**University of Lille 1**

Science and Technology

**PhD Thesis**

Engineering of Biological Functions

By

Walaa HUSSEIN AMIN MOHAMED ABDELWAHED

---

# STUDY ON THE REGULATION AND BIOSYNTHESIS OF FENGYCIN AND PLIPASTATIN PRODUCED BY *BACILLUS SUBTILIS*

---

Laboratory ProBioGEM, UPRES-EA 1026

Polytech'Lille

September 22 2011

| | |
|---|---|
| **Philippe Jacques**, Professor (University of Lille 1, France) | Supervisor |
| **Iordan Nikov,** Professor (University of Lille 1, France) | Co-supervisor |
| **Patrick Fickers,** Professor (University of Brusseles, Belgium) | Reporter |
| **Marc Ongena,** Doctor of University (University of Agricultural Sciences, Gembloux, Belgium) | Reporter |
| **Frédérique Gancel,** Doctor of University (University of Lille 1, France) | Examinator |
| **Maude Pupin**, Doctor of University (University of Lille 1, France) | Examinator |
| **Alain Brans,** researcher (University of Liege, Belgium) | Examinator |
| **Marie-Dominique Devignes**, Researcher CNRS (University of Henri Poincaré, Nancy1) | Examinator |

# Université Lille 1

Sciences et Technologies

# Thèse

Presentée par Walaa HUSSEIN AMIN MOHAMED ABDELWAHED

Pour l'obtention du titre de

Docteur de l'université Lille 1 Sciences et Technologies
Spécialité : Ingénierie des Fonctions Biologiques

---

## ETUDE DE LA REGULATION ET DE LA BIOSYNTHESE DES FENGYCINES ET PLIPASTATINES PRODUITES PAR *BACILLUS SUBTILIS*

---

Preparée au Laboratoire ProBioGEM, UPRES-EA 1026

Soutenue le 22 Septembre 2011 devant le Jury composé de :

| | |
|---|---|
| **Philippe Jacques**, Professeur (Université Lille 1, France) | Directeur |
| **Iordan Nikov,** Professor (Université Lille 1, France) | Co-directeur |
| **Patrick Fickers,** Professeur (Université Libre de Bruxelles, Belgique) | Rapporteur |
| **Marc Ongena,** Chercheur Qualifié FNRS (Gembloux AgroBioTech, Université de Liège) | Rapporteur |
| **Frédérique Gancel,** Docteur de l'Université (Université Lille 1, France) | Examinatrice |
| **Maude Pupin**, Docteur de l'Université (Université Lille 1, France) | Examinatrice |
| **Alain Brans,** Chercheur (Université de Liege, Belgique) | Examinateur |
| **Marie-Dominique Devignes**, Chargée de Recherche CNRS (Université Henri Poincaré, Nancy1) | Examinatrice |

# Dedication

to my children
with love and gratitude

## Acknowledgements

I would like to thank Professor Philippe Jacques for allowing me to work in his laboratory under his supervision and for his kind attention and continuous scientific guidance.

I would also like to thank Dr. Max Béchet, research engineer for his great support and for his follow-up throughout this work. I would like to thank him about all the knowledge I gained from him.

I express my deepest gratitude to Dr. Frédérique Gancel who supported me throughout this work, for giving me the benefit of her knowledge and for her valuable support during the writing of this thesis.

I would like to thank Dr. Valérie léclére for her help and support in the bioinformatics analyses encluded in this thesis.

I would like to thank all members of ProBioGEM laboratory for their friendly welcome during four years.

I thank my father, my mother, my stepmother, my brother and my sisters for all the support, trust and words encouragement they have given me.

I would like to thank you my husband for your great support, your love, your patience in a foreign country and for your encouragement.

It is important to me to thank my friends in Egypt who have always prayed for me.

I dedicate this work to my beloved country "Egypt"

# *List of contents*

# *List of tables*

# *List of figures*

## *Abbreviations*

| | |
|---|---|
| A-domain | adenylation |
| ACP | acyl carrier protein |
| AT | acyl transferase |
| BBG | *Bacillus* strain collection of ProBioGEM |
| C-domain | condensation domain |
| CoA | coenzyme |
| DNA | deoxyribonucleic acid |
| E-domain | epimerization domain |
| EDTA | ethylene di-amine tetra- acetate |
| ESI | electrospray ionization |
| HPLC | high performance liquid chromatography |
| HEPES | 4-(2- hydroxylethyl)-1- piperazineethane sulfonic acid |
| IPTG | isopropyl β-D- thiogalactopyranoside |
| KS | keto synthase |
| MOPS | 3-(N-morpholino)- propane sulfonic acid |
| NRPS | non ribosomal peptide synthetase |
| OD | optical density |
| ONGP | o-nitrophenyl-β- D-galactopyranoside |
| ORF | open reading frame |
| PCR | polymerase chain reaction |
| PCP (domain) | peptidyl carrier protein |
| PKS | polyketide synthetase |
| 4-PP | 4-phosphopantetheine |
| PPTase | 4-phosphopantetheinyl transferase |
| TE | thioesterase domain |
| UP | region upstream the promoter |
| UV | ultraviolet |
| W/V | weight per volume |
| X-Gal | 5-bromo-4- chloro-3- indolyl β-D- galactopyranoside |

# *Abstract*

In this study, 39 published genomes from *Bacillus* strains were analyzed by bioinformatics tools; eight of them do not harbour genes or operons coding for non ribosomally synthesis of peptides (NRPS), while the genetic potential for the production of a siderophore, bacillibactin, which is a molecule of NRP type, was revealed in 27 other strains. 14 strains showed: (1) the presence of genes/operons implied in the syntheses of the four families of lipopeptides [surfactin, pumilacidin, lichenysin (surfactins family), fengycin or plipastatin (fengycins family), bacillomycin, iturin and mycosubtilin (iturins family) and kurstakin (kurstakin family)], (2) the presence of genes implied in the synthesis of an antibiotic, bacitracine. In 16 *Bacillus* strains the presence genes/operons coding for other new NRPS and NRPS-PKS (Polyketide synthases) hybrid molecules were detected.

The analysis of operons implicated in synthesis of plipastatins and fengycins from the domesticated *Bacillus subtilis* strain 168, *B. subtilis* F29-3 and *B. amyloliquefaciens* FZB42 showed high homologies between these operons and strong similarities between those NRPS molecules. Moreover, the importance of sequencing the *B. subtilis* S499 genome was raised, because the structure of fengycin produced by this strain cannot be correlated with the structure of the synthetases described in other strains synthesizing fengycins/plipastatins, which are molecules difficult to differentiate.

A failure occurred in the obtention of a plipastatin mono-producer derivative from *B. subtilis* 168 by the replacement of the plipastatin native promoter by a constitutive one $P_{repu}$ linked to a neomycin resistance gene (*neo*). On the other hand, inactivation of plipastatin operon expression by the mutant BBG111, which produces constitutively surfactin, was obtained by inserting this $P_{repU}$-*neo* cassette inserted in the opposite direction to that of transcription of *pps* operon. The resulting BBG140 mutant was found to synthesize more surfactin that the parental strain.

The promoters $P_{pps}$ (*B. subtilis* 168) and $P_{fen}$ (*B. subtilis* BBG21) were cloned into the expression vector pDG1661, which bears the *lacZ* cassette, and inserted within the chromosome of strain 168. Under different growth conditions, expression and regulation of these $P_{pps}$ and $P_{fen}$ promoters, compared to those of $P_{srfA}$ promoter from *B. subtilis* 168, were studied in the respective derived mutants BBG142, BBG139 and BBG127. The results showed that the level of expression of these three promoters could be controlled by fermentation conditions. Temperature was found to be the first factor affecting the expression of $P_{srf}$ and $P_{fen}$ promoters, the second factor being the medium composition. On the contrary, the induction of the $P_{pps}$ promoter was different: not only it was affected by these two last parameters, but also by the "acidic stress" occurring during the first 24 h of the fermentation process and disappearing at 30 h.

# *Introduction*

Non-ribosomal synthesis system NRPS it is an alternative pathway that allows production of polypeptides other than through the traditional translation mechanism. The peptides are created here by multienzymatic complexes called *synthetases* and the resulting peptides are generally short, 2 to 50 residues. NRPS produces several pharmacologically important compounds, including antibiotics and immunosuppressors. This biosynthesis pathway is found in many bacteria and fungi. Peptides produced by NRPS show peculiar features compared to traditional proteins. First, they can contain standard as well as non-standard amino acids. Secondly, amino acids are linked not only by an amino-peptide link, but also by non-conventional links that form a non-linear peptide backbone. There exist iterative and nonlinear NRPS configurations that generate more complicated structures. Consequently, some peptides form cycles, unusual branching or repeats leading to various topological structures.

*Bacillus subtilis* strains produce many kinds of bioactive peptides as secondary metabolites. Some of them are synthesized non-ribosomally by a large multifunctional enzyme complex. Among them, surfactins, iturins, and fengycins or plipastatins are the main representatives. In this thesis, the lipopeptides under study are the fengycins and the plipastatins which are very similar in their chemical structure and close in their operon structure. Both fengycin and plipastatin operons include 5 genes (*fenA* to *fenE*), (*ppsA* to *ppsE*). Fengycin is a lipodecapeptide fungicide which consists of almost the same kind of amino acids and ß-hydroxy fatty acids as plipastatin and differing in the D and L- tyrosine positions. However, there is difficult to differentiate between the plipastatins and the fengycins; for example, *pps* operon was thought to be the fengycin operon, because significant homology was observed between the fengycin synthetase gene of fengycin-producing *B. subtilis* F29-3 and the operon from strain 168.

As an overview on the importance role of fengycin, it is considered as antifungal antibiotic by its effect on the membrane according to its concentration; at low concentration, fengycin aggregates to form pores leading to permeability changes of the membrane while, at high concentrations, it acts as a detergent by solubilizing the membrane. Fengycin stimulates phospholipase A2 enzyme which is involved in a number of physiologically important cellular processes such as, inflammation, acute hypersensitivity and blood platelet aggregation. It has surfactant strength (lower than surfactin) by reducing the interfacial

tension at the oil-water interface and have a haemolytic activity (less about 40-fold than surfactin). Fengycin plays a significant role in the biocontrol: it reduces grey mould disease caused by *Botrytis cinerea* on apple, reduces disease incidence caused by *Colletorrichum langenarim* and *Pythium aphanidermatum* on cucumber and tomato, protects vegetative cells of beans against damping-off caused by *Pythium ultimum*, and has antifungal activity against *Fusarium gramineavum* and *Sclerotinia sclerotiorum* in canola and wheat. Fengycins were added to the limited number of bacterial products that are described as elicitors of defence response in plants, and are likely to be widely applied in medicine, agriculture and the food industry.

The first aim of this thesis was to get an overview of the Non Ribosomal Peptide produced by Bacillus strains. The sequenced genome of 39 different strains were thus analysed by bioinformatics tools. We then focus our work on fengycins or plipastatins produced by the NRPS system in some *B. subtilis* strains, in order to relate the structure of these two molecules and the structure of operons responsible for their biosynthesis. Indeed, Schneider *et al*. (1999) confirmed the existence of fengycin molecule produced by *Bacillus subtilis* S499 strain, which cannot be correlated with the structure of the synthetases described in other fengycin- or plipastatin-producing strains. Thus, the sequencing of the whole genome of this strain is important and of interest, to elucidate the NRPS operon implied in this synthesis. The second aim was to study the regulation of the plipastatin and fengycin operons in *B. subtilis* 168 by determining the strength of both strain 168 $P_{pps}$ and strain BBG21 promoters using a gene reporter system. The last aim was to obtain new derivatives of *B. subtilis* 168 such as a plipastatin mono-producer, which will (i) facilitate purification of plipastatin only in significant amounts, and (ii) studies on plipastatin role as an antifungal antibiotic in biocontrol experiments, and its potential use as a biopesticide.

# *Review of Literature*

## 1. *Bacillus subtilis*

### 1.1. Organism description

*Bacillus subtilis* is a Gram-positive, rod-shaped bacterium commonly found in soil and vegetation. *B. subtilis* generaly grows in the mesophilic temperature range. The optimal temperature is 25-35°C. Stress and starvation are common in this environment; therefore, *B. subtilis* has evolved a set of strategies that allow survival under these harsh conditions. One strategy, for example, is the formation of stress-resistant endospores. *B. subtilis* bacteria use their flagella for a swarming motility. This motility occurs on surfaces, for example on agar plates, rather than in liquids. *B. subtilis* are arranged in single rods or chains. Cells arranged next to each other can only swarm together, not individually. These arrangements of cells are called 'rafts'. In order for *B. subtilis* bacteria to swarm, they need to secrete a slime layer which includes surfactin, a surface tension-reducing lipopeptide, as one of its components (Schaechter, 2006).

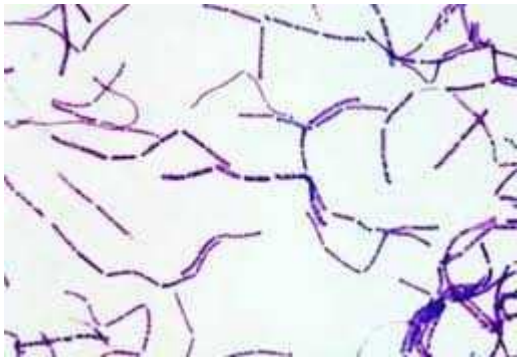


Figure 1: *Bacillus subtilis* gram staining.

### 1.2. *Bacillus subtilis* metabolism

*B. subtilis* bacteria have been considered strictly aerobic, meaning that they require oxygen to grow and they cannot undergo fermentation. However, recent studies showed that they can indeed grow under anaerobic conditions making them facultative anaerobes. The bacteria can make ATP in anaerobic conditions via butanediol fermentation as well as nitrate ammonification. *B. subtilis* can use nitrite or nitrate as a terminal acceptor of electrons. *B. subtilis* contains two nitrate reductases. One is used for nitrate assimilation and the other is used for nitrate respiration. However, there is only one nitrite reductase that serves both purposes. Nitrate reductase reduces nitrate to nitrite in nitrate respiration, which is then reduced to ammonia by nitrite reductase. *B. subtilis* is different from other facultative aerobes in that it undergoes fermentation without external acceptors of electrons (Nakano, 1998). During fermentation, the regeneration of NAD+ is chiefly mediated by lactate dehydrogenase, which is found in the cytoplasm. Lactate dehydrogenase converts pyruvate to lactate (Marino, 2001).

### 1.3. *Bacillus subtilis* genome structure

*B. subtilis* displays several features, which render this organism an interesting organism not only for the scientific community but also for industrial applications. These features include amongst others the capability to: take up exogenous DNA (so called competence, which facilitates genetic manipulations which are well established for this organism), form endospores (so called sporulation which offers a model for studying this relatively simple developmental process), secrete large quantities of proteins (also including heterologous proteins). In addition, it was the first Gram-positive bacterium for which the complete genome sequence became available in 1997 by Kunst *et al*. This revealed that the chromosomal DNA of *B. subtilis* 168 strain has a low G+C content (43.5%) and comprises around 4.1 million base pairs, corresponding to more than 4,100 genes. Remarkably, at that time only 30% of these genes had a well known function, while for approximately half of the remaining ones an annotated function could be assigned based on sequence homology. A great portion of the genome corresponds to carbon source applications (Kunst *et al*., 1997). 192 of the 4,100 genes are considered indispensable, and an additional 79 are thought to be essential. Most of the essential genes are involved in metabolism. Half of the essential genes are responsible for processing information, one-fifth of them are responsible for cell wall synthesis, cell division and shape, and one-tenth of them are responsible for the energetics of the cell. The essential

genes which code for unknown functions are 4% (Kobayashi, 2003). Most of the strains of *B. subtilis* synthesize numerous antibiotics (Kunst *et al*., 1997). Five signal peptidase genes were found to be important for this secretion function (Ara, 2007).

### 1.4. Regulation of transcription

Due to energy consumption reasons, only a fraction of the total number of genes so-called housekeeping genes is constitutively expressed (Rasmussen *et al*., 2009). The products of these genes are crucial for the cell at any circumstances, for example enzymes of the glycolytic pathway or ribosomal proteins.

On the other hand, many other proteins are only needed "upon request" in specific situations, such as the availability/lack of certain compounds or when being under stress conditions. To recognize environmental changes and adequately respond to them, bacteria possess regulatory molecules, so-called transcription factors (TFs) that control expression of one or more genes (van Hijum *et al*., 2009). The process of gene expression is initiated by binding of the RNA polymerase (RNAP) sigma factor complex to the promoter/operator site (DNA sequence in front of a gene) that is recognized by the sigma factor component (van Hijum *et al*., 2009). TFs bind to the promoter sequence as well and may either facilitate or hamper the attachment of the RNAP leading to activation or repression of transcription, respectively and as such they can be considered a key element in regulation of gene transcription. Although there are more than 250 predicted TFs and 17 sigma factors encoded in the genome sequence of *B. subtilis*, only a limited number of them have a defined function (Sierro *et al*., 2008; Grote *et al*., 2009). In general, global transcription regulators and specific regulators can be distinguished based on the number of regulated genes in their regulons, the former driving the expression of a large number of genes of different functional categories, whereas the latter regulates only a small number of genes in a particular metabolic pathway. Several factors contribute to the complexity of prokaryotic gene regulation by TFs: global regulators have the capability to act as either repressors or activators or both, a gene can be regulated by more than one TF and one TF can modify expression of another TF.

## 2. Non Ribosomal Peptides synthesis

Scientists in the 1960s realized that the ribosomal machinery is not suitable to produce antibiotic lipopeptides. Until recently, non ribosomal synthesis of peptides was thought to be restrictly produced of the gram-positive *Actinomycetes* and *Bacilli* genera, filamentous fungi, or marine microorganisms (Faulkner, 2002), but we know today that eukaryotes also harbor systems that resemble NRPSs (Finking and Maraheil, 2004).

NRPS is a multi enzyme system consisting of an arrangement of modules. A module is defined as a section of the NRPS's polypeptide chain that is responsible for the incorporation of one building block into the growing polypeptide chain (Marahiel, 1997; von Döhren *et al.,* 1997; Schwarzer and Marahiel, 2001; Mootz *et al.,* 2002). Thus, NRPSs are used simultaneously as template (because the amino acid to be incorporated is determined by the module) and biosynthetic machinery (it is the module that harbors all necessary catalytic functions). The enzymatic units that reside within a module are called domains. These domains catalyze at least the steps of substrate activation, covalent binding, and peptide bond formation of non ribosomal peptide synthesis (Stachelhaus and Marahiel., 1995). Domains of equal function share a number of highly conserved sequence motifs. These "core-motifs" allow the identification of individual domains at the protein level (Schwarzer *et al.*, 2003).

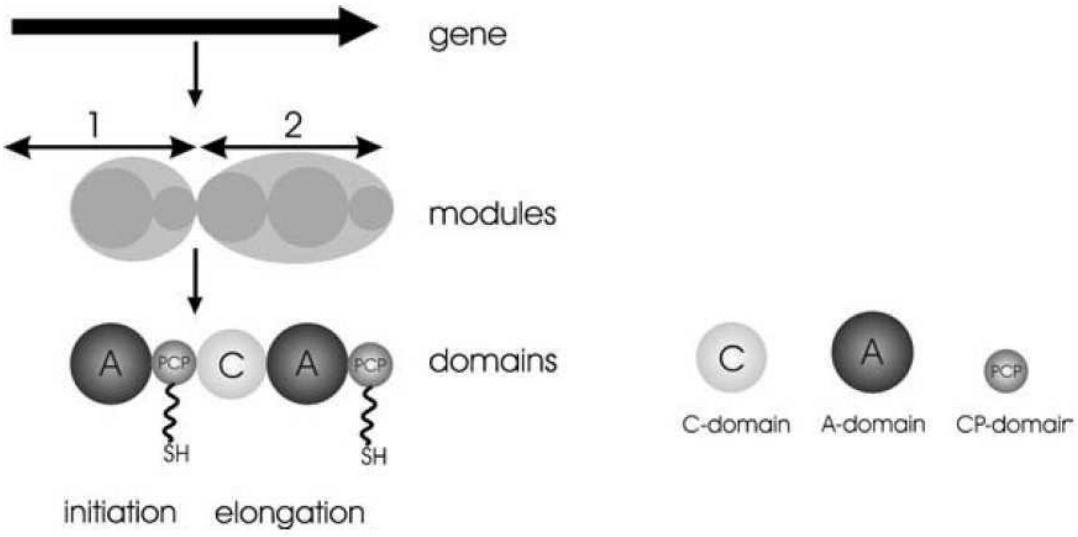



Figure 2: NRPS modules and domains coming from the gene (Finking and Marahiel, 2004)

### 2.1. NRPS modules

Modules can be subdivided into initiation, elongation and termination modules. Elongation modules contain at the minimum an A-domain for substrate recognition (Dieckmann *et al.,* 1995; Stachelhaus and Marahiel, 1995; Mootz and Marahiel, 1997; May *et al.*, 2001), a PCP that holds the activated substrate (Stachelhaus *et al.*, 1996; Ehmann et *al.,* 2000), and a C-domain for peptide bond formation (Stachelhaus et *al.*, 1998; Bergendahl et *al.*, 2002).

#### 2.1.1. Adenylation domain

The A-domain (~550 a.a) is responsible for the selection of the amino acids that makes up the product and thus controls its primary sequence. A-domains activate amino acid substrate as amino acyl adenylate, while ATP is consumed (Dieckmann *et al*., 1995; Stachelhaus and Marahiel, 1995; Mootz and Marahiel, 1997; May *et al*., 2001). Ten core-motifs were identified in polypeptidic sequence of A-domains. The residues lie in a 100-aa stretch between cores A4 and A5 of the A-domains analysis led to the introduction of the so-called non ribosomal code, which allows the prediction of A-domain selectivity on the basis of its primary sequence (Stachelhaus et *al*, 1999; Du *et al*, 2000). In addition to the prediction of A-domain selectivity, the non ribosomal code was also exploited for the creation of A-domains with altered selectivity (Eppelmann *et al.,* 2002). For exemple, site-directed mutagenesis within the core A4-A5 region changed the selectivity of *srfA-A1* (first adenylation domain of the first NRPS responsible for the surfactin biosynthesis) from Glu to Gln and that of *srfA-B2* (second adenylation domain of the second NRPS responsible for the surfactin biosynthesis) from Asp to Asn.

#### 2.1.2. Peptidyl Carrier Protein domain

Called PCP, the Peptidyl Carrier Protein domain is a small domain (80–100 a.a) which represents the transport unit that accepts the activated amino acid that is covalently tethered to its 4-Phosphopantetheinate (4-PP) cofactor as thioester (Stachelhaus and Marahiel, 1996; Ehmann *et al.,* 2000). This cofactor is post-translationally transferred by a phosphopantetheine transferase (PPTase) to a conserved serine residue of the carrier protein (CP) and acts as a flexible arm to allow the bound amino acyl and peptidyl substrate to travel between different catalytic centres.

### 2.1.3. Condensation domain

C-domains (~450 a.a) are the central entity of non ribosomal peptide synthesis because they are responsible for peptide bond formation between amino acyl substrates bound to PCPs of adjacent modules (Stachelhaus *et al.,* 1998; Bergendahl *et al.,* 2002). The enzyme catalyzes the nucleophilic attack of the amino (or imino, hydroxyl) group of the activated amino acid bound to the downstream (with respect to the C-domain) module onto the acyl group of the amino acid tethered to the upstream module, as shown in Figure 2. According to the multiple-carrier thiotemplate model (Stein *et al.*, 1996), the C-domain possesses a site for the nucleophile (acceptor site) and a position for the electrophile (donor site) as shown in Figure 3.

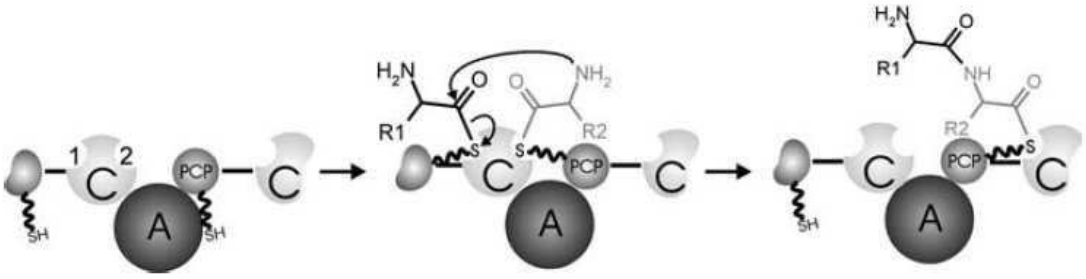



Figure 3: Reaction mechanism of the C-domain (Finking and Marahiel, 2004)

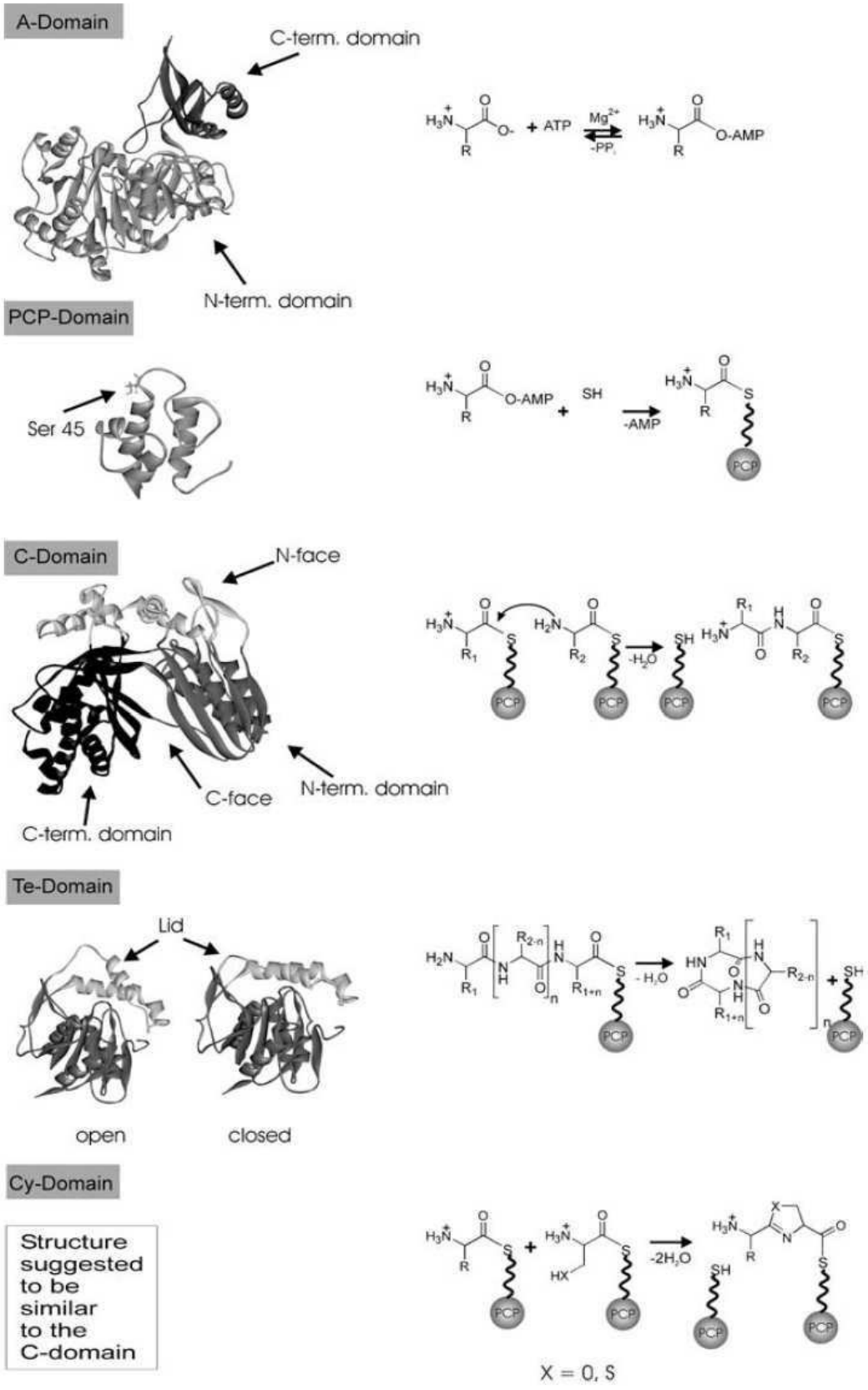



Figure 4: Crystal structures and reaction catalysed by the core-domains of NRPSs (Finking and Marahiel, 2004)

### 2.1.4. The thioesterase domain

During synthesis, the growing peptide chain is handed over from one module to the next until it reaches the final module's PCP-domain. This module contains in most cases a TE-domain (~250 aa) that is important for the liberation of the product (Schneider and Marahiel, 1998; Sieber and Marahiel, 2003) (Figure 3). Product release is achieved by a two-step process that involves an acyl-*O*-TE-enzyme intermediate that is subsequently attacked by either water or a peptide-internal nucleophile (Kohli *et al.*, 2001; Trauger *et al.*, 2001; Miller *et al.*, 2002) which results either in a macrocyclic product, as observed in the case of surfactin (Tseng *et al.*, 2002). Macrocyclic release seems to be the favoured mechanism and the mode of termination differs significantly between systems. TE-domains must hence have undergone diversification to catalyze reactions such as amide bond formation between the N-terminal amino group and the C terminus of the peptide (i.e., cyclosporine A), lactonization of the C-terminal carboxyl group thioester of the peptide and the hydroxyl group of an N-terminal *β*-hydroxy fatty acid (i.e., surfactin), or the formation of branched cyclic structures. Moreover, TE-domains also qualify for the oligomerization of peptide units. Gramicidin S, for instance, is composed of two identical pentapeptides bridged head to tail, whereas enterobactin represents a trimer that has been cyclized to give a macrolactone. This high degree of specialization is additionally reflected in the low sequence identity (10%–15%) among TE-domains.

### 2.1.5. Secondary or additional domains

The additional domains can be involved in the biosynthesis of the peptide to modify the structure of the monomer involved in the primary structure or to add some external compounds to the peptide. Among these tailoring domains, they are cyclisation (Cy), methylation (Me), oxidation (Ox), glycosylation, epimerisation (E) and addition of fatty acid chain. The last two are involved in lipopeptide biosynthesis in *Bacillus* spp. The E-domain catalyses the epimerisation of the PCP-bound L-amino acid of the growing polypeptide chain. The addition of the fatty acid chain to the first amino acid of the peptide moiety is catalysed by a first specific condensation domain. The added fatty acid chain can itself be partially synthesized by another main group of modular enzymes, called the polyketide synthases (PKS). In this last case, a hybrid PKS/NRPS is required for the synthesis of the biomolecules (Du *et al*., 2000).

## 2.2. Modes of biosynthesis

### 2.2.1. Linear NRPS

Surfactin synthetase is a linear NRPS (Peypoux *et al*., 1999) in which a total of 24 domains organized in seven modules are distributed over three NRPSs that act in the directional synthesis of surfactin, starting at module 1 and ending at the TE domain of module 7.

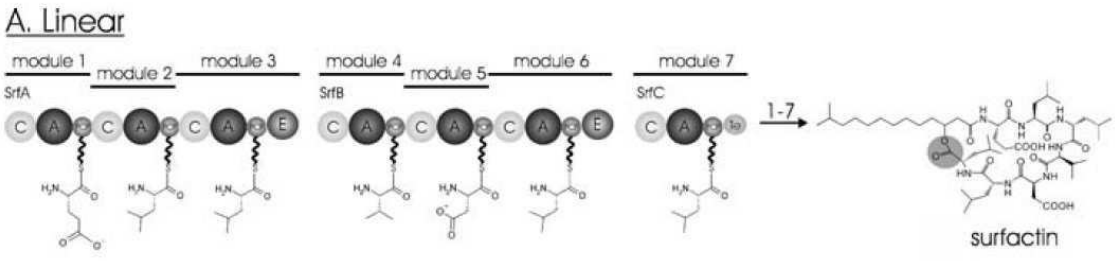



Figure 5: Surfactin operon a linear NRPS strategy (Finking and Marahiel, 2004)

### 2.2.2. Iterative NRPS

This mode is used for gramicidin biosynthesis. *GrsA* and *GrsB* first synthesize the pentapeptide monomer that is stalled on the active-site serine of the TE-domain. The regenerated NRPSs engage in a second round of synthesis and the TE-domain finally catalyzes the head-to-tail cyclization to give the mature product (Shaw-Reid *et al*, 1999; Kohli *et al.*, 2001).

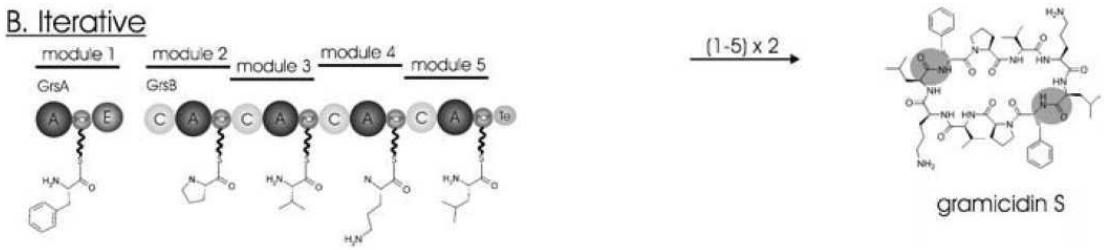



Figure 6: Gramicidin operon as an iterative NRPS strategy (Finking and Marahiel, 2004)

### 2.2.3. Non linear NRPS

This strategy is more complicated and characterized by an arrangement of modules that deviates from the classical (C-A-PCP)n arrangement or by incorporation of small molecules that are not covalently bound to the NRPS template during synthesis (Mootz *et al.*, 2002). Vibriobactin is synthesized using this complicated strategy by the condensation of norspermidine (NS) with dihydroxybenzoate (DHB) and with two molecules of DHB-mOx

(DHB-methyloxyzoline) by the second C-domain of *VibF*. Synthesis of DHB-mOx is realized by transfer of DHB to *VibF*, which condenses it with the heterocyclized threonine.

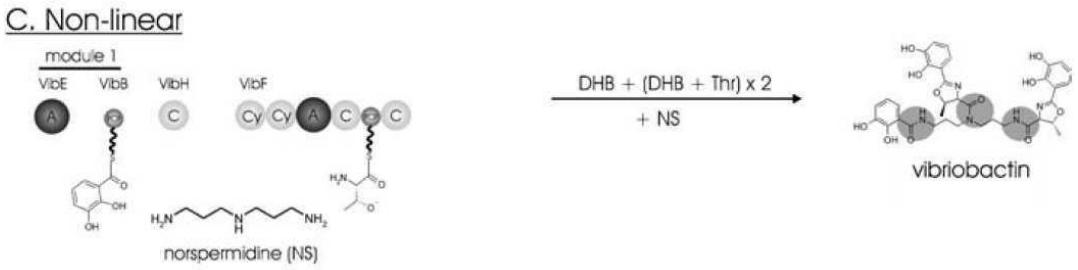



Figure 7: Vibriobactin operon as a non linear NRPS strategy (Finking and Marahiel, 2004).

### 2.3. Post-translational modification

All PCPs of NRPS systems need to be post-translationally modified to serve their function as transport proteins. For this purpose, the 4-PP moiety of CoA is transferred to a conserved serine residue of the *apo*-CP and is then converted to the active *holo*-CP. The transfer of 4-PP is catalyzed by a dedicated PPTase in an Mg2+-dependent reaction (Lambalot *et al.*, 1996; Walsh *et al.*, 1997). *B. subtilis* 168 was enabled to produce the lipopeptide antibiotic plipastatin after transformation with chromosomal DNA of the plipastatin producer *B. subtilis* YB8. The study revealed that, in addition to the *pps* operon, the PPTase (*lpa-8*) (*sfp*) was essential for the production of the antibiotic (Tsuge *et al.*, 1999).

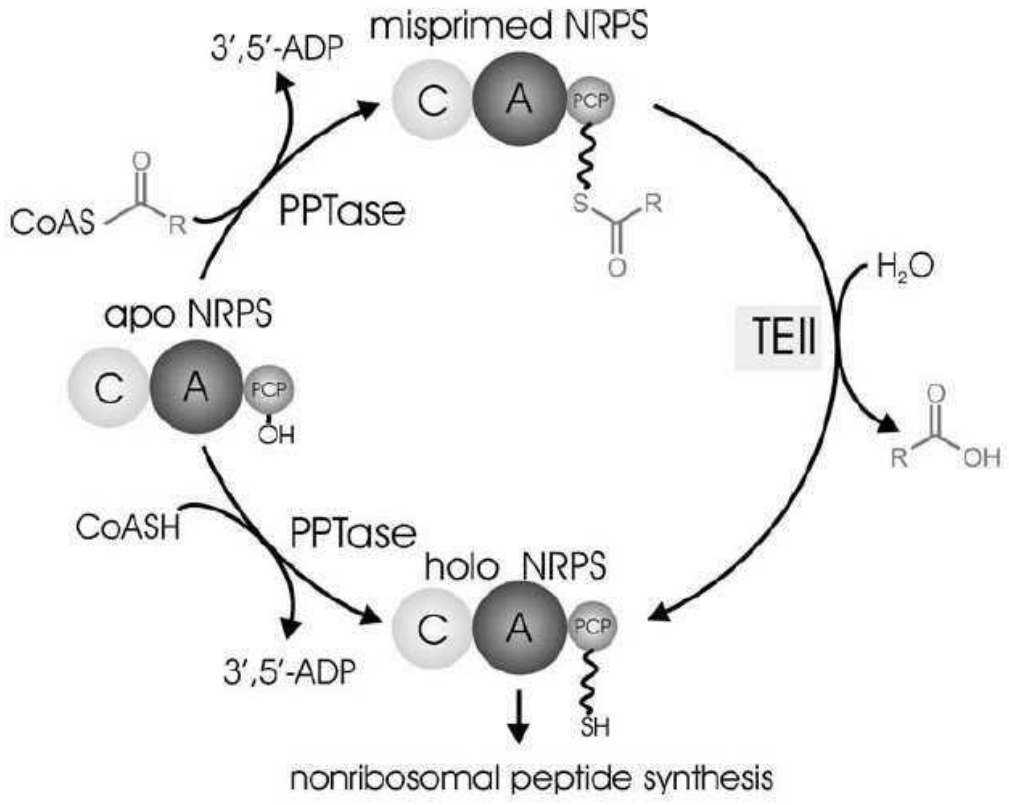



Figure 8: 4-Phosphopantetheinyl transferases catalyze the nucleophilic attack of the conserved serine's hydroxyl group of the *apo*-CP onto the *β*-phosphate of CoA. This results in the conversion of the CP to the *holo*-form and release of 3,5-ADP. *Apo*-PCPs of NRPSs can also be erroneously modified with acyl-CoA derivatives because PPTases exhibit a somewhat relaxed selectivity toward these CoA derivatives. To regenerate the inactive NRPS, a TEII-domain hydrolyzes the thioester and leaves only the 4-PPcofactor on the PCP, which can afterwards engage in non ribosomal peptide synthesis (Finking and Marahiel, 2004).

### 3. *Bacillus* lipopeptides

*Bacillus* lipopeptides are classified into three main different families: surfactins, iturins and fengycins (or plipastatins). In addition, several new lipopeptides have been discovered as Kurskatins, Antiadhesin, Bamylocin A, Circulocin 1, Circulocin 3 and Fusaricidin (Jacques, 2011; Roonsavang, 2011).

Table 1: Lipopeptides families of *Bacillus* spp. (Jacques, 2011)

| Surfactin family | Iturin family | Fengycin and Plipastatin family | Kurstakin family |
|---|---|---|---|
| Surfactin<br>Bamilocin A<br>Lichenysin<br>Esperin<br>Pumilacidin | Mycosubtilin<br>Bacillomycin F<br>Bacillomycin L<br>Bacillomycin D<br>Iturin A<br>Iturin AL<br>Iturin C | Fengycin<br>Plipastatin | Kurstakin |

Table 2: Primary structure of representative lipopeptides produced by *Bacillus* strains

| Name | Structure | Ref. |
|---|---|---|
| **Surfactin family** | | |
| Surfactin | FA-β-OH-L-Glu-L-Leu-D-Leu-L-Val-L-Asp-D-Leu-L-Leu | [31] |
| Lichenysin A/D | FA-β-OH-L-Gln-L-Leu-D-Leu-L-Val-L-Asp-D-Leu-L-Ile | [7,32] |
| (Lichenysin B) | FA-β-OH-L-Glu-L-Leu-D-Leu-L-Val-L-Asp-D-Leu-L-Leu | [33] |
| (Lichenysin C) | FA-β-OH-L-Glu-L-Leu-D-Leu-L-Val-L-Asp-D-Leu-L-Ile | [34] |
| Lichenysin G | FA-β-OH-L-Gln-L-[$A_2$]-D-Leu-L-[$A_4$]-L-Asp-D-Leu-L-[$A_7$]<br>$A_2$ = Leu/Ile, $A_4$ = Val/Ile, $A_7$ = Ile/Val | [6] |
| Surfactant BL86 | FA-β-OH-L-Glx-L-Leu-D-Leu-L-Val-L-Asx-D-Leu-L-[$A_7$]<br>$A_7$ = Ile/Val | [35] |
| Pumilacidin | FA-β-OH-L-Glu-L-Leu-D-Leu-L-Leu-L-Asp-D-Leu-L-[$A_7$]<br>$A_7$ = Ile/Val | [36] |
| **Fengycin family** | | |
| Fengycin | FA-β-OH<br>-L-Glu-D-Orn-D-Tyr-D-*a*Thr-L-Glu-D-[$A_6$]-L-Pro-L-Gln-L-Tyr-L-Ile<br>$A_6$ = Ala/Val | [37] |
| Plipastatin | FA-β-OH<br>-L-Glu-D-Orn-L-Tyr-D-*a*Thr-L-Glu-D-[$A_6$]-L-Pro-L-Gln-D-Tyr-L-Ile<br>$A_6$ = Ala/Val | [37] |
| **Iturin family** | | |
| Iturin A | FA-β-$NH_2$-L-Asn-D-Tyr-D-Asn-L-Gln-L-Pro-D-Asn-L-Ser | [38] |
| Iturin C | FA-β-$NH_2$-L-Asp-D-Tyr-D-Asn-L-Gln-L-Pro-D-Asn-L-Ser | [37] |
| Bacillomycin L | FA-β-$NH_2$-L-Asn-D-Tyr-D-Asn-L-Ser-L-Glu-D-Ser-L-Thr | [39] |
| Bacillomycin D | FA-β-$NH_2$-L-Asn-D-Tyr-D-Asn-L-Pro-L-Glu-D-Ser-L-Thr | [37] |
| Bacillomycin F | FA-β-$NH_2$-L-Asn-D-Tyr-D-Asn-L-Gln-L-Pro-D-Asn-L-Thr | [37] |
| Mycosubtilin | FA-β-$NH_2$-L-Asn-D-Tyr-D-Asn-L-Gln-L-Pro-D-Ser-L-Asn | [40] |
| **Other** | | |
| Antiadhesin | FA-β-OH-L-Asp-L-Leu-L-Leu-L-Val-L-Val-L-Glu-L-Leu | [41] |
| Bamylocin A | FA-β-OH-X-Glu-X-Leu-X-Met-X-Leu-X-Pro-X-Leu-X-Leu | [42] |
| Circulocin 1 | *g*FA-β-OH-X-Thr-X-Phe-X-Ile-X-Dba-X-Asp | [43] |
| Circulocin 3 | *g*FA-β-OH-X-Thr-X-Leu-X-Ile-X-Thr-X-Asn-X-Ala | [43] |
| Fusaricidin | *g*FA-β-OH-L-Thr-D-Val-L-Tyr-D-*a*Thr-D-Asn-D-Ala | [44] |
| Kurstakins | FA-X-Thr-X-Gly-X-Ala-X-Ser-X-His-X-Gln-X-Gln | [45] |

**4. Polyketide synthases and NRPS/PKS hybrids**

The synthesis of a polyketide requires at least three domains that can be organized in modules much like NRPSs. The AT (acyl transferase) domain is responsible for the selection of the substrate which can be malonyl-CoA, methyl-, ethyl-, or propylmalonyl-CoA units (Katz, 1997). The AT-domain transfers the substrate to the 4-PP of the corresponding *holo*-ACP (acyl carrier protein). The substrate is subsequently relocated to an active-site cysteine residue of the KS (keto synthase)-domain. The malonyl derivative tethered to the ACP of the adjacent downstream module is decarboxylated, which yields the free nucleophile necessary for the following Claisen-condensation with the KS-bound ketide. The chain starter (propionyl-CoA) is therefore always transferred to the extender molecule (methylmalonyl-CoA) until the product has travelled through all elongating steps and undergoes cleavage by macrocyclization catalyzed by a TE-domain.

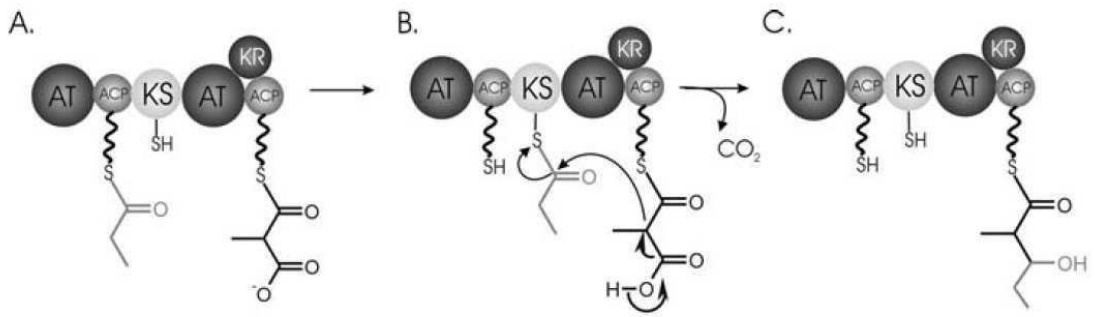



Figure 9: A fictitious dimodular PKS. (*A*) The ACP of the first module is loaded with propionyl by the AT-domain of the first module, and the second AT-domain loads its ACP with methylmalonyl. (*B*) The propionyl residue is translocated to an active-site cysteine of the KS-domain. The methylmalonyl is subsequently decarboxylated to yield the nucleophile for the condensation with the KS-bound propionyl. (*C*) The product of the condensation is now covalently tethered to the 4-PP of the second module's ACP and can participate in subsequent elongation steps. The presence of the KR-domain in this example has additionally led to the reduction of the *β*-carbonyl to a hydroxyl group (Finking and Marahiel, 2004).

Because of the close relationship between PKSs, FASs, and NRPSs, PKSs represent the mediator between primary (FAS) and secondary metabolism (NRPS) that allows continuous transition between these systems (Cane and Walsh., 1999). Du *et al.* (2000) showed that the same catalytic sites are conserved between the hybrid NRPS-PKS and normal NRPS or PKS systems, although the ketoacyl synthase domain in NRPS/PKS hybrids is unique, and that specific interpolypeptide linkers exist at both the C- and N-termini of the NRPS and PKS proteins, which presumably play a critical role in facilitating the transfer of the growing

peptide or polyketide intermediate between NRPS and PKS modules in hybrid NRPS-PKS systems.

## 5. Lipopeptide discovery

### 5.1. Surfactin discovery

In 1968 Arima *et al*. isolated an exo-cellular compound with an exceptional bio-surfactant activity from the supernatant of a culture of *B. subtilis*. Its structure was elucidated as that of a lipopeptide (Kakinuma *et al*., 1968). It was characterized as a valuable inhibitor of fibrin clot formation, an antibacterial, anti-tumour and hypo-cholesterolemic agent. Other strains or species producing surfactin derivatives were identified as *Bacillus coagulans* (Huszcza and Burczyk 2006) and *Bacillus mycoides* (Athukorala *et al*., 2009). Related compounds have also been found such as esperin (Thomas and Ito, 1969), halobacillin (Trischman *et al*., 1994), lichenysin from from *Bacillus licheniformis* (Horowitz *et al*., 1990), pumilacidin from *Bacillus pumilus* (Morikawa *et al*., 1992) or bamylocin A from *Bacillus amyloliquefaciens* (Lee *et al*., 2007).

### 5.2. Iturin discovery

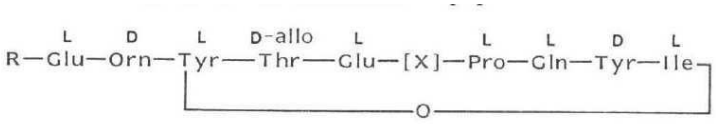
Mycosubtilin was the first antifungal lipopeptide from iturin family mentioned in literature (Walton and Woodruff, 1949). The molecule was isolated from *B. subtilis*. Then, iturin was characterized as a strong antifungal agent with a restricted antibacterial activity against *Micrococcus* and *Sarcina* strains. The precise structure of iturin, mycosubtilin and similar compounds from the same family was described during the 1970s and 1980s: mycosubtilin (Peypoux *et al*., 1976, 1986), bacillomycin L (Besson *et al*., 1977) identical to bacillomycin Lc or bacillopeptin (Eshita *et al*., 1995; Volpon *et al*., 2007), iturin A (Peypoux 1978), iturin C (Peypoux *et al*., 1978), bacillomycins D (Peypoux *et al*., 1981) and F (Peypoux *et al*., 1985).

Production of iturinic compounds was also identified in other species such as *Paenibacillus koreensis* (Chung *et al*., 2000), *B. amyloliquefaciens* (Yu *et al*., 2002) and *B. pumilus* (Cho *et al*., 2009). Mycocerein, an antifungal peptide produced by *Bacillus cereus* was partially described and could belong to the iturin family (Wakayama *et al*., 1984).

The genes involved in the biosynthesis of iturinic compounds have been first characterized for mycosubtilin in *B. subtilis* ATCC 6633 (Duitman *et al*., 1999) and then for iturin A (Tsuge *et al*., 2001a) and bacillomycin D (Koumoutsi *et al*., 2004; Moyne *et al*., 2004).

## 6. The difference between fengycins and plipastatins structure

Fengycins and plipastatins are very related lipo-decapeptides. They have two structures A and B (figures 14, 16), which differ by their amino acid residue in position 6 that can be Ala (form A) or Val (form B), and display an internal lactone ring in the peptidic moiety between the carboxyl terminal amino acid (Ile) and the hydroxyl group in the side chain of the tyrosine residue in position 3. Different β-hydroxy fatty acid chains (C14 to C18) are linked with an amide bond to the N-terminal amino acid residue (Glu) (Nishikiori *et al*., 1986b; Vanittanakom *et al*., 1986). The main representative fatty acid chains are C15, C16 and C17. They are saturated, except in the case of a single lipopeptide isolated from the supernatant of *Bacillus thuringiensis*, which has a fatty acid chain with one double bond between carbons 13 and 14 (Kim *et al*., 2004). The structure of the lipodecapeptide plipastatin was described first by (Nishikori *et al*., 1986a) from *B. cereus* BMG302-fF67. They reported that the plipastatins structure divided into two groups; the group A plipastatins had D-Ala at the 6-position of the amino acid sequence from the N-terminus, and the group B had D-Val in this position. The plipastatins numbered 'one' (i.e. A1 and B1) had 3(R)-hydroxyhexadecanoic acid (B-hydroxypalmitic acid) as the fatty acid residue and those numbered `two' (A2 and B2) had 14(S)-methyl-3(R)-hydroxyhexadecanoic acid (B-hydroxy-anteiso-palmitic acid) as the fatty acid residue.




| Plipastatin | | R | [X] |
|---|---|---|---|
| A1 | $CH_3(CH_2)_{12}CHCH_2CO-$ (OH) | 3(*R*)-hydroxyhexadecanoic acid | D-Ala |
| A2 | $CH_3CH_2CH(CH_2)_{10}CHCH_2CO-$ ($CH_3$, OH) | 14(*S*)-methyl-3(*R*)-hydroxyhexadecanoic acid | D-Ala |
| B1 | $CH_3(CH_2)_{12}CHCH_2CO-$ (OH) | 3(*R*)-hydroxyhexadecanoic acid | D-Val |
| B2 | $CH_3CH_2CH(CH_2)_{10}CHCH_2CO-$ ($CH_3$, OH) | 14(*S*)-methyl-3(*R*)-hydroxyhexadecanoic acid | D-Val |

Figure 10: The total structure of plipastatin (Nishikiori *et al*., 1986b)

A conformational model of plipastatin Al was proposed by (Nishikori. *et al*., 1986b) as shown in Figure 10. In 2000, Volpon *et al*. obtained two plipastatins from *B. subtilis* NCIB 8872, differing only by a D-Ala (plipastatin A)/D-Val (plipastatin B) substitution at the position 6.

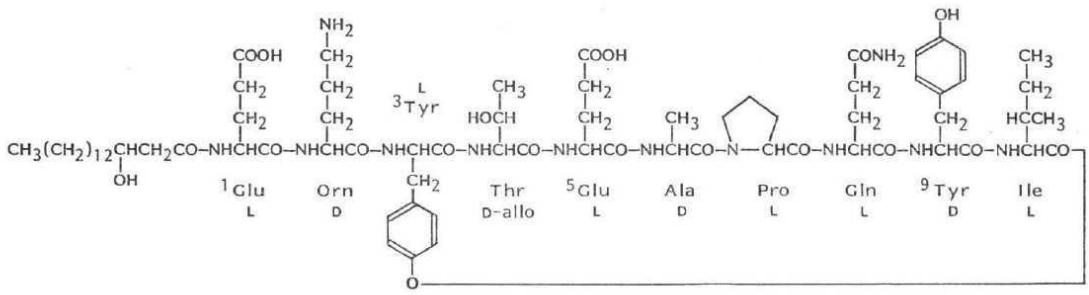


Figure 11: NMR spectrum of plipastatin $A_1$ (Nishikiori *et al*., 1986b)

In 2004, Wang *et al*. analysed the fengycin homologues produced by *B. subtilis* JA with electrospray ionization/Collision-induced dissociation (ESI/CID) mass spectrometry and ions representing characteristic fragmentations were detected.

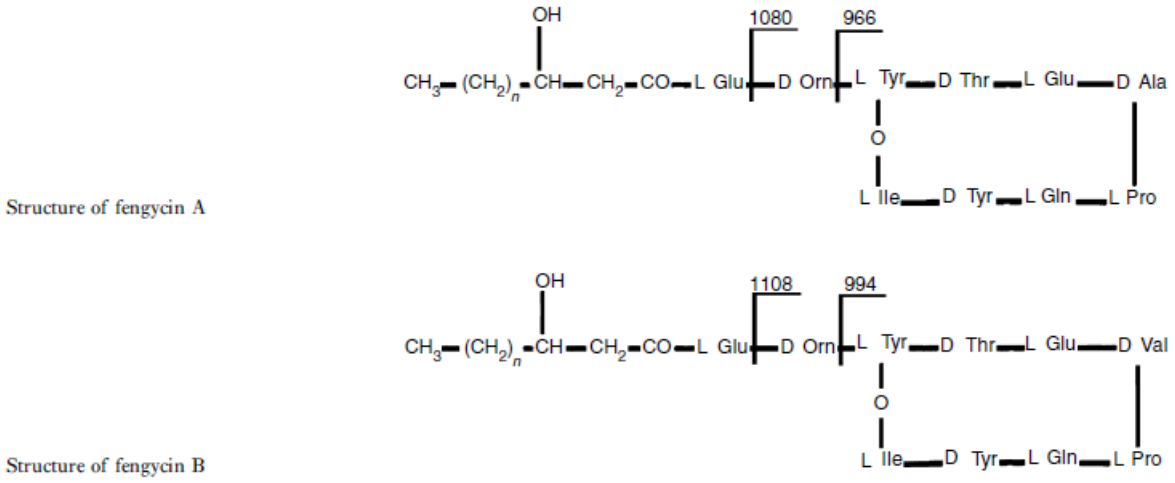



Figure 12: Plipastatins A, B structures from ESI/CID (Wang *et al*., 2004)

Only two differences were identified between fengycin and plipastatin: a Gln instead of a Glu in position 8 and the L and D forms of tyrosine, which are in position 3 and 9, respectively, for plipastatins and 9 and 3 for fengycins. In the different works describing fengycin structure since its discovery, the presence of a Glu in position 8 was never mentioned. However, (Schneider *et al*., 1999) confirmed the existence of fengycin molecule with a D-Tyr in position 3 from the supernatant of *B. subtilis* S499. This result cannot be correlated with the structure of the synthetases described in other fengycin- or plipastatin-producing strains.

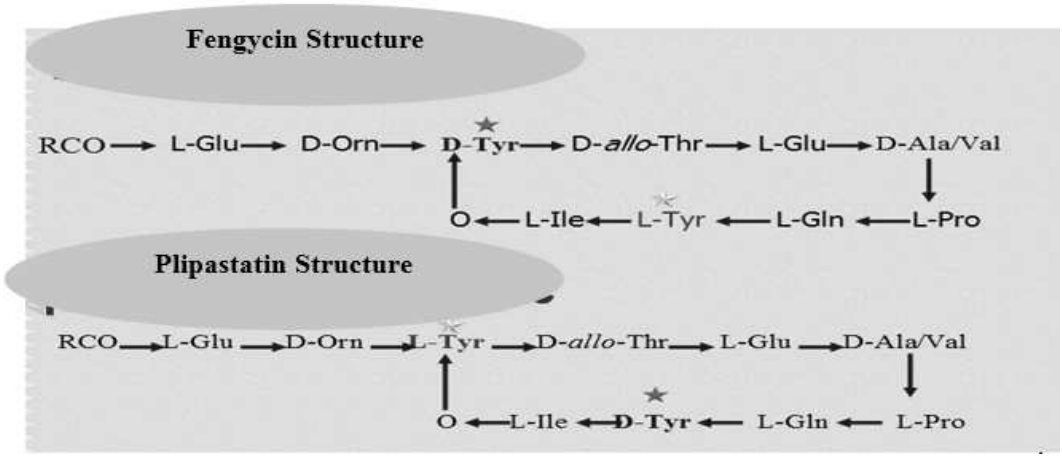



Figure 13: The difference between fengycin and plipastatin structures in the position of L- and D-Tyrosine in the molecules

## 7. Lipopeptide operons structure and regulation

### 7.1. Surfactin operon

Three large Open Reading Frames (ORFs) coding for surfactin synthetases are designated *srfAA*, *srfAB* and *srfAC* (Galli *et al*., 1994). They present a linear array of seven modules (one module per residue), three modules are present in the products of *srfA-A* and *srfA-B*, *srfA-C* respectively, and the last one in *srfA-D*. A high specificity was observed for adenylation domains of modules 1, 3, 5 and 6, which recognize L-Glu, L-Leu, L-Asp and L-Leu amino acid residues, respectively. Modules 3 and 6 contain an epimerase domain (E), which transforms the incorporated L-Leu in a D-Leu. *In vitro* studies of the specificity of adenylation domains of modules 2, 4 and 7 shown that they are able to accept several aliphatic amino acid residues as substrates. The β-hydroxylated fatty acid chain is added to the amino acid activated in the first module by the way of a starter condensation domain. A first thioesterase fused with the carboxy-terminal end of the final activation PCP domain is responsible for the release of the synthesized product from the enzymatic template.

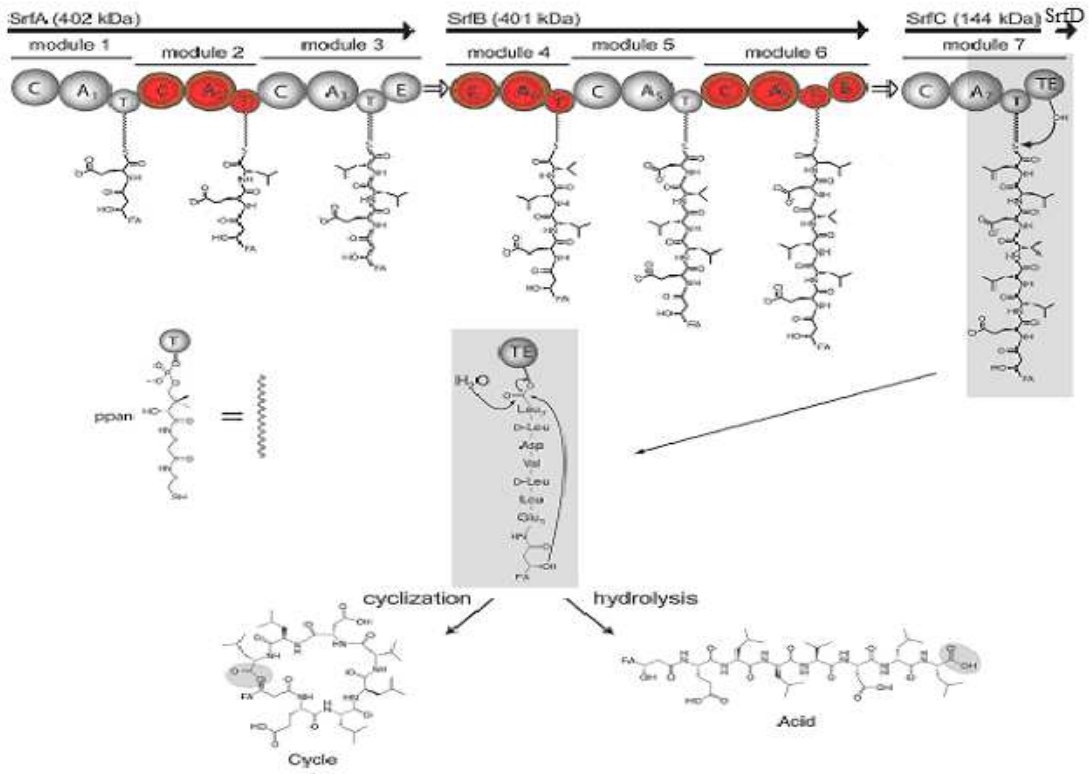



Figure 14: The regulation of NRPS modules and domains of surfactin operon (Seiber and Marahiel, 2003)

### 7.2. Iturin operon

Contrary to surfactin and fengycin, iturin derivatives are synthesized by a PKS–NRPS hybrid complex. The operon consists of four ORFs called *fenF*, *mycA*, *mycB* and *mycC* for mycosubtilin (Duitman *et al*., 1999), *itu*D, *itu*A, *itu*B and *itu*C for iturin (Tsuge *et al*., 2001) and *bmy*D, *bmy*A, *bmy*B and *bmy*C for bacillomycin D (Hofemeister *et al*., 2004; Koumoutsi *et al*., 2004). The last three genes encode for the three NRPSs, which are responsible for the incorporation of the first residue (Asn for *myc*A, *Itu*A and *Bmy*A), the following four residues (Tyr, Asn, Gln, Pro for *myc*B and *Itu*B; Tyr, Asn, Pro, Glu for *Bmy*B) and the two last residues (Ser, Asn for *myc*C; Asn, Ser for *Itu*C and Ser, Thr for *Bmy*C). The thioesterase present in the last module catalyses the release of the peptide and the formation of an amide bond between the carboxylic group of the last amino acid and the amino group of the fatty acid chain. Epimerisation domains were identified in modules 2, 3 and 6. D-form of the corresponding amino acid residues is observed in the final product.

### 7.3. Fengycin and plipastatin operons structure and regulation

#### 7.3.1. Fengycin and plipastatin operons synthetases systems

The operon encoding fengycin or plipastatin synthetases was first described in *B. subtilis* 168 by Tosato *et al*. (1997), and then in *B. subtilis* F29-3 by Lin *et al*. (1999) who cloned and sequenced five fengycin synthetase genes. Among these genes, *fenC* encodes a fengycin synthetase which activates L-glutamic acid and D-ornithine, respectively. A promoter is located upstream from the start site of *fenC*. Lin *et al.* (2005) analyzed and sequenced *fenD* gene that encodes an enzyme which activate Tyr and Thr, the third and the fourth amino acids in fengycin, respectively. Shu *et al.* (2002) analysed the functions of the enzyme encoded by *fenE,* which contains two amino acid activation modules that activate the three amino acids L-Glu, L-Ala/L-Val, and L-2-aminobutyric acid, indicating that *fenE* activates the fifth and the sixth amino acids in fengycin. Also, Lin *et al.* (1998) cloned and sequenced a fengycin synthetase gene, *fenB* which has substrate specificity toward L-Ile, *fenB* also consists of a thioesterase domain, suggesting that this protein may be involved in the activation of the last amino acid of fengycin. Finally, Wu *et al.* (2007) reported that the five fengycin synthetases, in *B. subtilis* F29-3, interlock to form a chain, which coils into a 14.5-nm structure. In this chain, fengycin synthetases are linked in the order FenC-FenD-FenE-FenA-FenB by interactions between the C-terminal region of an upstream enzyme and the N-terminal region of its downstream partner enzyme, with their amino acid activation modules arranged colinearly with the amino acids in fengycin. The third plipastatin or fengycin synthetases description was in *B. subtilis* b213 and A1/3 in 1999 (Steller *et al*., 2000) by purification of five multifunctional peptide synthetases (Fen 1-5) that catalyze biosynthesis of the peptide portion of the fengycin. This five gene cluster (*fen* 1-5) shows high homology to the *pps* and *fen* genes in *B. subtilis* strains 168 and F29-3. Finally, in *Bacillus amyloliquefaciens* FZB42, Koumoutsi *et al*. (2004) reported that the fengycin (*fen*) operon was organized and located as in *B. subtilis* 168.

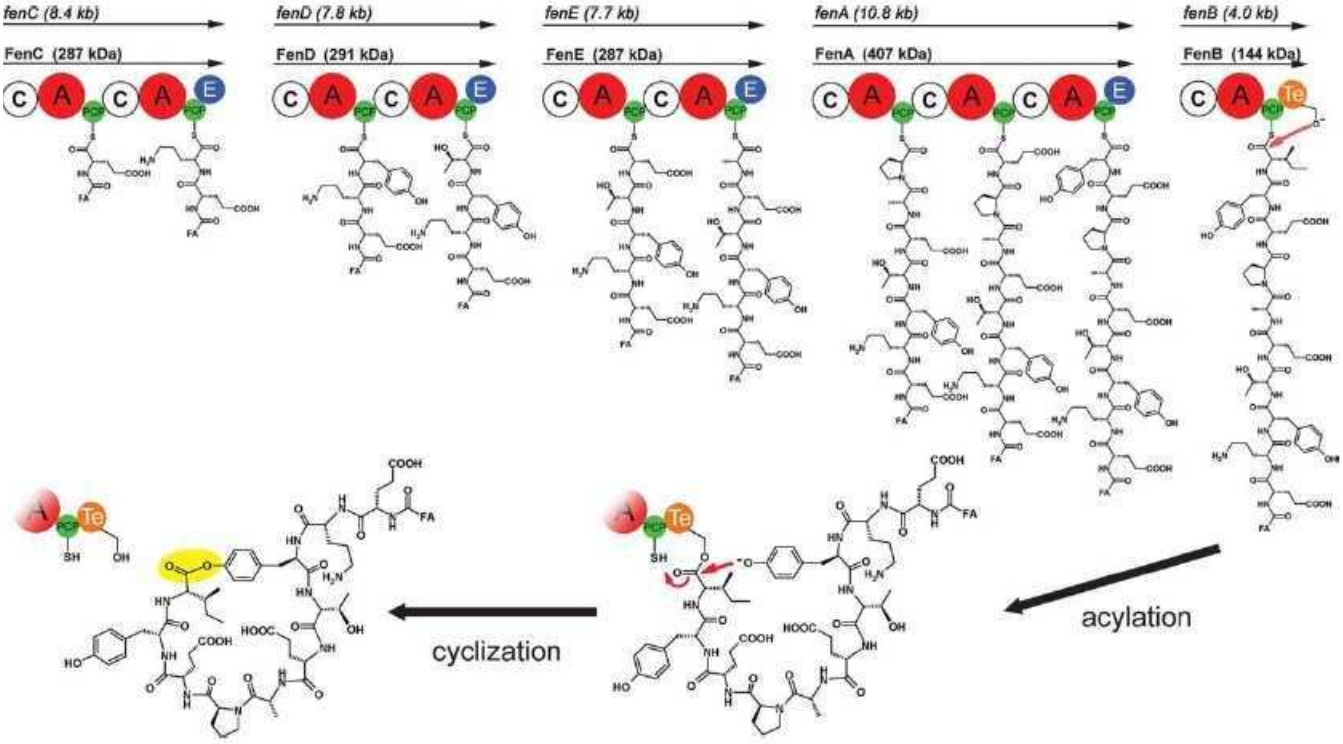



Figure 15: The regulation of the NRPS modules and domains of fengycin operon. (Samel *et al*., 2006)

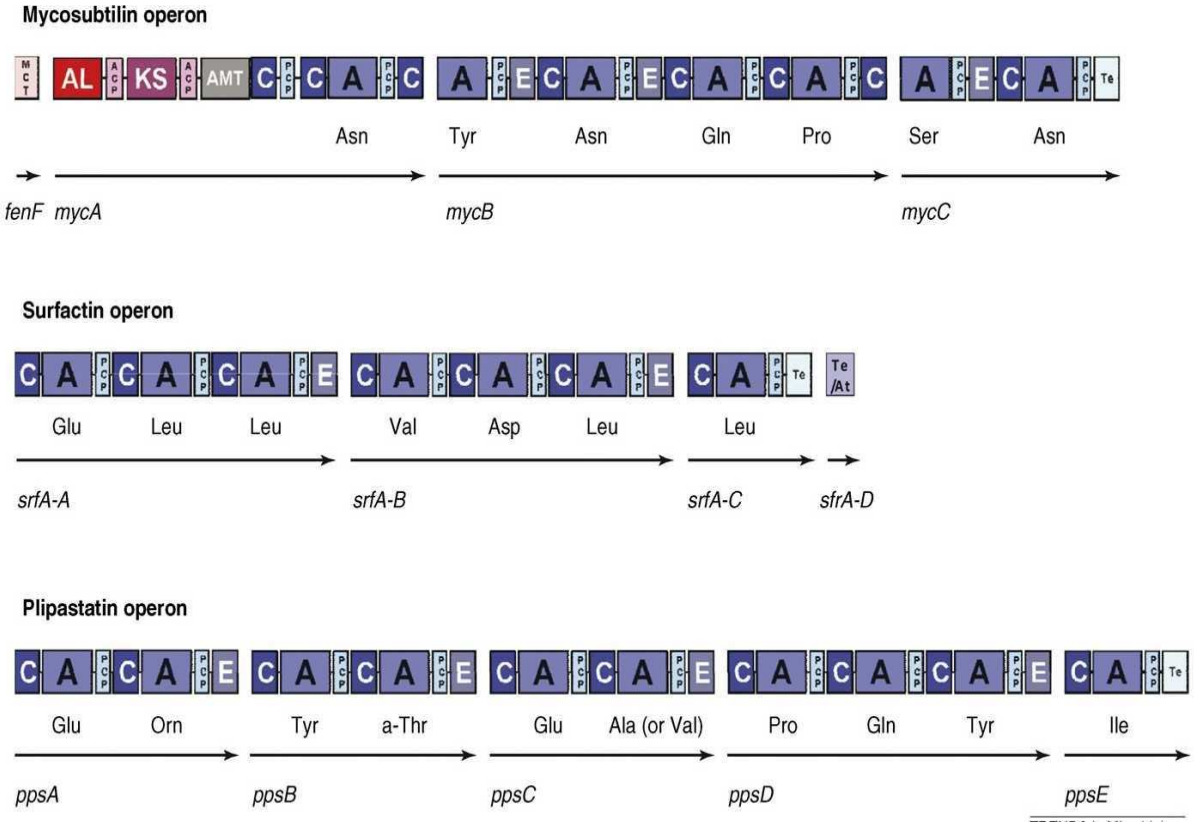



Figure 16: Operons of mycosubtilin, surfactin and plipastatin synthetases. Schematic representation of operons (ORFs, domains of NRPSs or PKSs and amino acid incorporated by the different modules) encoding catalytic machinery responsible for the biosynthesis of representative members of each family of lipopeptides produced by *Bacillus subtilis*: mycosubtilin for the iturin family, plipastatin for the fengycin family and surfactin. Domains are described in the box text (Ongena and Jacques, 2008)

### 7.3.2. Fengycin and plipastatin promoter expression

Since the promoter plays an important role in the regulation of the transcription, Ke *et al.* (2009) established that the promoter of the fengycin synthetase operon of *B. subtilis* F29-3 is located 86 nucleotides upstream from the translational initiation codon of *fenC*. This investigation involved transcriptional fusions with a DNA fragment that contains the region between positions -105 and -80 and determined that deleting the region between positions -55 and -42 reduces the promoter activity by 64.5%. Transcriptional fusions in the *B. subtilis* DB2 chromosome also indicated that mutating the sequence markedly reduces the promoter activity. An *in vitro* transcription analysis confirmed that the transcription is inefficient when the sequence in this region is mutated. Electrophoretic mobility shift and footprinting analyses demonstrated that the C-terminal domain of the RNA polymerase subunit binds to the region between positions -55 and -39. These results indicated that the sequence is an UP (region upstream the promoter) element. Finally, this UP element is critical for the production of fengycin, since mutating the UP sequence in the chromosome of *B. subtilis* F29-3 reduces the transcription of the *fen* operon by 85%.

To study the plipastatin or the fengycin promoter expression a reporter gene fusion system was used. By the early 1980's, the construction of *lacZ*-gene fusions was revolutionizing the field *Escherichia coli* genetics (Casadaban *et al*., 1983). The use of β-galactosidase, for which a simple colorimetric assay was available, to tag a gene product that was difficult to quantitate was a powerful new tool for studying gene regulation. Naturally, researchers in the *B. subtilis* community were eager to adapt the technique to their model system. In 1982, Donnelly and Sonenshein announced that they had created a *lacZ*-fusion with the promoter and first few codons of the *B. subtilis tms* gene. When they placed the construct on a high copy number plasmid in *B. subtilis,* copious amounts of β-galactosidase were produced. It had occurred that they might be able to integrate their fusion into the *tms* locus, but their attempt was unsuccessful. Fittingly, the first reported use of an integration vector to create a gene fusion within the chromosome was in Zuber and Losick's study of the "0.4 kb gene," *spoVG* (Zuber *et al*., 1983). Their vector allowed them to study the timing of *spoVG* expression in a natural, single-copy state and provided for an easy way to transfer the reporter gene fusion into a variety of genetic backgrounds. Reports of similar studies using integration vectors to create fusions with other genes quickly followed (Ferrari *et al*., 1985; Piggot *et al*., 1987; Lewandoski *et al*., 1988).

## 8. Fengycins and plipastatins production

### 8.1. The role of *sfp* and *degQ* genes in plipastatin or fengycin production

The 4-phosphopantetheinyl transferase *sfp* gene is required to promote the lipopeptide production, when this gene was inactivated in strain *B. subtilis* YB8, neither surfactin nor plipastatin $B_1$ was produced (Tsuge *et al*., 1996). Tsuge *et al*. (1999) claimed that strain 168 expressing only *lpa-8 (sfp)* can also produce plipastatin, although the yield is very low. However, the introduction of the pleiotropic regulator *degQ* from strain YB8 into strain 168 expressing *lpa-8* resulted in a ten-fold increase in the production of plipastatin. Also, Tsuge *et al*. (2007) reported that the biosynthesis of plipastatin from *B. subtilis* 168 requires three gene blocks at different genome loci, i.e. the peptide synthetase operon *ppsABCDE* (38kb), *degQ* (0.6 kb), and *sfp* (1.0 kb). On the other hand, Ongena *et al*. (2007) constructed three mutants of strain *B. subtilis*168 that produce specific amounts of each lipopeptide. After the introduction of active *sfp* from *B. subtilis* S499, the *B. subtilis* 2500 derivative was obtained which produces high levels of surfactin but still very low amounts of fengycin. Replacement of the fengycin promoter in *B. subtilis* 2500 by a stronger promoter sequence (P*amyQ*) yielded *B. subtilis* 2508 that efficiently produces both surfactin and fengycin in final concentrations higher than 400 mg/L without the introduction of active *degQ* gene. They also obtained *B. subtilis* 2504, a surfactin non-producer derivative of *B. subtilis* 2508 by interrupting the *srfA* operon, which also secretes fengycins in large amounts at a level similar to *B. subtilis* 2508 but is impaired in surfactin production.

### 8.2. Optimal growth conditions and culture medium for fengycin or plipastatin production

Lipopeptide production depends on the environmental and nutritional conditions to which the *B. subtilis* cells are exposed; the composition of the fermentation medium is one of these important factors, whereas Leenders *et al*. (1999) reported that *B. subtilis* 6633 showed a striking difference in lipopeptide production when it was cultivated under different culture media (only surfactin was detected with Landy medium, mycosubtilin was detected with DSM medium, while both mycosubtilin and surfactin were detected with ACS medium). Moreover, Volpon *et al*. (2000) observed that the modification of the culture conditions (nitrogen source, pH) for *B. subtilis* NCIB 8872 strain induced the production of fengycin or plipastatin. In 2001, Akpa *et al*. studied the effects of different media on the nature of the synthesis of fatty acid chains for each lipopeptide family from *B.subtilis* NT02 strain cultivated in optimized medium and Landy medium. Analysis of the lipopeptides showed that

*B. subtilis* NT02 yielded various homologous compounds when cultivated in Landy medium (which contains glucose, glutamic acid and mineral salts), but primarily one homologous product in high relative amounts when cultivated in the optimized medium (which contains saccharose, peptone, yeast extract and mineral salts), which was identified as bacillomycin L c15. However, Jacques *et al*. (1999) studied the fengycin production by *B. subtilis* S499 in fermentor with two different fermentation media. 863 medium yielded 31 mg/L fengycin; whereas the optimized medium yielded 95 mg/L. In addition, Wei *et al*., (2010) indicated that 26.2 g/L of mannitol, 21.9 g/L of soybean meal, 3.1 g/L of $NaNO_3$ and 0.2 g/L of $MnSO_4 \cdot 4H_2O$ represented the optimal medium composition, leading to the highest production of fengycin in *B. subtilis* F29-3. Furthermore, this optimization strategy increased the fengycin production from 1.2 g/L to 3.5 g/L.

## 9. The role of fengycins and plipastatins in biocontrol of plant pathogens

### 9.1. The mode of action of fengycin

The small size of the fengycin peptide part and its strong amphiphilic character facilitate its insertion into the membrane and thus speeds up the process. Deleu *et al*., (2008) investigated the fengycin binding to and/or insertion into a model biomembrane to reveal its perturbing effect and better understand its biological action. Monolayer experiments revealed that the fengycin insertion into a DPPC monolayer occurs very quickly. Also, Deleu *et al*., (2008) have confirmed the penetration ability of fengycin into lipid bilayers and they described that the mode of action of fengycin involved in the bilayer solubilization is dependant on the concentration as follows:

**a) Low fengycin concentration**

At low fengycin concentration (about 10 µM, fengycin/lipid molar ratio = 0.002), where the insertion of fengycin does not significantly affect the phospholipid interfacial assembly, fengycin is dispersed as a monomer into the hydrophobic core of the bilayer.

