Genome mining for new ribosomal antimicrobial peptides

Antti Mattila

Master's thesis

University of Helsinki

Department of Food and Environmental Sciences

Microbiology

Spring 2015

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta/Osasto Fakultet/Sektion – Faculty	Laitos/Institution– Department	
Maatalous-metsätieteellinen tiedekunta	Elintarvike- ja ympäristötieteiden laitos	
Tekijä/Författare – Author		
Antti Mattila		
Työn nimi / Arbetets titel – Title		
Ribosomaalisesti tuotettujen antimikrobisten peptidien genomilouhinta		
Oppiaine /Läroämne – Subject		
Mikrobiologia		
Työn laji/Arbetets art – Level	Aika/Datum – Month and year	Sivumäärä/ Sidoantal – Number of pages
Maisterin tutkielma	Toukokuu 2015	74 + 2 liitesivua

Tiivistelmä/Referat – Abstract

Antibioottiresistenssi on maailmanlaajuinen ongelma, joka uhkaa eri patogeenien aiheuttamien tautien hoitoa. Kaikki eliöt tuottavat ribosomaalisesti erilaisia pieniä peptidejä. Ne ovat laajalle levinneitä luonnossa ja niitä voidaan hyödyntää antimikrobisinä yhdisteinä lääkeaineina tai bioteknisissä sovelluksissa. Bakteerit tuottavat paljon erilaisia bioaktiivisia peptidejä, joista monet ovat antimikrobisia ja mikrobigenomien arvellaan koodaavan monia tuntemattomia antimikrobisia peptidejä. Genomilouhinta on laajentanut ribosomaalisesti tuotettujen peptidien kirjoa viime vuosina. Bioinformatiiviset menetelmät yhdessä molekyylibiologian ja kemiallisen analyysin avulla pyrkivät tunnistamaan uusia bioaktiivisia peptidejä.

Tietokannoissa julkaistuja bakteerigenomeita tutkittiin ja löydettiin uudenlainen geenijoukko, joka muistuttaa syanobaktiinien tuottogeenejä. Tuntematonta peptidiä koodaava geenijoukko löytyi syanobakteereista ja myös muista bakteereista. Työn tavoitteena oli tutkia geenijoukon esiintymistä eri bakteerien genomeissa ja niiden tuottamien peptidien mahdollisia rakenteita bioinformatiikan menetelmillä. Geenijoukon toiminta pyrittiin osoittamaan LCMS- ja Q-TOF -menetelmien avulla bioinformatiivisiin analyyseihin perustuen. Tavoitteena oli todistaa ennustettujen peptidien olemassaolo käyttämällä eri rikin isotooppeja kandidaattikannan kasvatusalustassa, jolla pyrittiin havaitsemaan kysteiiniä sisältävät peptidit. Lisäksi geenijoukon toiminta pyrittiin osoittamaan kloonaamalla peptidien tuottogeenit *E. coli* -bakteeriin ja tutkia peptidien antimikrobisia ominaisuuksia.

Bioinformatiivisen analyysin mukaan tämä geenijoukko koodaa 1–8 esipeptidiä ja proteaasia. Esipeptidi koostuu konservoituneesta johtosekvenssistä (LPxQxxPVxR) sekä hyvin vaihtelevasta ydinpeptidistä, jossa on usein parillinen määrä kysteiiniä. Lopullinen peptidi muodostuu ydinpeptidistä esipeptidin translaation jälkeisten muokkausreaktioiden kautta. Uuden peptidiryhmän tuottogeenit löytyivät 38 eri bakteerigenomista ja ne olivat levinneet laajalle bakteerien pääjaksojen välillä edustaen kuutta eri pääjaksoa; syanobakteerit, proteobakteerit, aktinobakteerit, bakteroidetekset, firmikuutit sekä planktomykeetit. Ydinpeptidejä tarkastelemalla havaittiin, että ne olisivat 8–131 aminohappoa pitkiä. Lisäksi peptidit voitiin jakaa kahteen eri ryhmään oletetun rakenteensa mukaan: Toisissa on 2–5 rikkisiltaa kysteiinien välillä ja osa vaikuttaisi muodostavan alfa-kierre-rakenteita. Tämän tyyppiset peptidit ovat yleisiä eukaryooteilla, mutta niitä on erittäin harvoin raportoitu bakteereista. Eukaryooteilla vastaavilla peptideillä on usein antimikrobisia ominaisuuksia ja ne ovat osana luontaista immuniteettia.

