Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biotechnology and Food Science

Hege Hetland Pedersen

CRISPR-Cas Based Genome Editing for *in-vivo* Biopolymer Engineering in *Paenibacillus polymyxa*

Master's thesis in Chemical Engineering and Biotechnology
Supervisor: Jochen Schmid
July 2020

NTNU
Norwegian University of
Science and Technology

Hege Hetland Pedersen

CRISPR-Cas Based Genome Editing for *in-vivo* Biopolymer Engineering in *Paenibacillus polymyxa*

Master's thesis in Chemical Engineering and Biotechnology
Supervisor: Jochen Schmid
July 2020

Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biotechnology and Food Science

NTNU
Norwegian University of
Science and Technology

Preface

This thesis has been a part of the molecular biology group, with Associate Professor Jochen Schmid as supervisor, at the Department of Biotechnology and Food Science at the University of Science and Technology. This thesis accounts for 30 credits and was conducted as a part of my Master's degree within the study program Chemical Engineering and Biotechnology. This project was conducted during the spring semester of 2020, and is a continuation of the specialization project which was completed during the fall of 2019.

I would like to thank my supervisor for excellent supervision during this work. I am thankful for all the guidance and advice he has given during the project. I would also like to thank Anne Kristin Haraldsvik for the useful insights and discussions along the way.

Finally, I would like to thank my friends and family for their support during the last five years.

Hege Hetland Pedersen

Trondheim, July 29, 2020

Abstract

Many bacteria have the ability to produce exopolysaccharides. These polysaccharides have important functions for the cell, but have also shown promising properties for industrial utilization. There are many different types of polysaccharides, and their diverse chemistry yields polysaccharides with different functions and thus putative novel applications. In addition, the use of genetic engineering can increase the production yields and give the possibility to create tailor-made polysaccharides. *Paenibacillus polymyxa* is able to produce an exopolysaccharide called Paenan, which holds many interesting properties and can give rise to several novel applications for the industry. A CRISPR-Cas9 plasmid called pCasPP has been constructed for targeted genetic modifications of *P. polymyxa*, and has been used to get a better understanding of the Paenan biosynthetic pathway.

In this study, CRISPR-Cas9 is used for two gene deletions and one gene replacement. The two genes, *pepM* and *pepN*, which encode glycosyl hydrolases in the Paenan cluster are deleted. The produced exopolysaccharides from the two deletion-strains are then compared to the wildtype. The aim of the gene replacement is to replace the *pepF* gene, which encodes a glycosyl transferase, with another glycosyl transferase from *Xanthomonas campestris*. This may give an indication on how the chemical structure can be modified. The possibility of other *Paenibacillus* strains to produce Paenan is also investigated by genome comparison. This is conducted using the tools BLAST, RAST and BRIG.

Finally, the *pepM* and *pepN* genes were successfully deleted. The produced exopolysaccharides from these two strains gave shorter fibres compared to the wildtype. This may indicate that glycosyl hydrolases play an important role in the biosynthesis, and possibly are involved in processing the polysaccharide for export or to function as a failsafe enzyme when the export fails. The gene replacement of *pepF* was attempted, but was not successful. The genome comparison of the Paenan cluster showed high similarity between several *Paenibacillus* strains, and indicates that Paenan can be produced by other strains.

Sammendrag

Mange bakterier har evnen til å produsere eksopolysakkarider. Disse polysakkaridene har viktige funksjoner for cellen, men har også vist lovende egenskaper for bruk i industrien. Det eksisterer mange forskjellige typer polysakkarider, med ulik kjemisk struktur. Ulikhetene i den kjemiske strukturen gir polysakkarider med forskjellige funksjoner og derfor også nye antatte anvendelser. Bruk av genredigering kan øke produksjonsutbyttet og gi muligheten til å lage skreddersydde polysakkarider. *Paenibacillus polymyxa* er i stand til å produsere en eksopolysakkarid som kalles Paenan. Denne polysakkariden har mange interessante egenskaper og kan gi opphav til flere nye anvendelser for industrien. Et CRISPR-Cas9 plasmid, som heter pCasPP, har blitt konstruert for å utføre målrettede genetiske endringer i *P. polymyxa*, og det har blitt brukt for å få en bedre forståelse av synteseveien til Paenan.

I denne studien har CRISPR-Cas9 blitt brukt for å slette to gener og erstatte ett gen. De to genene, *pepM* og *pepN*, som koder for glykosylhydrolaser i Paenan-klusteret blir slettet. Eksopolysakkarider ble produsert fra de to stammene med slettet gen, og dette ble sammenliknet med eksopolysakkarider produsert av villtypen. Det andre målet med oppgaven er å slette *pepF*-genet som koder for en glykosyltransferase. Dette genet blir byttet ut med et gen som koder for en annen glykosyltransferase i *Xanthomonas campestris*. Dette kan gi en indikasjon om hvordan den kjemiske strukturen til Paenan kan bli endret. Det blir også studert om andre stammer av *Paenibacillus* har muligheten til å produsere Paenan. Genomet til ulike stammer blir sammenliknet ved bruk av verktøyene BLAST, RAST og BRIG.

Slettingen av genene som kodet for glykosylhydrolaser, *pepM* og *pepN*, var vellykket. Eksopolysakkarider produsert fra disse stammene ga kortere fibre sammenliknet med villtypen. Dette kan tyde på at glykosylhydrolaser spiller en viktig rolle i biosyntesen av Paenan. Mulige roller kan være innen prosesseringen av polysakkarider for eksport eller å fungere som en sikkerhet hvis eksporten mislykkes. Utbyttingen av *pepF* ble forsøkt utført, man var ikke vellykket. Sammenlikningen av Paenan-klusteret viste høy likhet mellom de ulike stammene av *Paenibacillus*, noe som indikerer at Paenan også kan bli produsert av andre stammer.

Abbreviations

A summary of the abbreviations used in this thesis is given below:

ABC - ATP-binding cassette
BLAST - Basic Local Alignment Search Tool
bp - base pairs
BRIG - BLAST Ring Image Generator
Cas - CRISPR-associated
Cas9 - RNA-guided DNA endonuclease
CAZy - Carbohydrate-active enzymes
CRISPR - Clustered regulatory interspaced short palindromic repeats
crRNA - CRISPR RNA
DNA - Deoxyribonucleic acid
DS - Downstream
EPS - Exopolysaccharides
FW - Forward
GH - Glycosyl hydrolase
gRNA - guide RNA
GT - Glycosyl transferase
HDR - Homology-directed repair
IAA - Indole-3-acetic acid
kb - kilo base pairs
LB - Luria broth
LPS - Lipopolysaccharides
NC - Negative control
Neo - Neomycin
NHEJ - Non-homologous end joining
OE-PCR - Overlap extension PCR
OPX - Outer membrane polysaccharide export
ori - origin of replication
oriT - origin of transfer
PAM - Protospacer adjacent motif
PC - Positive control
PCP - Polysaccharide co-polymerase
PCR - Polymerase chain reaction
RAST - Rapid Annotation using Subsystem Technology
REV - Reverse
RNA - Ribonucleic acid
rpm - rounds per minute

rRNA - Ribosomal ribonucleic acid
tracrRNA - Trans-activating CRISPR RNA
US - Upstream
WT - wildtype

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Polysaccharides

Polysaccharides are the most common carbohydrates found in nature. There are many different types of polysaccharides, which differ from each other in the order of their monosaccharide units, types of bonds linking the units together, the length of the chain and branching [1]. Because of the diverse chemistries, the polysaccharides have many different biological functions and industrial applications [2]. Some polysaccharides serve as structural elements, while others are important for energy storage [1].

Some of the most abundant polysaccharides are cellulose and starch. Cellulose serves as a structural element and is commonly found in the cell walls of plants. The cellulose molecule is a linear polymer that consists of D-glucose units linked by β-1-4 glycosidic bonds [1]. Cellulose is highly stable, and exists in plant-based materials like wood and cotton [3, 4]. Starch is a polymer that mainly functions as an energy reserve in plant cells. It consists of amylose and amyplopectin, which both comprise only α-D-glucose. Starch is an important food and food additive, but is also used for industrial applications as fabric stiffeners, glues or in paper treatment [5].

In addition to serving as energy storage and giving stability to cells, some polysaccharides can be produced by microorganisms and secreted into the environment surrounding the cell. The production of cellulose and starch from plants is affected by both region and season, and may differ in structure, molecular weight

and physiochemical properties [6, 7]. Cellulose is an example of a polysaccharide that can be produced by microorganisms [8], where the culture conditions can be controlled in a bioreactor. This way, the product yield and composition is not affected by regional or seasonal factors. Under these culture conditions, it is possible to optimize the yield and manipulate the chemical properties of the polysaccharides [9]. Other advantages with polysaccharides from microorganisms are that they usually have shorter production times and can be extracted more easily, the bioprocesses are performed in bioreactors and therefore don't compete with land for food production and the waste products can be used as feedstocks. In addition, these polysaccharides show a greater variety in structure and properties [10]. Producing polysaccharides from microorganisms is therefore promising.

1.2 Microbial Polysaccharides

Microorganisms are single-cell or multicellular organisms that interact with the surrounding environment and other cells. Examples of microorganisms are bacteria, archea and fungi. Microorganisms have played important roles to humans in for example food industry, as biofuels or in bioremediation. Some microorganisms also produce commercial products such as industrially applied enzymes, antibiotics, chemicals and different polysaccharides [11].

Microbial polysaccharides are biopolymers that are synthesized by microorganisms. These polysaccharides can have important biological functions, such as energy storage, structure and stability. Some can also be secreted to the surrounding environment [12]. Microbial polysaccharides can be divided into three groups [13]:

- Structural polysaccharides
- Intracellular polysaccharides
- Exopolysaccharides

Structural polysaccharides constitute the cell wall, for example lipopolysaccharides (LPS), teichoid acids and peptidoglycans [13]. These polysaccharides are important for the structure of the cell [14].

Intracellular polysaccharides are located inside the cell or as a part of the membrane [15]. They are accumulated in the cell during periods of nutrient excess,

and serve as carbon and energy reserves. These polysaccharides are therefore important for the cell during times of starvation. Glycogen is an example of a major microbial and eucaryotic storage polysaccharide [16].

Exopolysaccharides (EPSs) are secreted into the environment surrounding the cell. Important functions of the EPSs are protection against xenobiotics, water retention, providing adherence to surfaces and they are also the main component of biofilms. For industrial purpose, the EPSs are the most important polysaccharides synthesized by microorganisms [12]. There are many different EPSs produced by microbes, but only a few are being used in the industry today. Table 1.1 shows the major commercial microbial polysaccharides.

1.2.1 Structure of Exopolysaccharides

Polysaccharides are composed of monosaccharide units that can be combined in infinite ways. This yields uncountable EPS variants with unique functions and structures. The monomeric units are linked to each other via glycosidic linkages [12]. Frequently used monomeric units include glucose, mannose, galactose, glucuronic acid, rhamnose and fucose [17]. In addition, these units can carry substituents that are attached through ester or ketal bonds. Carboxylic acids such as pyruvic acid or acetic acid are common substituents [12].

The EPSs can be divided into two classes based on their monosaccharide composition, homopolysaccharides and heteropolysaccharides [12]. Homopolysaccharides consist of only one type of monosaccharide units, while the heteropolysaccharides consist of several different types [17]. The exopolysaccharides are often linear or branched, but hyperbranching is uncommon [12]. The heteropolysaccharides are mainly built up of repeating units, which are repetitive elements that are composed of a specific sequence of the monomers [17]. Commonly, four to eight different monosaccharides build up the heteropolysaccharides [12]. Xanthan gum is an important polymer of industrial relevance and is used in various applications such as the food industry or in oil recovery. The structure of Xanthan is given in Figure 1.1 to present an example of the structure of bacterial EPSs.

Table 1.1: Major commercial polysaccharides produced by microbes. (Table adapted from Schmid et al. [12]) The table gives an overview of the major commercial polysaccharides produced by bacteria and fungi. The organism producing the polysaccharide, the main components and which industries they are used in are also shown. Glc = Glucose, Gal = Galactose, Man = Mannose, GlcA = Glucuronic acid, Rha = Rhamnose, GlcNAc = Glucosamine, mt = metric ton, N.A. = not available.

Polysaccharide	Microorganism	Main Components	Industries	Prize/Production volumes
Xanthan	*Xanthomonas campestris*	Glc, Man, GlcA	Food, oil, health care, personal care	165 000 mt, 3.5-10 US$/kg
Gellan	*Sphingomonas elodea*	Glc, Rha, GlcA	Food	4000 mt, 25-50 US$/kg
Dextran	*Leuconostoc mesenteroides*	Glc	Pharma	<500 mt
Welan	*Sphingomonas* sp.	Glc, Man, Rha, GlcA	Construction, oil	300 mt, 12-25 US$/kg
Diutan	*Sphingomonas* sp.	Glc, Rha, GlcA	Construction, oil	200 mt, 12-25 US$/kg
Curdlan	*Agrobacterium* sp.	Glc	Food	N.A.
Succinoglycan	*Sinorhizobium meliloti*	Glc, Gal	Oil	N.A.
Hyaluronic acid	*Streptomyces pyogenes*	GlcA, GlcNAc	Pharma, personal care	1 bn US$
Scleroglucan	*Sclerotium rolfsii* *Sclerotium glucanicum*	Glc	Oil, personal care	1500 mt, 12-20 US$/kg
Schizophyllan	*Schizophyllum commune*	Glc	Oil, personal care	N.A.
Pullulan	*Aureobasidium pullulans*	Glc	Pharma, personal care	N.A.

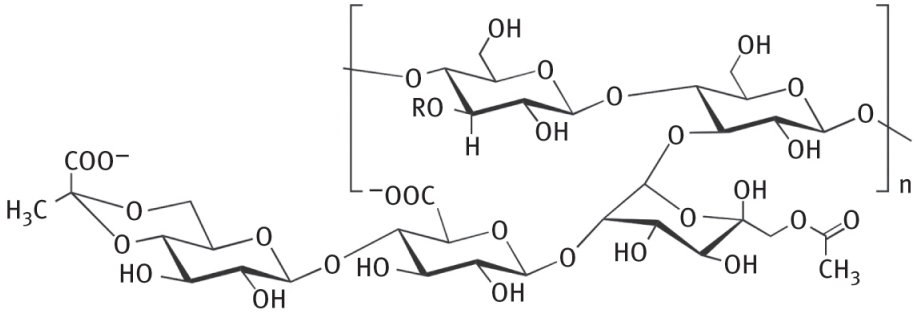


Figure 1.1: Chemical structure of the repeating unit of Xanthan. (Figure by Paderes et al. [18]) The backbone of Xanthan is β-1,4-D-glucoses with a trisaccharide attached to every second glucose molecule. The order of the monomer units in the trisaccharide is mannose-glucuronic acid-mannose [19]. The inner mannose unit is esterified with acetate, while the outer mannose can carry an acetate or a pyruvate group [12].

Xanthan gum is a branched polymer produced by the Gram-negative bacterium *Xanthomonas campestris* [19]. The chemical structure of the repeating unit of Xanthan consists of two D-glucoses, two D-mannoses and one D-glucuronic acid [12]. The backbone of Xanthan is β-1,4-D-glucoses with a trisaccharide attached to every second glucose molecule. The order of the monomer units in the trisaccharide is mannose-glucuronic acid-mannose [19]. The inner mannose unit may be esterified with acetate, while the outer mannose can carry an acetate or a puryvate. These variations depends on the growth conditions [12].

1.2.2 Biosynthesis of Exopolysaccharides

The biosynthesis of EPS consists of several steps, and each EPS is synthesized by its own way. The synthesis often appears in the stationary phase of the growth curve [13]. Most EPS biosynthetic mechanisms follows specific steps: They start with the nucleotide sugar precursors being synthesized, which are used to build up the repeating units. The repeating units are then assembled into the mature polysaccharide before the product is exported [20].

There are four general mechanisms for production of bacterial exopolysaccharides, and these are listed below. Three of them are intracellular and one is extracellular [21]. An overview of the three intracellular pathways is given in Figure 1.2.

- Wzx/Wzy dependent pathway

- ATP-binding cassette (ABC) transporter dependent pathway
- Synthase-dependent pathway
- Extracellular synthesis

The Wzx/Wzy dependent pathway is common for the production of complex heteropolysaccharides, such as Xanthan. In this pathway, the repeating units are assembled by several glycolsyltransferases (GTs) before they are exported through the inner membrane by a flippase called the Wzx protein. Then, the Wzy protein polymerizes the repeating units followed by the transport of the mature EPS to the cell surface. Two additional proteins, assigned the polysaccharide co-polymerase (PCP) and the outer membrane polysaccharide export (OPX) families, are involved in the transport to the cell surface [22].

The ABC transporter dependent pathway is mainly used for the biosynthesis of capsular polysaccharides. As for the Wzx/Wzy dependent pathway, GTs are important for the assembly of polysaccharides. For the assembly of homopolysaccharides, only one GT is involved, while several GTs are involved for the production of heteropolysaccharides [23]. The export to the cell surface is realized by a complex composed of ABC-transporters that spans the inner membrane and proteins of the PCP and OPX families in the periplasm and outer membrane [22].

The third intracellular pathway is the synthase dependent pathway. In this pathway, a synthase performs both the polymerization and transport across the membrane [24]. This is used for the production of homopolysaccharides, for instance cellulose and glucose [21].

Even though most bacterial polysaccharides are produced inside the cell, some are produced extracellular. In this case, specific GTs are secreted to the extracellular space or are covalently bound to the cell surface. The best known examples of polymers produced this way are dextran and levan [21]. The enzyme levansucrase is responsible for the biosynthesis of levan, and this enzyme is produced by a wide range of bacterial strains [25]. Levansucrase synthesizes levan from sucrose. At low concentrations of sucrose, levansucrase functions as a glycosyl hydrolase, while at high concentrations it adds fructosyl units to the growing chain [26].

The genes that are involved in the biosynthesis pathways mentioned above can be found clustered in the genome of EPS producing microbes or on large plasmids [24, 27]. The EPS operons can be identified by several genes encoding GTs, polymerases and secreting enzymes [21].

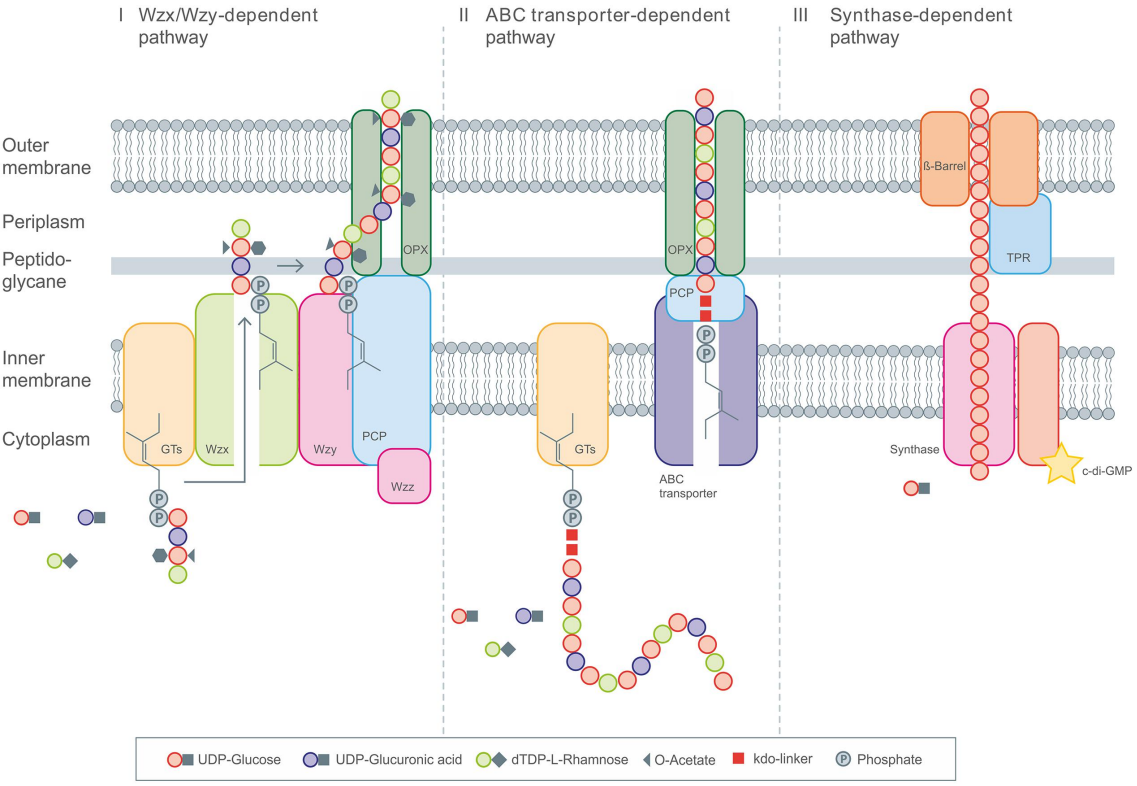


Figure 1.2: Overview of the three intracellular pathways for EPS biosynthesis. (Figure by Schmid et al.[21].) The Wzx/Wzy dependent pathway assemble the repeating units by several glycosyl transferase. The Wzx protein then export the repeating units to the periplasm where the Wzy protein polymerize the repeating units. Then the mature EPS is exported to the cell surface by the polysaccharide copolymerase (PCP) and outer-membrane polysaccharide export (OPX) protein families. The ABC transporter dependent pathway uses glycosyl transferases for assembly of the polymer before it is transported out of the cell by proteins from the PCP and OPX protein families. The synthase dependent pathway uses a synthase for both polymerization and transport across the membrane.

1.2.3 Exopolysaccharide Engineering

In order to improve production or create novel polysachharide variants, the producing organisms can be genetically engineered. Targets for this can be modifications of the molecular weight or the composition of monomers, or the addition of substituents. The strategies can be divided into two categories depending on the purpose. Some may aim for modifications that results in higher productivity and others may aim for new EPS variants. When aiming for a higher productivity, it is typical to try increasing the pool of sugar precursors. This can be done by overexpressing some genes involved in the biosynthesis of the precursors or genes involved in the EPS assembly [21]. Other ways to increase the productivity may be to target the regulatory proteins which may increase the transcription of the EPS gene clusters, or to disrupt the production of pathways using the same precursor molecules [28, 29].

