphorylation of a LanK histidine residue. A signal cascade is initiated and, subsequently, the phosphate group is transferred to the receiver domain of LanR that will mediate lanthipeptide transcriptional activation (**Figure 8**) [17,34]. After LanR binding to specific DNA sequences, RNA polymerase can interact properly and initiate transcription [35].

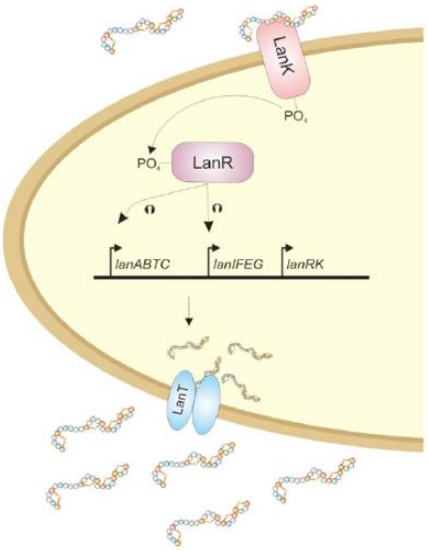

Figure 8: General overview of the regulatory mechanism for the biosynthesis of lanthipeptides [34].

For both nisin and subtilin, it was found that subinhibitory concentrations of the lanthipeptide itself can act as the molecule that triggers the regulatory system. Usually, this begins during mid-exponential growth of the cell and reach a peak of transcription at the transition from log to stationary phase [18]. In the regulation of nisin's production, the phosphoryl group from NisK is transferred to an Asp on NisR. Afterwards, NisR binds to promoters of the transcriptional units involved in nisin biosynthesis (*nisA*) as well as involved in self immunity (*nisF*). Nisin BGC also contains a separate *nisR* promoter for *nisRK* genes (**Figure 3B**) [15]. In subtilin, regulation depends on transcription of the *spaRK* operon (**Figure 9**), which in turn is regulated and dependent on the alternative sigma factor, σ^H, itself transcriptionally controlled by the transition state regulator AbrB [18].

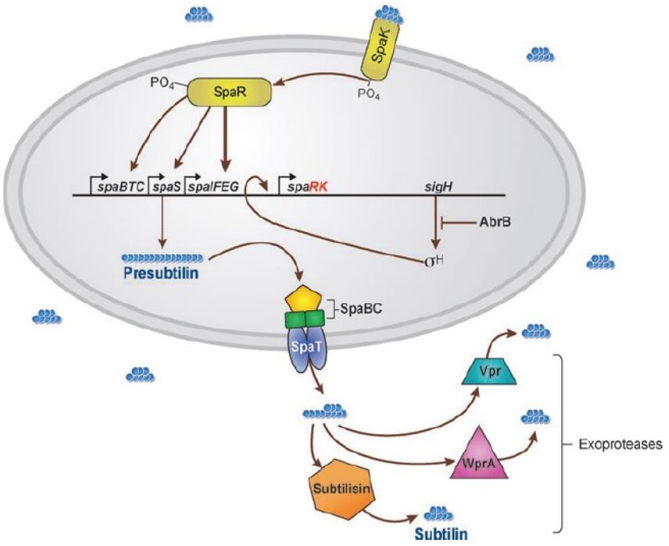

Figure 9: Subtilin regulation mechanism. The *spaS* gene encodes for presubtilin which is modified and transported by SpaB, SpaC, and SpaT. After export, the leader peptide is cleaved by exoproteases. Mature subtilin serves as ligand for the sensor kinase SpaK, which when activated phosphorylates SpaR to initiate transcription of *spaS*, biosynthetic and immunity operons. In turn, *spaRK* transcription is controlled by σ^H which is subject to AbrB regulation [18].

1.2.7 Self-immunity systems of lantibiotic producers

All the early lanthipeptides isolated displayed antibacterial activity. Historically, they were called lantibiotics as an abbreviation for "lanthionine containing antibiotic". Only when several lanthionine containing peptides without antibacterial activity had been described it became clear that lantibiotics were, in fact, a subgroup of a major family of lanthipeptides [13,31,36].

Usually, lantibiotics have bioactivity against strains closely related to their producer. Thus, producers should encode resistance/self-immunity mechanisms that protect them against the harmful effects of their own product. This immunity is normally provided by individual immunity proteins (LanI) and/or ABC transporters, composed by two or three subunits (LanFE(G)) [18]. For

example, nisin-producing *Lactococcus lactis* strains proved to have a mechanism that confers resistance to nisin based on NisFEG and NisI. Firstly, NisFEG functions by excreting the nisin molecules. Then, the lipoprotein NisI is orientated to the outside of the cytoplasmic membrane and is responsible for intercepting nisin to reduce its local concentration and/or preventing it from inserting into the membrane. *Bacillus subtilis* as a similar mechanism for self-immunity against subtilin. However, despite the structure similarities between the two lantibiotics, NisI is unable to protect *Lactococcus lactis* from subtilin [18,37].

1.2.8 The biotechnological interest of lanthipeptides

Lanthipeptides have "drug-like" properties, due to their thermal and protease stability, low toxicity risk, low tendency to generate resistance and immunogenicity [17]. The knowledge about lanthipeptides biosynthesis has accelerated tremendously in the recent years leading to the advent of strategies to enhance lanthipeptides application in many areas. However, certain limitations, such as, instability and/or insolubility at physiological pH and low production levels have prevented their widespread use [38]. Therefore, improvements in the biological and physicochemical properties still need to be deployed. This section highlights the biotechnological potential of lanthipeptides regarding the earlier and recent developments.

1.2.8.1 Current and potential applications of lanthipeptides

Currently, nisin is the only lantibiotic commercialized. It is used as a biological food additive (E234) since 1953 [17,39,40]. Besides nisin, several lantibiotics are in preclinical and clinical development (**Table 1**). Microbisporicin (NAI-107), a class I lanthipeptide produced by the actinomycetes *Microbispora corallina*, is the most active lantibiotic identified so far and has been tested for therapy of nosocomial infections, caused by multidrug-resistant (MDR) bacteria [17,41]. Mutacin 1140 showed great results *in vivo* against methicillin-resistant *Staphylococcus aureus* (MRSA), vancomycin resistant enterococci (VRE), *Clostridium difficile*, *Mycobacterium tuberculosis* and *Bacillus anthracis*. It is in preclinical development for the treatment of infections caused by Gram-positive bacteria [17]. This spectrum of bioactivity present in lantibiotics is highly relevant because, according to WHO, infections caused by MDR bacteria, are a threat to public health and, therefore, the search for novel products with antibiotic properties is an urgent necessity [42]. Also, for the treatment of *C. difficile* infections, the semisynthetic carboxy-amide derivative of the globular class II lantibiotic deoxy-actagardine B (also known as NVB302) is undergoing a

phase I clinical trial [17,43]. Class III and IV lanthipeptides have not showed any antibacterial activity yet. However, other important properties such as antiviral (anti-HIV and anti-HSV) and antiallodynic activity were exhibited by the class III labyrinthopetins A1 and A2, respectively (**Table 1**) [36,44,45].

1.2.8.2 Applications of the biosynthetic machinery of lanthipeptides

The use of lanthipeptides biosynthetic machinery is a promising technology to overcome inherent disadvantageous properties of existing therapeutic peptide drugs. Through bioengineering approaches several non-lanthipeptides have already been produced with improvements in protection against peptidase degradation and in rigidity of the peptide's receptor-binding-region, therefore enabling optimal interaction between the stabilized peptide and its receptor. This is possible due to the promiscuity of lanthipeptides biosynthetic machinery and their ability to display activity *in vitro* allowing to yield thioether-stabilized therapeutic peptides [46]. Thus, new compounds could be designed based on the introduction of thioether bridges in specific locations into existing pharmaceuticals. For example, a site-directed mutagenesis introduced a dehydratable Ser and a thiol-containing Cys in angiotensin [Ang-(1–7)]. Afterwards, one thioether ring in Ang-(1–7) was added to the heptapeptide using *L. lactis* as a host organism containing the nisin modification enzymes (NisB, NisC, NisT and a NisA leader peptide fused to the peptide of interest). The generated cyclic peptide [cAng-(1–7); **Figure 10**] showed enhanced pharmacodynamic properties via *in vitro* and *in vivo* assays [47]. Further exploration of cAng-(1–7) potential proven its efficiency in neonatal chronic lung disease (CLD) [48] (**Table 1**). In summary, the improvement of therapeutic peptide properties can create exciting new applications around known disease targets.

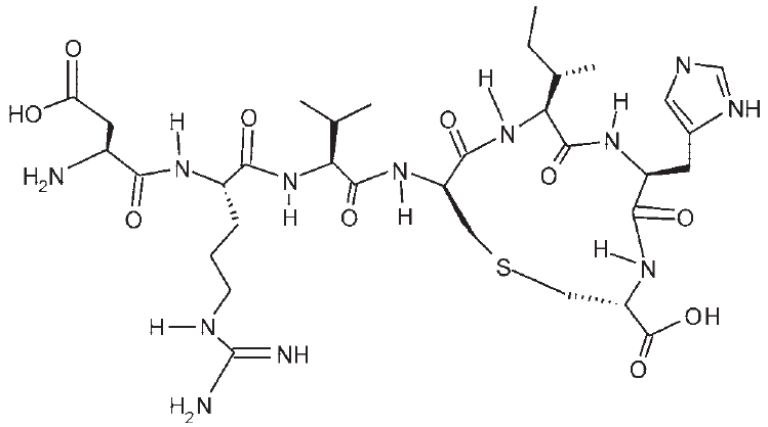

Figure 10: Chemical structure of cAng-(1–7) [47].

Table 1: Lanthipeptides with biotechnological interest. Overview of the development state, potential therapeutic applications, and relevant targets. Lanthipeptides are divided according to their classes.

Lanthipeptide	Producer/Origin	Development State	Applications	Target
		Class I lanthipeptides		
Nisin	*Lactococcus lactis*	Commercialized[17,39]	Biological food additive, veterinary medicine (prevention of bovine mastitis)	Gram-positive bacteria including, *Listeria monocytogenes*, MRSA
Microbisporicin	*Microbispora corallina*	*In vivo* experiments[17,41]	Systemic treatment of serious nosocomial infections caused by MDR pathogens	Gram-positive pathogens (including MRSA and VRE) Gram-negative bacteria (*Moraxella catarrhalis*, *Neisseria spp*., and *Haemophilus influenzae*)
Mutacin 1140	*Streptococcus mutans*	Preclinical development [17]	Treat dental caries and streptococcal throat infections	MRSA, VRE, *C. difficile*, *M. tuberculosis* and *B. anthracis*
OG716	Analog of mutacin 1140	Preclinical development[45,49]	*C. difficile* infection in enteritis	*C. difficile*
Pinensins	*Chitinophaga pinensis* DSM 28390	*In vitro* experiments[32]	Antifungal/yeast	Filamentous fungi and yeasts
		Class II lanthipeptides		
NVB302	Semisynthetic derivative of deoxy-actagardine B	Phase I clinical trial[43,50]	Treatment of *C. difficile* infections	*C. difficile*

Lanthipeptide	Producer/Origin	Development State	Applications	Target
Salivaricin A2 and B	*Streptococcus salivarius* K12	*In vivo* experiments[38,45]	Treatment of dental caries, and infection of the oral cavity	*S. pyogenes*
Mersacidin	*Bacillus amyloliquefaciens* HIL Y-85	*In vivo* experiments[38,51]	Treatment of staphylococcal (including MRSA) and enterococcal infections	Gram-positive bacteria including MRSA VRE, *C. difficile*, *S. pyogenes* and *B. subtilis*
Duramycin	*Streptomyces cinnamoneus*	Phase II clinical trial[17,38,45]	Treatment of cystic fibrosis	Stimulates chloride secretion of bronchial epithelia
		Class III lanthipeptides		
Labyrinthopeptins A1	*Actinomadura namibiensis* DSM 6313	*In vitro* experiments[44,45]	Antiviral activity against HIV and HSV	HIV and HSV
Labyrinthopeptins A2		*In vivo* experiments[36,45]	Treatment of neuropathic pain, allodynia	Nerve injury model of neuropathic pain
		Thioether-cyclized peptides		
MOR107	Angiotensin AT2 receptor agonist	Phase I clinical trial [52]	Diabetic nephropathy	Angiotensin AT2 receptor
cAng-(1–7)	Ang-(1–7)	*In vivo* experiments [48]	Neonatal CLD	MAS oncogene receptor
Lanthionine enkephalin analogues	Enkephalin	*In vivo* experiments [53]	Modulate antinociceptive responses	δ-opioid receptor

1.3 The genus *Pedobacter*

The genus *Pedobacter* belongs to the family *Sphingobacteriaceae* within the phylum Bacteroidetes. This genus, proposed by Steyn *et al*. [54], includes obligately aerobic, Gram-stain-negative rods able to use heparin as exclusive carbon and nitrogen source. The name *Pedobacter* was suggested since most species were isolated from soil. Notwithstanding, they can also be recovered from water, fish, sludge among others [55,56]. *Pedobacter heparinus* is the type species of this rapidly growing genus. Presently, 87 species have been described according to the *List of Prokaryotic names with Standing in Nomenclature* (accessed 20/10/2020).

1.3.1 The *Pedobacter lusitanus* NL19 strain

Pedobacter lusitanus (lu.si.ta´nus. L. masc. adj. *lusitanus* from Lusitania, the present-day Portugal) NL19 was isolated from a sludge collected in a deactivated mine at Quinta do Bispo, in the Viseu District (Portugal) [55–57]. Based on the sequence analysis of its 16S rRNA gene, this strain was included in the genus *Pedobacter* and is closely related to the strains *Pedobacter himalayensis* MTCC 6384^T, *Pedobacter cryoconitis* DSM 14825^T, *Pedobacter westerhofensis* DSM 19036^T and *Pedobacter hartonius* DSM 19033^T [55]. According to intergenic transcribed spacers (ITS) phylogenetic studies, strain NL19 was distinguished from these four species since it formed a unique branch.

NL19 strain cells are rod-shaped without motility and catalase and oxidase positive. On R2A and minimal medium, colonies are pale with shiny color whereas in NA, TSA and MacConkey agar medium the color of the colonies is light yellow. The optimal growth of NL19 strain in TSB ranges between 18-30 ºC. Nevertheless, the growth can occur between 4-30 ºC and at 26 ºC, the optimal pH for growth is from 6-8 [55].

As aforementioned, NL19 strain was isolated from a sludge. This extreme environment is rich in metals and radionuclides of uranium series but poor in nutrients. Although still poorly explored, the biotechnology potential of this strain is outstanding. The genome mining identified 17 BGCs encoding the production of SMs such as NRPs, siderophores, polyketides and lanthipeptides [57]. Moreover, NL19 strain exhibits potent antimicrobial activity *in vitro* against Gram-positive and Gram-negative bacteria relevant to the food, veterinary, clinical and aquaculture industry, including *Staphylococcus aureus* ATCC 29213, *Enterococcus faecalis* ATCC 29212, *Haemophilus influenzae* 121642 (clinical isolate), *Klebsiella pneumoniae* ATCC 700603 (an extended spectrum beta-lactamase producer) and *Salmonella enteritidis* ATCC 13076 [55]. Interestingly, this activity was identified in tryptic soy agar (TSA), but was absent in tryptic soy broth (TSB). This is congruent

with the knowledge that SMs are often produced in higher concentrations and quicker in agar, presumably due to the restricted access to nutrients and limited diffusion rates in such media [58]. The repression of antimicrobial production in TSB was attributed to the high concentration of peptone from casein (PC), since the production of antibacterial activity increase when the concentration of the nitrogen source PC was reduced [59]. Moreover, the bioactivity of NL19 was proved to be due to pedopeptins, a class of NRPs with a broad-spectrum activity [59].

1.3.1.1 The lanthipeptides of *Pedobacter lusitanus* NL19

Heretofore, the characterization of lanthipeptides in Gram-negative bacteria is scarce. Pinensins and prochlorosins belonging to class I and class II, respectively, are the lanthipeptides produced by Gram-negative bacteria better characterized so far.

Prochlorosins are produced by Cyanobacteria (genera *Prochlorococcus* and *Synechococcus*). The biosynthesis of the prochlorosins from these organisms represent a unique model in lanthipeptides production. Remarkably, their promiscuous lanthionine synthetase, ProcM, transforms up to 29 different precursor peptides (ProcAs) with highly variable sequences in the core region but highly conserved leader sequences, generating a library of lanthipeptides with diverse ring topologies (**Figure 11**) [60,61]. So far, no bioactivity has yet been assigned to prochlorosins [61].

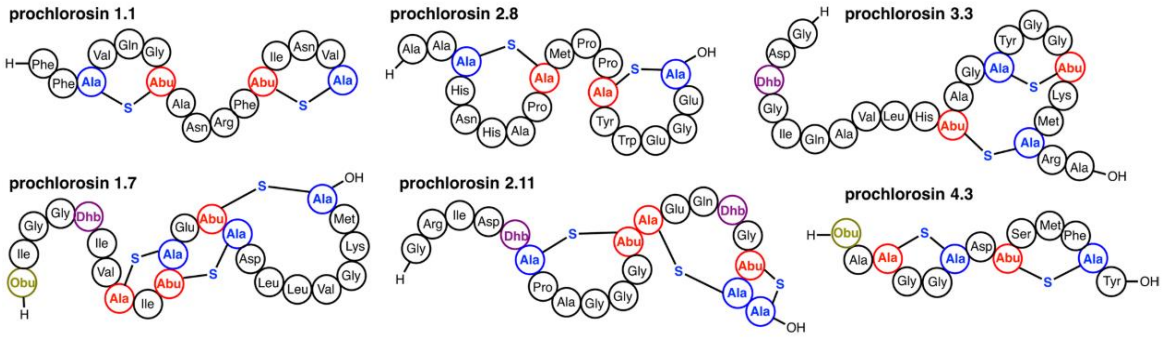

Figure 11: Representation of six structurally characterized prochlorosins that demonstrate the diverse ring topologies installed by a single ProcM [13].

Pinensins are produced by *Chitinophaga pinensis* (phylum Bacteroidetes) and have bioactivity against several filamentous fungi and yeast, thus being the first antifungal lanthipeptide described (**Figure 13B**) [32].

Recently, a search for *lanB* genes in the genomes of the phylum Bacteroidetes, revealed that the genus *Pedobacter* displayed the highest number of *lanB* genes per genome: 51 *lanB* genes

in 21 genomes [62]. In *P. lusitanus* NL19 genome, 7 *lanB* genes were identified, highlighting the potential of this strain for lanthipeptide production (**Figure 12**). Out of the 7 *lanB* genes identified, 4 of them (*pedB8*, *pedB14*, *pedB15* and *pedB17* in **Figure 12**) are embedded in BGCs that encode the other two genes fundamental for the biosynthesis of class I lanthipeptides: at least one gene encoding a precursor peptide (*pedA*), and a gene coding a cyclase enzyme (*pedC*). Moreover, a gene encoding a $LanT_p$ (*pedT* in **Figure 12**) - composed of an ATP-binding cassette, a transmembrane domain and a peptidase domain - was also identified. Previously, these transporters were only associated with the biosynthetic machinery of class II lanthipeptides. Nevertheless, *lanT*$_p$ have been recently encountered in gene clusters of other class I lanthipeptides, as abovementioned [13,32].

Other genes found in these BGCs encode proteins involved in: i) signal transduction, ii) transcriptional regulation and iii) transport. Such diversity in this "hybrid class I/class II" BGC suggests that the lanthipeptides produced will have unique biochemistries and structures, making them potentially suitable for biotechnological applications.

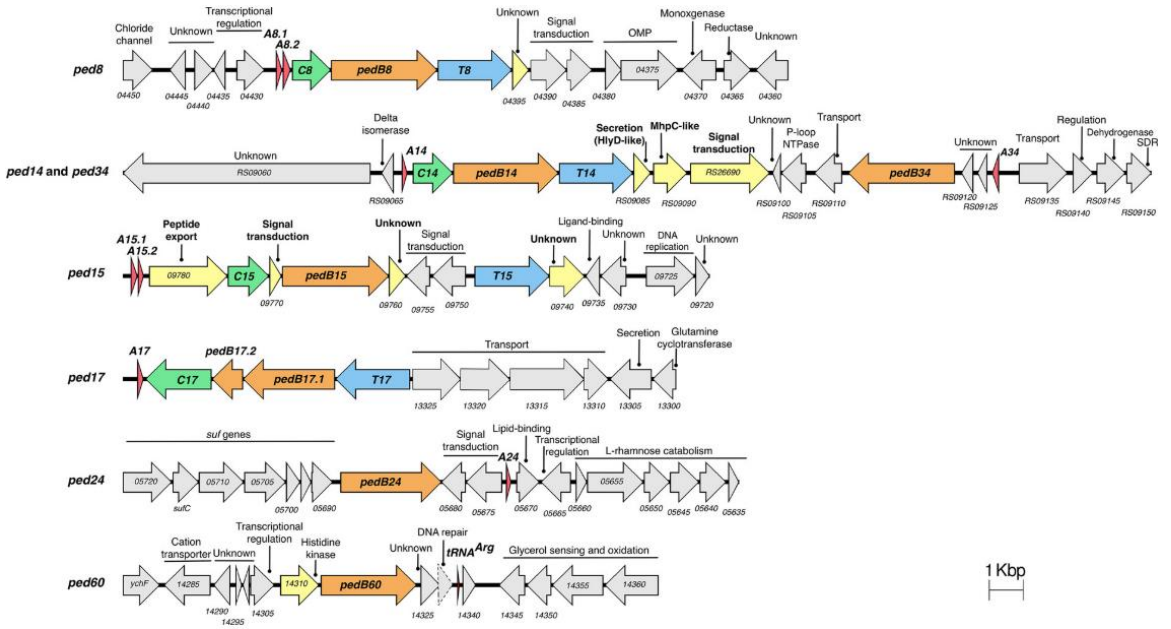

Figure 12: The *lanB* genes encountered in *P. lusitanus* NL19 embedded in their corresponding BGCs. The *lanB* are coloured in orange, structural *lanA* in red, *lanC* in green and *lanT* in blue. In yellow are highlighted the genes predicted as part of the same transcription unit of *lanA*, *lanC*, *lanB* or *lanT* [62].

All *pedA* genes found to be associated with a *pedB* gene encode peptides with a GG-motif in their leader peptide sequence (**Figure 13A**), suggesting the local at which the peptidase domain of $LanT_p$ removes the leader peptide. In the *ped8* and *ped15* BGCs, two putative *pedA* genes, with conserved residues in the leader peptides, were found. Interestingly, each of these BGCs features

only one *pedB* suggesting that the cognate dehydratase recognizes both leader peptides (**Figure 13A**). Only PedA8.2 does not contain Cys residues in the core peptide. In *ped14* cluster, only one precursor peptide was found, which includes three Cys residues in the core peptide. Considering the *ped17* cluster, comparative genomics analysis revealed that it is homologous to the pinensin cluster (**Figure 13B**). Both clusters have a precursor peptide gene and two operons: one containing a LanC, a $LanT_p$ and a split LanB and another encoding an ABC transporter (composed by two permeases and one ATPase) and an outer membrane protein. Moreover, the leader and core peptide of PedA17 are very similar with that of PinA [62].