Tässä työssä osoitettiin, että bioinformatiikan avulla ennustetut peptidit löytyivät syanobakteerikannasta *Pseudanabaena* sp. PCC 6802 käyttäen apuna molekyylibiologiaa sekä kemiallista analytiikkaa. Ilmentämällä geenijoukkoa heterologisesti todistettiin, että työssä tutkitut uudenlaiset geenijoukot tuottavat ennustettuja peptidejä. Lisäksi havaittiin, että lyhyet alfa-kierteiset kationiset peptidit, jotka ennustettiin *Oscillatoria* sp. PCC 10802-, *Dickeya zeae* Ech1591-, *Vibrio nigripulchritudo* SOn1-, *Agarivorans albus* MKT 106-, *Roseibium* sp. TrichSKD4- ja *Yersinia frederiksenii* ATCC 33641 -kannoista, olivat antimikrobisia pitoisuuksissa 0.8–100 µg/ml. Tämän tutkimuksen tulokset todistavat, että bakteerit tuottavat kysteiiniä sisältäviä peptidejä ja jotkut peptidit tästä uudesta peptidiryhmästä ovat antimikrobisia, joita voidaan mahdollisesti hyödyntää lähtökohtana uutta lääkettä kehiteltäessä.

Avainsanat – Nyckelord – Keywords
Genomilouhinta, antimikrobiset peptidit, syanobakteeri, ribosomaalinen peptidibiosynteesi, *Pseudanabaena* sp.

Säilytyspaikka – Förvaringställe – Where deposited
Elintarvike- ja ympäristötieteiden laitos, mikrobiologian ja biotekniikan osaston käsikirjasto

Muita tietoja – Övriga uppgifter – Additional information
Maisterin tutkielman ohjasivat dosentti David Fewer ja MMT Niina Leikoski. Työn rahoitti Suomen Akatemia (259505) ja Helsingin yliopisto (788/51/2010).

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta/Osasto Fakultet/Sektion – Faculty		Laitos/Institution– Department
Faculty of Agriculture and Forestry		Department of Food and Environmental Sciences
Tekijä/Författare – Author		
Antti Mattila		
Työn nimi / Arbetets titel – Title		
Genome mining for new ribosomal antimicrobial peptides		
Oppiaine /Läroämne – Subject		
Microbiology		
Työn laji/Arbetets art – Level	Aika/Datum – Month and year	Sivumäärä/ Sidoantal – Number of pages
Master`s thesis	May 2015	74 + 2 appendices

Tiivistelmä/Referat – Abstract

Antibiotic resistance is a worldwide problem and it threatens the prevention and treatment of infections caused by different pathogens. All living organisms produce natural products including ribosomal peptides with great variety. They are widely distributed in nature and they are playing more significant role in the search of new antimicrobial compounds used as therapeutical agents. Bacteria are a prolific source of peptides many of which are antimicrobial and microbial genomes are widely believed to encode new antimicrobial peptides. Genome mining has expanded the number of families of ribosomally synthesized natural products in recent years. These *In silico* approaches together with molecular biology and chemical analysis aim to identify novel compounds.

In this study an unknown cyanobactin-like gene cluster was discovered by genome mining from genomes of cyanobacteria and also other bacteria. The aim of this work was to study the occurrence of the gene clusters in various bacterial genomes and the structures of novel peptides. The active biosynthesis of these peptides was tested by LCMS- and Q-TOF -analyses based on bioinformatic predictions. The production of the predicted peptides was also tested with stable sulphur isotope labelling. The aim was also to clone the genes needed for peptide biosynthesis into *E. coli* and to study antimicrobial activities of these peptides.