**b) Medium fengycin concentration**

At medium fengycin concentration (about 133 µM, fengycin/lipid molar ratio = 0.026) the bilayer self-assembled structure is affected, fengycin accumulation within the bilayer leads to self-association of fengycin molecules.

**c) High fengycin concentration**

At high fengycin concentrations (about 2.4mM, fengycin/lipid molar ratio = 0.5) fengycin acts as a detergent by solubilizing the membrane. To conclude, the mechanism of fengycin action is probably based on a two-state transition controlled by the lipopeptide concentration. One state is the monomeric, not deeply anchored, and nonperturbing lipopeptide, and the other state is a buried, aggregated form, which is responsible for membrane leakage and bioactivity.

In addition, Eeman *et al*. (2009) reported that the fengycin insertion at the lipid interface is enhanced in the presence of cholesterol. Cholesterol strongly interacts with fengycin and undergoes specific molecular interactions.

### 9.2. Fengycin role in biocontrol

In the context of biocontrol of plant diseases, the three families of *Bacillus* lipopeptides surfactins, iturins and fengycins were at first mostly studied for their antagonistic activity for a wide range of potential phytopathogens, including bacteria, fungi and oomycetes. Recent investigations have shed light on the fact that these lipopeptides can also influence the ecological fitness of the producing strain in terms of root colonization (and thereby persistence in the rhizosphere) and also have a key role in the beneficial interaction of *Bacillus* species with plants by stimulating host defence mechanisms. The different structural traits and physicochemical properties of these effective surface and membrane active amphiphilic biomolecules explain their involvement in most of the mechanisms developed by bacteria for the biocontrol of different plant pathogens (Ongena *et al*., 2007).

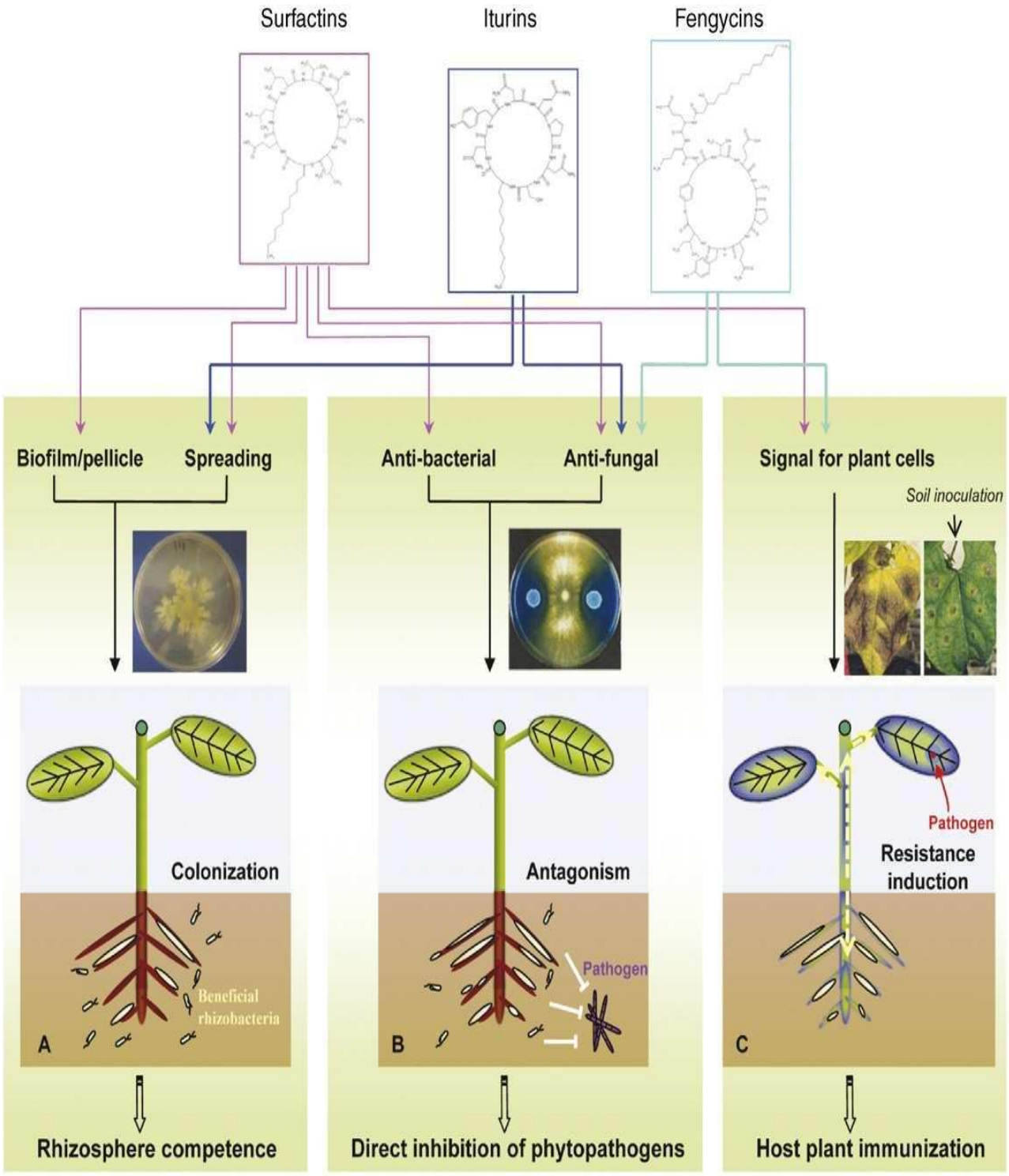



Figure 17: Overview of *Bacillus* lipopeptide interactions in the context of biological control of plant diseases. From left to right, the three photographs show bacterial spreading, fungal growth inhibition through the production of fungitoxic compounds by blue bacterial cells and leaf disease reduction following inoculation of the beneficial bacterium on roots. They illustrate how to get experimental indications about the potential involvement of one particular strain in the three phenomena schematically represented in (A), (B) and (C). Establishment in biofilm and/or microcolonies of the rhizobacterium is represented in (A), (B) represents direct antibiosis that can be exerted by the established biocontrol strain toward pathogens sharing the same microenvironment. In (C), the arrows illustrate the emission of a signal following perception of the rhizobacterium at the root level. This signal moves over the entire plant leading to some systemic reinforcement allowing pathogen restriction at distal sites of infections. (Ongena and Jacques, 2008)

In the control of phyllosphere diseases, a contribution of both iturins and fengycins was recently shown in the antagonism of *B. subtilis* towards *Podosphaera fusca* infecting melon leaves (Romero *et al*., 2007). This was demonstrated by identifying iturins and fengycins as the main antibiotic products excreted by the strains, by showing the strong inhibitory effect of these lipopeptides on *P. fusca* conidia germination, and by recovering lipopeptides from bacterial-treated leaves and using lipopeptide-deficient transformants. *B. subtilis* S499 efficiently produces lipopeptides from the three families, and notably produces a wide variety of fengycins (Jacques *et al*. 1999; Ongena *et al*., 2005). Direct evidence for a role of fengycins in disease reduction derives from experiments showing the ability of strain S499 to protect wounded apple fruits against gray mold disease caused by *Botrytis cinerea*. This role was demonstrated by the effective disease control provided by treatment of fruits with lipopeptide-enriched extracts and by in situ detection of fengycins in inhibitory amounts (Ongena *et al*., 2005). Also, fengycin from *B. subtilis* D1/2 was found to inhibit *Fusarium graminearum* (Chan *et al*., 2009).

In 2004, Ongena *et al.* described the ability of *B. subtilis* strain M4 to reduce disease incidence caused by *Colletotrichum lagenarium* and *Pythium aphanidermatum* on cucumber and tomato, respectively with mixture of lipopeptides from the surfactin, iturin and fengycin families. Also, Touré *et al*., (2004) tested the ability of *B. subtilis* GA1 to reduce post-harvest infection caused by *B. cinerea* was tested on apples. This strain was very effective at reducing disease incidence according to its wide variety production of antifungal lipopeptide isomers from the iturin, fengycin and surfactin families. It was marked also that the fengycins with longer hydrocarbon side chains (C17 and C18) are potentially more bioactive.

In 2006, Sun *et al*. isolated *Bacillus amyloliquefaciens* ES-2 strain, a co-producer of fengycin and surfactin, from the Chinese medicinal plant *Scutellaria baicalensis Georgi*. This strain was found to strongly inhibit the growth of all Gram-positive bacteria tested except *B. subtilis* and all Gram negative ones except *Ochrobactrum anthropi.* Moreover, the filtrate exhibited promising activities against phytopathogenic fungi, such as *Penicillium italicum* (apple rot fungus), *Fusarium culmorum*, *llliciun verum* Hook.f phytopathogenic fungus, *Botrytis cinerea* Pers (tomato phytopathogenic fungus), *Magnaporthe grisea* (rice blast), *Erysiphe graminis hordei* (rice sheath slight). Therefore, it is considered that the antimicrobial substances produced by ES-2 have potential for food preservation and fungal plant disease control. Moreover, Hu *et al*. (2007) isolated fengycin antibiotic from *B. subtilis* B-FS01 culture which

inhibit the growth of *Fusarium moniliforme* Sheldon ATCC38932. Also, Wang *et al*. (2007) detected inhibitory activity of *B. subtilis* strain IB against the *Fusarium* head blight disease fungus and the major inhibitory compound was fengycin. Whereas, Athokorala *et al*. (2009) detected the presence of fengycin synthetic gene in four *Bacillus* strains: *B. subtilis* H-08-02, *B. subtilis* 3057, *B. amyloliquefaciens B. subtilis* S6 and *Bacillus mycoides* 4079. All the strains showed > 50 % mycelia inhibition of *S. sclerotiorum* and (or) F. *graminearum* under laboratory conditions, in addition some have shown significant greenhouse and field control of stem rot disease in canola and fusarium head blight in wheat. In addition, Chen *et al*., (2009) shown that the fungicidal activities of bacillomycin D and fengycin are due to a synergistic action. *B. amyloliquefaciens* FZB42 inhibited growth of plant pathogenic fungi such as *Fusarium* spp. including *F. oxysporum*, *Gaeumannomyces graminis*, *Rhizoctonia solani*, *Alternaria alternatae*, *Botrytis cinerea* and *Pythium aphanidermatum*.

## 10. The role of fengycin in food applications

The increasing trend to limiting the use of chemical food preservatives has generated considerable interest in the use of ‘‘natural’’ alternatives. Several cationic antimicrobial peptides have been studied and applied commercially for food applications including protamine, nisin, and pediocin (Montville *et al*., 1995; Siragusa *et al*., 1999). Nevertheless, some antimicrobial lipopeptides including surfactin, fengycin, iturin, and bacillopeptins produced by *B. subtilis* also had natural characters as above (Kajimura *et al*., 1995; Shu et *al*., 2002; Cho *et al*., 2003). Therefore, their commercial applications would be feasible. Steller and Vater (2000) indicated that both surfactin and fengycin had a broad spectrum of antimicrobial activity, including Gram-positive and Gram-negative bacteria, fungi, protozoa, and viruses.

## 11. An overview of different works of ProBioGEM laboratory in relation with our study

### 11.1. Fengycin production from *B. subtilis* ATCC 21332

Firstly, *B. subtilis* ATCC 21332 a wild type strain was thought to be a mono- surfactin producer, but it was found to be co-producer for both surfactin and fengycin by Gancel *et al*., (2009) whereas they studied the surfactin and/or fengycin batch production by immobilized cells of *Bacillus subtilis* ATCC 21332. Influence of particles on lipopeptide production was analyzed and they found that the addition of carriers seemed to modify the ratio between surfactin and fengycin with an enhancement of the fengycin production. The highest concentration of fengycin was obtained with addition of pellets containing 0.35% of iron. Coutte *et al*. (2010) studied the lipopeptide production of this later strain in bubbleless bioreactors. In a bioreactor equipped with a submerged hollow fiber membrane air-liquid contactor in polypropylene, surfactin production was lower than in the other bioreactors equipped with external membrane contactor, but fengycin production was surprisingly high (982 mg after 72 h of culture). Also, Tapi *et al*. (2009) amplified a fragment similar (99%, e value=0.00) to the plipastatin synthetase gene in *B. subtilis* ATCC 21332, which has not yet been identified as a producer of plipastatin molecules. This result confirms the detection of fengycins in the supernatant of *B. subtilis* ATCC 21332 by Gancel et al. (2009). Recently, *B. subtilis* BBG21, a spontaneous mutant of the *B. subtilis* strain ATCC 21332, was obtained at ProBioGEM. This mutant differs from the wild type strain by ability to produce surfactin and fengycin in higher quantities.

### 11.2. Norine

This project results from a tight collaboration between the SEQUOIA bioinformatics research group of LIFL (Laboratoire d'Informatique Fondamentale de Lille) and INRIA (Institut National de Recherche en Informatique et en Automatique) and the NRPS team from the ProBioGem laboratory (Laboratoire des Procédés Biologiques Génie Enzymatique et Microbien). The researchers have developed several new tools for the study of NonRibosomal Peptides (NRPs) including a database (NORINE, Caboche *et al*., 2008), a tool for NRPs structure comparison (Caboche *et al*., 2009) and a tool for NRPs biological activity prediction (Caboche *et al*., 2010). This bioinformatics software server includes a database of nonribosomal peptides. The website of this project is: http://bioinfo.lifl.fr/norine.

# *Chapter III. Objectives of the work*

Numerous *Bacillus* genomes were recently sequenced and published, so it is interesting as a first aim of this work to get an overview of the Non Ribosomal Peptides produced or potentially produced by these *Bacillus* strains, especially lipopeptides. This work will be done by analysing their genomes using bioinformatics tools to detect the different NRPS genes, to decrypt their modular structure and to predict the structure of the peptide they could synthesize.

Since the concomitant discovery of fengycin and plipastatin in 1986 by the German (Vanittanakom et al. 1986) and Japanese teams (Nishikiori et al. 1986), a doubt still exists today about the differences in the structure of these lipopeptides and in their biological activities. Our second aim will be to focus our work on fengycins or plipastatins produced by the NRPS system in some *B. subtilis* strains, in order to relate the structure of these two molecules and the structure of operons responsible for their biosynthesis. Indeed, Schneider *et al*. (1999) confirmed the existence of fengycin molecule produced by *Bacillus subtilis* S499 strain, which cannot be correlated with the structure of the synthetases described in other fengycin- or plipastatin-producing strains such as *Bacillus subtilis* 168 or *Bacillus amyloliquefaciens* FZB48. Thus partial sequencing of the operon of *Bacillus subtilis* S499 will be carried out to elucidate its modular organization. Similar approaches will be performed to analyse the structure of the plipastatin/fengycin operons from *Bacillus subtilis* ATCC21332 and its spontaneous mutant, BBG21.

Since there are many similarities between fengycin and plipastatin, it could be interesting to produce and purify it from *B. subtilis* 168. A third objective will be thus to get a plipastatin overproducing strain by exchanging the native promoter of plipastatin operon with a strong and constitutive one $P_{repU}$, which will (i) facilitate purification of plipastatin only in significant amounts, and (ii) studies on plipastatin role as an antifungal antibiotic in biocontrol experiments, and its potential use as a biopesticide.

It was recently shown in ProBioGEM that *Bacillus subtilis* BBG21 is able to produce high amounts of fengycin in specific conditions. As a last objective, the strength of the fengycin promoter of this strain and its wild-type mother strain *B. subtilis* ATCC21332 will be compared in different cultural conditions by using a reporter gene.

# *Chaper IV. Materials and methods*

## 1. Materials

### 1.1. Strains

#### 1.1.1. *Bacillus subtilis* 168 and its derivatives

*B. subtilis* 168 was used in this study as its whole genome was sequenced and the data are available. Because of its high specificity and efficiency of genetic transformation, it is a widely used recipient for recombinant DNA. The *B. subtilis* 168 genome project revealed two large operons which encode for nonribosomal peptide synthetases: the surfactin and plipastatin operons. The strain 168 *trpC2* harbours an inactive *sfp* allele ($sfp^0$), and thereby cannot produce lipopeptides. It also bears an inactive *degQ* gene ($degQ^0$), thought to correspond to a pleiotropic regulator (Tsuge *et al.*, 1999). The *sfp* gene encodes a 4-phosphopantetheinyl transferase, which converts inactive apoenzyme peptide synthetases to their active holoenzyme forms by posttranslational transfer of the 4-phosphopantetheinyl moiety of coenzyme A to the synthetases.

Two *B. subtilis* 168 derivatives (BBG111 and BBG127) were from the collection of the ProBioGEM laboratory; strain 168 was genetically modified by the introduction of a functional *sfp* gene which resulting in both surfactin and plipastatin productions (strain BBG111). The strain BBG127 bears the *lacZ* ORF fused to the $P_{srfA}$ promoter of *B. subtilis* 168 (Coutte *et al*., 2010).

#### 1.1.2. *Bacillus subtilis* ATCC 21332

This strain is a wild type strain thought to be a mono- surfactin producer, but it was found to be co-producer for both surfactin and fengycin (Gancel *et al*., 2009; Coutte *et al*., 2010; Tapi *et al*., 2009).

#### 1.1.3. *Bacillus subtilis* BBG21

This strain is a spontaneous mutant of the *B. subtilis* strain ATCC 21332. It produces surfactin and fengycin in higher quantities than the mother strain ATCC 21332.

### 1.1.4. *Bacillus subtilis* S499

This strain was a gift from L. Delcambe (CNPEM, Liège, Belgium). It produces three lipopeptides; surfactin, iturin and fengycin. Schneider *et al*. (1999) confirmed that the fengycin molecule synthesized cannot be correlated with the structure of the synthetases described in other fengycin- or plipastatin-producing strains.

### 1.1.5. *Escherichia coli* JM109

This strain was used to the amplification and the conservation of the different plasmids.
JM109 is a high-efficiency competent strain (≥1 × 10 cfu/µg DNA) for transformations.
The genotype is *recA1*, *endA1*, *gyrA96*, *thi*, *hsdR17* (rK–,mK+), *relA1*, *supE44*, Δ(*lac-proAB*), [F′, *traD36*, *proAB*, *lac*IqZΔM15] (4)

Table 3: Different used and constructed *Bacillus subtilis* strains

| Name | Genotype | Phenotype | Lipopeptide production | Reference |
|---|---|---|---|---|
| **168** | trpC2, *sfp*$^0$ | Trp$^-$ , sfp$^-$ | - | ProBioGEM |
| **21332** | *sfp*$^+$ | - | Srf, Fen | ProBioGEM |
| **BBG21** | *sfp*+ | - | Srf, Fen | ProBioGEM |
| **S499** | *sfp*$^+$ | - | Srf, Fen, Itu | (Jacques *et al*., 1999) |
| **BBG111** | trpC$_2$, *amyE* ::*sfp-cat* | Trp$^-$, amy$^-$, Cm$^R$ | Srf, Fen | (Coutte *et al*., 2010) |
| **BBG127** | trpC$_2$, *amyE* ::P$_{srfA}$-*lacZ-cat* | Trp$^-$, amy$^-$, Cm$^R$ | - | (Coutte *et al*., 2010) |
| **BBG134** | trpC$_2$, *amyE* ::*sfp-cat*, *srfAA-Tet* | Trp$^-$, amy$^-$, Cm$^R$, Tet$^R$ | Fen | ProBioGem |
| **BBG142** | trpC$_2$, *amyE*:: *pps* 168-*lacZ-cat* | Trp$^{--}$, amy$^-$, Cm$^R$ | - | This study |
| **BBG139** | trpC$_2$, *amyE*:: *pps* BBG21-*lacZ-cat* | Trp$^{--}$, amy$^-$, Cm$^R$ | - | This study |
| **BBG140** | trpC$_2$, *amyE* :: *sfp- cat* , (*neo* -P$_{repU}$)::*ppsA* | Trp$^{--}$, amy$^-$, Cm$^R$, Nm$^R$ | Srf | This study |

## 1.2. Vectors and plasmids

All vectors and plasmids used are in this study were listed in table 2.

### 1.2.1. pGEM-T Easy Vector

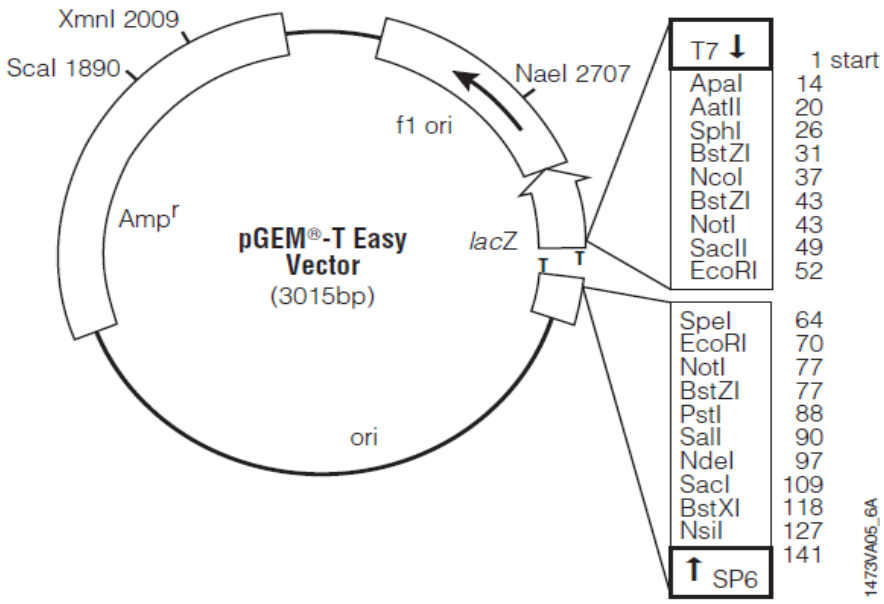



Figure 18: pGEM-T Easy vector with different restriction sites

pGEM-T Easy vector is a commercial plasmid from Promega (Leiden, The Netherlands). It is a 3 kb plasmid designed for the cloning of PCR products. pGEM®-T Easy Vector is high-copy-number vectors containing T7 and SP6 RNA polymerase promoters flanking a multiple cloning region within the α-peptide coding region of the enzyme β-galactosidase. Insertional inactivation of the α-peptide allows identification of recombinants by blue/white screening on indicator plates.

### 1.2.2. pDG1661 plasmid

The plasmid pDG1661 (figure 19) [Bacillus Genetic Stock Center (BGSC): ECE112, Gene Bank: U46196] was designed to integrate a *lacZ* fusion into the *B. subtilis 168* chromosome at the *amyE* locus. The user inserts the promoter-containing fragment into the multiple cloning site to create *lacZ* fusion. The plasmid is transformed into *B. subtilis* 168 derivatives with a selection for chloramphenicol resistance.

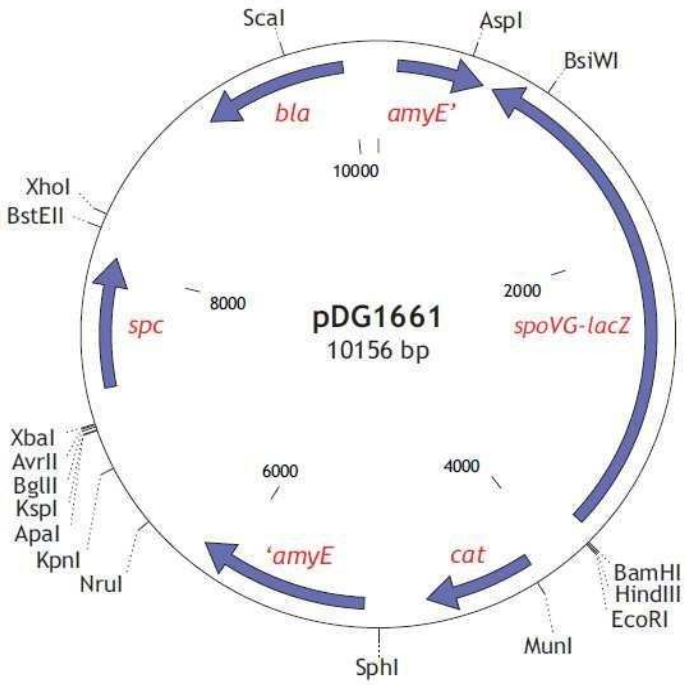



Figure 19: The integration plasmid pDG1661

### 1.2.3. pBG106 plasmid

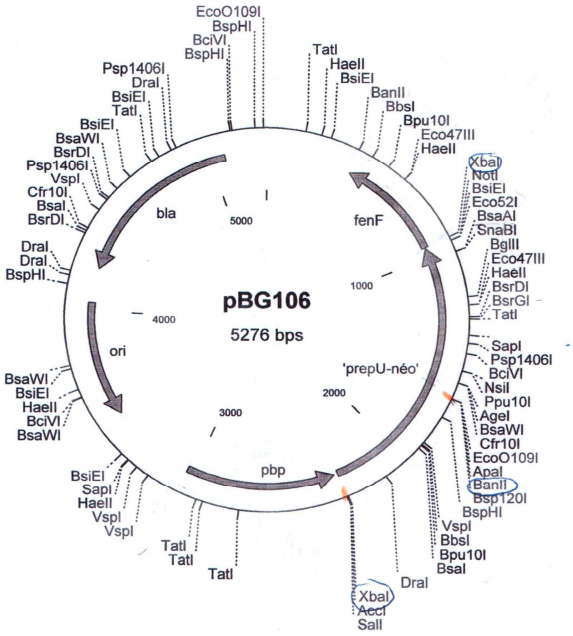



Figure 20: The pBG106 plasmid.

The plasmid pBG106 was derived from both *Staphylococcus aureus* plasmid pUB110 and vector pUC19. It was used to obtain the BBG100 derivative of *B. subtilis* ATCC6633, which overproduces mycosubtilin (Leclère *et al*., 2005).

Table 4: Different plasmids used and constructed in this study

| **Name** | **Resistance** | **Insert** | **Source** |
|---|---|---|---|
| pGEM -T Easy cloning vector | $Ap^R$ | - | Promega |
| pDG1661 integration vector | $Ap^R$ $Cm^R$ $Spc^R$ | - | (Guerout-Fleury *et al*., 1996) |
| pBG106 | $Ap^R$ $Nm^R$ | ε*pbp*, $P_{repU}$-*neo*, ε*fenF* | (Leclère *et al*. 2005) |
| pBG184 | $Ap^R$ $Cm^R$ $Spc^R$ | 0.34 kb *Hind*III-*EcoR*I P*fen* BBG21 fragment from pBG185 inserted into pDG1661 | This study |
| pBG177 | $Ap^R$ | 0.63 kb *ppsA* cassette 2 fragment cloned into pGEM-T Easy | This study |
| pBG178 | $Ap^R$ $Nm^R$ | 0.61 kb *dacC* cassette 1 fragment cloned into pGEM-T Easy | This study |
| pBG179 | $Ap^R$ | 0.61 kb *Aat*II-*Nco*I *dacC* cassette1 from pBG178 inserted into pBG177 | This study |
| pBG180 | $Ap^R$ $Nm^R$ | 1.34 kb *Xba*I $P_{repU}$-*neo* fragment from pBG106 inserted into pBG179 | This study |
| pBG185 | $Ap^R$ $Cm^R$ | 0.34 kb *Hind*III-*EcoR*I *pps* 168 from pBG184 inserted into pDG1661 | This study |
| pBG182 | $Ap^R$ $Cm^R$ | 0.621 kb *Xba*I-*Ban*II *ppsA* cassette 2 fragment from pBG177 inserted into pBG180 | This study |
| pBG183 | $Ap^R$ | 0.456 kb *ppsA* cassette 2N fragment cloned into pGEM-T Easy | This study |
| pBG176 | $Ap^R$ | 0.34 kb *Ppps* 168 fragment cloned into pGEM-T Easy | This study |
| pBG181 | $Ap^R$ | 0.34 kb P*fen* BBG21 fragment cloned into pGEM-T Easy | This study |
| pBG186 | $Ap^R$ | 0.72 kb *dacC*- P*pps* 168 fragment cloned into pGEM-T Easy | This study |
| pBG187 | $Ap^R$ | 0.604 kb *dacC* 168 cloned into pGEM-T Easy | This study |
| pBG188 | $Nm^R$ | 0.60 kb *Sph*I-*Sa1*I *dacC* 168 fragment from pBG186 cloned into pBG106 | This study |
| pBG189 | $Nm^R$ | 0.72 kb *Sph*I-*Sa1*I *dacC*-P*pps* 168 fragment from pBG186 cloned into pBG106 | This study |
| pBG190 | $Nm^R$ | 1.43 kb *Xba*I $P_{repU}$*neo* gene from pBG106 inserted into pBG189 | This study |

| | | | |
|---|---|---|---|
| pBG191 | $Nm^R$ | 0.456 kb *Sa1*I-*Nde*I *ppsA* cassette 2N 168 fragment from pBG183 cloned into pBG180 | This study |
| pBG192 | $Spc^R$ | 0.78 kb *Nco*I-*Sac*II *spc* gene from pBG214 inserted into pBG192 | This study |
| pBG193 | $Ap^R$ | 0.34 kb *EcoR*I P*fen* BBG21 fragment from pBG185 cloned into pBG191 | This study |
| pBG194 | $Ap^R$ | 0.45 kb fengycin primers BBG21 fragment cloned into pGEM-T Easy | This study |
| pBG195 | $Ap^R$ | 0.95 kb plipastatin primers BBG21 fragment cloned into pGEM-T Easy | This study |
| pBG196 | $Ap^R$ | 1.6 kb *ppsD* BBG21 fragment cloned into pGEM-T Easy | This study |
| pBG197 | $Ap^R$ | 1.2 kb *ppsB* BBG21 fragment cloned into pGEM-T Easy | This study |
| pBG198 | $Ap^R$ | 1.5 kb *ppsE-yngL* BBG21 fragment cloned into pGEM-T Easy | This study |
| pBG199 | $Ap^R$ | 0.76 kb *ppsA-ppsB* BBG21 fragment cloned into pGEM-T Easy | This study |
| pBG201 | $Ap^R$ | 1.074 kb *ppsC-ppsD* BBG21 fragment cloned into pGEM-T Easy | This study |
| pBG202 | $Ap^R$ | 1.2 kb *ppsB* 21332 fragment cloned into pGEM-T Easy | This study |
| pBG203 | $Ap^R$ | 1.6 kb *ppsD* 21332 fragment cloned into pGEM-T Easy | This study |
| pBG204 | $Ap^R$ | 1.5 kb *ppsE-ynLl* 21332 fragment cloned into pGEM-T Easy | This study |
| pBG205 | $Ap^R$ | 0.76 kb *pps*A-*ppsB* 21332 fragment cloned into pGEM-T Easy | This study |
| pBG206 | $Ap^R$ | 1.074 kb *ppsC-ppsD* 21332 fragment cloned into pGEM-T Easy | This study |
| pBG207 | $Ap^R$ | 0.85 kb plipastatin primers 21332 fragment cloned into pGEM-T Easy | This study |
| pBG213 | $Ap^R$ | 0.45 Kb fengycin primers 21332 fragment cloned into pGEM-T Easy | This study |
| pBG215 | $Ap^R$ | 0.45 kb fengycin primers S499 fragment cloned into pGEM-T Easy | This study |
| pBG216 | $Ap^R$ | 0.45 kb plipastatin primers S499 fragment cloned into pGEM-T Easy | This study |
| pBG217 | $Ap^R$ | 0.76 kb *ppsA-ppsB* S499 fragment cloned into | This study |

| | | | |
|---|---|---|---|
| | | pGEM-T Easy | |
| pBG218 | $Ap^R$ | 2.08 kb *ppsB-ppsC* S499 fragment cloned into pGEM-T Easy | This study |
| pBG219 | $Ap^R$ | 1.074 kb *ppsC-ppsD* S499 fragment cloned into pGEM-T Easy | This study |
| pBG220 | $Ap^R$ | 2.77 kb *ppsD-ppsE* S499 fragment cloned into pGEM-T Easy | This study |
| pBG221 | $Ap^R$ | 9 kb *ppsA- ppsC* S499 fragment cloned into pGEM-T Easy | This study |
| pBG222 | $Ap^R$ | 10 kb *pps*C- *ppsE* S499 fragment cloned into pGEM-T Easy | This study |

Table 5: Different *E. coli* strains constructed in this study

| Name of strain | Resistance | Cloned plasmid |
|---|---|---|
| EBG173 | $Ap^R$ | pBG177 |
| EBG174 | $Nm^R$ | pBG178 |
| EBG175 | $Ap^R$ | pBG179 |
| EBG176 | $Nm^R$ | pBG180 |
| EBG177 | $Cm^R$ | pBG184 |
| EBG178 | $Cm^R$ | pBG185 |
| EBG179 | $Nm^R$ | pBG182 |
| EBG180 | $Ap^R$ | pBG183 |
| EBG181 | $Ap^R$ | pBG176 |
| EBG182 | $Ap^R$ | pBG181 |
| EBG183 | $Ap^R$ | pBG187 |
| EBG184 | $Nm^R$ | pBG191 |
| EBG185 | $Ap^R$ | pBG192 |
| EBG186 | $Spc^R$ | pBG193 |
| EBG187 | $Ap^R$ | pBG194 |
| EBG188 | $Ap^R$ | pBG195 |
| EBG189 | $Ap^R$ | pBG196 |
| EBG190 | $Ap^R$ | pBG201 |
| EBG191 | $Ap^R$ | pBG202 |
| EBG192 | $Ap^R$ | pBG203 |
| EBG193 | $Ap^R$ | pBG204 |
| EBG194 | $Ap^R$ | pBG205 |
| EBG195 | $Ap^R$ | pBG206 |
| EBG196 | $Ap^R$ | pBG207 |
| EBG197 | $Ap^R$ | pBG213 |
| EBG198 | $Ap^R$ | pBG218 |
| EBG199 | $Ap^R$ | pBG219 |
| EBG200 | $Ap^R$ | pBG220 |
| EBG201 | $Ap^R$ | pBG221 |
| EBG202 | $Ap^R$ | pBG222 |

### 1.3. Culture media

#### 1.3.1. Liquid culture media

**Landy MOPS medium** (Landy *et al*., 1948)**:** 1 L is composed of: 20 g glucose, 5 g glutamic acid, 0.5 g $MgSO_4.7H_2O$, 0.5 g KCl, 1 g $K_2HPO_4$, 1 g yeast extract , 0.4 mg $MnSO_4$, 1.2 mg $Fe_2SO_4$and 1.6 mg $CuSO_4$, 100 mM 3-[N-morpholino]-propane sulfonic acid (MOPS) as a buffer. The pH was adjusted to 7.0 with 5 M KOH and all the solutions autoclaved by sterilization at 110°C for 30 min, except the glucose and the glutamic acid solutions which were sterilized by filtration.

**Landy Modified medium 1** (Chollet-Imbert *et al*., 2009): ammonium sulphate 2.2 g/L was added to Landy MOPS medium.

**Landy Modified medium 2:** Landy medium with only 2.5 g/L glutamic acid (not 5g) was complemented by 2.2 g/L of ammonium sulphate.

**Landy Modified medium 3:** Landy Mops medium modified for the nitrogen source (1.82 g $NH_4Cl$ instead of glutamic acid) Volpon *et al*., (2000).

**Optimized medium:** 20 g sucrose, 30 g peptone, 7 g yeast extract, 1.9 g $KH_2PO_4$, 0.450 g $MgSO_4$, 9 mL trace elements solution (composition of trace elements per litre of distilled water: 0.001 g of $CuSO_4$, 0.005 g of $FeCl_3$, 0.004 g of $NaMnO_4$, 0.002 g of KI, 0.014 g of $ZnSO_4$, 0.01 g of $H_3BO_3$, 0.0036 g of $MnSO_4$, 10 g of citric acid).

**LB medium** (Luria-Bertani) (Sambrook and Russel, 2001): composed of 10 g bactotryptone, 5 g yeast extract and 10 g NaCl. The pH is adjusted to 7.0 with NaOH. Then, it is sterilized at 121°C for 20 min.

**SOB medium:** is composed of 20 g bactotryptone, 5 g yeast extract, 10 mM NaCl and 2.5 mM KCl.

**SOC medium:** is composed of SOB medium complemented with 20 mM glucose, 20 mM $Mg^{+2}$ (10 mM $MgSO_4$ + 10 mM $MgCl_2$).

**MS1 medium:** is composed of 10 mL 10X medium [for 100 mL, 150 mM $(NH_4)_2SO_4$, 34 mM Tri-Na citrate, 800 mM $K_2HPO_4$, 440 mM $KH_2PO_4$], 9 mL distilled water, 100 µL 50% (w/v) glucose, 40 µL 5%,(w/v) casein hydrolysate, 200 µL 5% (w/v) yeast extract, 16.7 µL 1M $MgSO_4$, 50 µL 50 mM tryptophan.

**MS2 medium**: composed of MS1 medium + 50 mM $CaCl_2$ + 1 M $MgSO_4$

### 1.3.2. Solid culture media

**LB solid:** the same composition of liquid LB with the addition of 17 g of agar per litre.

**LB-Starch:** the same composition of LB liquid with the addition of 1% starch (w/v).

### 1.3.3. Buffers and solutions

**TBE buffer:** to prepare 5X TBE buffer in 1 litre, dissolve 53.9 g Tris (hydroxyl methyl amino-methane), 27.5 g Boric acid and 3.65 g EDTA (pH 8.8).

**MEB buffer**: A solution of 1 mM $MgCl_2$ and 1mM HEPES was adjusted the pH to 7.0; autoclaved at 120 ºC and stored at 4ºC.

**Phosphate buffer**: Mixture of 0.06 M $Na_2HPO4.7H_2O$, 0.04 M $NaH_2PO_4.H_2O$, adjusted to pH 7.0

**Z buffer**: the same composition as phosphate buffer complemented with 0.01 M KCl, 0.001 M $MgSO_4$, 0.05 M β- mercaptoethanol (BME), adjusted to pH 7.0 and stored at 4ºC.

**ONGP solution**: a fresh solution was prepared daily by dissolution of 4 mg/ mL o-nitrophenyl-β-D-galactopyranoside in phosphate buffer**.**

## 1.4. Bioinformatics tools

### 1.4.1. Amplifx software 1.4.4.

The main purpose of Amplifx is to seek in a collection of primers, those which can be used to amplify a fragment into a target sequence, and particularly, to design strategies to screen recombinant clones by PCR. Some information are associated with each primer-automatically computed by Amplifx (like TM, Quality, length) other data are given by the user (name, comments). This allows general aspects of primer management. Amplifx can also design new primers.

### 1.4.2. NRPS software (ProBioGEM laboratory)

NRPSsoftware is a tool dedicated to the recognition of different domains of synthetases and active sites based on the signature motifs of conserved protein. At the moment, this software is not accessible to everyone, because improvements are underway.

### 1.4.3. Web sites

**National Centre for Biotechnology Information (NCBI)**

This American site is the highly site in scientists citation because it contains the GenBank data base available with about 108,431,692 sequences recorded. It is accessible on **http://www.ncbi.nlm.nih.gov/genbank/**

**Primer 3**

This site for designing primers under the development of Primer3 and the Primer3 web site was funded by Howard Hughes Medical Institute and by the National Institute of Health and the National Human Genome Research Institute under grants R01-HG00257 (to David C. Page) and P50-HG00098 (to Eric S. Lander). It is accessible on **http://frodo.wi.mit.edu/primer3/**

**NEB cutter**

This tool will take a DNA sequence and find the large, non-overlapping open reading frames using the *E. coli* genetic code and the sites for all Type II and commercially available Type III restriction enzymes that cut the sequence. It is one of the services of the New England Bio Lab Insurance. It is accessible on **http://tools.neb.com/NEBcutter2/**

**PKS- NRPS analysis**

This project is still a work in progress of the Institute of Genome Sciences University of Maryland School of Medicine (Baltimore, MD 21201), to facilitate the research and analysis of the genes correspond to the NRPS or PKS systems (Bachman *et al*., 2009). It is accessible on **http://nrps.igs.umaryland.edu/nrps/**

**EMBOSS pairwise alignment**

The alignment of operons or sequences was realized by the help of EMBOSS (Needle, Global) Pairewise Alignment Algorithms (Needleman and Wunsch, 1970). The alignment was performed under the following conditions; DNA; open Gap: 10, Extend Gap: 0.5; Protein; Matrix: Blosum62, Open Gap: 10, Extend Gap: 0.5. This software is from the EMBL-EBI (European Molecular Biology Laboratory in European Bioinformatics Institute), The cornerstones of EMBL's mission are: to perform basic research in molecular biology, to train scientists, students and visitors at all levels, to offer vital services to scientists in the member states, and to develop new instruments and methods in the life sciences, and technology transfer. It is accessible on http://www.ebi.ac.uk/Tools/emboss/align/index.htmL

## 2. Methods

### 2.1. Molecular genetic techniques

#### 2.1.1. DNA Isolation

*B. subtilis* genomic DNAs strains were isolated using Wizard®Genomic DNA Purification Kit from Promega Corp (Madison, Wi, U.S.A). The isolation depends on four steps; the first step is to lyse the cells and the nuclei by nuclei lysis solution, then the RNA is digested by RNase. The cellular proteins are then removed by a salt precipitation step, which leaving the high molecular weight genomic DNA in solution. Finally, the genomic DNA is concentrated and desalted by isopropanol precipitation and dissolved in a TE buffer.

#### 2.1.2. PCR technique

##### 2.1.2.1. PCR mixture

PCR Master Mix (2X) from Fermentas GmbH (st. Leon. Rot, Germany) was used as a mixture of Taq DNA polymerase 0.5 units/ µL in reaction, PCR buffer; 4 mM $MgCl_2$, 0.4 mM of each dNTPs.

Table 6: PCR Mixture

| | |
|---|---|
| **PCR Master Mix (2X)** | 12.5 µL |
| **20 µM Fwd Primer** | 1.5 µL |
| **20 µM Rev Primer** | 1.5 µL |
| **DNA sample** | 50 ng |
| **$dH_2O$** | Up to 25 µL |

### 2.1.2.2. PCR Conditions and designed primers

Denaturation was performed at 94°C for 1 min and hybridization from 47.8 to 53°C for 30 sec. Time of elongation at 72°C depends on every fragment size, from 340 bp to 12283 bp. PCR was performed for 30 cycles.

Table 7: Different designed primers used for the detection of *B. subtilis* S499 fengycin operon

| Primers | Primer sequence | Fragment length (bp) | Hybridization temperature |
|---|---|---|---|
| ***ppsA-ppsB* sense** | 5′ TGCTGTTKGAAACVCAGCA′3 | 760 | 47.8 |
| ***ppsA- ppsB* antisense** | 5′GATAAMGGRTAAATATCTTG′3 | | |
| ***ppsB-ppsC* sense** | 5′TGTTYGARCATCAGACDCT′3 | 2084 | 50 |
| ***ppsB- ppsc* antisense** | 5′TGCTMCATSAGCCATTTAATATAC′3 | | |
| ***ppsC-ppsD* sense** | 5′GATACRCAAGATATTCTDYT′3 | 1073 | 51 |
| ***ppsC- ppsD* antisense** | 5′TCYTTTTGCARATGAAACCC′3 | | |
| ***ppsD-ppsE* sense** | 5′GAAMYGAAAGAAATTGAATC′3 | 2767 | 48.3 |
| ***ppsD- ppsE* antisense** | 5′TTTCAGCTTGTCYGTCAG′3 | | |
| ***ppsA-ppsB* sense** | 5′ TGCTGTTKGAAACVCAGCA′3 | 8990 | 52 |
| ***ppsB- ppsc* antisense** | 5′TGCTMCATSAGCCATTTAATATAC′3 | | |
| ***ppsB-ppsC* sense** | 5′TGTTYGARCATCAGACDCT′3 | 9578 | 53 |
| ***ppsC- ppsD* antisense** | 5′TCYTTTTGCARATGAAACCC′3 | | |
| ***ppsC-ppsD* sense** | 5′GATACRCAAGATATTCTDYT′3 | 12283 | 49.5 |
| ***ppsD- ppsE* antisense** | 5′TTTCAGCTTGTCYGTCAG′3 | | |

Table 8: Different designed primers used for the detection of *B. subtilis* 21332 and BBG21 fengycin or plipastatin operons

| Primers | Primer sequence | Fragment length (bp) | Hybridization temperature |
|---|---|---|---|
| ***dacC* complete sense**<br>***dacC* complete antisense** | 5′AAAGCTTTATGTTGCTGTTTTACTG′3<br>5′TTATTGATTTGCCAAAATGACA′3 | 1462 | 50 |
| ***dacC-ppsA* sense**<br>***dac*C- *ppsA* antisense** | 5′TCAAATCGCTGTCATTTTGG′3<br>5′TCAAGCTCAGCGTATGTCAG′3 | 1882 | 50 |
| **Promoter sense**<br>**Promoter antisense** | 5′TAATAGCGGGAGACCTGTTTTC′3<br>5′AAAAAGCAGAAAAATGACAATAAAAGA′3 | 340 | 50 |
| ***ppsD* sense**<br>***ppsD* antisense** | 5′AAAACGTTAAAAGCAATAACGTGTCA′3<br>5′GATCATCAAAATCACTGATCGTC′3 | 1200 | 50 |
| ***ppsE-yngL* sense**<br>***ppsE-yngL* antisense** | 5′GGTGCGTCTAGAGCGGATG′3<br>5′GGCAAGGGCTCAAATAAAGA′3 | 1510 | 50 |
| ***yngL* sense**<br>***yngL* antisense** | 5′ATGGACCGTCTTTCTTTTCTAA′3<br>5′TTAGGACTTTTTATCATTCATCATGG′3 | 393 | 50 |

Table 9: Different designed primers used for the new constructions

| Primers | Primer sequence | Fragment length (bp) | Hybridization temperature |
|---|---|---|---|
| ***dacC* sense**<br>***dacC* antisense** | 5′GCATGCAGCTTTATGTTGCT′3<br>5′GGGTCGACCTCAACAATCAC′3 | 604 | 50 |
| ***dacC* -Promoter sense**<br>***dacC* -Promoter antisense** | 5′GCATGCCGGATATGCCTTAG′3<br>5′GTCGACAAAGCGTTATTACAGACA′3 | 720 | 50 |
| ***dacC* cassette 1 sense**<br>***dacC* Cassette antisense** | 5′GACGTCAAGACGGGTGAAG′3<br>5′TCCCATGGAAAACAGGTCTC′3 | 617 | 50 |
| ***ppsA* Cassette 2 sense**<br>***ppsA* Cassette 2 antisense** | 5′TGGATTATCTAGACATATAATTTCTTT′3<br>5′GAGCTCAAGTAAGAAGGTTCC′3 | 621 | 50 |
| ***ppsA* Cassette 2N sense**<br>***ppsA* Cassette 2N antisense** | 5′GTCGACATTTTCTTATCCTCTTATTATG′3<br>5′CATATGATGGAATTTCGCAAATAG′3 | 456 | 50 |
| ***pps* Promoter sense**<br>***pps* Promoter antisense** | 5′TAATAGCGGGAGACCTGTTTTC′3<br>5′AAAAAGCAGAAAAATGACAATAAAAGA′3 | 340 | 50 |

All the primers were designed using the published sequence of *B. subtilis* 168 plipastatin operon (accession no. AL009126), except the four degenerated primers for *B. subtilis* S499

which were designed by the alignment of three *Bacillus* plipastatin and fengycin operons from *B. subtilis* 168, *B. subtilis* F29-3 (accessions no. AF023464, L42523, AF087452, AJ011849, AF023465) and *B. amyloliquefaciens* FZB42 (accession no. AJ576102).

### 2.1.3. DNA quantification

DNA samples concentrations were measured using Spectrophotometer WPA Lightwave II.

### 2.1.4. Agarose gel electrophoresis

Agarose was dissolved in 100 mL 0.5X TBE buffer pH 8.8 Gel concentrations depend on fragment sizes from 0.7% for large fragments (5-10 kb) to 1.2 % for the small fragments (0.5-2 kb). Gels were coloured by addition of Gel Red 10 µL/100 mL.

Gels were run in an electrophoresis unit using 0.5X TBE buffer pH 8.8 at 100 volts for two hours. Gels were photographed using Gel Documentation system.

### 2.1.5. Gel Extraction

The different obtained fragments were extracted and purified using Gel extraction QIAquick kit from QIAGEN (Hildon, Germany). The QIAquick system combines the convenience of spin-column technology with the selective binding properties of a uniquely-designed silica-gel membrane. Special buffers provided with each kit are optimized for efficient recovery of DNA and removal of contaminants in each specific application. DNA adsorbs to the silica-membrane in the presence of high salt while contaminants pass through the column. Impurities are efficiently washed away, and the pure DNA is eluted with Tris buffer or water.

### 2.1.6. Cloning

All obtained fragments were cloned in pGEM-T Easy vector from Promega Corp. (Madison, Wi, U.S.A) according to the following procedure:

#### 2.1.6.1. Optimizing insert: Vector molar ratios for ligation reaction

Ratio from 3:1 gave good initial parameters. pGEM-T Easy Vector is approximately 3 kb in size and is supplied at 50 ng/µl. The following equation was used to calculate the appropriate amount of PCR product (insert):

$$\frac{\text{Vector (ng)} \times \text{size of insert (kb)}}{\text{Size of vector (kb)}} \times (\text{insert (3): (1) vector molar ratio}) = \text{ng of insert}$$

The ligation mixtures used are in table 10.

Table 10: Ligation mixture of pGEM-T Easy vector system

| **Reaction Component** | **Standard Reaction** | **Positive Control** | **Background Control** |
|---|---|---|---|
| 2X Rapid Ligation Buffer, T4 DNA Ligase | 5 µl | 5 µl | 5 µl |
| pGEM-T Easy Vector (50 ng) | 1 µl | 1 µl | 1 µl |
| PCR product | X µl* | – | – |
| Control Insert DNA | – | 2µl | – |
| T4 DNA Ligase (3 Weiss units/µl) | 1 µl | 1 µl | 1 µl |
| Nuclease-free water to a final volume of | 10 µl | 10 µl | 10 µl |
| *Molar ratio of PCR product: vector may require optimization. | | | |

Ligation reactions were mixed well and incubated overnight at 16 ºC for the maximum number of transformants.

### 2.1.7. Transformation protocol

Transformation protocol was performed as described into pGEM®-T and pGEM®-TEasy Vector Systems Technical manual from QIAGEN as follows;

2 µL of each ligation reaction added to a sterile (17 × 100 mm) polypropylene tube or a 1.5 mL microcentrifuge tube on ice. Set up another tube on ice with 0.1 ng uncut plasmid for determination of the transformation efficiency of the competent cells. The JM109 High Efficiency Competent Cells tube removed of frozen from storage and placed in an ice bath until just thawed (about 5 min), then 50 µL of cells are transferred carefully into each tube prepared before. The tubes are flicked gently to mix and place them on ice for 20 min, then heat-shock for 45–50 s in a water bath at exactly 42°C. The tubes are returned immediately to ice for 2 minutes, after that, 950 µL room-temperature SOC medium are added to the tubes containing cells transformed with ligation reactions and 900 µL to the tube containing cells transformed with uncut plasmid. The tubes are incubated for 1.5 h at 37°C with shaking (~150 rpm) and 100 µL of each transformation culture are plated onto duplicate LB/ampicillin/IPTG/ X-Gal plates. For the transformation control, a 1:10 dilution with SOC medium is recommended for plating. The plates incubated overnight (16–24 h) at 37°C. If 100 µl are plated, approximately 100 colonies per plate are routinely seen using competent cells that are $1 \times 10^8$ cfu/µg DNA.

### 2.1.8. Transformation methods for *Bacillus subtilis*

#### 2.1.8.1. Transformation by natural competence

An overnight culture of the strain was prepared in LB (10 mL) and incubated at 37ºC under 140-200 rpm shaking. The next day, about (0.7 O.D 600nm) was calculated of the overnight culture, and was centrifuged at 4000 rpm for 10 min, and then the pellet was washed with sterilized $H_2O$ and re-centrifuged. 5 mL of MS1 solution was added to the pellet and incubated about 5 h at 37ºC (140- 200 rpm). The culture was diluted (1/10 volume) with MS2 solution and incubated 1 h 30 min at 37ºC (140- 200 rpm). About 5 µL of plasmid (100 ng/ µL) was added to the competent cells and the mixture incubated for 30 min minimum in a water bath at 37ºC. The samples were then plated on LB with the selective antibiotic(s) and incubated for 24 h at 37ºC; 50 µg/mL ampicillin, chloramphenicol 5 µg/mL, spectinomycin 50 µg/mL, 20 µg/mL neomycin.

#### 2.1.8.2. Transformation by electroporation

A culture in LB (10 mL) of the strain was prepared and incubated overnight at 37ºC at 140 rpm. 2 mL of this culture were inoculated in 100 mL of fresh LB and incubated at 37ºC and 140 rpm up to an O.D600 of 0.5-1 (about 3 hours). Then 20 mL of this culture were cooled on ice-water for 10 min and centrifuged in a sterile tube at 5240 g for 10 min at 4ºC. The supernatant was then discarded; the cells were suspended in 20 mL of cold MEB buffer and centrifuged. Previous step was repeated once and the supernatant discarded. The cells were re-suspended in 1 mL cold MEB and transferred into a sterile Eppendorf tube and centrifuged at 5240 g for 5 min at 4 ºC. The supernatant was discarded and the cells were suspended in 100 µL of cold MEB buffer. About 150 ng of the plasmid were added and mixed then, the tubes were placed on ice for 5 min. The tubes were transferred into a pre-chilled electroporation cuvette and exposed to a single electrical pulse for 3 ms using Bio-RAD Gene Pulser set at 2.5 KV, 25 µF and 200 Ω. 900 µL SOC medium were added and the mixture incubated for 90 min at 37 ºC. Then, the cells are spread on LB plates with the selective antibiotic for 24 h at 37 ºC.

### 2.1.9. Plasmid purification

Plasmids were isolated using QIAprep® Miniprep kit from QIAGEN. The QIAprep miniprep procedure is based on alkaline lysis of bacteria followed by adsorption of DNA onto silica in the presence of high salt solution. The unique silica membrane used in QIAprep Miniprep Kits completely replaces glass or silica slurries for plasmid minipreps. The procedure consists of three basic steps: preparation and clearing of a bacterial lysate, adsorption of DNA onto the QIAprep membrane, washing and elution of plasmid DNA. All steps are performed without the use of phenol, chloroform, CsCl, ethidium bromide, and without alcoholic precipitation.
Plasmid concentrations were measured using Spectrophotometer WPA LIGHTWAVE II and the hybrid plasmids were sent to sequencing at a concentration of 50-100 ng/µL (minimum 15 µL).

### 2.1.10. RNA extraction and RT-PCR

RNA extraction was performed by RNA-later kit of Ambion. RNA *later* ® (Applied Biosystems, Courtaboeuf, France) is an aqueous, nontoxic, tissue storage reagent that rapidly permeates most tissues to stabilize and protect RNA in fresh specimens. RNA *later* eliminates the need to immediately process or freeze samples; the specimen can simply be submerged in RNA *later* and stored for analysis at a later date. Samples in RNA *later* can be stored for extended periods under conditions where RNA degradation would normally take place rapidly. Tissues can be stored indefinitely in RNA *later* at –20°C or below.

#### 2.1.10.1. Culture preparation

The bacteria were inoculated in Landy MOPS at 37 ºC and under 160 rpm agitation rate.

#### 2.1.10.2. Transcriptome blockage

An equivalent volume of 2 x $10^9$ cells was obtained at each point of the kinetics previously defined, these volumes were centrifuged at 11,000g, -9 ºC for 5 min. The supernatant was discarded, while the pellet was stored in 1 mL RNA *later* at -20 ºC.

**2.1.10.3. RNA extraction**

The cells are thawed on ice and then resuspended. There were centrifuged for 1 min at -9 ºC, 12000 g. The supernatant was discarded and 50 µL of T10 E1 buffer was added to the pellet with 10 mg/mL lysozyme and then was incubated for 10 min at room temperature. The equivalent of 250 µL of zirconium beads which were provided in the kit was added to the tubes. Cells resuspended in 350 µL of RNA wiz (provided in the kit) were transferred into tubes with beads. The RNA wiz contains phenol, which allows fragilization of the cells and inhibits the RNases. The whole was homogenized 10 min by vortex to lyse the cells. The tubes were then centrifuged for 5 min at 4 ºC and 12000 g. A volume of 200 µL was obtained and 200 µL of chloroform were added, then the mix was agitated during 30 s and incubated 10 min at room temperature. The tubes were centrifuged for 5 min at 4 ºC and 12000 g. The aqueous phase was obtained, 0.5 vol of 100% ethanol was added and the whole was stirred vigorously. The sample was loaded on the column provided by the kit to purify the RNA. After that, the treatment by DNase I was started and the RNA dosage measured by Spectrophotometer WPA LIGHTWAVE II. 2.5 volume of ethanol 95 % was added to the RNA, which was stored at -80 ºC.

**2.1.10.4. The reverse transcription**

The reverse transcription which reverse transcribed the RNA into cDNA was performed with RevertAid First Strand cDNA Synthesis Kit (K1621, Fermentas).

**2.1.10.5. Amplification of gene under study**

The primers utilized are *ppsA* sense and antisense which correspond to a fragment of the *pps*A gene of plipastatin (Table 8).

These primers were amplified using AmpliTaq® Gold DNA polymerase from Roche.

**2.1.10.6. Expression quantification**

15µL of amplified fragment was mixed with 3 µL of 6x loading buffer provided with DNA O'GeneRuler™ ladder from Fermentas and was migrated in 1.2 % agarose gel colored with GelRed. The gel was photographed under UV in GelDoc.

### 2.1.11. Mutants and overproduction of fengycin and plipastatin operons

pBG106 plasmid was used to achieve the fengycin overproduction by the replacement of the native promoter of *B. subtilis* BBG134 strain by the constitutive promoter $P_{repU}$-neo. This promoter from *Staphylococcus aureus* is already inserted in pBG106 (Leclère *et al*., 2005).

Construction 1; pBG106 plasmid was used, *dacC* Cassette1 (fragment size 617 bp) and *ppsA* Cassette 2 (fragment size 621 bp) fragments were generated by PCR using PCR Master Mix (2 X) from Fermentas. The primers were designed using the sequence of the plipastatin operon of *B. subtilis* 168 $trpC_2$ (PubMed nucleotide accession no. AL009126, *B. subtilis* 168 complete genome sequence).

Construction 2; was performed by the replacement of the pBG106 by pGEM-T Easy vector to avoid the problems faced in construction 1.

Construction 3; was performed by the modification of construction 2 by the replacement of $P_{repU}$-*neo* gene and *ppsA* Cassette 2 by P*fen*-BBG21-*spc* and *ppsA* Cassette 2N, respectively.

#### 2.1.11.1. Plasmid construction

##### 2.1.11.1.1. Construction 1

Firstly, pBG106 plasmid was used to achieve the plipastatin overproduction by the replacement of the plipastatin promoter of the *B. subtilis* BBG134 and BBG111 by the constitutive promoter $p_{repU}$ which is already inserted in pBG106 plasmid and originated from *Staphylococcus aureus* plasmid pUB110 (Leclère *et al*., 2005).

The *dacC-pps* and *ppsA* Cassette 2 fragments from *B. subtilis* 168 strain were first cloned in pGEM-T Easy vector between the *Eco*RI sites of the vector, and *E. coli* JM109 was separately transformed. The resulting two plasmids were named pBG186 and pBG177, respectively. pBG106 and pBG186 were *Sph*I and *Sa*II double digested; then the ε*pbp* fragment was released from pBG106 and *pps-dacC* fragment was inserted between the *Sph*I and *Sa*II sites of pBG106 to obtain pBG189.

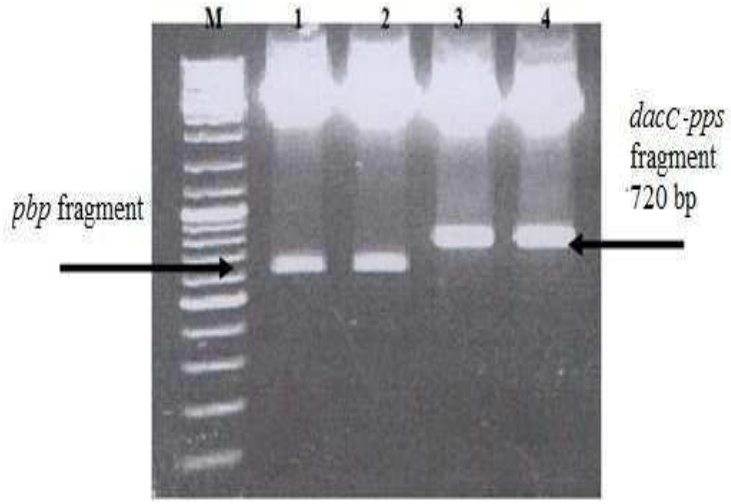



Figure 21: *Sph*I + *Sa1*I double digestion of pBG106 and pBG186 to release *pbp* fragment and replace it by the *dacC-pps* insert; M = O' Gene Ruler standard, 1, 2 = digested pBG106; 3, 4 = digested pBG186.

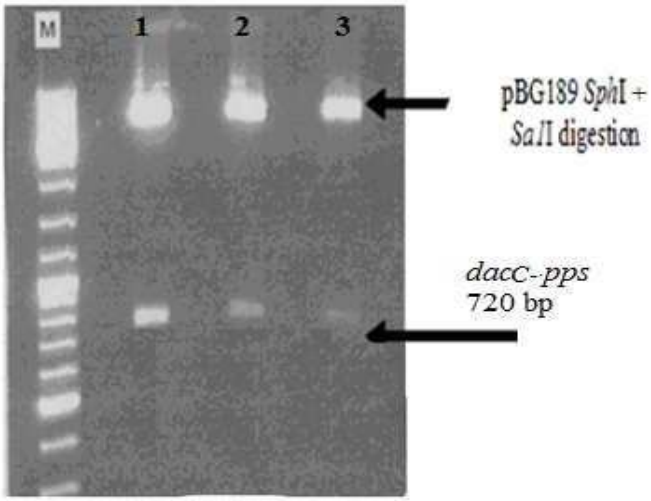



Figure 22: pBG189 *Sph*I + *Sa1*I double digestion to verify *dac*C-*pps* fragment insertion; M = O' Gene Ruler standard

pBG177 was easily digested *by Xba*I and *Ban*II, while it was difficult to digest pBG189 which bears two restriction sites for both *Xba*I and *Ban*II. A partial digestion for 70 min was effected, the reaction was stopped after 10, 20, 30, 40, 50, 60 and 70 min but many fragments were obtained and it was difficult to separate the required fragment (ε*fenF*) to replace it in the plasmid by *ppA* cassette 2 fragment. Another partial digestion was done (2, 4, 6, 8, 10, 20, 30, 40, 50, 60 and 70 min) with the same difficulty to separate the plasmid. So, it was difficult to complete with the plasmid.

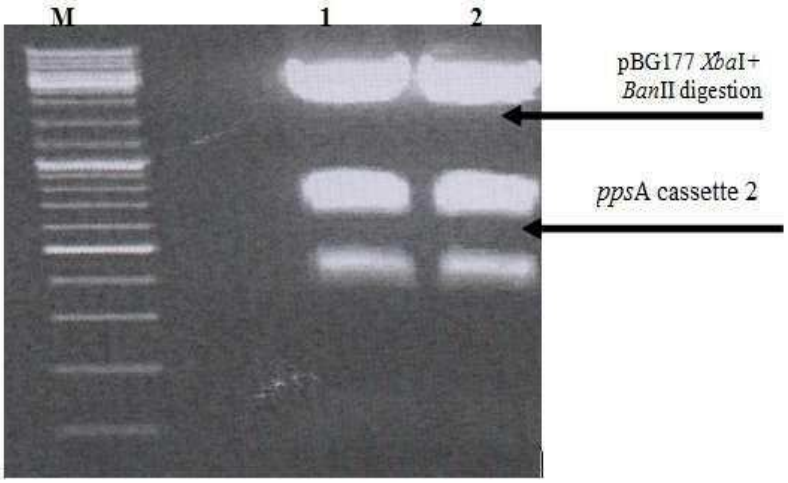



Figure 23: pBG177 *Xba*I + *Ban*II double digestion to insert *ppsA* Cassette 2 fragment into pBG189; M = O' Gene Ruler standard.

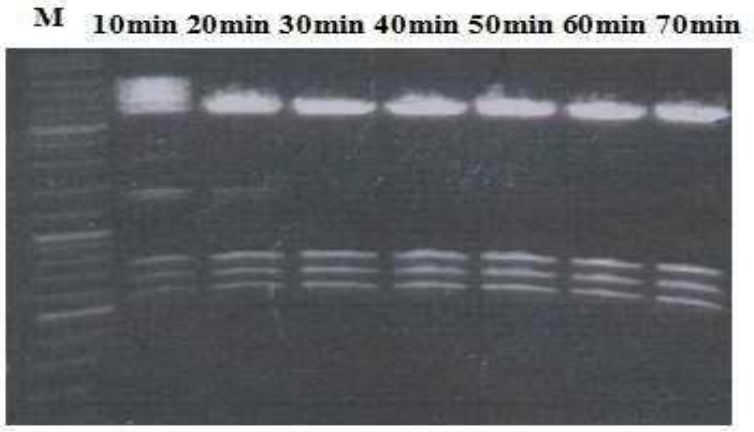



Figure 24: Partial double digestion of pBG189 by *Xba*I + *Ban*II up to 70 min, the reaction being stopped every 10 min.

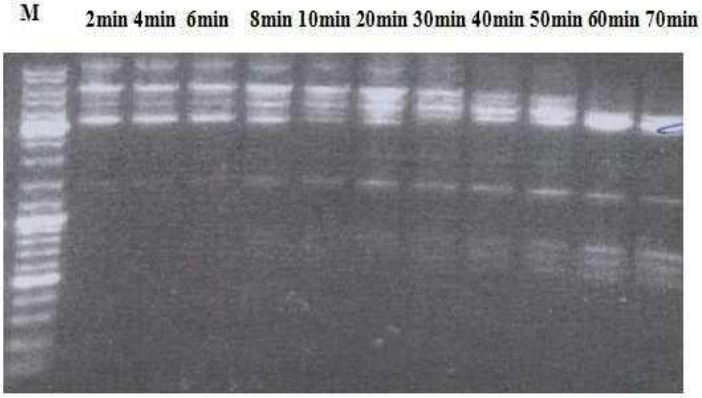



Figure 25: Partial double digestion of pBG189 by *Xba*I + *Ban*II for 70 min, the reaction being stopped every 2 min for the first five samples and every 10 min for the next six samples

For this reason, another system using the pGEM-T Easy vector was followed but, the *dacC-pps* fragment restriction sites wasn't compatible with the pGEM-T Easy vector. So, other primers were designed for the *dacC* Cassette 1 with two restriction sites for *Aat*II and *Nco*I.

#### 2.1.11.1.2. Construction 2

The construction was performed to solve the problems faced in construction 1 by replacing the pBG106 plasmid by the pGEM-T Easy vector. The complete construction plasmid was obtained with two types; (1) pBG180 (with $p_{repU}$-*neo* gene in sense) with *pps* operon and (2) pBG180* which bears an inverted $p_{repU}$-*neo* cassette. The two plasmids were used to transform *B. subtilis* BBG111 and BBG134, in order to verify the ability of the plasmid transformation and the success of the homologous recombination.

#### 2.1.11.1.3. Construction 3

The constructed plasmid pBG180 was used to design the construction 3 by the replacement of $P_{repU}$-*neo* cassette and *ppsA* Cassette 2 by $P_{fen}$-BBG21-spc and *ppsA* Cassette 2N, respectively. After the insertion of *ppsA*-Cassette 2N fragment in pBG180, pBG191 was obtained, and then pBG192 was further obtained by the insertion of *spc* gene into pBG191 plasmid. The insertion of $P_{fen}$-BBG21 into pBG192 revealed the pBG193 plasmid, which displays a weak growth on LB + spc. So, it was difficult to sequence and transform into *B. subtilis* BBG111 and BBG134.

### 2.1.12. The plipastatin or fengycin promoter expression in *B. subtilis* 168 derivatives

Plipastatin operon regulation in *B. subtilis* 168 derivatives was studied by analyzing the strength of the native P*pps* promoter using a gene reporter system, an extensive toolbox for the study of promoter regulation.

pDG1661 (BGSC accession: ECE112, GenBank U46196) plasmid was used to integrate *lacZ* fusion into *B. subtilis* 168 ($trpC_2$) $sfp^o$. Fragments were generated by PCR using PCR Master Mix (2 X) from Fermentas. Plipastatin promoter primers were designed using the sequence of the plipastatin operon of *B. subtilis* 168 (PubMed nucleotide accession no. AL009126 *B. subtilis* 168 complete genome sequence). For forward primer promoter; 5'-GC**A**TGCCGGATATGCCTTAG-3' (the underlined *Eco*RI artificial site was generated by the substitution of the one base in boldface type) and reverse primer 5'-TGTCTGTAATAACGCTTTGTCGA**C** -3' (the underlined *Hind*III artificial site was generated by the substitution of the one base in boldface type).

#### 2.1.12.1. BBG142 a new *B. subtilis* 168 derivative obtained by *lacZ* fusion to $P_{pps}$ 168

BBG142 strain was obtained by the transformation of pBG184 plasmid into *B. subtilis* 168 strain. pBG184 was obtained by the insertion of the $P_{pps}$168 fragment between *Eco*RI and *Hind* III sites of pDG1661 plasmid.

The new plasmid was transformed into *B. subtilis* 168 by the natural competence method. Two colonies were obtained on the LB medium supplemented by chloramphenicol, which were re-inoculated on another LB with Cm Petri dishes. DNA was then extracted as previously explained. The verification of the new strain BBG142 was done by PCR.

#### 2.1.12.2. BBG139 a new *B. subtilis* 168 derivative obtained by *lacZ* fusion to the $P_{fen}$ BBG21

BBG139 strain was obtained by the transformation of pBG185 plasmid into *B. subtilis* 168 strain. pBG185 was obtained by the insertion of $P_{fen}$ BBG21 between *Eco*RI and *Hind*III sites of pDG1661.

The new plasmid was transformed into *B. subtilis* 168 by the natural competence method. Four colonies were obtained on the LB medium supplemented with chloramphenicol, which were re-inoculated on other LB + Cm Petri dishes. DNA was then extracted as previously explained. The verification of the new strain BBG139 was done by PCR.

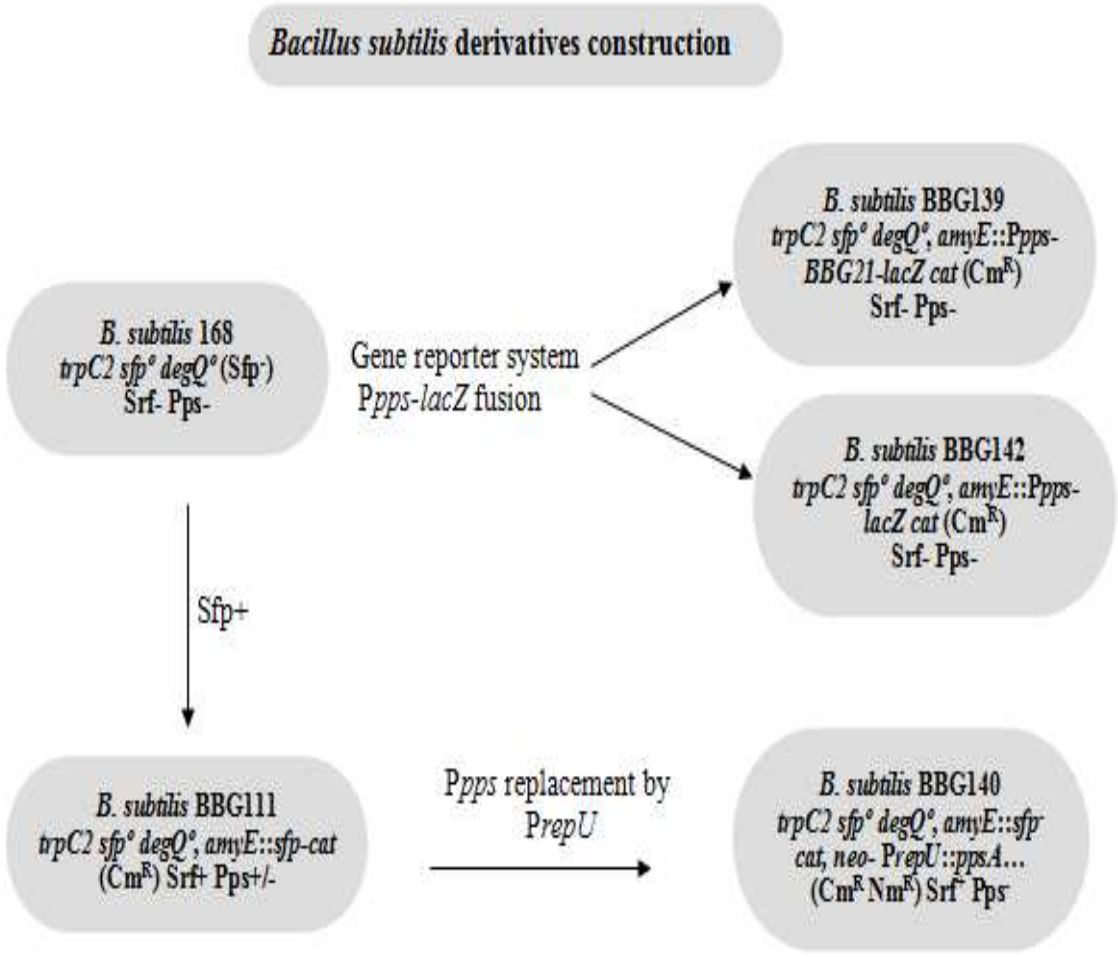



Figure 26: The different characteristics of constructed strains from *B. subtilis* 168

## 2.2. Biological tests

### 2.2.1. β-Galactosidase activity test

1 mL of culture medium was incubated for 20-60 min on ice to stop growth. After centrifuging 2 min at 12000 g, the pellet was suspended in 0.5 mL of Z buffer, and then was lysed with 30 µL of chloroform under strong agitation for 15 sec. Then both the tube and the ONGP fresh solution were incubated at 37ºC for 5 min.

200µL of the substrate (ONGP) was added to each tube and the time is measured, 0.5 mL 1 M $Na_2CO3$ was added to stop the reaction after appearance of yellow colour. Samples were Vortex and, the reaction time was time noted precisely.

1 mL of each sample was transferred to an Eppendorf tube and centrifuged 5 min at 12000 g to remove debris and chloroform. The optical density at 420 nm was recorded. The activity was calculated according to the following equation:

**Miller Units = 1000 x ($OD_{420}$) / (T x V x $OD_{600}$)**

$OD_{420}$ reflects subtilistrate hydrolysis

$OD_{600}$ reflects cell density

T = time of the reaction in minutes

V = volume of culture used in the assay in mL

### 2.2.2. Amylase test

This test was used for the verification of the insertion of a gene at *amy*E gene locus by the loss of amylase activity. Strains are inoculated on plates with LB complemented with 1% starch and incubated at 37ºC for 48 h. The plate are then covered with a potassium iodine solution. A halo of clarification around the colonies appeared if the amylase is active.

### 2.2.3. Antifungal test

5 µL of culture ($O.D_{600nm}$ = 0.2) were spotted onto the centre of PDA medium plates. Small squares of *Fusarium oxysporum*, *Botrytis cinerea* were cut-out and deposited on the plates, which were incubated at 25 ºC for 4 days before measuring the inhibition zone.

### 2.3. Quantitative analysis of lipopeptides

Both surfactin and fengycin were quantified using High Performance Liquid Chromatography (HPLC) after their extraction from 1 mL supernatant of culture by $C_{18}$ cartridges (Extract-clean SPE 500 mg, Grace Davison-Alltech, Deerfield, IL, U.S.A). The lipopeptides adsorbed on the $C_{18}$ were then eluted in 8 mL pure methanol. The eluted lipopeptides were dried by rotating vapour and speed vacuum followed by re-dissolving in 200 µL of pure methanol. The surfactin concentration was determined by reverse phase C18 HPLC (600 s, Waters, Milford, MA, U.S.A) equipped with a Merck C18 column (5 mm; Darmstadt, Germany) as previously described (Oka *et al*., 1993; Liu *et al*., 2007; Chollet-Imbert *et al*., 2009). The standard of surfactin was purchased from Sigma (St. Louis, MO, U.S.A). The fengycin was eluted during 40 min under an ACN/$H_2O$/TFA gradient from 45/55/0.1 to 55/45/0.1 at 0.6 mL/min. The standard of fengycins was kindly provided by Dr M. Deleu from the Agricultural University of Gembloux (Belgium) and gave peaks between 10 and 25 min. Peaks were selected through calibration and spectra were analyzed using values of second derivative which give two major peaks at 213 and 236 nm associated with a minor peak at 290 nm (Eeman *et al*., 2005; Chollet-Imbert *et al*., 2009; Gancel *et al*., 2009).

### 2.4. Bioinformatics analyses strategy

As the NCBI is the highly site in scientists citation because it contains the GenBank data base available, it was used to search the different *Bacillus* sequenced genomes.

Firstly, it was searched for genomes, and then for prokaryotes. In prokaryotes, 36 *Bacillus*, one *Brevibacillus brevis* and two *Paenibacillus polymexa* genomes were found in 06 mars 2011. One table was designed to all these strains with their name, genome size and sequence reference. For each genome and on NCBI site, the following keywords were used to search the different NRPS genes: ribosomal, synthetase, synthase, NRPS, antibiotic, antifungal, siderophore, lipopeptide and adenylation domain. The different proteins annotated with one of these keywords were added in one table for each strain. Each protein sequence for each gene was analyzed by NRPS-PKS analysis website (http://nrps.igs.umaryland.edu/nrps/) to detect the organisation of the different modules, domains and the specificity of the adenylation domains. Finally from these predictions, a potential primary structure of these peptides was obtained and then compared to all the known peptides annotated in the NORINE database (http://bioinfo.lifl.fr/norine/) using the Editor tool and pattern comparison.