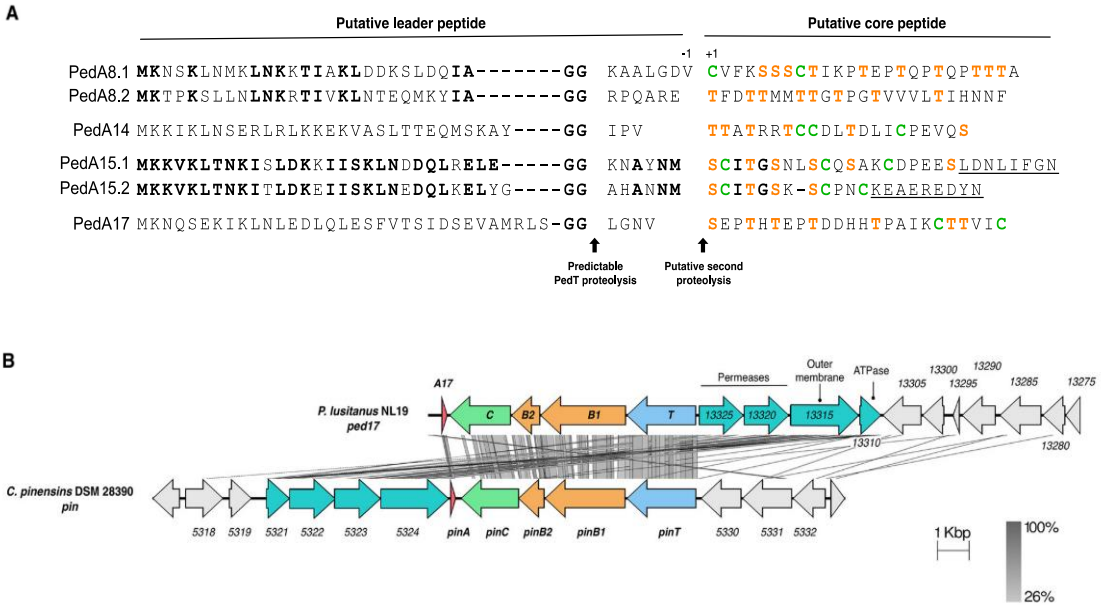

Figure 13: Amino acid alignment of the putative precursor peptides of NL19 (encoded by *ped8*, *ped14*, *ped15* and *ped17* BGCs). The residues highlighted in bold represent the conserved amino acids of peptides in the same cluster. The amino acids yielding the dehydrated residues are coloured in orange and cysteine residues are coloured in green. Also, a carboxypeptidase domain encountered in a putative protein encoded in *ped15* BGC is presumably responsible to remove the underlined residues [62] (A). Comparison of *ped17* and pinensins (*pin*) clusters (B).

1.4 Multi-omics approach

1.4.1 Genomics

The development of genome sequencing technologies and bioinformatic pipelines has empowered a rapid mining for BGCs encoding SMs in diverse microorganisms. Gilbert and Maxam were the first to successfully sequence DNA. Using a chemical degradation DNA sequencing technique, they reported, in 1973, the sequence of 24 base pairs (bp) of the *lac* operon [63]. Although being a remarkable accomplishment the method was laborious, and time consuming. Then, in 1977, Frederick Sanger, developed a more efficient and faster method called Sanger sequencing or chain termination method [64,65]. His pioneering work was a landmark in the DNA sequencing field, leading him to receive the Nobel Prize in Chemistry in 1980. This method is still widely used today, and the key principle of the experiment is the inclusion of dideoxynucleotide triphosphates (ddNTPs). The ddNTPs are identical to deoxynucleotides (dNTPs) with the difference being that ddNTPs lack the 3' hydroxyl group needed to form the phosphodiester bond between two nucleotides. In Sanger sequencing a reaction, composed by a mixture of dNTPs and ddNTPs, is catalyzed by DNA polymerase that adds dNTPs to the growing chain. However, when ddNTPs are inserted the synthesis is terminated, due to the lack of 3' hydroxyl group. Additionally, ddNTPs are labelled by a fluorophore molecule so the resulting DNA fragment will emit a signal corresponding to the last incorporated nucleotide (Adenine, Thymine, Cytosine, or Guanine). After an adequate number of cycles of amplification, the number of nucleotides in the template will correspond to the number of the resulting DNA fragments, that can be discriminated from each other by a single nucleotide. Electrophoresis, in a polyacrylamide gel or capillary tube gel, is used to arrange in increasing size order the resulting products. The sequence is read from the bottom of the gel, being that at each base position ddNTPs are identified through their attached fluorophore [66].

Nowadays, new technologies for genome sequencing have emerged. Known as next-generation sequencing (NGS), these methods allow low-cost and high-performance sequencing therefore, driving progress in genomics [66,67]. The first NGS sequencers created use a sequencing-by-synthesis concept to perform the DNA amplification. For example, Illumina is based on bridge PCR and the Roche 454 and Ion Torrent platforms use emulsion PCR [66]. Currently, aided by advances in the field of nanotechnology, new developed sequencers, such as Pacific Biosciences and Oxford Nanopore, are being used for real-time sequencing of a single DNA molecule [65–67].

1.4.2 Transcriptomics

The development of transcriptomic technologies has accelerated our understanding of biological processes in bacteria and has delivered many advances in the assignment of gene function. For example, the discovery that pedopeptins from NL19 were encoded in a nonribosomal peptide synthetase (NRPS) cluster was only possible due to transcriptomic analysis [59]. Besides, the increased access to genome sequencing and the observation that even well-studied bacteria still have an unexplored biosynthetic potential lead to a rise of interest in SMs.

Transcriptome is the sum of all RNA transcripts in an organism. Transcriptomic analysis is a key tool to better understand how gene expression changes in different cells under different conditions [68]. In 1979, the first attempts in studying messenger RNA (mRNA), through complementary DNA (cDNA) from silkmoth, were described [69]. However, it was in 1995, with the development of serial analysis of gene expression (SAGE), that transcriptomics reached a milestone. This method enables a rapid detailed analysis of many transcripts [68,70].

Another important highlight in transcriptomics was the development of reverse transcription quantitative PCR (RT-qPCR). RT-qPCR is a technique that allows reliable detection and quantification of gene expression, since the transcripts measurement can be performed against defined standards for the gene of interest as well as control genes [68]. Briefly, RT-qPCR uses detection probes, such as double-stranded DNA-intercalating agents (e.g., SYBR Green), hydrolysis probes (e.g., TaqMan probes), dual hybridization probes, molecular beacons or scorpion probe to quantify the expression levels of a specific gene by PCR. Regardless of the probe, a fluorescent signal is emitted when the probe hybridizes with the PCR product. This fluorescence can be detected in real time by a specific thermocycler. When the reporter fluorescence is greater than the background threshold level, we obtain the threshold cycle (Ct) level. The Ct is a basic principle of RT-qPCR since it depends on the initial amount of DNA. Hence, higher the number of copies of a gene in a sample, lower is the Ct values and vice-versa [71,72]. RT-qPCR has a higher degree in sensitivity, reproducibility, speed, throughput and potential automation compared to conventional quantification methods such as northern-blot analysis, for example [73].

Scientists have sought to develop high-throughput and accurate expression profiling techniques to better understand transcriptomics and provide insights to complex regulatory networks. Currently, microarrays and RNA-seq are the predominant methods in transcriptomics [68]. The former, measures the abundance of transcripts, fluorescently labelled, by hybridization to an array of complementary probes. The intensity of the fluorescence is directly related to the transcript abundance. On the other hand, RNA-seq, uses high-throughput sequencing of transcript cDNAs to quantify the amount of transcripts [74]. The NGS platforms used in RNA-seq are equal to the ones used for genomic data. Consequently, improvements in DNA sequencing technologies have

allowed direct sequencing of RNA using nanopore technologies [68,75]. Also, if validation of RNA-seq data is required, RT-qPCR might be utilized since it is an independent technique statistically assessable [76].

1.4.3 Proteomics

Proteomics is the large-scale study of proteins including identification, modification and quantification. Within a microbial cell the genomic information translates into more than 2000 proteins, several of which could be studied using proteomics. For genomes containing less than 1000 genes, it is estimated that 50% of predicted proteins could be identified through the genome [77]. Proteomics provides a complementary approach to genomics technology once it investigates biological phenomena at a protein level [78].

A wide range of technology is used to provide accurate and fast determination of the proteome. Due to biological and technical constrains, this task is often difficult. Since the proteome is the set of proteins expressed by a genome at a certain time, it is highly dynamic and can change with stress conditions, growth phase among others, making its analysis a complex task. Moreover, at the biological level, the existence of a high protein concentration range within a biological sample constitutes a technical problem. To overcome this, fractionation and enrichment of less abundant proteins is needed.

The most versatile and comprehensive tool in large-scale proteomics is mass spectrometry (MS). This technique uses mass analysis for protein characterization [77,79–81]. Briefly, MS is an analytical technique that, firstly, involve the ionization of the analyte in an ion source. The resulting gas phase ions then travel through the mass analyzer and are separated according to their mass-to-charge ratio (m/z). Lastly, the number of ions at each m/z value is recorded in a detector generating a mass spectrum by a computer system [80]. Throughout this process the analyzer and the detector are always under vacuum so that the background noise and fragmentation pattern of the molecules are not disturbed [82]. Besides, tandem mass spectrometry (MS/MS) is key for proteomic analysis, whereby mass analysis is carried out on intact molecular ions (full-scan MS) or on fragmented precursor ions (MS^n scans) [81]. A widely used method in MS is nano liquid chromatography electrospray ionization MS/MS (nano-LC-ESI-MS/MS) [83]. The sample fractionation and preconcentration, performed by nano LC, are important to increase protein identification rates and sequence coverage. In addition, the ESI allows proteins to be analysed by MS. In this method, a liquid containing the analyte flows through a metal capillary to which a high voltage is applied to produce ions without extensive degradation of the sample [81,83].

Protein identification can be achieved through analysis of whole-protein (‘top-down’ proteomics) or enzymatically produced peptides (‘bottom-up’ proteomics) [80,81]. Top-down

approach allows high-resolution mass measurement of intact protein ions and their fragments inside the mass spectrometer without prior digestion. This technique is better for characterization of PTMs and for higher sequence coverage of target proteins. On the other hand, bottom-up method is based on proteolytic digestion of proteins into peptides prior to mass analysis. This approach is preferred when addressing high-complexity samples. The resulting peptides masses and sequences are then compared against the predicted, *in silico*-generated fragmentation patterns of the peptides under investigation [80,81].

Summarizing, the integration of 'omics' approaches (e.g., genomics, transcriptomics and proteomics) is a powerful strategy that largely contributes to the renaissance of the discovery of new BGCs encoding SMs. Thereby, tremendous biosynthetic potential can be explored leading to a better elucidation of the complex mechanisms that control SMs regulation and production [84].

Objectives and structure

As aforementioned, *P. lusitanus* NL19 is a Gram-negative bacterium that belongs to the family *Sphingobacteriaceae*. It was isolated from an extreme environment and revealed the unprecedented potential to produce a great source of novel SMs possibly adequate to several applications. Among them are the lanthipeptides and the NRPs pedopeptins. Previous studies have shown that high concentration of PC can inhibit the production of pedopeptins, which are antimicrobial peptides that inhibit some antibiotic resistant bacteria listed by WHO as priorities to antibiotic development. Insights on the mechanisms underlying the production and the structure of these and other SMs produced by *P. lusitanus* NL19, such as lanthipeptides, need further investigation. As such, the main objectives of this dissertation were:

i) Evaluate the impact of high concentrations of PC in the transcription of NL19 lanthipeptides, using the BGC *ped*15 as case study;
ii) Following a multi-omics approach, evaluate the impact of high concentrations of PC in the proteome of NL19, especially proteins involved in the biosynthesis of SMs;
iii) Due to COVID-19 pandemics and confinement, a third objective was defined, which is unrelated with the influence of PC on the production of SMs: to identify and analyze lanthipeptide BGCs from the genomes of other genera of the family *Sphingobacteriaceae*.

The accomplishment of each of the abovementioned objectives will be presented in four chapters. Each of them includes a brief introduction to the subject, the results obtained accompanied by a discussion and the experimental procedures. To achieve objective i), it was necessary to sequence the upstream region of BGC *ped*15, since it was found in the beginning of a contig. This is described in chapter II and allowed to have access to the complete *ped*15 cluster, including the promoter region as well as all the *pedA*15 structural genes. The transcriptional analysis of cluster *ped*15 by RT-qPCR and the proteome analysis of NL19 grown in broth containing either high or lower concentrations of PC are presented in chapter III and chapter IV, respectively. Chapter V describes the analysis of lanthipeptide BGCs identified in other genera from the family *Sphingobacteriaceae*.

Chapter II – Sequencing of the upstream region of *ped15* cluster

2.1 Introduction

The genome of NL19 was previously sequenced using the platform Ion PGM system and then assembled establishing 201 contigs [57]. In total, 17 BGCs were identified within these contigs using antiSMASH. This bioinformatic tool detects Hidden Markov Models (HMMs) of protein motifs for key biosynthetic enzymes in the analyzed genomes [85]. One of the clusters predicted to be involved in the production of lanthipeptides was the *ped15* cluster (**Figure 14**). The analysis of this cluster revealed the presence of genes encoding characteristic enzymes of class I (PedB, PedC) and class II ($PedT_p$) lanthipeptides (**Figure 14**). Moreover, two genes encoding precursor peptides (PedA15.1 and PedA15.2) were identified in the beginning of this cluster (**Figure 14**). These peptides are mainly different in the amino acids sequence of their C-terminus (**Figure 13A**). Other genes found in *ped15* BGC encode: i) a histidine kinase (TH53_09750), ii) a DNA-binding response regulator (TH53_09755) and iii) an outer membrane protein with an N-terminal carboxypeptidase domain (TH53_09780). Some genes of unknow function were also identified (**Figure 14**) [62].

Since NL19 strain was recently described, its genome is still presented as draft genome rather than a finished genome. Hence, the genome profile has segments of contiguous base pairs - designated contigs - interspersed with gaps for which the sequence is unknown. In the case of *ped15*, the cluster is located at the beginning of the contig36 [57], and as such, the genetic information of its upstream region is still undefined (**Figure 14**). Accordingly, other putative precursor peptides can be encoded in this region. Moreover, the promoter of *pedA15* genes is still unknown and turns it difficult to perform studies on the regulation of their expression.

This chapter will be focused on the sequencing and analysis of the upstream region of contig36 as it is crucial to better understand the potential of *ped15* cluster as well as the mechanism underlying its expression. Sequencing was performed using a primer walking strategy with random flanking primers and two main objectives were then established: i) identify the total number of putative lanthipeptides precursor peptides and ii) perform the *in silico* analysis of the promoter region of the *pedA15* genes.

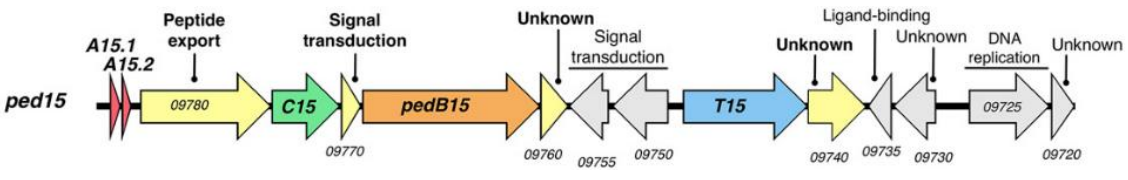

Figure 14: Representation of ped15 cluster in the beginning of contig36.

2.2 Results and Discussion

2.2.1 Assembly of *ped15* cluster upstream region

The final sequence of the upstream region of *ped15* cluster was obtained after three rounds of primer walking sequencing with random flanking primers (**Figure 15**). In particular cases, the PCR products needed to be cloned before sequencing. Primer walking has been efficiently used to amplify unknown regions of DNA [86]. In the first round following this methodology, contig36 and contig188 were successfully connected. The second round revealed the unknown upstream region of contig188 and the third round was pivotal to connect contig33 to the hitherto assembled sequence (**Figure 15**). In total, six new genes that can encode lanthipeptide precursor peptides were found, which were numbered after *pedA*15.2 (**Figure 15**).

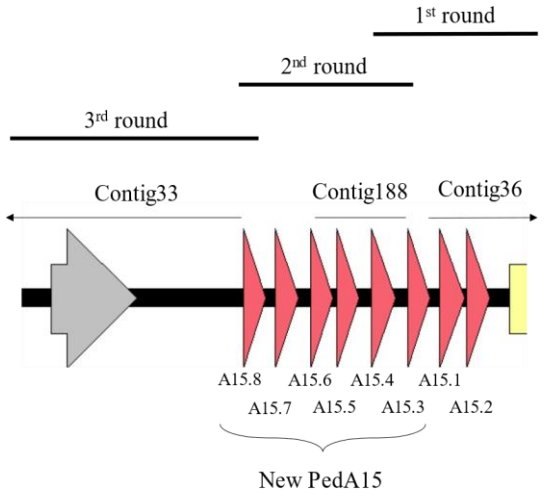

Figure 15: Representation of the upstream region of contig36. The sequencing performed allowed to connect contig188 and contig33 to contig36. The rounds illustrated in the figure represent the resulting amplicon after each primer walking procedure. Six new *pedA*15 genes were detected in the beginning of *ped15* cluster.

2.2.2 Analysis of PedA15 peptides found in the upstream region of *ped15*

Two structural genes had already been identified in cluster *ped15* (**Figure 14**). After sequencing the upstream region of contig36, six novel *pedA*15 genes were also identified to be associated with this cluster (**Figure 16**). The analysis of their encoded peptides allowed to recognize that all of them share a MKKVNLS(E/D)K(V/I)QLDKE(V/I)ISK consensus sequence in their leader peptides. This strongly suggests that their core peptides might be recognized and modified by the same PedB15 and PedC15 enzymes. In addition, the core peptides of all PedA15 have some degree of homology but are much more variable than the leader peptides (**Figure 16**). All the core peptides of the new PedA15 have 2 Cys residues (**Figure 16**). However, the already

known PedA15.1 and PedA15.1 possess three Cys residues each. Thus, the eight PedA15 peptides may have between two and three Lan/MeLan residues (**Figure 16**). Interestingly, the core peptide of PedA15.1, PedA15.2 and PedA15.8 have an unusual high number of additional amino acids in their C-terminus, after the last amino acid that can be involved in the formation of a thioether ring (Ser20 in PedA15.1, Cys12 in PedA15.2 and Ser14 in PedA15.8; **Figure 16**). As a result, these PedA15 have tails of 8/9 residues, which are not enfolded by Lan/MeLan bridges and are good candidates to be removed by the carboxypeptidase domain of the protein encoded by TH53_09780 gene in the *ped15* cluster (**Figure 14**) [62].

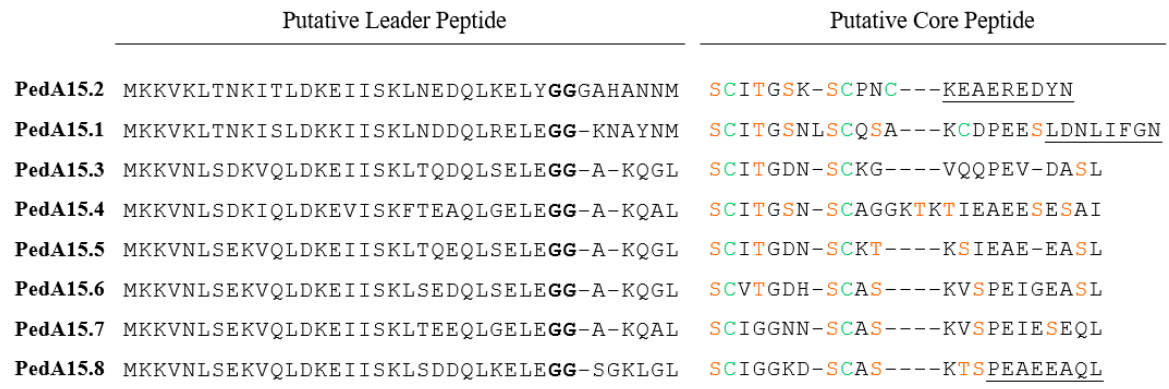

Figure 16: Alignment of *P. lusitanus* NL19 PedA15 peptides, with the double-glycine motif highlighted in bold in their leader sequences. The Cys residues present in the core peptide are coloured in green and Ser and Thr residues are both coloured in orange. The residues of PedA15.2, PedA15.1 and PedA15.8 in their C-terminus hypothesized to be removed by a carboxypeptidase are underlined.

Such as PedA15.1 and PedA15.2, all new PedA15 peptides possess a double-Gly motif which is likely recognized by the peptidase domain of PedT15 (**Figure 16**). The removal of the leader peptide by PedT15 will leave seven to five amino acids before the first Ser/Thr or Cys residue in the core peptide (**Figure 17**). Based on the structural knowledge, a second proteolysis step, independent of PedT15, is foreseen. This resembles what was found for the biosynthesis of some two-peptide (α and β) lantibiotics produced by Gram-positive bacteria, such as lichenicidin [12] (**Figure 17**). In the biosynthesis of this class II lantibiotic, the α and β peptides are modified by different LanM enzymes. In the case of the β peptide (Bliβ), the trimming step of an hexapeptide in the precursor peptide is catalyzed by a single uncharacterized serine protease (LicP) [12]. Besides, this proteolysis reaction is also predicted for pinensins. However, no gene encoding for a separate peptidase was encountered in the BGC of pinensins [32] (**Figure 13B**). In NL19, no protease is found encoded in *ped15* cluster and therefore it remains unknown. Still, it is assumed that the trimming of the N-terminal core peptide (after PedT15 intervention) is an essential step to obtain mature and active lanthipeptides.

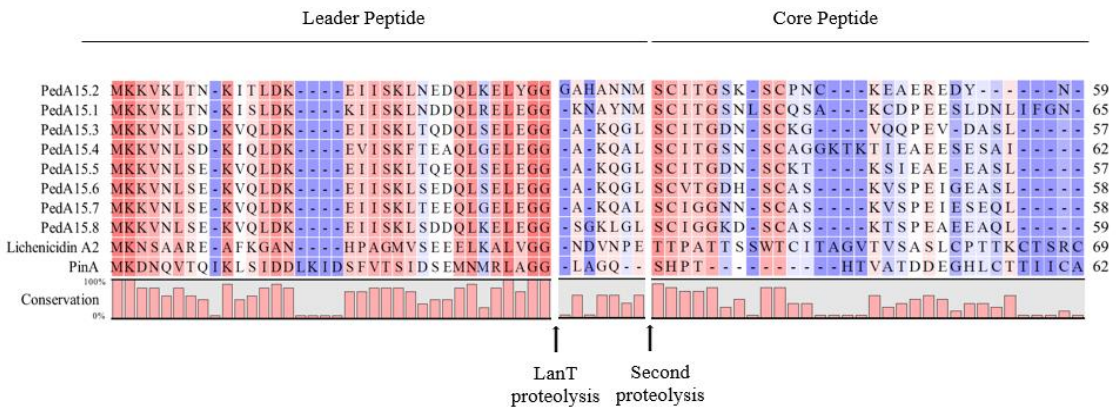

Figure 17: Alignment of PedA15 peptides with other two precursor peptides known to suffer a second proteolysis step. Lichenicidin A2 and PinA represent the Bliβ and the pinensins precursor peptides, respectively. The residues after which a presumable second proteolysis can occur were represented based on the information available for lichenicidin and pinensins. The color gradient, from blue (0%) to red (100%), indicates the conserved residues.