The second strategy for exopolysaccharide engineering gives the opportunity to create tailor-made EPS variants. Using this strategy, the aim is to alter the chemical structure and thereby the properties of the EPSs. This may, for instance, be to change the molecular weight or to modify the chain length of the polymer [30, 31]. For example, Hassler and Doherity performed several mutations of *Xanthomonas campestris* to modify the rheological properties of Xanthan. They found out that by changing the degree of acetylation and pyruvylation the viscosity was massively influenced. The presence of puryate increases the visocity, while the presence of acetate decreases the viscosity. They also found out that by targeting the GTs that catalyze the addition of sugar units to the side chain affected the viscosity. Xanthan without the terminal mannose of the sidechain showed to be a poor viscosifier. In contrast, when the polymer was lacking both the terminal mannose and glucuronic acid it was a good viscosifier [32]. This proves the importance of getting new insight about how the genes and proteins are involved in the EPS biosynthesis. By increasing this understanding, novel tailor-made EPS can be produced with new properties. Modern synthetic biology tools, for example CRISPR-Cas9, can be used for genetic engineering of EPS to produce novel EPS variants effectively. This gives rise to the production of economical competitive tailor-made EPS variants.

1.3 CRISPR-Cas Based Genome Editing

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) together with Cas (CRISPR associated) enzymes works as an adaptive immune system for many bacteria and archaea, by copying a part of the foreign DNA and storing it in the genome [33]. There are five types (type I to V) of CRISPR systems, and type II is most commonly used for genome editing [34]. The CRISPR system can be divided into three steps, as can be seen in Figure 1.3 for the type II system of *Streptococcus pyogenes*.

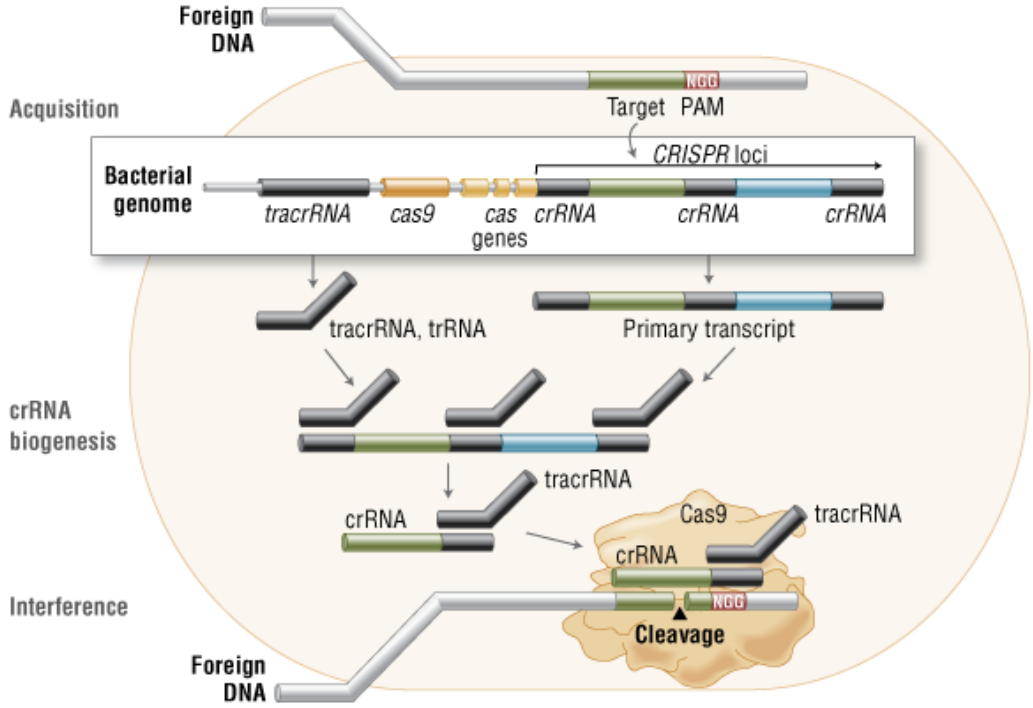


Figure 1.3: The three stages of type II CRISPR system for *Streptococcus pyogenes*. (Figure by New England Biolabs [35]). In stage one, aquisition, short sequences of the foreign DNA (spacers) are inserted into the CRISPR locus of the host's genome. In the crRNA biogenesis phase, the pre-crRNA is transcribed and bind to tracrRNA. This duplex is cleaved and gives mature crRNA. During interference, the duplex binds to Cas9 and guides it to the target sequence which is next to a PAM sequence. The Cas9 makes a double-strand break.

In the first stage, the acquisition stage, Cas proteins recognize the foreign DNA and cut it into short sequences of about 30 bp in length [36]. These sequences are then inserted into the CRISPR locus of the host's genome. This creates a chimeric sequence that consists of repeat sequences interspaced by the short sequences of foreign DNA, called spacers [37]. The CRISPR locus functions as a memory of previous invasions, and gives immunity for later exposure of the same DNA sequences. Different Cas proteins transcribe and process the CRISPR locus [38]. The primary transcript, the pre-CRISPR RNA (pre-crRNA), consists of repeats and spacers [39]. During the second stage, called crRNA biogenesis, the pre-crRNA is processed by the host's RNase III with help from the trans-activating CRISPR RNA (tracrRNA). The tracrRNA contains repeats that can bind to the pre-crRNA through homology. The duplex of pre-crRNA and tracrRNA is cleaved to become mature crRNA [40]. Mature crRNA consists of one spacer sequence and one repeat region [39]. In the last stage, the interference stage, the crRNA:tracrRNA duplex binds to the Cas9 protein. The Cas9 protein is a DNA endonuclease, and is guided by the crRNA to the target sequence. Through sequence homology between the spacer and the target sequence, the complex can bind to the target and create a double-strand break [41]. For the nuclease to be able to bind, a Protospacer Adjacent Motif (PAM) sequence must be present downstream of the target [42]. This is a short sequence of 3-5 bp, and for Cas9 this sequence is NGG where N can be any nucelotide [43]. The PAM sequence is not present in the CRISPR locus of the host's genome, and it therefore prevents the host from cutting it's own genome.

The use of CRISPR has made genome editing more precise, less expensive and faster. The CRISPR/Cas9 system used for genome editing only requires a single guide RNA (gRNA), which is a fusion of tracrRNA and crRNA. The gRNA conatins both a spacer region and a nuclease binding domain. The spacer region has a length of about 20 nucleotides and gives homology to the target sequence. In order to perform genome editing, the CRISPR system must be delivered to the target. Plasmids are commonly used as a delivery mechanism. It is also important that the plasmid can be replicated, and that an efficient selection is possible. An antibiotic resistance gene is often delivered with the plasmid as a selective marker [36].

In order to repair the double-strand break created in the target sequence, two repair mechanisms can be used: non-homologous end joining (NHEJ) and homology-directed repair (HDR). These two pathways are shown in Figure 1.4.

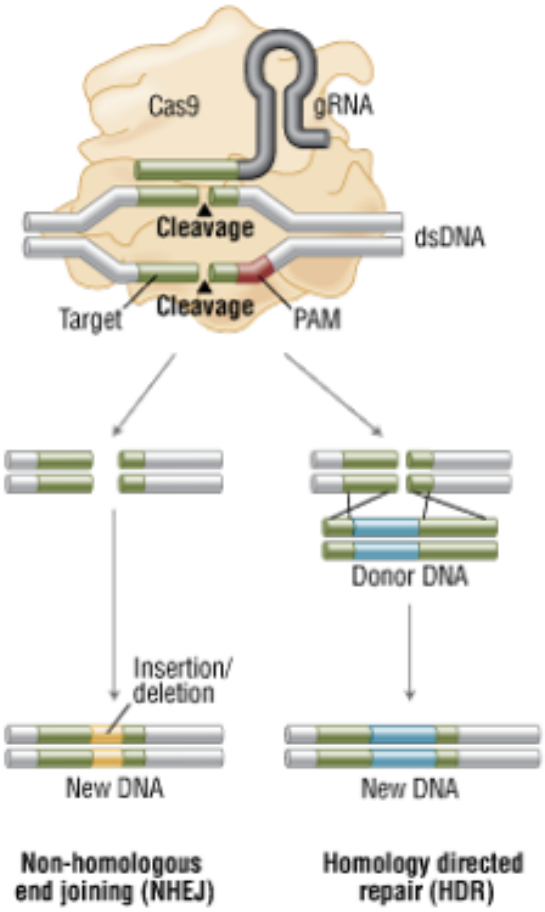


Figure 1.4: CRISPR-Cas9 cleavage and DNA repair mechanisms. (Figure by New England Biolabs [35]) A double-strand break is created by the Cas9 endonuclease. This activates a DNA repair mechanism of the cell. When a homologous repair template is not present, the non-homologous end joining mechanism is activated. This can result in both deletions and insertions, and cause frameshift mutations. When a homologous repair template is present, the cell can use the homology-directed repair mechanism. This can give point mutations, knock-outs and knock-ins.

For NHEJ, the double-strand break is religated directly and a homologous template is not required [44]. This mechanism can create both deletions and insertions of nucleotides, which causes frameshift mutations. However, most bacteria are not able to use NHEJ to repair the double-strand breaks [45]. Therefore, most bacteria must rely on HDR to repair the double-strand breaks. In this repair mechanism, a homologous strand is used as a template for repair [46]. A DNA

strand with homologous sequences upstream and downstream of the target gene can be recognized by the bacteria, and can thereby remove the target gene or replace it with another sequence. By providing a template for HDR, the cleaved sequence can be introduced to mutations, deletions or foreign genes.

A CRISPR-system for *Paenibacillus polymyxa* (*P. polymyxa*) has been developed by Rütering et al. [47], which has been constructed as a plasmid called pCasPP. The system consists of origin of replication (ori), origin of transfer (oriT), the *repU* gene and a neomycin (neo) resistance gene. The replication is initiated at the origin of replication, while the origin of transfer is important for conjugation. The *repU* gene is involved in the replication of the plasmid. For the gene encoding Cas9, the constitutive promoter *sgsE* is used. The constitutive *gapdh* promoter is used for guide expression. The unspecified system contains a lacZ casette that is surrounded by the cut sites of the restriction enzyme BbsI. This gives the opportunity to insert the gRNA using Golden Gate Cloning. An HDR-template can be inserted at the SpeI restriction cut site.

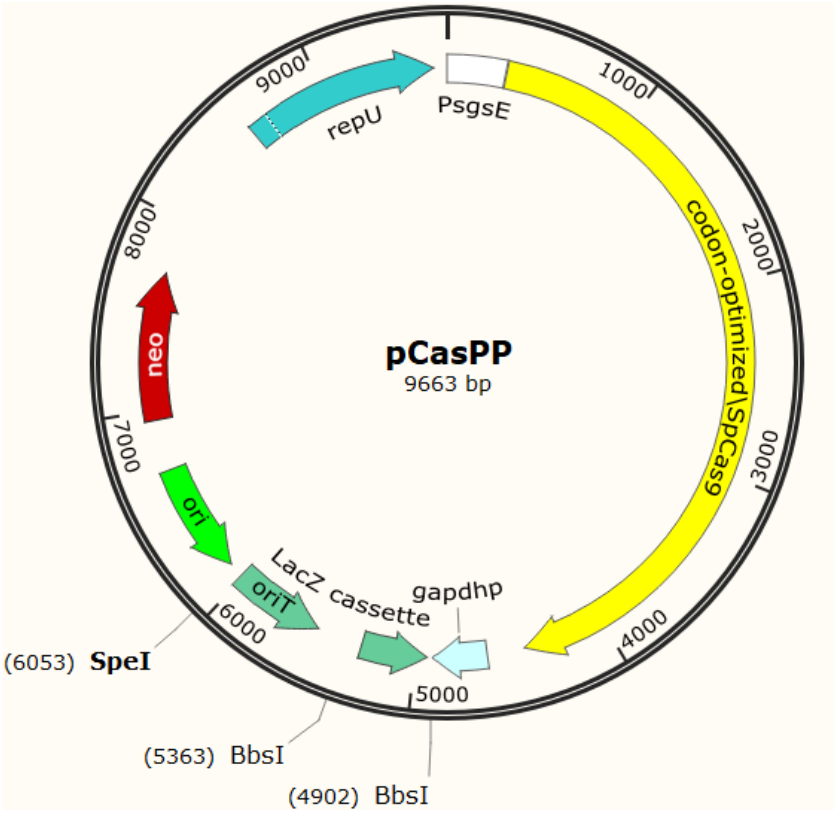


Figure 1.5: CRISPR-Cas9 system for *P. polymyxa* designed for *pepM* deletion. The system consists of origin of replication (ori), origin of transfer (oriT), the *repU* gene and a neomycin resistance gene. The promoters *sgsE* and *gapdh* are used for the Cas9 encoding gene and the gRNA, respectively. The restriction enzymes BbsI and SpeI gives the opportunity to insert the gRNA and the HDR-template.

1.4 *Paenibacillus* spp.

Bacteria that belongs to the genus *Paenibacillus* can be found in various environments. They exist in both polar and tropic regions, aquatic environments and dry deserts. There are about 200 different species of *Paenibacillus* known today, and most of them originate from soil [48]. In the beginning, the bacteria belonging to *Paenibacillus* were included in the genus *Bacillus* because of morphological similarities and the ability to form endospores. In 1991, *Bacillus* was divided into five highly divergent groups based on their 16S rRNA sequence [49]. This led to the phylogenetic distinction of *Paenibacillus* from other *Bacillus* groups [50].

Paenibacillus species are facultative anaerobic Gram-positive bacteria that have the ability to form endospores [51]. They can also grow on a large number of carbon sources because of the ability to break down complex polymers [11]. A scanning electrone microscope picture of *P. polymyxa* is given in Figure 1.6.

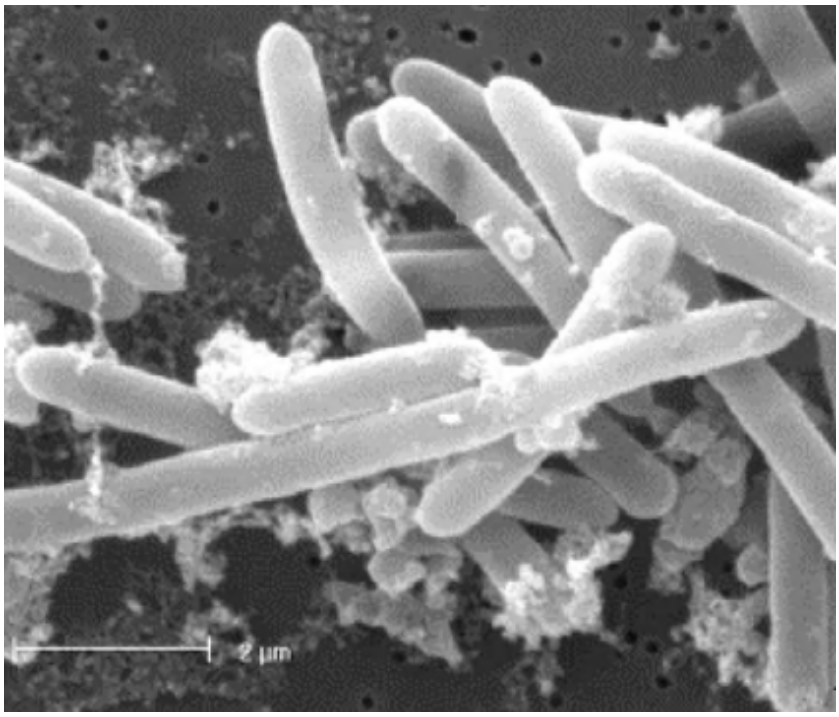


Figure 1.6: Picture of *P. polymyxa* (Figure by Raza et al [51]).

The species of the *Paenibacillus* genus can affect the environment both positively and negatively. Since most *Paenibacilli* originate from soil, they are often associated with plant roots. Many of them promote plant growth, and positive effects have been seen for different plants including rice and pumpkin [52, 53]. One way the *Paenibacilli* promote plant growth is by increasing the bioavailability of essential elements like phosphorus and nitrogen. More than 20 species have been shown to have nitrogen fixing abilities by reduction of N_2 to NH_3 [54]. Genomic analyses have also suggested that by producing glucuronic acid, most *Paenibacilli* can solubilize inorganic phosphorus [55]. The secretion of auxins also promotes plant growth, where indole-3-acetic acid (IAA) is the most abundant one [56]. Even though plants produce IAA themselves, they can also use exogenous sources. Since *Paenibacilli* can produce IAA, they can thereby contribute to plant growth [57]. *Paenibacillus* species exhibit several biocontrol capabilities that promote plant growth. This may both be to induce the plant's own resistance or to pro-

duce biocidal substances. For instance, *P. polymyxa* have been proven to enhance the resistance of peanut roots against crown root disease [58].

Paenibacilli are also relevant for the pharmaceutical industry because of their ability to produce various antimicrobial agents. The most known antibiotics are the polymyxins. However, polymyxins have shown to be toxic to the human central nervous system and kidneys, limiting their usage [59]. On the other hand, synthetic approaches aim to produce new variants of polymyxins with improved pharmacokinetic properties [60]. Other examples of antimicrobial substances produced by *Paenibacilli* are fusaricidins, pediocins and lantibiotics [61, 62, 63]. In addition to antimicrobial substances, *Paenibacilli* also produce many different enzymes including cellulases, amylases, lipases, hemicellulases, pectinases and lignin-modifying enzymes. These enzymes have potential to be used in industry for food, paper, detergents, textiles and biofuels [48]. By producing enzymes that catabolize aliphatic and aromatic pollutants, *Paenibacillus* can also be used in wastewater treatment [64, 65]. The production of EPSs is also an important property of the genus *Paenibacillus*, which will be further discussed in Section 1.4.1.

There are also some negative aspects about *Paenibacillus*. Some *Paenibacillus* species infect various organisms. For example, *P. larvae* is known to cause American Foulbrood that afflicts honeybees [66]. Another negative aspect, is that *Paenibacillus* has a role in dairy spoilage. *Paenibacilli* produce spoilage enzymes, and by being able to survive under extreme conditions, spores can be found in both raw and pasteurized milk [67].

1.4.1 Exopolysaccharides by *Paenibacilli*

Paenibacillus species produce various EPS, and each strain produces different EPSs with different properties[68]. There are huge differences between the structures of the different EPSs, and both homopolysaccharides and heteropolysaccharides are reported to be produced by *Paenibacilli* [69, 70, 71]. This results in many different physicochemical properties and gives rise to several potential applications. One proposed application is as thickeners or rheological agents because of the high viscosity and shear thinning behavior [72]. In addition, several EPSs from *Paenibacilli* have shown properties for medical applications, such as antioxidative potential, improvement of skin hydration, antitumor activity, and enhancement of immunity [68]. Other fields for the use of these EPSs may include emulsification, bioremediation and animal feed additives. The number of known EPS producing *Paenibacillus* species is quite small when compared to the number of described

species, and most of them are strains of *P. polymyxa* or species that are phylogenetically closely related.

The medium composition may influence the EPS production. Glucose and sucrose are the most common carbon sources. The use of sucrose in the medium gives high yields of levan-type EPSs because of the activity of levansucrase, which is present in many *P. polymyxa* strains. The EPSs produced by *Paenibacilli* usually consists of the monomers mannose, glucose, glucuronic acid, galactose, and sometimes pyruvate and fucose. Table 1.2 gives an overview of some of the EPSs produced by *Paenibacillus*, along with their carbon- and nitrogen-sources. Because of the use of different carbon- and nitrogen-sources, it can be difficult to directly compare the results.

1.4.2 Paenan

P. polymyxa DSM 365 is a producer of an EPS called Paenan. The molecular weight of Paenan is about $1.3 \cdot 10^7$ g/mol, and consists of the monomers glucose, mannose, galactose and glucuronic acid. The ratio of these monomers are 3.5:2:1:0.2, respectively. This polymer shows great thickening potential, and is therefore interesting for use as a rheological additive [83].

The genes involved in the Paenan biosynthesis are found in a cluster that spans almost 35 kb. This cluster consists of 28 genes and is shown in Figure 1.7. Several of the functional elements in this cluster are encoded twice. This could be because of a evolutionary development for a more reliable production of EPS, or it could be that the strain is able to produce another polymer where some of the genes are used in both pathways. For example polymerases and precursors may be involved in both pathways, while GTs are unique for each polymer. In addition, some of the precursor genes (*ugdH*, *manC*, *galU*) are also found in other parts of the genome, whereas the genes for fucose synthesis *fcl* and *gmd* are only present in the Paenan cluster [47].

Table 1.2: Summary of exopolysaccharides produced by *Paenibacillus* strains. The table shows the monomer composition of the EPS structure, the C-source and the N-source. Man=mannose. Gal= galactose. Glc= glucose. GlcA= glucuronic acid. Fuc= fucose. Pyr= pyruvate. Fru= fructose. UAs= uronic acids. Rha= rhamnose. Xyl= xylose. Tyr= tyrosine. Mal= maltose.

Strain	**Monomer composition**	**C-Source**	**N-Source**	**Reference**
P. polymyxa SQR-21	Man, Gal, Glc, GlcA	Gal	Yeast extract	[73]
P. polymyxa A49 KCTC 8648P	Man, Gal, Glc, GlcA, Fuc	Glc	NH_4NO_3, tryptone, yeast extract, soytone	[72, 74]
P. polymyxa NCIB 11429	Man, Gal, Glc, GlcA, Fuc, Pyr	Suc	Yeast extract	[75]
P. polymyxa NRRL B-18475	Fru	Suc	Yeast extract	[70]
P. polymyxa EJS-3	Man, Glc, Fru	Suc	Yeast extract	[76]
P. polymyxa JB115	Glc	Suc	NH_4Cl	[69]
P. jamilae CECT 5266	Man, Gal, Glc, UAs, Fuc, Rha	Olive-mill wastewater	NH_4Cl, yeast extract	[77]
P. elgii B69	Man, Glc, GlcA, Xyl	Suc	Peptone, yeast extract	[71]
P. polymyxa 1465	Man, Gal, Glc, UAs	Glc	$(NH_4)_2SO_4$	[78]
P. sp WN9 KCTC 8951P	Man, Glc, Tyr	Glc	Yeast extract	[79]
P. velaei sp. nov.	Man, Gal, Glc, Fuc	Glc	NH_4Cl	[80]
P. sp. TKU023	Glc, Mal	squid-pen powder		[81]
P. polymyxa No.271	Man, Gal, Glc, GlcA, Pyr	Glc	Peptone, urea	[82]

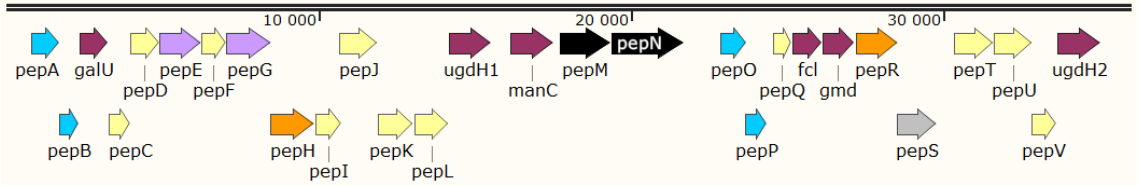


Figure 1.7: EPS cluster of *P. polymyxa* DSM 365. The gene cluster of *P. polymyxa* DSM 365 for production of Paenan. The blue arrows are genes involved in regulation of transport/chain length. Yellow arrows are genes encoding glycosyl transferases. The grey arrow is an uncharacterized gene. Black arrows are genes encoding glycosyl hydrolases. Dark purple arrows show genes responsible for synthesis of nucleotide sugars. Orange arrows represents genes that encodes flippases. Purple arrows are genes encoding polymerases.