# *Chapter V. Results and Discussion*

## 1. Bioinformatics analyses

### Introduction

The first objective of this work was to get an overview of the Non Ribosomal Peptides potentially produced by strains of the *Bacillus* group. The genome of 39 different *Bacillus*, *Paenibacillus* and *Brevibacillus* strains available in the NCBI database on the 06 mars 2011 were analysed using the bioinformatics tools described in the "Materials and Methods" section. NRPS genes/operons were detected in the sequenced genomes by keywords searching applied to the protein annotation (ribosomal, synthetase, synthase, antibiotic, antifungal, siderophore, lipopeptide, NRPS and adenylation domain). Then, some steps were followed; identification of NRPS genes, annotation, description of modules and domains, specificity prediction of A domains by PKS-NRPS website (http://nrps.igs.umaryland.edu/nrps/), prediction of the potential primary structure of the peptide and finally, comparison of peptide structure with those introduced in the NORINE database. From this site, we will know if the peptide is an existing molecule or a new one.

#### 1.1. Bioinformatics analyses on NRPS genes of *Bacillus*

This chapter will describe the different NRPS genes detected in sequenced genomes from 36 different *Bacillus,* 2 *Paenibacillus* and 1 *Brevibacillus* strains analyzed by PKS-NRPS analysis, NRPS predictor and NCBI web site. Table 11 shows the 39 different available genome sequences from NCBI web site with their genome size, genome accession (GenBank) and reference sequence (RefSeq). The largest genome size was detected for *Brevibacillus brevis* (6.3 Mb), followed by *Paenibacillus polymyxa* SC2 (6.21 Mb), *B. weihenstephanensis* KBAB4 (5.875 Mb), the *Bacillus cereus* group (5.43-5.84 Mb), *Paenibacillus polymyxa* E681 (5.4 Mb), the *Bacillus thuringiensis* group (5.3-5.6 Mb) and the *Bacillus anthracis* group (5.22-5.50 Mb). The *Bacillus tusciae* strain has the smallest genome size (3.4 Mb) and all the other strains genome sizes ranged from 3.6 Mb (*B. selenitireducens* MLS10) to 5.1 Mb (*B. megaterium* DSM 319). From the sequenced *Bacillus* strains in Table (10); 10 genomes of *B. cereus* group were found to be sequenced, followed by *B. anthracis* (5 genomes), *B. subtilis* and *B. thuringiensis* (3 genomes) and *B. megaterium* and *B. amyloliquefaciens* (2 genomes).

Table 11: Genomes sequenced for different *Bacillus, Brevibacillus* and *Paenibacillus* strains

| No | Organism | Genome Size | GenBank | RefSeq |
|---|---|---|---|---|
| 1 | *Bacillus amyloliquefaciens* DSM7 | 4.0 | FN597644.1 | NC_014551.1 |
| 2 | *Bacillus amyloliquefaciens* FZB42 | 3.9 | CP000560.1 | NC_009725.1 |
| 3 | *Bacillus anthracis* str. 'Ames Ancestor' | 5.50242 | AE017334.2 | NC_007530.2 |
| 4 | *Bacillus anthracis* str. A0248 | 5.50242 | CP001598.1 | NC_012kk659.1 |
| 5 | *Bacillus anthracis* str. Ames | 5.22729 | AE016879.1 | NC_003997.3 |
| 6 | *Bacillus anthracis* str. CDC 684 | 5.50511 | CP001215.1 | NC_012581.1 |
| 7 | *Bacillus anthracis* str. Sterne | 5.22866 | AE017225.1 | NC_005945.1 |
| 8 | *Bacillus atrophaeus* 1942 | 4.2 | CP002207.1 | NC_014639.1 |
| 9 | *Bacillus cellulosilyticus* DSM 2522 | 4.7 | CP002394.1 | NC_014829.1 |
| 10 | *Bacillus cereus* 03BB102 | 5.44963 | CP001407.1 | NC_012472.1 |
| 11 | *Bacillus cereus* AH187 | 5.59913 | CP001177.1 | NC_011658.1 |
| 12 | *Bacillus cereus* AH820 | 5.5841 | CP001283.1 | NC_011773.1 |
| 13 | *Bacillus cereus* ATCC 10987 | 5.43265 | AE017194.1 | NC_003909.8 |
| 14 | *Bacillus cereus* ATCC 14579 | 5.415 | AE016877.1 | NC_004722.1 |
| 15 | *Bacillus cereus* B4264 | 5.4 | CP001176.1 | NC_011725.1 |
| 16 | *Bacillus cereus* E33L | 5.8465 | CP000001.1 | NC_006274.1 |
| 17 | *Bacillus cereus* G9842 | 5.75 | CP001186.1 | NC_011772.1 |
| 18 | *Bacillus cereus* Q1 | 5.493 | CP000227.1 | NC_011969.1 |
| 19 | *Bacillus cereus* biovar *anthracis* str. CI | 5.488 | CP001746.1 | NC_014335.1 |
| 20 | *Bacillus clausii* KSM-K16 | 4.3 | AP006627.1 | NC_006582.1 |
| 21 | *Bacillus cytotoxicus* NVH 391-98 | 4.1071 | CP000764.1 | NC_009674.1 |
| 22 | *Bacillus halodurans* C-125 | 4.2 | BA000004.3 | NC_002570.2 |
| 23 | *Bacillus licheniformis* ATCC 14580 | 4.2226 | CP000002.3 | NC_006270.3 |
| 24 | *Bacillus megaterium* DSM 319 | 5.1 | CP001982.1 | NC_014103.1 |
| 25 | *Bacillus megaterium* QM B1551 | 5.5225 | CP001983.1 | NC_014019.1 |
| 26 | *Bacillus pseudofirmus* OF4 | 4.3 | CP001878.2 | NC_013791.2 |
| 27 | *Bacillus pumilus* SAFR-032 | 3.7 | CP000813.1 | NC_009848.1 |
| 28 | *Bacillus selenitireducens* MLS10 | 3.6 | CP001791.1 | NC_014219.1 |
| 29 | *Bacillus subtilis* subsp. subtilisp. *natto* BEST195 | 4.1058 | AP011541 | |
| 30 | *Bacillus subtilis* subsp. subtilisp. *spizizenii* str. W23 | 4.0 | CP002183.1 | NC_014479.1 |
| 31 | *Bacillus subtilis* subsp. subtilisp. *subtilis* str. 168 | 4.2 | AL009126.3 | NC_000964.3 |
| 32 | *Bacillus thuringiensis* BMB171 | 5.61 | CP001903.1 | NC_014171.1 |
| 33 | *Bacillus thuringiensis* serovar *konkukian* str. 97-27 | 5.31468 | AE017355.1 | NC_005957.1 |
| 34 | *Bacillus thuringiensis* str. Al Hakam | 5.31309 | CP000485.1 | NC_008600.1 |
| 35 | *Bacillus tusciae* DSM 2912 | *3.4 | CP002017.1 | NC_014098.1 |
| 36 | *Bacillus weihenstephanensis* KBAB4 | 5.87577 | CP000903.1 | NC_010184.1 |
| 37 | *Brevibacillus brevis* | 6.3 | AP008955.1 | NC_012491.1 |
| 38 | *Paenibacillus polymyxa* E681 | 5.4 | CP000154.1 | NC_014483.1 |
| 39 | *Paenibacillus polymyxa* SC2 | *6.21 | CP002213.1 | NC_014622.1 |

The detailed results obtained with the first strain of the list, *Bacillus amyloliquefaciens* DSM7 were given as an example in Table 12 (list of the different NRPS genes found) and Figures 27 to 29 (detailed organization and prediction of adenylation domain specificity). The other detailed results are collected in annexes. The biosynthetic genes of four different molecules were found in DSM7 strain: three lipopeptides, truncated fengycin (Figure 27), surfactin (Figure 28) and iturin and a siderophore, bacillibactin (Figure 29). Also, one potentially new unpredictable NRPS-PKS hybrid molecule was detected.

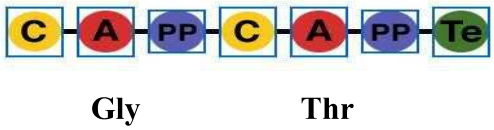
Table 12: Detected NRPS molecules in *Bacillus amyloliquefaciens* DSM7

| Product Name | Start | End | Strand | Length (a.a) | Accession | GeneID | Locus |
|---|---|---|---|---|---|---|---|
| non-ribosomal surfactin synthetase SrfAA | 333124 | 343878 | + | 3584 | YP_003918908.1 | 9779039 | *srfAA* |
| nonribosomal surfactin synthetase SrfAB | 343900 | 354660 | + | 3586 | YP_003918909.1 | 9780206 | *srfAB* |
| nonribosomal surfactin synthetase C | 354695 | 358534 | + | 1279 | YP_003918910.1 | 9780207 | *srfAC* |
| nonribosomal surfactin synthetase D | 358553 | 359284 | + | 243 | YP_003918911.1 | 9780208 | *srfAD* |
| bacillaene synthesis; hybrid NRPS/PKS protein | 1793732 | 1808683 | + | 4983 | YP_003920382.1 | 9778399 | *baeJ* |
| iturin A synthetase C (*B. subtilis*) | 1968515 | 1976365 | - | 2616 | YP_003920507.1 | 9780434 | *ituC* |
| iturin A synthetase B (*B. subtilis*) | 1976456 | 1992541 | - | 5361 | YP_003920508.1 | 9779079 | *ituB* |
| iturin A synthetase A (*B. subtilis*) | 1992590 | 2004538 | - | 3982 | YP_003920509.1 | 9779078 | *ituA* |
| malonyl-CoA transacylase ItuD | 2004558 | 2005799 | - | 413 | YP_003920510.1 | 9779077 | *ituD* |
| nonribosomal fengycin synthesis; COG1020; fengycin synthetase E | 2027434 | 2031240 | - | 1268 | YP_003920534.1 | 9781135 | *fenE* |
| nonribosomal fengycin synthesis; COG1020; fengycin synthetase D | 2031260 | 2038936 | - | 2558 | YP_003920535.1 | 9778781 | *fenD* |
| non-ribosomal peptide synthetase | 2526874 | 2532402 | - | 1842 | YP_003921077.1 | 9779368 | *nrpsE* |
| siderophore 2,3-dihydroxybenzoate-glycine-threonine trimeric ester bacillibactin synthetase | 3053649 | 3060776 | - | 2375 | YP_003921588.1 | 9779667 | *dhbF* |


Figure 27: Predicted modular organization and adenylation domain specificity of truncated fengycin *fenD*, *fenE* synthetases genes from *B. amyloliquefaciens* DSM7


Figure 28: Predicted modular organization and adenylation domain specificity of surfactin synthetases genes (*srfAA*, *srfAB* and *srfAC*) from *B. amyloliquefaciens* DSM7. The gene *srfaAD* is not shown but corresponds to a second thioesterase as found in the surfactin operon in *B. subtilis* 168.

A.




B.


C.


Figure 29: Predicted modular organization and adenylation domain specificity of bacillibactin synthetases gene (*dhbF*) (A), iturin (B) and *nrpsE* (C) from *B. amyloliquefaciens* DSM7

The other results are gathered within five categories:

- List of the strains which do not produce any NRPs
- Lipopeptide producers
- Siderophore producers
- Antibiotic producers
- New Non-Ribosomal peptides producers.

### 1.1.1. Detection of *Bacillus* strains with no NRPS genes

Following the strategy described in this study we couldn't have detected any NRPS genes from eight *Bacillus* strains as shown in Table 13. They belong to the species *Bacillus megaterium* or are isolated from selective environments (acidic for *B. tuscae* or alkaline for the other ones).

Table 13: *Bacillus* strains with no NRPS genes

| ***Bacillus* spp.** |
|---|
| *Bacillus cellulosilyticus* DSM 2522 |
| *Bacillus selenitireducens* MLS10 |
| *Bacillus clausii* KSM-K16 |
| *Bacillus halodurans* C-125 |
| *Bacillus megaterium* DSM 319 |
| *Bacillus megaterium* QM B1551 |
| *Bacillus pseudofirmus* OF4 |
| *Bacillus tusciae* DSM 2912 |

### 1.1.2. Detection of known lipopeptides

14 strains show the presence of genes/operons involved in the biosynthesis of members of the four families of lipopeptides [surfactin, pumilacidin, lichenysin from surfactin family, fengycin or plipastatin from fengycin/plipastatin family, bacillomycin and mycosubtilin from iturin family and kurstakin from kurstakin family], or other new lipodepsipeptide: fusaricidin.

Table 14: Different common lipopeptides detected by bioinformatics analysis from 14 *Bacillus* strains

| ***Bacillus* spp.** | **Lipopeptide** |
|---|---|
| *Bacillus amyloliquefaciens* DSM7 | Surfactin, iturin, truncated fengycin |
| *Bacillus amyloliquefaciens* FZB42 | Surfactin, fengycin, bacillomycin |
| *Bacillus atrophaeus* 1942 | Mycosubtilin, plipastatin, surfactin |
| *Bacillus licheniformis* ATCC 14580 | Lichenysin |
| *Bacillus pumilus* SAFR-032 | Variant from surfactin family |
| *Bacillus subtilis* subsp. *natto* BEST195 | Surfactin, truncated plipastatin |
| *Bacillus subtilis* subsp. *spizizenii* str. W23 | Surfactin, mycosubtilin |
| *Bacillus subtilis* subsp. *subtilis* str. 168 | Surfactin, plipastatin |
| *Bacillus thuringiensis* BMB171 | Kurstakin |
| *Bacillus weihenstephanensis* KBAB4 | Variant from Kurstakin family |
| *Bacillus cereus* G9842 | Kurstakin |
| *Bacillus cereus* B4264 | Kurstakin |
| *Bacillus cereus* ATCC 14579 | Truncated Kurstakin |
| *Paenibacillus polymyxa SC2* | Fusaricidin |

The detection of NRPS genes/operons in both *B. amyloliquefaciens* DSM7, *B. amyloliquefaciens* FZB42 and *B. atrophaeus* 1942 strains revealed as shown in Table 14 the potential for the biosynthesis of three different lipopeptides (surfactin, truncated fengycin, and iturin), (surfactin, fengycin and bacillomycin) and (mycosubtilin, plipastatin and surfactin) respectively. Three *Bacillus* strains; *B. subtilis* subsp. *natto* BEST195, *B. subtilis* subsp. *spizizenii* str. W23 and *B. subtilis* subsp. *subtilis* str. 168 have the genes/operons to synthesize two different lipopeptides (Surfactin, truncated plipastatin), (Surfactin, mycosubtilin) and (Surfactin, plipastatin), respectively. Also, *B. thuringiensis* BMB171, *B. weihenstephanensis* KBAB4, *Bacillus cereus* G9842, *Bacillus cereus* ATCC 14579, and *Bacillus cereus* B4264 have the operon involved in the lipopeptide kurstakin biosynthesis. Lichenysin, pumilacidin and fusaricidin lipopeptide biosynthetic genes were respectively detected in *B. licheniformis* ATCC 14580, *B. pumilus* SAFR-032 and *Paenibacillus polymyxa* SC2 genomes.

Table 15: Peptide sequence of the different variants of each lipopeptide family from 14 *Bacillus* spp.

| Lipopeptide | Predicted peptide structure | *Bacillus* spp. |
|---|---|---|
| **Surfactin family** | | |
| Surfactin | L-Glu/ Asp-L-Leu/Ile/Val-D-Leu-L-Val-L-Asp-D-Leu-L-Leu | *amyloliquefaciens* DSM7 |
| | L-Glu/ Asp-L-Leu/Ile/Val-D-Leu-L-Val-L-Asp-D-Leu-L-Leu/Ile/Val | *amyloliquefaciens* FZB42 |
| | L-Glu/Asp-L-Leu/Ile/Val-D-Leu-L-Val-L-Asp-D-Leu-L-Leu/Ile/Val | *subtilis* subsp. *subtilis* str. 168 |
| | L-Glu/ Asp-L-Leu/Ile/Val-D-Leu-L-Val-L-Asp-D-Leu-L-Leu | *subtilis* subsp. *natto* BEST195 |
| | L-Glu/Asp-L-Leu/Ile/Val-D-Leu-L-Val-L-Asp-D-Leu-L-Leu/Ile/Val | *subtilis* subsp. *spizizenii* str. W23 |
| New variant | L-Leu/Ile/Val-L-Leu-D-Glu/Asp-L-Asp-L-Leu-D-Val-L-Leu | *atrophaeus* 1942 |
| New variant | L-Glu/Asp-L-Leu-D-Leu-L-Phe/Trp-L-Asp-D-Leu-L-Leu | *pumilus* SAFR-032 |
| Lichenysin | L-Gln-L-Leu-D-Leu-L-Val-L-Asp-D-Leu-L-Leu | *licheniformis* ATCC 14580 |
| **Iturin family** | | |
| Iturin | L-Asx-D-Pro-D-Tyr/Trp-L-Asx-L-Gln-D-Asx-L-Ser | *amyloliquefaciens* DSM7 |
| Bacillomycin | L-Asx-D-Glu/Asp-D-Tyr/Trp-L-Asx-L-Pro-D-Ser-L-Thr | *amyloliquefaciens* FZB42 |
| Mycosubtilin | L-Asx-D-Pro-D-Tyr/Trp-L-Asx-L-Gln-D-Ser-L-Asx | *subtilis* subsp. *spizizenii* str. W23 |
| | L-Asx-D-Pro-D-Tyr/Trp-L-Asx-L-Gln-D-Ser-L-Asx | *atrophaeus* 1942 |
| **Fengycin family** | | |
| Truncated fengycin | L-X-D-Tyr-L-Ile | *amyloliquefaciens* DSM7 |
| Fengycin | L-Glu-D-Orn-L-X-D-Thr-L-Glu-D-Val-L-Pro-L-Glu-D-Tyr-L-Ile | *amyloliquefaciens* FZB42 |
| Plipastatin | L-Glu-D-X-L-Tyr-D-Thr-L-Glu-D-Val-L-Pro-L-Glu-D-Tyr-Ile | *subtilis* subsp. *subtilis* str. 168 |
| New variant | No Hit- D-Glu-L-Thr- D-Tyr- L-Val-D-Glu-L-Glu-L-Tyr-D-Pro-L-Ile | *atrophaeus* 1942 |
| **Kurstakin family** | | |
| Kurstakin | D-Thr-L-Gln-L-Gly-L-Ser-L-X-D-Gln-L-Gln | *thuringiensis* BMB171 |
| | D-Thr-L-Gln-L-Gly-L-Ser-L-X-D-Gln-L-Gln | *cereus* G9842 |
| Truncated kurstakin | D-Thr-L-Gln-L-Gly-L-No Hit-D-Gln-L-Gln | *cereus* ATCC 14579 |
| | D-Thr-L-Gln-L-Gly-L-Ser-L-X-D-Gln-L-Gln | *cereus* B4264 |
| New variant | D-Thr-L-Gln-L-Gly-L-Ser-L-X-D-Thr-L-Glu/Asp | *weihenstephanensis* KBAB4 |
| **Other** | | |
| Fusaricidin | L-Thr-D-Val/Ala-Tyr-D-Thr-D-Asx-L-X | *polymyxa* SC2 |

Table 15 shows the diversity of lipopeptides potentially produced by the different species of *Bacillus*. These lipopeptides are usually a mixture of compounds with different lengths and types of fatty acid (FA), β-hydroxy FA (FA-β-OH) for surfactins and fengycins, β-amino FA (FA-β-$NH_2$) for iturins or guanidylated-β-OH FA (*g*FA-β-OH) for fusaricidin. Surfactin and lichenysin are structurally related lipopeptides. It was reported that lichenysin is at least a twice more efficient biosurfactant than surfactin, probably due to the replacement of Glu1 by Gln1 [Grangemard *et al*., 1999; Grangemard *et al*., 2001]. Three interesting potential variants were detected:

- A new form of surfactin with variation in position 4 (replacement of Val by Phe/Trp) produced *B. pumilus*
- New forms of surfactin and plipastatin produced by *B. atrophaeus* with modified sequences

-*B. weihenstephanensis* KBAB4 showed kurstakin form that could differ than that of the other four kurstakin producing *Bacillus* strains; the structure has variation at position 6 and 7 (Thr, Glu/ Asp in spite of Gln, Gln).

### 1.1.3. Detection of siderophores

Genes responsible for the biosynthesis of 2,3-dihydroxybenzoate-glycine-threonine, the siderophore called bacillibactin, were detected in 27 different *Bacillus* strains (Tables 16 and 17), ten of them harboring no other NRPS genes (Table 16), while Table 17 shows 17 *Bacillus* strains which harbor bacillibactin and other NRPS genes.

Table 16: *Bacillus* strains which harbour only bacillibactin NRPS genes

| ***Bacillus* spp.** |
|---|
| *Bacillus thuringiensis* serovar *konkukian* str. 97-27 |
| *Bacillus thuringiensis* str. Al Hakam |
| *Bacillus anthracis* str. 'Ames Ancestor' |
| *Bacillus anthracis* str. A0248 |
| *Bacillus anthracis* str. Ames |
| *Bacillus anthracis* str. Sterne |
| *Bacillus cereus* 03BB102 |
| *Bacillus cereus* AH187 |
| *Bacillus cereus* ATCC 10987 |
| *Bacillus cereus* biovar *anthracis* str. CI |

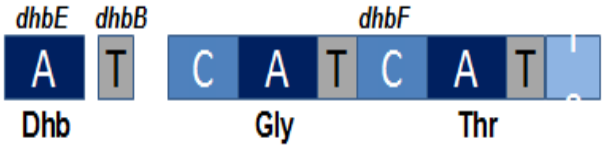



Figure 30: Bacillibactin NRPS synthetases operon

Table 17: Different *Bacillus* strains which harbor bacillibactin siderophore genes and other NRPS genes/operons.

| ***Bacillus* strains** | |
|---|---|
| *Bacillus licheniformis* ATCC 14580 | *Bacillus anthracis* str. CDC 684 |
| *Bacillus subtilis* subsp. *natto* BEST195 | *Bacillus atrophaeus* 1942 |
| *Bacillus subtilis* subsp. *spizizenii* str. W23 | *Bacillus cereus* AH820 |
| *Bacillus subtilis* subsp. *subtilis* str. 168 | *Bacillus cereus* B4264 |
| *Bacillus thuringiensis* BMB171 | *Bacillus cereus* E33L |
| *Bacillus weihenstephanensis* KBAB4 | *Bacillus cereus* G9842 |
| *Bacillus amyloliquefaciens* DSM7 | *Bacillus cereus* Q1 |
| *Bacillus amyloliquefaciens* FZB42 | *Bacillus cytotoxicus* NVH 391-98 |
| *Bacillus cereus* ATCC 14579 | |

Bacillibactin is a catecholate siderophore and a cyclic trimeric ester made of three units of 2,3-dihydroxybenzoate-glycine-threonine. It is synthesized by a complex process catalyzed by the bacillibactin synthase multienzyme complex. The process starts with the activation of 2,3-dihydroxybenzoate , catalyzed by 2,3-dihydroxybenzoate-AMP ligase , encoded by *DhbE* . The product is transferred onto the aryl carrier protein (ArCP) domain of the dhbB bi-functional protein (the second function being isochorismatase), where it is attached to the free thiol group of its cofactor, 4'-phosphopantetheine. The cofactor attachment is catalyzed by the 4-phosphopantetheinyl transferase protein, encoded by the *sfp* gene (Grossman *et al*., 1993). The next steps are the additions of glycine and L-threonine to the activated 2,3-dihydroxybenzoate. These amino acids are first bound to two peptidyl-carrier-protein domains of the seven-domain protein *holo*-DhbF protein , which must be activated first by the binding of two 4'-phosphopantetheine cofactors, presumably catalyzed by 4-phosphopantetheinyl transferase (May *et al*., 2001).

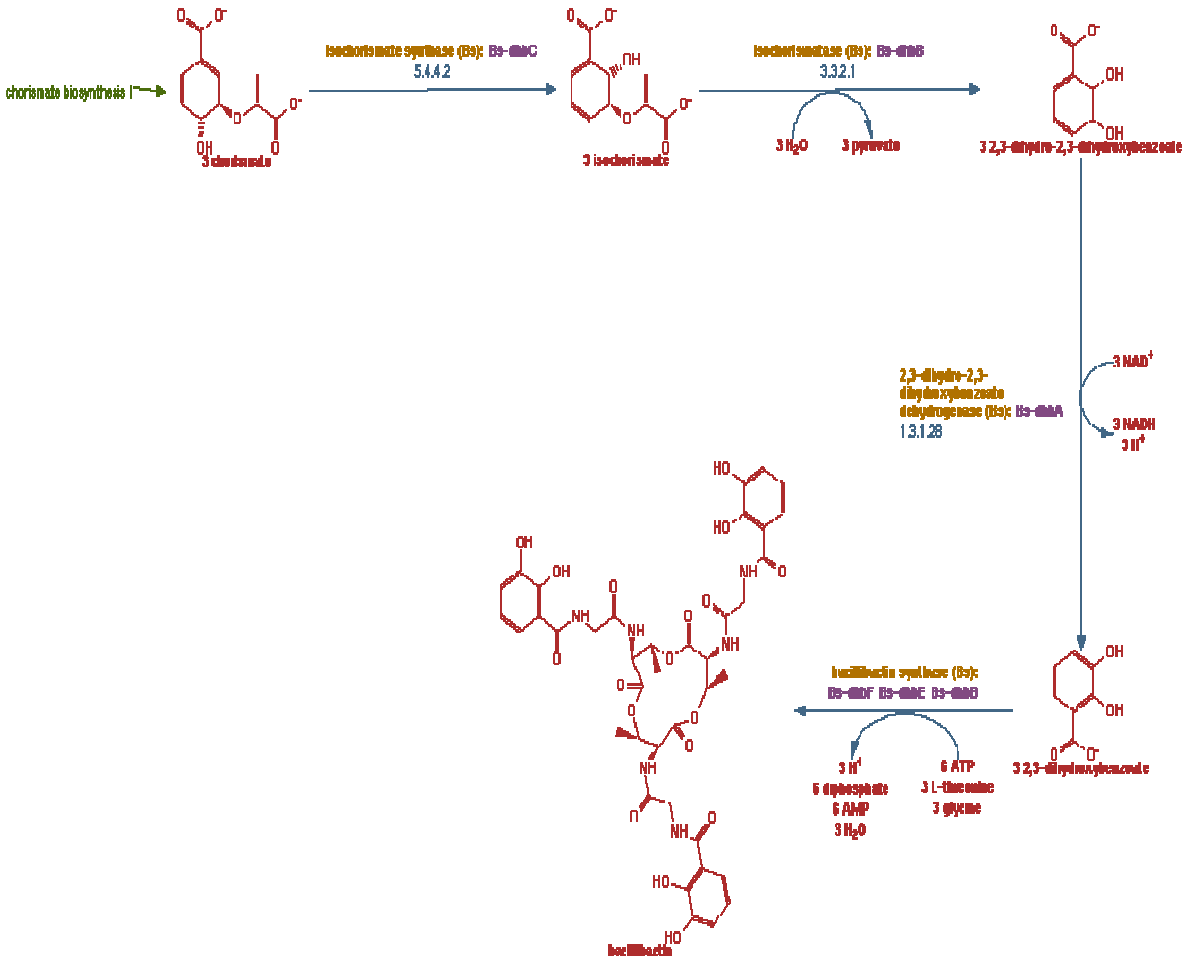


Figure 31: Bacillibactin biosynthesis pathway

### 1.1.4. Detected antibiotics

Genes responsible for the biosynthesis of three antibiotics were detected from five strains (Table 18); two antibiotics, gramicidin and tyrocidin from *Brevibacillus brevis* and one antibiotic, truncated bacitracin from four strains; *B. anthracis* str. CDC 684, *B. atrophaeus* 1942, *B. cereus* AH820 and *B. cereus* G9842.

Table 18: The different antibiotics for which biosynthetic genes were detected in *Brevibacillus brevis*

| ***Brevibacillus brevis* antibiotics structure** | |
|---|---|
| **Gramicidin** | **Tyrocidin** |
| **D-Asx_X_Gln/Gly/Ala/Val/Leu_Phe/Thr/Tyr_ Asx/Glx/Ser/Thr_Gly** | **Trp_Ser/Thr_Gly/Ala/Val/Leu_Gly/Ala/Val/Leu_ Asx/Glx_Asp/Arg/Asn_D-Orn/Lys** |

### 1.1.5. Detection of NRPS genes responsible for the biosynthesis of unknown molecules

Table 19: Predicted molecules from NRPS genes from 16 *Bacillus* strains

| ***Bacillus* strains** | **Structure of the predicted molecule** |
|---|---|
| *Bacillus amyloliquefaciens* DSM7 | X_X |
| *Bacillus pumilus* SAFR-032 | D-Asx<br>X<br>PKS_D-X _ PKS_Ile(Te) |
| *Bacillus subtilis* subsp. *natto* BEST195 | Leu/Ile/Val(Te) + other NRPS molecules |
| *Bacillus subtilis* subsp. *spizizenii* str. W23 | Gly_PKS<br>X_PKS (Bacillaene?) |
| *Bacillus thuringiensis* BMB171 | Thr _ Gln_Gly _ Gln_Ser_D-X_Gln(Te)<br>X_D-Pro _ Leu_Pro_Leu<br>Val |
| *Bacillus weihenstephanensis* KBAB4 | D-Thr _ Gln_Gly _ Thr_X_D-Ser_Glu/Asp(Te)<br>NRPS/PKS (X_X _ PKS_ Asx _ D-X_Orn_Asp)<br>X_X |
| *Brevibacillus brevis* | PKS-L-Gly-L-Gln-L-X-L-Gln<br>L-X-PKS<br>PKS-D-X<br>L-Phe-PKS-PKS-L-Gln-D-Gln<br>PKS-L-X-D-X-L-Val-L-X-L-Val-L-X-L-Trp |
| *Paenibacillus polymyxa* 681 | L-Thr-D-X-L-Val-D-Tyr-D-Asx-L-Thr (Te)<br>L-X<br>PKS-L-X-PKS |
| *Paenibacillus polymyxa* SC2 | PKS-L-Gln-D-X-L-Ser-D-Ser-L-Leu-D-Ser-L-Ser<br>L-Ile-D-Ser-L-Ile<br>PKS-D-3h4mPhe-D-Ile-L-X<br>D-Val-L-Val-L-Leu-PKS<br>PKS-L-Gly-PKS |
| *Bacillus cereus* B4264 | Val |
| *Bacillus cereus* G9842 | L-Ala-L-Phe<br>L-Val |
| *Bacillus cereus* Q1 | X |
| *Bacillus cytotoxicus* NVH 391-98 | X<br>Ser |
| *Bacillus cereus* E33L | D-Asx-L-Lys-D-X-L-X-D-HPG-L-X |
| *Bacillus cereus* 14579 | L-Val |
| *Bacillus cereus* AH178 | L-X<br>PKS-L-Val(Te)<br>PKS-L-X-D-X |

Table 19 shows the predicted product of NRPS genes detected in 16 *Bacillus* strains and which are responsible for the biosynthesis of unknown molecules: 73 genes were detected, 24 of them from 11 strains and 13 genes from only three strains; *Brevibacillus brevis*, *Paenibacillus polymyxa* 681 and *Paenibacillus polymyxa* SC2

Both *Paenibacillus polymyxa* SC2 and *Brevibacillus brevis* have five new molecules. Six *Bacillus* strains harbour three new molecules; *Paenibacillus polymyxa* 681, *Bacillus pumilus* SAFR-032, *Bacillus thuringiensis* BMB171, *Bacillus cereus* AH178, *Bacillus cereus* E33L and *Bacillus weihenstephanensis* KBAB4. Four other strains harbour two new molecules; *Bacillus subtilis* subsp. *spizizenii* str. W23, *Paenibacillus polymyxa* SC2, *Bacillus subtilis* subsp. *natto* BEST195 and *B. cytotoxicus* NVH 391-98. The other four strains harbor one new NRPS molecule.

Table 20: Thirty-nine *Bacillus, Brevibacillus* and *Paenibacillus* strains with all NRPS and PKS molecules types detected by bioinformatics tools and some comments about their origin or characteristic propertie

| ***Bacillus* strains** | **Siderophore** | **Lipopeptide** | **Antibiotics** | **New molecules** | **Comment** |
|---|---|---|---|---|---|
| *Bacillus clausii* KSM-K16 | 0 | 0 | 0 | 0 | alkaliphilic |
| *Bacillus cytotoxicus* NVH 391-98 | Bacillibactin | 0 | 0 | 2 | alkaliphilic |
| *Bacillus halodurans* C-125 | 0 | 0 | 0 | 0 | alkaliphilic |
| *Bacillus pseudofirmus* OF4 | 0 | 0 | 0 | 0 | alkaliphilic |
| *Bacillus selenitireducens* MLS10 | 0 | 0 | 0 | 0 | alkaliphilic |
| *Bacillus cellulosilyticus* DSM 2522 | 0 | 0 | 0 | 0 | alkaliphilic |
| *Bacillus tusciae* DSM 2912 | 0 | 0 | 0 | 0 | Acidophilic, thermophilic |
| *Bacillus subtilis* subsp. *natto* BEST195 | Bacillibactin | Surfactin | 0 | 1 | soil |
| *Bacillus subtilis* subsp. *spizizenii* str. W23 | Bacillibactin | Surfactin<br>Mycosubtilin | 0 | 2 | soil |
| *Bacillus subtilis* subsp. *subtilis* str. 168 | Bacillibactin | Plipastatin<br>Surfactin | 0 | 0 | soil |
| *Bacillus amyloliquefaciens* FZB42 | Bacillibactin | Surfactin<br>Fengycin<br>Bacillomycin | Bacillaene | 0 | soil, plants |
| *Bacillus amyloliquefaciens* DSM7 | Bacillibactin | Surfactin | Bacillaene | 1 | soil, plants |
| *Bacillus licheniformis* ATCC 14580 | Bacillibactin | Lichenysin | 0 | 0 | soil, humain |
| *Bacillus thuringiensis* str. Al Hakam | Bacillibactin | 0 | 0 | 0 | insecticide |
| *Bacillus thuringiensis* BMB171 | Bacillibactin | Kurstakin | 0 | 3 | insecticide |
| *Brevibacillus brevis* | 0 | 0 | Potential linear gramicidin<br>Tyrocidin | 4 | antibacterial |
| *Paenibacillus polymyxa* 681 | 0 | 0 | 0 | 5 | Antifungal, antibacterial |

| | | | | | |
|---|---|---|---|---|---|
| *Bacillus pumilus* SAFR-032 | 0 | Pumilacidin | 0 | 3 | antifungal |
| *Paenibacillus polymyxa* SC2 | 0 | Fusaricidin | 0 | 2 | Antifungal, antibacterial |
| *Bacillus anthracis* str. 'Ames Ancestor' | Bacillibactin | 0 | 0 | 0 | virulence |
| *Bacillus anthracis* str. A0248 | Bacillibactin | 0 | 0 | 0 | virulence |
| *Bacillus anthracis* str. Ames | Bacillibactin | 0 | 0 | 0 | virulence |
| *Bacillus anthracis* str. CDC 684 | Bacillibactin | 0 | Truncated bacitracin | 0 | virulence |
| *Bacillus anthracis* str. Sterne | Bacillibactin | 0 | 0 | 0 | virulence |
| *Bacillus megaterium* DSM 319 | 0 | 0 | 0 | 0 | Various habitats |
| *Bacillus megaterium* QM B1551 | 0 | 0 | 0 | 0 | Various habitats |
| *Bacillus cereus* 03BB102 | Bacillibactin | 0 | 0 | 0 | pneumonia |
| *Bacillus cereus* B4264 | Bacillibactin | Kurstakin | 0 | 1 | pneumonia |
| *Bacillus atrophaeus* 1942 | Bacillibactin | Mycosubtilin | Truncated bacitracin | 0 | decontamination |
| *Bacillus cereus* Q1 | Bacillibactin | 0 | 0 | 1 | Non pathogenic |
| *Bacillus cereus* AH820 | Bacillibactin | 0 | Truncated bacitracin | 0 | buccal |
| *Bacillus cereus* ATCC 10987 | Bacillibactin | 0 | 0 | 0 | cheese |
| *Bacillus cereus* ATCC 14579 | Bacillibactin | Kurstakin | 0 | 1 | |
| *Bacillus cereus* E33L | Bacillibactin | 0 | 0 | 3 | carcass |
| *Bacillus cereus* G9842 | Bacillibactin | Kurstakin | Truncated bacitracin | 2 | gastro |
| *Bacillus cereus* AH187 | Bacillibactin | 0 | 0 | 3 | cereulide, gastro |
| *Bacillus cereus* biovar *anthracis* str. CI | Bacillibactin | 0 | 0 | 0 | cereulide, gastro |
| *Bacillus thuringiensis* serovar *konkukian* str. 97-27 | Bacillibactin | 0 | 0 | 0 | human necrosis |
| *Bacillus weihenstephanensis* KBAB4 | Bacillibactin | Kurstakin | 0 | 3 | cold |

Table 20 shows all the NRPS and PKS molecules detected by bioinformatics tools from the 39 *Bacillus* sequenced genomes. The Table shows ecological origin or specific properties for each strain. Six strains were found to be alkaliphilic and to harbor no NRPS or PKS genes. Five *B. anthracis* strains have virulence character, four of them harbour only bacillibactin, while the fifth *Bacillus anthracis* str. CDC 684 harbour truncated bacitracin in addition to the bacillibactin. Among three *B. subtilis* strains found in soil, one of them *Bacillus subtilis* subsp. *subtilis* str. 168 harbours two lipopeptides; plipastatin and surfactin, while the other two strains harbour bacillibactin, two lipopeptides and new NRPS genes. From *B. cereus* group; two strains are pathogenic and cause pneumonia, one of them harbour only bacillibactin (*B. cereus* 03BB102), while *B. cereus* B4264 harbours two NRPS molecules. Six other *B. cereus* strains showed different ecological characters.

Two *Bacillus thuringiensis* strains have an insecticide effect, *Bacillus thuringiensis* BMB 171 harbour five different NRPS molecules, while *Bacillus thuringiensis* Al Hakam harbours only bacillibactin. *Bacillus thuringiensis* serovar *konkukian* str. 97-27 cause human necrosis and harbours bacillibactin. The antifungal and antimicrobial effect was mentioned for two *Paenibacillus polymyxa* strains, whereas *Paenibacillus polymyxa* 681 harbours five NRPS molecules and *Paenibacillus polymyxa* SC2 harbours three NRPS molecules. Both *Bacillus amyloliquefaciens* DSM7 and FZB42 strains, from soil and plants origin, harbour five and four NRPS molecules, respectively. *Bacillus tusciae* DSM 2912 is considered as acidophilic, thermophilic and has no NRPS gene. *Brevibacillus brevis* has antimicrobial effect and harbours six different NRPS molecules, while *Bacillus weihenstephanensis* KBAB4 may tolerate low temperatures and harbours five NRPS molecules.

### 1.1.6. Conclusion

In this study, the bioinformatics analyses by keywords searched among 36 *Bacillus*, *Brevibacillus brevis* and two *Paenibacillus polymexa* strains protein annotations revealed four main types of molecules; lipopeptides, siderophores, antibiotics, and new NRPS and PKS molecules. The analysis has to be completed trying to detect more potentially interesting NRPS and PKS genes by Blast approach to avoid missing a gene because of a poor or mis-annotation of the corresponding protein.

For the lipopeptides, the presence of members of the four main families was detected in fourteen different *Bacillus* strains [surfactin, pumilacidin, lichenysin from surfactin family, fengycin or plipastatin from fengycin/plipastatin family, bacillomycin and mycosubtilin from iturin family and kurstakin]. In addition, the presence of other lipohexapeptide from the fusaricidin lipopeptide group produced by *Paenibacillus polymyxa* (formerly *Bacillus polymyxa*) was also found. For the siderophores, the presence of bacillibactin in twenty-seven different *Bacillus* strains was detected by bioinformatics tools which confirmed that this siderophore is the main one used by this genus. For the antibiotics, five *Bacillus* strains showed the presence of three antibiotics; two antibiotics, gramicidin and tyrocidin, from *Brevibacillus brevis* and one antibiotic, truncated bacitracin, was detected from the four strains, *B. anthracis* str. CDC 684, *B. atrophaeus* 1942, *B. cereus* AH820 and *B. cereus* G9842. Concerning new NRPS and PKS molecules, fourteen *Bacillus* strains were found to harbour new NRPS and PKS genes, especially *Brevibacillus brevis*, *Paenibacillus polymyxa* E681 and *Paenibacillus polymyxa* SC2 which bears a high number of these new genes. Finally, eight *Bacillus* strains do not contain NRPS genes.

## 1.2. Bioinformatics analyses on plipastatin and fengycin sequenced operons

The following parts of the work will be focused on plipastatin and fengycin. The previous bioinformatics analyses showed that there are only two strains (*B. amyloliquefaciens* FZB42 and *B. subtilis* 168) carrying a complete fengycin and plipastatin operon sequence, in addition to the fengycin operon of *B. subtilis* F29-3 which was separately sequenced (Lin *et al*., 1998, 1999, 2005; Shu *et al*., 2002; Wu *et al*., 2007)). In this chapter, we will compare these available operons sequences to detect the differences between them.

Both fengycin and plipastatin operons are composed of five genes (*ppsA-ppsE*) in *B. subtilis* 168, (*fenA-fenE*) in *B. amyloliquefaciens* FZB42 and *fenC*, *D*, *E*, *A*, *B* in *B. subtilis* F29-3. The gene sequences of both fengycin and plipastatin operons of these three strains were analyzed by the GenBank database search tools provided by the National Center for Biotechnology Information (NCBI, USA) and Needle software (Needleman and Wunsch, 1970).

Table 21 shows the proteic and nucleic identity percentages between the genes included in plipastatin and fengycin operons of these strains. Proteic and nucleic identity % were found higher between *B. subtilis* 168 and *B. subtilis* F29-3 than between *B. subtilis* F29-3 and *B. amyloliquefaciens* FZB42 and than *B. subtilis* 168 and *B. amyloliquefaciens* FZB42, respectively. In the comparison between *B. subtilis* 168 and *B. subtilis* F29-3, the highest proteic and nucleic identity % were between the peptides *ppsC* and *fenE* (90.1% proteic identity and 80.5% nucleic identity) while the lowest identity was observed for the peptides *ppsA* and *fenC* (87.6% proteic identity and 73.3% nucleic identity). In the comparison between *B. subtilis* F29-3 and *B .amyloliquefaciens* FZB42, the highest proteic and nucleic identity % were between *fenE* and *fenC* (81.3% proteic identity and 68.6% nucleic identity), while the lowest identity was observed for the peptides *fenB* and *fenE* (77.5% proteic identity and 63.6% nucleic identity). The comparison between *B. subtilis* 168 and *B. amyloliquefaciens* FZB42 showed the highest identity % between *ppsC* and *fenC* (80.6% proteic identity and 68.3% nucleic identity), whereas the lowest identity was observed for the peptides *ppsE* and *fenE* (76.2% proteic identity and 65.0% nucleic identity).

Table 21: Comparative analysis of NRPS synthetases genes and proteins of *Bacillus* strains (*subtilis* 168, *subtilis* F29-3 and *amyloliquefaciens* FZB42) using Needle Global software

| Sequence Origins | Fengycin and Plipastatin genes | Protein identity (%) | DNA identity (%) |
|---|---|---|---|
| *B. subtilis* 168-*B. subtilis* F29-3 | *ppsA*/*fenC* | 87.6 | 73.3 |
| | *ppsB*/*fenD* | 89.7 | 79.4 |
| | *ppsC*/*fenE* | 90.1 | 80.5 |
| | *ppsD*/*fenA* | 88.2 | 80.0 |
| | *ppsE*/*fenB* | 88.4 | 78.0 |
| *B. subtilis* F29-3-*B. amyloliquefaciens* FZB42 | *fenC*/*fenA* | 78.0 | 60.9 |
| | *fend*/*fenB* | 78.1 | 65.3 |
| | *fenE*/*fenC* | 81.3 | 68.6 |
| | *fenA*/*fenD* | 78.6 | 66.6 |
| | *fenB*/*fenE* | 77.5 | 63.6 |
| *B. subtilis* 168-*B. amyloliquefaciens* FZB42 | *ppsA*/*fenA* | 78.1 | 66.3 |
| | *ppsB*/*fenB* | 77.7 | 65.3 |
| | *ppsC*/*fenC* | 80.6 | 68.3 |
| | *ppsD*/*fenD* | 78.9 | 65.5 |
| | *ppsE*/*fenE* | 76.2 | 65.0 |

The main differences identified between fengycin and plipastatin are the positions of the L- and D- forms of tyrosine, which are in position 3 and 9, respectively, for plipastatins and 9 and 3 for fengycins. For this reason, we should compare both modules 3 and 9 for the three operons. At position 3, no epimerization domain was detected, only adenylation and condensation domains were mentioned as shown in Tables 22. The first condensation and adenylation domains of M3 module (Tyr) shows the highest identity between *B. subtilis* 168 and *B. subtilis* F29-3 (90.4% for C domain and 82.5% for A domain), followed by *B. subtilis* 168 and *B. amyloliquefaciens* FZB42 (80.2% for C domain and 78.3% for A domain) and *B. subtilis* F29-3 and *B. amyloliquefaciens* FZB42 (79.9% for C domain and 71.5% for A domain).

Table 22: Comparative analysis of the first condensation (C) and adenylation (A) domains of the third module (M3) from *Bacillus* strains (*subtilis* 168, *subtilis* F29-3 and *amyloliquefaciens* FZB42) using Needle Global software (Data from NCBI) at nucleic level

| Domain | *B. subtilis* 168- *B. subtilis* F29-3 | *B. subtilis* F29-3- *B. amyloliquefaciens* FZB42 | *B. subtilis* 168- *B. amyloliquefaciens* FZB42 |
|---|---|---|---|
| C (1st) (M3) | | | |
| identity (%) | 90.4 | 79.9 | 80.2 |
| Gaps (%) | 0.7 | 0.0 | 0.7 |
| A (M3) | | | |
| identity (%) | 82.5 | 71.5 | 78.3 |
| Gaps (%) | 9.3 | 9.6 | 0.8 |

Table 23: Comparative analysis of the last condensation (C), adenylation (A) and epimerization (E) domains of the ninth module (M9) from *Bacillus* strains (*subtilis* 168, *subtilis* F29-3 and *amyloliquefaciens* FZB42) using Needle Global software (Data from NCBI) at nucleic level

| Domain | *Bacillus subtilis* 168- *Bacillus subtilis* F29-3 | *B. subtilis* F29-3- *B. amyloliquefaciens* FZB42 | *Bacillus subtilis* 168- *B. amyloliquefaciens* FZB42 |
|---|---|---|---|
| C (last) (M9) | | | |
| identity (%) | 86.7 | 80.3 | 80.6 |
| Gaps (%) | 0.2 | 0.0 | 0.2 |
| A (M9) | | | |
| identity (%) | 91.8 | 80.5 | 79.6 |
| Gaps (%) | 0.0 | 2.7 | 2.7 |
| **E** (M9) | | | |
| identity (%) | 85.9 | 75.3 | 75.1 |
| Gaps (%) | 1.6 | 1.6 | 1.6 |

As shown in Table 23, the M9 module (Tyr) differs from that of M3 by the presence of the E domain in addition to C and A domains and which refers to the presence of D-Tyr on the M9 of the *ppsD* (*B. subtilis* 168), *fenD* (*B. amyloliquefaciens* FZB42) and *fenA* (*B. subtilis* F29-3). This also may refer to the similarity of all these molecules which are plipastatins. The existence of fengycin molecule with a D-Tyr in position 3 should be then limited to the molecule which was analyzed by Schneider *et al*. (1999) from the supernatant of *B. subtilis* S499.

### 1.3. Conclusion

The analysis of plipastatin or fengycin operons sequences of *B. subtilis* F29-3, *B. subtilis* 168 and *B. amyloliquefaciens* FZB42 revealed the similarity of these molecules and raised the importance of sequencing the *B. amyloliquefaciens* S499 fengycin operon. This latter codes for a fengycin synthetases that cannot be correlated with the structure of the synthetases described in other fengycin- or plipastatin-producing strains. Deciphering this operon should allow to overcome the difficulties in the differentiation between fengycins and plipastatins.

### 1.4. Results valorization

The results allowed participating in the following international conferences:

Leclère V, Pupin M, Hussein W, Béchet M, Gancel F, Chollet M, Caboche S and Jacques P. 2011. The power of bionformatics strategy to decrypt NonRibosomal Peptides biosynthesis in *Bacilli*. 6$^{th}$ International Conference on Gram Positive Microorganisms, Montecatini, Italy

Leclère V, Pupin M, Hussein W, Béchet M, Gancel F, Chollet M, Caboche S and Jacques P. 2011. Leclère V, Pupin M, Hussein W, Béchet M, Gancel F, Chollet M, Caboche S and Jacques P. 2011. Bioinformatics tools to decrypt non-ribosomal peptide diversity in *Bacilli*. Bacell Conference, Göttingen, Germany

# 2. Differentiation between fengycin and plipastatin

## Introduction

In the previous chapter we described an overview of the Non Ribosomal Peptides produced by 39 *Bacillus* strains. We here will focus our work on fengycins or plipastatins operons produced by the NRPS system in some *B. subtilis* strains, in order to relate the structure of these two molecules with the structure of operons responsible for their biosynthesis.

Fengycins and plipastatins are very similar lipo-decapeptides and only two differences were identified between fengycin and plipastatin: a Gln instead of a Glu in position 8 and the L and D forms of tyrosine, which are in positions 3 and 9 for plipastatins and 9 and 3 for fengycins, respectively. In the different works describing fengycin structure since its discovery, the presence of a Glu in position 8 was never mentioned. However, Schneider *et al*. (1999) confirmed the existence of fengycin molecule with a D-Tyr in position 3 from the supernatant of *B. subtilis* S499. This result could not be correlated with the structure of the synthetases described in other fengycin- or plipastatin-producing strains. According to this information on *B. subtilis* S499 strain, it was interesting to study the sequence of its fengycin operon. Two other fengycin or plipastatin producing strains were identified in ProBioGEM Laboratory (Gancel *et al*., 2009, Coutte *et* al. 2010, Tapi *et al*. 2009): *B. subtilis* 21332 and *B. subtilis* BBG21, which is a spontaneous mutant of *B. subtilis* 21332 that was found to produce high amounts of fengycin, compared to the mother strain *B. subtilis* 21332. In this last case, it was interesting to discover the changes that happened in this new mutant at the fengycin or plipastatin operon level.

In order to analyse these operons, several primers were designed to amplify and sequence different fragments. As the three tested strains are *B. subtilis*, all these primers were designed using the published sequence of *B. subtilis* 168 plipastatin operon (accession no. AL009126). The list of the different primers used is given in Materials and Methods (chapter IV, 2.1.2.2). The different obtained amplified fragments were summarized in Figure 32.

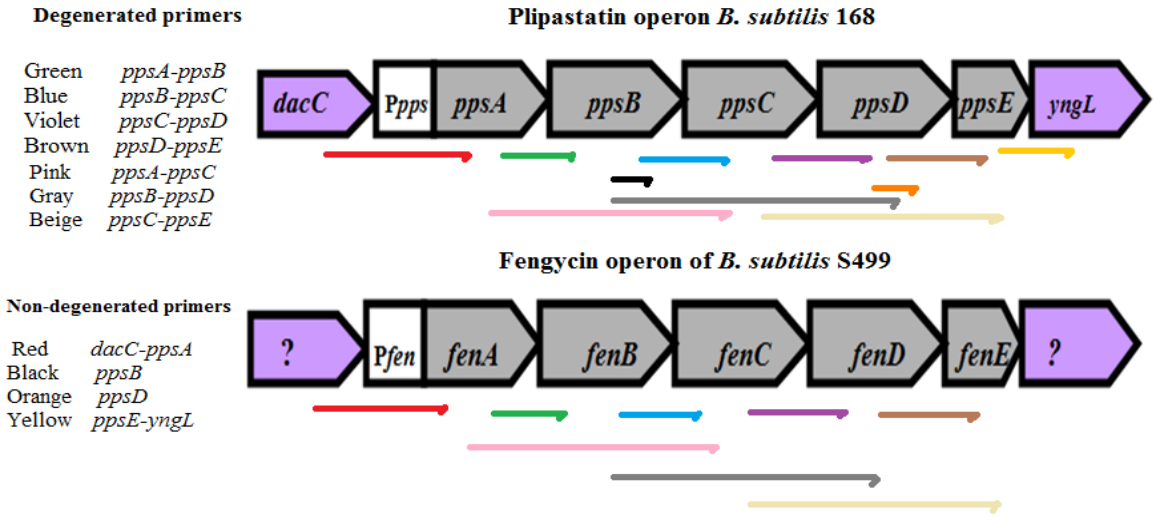



Figure 32: DNA fragments amplified by generated and non-degenerated primers in both plipastatin operon of *B. subtilis* 168 and potential fengycin operon of *B. subtilis* S499 strains.

## 2.1. Analysis of fengycin or plipastatin operons from *B. subtilis* 21332 and BBG21

### 2.1.1. Detection of the epimerization domain

The first primers pair was designed for the detection of the epimerization domain in module 3 (*ppsB* gene) and 9 (*ppsD* gene). The analysis was performed on *B. subtilis* 21332, BBG21 and 168. The results are shown in Figure 33.

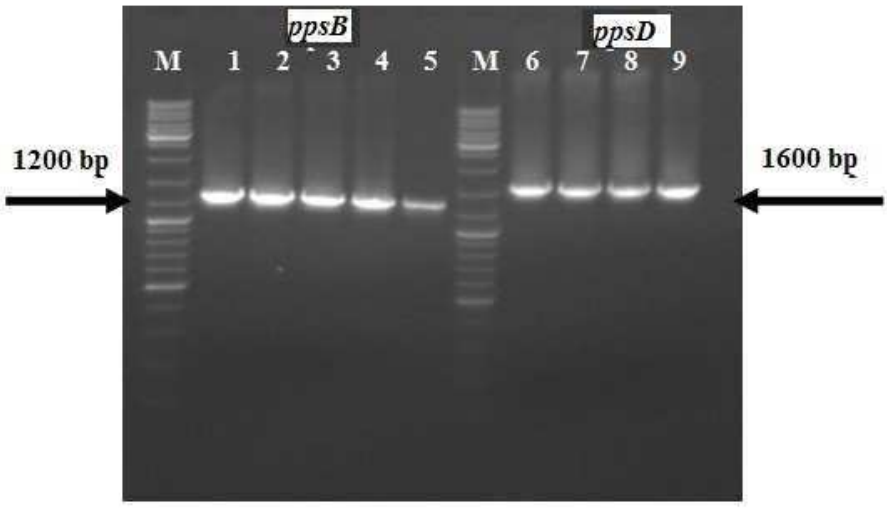



Figure 33: Amplicons generated by primers designed to detect the presence of an epimerisation domain in module 3 (*ppsB*) and 9 (*ppsD*) of the plipastatin operon of *B. subtilis* 168, BBG21 and ATCC21332 ; *ppsB* primers lanes (1-5), *ppsD* primers lanes (6-9) ; lanes (1,6) for *B. subtilis* 168 ; lanes (2,3,7,8) for *B. subtilis* BBG21 ; lanes (4,5,9) for *B. subtilis* ATCC21332. M = O'Gene Ruler standard.

As shown in Figure 33, both the two primers of *ppsB* and *ppsD* gave the same fragments length (1300 bp for *ppsB* and 1600 bp for *ppsD*), as for the control strain *B. subtilis* 168, in both strains *B. subtilis* BBG21 and 21332. The difference of the sizes of these fragments, 300 bp, correspond to the size of an epimerisation domain. After cloning of these fragments in pGEM-T Easy vector and sequencing, the fragments were not completely sequenced.

The Needle EMBOSS protein alignment of the sequenced E domain of *ppsD* of *B. subtilis* BBG21 and ATCC21332 with *B. subtilis* 168 showed high similarity, whereas 95.5% of identity was observed between E domain of *B. subtilis* 168 and BBG21 and 96.7% of identity between E domain of *B. subtilis* 168 and 21332, as shown in Figures 34 and 35. The epimerisation domains can be easily recognized by the presence of 6 consensus sequences (Table 24). Two of them were recovered in the fragment from strain 21332 and three of them in the fragment from strain BBG21.

Table 24 : Comparison of consensus sequences detected in epimerization domain regions of *ppsD* gene between *B. subtilis* 168 and *B. subtilis* BBG21 and 21332

| **(E) Domain** | Core sequences in *B. subtilis* 168 | Core sequences in *B. subtilis* BBG21 | Core sequences in *B. subtilis* 21332 |
|---|---|---|---|
| ***ppsD* gene** | E1 PVQKWF | ns | PVQKWF |
| | E2 GPLLQAGL | GPLLQAGL | GPLLQAGL |
| | E3 EGHGRE | EGHGRE | ns |
| | E4 RTIGWFTAIYPLLIK | RTIGWFTAIYPLLIK | ns |
| | E5 PDKGFGYG | ns | ns |
| | E6 FNYLGQ | ns | ns |

ns : non sequenced region


Figure 34: Needle EMBOSS alignment of the E domain protein of the sequenced *ppsD* of *B. subtilis* BBG21 (first line) and *B. subtilis* 168 (second line).


Figure 35: Needle EMBOSS alignment of the E domain protein of the sequenced *ppsD* of *B. subtilis* 21332 (first line) and *B. subtilis* 168 (second line).

### 2.1.2. Confirmation of the genes present on each side of the fengycin or plipastatin operons

Two sets of primers (*dacC-ppsA*) and (*ppsE-yngL*) were designed to confirm the genes present on each side of the fengycin or plipastatin operons in both strains ATCC21332 and BBG21: *dacC* and *yngL*.

#### 2.1.2.1. *dacC-ppsA* primers

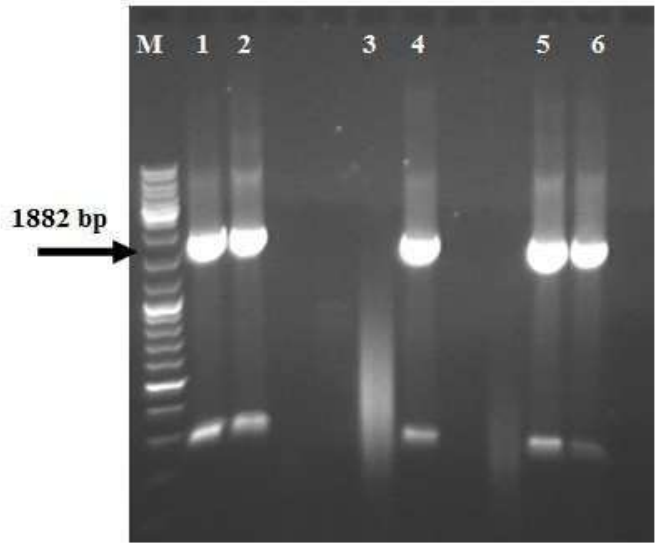



Figure 36: Amplicons generated by *dacC-ppsA* primers from *B. subtilis* BBG21 and 21332 ; lanes (1,2) from *B. subtilis* 168 ; lanes (3,4) from *B. subtilis* BBG21 ; lanes (5,6) from *B. subtilis* ATCC21332, M = O'Gene Ruler standard.

Figure 36 shows the PCR product of *B. subtilis* 168, BBG21 and ATCC21332 with the *dacC-ppsA* primers and which gave the same fragment size (1882 bp) with the three strains. The Needle EMBOSS protein alignment of the this fragments of *B. subtilis* 21332 with *B. subtilis* 168 showed high identity of 99% (Table 25) .

#### 2.1.2.2. *ppsE-yngL* primers

Figure 37 shows the PCR product from *B. subtilis* 168, BBG21 and ATCC21332 with the *ppsE-yngL* primers, which gave the same fragment size (961 bp) with the three strains. The Needle EMBOSS protein alignment of the of *B. subtilis* 21332 fragment with *B. subtilis* 168 showed a high identity of 99.4% (Table 25), while the fragment of *B. subtilis* BBG21 wasn't sequenced. In this experiment, PCR products from *B. subtilis* ATCC9943 and S499 were

obtained with the same primers. For both strains, the length of the amplicons is higher than for the three other strains indicating a different organization of the gene *ppsE* or the presence of a modified or another gene at the extermity of the NRPS operon.

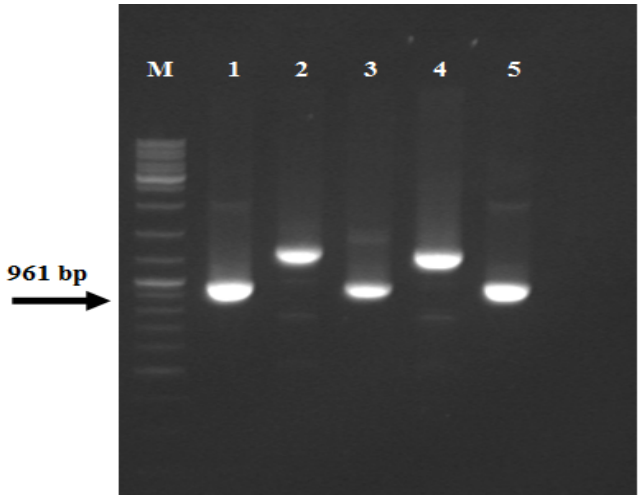



Figure 37: Amplicons generated by *ppsE-yngL* primers ; lane (1) from *B. subtilis* 168 ; lane (2) from *B. subtilis* ATCC9943; lane (3) from *B. subtilis* BBG21; lane (4) from *B. subtilis* S499; lane (5) from *B. subtilis* 21332, M = O'Gene Ruler standard.

### 2.1.4. Analysis with fengycin and plipastatin degenerated primers

Different degenerated primers were used to confirm the organization of fengycin/plipastatin operon in *B. subtilis* ATCC21332 and BBG21. Two sets of degenerated primers designed for the amplification of long DNA fragment from fengycin operon. These degenerated primers amplified fragment from *ppsA-ppsB* and *ppsC-ppsD* genes. They were designed by the alignement of fengycin/plipastatin operons of *B. subtilis* 168, F29-3 and *B. amyloliquefaciens* FZB48. The second set of degenerated primers were previously designed by Tapi *et al.* (2009) to screen for strains containing plipastatin or fengycin operons in their genome. These primers were designed by the alignment of the adenylation and thiolation domains of the different *Bacillus subtilis* F29-3 or 168 sequences of fengycin and plipastatin operons, respectively.

#### 2.1.4.1. *ppsA-ppsB* and *ppsC-ppsD*

Figure 38 shows the PCR products obtained with these fragments in *B. subtilis* 168, BBG21 and ATCC21332. Amplicons of the same size (760 and 1074 bp, respectively) were obtained with the three strains. After sequencing of these fragments, the Needle EMBOSS protein alignment of the *ppsA-ppsB* fragment of *B. subtilis* 21332 with *B. subtilis* 168 showed a high identity of 97.3%, the *ppsA-ppsB* fragment of *B. subtilis* BBG21 showed 98.1% identity with

*B. subtilis* 168, whereas the alignment of this fragment of *B. subtilis* BBG21 and *B. subtilis* 21332 showed a identity of 99.2%. Needle EMBOSS protein alignment of the *ppsC-ppsD* fragment of *B. subtilis* 21332 with *B. subtilis* 168 showed a high identity of 97.7%, the *ppsA-ppsB* fragment of *B. subtilis* BBG21 showed 97.6% identity with *B. subtilis* 168, while the alignment of this fragment of *B. subtilis* BBG21 and *B. subtilis* 21332 showed identity of 99.6% (Table 25).

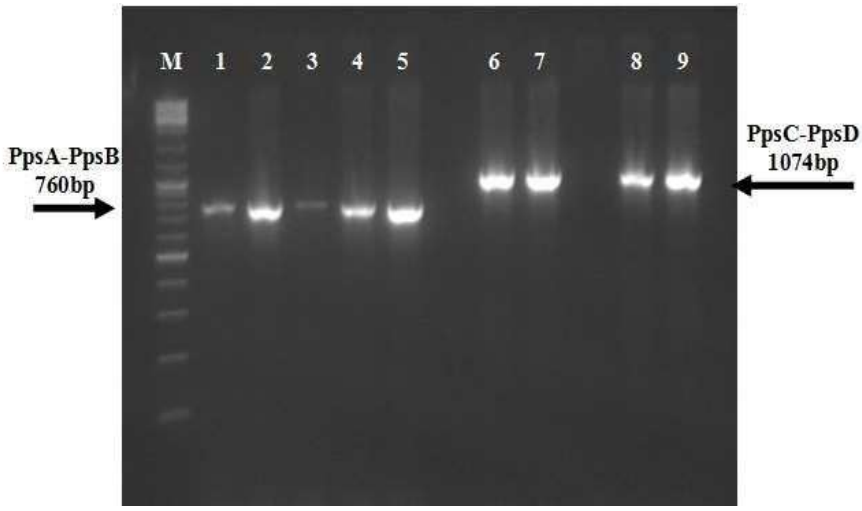



Figure 38: Amplicons generated by *ppsA-ppsB* primers lanes (1-5), *ppsC-ppsD* lanes (6-9) ; lane (1) from *B. subtilis* 168 ; lanes (2,3,6,7) from *B. subtilis* BBG21 ; lanes (4,5,8,9) from *B. subtilis* ATCC21332, M = O'Gene Ruler standard.

### 2.1.4.2. Degenerated primers for detection of fengycin/plipastatin operons

The results obtained with degenerated primers for specific detection of fengycin or plipastatin operons are shown in Figures 39 and 40.

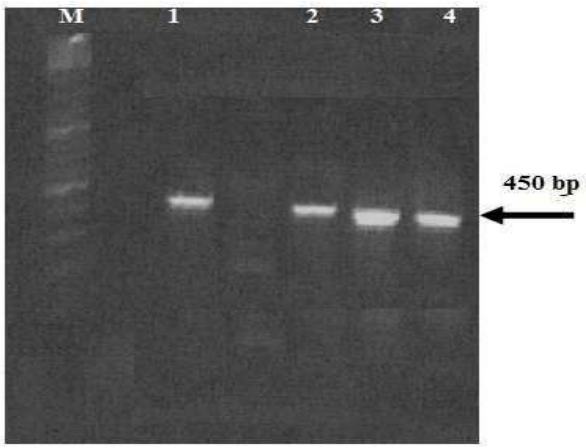



Figure 39: Amplicons revealed with the fengycin degenerated primers ; lane (1-4); lane (1) from *B. subtilis* 168 ; lane (2) from *B. subtilis* BBG21 ; lane (3) from *B. subtilis* S499 and lane (4) from *B. subtilis* 21332, M = O'Gene Ruler standard.

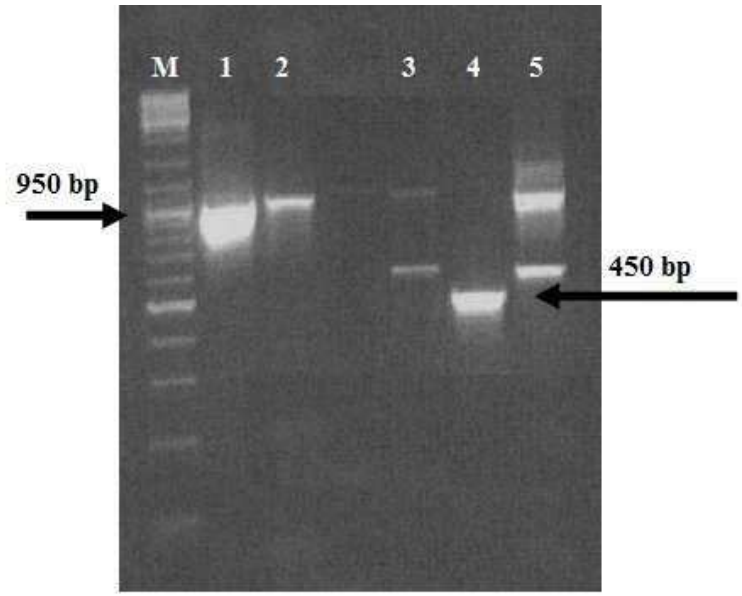
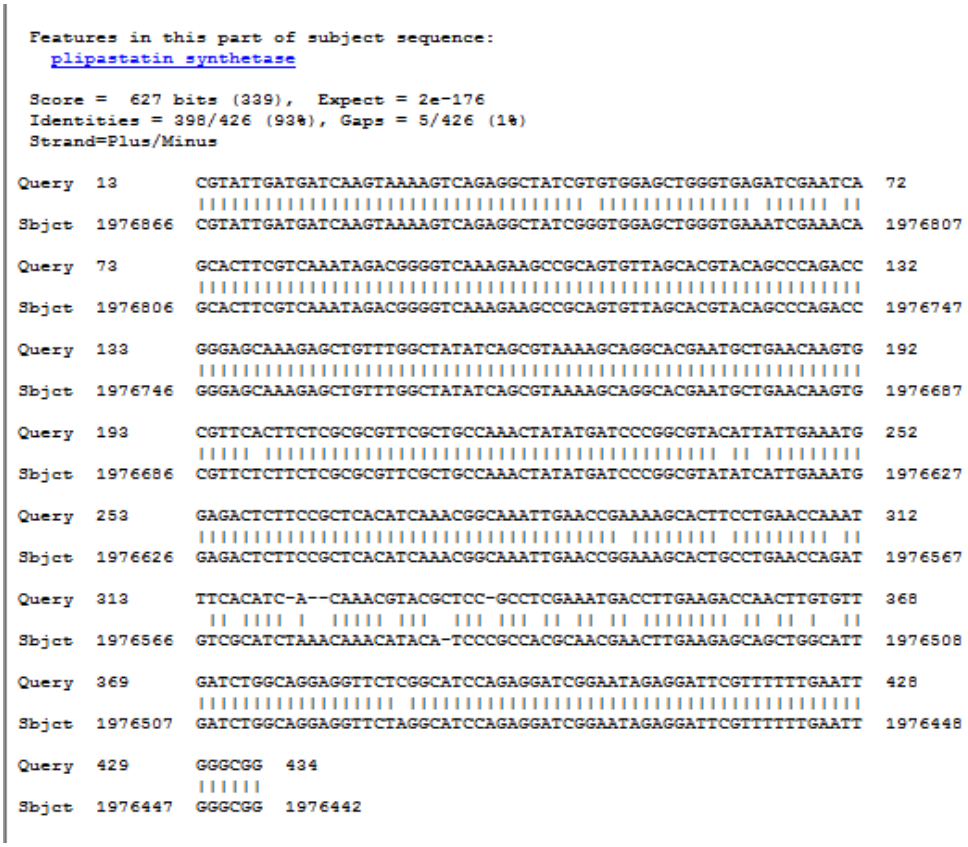



Figure 40: Amplicons revealed with the plipastatin degenerated primers ; lanes (1-5); lanes (1,2) from *B. subtilis* 168 ; lane (3) from *B. subtilis* BBG21 ; lane (4) from *B. subtilis* S499 and lane (5) from *B. subtilis* 21332, M = O'Gene Ruler standard.

The fengycin primers showed a similar fragment of size ranging from 450 bp for the three strains (the same result was obtained with *B. subtilis* S499), while the plipastatin primers showed a fragment of size ranging from 950 bp (the size of the amplified fragment is lower for *B. subtilis* S499). Plipastatin fragment alignment of *B. subtilis* 21332 with *B. subtilis* 168 showed 96% of identity, while the alignment of this fragment of *B. subtilis* BBG21 with *B. subtilis* 168 showed 99% of identity. Fengycin fragment alignment of *B. subtilis* 21332 with *B. subtilis* 168 showed 93% of identity (Figure 41), while the alignment of this fragment of *B. subtilis* BBG21 with *B. subtilis* 168 showed 99% of identity (Figure 42).



```
 Features in this part of subject sequence:
   plipastatin synthetase

 Score =  627 bits (339),  Expect = 2e-176
 Identities = 398/426 (93%), Gaps = 5/426 (1%)
 Strand=Plus/Minus

Query  13       CGTATTGATGATCAAGTAAAAGTCAGAGGCTATCGTGTGGAGCTGGGTGAGATCGAATCA  72
                ||||||||||||||||||||||||||||||||||| ||||||||||||| |||||| ||
Sbjct  1976866  CGTATTGATGATCAAGTAAAAGTCAGAGGCTATCGGGTGGAGCTGGGTGAAATCGAAACA  1976807

Query  73       GCACTTCGTCAAATAGACGGGGTCAAAGAAGCCGCAGTGTTAGCACGTACAGCCCAGACC  132
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1976806  GCACTTCGTCAAATAGACGGGGTCAAAGAAGCCGCAGTGTTAGCACGTACAGCCCAGACC  1976747

Query  133      GGGAGCAAAGAGCTGTTTGGCTATATCAGCGTAAAAGCAGGCACGAATGCTGAACAAGTG  192
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1976746  GGGAGCAAAGAGCTGTTTGGCTATATCAGCGTAAAAGCAGGCACGAATGCTGAACAAGTG  1976687

Query  193      CGTTCACTTCTCGCGCGTTCGCTGCCAAACTATATGATCCCGGCGTACATTATTGAAATG  252
                |||||  |||||||||||||||||||||||||||||||||||||||| || |||||||||
Sbjct  1976686  CGTTCTCTTCTCGCGCGTTCGCTGCCAAACTATATGATCCCGGCGTATATCATTGAAATG  1976627

Query  253      GAGACTCTTCCGCTCACATCAAACGGCAAATTGAACCGAAAAGCACTTCCTGAACCAAAT  312
                |||||||||||||||||||||||||||||||||||||||| |||||||| ||||||||| ||
Sbjct  1976626  GAGACTCTTCCGCTCACATCAAACGGCAAATTGAACCGGAAAGCACTGCCTGAACCAGAT  1976567

Query  313      TTCACATC-A--CAAACGTACGCTCC-GCCTCGAAATGACCTTGAAGACCAACTTGTGTT  368
                 || |||| |  ||||| |||  ||| ||| || || || |||||||| || || |  ||
Sbjct  1976566  GTCGCATCTAAACAAACATACA-TCCCGCCACGCAACGAACTTGAAGAGCAGCTGGCATT  1976508

Query  369      GATCTGGCAGGAGGTTCTCGGCATCCAGAGGATCGGAATAGAGGATTCGTTTTTTGAATT  428
                |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Sbjct  1976507  GATCTGGCAGGAGGTTCTAGGCATCCAGAGGATCGGAATAGAGGATTCGTTTTTTGAATT  1976448

Query  429      GGGCGG  434
                ||||||
Sbjct  1976447  GGGCGG  1976442
```

Figure 41: BLAST nucleotide alignment of the 21332 fengycin fragment with 168 plipastatin synthetase

```
 Features in this part of subject sequence:
   plipastatin synthetase

 Score =  811 bits (439),  Expect = 0.0
 Identities = 451/457 (99%), Gaps = 0/457 (0%)
 Strand=Plus/Minus

Query  56       CGATTGAATATATCGGCCGTCTTGATGATCAAGTGAAAATTAGGGGTTATCGTATTGAAC  115
                ||||||||||||| || ||| ||||||||||||||||||||||||||||||||||||||
Sbjct  1987696  CGATTGAATATATTGGACGTTTTGATGATCAAGTGAAAATTAGGGGTTATCGTATTGAAC  1987637

Query  116      TGAGAGAGATTGAAACAGTTCTTCGACAAGCACCGGGGGTAAAAGAAGCAGCGGTACTGG  175
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1987636  TGAGAGAGATTGAAACAGTTCTTCGACAAGCACCGGGGGTAAAAGAAGCAGCGGTACTGG  1987577

Query  176      CCCGTGATGTTTCTGCTGAGGAAAAGGAGCTCGTTGCTTATATCGTTCCGGAAAAGGGAA  235
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1987576  CCCGTGATGTTTCTGCTGAGGAAAAGGAGCTCGTTGCTTATATCGTTCCGGAAAAGGGAA  1987517

Query  236      ACAGCCTCCCAGATTTGTATCAGCATCTTGCCGGGACATTGCCGTCCTATATGATCCCAG  295
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1987516  ACAGCCTCCCAGATTTGTATCAGCATCTTGCCGGGACATTGCCGTCCTATATGATCCCAG  1987457

Query  296      CAAGTATTATCAACATCAGCCGGATGCCATTAACATCCAGCGGCAAGCTTGACAGGTTTG  355
                |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Sbjct  1987456  CAAGTATTATCAACATCAGCCAGATGCCATTAACATCCAGCGGCAAGCTTGACAGGTTTG  1987397

Query  356      CGCTTCCTGAACCAGAAAACAATACGAGTGTCACATATATGGCGCCCCGAACCTTAATCG  415
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1987396  CGCTTCCTGAACCAGAAAACAATACGAGTGTCACATATATGGCGCCCCGAACCTTAATCG  1987337

Query  416      AAGCTGATCTCGCACATATTTGGGAAGATGTGCTGAATAAACAGCACATCGGCATTCGTG  475
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1987336  AAGCTGATCTCGCACATATTTGGGAAGATGTGCTGAATAAACAGCACATCGGCATTCGTG  1987277

Query  476      ATGATTTCTTTCAGTTAGGCGGACATTCCCTAAAAGC  512
                |||||||||||||||||||||||| ||| |||||||
Sbjct  1987276  ATGATTTCTTTCAGTTAGGCGGACAGTCCTTAAAAGC  1987240
```

Figure 42: BLAST nucleotide alignment of the BBG21 fengycin fragment with 168 plipastatin synthetase

### 2.2. Analysis of fengycin or plipastatin operon organization of *B. subtilis* S499

Among the previous primer pairs tested which were designed with the published sequence of *B. subtilis* 168 plipastatin operon and tested with both *B. subtilis* BBG21 and 21332, most of them gave no results with *B. subtilis* S499. New series of degenerated primers designed by the alignment of three *Bacillus* plipastatin and fengycin operons from *B. subtilis* 168, *B. amyloliquefaciens* FZB42 and *B. subtilis* F29-3 were thus used to characterize the organisation of the *B. subtilis* S499 fengycin operon (*ppsA-ppsB* primers, *ppsB-ppsC* primers, *ppsC-ppsD* primers and *ppsD-ppsE* primers).