Bioinformatic analyses suggested that the gene clusters encoded 1–8 precursor peptides together with protease. The precursor peptides had conserved leader sequence (LPxQxxPVxR) and a highly variable core sequences, often encoding an even number of cysteines. The mature peptide is eventually formed from core sequence through post-translational changes in the precursor peptide. The gene cluster was present in 38 bacterial genomes representing a diverse selection of bacterial phyla including cyanobacteria, proteobacteria, actinobacteria, bacteroidetes, firmicutes and planctomycetes. Analyses of the precursor peptide core regions suggested that the products are 8–131 amino acids in length. These peptides could be divided into two groups based on their structures: They form a selection of disulphide-bridge stabilized peptides with 2–5 disulphide-bridges as well as short cationic peptides with an α-helical structure. Surprisingly, these types of peptides are common in eukaryotes and part of the innate immune system displaying potent antimicrobial properties but very rarely reported for bacteria.

The peptides predicted from bioinformatic analysis were detected from *Pseudanabaena* sp. PCC 6802 using a combination of molecular biology and structural chemistry. Heterologous expression of the gene cluster from *Pseudanabaena* sp. PCC 6802 in *E. coli* confirmed that the gene cluster is active. A set of short cationic synthetic peptides with α-helical structure predicted from *Oscillatoria* sp. PCC 10802, *Dickeya zeae* Ech1591, *Vibrio nigripulchritudo* SOn1, *Agarivorans albus* MKT 106, *Roseibium* sp. TrichSKD4 and *Yersinia frederiksenii* ATCC 33641 were shown to have potent antimicrobial activity between 0.8–100 µg/ml. These findings prove that predicted cysteine containing peptides are produced by bacteria and some peptides from this novel family have antimicrobial activity, which might pave the way for new possible drugs derived from natural products.

Avainsanat – Nyckelord – Keywords

Genome mining, natural products, antimicrobial peptides, cyanobacteria, ribosomal peptide biosynthesis, *Pseudanabaena* sp.

Säilytyspaikka – Förvaringställe – Where deposited

Department of Food and Environmental Sciences, the reference library of the Division of Microbiology and Biotechnology

Muita tietoja – Övriga uppgifter – Additional information

Master`s thesis was supervised by Docent David Fewer and PhD Niina Leikoski. The study was funded by the Academy of Finland (259505) and the University of Helsinki (788/51/2010).

Contents

ABBREVIATIONS

AMP	antimicrobial peptide
BLAST	Basic local alignment tool
bp	base pair
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen
EtBr	ethidium bromide
HAMBI	University of Helsinki Culture Collection
JCM	Japan Collection of Microorganisms
LB	Luria Bertani broth
LC-MS	liquid chromatography-mass spectrometry
LPS	lipopolysaccharide
mRNA	messenger RNA
m/z	mass to charge ratio
NRPS	non-ribosomal peptide synthesis
ORF	Open reading frame
PCC	The Pasteur Culture collection of *Cyanobacteria*
PCR	polymerase chain reaction
PKS	polyketide syntetase
Q-TOF	quadrapole-time of flight
RiPP	ribosomally synthesized and post-translationally modified peptide
TAE	Tris-Acetate-EDTA
TGY	tryptone glucose yeast
TSA	tryptone soya agar
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

LIST OF TABLES

LIST OF FIGURES

1. LITERATURE REVIEW

1.1 Natural products

Natural products are produced by primary or secondary metabolism of living organisms (microorganisms, insects, plants and animals) (Demain and Sanchez 2009). The primary metabolites are common in all biological systems (Berdy 2005) and they are formed during the stationary phase of microorganism. The secondary metabolites are chemically and taxonomically highly various compounds (Berdy 2005) with different and usually important functions, but they are not vital for growth and reproduction. Secondary metabolites have had a considerable impact on the control of infectious diseases and the development of pharmaceutical industry, for example clinically important antibiotics of natural origin (Zerikly and Challis 2009). Additionally, many chemically synthesized drugs originate from natural products (Demain and Sanchez 2009). Microorganisms are an abundant source of more than 20000 biologically active compounds (Demain and Sanchez 2009), many of these being peptides with exotic structures (Donia et al. 2008; Arnison et al. 2013).