Class I LanAs are normally characterized by the existence of a conserved FxLD sequence. This motif has proven to be essential for the recognition of LanBs and LanCs enzymes, hence, being crucial for inducing PTMs [33,87,88]. It was not identified in PedA15 peptides (**Figure 18**) but all have a LD sequence in the central region of their leader peptides. In class I lanthipeptides, the residue preceding the LD sequence (represented by x) is often an N (polar amino acid) or a D (negatively charged amino acid) [88]. In all PedA15 this residue is not conserved, however, it is always a polar amino acid (**Figure 18**). Instead of the FxLD sequence motif, a recent studied proposed the LxLxKx$_5$L motif for several class I LanAs from Bacteroidetes [33]. All PedA15 leader peptides have a motif – (V/I)xLxKx$_5$L – that is similar to the above mentioned (with the exception being PedA15.4, which has an F instead of an L in the last residue of the motif) (**Figure 18**).

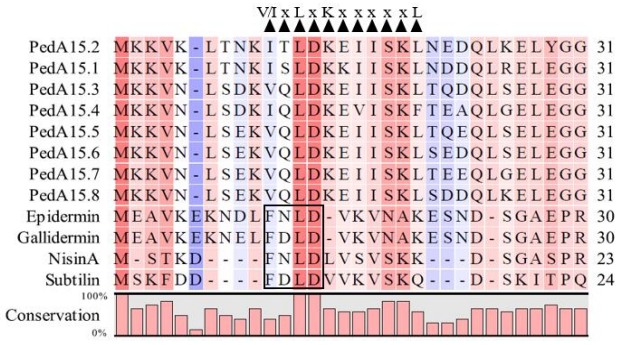

Figure 18: Alignment of leader peptides of PedA15 and four known class I lanthipeptides. The motif in the PedA15 peptides that resemble the LxLxKx$_5$L motif proposed by Walker *et al.* (2020) is emphasized with triangles. The class I FxLD conserved sequence was represented with a black box. The color gradient, from blue (0%) to red (100%), indicates the conserved residues.

2.2.3 Comparison with other class I/class II hybrids leader peptides

These novel "class I/class II hybrid" biosynthetic clusters with class I – LanC, LanB – and class II $LanT_p$ enzymes are rapidly being discover as genome mining and sequencing efforts help to uncover BGCs [62]. In the phylum Bacteroidetes, several genomes from a wide range of biological classes encode such hybrids. Likewise, some of these clusters also have a high number of potential precursor peptides as *ped15*. The sequences of these precursor peptides from selected Bacteroidetes were align (**Figure 19**). The alignment showed the presence of the conserved motif GG-motif characteristic of class II lanthipeptides, except for *Lewinella agarilytica* DSM 24740 in which one of the precursor peptides presented a GA motif, that is also recognized by $LanT_p$ transporters. Moreover, taking into account all but the *ped15* precursor peptides, the previously described $LxLxKx_5L$ motif is highly conserved. The only exception being the sequences from the precursor peptides of *Chryseobacterium viscerum* 687B-08 which presented an Ile instead of a Leu (isomers) in one of the residues of the motif. This shared motif is presumably relevant for the recognition of the biosynthetic machinery responsible for the PTMs[33]. A deeper analysis into the leader peptide sequences from the different LanAs within the same clusters proved a high degree of homology. Hence, as in PedA15 leader peptides, this suggest that they are substrates for the same biosynthetic enzymes.

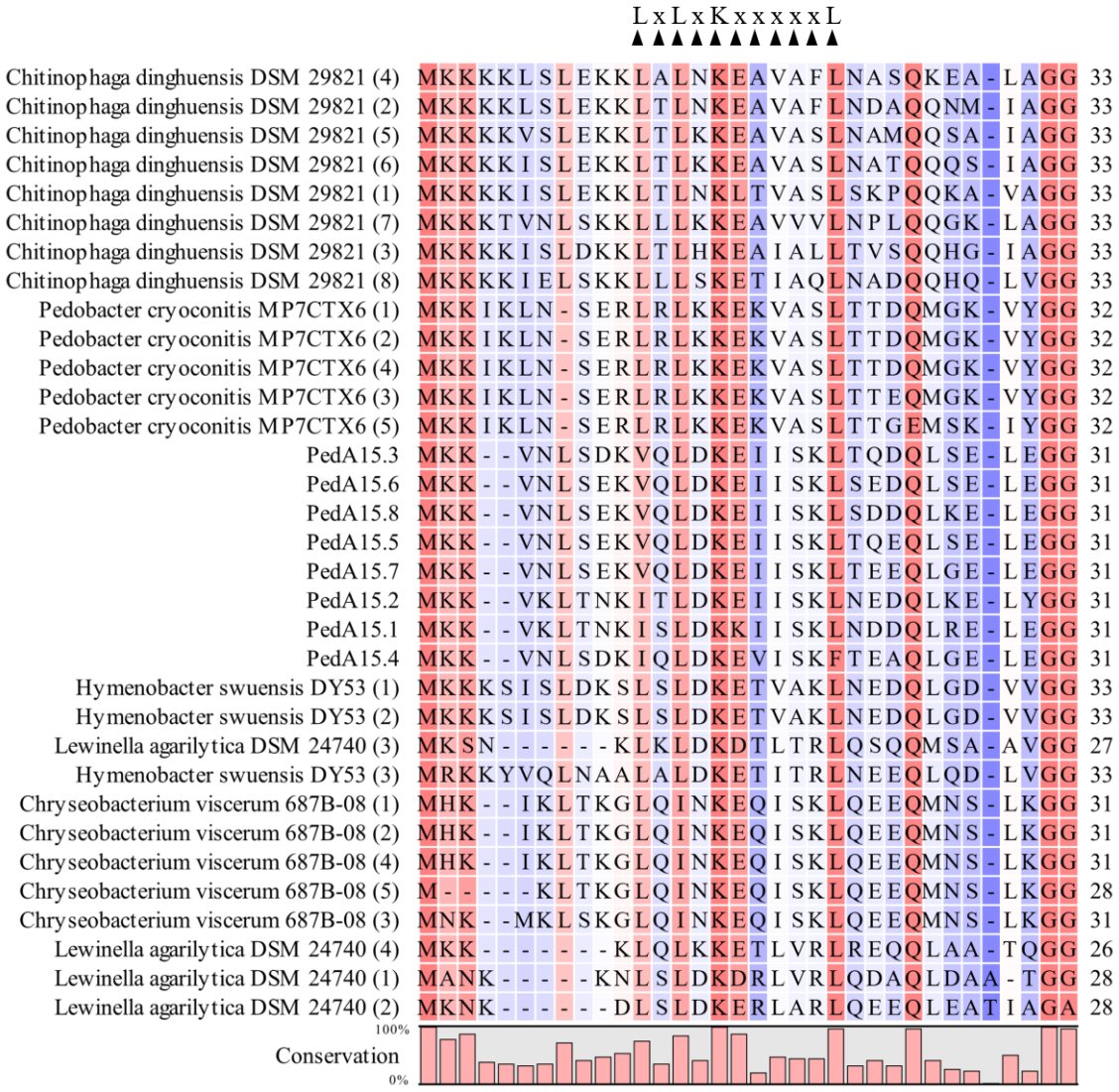

Figure 19: Alignment of the leader peptides of PedA15 and other LanAs from different species within the phylum Bacteroidetes with multiple precursor peptides in the same class I cluster. The motif in the leader peptides that resemble LxLxKx$_5$L motif is emphasized with triangles. The color gradient, from blue (0%) to red (100%), indicates the conserved residues.

2.2.4 Analysis of the *ped15* cluster promoter region

Prochlorosins and pinensins are the only characterized lanthipeptides produced by Gram-negative [32,89]. Nonetheless, the processes involved in the regulation of lanthipeptides biosynthetic machinery in Gram-negative bacteria are yet unknow. The bioinformatic analysis of *ped15* cluster revealed the presence of genes encoding for proteins involved in a two-component signal transduction system – a histidine kinase (TH53_09750) and the cognate DNA binding response regulator (TH53_09755) (**Figure 14**). The DNA binding response regulator belongs to the AlgR/AgrA/LytR family [90], that has a signal receiver (REC) and a DNA-binding domain (LytTR).

These type of transcriptional regulators are usually part of a two-component regulatory system [91,92], as discussed in section **1.2.6**. This allows us to hypothesize that the mechanism for regulating the transcription of PedA15 peptides may involve these two proteins. To further investigate this hypothesis, after sequencing the upstream region of contig36, the nucleotide region in the beginning of *ped15* cluster was analyzed to identify the promoter of *pedA*15 genes and putative binding sites for transcription factors (**Figure 20**).

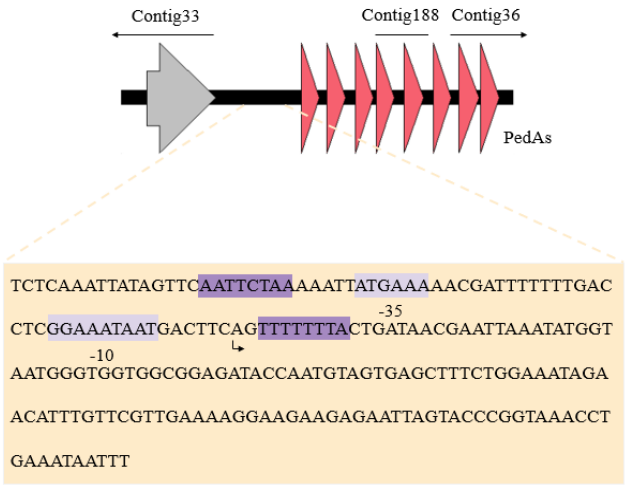

Figure 20: The nucleotide sequence presented in the light purple box represents the promotor region found upstream of *pedA*15 genes according to BPROM analysis; -35 and -10 (Pribnow box) - transcription regulatory regions; the arrow marks the transcription starting site; the dark purple box represents the sequence of binding sites for transcriptional repressors.

As shown in **Figure 20**, it was possible to identify a promoter upstream to the *pedA15* genes containing the -35 and -10 boxes (P_{pedA15} promoter). Furthermore, the search for transcription factors binding sites in the sequence showed two putative binding sites for the transcriptional regulators DeoR and LexA. Both of them are repressors that bind to their recognition sites and thus sterically hinders open complex formation between the promoter and the RNA polymerase [93–95]. This analysis suggests that all *pedA15* genes are transcribed in the same transcription unit (operon), under the control of the P_{pedA15} promoter. The binding sites for transcriptional repressors identified support that a DNA binding response regulator such as the one encoded by the TH53_09755 gene can act as a repressor. However, to confirm this hypothesis, further studies should be conducted to evaluate the role of DNA binding response regulator in the expression of *ped15* cluster, also including other promoters besides P_{pedA15}.

2.3 Conclusions

Within NL19 genome, *ped15* cluster, located at the end of a contig, encodes the essential genes for the biosynthesis of putative lanthipeptides. Herein, the sequencing and analysis of the upstream region of this cluster allowed to find six new *pedA15* genes encoding a wide diversity of core peptides, supporting the singularity and potential of *ped15* cluster. Moreover, a novel class I/II hybrid, characterized with class I – LanC, LanB – and class II $LanT_p$ enzymes as well as typical GG-motifs of class II precursor peptides, was uncovered within *ped15* and other clusters in Bacteroidetes phylum. As so, distinct structures and exciting new biotechnological attributes are expected from this class of lanthipeptides. Nonetheless, the mechanism underlying lanthipeptides production in Gram-negative bacteria is not yet established and further studies need to be performed. The *in silico* analysis of P_{pedA15} region revealed the potential effect of DNA binding response regulator as a repressor creating an exciting bridge between the knowledge heretofore acquired regarding lanthipeptides biosynthetic machinery and their regulation in NL19.

In short, the results obtained so far encourage the further investigation of lanthipeptides in *P. lusitanus* NL19 to better understand their potential as well as their transcription/production regulation.

2.4 Experimental procedures

2.4.1 Sequencing the upstream region of contig36 in strain NL19

The determination of the unknown DNA sequence adjacent to contig36 was determined following the random flanking primer 2 step PCR methodology, which steps are described in the following subsections.

2.4.1.1 DNA extraction

DNA was extracted from an overnight culture of strain NL19 using the Wizard® Genomic DNA Purification Kit (Promega), according to manufacturer's instructions (**Appendix 1**).

2.4.1.2 Amplification of contig36 upstream region

The upstream region of contig36 was amplified using random flanking primers as previously described by Toleman *et al* (**Figure 21**) [86]. In this methodology, 4 different random primers were used as forward primers. They were designed with a sequence tag of 24 bp preceded by seven random bases plus four bases in different combinations of two G and C residues and a T

residue in the end [86]. Two primers were designed to function as reverse primers with Primer3 [96] based on the sequence of contig36 already known. These were the biotinylated (containing biotin in its 5-terminal) and the nested primer and amplified fragments with approximately 100bp of difference (**Figure 21**).

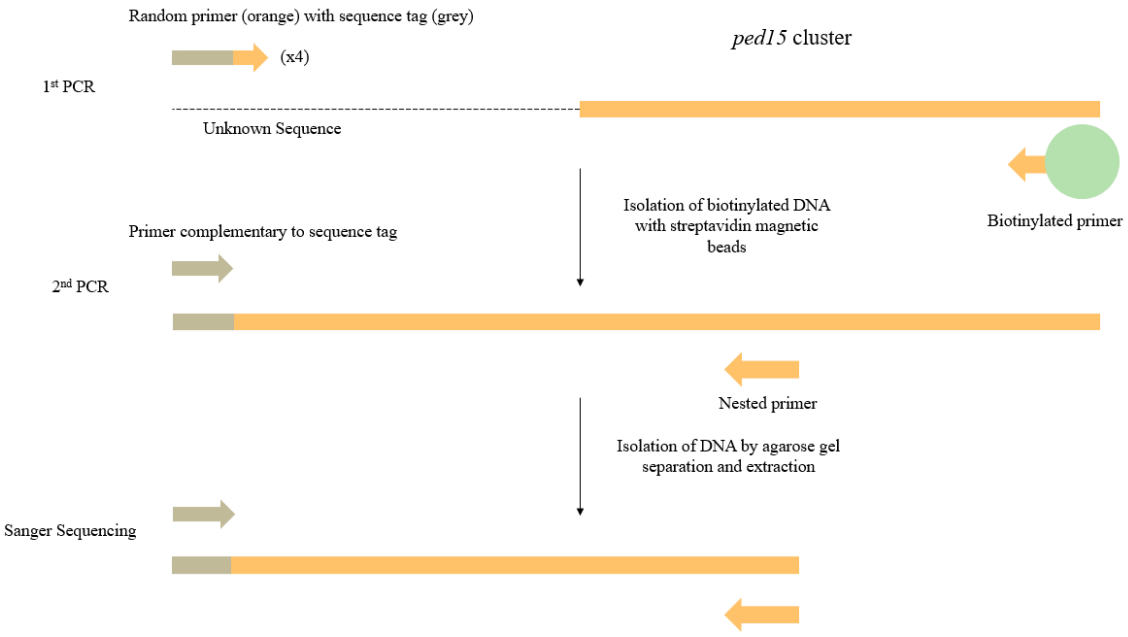

Figure 21: Random flanking primers strategy used to amplify the unknown DNA sequence in the upstream region of *ped15* cluster.

Briefly, 4 PCR reactions were performed, each of them containing the biotinylated primer and one of the four random primers (**Table 2**). The amplification was carried out by Platinum™ SuperFi™ DNA Polymerase (Invitrogen) in a 50 µL reaction containing the reagents described in **Table 3** and the amplification parameters of **Table 4**. The PCR products were purified using the NZYGelpure kit (NZYTech), according to the manufacturer's instructions (**Appendix 2**). Since the reverse primer had biotin, after purification, the PCR products were incubated with Dynabeads™ M-280 Streptavidin (Invitrogen) and the recovery of PCR products containing biotin was performed according to the manufacturer's instructions (**Appendix 2**). Thereafter, these products were used as DNA template to the second PCR that contained a forward primer complimentary to the sequence tag of the random primers (Random_2sd PCR; **Figure 21**) and the nested reverse primer targeting the contig36 sequence (**Figure 21**). The PCR reaction used the conditions and amplification parameters described in **Table 3** and **Table 4**, respectively. The resulting amplicons were analyzed by gel agarose electrophoresis (1%) and purified with the NZYGelpure kit (NZYTech), according to manufacturer's instructions (**Appendix 3**). DNA concentration was determined using Qubit® (**Appendix 4**). Afterwards, the products from the PCR reactions were

submitted to nucleotide sequence determination (STABVIDA-Portugal). The sequence reads were, then, assembled using CLC Main Workbench 8.1.3 program (Qiagen) as well as SnapGene Viewer 5.0.2 and new ORFs were sought after using NCBI ORFfinder.

In total, this procedure was performed 3 times for different nucleotide regions, in order to obtain a nucleotide sequence of sufficient length to obtain additional characterization of *ped15* cluster.

Table 2: List of primers used for random flanking 2 step PCR reactions.

Primer	Sequence (5´→ 3´)	Melting Temperature (°C)	Modification
Reverse1_biotin	TTGCGTTTGGCAAGTGTATC	55.3	5'-BIO
Reverse1_2sd	CTGCTCGTCCACAACTTTACC	59.8	-
Reverse2_biotin	CTTCTGGTTGCTGAACACCTT	57.9	5'-BIO
Reverse2_2sd	AATTCGCTCAATTGGTCCTG	55.3	
Reverse3_biotin	GATTTACTAAAGTTGCTCTG	51.1	5'-BIO
Reverse3_2sd	ATTGATATGCAAATGCCGG	52.4	
Random Primer 1	CAGTTCAAGCTTGTCCAGGAATTC$(N)_7$GGCCT	72.3	-
Random Primer 2	CAGTTCAAGCTTGTCCAGGAATTC$(N)_7$GCGCT	72.3	-
Random Primer 3	CAGTTCAAGCTTGTCCAGGAATTC$(N)_7$CCGGT	72.3	-
Random Primer 4	CAGTTCAAGCTTGTCCAGGAATTC$(N)_7$CGCGT	72.3	-
Random_2sd PCR	CAGTTCAAGCTTGTCCAGGAATTC	61.0	-

Table 3: PCR reaction used for amplification of contig36 upstream region.

Component	Volume
2X Platinum™ SuperFi™ PCR Master Mix	25 µL
10 pmol/µL Forward Primer	2.5 µL
10 pmol/µL Reverse Primer	2.5 µL
Template DNA	1 µL
Water, nuclease-free	19 µL

Table 4: Amplification parameters of PCR reaction.

Step		Temperature	Time
Initial Denaturation		98 ºC	30 sec
30 PCR Cycles	Denaturation	98 ºC	10 sec
	Annealing	51 ºC – 57 ºC	10 sec
	Extension	72 ºC	45 sec - 90 sec
Final Extension		72 ºC	5 min

2.4.1.3 Transformation of *E. coli* DH5α cells with plasmid pUC19

In some cases, the fragments amplified from PCR had ambiguous singularity. Therefore, before sequencing, amplicons were cloned into digested pUC19 and transformed in *E. coli* DH5α to allow the blue and white selection system and the analysis of the desired PCR products.

2.4.1.3.1 Digestion of amplicons and pUC19

Plasmid pUC19 was digested with *Sma*I in a reaction of 40 µL containing 4 µL of 10x Buffer Tango, 4 µL of pUC19 (500 ng/µL), and 4 µL of *Sma*I that was incubated at 30 ºC for 16 hours. Afterwards, the *Sma*I was inactivated by incubation at 65 ºC for 25 minutes. 60 µL of distilled water were added and plasmid was purified with NZYGelpure kit (NZYTech), according to the manufacturer's instructions (**Appendix 2**). Then, the plasmid concentration was determined using Qubit® (**Appendix 4**). This allowed to obtain a linear blunt-ended vector that could be ligated with PCR products obtained with a proofreading DNA polymerase that also produces blunt-ended amplicons. Ligation was performed with the Anza™ T4 DNA Ligase Master Mix (Invitrogen) in a final volume of 20 µL as described in **Table 5**. The adequate quantities of the purified and linearized vector and DNA insert needed for DNA ligation protocol were calculated. The calculation had in consideration the vector size (2686 bp), insert size (approximately 2000 bp) and vector/insert ratio of 1:3. The reagents were mixed, briefly centrifugated and incubated at room temperature for 30 min.

Table 5: Components used in DNA ligation protocol.

Reagent	Volume (µL)	Quantity (ng)
Linearized vector DNA	0.6	30
DNA insert	9.3	67
Anza™ T4 DNA Ligase Master Mix	5.0	-
Nuclease-free water	5.1	

2.4.1.3.2 Transformation of *E. coli* DH5α cells

After ligation, transformation was performed as follows: 50 µL of competent *E. coli* DH5α cells from -80 ºC were thawed on ice and 5 µL of the ligation were added in a microcentrifuge tube. The mixture was incubated on ice for 15 min. Heat shock was performed by placing the tube into a 42 ºC dry bath incubator for 45 sec. Immediately after, the tube was transferred to ice and

incubated for 2 min. 950 µL of Luria-Bertani broth (LB) were added and *E. coli* DH5α cells were incubated at 37 ºC with shaking, for 60 min. 100 µL of the transformation was plated onto Luria-Bertani agar (LA) plates containing 100 µg/mL of Amp, 0.2 mg/mL of X-Gal and 1 mM of IPTG. The plates were incubated overnight at 37 ºC.

2.4.1.3.3 Colony-PCR

Colony-PCR was used to confirm the insertion of an amplicon in white colonies. The amplification reaction was performed using *Taq* DNA polymerase (NZYTech) in a final volume of 12.5 µL containing 0.75 µL of $MgCl_2$ (50 mM), 1.25 µL of 10x reaction buffer, 0.25 µL of dNTPs (10 mM), 0.375 µL of each primer (10 mM) and 0.25 µL of *Taq* polymerase (5U/µL). For colony-PCR, only white-colonies were selected and each isolated colony was picked to the mixture as the template instead of purified DNA. The amplification conditions were as follows: 3 min at 95 ºC, 30 cycles of 30 sec at 94 ºC, 30 sec at 54 ºC, 1 min and 30 sec at 72 ºC. Lastly, the final extension step at 10 min at 72 ºC. The primers used are described in **Table 6**. PCR products were analyzed by agarose gel electrophoresis (1%). Positive colonies were grown overnight at 37 ºC in LB supplemented with ampicillin and plasmids were purified with the NZYMiniprep (NZYTech), according to the manufacturer's instructions (**Appendix 2**) and sent for Sanger sequencing (StabVida).