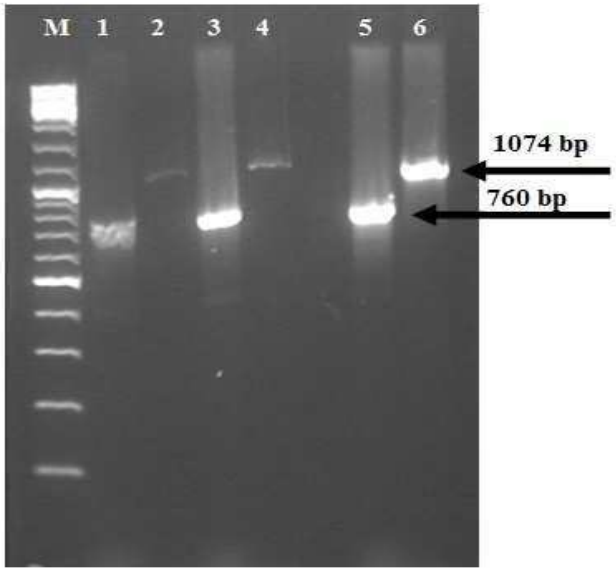



Figure 43: Amplicons generated by *ppsA-ppsB* primers lanes (1,3,5), *ppsC-ppsD* lanes (2,4,6) ; lanes (5,6) from *B. subtilis* 168 ; lanes (1,2,3,4) from *B. subtilis* S499, M = O'Gene Ruler standard

Figure 43 shows the PCR products of *B. subtilis* 168 and S499 with the *ppsA-ppsB* and *ppsC-ppsD* primers, which gave the same fragment size (760 and 1074 bp, respectively). After sequencing of these fragments, BLAST nucleotide alignment of the *ppsA-ppsB* fragment of *B. subtilis* S499 with *B. subtilis* 168 showed identity of 65.2%, while a high identity of 96% with *B. amyloliquefaciens* FZB42 was found. The alignment of *ppsC-ppsD* fragment of *B. subtilis* S499 with *B. subtilis* 168 showed identity of 72.8%, while the aligment with *B. amyloliquefaciens* FZB42 revealed a high identity of 97% (Table 25).

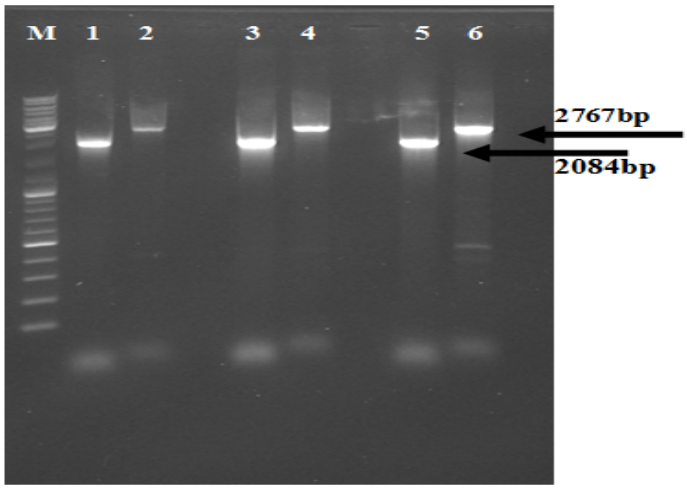



Figure 44: Amplicons generated by *ppsB-ppsC* primers lanes (1,3,5), *ppsD-ppsE* lanes (2,4,6) ; lanes (5,6) from *B. subtilis* 168 ; lanes (1,2,3,4) from *B. subtilis* S499, M = O'Gene Ruler standard

Figure 44 shows the PCR products of *B. subtilis* 168 and S499 with the *ppsB-ppsC* and *ppsD-ppsE* primers which gave the same fragment size (2084 and 2767 bp, respectively). After sequencing of these fragments, BLAST nucleotide alignment of the *ppsB-ppsC* fragment of *B. subtilis* S499 with *B. subtilis* 168 showed identity of 64.5%, while a high identity of 97% with *B. amyloliquefaciens* FZB42 was found. The alignment of *ppsD-ppsE* fragment of *B. subtilis* S499 with *B. subtilis* 168 showed identity of 69%, while the alignment with *B. amyloliquefaciens* FZB42 revealed a high identity of 97% (Table 25).

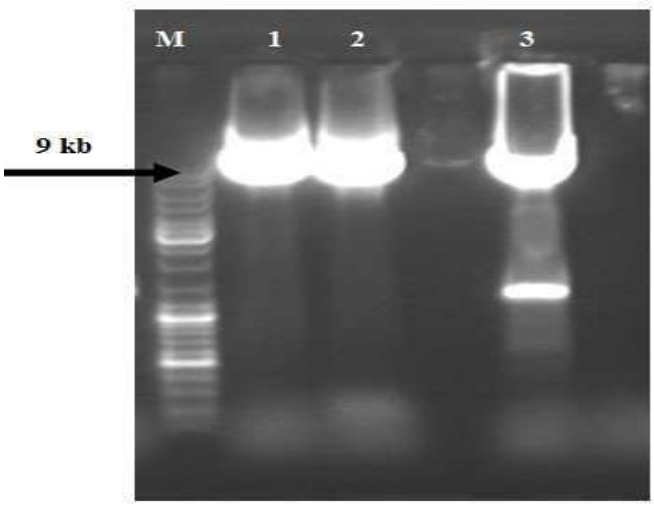



Figure 45: Amplicons generated with *ppsA-ppsC* primers ; lanes (1,2) from *B. subtilis* S499; lane (3) from *B. subtilis* 168, M = O'Gene Ruler standard

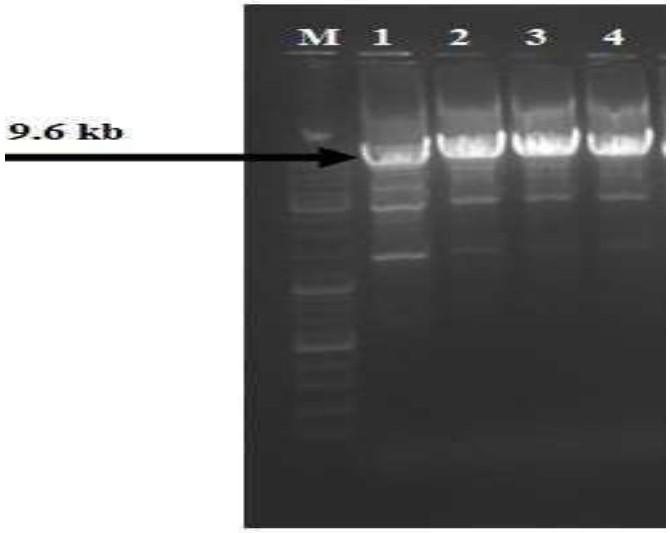



Figure 46: PCR gradient for *ppsB-ppsD* primers lane (1) from *B. subtilis* 168, lane (2-11) from *B. subtilis* S499, M = O'Gene Ruler standard

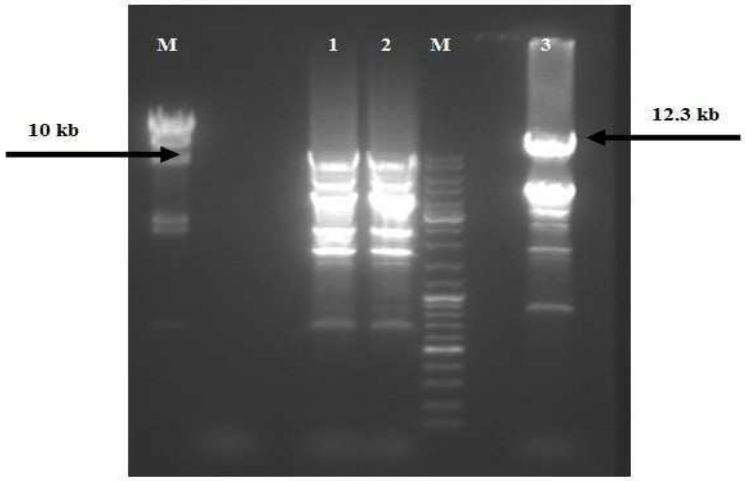



Figure 47: *ppsC-ppsE* primers lanes (1,2) from *B. subtilis* S499, lane (3) from *B. subtilis* 168, M = λ DNA + *Hin*dIII marker

Figure 45 shows the long PCR product of *B. subtilis* 168 and S499 with the *ppsA-ppsC* primers which gave the same fragment size (9 kb). Figure 46 shows the long PCR product of *B. subtilis* 168 and S499 with the *ppsB-ppsD* which gave a convergent fragment size (9.6 kb for *B. subtilis* 168) and about 10 kb for *B. subtilis* S499, while Figure 47 shows the PCR product of *B. subtilis* 168 and S499 with the *ppsC-ppsE* which gave fragment size (12.3 kb) with *B. subtilis* 168 while S499 showed a fragment (10 kb) smaller than that of *B. subtilis* 168. This difference between the length of this fragment between *B. subtilis* 168 and S499 could be referred to the absence of epimerization domain in position 9 from *ppsD* or *fenD* of *B. subtilis* S499, which could confirm the absence of a D-Tyr in position 9 from the

supernatant of *B. subtilis* S499 as reported by Schneider *et al*. (1999). However here, the presence of D-Tyr (E domain) couldn't be detected in position 3 (*ppsB* or *fenB*) because the difference of the fragment size in this region of S499 was not long enough than that of *B. subtilis* 168 to say the E domain is present in position 3. The absence of this domain was detected in position 9 (*ppsD* or *fenD*) but unfortunately, the large fragment detected for S499 couldn't be cloned or sequenced.

Table 25: Needle alignment of the different sequenced fragments detected for fengycin and plipastatin operons of *B. subtilis* S499, 21332 and BBG21

| Sequence Origins | Fengycin and Plipastatin sequenced fragments alignment | DNA identity (%) |
|---|---|---|
| *B. subtilis* S499-<br>*B. amyloliquefaciens*<br>FZB42 | *ppsA-ppsB*<br>*ppsB-ppsC*<br>*ppsC-ppsD*<br>*ppsD-ppsE* | 96<br>97<br>97<br>97 |
| *B. subtilis* S499-<br>*B. subtilis* 168 | *ppsA-ppsB*<br>*ppsB-ppsC*<br>*ppsC-ppsD*<br>*ppsD-ppsE* | 65.2<br>64.5<br>72.8<br>69 |
| *B. subtilis* 21332-<br>*B. subtilis* 168 | *dacC-ppsA*<br>*ppsA-ppsB*<br>*ppsB*<br>*ppsC-ppsD*<br>*ppsD E* domain<br>*ppsE-yngL* | 99<br>97.7<br>100<br>97.7<br>96.7<br>99.4 |
| *B. subtilis* BBG21-<br>*B. subtilis* 168 | *ppsA-ppsB*<br>*ppsC-ppsD*<br>*ppsD E* domain | 98.1<br>97.6<br>95.5 |

## 2.3. Conclusion

The DNA identity observed between sequences of *Bacillus* S499 and *B. amyloliquefaciens* was more important than that *Bacillus* S499 and *B. subtilis* 168. This result showed that S499 strain could belong to the *amyloliquefaciens* species, which was recently confirmed by researchers of Gembloux Agro-Biotech (Marc Ongena, personal communication). The study of the organization of fengycin/plipastatin operons in *B. subtilis* BBG21 and 21332 is probably identical to this of *Bacillus subtilis* 168. For *Bacillus subtilis* S499, several differences were observed and the absence of some sequences did not allow to conclude about the definite organisation of this operon. The S499 genome sequencing which has started now should give the answer to the question: "fengycin and plipastatin, who is who?"

# 3. Plipastatin overproduction in *B. subtilis* 168 derivatives

The next objective of our work was to study the effect of the replacement of plipastatin operon promoter of two *B. subtilis* 168 derivatives named BBG111 and BBG134. In a previous work, introduction of an active *sfp* gene from *B. subtilis* 21332 to *B. subtilis* 168 gave rise to strain BBG111 which has an active *sfp* gene required to promote lipopeptide production. This derivative is co-producer for both surfactin and plipastatin (Coutte *et al.*, 2010). In addition, the interruption of surfactin operon in BBG111 gave the strain BBG134 a plipastatin mono-producer strain (Deravel, ProBioGEM). These two derivatives were used in attempts to replace their plipastatin promoter by a constitutive one $P_{repU}$.

## 3.1. The different constructions used to replace plipastatin native promoter of *B. subtilis* 168 derivatives

### 3.1.1. Plasmid construction 1

In a previous work, $P_{repU}$, a promoter originating from *Staphylococcus aureus*, was inserted in the plasmid named pBG106 (Leclère *et al.*, 2005).

*dacC* Cassette1 (621 bp) and *ppsA* Cassette2 (617 bp) from *B. subtilis* 168 were cloned in pGEM-T Easy vector. The resulting plasmids were named pBG186 and pBG177 respectively (see chapter IV, Materials and Methods, 2.1.11.1. Plasmids construction**).**

pBG106 was first used to achieve the plipastatin overproduction in *B. subtilis* BBG134 and BBG111 derivatives. The expected construction is shown below in Figure 48.

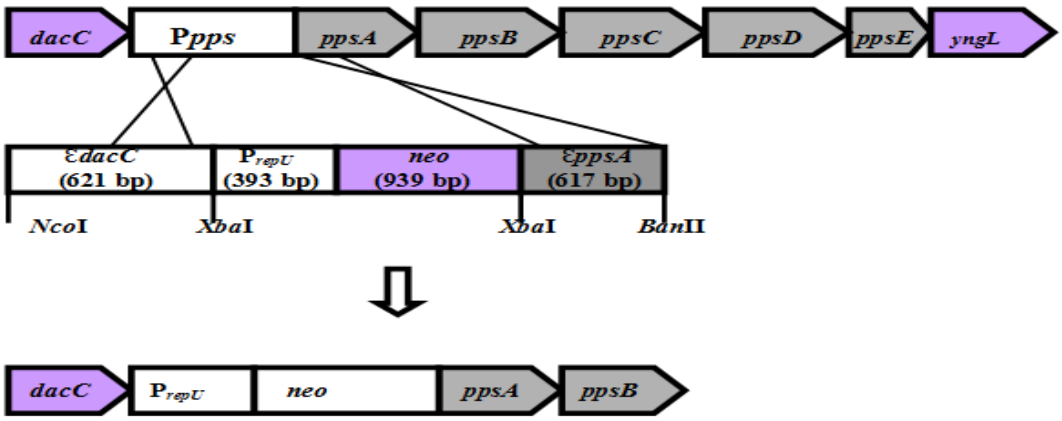



Figure 48: Expected construction 1; *pps* promoter replacement by the $P_{repU}$-*neo* cassette.

*dacC* cassette1 (from pBG186) was then inserted in pBG106 to obtain pBG189. Unfortunately, pBG189 was difficult to digest and a new strategy was elaborated to obtain the final plasmid.

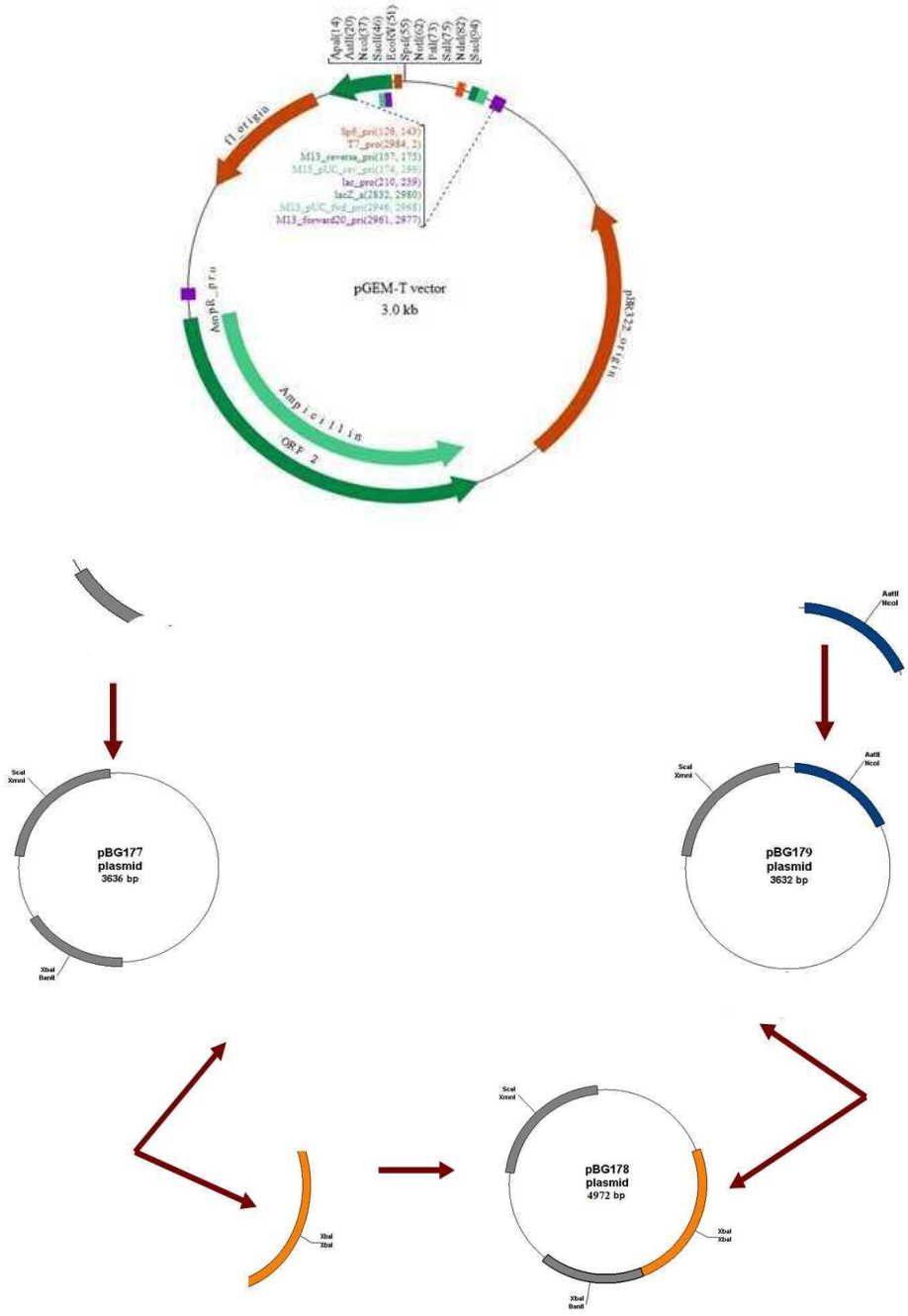



Figure 49: Different construction steps of pBG180 (P*repU- neo* cassette in sense direction) and pBG180*(P*repU- neo* cassette in antisense direction) plasmids

### 3.1.2. Plasmid construction 2

This construction was performed using pGEM-T Easy vector. A new fragment *dacC* Cassette1, which have two artificial sites *Aat*II and *Nco*I was first cloned in pGEM-T Easy vector between the *Eco*RI sites of the vector. The ligation mixture was transformed into *E. coli* JM109 cells. The resulting plasmid was named pBG178.

Both pBG177 (previously prepared in construction 1) and pBG178 were *Aat*II and *Nco*I double digested, the *dacC* Cassette1 fragment was then inserted between the *Aat*II and *Nco*I sites of pBG178 to obtain pBG179.

Both pBG106 and pBG179, which have two artificial sites of *Xba*I and *Ban*II were *Xba*I digested. The $P_{repU}$-neo cassette was released from pBG106 and inserted between *Xba*I sites of pBG179 to obtain pBG180. The size of thenew plasmid pBG180 was 5559 bp (Figure 49).

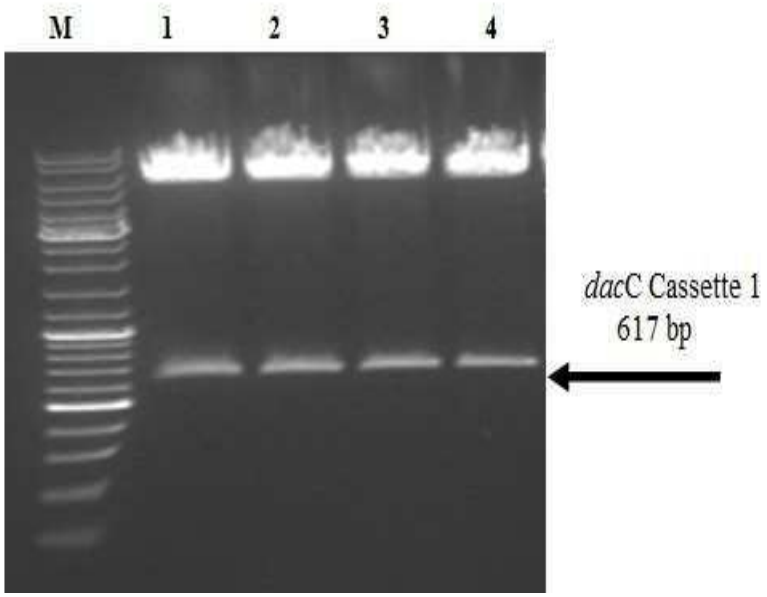



Figure 50: pBG180 *Aat*II + *Nco*I double digestion to verify *dac*C insertion; M = O' Gene Ruler standard, 1, 2, 3, 4 = *dacC* Cassette 1

Ten colonies were obtained, after the inoculation of two colonies with pBG180 plasmid in LB liquid with neomycin; both sequencing and the transformation were performed at the same time for the two plasmids. One of the two plasmids was successfully transformed into *B. subtilis* BBG111 and gave the derivative BBG140, while no transformants were obtained for the *B. subtilis* BBG134. The sequencing results were then obtained and surprisingly, it showed that the plasmid transformed in BBG111 (BBG140) harbour the

P*repU*$_{neo}$ cassette in the antisense direction, while the other plasmid harbour the P*repU*$_{neo}$ cassette in the sense direction.

These two types of final plasmid were obtained as the P$_{repU}$-*neo* cassette was digested with *Xba*I enzyme from the two sides, the type (1) was named pBG180 and harbour P$_{repU}$-*neo* cassette in sense direction, while type (2) plasmid was named pBG180* and harbour P*repU*-*neo* cassette in antisense direction (inverted).

The ten colonies obtained were inoculated in LB liquid with neomycin. The plasmids were purified and tested by PCR to verify the construction with *dacC-pps* sense and *ppsA* cassette 2 antisense primers and with P$_{repU}$-*neo* fwd2 + *ppsA* cassette 2 antisense primers (to detect the sense of P$_{repU}$-*neo* cassette).

According to that, six of the ten tested plasmids were found to harbour P$_{repU}$-neo cassette in sense direction, while four plasmids were found to have P$_{repU}$-*neo* cassette in antisense direction. Type (1) plasmid was used to transform by the natural competence method both of *B. subtilis* BBG134 and BBG111.

The type (1) plasmid couldn't be introduced in *B. subtilis* BBG111 and BBG134 after five times by the natural competence method and four times by electroporation. Every time many colonies were observed which were resistant to neomycin. Many of them were tested by PCR and for lipopeptide production. Neither positive specific PCR amplification nor significant plipastatin production were observed.

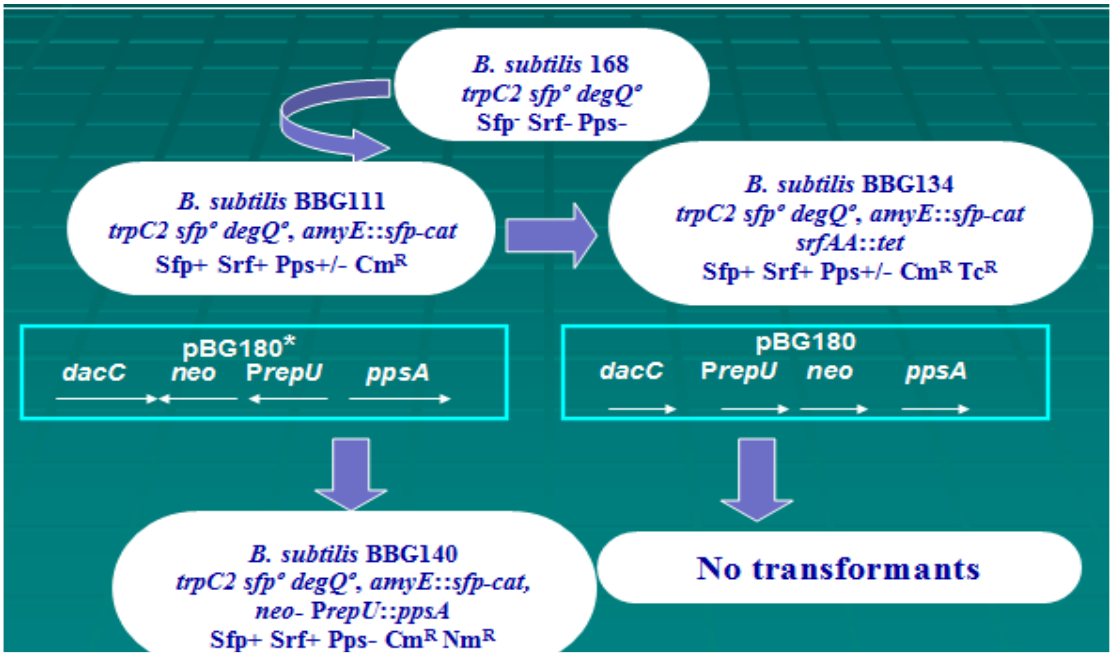



Figure 51: *B. subtilis* 168 derivatives constructions; BBG111 (Coutte *et al*. 2010), BBG134 (Deravel *et al*. 2011) and BBG140 (this study) by pBG180* plasmid transformation

### 3.1.2.1. Verification of the new constructed *B. subtilis* BBG140

While the obtained construction was not the requested one, it was genetically analysed and the phenotype of the resulting mutant was analysed. The verification of the new strain BBG140 construction was performed by PCR using two primers located up and downstream from the inserted cassette *dacC-pps* sense and *ppsA* Cassette 2 antisense. The PCR generated large fragment of about 2.5 kb for the new *B. subtilis* BBG140 strain, compared to 1.2 kb for the mother strain *B. subtilis* 168. The fragment was sent to sequencing and the result showed that the insertion derivative BBG140 had the $P_{repU}$-*neo* cassette inverted and the other two cassettes were identical (100 %) to that of the mother strain *B. subtilis* 168.

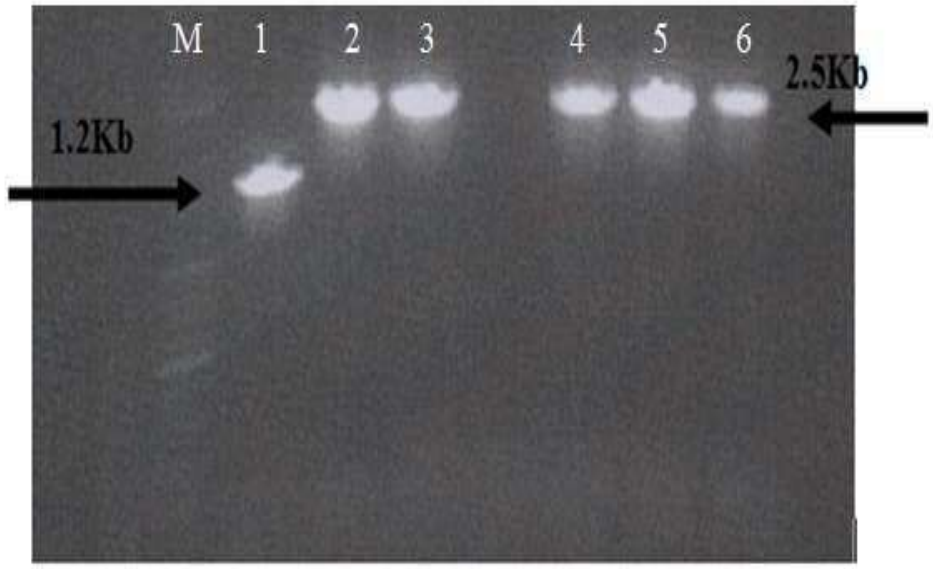



Figure 52: Verification of pBG180 insertion in *B. subtilis* BBG140; M = O' Gene Ruler standard, 1 = *B. subtilis* 168, 2 to 6 = BBG140

### 3.1.2.3. Plipastatin operon expression of BBG140 strain

Expression of the plipastatin operon was checked for both BBG140 and its mother strain BBG111. It was measured during the growth in Landy MOPS medium at pH 7.0 in 500 mL flask at 37 ºC and under an agitation rate of 200 rpm.

Three samples were withdrawn during the log phase after 4 h, 10 h and 24 h of growth. The transcriptome was blocked by the *RNAlater* solution and the RNA was extracted and stored at -20 ºC (see chapter IV, material and methods, 2.1.10.3). The day after, the total RNA was inverse transcribed to cDNA by RevertAid First Strand kit from Fermentas. The cDNA concentrations were measured using Biowave II, WPA spectrophotometer.

As shown in Figure 53, the plipastatin operon of BBG111 has been amplified. The fragment of 760 bp indicates the expression of the plipastatin operon in this strain. No amplicon was observed for BBG140 with the *ppsA* sense and anti-sense primers which were affected with the cDNA to detect the expression of the the plipastatin operon. The absence of plipastatin operon expression in the derivative BBG140 is thus a result of integration of the inverted $P_{repu}$-*neo* cassette.

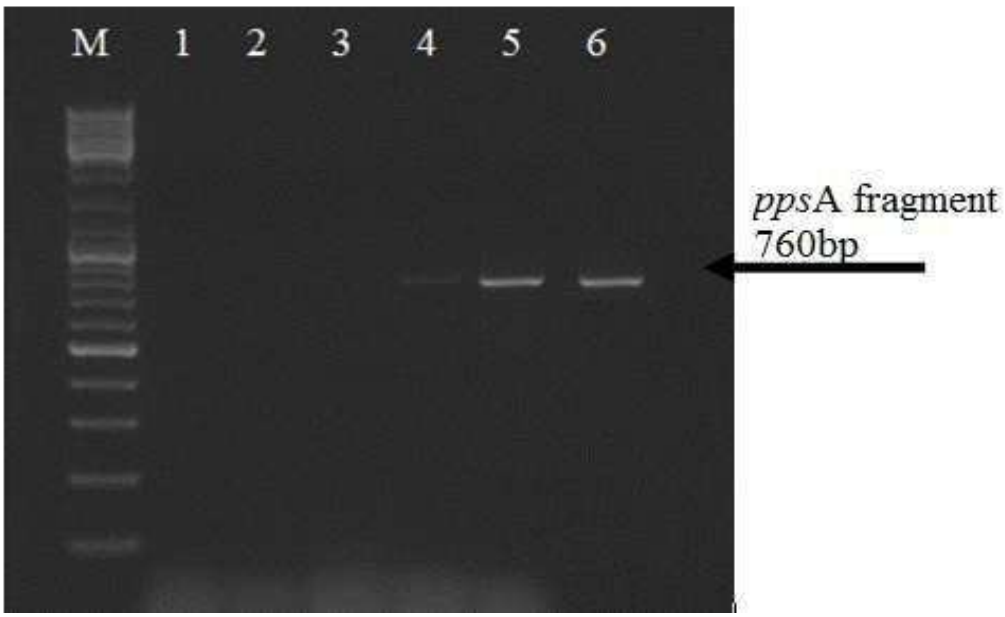



Figure 53: *ppsA* primers for the cDNA samples of BBG140 and BBG111, M = O' Gene Ruler standard, 1 = 4 h sample of BBG140, 2 = 10 h sample of BBG140, 3 = 24 h sample of BBG140, 4 = 4 h sample of BBG111, 5 = 10 h sample of BBG111, 6 = 24 h sample of BBG111

### 3.1.2.4. Lipopeptide production by the new strain *B. subtilis* BBG140

Cultures were performed with the strain BBG140 and its mother strain BBG111 to verify the interruption of plipastatin operon by the absence of plipastatin production. A lot of studies have pointed out different environmental factors for their effect on lipopeptide production and several experiments performed in our laboratory have shown that this effect can be strain-dependent. Carbon and nitrogen sources as well as mineral requirements could drastically affect the synthesis (Landy *et al*., 1948; Cooper *et al*., 1981; Jacques *et al*. 1999; Guez *et al*., 2007; Wei *et al*., 2007). To take into account these factors, two set of different cultures conditions (series A and B) were realized.

### 3.1.2.2.1. Series A

Three media were analysed which differ by carbon and nitrogen sources. The experiments were carried out at 30°C, in 1 L flasks, 10% filling volume under 160 rpm. Growth of the culture, pH of media and lipopeptide production were followed during 140 h as explained in Materials and Methods. At least two replicates were realized and the values shown in Figures 54, 55 and 56 are averages.

#### a) Modified Landy medium

As shown in Figure 54, growth of the two strains remains low in the medium. BBG111 synthesized a small amount of plipastatin at 24 h (9.57 mg/L) and no production was observed in BBG 140. At 72 h, the specific productions (in mg/L/U.D.O) for surfactin are 47 and 290 for BBG 111 and BBG 140 respectively.

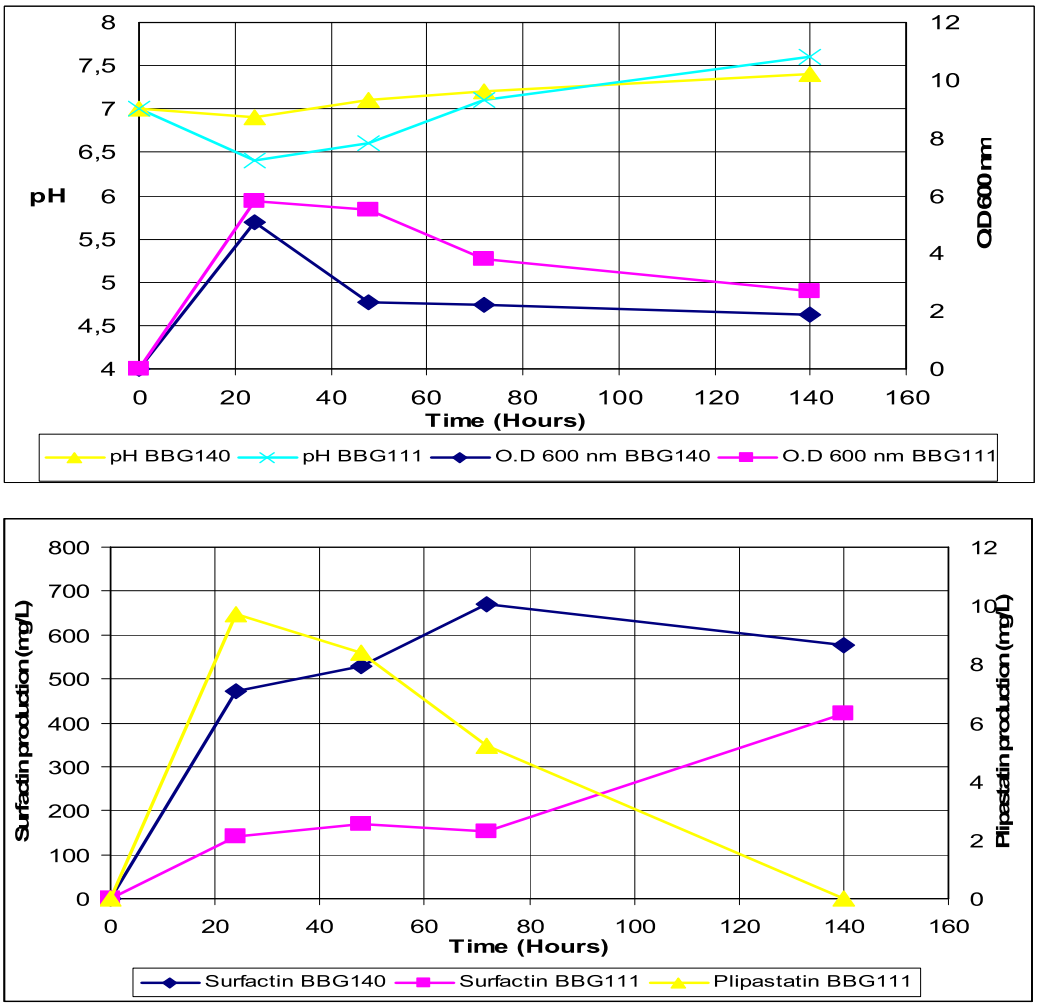



Figure 54: Microbial growth at 600nm, pH, surfactin production for BBG140, surfactin and plipastatin production for BBG111 in modified Landy medium

**b) Modified Landy 3 medium**

As shown in Figure 55, the plipastatin production for BBG111 gave a small level of 6.9 mg/L at 24 h (9.6 $O.D_{600nm}$ and pH 5.4) which decreased to reach 3.6 mg/L at 144 h. No production could be detected for the strain BBG140. In this medium, BBG140 acidified less than BBG111. The decrease in pH for BBG111 after 24h could be the reason for the low detection of surfactin, due to precipitation. In this medium, at 72 h, the specific productions (in mg/L/U.D.O) for surfactin were 0 and 16 for BBG111 and BBG140, respectively.

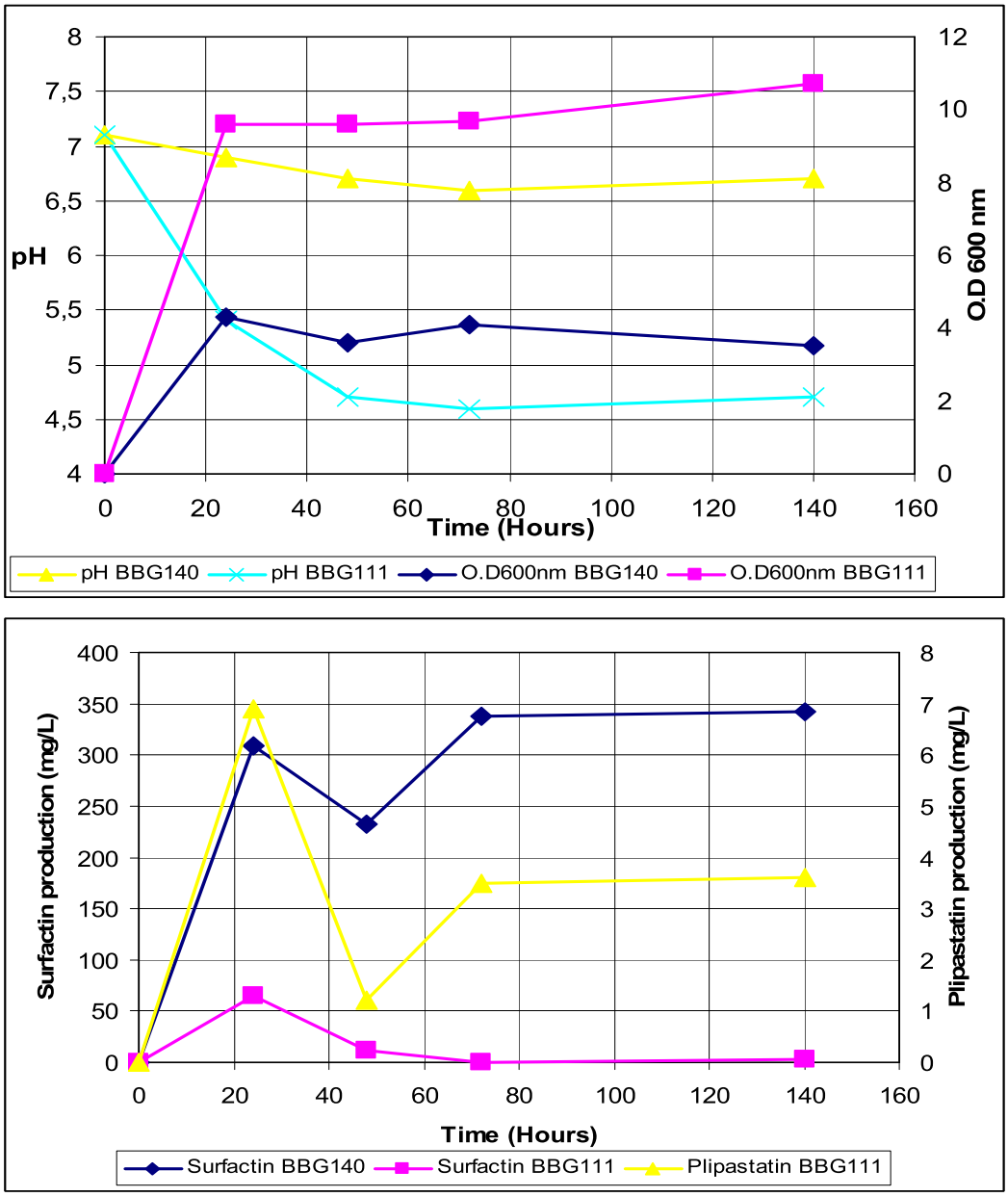



Figure 55: Microbial growth at 600 nm, pH, surfactin production for BBG140, surfactin and plipastatin production for BBG111 in modified Landy 3 medium

### c) Optimalized medium

Contrary to modified Landy medium and modified Landy 3 medium the optimalized medium promoted an important growth of the two strains after 72 h of cultivation (Figure 56). As in the previous experiments, a low amount of plipastatin could be observed for BBG111 whereas no production was observed for BBG 140. In this medium, at 72 h, the specific productions (in mg/L/U.D.O) for surfactin were 81 and 216 for BBG 111 and BBG 140, respectively.

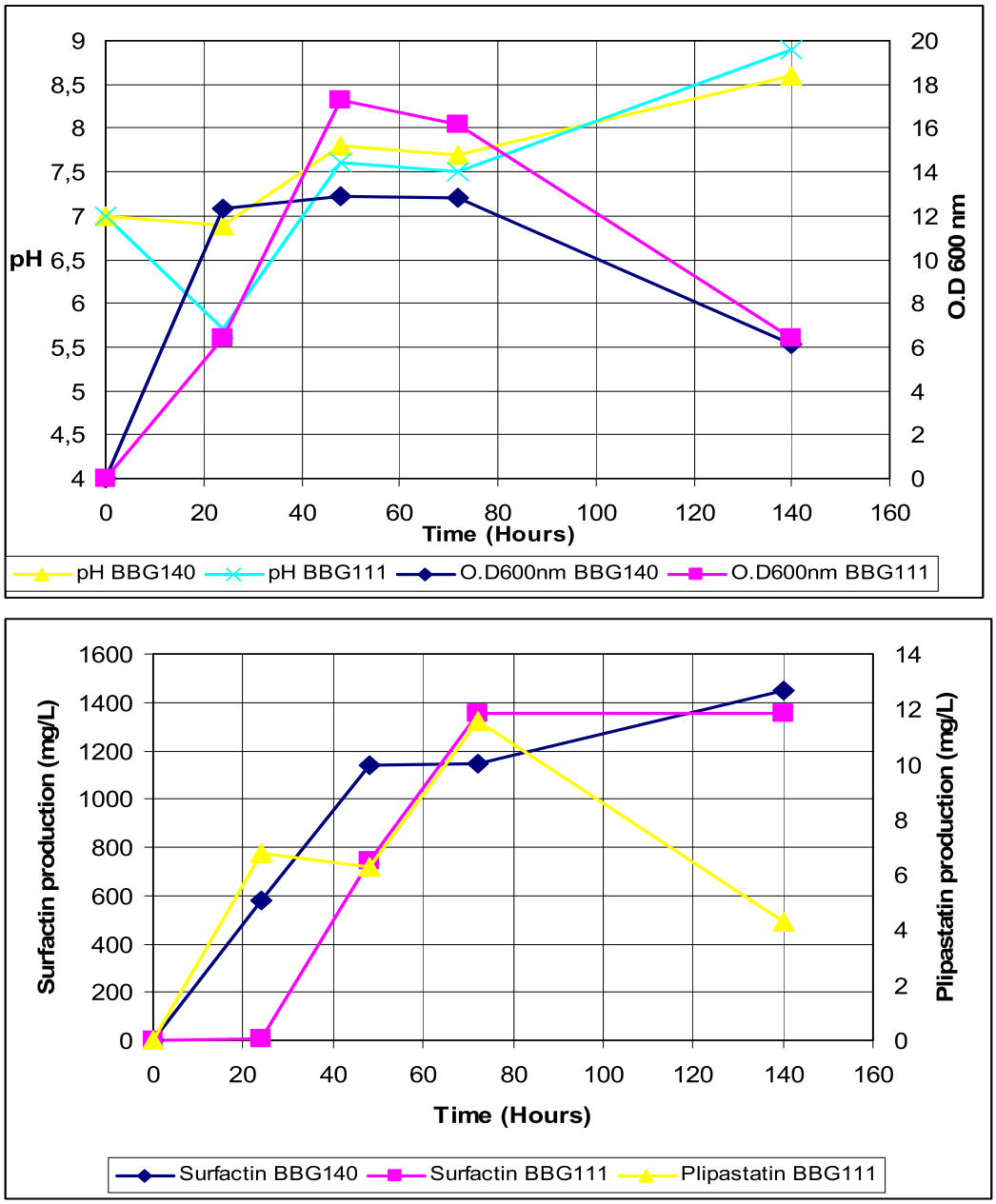



Figure 56: Microbial growth at 600 nm, pH, surfactin production for BBG140, surfactin and plipastatin production for BBG111 in optimalized medium

In conclusion, no production of plipastatin was observed in BBG140 whatever the medium. BBG111 and BBG140 showed different regulation of surfactin synthesis, as the specific production is always higher in BGG140.

**3.1.2.2.2. Series B**

These series of experiments were performed in Landy MOPS medium with addition of organic acids (fumaric acid and pyruvic acid) or by replacement of glutamic acid with proline and tyrosine. Productions in Landy modified 3 and optimalized medium were used as control. The experiments were carried out at 30°C, in 1 L flasks, 10% filling volume under 160 rpm. Lipopeptide productions were quantified by HPLC after 72 h of fermentation. At least two replicates were realized and the values shown in Figure 57 are averages.

As expected for BBG140, no plipastatin production was observed. In Landy MOPS, BBG111 gave a small production of plipastatin (9.5 mg/L) and in Landy modified 3 a more important synthesis (55 mg/L).

The highest plipastatin production was observed in Landy MOPS with fumaric acid (62 mg/L). Landy MOPS with tyrosine gave 25 mg/L, optimalized medium, 12 mg/L and finally Landy with pyruvic acid gave no plipastatin production.

The highest surfactin production was observed in optimalized medium (1450 mg/L for BBG140 and 1360 mg/L for BBG111) followed by Landy MOPS with tyrosine (860 mg/L for BBG140 and 380 mg/L for BBG111), Landy MOPS with proline (700 mg/L for BBG140 and 300 mg/L for BBG111), Landy MOPS with fumaric acid (700 mg/L for BBG140 and 380 mg/L for BBG111), Landy MOPS (680 mg/L for BBG140 and 410 mg/L for BBG111), Landy modified 3 (580 mg/L for BBG140 and 200 mg/L for BBG111) and finally Landy MOPS with pyruvic acid (370 mg/L for BBG140 and 290 mg/L for BBG111).

The overall effect of all these conditions was on the surfactin production which was observed to be higher in BBG140 than in BBG111. Interestingly, these results revealed that the non-expression of the plipastatin operon in BBG140 affects positively the production of surfactin.

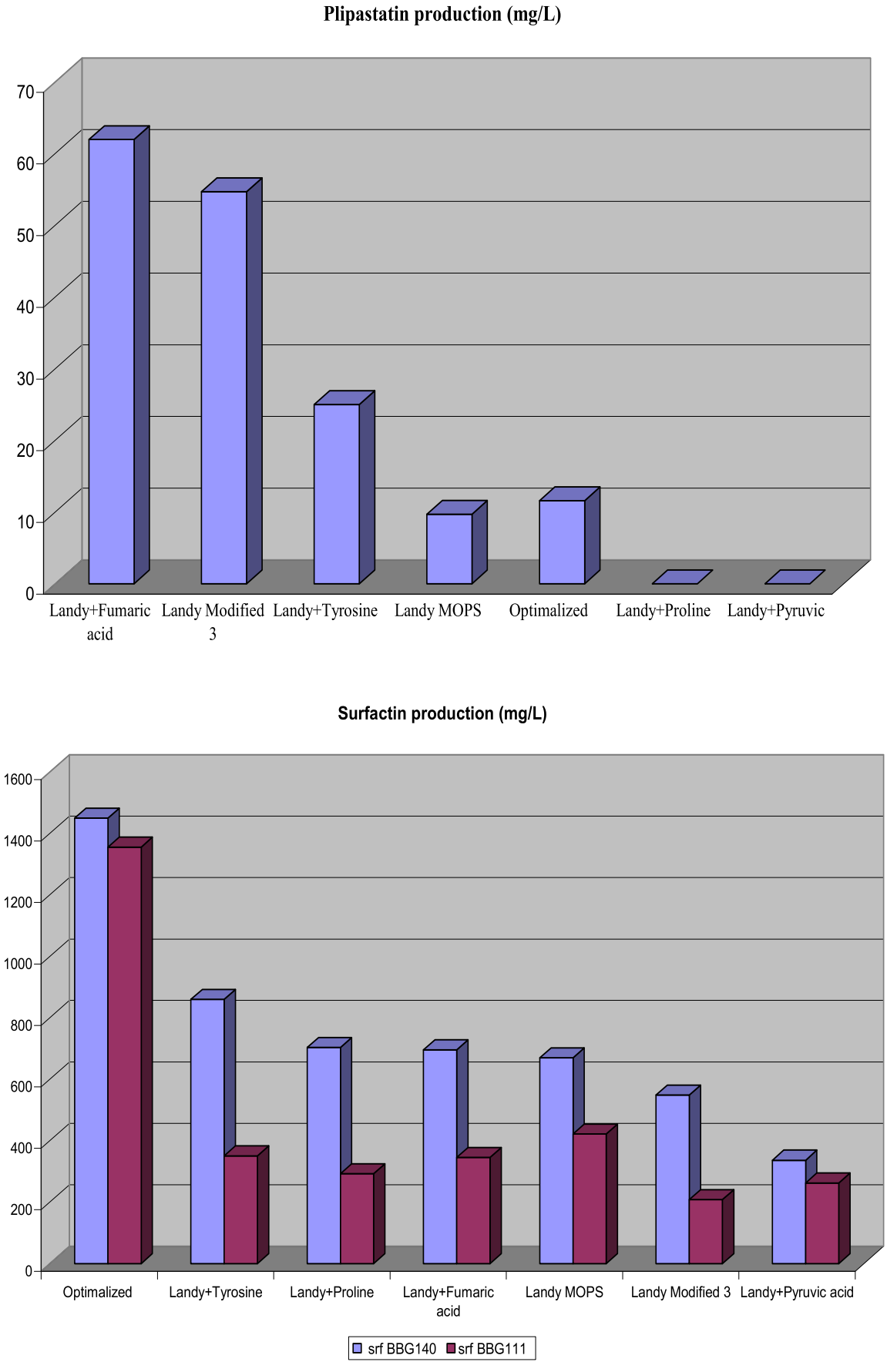



Figure 57: Plipastatin and surfactin productions (mg/L) for BBG140 and BBG111 in different culture media

### 3.1.3. Plasmid construction 3

The constructed plasmid pBG180 was used to design the construction 3. The expected construction is shown in Figure 58.

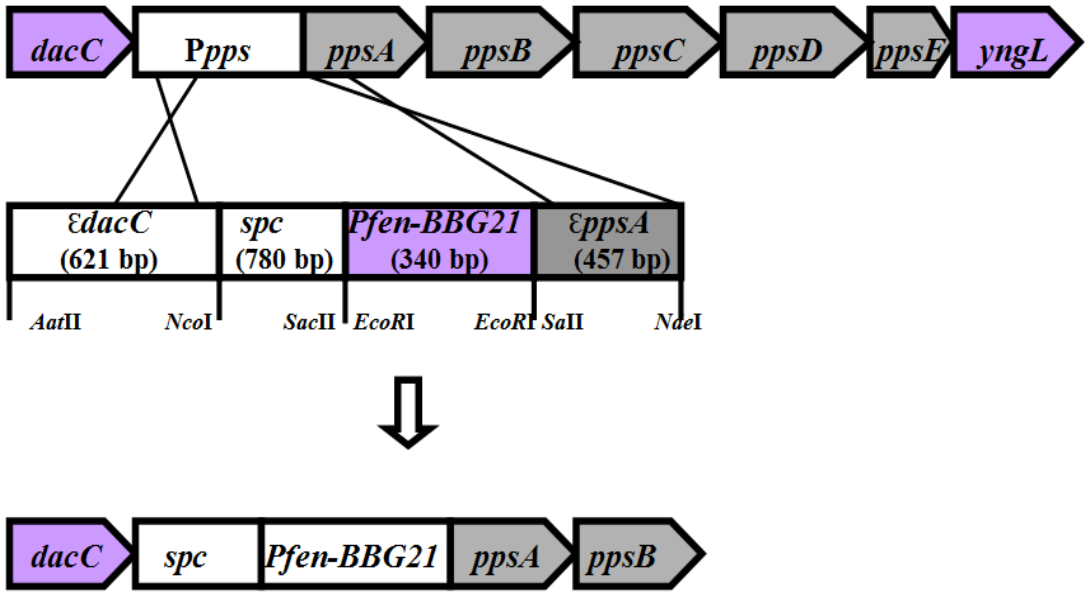



Figure 58: Expected construction 3; *pps*/*fen* promoter replacement by *spc*-P*fen*-BBG21 cassette

Firstly, the P*fen*-BBG21 and *ppsA*-Cassette2N fragments were inserted in pGEM-T Easy vector between the *Eco*RI sites of the vector. The ligation mixtures were then transformed into *E. coli* JM109 cells. The obtained plasmids were named pBG185 and pBG183, respectively. Both pBG180 and pBG183 were *Sa1*I and *Nde*I double digested, and then *ppsA*-Cassette 2N was inserted between *Sa1*I and *Nde*I sites of pBG180 to obtain pBG191. Both pBG191 and pBG214 (bearing a spectinomycin resistance gene) were *Nco*I and *Sac*II double digested, the *spc* gene then released and inserted between the *Nco*I and *Sac*II sites of pBG191 to obtain pBG192. Both pBG185 and pBG192 were *EcoR*I digested, and then the previous fragments $P_{repU}$-*neo* cassette and *ppsA* Cassette 2 were released (Figure 59).

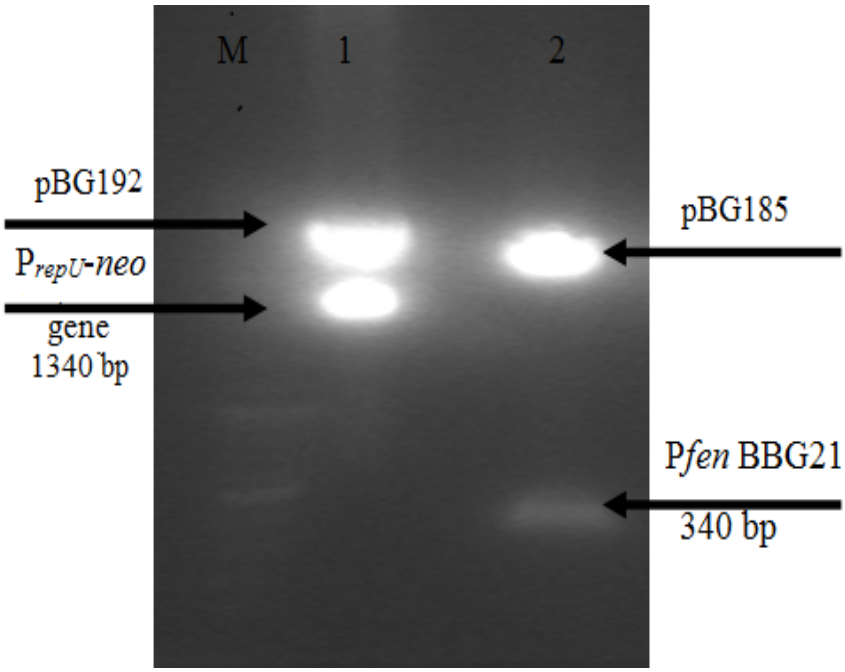



Figure 59: *Eco*RI digestion for both pBG191 and pBG185 plasmids

The new promoter fragment P*fen*-BBG21 was then inserted between the *EcoR*I sites of pBG192 to obtain pBG193.

Only five colonies were observed on LB plates with spectinomycin after the transformation of pBG193 into *E. coli*. The plasmid was purified and sent to sequencing but sequencing failed. To date, transformation of *B. subtilis* BBG111 and BBG134 with this new plasmid pBG193 did not succeed.

### 3.1.4. Conclusion

Achievement of plipastatin mono-producer or plipastatin + surfactin co-producer derivatives (with high amounts of plipastatin) by replacement of native promoter by the constitutive one $p_{repU}$ faced several difficulties. Two conclusions arose:

(1) The first difficulty was to place the plipastatin operon under the control of a constitutive promoter in spite of the previous obtention of a *B. subtilis* 168 derivative called *B. subtilis* 2504 (Ongena *et al*., 2007). *B. subtilis* 2504 was constructed by the replacement of the native promoter by the $P_{amyQ}$, whereas we worked with $P_{repU}$ promoter. Difference between $P_{repU}$ and $P_{amyQ}$ or incompatibility of the $P_{repU}$ with the plipastatin operon could explain the difficulty.

**(2)** The construction of a non-producing plipastatin strain (obtained by the insertion of the P*repU-neo* cassette in the antisense direction) led to an increase of the surfactin production.

# 4. Plipastatin and fengycin promoter region expression of *B. subtilis* 168 derivatives

## Introduction

In this last part of the work, we wanted to determine in which environmental conditions the genes involved in the biosynthesis of fengycin/plipastatin are expressed. We first compared the level of fengycin production of several mutant strains obtained in ProbioGEM. Then, we constructed new mutant strains with *lacZ* ORF fused to their fengycin/plipastatin operon promoter. We used this reporter gene to compare the expression ofoperons in different culture conditions. We were especially interested in expression of *B. subtilis* BBG21 a spontaneous mutant of *B. subtilis* 21332 found to produce high concentrations of fengycin, in non well identified conditions.

### 4.1. Plipastatin and fengycin productions by the different *B. subtilis* strains

The production of fengycin or plipastatin was first analysed in four strains of ProBioGEM collection : *B. subtilis* 168, *B. subtilis* BBG111, a *sfp*$^+$ mutant derivatives of 168 strain (Coutte *et al*., 2010), *B. subtilis* BBG134, a srf$^-$ mutant of *B. subtilis* BBG111 and *B. subtilis* BBG21. These strains were cultivated in Landy MOPS medium (pH 7.0, 160 rpm, in 1 L flask and 10 % filling volume). The lipopeptide production was measured after 72h as shown in Figure 60.

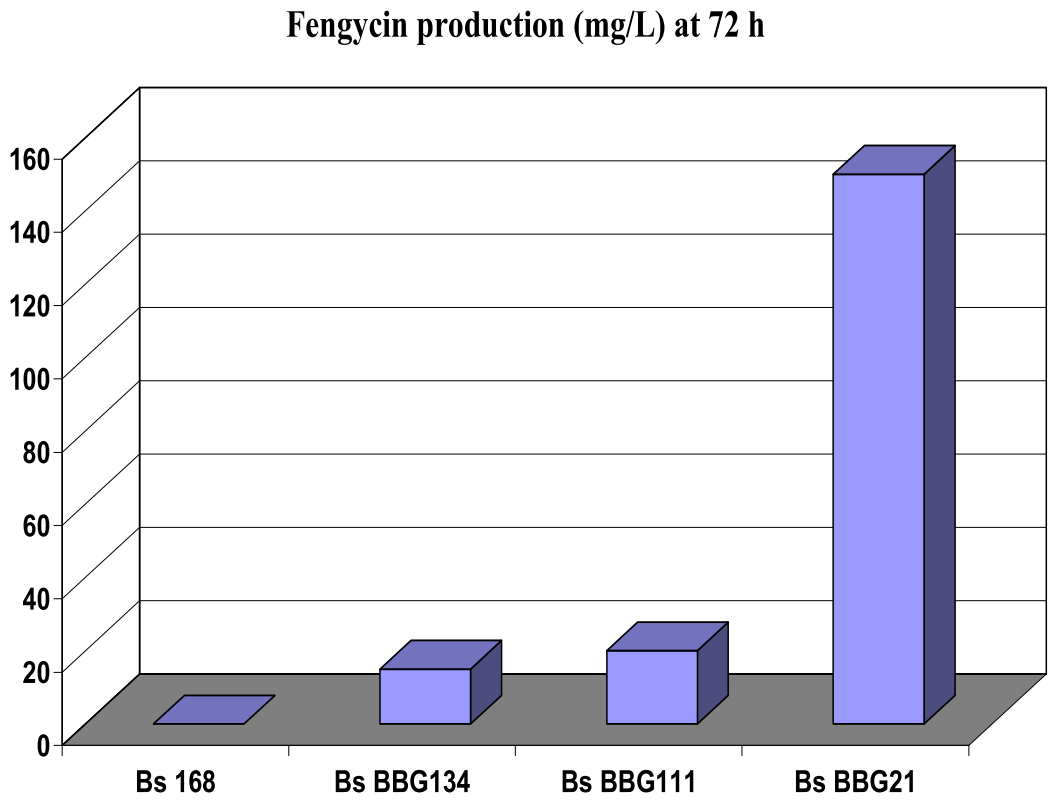



Figure 60: Fengycin or plipastatin production after 72 h of fermentation in Landy MOPS for *B. subtilis* 168, BBG134, BBG111 and BBG21

*B. subtilis* 168 gave no lipopeptide production, while *B. subtilis* BBG 111 and BBG 134 revealed a poor production (20 mg/L and 15 mg/L respectively). In contrast, *B. subtilis* BBG21 ~~which~~ produced about 150 mg/L of fengycin. These results confirm the higher level of production of *Bacillus subtilis* BBG21.

### 4.2. Expression analyses of plipastatin/fengycin promoters by fusion to *lacZ* gene

Plipastatin and fengycin operons regulation in *B. subtilis* 168 and BBG21 was studied by analyzing the strength of native promoters using a gene reporter system which constitutes an extensive toolbox for the study of promoter regulation. In this study, we utilized *lacZ* reporter system. The $P_{pps}$ promoter of *B. subtilis* 168 and $P_{fen}$ promoter of *B. subtilis* BBG21 were fused to the *lacZ* gene from pDG1661 plasmid. The resulting plasmids were used to transform *B. subtilis* 168, leading to strains BBG142 and BBG139, respectively (see chapter IV of material and methods, paragraph 2.1.12.1). In these strains, β-galactosidase activities (expressed in Miller units) were used to study the expression levels of the lipopeptides fengycin and plipastatin synthetases operons

### 4.2.1. *B. subtilis* 168 and BBG21 plipastatin and fengycin promoter sequence comparison

Both plipastatin and fengycin promoters from *B. subtilis* 168 and BBG 21 strains were sequenced after PCR amplification. The primers were designed according to the published sequence of *B. subtilis* 168 plipastatin operon. The product size was 340 bp from total promoter region 382 bp. The sequencing and alignment results revealed 99.1% of identity.

**$P_{pps}$168 and $P_{Fen}$ BBG21 alignment**

```
Length: 340
Identity:     337/340 (99.1%)
Similarity:   337/340 (99.1%)
Gaps:           1/340 (0.3%)
Score: 1667.0
=================================

EMBOSS_001         1 TAATAGCGGGAGACCTGAATTCCAAGAGAAAACAGGTTTTTTTATGTCTG     50
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001         1 TAATAGCGGGAGACCTGAATTCCAAGAGAAAACAGGTTTTTTTATGTCTG     50

EMBOSS_001        51 TAATAACGCTTTGTCGAACCTTCCAACAATTAGAAGAATGATGCACAAAA    100
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001        51 TAATAACGCTTTGTCGAACCTTCCAACAATTAGAAGAATGATGCACAAAA    100

EMBOSS_001       101 GAAAATTAGTCCGATTAAAAAGTAGTGTATAAGACCTTATGTAAAACAAT    150
                     ||||||||||.||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       101 GAAAATTAGTTCGATTAAAAAGTAGTGTATAAGACCTTATGTAAAACAAT    150

EMBOSS_001       151 ACTTTTATCACTTTATATCCGGAAATTTGATAAAAAAAGAAACGAAAAAT    200
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       151 ACTTTTATCACTTTATATCCGGAAATTTGATAAAAAAAGAAACGAAAAAT    200

EMBOSS_001       201 AAAGAAATAATACTATATTCAAAAAAAATGTTATAAAAATTTAATGTTTT    250
                     |||||||||||||||||| ||||||||||||||.||||||||||||||||
EMBOSS_001       201 AAAGAAATAATACTATATTC-AAAAAAATGTTATAAGAATTTAATGTTTT    249

EMBOSS_001       251 TTATTGTACTTTAAAAAATATCCGTTAATATAGTGCATATATGGATTATA    300
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       250 TTATTGTACTTTAAAAAATATCCGTTAATATAGTGCATATATGGATTATA    299

EMBOSS_001       301 TAGTCATATAATTTCTTTTATTGTCATTTTTAAGCTTTTT    340
                     |||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       300 TAGTCATATAATTTCTTTTATTGTCATTTTTAAGCTTTTT    339
```

Figure 61: Needle alignment of $P_{fen}$ BBG21 sequenced promoter with $P_{pps}$ 168, three bases differences (in red colour)

### 4.2.2. Verification of *B. subtilis* BBG142 and BBG139 strains construction

To verify that the promoter was inserted in the appropriate sense within the plasmid pDG1661, *Sph*I and *EcoR*I double digestion was performed (Figures 62 and 63) to release the *cat* gene (1280 bp). If the promoter was in the anti-sense direction, a fragment of 1620 bp (*cat* gene + $P_{pps}$ 168) and (*cat* gene + $P_{fen}$ BBG21) would have been obtained. As shown in figures 62 and 63, the constructions of both plasmids are in the right direction.

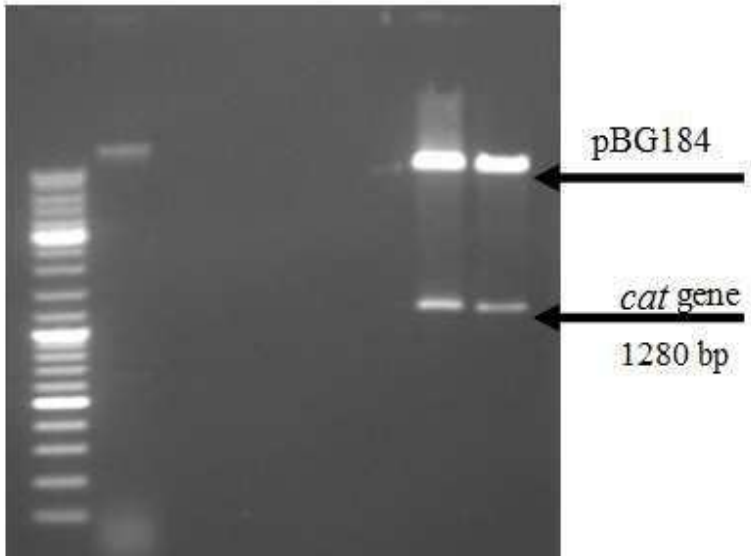



Figure 62: *Sph*I + *EcoR*I double digestion of pBG184

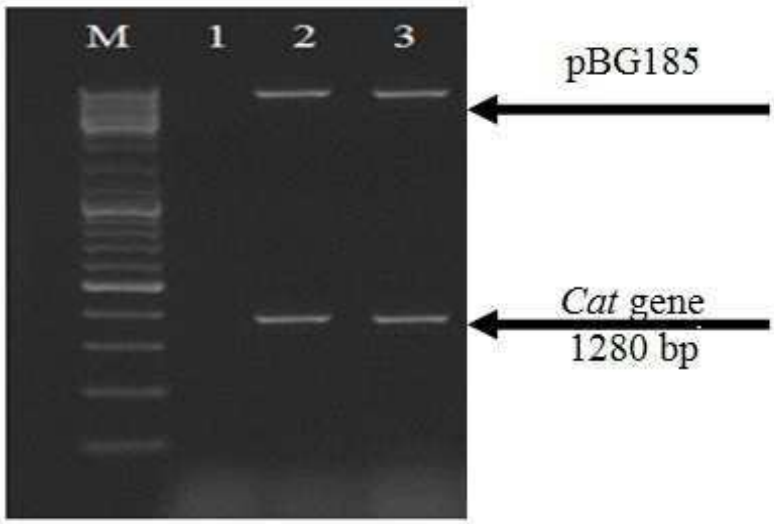



Figure 63: *Sph*I + *EcoR*I double digestion for pBG185

The new plasmids pBG184 and pBG185 were used to transform *B. subtilis* 168 by the natural competence method. Two colonies (for pBG184) and four colonies (for pBG185) were obtained on the LB Petri dishes supplemented with chloramphenicol. One mutant strain was selected for each construction and called BBG142 (for the strain containing the $P_{pps}$ promoter of *B. subtilis* 168) and BBG139 (for the strain containing $P_{fen}$ promoter of *B. subtilis* BBG21) respectively. After the extraction of genomic DNA, the verification of the new strains BBG142 and BBG139 was effected by PCR using *amy*E primers to detect an insertion within the amylase gene of the recombinant plasmid. An amplicon of about 5 kb from both *B. subtilis* BBG142 and BBG139 was detected compared to the 3 kb fragment from the mother strain 168, indicating that the fragment bearing $P_{pps}$ - *lacZ and* $P_{fen}$ - *lacZ* plasmid was successfully integrated into the new strain BBG142 and BBG139, respectively.

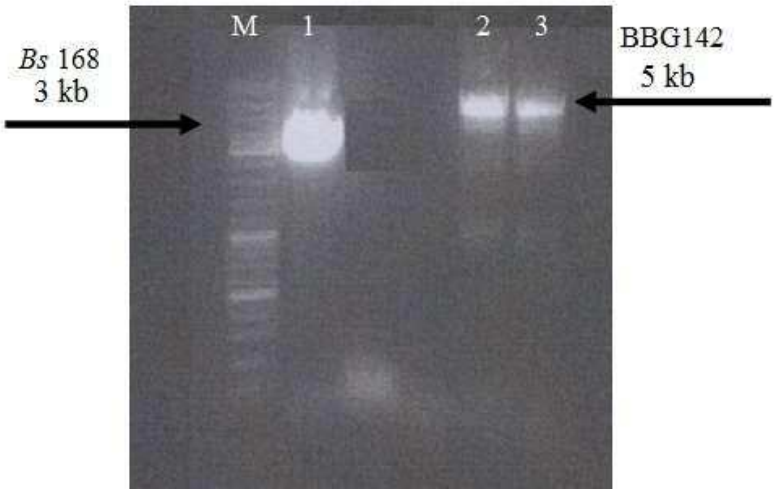



Figure 64: Verification of *pps* 168 insertion in *amyE* locus in *B. subtilis* BBG142 (*B. subtilis* 168 derivative)

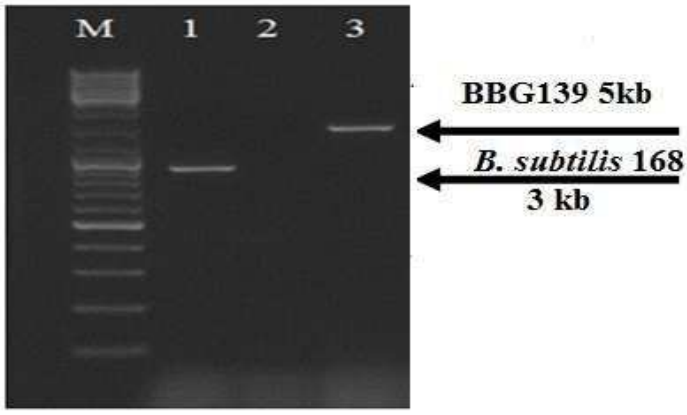



Figure 65: Verification of *Pfen* BBG21 insertion in *amyE* locus in *B. subtilis* BBG139 (*B. subtilis* 168 derivative)

As the insertion of 3kb fragment occurred in the *amy*E gene locus, amylase activity was checked. Another mutant strain containing $P_{srf}$-LacZ fusion also inserted in AmyE locus was also used in this experiment. As shown in figure 66, *B. subtilis* 168 showed an amylase activity whereas BBG 127, BBG 139 and BBG 142 can't no more hydrolyse the starch.

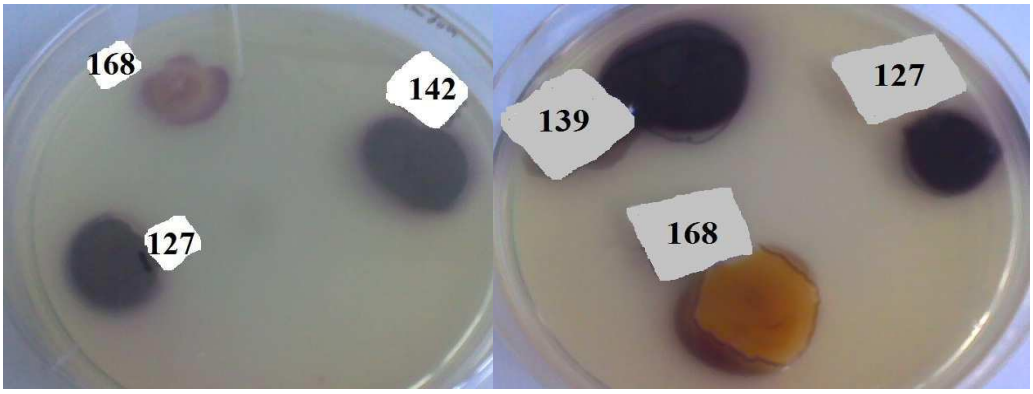



Figure 66: Amylase test on the new strains *B. subtilis* BBG127, BBG142 and BBG139 compared to mother strain *B. subtilis* 168. Amylase activity is shown as yellow colony. Lack of activity is shown as dark blue colony.

### 4.3. Comparison of β-Galactosidase activity between BBG142 and BBG139 strains

The plipastatin and the fengycin promoters expression of BBG142 and BBG139 strains were compared under different growth conditions. The strain BBG127 was added to this study to get, in parallel, results about the surfactin promoter expression as this promoter is known to be efficient and can be considered as a positive control. Eleven different experiments were performed in which we modified several cultivation parameters which could, on the basis of previous results obtained in the laboratory, influence plipastatin/fengycin promoter expression. These parameters were: temperatures (25, 30 and 37ºC); culture media with different nitrogen sources (Landy Mops (glutamic acid 5g/L), Landy modified (glutamic acid 5 g/L and ammonium sulfate 2.2 g/L) and Landy modified 2 (glutamic acid 2.5 g/L and ammonium sulfate 2.2 g/L)), oxygen transfer conditions (modified by changing the flask sizes (1L, 500 mL and 100 mL)), the agitation rates (200 and 160 rpm) and the filling volume of the flasks (20% and 30 %). All the experiments were carried with an initial pH of 7.0.

Table 26 shows the different experiments with their cultivation conditions used to determine the $P_{pps}$, $P_{fen}$ and $P_{srf}$ expression levels. The average of the expression level determined by Miller units at 78 h shows that the highest expression levels of $P_{pps}$ were induced by conditions of experiment 11 (79 Miller units) followed by experiment 3 (32.5 Miller units)

and experiment 10 (29 Miller units). These three experiments were performed at 30 ºC. In experiment 11 (maximum expression level), the oxygen transfer rate [the filling volume (30 %), the agitation rate (160 rpm) and the flask size (100 mL)] play also an important role in the induction of this promoter expression.

The average of the expression level determined by Miller units at 78 h shows that the highest expression levels of $P_{fen}$ were induced under the conditions of experiment 6 followed by experiment 5 and experiment 8. These three experiments were performed at 37 ºC. In experiment 6 (maximum expression level), the culture medium play an important role in the expression induction of this promoter. The Landy modified 2 medium have the more effect on expression induction followed by Landy MOPS and Landy modified medium.

The average of the expression level determined by Miller units at 78 h shows that the expression level of $P_{srf}$ has close expression levels under several experiments.

**Table 26: Different experiments cultivation conditions used to determine the $P_{pps}$, $P_{fen}$ and $P_{srf}$ expression levels**

| Experiment No | Culture media | Temperature | Flask size | Filling volume | Agitation rate | $P_{pps}$ expression level (Miller units) | $P_{fen}$ expression level (Miller units) | $P_{srf}$ expression level (Miller units) |
|---|---|---|---|---|---|---|---|---|
| **1** | Landy MOPS | 25°C | 1 L | 20% | 200 rpm | 2.5 | 20 | 7 |
| **2** | Landy modified 2 | 25°C | 1 L | 20% | 200 rpm | 27 | 41 | 39 |
| **3** | Landy MOPS | 30°C | 1 L | 20% | 200 rpm | 32.5 | 5 | 12 |
| **4** | Landy MOPS | 30°C | 100 mL | 20% | 200 rpm | 3 | 12 | 22.5 |
| **5** | Landy MOPS | 37°C | 1 L | 20% | 200 rpm | 3.2 | 93 | 58 |
| **6** | Landy modified 2 | 37°C | 1 L | 20% | 200 rpm | 2.8 | 132 | 11 |
| **7** | Landy modified | 25°C | 1 L | 20% | 200 rpm | 1.6 | 11.5 | 6 |
| **8** | Landy modified | 37°C | 1 L | 20% | 200 rpm | 0.2 | 72 | 61 |
| **9** | Landy modified | 30°C | 1 L | 20% | 160 rpm | 4.3 | 45 | 37 |
| **10** | Landy modified | 30°C | 500 mL | 20% | 160 rpm | 29 | 30 | 60 |
| **11** | Landy modified | 30°C | 100 mL | 30% | 160 rpm | 79 | 21 | 59 |

Expression levels were performed after 78h.

All the expression levels of these 11 experiments during 100 hours, microbial growth at $O.D_{600nm}$ and pH figures are listed in index 7.