The EPS production of Paenan is believed to follow the Wzx/wzy pathway [47]. This pathway can be divided into three parts as illustrated in Figure 1.8. First is the nucleotide sugar biosynthesis, followed by the assembly of the repeating units before polymerization and export. This is described in more detail in Section 1.2.2. In the figure below, several GTs are involved in the assembly of the polymer adding monomers to the repeating units. These genes can be targeted to alter the structure and chemical properties of Paenan. The functional role of all GTs are not currently known, and the final structure of Paenan is therefore not known either.

Rütering *et al.* did several deletions in the Paenan cluster, which revealed information about some of the genes in the cluster. For example, they deleted the *pepF* gene. By studying the monomer composition of the Δ*pepF* variant, they found out that galactose content was reduced by 50 % compared to the wildtype, as well as an almost complete loss of pyruvate. This indicates that half of the galactoses are pyruvylated, and that *pepF* is responsible for the transfer of the pyruvylated galactoses. By targeting genes like *pepF*, more information about the biosynthesis and also the opportunity to make tailor-made EPSs with modified properties important for the industry can be obtained. The deletion of *pepF* clearly showed different physicochemical properties compared to the wildtype, which may give rise to an EPS that can be used for other applications [47].

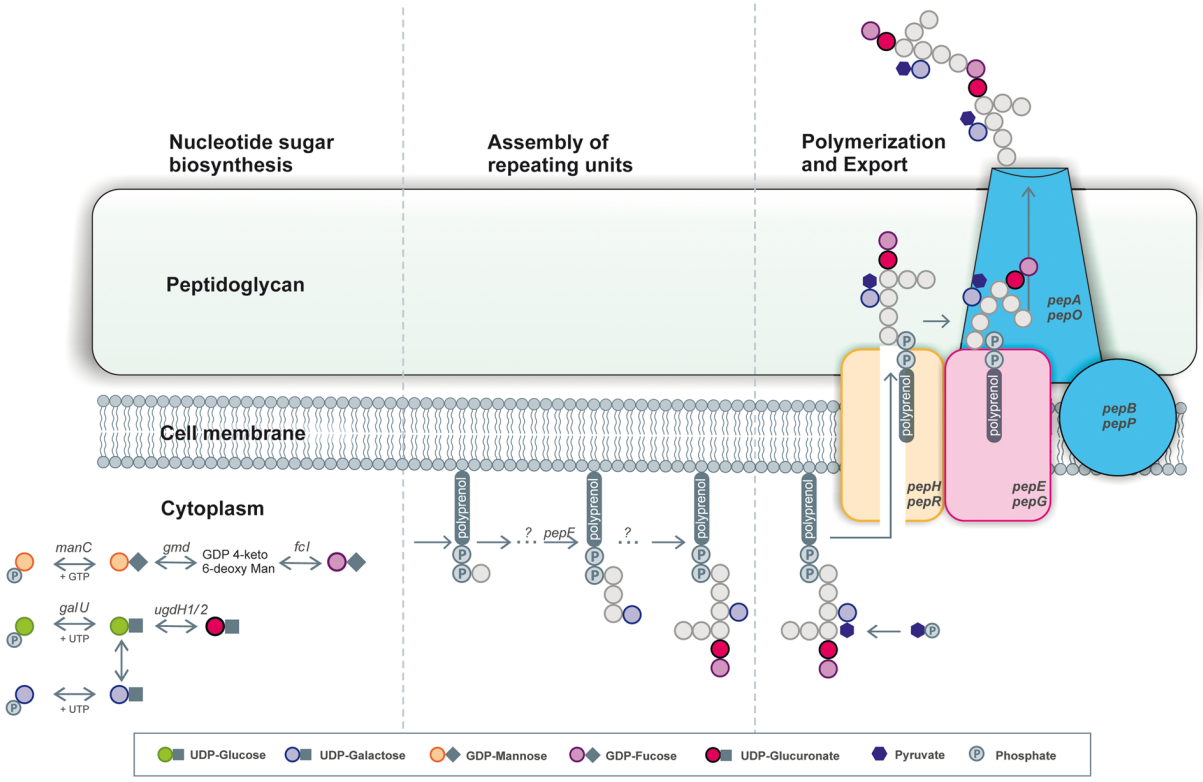


Figure 1.8: Description of the Wzx/Wzy-pathway for biosynthesis of Paenan. (Figure by Rütering et al.[47]). The Wzx/Wzy-pathway can be divided into three parts. First is the nucleotide sugar biosynthesis, followed by the assembly of the repeating units before polymerization and export. This figure shows only a hypothetical structure of Paenan.

1.5 Glycosyl Hydrolases

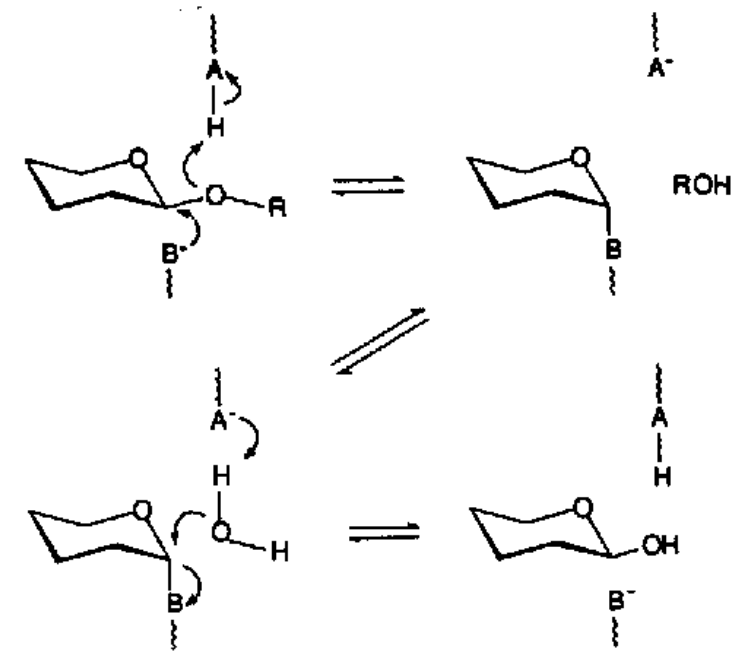
Hydrolases belong to the enzyme class that break chemical bonds by the use of water, a reaction which is called hydrolysis [1]. This typically means that a larger molecule is divided into a smaller molecule. Glycosyl hydrolases (GHs) are an example of a group of hydrolases. These are enzymes that catalyze the hydrolysis of glycosidic bonds. In 1991, a classification of GHs based on their amino acid sequence was started [84]. This classification is updated regularly in The Carbohydrate-Active Enzymes database (CAZy), which currently contains 167 registered GH families [85].

Two residues are important for hydrolysis by GHs: a nucleophile/base and a proton

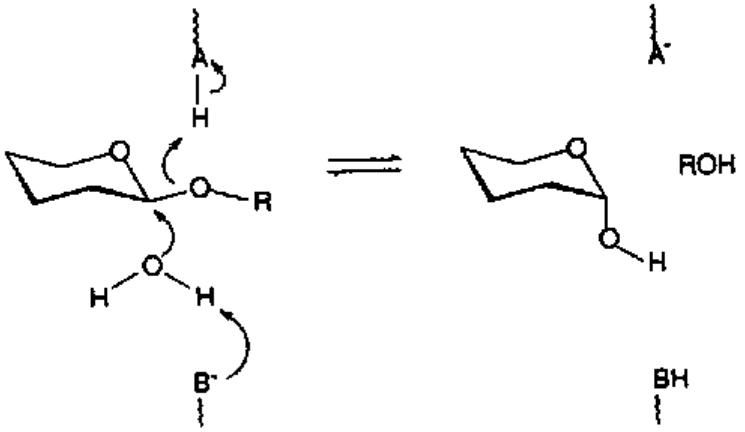
donor. In general, there are to main mechanisms which either give a retention or an inversion of the anomeric configuration. The proton donor is in hydrogen-bonding distance to the glycosidic oxygen for both inverting and retaining enzymes. However, the nucleophilic base is more distant for inverting enzymes than retaining enzymes [86]. Figure 1.9 shows the two mechanisms. As can bee seen from the figure, the retaining mechanism gives a product with the same stereochemistry as the substrate. The inverting mechanism gives a product with the opposite stereochemistry.

GHs are found in both prokaryotes and eukaryotes [84]. In prokaryotes, these enzymes are found both extracellular and intracellular and are important for nutrient acquisition. One example of a GH in prokaryotes is the β-galactosidase (LacZ) that is involved in the regulation of the *lacZ* operon in *Escherichia coli* (*E. coli*) [1]. In eukaryotes, they are found in the endoplasmic reticulum and Golgi apparatus where they are important for the processing of N-linked glycoproteins. They are also found in the lysosomes where they degrade carbohydrates [87].

Some studies have also found GHs to be essential for EPS production, but the function of these enzymes are not fully understood. For example in the cellulose biosynthesis in *Acetobacter xylinum*, there are two glycosyl hydrolases that are important for the production. BglxA, a β-glucosidase, may be a part of the regulation of the material for the substrate for the production of cellulose. It may also be a part of the regulation of other genes and proteins [88]. Another study identified that PslG from *Pseudomonas aeruginosa* , a β-D-xylosidase, is involved in the disassembly of biofilms. PslG was found to be important for the production of Psl, which is an EPS important for biofilm formation. By deleting the *pslG* gene, less Psl was produced. However, overexpressing the same gene, led to an increased degradation of the Psl polysaccharide [89]. It has been suggested that the PslG enzyme is responsible for processing the polysaccharide for export, or to act as a failsafe enzyme by clearing the periplasm of abnormal polysaccharides when the export fails. PelA is another hydrolase involved in EPS production. The function is not known, but it is predicted to have similar roles as PslG [90].



(a) Retension



(b) Inversion

Figure 1.9: Main mechanisms of glycosyl hydrolases. The two main mechanisms are retension (a) and inversion (b). Retension gives a product with the same stereochemistry as the substrate. Inversion gives a product with the opposite stereochemistry as the substrate.

1.6 Aim of the Thesis

The aim of this thesis is to do two targeted gene deletions and one gene replacement in *P. polymyxa* using the CRISPR-Cas9 based pCasPP system. For the gene replacement, the *pepF* gene in the genome of *P. polymyxa*, should be replaced by the *gumK* gene from *Xanthomonas campestris*. The goal is to change the chemical structure of Paenan. For this, a plasmid constructed during the specialization project will be used [91].

According to Rütering *et al.*, *pepF* is believed to encode a glycosyl transferase that transfer a galactose, which is further pyruvylated, to a mannose unit [47]. The *gumK* gene, which encodes a glycosyl transferase, is involved in the biosynthesis of Xanthan [20]. GumK adds a glucuronic acid to a mannose. Both *pepF* and *gumK* encodes glycosyl transferases that adds a sugar unit to a mannose. In both cases, the mannose is connected to a glucose. Because of this similarity, it might be possible to alter the chemical structure of Paenan. By replacing *pepF* with *gumK*, the pyruvylated galactose might be exchanged with a glucuronic acid. Figure 1.10 shows the putative chemical modification.

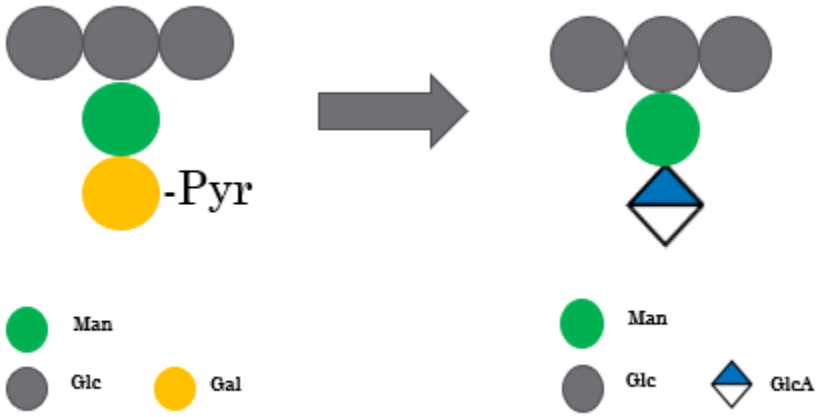


Figure 1.10: Chemical modification of Peanan by gene replacement. By replacing the *pepF* gene with the *gumK* gene, a modification of the chemical structure of Paenen may occur. The galactose and pyruvate which is added to Paenan by a GT encoded by *pepF* may be replaced with a glucuronic acid which is added by a GT encoded by the *gumK* gene from *X. campestris*.

The role of glycosyl hydrolases will also be explored in this thesis, by deletion of

pepM and *pepN*. These genes are believed to encode glycosyl hydrolases. The two genes will be deleted separately, and the EPS produced from these strains will be studied and compared to the wildtype. The plasmids will be constructed using Golden Gate cloning for guide insertion and overlap-extension PCR for insertion of the HDR-template.

In addition, the Paenan gene cluster and the whole genome of *P. polymyxa* DSM 365 will be explored further using different genome comparison tools. This can give information about the similarity between different *Paenibacillus* species, and which other species that may have the ability to produce Paenan.

Chapter 2

Material and Methods

2.1 Media and Solutions

Luria Broth (LB) medium

10 g/L Tryptone
5 g/L Yeast Extract
5 g/L NaCl
The components were added to sterile water and autoclaved.

LB agar medium

15 g/L Agar Extract was added to the LB medium before autoclavation.

Vitamin solution

0.02 g/L D-Biotin
0.3 g/L Choline chlorid
0.1 g/L Folic acid
3.5 g/L Myo-inositol
0.1 g/L Niacinamide

0.1 g/L p-Amino benzoic acid
0.025 g/L D-Pantothenic acid $\cdot$0.5 Ca
0.1 g/L Pyridoxal $\cdot$HCl
0.02 g/L Riboflavin
0.1 g/L Thiamine $\cdot$HCl
0.0005 g/L Vitamin B12
0.2 g/L KCl
0.2 g/L KH_2PO_4 (Anhydrous)
8.0 g/L NaCl
1.15 g/L Na_2HPO_4

Trace element solution

1.80 g/L $MnCl_2 \cdot 4H_2O$
2.50 g/L $FeSO_4 \cdot 7H_2O$
258 mg/L Boric acid
31.0 mg/L $CuSO_4 \cdot 7H_2O$
21.0 mg/L $ZnCl_2$
75.0 mg/L $CoCl_2 \cdot 6H_2O$
23.0 mg/L $MgMoO_4$
2.10 g/L Sodium tartrate $\cdot 2H_2O$
All the elements were dissolved and then sterile filtered.

EPS medium

30.0 g/L Glucose
5.0 g/L Casein peptone
1.33 g/L $MgSO_4 \cdot H_2O$
20.0 mL/L NaOH [2.0 M] (added to adjust pH)
20.0 mL/L KH_2PO_4 [83.5 g/L]
1.0 mL/L $CaCl_2 \cdot 2H_2O$ [50 g/L]
2.0 mL/L Vitamins solution
1.0 mL/L Trace elements solution
Glucose and Casein peptone were autoclaved separately. Stock-solutions of magnesium sulphate, potassium phosphate and calcium chloride were autoclaved separately. All solutions were then mixed together under sterile conditions and pH was adjusted to neutral.

Neomycin stock solution [50 mg/mL]

0.50 g Neomycin
10 mL Sterile water
The Neomycin was dissolved completely before it was filtered through a 0.22 µm syringe filter.
For cultivation, neomycin was added to LB medium with a concentration 1:1000.

Polymyxin stock solution [20 mg/mL]

0.20 g Polymyxin
10 mL Sterile water
The Polymyxin was dissolved completely before it was filtered through a 0.22 µm syringe filter.
For cultivation, polymyxin was added to LB medium with a concentration 1:1000.

2.2 Preparation of Cryo Cultures

For long time storage of cells, 400 µl of 60 % glycerol was added to 600 µl cell culture. The tubes were stored in a -80 °C freezer.

2.3 *In silico* design

Two softwares, SnapGene and Benchling, were used for designing primers, guide RNA design and *in-silico* design of the plasmids.

2.3.1 Benchling

Benchling is an online tool for DNA sequence editing, analyzing data, sharing research and running experiments. In this thesis, it was mainly used for designing CRISPR guide RNA. By using this tool, the guide RNA sequences can be visualized, optimized and annotated. The possible guide sequences are listed with on- and off-target scores. These sequences can then be assembled into plasmids [92].

2.3.2 SnapGene

SnapGene is a software that can be used to plan, visualize and document molecular biology procedures. During this thesis it has been used to design primers, insert restriction enzyme cut sites, visualize assembly reactions and check sequencing results [93].

2.4 Bacterial Strains

In this thesis, *Escherichia coli* (*E. coli*) was used for cloning procedures and *P. polymyxa* was the target organism for the CRISPR-system. Table 2.1 shows the bacterial strains used in this thesis.

Table 2.1: Overview of the bacterial strains used in this thesis.

Bacterial strain	**Description**	**Source**
P. polymyxa DSM 365	Target strain of CRISPR-Cas9 system.	DSMZ
P. polymyxa rep_pepF	The *pepF* gene replaced by the *gumK* gene.	This work
P. polymyxa ΔpepM	The *pepM* gene deleted.	This work
P. polymyxa ΔpepN	The *pepN* gene deleted.	This work
E. coli DH5α	Used for standard cloning purposes. Chemically competent, rapid growth.	NEB 5-alpha Competent *E. coli* (New England Biolabs, catalog number: C2987)
E. coli S17-1	F^+ strain. Used for conjugal transfer.	*Escherichia coli* (ATCC©47055™)

2.5 Plasmids

The plasmids used in this thesis are presented in Table 2.2.

Table 2.2: Plasmids used in this project.

Plasmid	Description	Source
pCasPP	CRISPR-Cas system for gene editing in *P. polymyxa*.	[47]
pCasPP_sg1_pepM	CRISPR-Cas system with spacer sequence for *pepM* deletion inserted.	This work
pCasPP_sg1_pepN	CRISPR-Cas system with spacer sequence for *pepN* deletion inserted.	This work
pCasPP_ΔpepM	CRISPR-Cas system for deletion of *pepM* in *P. polymyxa* by homologous template inserted.	This work
pCasPP_ΔpepN	CRISPR-Cas system for deletion of *pepN* in *P. polymyxa* by homologous template inserted.	This work
pCasPP_rep_pepF	CRISPR-Cas system for replacing *pepF* with *gumK* in *P. polymyxa* by homologous template inserted.	[91]

2.6 DNA Isolation

Isolating plasmid DNA

The GeneJET Plasmid Miniprep Kit (K0502) was used to extract the plasmid DNA from cells. The plasmids were extracted from *E. coli* cultures that had been incubated at 37 °C for 12-16 hours while shaking at 225 rpm. The bacterial culture was harvested by centrifuging 5 ml of the culture at 8000 rpm for 2 minutes at room temperature. The supernatant was removed by decanting, and the cell pellet was resuspended in 250 µl Resuspension Solution and transferred to a 1,5 ml microcentrifuge tube. Then, 250 µl Lysis Solution was added to the tube. In this step the cells were lysed under strong alkali conditions, and the cell contents were released. The lysate is then neutralized with 350 µl Neutralization Solution. This gives appropriate conditions for the plasmids to bind to the membrane of the spin column. The tube was then centrifuged for 5 minutes at 14000 rpm to pellet the chromosomal DNA and cells. The plasmids were then in the supernatant which was transferred to a spin column in a collection tube. This was centrifuged for 1 minute at 14000 rpm for the plasmid DNA to bind to the membrane. The flow-through was discarded. To remove contaminants, 500 µl of Wash Solution

was added and centrifuged for 1 minute at 14000 rpm. The flow-through was discarded, and the wash step was repeated to remove residual contaminants. The spin column was then centrifuged for an additional minute to remove the remaining Wash Solution. The column was then transferred to a 1.5 ml Eppendorf tube. To elute the DNA, 50 µl Elution Buffer was added and incubated for 2 minutes at room temperature before centrifugation at 14000 g for 2 minutes.

Isolating genomic DNA

To extract genomic DNA from *P. polymyxa*, the DNeasy Blood and Tissue Kit from QIAGEN was used. First, the cells were harvested by centrifugation for 10 minutes at 5000 g. The cell pellet was then resuspended in 180 µl Lysis Buffer, and incubated at 37 °C for 30 minutes. 25 µl proteinase K and 200 µl buffer AL was then added to the soultion and incubated at 56 °C for 30 minutes. This step releases the DNA and the cellular proteins are dissolved. 200 µl of ethanol was then added, and the mixture was transferred to a spin column that was placed in a collection tube. This was centrifuged for 1 minute at 6000 g. In this step, the DNA binds to the spin column. Next, the spin column was placed in a new collection tube. 500 µl Buffer AW1 was added to the column and centrifuged for 1 minute at 6000 g. The spin column was transferred to a new collection tube and 500 µl Buffer AW2 was added and centrifuged for 3 minutes at 20000 g. This step removes residual ethanol. The spin column was placed in a 1.5 ml Eppendorf tube and 200 µl Buffer AE was added. The column was incubated at room temperature for 1 minute before the DNA was eluted by centrifugation at 6000 g for 1 minute.

2.7 Molecular Cloning

2.7.1 Polymerase Chain Reaction

Polymerase chain reaction (PCR) is a method used for amplifying DNA sequences. In this thesis the Phusion High-Fidelity DNA Polymerase was used to amplify the sequences for cloning purposes, while Standard Taq DNA Polymerase was used to verify insertions.

PCR Protocol for Phusion High-Fidelity DNA Polymerase

The PCR Protocol for Phusion High-Fidelity DNA Polymerase (M0530) from NEB was used in this thesis [94]. The components shown in Table 2.3 were mixed in a PCR tube. The polymerase was added last to prevent degradation of the primers.

Table 2.3: Components for the Phusion DNA Polymerase reaction.