1.2 Genome mining for new natural products

The proliferation of genome sequence data in public databases combined with research has enabled the discoveries of novel natural products from a broad diversity of micro-organisms (Challis 2008; Velázquez and van der Donk 2011; Bachmann et al. 2014). This huge quantity of accessible genome data works as a base for examination and discovery of new natural products (Zerikly and Challis 2009) with *in silico* approaches (Micallef et al. 2014). Genome mining includes examination of the sequenced genome of microorganism to conclude if gene clusters involved in the production of novel natural products can be found in these organisms (Scheffler et al. 2013). The classical approach for natural product discovery usually includes accurate screening of crude extracts from natural sources, such as bacteria, followed by chemical screening and bioactivity testing (Winter et al. 2011). This method is however often strenuous because of probable rediscovery of known natural products (Winter et al. 2011). Genome mining combined with modern techniques, such as LC-MS and Q-TOF, aims to product identification of novel gene clusters (Scheffler et al. 2013).

Genome mining is a powerful tool in the search for new secondary metabolites, because it links directly the discovery of new natural products with the genes encoding them (Zerikly and Challis 2009) and it has the theoretical potential to eliminate all chance from natural product discovery (Bachmann et al. 2014). Prediction of the amino acids included in ribosomal peptides is relatively straightforward by sequence comparison of similar systems and prediction of excision points of mature natural products encoded in larger precursor peptides (Challis 2008). Recognition of the protease cleavage sites in the precursor peptides enable detailed predictions of the mature peptide products (Velásquez and van der Donk 2011). Once the bioinformatics search is done, it is important for the user to manually check the output to inspect the authenticity of the identified gene clusters and to identify possible tailoring enzymes in the gene cluster. Eventually, novel gene clusters can be used for heterologous expression and bioactivity testing (Micallef et al. 2014).

1.3 Non-ribosomal peptide synthesis (NRPS)

Nonribosomal peptides and polyketides (PKS) have long been recognized as a source of a wide array of therapeutic agents (Hancock and Sahl 2006; Kopp and Marahiel 2007). These compounds are produced by large megaenzymes: Multimodular non-ribosomal peptide synthetases (NRPSs) and polyketide synthases (PKSs) (Finking and Marahiel 2004; Kopp and Marahiel 2007; Marahiel 2009). NRPSs assemble either proteinogenic or nonproteinogenic amino acids into the final peptide structure (Fischbach and Walsh 2006). PKSs assemble polyketides from acyl-CoA in a sequential aspect (Fischbach and Walsh 2006). The biggest portion of known complex secondary metabolites biosynthesis can happen by nonribosomal peptide synthetase (NRPS) or the mixed polyketide-NRP pathway (Tan 2007). In contrast to ribosomal machinery, non-ribosomal peptides are synthesized independently without mRNA (Finking and Marahiel 2004). While ribosomal peptide synthesis is usually limited to a set of 20 amino acids as components of proteins, several hundred substrates of NRPSs are known (Marahiel et al. 1997). Some non-ribosomal peptides include unnatural amino acids and other molecules not found in ribosomal peptides (Finking and Marahiel 2004). Hence, structural variety is a predominant element of peptides produced by non-ribosomal machinery (Finking and Marahiel 2004).