## 4.4. Maximum expression levels for $P_{fen}$ BBG139

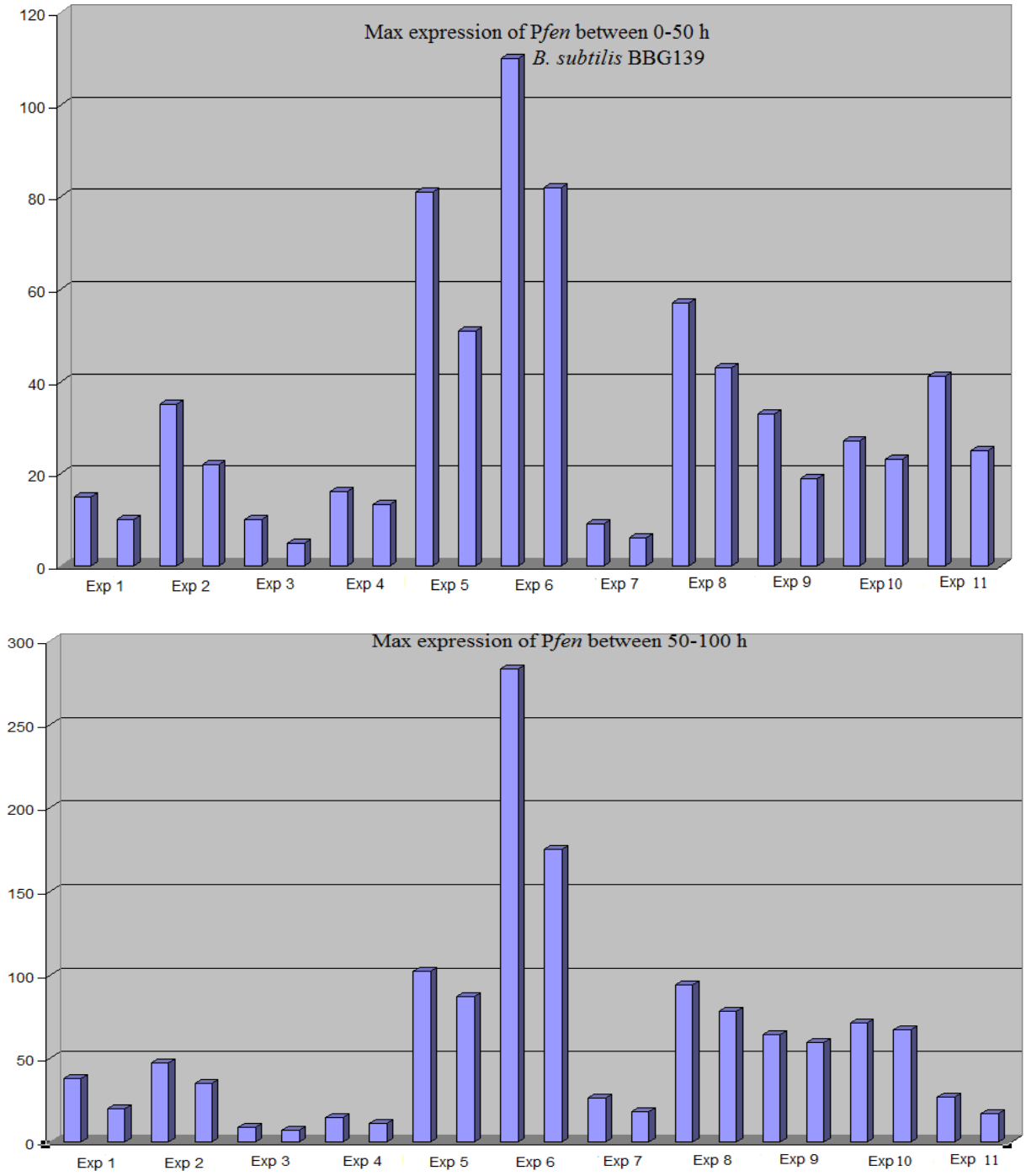



Figure 67: Maximum expression levels for $P_{fen}$ BBG139 at 0-50 h and 50-100 h under the different experiments.
**Exp 1**: Landy MOPS pH 7.1, 20% f.v, 1 L flask, 200 rpm and 25°C; **Exp 2**: Landy modified 2 pH 7.0, 20% f.v, 1 L flask, 200 rpm and 25°C; **Exp 3**: Landy MOPS pH 7.0, 20% f.v, 1 L flask, 200 rpm and 30°C; **Exp 4**: Landy MOPS pH 7.0, 20% f.v, 100 mL flask, 200 rpm and 30°C; **Exp 5**: Landy MOPS pH 7.0, 20% f.v, 1 L flask, 200 rpm and 37°C; **Exp 6**: Landy modified 2 pH 7.0, 20% f.v, 1 L flask, 200 rpm and 37°C ; **Exp 7** : Landy modified pH 7.0, 20% f.v, 1 L flask, 200 rpm and 25°C; **Exp 8**: Landy modified pH 7.0, 20% f.v, 1 L flask, 200 rpm and 37°C; **Exp 9**: Landy modified pH 7.0, 20% f.v, 1 L flask, 160 rpm and 30°C; **Exp 10**: Landy modified pH 7.0, 20% f.v, 500 mL flask, 160 rpm and 30°C; **Exp 11**: Landy modified pH 7.0, 30% f.v, 100 mL flask, 160 rpm and 30°C.

Figure 67 summarizes all the expression levels observed for $P_{fen}$ BBG139 under the different experiments performed t 0-50 h and 50-100 h with two replicates of each experiment.

At 0-50 h, $P_{fen}$ BBG139 showed the highest expression level for experiment 6 (110 and 82 Miller units) followed by experiment 5 (81 and 51 Miller units), 8 (57 and 43 Miller units) and 11 (41 and 25 Miller units), respectively. All the other experiments showed expression levels less than 40 Miller units.

At 50-100 h, the highest expression level was also observed for the experiment 6 (283 and 175 Miller units). Four experiments (5, 8, 9 and 10) showed expression levels ranged from 100 to 50 Miller units, while six experiments (1, 2, 3, 4, 7, and 11) showed expression level less than 50 Miller units.

It was observed that the expression levels obtained from 50-100 h was more important than that from 0-50 h.

A temperature of 37°C and the use of Landy modified 2 medium which contains glutamic acid at 2.5 g/L and ammonium sulphate at 2.2 g/L are the most favourable conditions for the expression of $P_{fen}$.

## 4.5. Maximum expression levels for $P_{pps}$ BBG142

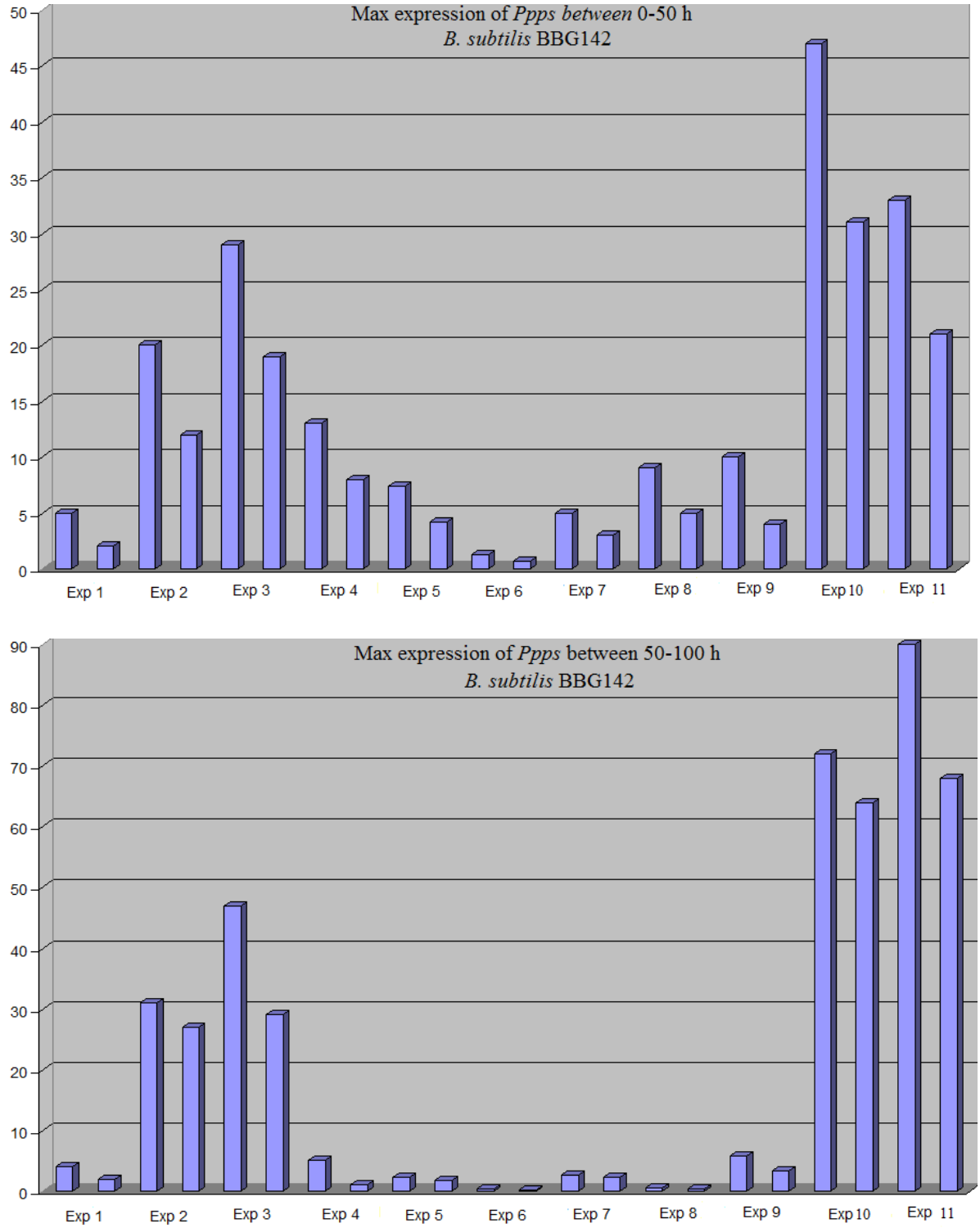



Figure 68: Maximum expression levels for $P_{pps}$ BBG142 at 0-50 h and 50-100 h under the different experiments. **Exp 1**: Landy MOPS pH 7.1, 20% f.v, 1 L flask, 200 rpm and 25°C; **Exp 2**: Landy modified 2 pH 7.0, 20% f.v, 1 L flask, 200 rpm and 25°C; **Exp 3**: Landy MOPS pH 7.0, 20% f.v, 1 L flask, 200 rpm and 30°C; **Exp 4**: Landy MOPS pH 7.0, 20% f.v, 100 mL flask, 200 rpm and 30°C; **Exp 5**: Landy MOPS pH 7.0, 20% f.v, 1 L flask, 200 rpm and 37°C; **Exp 6**: Landy modified 2 pH 7.0, 20% f.v, 1 L flask, 200 rpm and 37°C ; **Exp 7** : Landy modified pH 7.0, 20% f.v, 1 L flask, 200 rpm and 25°C; **Exp 8**: Landy modified pH 7.0, 20% f.v, 1 L flask, 200 rpm and 37°C; **Exp 9**: Landy modified pH 7.0, 20% f.v, 1 L flask, 160 rpm and 30°C; **Exp 10**: Landy modified pH 7.0, 20% f.v, 500 mL flask, 160 rpm and 30°C; **Exp 11**: Landy modified pH 7.0, 30% f.v, 100 mL flask, 160 rpm and 30°C.

Figure 68 summarizes all the expression levels detected for $P_{pps}$ BBG142 the different experiments performed at 0-50 h and 50-100 h with two replicates of each experiment.

At 0-50 h, the highest expression levels was detected for experiment 10 (47 and 31 Miller units) followed by experiment 11 (33 and 21 Miller units), experiment 3 (29 and 19 Miller units) and experiment 2 (20 and 12 Miller units). All the other experiments showed expression levels less than 15 Miller units.

At 50-100 h, the highest expression levels was detected for experiment 11(90 and 68 Miller units) followed by experiment 10 (72 and 64 Miller units), experiment 3 (47 and 29 Miller units) and experiment 2 (31 and 27 Miller units). All the other experiments showed expression levels less than 10 Miller units.

The expression levels obtained from 50-100 h was more important than 0-50 h for experiment 2, 3, 10 and 11. The maximum expression level from 50 -100 h was 90 Miller units and 47 Miller units at 0-50 h.

For the expression of this promoter, the best conditions are thus 30 °C, Landy modified medium, 160 rpm and a flasl size of 100 or 500 ml.

As observed from these results that the expression of both $P_{fen}$ and $P_{pps}$ promoters are quiet different in spite of the high similarity of their DNA sequences.

## 4.6. Maximum expression levels for $P_{srf}$ BBG127

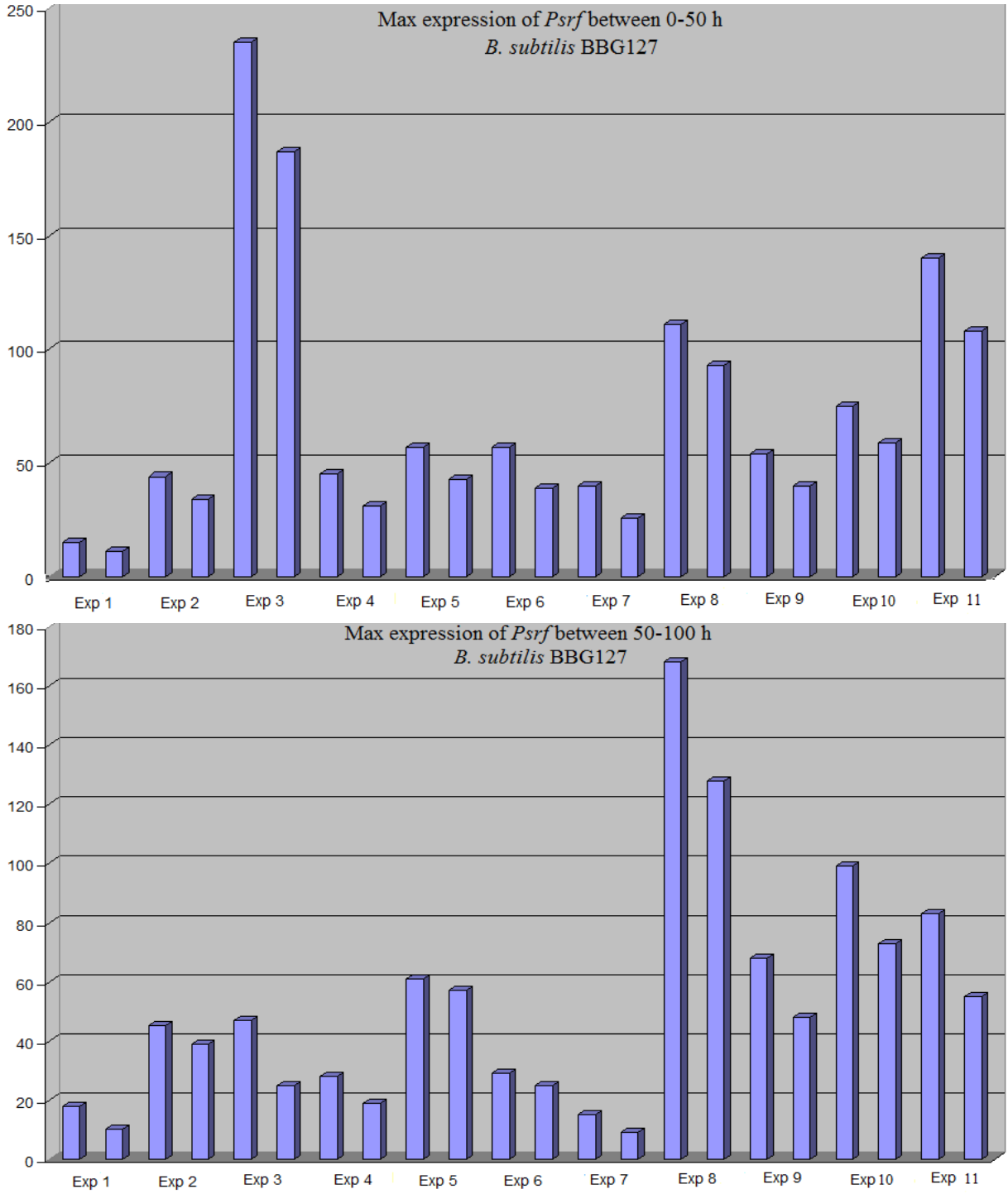



Figure 69: Maximum expression levels for $P_{srf}$ BBG127 at 0-50 h and 50-100 h under the different experiments. **Exp 1**: Landy MOPS pH 7.1, 20% f.v, 1 L flask, 200 rpm and 25°C; **Exp 2**: Landy modified 2 pH 7.0, 20% f.v, 1 L flask, 200 rpm and 25°C; **Exp 3**: Landy MOPS pH 7.0, 20% f.v, 1 L flask, 200 rpm and 30°C; **Exp 4**: Landy MOPS pH 7.0, 20% f.v, 100 mL flask, 200 rpm and 30°C; **Exp 5**: Landy MOPS pH 7.0, 20% f.v, 1 L flask, 200 rpm and 37°C; **Exp 6**: Landy modified 2 pH 7.0, 20% f.v, 1 L flask, 200 rpm and 37°C ; **Exp 7** : Landy modified pH 7.0, 20% f.v, 1 L flask, 200 rpm and 25°C; **Exp 8**: Landy modified pH 7.0, 20% f.v, 1 L flask, 200 rpm and 37°C; **Exp 9**: Landy modified pH 7.0, 20% f.v, 1 L flask, 160 rpm and 30°C; **Exp 10**: Landy modified pH 7.0, 20% f.v, 500 mL flask, 160 rpm and 30°C; **Exp 11**: Landy modified pH 7.0, 30% f.v, 100 mL flask, 160 rpm and 30°C.

Figure 69 summarizes all the expression levels detected for $P_{srf}$ BBG127 the different experiments performed previously at 0-50 h and 50-100 h with two replicates of each experiment.

At 0-50 h, the highest expression levels was detected for experiment 3 (235 and 187 Miller units) followed by experiment 11 (140 and 108 Miller units), experiment 8 (111 and 93 Miller units) and experiment 10 (75 and 59 Miller units). All the other experiments showed expression levels less than 57 Miller units.

At 50-100 h, the highest expression levels was detected for experiment 8 (168 and 128 Miller units) followed by experiment 10 (99 and 73 Miller units), experiment 11 (83 and 55 Miller units), experiment 9 (68 and 48 Miller units) and experiment 5 (61 and 57 Miller units). All the other experiments showed expression levels less than 47 Miller units.

As observed from these results the $P_{srf}$ BBG127 behave different than both $P_{fen}$ BBG139 and $P_{pps}$ BBG142 expression levels. $P_{srf}$ BBG127 showed more higher expression levels from 0-50 h than from 50-100 h, while both $P_{fen}$ BBG139 and $P_{pps}$ BBG142 showed expression levels from 50-100 h than from 0-50 h.

Contrary to the two other promoters, this promoter shows maximal expression at two different period of the culture in two different conditions. The best expression is observed between 0 and 50h in Landy MOPS, 1L flask and 30°C. Another important expression was observed between 50 and 100h in Landy modified and 37°C.

### 4.7. Discussion

In this work described here, we studied expression and regulation of $P_{pps}$, $P_{Fen}$ and $P_{srf}$ in *B. subtilis* 168 derivatives: BBG142, BBG139 and BBG127 respectively. The results revealed that the level of expression of $P_{pps}$, $P_{fen}$ and $P_{srf}$ could be controlled by fermentation conditions. By fixing the following factors: 20% filling volume in flask of 1 L and an agitation rate of 200 rpm, the temperature was found to be the first factor affected on the expression of $P_{srf}$ and $P_{fen}$ promoters. 30º C was found to highly induce $P_{srf}$ expression levels, while 37º C favored the highest expression levels of $P_{fen}$ promoter.

The second factor is the medium composition. Landy MOPS medium was the best for inducing $P_{srf}$ promoter, while Landy modified (Landy MOPS supplemented with ammonium sulphate) was the best for inducing $P_{fen}$ in contrast to $P_{srf}$ promoter, whose expression levels were reduced in the presence of ammonium sulphate.

Interestingly, the induction of the $P_{pps}$ promoter was found to be different, whereas it was affected not only by temperature and medium composition, but also by the acidity stress occurring during 24 h of the fermentation process and consumed at 30 h. $P_{pps}$ promoter showed the highest expression level under 30 ºC, in Landy modified and in flask of 100 mL with 30% filling volume and an agitation rate of 160 rpm.

Here, we didn't study all the transcription regulators of fengycin, plipastatin and surfactin, but earlier studies in *B. subtilis* 168 have shown that expression of *srfA* is mainly dependent on ComA (Nakano and Zuber, 1991; Roggiani and Dubnau, 1993). Regulation of *srfA* in *B. subtilis* 168 has been extensively studied because *srfA* also harbours the *comS* gene, a small protein involved in the development of genetic competence (Cosmina *et al.*, 1993; Hamoen *et al.*, 1995). For fengycin promoter of *B. subtilis* F29-3, Ke *et al*. (2009) reported that the UP element between positions -55 and -39 in *fenp* is critical for fengycin synthesis, but in this study the three differences observed for the promoter of BBG21 are not in this area. *B. subtilis* F29-3 mutant with a mutated UP element in *fenp* cannot transcribe the *fen* operon efficiently or produce sufficient fengycin to inhibit the germination of fungal spores. Since factors other than the UP element may regulate the transcription of the fengycin synthetase operon, further work should be performed to elucidate fully how these factors affect the synthesis of fengycin. Ke *et al*. (2009) also reported that the sporulation genes probably do not regulate the transcription of the *fen* operon.

### 4.8. Results valorization

A part of this work led to the following oral presentation:

Hussein W, Béchet M, Gancel F and Jacques P. 2010. Plipastatin operon regulation in *B. subtilis* 168 derivatives**.** 1st Bio-Processing and Application of Microbial Biotechnology in Agriculture, Cairo, Egypt.

# *Chapter V. General discussion and perspectives*

The characterization of the genes/operons encoding Non Ribosomal Peptide Synthetases from 39 *Bacillus* strains by bioinformatics tools has given, for the first time, an overview of the biodiversity of peptides and lipopeptides potentially produced by strains of this genus. Four main types of molecules were detected; lipopeptides, siderophores, antibiotics and new NRPS and PKS molecules. For the first molecules type 'lipopeptides', the presence of at least one member of the four main families of lipopeptides was detected in 15 *Bacillus* strains [surfactin, pumilacidin, lichenysin from surfactin family, fengycin or plipastatin from fengycin/plipastatin family, bacillomycin and mycosubtilin from iturin family and kurstakin]. Roongsawang *et al*. (2010) detected the presence of a wide range of lipopeptide synthetases from *Bacillus* strains; surfactin and lichenysin synthetases, fengycin synthetase, bacillomycin, iturin, and mycosubtilin synthetases. Also, Abderrahmani *et al*. (2010) confirmed the detection of NRPS genes involved in the biosynthesis of kurstakin from *Bacillus thuringiensis*.

In addition, the presence of other new lipohexapeptide fusaricidin from fusaricidins lipopeptide antibiotics group produced by *Paenibacillus polymyxa* (formerly *Bacillus polymyxa*) was also found. This result was in agreement with those of Bealty and Jensen (2002) who extracted the fusaricidin from *Paenibacillus polymyxa* PKB1 as an antifungal antibiotic active against *Leptosphaeria maculans*, the causative agent of blackleg disease of canola. Also, Roongsawang *et al*. (2010) confirmed the fusaricidin synthetase presence from *Paenibacillus polymyxa* PKB1.

The detection of the fengycin, surfactin and bacillomycin synthetases was also detected by Koumoutsi *et al*. (2004) from *Bacillus amyloliquefaciens* strain FZB42.

For the second molecules type 'siderophores', the presence of bacillibactin in 27 different *Bacillus* strains was detected. This result demonstrates that this compound is the main siderophore used by this genus. However, for ten of the analysed strains, the presence of genes involved in the biosynthesis of this siderophore was not detected. This strain should potentially use a siderophore synthesized by another mechanism.

For the third molecules type 'antibiotics', five *Bacillus* strains shown the presence of three antibiotics : gramicidin and tyrocidin from *Brevibacillus brevis* and one,truncated bacitracin

from four strains (*B. anthracis* str. CDC 684, *B. atrophaeus* 1942, *B. cereus* AH820 and *B. cereus* G9842). Bas Vogt *et al*. (2003) investigated the biosynthesis of isotopically labeled gramicidins and tyrocidins by *Bacillus brevis*.

For the fourth molecules type 'new NRPS and PKS molecules', thirteen *Bacillus* strains were found to harbour new NRPS and PKS genes, especially *Brevibacillus brevis*, *Paenibacillus polymyxa* E681 and *Paenibacillus polymyxa* SC2 which harbor high number of these new genes.

Finally, eight *Bacillus* strains were found to harbor no NRPS genes. These strains belong to the species *Bacillus megaterium* or grow in specific environments. All the results should be confirmed by blast analysis with a known NRPS gene.

The second part of the work was focused on fengycins or plipastatins produced by the NRPS system in some *B. subtilis* strains, in order to relate the structure of these two molecules and the structure of operons responsible for their biosynthesis. The analysis of plipastatin or fengycin operons sequences of *B. subtilis* F29-3, *B. subtilis* 168 and *B. amyloliquefaciens* FZB42 revealed the similarity of these molecules, especially in the two modules M9 and M3 (Tyr), which refers to the presence of D-Tyr on the M9 of the *ppsD* (*B. subtilis* 168), (*B. amyloliquefaciens* FZB42) and *fenA* (*B. subtilis* F29-3). This also may refer to the similarity of all these molecules which are plipastatins, while Schneider *et al*. (1999) reported the existence of fengycin molecule with a D-Tyr in position 3 from the supernatant of *B. subtilis* S499. For the importance of the fengycin molecule from *B. subtilis* S499, it was interesting to work on this strain.

In this study, we thus tried to amplify and sequence genes involved in the fengycin biosynthesis in *B. subtilis* S499. Firstly, primers designed using the sequence of plipastatin operon of *B. subtilis* 168 were used, then degenerated primers were tested. Amplified sequences from S499 strain showed a highest similarity with sequences from *B. amyloliquefaciens* than from *B. subtilis*. These results have shown that S499 strain belongs to *amyloliquefaciens* species. This result was recently confirmed by researchers of Gembloux Agro-Biotech (Marc Ongena, Personal Communication). Secondly, a long PCR product was detected from *B. subtilis* 168 and S499 with the *ppsB-ppsD* primers which gave a convergent fragment size (9.6 kb for *B. subtilis* 168) and about 10 kb for *B. subtilis* S499, while with the *ppsC-ppsE* primers, *B. subtilis* 168 gave PCR product of (12.3 kb), while *B. subtilis* S499 showed a fragment (10 kb) smaller than that of *B. subtilis* 168. This difference between the

length of this fragment between *B. subtilis* 168 and S499 could be referred to the absence of epimerization domain in position 9 from *ppsD* or *fenD* of *B. subtilis* S499, which could confirm the existence of a fengycin molecule with a D-Tyr in position 3 from the supernatant of *B. amyloliquefaciens* S499 reported by Schneider *et al*. (1999). However here, the presence of thisE domain couldn't be confirmed in position 3 (*ppsB* or *fenB*) because the fragment size in this region of S499 was not long enough (about 0.4 kb more) to say the E domain is present in position 3. The absence of this domain in position 9 was detected (*ppsD* or *fenD*) but unfortunately, the large fragment detected for S499 couldn't be cloned or sequenced. Then, these results highlighted the importance of sequencing the *B. amyloliquefaciens* S499 fengycin operon as it code for a synthetase responsible for the biosynthesis of a fengycin molecule that cannot be correlated with the structure of any of the synthetases described in other fengycin- or plipastatin-producing strains. Deciphering this operon should allow to overcome the difficulties in the differentiation between fengycins and plipastatins. Thus, the sequencing of the whole genome of this strain is important and of interest, to elucidate the NRPS operon implied in this synthesis. Meanwhile, our Laboratory ProBioGEM sent the *B. amyloliquefaciens* S499 strain to genome sequencing.

Obtaining plipastatin mono-producer or plipastatin- surfactin co-producer derivatives with high amounts of plipastatin by plipastatin native promoter replacement with the constitutive one $p_{repU}$ faced several difficulties. Two conclusions arose: the first is the difficulty of placing the plipastatin operon under the control of $P_{repU}$ constitutive promoter in spite of the previous obtention of a *B. subtilis* 168 derivative called *B. subtilis* 2504 (Ongena *et al*., 2007), which was constructed by the replacement of the plipastatin native promoter by the $P_{amyQ}$. This could be due to a difference between $P_{repU}$ and $P_{amyQ}$ or to an incompatibility of the $P_{repU}$ with the plipastatin operon. The second conclusion is that, the interruption of the plipastatin operon promoter obtained here led to increasing surfactin production, in spite of interrupting the surfactin operon reveals no increasing in plipastatin production as mentioned by Ongena *et al.* 2007, which reported that *B. subtilis* 2508 (*B. subtilis* 168 derivative) produced 697 mg/L of surfactin and 434 mg/L of plipastatin in optimalized medium and after the interruption of its surfactin operon, *B. subtilis* 2504 strain was obtained which produced 452 mg/L of plipastatin.

The last part of this study was dedicated to the regulation of the plipastatin, fengycin and surfactin operons in *B. subtilis* 168 by determining the strength of their promoters $P_{pps}$, $P_{Fen}$

and $P_{srf}$ using a gene reporter system. The results revealed that the level of expression of $P_{pps}$, $P_{fen}$ and $P_{srf}$ could be differently controlled by fermentation conditions. The temperature was found to be the first factor modifying the expression of $P_{srf}$ and $P_{fen}$ promoters but with fixing the following factors: 20% filling volume in flask of 1 L and an agitation rate of 200 rpm. Interestingly, the induction of the $P_{pps}$ promoter was found to be different, whereas it was affected not only by temperature and medium composition, but also by the acidity stress occurring during the first 24 h of the fermentation process. $P_{pps}$ promoter showed the highest expression level under 30 ºC, in Landy modified and in flask of 100 mL with 30% filling volume and an agitation rate of 160 rpm.

No available publications studied the effect of the different growth conditions on fengycin or plipastatin promoter expression. On the other hand, Ke *et al*. (2009) had studied the fengycin synthetase operon promoter of *Bacillus subtilis* F29-3 and determined that deleting the region between positions -55 and -42 reduces the promoter activity by 64.5 and nHA medium was used for production of fengycin. Also, Duitman *et al*. (2007) studied the transcriptional regulation of the surfactin- and mycosubtilin synthetase operons in *B. subtilis* ATCC6633 compared to the transcriptional regulation of the surfactin synthetase operon in *B. subtilis* 168 and reported that the expression of *srfA* is more than 100 times lower in strain ATCC 6633 than in the laboratory strain *B. subtilis* 168. Expression of the *myc* operon was highest in rich medium, whereas *srfA* expression reached maximal levels in minimal medium. In our study the expression of the three promoters $P_{pps,}$ $P_{fen}$ and $P_{srf}$ were studied during 96 hours of fermentation time, while Ke *et al*. (2009) studied the fengycin synthetase operon promoter of *Bacillus subtilis* F29-3 during 32 hours and Duitman *et al*. (2007) studied the transcriptional regulation of the surfactin- and mycosubtilin synthetase operons during the first eight hours of culture.

Regarding results obtained in this study, the following perspectives are proposed:

1- Bioinformatics analyses of 39 *Bacillus* strains revealed high number of new NRPS and PKS molecules, which could help in the future the researchers interesting of this domain to study and discover the properties and the industrial applications of these new molecules.

2- A deeper study of the organization of *B. subtilis* BBG21, 21332 and S499 is correlated to the promising expected results of S499 genome sequencing which would answer

the question of fengycin and plipastatin ″who is who″? and help in finding the main differences in their biosynthesis, structures, biological activities and applications.

3- The difficulty of obtaining plipastatin over and mono-producer strain could be faced without genetic tools, this was observed after the study of the plipastatin promoter expression of *B. subtilis* 168, which was not only induced by temperature and medium composition, but also by the acidity stress occurring during 24 h of the fermentation process.It showed the highest expression level under 30 ºC, in Landy modified and in flask of 100 mL with 30% filling volume and an agitation rate of 160 rpm. So, if we could link between the expression and the production of the plipastatin operon of *B. subtilis* 168 derivatives (bearing active *sfp* gene), such as BBG134 and BBG111 strains by following these fermentation conditions mentioned before, we could obtain good production of plipastatin.

# *References*

Abderrahmani A, Tapi A, Nateche F, Chollet M, Leclère V, Wathelet B, Hacene H, Jacques P (2011) Bioinformatics and molecular approaches to detect NRPS genes involved in the biosynthesis of kurstakin from *Bacillus thuringiensis*. Appl Microbiol Biotechnol **11**:3453-3456

Akpa E, Jacques P, Wathelet B, Paquot M, Fuchs R, Budzikiewicz H, Thonart P (2001) Influence of culture conditions on lipopeptide production by *Bacillus subtilis*. Appl Biochem Biotechnol **91**–93: 551-561

Ara K, Ozaki K, Nakamura K, Yamane K, Sekiguchi J, Ogasawara N (2007) *Bacillus* minimum genome factory: effective utilization of microbial genome information. Biotechnol Appl Biochem **46**:169-178

Arima K, Kakinuma A, Tamura G (1968) Surfactin, a crystalline peptidelipid surfactant produced by *Bacillus subtilis*: isolation, characterization and its inhibition of fibrin clot formation. Biochem Biophys Res Commun **31**:488-494

Athukorala SNP, Fernando WGD, Rashid KY (2009) Identification of antifungal antibiotics of *Bacillus* species isolated from different microhabitats using polymerase chain reaction and MALDI-TOF mass spectrometry. Can J Microbiol **55**:1021-1032

Bachmann BO, Ravel J (2009) In silico prediction of microbial secondary metabolic pathways from DNA sequence data. Meth Enzymol **458**:181-217

Bas Vogt TC, Schinzel S, Bechinger B (2003) Biosynthesis of isotopically labeled gramicidins and tyrocidins by *Bacillus brevis*. J Biomol NMR **26**: 1-11

Beatty PH, Jensen SE (2002) *Paenibacillus polymyxa* produces fusaricidin-type antifungal antibiotics active against *Leptosphaeria maculans*, the causative agent of blackleg disease of canola. Canad J Microbiol **48**: 159-169

Bergendahl V, Linne U, Marahiel MA (2002) Mutational analysis of the C domain in nonribosomal peptide synthesis. Eur J Biochem **269**:620-629

Besson F, Peypoux F, Michel G, Delcambe L (1977) The structure of bacillomycin L, an antibiotic from *Bacillus subtilis*. Eur J Biochem **77**:61-67

Bonmatin JM, Laprevote O, Peypoux F (2003) Diversity among microbial cyclic lipopeptides: iturins and surfactins. Activity-structure relationships to design new bioactive agents. Comb Chem High Throughput Screen **6**:541–556

Cane DE, Walsh CT (1999) The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases. Chem Biol **6**:R319-325

Casadaban MJ, Martinez-Arias A, Shapira SK, Chou J **(**1983) Beta-galactosidase gene fusions for analyzing gene-expression in *Escherichia coli* and yeast. Method Enzymol **100**:293-308

Chan YK, Marc E, Savard EM, Reid ML, Cyr T, McCormick AW, Seguin C (2009) Identification of lipopeptide antibiotics of a *Bacillus subtilis* isolate and their control of *Fusarium graminearum* diseases in maize and wheat. BioControl **54**:567-574

Chen XH, Koumoutsi A, Scholz R, Schneider K, Vater J, Süssmuth R, Piel CJ, Borriss R (2009) Genome analysis of *Bacillus amyloliquefaciens* FZB42 reveals its potential for biocontrol of plant pathogens. J Biotechnol **140**:27-37

Cho KM, Math RK, Hong SY, Islam SMA, Mandanna DK, Cho JJ, Yun MG, Kim JM, Yun HD (2009) Iturin produced by *Bacillus pumilus* HY1 from Korean soybean sauce (kanjang) inhibits growth of aflatoxin producing fungi. Food Control **20**:402-406

Cho SJ, Lee SK, Cha BJ, Kim YH, Shin KS (2003) Detection and characterization of the *Gloeosporium gloeosporioides* growth inhibitory compound iturin A from *Bacillus subtilis* strain KS03. FEMS Microbiol Lett **223**:47-51

Chollet-Imbert M, Gancel F, Slomianny C, Jacques P (2009) Differentiated pellicle organization and lipopeptide production in standing culture of *Bacillus subtilis* strains. Arch Microbiol **191**:63-71

Chung YR, Kim CH, Hwang I, Chun J (2000) *Paenibacillus koreensis* sp. nov., a new species that produces an iturin-like antifungal compound. Int J Syst Evol Microbiol **50**:1495-1500

Cooper DG, Macdonald CR, Duff JP, Kosaric N (1981) Enhanced production of surfactin from *Bacillus subtilis* by continuous product removal and metal cation additions. Appl Environm Microbiol **42**: 408-412

Cosmina P, Rodriguez F, de Ferra F, Grandi G, Perego M, Venema G, van Sinderen D (1993) Sequence and analysis of the genetic locus responsible for surfactin synthesis in *Bacillus subtilis*. Mol Microbiol **8**:821-831

Coutte F, Lecouturier D, Ait Yahia S, Leclère V, Béchet M, Jacques P, Dhulster P (2010) Production of surfactin and fengycin by *Bacillus subtilis* in a bubbleless membrane bioreactor. Appl Microbiol Biotechnol **87**:499-507

Deleu M, Paquot M, Nylander T (2008) Effect of fengycin, a lipopeptide produced by *Bacillus subtilis*, on model biomembranes. Biophys J **94**:2667-2679

Dieckmann R, Lee YO, van Liempt H, von Döhren H, Kleinkauf H (1995) Expression of an active adenylate-forming domain of peptide synthetases corresponding to acyl-CoA-synthetases. FEB SUBTILIS Lett **357**:212-16

Donnelly CE, Sonenshein AL **(**1982) Genetic fusion of *E. coli lac* genes to a *B. subtilis* promoter. p. 63-72. In A. Ganesan, J. Hoch, and S. Chang (ed.), Molecular cloning and gene regulation in *Bacilli*. Academic Press, Inc., New York

Du L, Sanchez C, Chen M, Edwards DJ, Shen B (2000) The biosynthetic gene cluster for the antitumor drug bleomycin from *Streptomyces verticillus* ATCC 15003 supporting functional

interactions between nonribosomal peptide synthetases and a polyketide synthase. Chem Biol **7**:623-642

Duitman EH, Hamoen LW, Rembold M, Venema G, Seitz H, Saenger W, Bernhard F, Reinhardt R, Schmidt M, Ullrich C, Stein T, Leenders F, Vater J (1999) The mycosubtilin synthetase of *Bacillus subtilis* ATCC6633: a multifunctional hybrid between a peptide synthetase, an amino transferase, and a fatty acid synthase. Proc Natl Acad Sci USA **96**:13294-13299

Duitman EH, Wyczawski D, Boven LG, Venema G, Kuipers OP, Hamoen LW ( 2007) Novel methods for genetic transformation of natural *Bacillus subtilis* isolates used to study the regulation of the mycosubtilin and surfactin synthetases. Appl Environ Microbiol **73**:3490-3496

Eeman M, Deleu M, Paquot M, Thonart P, Dufrêne YF, (2005) Nanoscale properties of mixed fengycin-ceramide monolayers explored using atomic force microscopy. Langmuir **21**:2505- 2511

Eeman M, Francius G, Dufrêne YF, Nott K, Paquot M, Deleu M (2009) Effect of cholesterol and fatty acids on the molecular interactions of fengycin with *Stratum corneum* mimicking lipid monolayers. Langmuir **25**:3029-3039

Ehmann DE, Shaw-Reid CA, Losey HC, Walsh CT (2000) The EntF and EntE adenylation domains of *Escherichia coli* enterobactin synthetase: sequestration and selectivity in acyl-AMP transfers to thiolation domain co-substrates. Proc Natl Acad Sci USA **97**:2509-2514

Eppelmann K, Stachelhaus T, Marahiel MA (2002) Exploitation of the selectivity conferring code of nonribosomal peptide synthetases for the rational design of novel peptide antibiotics. Biochem **41**:9718-9726

Eshita SM, Roberto NH, Beale JM, Mamiya BM, Workman RF (1995) Bacillomycin Lc, a new antibiotic of the iturin group: isolations, structures, and antifungal activities of the congeners. J Antibiot (Tokyo) **48**:1240-1247

Faulkner DJ (2002) Marine natural products. Nat Prod Rep **19**:1-48

Ferrari E, Howard SMH, Hoch JA **(**1985) Effect of sporulation mutations on subtilisin expression, assayed using a subtilisin-β-galatosidase gene fusion, p. 180-184. In J. A. Hoch and P. Stelow (ed.)., Molecular biology of microbial differentiation. American Society for Microbiology, Washington, D. C.

Finking R, Marahiel MA (2004) Biosynthesis of nonribosomal peptides. Annu Rev Microbiol **58**:453-88

Galli G, Rodriguez F, Cosmina P, Pratesi C, Nogarotto R, de Ferra F, Grandi G (1994) Characterization of the surfactin synthetase multi-enzyme complex. Biochim Biophys Acta **1205**:19-28

Gancel F, Montastruc L, Liu T, Zhao L, Nikov I (2009) Lipopeptide overproduction by cell immobilization on iron-enriched light polymer particles. Process Biochem **44**:975-978

Grangemard, I, Bonmatin, JM, Bernillon J, Das BC, Peypoux F (1999) Lichenysins G, a novel family of lipopeptide biosurfactants from *Bacillus licheniformis* IM 1307: production, isolation and structural evaluation by NMR and mass spectrometry. J Antibiot **52**:363-373

Grangemard I, Wallach J, Maget-Dana R, Peypoux F (2001) Lichenysin: a more efficient cation chelator than surfactin. Appl Biochem Biotechnol **90**:199-210

Grossman TH, Tuckman M, Ellestad S, Osburne MS (1993). "Isolation and characterization of *Bacillus subtilis* genes involved in siderophore biosynthesis: relationship between *B. subtilis sfpo* and *Escherichia coli entD* genes." J Bacteriol 175:6203-6211

Grote A, Klein J, Retter I, Haddad I, Behling S, Bunk B, Biegler I, Yarmolinetz S, Jahn D, Munch R (2009) A database and tool platform for the analysis of gene regulation in prokaryotes**.** Nucl Acids Res **37:**D61-D65

Guerout-Fleury AM, Frandsen N, Stragier P (1996) Plasmids for ectopic integration in *Bacillus subtilis*. Gene **180**:57-61

Guez JS, Chenikher S, Cassar JP, Jacques P (2007) Setting up and modelling of overflowing fed-batch cultures of *Bacillus subtilis* for the production and continuous removal of lipopeptides. J Biotechnol **131**:67-75

Hamoen LW, Eshuis H, Jongbloed J, Venema G, van Sinderen D **(**1995) A small gene, designated *comS*, located within the coding region of the fourth amino acid-activation domain of *srfA*, is required for competence development in *Bacillus subtilis*. Mol Microbiol **15**:55-63

Hofemeister J, Conrad B, Adler B, Hofemeister B, Feesche J, Kucheryava N, Steinborn G, Franke P, Grammel N, Zwintscher A, Leenders F, Hitzeroth G, Vater J (2004) Genetic analysis of the biosynthesis of nonribosomal peptide- and polyketide-like antibiotics, iron uptake and biofilm formation by *Bacillus subtilis* A1/3. Mol Genet Genom **272**:363-378

Horowitz S, Gilbert JN, Griffin WM (1990) Isolation and characterization of a surfactant produced by *Bacillus licheniformis* 86. J Ind Microbiol Biotechnol **6**:243-248

Hu LB, Shi QZ, Zhang T, Yang ZM (2007) Fengycin antibiotics isolated fromB-FS01culture inhibit the growth of *Fusarium moniliforme* Sheldon ATCC 38932. FEMS Microbiol Lett **272**:91-98

Huang X, Wei Z, Zhao G, Gao X, Yang S, Cui Y (2008) Optimization of sterilization of *Escherichia coli* in milk by surfactin and fengycin using a response surface method. Curr Microbiol **56**:376-381

Huszcza E, Burczyk B (2006) Surfactin isoforms from *Bacillus coagulans*. Z Naturforsch C **61**:727-733

Jacques P, Hbid C, Destain J, Razafindralambo H, Paquot M, De Pauw E, Thonart P (1999) Optimization of biosurfactant lipopeptide production from *Bacillus subtilis* S499 by Plackett-Burman design. Appl Biochem Biotechnol **77**:223-233

Jacques P. (2011) Surfactin and other lipopeptides from *Bacillus* spp. In Biosurfactants Microbiology Monographs, Soberon-Chavez G, ed., Volume 20, Chapter 3**,** pp. 57-91, Springer

Joshi R, McSpadden Gardener B (2006) Identification and characterization of novel genetic markers associated with biological control activities in *Bacillus subtilis*. Phytopathology **96**:145-154

Kajimura Y, Sugiyama M, Kaneda M (1995) Bacillopeptins, new cyclic lipopeptide antibiotics from *Bacillus subtilis* FR-2. J Antibiot (Tokyo) **48**:1095-1103

Kakinuma A, Tamura G, Arima K (1968) Wetting of fibrin plate and apparent promotion of fibrinolysis by surfactin, a new bacterial peptidelipid surfactant. Experientia **24**:1120-1121

Katz L (1997) Manipulation of modular polyketide synthases. Chem Rev **97**:2557-2576

Ke W-J, Chang BY, Lin TP, Liu ST (2009) Activation of the promoter of the fengycin synthetase operon by the UP element. J Bacteriol **191**:4615-4623

Kim PI, Bai H, Bai D, Chae H, Chung S, Kim Y, Park R, Chi YT (2004) Purification and characterization of a lipopeptide produced by *Bacillus thuringiensis* CMB26. J Appl Microbiol **97**:942-949

Kobayashi K, *et al* (2003) Essential *Bacillus subtilis* genes. Proc Natl Acad Sci U S A **100**: 4678-4683

Kohli RM, Trauger JW, Schwarzer D, Marahiel MA, Walsh CT (2001) Generality of peptide cyclization catalyzed by isolated thioesterase domains of nonribosomal peptide synthetases. Biochem **40**:7099-7108

Koumoutsi A, Chen XH, Henne A, Liesegang H, Hitzeroth G, Franke P, Vater J, Borriss R (2004) Structural and functional characterization of gene clusters directing nonribosomal synthesis of bioactive cyclic lipopeptides in *Bacillus amyloliquefaciens* strain FZB42. J Bacteriol **186**:1084-1096

Kunst F, Ogasawara N, Moszer I, Albertini AM, Alloni G, Azevedo V, Bertero MG, Bessieres P, Bolotin A, Borchert S *et al*. (1997) The complete genome sequence of the gram positive bacterium *Bacillus subtilis***.** Nature **390:** 249-256

Lambalot RH, Gehring AM, Flugel RS, Zuber P, LaCelle M (1996) A new enzyme superfamily: the phosphopantetheinyl transferases. Chem Biol **3**:923-936

Landy M, Warren GH, Rosenman SB, Colio LG (1948) Bacillomycin an antibiotic from *Bacillus subtilis* active against pathogenic fungi. Proc Soc Exp Biol Med **67**:530-541

Leclere V, Bechet M, Adam A, Guez J-S, Wathelet B, Ongena M, Thonart P, Gancel F, Chollet-Imbert M, Jacques P (2005) Mycosubtilin overproduction by *Bacillus subtilis* BBG100 enhances the organism's antagonistic and biocontrol activities. Appl Envirom Microbiol **71**:4577-4584

Lee SC, Kim SH, Park IH, Chung SY, Choi YL (2007) Isolation and structural analysis of bamylocin A, novel lipopeptide from *Bacillus amyloliquefaciens* LP03 having antagonistic and crude oil-emulsifying activity. Arch Microbiol **188**:307-312

Leenders F, Stein TH, Kablitz B, Franke P, Vater J (1999) Rapid typing of *Bacillus subtilis* strains by their secondary metabolites using matrix-assisted laser desorption/ionization mass spectrometry of intact cells. Rap Commun Mass Spectrom **13**: 943-949

Lewandoski M, Smith I (1988) Use of a versatile *lacZ* vector to analyze the upstream region of the *Bacillus subtilis spoOF* gene. Plasmid **20**:148-154

Lin GH, Chen CL, Tschen JSM, Tsay SS, Chang YS, Liu ST (1998) Molecular cloning and characterization of fengycin synthetase gene *fenB* from *Bacillus subtilis.* J Bacteriol **180**: 1338-1341

Lin TP, Chen CL, Chang LK, Tschen JSM, Liu ST (1999) Functional and transcriptional analyses of a fengycin synthetase gene, *fenC*, from *Bacillus subtilis.* J Bacteriol **181**: 5060-5067

Lin TP, Chen C-L, Fu H-C, Wu C-Y, Lin G-H, Huang S-H, Chang LK, Liu S-T (2005) Functional analysis of fengycin synthetase FenD. Biochimica Biophysica Acta **1730**:159-164

Liu T, Montastruc L, Gancel F, Zhao L, Nikov I (2007) Integrated process for production of surfactin (Part 1): adsorption rate of pure surfactin onto activated carbon. Biochem Engin J **33**:333-340

Marahiel MA (1997) Protein templates for the biosynthesis of peptide antibiotics. Chem Biol **4**:561-567

Marino M, Ramos HC, Hoffmann T, Glaser P, Jahn D (2001) Modulation of anaerobic energy metabolism of *Bacillus subtilis* by *arfM* (*ywiD*). J Bacteriol **183**:6815-6821

May JJ, Wendrich TM, Marahiel MA (2001) The *dhb* operon of *Bacillus subtilis* encodes the biosynthetic template for the catecholic siderophore 2,3-dihydroxybenzoate-glycine-threonine trimeric ester bacillibactin. J Biol Chem **276**:7209-7217

Miller DA, Luo L, Hillson N, Keating TA, Walsh CT (2002) Yersiniabactin synthetase. A four-protein assembly line producing the nonribosomal peptide/polyketide hybrid siderophore of *Yersinia pestis*. Chem Biol **9**:333-344

Montville TJ, Chen Y (1995) Efflux of ions and ATP depletion induced by pediocin PA-1 are concomitant with cell death in *Listeria monocytogenes* Scott A. J Appl Microbiol **70**:684-690

Mootz HD, Marahiel MA (1997) The tyrocidine biosynthesis operon of *Bacillus brevis*: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. J Bacteriol **179**:6843-6850

Mootz HD, Kessler N, Linne U, Eppelmann K, Schwarzer D, Marahiel MA (2002) Decreasing the ring-size of a nonribosomal peptide antibiotic by in frame module deletion in the biosynthetic genes. J Am Chem Soc **124**:10980-10981

Morikawa M, Ito M, Imanaka T (1992) Isolation of a new surfactin producer *Bacillus pumilus* A-1, and cloning and nucleotide sequence of the regulator gene, *psf-1*. J Ferment Bioeng **74**:255-261

Moyne AL, Cleveland TE, Tuzun S (2004) Molecular characterization and analysis of the operon encoding the antifungal lipopeptide bacillomycin D. FEMS Microbiol Lett **234**:43-49

Nakano MM, Zuber P (1991) The primary role of ComA in stablishment of the competent state in *Bacillus subtilis* is to activate expression of *srfA*. J Bacteriol **173**:7269-7274

Nakano MM, Zuber P (1998) Anaerobic growth of a strict aerobe (*Bacillus subtilis*). Ann Rev Microbiol **52**:165-190

Naruse N, Tenmyo O, Kobaru S, Kamei H, Miyaki T, Konishi M, Oki T (1990) Pumilacidin, a complex of new antiviral antibiotics. Production, isolation, chemical properties, structure and biological activity. J Antibiotics **XLIII**: 267-280

Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol **48**:443-453

Nishikiori T, Naganawa H, Muraoka Y, Aoyagi T, Umezawa H (1986a) Plipastatins: new inhibitors of phospholipase A2, produced by *Bacillus cereus* BMG302-fF67. III. Structural elucidation of plipastatins. J Antibiot **39**:755-761

Nishikiori T, Naganawa H, Muraoka Y, Aoyagi T, Umezawa H (1986b) The conformational studies on plipastatin A1 by 400 MHZ Proton Magnetic Resonance. J Antibiot **6**:860-863

Oka K, Hirano T, Homma M, Ishii H, Murakami K, Mogami S *et al*. (1993) Satisfactory separation and MS–MS spectrometry of six surfactin isolated from *Bacillus subtilis* natto. Chem Pharmac Bull **41**:1000-1002

Ongena M, Duby F, Jourdan E, Beaudry T, Jadin V. Dommes J, Thonart P (2004) *Bacillus subtilis* M4 decreases plant susceptibility towards fungal pathogens by increasing host

resistance associated with differential gene expression. Appl Microbiol Biotechnol **67**:692-698

Ongena M, Jacques P (2008) *Bacillus* lipopeptides: versatile weapons for plant disease biocontrol. Trends Microbiol **16**:115-125

Ongena M, Jacques P, Touré Y, Destain J, Jabrane A, Thonart P (2005) Involvement of fengycin-type lipopeptides in the multifaceted biocontrol potential of *Bacillus subtilis*. Appl Microbiol Biotechnol **69**:29-38

Ongena M, Jourdan E, Adam A, Paquot M, Brans A, Joris B, Arpigny JL, Thonart P (2007) Surfactin and fengycin lipopeptides of *Bacillus subtilis* as elicitors of induced systemic resistance in plants. Environ Microbiol **9**:1084-1090

Peypoux F, Michel G, Delcambe L (1976) The structure of mycosubtilin, an antibiotic isolated from *Bacillus subtilis*. Eur J Biochem **63**:391-398

Peypoux F, Besson F, Michel G (1978) Structure of iturin C from *Bacillus subtilis*. Tetrahedron **34**:1147-1152

Peypoux F, Besson F, Michel G, Delcambe L (1981) Structure of bacillomycin D, a new antibiotic of the iturin group. Eur J Biochem **118**:323-327

Peypoux F, Marion D, Maget-Dana R (1985) Structure of bacillomycin F, a new peptidolipid antibiotic of the iturin group. Eur J Biochem **153**:335-340

Peypoux F, Pommier MT, Marion D, Ptak M, Das BC, Michel G (1986) Revised structure of mycosubtilin, a peptidolipid antibiotic from *Bacillus subtilis*. J Antibiot (Tokyo) **39**:636-641

Peypoux F, Bonmatin JM, Wallach J (1999) Recent trends in the biochemistry of surfactin. Appl Microbiol Biotechnol **51**:553-563

Piggot PJ, Curtiss CAM **(**1987) Analysis of the regulation of gene expression during *Bacillus subitlis* sporulation by manipulation of the copy number of *spo-lacZ* fusions. J Bacteriol **169**:1260-1266

Rasmussen S, Nielsen HB, Jarmer H (2009) The transcriptionally active regions in the genome of *Bacillus subtilis***.** Mol Microbiol **73**:1043-1057

Roggiani M, Dubnau D (1993) ComA, a phosphorylated response regulator protein of *Bacillus subtilis*, binds to the promoter region of *srfA*. J Bacteriol **175**:3182-3187

Romero D, de Vicente A, Rakotoaly RH, Dufour SE, Veening JW, Arrebola E, Cazorla FM, Kuipers OP, Paquot M, Pérez-García A ( 2006) The iturin and fengycin families of lipopeptides are key factors in antagonism of *Bacillus subtilis* toward *Podosphaera fusca.* Mol Plant Microbe Interactions **20**:430-440

Roongsawang N, Washio K, Morikawa M (2011) Diversity of nonribosomal peptide synthetases involved in the biosynthesis of lipopeptide biosurfactants. Int J Mol Sci **12**:141-172

Sambrook J, and Russel DW (2001) Molecular cloning: a laboratory manual, 3rd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Samel SA, Wagner B, Marahiel MA, Essen LV (2006) The thioesterase domain of the fengycin biosynthesis cluster: A structural base for the macrocyclization of a non-ribosomal lipopeptide. J Mol Biol **359**:876-889

Schaechter M, Ingraham JL, Neidhardt FC (2006) Microbe

Schneider A, Marahiel MA (1998) Genetic evidence for a role of thioesterase domains, integrated in or associated with peptide synthetases, in non-ribosomal peptide biosynthesis in *Bacillus subtilis*. Arch Microbiol **169**:404-410

Schneider J, Taraz K, Budzikiewicz H, Deleu M, Thonart P, Jacques P (1999) The structure of two fengycins from *Bacillus subtilis* S499. Z Naturforsch C **54**:859-865

Schwarzer D, Marahiel MA (2001) Multimodular biocatalysts for natural product assembly. Naturwissenschaften **88**:93-101

Schwarzer D, Finking R, Marahiel MA (2003) Nonribosomal peptides: from genes to products. Nat Prod Rep **20**:275-287

Shaw-Reid CA, Kelleher NL, Losey HC, Gehring AM, Berg C, Walsh CT (1999) Assembly line enzymology by multimodular nonribosomal peptide synthetases: The thioesterase domain of *E. coli* EntF catalyzes both elongation and cyclolactonization. Chem Biol **6**:385-400

Shu HY, Lin GH, Wu YC, Tschen JS, Liu ST (2002) Amino acids activated by fengycin synthetase FenE. Biochem Biophys Res Commun **292**:789-793

Sieber SA, Marahiel MA (2003) Learning from nature's drug factories: nonribosomal synthesis of macrocyclic peptides. J Bacteriol **185**:7036-7043

Sierro N, Makita Y, de HM, Nakai K (2008) DBTB: a database of transcriptional regulation in *Bacillus subtilis* containing upstream intergenic conservation information**.** Nucl Acids Res **36:** D93-D96

Siragusa GR, Cutter CN, Willett JL (1999) Incorporation of bacteriocin in plastic retains activity and inhibits surface growth of bacteria on meat. Food Microbiol **16**:229-235

Stachelhaus T, Marahiel MA (1995) Modular structure of peptide synthetases revealed by dissection of the multifunctional enzyme GrsA. J Biol Chem **270**: 6163-6169

Stachelhaus T, Hüser A, Marahiel MA (1996) Biochemical characterization of peptidyl carrier protein (PCP) the thiolation domain of multifunctional peptide synthetases. Chem Biol **3**:913-921

Stachelhaus T, Mootz HD, Bergendahl V, Marahiel MA (1998) Peptide bond formation in nonribosomal peptide biosynthesis catalytic role of the condensation domain. J Biol Chem **273**:22773-22781

Stachelhaus T, Mootz HD, Marahiel MA (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. Chem Biol **6**:493-505

Stein T, Vater J, Kruft V, Otto A, Wittmann-Liebold B, Franke P, Panico M, McDowell R, Morris HR (1996) The multiple carrier model of nonribosomal peptide biosynthesis at modular multienzymatic templates. J Biol Chem **271**:15428-15435

Steller S, Vater J (2000) Purification of the fengycin synthetase multienzyme system from *Bacillus subtilis* b213. J Chromatog Biomed Sci Appl **14**:267-275

Sun L, Zhaoxin Lu Z, Bie X, Lu F, Yang S (2006) Isolation and characterization of a co-producer of fengycins and surfactins, endophytic *Bacillus amyloliquefaciens* ES-2, from *Scutellaria baicalensis* Georgi. World J Microbiol Biotechnol **22**:1259-1266

Tapi A, Chollet-Imbert M, Scherens B, Jacques P (2009) New approach for the detection of non-ribosomal peptide synthetase genes in *Bacillus* strains by polymerase chain reaction. Appl Microbiol Biotechnol **85**:1521-1531

Thomas DW, Ito T (1969) The revised structure of the peptide antibiotic esperin, established by mass spectrometry. Tetrahedron **25**:1985-1990

Tosato V, Albertini AM, Zotti M, Sonda S, Bruschi CV (1997) Sequence completion, identification and definition of the fengycin operon in *Bacillus subtilis* 168. Microbiology **143**:3443-3450

Touré Y, Ongena M, Jacques P, Guiro A, Thonart P (2004) Role of lipopeptides produced by *Bacillus subtilis* GA1 in the reduction of grey mould disease caused by *Botrytis cinerea* on apple. J Appl Microbiol **96**:1151-1160

Trauger JW, Kohli RM, Walsh CT (2001) Cyclization of backbone-substituted peptides catalyzed by the thioesterase domain from the tyrocidine nonribosomal peptide synthetase. Biochemistry **40**:7092-7098

Trischman JA, Jensen PR, Fenical W (1994) Halobacillin: a cytotoxic cyclic acylpeptide of the iturin class produced by a marine *Bacillus*. Tetrahedron Lett **35**:5571-5574

Tseng CC, Bruner SD, Kohli RM, Marahiel MA, Walsh CT, Sieber SA (2002) Characterization of the surfactin synthetase C-terminal thioesterase domain as a cyclic depsipeptide synthase. Biochemistry **41**:13350-13359

Tsuge K, Ano T, Shoda M (1996) Isolation of a gene essential for biosynthesis of the lipopeptide antibiotics plipastatin B1 and surfactin in *Bacillus subtilis* YB8. Arch Microbiol **165**:243-251

Tsuge K, Ano T, Hirai M, Nakamura Y, Shoda M (1999) The genes *degQ*, *pps*, and *lpa-8* (sfp) are responsible for conversion of *Bacillus subtilis* 168 to plipastatin production. Antimicrob Agents Chemother **43**:2183-2192

Tsuge K, Akiyama T, Shoda M (2001) Cloning, sequencing, and characterization of the iturin operon. J Bacteriol **183**:6265-6273

Tsuge K,Matsui K, Itaya M (2007) Production of the non-ribosomal peptide plipastatin in *Bacillus subtilis* regulated by three relevant gene blocks assembled in a single movable DNA segment. J Biotechnol **129**:592-603

van Hijum SA, Medema MH, Kuipers OP (2009) Mechanisms and evolution of control logic in prokaryotic transcriptional regulation**.** Microbiol Mol Biol Rev **73**:481-509

Vanittanakom N, Loeffler W, Koch U, Jung G (1986) Fengycin – a novel antifungal lipopeptide antibiotic produced by *Bacillus subtilis* F 29-3. J Antibiot (Tokyo) **39**:888-901

Volpon L, Besson F, Lancelin JM (2000). NMR structure of antibiotics plipastatins A and B from *Bacillus subtilis* inhibitors of phospholipase A2. FEB Letters **485**:76-80

Volpon L, Tsan P, Majer Z, Vass E, Hollosi M, Noguera V, Lancelin JM, Besson F (2007) NMR structure determination of a synthetic analogue of bacillomycin Lc reveals the strategic role of L-Asn1 in the natural iturinic antibiotics. Spectrochim Acta A Mol Biomol Spectrosc **67**:1374-1381

von Döhren H, Keller U, Vater J, Zocher R (1997) Multifunctional peptide synthetases. Chem Rev **97**:2675-2706

Wakayama S, Ishikawa F, Oishi K (1984) Mycocerein, a novel antifungal peptide antibiotic produced by *Bacillus cereus*. Antimicrob Agents Chemother **26**:939-940

Walsh CT, Gehring AM, Weinreb PH, Luis EN, Flugel RS (1997) Posttranslational modification of polyketide and nonribosomal peptide synthetases. Curr Opin Chem Biol **1**:309-315

Walton RB, Woodruff HB (1949) A crystalline antifungal agent, mycosubtilin, isolated from subtilin broth 1. J Clin Invest **28**:924-926

Wang J, Liu J, Wang X, Yao J, Yu Z (2004) Application of electrospray ionization mass spectrometry in rapid typing of fengycin homologues produced by *Bacillus subtilis*. Lett Appl Microbiol **39**:98-102

Wang J, Liu J, Chen H, Yao J (2007) Characterization of *Fusarium graminearum* inhibitory lipopeptide from *Bacillus subtilis* IB. Appl Microbiol Biotechnol **76**:889-894

Wei YH, Wang LC, Chen WC, Chen SY (2010) Production and characterization of fengycin by indigenous *Bacillus subtilis* F29-3 originating from a potato farm. J Mol Sci **11**:4526-4538

Wu CY, Chen CL, Lee YH, Cheng YC, Wu YC, Shu HY, Gotz F, Liu ST (2007) Nonribosomal synthesis of fengycin on an enzyme complex formed by fengycin synthetases. J Biol Chem **282**: 5608-5616

Yu GY, Sinclair JB, Hartman GL, Bertagnolli BL (2002) Production of iturin A by *Bacillus amyloliquefaciens* suppressing *Rhizoctonia solani*. Soil Biol Biochem **34**:955-963

Zuber P, Losick R **(**1983) Use of a *lacZ* fusion to study the role of the *spo0* genes of *Bacillus subtilis* in developmental regulation. Cell **35**:275-283

**List of indexes**

## Index 1: International conferences participation abstracts

1st Bio-Processing and Application of Microbial Biotechnology in Agriculture, Cairo, Egypt

### Plipastatin operon regulation in *Bacillus subtilis* 168 derivatives

Walaa Hussein, Max Béchet, Frédérique Gancel and Philippe Jacques
Laboratoire ProBioGEM, Polytech'Lille, Université Lille 1 Sciences et Technologies,
F-59655 Villeneuve d'Ascq Cedex, France

*Bacillus subtilis* 168 is a Sfp- (phosphopantetheine-transferase deficient) strain harbouring the surfactin (Srf) and plipastatin (Pps) operons. These large operons (of lengths about 26 and 38 kb, respectively) code for non-ribosomally synthesized lipopeptides with biosurfactant and antifungal as main properties. The syntheses of these lipopeptides are driven by large multienzymatic proteins called the Non-Ribosomal Peptides Synthetases (NRPS). Plipastatin operon expression was studied in derivatives of this strain by analyzing the strength of the native *pps* promoter using a gene reporter system. For this purpose, the plipastatin promoter was inserted upstream from the *lacZ* gene cassette of pDG1661 plasmid and, after transformation of *B. subtilis* 168, ß-galactosidase activity was measured. Under different growth conditions, a very low ß-gal activity could be detected. Taking into account this weakness of the plipastatin promoter, a new 168 derivative (BBG140) was obtained by replacement of the native *pps* promoter of *B. subtilis* BBG111 (*B. subtilis* 168 Sfp+) by a constitutive promoter originating from the replication gene *repU* of plasmid pUB110, associated to a neomycin resistance (*neo*) gene cassette. This modification did not change significantly the plipastatin production but seems to influence other phenotypic traits of the strain such as growth and surfactin production kinetics. The surfactin operon of BBG140 was then interrupted to obtain the new derivative BBG143, a plipastatin monoproducer. In addition, antagonistic and haemolytic activities for both strains BBG140 and BBG143 were tested.

16th International Conference on Gram Positive Microorganisms, Montecatini, Italy

**The power of bionformatics strategy to decrypt NonRibosomal Peptides biosynthesis in *Bacilli***

Valérie Leclère[1], Maude Pupin[2], Walaa Hussein[1], Max Béchet[1], Frédérique Gancel[1], Marlène Chollet[1], Ségolène Caboche[2] and Philippe Jacques[1]

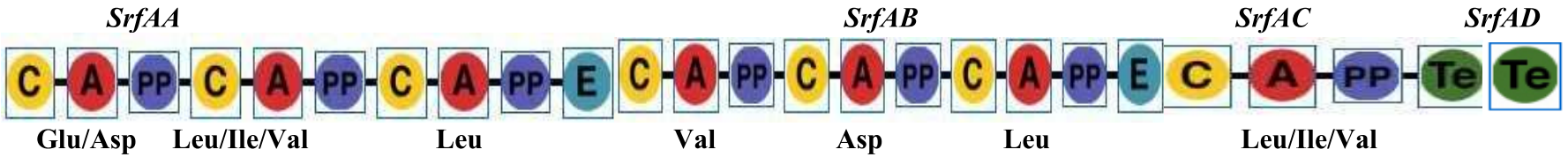
*[1]ProBioGEM, UPRES EA 1026, Polytech'Lille/IUTA, Avenue Langevin, Université de Lille Nord de France, Sciences et Technologies, F-59655 Villeneuve d'Ascq cedex, France*
*[2] Laboratoire d'Informatique Fondamentale de Lille (LIFL), UMR8020 CNRS, INRIA, Université de Lille Nord de France, Sciences et Technologies F-59655 Villeneuve d'Ascq cedex, France*

NonRibosomal Peptide Synthetases (NRPS) are multifunctional enzymes organized in sets of catalytic domains which constitute modules containing the information needed to complete an elongation step in peptide biosynthesis. The main catalytic functions are responsible for the activation of an amino acid residue (adenylation domain), the transfer of the corresponding adenylate to the enzyme-bound 4'-phosphopantetheinyl cofactor (peptidyl carrier protein domain) and the peptide bond formation (condensation domain). Additional domains can lead to modification of the substrates if required in the peptide synthesis. A thioesterase domain is usually present in final position to ensure the cleavage of the thioester bond between the nascent peptide and the last PCP-domain and, in several cases, to cyclise the peptide. Several bioinformatics tools were developed during the past decade to predict, from the genome sequence, the modular organization of the encoded NRPS and the potential structure of the synthesized peptidic product. In addition, for the first time, the peptides synthesized or putatively synthesized by this original pathway were gathered in a database called NORINE (1). In this work, analyses were performed on the genomes of different *Bacillus* strains in order to detect NRPS genes and to predict their products. Two new bioinformatics tools developed by our team were used to compare the structure of potentially new NRPs to all the existing ones (2) and to predict their biological activity (3). This work led to the evaluation of the potential of NonRibosomal Peptides (NRPs) production of the different strains, the discovery of biosynthetic pathways of putative NRPs, the correlation of mis-annotated genes with their products and the set up of new potential NRPs produced by these strains.

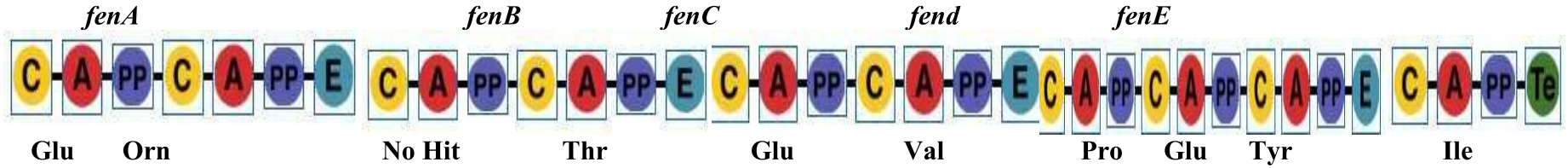
## Index 2: Bioinformatic analyses of *Bacillus* spp.