Table 6: List of primers used in colony-PCR screening of PCR products from upstream region of contig36. The respective sequence and annealing temperature are also indicated.

Primer	Sequence (5'→3')	Annealing Temperature (ºC)
pUC19Fw	AGGGTTTTCCCAGTCACGAC	54
pUC19_TC_rv	CTTCCGGCTCGTATGTTG	

2.4.2 Identification and characterization of *ped15* promoter region

BPROM (Softberry Inc., Mount Kisco, NY, USA; http://www.softberry.com/), a pipeline responsible for prediction of bacterial promoters, within Softberry (a web service that encompasses several software tools for genomic research), was used to search promoter regions within the previously sequenced and assembled region. As input, the nucleotide sequence determined from the amplification of the upstream region of contig36 was used. During processing, BPROM identified positions of the promoter, i.e., transcription start site, -10 box and -35 box and predicted binding sites for possible transcription factors.

Chapter III- Transcriptional analysis of *ped15* cluster

3.1 Introduction

Under standard laboratory conditions, expression of BGCs, that typically yielded numerous molecular scaffolds with biotechnological importance, is tightly regulated [97]. As so, certain BGCs encoded in NL19 genome might not be expressed at all, owing to the seemingly absence of activating signals in the experimental conditions tested so far. These type of clusters are normally designated as "cryptic" or silent BGCs, since their transcription does not exceed the threshold at which their products would be detected [97,98].

Lanthipeptides were formerly known as lantibiotics due to their antibacterial activities. As novel compounds containing the typical lanthionine rings but without antimicrobial activities were discovered, the family name was broadened to lanthipeptides [9,13]. Even though NL19 presented antimicrobial activity, this trait was not assigned to lanthipeptides. However, so far, the production of these compounds, including their transcription was not investigated. As such, their BGCs can be cryptic and therefore their bioactivity, if any, could never have been detected.

The production of pedopeptins, a class of SMs belonging to the NRPs group, proved to be the main responsible for the antimicrobial activity of NL19 [59]. RT-qPCR results showed that the expression of the NRPS cluster responsible for pedopeptins production varied significantly in broth with different concentrations of peptone from casein (PC). As a consequence, pedopeptins production in medium with high concentration of PC (TSB100%) was lower than in a medium with a reduced concentration of PC (PC25%) [59]. Therefore, it is possible that the concentration of PC can modulate the expression of many other genes in NL19, including those that are involved in the production of SMs such as lanthipeptides. To better understand the effects of PC in the transcriptome of NL19, a high-throughput RNA-seq analysis was performed in the presence and absence of high concentrations of PC (unpublished data). Regarding lanthipeptide BGCs, results showed that, in general, their transcription was not affected by the concentration of PC in the broth (unpublished data). One exception was one of the *pedA*15 precursor peptides (*pedA15.2*) which was found to be about 6X more expressed in high concentrations of PC (TSB100%; **Table 7**). Also considering *ped15* cluster, it was found that, independently of the broth, the genes TH53_09755 and TH53_09750 (**Figure 14**) presented higher expression values than the other genes of the cluster, particularly the first (**Table 7**). Since their predicted molecular functions are related with signal transduction (as discussed in section **2.2.4**), these results support their possible involvement in transcriptional repression, probably by binding to the promoter regions of other genes of the cluster, including that of *pedA* genes.

Table 7: RNA-seq results obtained for the genes composing the *ped15* cluster when *P. lusitanus* NL19 was grown in broth with high or low concentrations of PC (TSB100% and PC25%, respectively). The locus tag, gene designation and fold changes ≥ 1 between the expression values in the different media are described. The predicted transcription unit as well as the conserved domains of the proteins encoded in this gene cluster are also represented. ND stands for not detected.

Predicted Tran--scription unit	Locus tag	Gene	Conserved domains	Expression value in TSB100%	Expression value in PC25%	Fold change	
						PC25%/ TSB100%	TSB100%/ PC25%
1	TH53_09790	*pedA*15.1	ND	3.59	5.16	1.44	-
	TH53_09785	*pedA*15.2	ND	3.66	0.60	-	6.10
2	TH53_09780	-	Carboxy--peptidase, CirA and OMP	2.01	2.83	1.41	-
	TH53_09775	*pedC*15	Putative LanC	4.83	6.44	1.33	-
	TH53_09770	-	CitB superfamily	9.84	8.15	-	1.21
	TH53_09765	*pedB*15	Putative LanB	27.15	33.78	1.24	-
	TH53_09760	-	ND	16.30	15.57	-	1.05
3	TH53_09755	-	DNA binding response regulator	185.49	304.15	1.64	-
	TH53_09750	-	Histidine kinase	65.67	80.60	1.23	-
4	TH53_09745	*pedT*15	SunT	4.86	4.29	-	1.13
	TH53_09740	-	ND	8.48	8.93	1.05	-

However, the RNA-seq analysis was based only in one biological replica and, as such, the obtained results need further validation and *ped15* cluster was used as case study. Gene expression of some of the *ped15* genes was evaluated in PC25% (lower PC concentrations) and TSB100% (higher PC concentrations) broth at different growth phases (exponential and stationary) by RT-qPCR.

The main objective of the work presented in this chapter was to analyze the expression of *ped15* cluster under different conditions (media and growth phase) in order to understand if its expression is affected by PC concentrations (as for pedopeptins) and/or if it is a cryptic cluster.

3.2 Results and Discussion

3.2.1 Influence of PC concentration in the transcription of *ped15* genes

To determine the influence of PC concentration in the transcription of *ped15*, the expression of seven genes from the cluster was evaluated in PC25% and TSB100% broth media. A

RT-qPCR experiment was implemented to quantify, in exponential and stationary phase, the expression of the genes encoding: i) the precursor peptide (TH53_09785), ii) the protein with a carboxypeptidase domain (TH53_09780), iii) the lanthionine synthetase enzymes (TH53_09775 and TH53_09765), iv) the signal transduction proteins (TH53_09755 and TH53_09750) and v) the bifunctional transporter/protease (TH53_09745) (**Figure 14**).

The RT-qPCR revealed that only four genes had a fold change ≥ 2 or ≤ -2 that include (**Figure 22**): i) the *pedA15*.2 (2.3X fold change in PC25% in stationary phase), ii) the carboxypeptidase domain protein (3.7X fold change in TSB100% in stationary phase), iii) the *pedC15* (2.1X fold change in TSB100% in stationary phase) and iv) the DNA binding protein (2.0X fold change in TSB100%, but only in the exponential phase). From these, only the expression of the protein with a carboxypeptidase domain was found to be statistically different (with a p-value ≤ 0.05) (**Figure 22**). Thus, only this gene was identified as being differentially expressed (DEG; fold change ≥ 2 or ≤ -2, with a p-value ≤ 0.05) in the stationary phase.

Noteworthy that the RNA-seq analysis was performed only for the exponential growth phase. Apart from *pedA15.2*, all the RT-qPCR results (considering only those obtained for exponential phase) were in accordance with the RNA-seq results previously obtained (**Table 7**; see section **3.1**). As such, the RT-qPCR findings proved to be a helpful technology to validate the RNA-seq results. The validation of DEGs identified using RNA-seq is often performed using RT-qPCR [76]. For example, a study with the Gram-negative *Acinetobacter baumannii* showed that the DEGs identified with RNA-seq had high correlation with that from the RT-qPCR [99]. However, in the case of *pedA15.2*, RNA-seq and RT-qPCR results were inconsistent. A possible explanation may be that RNA-seq was based only in one biological replica and, therefore, lack repeatability, which is always pivotal to guarantee a valid interpretation of the results. In fact, a highly replicated RNA-seq experiment highlights the primary effect of increasing replicates numbers to increase the sensitivity of these tools [100,101]. RT-qPCR results were based on three biological replicas, each with two technical replicas, for PC25% as well as TSB100%. Consequently, the biological conclusions obtained from the RT-qPCR experiments are more meaningful [76].

Overall, having in consideration the RT-qPCR and RNA-seq results, the concentration of PC had no major effects in the transcription of most of the genes of *ped15* cluster from *P. lusitanus* NL19. This suggests that, unlike pedopeptins [59], the concentration of PC has no effect in the transcription of lanthipeptides biosynthesis machinery of *ped15* cluster in *P. lusitanus* NL19.

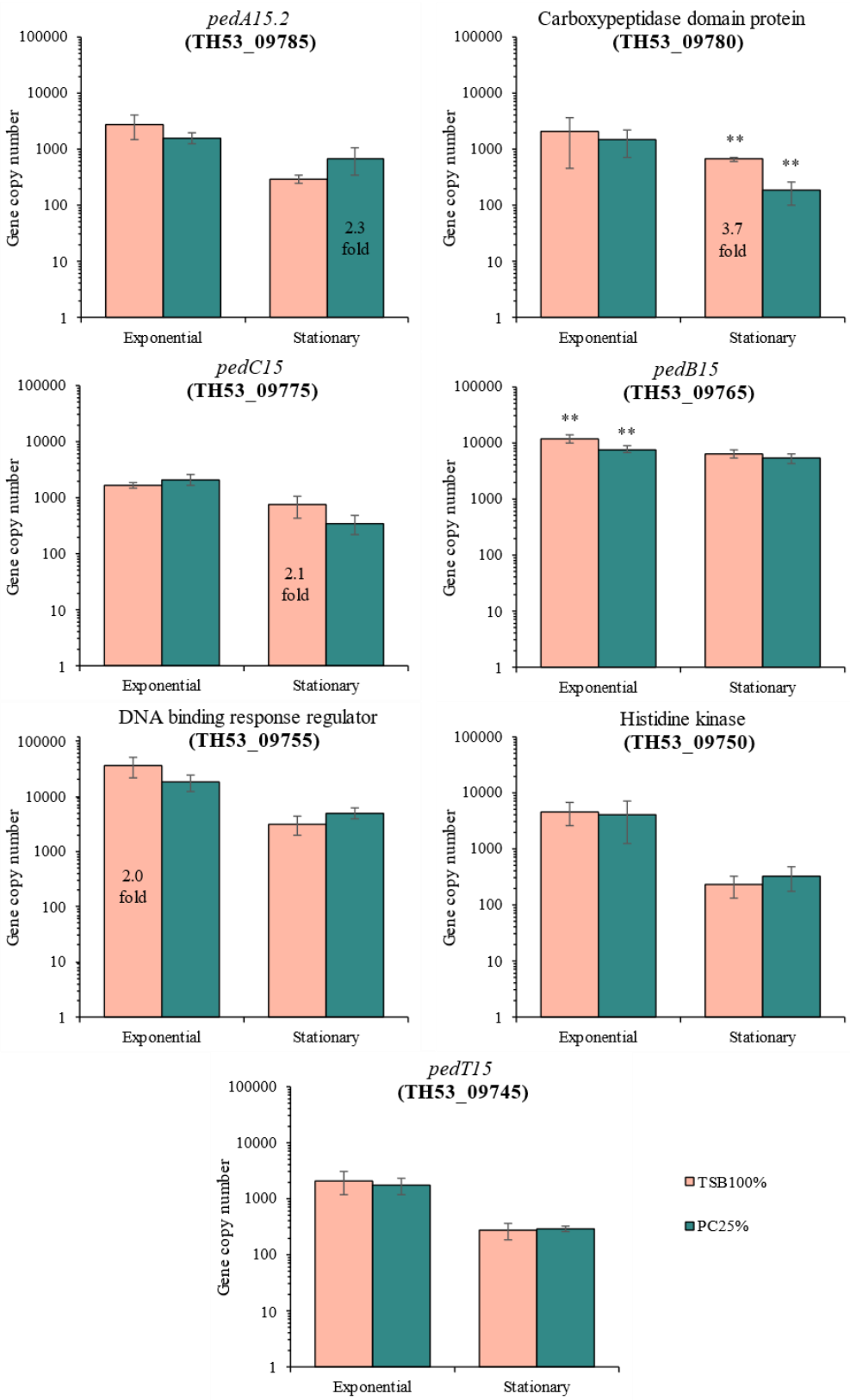

Figure 22: RT-qPCR absolute quantification results of the expression of *ped15* cluster genes when *P. lusitanus* NL19 was grown in different media (PC25% and TSB100%) and at different growth phases (exponential and stationary). The statistical analysis showed significant statistical differences in two genes (**$p \leq 0.05$).

3.2.2 Influence of growth phase in the transcription of *ped15* genes

The transcription of all *ped15* genes was higher in exponential phase and decreased in the stationary phase in both media tested (**Figure 22**; **Table 8**). In TSB100%, the highest differences were identified for the histidine kinase, the DNA binding response regulator, the *pedA*15.2 and the *pedT15* genes (**Table 8**). In PC25%, the highest differences were identified for the histidine kinase, the carboxypeptidase domain protein, the *pedC*15 and the *pedT15* genes (**Table 8**). Among them, the major difference was in the histidine kinase gene since its transcription decreased 20.4X and 12.6X from the exponential to the stationary phase in TSB100% and PC25%, respectively (**Table 8**). The same magnitude of decrease was observed for the gene encoding the transcriptional repressor in TSB100% (11.5X; **Table 8**). The transcription of *pedC15*, *pedB15* and the gene encoding a protein with a carboxypeptidase domain was found to be more constant in TSB100%, albeit slightly decreased in the stationary phase (**Table 8**). Likewise, in PC25%, the transcription of *pedB15, pedA15.2* and the DNA regulator gene was consistent, although it is higher in the exponential phase (**Table 8**). Taken together, these results indicate that the growth phase can influence the transcription of the genes encoded by the *ped15* cluster.

The regulation of lanthipeptide synthesis is often associated with cellular events occurring in late exponential growth phase [15]. In fact, in the nisin and subtilin producers, the transcription of biosynthetic genes was found to begin during mid-exponential growth, reaching a peak at the log- to stationary phase transition. Interestingly, these lantibiotics trigger their own transcription [15,18]. Medium acidification, quorum sensing and solid substrates proved to stimulate gene expression in others lanthipeptides [18]. Nevertheless, the abovementioned examples are related to lanthipeptides produced by Gram-positive bacteria. The factors and environmental conditions that mediate transcription in Gram-negative bacteria might be different. The Gram-negative *Prochlorococcus* MIT9313, which produces prochlorosins, revealed the transcription of *procM* and several *procA* genes in exponential growth and prochlorosins production was detected in late-exponential phase [61].

Table 8: Fold change of *ped15* genes transcription between stationary phase and exponential phase of NL19 grown in broth with high or low concentrations of PC (TSB100% and PC25%, respectively).

Gene	TSB100%	PC25%
Histidine kinase	**-20.4**	**-12.6**
DNA binding response regulator	**-11.5**	-3.6
pedT15	**-7.7**	**-6.0**
pedA15.2	**-9.3**	-2.3
Carboxypeptidase domain protein	-3.2	**-8.2**
pedC15	-2.2	**-6.1**
pedB15	-1.9	-1.4

3.2.1 Overall transcription of *ped15* genes

The genes with higher transcription were found to be comparable independently of the media, growth phase or molecular method used in the analysis (RT-qPCR or RNA-seq) (**Figure 23**). In all, either the DNA-binding response regulator or the *pedB15* were always the more expressed genes (**Figure 23**). According to RT-qPCR results, in average, the transcriptional repressor is 11X more expressed than all the other genes (not considering *pedB15*) and *pedB15* is 9X (not considering the transcriptional repressor). Surprisingly, the transcriptional levels of *pedB15* and *pedC15* genes were always distinct (about 10X), despite the fact that they were predicted in the same transcription unit (**Table 7**). Thus, these results can indicate that their transcription is not directed by the same promoter. Similar results were obtained for the transcriptional repressor and the histidine kinase.

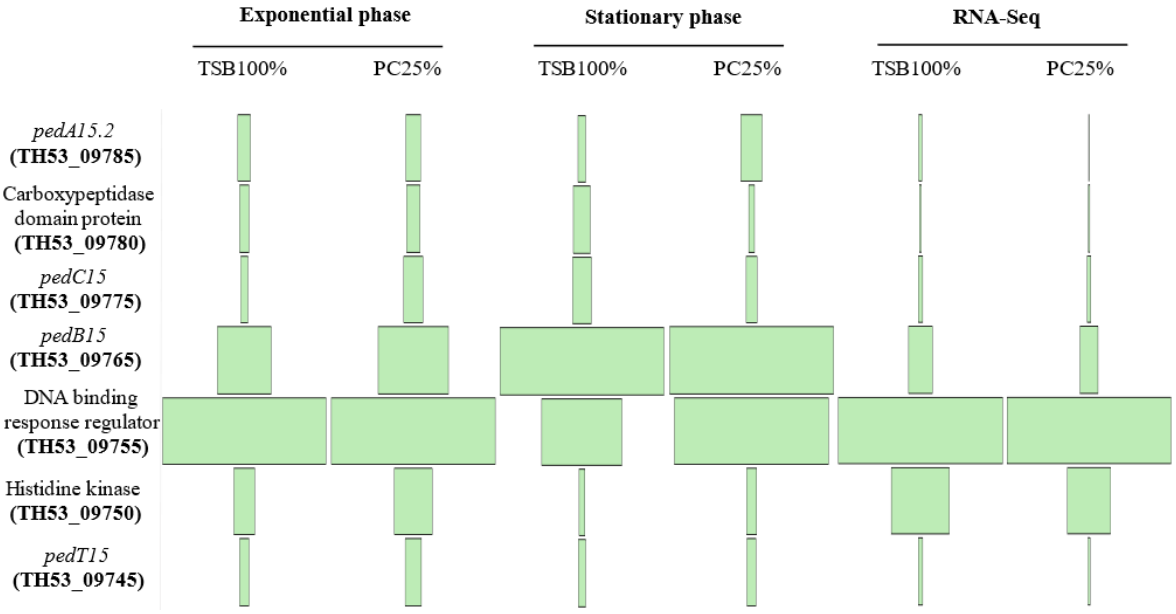

Figure 23: Funnel chart showing the transcription level of *ped15* genes determined in each condition that was evaluated: different media (TSB100% and PC25%), different growth phase (exponential and stationary) and the two methodologies used (RT-qPCR and RNA-seq).

The average number of transcripts of a housekeeping gene such as *recA* (encoding a multifunctional protein involved in homologous recombination and DNA repair) [102] determined in the exponential growth phase of NL19 were between 2 and $3x10^5$ copies/µL (in PC25% and TSB100%, respectively). Taking this into consideration, the transcription of all genes, with the exception of *pedB15* and the DNA-binding regulator, is very low (**Figure 24**). The high expression

of the DNA binding response regulator, as opposed to the low expression of the majority of the other genes in the *ped15* cluster, strengthens the hypothesis that it can act as a repressor.

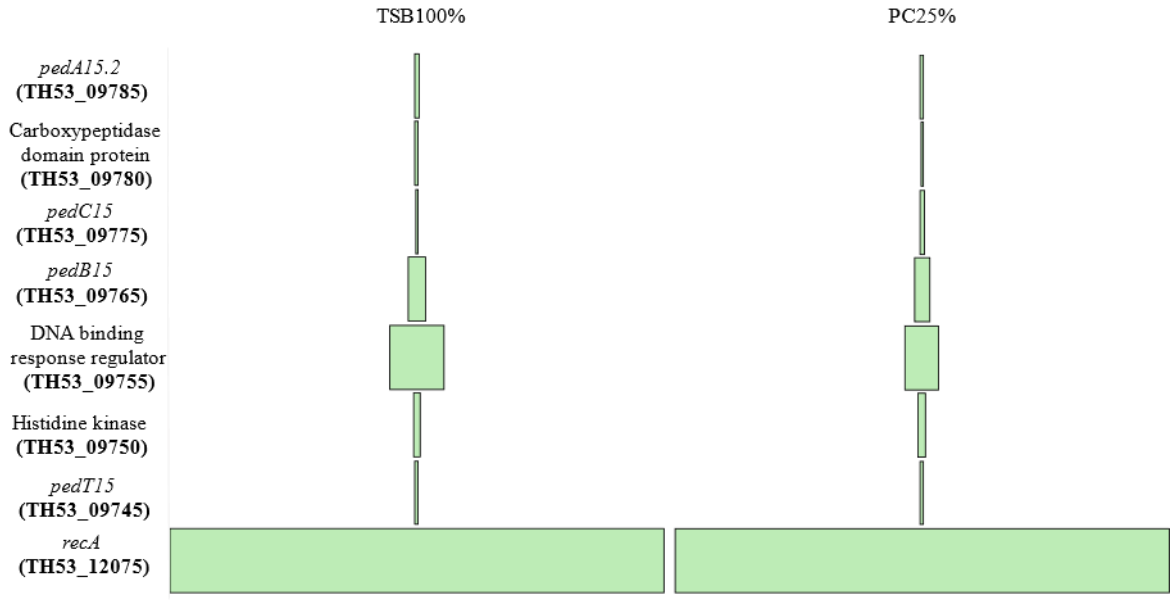

Figure 24: Funnel chart showing the transcription level of *ped15* and *recA* genes determined for TSB100% and PC25% in the exponential phase of growth by RT-qPCR.

Altogether, these results suggest that *ped15* is a cryptic cluster because most of the genes that are essential for the production of the mature lanthipeptides, such as the precursor peptide (*pedA15.2*), the cyclase (*pedC15*) and exporter (*pedT15*), have very low transcription rates in the media tested. Therefore, strategies that will increase their production are needed to allow their purification for structural characterization and bioactivity testing. For instance, the transcription of *procM* and several *procA* genes in the Gram-negative *Prochlorococcus* MIT9313 was downregulated under nitrogen starvation suggesting that nutrient depletion can influence its gene expression [61]. Pinensins, the first antifungal lantibiotic, were only produced by the Gram-negative *C. pinensis* cultivated in MYC-medium, which has glucose and phytone peptone, an enzymatic digest of soy with high vitamin and high carbohydrate content [32]. The computational analysis of pinensins BGC and corresponding cognate peptide revealed resemblances with *ped17* cluster, a lanthipeptide BGC encoded in NL19 genome (**Figure 13B;** section **1.3.1.1**) [62]. Like the pedopeptins produced by the NL19 strain, pinensins have activity against *Saccharomyces cerevisiae* [59]. However, as *ped15* cluster, *ped17* of NL19 can be a cryptic cluster. Therefore, if NL19 is grown under culture conditions that allow the transcription of *ped15*, as well as *ped17*, there may be an increase in its antifungal activity.