Component	**20 µl reaction**
5x Phusion HF-buffer	4 µl
Phusion DNA Polymerase	0.2 µl
dNTPs	0.4 µl
DMSO	0.6 µl
10 µM Forward Primer	1 µl
10 µM Reverse Primer	1 µl
Template DNA	50 ng
Water	add to 20 µl

The PCR tubes were then transferred to a PCR machine running the program shown in Table 2.4:

Table 2.4: The PCR reaction for Phusion DNA Polymerase. The annealing temperature depends on the primers, and uses the melting temperature of the primer with the lowest melting temperature. The extension time is dependent on the amplicon length.

Step	**Temperature** [°C]	**Time**
Initial denaturation	98	30 seconds
30 cycles	98	10 seconds
	60	30 seconds
	72	30 seconds/kb
Final Extension	72	10 minutes
Hold	4	∞

PCR Protocol for *Taq* DNA Polymerase

The Protocol for *Taq* DNA Polymerase (M0273) from NEB was used in this thesis [95]. The components shown in Table 2.5 were mixed together in a PCR tube where the polymerase was added last.

Table 2.5: Components for the *Taq* DNA Polymerase reaction.

Component	**50 µl reaction**
10X Standard *Taq* Reaction Buffer	5 µl
Taq DNA Polymerase	0.25 µl
10 mM dNTPs	1 µl
10 µM Forward Primer	1 µl
10 µM Reverse Primer	1 µl
Template DNA	<1000 ng
Water	add to 50 µl

Then, the program shown in Table 2.6 was run by a PCR machine.

Table 2.6: The PCR reaction for *Taq* DNA Polymerase. The annealing temperature depends on the primers, and uses the melting temperature of the primer with the lowest melting temperature. The extension time is dependent on the amplicon length.

Step	**Temperature** [°C]	**Time**
Initial denaturation	95	30 seconds
30 cycles	95	30 seconds
	60	1 minute
	68	1 minute/kb
Final Extension	68	5 minutes
Hold	4	∞

2.7.2 Golden Gate Cloning

Golden Gate Cloning was used to insert the gRNA into the pCasPP plasmid. First, the gRNA was phosphorylated and annealed. This was done by mixing the components shown in Table 2.7.

Table 2.7: Components for phosphorylating and annealing gRNA. FW = forward. REV = reverse.

Component	**Volume**
Primer FW 100 µM	1 µl
Primer REV 100 µM	1 µl
Polynucleotide kinase	1 µl
T4 ligase buffer	1 µl
Water	6 µl

The mixture was incubated at 37 °C for 30 minutes. Then it was heated up to 95 °C for 5 minutes, before it was cooled down to 25 °C by decreasing the temperature at 0.1 °C/sec. The mixture with the annealed gRNA was then diluted 1:100. To insert the gRNA into the pCasPP plasmid, the Golden Gate reaction mixture shown in Table 2.8 was prepared.

Table 2.8: Components for the Golden Gate Assembly.

Component	**Volume**
BbsI	0.75 µl
T7 ligase	0.25 µl
10X CutSmart Buffer	1.0 µl
ATP 10 mM	1.0 µl
DTT 10 mM	1.0 µl
Plasmid backbone	100 ng
Annealed gRNA 1:100	1.0 µl
Water	Up to 10 µl

This mixture was then put into a PCR machine with the program shown in Table 2.9:

Table 2.9: PCR program for Golden Gate Assembly.

20 cycles	37 °C	5 minutes
	25 °C	5 minutes
Heat inactivation	80 °C	20 minutes
Hold	12 °C	

After Golden Gate Cloning, the PCR product was transformed into *E. coli* DH5α. The obtained colonies were then screened with colony PCR.

2.7.3 Overlap Extension PCR

Overlap extension PCR (OE-PCR) was used to combine DNA fragments by creating overlaps complementary to each other. The strands will this way function as primers for each other when mixed. In this thesis, OE-PCR was used to combine the homologous arms.

First, the two homologous arms were amplified by PCR with Phusion DNA polymerase in separate tubes as described in section 2.7.1. The PCR products were run on a gel, and the correct fragments were extracted from the gel and purified. To fuse the fragments, the reaction shown in Table 2.10 was used:

Table 2.10: Components for the fusion of the fragments. DS: homologous arm down stream. US: homologous arm up stream.

Component	**Volume /Amount**
Phusion HF-buffer	6 µl
Phusion DNA Polymerase	0.3 µl
dNTPs	0.6 µl
DMSO	0.6 µl
DS	50 ng
US	50 ng
Water	add to 30 µl

The PCR program shown in Table 2.11 was run to fuse the fragments.

Table 2.11: PCR program used for fusion of the fragments.

Temperature [°C]	**Time**
98	1 min
98	10 s
55	30 s → 10 cycles
72	15 s
12	∞

The next step of the overlap extension PCR was to amplify the fused sequence. Table 2.12 shows the components added to the reaction in Table 2.10.

Table 2.12: Components added to the PCR mixture where the fragments were fused.

Component	**Volume** [µl]
Phusion HF-Buffer	4
Phusion DNA Polymerase	0.2
dNTPs	0.4
DMSO	0.4
Primer US FWD	5
Primer DS REV	5
Water	5

For this reaction, the PCR program shown in Table 2.13 was used.

Table 2.13: PCR program used to amplify the fused fragment.

Temperature [°C]	**Time**
98	10 s
T_m (60)	30 s → 20 cycles
72	15 s/kb
72	10 min
12	∞

2.7.4 Restriction Digest

In order to insert the homologous arms into the pCasPP plasmid, both the fragment and the vector were digested with restriction enzymes giving compatible ends. The plasmid was digested with the SpeI restriction enzyme, while the fragment was digested with the XbaI resitrction enzyme. Figure 2.1 shows the cut sites of these two enzymes. As can be seen from the figure, both enzymes give the same sticky ends.

(a) XbaI **(b)** SpeI

Figure 2.1: Cut sites of the resitriction enzymes XbaI and SpeI. (Pictures by New England Biolabs [96, 97])

The components shown in Table 2.14 were mixed.

Table 2.14: Components for the restriction digest reaction.

Component	Volume/Amount
Enzyme	1 µl
DNA	1000 ng
CutSmart Buffer	5 µl
Water	up to 50 µl

Digestion of DNA using restriction enzymes will give 5' phosphate ends. For ligation to occur, at least one of the ends should contain a 5' phosphate. Therefore, 1 µl Quick CIP was added to the mix for plasmid digestion to prevent self ligation of the plasmid. The mixtures were incubated at 37 °C for 1 hour. After incubation, the enzymes were heat inactivated for 20 minutes. The SpeI enzyme was heat inactivated at 80 °C, while XbaI was heat inactivated at 65 °C. Subsequently, the mixtures were purified using the E.Z.N.A. Cycle Pure Kit described in Section 2.10.

2.7.5 Ligation

After purification of the digestion mix, the plasmids and the fragments were ligated. The T4 ligase from New England Biolabs (NEB) was used. The molar ratio 1:3 of vector to insert was used, and calculated using NEBioCalculator [98]. Table 2.15 shows the components for the ligation reaction:

Table 2.15: Components for the ligation reaction. The amount of vector and insert DNA was calculated using NEBioCalculator [98].

Component	Volume
T4 DNA Ligase	1.0 µl
T4 DNA Ligase Buffer	2.0 µl
Vector DNA	0.020 pmol
Insert DNA	0.060 pmol
Water	up to 20 µl

The mixtures were incubated at 16 °C overnight. The ligated plasmid was then ready for transformation.

2.7.6 Colony PCR

Colony PCR was used to verify that the desired DNA construct was present in the bacteria after transformation or conjugation. After incubation on agar plates, up to 20 colonies were picked and transferred to a new plate and then transferred to the PCR tube using the same toothpick. The *Taq* DNA Polymerase was used for this reaction. A master mix of the PCR mix was prepared by the following protocol:

Table 2.16: Components of the master mix for Colony PCR.

Component	**150 μl reaction**
10X Standard *Taq* Reaction Buffer	15 μl
10 mM dNTPs	3.0 μl
10 μM Forward Primer	3.0 μl
10 μM Reverse Primer	3.0 μl
Taq DNA Polymerase	0.75 μl
Nuclease-free water	125.5 μl

10 µl of the master mix was added to each PCR tube. Table 2.17 shows the PCR program that was run. The initial heating step is important for the DNA to be released from the cells.

Table 2.17: The PCR reaction for colony PCR using *Taq* DNA Polymerase. The annealing temperature depends on the primers, and uses the melting temperature of the primer with the lowest melting temperature. The extension time is dependent on the amplicon length.

Step	**Temperature** [°C]	**Time**
Initial denaturation	95	5 minutes
30 cycles	95	30 seconds
	60	1 minute
	68	1 minute/kb
Final Extension	68	5 minutes
Hold	4	∞

After the PCR, the samples were loaded on a gel for gel electrophoresis.

2.8 DNA Analysis

2.8.1 Nanodrop Spectrophotometer

A Nanodrop spectrophotometer was used for DNA concentration and purity measurements. 1-2 µl of the DNA sample was loaded to the Nanodrop. The same elution buffer that was used for extracting or purifying the DNA was used as a blank. After measuring, three values are given: concentration, A260/280 and A260/230. The concentration is given in ng/µl. The A260/280 ratio gives information about protein contamination where a ratio of 2.0 indicates a pure DNA sample. The A260/230 ratio gives information about RNA contamination, and this value should be close to 1.8 [99].

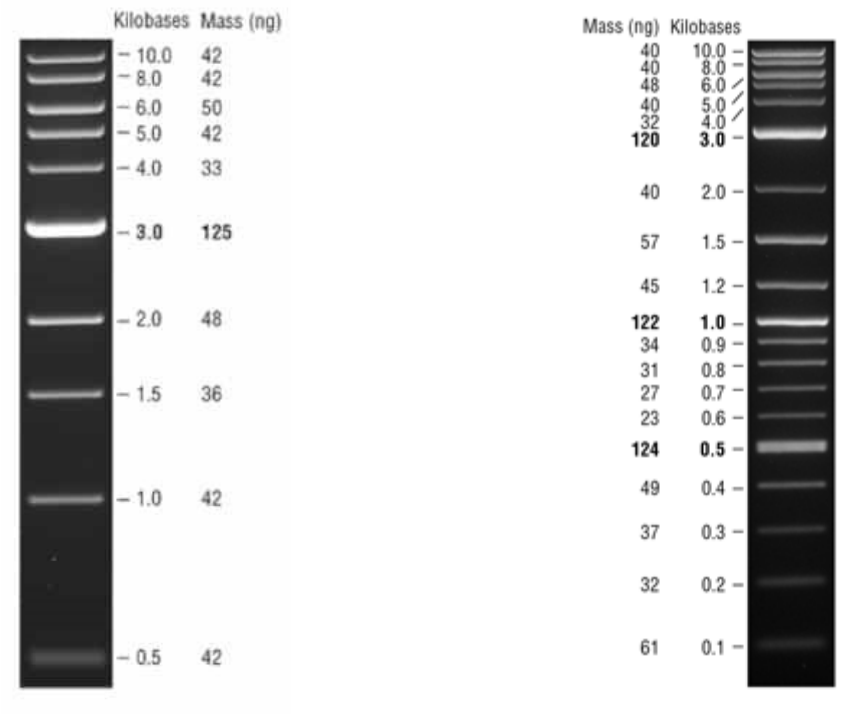
2.8.2 Gel Electrophoresis

Gel electrophoresis is used to separate different DNA fragments based on their size. In this thesis, 0.8 % agarose in 1xTAE buffer was used. GelRed was used to visualize DNA fragments, while GelGreen was used when the fragments were extracted from the gel. Purple 6x Gel Loading Dye from NEB was used to dye the DNA fragments. This was mixed using the amounts shown in Table 2.18.

Table 2.18: Amounts of the components for dying samples and ladder.

Dying of sample
10 µl sample
2 µl loading dye
Dying of ladder
9 µl dH_2O
2 µl loading dye
1 µl ladder

In this thesis, 1 kb DNA ladder and 1 kb Plus DNA ladder from NEB used. These are shown in Figure 2.2.



(a) 1 kb Ladder **(b)** 1 kb Plus Ladder

Figure 2.2: The ladders used in this thesis: 1 kb and 1 kb Plus DNA ladder from NEB.

The ladder and the samples were loaded on the gel and run for 40-50 minutes at 90 V. After running the gel, the fragments were visualized under UV light. Shorter fragments will migrate further down the gel than the longer fragments, and can thereby be separated.

2.9 Isolation of DNA from Agarose Gels

The E.Z.N.A. Gel Extraction Kit was used to isolate the DNA fragments that were extracted from the agarose gel. First, the gel fragments were weighed. One volume of Binding Buffer (XP2) was added by assuming a density of 1 g/ml of the gel fragments. To dissolve the gel piece, the tube was incubated at 60 °C for 7 minutes. Then, the solution was transferred to a HiBind DNA Mini Column in a Collection Tube and centrifuged for 1 minute at 10000 g. The flowthrough was discarded, and 300 µl of Binding Buffer was added to the same column. This was then centrifuged for 1 minute at 16000 g. The flowthrough was discarded,

and 700 µl SPW Wash Buffer was added to the column. This was centrifuged at 16000 g for one minute. The wash buffer was added once more and centrifuged for an additional washing step. The flowthrough was discarded. To remove all the ethanol left, the dry column was centrifuged at 16000 g for 2 minutes. The column was then placed in a 1.5 ml Eppendorf tube. To elute the DNA, 30-50 µl Elution Buffer was added to the column and incubated at room temperature for 2 minutes. Then a last centrifugation step at 16000 g for 1 minute was done, and DNA was in the filtrate. The concentration was measured using Nanodrop, and the tubes were then stored at 4 °C.

2.10 Purification of PCR Products

The Thermo Scientific GeneJET PCR Purification Kit was used the purify to DNA after ligation. First, one volume of Binding Buffer was added to the ligation mixture. The mixture was then mixed and added to a purification column. The binding buffer denaturates proteins and promotes the DNA to bind to the membrane of the column. After 1 minute of centrifugation, the flowthrough was discarded and 700 µl Wash Buffer was added. The column was then centrifuged for one more minute. In this step, all impurities are removed. The flowthrough was discarded and the column was centrifuged for one more minute to remove any remaining wash buffer. The column was then transferred to a 1.5 ml Eppendorf tube. 50 µl of Elution Buffer was then added to the column, and the column was centrifuged for 1 minute. Here, the purified DNA is eluted from the column. The concentration of the recovered DNA was measured using Nanodrop, and the tubes were stored at 4 °C.

2.11 Chemical Transformation

50 µl competent cells stored in a 1.5 ml tube were taken out of a -80 °C freezer and thawed on ice. Then approximately 100 ng DNA was added. The tubes were incubated on ice for 30 minutes. After incubation, the tubes were placed in a 42 °C water bath for 45 seconds. This step increases the permeability of the membrane, allowing the DNA to enter the cell. After heat shock the tubes were incubated on ice for 1 minute. Subsequently, 1 ml LB medium was added to the tubes and incubated at 37 °C with shaking for 1 hour. This step is important for the bacteria to gain antibiotic resistance. The tubes were then centrifuged at 6000 g. Most of

the supernatant was removed by decanting, and the cells were resuspended in the remaining LB medium. The cells were then plated on LB plates with neomycin preheated to 37 °C, and stored overnight.

2.12 Conjugation

Conjugation was used to transfer the constructed plasmids from *E. coli* S17-1 to *P. polymyxa*. Recipient (*P. polymyxa*) and donor (*E. coli* S17-1) was inoculated in non-selective and selective LB medium, respectively. *P. polymyxa* was incubated at 30 °C and *E. coli* was incubated at 37 °C. After overnight incubation, the cultures were diluted 1:100 and grown for four hours. Then, 900 µl of the recipient cells were heat shocked at 42 °C for approximately 15 minutes and mixed with 300 µl of the donor strain culture. Afterwards, the cells were centrifuged at 8000 g for 3 minutes to separate the cells from the LB medium. The medium was discarded and the cell pellet was resuspended and dropped on non-selective LB agar plates. The plates were incubated at 30 °C overnight. After incubation, the cells were scraped from the agar and resuspended in 500 µl 0.9 % NaCl. 200-300 µl of this was then plated on selective LB agar with both polymyxin and neomycin. The plates were incubated at 30 °C for 48 hours.

2.13 Plasmid Curing

Plasmid curing is the process of eliminating the plasmid from a cell. After the targeted genes had been deleted or replaced, the pCasPP plasmid could be removed. High-temperature growth was used for this. Cells were inoculated in 3 ml LB medium, and incubated overnight. Some tubes were incubated at 37 °C, and one tube was incubated at 50 °C. After incubation, 10 µl cell culture was transferred to a selective LB plate (LBneo), 10 µl was transferred to a plate with no antibiotics and 10 µl was transferred to a tube with 3 ml fresh LB medium. This procedure was repeated until there were no colonies on the plate with neomycin.

2.14 EPS Production and Purification

For EPS production, 250 ml baffled shake flasks were filled with 100 ml EPS medium and inoculated with 1 ml of an overnight culture of *P. polymyxa* grown in LB medium. The cultures were incubated at 30 °C with shaking at 170 rpm for 48 hours. After incubation, the cultures were diluted with 100 ml sterile water to decrease the viscosity and centrifuged at 30000 g room temperature for 30 minutes. This collects the cells in the bottom of the centrifugation flask, and the EPS can be extracted more easily from the supernatant. The supernatant was slowly poured into 400 ml ethanol to precipitate the EPS. The EPS was then collected using a spatula and a strainer, and transferred to a 50 ml falcon tube. The tubes were left open overnight to evaporate the remaining ethanol. The next day, the samples were shock-freezed with liquid nitrogen and placed in a freeze dryer overnight. The samples were then weighted.

2.15 Bioinformatics

2.15.1 BLAST

Basic Local Alignment Search Tool (BLAST) is a tool used to find local similarities between nucleotide or protein sequences [100]. A nucleotide or protein sequence is used as a query sequence and is searched for in the NCBI database [101]. There are many different BLAST tools depending on what kind of comparisons that are desired, but only BLASTn was used in this thesis. BLASTn uses a nucleotide query sequence and search in a nucleotide database [102]. For each alignment, BLAST gives five values that describes the quality of the match:

- Max Score
- Total Score
- Query Cover
- Expect value (E-Value)
- Percent Identity (Per. ident)

There may be several separate alignments in the sequences that match the query sequence. Therefore, both the max score and total score are given. Max score gives a value for the highest scoring alignment, while the total score gives a value for all found alignments. The query cover gives the percentage of how much of the query sequence that matches the found sequence. The E-value estimates the probability of an alignment to happen by chance. The lower this score is, the better. Per. ident is calculated by dividing the number of identical nucleotides by the number of nucleotides in the alignment [103].

In addition to the table with the values described above, BLAST also provides a graphical summary of the results. This gives a view of which parts of the query sequence that are present in the selected sequences. There are five different colors, where each color represent an alignment score. BLAST also shows the alignments of the subject sequences, where the whole query sequence can be compared to the subject sequence, and every mismatch is displayed. A distance tree of the results can also be made using BLAST, which is based on pairwise alignment. This can show the relationship between all related sequences that shows up in a BLAST search. When generating the tree, a maximum sequence distance can be chosen. This value is the maximum allowed fraction of bases that mismatches in the aligned part between any pair of the sequences.

2.15.2 RAST

RAST (Rapid Annotation using Subsystem Technology) is a server that annotates the genome of bacteria and archaea. The genomes can be uploaded to RAST in both FASTA and GenBank format. Two classes of asserted gene functions are being produced by RAST: subsystems-based assertions and nonsubsystem-based assertions. Subsystem-based assertions are based on the identification of functional variants of subsystems, while nonsubsystem-based assertions uses evidence from a number of other common tools. A subsystem, which is manually curated, can be defined as a set of abstract functional roles. This means that proteins with similar functions can be a part of the same subsystem. In addition, RAST uses FIGfams, which is a collection of protein families that are derived from the subsystem technology. The collection of proteins in each FIGfam are isofunctional homologs. This means that they probably have the same function and come from a common ancestor. RAST predicts tRNA, rRNA, protein-encoding genes and which subsystems that are present in the genome and assigns the function of genes [104].

2.15.3 BRIG

BLAST Ring Image Generator (BRIG) is a platform that can be used to show images of multiple genome comparisons. The comparison is shown in rings that compares a reference sequence to the uploaded sequences. The comparisons are shown in a sliding color scale depending on the chosen percent identities [105].

Chapter 3

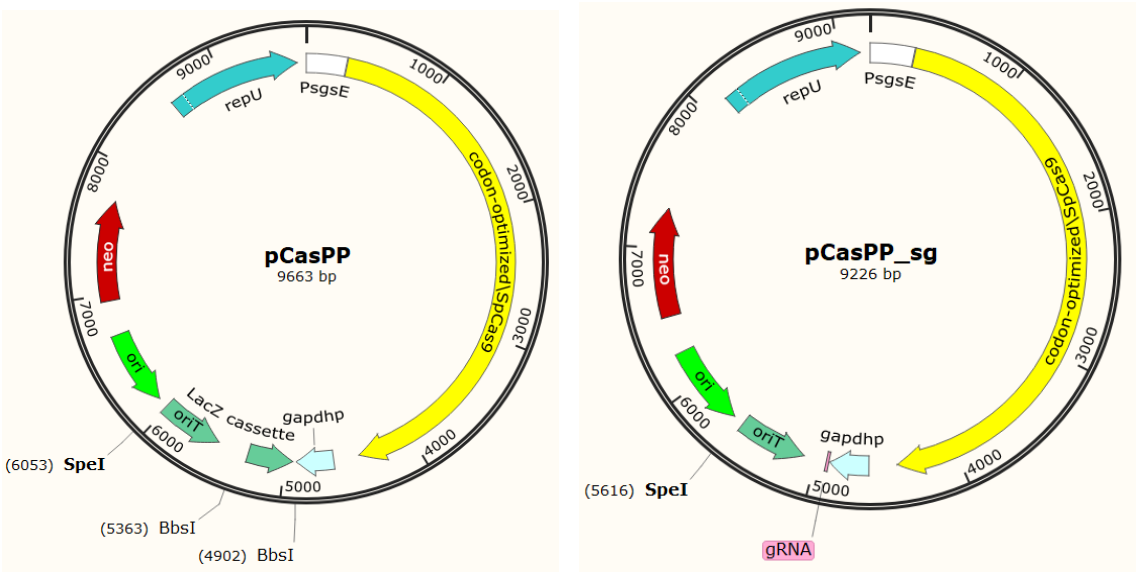
Results

Section 3.1 describes the results from the cloning procedure for construction of pCasPP_ΔpepM and pCasPP_ΔpepN. Section 3.2 shows the results for the conjugations of the plasmids used for gene deletions and gene replacement, while Section 3.3 describes the EPS produced by the deletion strains. The results from the genome comparison are shown in Section 3.4. The primers used in this project are shown in Table 1 in Appendix A.