### Detected NRPS genes in *Bacillus amyloliquefaciens* FZB42

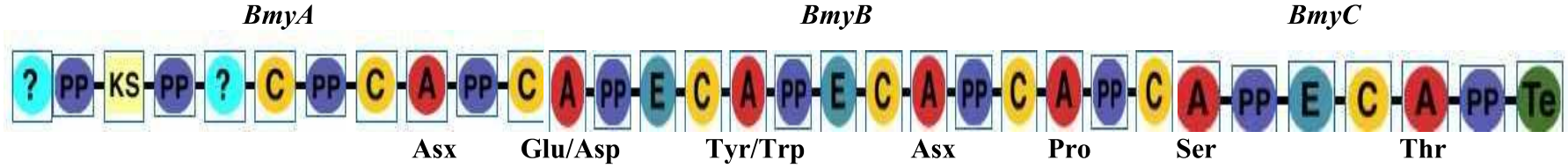
| Product Name | Start | End | Strand | Length (a.a) | Accession | GeneID | Locus |
|---|---|---|---|---|---|---|---|
| SrfAA | 342618 | 353372 | + | 3584 | YP_001419995.1 | 5462766 | *srfAA* |
| SrfAB | 353394 | 364154 | + | 3586 | YP_001419996.1 | 5462767 | *srfAB* |
| ComS | 356503 | 356667 | + | 54 | YP_001419997.1 | 5460070 | *comS* |
| SrfAC | 364189 | 368025 | + | 1278 | YP_001419998.1 | 5460071 | *srfAC* |
| SrfAD | 368045 | 368776 | + | 243 | YP_001419999.1 | 5460072 | *srfAD* |
| FenE | 1931328 | 1935131 | - | 1267 | YP_001421436.1 | 5460038 | *fenE* |
| FenD | 1935150 | 1945925 | - | 3591 | YP_001421437.1 | 5459676 | *fenD* |
| FenC | 1945951 | 1953600 | - | 2549 | YP_001421438.1 | 5459677 | *fenC* |
| FenB | 1953616 | 1961313 | - | 2565 | YP_001421439.1 | 5459678 | *fenB* |
| FenA | 1961339 | 1968997 | - | 2552 | YP_001421440.1 | 5459679 | *fenA* |
| BmyC | 1871172 | 1879031 | - | 2619 | YP_001421410.1 | 5461147 | *bmyC* |
| BmyB | 1879116 | 1895207 | - | 5363 | YP_001421411.1 | 5461148 | *bmyB* |
| BmyA | 1895252 | 1907200 | - | 3982 | YP_001421412.1 | 5461149 | *bmyA* |
| BmyD | 1907220 | 1908422 | - | 400 | YP_001421413.1 | 5462654 | *bmyD* |
| BaeJ | 1708756 | 1723704 | + | 4982 | YP_001421292.1 | 5462451 | baeJ |
| BaeL | 1723706 | 1737133 | + | 4475 | YP_001421293.1 | 5461785 | baeL |
| BaeM | 1737151 | 1747686 | + | 3511 | YP_001421294.1 | 5461786 | baeM |
| BaeN | 1747676 | 1763977 | + | 5433 | YP_001421295.1 | 5461787 | baeN |




Surfactin synthetases genes prediction from *B. amyloliquefaciens* FZB42




Fengyin synthetases genes prediction from *B. amyloliquefaciens* FZB42

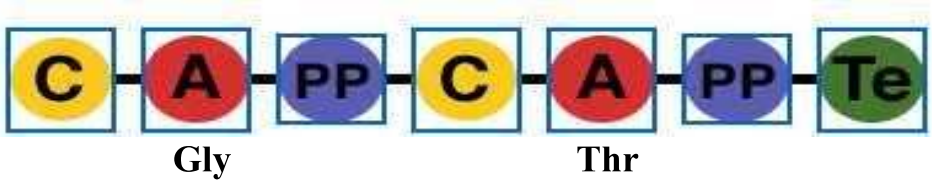



Bacillomycin synthetases genes prediction from *B. amyloliquefaciens* FZB42


Bacillaene synthetases genes prediction from *B. amyloliquefaciens* FZB42

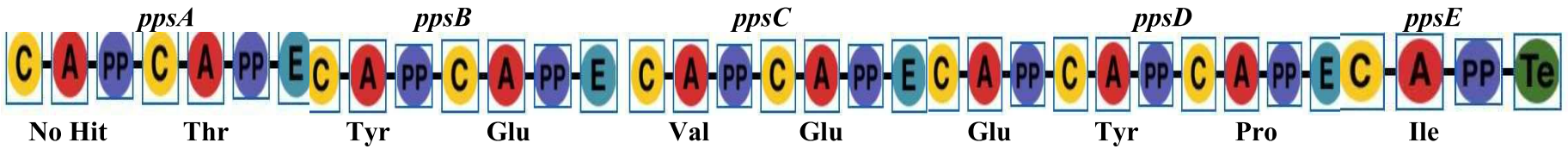
## Detected NRPS genes in *Bacillus atrophaeus* 1942

| Product Name | Start | End | Strand | Length (a.a) | Accession | GeneID |
|---|---|---|---|---|---|---|
| siderophore 2,3-dihydroxybenzoate-glycine-threonine trimeric ester bacillibactin synthetase | 2705675 | 2712859 | - | 2394 | YP_003974613.1 | 9892747 |
| isochorismatase | 2712874 | 2713809 | - | 311 | YP_003974614.1 | 9892748 |
| bacitracin synthetase 1 | 3832904 | 3839914 | + | 2336 | YP_003975767.1 | 9893932 |
| Mycosubtilin synthase subunit C | 1514757 | 1522565 | - | 2602 | YP_003973350.1 | 9891455 |
| Mycosubtilin synthase subunit B | 1522645 | 1538736 | - | 5363 | YP_003973351.1 | 9891456 |
| Mycosubtilin synthase subunit A | 1538782 | 1550724 | - | 3980 | YP_003973352.1 | 9891457 |
| plipastatin synthetase | 1450287 | 1454114 | - | 1275 | YP_003973321.1 | 9891425 |
| plipastatin synthetase | 1454143 | 1464969 | - | 3608 | YP_003973322.1 | 9891426 |
| plipastatin synthetase | 1464995 | 1472653 | - | 2552 | YP_003973323.1 | 9891427 |
| plipastatin synthetase | 1472672 | 1480405 | - | 2577 | YP_003973324.1 | 9891428 |
| plipastatin synthetase | 1480429 | 1488111 | - | 2560 | YP_003973325.1 | 9891429 |
| surfactin synthetase | 4010450 | 4021222 | + | 3590 | YP_003975931.1 | 9894096 |
| surfactin synthetase | 4021245 | 4032014 | + | 3589 | YP_003975932.1 | 9894097 |
| surfactin synthetase | 4032033 | 4035863 | + | 1276 | YP_003975933.1 | 9894098 |
| surfactin synthetase | 4035891 | 4036625 | + | 244 | YP_003975934.1 | 9894099 |

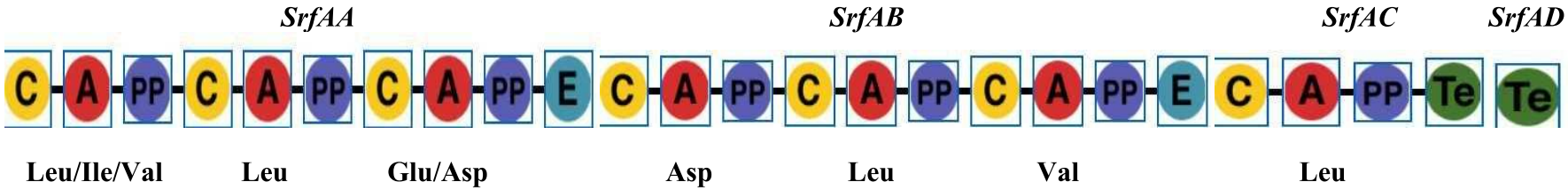



Bacillibactin siderophore NRPS prediction from *Bacillus atrophaeus* 1942


Mycosubtilin siderophore NRPS prediction from *Bacillus atrophaeus* 1942




Plipastatin NRPS prediction from *Bacillus atrophaeus* 1942

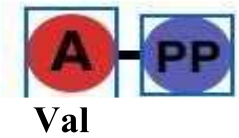



Surfactin NRPS prediction from *Bacillus atrophaeus* 1942

## Detected NRPS genes in *Bacillus cereus* ATCC 14579

| Product Name | Start | End | Strand | Length (a.a) | Accession | GeneID | Locus |
|---|---|---|---|---|---|---|---|
| peptide synthetase | 410851 | 413832 | + | 993 | NP_830262.1 | 1202776 | - |
| peptide synthetase | 2396036 | 2399371 | + | 1111 | NP_832214.1 | 1204801 | - |
| peptide synthetase | 2399374 | 2405844 | + | 2156 | NP_832215.1 | 1204802 | - |
| peptide synthetase | 2405841 | 2406539 | + | 232 | NP_832216.1 | 1204803 | - |
| peptide synthetase | 2406529 | 2408079 | + | 516 | NP_832217.1 | 1204804 | - |
| peptide synthetase | 2410451 | 2420725 | + | 3424 | NP_832218.1 | 1204805 | - |
| 4'-phosphopantetheinyl transferase | 2421047 | 2421751 | + | 234 | NP_832219.1 | 1204806 | - |

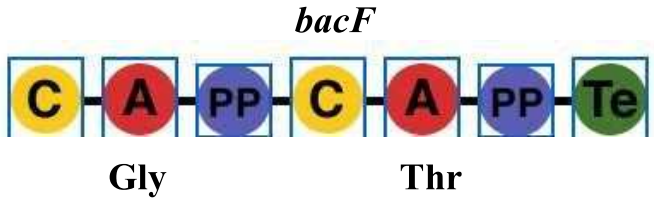

NRPS prediction of truncated kurstakin synthetases detected from *Bacillus cereus* ATCC 14579



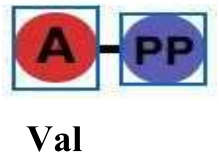

NRPS prediction of unknown synthetase gene detected from *Bacillus cereus* ATCC 14579

## Detected NRPS genes in *Bacillus cereus* B4264

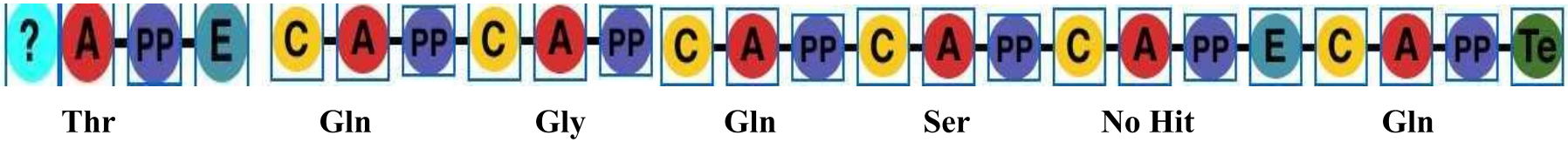
| Product Name | Start | End | Strand | Length (a.a) | Accession | GeneID | Locus |
|---|---|---|---|---|---|---|---|
| isochorismatase | 2235831 | 2236724 | + | 297 | YP_002367053.1 | 7101477 | *bacB* |
| nonribosomal peptide synthetase DhbF | 2236758 | 2243915 | + | 2385 | YP_002367054.1 | 7097461 | *bacF* |
| linear gramicidin synthetase subunit C | 428467 | 432048 | + | 1193 | YP_002365222.1 | 7097392 | - |
| linear gramicidin synthetase subunit C | 2361088 | 2365644 | + | 1518 | YP_002367177.1 | 7096783 | - |
| linear gramicidin synthetase subunit D | 2365647 | 2372117 | + | 2156 | YP_002367178.1 | 7097346 | - |
| bacitracin synthetase 1 | 2372114 | 2386996 | + | 4960 | YP_002367179.1 | 7101407 | - |
| 4'-phosphopantetheinyl transferase family protein | 2387438 | 2388064 | + | 208 | YP_002367180.1 | 7098457 | - |




NRPS prediction of the bacillibactin ( nonribosomal peptide synthetase DhbF) detected from *Bacillus cereus* B4264



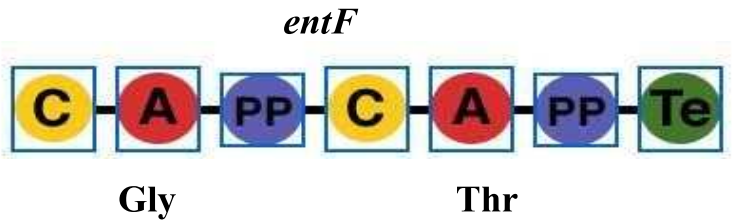

NRPS prediction of an unknown synthetase gene (linear gramicidin synthetase subunit C) detected from *Bacillus cereus* B4264



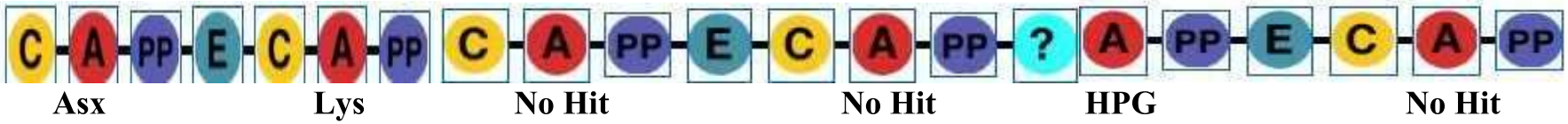

NRPS prediction of the kurstain synthetases (gramicidin synthetases subunits C, D and bacitracin synthetase 1) detected from *Bacillus cereus* B4264

**Detected NRPS genes in *Bacillus cereus* E33L**

| Product Name | Start | End | Strand | Length (a.a) | Accession | GeneID | Locus |
|---|---|---|---|---|---|---|---|
| multifunctional nonribosomal peptide synthetase (mycosubtilin synthetase) | 1814544 | 1822175 | + | 2543 | YP_083310.1 | 3022893 | *dhbF* |
| multifunctional nonribosomal peptide synthetase (mycosubtilin synthetase) | 1822244 | 1831378 | + | 3044 | YP_083311.1 | 3025745 | *dhbF* |
| multifunctional nonribosomal peptide synthetase (mycosubtilin synthetase) | 1831431 | 1838318 | + | 2295 | YP_083312.1 | 3027044 | *dhbF* |
| isochorismatase | 2253873 | 2254760 | + | 295 | YP_083721.1 | 3024313 | *entB* |
| nonribosomal peptide synthetase | 2254792 | 2261949 | + | 2385 | YP_083722.1 | 3024779 | *entF* |



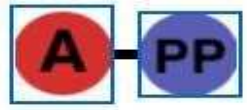

NRPS prediction of the bacillibactin (nonribosomal peptide synthetase) detected from *Bacillus cereus* E33L




NRPS prediction of new synthetases genes (mycosubtilin synthetases) detected from *Bacillus cereus* E33L

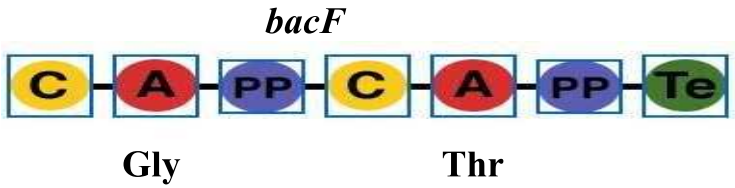
**Detected NRPS genes in *Bacillus cereus* AH178**

| Product Name | Start | End | Strand | Length (a.a) | Accession | GeneID | Locus |
|---|---|---|---|---|---|---|---|
| linear gramicidin synthetase subunit B, putative | 388335 | 391523 | + | 1062 | YP_002336454.1 | 7077011 | - |
| nonribosomal peptide synthetase DhbF | 2305916 | 2313073 | + | 2385 | YP_002338425.1 | 7073158 | bacF |




**No Hit**

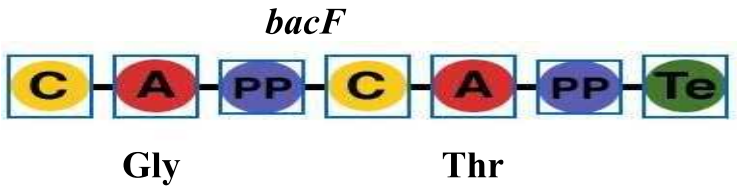
Unknown synthetase gene (linear gramicidin synthetase subunit B, putative)




Bacillibactin predicted by PKS-NRPS website from *B. cereus* AH178

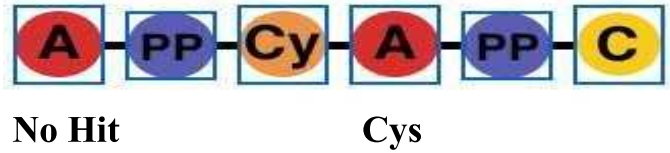
## Detected NRPS genes in *Bacillus cereus* G9842

| Product Name | Start | End | Strand | Length (a.a) | Accession | GeneID | Locus |
|---|---|---|---|---|---|---|---|
| isochorismatase | 2194542 | 2195435 | + | 297 | YP_002445732.1 | 7183375 | *bacB* |
| nonribosomal peptide synthetase DhbF | 2195469 | 2202626 | + | 2385 | YP_002445733.1 | 7184080 | *bacF* |
| bacitracin synthetase 1 | 2629146 | 2636162 | - | 2338 | YP_002446163.1 | 7186245 | - |
| linear gramicidin synthetase subunit C | 2307022 | 2311578 | + | 1518 | YP_002445845.1 | 7183669 | - |
| non-ribosomal peptide synthase | 2311581 | 2318051 | + | 2156 | YP_002445846.1 | 7182828 | - |
| bacitracin synthetase 1 | 2318048 | 2332930 | + | 4960 | YP_002445847.1 | 7182289 | - |
| 4'-phosphopantetheinyl transferase Sfp | 2333371 | 2333997 | + | 208 | YP_002445848.1 | 7183004 | - |
| linear gramicidin synthetase subunit C | 406592 | 410173 | + | 1193 | YP_002443907.1 | 7183087 | - |
| linear gramicidin synthetase subunit D | 893791 | 899955 | + | 2054 | YP_002444361.1 | 7182896 | - |

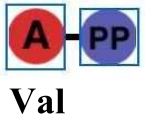



NRPS prediction of bacillibactin (nonribosomal peptide synthetase DhbF) from *Bacillus cereus* G9842

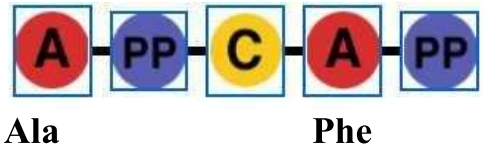

NRPS prediction of kurstain genes from *Bacillus cereus* G9842




NRPS prediction of truncated bacitracin synthetase gene (bacitracin synthetase 1) from *Bacillus cereus* G9842



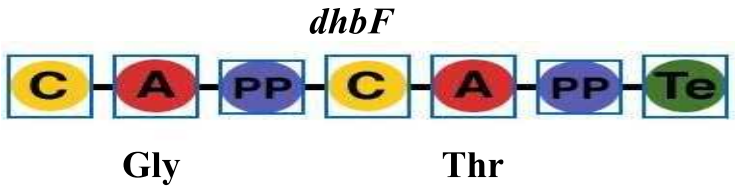

NRPS prediction of unknown synthetase gene (linear gramicidin synthetase subunit C) from *Bacillus cereus* G9842




NRPS prediction of unknown synthetase gene(linear gramicidin synthetase subunit D) from *Bacillus cereus* G9842

## Detected NRPS genes in *Bacillus licheniformis* ATCC 14580

| Product Name | Start | End | Strand | Length (a.a) | Accession | GeneID | Locus |
|---|---|---|---|---|---|---|---|
| lichenysin synthetase A | 378877 | 389619 | + | 3580 | YP_077640.1 | 3029783 | *lchAA* |
| lichenysin synthetase B | 389639 | 400405 | + | 3588 | YP_077641.1 | 3029845 | *lchAB* |
| lichenysin synthetase C | 400463 | 404311 | + | 1282 | YP_077642.1 | 3029681 | *lchAC* |
| thioesterase LchAD | 404315 | 405055 | + | 246 | YP_077643.1 | 3029686 | *lchAD* |
| nonribosomal peptide synthetase DhbF | 3718288 | 3725445 | - | 2385 | YP_080976.1 | 3027963 | *dhbF* |
| Isochrismatase | 3725460 | 3726386 | - | 308 | YP_080977.1 | 3028505 | *dhbB* |




NRPS prediction of bacillibactin gene (nonribosomal peptide synthetase DhbF) from *Bacillus licheniformis* ATCC 14580


NRPS prediction of lichenysin synthetases genes from *Bacillus licheniformis* ATCC 14580

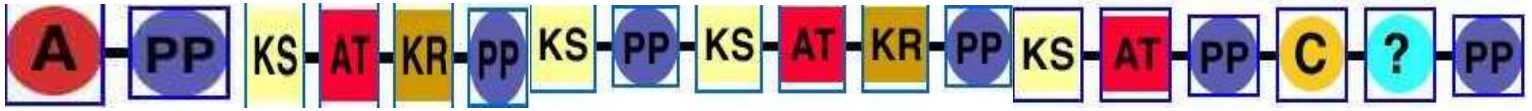
## Detected NRPS genes in *Bacillus pumilus* SAFR-032

| Product Name | Start | End | Strand | Length (a.a) | Accession | GeneID | Locus |
|---|---|---|---|---|---|---|---|
| nonribosomal peptide synthetase subunit | 669948 | 674360 | + | 1470 | YP_001485881.1 | 5619877 | - |
| beta-lactamase | 674357 | 675898 | + | 513 | YP_001485882.1 | 5619878 | *pbpX* |
| nonribosomal peptide synthetase subunit | 690025 | 692694 | + | 889 | YP_001485888.1 | 5619884 | - |
| polyketide synthase subunit | 692687 | 697192 | + | 1501 | YP_001485889.1 | 5619885 | - |
| polyketide synthase subunit | 697193 | 704443 | + | 2416 | YP_001485890.1 | 5619886 | - |
| polyketide synthase subunit | 704436 | 710846 | + | 2136 | YP_001485891.1 | 5619887 | - |
| nonribosomal peptide synthetase subunit | 318695 | 329407 | + | 3570 | YP_001485576.1 | 5619546 | - |
| nonribosomal peptide synthetase subunit | 329428 | 340116 | + | 3562 | YP_001485577.1 | 5619547 | - |
| nonribosomal peptide synthetase subunit | 340161 | 343994 | + | 1277 | YP_001485578.1 | 5619548 | - |
| nonribosomal peptide synthetase subunit | 344058 | 354188 | + | 3376 | YP_001485579.1 | 5619549 | - |
| nonribosomal peptide synthetase subunit | 354202 | 362397 | + | 2731 | YP_001485580.1 | 5619550 | - |


NRPS prediction of pumilacidin synthetases genes detected from *Bacillus pumilus* SAFR-032

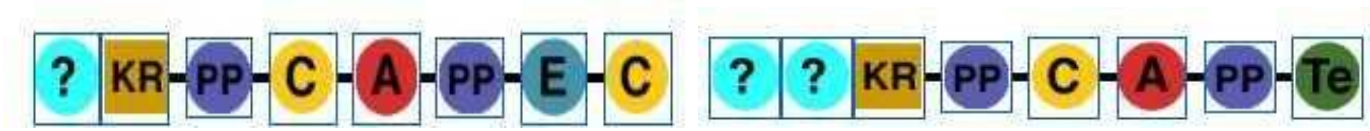

**Asx**

NRPS prediction of unknown synthetase gene for *Bacillus pumilus* SAFR-032




**No Hit**

NRPS prediction of four unknown synthetases NRPS-PKS from *Bacillus pumilus* SAFR-032

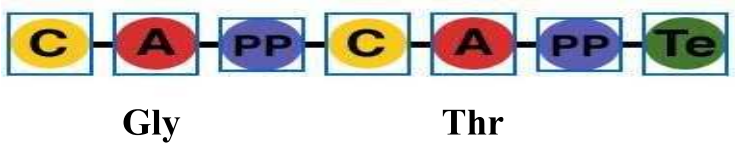



**No Hit** **Ile**

NRPS prediction of unknown PKS-NRPS molecule detected from *Bacillus pumilus* SAFR-032

**Detected NRPS genes in *Bacillus subtilis* subsp. *natto* BEST195**

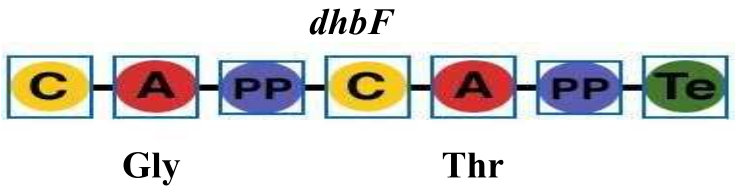
| Product Name | Start | End | Strand | Locus |
|---|---|---|---|---|
| **Surfactin synthetase** | **374033** | **383585** | + | ***srfAA*** |
| **Surfactin synthetase** | **383598** | **394827** | + | ***srfAB*** |
| **Surfactin synthetase** | **394897** | **398691** | + | ***srfAC*** |
| **Surfactin synthetase** | **398720** | **399448** | + | ***srfAD*** |
| **Plipastatin synthetase** | **1963356** | **1966521** | - | ***ppsE*** |
| **Plipastatin synthetase** | **1966795** | **1974497** | - | ***ppsAD*** |
| **siderophore 2,3-dihydroxybenzoate-glycine-threonine trimeric ester bacillibactin synthetase** | **3043147** | **3050283** | - | - |
| **enterobactin synthetase component E** | **3052917** | **3054113** | - | - |



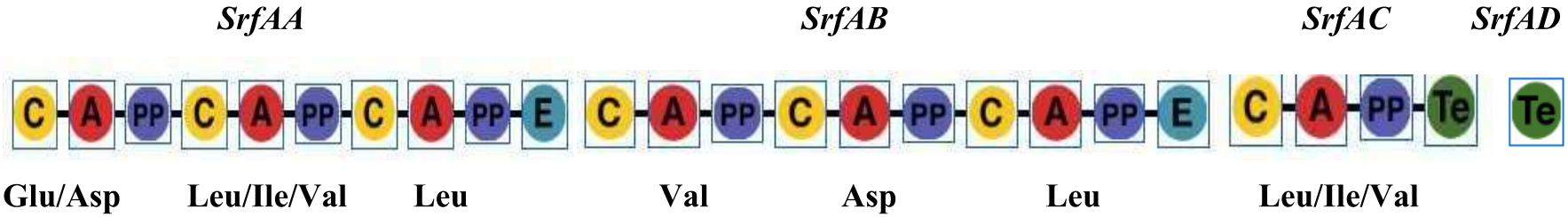

NRPS prediction of bacillibactin synthetase gene from *Bacillus subtilis* subsp. *natto* BEST195

## Detected NRPS genes in *Bacillus subtilis* subsp. *spizizenii* str. W23

| Product Name | Start | End | Strand | Length (a.a) | Accession | GeneID | Locus |
|---|---|---|---|---|---|---|---|
| siderophore 2,3-dihydroxybenzoate-glycine-threonine trimeric ester bacillibactin synthetase | 3063946 | 3071085 | - | 2379 | YP_003867454.1 | 9724116 | *dhbF* |
| isochorismatase | 3071105 | 3072040 | - | 311 | YP_003867455.1 | 9724117 | *dhbB* |
| surfactin synthetase | 364107 | 374870 | + | 3587 | YP_003864726.1 | 9721321 | *srfAA* |
| surfactin synthetase | 374883 | 385634 | + | 3583 | YP_003864727.1 | 9721322 | *srfAB* |
| regulator of genetic competence | 378019 | 378159 | + | 46 | YP_003864728.1 | 9721323 | *comS* |
| surfactin synthetase | 385671 | 389498 | + | 1275 | YP_003864729.1 | 9721324 | *srfAC* |
| surfactin synthetase | 389526 | 390254 | + | 242 | YP_003864730.1 | 9721325 | *srfAD* |
| Mycosubtilin synthase subunit C | 1911931 | 1919763 | - | 2610 | YP_003866243.1 | 9722874 | *mycC* |
| Mycosubtilin synthase subunit B | 1919861 | 1935958 | - | 5365 | YP_003866244.1 | 9722875 | *mycB* |
| Mycosubtilin synthase subunit A | 1936015 | 1947930 | - | 3971 | YP_003866245.1 | 9722876 | *mycA* |
| Malonyl CoA-acyl carrier protein transacylase | 1947951 | 1949153 | - | 400 | YP_003866246.1 | 9722877 | *fenF* |
| polyketide synthase of type I | 1750201 | 1765341 | + | 5046 | YP_003866122.1 | 9722753 | *pksJ* |
| polyketide synthase of type I | 1765343 | 1779013 | + | 4556 | YP_003866123.1 | 9722754 | *pksL* |
| polyketide synthase | 1779029 | 1791811 | + | 4260 | YP_003866124.1 | 9722755 | *pksM* |
| polyketide synthase of type I | 1791801 | 1808345 | + | 5514 | YP_003866125.1 | 9722756 | *pksN* |




NRPS prediction of bacillibactin from *Bacillus subtilis* subsp. *spizizenii* str. W23



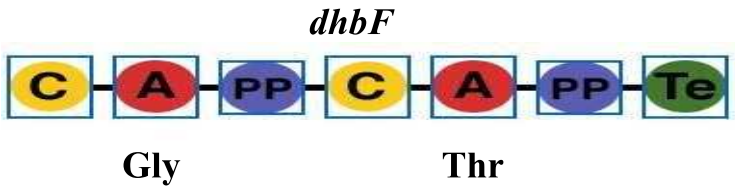

NRPS prediction of surfactin from *Bacillus subtilis* subsp. *spizizenii* str. W23


NRPS prediction of three mycosubtilin synthetases genes from *Bacillus subtilis* subsp. *spizizenii* str. W23

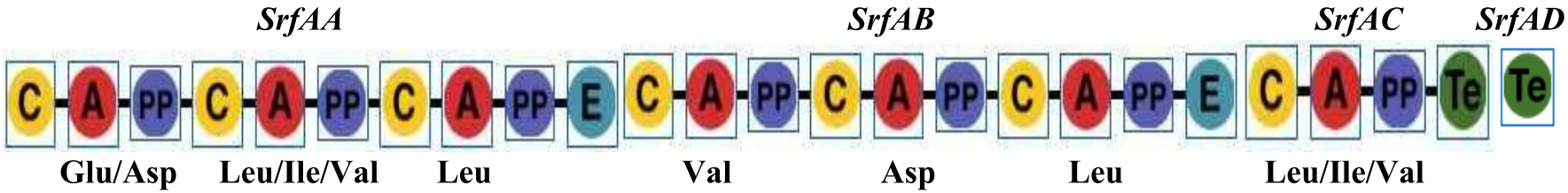

NRPS prediction of new four syynthetases genes from *Bacillus subtilis* subsp. *spizizenii* str. W23

## Detected NRPS genes in *Bacillus subtilis* subsp. *subtilis* str. 168

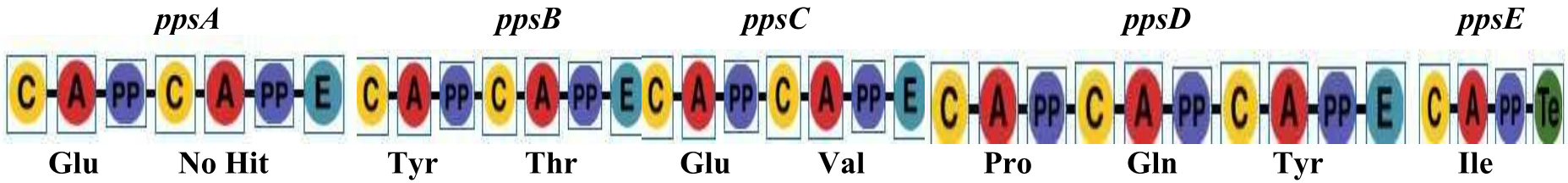
| Product Name | Start | End | Strand | Length (a.a) | Accession | GeneID | Locus |
|---|---|---|---|---|---|---|---|
| surfactin synthetase | 376968 | 387731 | + | 3587 | NP_388230.2 | 938306 | *srfAA* |
| surfactin synthetase | 387744 | 398495 | + | 3583 | NP_388231.2 | 938303 | *srfAB* |
| regulator of genetic competence | 390880 | 391020 | + | 46 | NP_388232.1 | 938310 | *comS* |
| surfactin synthetase | 398532 | 402359 | + | 1275 | NP_388233.2 | 938308 | *srfAC* |
| surfactin synthetase | 402388 | 403116 | + | 242 | NP_388234.1 | 938300 | *srfAD* |
| plipastatin synthetase | 1960198 | 1964037 | - | 1279 | NP_389712.1 | 939956 | *ppsE* |
| plipastatin synthetase | 1964045 | 1974856 | - | 3603 | NP_389713.2 | 940013 | *ppsD* |
| plipastatin synthetase | 1974881 | 1982548 | - | 2555 | NP_389714.1 | 940102 | *ppsC* |
| plipastatin synthetase | 1982565 | 1990247 | - | 2560 | NP_389715.1 | 939993 | *ppsB* |
| plipastatin synthetase | 1990272 | 1997957 | - | 2561 | NP_389716.2 | 939983 | *ppsA* |
| siderophore 2,3-dihydroxybenzoate-glycine-threonine trimeric ester bacillibactin synthetase | 3280519 | 3287655 | - | 2378 | NP_391076.3 | 936569 | *dhbF* |
| isochorismatase | 3287675 | 3288613 | - | 312 | NP_391077.1 | 936582 | *dhbB* |




NRPS prediction of bacillibactin from *Bacillus subtilis* subsp. *subtilis* str. 168




NRPS prediction of four surfactin synthetases from *Bacillus subtilis* subsp. *subtilis* str. 168

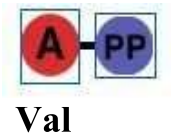



NRPS prediction of five plipastatin synthetases genes from *Bacillus subtilis* subsp. *subtilis* str. 168

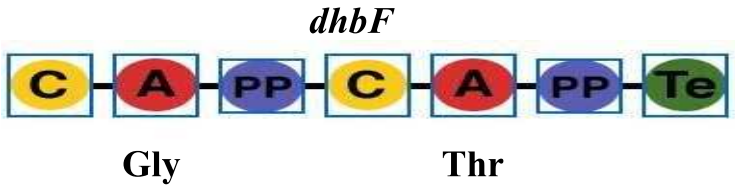
## Detected NRPS genes in *Bacillus thuringiensis* BMB171

| Product Name | Start | End | Strand | Length (a.a) | Accession | GeneID | Locus |
|---|---|---|---|---|---|---|---|
| peptide synthetase | 2346120 | 2350676 | + | 1518 | YP_003664738.1 | 9193247 | - |
| peptide synthetase | 2350679 | 2357149 | + | 2156 | YP_003664739.1 | 9193248 | - |
| peptide synthetase | 2357146 | 2372028 | + | 4960 | YP_003664740.1 | 9193249 | - |
| peptide synthetase | 430607 | 434188 | + | 1193 | YP_003662902.1 | 9191405 | - |
| glycine-AMP ligase | 2208917 | 2216074 | + | 2385 | YP_003664610.1 | 9193118 | *bacF* |
| Hypothetical protein | | | | 2540 | YP_003667784.1 | | - |
| Non ribosomal peptide synthetase C | | | | 3484 | YP_003667785.1 | | - |

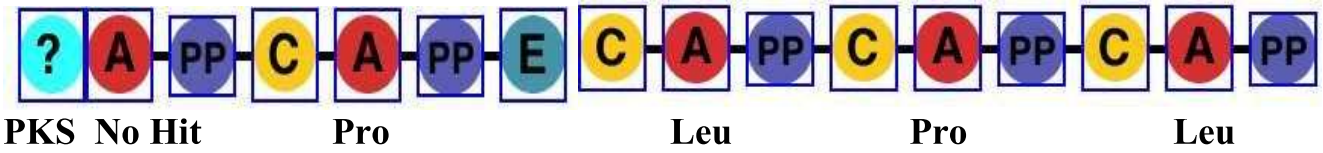

NRPS prediction of three kurstakin synthetases genes detected by PKS-NRPS website from *Bacillus thuringiensis* BMB171




NRPS prediction of unknown synthetase gene detected by PKS-NRPS website from *Bacillus thuringiensis* BMB171



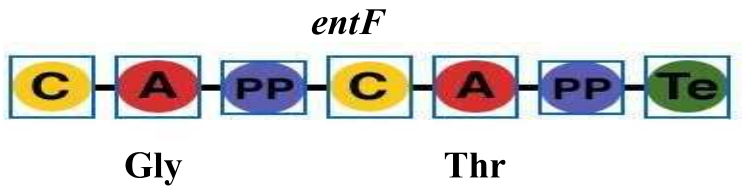

NRPS prediction of bacillibactin detected by PKS-NRPS website from *Bacillus thuringiensis* BMB171




NRPS prediction of two unknown synthetases genes by PKS-NRPS website from *Bacillus thuringiensis* BMB171

## Detected NRPS genes in *Bacillus thuringiensis* serovar *konkukian* str. 97-27

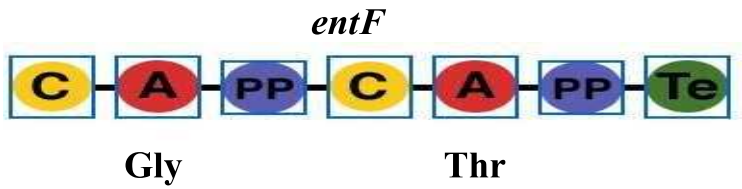
| Product Name | Start | End | Strand | Length (a.a) | Accession | GeneID | Locus |
|---|---|---|---|---|---|---|---|
| isochorismatase | 2226837 | 2227724 | + | 295 | YP_036473.1 | 2855946 | *entB* |
| nonribosomal peptide synthetase | 2227756 | 2234913 | + | 2385 | YP_036474.1 | 2857779 | *entF* |




NRPS prediction of bacillibactin synthetase gene detected by PKS-NRPS website for *Bacillus thuringiensis* serovar *konkukian* str. 97-27

## Detected NRPS genes in *Bacillus thuringiensis* str. Al Hakam

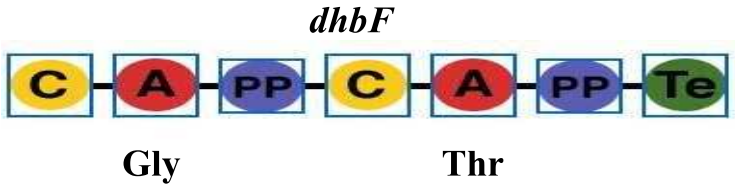
| Product Name | Start | End | Strand | Length (a.a) | Accession | GeneID | Locus |
|---|---|---|---|---|---|---|---|
| isochorismatase | 2261569 | 2262462 | + | 297 | YP_894923.1 | 4546333 | - |
| nonribosomal peptide synthetase | 2262494 | 2269651 | + | 2385 | YP_894924.1 | 4546334 | *entF* |




NRPS prediction of bacillibactin synthetase gene detected by PKS-NRPS website from *Bacillus thuringiensis str.* Al Hakam

**Detected NRPS genes in *Bacillus anthracis* str. 'Ames Ancestor'**

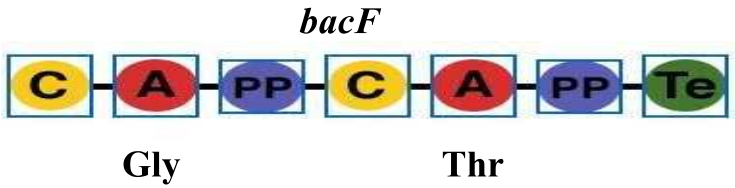
| Product Name | Start | End | Strand | Length (a.a) | Accession | GeneID | Locus |
|---|---|---|---|---|---|---|---|
| isochorismatase | 2205715 | 2206602 | + | 295 | YP_019015.1 | 2816570 | *dhbB* |
| nonribosomal peptide synthetase DhbF | 2206634 | 2213791 | + | 2385 | YP_019016.1 | 2816553 | *dhbF* |




NRPS prediction of bacillibactin synthetase gene detected by PKS-NRPS website from *Bacillus anthracis* str. 'Ames Ancestor'

**Detected NRPS genes in *Bacillus anthracis* str. A0248**

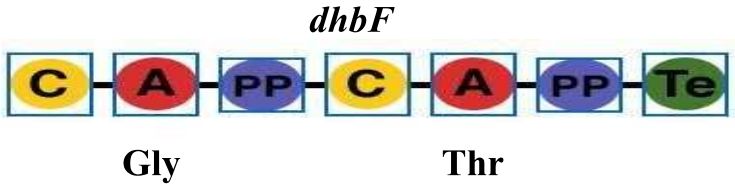
| Product Name | Start | End | Strand | Length (a.a) | Accession | GeneID | Locus |
|---|---|---|---|---|---|---|---|
| isochorismatase | 2205615 | 2206502 | + | 295 | YP_002866708.1 | 7852557 | *bacB* |
| nonribosomal peptide synthetase DhbF | 2206534 | 2213691 | + | 2385 | YP_002866709.1 | 7852575 | *bacF* |




NRPS prediction of bacillibactin synthetase gene detected by PKS-NRPS website from *Bacillus anthracis* str. A0248

## Detected NRPS genes in *Bacillus anthracis* str. Ames

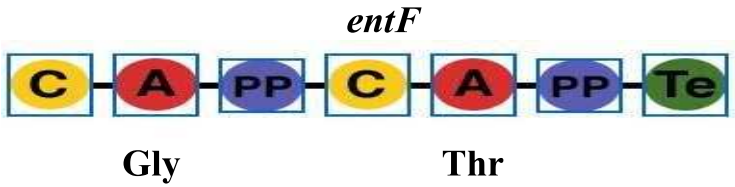
| Product Name | Start | End | Strand | Length (a.a) | Accession | GeneID | Locus |
|---|---|---|---|---|---|---|---|
| isochorismatase | 2205591 | 2206478 | + | 295 | NP_844753.1 | 1083815 | *dhbB* |
| nonribosomal peptide synthetase DhbF | 2206510 | 2213667 | + | 2385 | NP_844754.1 | 1083894 | *dhbF* |




NRPS prediction of bacillibactin synthetase gene detected by PKS-NRPS website from *Bacillus anthracis* str. Ames

## Detected NRPS genes in *Bacillus anthracis* str. Sterne

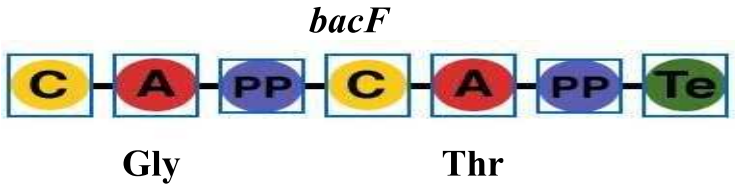
| Product Name | Start | End | Strand | Length | Accession | GeneID | Locus |
|---|---|---|---|---|---|---|---|
| isochorismatase | 2205697 | 2206584 | + | 295 | YP_028468.1 | 2852177 | - |
| nonribosomal peptide synthetase DhbF | 2206616 | 2213773 | + | 2385 | YP_028469.1 | 2853116 | - |




NRPS prediction of bacillibactin synthetase gene detected by PKS-NRPS website from *Bacillus anthracis* str. Sterne

## Detected NRPS genes in *Bacillus cereus* 03BB102

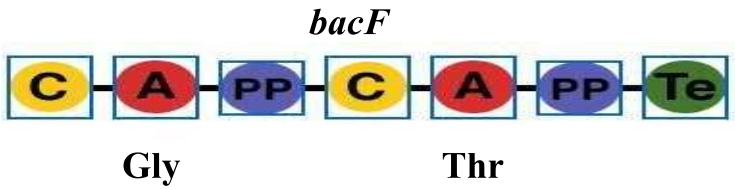
| Product Name | Start | End | Strand | Length (a.a) | Accession | GeneID | Locus |
|---|---|---|---|---|---|---|---|
| isochorismatase | 2268498 | 2269391 | + | 297 | YP_002749712.1 | 7689816 | *bacB* |
| nonribosomal peptide synthetase DhbF | 2269423 | 2276580 | + | 2385 | YP_002749713.1 | 7691247 | *bacF* |




NRPS prediction of bacillibactin synthetase gene detected by PKS-NRPS website from *Bacillus cereus* 03BB102

## Detected NRPS genes in *Bacillus cereus* AH187

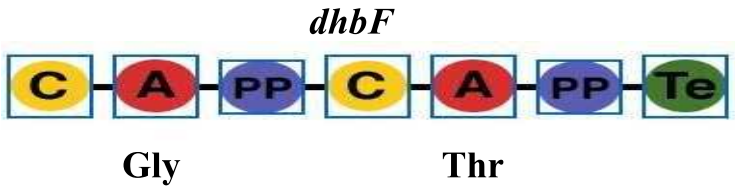
| Product Name | Start | End | Strand | Length (a.a) | Accession | GeneID | Locus |
|---|---|---|---|---|---|---|---|
| isochorismatase | 2304997 | 2305884 | + | 295 | YP_002338424.1 | 7076315 | *bacB* |
| nonribosomal peptide synthetase DhbF | 2305916 | 2313073 | + | 2385 | YP_002338425.1 | 7073158 | *bacF* |




NRPS prediction of bacillibactin synthetase gene detected by PKS-NRPS website from *Bacillus cereus* AH187

## Detected NRPS genes in *Bacillus cereus* ATCC 10987

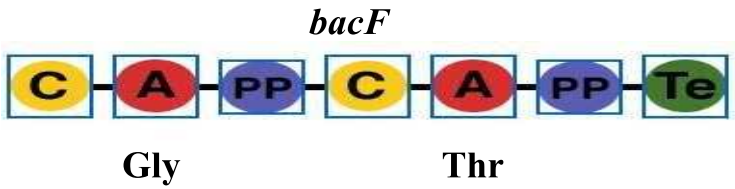
| Product Name | Start | End | Strand | Length (a.a) | Accession | GeneID | Locus |
|---|---|---|---|---|---|---|---|
| isochorismatase | 2288736 | 2289629 | + | 297 | NP_978711.1 | 2747590 | *dhbB* |
| nonribosomal peptide synthetase DhbF | 2289663 | 2296820 | + | 2385 | NP_978712.1 | 2748620 | *dhbF* |



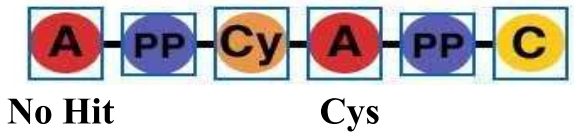

NRPS prediction of bacillibactin synthetase gene detected by PKS-NRPS website from *Bacillus cereus* ATCC 10987

## Detected NRPS genes in *Bacillus anthracis* str. CDC 684

| Product Name | Start | End | Strand | Length (a.a) | Accession | GeneID | Locus |
|---|---|---|---|---|---|---|---|
| nonribosomal peptide synthetase DhbF | 2025419 | 2032576 | - | 2385 | YP_002814826.1 | 7783437 | *bacF* |
| isochorismatase | 2032608 | 2033495 | - | 295 | YP_002814827.1 | 7783527 | *bacB* |
| bacitracin synthetase 1 (BA1) | 1699287 | 1706297 | + | 2336 | YP_002814470.1 | 7787255 | - |



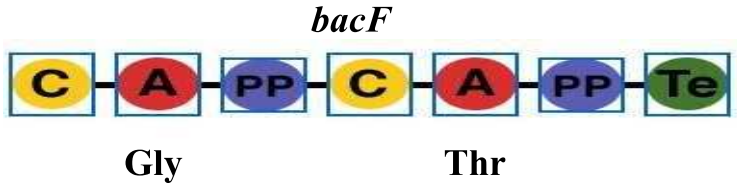

NRPS prediction of bacillibactin synthetase gene detected by PKS-NRPS website from *Bacillus anthracis* str. CDC 684



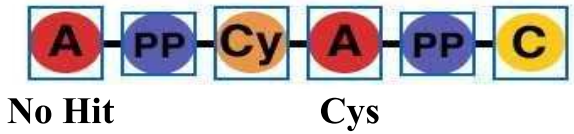

NRPS prediction of trunated bacitracin synthetase gene detected by PKS-NRPS website from *Bacillus anthracis* str. CDC 684

## Detected NRPS genes in *Bacillus cereus* AH820

| Product Name | Start | End | Strand | Length (a.a) | Accession | GeneID | Locus |
|---|---|---|---|---|---|---|---|
| isochorismatase | 2269542 | 2270429 | + | 295 | YP_002451339.1 | 7190404 | *bacB* |
| nonribosomal peptide synthetase DhbF | 2270461 | 2277618 | + | 2385 | YP_002451340.1 | 7191369 | *bacF* |
| bacitracin synthetase 1 | 2598466 | 2605476 | - | 2336 | YP_002451685.1 | 7189494 | - |



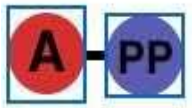

NRPS prediction of bacillibactin synthetase gene detected by PKS-NRPS website from *Bacillus cereus* AH820




NRPS prediction of trunated bacitracin synthetase gene detected by PKS-NRPS website from *Bacillus cereus* AH820

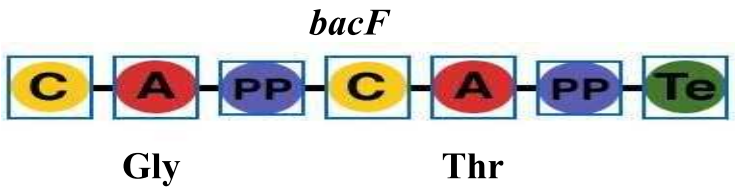
## Detected NRPS genes in *Bacillus cereus* Q1

| Product Name | Start | End | Strand | Length (a.a) | Accession | GeneID | Locus |
|---|---|---|---|---|---|---|---|
| multifunctional nonribosomal peptide synthetase (mycosubtilin synthetase) | 387368 | 390556 | + | 1062 | YP_002528166.1 | 7374198 | *dhbF* |
| isochorismatase | 2260387 | 2261274 | + | 295 | YP_002530014.1 | 7373284 | *entB* |
| nonribosomal peptide synthetase dhbf | 2261306 | 2268463 | + | 2385 | YP_002530015.1 | 7373285 | *dhbF* |




## No Hit

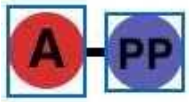
NRPS prediction of unknown synthetase gene detected by PKS-NRPS website from *Bacillus cereus* Q1




NRPS prediction of bacillibactin synthetase gene detected by PKS-NRPS website from *Bacillus cereus* Q1

## Detected NRPS genes in *Bacillus cytotoxicus* NVH 391-98

| Product Name | Start | End | Strand | Length (a.a) | Accession | GeneID | Locus |
|---|---|---|---|---|---|---|---|
| amino acid adenylation domain-containing protein | 422965 | 426549 | + | 1194 | YP_001373721.1 | 5344843 | - |
| amino acid adenylation domain-containing protein | 1868738 | 1875898 | + | 2386 | YP_001375053.1 | 5343668 | - |
| amino acid adenylation domain-containing protein | 3257499 | 3263594 | - | 2031 | YP_001376413.1 | 5347172 | - |

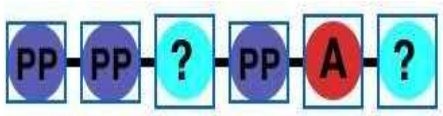



**No Hit**

NRPS prediction of unknown synthetase gene detected by PKS-NRPS website from *Bacillus cytotoxicus* NVH 391-98


**Gly** **Thr**

NRPS prediction of bacillibactin synthetase gene detected by PKS-NRPS website from *Bacillus cytotoxicus* NVH 391-98

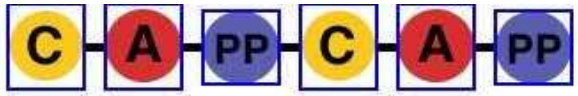



**PKS Ser PKS**

NRPS prediction of unknown synthetase gene detected by PKS-NRPS website from *Bacillus cytotoxicus* NVH 391-98

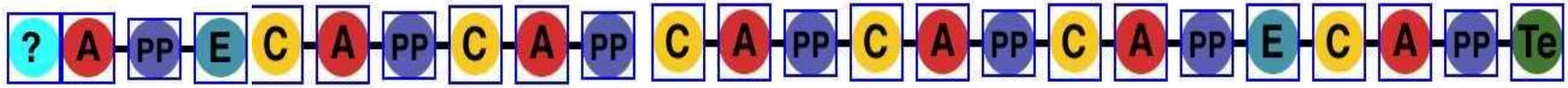
## Detected NRPS genes in *Bacillus weihenstephanensis* KBAB4

| Product Name | Start | End | Strand | Length (a.a) | Accession | GeneID | Locus |
|---|---|---|---|---|---|---|---|
| amino acid adenylation domain-containing protein | | | + | 2439 | YP_001642453.1 | 5839703 | - |
| amino acid adenylation domain-containing protein | | | - | 4968 | YP_001642560.1 | 5839712 | - |
| amino acid adenylation domain-containing protein | | | - | 2156 | YP_001642561.1 | 5839713 | - |
| amino acid adenylation domain-containing protein | | | - | 1518 | YP_001642562.1 | 5839714 | - |
| amino acid adenylation domain-containing protein | 1823934 | 1829702 | + | 1922 | YP_001644642.1 | 5841993 | - |
| Beta-ketoacyl synthase | | | + | 1408 | YP_001644643.1 | 5841994 | - |
| amino acid adenylation domain-containing protein | 1833994 | 1838424 | + | 1476 | YP_001644644.1 | 5841995 | - |
| amino acid adenylation domain-containing protein | 1841596 | 1853838 | + | 4080 | YP_001644646.1 | 5841997 | - |
| amino acid adenylation domain-containing protein | 2239637 | 2246797 | + | 2386 | YP_001645028.1 | 5842392 | - |



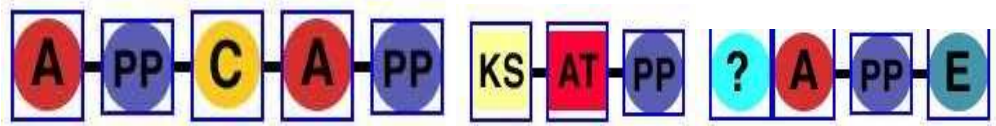

**No Hit** **No Hit**

NRPS prediction of unknown synthetase gene detected by PKS-NRPS website from *Bacillus weihenstephanensis* KBAB4



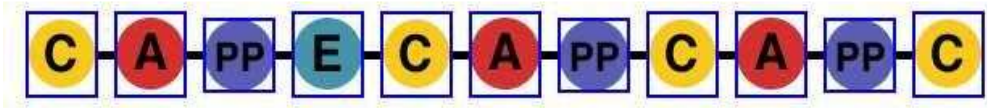

**PKS Thr** **Gln** **Gly** **Glu/Asp** **Ser** **No Hit** **Thr**

NRPS prediction of three kurstakin synthetases genes detected by PKS-NRPS website from *Bacillus weihenstephanensis* KBAB4




**No Hit** **No Hit** **PKS** **Asx**

NRPS prediction of three unknown synthetases genes detected by PKS-NRPS website from *Bacillus weihenstephanensis* KBAB4




**No Hit** **5hOrn** **Asp**

NRPS prediction of unknown unknown synthetase gene detected by PKS-NRPS website from *Bacillus weihenstephanensis* KBAB4


**Gly** **Thr**

NRPS prediction of bacillibactin synthetase gene detected by PKS-NRPS website from *Bacillus weihenstephanensis* KBAB4

**Detected NRPS genes in *Bacillus cereus* biovar *anthracisstr.*C1**

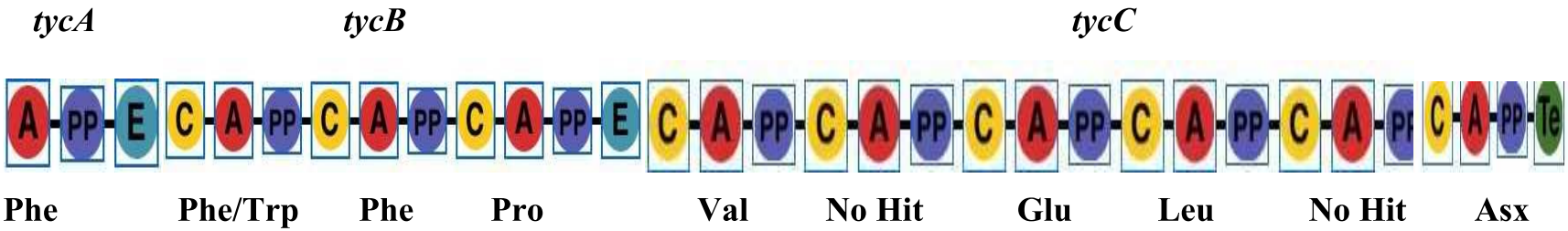
| Product Name | Start | End | Strand | Length | Accession | GeneID |
|---|---|---|---|---|---|---|
| 2,3-dihydroxybenzoate-AMP ligase | 2179267 | 2180883 | + | 538 | YP_003792100.1 | 9454797 |
| isochorismatase | 2180915 | 2181802 | + | 295 | YP_003792101.1 | 9454798 |
| ribosomal-protein-alanine acetyltransferase | 2181834 | 2188991 | + | 2385 | YP_003792102.1 | 9454799 |
| MbtH-like protein | 2188988 | 2189212 | + | 74 | YP_003792103.1 | 9454800 |


**Gly** **Thr**

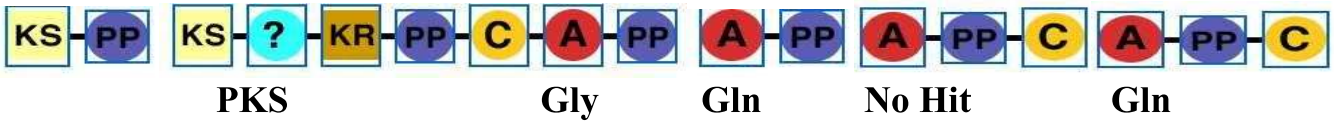
NRPS prediction of bacillibactin NRPS molecule detected by PKS-NRPS website from *Bacillus cereus* biovar *anthracis str.*C1

**Detected NRPS genes from *Brevibacillus brevis***

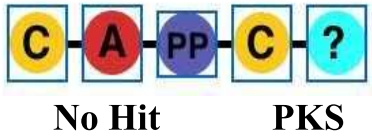
| Product Name | Start | End | Strand | Length | Accession | GeneID | Locus |
|---|---|---|---|---|---|---|---|
| tyrocidine synthetase I | 2876081 | 2879359 | + | 1092 | YP_002772268.1 | 7715463 | tycA |
| tyrocidine synthetase II | 2879454 | 2890223 | + | 3589 | YP_002772269.1 | 7720771 | tycB |
| tyrocidine synthetase III | 2890329 | 2909792 | + | 6487 | YP_002772270.1 | 7720772 | tycC |
| ABC transporter ATP-binding protein/permease | 2909876 | 2911657 | + | 593 | YP_002772271.1 | 7720773 | tycD |
| ABC transporter ATP-binding protein/permease | 2911708 | 2913459 | + | 583 | YP_002772272.1 | 7720774 | tycE |
| 4'-phosphopantetheinyl transferase | 2913456 | 2914148 | + | 230 | YP_002772273.1 | 7720775 | - |
| polyketide synthase | 3364909 | 3368277 | - | 1122 | YP_002772674.1 | 7720171 | - |
| malonyl-CoA transacylase | 3368319 | 3369587 | - | 422 | YP_002772675.1 | 7720172 | - |
| mixed polyketide synthase/non-ribosomal peptide synthetase | 3369604 | 3378261 | - | 2885 | YP_002772676.1 | 7720173 | - |
| non-ribosomal peptide synthetase | 3378263 | 3380953 | - | 896 | YP_002772677.1 | 7720174 | - |
| non-ribosomal peptide synthetase | 3380984 | 3384967 | - | 1327 | YP_002772678.1 | 7720175 | - |
| non-ribosomal peptide synthetase | 3385001 | 3388291 | - | 1096 | YP_002772679.1 | 7720176 | - |
| non-ribosomal peptide synthetase | 3389488 | 3396009 | - | 2173 | YP_002772681.1 | 7720178 | - |
| non-ribosomal peptide synthetase | 3397358 | 3401749 | - | 1463 | YP_002772683.1 | 7720180 | - |
| non-ribosomal peptide synthetase | 3457566 | 3462290 | - | 1574 | YP_002772737.1 | 7720225 | - |
| non-ribosomal peptide synthetase | 3462307 | 3477903 | - | 5198 | YP_002772738.1 | 7720226 | - |
| non-ribosomal peptide synthetase | 5470251 | 5479013 | - | 2920 | YP_002774676.1 | 7721045 | - |
| non-ribosomal peptide synthetase | 5479032 | 5481056 | - | 674 | YP_002774677.1 | 7721046 | - |
| non-ribosomal peptide synthetase | 5481139 | 5490681 | - | 3180 | YP_002774678.1 | 7721047 | - |
| non-ribosomal peptide synthetase | 5490692 | 5497795 | - | 2367 | YP_002774679.1 | 7721048 | - |



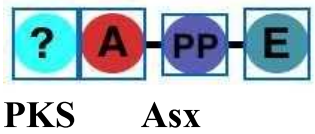

NRPS prediction of three tyrocidine synthetases genes detected by PKS-NRPS website from *Brevisbacillus brevis*




NRPS prediction of five unknown synthetases genes detected by PKS-NRPS website from *Brevisbacillus brevis*




NRPS prediction of unknown synthetase gene detected by PKS-NRPS website from *Brevisbacillus brevis*



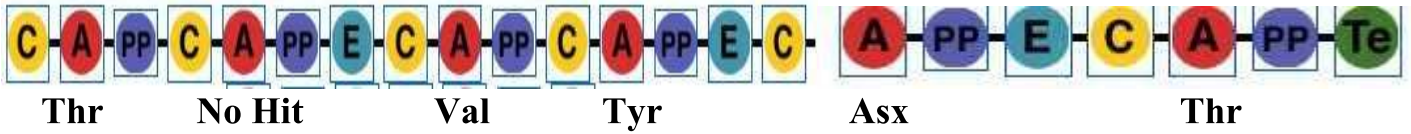

NRPS prediction of unknown synthetase gene detected by PKS-NRPS website from *Brevisbacillus brevis*


NRPS prediction of two unknown synthetases genes detected by PKS-NRPS website from *Brevisbacillus brevis*

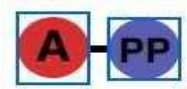

NRPS prediction of four unknown synthetases genes detected by PKS-NRPS website from *Brevisbacillus brevis*

## Detected NRPS genes from *Paenibacillus polymyxa* E681

| Product Name | Start | End | Strand | Length | Accession | GeneID | Locus |
|---|---|---|---|---|---|---|---|
| bacitracin synthetase 3 | 88134 | 111860 | + | 7908 | YP_003868523.1 | 9772842 | - |
| Non-ribosomal peptide synthetase module | 1926164 | 1929937 | - | 1257 | YP_003870182.1 | 9774561 | - |
| polyketide synthase modules containing protein | 3279068 | 3280045 | - | 325 | YP_003871368.1 | 9775755 | - |
| polyketide synthase module containing protein | 3281019 | 3289556 | - | 2845 | YP_003871370.1 | 9775757 | - |
| polyketide synthase module containing protein | 3289588 | 3302988 | - | 4466 | YP_003871371.1 | 9775758 | - |
| polyketide synthase module containing protein | 3303039 | 3322760 | - | 6573 | YP_003871372.1 | 9775759 | - |
| Non-ribosomal peptide synthetase module containing protein | 3322744 | 3333888 | - | 3714 | YP_003871373.1 | 9775760 | - |
| Non-ribosomal peptide synthetase module containing protein | 4558724 | 4562032 | - | 1102 | YP_003872505.1 | 9776927 | - |
| Gramicidin S synthetase II | 4562187 | 4577048 | - | 4953 | YP_003872506.1 | 9776928 | - |




NRPS prediction of unknown synthetase gene detected by PKS-NRPS website from *Paenibacillus polymyxa* 681



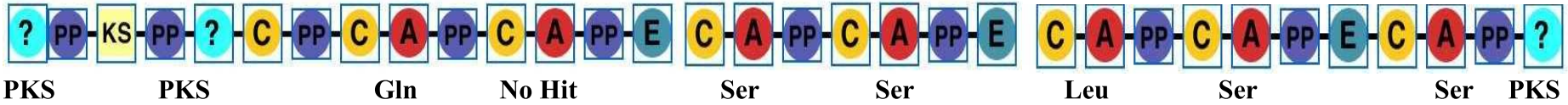
**No Hit**

NRPS prediction of unknown synthetase gene detected by PKS-NRPS website from *Paenibacillus polymyxa* 681

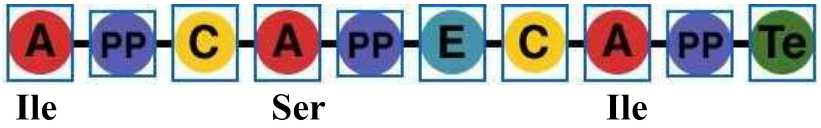

NRPS prediction of four unknown synthetases genes detected by PKS-NRPS website from *Paenibacillus polymyxa* 681

**Detected NRPS genes from *Paenibacillus polymyxa* SC2**

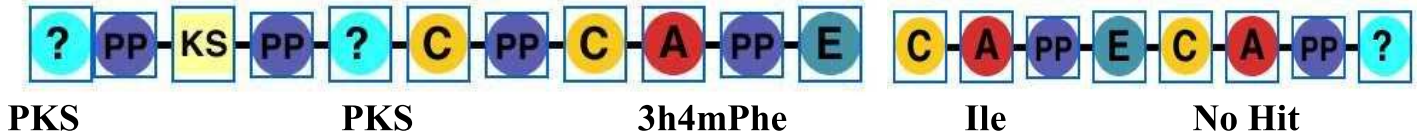
| Product Name | Start | End | Strand | Length | Accession | GeneID | Locus |
|---|---|---|---|---|---|---|---|
| ATP-dependent asparagine adenylase 1 | 1091648 | 1107397 | + | 5249 | YP_003945216.1 | 9849314 | - |
| nonribosomal peptide synthetase | 1107414 | 1115180 | + | 2588 | YP_003945217.1 | 9849315 | - |
| non-ribosomal peptide synthase, pvdd/pvdj | 1115191 | 1127733 | + | 4180 | YP_003945218.1 | 9849316 | - |
| phenylalanine racemase | 2509308 | 2520554 | + | 3748 | YP_003946635.1 | 9850733 | - |
| pks acyltransferase domain protein | 2793691 | 2794902 | + | 403 | YP_003946861.1 | 9850959 | - |
| iturin a synthetase a | 2794889 | 2807536 | + | 4215 | YP_003946862.1 | 9850960 | - |
| non-ribosomal peptide synthase, pvdj(2) | 2807560 | 2816919 | + | 3119 | YP_003946863.1 | 9850961 | - |
| non-ribosomal peptide synthase, pvdj(2) | 2956671 | 2965148 | - | 2825 | YP_003946998.1 | 9851096 | - |
| phenylalanine racemase | 2965263 | 2968694 | - | 1143 | YP_003946999.1 | 9851097 | - |
| polyketide synthase | 3570568 | 3579456 | - | 2962 | YP_003947587.1 | 9851685 | - |
| polyketide synthase | 3579488 | 3592885 | - | 4465 | YP_003947588.1 | 9851686 | - |
| polyketide synthase | 3592936 | 3612693 | - | 6585 | YP_003947589.1 | 9851687 | - |
| non-ribosomal peptide synthase/polyketide synthase ta1 | 3612677 | 3623809 | - | 3710 | YP_003947590.1 | 9851688 | - |



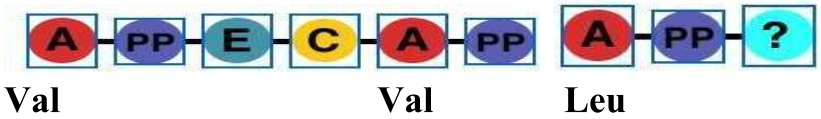

NRPS prediction of three unknown synthetases genes detected by PKS-NRPS website from *Paenibacillus polymyxa* SC2




NRPS prediction of unknown synthetase gene detected by PKS-NRPS website from *Paenibacillus polymyxa* SC2




NRPS prediction of two unknown synthetase gene detected by PKS-NRPS website from *Paenibacillus polymyxa* SC2




NRPS prediction of two unknown synthetase gene detected by PKS-NRPS website from *Paenibacillus polymyxa* SC2


NRPS prediction of four unknown synthetase gene detected by PKS-NRPS website from *Paenibacillus polymyxa* SC2

## Index 3: Different alignments of sequenced fragments from *B. subtilis* 21332

### *ppsB* fragment Needle alignment of *B. subtilis* 21332 and 168

```
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 1303
# Identity:    1303/1303 (100.0%)
# Similarity:  1303/1303 (100.0%)
# Gaps:           0/1303 ( 0.0%)
# Score: 6515.0
#
#
#=======================================

EMBOSS_001          1 TCAGACTCTTGGTGAACTGTCCGCCTGGGTTCGAAAAGACGTACGCGCCA     50
                      ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001          1 TCAGACTCTTGGTGAACTGTCCGCCTGGGTTCGAAAAGACGTACGCGCCA     50

EMBOSS_001         51 TTGACCAAGGCCCGGTCGAAGGCGAGATTACTTGGACACCTATCCAGCAG    100
                      ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001         51 TTGACCAAGGCCCGGTCGAAGGCGAGATTACTTGGACACCTATCCAGCAG    100

EMBOSS_001        101 TGGTTTTTCTCGCAATCTCTGGAAAGCCATCATTTTAACCAGTCCGTTAT    150
                      ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001        101 TGGTTTTTCTCGCAATCTCTGGAAAGCCATCATTTTAACCAGTCCGTTAT    150

EMBOSS_001        151 GATCTACCGGGCTGAACGATTTGATGAGGCTGCGCTTAGAAAAGTACTAA    200
                      ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001        151 GATCTACCGGGCTGAACGATTTGATGAGGCTGCGCTTAGAAAAGTACTAA    200

EMBOSS_001        201 AAAGCCTTGTGACTCATCATGATGCACTGAGAATCGTATGCCGGCATGAA    250
                      ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001        201 AAAGCCTTGTGACTCATCATGATGCACTGAGAATCGTATGCCGGCATGAA    250

EMBOSS_001        251 GATGGCAGGCACGTGCAGATAAACAGAGGGATAGATCTTTCAGATGAGGA    300
                      ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001        251 GATGGCAGGCACGTGCAGATAAACAGAGGGATAGATCTTTCAGATGAGGA    300

EMBOSS_001        301 GCTATATGCACTTGAACTGTTTGATGTAAAGGATAGCCTAACAGAAGCAC    350
                      ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001        301 GCTATATGCACTTGAACTGTTTGATGTAAAGGATAGCCTAACAGAAGCAC    350

EMBOSS_001        351 GGTACACGATTGAGGAAGCAGCCAGCCGGATGCAAGAACATATTCGTTTG    400
                      ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001        351 GGTACACGATTGAGGAAGCAGCCAGCCGGATGCAAGAACATATTCGTTTG    400

EMBOSS_001        401 GAAACCGGCCCCTTACTTCACGCTGGATTGTTTAGAACCGAGAATGGTGA    450
                      ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001        401 GAAACCGGCCCCTTACTTCACGCTGGATTGTTTAGAACCGAGAATGGTGA    450

EMBOSS_001        451 CCATTTATTTCTGACGATTCATCACTTAGTTGTTGATGCGGTTTCCTGGC    500
                      ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001        451 CCATTTATTTCTGACGATTCATCACTTAGTTGTTGATGCGGTTTCCTGGC    500

EMBOSS_001        501 GCATTTTGTTCGAGGATTTTTCAGCTGCTTACAAACAGGCAGTCTCAGGA    550
                      ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001        501 GCATTTTGTTCGAGGATTTTTCAGCTGCTTACAAACAGGCAGTCTCAGGA    550

EMBOSS_001        551 GAATCTATTAAATTGCCGAAAAAAACAGATTCATATTTAACGTATTCACA    600
                      ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001        551 GAATCTATTAAATTGCCGAAAAAAACAGATTCATATTTAACGTATTCACA    600

EMBOSS_001        601 AAGAATAGCTGACTACTCTAAAAGCCGCCAGGTGCAGCGCGAAGCAGCGT    650
                      ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001        601 AAGAATAGCTGACTACTCTAAAAGCCGCCAGGTGCAGCGCGAAGCAGCGT    650

EMBOSS_001        651 ACTGGGATGAGTGCGAGAACCGTCAAATCCAGCCTATTCCAAAAGACAAT    700
                      ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001        651 ACTGGGATGAGTGCGAGAACCGTCAAATCCAGCCTATTCCAAAAGACAAT    700

EMBOSS_001        701 GACGCAGCATCTAACACATTCAAAGATACAGAAGTGATTGATTTTGAGCT    750
                      ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001        701 GACGCAGCATCTAACACATTCAAAGATACAGAAGTGATTGATTTTGAGCT    750

EMBOSS_001        751 GTCTCGCCATCACACAGAATTACTATTAACGGCTGCACATAAAGCTTACA    800
                      ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001        751 GTCTCGCCATCACACAGAATTACTATTAACGGCTGCACATAAAGCTTACA    800

EMBOSS_001        801 GCACGGAGATGAATGATATCCTTCTGACAGCTTTAGGTCTTGCTTTGCAG    850
                      ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001        801 GCACGGAGATGAATGATATCCTTCTGACAGCTTTAGGTCTTGCTTTGCAG    850

EMBOSS_001        851 AAATGGACGGGAAATAACCAATTTAAAATCAGCATGGAGGGCCATGGAAG    900
                      ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001        851 AAATGGACGGGAAATAACCAATTTAAAATCAGCATGGAGGGCCATGGAAG    900
```

```
EMBOSS_001       901 GGAATCATATCTTGAGGACATCGATATTAGCAGAACTGTTGGCTGGTTCA    950
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       901 GGAATCATATCTTGAGGACATCGATATTAGCAGAACTGTTGGCTGGTTCA    950

EMBOSS_001       951 CCTCCATTTATCCTGTGTGGCTTGATATGAGAGATTCCGATCATAAGGAC   1000
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       951 CCTCCATTTATCCTGTGTGGCTTGATATGAGAGATTCCGATCATAAGGAC   1000

EMBOSS_001      1001 AAAGAAGAGCGGCTGGGCCACCTTATTAAACAGACAAAAGATATGCTGCA   1050
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      1001 AAAGAAGAGCGGCTGGGCCACCTTATTAAACAGACAAAAGATATGCTGCA   1050

EMBOSS_001      1051 CCGTATACCGCATAAAGGCGCGGGCTACGGCGTGCTGAAATATATCAGCA   1100
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      1051 CCGTATACCGCATAAAGGCGCGGGCTACGGCGTGCTGAAATATATCAGCA   1100

EMBOSS_001      1101 AAAGATGGGGCTCTCAGAAAAACAGCCCAGAAATCAGCTTTAATTATTTA   1150
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      1101 AAAGATGGGGCTCTCAGAAAAACAGCCCAGAAATCAGCTTTAATTATTTA   1150

EMBOSS_001      1151 GGCCAGTTTGATCAGGATATTCAATCAAATGCCTTTGAGGTGTCTGACAT   1200
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      1151 GGCCAGTTTGATCAGGATATTCAATCAAATGCCTTTGAGGTGTCTGACAT   1200

EMBOSS_001      1201 CAAGCCAGGAAATGAGATCAGCCCGAACTGGAAGCGTCCTTATGCACTGG   1250
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      1201 CAAGCCAGGAAATGAGATCAGCCCGAACTGGAAGCGTCCTTATGCACTGG   1250

EMBOSS_001      1251 ATATCAGCGGCGCCGTTTCTTCGGGGTGCCTGAACATGCACATCATCTAT   1300
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      1251 ATATCAGCGGCGCCGTTTCTTCGGGGTGCCTGAACATGCACATCATCTAT   1300

EMBOSS_001      1301 AAT   1303
                     |||
EMBOSS_001      1301 AAT   1303
```

## *dacC-ppsA* fragment BLAST alignment of *B. subtilis* 21332 and 168

```
>emb|AL009126.3|  Bacillus subtilis subsp. subtilis str. 168 complete genome
Length=4215606



                                                    Sort alignments for this subject sequence by:
                                                      E value  Score  Percent identity
                                                      Query start position  Subject start position

 Features in this part of subject sequence:
   plipastatin synthetase

 Score = 1356 bits (734),  Expect = 0.0
 Identities = 751/759 (99%), Gaps = 1/759 (0%)
 Strand=Plus/Minus

Query  860      ATCTTCTGCGAAGAACACAAGATCAATATCATTTCCCTTATTTATGGCTTCATTTTACAT  919
                ||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Sbjct  1997246  ATCTTCTGCGAAGAACACAAGATCAATATCATTT-CCTTATTTATGGCTTCATTTTACAT  1997188

Query  920      ATGCATCAGCCGAATCACTTCTAAAAAAGATCTTGCCATTGGCACCTACTATGGCAACAG  979
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1997187  ATGCATCAGCCGAATCACTTCTAAAAAAGATCTTGCCATTGGCACCTACTATGGCAACAG  1997128

Query  980      AGGGTCAAAGGCTGAAAAAGAAATGCTCGGCATGTTCGTTAGTTCTCTTCCAATCAGAAT  1039
                ||| ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Sbjct  1997127  AGGTTCAAAGGCTGAAAAAGAAATGCTCGGCATGTTCGTTAGTTCCCTTCCAATCAGAAT  1997068

Query  1040     AACGGTTGATCCTGACACAGATTTTCTTTCCTTTGTCCGCACAATAGGCAGGGAACAGCT  1099
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1997067  AACGGTTGATCCTGACACAGATTTTCTTTCCTTTGTCCGCACAATAGGCAGGGAACAGCT  1997008

Query  1100     TTCTGTTATGCGTCATCAACGGTTTCCTTATAATTTGCTTGTAAATGAATTAAGAAATGA  1159
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1997007  TTCTGTTATGCGTCATCAACGGTTTCCTTATAATTTGCTTGTAAATGAATTAAGAAATGA  1996948

Query  1160     ACAGAAGGATTTACACAACCTTATCGGTATATCTATGCAGTATCAGCCACTTCAGTGGCA  1219
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1996947  ACAGAAGGATTTACACAACCTTATCGGTATATCTATGCAGTATCAGCCACTTCAGTGGCA  1996888

Query  1220     TAATGCAGATGATTTTGACTATGAAACAGCGCTTTACTTCAGCGGATACACCGCGAACGA  1279
                 |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Sbjct  1996887  CAATGCAGATGATTTTGACTATGAAACAGCGCTTTATTTCAGCGGATACACCGCGAACGA  1996828

Query  1280     GCTTTCAGTTCAAATACAAGAACGGATAGACAATGGAACCATTCAATTAAATTTCGACTA  1339
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1996827  ACTTTCAGTTCAAATACAAGAACGGATAGACAATGGAACCATTCAATTAAATTTCGACTA  1996768

Query  1340     TCAAAATACCTTGTTTTCTCTTGAGGATATTAAACGAATACAATCTCATCTGCTTACCAT  1399
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1996767  TCAAAATACCTTGTTTTCTCTTGAGGATATTAAACGAATACAATCTCATCTGCTTACCAT  1996708

Query  1400     ACTGGAAAACGCCCTTCAACATCCGCACAGCTTTATAAGAGAATTAGATATGACGAATAC  1459
                |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Sbjct  1996707  ACTGGAAAACGCCCTTAAACATCCGCACAGCTTTATAAGAGAATTAGATATGACGAATAC  1996648

Query  1460     AAGAGAAAAGCAAAAATTGTTGTATGAGTTTAACAAGACGGAAGCTGTTTCGCCTAAAGC  1519
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1996647  AAGAGAAAAGCAAAAATTGTTGTATGAGTTTAACAAGACGGAAGCTGTTTCGCCTAAAGC  1996588

Query  1520     TTTCACCCTGCATGGGCTGTTTGAGCGACAGGCAGCTTTCACACCTGAGCGGCTGGCCAT  1579
                ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Sbjct  1996587  TTTCACCCTGCATGGGCTGTTTGAGCGACAGGCAGCTTTCACACCGGAGCGGCTGGCCAT  1996528

Query  1580     TCGCTTTTCAGGCGGTTCGCTGACATACGCTGAGCTTGA  1618
                ||||||||||||||||||||||||||||||||||||||
Sbjct  1996527  TCGCTTTTCAGGCGGTTCGCTGACATACGCTGAGCTTGA  1996489
```

## *ppsA-ppsB* fragment Needle alignment of *B. subtilis* 21332 and 168

```
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 754
# Identity:     734/754 (97.3%)
# Similarity:   734/754 (97.3%)
# Gaps:           2/754 ( 0.3%)
# Score: 3578.0
#
#
#=======================================

EMBOSS_001         1 Tgctgtttgaaacccagcaagcctatggtactgatgccaacgaacttttg     50
                     ||||| |||.|||.||||||||||||||||||||||||||||||||||||
EMBOSS_001         1 TGCTG-TTGGAACGCAGCAAGCCTATGGTACTGATGCCAACGAACTTTTG     49

EMBOSS_001        51 cttacggctttagg-aatggcgctttctgaatggactgggcatgaccaaa     99
                     ||||||||||||||| |||||||||||||||||||||.|||||||||||||
EMBOSS_001        50 CTTACGGCTTTAGGAAATGGCGCTTTCTGAATGGACAGGGCATGACCAAA     99

EMBOSS_001       100 tcgtgatcagcactgagggacacggtagagaagggcatgtgccaaacatt    149
                     ||||||||||||||||||||||||||.||||||||||||||||||||||||
EMBOSS_001       100 TCGTGATCAGCACTGAGGGACACGGCAGAGAAGGGCATGTGCCAAACATT    149

EMBOSS_001       150 gatatttcaaggacagttggatggtttacctccatttatcctattctgct    199
                     ||||||||||||||.||||||||||||||||||||||||||||||||||||
EMBOSS_001       150 GATATTTCAAGGACCGTTGGATGGTTTACCTCCATTTATCCTATTCTGCT    199

EMBOSS_001       200 agacatgggtattcccgagcctttcgaggatcagctcgcataccgaatta    249
                     ||||||||||||||.||||||||||||||||||||||||||||||||||||
EMBOSS_001       200 AGACATGGGTATTTCCGAGCCTTTCGAGGATCAGCTCGCATACCGAATTA    249

EMBOSS_001       250 agacgacaaaagatatgctgagaagagtaccgaacaaaggaactggatac    299
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       250 AGACGACAAAAGATATGCTGAGAAGAGTACCGAACAAAGGAACTGGATAC    299

EMBOSS_001       300 ggattgttaacacacataggagaactgcggcataaagagcctgaggtcag    349
                     ||||||||||||||||||||.||||||||||||||||||||||||||||||
EMBOSS_001       300 GGATTGTTAACACACATAGGAAAACTGCGGCATAAAGAGCCTGAGGTCAG    349

EMBOSS_001       350 ctttaattatctcggtcaatttagcgaagagaaggaagttgaaacgtttc    399
                     ||||||||||||||||||||||||||||.|||||.||||.|||.|||||||
EMBOSS_001       350 CTTTAATTATCTCGGTCAATTTAGCGAGGAGAAAGAAGCTGATACGTTTC    399

EMBOSS_001       400 aattatcatattaccagccccgttacgaaatagcgggtgagagagaacga    449
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       400 AATTATCATATTACCAGCCCCGTTACGAAATAGCGGGTGAGAGAGAACGA    449

EMBOSS_001       450 gagtatgagctggacatcaatgctctcattacggacggacggcttcacgt    499
                     ||||||||||||||||||||.||||||||||||||||||||||||||.||
EMBOSS_001       450 GAGTATGAGCTGGACATCAACGCTCTCATTACGGACGGACGGCTTCAAGT    499

EMBOSS_001       500 aaaagcggtatacacccaagtgttcagcaaacactcaattgagtgctttа    549
                     ||||||||||||||||||.|.||||||||||||||||||||||||||||||
EMBOSS_001       500 AAAAGCGGTATACACCCAGGCGTTCAGCAAACACTCAATTGAGTGCTTTA    549

EMBOSS_001       550 tggatcgttttcacaggcatcttatcgagacgattgagcattgttcgcag    599
                     ||||||||||||||.||||||||||||||||||||||||||||||||||||
EMBOSS_001       550 TGGATCGTTTTCATAGGCATCTTATCGAGACGATTGAGCATTGTTCGCAG    599

EMBOSS_001       600 aaaaaagcacgcgaaaaaacattgagcgatttcagcaataaagaattaac    649
                     ||||||||||||||||||||||||||||||||||||||||||.||||||||
EMBOSS_001       600 AAAAAAGCACGCGAAAAAACATTGAGCGATTTCAGCAATAAGGAATTAAC    649

EMBOSS_001       650 gttgtctgcattatctagcatcgaagatctcgtgaaagatctttaacttt    699
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       650 GTTGTCTGCATTATCTAGCATCGAAGATCTCGTGAAAGATCTTTAACTTT    699

EMBOSS_001       700 tgaaaaggtgtgtggaattgatggcacaatcggcacaaattcaagatatt    749
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       700 TGAAAAGGTGTGTGGAATTGATGGCACAATCGGCACAAATTCAAGATATT    749

EMBOSS_001       750 tatc    753
                     ||.|
EMBOSS_001       750 TACC    753
```