3.3 Conclusions

Different concentration of PC had a tremendous effect in the expression of the NRPS cluster responsible for the biosynthesis of pedopeptins. As so, it was hypothesized that other SM BGCs, such as *ped15*, could have significant differences in their transcription in the same conditions. RT-qPCR results revealed that only the gene encoding the carboxypeptidase domain protein, when evaluated in stationary phase, was considered a DEG in media with different concentrations of PC. Therefore, the concentration of PC did not have a key effect in the transcription of *ped15* cluster. On the contrary, the growth phase proved to have an impact in the transcription of the *ped15* genes studied since the gene expression was always higher in the exponential phase. Nevertheless, our results indicate that *ped15* is a cryptic cluster as the transcription levels of the genes encoding the biosynthetic machinery of lanthipeptides are very low, when compared to the housekeeping gene *recA*. The only exception being the DNA binding response regulator presumably due to its function as a repressor of *ped15* cluster. In conclusion, more comprehensive laboratory studies focused on *ped15* cluster transcription are needed to better understand the influence of medium components and also to identify possible trigger factors that turn on the expression of this lanthipeptide cluster. For example, other culture medium, like MYC (the medium that boosted the production of pinensins) might have this effect.

3.4 Experimental procedures

3.4.1 Media and growth conditions

P. lusitanus NL19 was grown in TSA plates at 26 ºC. When single colonies were visible, TSB100% and PC25% (**Table 9**) were used as broth to grow *P. lusitanus* NL19 at 26 ºC, with aeration at 180 rpm. For both media, the final pH was adjusted with NaOH to 7.0.

Table 9: Medium components and their respective concentration, used to test the effect of peptone of casein in the transcription of *ped15* in NL19.

	Peptone of casein	Peptone of soy	NaCl
TSB100%	15 g/L	5 g/L	5 g/L
PC25%	3.75 g/L	5 g/L	5 g/L

3.4.2 Preparation of NL19 cells for RNA extraction

P. lusitanus NL19 was grown overnight in 5 mL of TSB100% or PC25% broth as described in section **3.4.1**. The OD_{600nm} of the cultures was measured and normalized to 0.7. 500 µL of this culture was used to inoculate 100 mL erlenmeyers containing 50 mL of either TSB100% or PC25% broth and incubated at 26 ºC with aeration at 180 rpm. Three replicas for each media were performed.

At an OD_{600nm} of approximately 0.5-0.6 (between 11 to 12 hours; exponential phase samples), 1 mL of each culture was transferred to a microcentrifuge tube. After 48 hours of growth (stationary phase samples, according to Covas *et al* [59]), the cultures were diluted to an OD_{600nm} of 0.5-0.6 in a final volume of 1 mL in a microcentrifuge tube. All samples were centrifuged at 4 ºC and 13200 rpm for 5 min. The supernatant was discarded, and the cell pellets were resuspended in 1 mL of cold 0.9% NaCl. The samples were centrifuged again at 4 ºC and 13200 rpm for 5 min. The resulting cell pellet was conserved in RNAlater™ (QIAGEN, USA) at -80 ºC for the correct stabilization and protection of cellular RNA, until further processing.

3.4.3 RNA purification and cDNA synthesis

Total RNA was purified with the Invitrogen™ PureLink® RNA Mini Kit, according to manufacturer's instructions (**Appendix 5**) from each of the tubes prepared in section **3.4.2**. RNA concentration was determined with Qubit® and integrity was checked with a 1% agarose gel containing 1% of bleach. 500 ng of RNA was used to synthesize cDNA with the SuperScript™ IV VILO™ Master Mix (Invitrogen), according to manufacturer's instructions. Briefly, a reaction of 10 µL containing 1 µL of 10X ezDNase Buffer, 1 µL of ezDNase enzyme and 500 ng of RNA was gently mixed and incubated at 37 ºC for 2 minutes. Afterwards, the reaction was briefly centrifuged and placed on ice. Then, 4 µL of SuperScript™ IV VILO™ Master Mix and 6 µL of nuclease-free water were added to the tube containing the 10 µL reaction mix. The reaction was gently mixed and cDNA synthesis was as follows: 10 min at 25 ºC to anneal the primers, 10 min at 50 ºC to reverse transcribe the RNA and 5 min at 85 ºC to inactivate the enzyme. In parallel, a no RT (reverse transcriptase) control reaction was performed by substituting SuperScript™ IV VILO™ Master Mix with the SuperScript™ IV VILO™ No RT Control. Lastly, the synthesized cDNA was stored at -20 ºC.

3.4.4 RT-qPCR analysis of *ped15* genes

To evaluate gene expression of *ped15* cluster, a RT-qPCR experiment targeting the genes encoded in the cluster was performed. The genes selected were listed in section **3.2.1**. 10-fold dilution series, ranging from 3.09 x 10^6 to 3.09 x 10^2 copy number/µL of total DNA of *P. lusitanus* NL19 were used as a DNA template to produce a quantification standard curve. This procedure was performed in two independent series, each one with three technical replicas. For cDNA quantification, 1/10 or 1/100 dilutions were prepared and used as DNA template for the RT-qPCR reaction depending on the gene analyzed. Three biological replicas and two technical replicates for each gene and condition tested (PC25%, TSB100% and two different growth phases). RT-qPCR reactions were performed at a final volume of 20 µL containing 1 µL of cDNA or DNA (standard curve), 300 nM of each primer (**Table 10**), 10 µL of PowerUp™ SYBR™ Green Master Mix (Thermo Fisher Scientific) and nuclease-free water. Quantification was performed with the CFX96 real-time PCR system (Bio-Rad, Hercules, CA) with the following parameters: 2 min at 50 °C, 2 min at 95 °C, and 50 cycles of 15 sec at 95 °C, 15 sec at the appropriate annealing temperature depending on each pair of primers (**Table 10**) and 1 min at 72 °C, with a plate read at the end of each cycle. At the end of the 50 cycles, a melting curve, ranging from 65 °C to 95 °C (increments of 0.5 °C at each 5 sec), was performed to assess the presence of nonspecific products or primer-dimers in the reaction. For each reaction, non-template controls (NTC) were included (without cDNA or DNA). PCR efficiency and quantification were calculated by the CFX Manager software (Bio-Rad). The RT-qPCR amplification was considered valid only when the PCR efficiency was between 90–110% and the standard curves had a correlation coefficient greater than 0.990.

Table 10: List of primers used in RT-qPCR and annealing temperatures used for each pair in RT-qPCR. The target and expected size of each amplicon are also indicated.

Primer	Sequence (5'→ 3')	Annealing Temperature (°C)	Target	Amplicon Size (bp)
Forward	AGAGTTATATGGTGGTGGTGCA	59	TH53_09785	101
Reverse	AGTTATAATCTTCCCGCTCTGCT			
Forward	TCGAATGTTAGCGGGAATGT	56	TH53_09780	106
Reverse	ATCCTGCAGAAACTCCGAAAT			
Forward	TCACTGCGTTGTCTATGGTGA	57	TH53_09775	103
Reverse	GGCCAGATCAATTCCTTCAA			
Forward	TCGCTTTAATGATCCGGAAG	56	TH53_09765	101
Reverse	TCAGCGCCTGTTGTATTCTG			

Primer	Sequence (5'→ 3')	Annealing Temperature (°C)	Target	Amplicon Size (bp)
Forward	CAACCCGTTTCATCGTAAGG	56	TH53_09755	110
Reverse	ACACCGGGTAATGTTTGATGA			
Forward	AACACGGGATTGATGCAACT	57	TH53_09750	114
Reverse	CAGTTGGGTAGCCATTACCG			
Forward	CAGGCCAGATTGTTCAGGTT	57	TH53_09745	116
Reverse	CTGGCTGCAAGAAACAGGAT			

Chapter IV- Influence of peptone from casein in the proteome of NL19

4.1 Introduction

The genome of NL19 has 17 clusters encoding the biosynthesis of a wide range of SMs (including NRPs, siderophores, polyketide and lanthipeptides) [57,59,62]. Recently, it was proved that, in NL19, high concentration of PC induced the repression of one of those SMs: the pedopeptins, a type of NRP synthesized by a NRPS BGC (**Figure 25**) [59].

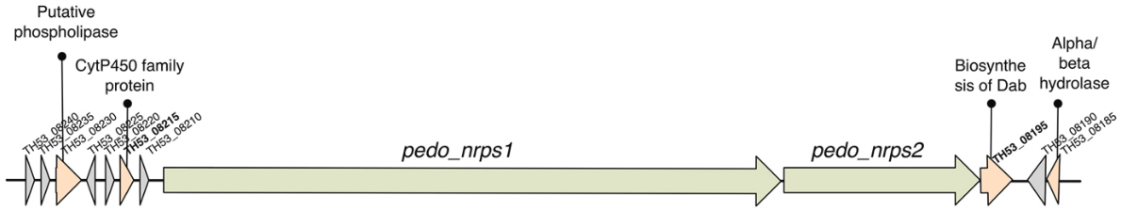

Figure 25: Schematic representation of the NRPS cluster responsible to produce pedopeptins in NL19.

High concentrations of nitrogen sources (such as PC) in medium usually spur bacterial growth but can also repress/suppress the biosynthesis of many SMs [103]. The seemingly increase of the production of SMs when the concentration of a nitrogen source was reduced, was observed for the biosynthesis of the polyketide actinorhodin and the alkaloid undecylprodigiosine by *Streptomyces* spp. [104]. Moreover, *Pseudoalteromonas* spp., showed maximum antimicrobial activity in media with low nitrogen concentration, due to a higher production of SMs [103]. Nonetheless, the Gram-negative bacterium *Serratia marcescens* demonstrated enhanced production of prodigiosin, a natural red pigment and bioactive alkaloid, in medium supplemented with peptone as the nitrogen source [105].

However, in NL19, the concentration of PC does not seem to affect the transcription of the genes associated with the production of other SMs, such as the lanthipeptides encoded by cluster *ped15* (see section **3**). Thus, the repressive effect of PC towards the transcription of SMs biosynthetic machinery is not ubiquitous. To better understand the impact of PC in the metabolism of NL19, its proteome was determined in the two different media TSB100% and PC25%. For this purpose, total proteins were identified and quantified by a combination of nano LC-ESI-MS/MS and UniProt search using Proteome Discoverer software. Accordingly, the objectives of this chapter were i) to identify the proteins that are differentially produced by *P. lusitanus* in different concentrations of PC and ii) among the differentially produced proteins, identify those that are associated with the production of SMs.

4.2 Results and Discussion

The proteome of NL19 grown in high concentrations of PC (medium TSB100%) was analyzed and compared with that of NL19 grown in lower concentrations of PC (medium PC25%) in the exponential phase of growth. In total, 96 proteins were detected as differentially expressed proteins (DEPs; **Appendix 7**) according to the criteria: abundance ratio TSB100%/PC25% ≥ 2 or ≤ 0.5 and *p*-value ≤ 0.05. From these, 23 and 73 proteins were found to be overexpressed and underexpressed, respectively, in high concentrations of PC (TSB100% broth).

4.2.1 Proteins overexpressed in high concentrations of PC

The up-regulated proteins were successfully distributed into 4 biological processes and 6 molecular functions. The most abundant biological processes (**Figure 26A**) were transport (24.00%) and metabolic process (24.00%), both with 6 proteins. Equally, catalytic activity (5 proteins; 20.83%) and transporter activity (3 proteins; 12.50%) were the molecular functions (**Figure 26B**) found more regularly. Nonetheless, the classification of approximately 50% of the DEPs into a biological process (10 DEPs) and a molecular function (12 DEPs) was not possible (**Figure 26**).

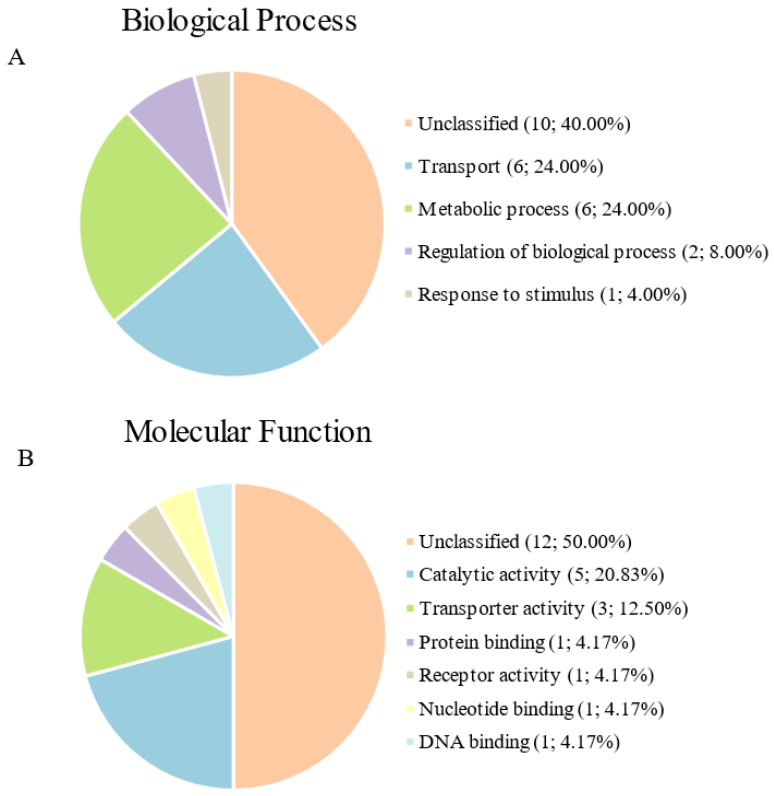

Figure 26: Distribution of the biological processes (A) and molecular functions (B) found in the overexpressed proteins in TSB100% analyzed according to Proteome Discoverer.

4.2.1 Underexpressed proteins in high concentrations of PC

Results showed that 73 proteins were significantly underexpressed in TSB100%, when compared to PC25%. These DEPs were successfully categorized into 5 biological processes and 8 molecular functions. In the case of the biological processes (**Figure 27A**), the majority of proteins are associated with metabolic and transport processes (19 proteins; 22.62% and 10 proteins; 11.90%, respectively). Catalytic activity and receptor activity (19 proteins; 21.84% and 6 proteins; 6.90% respectively) were the most common molecular functions (**Figure 27B**), followed by metal ion binding and transporter activity (each represented with 5 proteins; 5.75%). However, Proteome Discoverer software was unable to predict the biological process for 45 proteins and the molecular function for 42 proteins (**Figure 27**). As so, a large percentage of the downregulated proteins identified (more than 50%) is still not characterized.

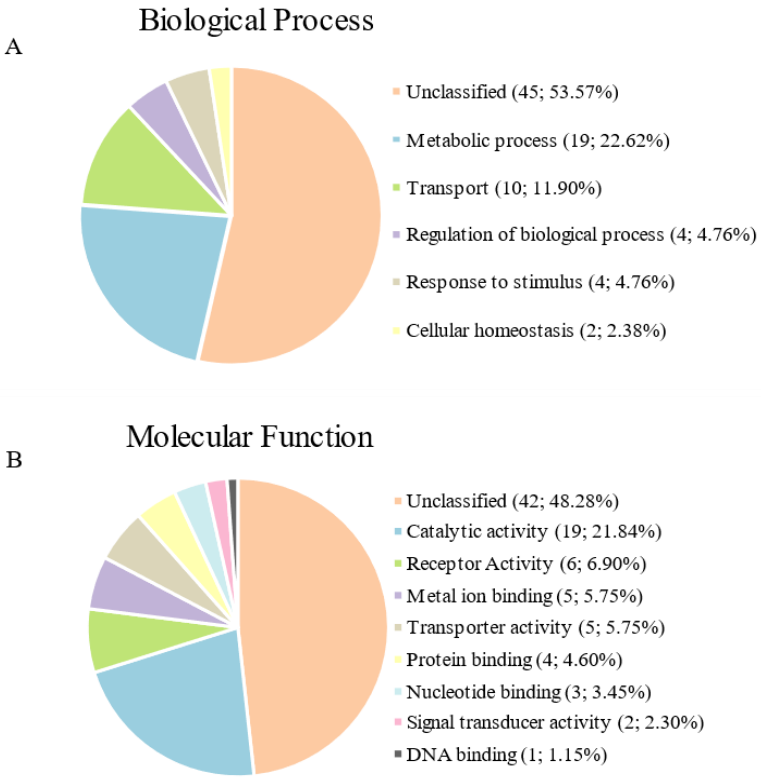

Figure 27: Distribution of the biological processes (A) and molecular functions (B) found in the overexpressed proteins in PC25% analyzed according to Proteome Discoverer.

4.2.1 DEPs possibly related with the production of SMs

The 96 DEPs identified were analyzed for their possible enrollment in the production of SMs encoded by BGCs previously identified in the genome of NL19. A total of 12 proteins were selected that included two overexpressed (**Figure 28**) and 10 underexpressed (**Figure 29**) DEPs in TSB100% when comparing to PC25%.

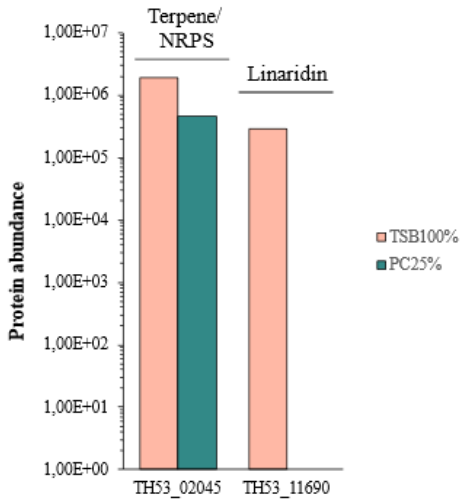

Figure 28: Abundance values of the overexpressed proteins in TSB100% encoded in a terpene/NRPS (TH53_02045) and a linaridin (TH53_11690) BGC. Each protein is represented by the locus tag number and the type of cluster.

The overexpressed proteins identified were found in a hybrid terpene/NRPS BGC (TH53_02045) and in a linaridin BGC (TH53_11690) that is a subtype of RiPPs (**Table 11**). It was not possible to predict the function of the protein encoded by TH53_02045 due to the absence of domains with known function. The protein encoded by TH53_11690 encodes an AcrB-like protein and its production in PC25% was very reduced (**Figure 28**). This protein is predicted as a component of a multidrug efflux system involved in resistance processes to a wide range of drugs [106]. Thus, this protein is not part of the core biosynthetic machinery of linaridins and can be involved in the extrusion of these or other compounds that can be toxic to the cells [106].

Table 11: Proteins involved in biosynthesis of SMs overexpressed in TSB100%. The type of cluster, locus tag of the corresponding gene and the accession number of the protein are presented. Conserved domains and molecular function are according to InterPro analysis. ND stands for not detected.

Type of cluster	Locus tag	Accession	Conserved Domains	Molecular Function
Terpene/NRPS	TH53_02045	A0A0D0GNE9	Domain of unknown function DUF892	ND
Linaridin	TH53_11690	A0A0D0FWY1	Multidrug efflux transporter AcrB TolC docking domain	transmembrane transporter activity; efflux transmembrane transporter activity

The 10 proteins that were found to be downregulated in TSB100% that can be associated to the production of SMs by NL19 are related with NRPS, lanthipeptide, siderophore and a terpene/NRP hybrid BGCs (**Figure 29** and **Table 12**).

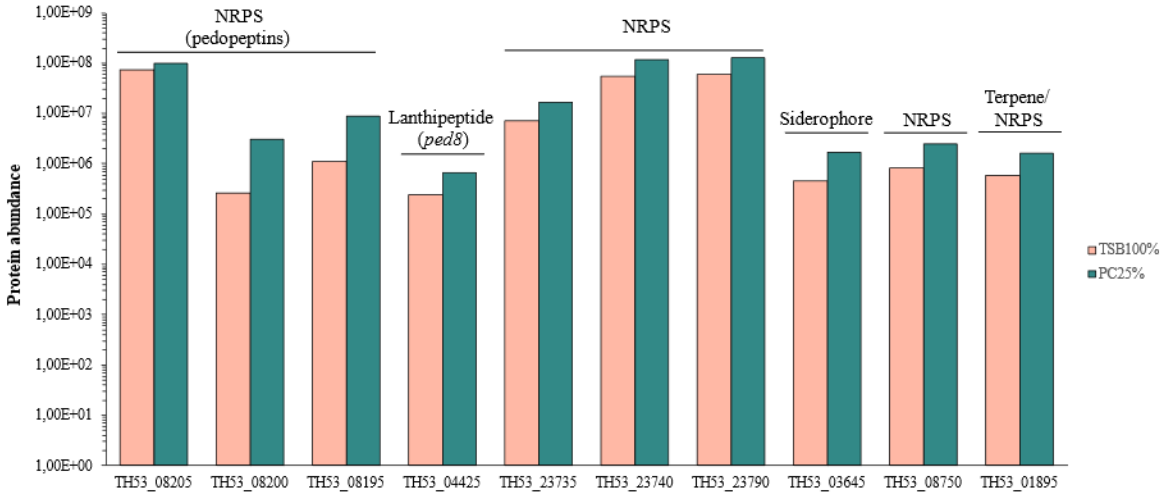

Figure 29: Abundance value of the proteins possibly involved in the biosynthesis of SMs with overexpression in PC25% medium. Each protein is represented by the locus tag number and the type of cluster.

Results confirmed that the two NRPSs responsible for the production of pedopeptins were overexpressed in PC25%, as previously described [59] (**Figure 29**; TH53_08200 and TH53_08205). Other protein encoded in the pedopeptin's operon (**Figure 25**; TH53_08195), also had their production increased in PC25% (**Figure 29**).

Regarding the lanthipeptides, PedA8.1 which is encoded in the *ped8* lanthipeptide cluster (**Figure 12**), was significatively more abundant in PC25% than in TSB100% (**Figure 29**; TH53_04425). As so, high concentrations of PC repressed the production of this putative LanA. Interestingly, taking into account all the overexpressed proteins in TSB100% or PC25%, none was linked to the biosynthetic machinery of the lanthipeptides encoded by *ped15* cluster. Consequently, this proteomic approach corroborates the RNA-seq and RT-qPCR results in the sense that PC has no interference with the transcription of *ped15* genes.

The other 6 downregulated proteins in TSB100% were found to be associated with two NRPS, one siderophore and one terpene/NRPS BGCs (**Table 12**). Such clusters, as well as the others BGCs predicted for *P. lusitanus* NL19, were identified through antiSMASH. This bioinformatic pipeline uses a built-in library of profile Hidden Markov Models (pHMMs) of signature proteins involved in the biosynthesis of specialized metabolites to detect BGC on a query sequence. When antiSMASH defines a signature gene pHMM hit, also known as core biosynthetic enzyme, the BGC is extended to either side by 5, 10 or 20 kb depending on the gene cluster type detected. This allows to implement the incorporation of tailoring enzymes. However, with this approach, it is common that genes not involved in the biosynthesis of BGCs are also included in the clusters [107,108]. Indeed, none of the proteins mentioned in the beginning of this paragraph are core biosynthetic proteins and therefore their association with the production of such SMs is dubious and should be further investigated.