3.1 Assembly of plasmids

The plasmids pCasPP_ΔpepM and pCasPP_ΔpepN were used to delete the *pepM* and *pepN* genes in the genome of *P. polymyxa*. This was done using the CRISPR-Cas9 based pCasPP system. First, Golden Gate cloning was used to insert the gRNA. Using Benchling, a 20 nucleotides long guide sequence was selected. The PAM sequence used was NGG where N is any nucleotide. Sticky ends were added to the guide sequence to give ends complementary to the pCasPP plasmid digested with the restriction enzyme BbsI. ACGC was added to the forward primer, while AAAC was added to the reverse primer. Primer 1 and 2 in Table 1 were used for the *pepM* guide sequence, while primer 7 and 8 were used for the *pepN* guide sequence.

After annealing the guide sequences, the pCasPP plasmid was digested with BbsI. This leads to the loss of a 461 bp long sequence, the lacZ casette. Figure 3.1 shows the difference between the plasmid before and after insertion of the gRNA.



(a) pCasPP **(b)** pCasPP_sg

Figure 3.1: Figure (a) shows the pCasPP plasmid without the gRNA. Figure (b) shows the plasmid with the gRNA inserted, which led to the loss of LacZ casette.

Colony PCR was conducted after the insertion of the guide sequence. Primer 7 and 8 in Table 1 were used. The results for the instertion of gRNA for *pepM* deletion are given in Figure 3.2.

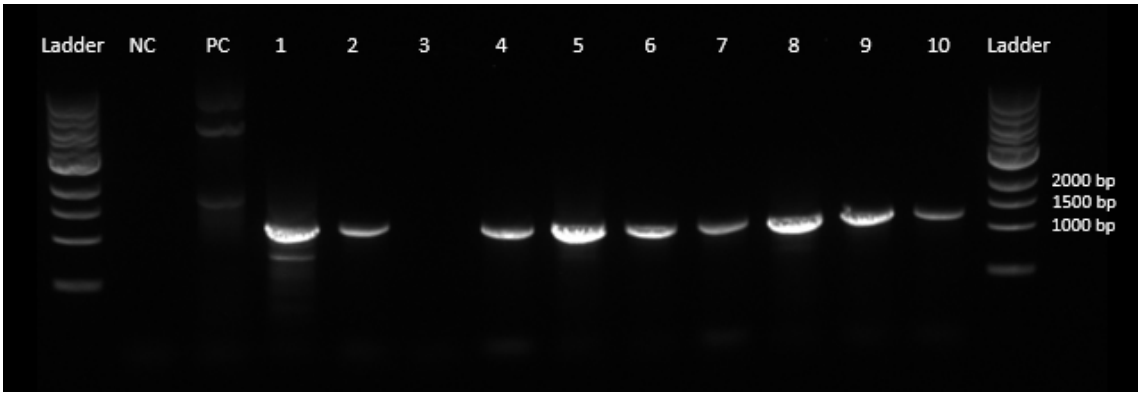


Figure 3.2: Colony PCR to verify insertion of spacer sequence for deletion of *pepM*. PC: positive control. NC: negative control. The positive control is an empty pCasPP plasmid, which should give a length of 1717 bp. Colony 1-10 should give the length 1280 bp if the gRNA has been inserted.

The positive control shows the empty pCasPP plasmid and gives a fragment with the length 1717 bp. In Figure 3.2, a weak band can be seen at approximately 1717 bp. Colony 1-2 and 4-10 shows a band that is smaller than the positive control. A band of 1280 bp is expected for the colonies with the gRNA inserted. This seems to be correct for all of the colonies except colony 3 that did not show any band. Colony 2 and 8 were inoculated for further cloning. The plasmid from colony 2 was sequenced, which verified that the gRNA had been inserted.

Primer 7 and 8 in Table 1 were used for verifying insertion of the *pepN* spacer sequence. The results from the gel electrophoresis are shown in Figure 3.3.

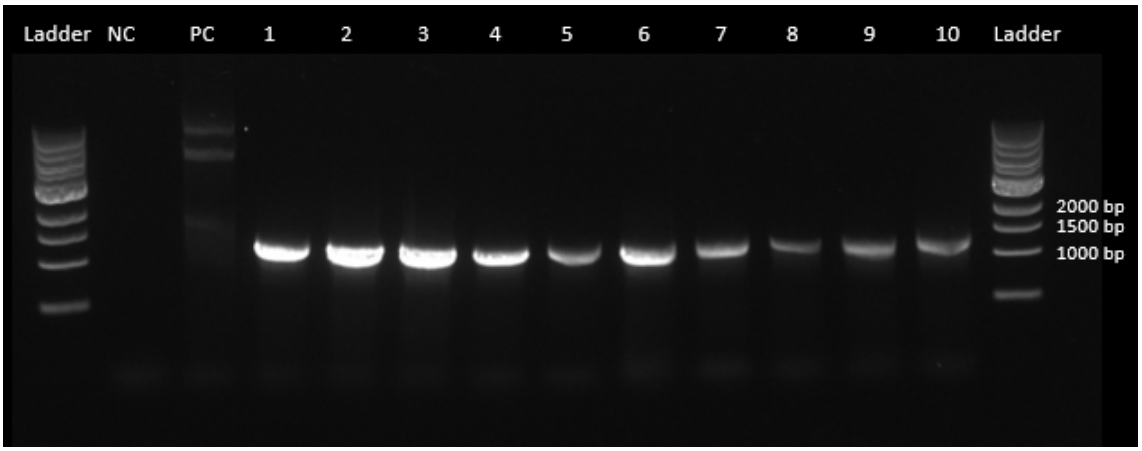


Figure 3.3: Colony PCR for verifying insertion of spacer sequence for deletion of *pepN*. PC: positive control. NC: negative control. The positive control is an empty pCasPP plasmid, which should give a length of 1717 bp. Colony 1-10 should give the length 1280 bp if the gRNA has been inserted.

A weak band can be seen for the positive control at the expected size of 1717 bp. For all colonies 1-10, the fragments are smaller than this and at the expected size of 1280 bp. This indicates that the guide sequence has been inserted. Colony 1 and 5 were picked. The insertion was verified by sequencing of colony 5.

A template for homology directed repair was constructed using overlap extension PCR. Approximately 1000 bp upstream (US) and 1000 bp downstream (DS) of the *pepM* gene was amplified using primers 3-6 in Table 1. For *pepN*, the primers 9-12 were used to amplify the homologous arms. The primers were designed with an overlapping region, which enables the fusion of the US and DS flanks. In addition, XbaI restriction cut sites were added to the ends of the fused fragment. These cut sites gives ends complementary to the pCasPP plasmid digested with SpeI. Hence,

this can be used to insert the fused fragment into the plasmid.

First, the US and DS fragments were amplified separately using the genome of *P. polymyxa* DSM 365 as template. The results from from the gel electrophoresis for the amplified homologous arms for *pepM* are shown in Figure 3.4. Two parallels of each flank were amplified. One of each of the fragments (US 1 and DS 1) gave the correct size at approximately 1000 bp as can be seen from the figure below.

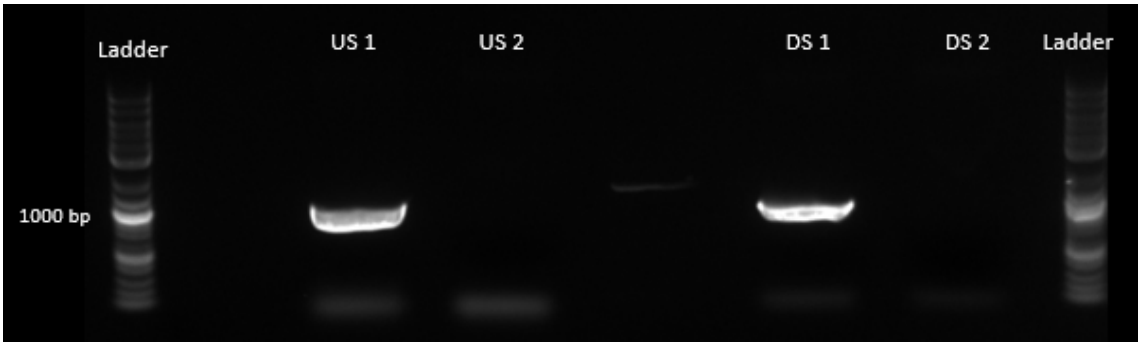


Figure 3.4: Gel electrophoresis of PCR product after amplifying the upstream and downstream fragments of *pepM*. The upstream (US) flank should give a length of 1029 bp. The downstream (DS) flank should give a length of 1051 bp.

The results after amplifying the US and DS fragments for *pepN* are shown in Figure 3.5.

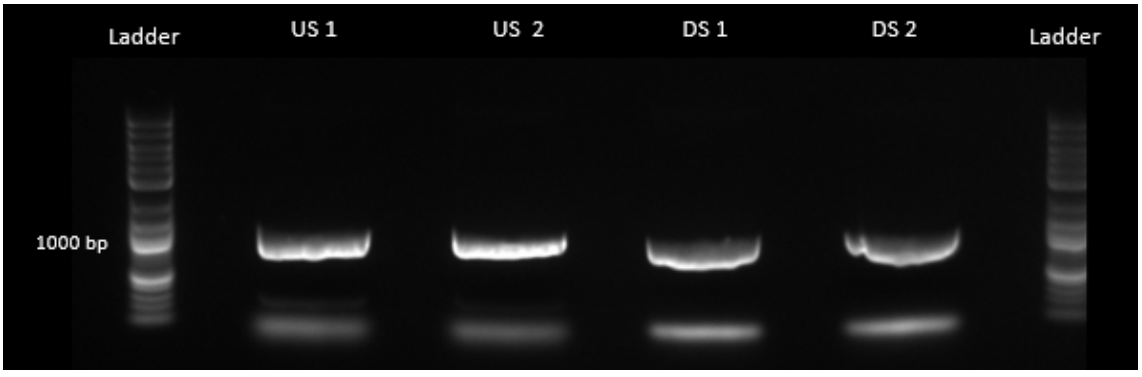


Figure 3.5: Gel electrophoresis of PCR product after amplifying the upstream and downstream fragments of *pepN*. The upstream (US) flank should give a length of 1046 bp. The downstream (DS) flank should give a length of 971 bp.

As can be seen from the figure, both parallels of both US and DS gave the expected fragment size of about 1000 bp. The DS flank (971 bp) is also a bit shorter than the US flank (1046 bp), which was expected.

The US and DS flanks were fused together using overlap extension PCR. Primer 2 and 6 in Table 1 for *pepM*, and primer 9 and 12 for *pepN* were used. The results from the gel electrophoresis after overlap extension PCR are shown in Figure 3.6

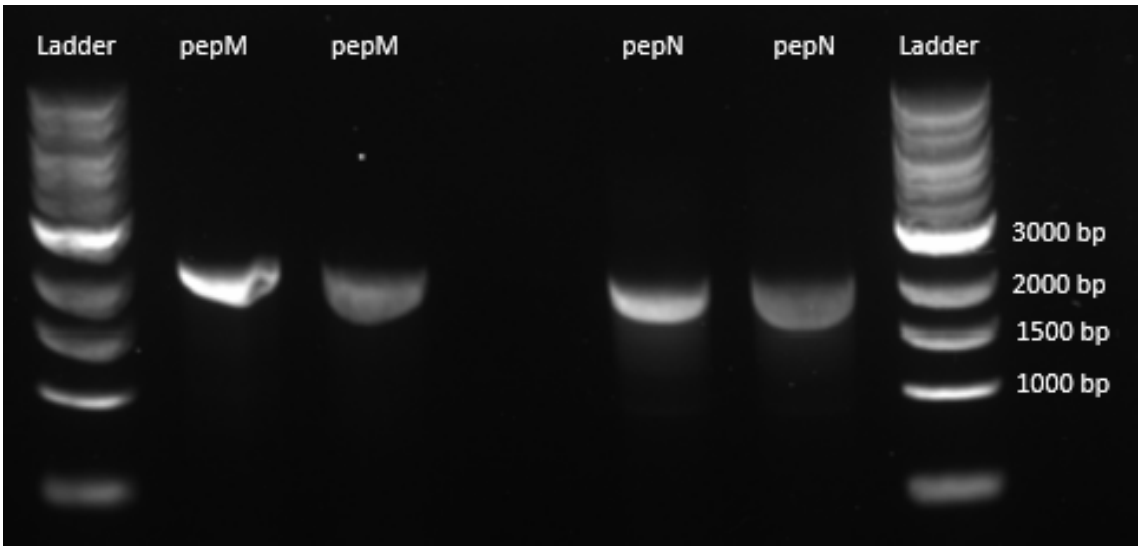


Figure 3.6: Gel electrophoresis of PCR product after overlap extension PCR The fused homologous arms of *pepM* gives a length of 2062 bp. For *pepN* the length is 1996 bp.

The figure above shows the expected fragment lengths, and it is therefore assumed that the OE-PCR was successful. The fused fragments were excised from the gel and purified. After digestion and ligation, the ligated plasmids were transferred into *E. coli* DH5α. The bacteria were plated on selective LB plates. Only a few colonies were obtained the next day, and gel electrophoresis was used to check if the correct plasmid was inserted. The same primers as for the fused homologous arms were used (primer 2 and 6 for *pepM* and primer 9 and 12 for *pepN*). It took several attempts to get a successful ligation and transformation.

Figure 3.7 shows the results of a gel electrohporesis of the pCasPP_ΔpepM plasmid where 5 of 6 colonies showed the correct fragment size. Colony 1 does not show any band, and the correct plasmid was therefore probably not inserted. Colony 2-6 shows a fragment of about 2000 bp as expected, and all fragments are also the same size as the positive control. The plasmids from colony 2-6 were extracted

and sent for sequencing, which verified that the insertion was correct.

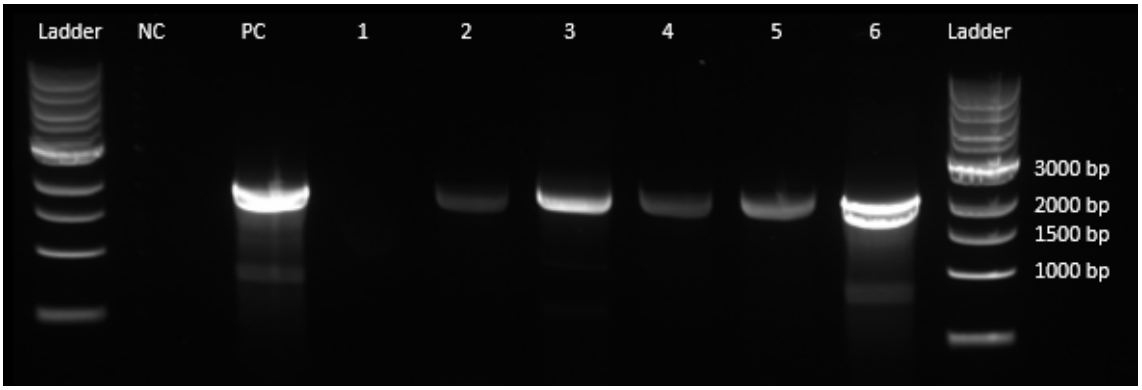


Figure 3.7: Gel electrophoresis of pCasPP_ΔpepM. The pCasPP_ΔpepM plasmid after ligations and transformation into *E. coli* DH5α was checked on the gel. The expected fragment size was 2062 bp. Colony 2-6 shows the expected fragment size.

Figure 3.8 shows a gel electrophoresis for the pCasPP_ΔpepN plasmid where colony 1 and 3 shows the correct fragment size. The expected fragment size was 1996 bp, but compared to the ladder, it looks like the fragments are closer to 1500 bp. However, fragment 1 and 3 are at the same size as the positive control. Since the positive control is the fused homologous arms, the insertion was therefore regarded as correct. This was also verified by sequencing.

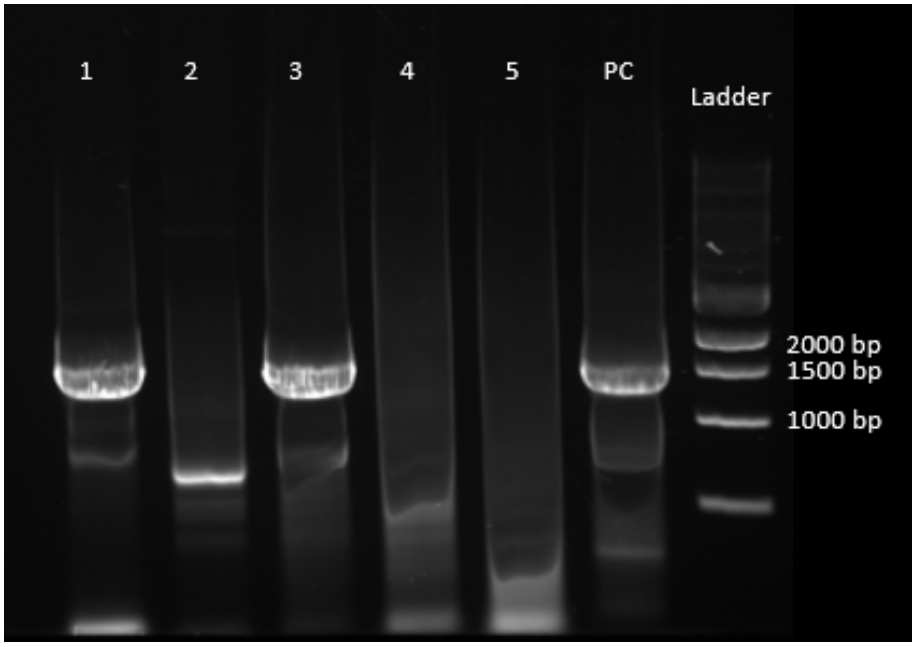


Figure 3.8: Gel electrophoresis of pCasPP_ΔpepN. The pCasPP_ΔpepN plasmid after ligations and transformation into *E. coli* DH5α was checked on the gel. The expected fragment size was 1996 bp. Colony 1 and 3 show the expected fragment size.

The plasmids were extracted from *E. coli* DH5α and transformed into *E. coli* S17-1 for conjugations.

3.2 Gene Deletions and Replacement

Replacement of *pepF*

The conjugation steps were followed to insert the pCasPP_rep_pepF plasmid into *P. polymyxa*. Two types of selective agar plates were used: LB agar with neomycin and polymyxin, and LB agar with neomycin and polymyxin and 60 mM $CaCl_2$ added. After the 48 hours incubation, nine colonies were obtained on the plates without $CaCl_2$, while the plates with $CaCl_2$ obtained a much higher number of colonies (>300 colonies). Nine colonies were picked from each plate and a colony PCR was conducted using primer 16 and 17 in Table 1.

The results from the gel electrophoresis showed no band from any of the colonies.

A band of 1203 bp was expected. The colony PCR was conducted before curing, and a band would be expected both with *gumK* as a part of the genome, and as a part of the inserted plasmid. The positive control showed the expected size of 1203 bp, and this indicated that the PCR reaction was successful. From this it looks like the plasmid had not been transferred to *P. polymyxa*. The conjugation was repeated several times, and did not show any positive result.

Deletion of *pepM*

For the conjugation for deletion of *pepM*, ten colonies were observed growing on plates without $CaCl_2$, while the plates with $CaCl_2$ seemed to be contaminated. A colony PCR was conducted for these ten colonies, and checked using gel electrophoresis. The primers 18 and 19 in Table 1 were used.

The predicted size was 3923 bp for cells without the deletion, and 2315 bp for colonies with the desired deletion. The results from the colony PCR can be seen in Figure 3.9

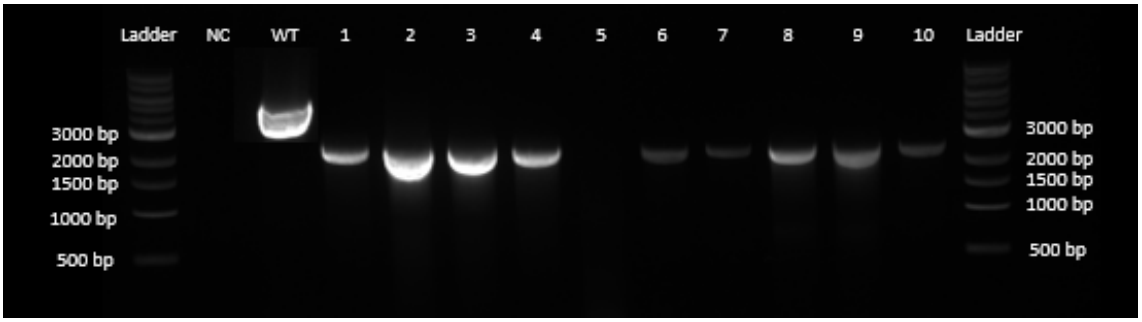


Figure 3.9: Colony PCR after insertion of pCasPP_ΔpepM. The expected band for a successful deletion is 2315 bp, while no deletion gives a band of 3923 bp. 9 out of the 10 colonies have the correct fragment size. PC: positive control. WT: wildtype.

As can be seen form the gel picture above, 9 of the 10 colonies show a band between 2000 bp and 3000 bp. All of these colonies therefore probably have the correct deletion, and the fragment size is then 2315 bp. Colony 5 does not show any bands. The wildtype *P. polymyxa* shows a band above 3000 bp, which is the expected size for no deletion at 3923 bp.

Deletion of *pepN*

After conjugation of pCasPP_ΔpepN, there were five colonies on the plates without $CaCl_2$ and a lot of growth on the plates with $CaCl_2$. Five colonies were picked from each plate, and a colony PCR was conducted using primer 20 and 21 in Table 1.

The results from the colony PCR is shown in Figure 3.10. Sample 1-5 show colonies picked from the plate without $CaCl_2$, while sample 6-10 are colonies picked from a plate with $CaCl_2$.