## *ppsC-ppsD* fragment Needle alignment of *B. subtilis* 21332 and 168

```
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 1066
# Identity:    1041/1066 (97.7%)
# Similarity:  1041/1066 (97.7%)
# Gaps:           2/1066 ( 0.2%)
# Score: 5093.0
#
#
#=======================================

EMBOSS_001         1 gatacacaagatattctgttaaccgcagcatccttggcgatttgtgaatg     50
                     |||||.||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001         1 GATACGCAAGATATTCTGTTAACCGCAGCATCCTTGGCGATTTGTGAATG     50

EMBOSS_001        51 gacaggaggaagcaagctccgtatcgctatggaaggacacgggagagaac    100
                     ||||||||||||.|.||||||||||||||||||||||||||||||||||||
EMBOSS_001        51 GACAGGAGGAAGTATGCTCCGTATCGCTATGGAAGGACACGGGAGAGAAC    100

EMBOSS_001       101 acattctgccagagttagatataagcaggacggtcggctggtttacatcc    150
                     |.|||||||||||||||||||.||||||||||||||||||||||||||||
EMBOSS_001       101 ATATTCTGCCAGAGTTAGATATCAGCAGGACGGTCGGCTGGTTTACATCC    150

EMBOSS_001       151 atgtacccggccctcataagctttgaaaatcatagggatgagctcggaac    200
                     |||||.||||||||||||||||||||||||||||.|||||||||||||||
EMBOSS_001       151 ATGTATCCGGCCCTCATAAGCTTTGAAAATCATAGGTATGAGCTCGGAAC    200

EMBOSS_001       201 ctcagtaaaaaccgttaaagatacactgggccggataccaaacaaagggg    250
                     |.||.||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       201 CGCATTAAAAACCGTTAAAGATACACTGGGCCGGATACCAAACAAAGGGG    250

EMBOSS_001       251 taggttacggtatgctgaaataccttaca-caccctgaaaataaaagcat    299
                     |||||||||||||||||||||||||||||| ||||||||||||||||||||
EMBOSS_001       251 TAGGTTACGGTATGCTGAAATACCTTACACCACCCTGAAAATAAAAGCAT    300

EMBOSS_001       300 tacgttttcaaaaacacctgaaatcagctttaactatttgggccagttca    349
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       301 TACGTTTTCAAAAACACCTGAAATCAGCTTTAACTATTTGGGCCAGTTCA    350

EMBOSS_001       350 acgacatcgaaagacaggacaccttccgcccatcaagccttggtagcgga    399
                     |||||||||||||||||||||||||||||||||||||||||||.||||||
EMBOSS_001       351 ACGACATCGAAAGACAGGACACCTTCCGCCCATCAAGCCTTGGGAGCGGA    400

EMBOSS_001       400 aaggatatcactcatacatggaagcgcgagcaaatcatcgagatgagtgc    449
                     |||||||||||||||||||||||||.||||||||||||||||||.||||||
EMBOSS_001       401 AAGGATATCACTCATACATGGAAGCGTGAGCAAATCATCGAGATGAGCGC    450

EMBOSS_001       450 aatggcagcagacaaaaagctccatttcaatcttagttatccgccagccc    499
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       451 AATGGCAGCAGACAAAAAGCTCCATTTCAATCTTAGTTATCCGCCAGCCC    500

EMBOSS_001       500 gctttcacagaaacacaatggaacagcttattaacaggattgaacacttt    549
                     |||||||||||||||||||||||.||||||||||||||||||||||||||||
EMBOSS_001       501 GCTTTCACAGAAACACAATGGAGCAGCTTATTAACAGGATTGAACACTTT    550

EMBOSS_001       550 ttgctggacattatgaaacactgcgccggtcaacagaaagcagaaaagac    599
                     ||||||||||||||||||||||||||||||||||||||||||.||||||||||
EMBOSS_001       551 TTGCTGGACATTATGAAACACTGCGCCGGTCAACAGAAAACAGAAAAGAC    600

EMBOSS_001       600 actgagcgatttcagcagtcaatcattaacggcagaggacttggacagca    649
                     ||||||||||||||||||||||||||||||||||.||||||||||||||||
EMBOSS_001       601 ACTGAGCGATTTCAGCAGTCAATCATTAACGGCTGAGGACTTGGACAGCA    650

EMBOSS_001       650 tatccagcttggtggaagaattgtagaaaatcgtgaca-ggagacatgaa    698
                     |||||||||||||||||||||||||||||||||||||| ||||||||||||
EMBOSS_001       651 TATCCAGCTTGGTGGAAGAATTGTAGAAAATCGTGACAGGGAGACATGAA    700

EMBOSS_001       699 catgacaaaagcgaattcaattcaggatatttatccgctgtcttacatgc    748
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       701 CATGACAAAAGCGAATTCAATTCAGGATATTTATCCGCTGTCTTACATGC    750

EMBOSS_001       749 aggaagggatgcttttccattccctcctgcaaaaagattcacaggcgtat    798
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       751 AGGAAGGGATGCTTTTCCATTCCCTCCTGCAAAAAGATTCACAGGCGTAT    800

EMBOSS_001       799 gttgagcaagcttctttttacaatagaaggaaaagtcacccccgcagttttt    848
                     ||||||||||||||||||||||.|||||||||||||.|||||||||||||||
EMBOSS_001       801 GTTGAGCAAGCTTCTTTTACAATTGAAGGAAAAGTCAACCCGCAGTTTTT    850

EMBOSS_001       849 tcaaaacagcatcaacgctcttgtagaaagacatgatattttcagaacga    898
                     ||||||||||||||||||||||||||||||||||||||||||||||||||.|
EMBOSS_001       851 TCAAAACAGCATCAACGCTCTTGTAGAAAGACATGATATTTTCAGAACCA    900
```

```
EMBOSS_001        899 tttttatcagccaaaatgtttcctccccccagcaggttgttctgagagaa    948
                      ||||||||||||||||||||||||||||||||||||||||||||.|||||||
EMBOSS_001        901 TTTTTATCAGCCAAAATGTTTCCTCCCCCCAGCAGGTTGTTCTAAGAGAA    950

EMBOSS_001        949 cgaaatgtcatcgtactggaagaagacattactcatttaaacgaggcgga    998
                      |||||||||||||||.||||||||||||||||||||.|||||||||||||||
EMBOSS_001        951 CGAAATGTCATCGTGCTGGAAGAAGACATTACTCAATTAAACGAGGCGGA   1000

EMBOSS_001        999 gcaatcacagtttatcgagcaatggaaagaaaaggaccgggaccgggggt   1048
                      |||||||||||||||.||||||||||||||||||||.|||||||||||||||
EMBOSS_001       1001 GCAATCACAGTTTATTGAGCAATGGAAAGAAAAAGACCGGGACCGGGGGT   1050

EMBOSS_001       1049 ttcatttgcaaaagga   1064
                      ||||||||||||||||
EMBOSS_001       1051 TTCATTTGCAAAAGGA   1066
```

## Plipastatin fragment BLAST alignment of *B. subtilis* 21332 and 168

```
 Features in this part of subject sequence:
   plipastatin synthetase

 Score = 1458 bits (789),  Expect = 0.0
 Identities = 864/900 (96%), Gaps = 6/900 (1%)
 Strand=Plus/Minus

Query  12       TGAAGGCCGGCGGGGGCGTATTTGCCGCTTGATCCTGCGTATCCAAAGGAGCGGCTGAGCT  71
                ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1980945  TGAAGGCCGGAGGGGCGTATTTGCCGCTTGATCCTGCGTATCCAAAGGAGCGGCTGAGCT  1980886

Query  72       ACATGCTGAAAGACAGCGGGGCATCGCTCTTGCTGACACAGCCAGGATGCTCCGCACCGA  131
                |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||| |
Sbjct  1980885  ACATGCTGAAAGACAGCGGGGCATCGCTCTTGCTGACACAGCCGGGATGCTCCGCACCAA  1980826

Query  132      ACTTTTCTGGAGAGACGCTTGAAGTTGACATGACATCTCTTGAGCGTGAAGAAGTAAAAC  191
                |||||||||| || |||||||||| |||||||||||||||| |||||||||||||||||
Sbjct  1980825  ACTTTTCTGGCGAAACGCTTGAAGTGGACATGACATCTCTTGAGTGTGAAGAAGTAAAAC  1980766

Query  192      GTCATGTGTCTGCTTCTGTTAGCGACGGATCTCTCGCCTATGTCATCTATACATCGGGCT  251
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1980765  GTCATGTGTCTGCTTCTGTTAGCGACGGATCTCTCGCCTATGTCATCTATACATCGGGCT  1980706

Query  252      CTACCGGGCAGCCGAAAGGAGTGGCCGTTGAGCACCGTCAGGCTGTTTCCTTTCTGACCG  311
                |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Sbjct  1980705  CTACCGGGCAGCCGAAAGGAGTGGCCGTTGAGCACCGCCAGGCTGTTTCCTTTCTGACCG  1980646

Query  312      GCATGCAGCACCAGTTTCCGCTTTCGGAAGATGACATTGTCATGGTGAAAACCTCTTTTT  371
                |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Sbjct  1980645  GCATGCAGCACCAGTTTCGGCTTTCGGAAGATGACATTGTCATGGTGAAAACCTCTTTTT  1980586

Query  372      CCTTTGATGCGTCTGTGTGGCAGCTGTTTTGGTGGACTCTTTCCGGTGCTTCGGCGTATC  431
                |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Sbjct  1980585  CCTTTGATGCGTCTGTGTGGCAGCTGTTTTGGTGGGCTCTTTCCGGTGCTTCGGCGTATC  1980526

Query  432      TGCTTCCGCCCGGTTGGGAAAAAGACCCCGCGTTGATCGTACAAGCCATTCATCAGGAAA  491
                |||||||||||||| |||||||||||| |||| ||||||||||||||||||||||||||
Sbjct  1980525  TGCTTCCGCCCGGCTGGGAAAAAGACTCCGCATTGATCGTACAAGCCATTCATCAGGAAA  1980466

Query  492      ACGTGACAACGGCTCATTTTATTCCGGCTATGCTGAACAGCTTTCTAGATCAAGCGGAAA  551
                |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Sbjct  1980465  ACGTGACAACGGCTCATTTTATTCCGGCTATGCTGAACAGCTTTCTCGATCAAGCGGAAA  1980406

Query  552      T-GTTAGCGCTCG-GTGACGGGACAAACCTAAAGCGTGTATTCGCCGGAGGTGAACCGCT  609
                | |  || ||| | ||||| |||||| ||||||||||||||||||||||||||||||||
Sbjct  1980405  TCGAAAG-GCT-GAGTGACAGGACAAGCCTAAAGCGTGTATTCGCCGGAGGTGAACCGCT  1980348

Query  610      TGCGCCGCGTACGGCAGCCCGTTTTGCTTCTATCTTGCCACAGGTATCATTAATCCATGG  669
                |||||||||||||||||||||||||||||| | ||||| ||||||||||||||||||||
Sbjct  1980347  TGCGCCGCGTACGGCAGCCCGTTTTGCTTCTGTATTGCCGCAGGTATCATTAATCCATGG  1980288

Query  670      TTACGGGCCGACAGAAGCAACGGTAGACGCGGCATTTTACGTGTTGGACCCAGAACGGGA  729
                 |||||||||||||||| |||||||||||||||||||||||| || ||||||||||||
Sbjct  1980287  CTACGGGCCGACAGAAGCCACGGTAGACGCGGCATTTTACGTGTTAGATCCAGAACGGGA  1980228

Query  730      CAGGGACCGTTTGCGGATTCCGATTGGGAAGCCCGTACCCGGTGCGCGTTTGTATGTTTT  789
                ||| |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Sbjct  1980227  CAGAGACCGTTTGCGGATTCCGATTGGGAAGCCCGTACCTGGTGCGCGTTTGTATGTTTT  1980168

Query  790      AGATCCACATTTAGCCGTACAGCCTTCTGGTGTCGCCCGGCGAACTGTACATTGCCGGAG  849
                |||||| |||||||||||||||||| ||||||||| |||||||| ||||||||||||||
Sbjct  1980167  AGATCCGCATTTAGCCGTACAGCCTTCCGGTGTCGCC-GGCGAACTATACATTGCCGGAG  1980109

Query  850      CGGGTGTTGCCAGAGGTTATTTGAATCGGCCGGCATTAACGGAGGAGCGTTTTCTT-AAG  908
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Sbjct  1980108  CGGGTGTTGCCAGAGGTTATTTGAATCGGCCGGCATTAACGGAGGAGCGTTTTCTTGAAG  1980049
```

## Index 4: Different alignments of sequenced fragments from *B. subtilis* BBG21

### *ppsA-ppsB* fragment Needle alignment of *B. subtilis* BBG21 and 168

```
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 753
# Identity:     739/753 (98.1%)
# Similarity:   739/753 (98.1%)
# Gaps:           0/753 ( 0.0%)
# Score: 3639.0
#
#
#=======================================

EMBOSS_001         1 Tgctgtttgaaacccagcaagcctatggtactgatgccaacgaacttttg     50
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001         1 TGCTGTTTGAAACCCAGCAAGCCTATGGTACTGATGCCAACGAACTTTTG     50

EMBOSS_001        51 cttacggctttaggaatggcgctttctgaatggactgggcatgaccaaat    100
                     ||||||||||||||||||||||||||||||||||.||||||||||||||||
EMBOSS_001        51 CTTACGGCTTTAGGAATGGCGCTTTCTGAATGGACAGGGCATGACCAAAT    100

EMBOSS_001       101 cgtgatcagcactgagggacacggtagagaagggcatgtgccaaacattg    150
                     ||||||||||||||||||||||||.||||||||||||||||||||||||||
EMBOSS_001       101 CGTGATCAGCACTGAGGGACACGGCAGAGAAGGGCATGTGCCAAACATTG    150

EMBOSS_001       151 atatttcaaggacagttggatggtttacctccatttatcctattctgcta    200
                     ||||||||||||.||||||||||||||||||||||||||||||||||||||
EMBOSS_001       151 ATATTTCAAGGACCGTTGGATGGTTTACCTCCATTTATCCTATTCTGCTA    200

EMBOSS_001       201 gacatgggtattcccgagcctttcgaggatcagctcgcataccgaattaa    250
                     ||||||||||||.||||||||||||||||||||||||||||||||||||||
EMBOSS_001       201 GACATGGGTATTTCCGAGCCTTTCGAGGATCAGCTCGCATACCGAATTAA    250

EMBOSS_001       251 gacgacaaaagatatgctgagaagagtaccgaacaaaggaactggatacg    300
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       251 GACGACAAAAGATATGCTGAGAAGAGTACCGAACAAAGGAACTGGATACG    300

EMBOSS_001       301 gattgttaacacacataggagaactgcggcataaagagcctgaggtcagc    350
                     ||||||||||||||||||||.||||||||||||||||||||||||||||||
EMBOSS_001       301 GATTGTTAACACACATAGGAAAACTGCGGCATAAAGAGCCTGAGGTCAGC    350
EMBOSS_001       351 tttaattatctcggtcaatttagcgaagagaaggaagttgaaacgtttca    400
                     ||||||||||||||||||||||||||||||||.||||.|||.||||||||
EMBOSS_001       351 TTTAATTATCTCGGTCAATTTAGCGAAGAGAAAGAAGCTGATACGTTTCA    400

EMBOSS_001       401 attatcatattaccagccccgttacgaaatagcgggtgagagagaacgag    450
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       401 ATTATCATATTACCAGCCCCGTTACGAAATAGCGGGTGAGAGAGAACGAG    450

EMBOSS_001       451 agtatgagctggacatcaatgctctcattacggacggacggcttcacgta    500
                     ||||||||||||||||||||.||||||||||||||||||||||||.|||||
EMBOSS_001       451 AGTATGAGCTGGACATCAACGCTCTCATTACGGACGGACGGCTTCAAGTA    500

EMBOSS_001       501 aaagcggtatacacccaagtgttcagcaaacactcaattgagtgctttat    550
                     ||||||||||||||||||.|.||||||||||||||||||||||||||||||
EMBOSS_001       501 AAAGCGGTATACACCCAGGCGTTCAGCAAACACTCAATTGAGTGCTTTAT    550

EMBOSS_001       551 ggatcgttttcacaggcatcttatcgagacgattgagcattgttcgcaga    600
                     ||||||||||||.||||||||||||||||||||||||||||||||||||||
EMBOSS_001       551 GGATCGTTTTCATAGGCATCTTATCGAGACGATTGAGCATTGTTCGCAGA    600

EMBOSS_001       601 aaaaagcacgcgaaaaaacattgagcgatttcagcaataaagaattaacg    650
                     ||||||||||||||||||||||||||||||||||||||||.|||||||||
EMBOSS_001       601 AAAAAGCACGCGAAAAAACATTGAGCGATTTCAGCAATAAGGAATTAACG    650

EMBOSS_001       651 ttgtctgcattatctagcatcgaagatctcgtgaaagatctttaacttttt    700
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       651 TTGTCTGCATTATCTAGCATCGAAGATCTCGTGAAAGATCTTTAACTTTT    700

EMBOSS_001       701 gaaaaggtgtgtggaattgatggcacaatcggcacaaattcaagatattt    750
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       701 GAAAAGGTGTGTGGAATTGATGGCACAATCGGCACAAATTCAAGATATTT    750

EMBOSS_001       751 atc    753
                     |||
EMBOSS_001       751 ATC    753
```

## *ppsC-ppsD* fragment Needle alignment of *B. subtilis* BBG21 and 168

```
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 901
# Identity:     879/901 (97.6%)
# Similarity:   879/901 (97.6%)
# Gaps:           3/901 ( 0.3%)
# Score: 5015.0
#
#
#=======================================

EMBOSS_001         1 gatacacaagatattctgttaaccgcagcatccttggcgatttgtgaatg     50
                     ||||| ||||||||||| |||||||||||||||||||||||||||||||
EMBOSS_001         1 GATACGCAAGATATTCTTTTAACCGCAGCATCCTTGGCGATTTGTGAATG     50

EMBOSS_001        51 gacaggaggaagcaagctccgtatcgctatggaaggacacgggagagaac    100
                     ||||||||||||.|.|||||||||||||||||||||||||||||||||||
EMBOSS_001        51 GACAGGAGGAAGTATGCTCCGTATCGCTATGGAAGGACACGGGAGAGAAC    100

EMBOSS_001       101 acattctgccagagttagatataagcaggacggtcggctggtttacatcc    150
                     |.|||||||||||||||||||| |||||||||||||||||||||||||||
EMBOSS_001       101 ATATTCTGCCAGAGTTAGATATCAGCAGGACGGTCGGCTGGTTTACATCC    150

EMBOSS_001       151 atgtacccggccctcataagctttgaaaatcatagggatgagctcggaac    200
                     |||||.|||||||||||||||||||||||||||||.|||||||||||||||
EMBOSS_001       151 ATGTATCCGGCCCTCATAAGCTTTGAAAATCATAGGTATGAGCTCGGAAC    200

EMBOSS_001       201 ctcagtaaaaaccgttaaagatacactgggccggataccaaacaaagggg    250
                     |.||.|||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       201 CGCATTAAAAACCGTTAAAGATACACTGGGCCGGATACCAAACAAAGGGG    250

EMBOSS_001       251 taggttacggtatgctgaaatacctacacaccctgaaaataaaagcatt    300
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       251 TAGGTTACGGTATGCTGAAATACCTTACACACCCTGAAAATAAAAGCATT    300

EMBOSS_001       301 acgttttcaaaaacacctgaaatcagctttaactatttgggccagttcaa    350
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       301 ACGTTTTCAAAAACACCTGAAATCAGCTTTAACTATTTGGGCCAGTTCAA    350

EMBOSS_001       351 cgacatcgaaagacaggacaccttccgcccatcaagccttggtagcggaa    400
                     ||||||||||||||||||||||||||||||||||||||||||.|||||||||
EMBOSS_001       351 CGACATCGAAAGACAGGACACCTTCCGCCCATCAAGCCTTGGGAGCGGAA    400

EMBOSS_001       401 aggatatcactcatacatggaagcgcgagcaaatcatcgagatgagtgca    450
                     |||||||||||||||||||||||.|||||||||||||||||||||.|||||
EMBOSS_001       401 AGGATATCACTCATACATGGAAGCGTGAGCAAATCATCGAGATGAGCGCA    450

EMBOSS_001       451 atggcagcagacaaaaagctccatttcaatcttagttatccgccagcccg    500
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       451 ATGGCAGCAGACAAAAAGCTCCATTTCAATCTTAGTTATCCGCCAGCCCG    500

EMBOSS_001       501 ctttcacagaaacacaatggaacagcttattaacaggattgaacacttt    550
                     ||||||||||||||||||||||.|||||||||||||||||||||||||||
EMBOSS_001       501 CTTTCACAGAAACACAATGGAGCAGCTTATTAACAGGATTGAACACTTTT    550

EMBOSS_001       551 tgctggacattatgaaacactgcgccggtcaacagaaagcagaaaagaca    600
                     |||||||||||||||||||||||||||||||||||||||.|||||||||||
EMBOSS_001       551 TGCTGGACATTATGAAACACTGCGCCGGTCAACAGAAAACAGAAAAGACA    600

EMBOSS_001       601 ctgagcgatttcagcagtcaatcattaacggcagaggacttggacagcat    650
                     |||||||||||||||||||||||||||||||||.||||||||||||||||||
EMBOSS_001       601 CTGAGCGATTTCAGCAGTCAATCATTAACGGCTGAGGACTTGGACAGCAT    650

EMBOSS_001       651 atccagcttggtggaagaattgtagaaaatcgtgaca-ggagacatgaac    699
                     |||||||||||||||||||||||| |||||||||||| ||||||||||||||
EMBOSS_001       651 ATCCAGCTTGGTGGAAGAATTGTAG-AAATCGTGACAGGGAGACATGAAC    699

EMBOSS_001       700 atgacaaaagcgaattcaattcaggatatttatccgctgtcttacatgca    749
                     |||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       700 ATGACAAAAGCGAATTCAATTCAGGATATTTATCCGCTGTCTTACATGCA    749

EMBOSS_001       750 ggaagggatgcttttccattccctcctgcaaaaagattcacaggcgtatg    799
                     |||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       750 GGAAGGGATGCTTTTCCATTCCCTCCTGCAAAAAGATTCACAGGCGTATG    799

EMBOSS_001       800 ttgagcaagcttctttttacaatag-aaggaaaagtcaccccgcagttttt    848
                     |||||||||||||||||||||||.| ||||||||||||.|||||||||||||
EMBOSS_001       800 TTGAGCAAGCTTCTTTTACAATTGAAAGGAAAAGTCAACCCGCAGTTTTT    849

EMBOSS_001       849 tcaaaacagcatcaacgctcttgtagaaagacatgatattttcagaacga    898
                     |||||||||||||||||||||||||||||||||||||||||||||||||.|
EMBOSS_001       850 TCAAAACAGCATCAACGCTCTTGTAGAAAGACATGATATTTTCAGAACCA    899

EMBOSS_001       899 t     899
                     |
EMBOSS_001       900 T     900
```

## Plipastatin fragment BLAST alignment of *B. subtilis* BBG21 and 168

```
 Features in this part of subject sequence:
   plipastatin synthetase

 Score = 1544 bits (836),  Expect = 0.0
 Identities = 858/869 (99%), Gaps = 0/869 (0%)
 Strand=Plus/Minus

Query  60       TGAAGACCGGCGGGGCGTATTTGCCGCTTGATCCTGCGTATCCAAAGGAGCGGCTGAGCT  119
                ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1996372  TGAAGGCCGGCGGGGCGTATTTGCCGCTTGATCCTGCGTATCCAAAGGAGCGGCTGAGCT  1996313

Query  120      ACATGCTGAAAGACAGCGGGGCATCGCTCTTGCTGACACAGCCAGGATGCTCCGCACCAA  179
                |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Sbjct  1996312  ACATGCTGAAAGACAGCGGGGCATCGCTCTTGCTGACACAGCCGGGATGCTCCGCACCAA  1996253

Query  180      ACTTTTCTGGCGAAACGCTTGAAGTGGACATGACATCTCTTGCGAGCGAAAAAGCAGAAA  239
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1996252  ACTTTTCTGGCGAAACGCTTGAAGTGGACATGACATCTCTTGCGAGCGAAAAAGCAGAAA  1996193

Query  240      ACCATGAGTTTACTCCTGCTGACGGCGGATCTCTCGCCTATGTCATCTATACATCGGGCT  299
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1996192  ACCATGAGTTTACTCCTGCTGACGGCGGATCTCTCGCCTATGTCATCTATACATCGGGCT  1996133

Query  300      CTACCGGGCAGCCGAAAGGAGTGGCCGTTGAGCACCGCCAGGCTGTTTCCTTTCTGACCG  359
                |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Sbjct  1996132  CTACCGGGCAGCCGAAAGGAGTGGCCGTTGAGCACCGTCAGGCTGTTTCCTTTCTGACCG  1996073

Query  360      GCATGCAGCACCAGTTTCCGCTTTCGGAAGATGACATTGTCATGGTGAAAACCTCTTTTT  419
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1996072  GCATGCAGCACCAGTTTCCGCTTTCGGAAGATGACATTGTCATGGTGAAAACCTCTTTTT  1996013

Query  420      CCTTTGATGCGTCTGTGTGGCAGCTGTTTTGGTGGACTCTTTCCGGTGCTTCGGCGTATC  479
                |||||||||||||||||||||||||||||||||| | ||||||||||||||||||||||
Sbjct  1996012  CCTTTGATGCGTCTGTGTGGCAGCTGTTTTGGTGGTCGCTTTCCGGTGCTTCGGCGTATC  1995953

Query  480      TGCTTCCGCCCGGCTGGGAAAAAGACCCCGCATTGATCGTACAAGCCATTCATCAGGAAA  539
                |||||||||||||||||| |||||| ||||||||||||||||||||||||||||||||||
Sbjct  1995952  TGCTTCCGCCCGGCTGGGAGAAAGACTCCGCATTGATCGTACAAGCCATTCATCAGGAAA  1995893

Query  600      TCGAAAGGCTGAGTGACAGGACAAGCCTGAAGCGTGTATTCGCCGGAGGTGAACCGCTTG  659
                |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Sbjct  1995832  TCGAAAGGCTGAGTGACAGGACAAGCCTAAAGCGTGTATTCGCCGGAGGTGAACCGCTTG  1995773

Query  660      CGCCGCGTACGGCAGCCCGTTTTGCTTCTGTATTGCCGCAGGTATCATTAATCCATGGTT  719
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1995772  CGCCGCGTACGGCAGCCCGTTTTGCTTCTGTATTGCCGCAGGTATCATTAATCCATGGTT  1995713

Query  720      ACGGGCCGACAGAAGCAACGGTAGACGCGGCATTTTACGTATTGCATCCAGAACGGGACA  779
                |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Sbjct  1995712  ACGGGCCGACAGAAGCAACGGTAGACGCGGCATTTTACGTATTGGATCCAGAACGGGACA  1995653

Query  780      GGGATCGTTTGCGGATTCCGATCGGGAAGCCCGTACCTGGTGCGCGTTTGTATGTTTTAG  839
                |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Sbjct  1995652  GGGATCGTTTGCGGATTCCGATTGGGAAGCCCGTACCTGGTGCGCGTTTGTATGTTTTAG  1995593

Query  840      ATCCGCATTTAGCCGTACAGCCTTCCGGTGTCGCCGGCGAACTGTACATTGCCGGAGCGG  899
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1995592  ATCCGCATTTAGCCGTACAGCCTTCCGGTGTCGCCGGCGAACTGTACATTGCCGGAGCGG  1995533

Query  900      GTGTTGCCAGAGGTTATTTGAATCGGCCG  928
                |||||||||||||||||||||||||||||
Sbjct  1995532  GTGTTGCCAGAGGTTATTTGAATCGGCCG  1995504
```

**Index 5: Different alignments of sequenced fragments from *B. subtilis* S499 *ppsA-ppsB* fragment BLAST alignment of *B. subtilis* S499 and *B. amyloliquefaciens* FZB42**

```
>gb|CP000560.1|  Bacillus amyloliquefaciens FZB42, complete genome
Length=3918589

 Features in this part of subject sequence:
   FenB
   FenA

 Score = 1258 bits (681),  Expect = 0.0
 Identities = 731/756 (96%), Gaps = 0/756 (0%)
 Strand=Plus/Minus

Query  5        TGCTGTTGGAAACCCAGCAGGCGTACGGAACGGATGTCCATGAATTGCTTTTAACGGCGC  64
                |||||||||||| ||||||| |||||||||||||| ||||||||||||||||||||||
Sbjct  1962027  TGCTGTTGGAAACACAGCAGGCATACGGAACGGATGTTCATGAATTGCTTTTAACGGCGC  1961968

Query  65       TTGGAATGGCGCTTCATGAGTGGACGGGTTCCGAACAGATCCTTATCAGCACTGAAGGAC  124
                |||||||||||||||||||||||||||||| |||||||||||||||||||||| || |
Sbjct  1961967  TTGGAATGGCGCTTCATGAGTGGACGGGTTCTGAACAGATCCTTATCAGCACTGAGGGGC  1961908

Query  125      ACGGACGCCAGGGGCATATTCCTGACATTGATATGTCAAGAACCGCAGGCCGGTTTACAT  184
                |||| |||||||||||||| |||||||||||||||||||||| |||||  |||||||||
Sbjct  1961907  ACGGGCGCCAGGGGCATATGCCTGACATTGATATGTCAAGAACGGCAGGATGGTTTACAT  1961848

Query  185      CTGTTTACCCGGTGCTTTTGGATATGAGAGTGCCGGACGGCTCAGGCGAAACAACCGCGC  244
                |||||||||||||||||||||||||||||||||||||||||| ||||||||| |||||| |
Sbjct  1961847  CTGTTTACCCGGTGCTTTTGGATATGAGAGTGCCGGACGGCACAGGCGAAAAAACCGCTC  1961788

Query  245      ACCGCATAAAGACGGTAAAAGACATGATGAGACGGGTTCCTCATCACGGAACCGGCTACG  304
                || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1961787  ACAGCATAAAGACGGTAAAAGACATGATGAGACGGGTTCCTCATCACGGAACCGGCTACG  1961728

Query  305      GGCTGCTGAGATATTCGCGAAACATTCTTTATAAAGACCCAGAAATAAGCTTTAACTATC  364
                |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1961727  GGCTGCTGAGGTATTCGCGAAACATTCTTTATAAAGACCCAGAAATAAGCTTTAACTATC  1961668

Query  365      TCGGTGCGTTTGACGGTCTTGAGGGGACGCTGAAGGTGTCACCGTTCAAGCCGCAGCATG  424
                |||||||||||||||||||||||| ||||||||| |||||||||||||||| |||||||
Sbjct  1961667  TCGGTGCGTTTGACGGTCTTGAGGGAACGCTGAAGATGTCACCGTTCAAGCCACAGCATG  1961608

Query  425      ATATAGCGGCCGGGCGCATCCGCGAGTTTGATCTGGATATTAATGCGATGGTGGCGGCCG  484
                |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Sbjct  1961607  ATATAGCGGCCGGGCGCATCCGCGAGTTTGATCTGGATATTAACGCGATGGTGGCGGCCG  1961548
```

## *ppsB-ppsC* fragment BLAST alignment of *B. subtilis* S499 and *B. amyloliquefaciens* FZB42

```
>gb|CP000560.1| Bacillus amyloliquefaciens FZB42, complete genome
Length=3918589

                                                    Sort alignments for this subject sequence by:
                                                      E value  Score  Percent identity
                                                      Query start position  Subject start position

 Features in this part of subject sequence:
   FenC
   FenB

 Score = 1533 bits (830),  Expect = 0.0
 Identities = 878/902 (97%), Gaps = 0/902 (0%)
 Strand=Plus/Plus

Query  924      TGCTACATGAGCCATTTAATATACTTCCCATATGGCTGAACAGGCTCGAGAGACGGAAGC  983
                ||||  ||||||||||||||||||||| |||||||||||||||||||||||||||| |||
Sbjct  1953029  TGCTTCATGAGCCATTTAATATACTTACCATATGGCTGAACAGGCTCGAGAGACGGGAGC  1953088

Query  984      CGTCCGTCTCCGAGAGATTGATAGATCACCAGAAATTCCTTCATGACAATTCCGAAGCAC  1043
                |  |||||||||||||||||||||||| |||||||||||||||||||| |||||||||
Sbjct  1953089  CACCCGTCTCCGAGAGATTGATAGATCATCAGAAATTCCTTCATGACAATCCCGAAGCAC  1953148

Query  1044     CAGCCGTCCATGACAATATGATGATGGCTCCATATACAAACATAGCGTTCAGGCGCTTTT  1103
                |||||||||||||| |||||||||||||||||||||||||||||||||||||| |||||
Sbjct  1953149  CAGCCGTCCATGACGATATGATGATGGCTCCATATACAAACATAGCGTTCAGGCCCTTTT  1953208

Query  1104     TTCAAAATGGAGACTCTCATGAGAGGGTCTGATTGCAAATCAAAGCCTTTATCTTTATCT  1163
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1953209  TTCAAAATGGAGACTCTCATGAGAGGGTCTGATTGCAAATCAAAGCCTTTATCTTTATCT  1953268

Query  1164     TCTTTTCTGAACCGGTCGGCATATTTGTCTTGCGCCTTCTCGTCCAAATGAGAAATATCA  1223
                || |||||||| ||||||||||||||| ||||| ||||||||||||||||||||||||||
Sbjct  1953269  TCCTTTCTGAATCGGTCGGCATATTTATCTTGTGCCTTCTCGTCCAAATGAGAAATATCA  1953328

Query  1224     TGAAACTGAACGCGGCTCGGCCTGTTTTTCAACACGACTTGTCGGGGCTTTGCCACATTC  1283
                |||||||||||||||||||||||||||||||||||||||||| ||||||||| |||||||||
Sbjct  1953329  TGAAACTGAACGCGGCTCGGCCTGTTTTTCAACACGACTTGGCGGGGCTTCGCCACATTC  1953388

Query  1284     TTGTGAATAAATGTCGTACGGAAAATATCGTATCGTTCAAAAACAGTGTCTATACTTTTT  1343
                || |||||||||||||||||||||||||||||||||||||||||||  |||||||||||||
Sbjct  1953389  TTATGAATAAATGTCGTACGGAAAATATCGTATCGTTCAAAAACAGCATCTATACTTTTT  1953448
Query  1344     TGGAACCGTTCCCGGTCAAGCTTTCCGCTGATCGTAAACACCGCTTGTTCGAAGTAGGCT  1403
                ||||| ||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Sbjct  1953449  TGGAATCGTTCCCGGTCAAGCTGTCCGCTGATCGTAAACACCGCTTGTTCGAAGTAGGCT  1953508

Query  1404     CTTGAGTCATGATCCAACAGAGAATGAAAAAGCATGCCCTCCTGCATAAAAGACAAAGGA  1463
                |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Sbjct  1953509  CTTGAGTCATGATCCAACAGAGAATGAAAAAGCATGCCCTCCTGCATAAAAGATAAAGGA  1953568

Query  1464     TATATATCTTGTATTTCCGGTTGCTGCCCCATTGCGCCACTCCTTTTCTATAGTTTGTTG  1523
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1953569  TATATATCTTGTATTTCCGGTTGCTGCCCCATTGCGCCACTCCTTTTCTATAGTTTGTTG  1953628

Query  1524     ACGGCTCCCATAATGTCACTCAGATCTTCCAGCGTTAACTCCTCGTCACTGAAATCAGCG  1583
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1953629  ACGGCTCCCATAATGTCACTCAGATCTTCCAGCGTTAACTCCTCGTCACTGAAATCAGCG  1953688

Query  1584     GCGCTCCATTCACGCTCCTCTTTGCCTGTGCAATGCGTGATGATCTGCTTGAGAAAGGTA  1643
                |||||||||||||||||||||||||||||||||||| |||||||||||||||||||| |
Sbjct  1953689  GCGCTCCATTCACGCTCCTCTTTGCCTGTGCAATGCGAGATGATCTGCTTGAGAAAGGCA  1953748

Query  1644     TGAAAATATCCCGTCAAAACTTCAATCGTTTCTTTTTGATACTGAAAACGGTTGTACACG  1703
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1953749  TGAAAATATCCCGTCAAAACTTCAATCGTTTCTTTTTGATACTGAAAACGGTTGTACACG  1953808

Query  1704     ACATGCATGCTGAGACATTCCGACGATACGGAGCCGCTGATATCAAGTGTGTACGGACGT  1763
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1953809  ACATGCATGCTGAGACATTCCGACGATACGGAGCCGCTGATATCAAGTGTGTACGGACGT  1953868

Query  1764     TCCCAGTCTGGGCTGACTTCATTTCCCGCTTTTACGGGAGAAACACCGAAGCCGCTGGAA  1823
                |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Sbjct  1953869  TCCCAGTCTGGGCTGACTTCATTTCCCGCTTTTACGGGAGAAACACCAAAGCCGCTGGAA  1953928

Query  1824     CG  1825
                ||
Sbjct  1953929  CG  1953930
```

## *ppsC-ppsD* fragment BLAST Alignment of *B. subtilis* S499 and *B. amyloliquefaciens* FZB42

```
>gb|CP000560.1| Bacillus amyloliquefaciens FZB42, complete genome
Length=3918589

                                                    Sort alignments for this subject sequence by:
                                                      E value  Score  Percent identity
                                                      Query start position  Subject start position

 Features in this part of subject sequence:
   FenD
   FenC

 Score = 1620 bits (877),  Expect = 0.0
 Identities = 916/935 (97%), Gaps = 1/935 (0%)
 Strand=Plus/Minus

Query  914      GATACGCAAGATATTCTACTGACTGCCGCAGCTTTGGCTTTACGGGACTGGACAGGGGGC  973
                ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Sbjct  1946625  GATACGCAAGATATTCTTCTGACTGCCGCAGCTTTGGCTTTACGGGACTGGACAGGGGGC  1946566

Query  974      GGCAGACTGCGCATCGCGATGGAAGGGCACGGGAGAGAGCACATCATGCCGGAGCTCGAC  1033
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1946565  GGCAGACTGCGCATCGCGATGGAAGGGCACGGGAGAGAGCACATCATGCCGGAGCTCGAC  1946506

Query  1034     ATCAGCCGCACCGTAGGCTGGTTTACCTCCATGTACCCGGTGCTCATTGACTTGAACACA  1093
                |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Sbjct  1946505  ATCAGCCGCACCGTAGGCTGGTTTACCTCCATGTATCCGGTGCTCATTGACTTGAACACA  1946446

Query  1094     GCAGGCGCTGAGCTTGGAACGGCTGTAAAAACGGTCAAAGACACACTCGGAAGAATACCG  1153
                ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Sbjct  1946445  GCAGGCGCTGAGCTTGGAACGGCGGTAAAAACGGTCAAAGACACACTCGGAAGAATACCG  1946386

Query  1154     GATAAAGGAATAGGCTACGGCATTTTAAAATACATGACACCGCCTGAACAAAAAACAATC  1213
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1946385  GATAAAGGAATAGGCTACGGCATTTTAAAATACATGACACCGCCTGAACAAAAAACAATC  1946326

Query  1214     CGCTTCAGACAAGCGCCTGAAATCAGTTTTAACTATTTAGGGCAATTCAATGATACAGAA  1273
                |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Sbjct  1946325  CGCTTCAGACAAGCGCCTGAAATCAGCTTTAACTATTTAGGGCAATTCAATGATACAGAA  1946266

Query  1274     GAGCAGCACACATTCAGTTTATCCGGTCTTGCGAGCGGGCACGACATCACGCCGACATGG  1333
                || |||||||| |||||| | |||||||||||| ||||||||||||||||||||||||||
Sbjct  1946265  GATCAGCACACGTTCAGTCTGTCCGGTCTTGCAAGCGGGCACGACATCACGCCGACATGG  1946206

Query  1334     CAGCGGGAACAAGCGGTGGAAATGAGCGCGATGGCCGCTCAACATAAACTGCACTTCAGT  1393
                |||||||||||||||||||||||||||||||||||||||| ||| | ||||| |||||||||
Sbjct  1946205  CAGCGGGAACAAGCGGTGGAAATGAGCGCGATGGCCGCACAAAACAAACTTCACTTCAGT  1946146

Query  1394     TTGAGCTATCCGCCGAGCCGCTTCCGCAAAGAGACGATGGAACAGCTGCTGCAAACCTCA  1453
                |||||||||||||||||||||||||||||||| || |||||||||||||||||| | | |
Sbjct  1946145  TTGAGCTATCCGCCGAGCCGCTTCCGCAAAGAAACAATGGAACAGCTGCTGCAAACTTTA  1946086

Query  1454     CAGCAATATTTGCGTGATATTATGGCGCACTGCACAGGCAAAGAGGAAGCAGAAAAAACC  1513
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1946085  CAGCAATATTTGCGTGATATTATGGCGCACTGCACAGGCAAAGAGGAAGCAGAAAAAACC  1946026

Query  1514     GTTAGTGATTTTAGCAGCAAAACGTTAACATCTGATGATTTGGACAGCATAGCGAGTTTT  1573
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1946025  GTTAGTGATTTTAGCAGCAAAACGTTAACATCTGATGATTTGGACAGCATAGCGAGTTTT  1945966

Query  1574     GTCGAGGAACTGTAGAAAAAGGCGACGGGGAGATTATAACATGACAAAAAAGAATGCAAT  1633
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1945965  GTCGAGGAACTGTAGAAAAAGGCGACGGGGAGATTATAACATGACAAAAAAGAATGCAAT  1945906

Query  1634     TCAGGATATTTATCCGCTGTCTTATATGCAGGAAGGAATGCTTTTTCATTCCCTCCTGCA  1693
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1945905  TCAGGATATTTATCCGCTGTCTTATATGCAGGAAGGAATGCTTTTTCATTCCCTCCTGCA  1945846

Query  1694     AAAGGAATCACAGGCTTACGCCGAGCAGGCCTCCTTTTCAATAACGGGGAAAGGTTGACA  1753
                |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Sbjct  1945845  AAAGGAATCACAGGCTTACGCCGAGCAGGCCTCCTTTTCAATAACGGGGAA-GGTTGACA  1945787

Query  1754     CCCGCGTGTTTGAAGAGAGCGTTCATGCCCTTTTCGAAAGACATGATATTTTCAGAACGA  1813
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1945786  CCCGCGTGTTTGAAGAGAGCGTTCATGCCCTTTTCGAAAGACATGATATTTTCAGAACGA  1945727

Query  1814     TTTTTATCAGCCAGAACGTGTCAGTGCCGCAGCAG  1848
                | ||||||||||||||||||||||||||||||||
Sbjct  1945726  TCTTTATCAGCCAGAACGTGTCAGTGCCGCAGCAG  1945692
```

## *ppsD-ppsE* fragment BLAST alignment of *B. subtilis* S499 and *B. amyloliquefaciens* FZB42

```
 Features in this part of subject sequence:
   FenE
   FenD

 Score = 1563 bits (846),  Expect = 0.0
 Identities = 887/907 (97%), Gaps = 1/907 (0%)
 Strand=Plus/Plus

Query  10       TTTTCAGCTTGTCTGTCAGATTTTGATCTAAAGAAAACAGGAGCTCTTTCTTACGGTAGC  69
                ||||||||||||| ||||| |||||||||| ||||||||| ||||||||||||||||||
Sbjct  1934396  TTTTCAGCTTGTCCGTCAGGTTTTGATCTATAGAAAACAGCAGCTCTTTCTTACGGTAGC  1934455

Query  70       CTTTTTCCTGGGGACGGGCCGATATTTTCGGGAGTCCTGAGGGTGTCTCGCAGCCTGCTA  129
                ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Sbjct  1934456  CTTTTTCCTGGGGACGGGCCGATATTTTCGGGAGTCCTGAGGGTGTCTCACAGCCTGCTA  1934515

Query  130      AATATCCTTTCCAATACTCAGAAGCCGCGTCTTTGTCCTGATCAGTCAGCCATGTAATAT  189
                |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Sbjct  1934516  AATATCCTTTCCAATACTCAGAAGCCGCGTCTTTGTCCTGATCCGTCAGCCATGTAATAT  1934575

Query  190      ACGTGCTGTAAGGCTTTGCGGGCGCCAATGTGACGGGTGAGCCTGATTCAATCGCCCGGT  249
                | ||||||||||||||||||||||||| |||||||||||||||| ||||||||||||||
Sbjct  1934576  AAGTGCTGTAAGGCTTTGCGGGCGCCAGTGTGACGGGTGAGCCTGTTTCAATCGCCCGGT  1934635

Query  250      ACATATGCATGAATTCCTGAAGGACGATGCCGAGGCACCAGCCGTCCATAAGGATGTGAT  309
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1934636  ACATATGCATGAATTCCTGAAGGACGATGCCGAGGCACCAGCCGTCCATAAGGATGTGAT  1934695

Query  310      GGTGGCTCCAGACGCATACATATTTTTCGTCAGCCGTTTTAAACAAAGACACACGCATCA  369
                |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Sbjct  1934696  GGTGGCTCCAGACGCATACATATTTTTCGTCAGCCGTTTTAAACAAAGAAACACGCATCA  1934755

Query  370      GCATATCCTTTTGCAAGTCAAAGCCTTTTAAGCGGTCACGTTCCTTAAATCGCTGTAAGT  429
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Sbjct  1934756  GCATATCCTTTTGCAAGTCAAAGCCTTTTAAGCGGTCACGTTCCTTAAATCGCTGTAAAT  1934815

Query  430      AGGCGCTCTGTTCCGCTTCATCCAAATGTGTAAGATCTTCGCGGTGCAGCCGGAATTCCC  489
                ||  |||||||||||| ||| ||||||||||||||||||||||||||||||||||||||
Sbjct  1934816  AGATGCTCTGTTCCGGTTCGTCCAAATGTGTAAGATCTTCGCGGTGCAGCCGGAATTCCC  1934875

Query  490      GCTCACGCAGCACGACTTGGCGGGGGCCGTTCAGCTGAGCGACATGAGGTaaaaaaaCCG  549
                |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1934876  GCTCACGCAGCACAACTTGGCGGGGGCCGTTCAGCTGAGCGACATGAGGTAAAAAAACCG  1934935

Query  550      TTCTGAAAATATCGTGACGGCTGATGACAGACTGTACGCTTTGCTCAAACAATTCCGGAC  609
                |||||||||||||||| || ||||||||||||||||||||||||||||||||||||||||
Sbjct  1934936  TTCTGAAAATATCGTGGCGTCTGATGACAGACTGTACGCTTTGCTCAAACAATTCCGGAC  1934995

Query  610      GAAGCCGGCCTTTTATGGTGAAGACGGATTGCTCGATATATGCAGCCCCCTCCTTTTGCA  669
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1934996  GAAGCCGGCCTTTTATGGTGAAGACGGATTGCTCGATATATGCAGCCCCCTCCTTTTGCA  1935055

Query  670      GAAAGGAATGAAAAAGCATTCCTTCCTGCATTCGGCTGAGGGGATAAATATTTTGGATAT  729
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1935056  GAAAGGAATGAAAAAGCATTCCTTCCTGCATTCGGCTGAGGGGATAAATATTTTGGATAT  1935115

Query  730      TTTTTGTTTTGTCCATGTGTTCCTCCCGCTCGGTTTAGTGGAAGCTGAGTAAGTCGGCAA  789
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1935116  TTTTTGTTTTGTCCATGTGTTCCTCCCGCTCGGTTTAGTGGAAGCTGAGTAAGTCGGCAA  1935175

Query  790      TATCCTGCAAGGCTTCCTCAGTCAGCTCCCGATCATCAAAGTCACTGACGGTCTTCTCCG  849
                |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1935176  TATCCTGCAATGCTTCCTCAGTCAGCTCCCGATCATCAAAGTCACTGACGGTCTTCTCCG  1935235

Query  850      TATCGGTTTTGCTGAGGCAGTGTTCCGACAGTTTCAGGAGATA-TACCGGCAGCTTTCGG  908
                |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Sbjct  1935236  TATCGGTTTTGCTGAGGCAGTGTTCCGACAGTTTCAGGAGATAATACCGGCAGCTTTCGG  1935295

Query  909      ACAGCCG  915
                |||||||
Sbjct  1935296  ACAGCCG  1935302
```

## *ppsA-ppsB* fragment alignment of *B. subtilis* S499 and 168

```
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 794
# Identity:     518/794 (65.2%)
# Similarity:   518/794 (65.2%)
# Gaps:          88/794 (11.1%)
# Score: 1585.5
#
#
#=======================================

EMBOSS_001         1 TGCTGTTGGAAACCCAGCAGGCGTACGGAACGGATGTCCATGAA---TTG     47
                     |||||||.||||||||||.||.||.||.||.||||.|.|.|||   |||
EMBOSS_001         1 tgctgtttgaaacccagcaagcctatggtactgatgccaacgaacttttg     50

EMBOSS_001        48 CTTTTAACGGCGCTTGGAATGGCGCTTCATGAGTGGACGGGTTCCGAACA     97
                     |||   |||||..|.||||||||||||..|||.|||||.||....||.||
EMBOSS_001        51 ctt---acggctttaggaatggcgctttctgaatggactgggcatgacca     97

EMBOSS_001        98 GATCCTTATCAGCACTGAAGGACACGGACGCCAGGGGCATATTCCTGACA    147
                     .|||.|.||||||||||.||||||||..|..|.||||||.|.||..|||
EMBOSS_001        98 aatcgtgatcagcactgagggacacggtagagaagggcatgtgccaaaca    147

EMBOSS_001       148 TTGATATGTCAAGAACCGCAGGCCGGTTTACATCTGTTTACCCGGTGCTT    197
                     |||||||.|||||.||.|..||..|||||||.||..||||.||..|.||.
EMBOSS_001       148 ttgatatttcaaggacagttggatggtttacctccatttatcctattctg    197

EMBOSS_001       198 TTGGATATGAGAGTGCCGGACGGC---TC-AGGCGAAACAAC-CGCGCAC    242
                     .|.||.|||.|..|.||.||  ||   || |||   |.||.| |||..||
EMBOSS_001       198 ctagacatgggtattcccga--gcctttcgagg---atcagctcgcatac    242

EMBOSS_001       243 CGCATAAAGACGGTAAAAGACATGATGAGACGGGTTCCTCATCACGGAAC    292
                     ||.||.||||||..||||||.|||.|||||.|.||.||..|..|.|||||
EMBOSS_001       243 cgaattaagacgacaaaagatatgctgagaagagtaccgaacaaaggaac    292

EMBOSS_001       293 CGGCTACGGGCTGCT-----------GAGATATTC-GCGAAACATTCTTT    330
                     .||.|||||..||.|           ||||..|.| ||
EMBOSS_001       293 tggatacggattgttaacacacataggagaactgcggc------------    330

EMBOSS_001       331 ATAAAGACCCAGAAATAAGCTTTAACTATCTCGGTGCGTTTGACGGTCTT    380
                     |||||||.||.||..|.||||||||.|||||||||...|||..|
EMBOSS 001       331 ataaagaccctgaagtcagctttaattatctcggtcaatttagc------    374
EMBOSS_001       381 GAGGGGACGCTGAAGGTGTCACCGTTCA--------------AGCCGCAG    416
                     ||.|.||.|  ||||.||..||..||||              ||||.|..
EMBOSS_001       375 gaagagaag--gaagttgaaacgtttcaattatcatattaccagccccgt    422

EMBOSS_001       417 CATGATATAGCGGCCGGGCGCATCCGCGAGTTTGATCTGGATATTAATGC    466
                     .|.||.|||||||..|.|.|....||.||||.|||.|||||.||.|||||
EMBOSS_001       423 tacgaaatagcgggtgagagagaacgagagtatgagctggacatcaatgc    472

EMBOSS_001       467 GATGGTGGCGGCCGGGCAACTGGAGGTAAAAGCCGTTTATACCGATGTAT    516
                     ..|..|..|||.|||.|..||..|.||||||||.||.||.|||.|.||.|
EMBOSS_001       473 tctcattacggacggacggcttcacgtaaaagcggtatacacccaagtgt    522

EMBOSS_001       517 T-----TGCACCGGGGGCAATCGAGTTCTTCATGAATCGTTTCCATACTC    561
                     |     ..|||     .||||.||||.|||.|||.|||||||.||.|..|
EMBOSS_001       523 tcagcaaacac-----tcaattgagtgctttatggatcgttttcacaggc    567

EMBOSS_001       562 ACCTCGGGGACATCATTGCCCACTGTTCCGGC--AAAAAAGAGCGGGAAA    609
                     |.||....||.|..||||..|||.|||||  ||  |||||||..||.||||
EMBOSS_001       568 atcttatcgagacgattgagcattgttc--gcagaaaaaagcacgcgaaa    615

EMBOSS_001       610 AAACATTAACGGATTACAGCCATACAGAATTGACCG----CTCAAGCCCT    655
                     |||||||.|..||||.||||.|||.|||||| |.||    ||   ||..|
EMBOSS_001       616 aaacattgagcgatttcagcaataaagaatt-aacgttgtct---gcatt    661

EMBOSS_001       656 ATCAAGCATTGAAGATCTCGTAAAAGGACTTTAA-TTTCATAAAAAGGTG    704
                     |||.|||||.||||||||||.||||..|||||| |||  |.|||||||||
EMBOSS_001       662 atctagcatcgaagatctcgtgaaagatctttaacttt--tgaaaaggtg    709

EMBOSS_001       705 TGTGGCATTCGTGACACAAGCGACAGAAATTCAAGATATTTAC-    747
                     |||||.|||..||.||||||.||.||.||||||||||||||||.
EMBOSS_001       710 tgtggaattgatggcacaatcggcacaaattcaagatatttatc    753
```

## *ppsB-ppsC* fragment alignment of *B. subtilis* S499 and 168

```
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 1917
# Identity:    1237/1917 (64.5%)
# Similarity:  1237/1917 (64.5%)
# Gaps:         250/1917 (13.0%)
# Score: 3741.5
#
#
#=======================================

EMBOSS_001          1 TGTTCGAGCATCAGACTCTGGGAGAAGTGTCGATGCGGCTCCGGAAAGTC      50
                      ||||.||.||||||||||||.||.|||.||||....|||.|.||.||||.|
EMBOSS_001          1 tgtttgaacatcagactcttggtgaactgtccgcccgggttcgaaaagac      50

EMBOSS_001         51 CGCCGCGTGATTGATC-AGGCTCCCGTC--ACGGGAG---ACGTGCCGCT      94
                      ...|||||..||||||.| |||| ||.|||  |.|.|||    ||.||
EMBOSS_001         51 gtacgcgccattgaccaaggc-ccggtcgaaggcgagattacttg-----      94

EMBOSS_001         95 GACGCCGATCCAGCACTGGTTCTTTTCGCAATCATTGCCGCATG--ATCA     142
                      |||.||.||||||||.||||.||.||||||||..||  |.|.|  ||||
EMBOSS_001         95 gacacctatccagcagtggtttttctcgcaatctctg--gaaagccatca     142

EMBOSS_001        143 TTTTAATCAGTCCGTTATGCTCAGCAGTGCGGACAGAATCAATGAATCTG     192
                      ||||||.||.||||||||.||..|.|.||.||..||.|..||||..|||
EMBOSS_001        143 ttttaaccaatccgttatgatctaccgggctgaacgatttgatgaggctg     192

EMBOSS_001        193 CTATGAGAAAAGCGCTGGCACAG-CTCGCGGTTCATCATGACGCATTACG     241
                      |..|.|||||||..|| ..|.|| ||.|.|..|||||||||.|||.|..|
EMBOSS_001        193 cgcttagaaaagtact-aaaaagccttgtgactcatcatgatgcactgag     241

EMBOSS_001        242 GATGGTATGC------------GGCACACAAAACGGATCAGTGATTCA-A     278
                      .||.||||||            ||||..||          ||||   || |
EMBOSS_001        242 aatcgtatgccggcatgaagatggcaggca---------agtg---caga     279

EMBOSS_001        279 TATAACAGAGCGGACA-ACCTTTCTGAAGAAGAACTCTTTACGTTTGAAA     327
                      || ||||||| |||.| |.|||||.||.||.||.||.|.|.|..||||||.
EMBOSS_001        280 ta-aacagag-ggatagatctttcagatgaggagctatatgcacttgaac     327

EMBOSS_001        328 CGCATGACGTGAGAGGGAAAGGGT--CATTACAAGA-----GGAACACGC     370
                      .|..|||.||        ||||||.|  |.|.|| |||      ||||||||
EMBOSS_001        328 tgtttgatgt-------aaaggatagcctaac-agaagcccggaacacg-     368
EMBOSS_001        371 CGCAATTGAACAGGCGGCCGCCCGTATTCA--AACTTCCATTCGGCTGGA     418
                          ||||||..|.||.|||..|||.||.||  |||.|  ||||||..||||
EMBOSS_001        369 ----attgaggaagcagccagccggatgcaagaacat--attcgtttgga     412

EMBOSS_001        419 GACGGGTCCGCTCGTA------GCTGTCGGTTTATTTCATGCCGCTGATG     462
                      .||.||.||   |.||       |||   ||.||.|||....|||...|||
EMBOSS_001        413 aaccggccc---cttacttcacgct---ggattgtttagaaccgagaatg     456

EMBOSS_001        463 GCGATCATCTGCTGCTTTCCATTCATCACCTTGTCATTGACGGAGTGTCA     512
                      |.||.|||.|..|.||..|.|||||||||.|.||..||||.|..||.||.
EMBOSS_001        457 gtgaccatttatttctaacgattcatcacttagttgttgatgcggtttct     506

EMBOSS_001        513 TGGCGGATTTTATTTGAAGATCT---CACCGCGTGC-TACAGGCAGGCGC     558
                      |||||.|||||.|||||.|||.|   .||.||.|.| .|||||||||
EMBOSS_001        507 tggcgcattttgtttgaggatttttcaacagcttacaaacaggcag----     552

EMBOSS_001        559 TCGAAGGAAAAGAAGCCGCT------CTTCCGGCGAAAACCGATTCCTAT     602
                      ||..||||.||       .||       ||.|||...|||||.|||||.|||
EMBOSS_001        553 tctcaggagaa------tctattaaactgccgcaaaaaacagattcatat     596

EMBOSS_001        603 CAGACTTAT--GCAAAGCAAATCTCTGATTATGCAAAAAGCCGCCGGCTG     650
                      ...||.|||  .|||||  |||..||||.||..|.|.|||||||.|.||
EMBOSS_001        597 ttaacgtattcacaaag--aatagctgactactctataagccgccaggtg     644

EMBOSS_001        651 TTACAG---GAAGCAGACTACTGGTCCGAAAGAGAAAAAGCGGCGGTCA-     696
                         |||   |||||||..||||||...||..|.||.||      |.||||
EMBOSS_001        645 ---cagcgcgaagcagcgtactgggatgagtgcgagaa-----ccgtcac     686

EMBOSS_001        697 ----AACCGCTTCCGAA-------GATGCCCGCATCTCTTCCAATCTTCT     735
                          |.||..||||.||        ||.| |.||||||....||.||
EMBOSS_001        687 atccagcctattccaaaagacaatgacg-cagcatctaacacattc----     731

EMBOSS_001        736 GAAAGATACAGATGTGA-TGACCGTTACATTGA----------CAAAACA     774
                       ||||||||||.|||| |||   ||   ||||          ||..|||
EMBOSS_001        732 -aaagatacagaagtgattga---tt---ttgagctatctcgccatcaca     774

EMBOSS_001        775 GGAGACCGAGCAG-----CTGTTAACGGAAGCGAACAGAGCATACACAAC     819
                              |||      ||.|||||||..||..|.|.|||.||||..||
EMBOSS_001        775 ----------cagaattactattaacggctgcacataaagcttacagcac     814

EMBOSS_001        820 AGAGACGGGCGAATTAT----TGCTTGCCGCGTTAAGCCTTGCGCTGAAC     865
                      .||||    .||||.||    |.||..|.||.|||.|.||||..||.|.
EMBOSS_001        815 ggaga----tgaatgatatccttctgacagctttaggtcttgctttgcag     860

EMBOSS_001        866 AGATGGACCGGAAACGAGACATTCAAAATCAGCATGGAAGGCCACGTTCC     915
                      |.||||||.||||||..|...|||.|||||||||||||.||||
EMBOSS_001        861 aaatggacgggaaataaccaatttaaaatcagcatggaggcca------     904
```

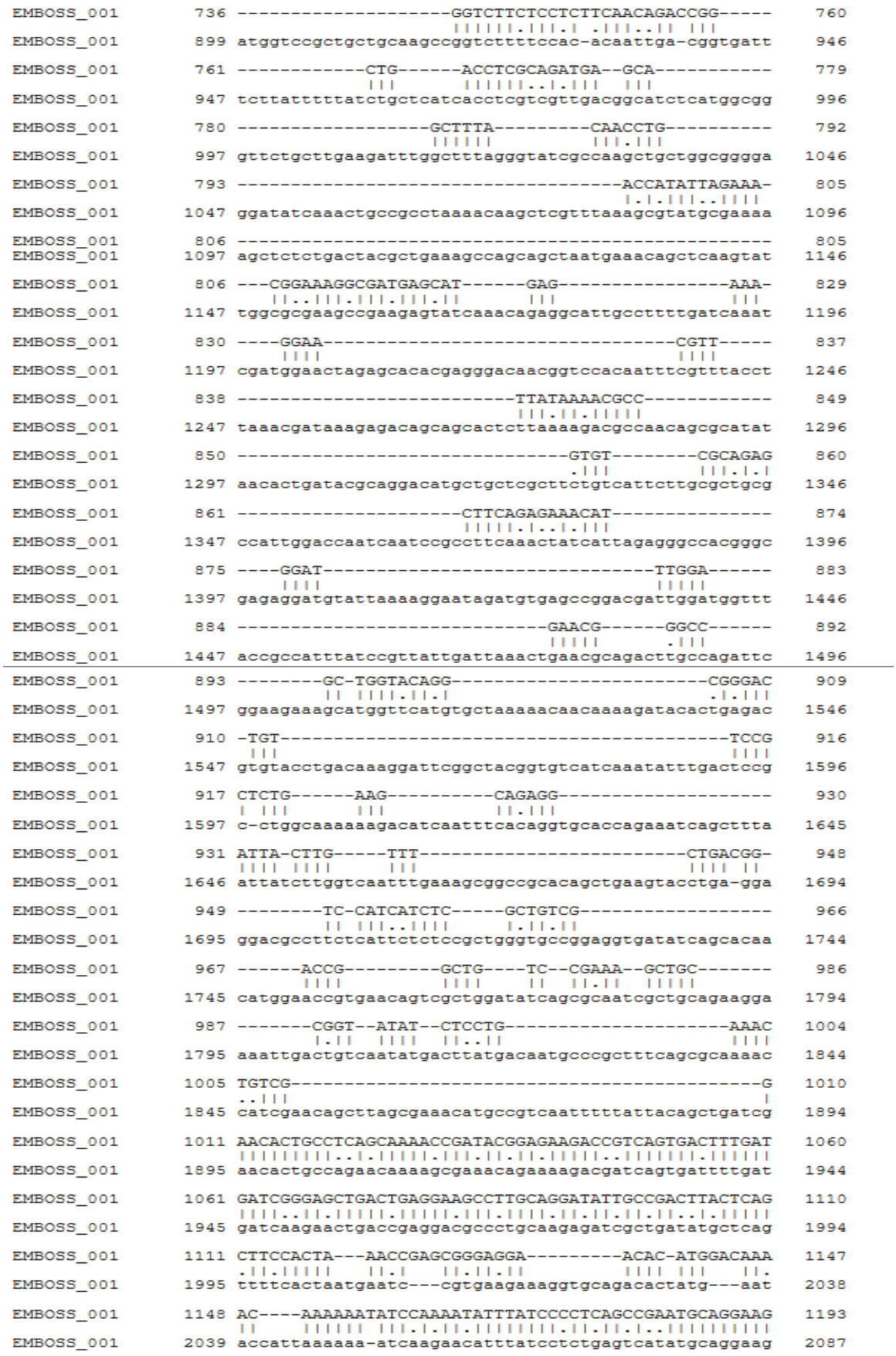
```
EMBOSS_001       916 AGCGGCTTCGGTGTTTCTCCCGTAAAAGCGGGAAATGAAGTCAGCCCAGA    965
                                                 .|||||||||..|||||||..|
EMBOSS_001       905 ----------------------------tgggaaatgagatcagcccgaa    926

EMBOSS_001       966 CTGGGAACGTCCGTACACACTTGATATCAGCGGCTCCGTATCGTCGGAAT   1015
                     ||||||.|||||.||..||||.||||||||||||.||||.||.|||||..|
EMBOSS_001       927 ctgggagcgtccttatgcactggatatcagcggcgccgtttcttcggggt    976

EMBOSS_001      1016 GTCTCAGCATGCATGTCGTGTACAACCGTTTTCAGTATCAAAAAGAAACG   1065
                     |.||.|.||||||..||.|.||.|||.|.|||||.|.|.|..||.||||.
EMBOSS_001       977 gcctgaacatgcacatcatctataacaggtttcaatttgaggaaaaaaca   1026

EMBOSS_001      1066 ATTGAA--GTTTTGACGGGATATTTT---CATACCTTTCTCAAGCAGATC   1110
                     ||..||  .||||  |..|..|||||   ||.||   |||..|..|.|||
EMBOSS_001      1027 atccaaacatttt--ccaggcattttaaacagac---tctggaaaacatc   1071

EMBOSS_001      1111 ATCACGCATTGCACAGGCAAAGAGGAGCGTGAATGGAGCGCCGCTGATTT   1160
                     ||...||||||.||.||||||||..|.|..||.||||||||..|||||||
EMBOSS_001      1072 attgagcattgtaccggcaaagaaaaccaagagtggagcgcatctgattt   1121

EMBOSS_001      1161 CAGTGACGAGGAGTTAACGCTGGAAGATCTGAGTGACATTATGGGAGCCG   1210
                     ||.|||.||.||.||||||||.||.||..||||||||.||.||||||||||
EMBOSS_001      1122 cactgatgaagacttaacgcttgatgaattgagtgagatcatgggagccg   1171

EMBOSS_001      1211 TCAACAAACTATAGAAAAGGAG------TGGCGCAATGGGGCAGCAACCG   1254
                     ||||||||||||||.|.|||||       ||     ||||..|||||||||.
EMBOSS_001      1172 tcaacaaactataggagaggagattcactg-----atgccgcagcaacct   1216

EMBOSS_001      1255 GAAATACAAGATATATATCCTTTGTCTTTTATGCAGGAGGGCATGCTTTT   1304
                     |||||||.||||||||.||||||||.||||||||||||.||.||||||||
EMBOSS_001      1217 gaaatacaggatatataccctttgtcatttatgcaggaagggatgctttt   1266

EMBOSS_001      1305 TCATTCTCTGTTGGATCATGACTCAAGAGCCTACTTCGAACAAGCGGTGT   1354
                     |||.||.||||..|||.|..|.||.|||||.||.||.|||||.||....|
EMBOSS_001      1267 tcactcactgtatgatgaacagtcgagagcttattttgaacaggcttctt   1316

EMBOSS_001      1355 TTACGATCAGCGGAAAGCTTGACCGGGAACGGTTCCAAAAAAGTATAGAC   1404
                     ||||||||...|||.||||||||..|||.||.|||||||||||||.||.
EMBOSS_001      1317 ttacgatccatggacagcttgacttggagcgtttccaaaaaagtatggat   1366

EMBOSS_001      1405 ACTGTTTTTGAACGATACGATATTTTCCGTACGACATTTATTCACAAGAA   1454
                     .||||||||..|.||||||||.||.||.||..||||.||..|.||.||||
EMBOSS_001      1367 gctgtttttgacaggtacgatatctttcgaacagcattcatctataaaaa   1416

EMBOSS_001      1455 TGTGGCAAAGCCCCGACAAGTCGTGTTGAA-AAACAGGCCGAGCCGCGTT   1503
                     |||.||.||.||.||.|||||||||.|.|| .||| |.|...|||...||
EMBOSS_001      1417 tgtagccaaacctcgtcaagtcgtgctcaagcaac-gtcattgccctatt   1465

EMBOSS_001      1504 CAGTTTCATGATATTTCTCATTTGGACGAGAAGGCGCAAGACAAATATGC   1553
                     ||..||.|.||||||||.|||.|..|.||.|.||...||||..|.|.|.|
EMBOSS_001      1466 catattgaagatatttcccatctcaatgaaagggataaagaacattgtac   1515

EMBOSS_001      1554 CGACCGG--TTCAGAAAAG-AAGATAAAGATAAAGGCTTTGATTTGCAAT   1600
                     |||  ||  |||| ||.|| |.||.|||...||||||||||.|.|||.
EMBOSS_001      1516 cga--ggcattca-aagagcaggacaaatccaaaggctttgatctccaaa   1562

EMBOSS_001      1601 CAGACCCTCTCATGAGAGTCTCCATTTTGAAAAAAGCGCCTGAACGCTAT   1650
                     |.||....||.||||||.|.||.|||||.|||...||||||||.|..|||
EMBOSS_001      1563 cggatgtactgatgagaatatctattttaaaatgggcgcctgatcattat   1612

EMBOSS_001      1651 GTTTGTATATGGAGCCATCATCATATTGTCATGGACGGCTGGTGCTTCGG   1700
                     ||.|||||.||||||||.|||||||||.|.|||||.||||||||.||.||
EMBOSS_001      1613 gtctgtatctggagccaccatcatattttaatggatggctggtgtttagg   1662

EMBOSS_001      1701 AATTGTCATGAAGGAATTTCTGGTGATCTATCAATCTCTCGGAGACGGAC   1750
                     .||.|||||.||.||.||||||....||.||||||.|.||.||..|.||.|
EMBOSS_001      1663 catcgtcattaaagactttctccatatttatcaagcgctgggtaaaggcc   1712

EMBOSS_001      1751 GGCTTCC------GTCTCTCGAGCCTGTTCAGCCATATGGGAAGTATATT   1794
                     .||||||      |.||      ||.||.|||||.||.||||.|||||||
EMBOSS_001      1713 agcttcctgatttgcct------cccgtacagccgtacgggacgtatatt   1756

EMBOSS_001      1795 AAATGGCTCATGTAGCA   1811
                     ||||||||.|||.||||
EMBOSS_001      1757 aaatggctgatgcagca   1773
```

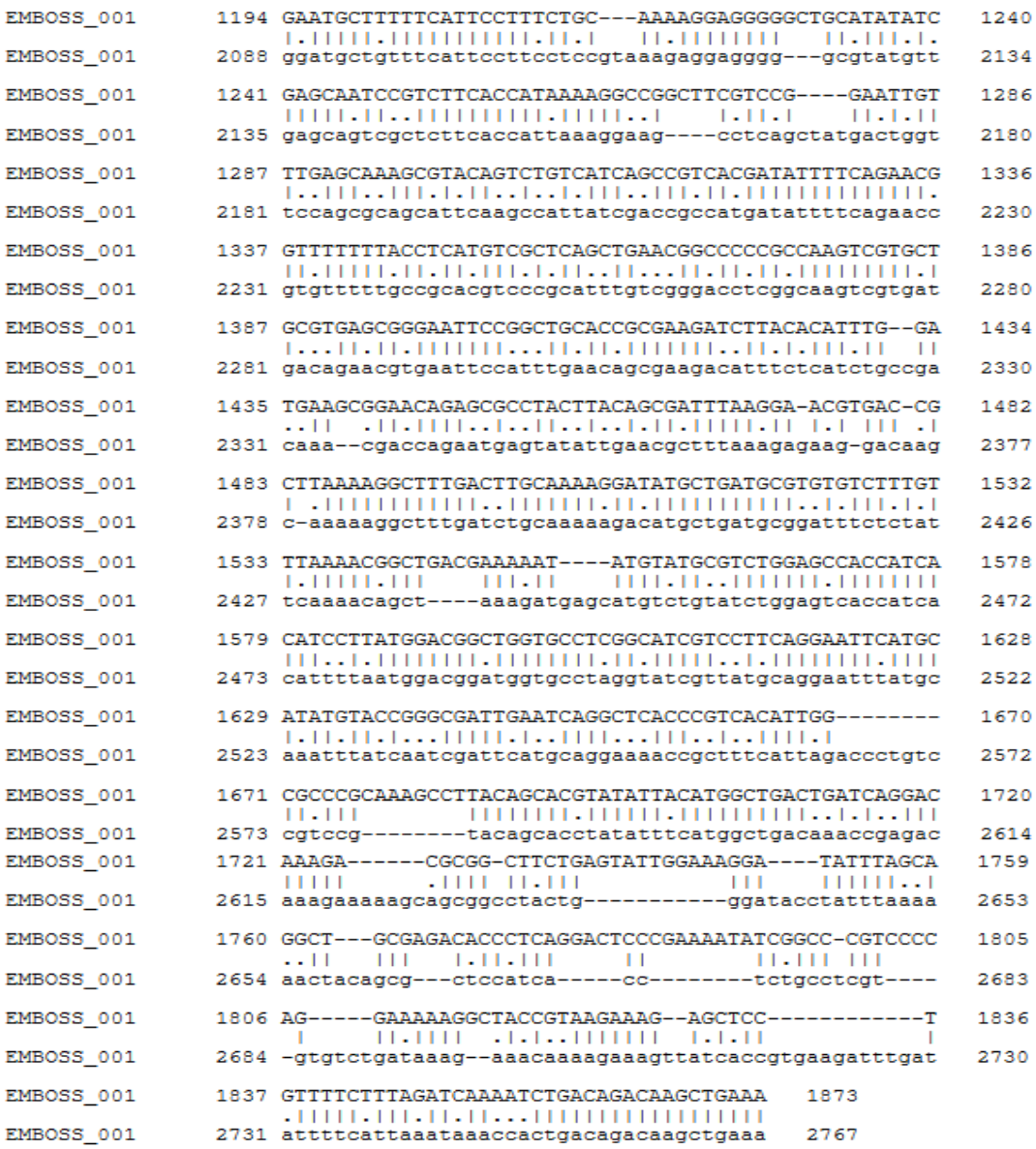
## *ppsC-ppsD* fragment alignment of *B. subtilis* S499 and 168

```
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 969
# Identity:     705/969 (72.8%)
# Similarity:   705/969 (72.8%)
# Gaps:          59/969 ( 6.1%)
# Score: 2428.5
#
#
#=======================================

EMBOSS_001          1 GATACGCAAGATATTCTACTGACTGCCGCAGCTTTGGC--TTTACGGGAC      48
                      |||||.|||||||||||..|.||.||.|||.|.|||||  |||  |.||.
EMBOSS_001          1 gatacacaagatattctgttaaccgcagcatccttggcgattt--gtgaa      48

EMBOSS_001         49 TGGACAGGGGGCGGCAGACTGCGCATCGCGATGGAAGGGCACGGGAGAGA      98
                      ||||||||.||..|||..||.||.|||||.||||||||.||||||||||
EMBOSS_001         49 tggacaggaggaagcaagctccgtatcgctatggaaggacacgggagaga      98

EMBOSS_001         99 GCACATCATGCCGGAGCTCGACATCAGCCGCACCGTAGGCTGGTTTACCT     148
                      .|||||..||||.|||.|.||.||.|||.|.||.||.||||||||||.|
EMBOSS_001         99 acacattctgccagagttagatataagcaggacggtcggctggtttacat     148

EMBOSS_001        149 CCATGTACCCGGTGCTCAT-TGACTTGAACA-CAGCAGGCGCTGAGCTTG     196
                      ||||||||||||..|||||  .|..||||||.|  ||  .|||  |.|||||.|
EMBOSS_001        149 ccatgtacccggccctcataagctttgaaaatca-tagg-gatgagctcg     196

EMBOSS_001        197 GAACGGCTGTAAAAACGGTCAAAGACACACTCGGAAGAATACCGGATAAA     246
                      ||||..|.||||||||.||.||||||.|||||.||..|.||||||..|.|||
EMBOSS_001        197 gaacctcagtaaaaaccgttaaagatacactgggccggataccaaacaaa     246

EMBOSS_001        247 GGAATAGGCTACGGCATTTTAAAATACATGACACCGCCTGAACA-AAAAA     295
                      ||..||||.|||||.||..|.||||||.|.||||..|||||||.|  ||||.
EMBOSS_001        247 ggggtaggttacggtatgctgaaataccttacacaccctgaaaataaaag     296

EMBOSS_001        296 CAATCCG-CTTCAGACAAGCGCCTGAAATCAGTTTTAACTATTTAGGGCA     344
                      ||.|.||  .|||  |.||.|.||||||||||.||||||||||.||.||
EMBOSS_001        297 cattacgttttc--aaaaacacctgaaatcagctttaactatttgggcca     344

EMBOSS_001        345 ATTCAATGATACAGAAGAG-CAGCACACATTCAGTTTATCCGGTCTTGCG     393
                      .|||||.||.|..||| || |||.||||.|||.|...|||..|.||||..
EMBOSS_001        345 gttcaacgacatcgaa-agacaggacaccttccgcccatcaagccttggt     393
EMBOSS_001        394 AGCGGGCACGACATCACGCCGACATGGCAGCGGGAACAAGCGGTGGAAAT     443
                      ||||||..|.||.|||||.|..|||||||.||||.||.|||....|.||.||
EMBOSS_001        394 agcggaaaggatatcactcatacatggaagcgcgagcaaatcatcgagat     443

EMBOSS_001        444 GAGCGCGATGGCCGCTCA-ACATAAA-CTGCACTTCAGTTTGAGCTATCC     491
                      |||.||.||||||.|  || |||.||| ||.||.||||.|.|.||.|||||
EMBOSS_001        444 gagtgcaatggcag--cagacaaaaagctccatttcaatcttagttatcc     491

EMBOSS_001        492 GCCGAG-CCGCTTCCGCAAAGAGACGATGGAACAGCTGCTGCAAACCTCA     540
                      ||| || ||||||.|.||.|.|.||.||||||||||..|        |.|
EMBOSS_001        492 gcc-agcccgctttcacagaaacacaatggaacagcttat-------taa     533

EMBOSS_001        541 CAG-------CAATATTTGCGTGATATTATGGCGCACTGCACAGGCAAAG     583
                      |||       ||.|.||||||..||.||||||...||||||.|.||..||.
EMBOSS_001        534 caggattgaacacttttgctggacattatgaaacactgcgccggtcaac     583

EMBOSS_001        584 AGGAAGCAGAAAAAACCGTTAGTGATTTTAGCAGCAAAACGTTAACATCT     633
                      ||.||||||||||.||..|.||.||||||.||||||..||.|.||||||.|.
EMBOSS_001        584 agaaagcagaaaagacactgagcgatttcagcagtcaatcattaacggca     633

EMBOSS_001        634 GATGATTTGGACAGCATAGCGAGTTTTGTCGAGGAACTGTAGAAAAAGGC     683
                      ||.||.|||||||||||||.|.||.||.||.||.|||.|||||||||..|.
EMBOSS_001        634 gaggacttggacagcatatccagcttggtggaagaattgtagaaaatcgt     683

EMBOSS_001        684 GAC-GGGGAGATTATAACATGACAAAAA---------AGAATGCAATTCA     723
                      |||  ||.||.||  .||||||||||||         .||||.|||||||
EMBOSS_001        684 gacaggagacat--gaacatgacaaaaatgacaaaagcgaattcaattca     731

EMBOSS_001        724 GGATATTTATCCGCTGTCTTATATGCAGGAAGGAATGCTTTTTCATTCCC     773
                      ||||||||||||||||||.||||||||||.|||||||.|||||||.|||||||
EMBOSS_001        732 ggatatttatccgctgtcttacatgcaggaagggatgcttttccattccc     781

EMBOSS_001        774 TCCTGCAAAAGGAATCACAGGCTTACGCCGAGCAGGCCTCCTTTTCAATA     823
                      ||||||||||.||.||||||||.||.|..||||||.||.||.|||.||||||
EMBOSS_001        782 tcctgcaaaaagattcacaggcgtatgttgagcaagcttcttttacaata     831

EMBOSS_001        824 ACGG--GGAAAGGTTGACACCCGCG----TGTTTGAAGAGAGCGTTCATG     867
                          |  |||||.||   ||||| ||    |.|||.||.|.|||.|..|.|
EMBOSS_001        832 ---gaaggaaaagt---caccc-cgcagtttttcaaaacagcatcaacg     874

EMBOSS_001        868 CCCTTTTCGAAAGACATGATATTTTCAGAACGATTTTTATCAGCCAGAAC     917
                      |.|||.|.||||||||||||||||||||||||||||||||||||||||.||.
EMBOSS_001        875 ctcttgtagaaagacatgatattttcagaacgatttttatcagccaaaat     924

EMBOSS_001        918 GTGTCAGTGCCGCAGCAGT    936
                      ||.||....||.||||||.
EMBOSS_001        925 gtttcctccccccagcagg    943
```