Wang et al. (2014) studied 2699 genomes and cataloged 3339 NRPS,- and PKS gene clusters from the three domains of life. They suggest that these gene clusters show variable distribution in bacteria. They are widespread in the phyla Proteobacteria, Actinobacteria, Firmicutes and Cyanobacteria. They are also present in Ascomycota phylum in fungi (Wang et al. 2014). These peptides which are synthesized by fungi, bacteria and lower eukaryotes have been used as pharmacological agents including infection diseases, cancer and immunosuppression (Kopp and Marahiel 2007; Marahiel 2009).

1.4 Ribosomally synthesized and post-translationally modified peptides (RiPP).

Ribosomal machinery is another strategy how bacteria synthesize modified peptides. Ribosomal peptides are natural products which are synthesized by translation of mRNA. According to Arnison et al. (2013) more structurally diverse peptides are emerging from the ribosomally synthesized and post-translationally modified peptide pathway (RiPP). This group is rapidly increasing and 20 families are known (Arnison et al. 2013). The precursor peptide consists of signal sequence, a leader peptide, a core peptide and a recognition sequence. The leader peptides are believed to have numerous roles in post-translational modifications and export (Arnison et al. 2013) and the core peptide encoded into the precursor peptide is eventually produced to form the mature RiPP through post-translational modifications (Figure 1.3). Comprehensive post-translational modifications allow ribosomal peptides for better target recognition, to augment metabolic and chemical stability and to increase chemical functionality. RiPPs undergo a range of post-translational modifications which usually result in the restriction of the peptides conformational flexibility (Arnison et al. 2013). Common features for different classes of ribosomal peptides are for example macrocyclization and diminish conformational flexibility, and modifications to the N- and C-termini to restrain the sensitivity to degradation by exoproteases (Arnison et al. 2013). In all ribosomal peptides, the precursor peptide undergoes modification by tailoring enzymes which are usually found in a single genomic region (Dunbar and Mitchell 2013). Post-translational modifications include, among others, miscellaneous cyclizations, dehydrations, rearrangements, terminal capping, hydroxylations, oxidations, phosphorylations, glycosylations, reductions, targeting moieties and formation of disulphide bridges (Wang 2014). All these modifications endow the peptide with a rigidified structure and bioactivity (Dunbar and Mitchell 2013). Bacteria produce a wide

range of antimicrobial peptides through the posttranslational modification of short precursor proteins (Hassan et al. 2012; Arnison et al. 2013; Cotter et al. 2013).