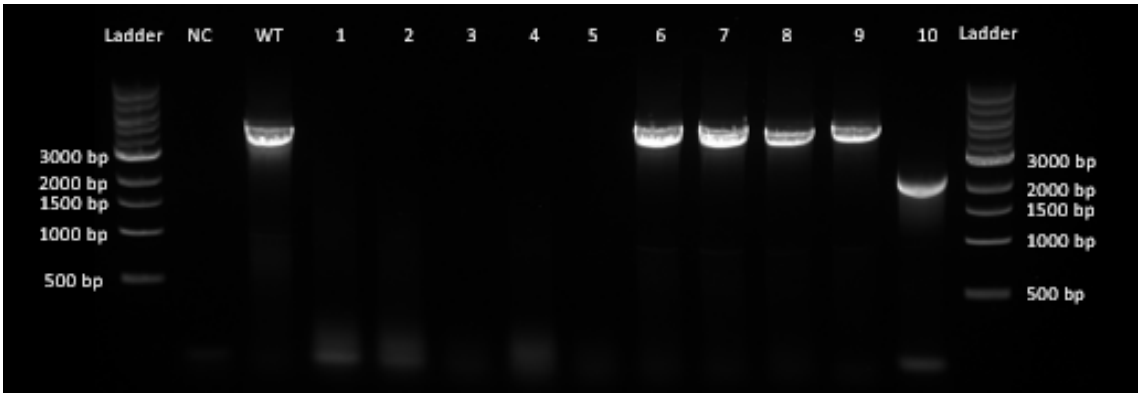


Figure 3.10: Colony PCR after insertion of pCasPP_ΔpepN. The expected band for a successful deletion of *pepN* is 2198 bp, while for the wild-type (WT) is 4496 bp. Colony 5-10 shows a band, but only colony shows the correct size.

The expected size for colonies with the correct deletion is 2198 bp, while colonies without the deletion will give a size of 4496 bp. The wildtype *P. polymyxa* shows a band above 3000 bp, which is most likely the expected band for no deletion of 4496 bp. The gel picture shows no bands for colony 1-5. Bands are visible for all colonies from the plates with $CaCl_2$, but only colony 10 shows a band around 2000 bp. It is therefore likely that this colony has the correct deletion.

3.3 EPS Production

The EPS production of the successful deletions of the *pepM* and *pepN* genes were analysed and compared to the wildtype. When precipitating the EPS, clear

differences could be seen between the three tested strains. Figure 3.11 shows the EPS from the three different strains in ethanol.

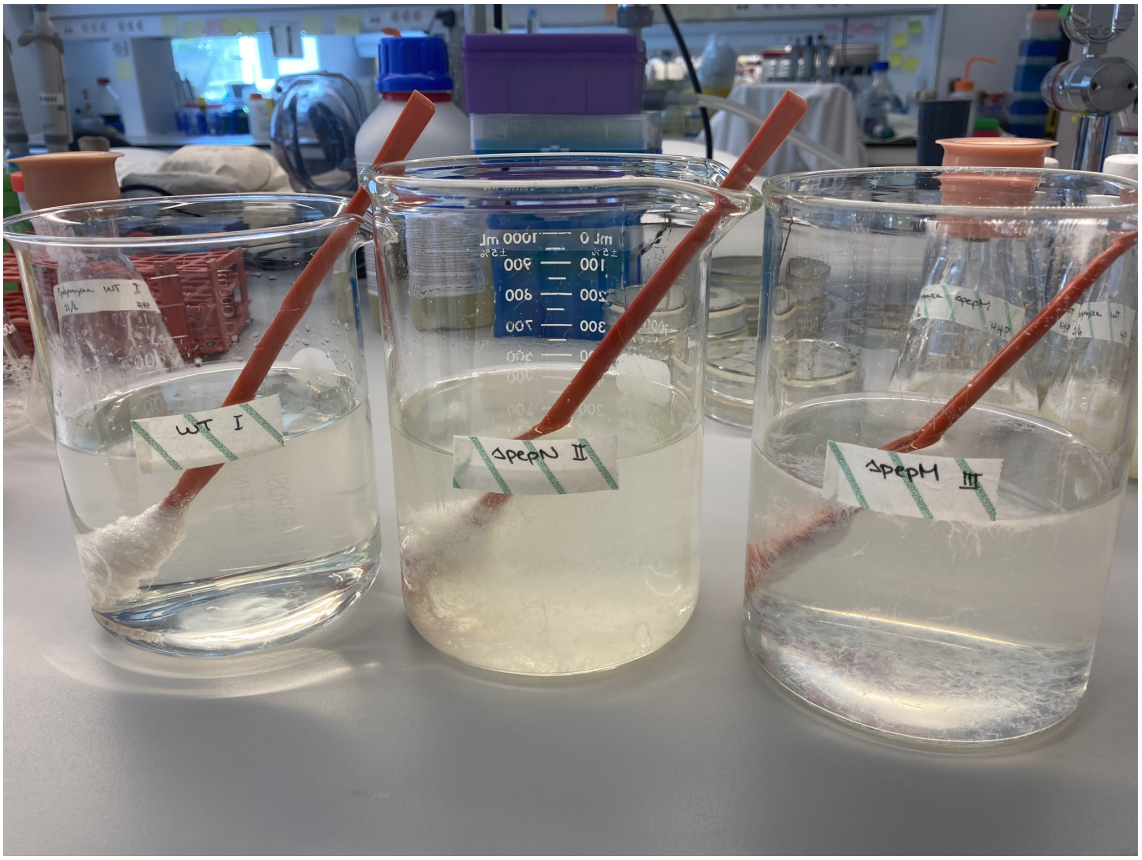


Figure 3.11: EPS produced from three different strains in ethanol. EPS precipitation with with 2 volumes of ethanol and 1 volume EPS. The beaker to the left shows EPS from *P. polymyxa* wildtype, *P. polymyxa*_ΔpepN in the middle and *P. polymyxa*_ΔpepM to the right.

The wildtype strain produced a fibrous precipitate, which could be easily collected with the spatula. The strain with the *pepN* deletion, on the other hand, produces short fibres that floats around in the ethanol. The *pepM* deletion strand also produces the shorter strands that could not be collected with the spatula. This strain also appeared to be less viscous in the fermentation broth compared to the other strains. After freezedrying, the EPS yield was analysed by gravimetrical measurement of the dried EPS. The results are shown in Table 3.1.

Table 3.1: Amount of EPS produced from the strains *P. polymyxa* WT, *P. polymyxa* _ΔpepN and *P. polymyxa*_ΔpepM.

Strain	**Weight [g]**
P. polymyxa WT	0.30
P. polymyxa ΔpepN	0.29
P. polymyxa ΔpepM	0.10

3.4 Genome comparison

3.4.1 BLAST

A BLASTn search was used in order to see in which *Paenibacillus* strains the EPS cluster of Paenan is present. Table 3.2 shows the top results from this search. All 21 strains show high similarity to the EPS cluster of *P. polymyxa* DSM 365 with a query cover of 99 % for all strains and an identity match between 89.52 % and 97.48 %. The E-value was 0.0 for all strains. The maximun score differs from 18746 to 49879, and the total score differs from 42568 to 61178.

When setting the maximum sequence distance to 0.2, the distance tree shown in Figure 3.12 was generated. The distance tree was used to select strains for more detailed comparison. Eighth strains were chosen from different branches of the distance tree to be compared to *P. polymyxa* DSM 365.

Table 3.2: BLASTn results. The BLASTn results of the Paenan cluster of *P. polymyxa* DSM 365. Max score gives a value for the highest scoring alignment. The total score gives a value for all found alignments. The query cover gives how much of the query sequence that matches the found sequence. The E-value estimates the probability of an alignment to happen by chance. Per.ident is calculated by dividing the number of identical nucleotides by the number of nucleotides in the alignment.

Strain	**Max Score**	**Total Score**	**Query Cover [%]**	**E value**	**Per. Ident [%]**
Paenibacillus polymyxa E681	49879	60168	99	0.0	96.86
Paenibacillus polymyxa J	46412	61178	99	0.0	97.48
Paenibacillus polymyxa Sb3-1	38985	49947	99	0.0	92.56
Paenibacillus sp. Izh-N1	38941	50100	99	0.0	92.53
Paenibacillus polymyxa SC2	38882	49957	99	0.0	92.49
Paenibacillus polymyxa M1	38882	49957	99	0.0	92.49
Paenibacillus polymyxa YC0136	26679	60891	99	0.0	97.18
Paenibacillus polymyxa CR1	26526	60219	99	0.0	96.99
Paenibacillus polymyxa ATCC 15970	26509	60323	99	0.0	96.97
Paenibacillus peoriae HS311	26354	60537	99	0.0	96.80
Paenibacillus polymyxa YC0573	26332	60620	99	0.0	96.75
Paenibacillus polymyxa ZF129	23261	50267	99	0.0	93.24
Paenibacillus polymyxa CF05	23224	50111	99	0.0	93.21
Paenibacillus polymyxa SQR-21	23219	50105	99	0.0	93.21
Paenibacillus sp. M-152	23156	50086	99	0.0	93.14
Paenibacillus polymyxa HY96-2	23095	50055	99	0.0	93.05
Paenibacillus polymyxa ZF197	22521	48732	99	0.0	92.39
Paenibacillus kribbensis AM49	20447	44359	99	0.0	90.04
Paenibacillus brasilensis KACC 13842	20197	46874	99	0.0	89.78
Paenibacillus terrae HPL.003	19293	43097	99	0.0	89.79
Paenibacillus sp. IHB B 3084	18746	42568	99	0.0	89.52

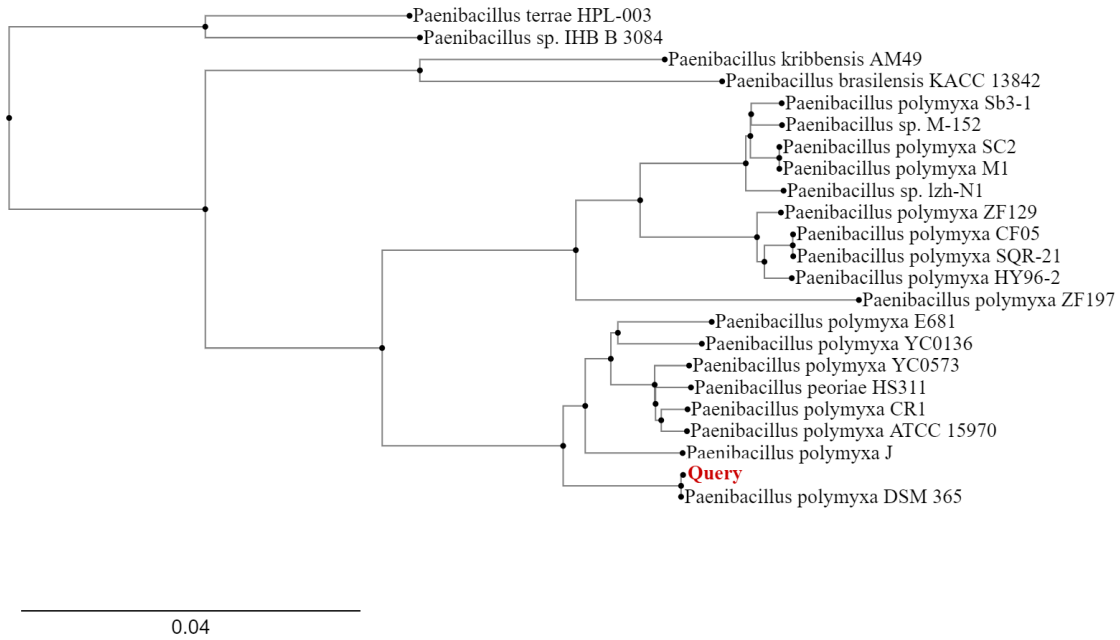


Figure 3.12: Distance tree. The distance tree shows the relationship between the results of a BLAST search, and is based on pairwise alignment between the query sequence and the subject sequence. The query sequence is marked in red. The maximum sequence distance in this case was set to 0.2, which is the maximum allowed fraction of bases that mismatches in the aligned part between any pair of sequences.

The eight chosen strains are shown in the graphical summary of the Paenan cluster in Figure 3.13. The red lines are the parts that are conserved in the different strains. The black and white parts are gaps. As can be seen from the figure, most of the cluster is conserved. However, some gaps can bee seen. Most of the strains show a gap around 27000 bp and 30000 bp. Four of the strains also show a gap around 15000 bp. These eight strains were studied further using BRIG.

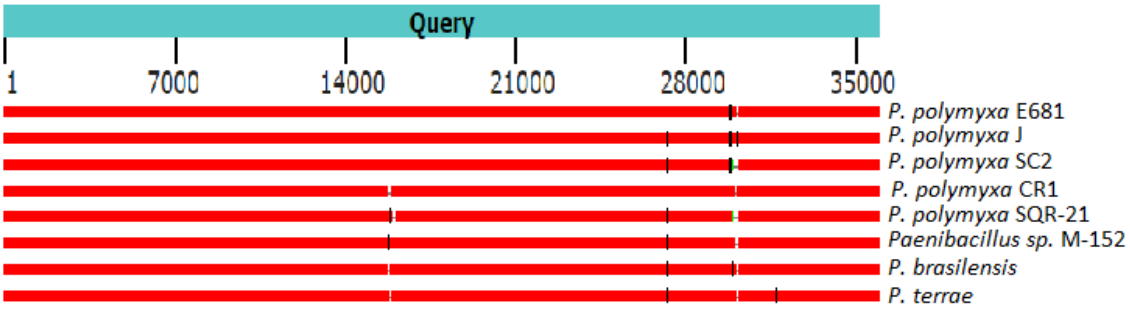


Figure 3.13: Graphical summary. This gives a view of which parts of the query sequence that are present in the selected sequences. The red parts are the parts with a high alignment score. The white and black parts have low alignment scores.

3.4.2 BRIG

BRIG was used to study the homology between the eight strains and *P. polymyxa* DSM 365 which was used as the reference sequence. Figure 3.14 shows the genome comparison of the whole genome with 70 % as the upper identity threshold and 50 % as the lower identity threshold. The figure shows high similarity between the genomes of the different strains. Most parts of the ring has a strong color, showing that the identity match is 70 % or more. Some white gaps can be seen in the ring.

A comparison of the Paenan cluster was also generated. This shows how identical the different genes are, and which genes that are conserved the most. First, a ring with upper threshold at 90 % and lower threshold at 70 % was generated. This is shown in Figure 3.15.

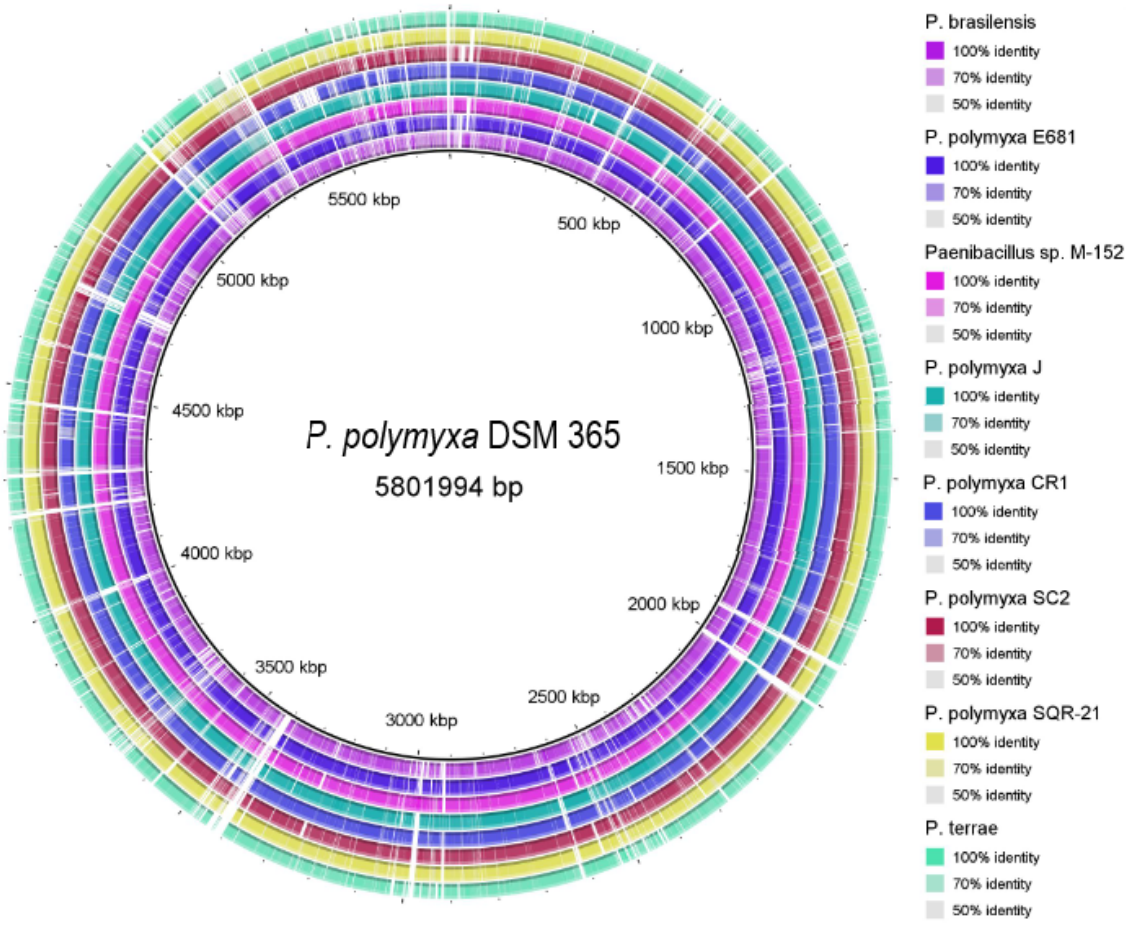


Figure 3.14: BRIG image of genome comparison. The genome of *P. polymyxa* DSM 365 was used as a reference. Upper identity threshold was set to 70 %. Lower identity threshold was set to 50 %

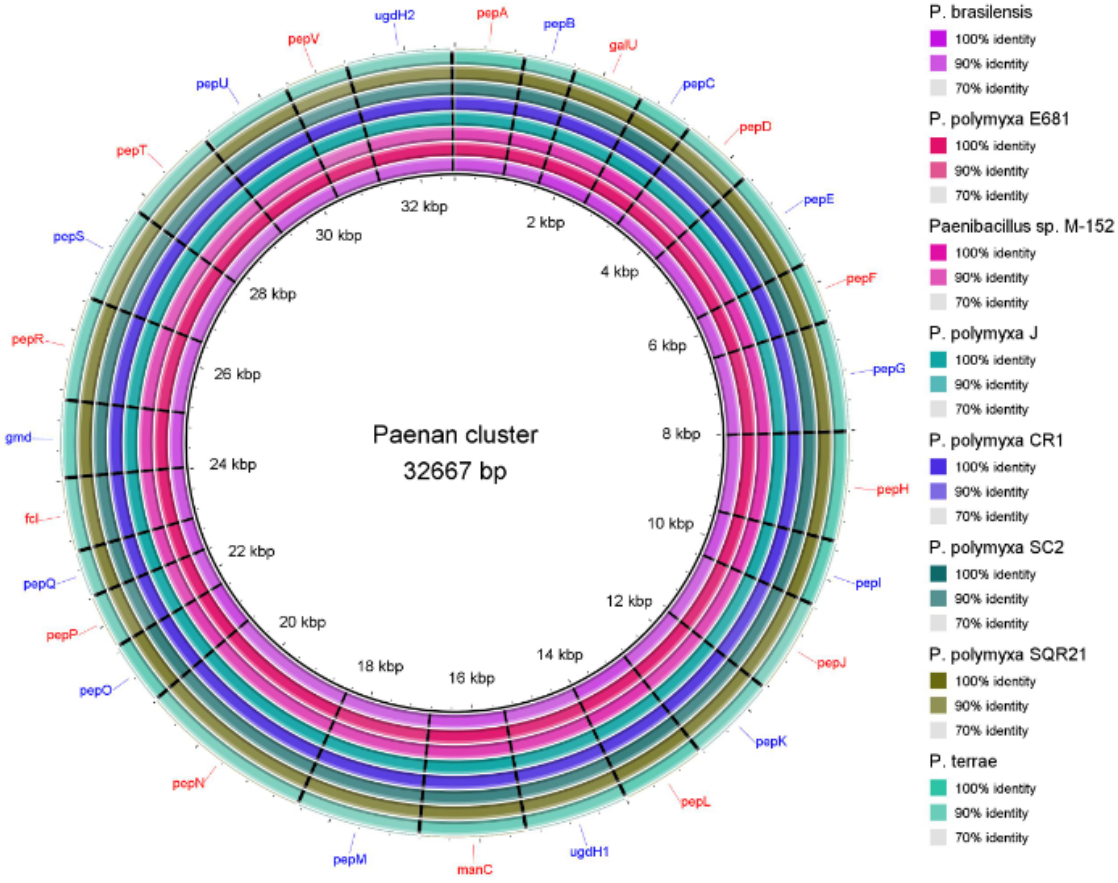


Figure 3.15: BRIG image of Paenan cluster. The upper threshold value was set to 90 % and the lower threshold value was set to 70 %.

The generated ring show strong colors, and shows an identity match of at least 70 % for all genes. By that, the Paenan cluster seems to be highly conserved in all strains. To see if there are any differences between the genes and the different strains, a ring with a higher threshold was created. This ring is shown in Figure 3.16.

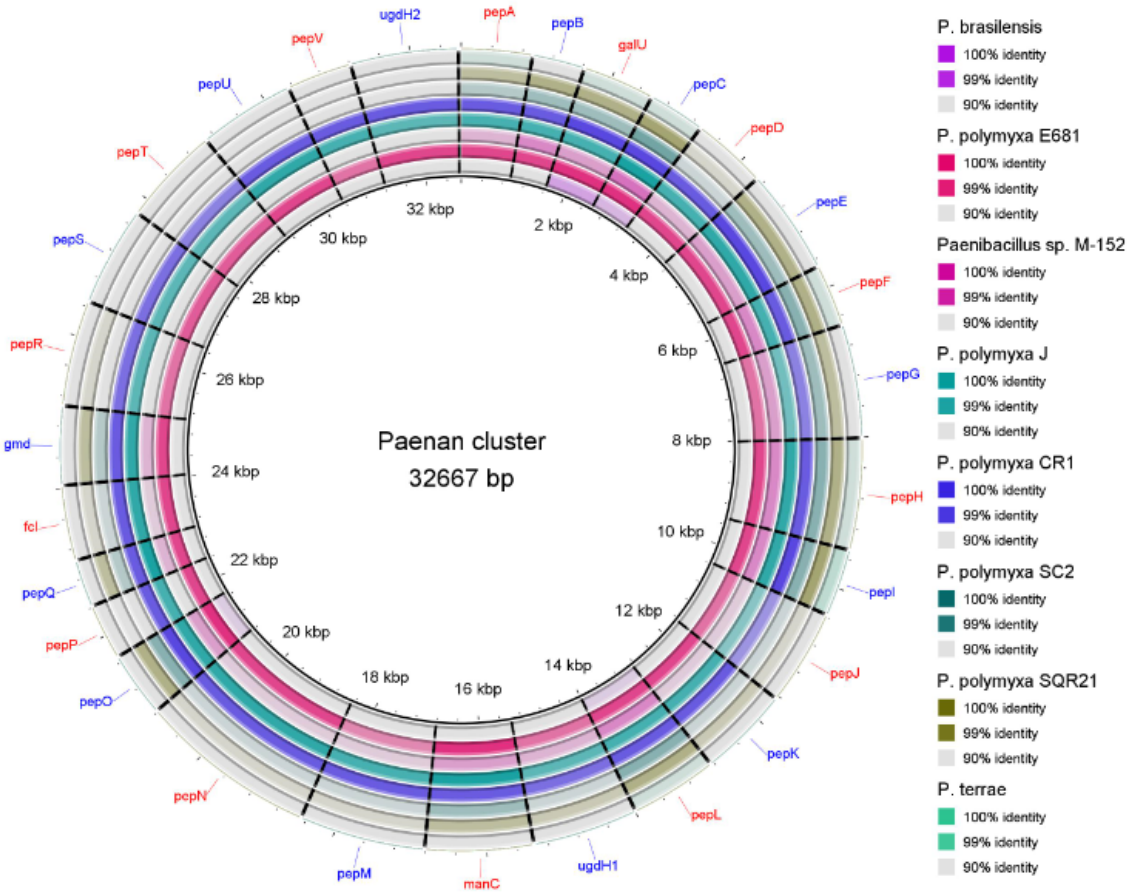


Figure 3.16: BRIG image of Paenan cluster with higher identity thresholds. The upper identity threshold value was set to 99 % and the lower identity threshold value was set to 90 %.