Table 12: Proteins involved in the biosynthesis of SMs that were overexpressed in PC25% when compared to TSB100%. The type of cluster, locus tag of the corresponding gene and the accession number of the protein are presented. Conserved domains and molecular function are according to InterPro analysis. ND stands for not detected.

Type of cluster	Locus tag	Accession	Conserved Domains	Molecular Function
NRPS	TH53_08205	A0A0D0GT10	AMP-dependent synthetase/ligase; Condensation domain; PP binding ACP domain; NRPS synthase; AA adenylation domain; PKS-PP binding domain; AMP-binding enzyme, C-terminal domain	phosphopantetheine binding catalytic activity
	TH53_08200	A0A0D0GN55	AMP-dependent synthetase/ligase; Thioesterase; Condensation domain; PP binding ACP domain; AA adenylation domain; PKS-PP binding domain; AMP-binding enzyme, C-terminal domain	phosphopantetheine binding catalytic activity hydrolase activity, acting on ester bonds
	TH53_08195	A0A0D0FYJ6	Aminotransferase class-III family	transaminase activity; pyridoxal phosphate binding; diaminobutyrate-pyruvate transaminase activity; catalytic activity
Lanthipeptide	TH53_04425	A0A0D0F951	ND	PedA8.1; Unknown
NRPS	TH53_23735	A0A0D0EZV8	Tryptophan/Indoleamine 2,3-dioxygenase family	heme binding; tryptophan 2,3-dioxygenase activity; metal ion binding
	TH53_23740	A0A0D0FR44	Peptidase_M28	ND
	TH53_23790	A0A0D0GFJ6	CmlA, N-terminal domain;	ND
Siderophore	TH53_03645	A0A0D0F9J1	Histidine kinase	phosphorelay sensor kinase activity
NRPS	TH53_08750	A0A0D0GMW0	5-methylcytosine restriction system component	ND
Terpene/ NRPS	TH53_01895	A0A0D0FA54	Peptidase S8/S53 domain	serine-type peptidase and endopeptidase activity

4.3 Conclusions

Proteome comparison of *P. lusitanus* NL19 grown in TSB100% and PC25% media revealed that most of DEPs were involved in transport and metabolic processes as well as in catalytic activities. A group of 23 proteins had their production incremented when the concentration of PC in the broth was higher. A more detailed analysis revealed that two of those were associated with BGCs of SMs, more specifically a linaridin and a terpene/NRPS hybrid, but do not correspond to core biosynthetic proteins. On the other hand, 73 proteins were found to be overexpressed in medium with lower concentrations of PC. Some core proteins involved in the biosynthesis of SMs were encounter within this group that included the pedopeptins NRPSs, which is in accordance with previous studies, and a putative lanthipeptide precursor peptide (PedA8.1). Thus, in the future, the study of *ped8* cluster and the characterization of its peptides should be performed in PC25%. No DEPs encoded by *ped15* cluster were detected, which validates the RNA-seq and RT-qPCR results obtained in this study (chapter III) and confirms that the production of lanthipeptides encoded by *ped15* is not affect by the PC concentration. Accordingly, the media and/or additives that can allow the production of PedA15 peptides in laboratory remain unknow. Thus, it was confirmed that the decrease of the nitrogen source in media does not have the same effect in the biosynthesis of all SMs produced by NL19.

To conclude, further studies should be conducted, as it seems that, under the appropriate culture conditions, this strain might be a relevant organism to produce desired compounds.

4.4 Experimental procedures

4.4.1 Preparation of total proteins for mass spectrometry analysis

Three biological replicates of *P. lusitanus* NL19 cultures in TSB100% and PC25% media were prepared as described in section **3.4.2**. When the cultures reached an OD_{600nm} of 0.5-0.6, 1 mL was collected to microcentrifuge tubes and centrifuged at 4 ºC for 5 min at 13200 rpm. The pellet was washed with 1 mL of cold NaCl at 0.9% and centrifuged at 4 ºC for 5 min at 13200 rpm. The final pellet was stored at -80ºC.

The total protein concentration was determined for two biological replicas of PC25% with the Bradford assay. For this, firstly, the cells were lysed. Each pellet was resuspended in 300 µL of Protein Extraction Solution (100 mM Tris, pH 8.5; 1% (w/v) sodium deoxycholate; 10 mM TCEP and 40 mM 2-chloroacetamide) and then incubated at 95 ºC for 10 min. When the lysis ended, the microtubes were centrifuged at 4 ºC for 5 min at 13200 rpm. The supernatant was transferred to a

different tube and used to measure the protein concentration. The concentration obtained was within the limits required for proteome analysis (**Appendix 8**)

4.4.2 Proteome Analysis

Protein preparation, identification and quantitation was performed by nanoLC-ESI-MS/MS at i3S (Porto, Portugal). The equipment is composed by an Ultimate 3000 liquid chromatography system coupled to a Q-Exactive Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Scientific, Bremen, Germany). The analysis of proteome differences caused by PC in *P. lusitanus* NL19 was performed using Proteome Discoverer 2.4 application. The DEPs were identified with an abundance ratio TSB/PC25% ≥ 2 or ≤ 0.5, with a *p*-value ≤ 0.05. To enhance data quality, the protein false discovery rate confidence considered was at least of medium level. The classification of DEPs used in this study was based on Proteome Discoverer 2.4 software.

Chapter V – Lanthipeptides mining in *Mucilaginibacter* and *Sphingobacterium* genomes

5.1 Introduction

P. lusitanus NL19 belongs to the family *Sphingobacteriaceae*. This family of bacteria are widely distributed and have been shown to contribute to bioremediation, improvement of crop yields and production of some value-added compounds such as carotenoids [109–112]. For example, some species from *Pedobacter* genus are associated with bacterial consortia that assist in the degradation of environmental contaminants in the soil, as is the case of some phenanthrene-degraders [111]. Moreover, in this genus some strains have already demonstrated antimicrobial activity [57,113,114]. Some strains of *Mucilaginibacter* and *Sphingobacterium* genera promote efficient enzymatic degradation of cellulose for bioconversion and biofuel production [115–118]. Certainly, their SMs should play a central role in many of these processes. However, the SMs produced by *Sphingobacteriaceae* are still poorly characterized. The advent of genome mining is a milestone for the rapidly discovery of novel promising BGCs. As so, the exploitation of *Sphingobacteriaceae* genomes and its diversity, taking advantage of genomics and bioinformatic pipelines, is pivotal to uncover unique SMs that can have important biotechnological qualities within this family. We analyzed the 133 *Sphingobacteriaceae* genomes available and concluded that 64% have, at least, one BGC encoding the biosynthesis of a SM. In total, through antiSMASH and BiG-SCAPE tools, 446 BGCs were identified, distributed by all the genera of family *Sphingobacteriaceae*. The genus with the highest number of BGCs was *Pedobacter* (232 BGCs), followed by the *Mucilaginibacter* (122 BGCs) and *Sphingobacterium* (55 BGCs). If *Pedobacter* is excluded, *Mucilaginibacter* and *Sphingobacterium* genera represent 88% of all lanthipeptides BGCs detected in this family. Therefore, and due to the growing interest in these peptides, the main objectives of this chapter were to: i) characterize the lanthipeptide BGCs of genera *Mucilaginibacter* and *Sphingobacterium*, ii) characterize the diversity of lanthipeptide synthetases encoded in these BGCs and iii) characterize the diversity of lanthipeptide encoded in these BGCs.

5.2 Results and discussion

5.2.1 Class I lanthipeptides of *Mucilaginibacter* and *Sphingobacterium*

The genomes of *Mucilaginibacter* spp. and *Sphingobacterium* spp. strains contain lanthipeptide BGCs, any of which generated hits with the MIBiG (Minimum Information about a Biosynthetic Gene Cluster) database, which encompasses BGCs of known function. The *Mucilaginibacter* and *Sphingobacterium* genera encode the biosynthesis of class I and class III lanthipeptides (**Figure 30A** and **Figure 31A**).

The biosynthetic enzymes of class I lanthipeptides, LanBs and LanCs, were found in *Sphingobacterium* spp. and some *Mucilaginibacter* spp. (**Figure 30A**). Even though all BGCs are different, some genes in the vicinity of *lanB* and *lanC* are always present and include *lanAs* and *lanT$_p$s*. The last encodes the bifunctional proteins LanT$_p$ that are responsible for leader peptide removal and secretion of active and modified lanthipeptides [13]. Some clusters also have other transport-related genes encoding outer membrane proteins (OMPs) with carboxypeptidase domains and/or HlyD-like proteins, which is the component of type I secretion systems of Gram-negatives, linking the inner membrane transporters to OMPs [119] (**Figure 30A**). Moreover, all LanB and LanC clusters have, at least, one gene encoding additional putative tailoring enzymes that certainly expand the structural diversity of *Mucilaginibacter* spp. and *Sphingobacterium* spp. class I lanthipeptides. The most frequently found were glycosyltransferases, which can be involved in peptide's glycosylation [120] (**Figure 30A**). Putative LanAs were found in all clusters and have the characteristic lanthipeptides residues, except *S. mizutaii* LanAs that lack Cys (**Figure 30B**). No consensus sequence was identified in the leader peptides of LanAs but they are rich in branched-chain amino acids (BCAA) (**Figure 30B**). However, five of these LanAs have the LxLxKx$_5$L motif that was recently described by Walker *et al*. This motif is found in a family of precursor peptides (termed I6) composed only by LanAs from phylum Bacteroidetes that do not have characterized members [33]. Furthermore, all the LanA leaders possess the GG/GS/GA motif, which is the expected proteolysis substrate of the abovementioned LanT$_p$ transporters [62]. Together, these results show that *Mucilaginibacter* and *Sphingobacterium* strains should be able to produce class I and class III lanthipeptides with novel structures and possibly exciting bioactivities, especially the latter.

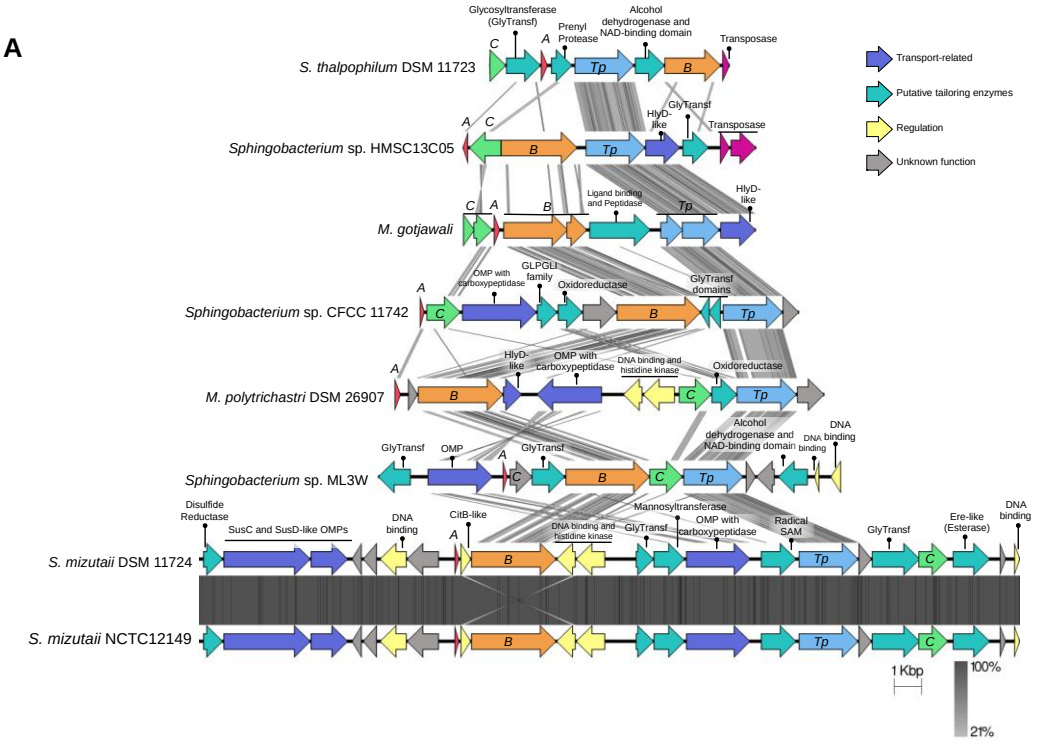

Figure 30: Tblastx analysis of class I lanthipeptide clusters found in *Sphingobacterium spp*. and *Mucilaginibacter spp*. (A) and respective amino acid sequences of the putative LanAs (precursor peptides; B) identified in each BGC.

5.2.2 The class III lanthipeptides of *Mucilaginibacter* spp.

The LanKCs identified in *Mucilaginibacter* spp. have the kinase and cyclization domains (as suggested by InterPro), but the lyase domain residues are mostly absent (**Figure 31B**). This finding raises doubts on the dehydration mechanism of *Mucilaginibacter* spp. LanKC proteins, which should be investigated in more detail. This will allow to confirm if these proteins share the same biochemistry as LanKC proteins, or if, in alternative they are novel LanKC variants. The *Mucilaginibacter* spp. BGCs with lanKC genes also have ORFs that encode peptides composed by several Ser/Thr and Cys residues, which makes them good LanA candidates (**Figure 31C**). This hypothesis is reinforced by the fact that their N-termini are rich in hydrophobic amino acids that were found to have a pivotal role in substrate recognition by LanKCs [26,28] (**Figure 31C**). Still, other features such as length and the presence of Cys only in the peptide's central region can discourage their involvement as cognate precursors (**Figure 31C**). *M. paludis* DSM 18603 has two structurally identical BGCs with a lanKC gene (**Figure 31A**). One of the lanKC genes is associated with the abovementioned LanA (**Figure 31A** and **Figure 31C**), whereas the second gene (MUCPA_RS14405) is associated with two other peptides containing four Cys, two of them involved in a motif that resembles the precursor residues of labionin (**Figure 31C**: Tx_2SxCys and $TxTx_6Cys$ versus $Sx_2Sx_{2\text{-}5}Cys$). Labionins are PTMs that have exclusively been found in class III lanthipeptides thus far [26]. Comparative genomics of the class III BGCs showed that, additionally to lanKC and lanAs, other genes in common encode (**Figure 31A**): i) a peptide of unknown function, ii) a disulphide reductase, which can participate in the formation of disulphide bridges in the LanA precursors that is another PTM found in some lanthipeptides (including class III) together with Lan/MeLan or labionins [121] and iii) a RagAB-like complex (porin and surface protein) that can function as a dynamic outer membrane importer for peptides [122]. This type of gene association is singular among class III BGCs (especially the association with a RagAB complex) and suggests that, from an ecological standpoint, lanthipeptide biosynthesis may be a more dynamic and diverse process than previously predicted.

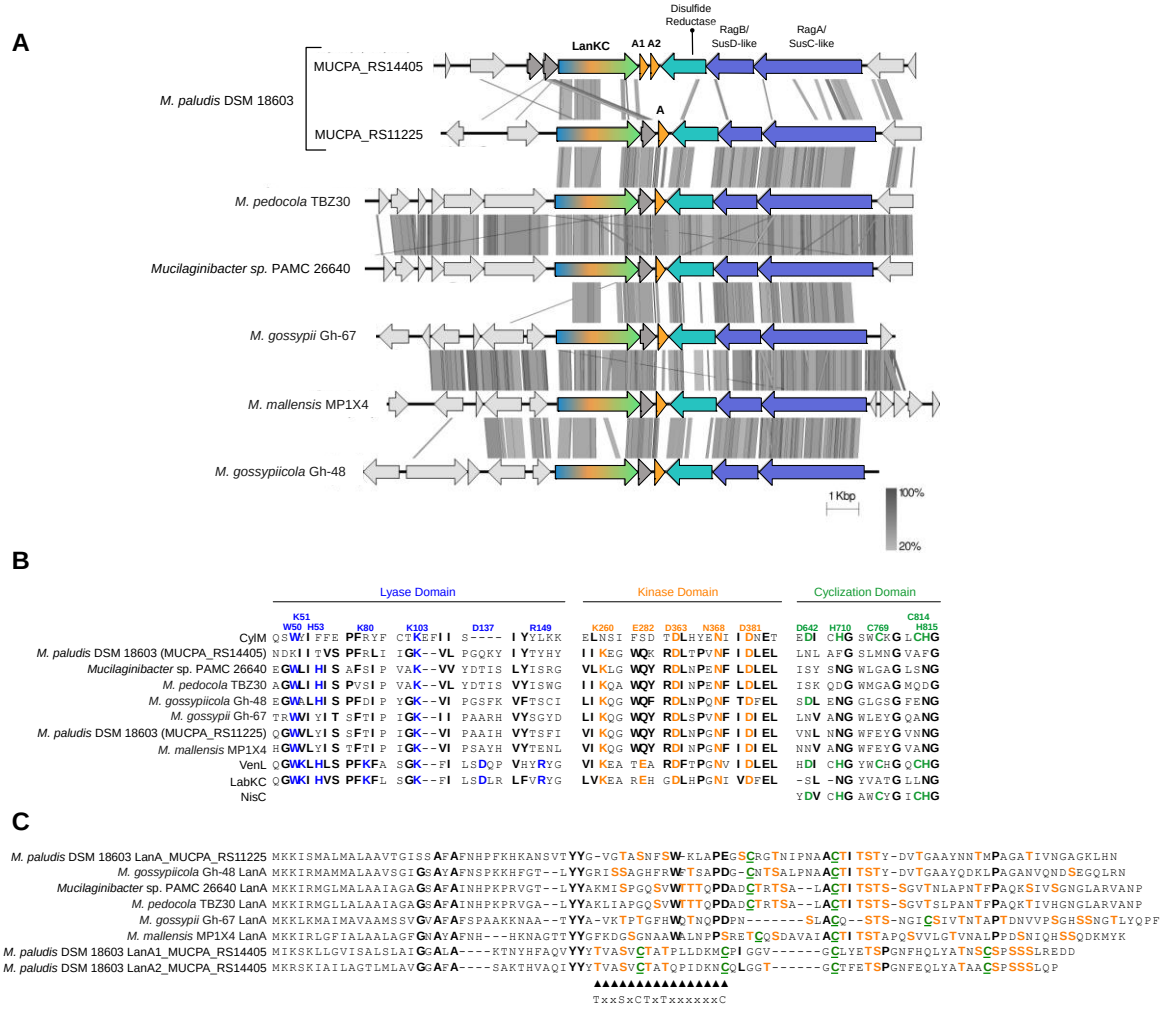

Figure 31: Tblastx analysis of the gene clusters encoding putative LanKC enzymes in *Mucilaginibacter spp.* strains (A). Alignment of lyase, kinase and cyclization domains of LanKCs from *Mucilaginibacter spp*. with CylM, VenL, LabKC and the cyclization domain of NisC (B). The catalytic residues of lyase, kinase and cyclization domains are highlighted in blue, orange and green, respectively. Conserved residues among enzymes are highlighted in bold. In C, it is shown the alignment of putative LanAs encoded in the vicinity of LanKCs, where the conserved residues were highlighted in bold. The amino acids involved in lanthionine and/or methyllanthionine are coloured in orange (Ser/Thr) and green (Cys). The two motifs that resemble labionin-formation motif ($Sx_2Sx_{2\text{-}5}Cys$) in the two peptides associated with MUCPA_RS14405 enzyme are emphasized with triangles.

5.3 Conclusions

The increasing amount of DNA sequence data and the use of pipelines, such as antiSMASH and BiG-SCAPE, have led to an exponential growing interest in genome mining for specialized metabolites, like the lanthipeptides.

Herein, we concluded that some *Mucilaginibacter* and *Sphingobacterium* strains encode class I and III lanthipeptides. Interestingly, the LanKC-like proteins, found in some *Mucilaginibacter* genomes, with different lyase domains can represent a novel mechanism of dehydration by class III lanthipeptide synthetases. Therefore, the resulting lanthipeptides might have unprecedented PTMs. Since class III lanthipeptides previously showed antiallodynic and antiviral activities, further studies should be conducted to characterize, structurally and functionally, the lanthipeptides identified in this study.

5.4 Experimental procedures

5.4.1 *Sphingobacteriaceae*'s BGCs database

A total of 446 BGCs from 85 genomes belonging to the family *Sphingobacteriaceae* were downloaded from the publicly available antiSMASH database v.4.2.1 (accessed in 15/10/2019) [123]. Draft or complete genomes of *Sphingobacteriaceae* strains can be accessed at the NCBI microbial genome database (accessed on 15/10/2019). The homology between *Sphingobacteriaceae*'s BGCs was analyzed with BiG-SCAPE [124,125].

5.4.2 Analysis of the lanthipeptides BGCs from *Mucilaginibacter* and *Sphingobacterium* genera

The network files produced by BiG-SCAPE were used to further investigate the BGCs encoding the biosynthesis of lanthipeptides in *Mucilaginibacter* and *Sphingobacterium* genera. The BGCs final annotation was performed with Artemis [126] and the functional analysis of the encoded ORFs was performed with InterPro and the NCBI Conserved Domain Database. Comparison between lanthipeptides' BCGs was performed by tblastx analysis with Easyfig [127]. Protein sequences of characterized dehydratases and cyclases were obtained in UniProt database. Putative *lanA* genes were searched in NCBI's ORFfinder. Amino acid sequence alignment was performed in CLC sequence viewer.

Chapter VI - Synopsis and Future perspectives

Extreme environments can be an endless source of microorganisms with high biotechnological potential. *P. lusitanus* NL19 was isolated from one such environment and its genome analysis revealed the presence of several BGCs including NRPS and lanthipeptides.

In fact, NL19 can produce bioactive compounds such as pedopeptins. These peptides are biosynthesized by a NRPS cluster whose transcription is highly affected by high concentrations of PC. Herein, the objective was to evaluate the effect of PC in i) the transcription of *ped15*, a cluster that encodes the essential genes for the biosynthesis of putative lanthipeptides and ii) the proteome of NL19.

Firstly, the upstream region of *ped15* cluster was sequenced, through primer walking, and its analysis uncovered six new *pedA15* genes. The alignment of these precursor peptides sequences shed light on a novel hybrid class I/II of lanthipeptides with extraordinary and exciting new characteristics.

The RT-qPCR results obtained revealed that, unlike pedopeptins, the transcription of *ped15* genes is not affected in cells cultivated in high concentrations of PC. In general, the results were consistent with those obtained in a previous RNA-seq experiment and in the proteomic analysis. Moreover, the gene expression levels of *ped15* cluster were always low, suggesting that this lanthipeptide BGC is cryptic. Therefore, the structure of PedA15 peptides and their bioactivity is still unknown. Furthermore, the acknowledgment of binding sites for transcriptional repressors in the predicted promoter region of *pedA15* genes as well as the fact that DNA binding response regulator is, in average, the most expressed gene of the *ped15* cluster, might indicate its function as a repressor.

Proteome results showed that a total of 12 proteins associated with the production of SMs were identified as DEPs in the media with different concentration of PC. Nonetheless, only 3 of them were considered core biosynthetic enzymes: two proteins responsible for the production of the pedopeptins and a LanA encoded by cluster *ped8*. The exact involvement of the other proteins in the biosynthesis of the respective SMs is yet unknown and a more in-depth characterization should be done.