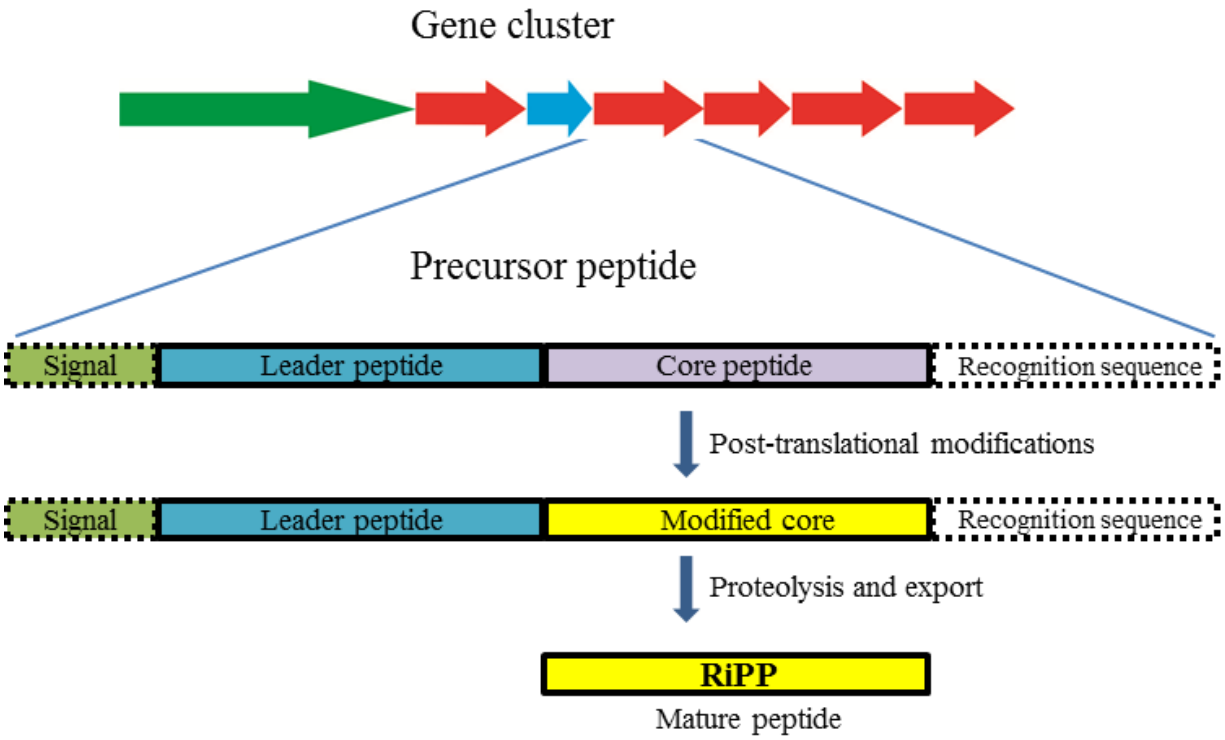


Figure 1.4. Biosynthesis for RiPPs from the precursor peptide (modified from Arnison et al 2013).

1.5 Antimicrobial peptides

The antibiotic resistance is a global health problem and also lack of new antimicrobial medicines is another large concern (Tossi et al. 2000; Giuliani et al. 2007; Hassan et al. 2012). As a consequence, there is a compulsion for the development of new antimicrobial agents for clinical settings that are extensively effective and less likely to inflict antimicrobial resistance (Sang and Blecha 2008; Cotter et al. 2013). An antimicrobial peptide (AMP) is an agent that kills micro-organisms or hinders their growth. Antimicrobial peptides form four wide structural groups: 1) β-sheet peptides stabilized by two to four disulphide bonds (for example human α- and β-defensins); 2) α-helical peptides (for example magainin); 3) extended structures rich in glycine, tryptophan, proline, arginine and/or histidine and 4) cyclic peptides (Boman 1998; Hancock and Lehrer 1998; Hancock and Sahl 2006). From these, β-sheet peptides and α-helical peptides are most widespread in nature (Giuliani et al. 2007).

Natural gene-encoded antimicrobial peptides are produced in all forms of life, from micro-organisms to multicellular organisms (Sang and Blecha 2008; Hassan et al. 2012) and they possess widely recognized activity against Gram-positive and Gram-negative bacteria and also fungi (Wimley and Hristova 2011). Most naturally appearing antimicrobial peptides are cationic and amphipathic (Jin et al. 2005). Cationic peptides are being examined as a new generation of antibiotics, in addition to innate immune modulators (Hancock and Sahl 2006). Most antimicrobial peptides are evolved from bigger precursors, including a signal (leader) sequence (Zasloff 2002), through one or more proteolytic activation steps (Hancock and Sahl 2006). Cationic antimicrobial peptides can be described as being short (10–50 amino acids), with a general positive charge (usually +2 to +9) and a significant portion (≥30%) of hydrophobic amino acid residues (Hancock and Lehrer 1998; Zasloff 2002; Giuliani et al 2007). These characteristics allow the peptide to fold into amphiphilic conformations in three dimensions, forming separate areas rich in positively charged and hydrophobic amino acids, often at the time of contact with membranes (Hancock and Sahl 2006; Giuliani et al. 2007).