The figure above better highlights the differences between the strains and the genes. Three of the strains seems to have the whole cluster very well conserved with a strong color in all of the genes. This is for the strains *P. polymyxa* E681, *P. polymyxa* J and *P. polymyxa* CR1. These three strains has an identity match of at least 90 % for all genes. The rest of the strains do not have the same similarity between all of the genes. The first part of the cluster shows a higher identity match than the last part, and the six last genes have only 90 % identity in the three strains mentioned above. The *galU* and *pepC* genes are the only two genes that have above 90 % identity match for all strains.

P. brasilensis has over 90 % identity match for only the two genes *galU* and *pepC*. To study the possibility for this strain to produce Paenan, the genome was

annotated in RAST. This was used to check the function of the genes located downstream of *galU* and *pepC*. In the genome of *P. brasilensis*, the aligned cluster is in the range 3531765 to 3566155. Table 3.3 shows the presumed function of the coding sequences, the percentage of bases that matches, and to what gene in the Paenan cluster it matches.

All genes have a high similarity to the genes in the Paenan cluster. The coding sequence that is aligned to *pepT* has the lowest similarity with 86.3 %. The table also shows that there are many glycosyl transferases, and the putative function is similar to most of the genes in the Paenan cluster.

Table 3.3: Putative function of Paenan cluster in *P. brasilensis*. The percentage alignment shows the percentage of nucleotides aligned compared to the total number of nucleotides in the alignment. The Gene in the Paenan cluster column shows the gene it was aligned to in the Paenan cluster.

Function	**Percentage alignment**	**Gene in Paenan cluster**
Tyrosine-protein kinase transmembrane modulator EpsC	90.3	*pepA*
Tyrosine-protein kinase EpsD	89.0	*pepB*
UTP-glucose-1-phosphate uridylyltransferase	93.0	*galU*
Undecaprenyl-phosphate galactosephosphotransferase	92.3	*pepC*
Dolichol-phosphate mannosyltransferase	90.5	*pepD*
Polysaccharide polymerase	90.7	*pepE*
N-acetylmannosaminyltransferase	89.2	*pepF*
Glucose-6-phosphate isomerase	88.7	*pepG*
Polysaccharide biosynthesis protein	88.1	*pepH*
Glycosyl transferase	91.0	*pepI*
Glycosyl transferase	86.9	*pepJ*
Glycosyl transferase	89.3	*pepK*
Glycosyl transferase	91.1	*pepL*
UDP-glucose dehydrogenase	88.0	*ugdH1*
Mannose-1-phosphate guanylyltransferase	90.4	*manC*
Glucosesorbosone dehydrogenases	88.9	*pepM*
Hypothetical protein	88.0	*pepN*
Tyrosine-protein kinase transmembrane modulator EpsC	91.3	*pepO*
Tyrosine-protein kinase EpsD	89.3	*pepP*
Undecaprenyl-phosphate galactosephosphotransferase	90.0	*pepQ*
GDP-L-fucose synthetase	89.2	*fcl*
GDP-mannose 4,6-dehydratase	90.4	*gmd*
Membrane protein	88.3	*pepR*
Hypothetical protein	88.4	*pepS*
Glycosyl transferase	86.3	*pepT*
Glycosyl transferase	88.5	*pepU*
N-acetylmannosaminyltransferase	87.5	*pepV*
UDP-glucose dehydrogenase	88.1	*ugdH2*

Chapter 4

Discussion

4.1 Deletion of *pepM* and *pepN*

As seen in Section 3.2, the deletions of *pepM* and *pepN* in *P. polymyxa* were successful. Compared to the wildtype, both *P. polymyxa* Δ*pepM* and *P. polymyxa* Δ*pepN* produced EPS with shorter fibres. Since hydrolases break the chemical bonds of polysaccharides, one could expect longer fibres when hydrolases are deleted. From this, one could argue that *pepM* and *pepN* do not encode hydrolyses the cleave the chemical bond of Paenan. As mentioned in Section 1.5, glycosyl hydrolases have been found to be essential to the production of some EPSs. The production of cellulose is one example that uses hydrolases. The hydrolases involved in cellulose production are believed to have a role in the regulation of the material for the substrate [88]. If this is the role for the *pepM* and *pepN* genes, it might explain why the produced fibres are shorter. A limited supply of substrates may also limit the production of Paenan.

Another study deleted the *pslG* gene which is a hydrolase involved in the production of Psl [89]. This led to less production of Psl, a decrease which was also observed for the production of EPS from the Δ*pepM* strain. The glycosyl hydrolases in the Paenan cluster may therefore have the same role as PslG in the Psl production. One putative function of PslG is, as mentioned in Section 1.5, to be responsible for processing the polysaccharide for export. This may explain why the deletion of *pepM* led to only one-third of the EPS produced compared to the wildtype. Another putative role of the PslG is to act as a failsafe enzyme when the export fails. It does this by clearing the periplasm of abnormal polysaccharides.

This may explain why many short fibres were produced. If one of the glycosyl hydrolases is responsible for the processing of the EPS, then a deletion of this hydrolase will result in a not completely processed polysaccharide. The other hydrolase may then have a role of clearing the periplasm of abnormal polysaccharides through export, which may lead to shorter fibres being exported. One could then argue that *pepM* is responsible for clearing the periplasm, since the deletion of this gene led to less EPS being exported from the cell. The *pepN* gene could have a role in the processing of EPS because the deletion of this gene resulted in shorter fibres, but it produced more EPS than the Δ*pepM* strain. In the Δ*pepN* strain, the *pepM* gene was still functional, and therefore it might be able to export the unprocessed polymer.

Another enzyme that is involved in the synthesis of EPS is AlgL, which is important for the synthesis of alginate. This is not a hydrolase, but a lyase. Lyases are also enzymes that degrade polymers. A hypothetical function of this enzyme is to handle errors during the biosynthesis. It is believed that the enzyme is responsible for clearing the periplasm of alginates that are not exported. When this gene is deleted, it leads to an accumulation of alginate in the periplasmic space. This seemed to be toxic to the cell at high concentrations of alginate [106]. The deletion of *pepM* and *pepN* did not seem to affect the growth of *P. polymyxa*, and this therefore contradicts the role in clearing the periplasmic space. However, the Paenan cluster has two genes that encodes glycosyl hydrolases. This may be a safety if one of the genes is disrupted. Since the deletion of *pepM* resulted in less EPS produced, this gene seems to be more important for the EPS production than *pepN*.

4.2 Gene replacement of *pepF*

Another aim of this thesis was to replace the *pepF* gene with the *gumK* gene. Section 3.2 shows that this was unsuccessful. Since the plasmid was able to replicate in *E. coli* DH5α, it is unlikely that the error lies with the assembled plasmid. In addition, *E. coli* S17-1 was able to grow on selective LB plates after transformation, which indicates that the plasmid could replicate in this strain as well. Plasmids of large sizes can give problems with conjugations, but this should not be a problem with a plasmid of size 12310 bp. This is of similar size as the plasmids used for the deletions. Even if the plasmid did not seem to be inserted, colonies were obtained on the selective plates. Since *P. polymyxa* is not resistant to neomycin, there should not have been colonies if the transfer was unsuccessful. More colonies were obtained on the plates with $CaCl_2$. $CaCl_2$ had previously shown to increase the

number of colonies obtained after conjugation. However, A. K. Haraldsvik found out that this also seems to reduce the recombination efficiency [107]. The $CaCl_2$ may inhibit the antibiotics, and bacteria without the plasmid may thereby be able to grow on the plates. Therefore a potential issue is contamination on the plates. Another aspect to consider is that a heterologous gene has never been integrated into the cluster. Some regions are easier to modify than others, and a heterologous gene may not be easily inserted into the Paenan cluster.

4.3 Genome comparison

The genome and Paenan gene cluster were also studied using different bioinformatic tools. Figure 3.14 shows that the genome is well conserved through the different *Paenibacillus* strains. Only a few gaps were found in the genome. The Paenan cluster was also well conserved. The graphical summary in Figure 3.13 shows some gaps in the cluster for several strains. However, Figure 3.15 shows that the whole cluster is highly conserved. This may indicate that the gaps observed in the graphical summary are not in the genes, but between them. Figure 3.16 shows that *P. brasilensis* was the strain of those investigated with the lowest similarity to the *P. polymyxa* DSM 365 Paenan cluster. A RAST annotation was conducted to analyze the differences in more detail. By aligning the putative genes to the Paenan cluster in SnapGene, it was observed that the genes in grey color in the BRIG image were also highly conserved. The annotation showed that the aligned genes served similar functions as the genes in the Paenan cluster. It is therefore likely that these strains are able to produce an EPS similar to Paenan.

Table 1.2 shows different EPS produced by *Paenibacillus* strains. As can be seen, the nitrogen and carbon sources vary a lot between the different characterized EPSs. The structure of EPSs is highly dependent on the growth conditions, and since the different characterized EPSs were produced under different growth conditions, it might be hard to compare the EPSs between the different strains. For example, *P. polymyxa* SQR-21, which showed high similarity of the Paenan cluster, is known to produce a polymer of the same monomers as Paenan. However, the ratio of these monomers are different compared to Paenan. This therefore makes it difficult to analyze if this strain is able to produce Paenan under the same growth conditions as *P. polymyxa* DSM 365. In addition the results might highly depend on the analytics applied, and the HT-PMP method for the analysis of Paenan might be more sensitive and correct then the techniques used in the other studies. By that, a direct comparison is hampered by various aspects and should be performed under controlled and comparable conditions.

4.4 Further work

To get a better understanding on how the glycosyl hydrolases affect the production of Paenan, there are several aspects that should be investigated further. The produced Paenan from the two deletion-strains in this thesis should be analyzed in more detail to find out the monomer composition, molecular weight and the rheological behavior. Both *pepM* and *pepN* should also be deleted in one strain. This will give more information about the roles of the hydrolases. Another aspect of interest is to investigate what happens to the EPS production when the hydrolases are overexpressed. The Psl production has been reported to decrease when the hydrolase PslG was overexpressed [89]. It is of interest to see if the same result applies to the Paenan production.

The reason why the insertion of the *gumK* gene into *P. polymyxa* was not succesful should also be further investigated. The conjugation protocol may be modified to increase the conjugation efficiency. Other genes in the Paenan cluster may also be of interest to target, along with insertion of other genes than *gumK*. This might also give useful insight about the problems of the gene replacement targeted in this thesis. If the conjugation is successful with another plasmid assembled for another gene replacement, then this may indicate that the constructed plasmid is the problem of the *pepF* replacement. If it is not successful, the problem might be that the strain is receptive for gene replacements in this cluster.

The results from the genome comparison should also be further investigated. It would be interesting to check what kind of genes are in the gaps as identified by Figure 3.14. Even though the Paenan cluster showed a high similarity between the different strains chosen in this thesis, a multiple alignment of the amino acid sequence can be conducted to see where the differences are. One could also investigate the EPS production of these strains in the lab and compare the monomer composition of the EPS produced by the different strains. It is then important to grow all strains under the same conditions. This might give useful information about which parts of the gene sequences that are important for the Paenan production.

Chapter 5

Conclusion

Two strains of *P. polymyxa* were constructed during this project by deleting the two genes *pepM* and *pepN*, which encode glycosyl hydrolases. This was obtained using the CRISPR-Cas9 system pCasPP. A gRNA and a homologous directed repair template were inserted into the plasmid. Compared to the wildtype, both deletion-strains produced shorter EPS fibres. The function of the glycosyl hydrolases in EPS production is not fully understood, and needs to be further studied. Possible explanations for the short fibres may be the that the glycosyl hydrolases have roles in processing the polysaccharide for export, or to clear the periplasm for abnormal polysaccharides when the export fails.

The *pepF* gene, which encodes a glycosyl transferase, was also targeted in this project. The aim was to replace this gene with the *gumK* gene which encodes another glycosyl transferase. This replacement was not successful. The reason for this is not known, and needs to be further investigated.

The tools BLAST, RAST and BRIG were used for genome comparison analysis. The Paenan cluster was found in other *Paenibacillus* strains with high similarity, and it is therefore likely that Paenan can be produced by these strains.

Bibliography

[1] David L Nelson, Michael M Cox, and Albert L Lehninger. *Principles of biochemistry*. Freeman New York, 2017.

[2] Christoph Schilling, Abinaya Badri, Volker Sieber, Mattheos Koffas, and Jochen Schmid. Metabolic engineering for production and functional polysaccharides. *Current Opinion in Biotechnology*, 2020.

[3] Ute Römling and Michael Y Galperin. Bacterial cellulose biosynthesis: diversity of operons, subunits, products, and functions. *Trends in microbiology*, 23(9):545–557, 2015.

[4] Mona Moniri, Amin Boroumand Moghaddam, Susan Azizi, Raha Abdul Rahim, Arbakariya Bin Ariff, Wan Zuhainis Saad, Mohammad Navaderi, and Rosfarizan Mohamad. Production and status of bacterial cellulose in biomedical engineering. *Nanomaterials*, 7(9):257, 2017.

[5] Bjørn E. Christensen. *Biopolymers*. NTNU Trondheim, 2018.

[6] Tesfaye Abebe Desta and Yalfal Temesgen Tigabu. Starch production, consumption, challenges and investment potentials in ethiopia: The case of potato starch. 2015.

[7] R Malcolm Brown Jr, Inder M Saxena, and Krystyna Kudlicka. Cellulose biosynthesis in higher plants. *Trends in plant science*, 1(5):149–156, 1996.

[8] Woo Dae Jang, Ji Hyeon Hwang, Hyun Uk Kim, Jae Yong Ryu, and Sang Yup Lee. Bacterial cellulose as an example product for sustainable production and consumption. *Microbial biotechnology*, 10(5):1181–1185, 2017.

[9] Robert J Seviour, Brian McNeil, Mariana L Fazenda, and Linda M Harvey. Operating bioreactors for microbial exopolysaccharide production. *Critical reviews in biotechnology*, 31(2):170–185, 2011.

[10] Filomena Freitas, Cristiana AV Torres, and Maria AM Reis. Engineering aspects of microbial exopolysaccharide production. *Bioresource technology*, 245:1674–1683, 2017.

[11] MD Madigan, JM Martinko, KS Bender, DH Buckley, and DA Stahl. Brock biology of microorganisms 14th ed., glenview, il, 2014.

[12] Volker Sieber, Jochen Schmid, and Gerd Hublik. Microbial polysaccharides. *Industrial Microbiology*, 2019.

[13] Mayara CS Barcelos, Kele AC Vespermann, Franciele M Pelissari, and Gustavo Molina. Current status of biotechnological production and applications of microbial exopolysaccharides. *Critical Reviews in Food Science and Nutrition*, 60(9):1475–1495, 2020.

[14] Chris Whitfield and Miguel A Valvano. Biosynthesis and expression of cell-surface polysaccharides in gram-negative bacteria. In *Advances in microbial physiology*, volume 35, pages 135–246. Elsevier, 1993.

[15] JF Wilkinson. The extracellular polysaccharides of bacteria. *Bacteriological Reviews*, 22(1):46, 1958.

[16] Wayne A Wilson, Peter J Roach, Manuel Montero, Edurne Baroja-Fernández, Francisco José Muñoz, Gustavo Eydallin, Alejandro M Viale, and Javier Pozueta-Romero. Regulation of glycogen metabolism in yeast and bacteria. *FEMS microbiology reviews*, 34(6):952–985, 2010.

[17] Ian W Sutherland. Structure-function relationships in microbial exopolysaccharides. *Biotechnology advances*, 12(2):393–448, 1994.

[18] Monissa Paderes, Deepak Ahirwal, and Susana Fernández Prieto. Natural and synthetic polymers in fabric and home care applications. *Physical Sciences Reviews*, 2(9), 2017.

[19] F Garcıa-Ochoa, VE Santos, JA Casas, and E Gómez. Xanthan gum: production, recovery, and properties. *Biotechnology advances*, 18(7):549–579, 2000.

[20] Anke Becker. Challenges and perspectives in combinatorial assembly of novel exopolysaccharide biosynthesis pathways. *Frontiers in microbiology*, 6:687, 2015.

[21] Jochen Schmid, Volker Sieber, and Bernd Rehm. Bacterial exopolysaccharides: biosynthesis pathways and engineering strategies. *Frontiers in microbiology*, 6:496, 2015.

[22] Leslie Cuthbertson, Iain L Mainprize, James H Naismith, and Chris Whitfield. Pivotal roles of the outer membrane polysaccharide export and polysaccharide copolymerase protein families in export of extracellular polysaccharides in gram-negative bacteria. *Microbiol. Mol. Biol. Rev.*, 73(1):155–177, 2009.

[23] JC Whitney and PL Howell. Synthase-dependent exopolysaccharide secretion in Gram-negative bacteria. *Trends in microbiology*, 21(2):63–72, 2013.

[24] Bernd HA Rehm. Bacterial polymers: biosynthesis, modifications and applications. *Nature Reviews Microbiology*, 8(8):578–592, 2010.

[25] Rapala Srikanth, Chinta HSS Sundhar Reddy, Gudimalla Siddartha, M Janaki Ramaiah, and Kiran Babu Uppuluri. Review on production, characterization and applications of microbial levan. *Carbohydrate polymers*, 120:102–114, 2015.

[26] Guoyu Meng and Klaus Fütterer. Structural framework of fructosyl transfer in *Bacillus subtilis* levansucrase. *Nature Structural & Molecular Biology*, 10(11):935–941, 2003.

[27] TURLOUGH M Finan, BARBARA Kunkel, GUIDO F De Vos, and ETHAN R Signer. Second symbiotic megaplasmid in *Rhizobium meliloti* carrying exopolysaccharide and thiamine synthesis genes. *Journal of bacteriology*, 167(1):66–72, 1986.

[28] C Pena, L Miranda, D Segura, C Nunez, G Espin, and E Galindo. Alginate production by *Azotobacter vinelandii* mutants altered in poly-β-hydroxybutyrate and alginate biosynthesis. *Journal of Industrial Microbiology and Biotechnology*, 29(5):209–213, 2002.

[29] Itzel Gaytán, Carlos Peña, Cinthia Núñez, María S Córdova, Guadalupe Espín, and Enrique Galindo. *Azotobacter vinelandii* lacking the Na+-NQR activity: a potential source for producing alginates with improved properties and at high yield. *World Journal of Microbiology and Biotechnology*, 28(8):2731–2740, 2012.

[30] Estela M Galván, María V Ielmini, Yamini N Patel, María I Bianco, Esteban A Franceschini, Jane C Schneider, and Luis Ielpi. Xanthan chain length

is modulated by increasing the availability of the polysaccharide copolymerase protein GumC and the outer membrane polysaccharide export protein GumB. *Glycobiology*, 23(2):259–272, 2013.

[31] Alvaro Díaz-Barrera, Paulina Silva, Julio Berrios, and Fernando Acevedo. Manipulating the molecular weight of alginate produced by *Azotobacter vinelandii* in continuous cultures. *Bioresource technology*, 101(23):9405–9408, 2010.

[32] Randal A Hassler and Daniel H Doherty. Genetic engineering of polysaccharide structure: production of variants of xanthan gum in *Xanthomonas campestris*. *Biotechnology Progress*, 6(3):182–187, 1990.

[33] Rodolphe Barrangou, Christophe Fremaux, Hélène Deveau, Melissa Richards, Patrick Boyaval, Sylvain Moineau, Dennis A Romero, and Philippe Horvath. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 315(5819):1709–1712, 2007.

[34] Kira S Makarova, Yuri I Wolf, Omer S Alkhnbashi, Fabrizio Costa, Shiraz A Shah, Sita J Saunders, Rodolphe Barrangou, Stan JJ Brouns, Emmanuelle Charpentier, Daniel H Haft, et al. An updated evolutionary classification of CRISPR–Cas systems. *Nature Reviews Microbiology*, 13(11):722–736, 2015.

[35] CRISPR/Cas9 & Targeted Genome Editing: New Era in Molecular Biology. `https://international.neb.com/tools-and-resources/feature-articles/crispr-cas9-and-targeted-genome-editing-a-new-era-in-molecular-biology`. Accessed: 11.07.2020.

[36] Mary Gearing. *CRISPR 101: A Desktop Resource*. Addgene, 2017.

[37] Alexander Martynov, Konstantin Severinov, and Iaroslav Ispolatov. Optimal number of spacers in CRISPR arrays. *PLoS computational biology*, 13(12):e1005891, 2017.

[38] Rodolphe Barrangou and Luciano A Marraffini. CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. *Molecular cell*, 54(2):234–244, 2014.

[39] Tautvydas Karvelis, Giedrius Gasiunas, Algirdas Miksys, Rodolphe Barrangou, Philippe Horvath, and Virginijus Siksnys. crRNA and tracrRNA guide Cas9-mediated DNA interference in *Streptococcus thermophilus*. *RNA biology*, 10(5):841–851, 2013.

[40] Elitza Deltcheva, Krzysztof Chylinski, Cynthia M Sharma, Karine Gonzales, Yanjie Chao, Zaid A Pirzada, Maria R Eckert, Jörg Vogel, and Emmanuelle Charpentier. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 471(7340):602–607, 2011.

[41] Martin Jinek, Krzysztof Chylinski, Ines Fonfara, Michael Hauer, Jennifer A Doudna, and Emmanuelle Charpentier. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *science*, 337(6096):816–821, 2012.

[42] Francisco JM Mojica, César Díez-Villaseñor, Jesús García-Martínez, and Cristóbal Almendros. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, 155(3):733–740, 2009.