## *ppsD-ppsE* fragment alignment of *B. subtilis* S499 and 168

```
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 2887
# Identity:    1350/2887 (46.8%)
# Similarity:  1350/2887 (46.8%)
# Gaps:        1134/2887 (39.3%)
# Score: 3469.0
#
#
#=======================================

EMBOSS_001          1 GAAATGAAAGAAATTGAATCGGTCTTAACAG--GCATAAAAGGTGTCAAA      48
                      ||||.||||||||||||||||||  |.|.|  |.|||||.|||||.|||
EMBOSS_001          1 gaaacgaaagaaattgaatcggtc--atccggtgtataaagggtgtaaaa      48

EMBOSS_001         49 GAAGCAGCCGTCACCGCTCATACCGCATCAG---CAGGCCAAACGGAACT      95
                      ||.||.||.||....||||||   |.|.|||   ||||.|||||.|||.|
EMBOSS_001         49 gacgccgcagtagttgctcat---gtaacagcgtcaggacaaaccgaatt      95

EMBOSS_001         96 ATGCGCTTATATCGTGACAGAAGAAG----GAACAGAAAGTAAAACCGTG     141
                      |.|||||||.|.||||||.||..||    ||.||.|||          ||
EMBOSS_001         96 aagcgcttatgtggtgacaaaaccagggttgagcacaaa---------tg     136

EMBOSS_001        142 CAGCAGGCGCTCAGAAATGAAA-------TGCCCGCTTATATGGTGCCGG     184
                      |.|.|  ||||||||||.|..||       ||||.|..|.||||...||.|
EMBOSS_001        137 ctgta--cgctcagaattacaaaacaagctgccggtgtttatgcaccctg     184

EMBOSS_001        185 CGTTCTTTGAAACACTTGAAGCCCTGCC--TGTCACACCGAACGGGAAGC     232
                      ||||..||||||.|||.||..|.||.||  |||  ||||.||||||.||||
EMBOSS_001        185 cgtttattgaaaaactggactcacttccattgt--caccaaacggcaagc     232

EMBOSS_001        233 TTGACCGCCGCGCGCTTCCTGAGCCGCGGAAAAAAGC-----GCACACC-     276
                      |.||.|||.|.||.|||||             ||||.|     .|| |||
EMBOSS_001        233 tggatcgcggagcacttcc------------aaaacctgtataca-acca     269

EMBOSS_001        277 -------GGACGGGCATT--TACGG--AGCCTGAAAGCGACATGGAA-AA     314
                          |.|||..||||  |.|.|  |.||    |||.|.||||||  ||
EMBOSS_001        270 cgaaggtgaacgtccattcctcccgccatcc----agcaaaatggaacaa     315

EMBOSS_001        315 AGAGCTTTCCGCGATTTGGAGCGAAGTGCTCGGAACTGAAAACATTG---     361
                      |.| ||..|.|..||.||||..|||||.||.||..|.|||||.||.|
EMBOSS_001        316 ata-ctggcagatatatggaaagaagtactgggggcagaaaagatcggaa     364
EMBOSS_001        362 CCGCGGATCAATCATTTTTTGAACTGGGCGGCGATTCCATCAAAGCCCTG     411
                      |.||.|||   ||.||.|||||||.|||||.|||||.|||||||..|.
EMBOSS_001        365 cagctgat---tcgttctttgaacttggcggagattcaatcaaagcgtta     411

EMBOSS_001        412 CAAGTATCAGCCCGATTACATCAGGCAGGAAAACAGATCGCCGTAAAAGA     461
                      ||.|||||.||.||..|.||.|.|..|||.||.||.||.||.||.|||||
EMBOSS_001        412 caggtatctgcgcggcttcaccggataggcaagcaaatggcagtgaaaga     461

EMBOSS_001        462 TATTTTCAGCCGGCCGACAATTAGAGAGCTGGCTCCTTATGTCCG---TA     508
                      |.|.||.||||..|||||.|||..|||..||||..|||||.||||   |.
EMBOSS_001        462 tttgtttagccatccgaccattcaagaattggcagcttatatccgtgatt     511

EMBOSS_001        509 CAGAGCGGCAGCCGGT--------CAGCCAAGCTCCTGCAGAAGGAGAAG     550
                      ||||           |        ||||||||||.|.|..||.|||||.|
EMBOSS_001        512 caga-----------tacaagttccagccaagctgcggttgagggagacg     550

EMBOSS_001        551 TGAAGTGGTCGCCGGTTCACAAATGG-TTCTTCACACAAGACATGAAGGA     599
                      |..|.||||||||.||||||.|||||  |||||  .|.|||||.||.||.||
EMBOSS_001        551 tacaatggtcgcctgttcaaaaatggtttctt-tctcaagatataaaaga     599

EMBOSS_001        600 AGCAAATCATTTTAACCAGTCCGTCATGCTGACGAGAACAAATTCAATAG     649
                      |..|.|.||||||||.||||||||.|||||||....|||..|..||..||.
EMBOSS_001        600 aaaacaccattttaatcagtccgtaatgctgcatcgaagcacctctgtac     649

EMBOSS_001        650 ATGAAGAGGCTCTCAGGAAAACGTTAAAAGCCATTACCGT-TCATCATGA     698
                      |.||||||.||.||..|.|||||||||||||  ||.||.|||  |||||||||
EMBOSS_001        650 aagaagacgcactgcgaaaaacgttaaaag-caataacgtgtcatcatga     698

EMBOSS_001        699 TGCCTTGC-----------------------------------------     706
                      |||..|||
EMBOSS_001        699 tgcgctgcgaatggtgtttacacagaatgaacaaggaaagtgggatcaat     748

EMBOSS_001        707 -------GCC-----------------------------------------     709
                             |||
EMBOSS_001        749 ataatcggccgctcagtcattcggacgacgcgttgtacggccttcagatg     798

EMBOSS_001        710 ------TTGTCTGC-----------------AAAAAAGAC-----GAA--     729
                            |||||.||                .|||.||||     |||
EMBOSS_001        799 attgatttgtcagctccggacggcactgatggaaatagaccgtatgaacc     848

EMBOSS_001        730 --------------------------------GAGAAA-------------     735
                                                     .||||||
EMBOSS_001        849 tcttatcaagcgccatgtacttgatattcagcagaaaatggatttgaaga     898
```





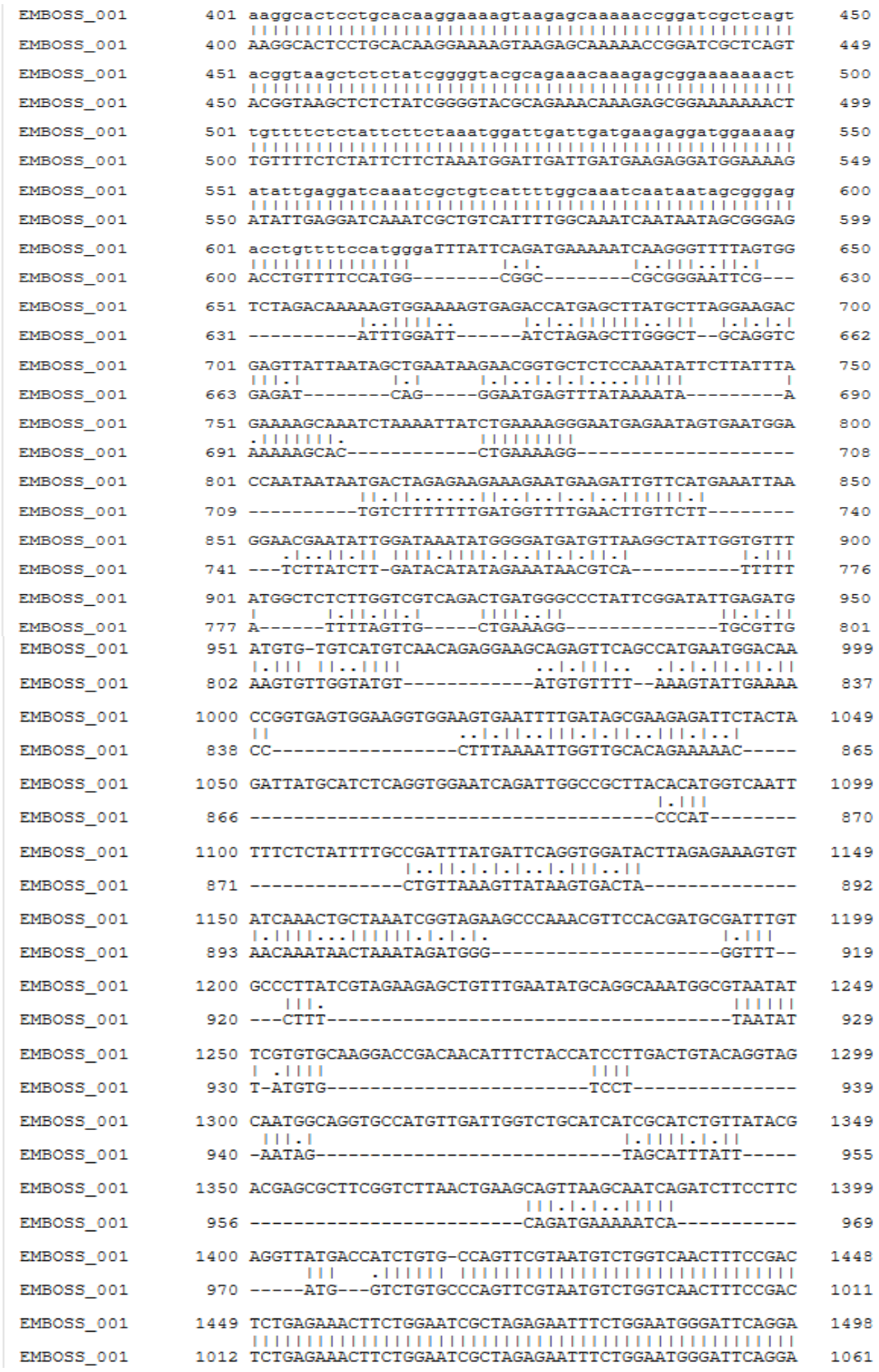
## Fengycin fragment BLAST alignment of *B. subtilis* S499

```
>gb|GQ489109.1|  Bacillus amyloliquefaciens strain 96-82 fengycin synthetase E
(fenE) gene, partial cds
Length=1214

 Score =  785 bits (425),  Expect = 0.0
 Identities = 427/428 (99%), Gaps = 0/428 (0%)
 Strand=Plus/Plus

Query  13    CGTATTGACCGTCAATTCAAGATTCGCGGGAAGCGTATCGAGCCGGCCGAAATCGAAGCG  72
             ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  619   CGTATCGACCGTCAATTCAAGATTCGCGGGAAGCGTATCGAGCCGGCCGAAATCGAAGCG  678

Query  73    CGGCTGACTGAAATTGATGGAATCAGGGAAGCGGCAGTCATCGTGTCAGACGAAGGACAA  132
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  679   CGGCTGACTGAAATTGATGGAATCAGGGAAGCGGCAGTCATCGTGTCAGACGAAGGACAA  738

Query  133   GAAGCAACCCTTTGTGCGTATTATACAGGCGCACATGCTGAAGAAAGAGCCATTCGATCC  192
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  739   GAAGCAACCCTTTGTGCGTATTATACAGGCGCACATGCTGAAGAAAGAGCCATTCGATCC  798

Query  193   CTATTGGCCAGGTCGCTTCCCGACTATATGATTCCGCAGTATGTGATGAAGCTTGACCGT  252
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  799   CTATTGGCCAGGTCGCTTCCCGACTATATGATTCCGCAGTATGTGATGAAGCTTGACCGT  858

Query  253   ATGCCGCTTACCGCAAACGGCAAAGTGGACCGCCGCGGGCTGCCTGCGCCGGAACGGAAA  312
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  859   ATGCCGCTTACCGCAAACGGCAAAGTGGACCGCCGCGGGCTGCCTGCGCCGGAACGGAAA  918

Query  313   AAAGCTTCATCAAAAGCCGTTCCGCCACGAAACTGGGTAGAGCGTGAGCTGATCGACATT  372
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  919   AAAGCTTCATCAAAAGCCGTTCCGCCACGAAACTGGGTAGAGCGTGAGCTGATCGACATT  978

Query  373   TGGACGCCGATCCTCGGCTGCGGGGAGCTGGGCATACAGGATGACTTCTTTGAGCTGGGC  432
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  979   TGGACGCCGATCCTCGGCTGCGGGGAGCTGGGCATACAGGATGACTTCTTTGAGCTGGGC  1038

Query  433   GGCCATTC  440
             ||||||||
Sbjct  1039  GGCCATTC  1046
```

## Index 6: Alignment of pBG180 constructed plasmid sequence

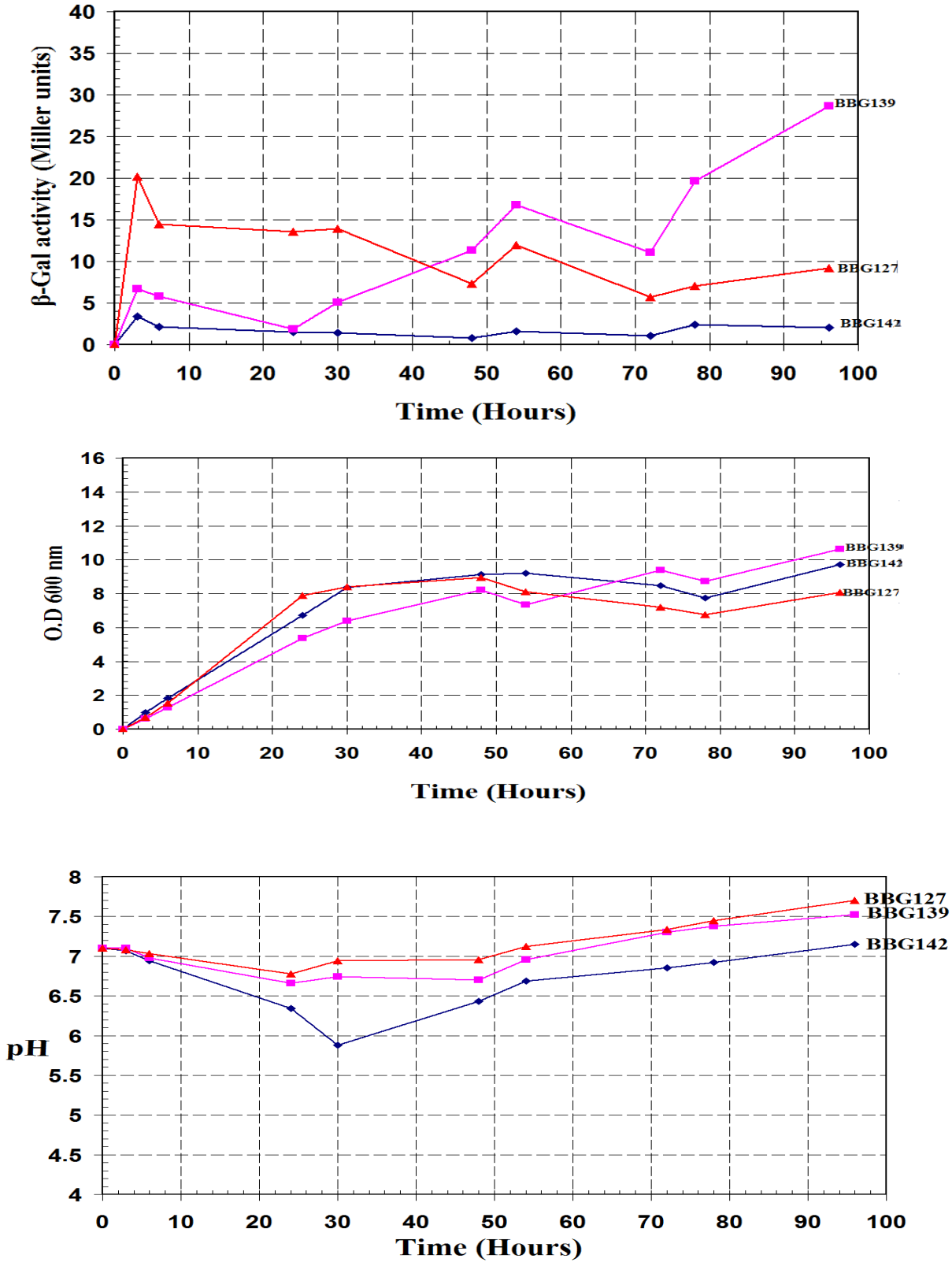
## pBG180 plasmid Needle sequencing result

```
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 2299
# Identity:    1702/2299 (74.0%)
# Similarity:  1702/2299 (74.0%)
# Gaps:         473/2299 (20.6%)
# Score: 8956.5
#
#
#=======================================

EMBOSS_001         1 Gacgtcaagacgggtgaagcacctagttcttcagatgtgctgctctctca     50
                     ||||||  |||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001         1 GACGTC-AGACGGGTGAAGCACCTAGTTCTTCAGATGTGCTGCTCTCTCA     49

EMBOSS_001        51 tcgctcaatgcctttatcgaaacttttgtgccgttcatgaaattaagta    100
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001        50 TCGCTCAATGCCTTTATCGAAACTTTTGTGCCGTTCATGAAATTAAGTA     99

EMBOSS_001       101 ataacggacatgccgaagtgctcgtaaaagagatggggaaagtgaaaaaa    150
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       100 ATAACGGACATGCCGAAGTGCTCGTAAAAGAGATGGGGAAAGTGAAAAAA    149

EMBOSS_001       151 ggagaagggagctgggaaaaggggctcgaagtgttaaacagtacgcttcc    200
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       150 GGAGAAGGGAGCTGGGAAAAGGGGCTCGAAGTGTTAAACAGTACGCTTCC    199

EMBOSS_001       201 ggaatttggtgttgattctaaatcactcgtcttaagggacggatcaggaa    250
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       200 GGAATTTGGTGTTGATTCTAAATCACTCGTCTTAAGGGACGGATCAGGAA    249

EMBOSS_001       251 tctcacatattgatgctgtatcctcagatcaactttcacagctgttatat    300
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       250 TCTCACATATTGATGCTGTATCCTCAGATCAACTTTCACAGCTGTTATAT    299

EMBOSS_001       301 gacattcaggatcagagttggttctcggcttatctaaattctttacctgt    350
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       300 GACATTCAGGATCAGAGTTGGTTCTCGGCTTATCTAAATTCTTTACCTGT    349

EMBOSS_001       351 tgcggggaatcctgacagaatggtgggcggaacgctgagaaaccgcatga    400
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       350 TGCGGGGAATCCTGACAGAATGGTGGGCGGAACGCTGAGAAACCGCATGA    399
```

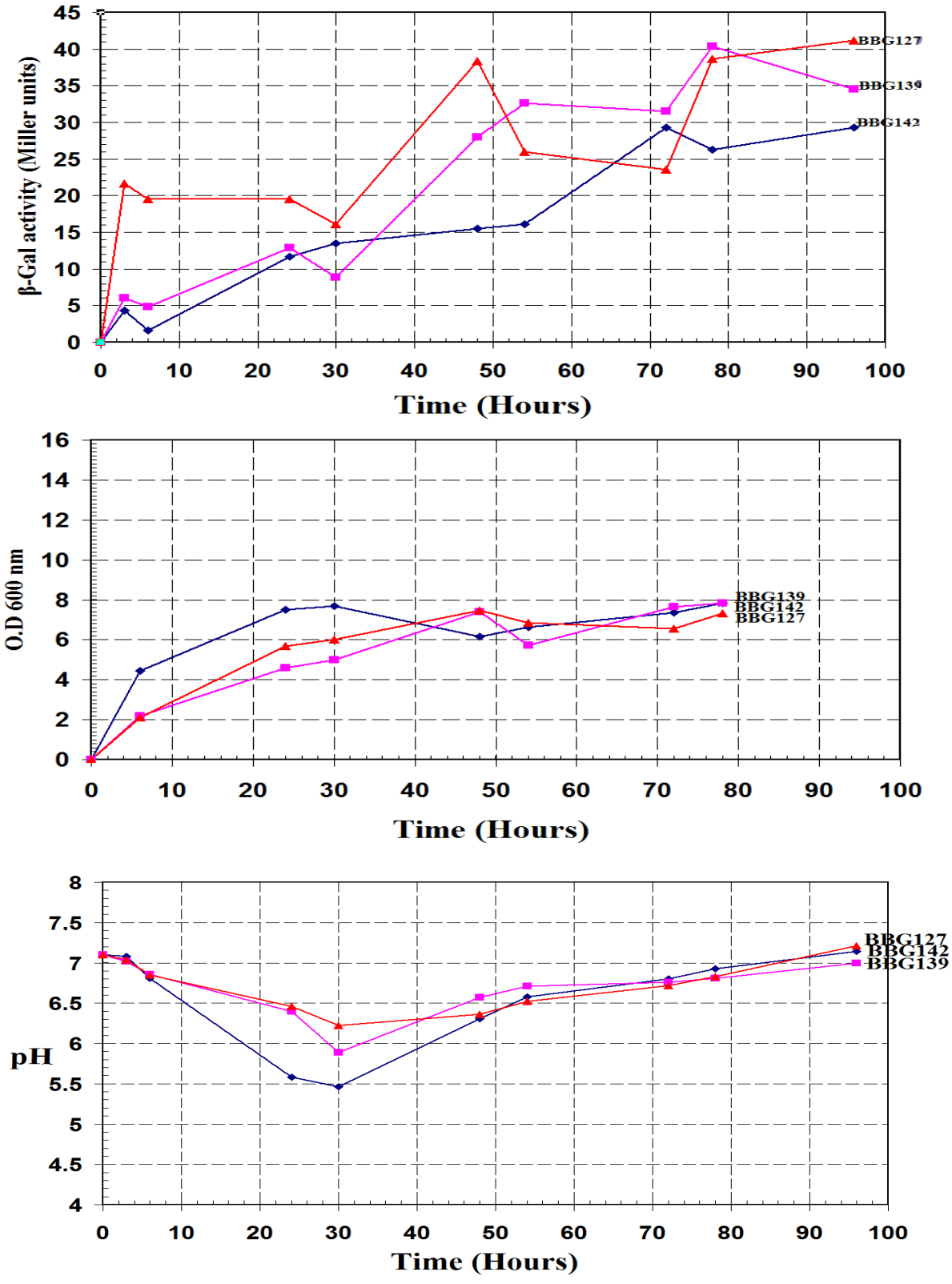


```
EMBOSS_001      1499 GTGGACAGAACGACACGGATATATAGTGGATGTGTCAAAACGCATACCAT   1548
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      1062 GTGGACAGAACGACACGGATATATAGTGGATGTGTCAAAACGCATACCAT   1111

EMBOSS_001      1549 TTTGAACGATGACCTCTAATAATTGTTAATCATGTTGGTTACGTATTTAT   1598
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      1112 TTTGAACGATGACCTCTAATAATTGTTAATCATGTTGGTTACGTATTTAT   1161

EMBOSS_001      1599 TAACTTCTCCTAGTATTAGTAATTATCAGAATTGATCTGCGGCCGCGAAT   1648
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      1162 TAACTTCTCCTAGTATTAGTAATTATCAGAATTGATCTGCGGCCGCGAAT   1211

EMBOSS_001      1649 TCAAGCTCTAGAGGATCCCGGGTACCGAGCTCGAATTCtctagacatata   1698
                     |||||                                 ||||||||||||||
EMBOSS_001      1212 TCAAG-------------------------------CTCTAGACATATA   1229

EMBOSS_001      1699 atttcttttattgtcatttttctgctttttctacattttcttatcctct   1748
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      1230 ATTTCTTTTATTGTCATTTTTCTGCTTTTTTCTACATTTTCTTATCCTCT   1279

EMBOSS_001      1749 tattatgagaactggagggaatccgttgagcgaacatacttattctttaa   1798
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      1280 TATTATGAGAACTGGAGGGAATCCGTTGAGCGAACATACTTATTCTTTAA   1329

EMBOSS_001      1799 cccatgcccaaaggagagtgtggtttaccgaactgctggagccagatacc   1848
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      1330 CCCATGCCCAAAGGAGAGTGTGGTTTACCGAACTGCTGGAGCCAGATACC   1379

EMBOSS_001      1849 agcatttgcaatcttacagcttgtgtgaaatttaaaggcaatattgagct   1898
                     ||||||||||||||||.||||||||||||||||||||||||||||||||||
EMBOSS_001      1380 AGCATTTGCAATCTTATAGCTTGTGTGAAATTTAAAGGCAATATTGAGCT   1429

EMBOSS_001      1899 tgatactctggagggagcactgaatcattccatctcccgcaacgacgcca   1948
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      1430 TGATACTCTGGAGGGAGCACTGAATCATTCCATCTCCCGCAACGACGCCA   1479

EMBOSS_001      1949 ttagatttcagctgttagaaggagaagagcttgagcctcgattgcacctt   1998
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      1480 TTAGATTTCAGCTGTTAGAAGGAGAAGAGCTTGAGCCTCGATTGCACCTT   1529

EMBOSS_001      1999 actgaatataaatattatccacttagaataattgatttttcaaatgttga   2048
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      1530 ACTGAATATAAATATTATCCACTTAGAATAATTGATTTTTCAAATGTTGA   1579

EMBOSS_001      2049 aatgatagaaatagagcaatggattcaggatcaagcgagcattccattta   2098
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      1580 AATGATAGAAATAGAGCAATGGATTCAGGATCAAGCGAGCATTCCATTTA   1629

EMBOSS_001      2099 aactgttcaactcgcctctttttcaattttacctccttagaattgactct   2148
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      1630 AACTGTTCAACTCGCCTCTTTTTCAATTTTACCTCCTTAGAATTGACTCT   1679

EMBOSS_001      2149 catgaagtttggctatttgcgaaattccatcatatcataatggatggaat   2198
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      1680 CATGAAGTTTGGCTATTTGCGAAATTCCATCATATCATAATGGATGGAAT   1729

EMBOSS_001      2199 ctctttaaatgtgatgggaaatcaaattatagatctttatcaaaaaatga   2248
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      1730 CTCTTTAAATGTGATGGGAAATCAAATTATAGATCTTTATCAAAAAATGA   1779

EMBOSS_001      2249 aaaagaaagatcctttacctgatcagccggaaccttcttacttgagctc   2297
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      1780 AAAAGAAAGATCCTTTACCTGATCAGCCGGAACCTTCTTACTTGAGCTC   1828
```

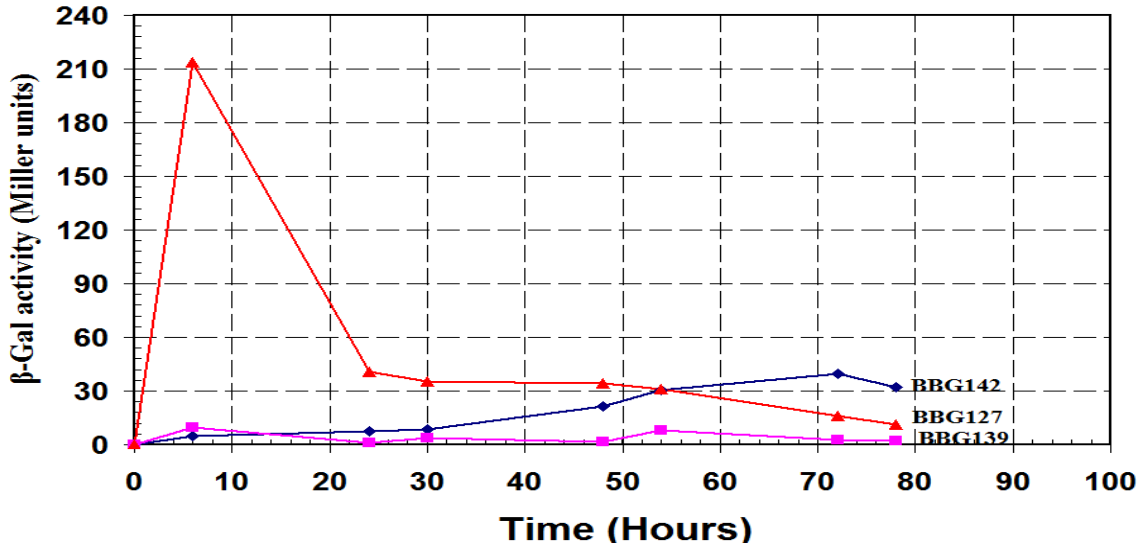
**Index 7: β-Gal activity expressed in Miller units, microbial growth at ($O.D_{600\ nm}$) and pH for *B. subtilis* BBG142, BBG139 and BBG127 under different growth conditions**

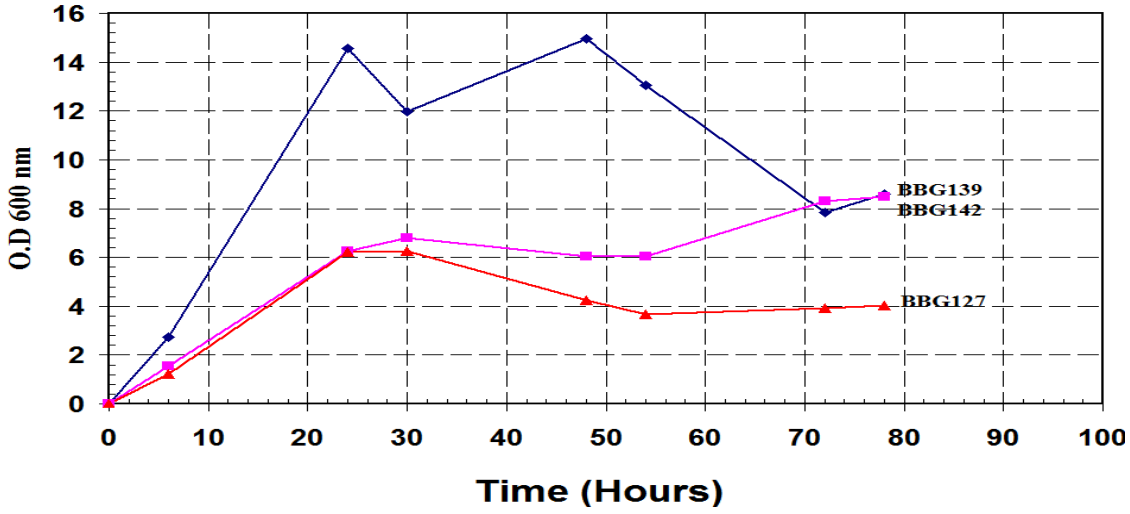


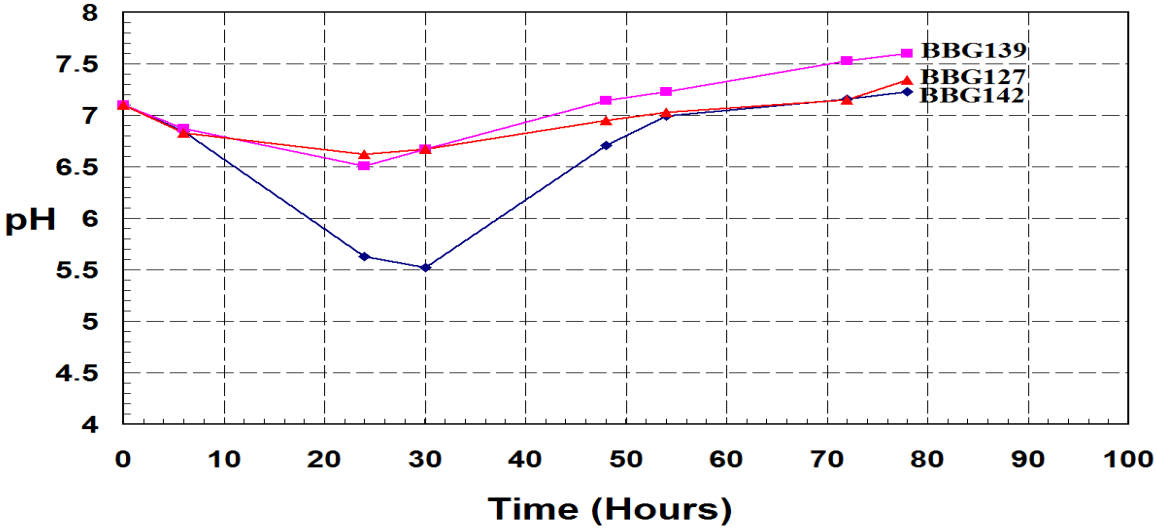

Experiment 1: β-Gal activity expressed in Miller units microbial growth at ($O.D_{600\ nm}$) and pH for *B. subtilis* BBG142, BBG139 and BBG127 under growth in Landy MOPS pH 7.1, 20% f.v, 1 L flask, 200 rpm and 25°C



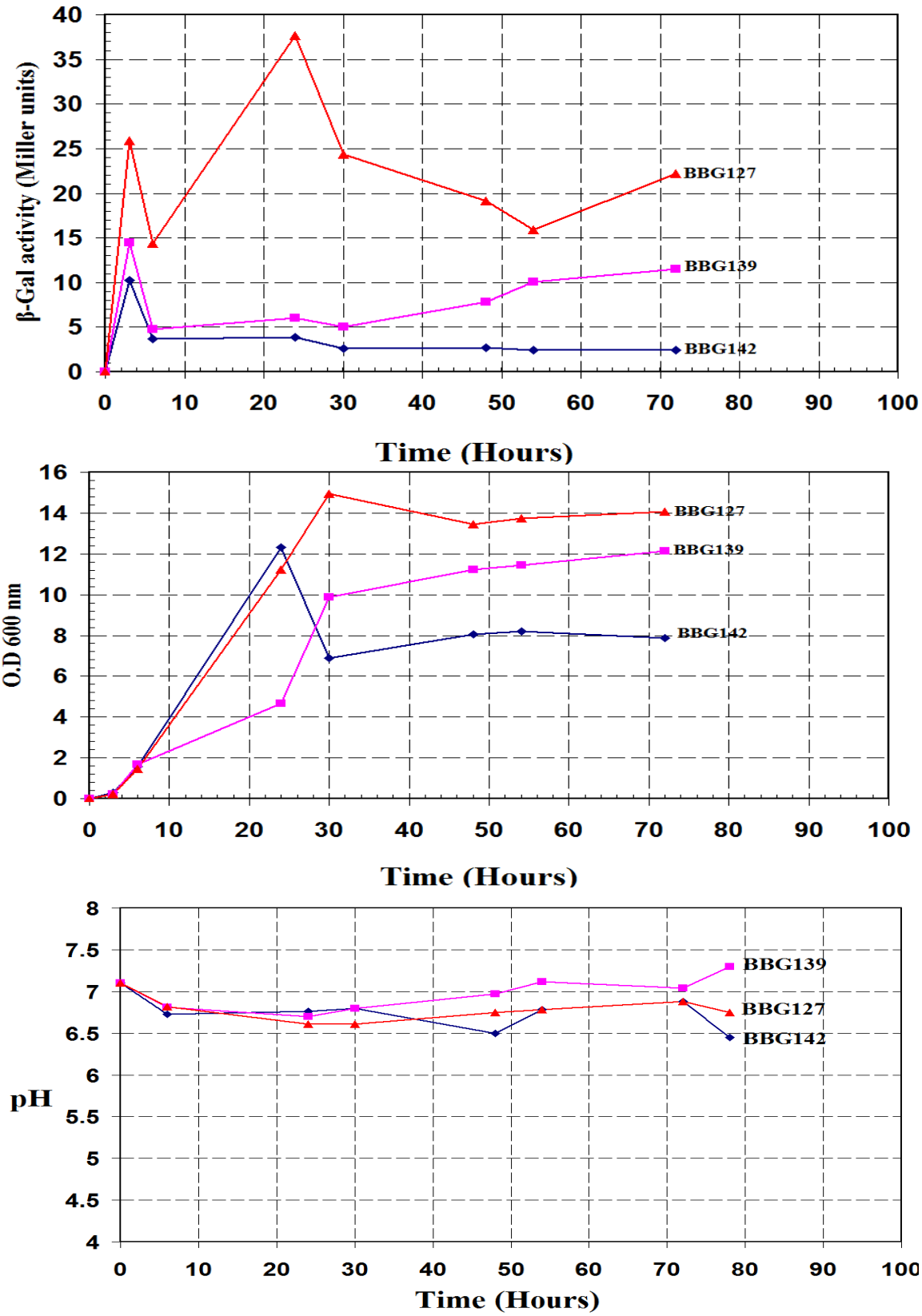

Experiment 2: β-Gal activity expressed in Miller units, microbial growth at ($O.D_{600\ nm}$) and pH for *B. subtilis* BBG142, BBG139 and BBG127 under growth in Landy modified 2 pH 7.0, 20% f.v, 1 L flask, 200 rpm and 25°C



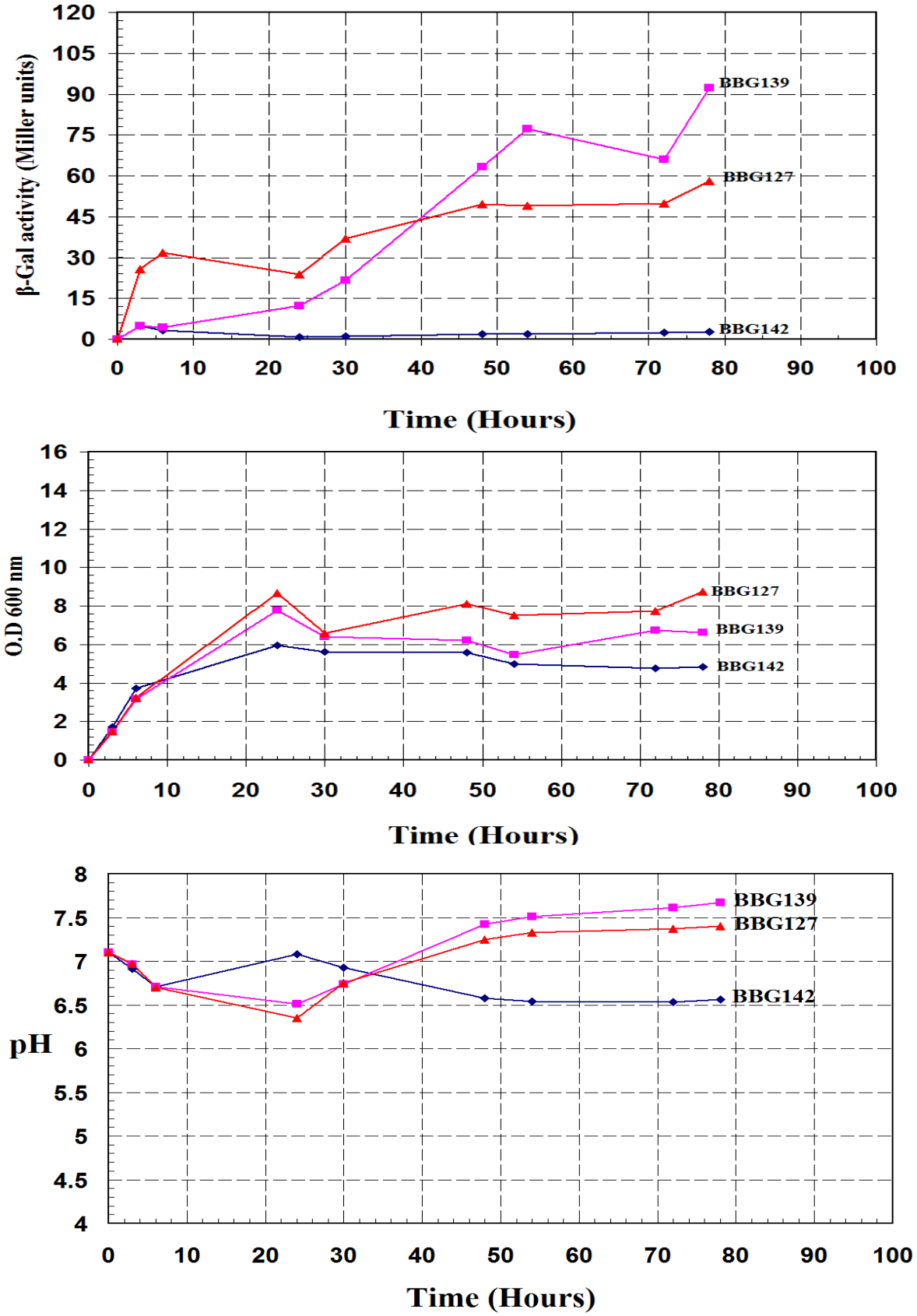



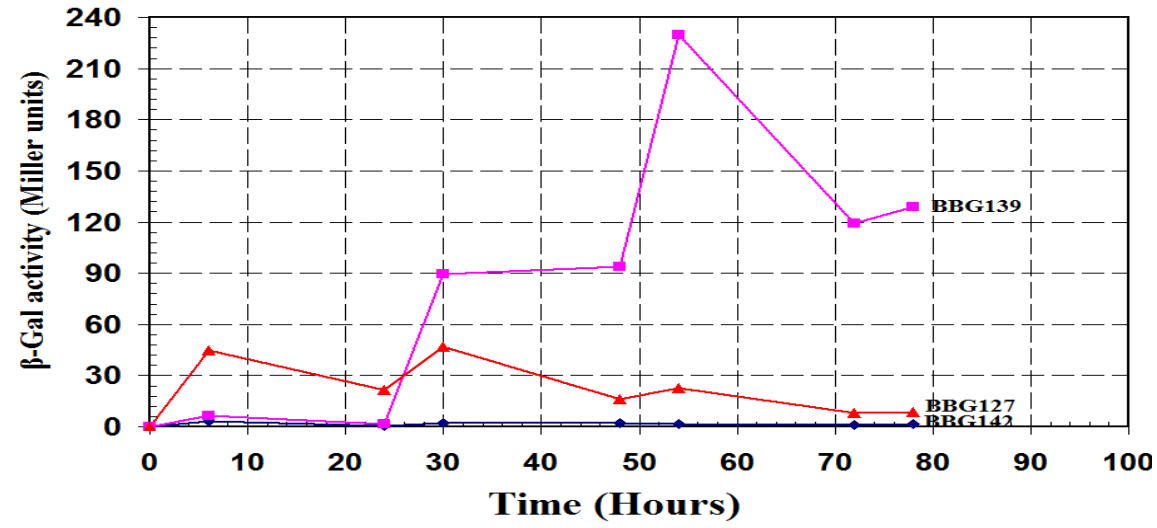

Experiment 3: β-Gal activity expressed in Miller units, microbial growth at ($O.D_{600\ nm}$) and pH for *B. subtilis* BBG142, BBG139 and BBG127 under growth in Landy MOPS pH 7.0, 20% f.v, 1 L flask, 200 rpm and 30°

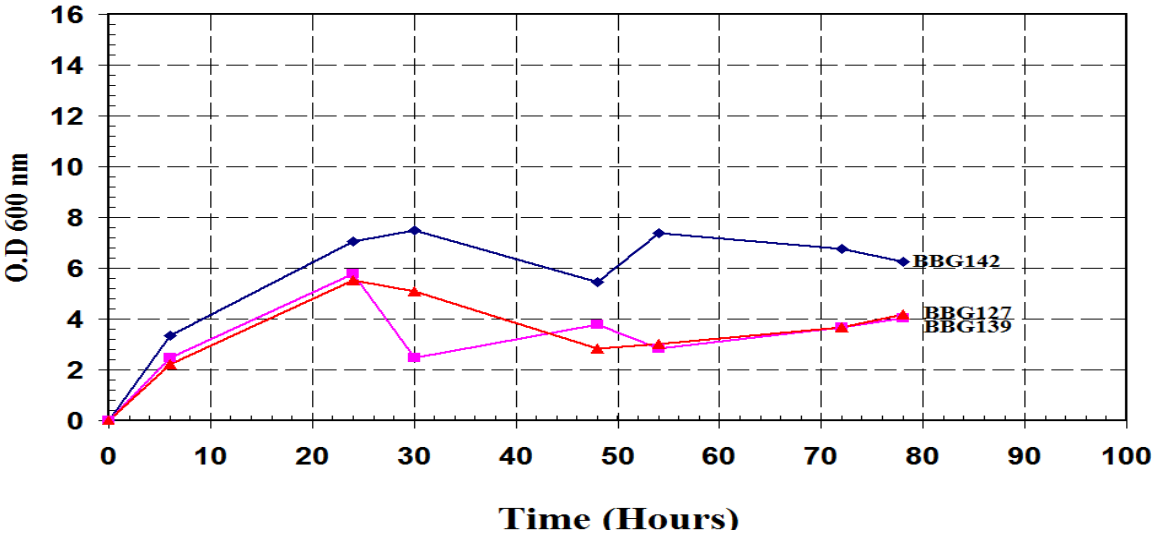


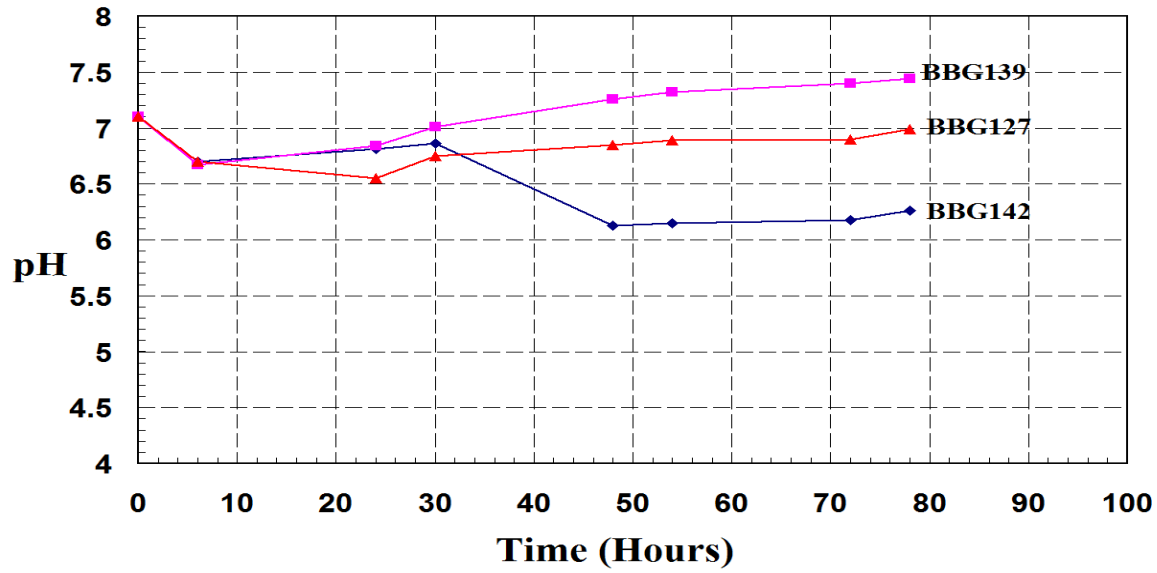

Experiment 4: β-Gal activity expressed in Miller units, microbial growth at ($O.D_{600\ nm}$) and pH for *B. subtilis* BBG142, BBG139 and BBG127 under growth in Landy MOPS pH 7.0, 20% f.v, 100 mL flask, 200 rpm and 30°C



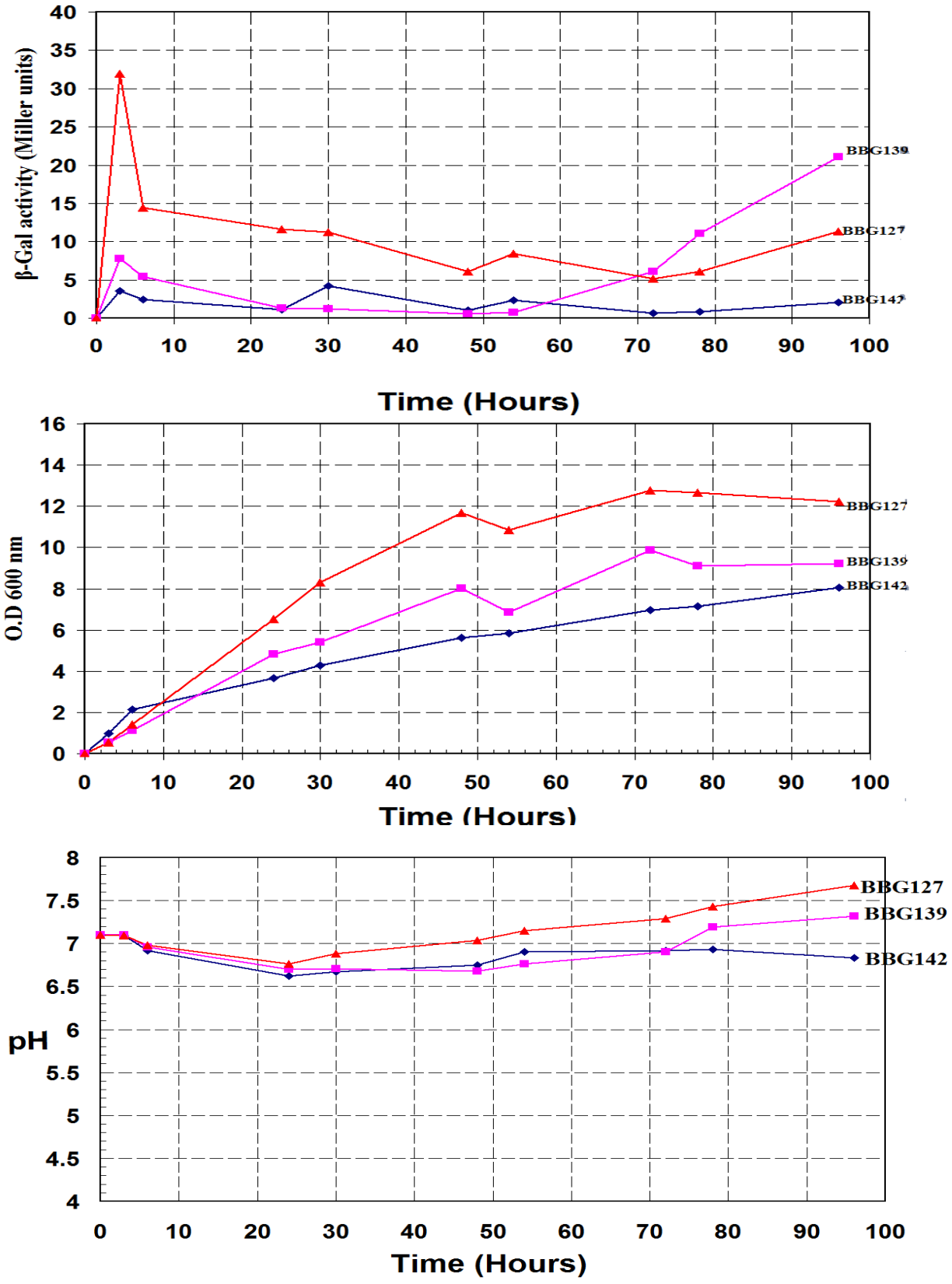

Experiment 5: β-Gal activity expressed in Miller units, microbial growth at ($O.D_{600\ nm}$) and pH for *B. subtilis* BBG142, BBG139 and BBG127 under growth in Landy MOPS pH 7.0, 20% f.v, 1 L flask, 200 rpm and 37°C



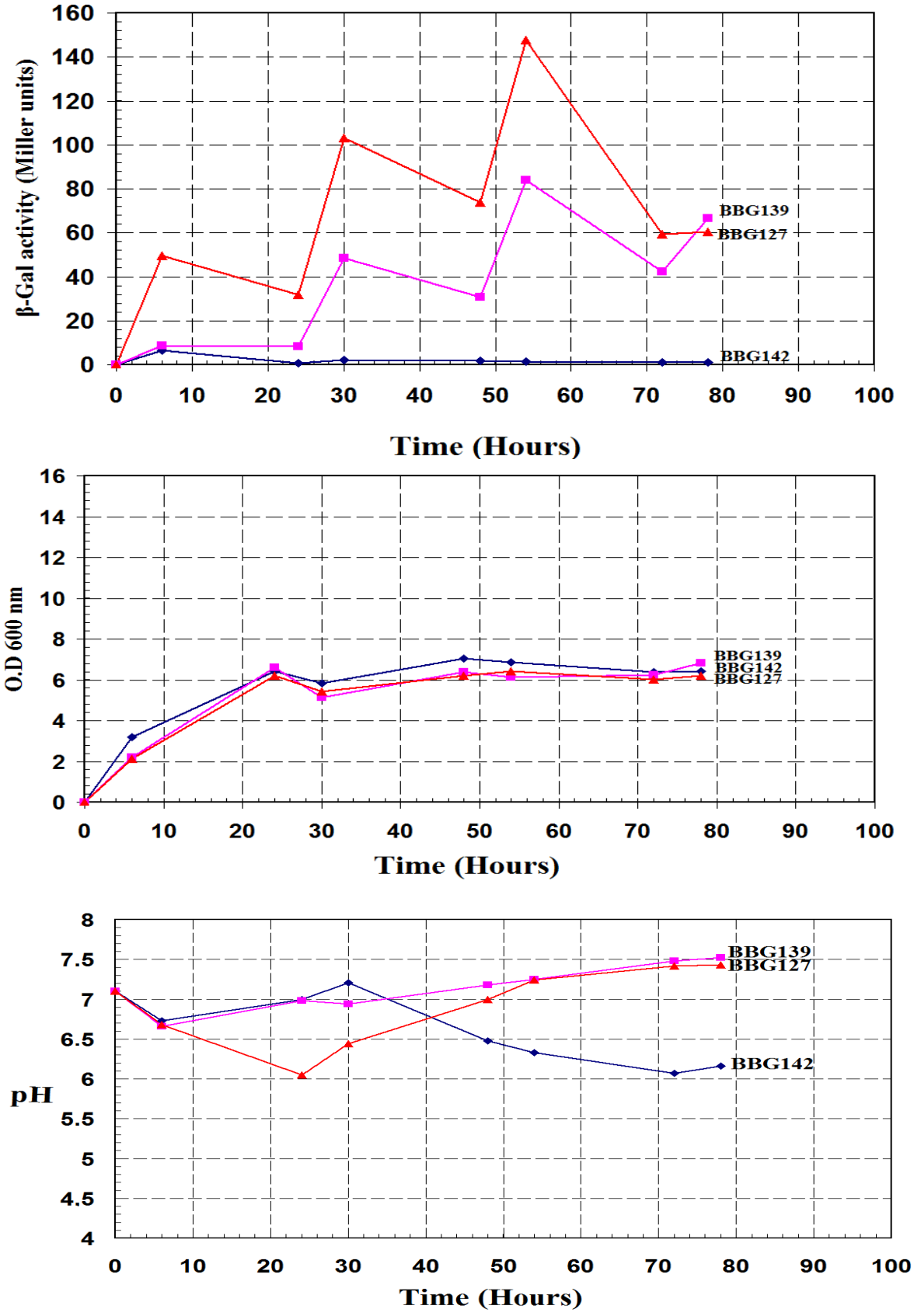



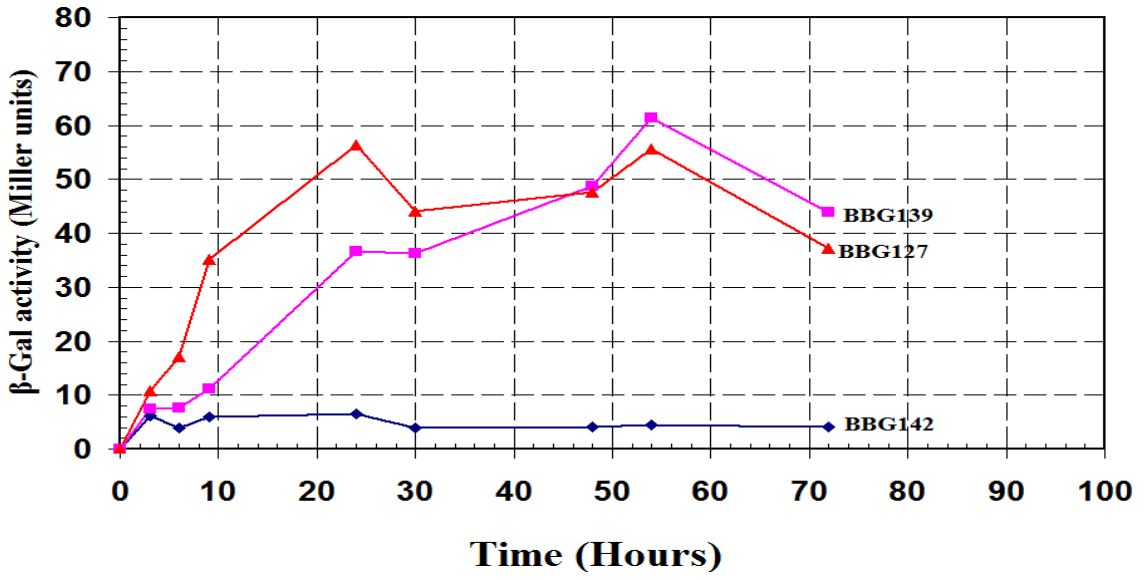

Experiment 6: β-Gal activity expressed in Miller units, microbial growth at ($O.D_{600\ nm}$) and pH for *B. subtilis* BBG142, BBG139 and BBG127 under growth in Landy modified 2, pH 7.0, 20% f.v, 1 L flask, 200 rpm and 37°C

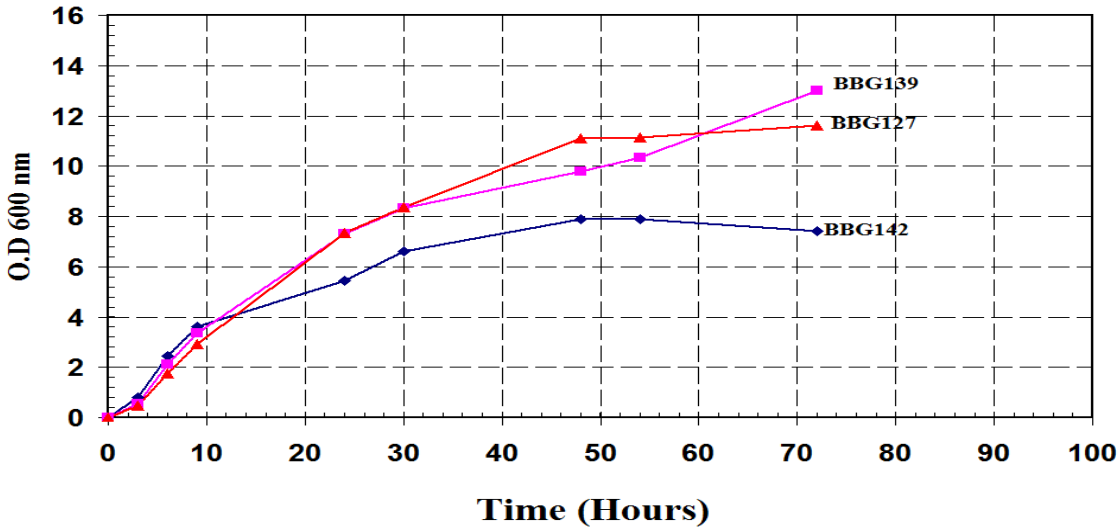



Experiment 7: β-Gal activity expressed in Miller units, microbial growth at ($O.D_{600\ nm}$) and pH for *B. subtilis* BBG142, BBG139 and BBG127 under growth in Landy modified pH 7.0, 20% f.v, 1 L flask, 200 rpm and 25°C

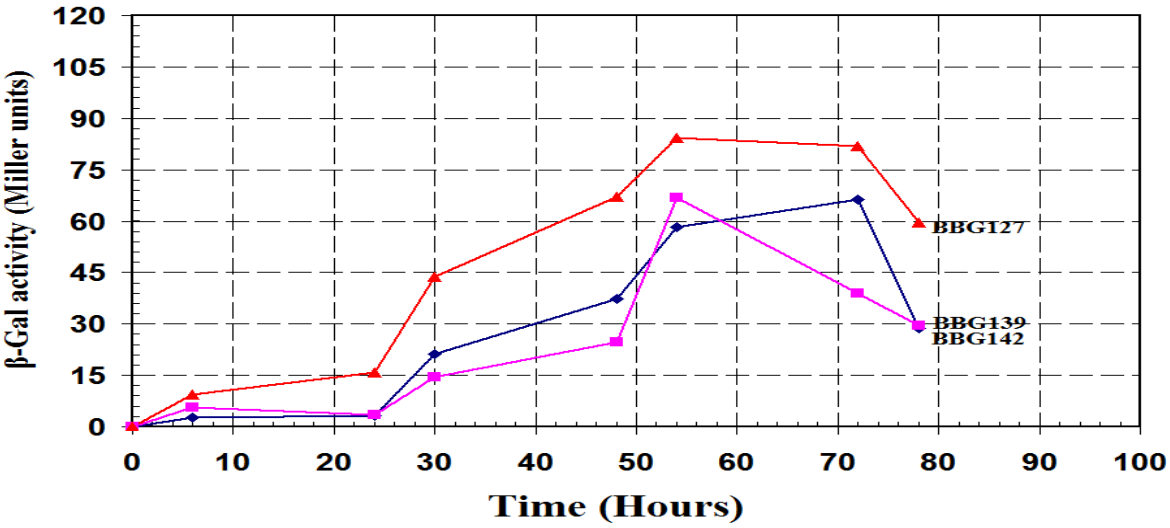


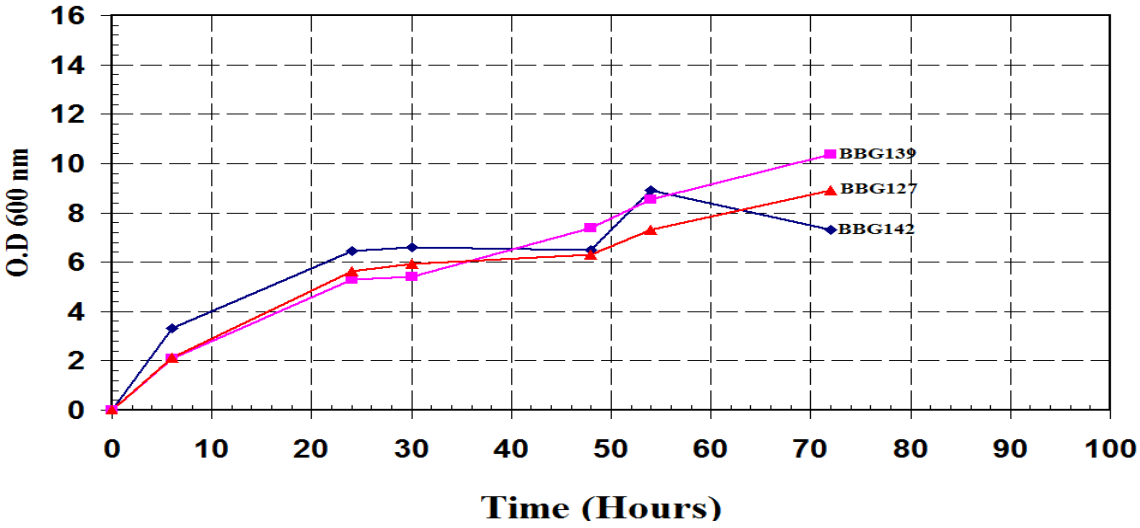

Experiment 8: β-Gal activity expressed in Miller units, microbial growth at ($O.D_{600\ nm}$) and pH for *B. subtilis* BBG142, BBG139 and BBG127 under growth in Landy modified pH 7.0, 20% f.v, 1 L flask, 200 rpm and 37°C

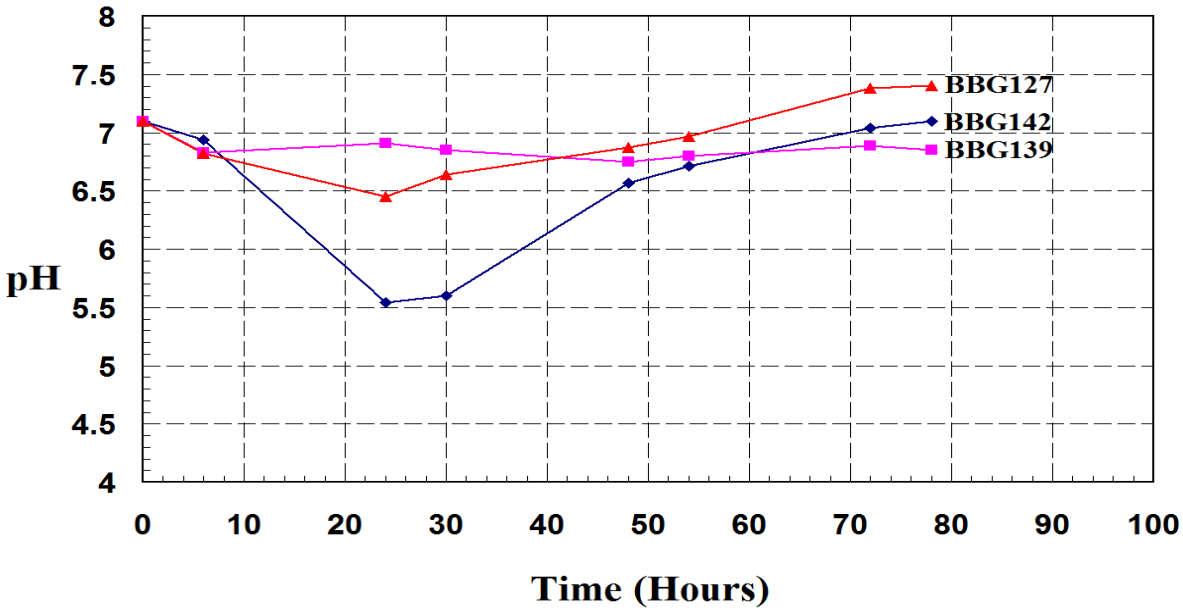


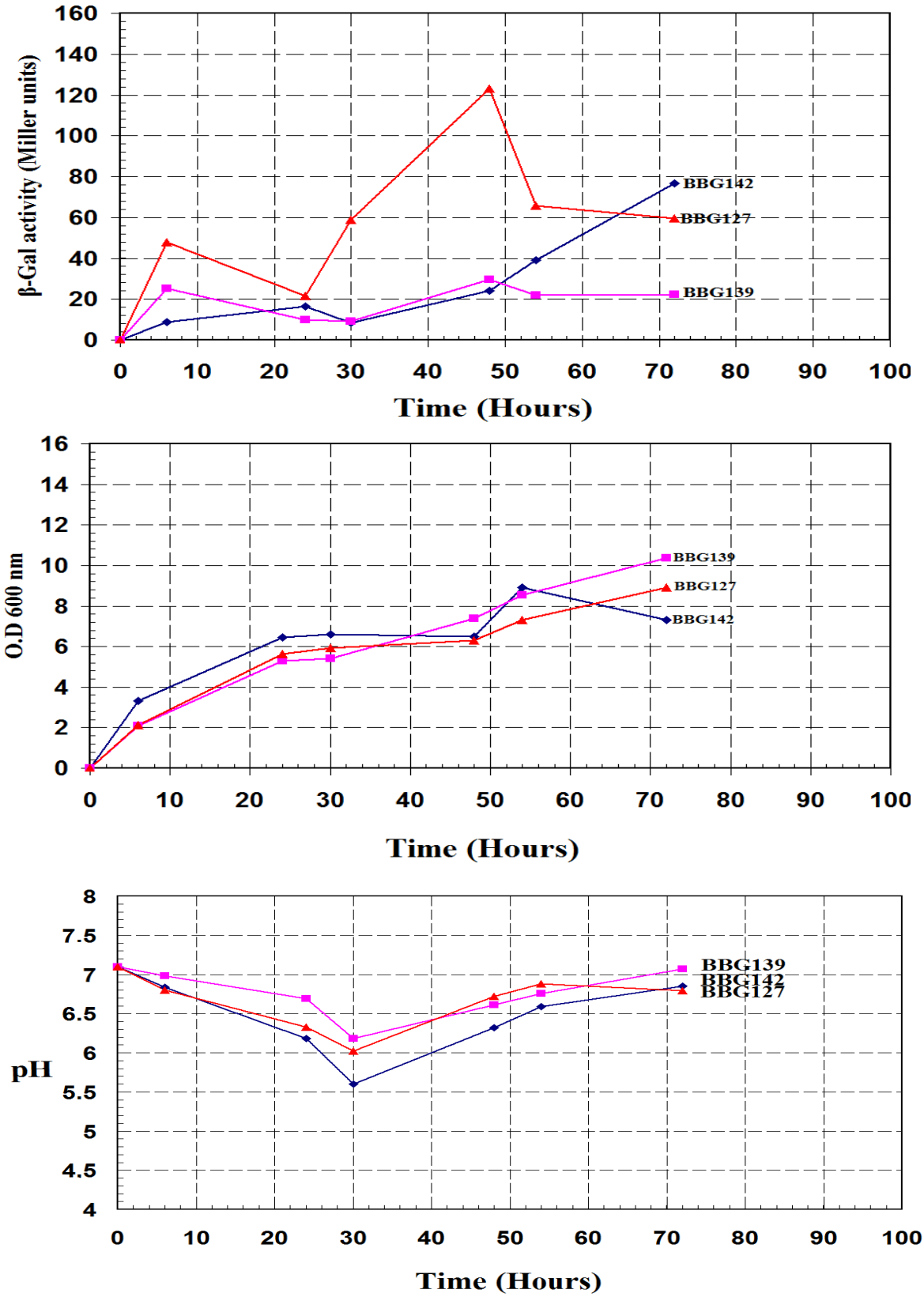

Experiment 9: β-Gal activity expressed in Miller units and microbial growth at ($O.D_{600\ nm}$) for *B. subtilis* BBG142, BBG139 and BBG127 under growth in Landy modified pH 7.0, 20% f.v, 1 L flask, 160 rpm and 30°C



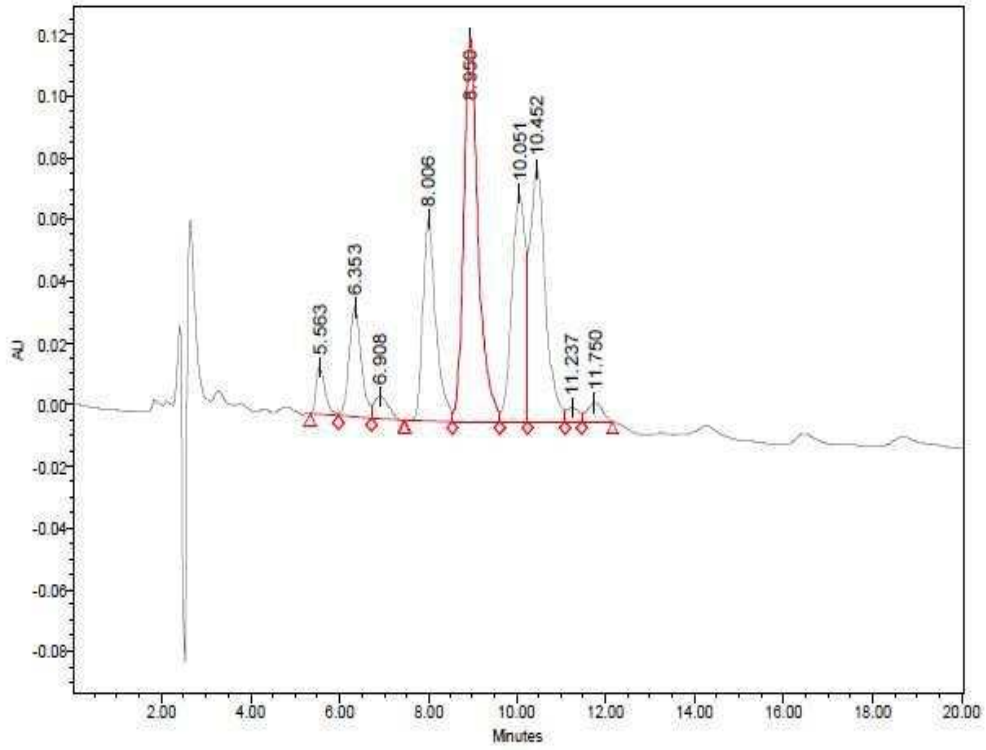

Experiment 10: β-Gal activity expressed in Miller units, microbial growth at ($O.D_{600\ nm}$) and pH for *B. subtilis* BBG142, BBG139 and BBG127 under growth in Landy modified pH 7.0, 20% f.v, 500 mL flask, 160 rpm and 30°



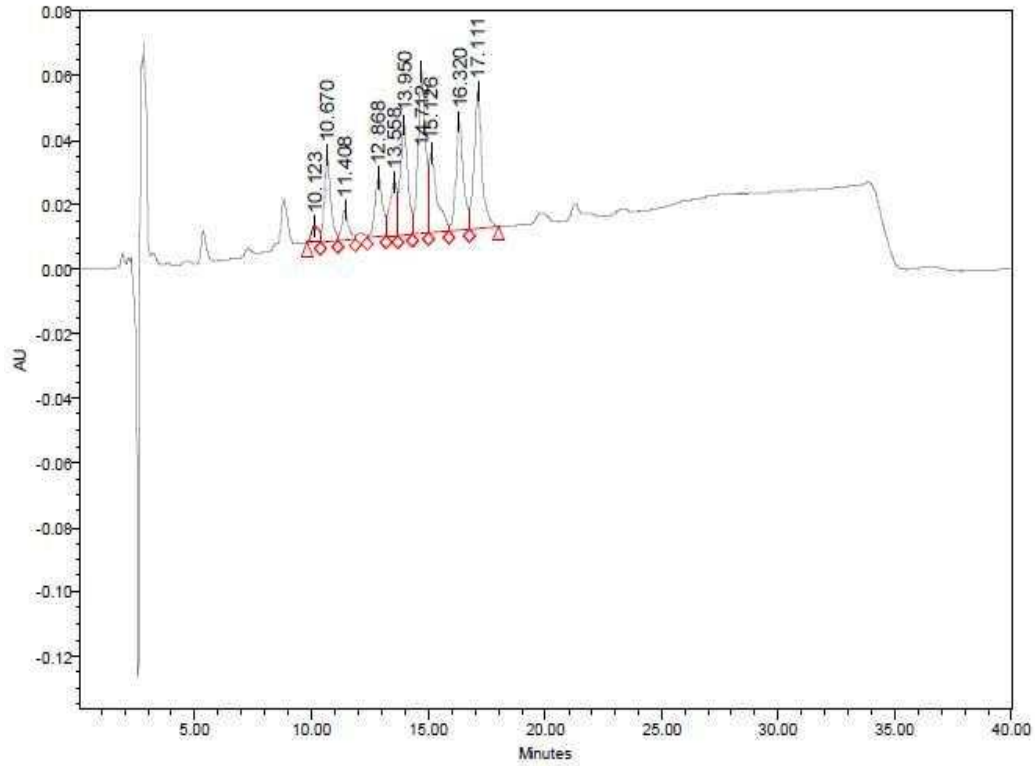

Experiment 11: β-Gal activity expressed in Miller units, microbial growth at ($O.D_{600\ nm}$) and pH for *B. subtilis* BBG142, BBG139 and BBG127 under growth in Landy modified pH 7.0, 30% f.v, 100 mL flask, 160 rpm and 30°C

## Index 8: Chromatograms of lipopeptides (standard peaks)

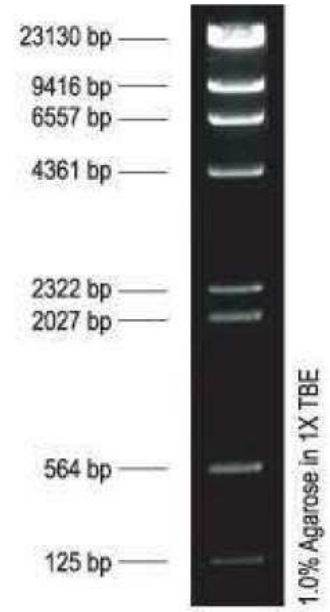
### Surfactin standard spectrum



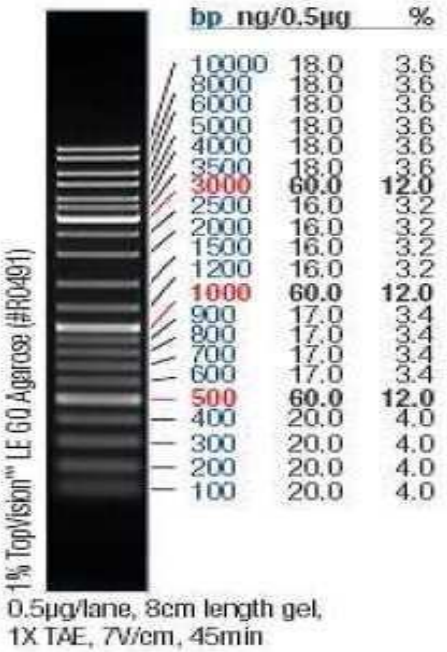

### Fengycin standard spectrum



**Index 9**: **Description of used markers sizes**




Lambda+*Hin*dIII




O 'GeneRuler

# Résumé

Dans un premier temps, 39 génomes publiés de souches de *Bacillus* ont été analysés par des outils bioinformatiques afin de mettre en évidence et caractériser les gènes ou les opérons codant des synthétases de peptides par la voie non ribosomiale (NRPS). Parmi ces 39 souches, 8 ne possèdent aucun gène de ce type. 27 ont le potentiel génétique de production d'un sidérophore, la bacillibactine. 14 souches possèdent des gènes ou opérons codant pour la synthèse d'un ou plusieurs lipopeptides [famille des surfactines, des fengycines, des iturines et des kurstakines] ou d'un antibiotique la bacitracine. L'existence de gènes/opérons codant pour des molécules de type NRPS nouvelles et des molécules issues d'une synthèse hybride NRPS/PKS (Polycétide synthases) a été détectée dans 16 souches. La suite du travail s'est focalisée sur la synthèse des lipopeptides de type fengycines et plispastatines. L'analyse des opérons impliqués dans la synthèse de ces molécules chez les souches *B. subtilis* 168, *B. subtilis* F29-3 et *B. amyloliquefaciens* FZB42, a montré des homologies élevées entre ces opérons. Des premiers éléments d'analyse de l'opéron impliqué dans la synthèse de ces molécules dans la souche *B. subtilis* S499 non séquencée, ont montré une différence qui pourrait être responsable de la structure de la fengycine produite chez cette souche qui ne peut pas être corrélée avec l'organisation des synthétases NRPS déjà décrites. L'obtention d'un mutant mono-producteur de plipastatine, dérivé de *B. subtilis* 168, par le remplacement du promoteur natif $P_{pps}$ par un promoteur constitutif $P_{repU}$ associé à un gène de résistance à la néomycine (*neo*), n'a pas abouti. Par contre, l'inactivation de l'expression de l'opéron plipastatine chez le mutant BBG111 par cette cassette $P_{repU}$-*neo* située en sens inverse de la transcription de l'opéron *pps*, a conduit à l'isolement du dérivé BBG140 montrant une production accrue de surfactine par rapport à la souche parentale. Les promoteurs $P_{pps}$ (*B. subtilis* 168) et $P_{fen}$ (*B. subtilis* BBG21, un mutant isolé au laboratoire et capable de produire des quantités importantes de plipastatines) ont été clonés dans le vecteur d'expression pDG1661, portant la cassette *lacZ*, et intégrés dans le chromosome de la souche 168. Le fonctionnement de ces promoteurs $P_{pps}$ et $P_{fen}$, comparé à celui du promoteur $P_{srfA}$ de *B. subtilis* 168, a été étudié, dans différentes conditions de croissance. Les résultats ont montré que le niveau d'expression de ces trois promoteurs varie en foction des paramètres environnementaux testés.