1.5.1 Mode of action

AMPs bactericidal activity is expected to happen by interaction with cell membranes resulting in cell membrane integrity disruption. Cell death is being caused by efflux of metabolites and influx of ions (Nolan and Walsh 2009). Factors that affect AMPs molecular mechanism of membrane disruption are among others, amino acid sequence, peptide concentration and membrane lipids (Giuliani et al. 2007). AMPs can be divided into two mechanistic classes by their mode of action: Membrane disruptive (“barrel stave”, “micellar aggregate” and “carpet”) and non-membrane disruptive (intracellular targets) (Powers and Hancock 2003). In order to kill cells, AMPs must first have interaction with bacterial surfaces, and one mechanism is electrostatic bonding between cationic peptides and compounds on the bacterial cell wall (Brogden 2005). These compounds are lipopolysaccharide (LPS) in Gram-negative bacteria and polysaccharides, teichoic acids and lipoteichoic acids in Gram-positive bacteria (Brogden 2005). Once peptides have traversed outer membrane and achieved access to the cytoplasmic membrane they can interact with lipid bilayers (Brogden 2005). It is possible that cationic peptides are attached to anionic phospholipids that are plenteous in bacterial membranes (Powers and Hancock 2003). It is suggested that peptide interacts with target cells membrane,

this causes displacement of lipids and followed by change of membrane structure. Sometimes peptide enters the cytoplasm of target cell (Zasloff 2002).

1.6 Antimicrobial peptides from eukaryotes

1.6.1 α-helical peptides

Linear α-helical peptides are one of the most plenteous and widespread in nature and seem to represent a successful structure in innate defense (Tossi et al. 2000), particularly in higher organisms including invertebrates, amphibians, fish and mammals (Tossi et al. 2000; Hassan et al. 2012). α-helical peptides adopt amphipathic structures only when in contact with membranes, meanwhile they are highly flexible in solution (Hancock and Sahl 2006). According to Tossi et al. (2000) they are most likely short (<40 amino acid residues), which makes them accessible to chemical synthesis. They are differently active against various pathogens, including Gram-positive and Gram-negative bacteria, protozoa, fungi (Tossi et al. 2000), parasites, enveloped viruses and some cancer cells (Sang and Blecha 2008). Structure/activity relationship-studies have revealed at least seven factors that can impact the potency and activity of α-helical anti-microbial peptides: the size, the sequence, the helical content (%), the charge of the compound, the general hydrophobicity, the amphipathicity and the widths of the hydrophobic and hydrophilic sites of the helix respectively. All of these factors are closely related, so modifying one factor can result in remarkable changes to one or more of the other factors (Tossi et al. 2000). Antimicrobial capability is correlated with an amphipathic structure, an α-helix or a β-sheet. Moreover, β-sheet peptides occur to be more selective in contrast with amphipathic α-helical peptides. (Jin et al. 2005).

1.6.1.1 Magainins

Zasloff (1987) reported two 23 amino acid residues long peptides with broad spectrum antimicrobial activity from the skin of African clawed frog *Xenopus laevis*, called magainins. These peptides were shown to be active against Gram-negative and Gram-positive bacteria, fungi and protozoa (Zasloff 1987). A few years later Moore et al (1991) found out that magainins are also synthesized in the stomach of *X. laevis* and they undergo the same post-

translational modifications as do those ones found in the skin (Moore et al 1991). Magainin peptides are amphiphilic α-helices by their structure (Zasloff 1987).

Pexiganan, a synthetic analog of magainin, made it to Phase 3 clinical trials and was indicated for the treatment of diabetic foot ulcers. However, it was not approved by the US Food and Drug Administration (FDA) in 1999. This was a major setback for the whole industry of antimicrobial peptides as drugs. Pexiganan is, however, getting another chance entering clinical trials by another pharmaceutical company (Fox 2013). The major disadvantage of antimicrobial peptides in general is their high production cost compared to other types of antibiotics (Wimley and Hristova 2011).