[43] Wenyan Jiang, David Bikard, David Cox, Feng Zhang, and Luciano A Marraffini. CRISPR-assisted editing of bacterial genomes. *Nature biotechnology*, 31(3):233, 2013.

[44] Anthony J Davis and David J Chen. DNA double strand break repair via non-homologous end-joining. *Translational cancer research*, 2(3):130, 2013.

[45] Robert S Pitcher, Nigel C Brissett, and Aidan J Doherty. Nonhomologous end-joining in bacteria: a microbial perspective. *Annu. Rev. Microbiol.*, 61:259–282, 2007.

[46] Xuan Li and Wolf-Dietrich Heyer. Homologous recombination in DNA repair and DNA damage tolerance. *Cell research*, 18(1):99–113, 2008.

[47] Marius Rütering, Brady F Cress, Martin Schilling, Broder Rühmann, Mattheos AG Koffas, Volker Sieber, and Jochen Schmid. Tailor-made exopolysaccharides—CRISPR-Cas9 mediated genome editing in *Paenibacillus polymyxa*. *Synthetic Biology*, 2(1):ysx007, 2017.

[48] Elliot Nicholas Grady, Jacqueline MacDonald, Linda Liu, Alex Richman, and Ze-Chun Yuan. Current knowledge and perspectives of *Paenibacillus*: a review. *Microbial cell factories*, 15(1):203, 2016.

[49] Carol Ash, JAE Farrow, Sally Wallbanks, and MD Collins. Phylogenetic heterogeneity of the genus *Bacillus* revealed by comparative analysis of small-subunit-ribosomal RNA sequences. *Letters in Applied Microbiology*, 13(4):202–206, 1991.

[50] Carol Ash, Fergus G Priest, and M David Collins. Molecular identification of rRNA group 3 *bacilli* (Ash, Farrow, Wallbanks and Collins) using a PCR probe test. *Antonie van Leeuwenhoek*, 64(3-4):253–260, 1993.

[51] W Raza, W Yang, and QR Shen. *Paenibacillus polymyxa*: antibiotics, hydrolytic enzymes and hazard assessment. *Journal of Plant Pathology*, pages 419–430, 2008.

[52] Rocheli de Souza, Jacqueline Meyer, Rodrigo Schoenfeld, Pedro Beschoren da Costa, and Luciane MP Passaglia. Characterization of plant growth-promoting bacteria associated with rice cropped in iron-stressed soils. *Annals of microbiology*, 65(2):951–964, 2015.

[53] Michael Fürnkranz, Eveline Adam, Henry Müller, Martin Grube, Herbert Huss, Johanna Winkler, and Gabriele Berg. Promotion of growth, health and stress tolerance of styrian oil pumpkins by bacterial endophytes. *European journal of plant pathology*, 134(3):509–519, 2012.

[54] Jian-Bo Xie, Zhenglin Du, Lanqing Bai, Changfu Tian, Yunzhi Zhang, Jiu-Yan Xie, Tianshu Wang, Xiaomeng Liu, Xi Chen, Qi Cheng, et al. Comparative genomic analysis of n 2-fixing and non-n 2-fixing *Paenibacillus* spp.: organization, evolution and expression of the nitrogen fixation genes. *PLoS Genet*, 10(3):e1004231, 2014.

[55] Jianbo Xie, Haowen Shi, Zhenglin Du, Tianshu Wang, Xiaomeng Liu, and Sanfeng Chen. Comparative genomic and functional analysis reveal conservation of plant growth promoting traits in *Paenibacillus polymyxa* and its closely related species. *Scientific reports*, 6(1):1–12, 2016.

[56] Carolin Delker, Anja Raschke, and Marcel Quint. Auxin dynamics: the dazzling complexity of a small molecule's message. *Planta*, 227(5):929–941, 2008.

[57] Quyet-Tien Phi, Yu-Mi Park, Choong-Min Ryu, Seung-Hwan Park, Sa-Youl Ghim, et al. Functional identification and expression of indole-3-pyruvate decarboxylase from *Paenibacillus polymyxa* e681. *J Microbiol Biotechnol*, 18(7):1235–44, 2008.

[58] Wafaa M Haggag. Colonization of exopolysaccharide-producing *Paenibacillus polymyxa* on peanut roots for enhancing resistance against crown rot disease. *African Journal of Biotechnology*, 6(13), 2007.

[59] Stephen A Cochrane and John C Vederas. Lipopeptides from *Bacillus* and *Paenibacillus* spp.: a gold mine of antibiotic candidates. *Medicinal Research Reviews*, 36(1):4–31, 2016.

[60] Tony Velkov, Kade D Roberts, Roger L Nation, Jiping Wang, Philip E Thompson, and Jian Li. Teaching 'old'polymyxins new tricks: new-generation lipopeptides targeting gram-negative 'superbugs'. *ACS chemical biology*, 9(5):1172–1177, 2014.

[61] Yoshio Kajimura and Miyuki Kaneda. Fusaricidin A, a new depsipeptide antibiotic produced by *Bacillus polymyxa* KT-8. *The Journal of antibiotics*, 49(2):129–135, 1996.

[62] Hikmate Abriouel, Charles MAP Franz, Nabil Ben Omar, and Antonio Gálvez. Diversity and applications of *Bacillus* bacteriocins. *FEMS microbiology reviews*, 35(1):201–232, 2011.

[63] Zengguo He, Duygu Kisla, Liwen Zhang, Chunhua Yuan, Kari B Green-Church, and Ahmed E Yousef. Isolation and identification of a *Paenibacillus polymyxa* strain that coproduces a novel lantibiotic and polymyxin. *Applied and Environmental Microbiology*, 73(1):168–178, 2007.

[64] Firouz Abbasian, Robin Lockington, Megharaj Mallavarapu, and Ravi Naidu. A comprehensive review of aliphatic hydrocarbon biodegradation by bacteria. *Applied biochemistry and biotechnology*, 176(3):670–699, 2015.

[65] AK Haritash and CP Kaushik. Biodegradation aspects of polycyclic aromatic hydrocarbons (PAHs): a review. *Journal of hazardous materials*, 169(1-3):1–15, 2009.

[66] Barbara J Morrissey, Thorunn Helgason, Lena Poppinga, Anne Fünfhaus, Elke Genersch, and Giles E Budge. Biogeography of *Paenibacillus larvae*, the causative agent of american foulbrood, using a new multilocus sequence typing scheme. *Environmental microbiology*, 17(4):1414–1424, 2015.

[67] Nidhi Gopal, Colin Hill, Paul R Ross, Tom P Beresford, Mark A Fenelon, and Paul D Cotter. The prevalence and control of *Bacillus* and related spore-forming bacteria in the dairy industry. *Frontiers in microbiology*, 6:1418, 2015.

[68] Tzu-Wen Liang and San-Lang Wang. Recent advances in exopolysaccharides from *Paenibacillus spp.*: production, isolation, structure, and bioactivities. *Marine drugs*, 13(4):1847–1863, 2015.

[69] Hee-Kyoung Jung, Joo-Heon Hong, Seung-Chun Park, Byung-Kwon Park, Doo-Hyun Nam, and Sang-Dal Kim. Production and physicochemical characterization of β-glucan produced by *Paenibacillus polymyx*a JB115. *Biotechnology and Bioprocess Engineering*, 12(6):713–719, 2007.

[70] Youn W Han and Margaret A Clarke. Production and characterization of microbial levan. *Journal of Agricultural and Food Chemistry*, 38(2):393–396, 1990.

[71] Ou Li, Cui Lu, Ao Liu, Liang Zhu, Pin-Mei Wang, Chao-Dong Qian, Xin-Hang Jiang, and Xue-Chang Wu. Optimization and characterization of polysaccharide-based bioflocculant produced by *Paenibacillus elgii* b69 and its application in wastewater treatment. *Bioresource technology*, 134:87–93, 2013.

[72] Sung-Gu Ahn, Hyun-Hyo Suh, Chang-Ho Lee, Seong-Hoon Moon, Hee-Sik Kim, Keug-Hyun Ahn, Gi-Seok Kwon, Hee-Mock Oh, and Byung-Dae Yoon. Isolation and characterization of a novel polysaccharide producing *Bacillus polymyxa* A49 KCTC 4648P. *Journal of microbiology and biotechnology*, 8(2):171–177, 1998.

[73] Waseem Raza, Kousar Makeen, Yang Wang, Yangchun Xu, and Shen Qirong. Optimization, purification, characterization and antioxidant activity of an extracellular polysaccharide produced by *Paenibacillus polymyxa* SQR-21. *Bioresource technology*, 102(10):6095–6103, 2011.

[74] Seon-Won Kim, Seung-Gu Ahn, Weon-Taek Seo, Gi-Seok Kwon, and Young-Hoon Park. Rheological properties of a novel high viscosity polysaccharide, a49-pol, produced by *Bacillus polymyxa*. *Journal of microbiology and biotechnology*, 8(2):178–181, 1998.

[75] JK Madden, ICM Dea, and DC Steer. Structural and rheological properties of the extracellular polysaccharides from *Bacillus polymyxa*. *Carbohydrate polymers*, 6(1):51–73, 1986.

[76] Jun Liu, Jianguang Luo, Hong Ye, Yi Sun, Zhaoxin Lu, and Xiaoxiong Zeng. Production, characterization and antioxidant activities in vitro of exopolysaccharides from endophytic bacterium *Paenibacillus polymyxa* EJS-3. *Carbohydrate polymers*, 78(2):275–281, 2009.

[77] José Antonio Morillo, Victor Guerra del Aguila, Margarita Aguilera, Alberto Ramos-Cormenzana, and Mercedes Monteoliva-Sánchez. Production and characterization of the exopolysaccharide produced by *Paenibacillus jamilae* grown on olive mill-waste waters. *World Journal of Microbiology and Biotechnology*, 23(12):1705, 2007.

[78] IV Yegorenkova, KV Tregubova, L Yu Matora, GL Burygin, and VV Ignatov. Composition and immunochemical characteristics of exopolysaccha-

rides from the rhizobacterium *Paenibacillus polymyxa* 1465. *Microbiology*, 77(5):553–558, 2008.

[79] Weon-Taek Seo, Goon-Gjung Kahng, Sang-Hae Nam, Sang-Do Choi, Hyun-Hyo Suh, Seon-Won Kim, and Yong-Ha Park. Isolation and characterization of a novel exopolysaccharide-producing *Paenibacillus sp.* WN9 KCTC 8951P. *Journal of microbiology and biotechnology*, 9(6):820–825, 1999.

[80] Patamaporn Sukplang. Production and characterization of a novel extracellular polysaccharide produced by *Paenibacillus velaei* sp. nov. 2003.

[81] Chuan-Lu Wang, Tzu-Huang Huang, Tzu-Wen Liang, Chun-Yong Fang, and San-Lang Wang. Production and characterization of exopolysaccharides and antioxidant from *Paenibacillus sp.* TKU023. *New biotechnology*, 28(6):559–565, 2011.

[82] Eizi Ninomiya and Tadashige Kizaki. Bacterial polysaccharide from *Bacillus polymyxa* no. 271. *Die Angewandte Makromolekulare Chemie: Applied Macromolecular Chemistry and Physics*, 6(1):179–185, 1969.

[83] Marius Rütering, Jochen Schmid, Moritz Gansbiller, Andre Braun, Jochen Kleinen, Martin Schilling, and Volker Sieber. Rheological characterization of the exopolysaccharide paenan in surfactant systems. *Carbohydrate polymers*, 181:719–726, 2018.

[84] Bernard Henrissat. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochemical journal*, 280(2):309–316, 1991.

[85] Glycoside Hydrolase family classification. `http://www.cazy.org/Glycoside-Hydrolases.html`. Accessed: 23.07.2020.

[86] Gideon Davies and Bernard Henrissat. Structures and mechanisms of glycosyl hydrolases. *Structure*, 3(9):853–859, 1995.

[87] Erica E Rosenbaum, Eva Vasiljevic, Kimberley S Brehm, and Nansi Jo Colley. Mutations in four glycosyl hydrolases reveal a highly coordinated pathway for rhodopsin biosynthesis and n-glycan trimming in *Drosophila melanogaster*. *PLoS Genet*, 10(5):e1004349, 2014.

[88] Shin Kawano, Kenji Tajima, Yukako Uemori, Hitomi Yamashita, Tomoki Erata, Masanobu Munekata, and Mitsuo Takai. Cloning of cellulose synthesis related genes from *Acetobacter xylinum* ATCC23769 and ATCC53582: comparison of cellulose synthetic ability between strains. *DNA research*, 9(5):149–156, 2002.

[89] Shan Yu, Tiantian Su, Huijun Wu, Shiheng Liu, Di Wang, Tianhu Zhao, Zengjun Jin, Wenbin Du, Mei-Jun Zhu, Song Lin Chua, et al. PslG, a self-produced glycosyl hydrolase, triggers biofilm disassembly by disrupting exopolysaccharide matrix. *Cell research*, 25(12):1352–1367, 2015.

[90] Michael J Franklin, David E Nivens, Joel T Weadge, and P Lynne Howell. Biosynthesis of the *Pseudomonas aeruginosa* extracellular polysaccharides, alginate, pel, and psl. *Frontiers in microbiology*, 2:167, 2011.

[91] Hege H Pedersen. CRISPR-Cas based gene replacement for *in-vivo* biopolymer engineering in *Paenibacillus polymyxa*, 2020.

[92] Benchling [biology software]. `https://www.benchling.com`. Accessed: 2020.

[93] Snapgene software (from gsl biotech). `https://www.snapgene.com/`. Accessed: 2020.

[94] PCR Protocol for Phusion High-Fidelity DNA Polymerase (M0530). `https://www.https://www.international.neb.com/protocols/0001/01/01/pcr-protocol-m0530`. Accessed: 14.05.2020.

[95] PCR Protocol for Taq DNA Polymerase with Standard Taq Buffer (M0273). `https://international.neb.com/protocols/0001/01/01/taq-dna-polymerase-with-standard-taq-buffer-m0273`. Accessed: 22.06.2020.

[96] New England Biolabs. XbaI. `https://international.neb.com/products/r0145-xbai#Product%20Information`. Accessed: 10.05.2020.

[97] New England Biolabs. SpeI. `https://international.neb.com/products/r0133-spei#Product%20Information`. Accessed: 10.05.2020.

[98] New England Biolabs. NEBioCalculator. `https://nebiocalculator.neb.com/#!/ligation`. Accessed: 12.05.2020.

[99] Oxford Gene Technology. Understanding and measuring variations in DNA sample quality. `https://www.ogt.com/resources/literature/483_understanding_and_measuring_variations_in_dna_sample_quality`. Accessed: 28.01.2020.

[100] Stephen F Altschul, Warren Gish, Webb Miller, Eugene W Myers, and David J Lipman. Basic local alignment search tool. *Journal of molecular biology*, 215(3):403–410, 1990.

[101] T Madden. Chapter 16: The BLAST Sequence Analysis Tool, The NCBI Handbook (2002), 2012.

[102] B Rehm. Bioinformatic tools for DNA/protein sequence analysis, functional assignment of genes and protein classification. *Applied microbiology and biotechnology*, 57(5-6):579–592, 2001.

[103] Grzegorz M Boratyn, Christiam Camacho, Peter S Cooper, George Coulouris, Amelia Fong, Ning Ma, Thomas L Madden, Wayne T Matten, Scott D McGinnis, Yuri Merezhuk, et al. BLAST: a more efficient report with usability improvements. *Nucleic acids research*, 41(W1):W29–W33, 2013.

[104] Ramy K Aziz, Daniela Bartels, Aaron A Best, Matthew DeJongh, Terrence Disz, Robert A Edwards, Kevin Formsma, Svetlana Gerdes, Elizabeth M Glass, Michael Kubal, et al. The RAST Server: rapid annotations using subsystems technology. *BMC genomics*, 9(1):1–15, 2008.

[105] Nabil-Fareed Alikhan, Nicola K Petty, Nouri L Ben Zakour, and Scott A Beatson. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC genomics*, 12(1):402, 2011.

[106] Karianne Bakkevig, Håvard Sletta, Martin Gimmestad, Randi Aune, Helga Ertesvåg, Kristin Degnes, Bjørn Erik Christensen, Trond E Ellingsen, and Svein Valla. Role of the *Pseudomonas fluorescens* alginate lyase (AlgL) in clearing the periplasm of alginates not exported to the extracellular environment. *Journal of bacteriology*, 187(24):8375–8384, 2005.

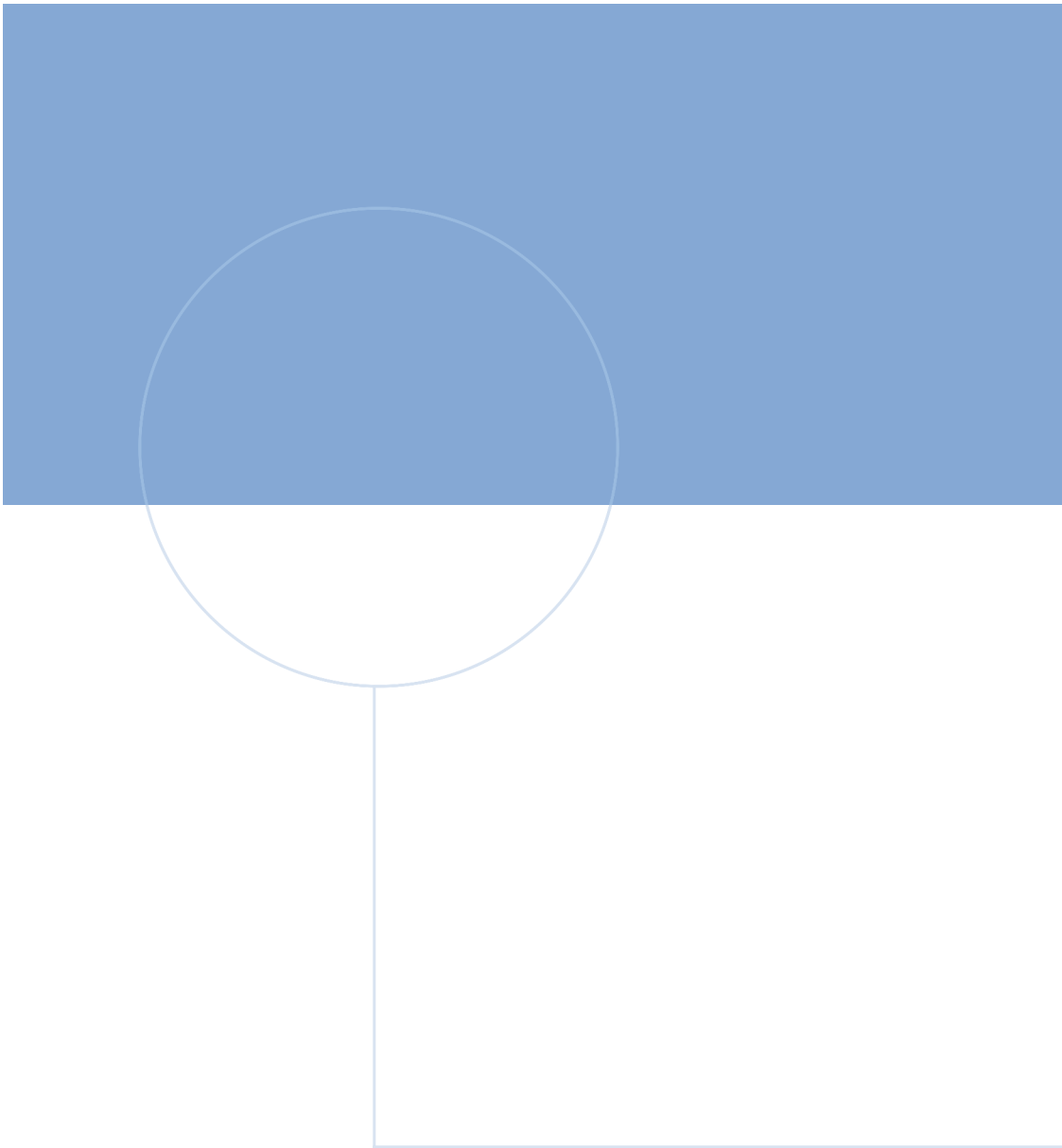
[107] Anne K Haraldsvik. Development of a chassis organism based on CRISPR-Cas engineering in *Paenibacillus polymyxa*, 2020.

Appendix

A Primers

Table 1: Primers used for the construction of pCasPP_ΔpepM and pCasPP_ΔpepN. Underlined letters define cut sites for restriction enzymes. Small letters define overlaps between primers. R=Reverse. F=Forward.

	Name	**Sequence (5'-3')**
1	pepM_sg1 F	ACGCGTATACCCCGAAATACCCGG
2	pepM_sg1 R	AAACCCGGGTATTTCGGGGTATAC
3	OE_US_pepM_XbaI F	TATATCTAGAGAACTGCGGGGTTTTTGCATTC
4	OE_US_pepM R	ggaagaaacaaagcatccTTGTAGAAGGCCCTTTAGC
5	OE_DS_pepM F	ggatgctttgtttcttccTTGATTGATACTCTTTCCCG
6	OE_DS_pepM_XbaI R	TATATCTAGAGAACCATCTTTTTGCAGCTTAATATCACC
7	pepN_sg1 F	ACGCGGAGACGTTAGCTTCACCTG
8	pepN_sg1 R	AAACCAGGTGAAGCTAACGTCTCC
9	OE_US_pepN_XbaI F	TATATCTAGAGAATGGTGACTGGTTCTGGTGG
10	OE_US_pepN R	cgatctgttattaagcatatgATCCACCCTCTCTCCTCA
11	OE_DS_pepN F	catatgcttaataacagatcgGAATGTATATAGGACCTGAG
12	OE_DS_pepN_XbaI R	TATATCTAGACTTTCTTAACCTTGCCTTCAGGAGG
13	seq_sg_pCasMC.R	CCAGGGGGAAACGCCTGG
14	seq_sg_pCasMC.F	CCGAATATATCGGTTATGCGTGG
15	seq_PSG5_F	GGAAAGTCTACACGAACCCTTTGGC
16	OE_gumK_F	AGAAGTGATACGATGAGCGTCTCTCCCGCAG
17	OE_gumK_R	CCGACTCATGTTTCTCAATGTGAGAGCGCTGCCTCC
18	pepM_KO_proof_F	ACCGCGAAGTGAGCCATTTTCAAG
19	pepM_KO_proof_R	GGTAGAGTACGTTGGCCAGTAGGC
20	pepN_KO_proof_F	CTCCTCATTGCCGATCTGGATAAGG
21	pepN_KO_proof_R	GCTCAATCATCGGTGGCCTCTTAGG



NTNU

Norwegian University of
Science and Technology