Finally, in the chapter not devoted to the study on the influence of PC in the production of SMs by NL19, the *in silico* analysis of 446 genomes of the family *Sphingobacteriaceae* revealed the presence of class I and class III lanthipeptide BGCs in the genera *Mucilaginibacter* and *Sphingobacterium*. Interestingly, the LanKC enzymes encoded in the class III BGCs of *Mucilaginibacter* spp. lack the typical lyase domain residues suggesting a different mechanism for Lan/MeLan installation.

Considering all the results obtained, it is clear that some aspects need further investigation. Therefore, studies and laboratorial analysis are required to:

i) Increase the levels of lanthipeptides transcription/production in NL19 using different medium, such as MYC.

ii) Assess the full potential of *ped15* cluster as responsible for the production of class I/II hybrid lanthipeptides with unique characteristics, mainly by characterization of PedA15 structure and bioactivity.

iii) Understand the exact role of DNA binding response regulator in the regulation of PedA15 biosynthesis.

iv) Evaluate the effects of PC concentration in the transcription of others BGCs encoded in NL19 genome, such as the lanthipeptide *ped8* cluster.

v) Characterize the potential lanthipeptides produced by class III BGCs of *Mucilaginibacter* and their respective LanKC enzymes.

Outputs of this dissertation

7.1 ISI Papers

Figueiredo, G.*****, **Gomes, M.***, Covas, C., Mendo S. & Caetano T. (Submitted) The unexplored wealth of environmental secondary metabolites: the *Sphingobacteriaceae* case study. *Microbial Ecology*. ***Both first authors**

Covas, C., Figueiredo, G., **Gomes, M**., Santos T., Mendo, S. & Caetano, T. (Under Review) The pangenome of environmental Gram-negative bacteria and its hidden, but promising, biotechnological potential. *Molecular Ecology*.

7.2 Outreach activities

Participation as assistant monitor in the workshop "Técnicas de cultura de procariotas (Bacteria e Archaea) e avaliação de características fenotípicas" at Science Week of the Secondary School Adolfo Portela, Águeda, Portugal. 6th March 2020

References

1. Dias, D. A., Urban, S. & Roessner, U. A Historical overview of natural products in drug discovery. *Metabolites* **2**, 303–336 (2012).
2. O'Connor, S. E. Engineering of Secondary Metabolism. *Annu. Rev. Genet.* **49**, 71–94 (2015).
3. Challinor, V. L. & Bode, H. B. Bioactive natural products from novel microbial sources. *Ann. N. Y. Acad. Sci.* **1354**, 82–97 (2015).
4. Jenke-Kodama, H., Müller, R. & Dittmann, E. Evolutionary mechanisms underlying secondary metabolite diversity. *Progress in Drug Research* **65**, 120–140 (2008).
5. Watve, M. G., Tickoo, R., Jog, M. M. & Bhole, B. D. How many antibiotics are produced by the genus *Streptomyces*? *Arch. Microbiol.* **176**, 386–390 (2001).
6. Cragg, G. M. & Newman, D. J. Natural products: A continuing source of novel drug leads. *Biochim. Biophys. Acta - Gen. Subj.* **1830**, 3670–3695 (2013).
7. Patrick, F. Antibiotic compounds from *Bacillus*: Why are they so amazing? *Am. J. Biochem. Biotechnol.* **8**, 40–46 (2012).
8. Weissman, K. J. & Müller, R. Myxobacterial secondary metabolites: Bioactivities and modes-of-action. *Natural Product Reports* **27**, 1276–1295 (2010).
9. Arnison, P. G. *et al.* Ribosomally synthesized and post-translationally modified peptide natural products: Overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* **30**, 108–160 (2013).
10. McIntosh, J. A., Donia, M. S. & Schmidt, E. W. Ribosomal peptide natural products: Bridging the ribosomal and nonribosomal worlds. *Nat. Prod. Rep.* **26**, 537–559 (2009).
11. Knerr, P. J. & van der Donk, W. A. Discovery, Biosynthesis, and Engineering of Lantipeptides. *Annu. Rev. Biochem.* **81**, 479–505 (2012).
12. Barbosa, J., Caetano, T. & Mendo, S. Class I and Class II Lanthipeptides Produced by *Bacillus* spp. *J. Nat. Prod.* **78**, 2850–2866 (2015).
13. Repka, L. M., Chekan, J. R., Nair, S. K. & Van Der Donk, W. A. Mechanistic Understanding of Lanthipeptide Biosynthetic Enzymes. *Chem. Rev.* **117**, 5457–5520 (2017).
14. Zhang, Q., Doroghazi, J. R., Zhao, X., Walker, M. C. & van der Donka, W. A. Expanded natural product diversity revealed by analysis of lanthipeptide-like gene clusters in Actinobacteria. *Appl. Environ. Microbiol.* **81**, 4339–4350 (2015).
15. Chatterjee, C., Paul, M., Xie, L. & van der Donk, W. A. Biosynthesis and mode of action of lantibiotics. *Chem. Rev.* **105**, 633–683 (2005).
16. Sheng, W. *et al.* Substrate tolerance of the biosynthetic enzymes of glycosylated lanthipeptide NAI-112. *Org. Biomol. Chem.* **18**, 6095–6099 (2020).
17. Dischinger, J., Basi Chipalu, S. & Bierbaum, G. Lantibiotics: Promising candidates for

future applications in health care. *Int. J. Med. Microbiol.* **304**, 51–62 (2014).

18. Willey, J. M. & van der Donk, W. A. Lantibiotics: Peptides of Diverse Structure and Function. *Annu. Rev. Microbiol.* **61**, 477–501 (2007).
19. Rogers, L. A. the Inhibiting Effect of *Streptococcus Lactis* on *Lactobacillus Bulgaricus*. *J. Bacteriol.* **16**, 321–325 (1928).
20. Lubelski, J., Rink, R., Khusainov, R., Moll, G. N. & Kuipers, O. P. Biosynthesis, immunity, regulation, mode of action and engineering of the model lantibiotic nisin. *Cellular and Molecular Life Sciences* **65**, 455–476 (2008).
21. Ortega, M. A. *et al.* Structure and mechanism of the tRNA-dependent lantibiotic dehydratase NisB. *Nature* **517**, 509–512 (2015).
22. Garg, N., Salazar-Ocampo, L. M. A. & Van Der Donk, W. A. *In vitro* activity of the nisin dehydratase NisB. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 7258–7263 (2013).
23. Bothwell, I. R. *et al.* Characterization of glutamyl-tRNA–dependent dehydratases using nonreactive substrate mimics. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 17245–17250 (2019).
24. Ortega, M. A. *et al.* Structure and tRNA Specificity of MibB, a Lantibiotic Dehydratase from Actinobacteria Involved in NAI-107 Biosynthesis. *Cell Chem. Biol.* **23**, 370–380 (2016).
25. Okeley, N. M., Paul, M., Stasser, J. P., Blackburn, N. & Van Der Donk, W. A. SpaC and NisC, the Cyclases Involved in Subtilin and Nisin Biosynthesis, Are Zinc Proteins. *Biochemistry* **42**, 13613–13624 (2003).
26. Hegemann, J. D. & Süssmuth, R. D. Matters of class: coming of age of class III and IV lanthipeptides. *RSC Chem. Biol.* **1**, 110–127 (2020).
27. Zhang, Q., Yu, Y., Vélasquez, J. E. & Van Der Donk, W. A. Evolution of lanthipeptide synthetases. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 18361–18366 (2012).
28. Müller, W. M., Ensle, P., Krawczyk, B. & Süssmuth, R. D. Leader peptide-directed processing of labyrinthopeptin A2 precursor peptide by the modifying enzyme LabKC. *Biochemistry* **50**, 8362–8373 (2011).
29. Iorio, M. *et al.* A Glycosylated, Labionin-Containing Lanthipeptide with Marked Antinociceptive Activity. *ACS Chem. Biol* **9**, 30 (2014).
30. Chen, S. *et al.* Zn-dependent bifunctional proteases are responsible for leader peptide processing of class III lanthipeptides. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 2533–2538 (2019).
31. Goto, Y. *et al.* Discovery of unique lanthionine synthetases reveals new mechanistic and evolutionary insights. *PLoS Biol.* **8**, 1000339 (2010).
32. Mohr, K. I. *et al.* Pinensins: The First Antifungal Lantibiotics. *Angew. Chemie - Int. Ed.* **54**,

11254–11258 (2015).

33. Walker, M. C. *et al.* Precursor peptide-targeted mining of more than one hundred thousand genomes expands the lanthipeptide natural product family. *BMC Genomics* **21**, 1–17 (2020).
34. Caetano, T. Lichenicidin biosynthesis and search for novel antibacterial peptides. (Universidade de Aveiro, 2011).
35. Chandrapati, S. & O'Sullivan, D. J. Characterization of the promoter regions involved in galactose- and nisin-mediated induction of the *nisA* gene in *Lactococcus lactis* ATCC 11454. *Mol. Microbiol.* **46**, 467–477 (2002).
36. Meindl, K. *et al.* Labyrinthopeptins: A new class of carbacyclic lantibiotics. *Angew. Chemie - Int. Ed.* **49**, 1151–1154 (2010).
37. Stein, T., Heinzmann, S., Solovieva, I. & Entian, K.-D. Function of *Lactococcus lactis* Nisin Immunity Genes *nisI* and *nisFEG* after Coordinated Expression in the Surrogate Host *Bacillus subtilis*. *J. Biol. Chem.* **278**, 89–94 (2002).
38. Field, D., Cotter, P. D., Hill, C. & Ross, R. P. Bioengineering lantibiotics for therapeutic success. *Front. Microbiol.* **6**, 1–8 (2015).
39. Denny, C. B., Sharpe, L. E. & Bohrer, C. W. Effects of tylosin and nisin on canned food spoilage bacteria. *Appl. Microbiol.* **9**, 108–110 (1961).
40. Shin, J. M. *et al.* Biomedical applications of nisin. *Journal of Applied Microbiology* **120**, 1449–1465 (2016).
41. Castiglione, F. *et al.* Determining the Structure and Mode of Action of Microbisporicin, a Potent Lantibiotic Active Against Multiresistant Pathogens. *Chem. Biol.* **15**, 22–31 (2008).
42. WHO publishes list of bacteria for which new antibiotics are urgently needed. Available at: https://www.who.int/news/item/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed. (Accessed: 11th January 2021)
43. Boakes, S., Appleyard, A. N., Cortés, J. & Dawson, M. J. Organization of the biosynthetic genes encoding deoxyactagardine B (DAB), a new lantibiotic produced by *Actinoplanes liguriae* NCIMB41362. *J. Antibiot. (Tokyo).* **63**, 351–358 (2010).
44. Férir, G. *et al.* The Lantibiotic Peptide Labyrinthopeptin A1 Demonstrates Broad Anti-HIV and Anti-HSV Activity with Potential for Microbicidal Applications. *PLoS One* **8**, (2013).
45. Geng, M. & Smith, L. Improving the attrition rate of Lanthipeptide discovery for commercial applications. *Expert Opinion on Drug Discovery* **13**, 155–167 (2018).
46. Moll, G. N., Kuipers, A., De Vries, L., Bosma, T. & Rink, R. A biological stabilization technology for peptide drugs: Enzymatic introduction of thioether-bridges. *Drug Discov. Today Technol.* **6**, e13–e18 (2009).
47. Kluskens, L. D. *et al.* Angiotensin-(1-7) with thioether bridge: An angiotensin- converting

enzyme-resistant, potent angiotensin-(1-7) analog. *J. Pharmacol. Exp. Ther.* **328**, 849–855 (2009).

48. Wagenaar, G. T. M. *et al.* Agonists of MAS oncogene and angiotensin II type 2 receptors attenuate cardiopulmonary disease in rats with neonatal hyperoxia-induced lung injury. *Am. J. Physiol. - Lung Cell. Mol. Physiol.* **305**, L341 (2013).
49. OG716 :: Oragenics, Inc. (OGEN). Available at: https://www.oragenics.com/technology-pipeline/lantibiotics/og716. (Accessed: 27th December 2019)
50. Crowther, G. S. *et al.* Evaluation of NVB302 versus vancomycin activity *in* an *in vitro* human gut model of *Clostridium difficile* infection. *J. Antimicrob. Chemother.* **68**, 168–176 (2013).
51. Chatterjee, S. *et al.* Mersacidin, a new antibiotic from *Bacillus* fermentation, isolation, purification and chemical characterization. *J. Antibiot. (Tokyo).* **45**, 832–838 (1992).
52. First in Human Single Ascending Dose Study of MOR107 - Full Text View - ClinicalTrials.gov. Available at: https://clinicaltrials.gov/ct2/show/study/NCT03067363. (Accessed: 20th October 2020)
53. Rew, Y. *et al.* Synthesis and biological activities of cyclic lanthionine enkephalin analogues: δ-opioid receptor selective ligands. *J. Med. Chem.* **45**, 3746–3754 (2002).
54. Steyn, P. L. *et al.* Classification of heparinolytic bacteria into a new genus, *Pedobacter*, comprising four species: *Pedobacter heparinus* comb. nov., *Pedobacter piscium* comb. nov., *Pedobacter africanus* sp. nov. and *Pedobacter saltans* sp. nov. proposal of the family *Sphingobacteriaceae* fam. nov.. *Int. J. Syst. Bacteriol.* **48**, 165–177 (1998).
55. Covas, C. *et al. Pedobacter lusitanus* sp. nov., isolated from sludge of a deactivated uranium mine. *Int. J. Syst. Evol. Microbiol.* **67**, 1339–1348 (2017).
56. Viana, A. T., Caetano, T., Covas, C., Santos, T. & Mendo, S. Environmental superbugs: The case study of *Pedobacter* spp. *Environ. Pollut.* **241**, 1048–1055 (2018).
57. Santos, T., Cruz, A., Caetano, T., Covas, C. & Mendo, S. Draft genome sequence of *Pedobacter* sp. strain NL19, a producer of potent antibacterial compounds. *Genome Announc.* **3**, e00184-15 (2016).
58. Barrios-González, J. Solid-state fermentation: Physiology of solid medium, its molecular basis and applications. *Process Biochemistry* **47**, 175–185 (2012).
59. Covas, C. *et al.* Peptone from casein, an antagonist of nonribosomal peptide synthesis: a case study of pedopeptins produced by *Pedobacter lusitanus* NL19. *N. Biotechnol.* **60**, 62–71 (2021).
60. Tang, W. & Van Der Donk, W. A. Structural characterization of four prochlorosins: A novel class of lantipeptides produced by planktonic marine cyanobacteria. *Biochemistry* **51**,

4271–4279 (2012).

61. Li, B. *et al.* Catalytic promiscuity in the biosynthesis of cyclic peptide secondary metabolites in planktonic marine cyanobacteria. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 10430–10435 (2010).
62. Caetano, T., van der Donk, W. & Mendo, S. Bacteroidetes can be a rich source of novel lanthipeptides: The case study of *Pedobacter lusitanus*. *Microbiol. Res.* **235**, 126441 (2020).
63. Gilbert, W. & Maxam, A. The nucleotide sequence of the *lac* operator. *Proc. Natl. Acad. Sci. U. S. A.* **70**, 3581–3584 (1973).
64. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467 (1977).
65. Pareek, C. S., Smoczynski, R. & Tretyn, A. Sequencing technologies and genome sequencing. *Journal of Applied Genetics* **52**, 413–435 (2011).
66. Garrido-Cardenas, J. A., Garcia-Maroto, F., Alvarez-Bermejo, J. A. & Manzano-Agugliaro, F. DNA sequencing sensors: An overview. *Sensors* **17**, 1–15 (2017).
67. Thermes, C. Ten years of next-generation sequencing technology. *Trends in genetics : TIG* **30**, 418–426 (2014).
68. Lowe, R., Shirley, N., Bleackley, M., Dolan, S. & Shafee, T. Transcriptomics technologies. *PLoS Comput. Biol.* **13**, 1–23 (2017).
69. Sim, G. K. *et al.* Use of a cDNA library for studies on evolution and developmental expression of the chorion multigene families. *Cell* **18**, 1303–1316 (1979).
70. Hu, M. & Polyak, K. Serial analysis of gene expression. *Nat. Protoc.* **1**, 1743–1760 (2006).
71. Arya, M. *et al.* Basic principles of real-time quantitative PCR. *Expert Review of Molecular Diagnostics* **5**, 209–219 (2005).
72. Jozefczuk, J. & Adjaye, J. *Quantitative real-time PCR-based analysis of gene expression*. *Methods in Enzymology* **500**, (Elsevier Inc., 2011).
73. Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**, 1–12 (2002).
74. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10**, 57–63 (2009).
75. Garalde, D. R. *et al.* Highly parallel direct RNA sequencing on an array of nanopores. *Nat. Methods* **15**, 201–206 (2018).
76. Fang, Z. & Cui, X. Design and validation issues in RNA-seq experiments. *Brief. Bioinform.* **12**, 280–287 (2011).
77. Singhal, N., Kumar, M., Kanaujia, P. K. & Virdi, J. S. MALDI-TOF mass spectrometry: An emerging technology for microbial identification and diagnosis. *Frontiers in Microbiology*

6, (2015).

78. Mallick, P. & Kuster, B. Proteomics: A pragmatic perspective. *Nat. Biotechnol.* **28**, 695–709 (2010).
79. Domon, B. & Aebersold, R. Mass spectrometry and protein analysis. *Science.* **312**, 212–217 (2006).
80. Han, X., Aslanian, A. & Yates, J. R. Mass spectrometry for proteomics. *Curr. Opin. Chem. Biol.* **12**, 483–490 (2008).
81. Yates, J. R., Ruse, C. I. & Nakorchevsky, A. Proteomics by Mass Spectrometry: Approaches, Advances, and Applications. *Annu. Rev. Biomed. Eng.* **11**, 49–79 (2009).
82. Domingues, P., García, A. & Skrzydlewska, E. *AACLifeSci Course Companion Manual*. (2018).
83. Silva, A. M. N., Vitorino, R., Domingues, M. R. M., Spickett, C. M. & Domingues, P. Post-translational modifications and mass spectrometry detection. *Free Radic. Biol. Med.* **65**, 925–941 (2013).
84. Palazzotto, E. & Weber, T. Omics and multi-omics approaches to study the biosynthesis of secondary metabolites in microorganisms. *Current Opinion in Microbiology* **45**, 109–116 (2018).
85. Blin, K. *et al.* antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* **47**, W81–W87 (2019).
86. Toleman, M. A., Bennett, D. M., Jones, R. N. & Walsh, T. R. Multi-Resistance Genomic Islands May Be Present in Metallo- § -Lactamase Producing *Pseudomonas aeruginosa* Isolates from Italian Hospitals : Report from the SENTRY Antimicrobial Surveillance Program. 10117 (2004).
87. Abts, A., Montalban-Lopez, M., Kuipers, O. P., Smits, S. H. & Schmitt, L. NisC binds the FxLx motif of the nisin leader peptide. *Biochemistry* **52**, 5387–5395 (2013).
88. Plat, A., Kluskens, L. D., Kuipers, A., Rink, R. & Moll, G. N. Requirements of the Engineered Leader Peptide of Nisin for Inducing Modification, Export, and Cleavage †. *Appl. Environ. Microbiol.* **77**, 604–611 (2011).
89. Cubillos-Ruiz, A., Berta-Thompson, J. W., Becker, J. W., Van Der Donk, W. A. & Chisholm, S. W. Evolutionary radiation of lanthipeptides in marine cyanobacteria. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E5424–E5433 (2017).
90. Nikolskaya, A. N. & Galperin, M. Y. A novel type of conserved DNA-binding domain in the transcriptional regulators of the AlgR/AgrA/LytR family. *Nucleic Acids Res.* **30**, 2453–2459 (2002).
91. West, A. H. & Stock, A. M. Histidine kinases and response regulator proteins in two-

component signaling systems. *Trends in Biochemical Sciences* **26**, 369–376 (2001).

92. Stock, A. M., Robinson, V. L. & Goudreau, P. N. Two-Component Signal Transduction. *Annu. Rev. Biochem.* **69**, 183–215 (2000).
93. Mortensen, L., Dandanell, G. & Hammer, K. Purification and characterization of the *deoR* repressor of *Escherichia coli*. *EMBO J.* **8**, 325–331 (1989).
94. Zeng, X., Saxild, H. H. & Switzer, R. L. Purification and characterization of the DeoR repressor of *Bacillus subtilis*. *J. Bacteriol.* **182**, 1916–1922 (2000).
95. Butala, M., Žgur-Bertok, D. & Busby, S. J. W. The bacterial LexA transcriptional repressor. *Cell. Mol. Life Sci.* **66**, 82–93 (2009).
96. Untergasser, A. *et al.* Primer3-new capabilities and interfaces. *Nucleic Acids Res.* **40**, (2012).
97. Gupta, A. *et al.* Global Awakening of Cryptic Biosynthetic Gene Clusters in *Burkholderia thailandensis*. *ACS Chem. Biol.* **12**, 3012–3021 (2017).
98. Amos, G. C. A. *et al.* Comparative transcriptomics as a guide to natural product discovery and biosynthetic gene cluster functionality. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E11121–E11130 (2017).
99. Camarena, L., Bruno, V., Euskirchen, G., Poggio, S. & Snyder, M. Molecular mechanisms of ethanol-induced pathogenesis revealed by RNA-sequencing. *PLoS Pathog.* **6**, 1–14 (2010).
100. Schurch, N. J. *et al.* How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA* **22**, 839–851 (2016).
101. Tarazona, S., García-Alcalde, F., Dopazo, J., Ferrer, A. & Conesa, A. Differential expression in RNA-seq: A matter of depth. *Genome Res.* **21**, 2213–2223 (2011).
102. Chen, I. P. & Michel, H. Cloning, sequencing, and characterization of the *recA* gene from *Rhodopseudomonas viridis* and construction of a *recA* strain. *J. Bacteriol.* **180**, 3227–3232 (1998).
103. López, R., Trueba, P., Lindavista, C., 24090, C. P. & Monteón, V. Influence of different organic nitrogen sources on antimicrobial activity of *Pseudoalteromonas* sp. *BioMedRx An Int. J.* **1**, 952–956 (2013).
104. Martín, J. F., Rodríguez-García, A. & Liras, P. The master regulator PhoP coordinates phosphate and nitrogen metabolism, respiration, cell differentiation and antibiotic biosynthesis: Comparison in *Streptomyces coelicolor* and *Streptomyces avermitilis*. *Journal of Antibiotics* **70**, 534–541 (2017).
105. Kurbanoglu, E. B., Ozdal, M., Ozdal, O. G. & Algur, O. F. Enhanced production of prodigiosin by *Serratia marcescens* MO-1 using ram horn peptone. *Brazilian J. Microbiol.*

46, 631–637 (2015).

106. Sennhauser, G., Amstutz, P., Briand, C., Storchenegger, O. & Grütter, M. G. Drug Export Pathway of Multidrug Exporter AcrB Revealed by DARPin Inhibitors. *PLoS Biol.* **5**, e7 (2006).
107. Chavali, A. K. & Rhee, S. Y. Bioinformatics tools for the identification of gene clusters that biosynthesize specialized metabolites. *Briefings in bioinformatics* **19**, 1022–1034 (2018).
108. Medema, M. H. *et al.* AntiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* **39**, W339 (2011).
109. Asker, D., Beppu, T. & Ueda, K. *Nubsella zeaxanthinifaciens* gen. nov., sp. nov., a zeaxanthin-producing bacterium of the family *Sphingobacteriaceae* isolated from freshwater. *Int. J. Syst. Evol. Microbiol.* **58**, 601–606 (2008).
110. Chen, W. M., Chen, Y. L. & Sheu, S. Y. *Mucilaginibacter roseus* sp. nov., isolated from a freshwater river. *Int. J. Syst. Evol. Microbiol.* **66**, 1112–1118 (2016).
111. Li, J. *et al.* Diversity of the active phenanthrene degraders in PAH-polluted soil is shaped by ryegrass rhizosphere and root exudates. *Soil Biol. Biochem.* **128**, 100–110 (2019).
112. Fan, D., Subramanian, S. & Smith, D. L. Plant endophytes promote growth and alleviate salt stress in *Arabidopsis thaliana*. *Sci. Rep.* **10**, 1–18 (2020).
113. Song, Y. S., Seo, D. J. & Jung, W. J. Identification, purification, and expression patterns of chitinase from psychrotolerant *Pedobacter* sp. PR-M6 and antifungal activity in vitro. *Microb. Pathog.* **107**, 62–68 (2017).
114. Bitzer, A. S., Garbeva, P. & Silby, M. W. Draft Genome Sequence of *Pedobacter* sp. Strain V48, Isolated from a Coastal Sand Dune in the Netherlands. *Genome Announc.* **2**, (2014).
115. Janbandhu, A. & Fulekar, M. H. Biodegradation of phenanthrene using adapted microbial consortium isolated from petrochemical contaminated environment. *J. Hazard. Mater.* **187**, 333–340 (2011).
116. Zheng, G. *et al.* Degradation of rice straw at low temperature using a novel microbial consortium LTF-27 with efficient ability. *Bioresour. Technol.* **304**, 123064 (2020).
117. Jiménez, D. J., Dini-Andreote, F. & Van Elsas, J. D. Metataxonomic profiling and prediction of functional behaviour of wheat straw degrading microbial consortia. *Biotechnol. Biofuels* **7**, 92 (2014).
118. Wang, Z. Y., Wang, R. X., Zhou, J. S., Cheng, J. F. & Li, Y. H. An assessment of the genomics, comparative genomics and cellulose degradation potential of *Mucilaginibacter polytrichastri* strain RG4-7. *Bioresour. Technol.* **297**, 122389 (2020).
119. Kim, J.-S. *et al.* Crystal Structure of a Soluble Fragment of the Membrane Fusion Protein

HlyD in a Type I Secretion System of Gram-Negative Bacteria. *Structure* **24**, 477–485 (2016).

120. Breton, C., Šnajdrová, L., Jeanneau, C., Koča, J. & Imberty, A. Structures and mechanisms of glycosyltransferases. *Glycobiology* **16**, 29–37 (2006).
121. Wang, J. *et al.* Bovicin HJ50-Like Lantibiotics, a Novel Subgroup of Lantibiotics Featured by an Indispensable Disulfide Bridge. *PLoS One* **9**, e97121 (2014).
122. Madej, M. *et al.* Structural and functional insights into oligopeptide acquisition by the RagAB transporter from *Porphyromonas gingivalis*. *Nat. Microbiol.* **5**, 1016–1025 (2020).
123. Blin, K. *et al.* The antiSMASH database version 2: a comprehensive resource on secondary metabolite biosynthetic gene clusters. *Nucleic Acids Res.* **47**, 625–630 (2018).
124. Yeong, B. M. BiG-SCAPE. (2016).
125. Navarro-Muñoz, J. C. *et al.* A computational framework to explore large-scale biosynthetic diversity. *Nat. Chem. Biol.* **16**, 60–68 (2020).
126. Rutherford, K. *et al.* Artemis: Sequence visualization and annotation. *Bioinformatics* **16**, 944–945 (2000).
127. Sullivan, M. J., Petty, N. K. & Beatson, S. A. Easyfig: A genome comparison visualizer. *Bioinformatics* **27**, 1009–1010 (2011).

Appendices

Appendix 1

Genomic DNA extraction

In *P. lusitanus* NL19, the extraction of genomic DNA was performed using Wizard® Genomic DNA Purification Kit (Promega). 2 mL of overnight culture were used for genomic DNA extraction. The cells were harvested for 2 min at 13,000 x g and the supernatant was discarded. The pellet was resuspended in 600 µL of Nuclei Lysis Solution in the 2 mL microcentrifuge tube. The tube was incubated for 5 min at 80 °C, then cool to room temperature. To this mixture, 3 µL of RNase Solution was added. The sample was gently mixture and incubated at 37 °C for 60 min, then cooled to room temperature. Afterwards, 200 µL of Protein Precipitation Solution were added and mixed by vortexing. The sample was incubated on ice for 5 min and then centrifuged at 13,000 x g for 3 min. The supernatant was transferred to a clean tube containing 600 µL of isopropanol at room temperature. After mixing, the sample was centrifuge for 2 min at 13,000 x g and the supernatant was decanted. 600 µL of room temperature 70% ethanol were added and mixed. A centrifugation of 2 min at 13,000 x g took place and after which the ethanol was aspirated and the pellet was air-dried for 15 min. The DNA pellet was rehydrated in 100 µL of Rehydration Solution and stayed at 4 °C overnight.

Appendix 2

Purification of PCR products and DNA digestions

The volume of the PCR reaction mixture was transferred to a 1.5 mL microcentrifuge tube and five volumes of Binding Buffer were added, mixed well, and centrifuged briefly. Then, the mixture was applied to an NZYTech spin column incubated at room temperature for 2 min and centrifuged for 1 min. The flow-through was discarded and 600 μL of Wash Buffer were added to the spin column. The column was centrifuged for 1 min and the flow-through was discarded. An additional centrifugation was performed for 1 min to remove residual ethanol. The NZYTech spin column was placed into a clean 1.5 mL microcentrifuge tube and 50 μL of sterile distilled water were added to the center of the column followed by an incubation of 1 min at room temperature. Lastly, the column was centrifuged for 1 min to elute the DNA and the purified DNA was stored at -20 °C until further use.

Purification of plasmid DNA

Briefly, 2 mL of bacterial culture were centrifuged for 30 sec. The supernatant was discarded, and the remaining cell pellet was resuspended in 250 μL of Buffer A1 by vigorous vortexing. 250 μL of Buffer A2 were then added and mixed gently by inverting the tube for 6-8 times (without vortexing). An incubation at room temperature for 4 min was performed. Then, 300 μL of Buffer A3 were added and the suspension was mixed thoroughly by inverting the tube 6-8 times (without vortexing). The lysate was centrifuged for 10 min at room temperature at top-speed. The supernatant was loaded to a NZYTech spin column in a 2 mL collection tube. After a 1 min centrifugation at 11,000 x g the flow-trough was discarded. The column was placed back into the same collection tube and washed by adding 500 μL of Buffer AY and centrifuging for 1 min. The flow-through was discarded and 600 μL of Buffer A4 were added. A centrifugation was performed as described. The flow-through was again discarded and the NZYTech column was inserted in the same collection tube and centrifuged for 2 min to remove residual ethanol. The column was placed into a clean 1.5 mL microcentrifuge tube and 50 μL of Buffer AE were added to the center of the column. An incubation of 1 min at room temperature and a centrifugation for 1 min at top speed was performed to elute de plasmid DNA.

Purification of DNA containing biotinylated molecules

Wash Dynabeads™ magnetic beads

The beads in the vial were resuspended for approximately 30 sec in the vortex. 500 μL of beads were transferred to a tube and an equal volume of Washing Buffer was added. The tube was mixed then placed on a magnet for 1 min and the supernatant was discarded. Subsequently, the tube was removed from the magnet and the washed beads were resuspended in 500 μL of Washing Buffer.

Immobilize nucleic acids

To immobilize the nucleic acids, firstly 100 μL of the PCR products (50 μL from the PCR reaction + 50 μL of sterile distilled water) were added to 100 μL of the binding and washing (B&W) Buffer (2x) and the mixture was resuspended. An incubation was performed for 15 min at room temperature using gentle rotation. Afterwards, the tube was placed on a magnet for 3 min to separate the biotinylated DNA coated beads. Then, 200 μL of B&W Buffer (1x) were used to wash the biotinylated DNA. This step was repeated twice. Finally, the DNA was resuspended in 50 μL of sterile distilled water.

Appendix 3

Agarose gel electrophoresis

Analysis of PCR products was performed on agarose gel electrophoresis. The samples were mixed with 6X loading buffer in proportion of 1:6 (v/v) and loaded in a 0.8% or 1% agarose gel. The gel was prepared with 1X of TAE buffer (Bio-Rad) and EtBr (AppliChem) to a final concentration of 0.5 μg/mL added before pouring the melted agarose in the running tray. All gels included 0.5 μg of the DNA marker DNA Ladder Mix (Fermentas). Electrophoresis was usually performed at 120 V for the desired time and the DNA was analyzed under UV light.

Solutions:

Loading buffer 6X: 2.5 mg/mL of bromophenol blue, 2.5 mg/mL of xylene cyanol FF and 30 % (v/v) glycerol; stored at 4 °C.

Purification of DNA from agarose

The purification of DNA from agarose gels was performed using the kit NZYGelpure (NZYTech), according to manufacturer's instructions. Carefully, the desired DNA fragment was excised from the agarose gel with a clean scalpel and placed in a 1.5 mL microcentrifuge tube. The gel slice was weighted and 300 μL of Binding Buffer were added for each 100 mg of gel. The tube was incubated at 60 °C for 10 min (or at room temperature for 1 hour) and occasionally shacked until agarose was completely dissolved. The mixture was loaded into a NZYTech spin column placed in a 2 mL collection tube and centrifuged at > 12,000 x g for 1 min. The flow-through was discarded and the column was placed back to the collection tube. When the DNA was intended to be used for direct sequencing, 500 μL of Wash Buffer were added to the column, which was centrifuged for 1 minute and the flow-through was discarded. Next, the column was washed with 600 μL of Wash Buffer for 1 min. The flow-through was discarded and the column was centrifuged for an additional minute to ensure the complete removal of residual ethanol. The NZYTech spin column, was placed into a clean 1.5 mL microcentrifuge tube and 30 to 50 μL of sterile distilled water were added, depending on the subsequent application. After 1 min of incubation at room temperature, the column was centrifuged for 1 minute to elute the DNA. The purified DNA was stored at -20 °C until further use.

Appendix 4

DNA concentration QUBIT® (Invitrogen)

To quantify the DNA amount in a sample, a master mix containing 199 μL of Qubit® dsDNA HS Assay buffer and 1 μL of Qubit® dsDNA HS Reagent was performed in a 1.5 mL tube. The resulting master mix was vortex followed by a fast spin step. Then, 199 μL of the master mix were introduced into an appropriate tube for the Qubit® and 1 μL of DNA from the sample was also added. Finally, the tubes were vortex for 2-3 seconds, incubated at room temperature for 2 min and the concentration was read in the Qubit® Flurometer.

Appendix 5

RNA purification

In *P. lusitanus* NL19, the extraction of total RNA was performed using Invitrogen™ PureLink® RNA Mini Kit. Cell pellets were centrifuged at 5000 rpm for 10 min at 4 ºC. The supernatant was discarded and 100 µL of prepared lysozyme solution were added to the cell pellet. The reaction was resuspended by vortex. Afterwards, 0.5 µL of 10% SDS solution were added and vortexed to mixed well. The cells were incubated at room temperature for 5 minutes. 350 µL of Lysis Buffer prepared with 1% of 2-mercaptoethanol were added and the reaction was mixed well. The cell lysate was then homogenized by pipetting. After pipetting, 250 µL of 100% ethanol were added to the bacterial cell homogenate. The reaction was vortexed to disperse any visible precipitate that may have formed after adding ethanol. The sample was then transferred to a Spin Cartridge (with a Collection Tube) and centrifuged at 12,000 x g for 15 sec at room temperature. The flow-through was discarded and the Spin Cartridge reinserted in the same Collection Tube. During this extraction process a treatment with DNase I was proceed according to On-column PureLink® DNase Treatment Protocol. As so, 350 µL of Wash Buffer I were added to the Spin Cartridge containing the bound RNA. A centrifugation at 12,000 x g for 15 sec at room temperature was performed and the flow-through and Collection Tube discarded. The Spin Cartridge was inserted into a new Collection Tube and 80 µL of PureLink® DNase mixture was directly added onto the surface of the Spin Cartridge membrane. This PureLink® DNase mixture contained 8 µL of 10X DNase I Reaction Buffer, 10 µL of Resuspended DNase (~3U/µL) and 62 µL of RNase Free Water. The column was incubated at room temperature for 15 min. Afterwards, 350 µL of Wash Buffer I were added to the Spin Cartridge followed by a centrifugation at 12,000 x g for 15 sec at room temperature. The flow-through and the Collection Tube were discarded and the Spin Cartridge inserted into a new Collection Tube. 500 µL of Wash Buffer II prepared with ethanol were added to the Spin Cartridge and a centrifugation at 12,000 x g for 15 sec at room temperature was performed. The flow-through was discarded and the Spin Cartridge reinserted into the same Collection Tube. The addiction of 500 µL Wash Buffer II with ethanol and centrifugation was repeated once. Then the Spin Cartridge was centrifugated at 12,000 x g for 1 min to dry the membrane with bound RNA. The Collection Tube was discarded, and the Spin Cartridge inserted into a Recovery Tube. 100 µL of RNase–Free Water were added to the centre of the Spin Cartridge. After an incubation of 1 min, the Spin Cartridge and Recovery Tube were centrifuged for 1 min at ≥12,000 x g at room temperature to elute the RNA. The purified RNA was stored at -20 °C until further use.

Appendix 6

Table 13: Accession numbers of the structural genes used in this thesis.

Strain	Protein ID
Pedobacter cryoconitis MP7CTX6	WP_183867758.1
	WP_183867757.1
	WP_157287913.1
	WP_183867756.1
	WP_183867755.1
Lewinella agarilytica DSM 24740	WP_090170703.1
	WP_090170706.1
	WP_090170708.1
	WP_090170711.1
Chitinophaga dinghuensis DSM 29821	WP_111594009.1
	WP_111594010.1
	WP_111594011.1
	WP_111594012.1
	WP_111594013.1
	WP_111594014.1
	WP_111594015.1
	WP_111594016.1
Hymenobacter swuensis DY53	WP_044000688.1
	WP_044000689.1
	WP_044000690.1
Chryseobacterium viscerum 687B-08	WP_076595230.1
	WP_076595229.1
	WP_109737881.1
	WP_076595230.1
	WP_175579922.1
Bacillus licheniformis DSM 13	P86720
Lactococcus lactis subsp. *lactis* CV56	ADZ63249.1
Bacillus subtilis subsp. *spizizenii* ATCC 6633	AAB91589.1
Chitinophaga pinensis DSM 2588	ACU62756.1
Staphylococcus epidermidis strains	P08136.1
Staphylococcus gallinarum strains	P21838.2

Appendix 7

Table 14: Accession number, biological process and molecular function of proteins overexpressed in high concentrations of PC.

Locus Tag	Accession	Biological Process	Molecular Function
TH53_21985	A0A0D0GGB4		
TH53_22175	A0A0D0GCP9		
TH53_22780	A0A0D0GKU7		
TH53_23040	A0A0D0FRC0		protein binding
TH53_23055	A0A0D0GKP5	metabolic process	catalytic activity
TH53_23690	A0A0D0GFI0	metabolic process; regulation of biological process	DNA binding
TH53_24550	A0A0D0EZI2	transport	
TH53_24835	A0A0D0GBQ5	transport	
TH53_24840	A0A0D0EZA0	transport	transporter activity
TH53_05430	A0A0D0GLM3	metabolic process	
TH53_11690	A0A0D0FWY1	transport	transporter activity
TH53_00810	A0A0D0GNT9		catalytic activity
TH53_12195	A0A0D0F5N1		
TH53_13055	A0A0D0FWC3	transport	receptor activity
TH53_14345	A0A0D0GPU3		
TH53_14350	A0A0D0GGV5	transport	transporter activity
TH53_14360	A0A0D0FVP7	metabolic process	catalytic activity
TH53_14760	A0A0D0GPQ0		
TH53_02045	A0A0D0GNE9		
TH53_16835	A0A0D0GJ08	regulation of biological process; response to stimulus	
TH53_17220	A0A0D0GIX0		
TH53_18915	A0A0D0FTG6	metabolic process	catalytic activity
TH53_14355	A0A0D0F4I5	metabolic process	catalytic activity; nucleotide binding

Table 15: Accession number, biological process and molecular function of proteins underexpressed in high concentrations of PC.

Locus Tag	Accession	Biological Process	Molecular Function
TH53_26180	A0A0D0FQ10	metabolic process	catalytic activity; metal ion binding
TH53_25565	A0A0D0GJJ7	metabolic process	catalytic activity; metal ion binding; nucleotide binding
TH53_25560	A0A0D0GEK2	metabolic process	catalytic activity; metal ion binding
TH53_02820	A0A0D0G1B0		
TH53_20830	A0A0D0GH44		
TH53_21200	A0A0D0GGV6		protein binding
TH53_22380	A0A0D0GCU6		
TH53_23490	A0A0D0EZX4		
TH53_23735	A0A0D0EZV8	metabolic process	catalytic activity; metal ion binding
TH53_23740	A0A0D0FR44		
TH53_23790	A0A0D0GFJ6		
TH53_03530	A0A0D0GMT0	transport	receptor activity
TH53_03645	A0A0D0F9J1	metabolic process; regulation of biological process; response to stimulus	catalytic activity; receptor activity; signal transducer activity
TH53_04060	A0A0D0G0T2	transport	receptor activity
TH53_25555	A0A0D0FQ68	metabolic process	catalytic activity
TH53_25575	A0A0D0EYZ0		
TH53_04300	A0A0D0GQE0		
TH53_25760	A0A0D0GBB3		
TH53_04425	A0A0D0F951		
TH53_04465	A0A0D0GQ33	metabolic process; transport	catalytic activity; nucleotide binding
TH53_04470	A0A0D0GVA3	metabolic process	catalytic activity
TH53_04475	A0A0D0GV45	cellular homeostasis; metabolic process; regulation of biological process	catalytic activity
TH53_25985	A0A0D0FQ20		
TH53_04880	A0A0D0G070		

Locus Tag	Accession	Biological Process	Molecular Function
TH53_05150	A0A0D0GLP4		
TH53_05890	A0A0D0FZJ6		
TH53_07450	A0A0D0GKL2		
TH53_08200	A0A0D0GN55	metabolic process	catalytic activity
TH53_08205	A0A0D0GT10	metabolic process	catalytic activity
TH53_08750	A0A0D0GMW0		
TH53_09160	A0A0D0F787	transport	receptor activity
TH53_09165	A0A0D0FY28		protein binding
TH53_09685	A0A0D0GML6		
TH53_09690	A0A0D0GMJ9	transport	catalytic activity; nucleotide binding; transporter activity
TH53_09695	A0A0D0GSJ4	metabolic process	catalytic activity
TH53_09700	A0A0D0GJH3		
TH53_09880	A0A0D0F716		
TH53_10200	A0A0D0GJ93		
TH53_10410	A0A0D0F6M6	transport	receptor activity
TH53_10625	A0A0D0FXI0		catalytic activity; protein binding
TH53_10975	A0A0D0FXF9		
TH53_00950	A0A0D0FAQ4	metabolic process; regulation of biological process	DNA binding
TH53_13005	A0A0D0GKX7		
TH53_13240	A0A0D0GQE9		
TH53_13415	A0A0D0GQ47		
TH53_13420	A0A0D0GHB1		
TH53_13930	A0A0D0GQ35		
TH53_13935	A0A0D0GQ25		
TH53_13980	A0A0D0GQ27	metabolic process; regulation of biological process; response to stimulus	catalytic activity; protein binding; receptor activity; signal transducer activity
TH53_14120	A0A0D0GGZ6		
TH53_14340	A0A0D0GK40		
TH53_14655	A0A0D0GJZ8	transport	transporter activity
TH53_14830	A0A0D0GPT1	transport	transporter activity
TH53_14960	A0A0D0F4C7		

Locus Tag	Accession	Biological Process	Molecular Function
TH53_15375	A0A0D0GJN6		
TH53_15820	A0A0D0FV43		
TH53_01895	A0A0D0FA54	metabolic process	catalytic activity
TH53_02165	A0A0D0GNF4		
TH53_16320	A0A0D0GNV3	transport	transporter activity
TH53_16590	A0A0D0GJ24		
TH53_17600	A0A0D0F337		
TH53_02340	A0A0D0FA50		
TH53_02595	A0A0D0G1H3		
TH53_18385	A0A0D0GIH0	cellular homeostasis; metabolic process; response to stimulus	metal ion binding
TH53_18390	A0A0D0GN77		
TH53_19130	A0A0D0GEF0		
TH53_20300	A0A0D0FSU1		
TH53_20350	A0A0D0F1I9		
TH53_20355	A0A0D0FSV8		catalytic activity
TH53_08195	A0A0D0FYJ6	metabolic process	catalytic activity
TH53_10640	A0A0D0GM08	metabolic process; response to stimulus	
TH53_04885	A0A0D0GPW9	metabolic process	catalytic activity
TH53_02680	A0A0D0FA45	transport	transporter activity

Appendix 8

Quantification of protein concentration using the Bradford assay

The total protein concentration was determined with the Bradford assay. For this, a standard curve was obtained using four solutions of BSA (1 mg/mL; 0,5 mg/mL; 0,25 mg/mL and 0,125 mg/mL). Three replicas of each solution were prepared to a final volume of 10 µL. Afterwards, 790 µL of dH_20 and 200 µL of Bradford reagent were added. Moreover, a blank was also prepared containing 800 µL of dH_2O and 200 µL of Bradford reagent.

To prepare the samples firstly, they were diluted ten times to a final volume of 10 µL. Then, 790 µL of dH_2O and 200 µL of Bradford reagent were added. Afterwards, the samples and standard curve reactions were incubated at room temperature for 5 min and the OD_{595nm} was measured.

Table 16: Values of absorbance at 595 nm of BSA reactions with different concentrations.

[BSA] (mg/mL)	OD_{595nm}
0	0,000 0,000 0,000
1,25	0,003 0,005 0,002
2,5	0,008 0,009 0,010
5	0,022 0,024 0,021
10	0,048 0,043 0,046

Table 17: Values of absorbance at 595 nm and protein concentration in the samples of PC25%.

Sample	**OD_{595nm}**	**[Protein] (µg/µL)**
PC25% (1)	0,068	14,809
PC25% (2)	0,043	9,489