1.6.2 Defensins

Defensins are antibacterial peptides which are produced in mammals in the skin, lungs and gut. They are active against Gram-positive and Gram-negative bacteria and some defensins are even potent against viruses, fungi and protozoa (Nolan and Walsh 2009). Defensins form one of the largest group of antimicrobial peptides, the cationic peptides, which exist widely in animals and plants (Papagianni 2003). They have complex structural motifs and represent a large group of 4-kDa open-ended, cysteine containing, amphiphilic peptides (Papagianni 2003). Defensins are divided into three groups: the α-defensins, the β-defensins and the insect defensins (White et al. 1995). The α-defensins include 29–35 amino acids, the β-defensins 38–42 amino acids (White et al. 1995, Papagianni 2003) and the insect defensins 29–34 amino acid residues (White et al. 1995). They are ribosomally produced and the active peptide is derived from bigger precursor (~60–100 amino acids) by post-translational proteolytic cleavage (Nolan and Walsh 2009). When aligned optimally, α-, and β-defensins share eight identical amino acid residues in their sequences, containing the six cysteines (Nissen-Meyer and Nes 1997; Papagianni 2003). About 30 % of the amino acid residues are substantially conserved among the α-defensins, and 60-70 % are well conserved among the β-defensins (Nissen-Meyer and Nes 1997). By their structure they are all cationic with arginine being the predominant cationic residue, and all defensins are stabilized by three disulfide bridges (White et al. 1995). Mammalian defensins (α-, and β) are rather different from plant defensins in their sequence and structure (Papagianni 2003) and they use an anti-parallel β-sheet, which is surrounded by cationic and hydrophobic residues (Zasloff 2002). The predominant β-sheet with three stabilizing disulfide bridges is an integrating feature

that differentiates them from some other antimicrobial peptides, many of which are formed by amphiphilic α-helices (Nissen-Meyer and Nes 1997).

1.6.3 Conus peptides and conotoxins

According to Wang and Chi (2004) conus species have many venoms for feeding and defense as a survival strategy. Their venoms are mostly a mixture of bioactive peptides which are usually small and structurally constricted. These peptides have very high affinities and selectivities and they are optimized, among others, to target specific ion channels. Conus peptides are classified into two major classes: 1) majority with multiple disulphide bonds, indicated as conotoxins and 2) minority with a single or no disulphide bond (Kaas et al. 2010; Wang and Chi 2004; Akondi et al. 2014), indicated as conopeptides (Terlau and Olivera 2004). It is estimated that there are 5–10 conotoxin gene superfamilies that encode the majority of all conus peptides. Superfamilies of these peptides can be divided into various families based on functionality and structure, each with a specified pharmacological targeting precision (Olivera and Cruz 2001). Most of bioactive conus venom substances are small (sequence length 12–35 amino acids) structured peptides with exceptional potency and selectivity that target ion channels, and most likely these peptides have a specific ion channel as a suited target (Terlau and Olivera 2004; Kaas et al. 2010; Akondi et al. 2014). According to Kaas et al. (2010) conopeptides are ribosomal peptides which are synthesized by mRNA and translated into protein precursors. Precursors are later cleaved and they undergo various maturation steps to form the active peptide. They have a high prevalence of post-translational modifications which ends up in rich chemical diversities. Mature conopeptides are 8–86 amino acids in length, with an average of 26 amino acid residues (Kaas et al. 2010).

Cone snail venom peptides, especially conotoxins, hold tremendous promise for the advancement of peptides as pharmacological drugs (Akondi et al. 2014). The small size, relatively easy synthesis route and target specificity of conotoxins and conopeptides makes them meaningful pharmacological probes (Lewis and Garcia 2003). It has been estimated that less than 0.1 % pharmacologically valuable conopeptides have been characterized out of 50000 (Lewis and Garcia 2003). Conopeptides have eminent specificity for macromolecular targets which can be explained, among other reasons, by their high density of post-translational modifications (Kaas et al. 2010). The formation of disulphide bridges is the most common and

notably important because the protein 3D-structure can be rigidified and stabilized by cross-linking the peptide between cysteine residues (Kaas et al. 2010). Other modifications include for example C-terminal amidation, proline hydroxylation, glutamate γ-carboxylation (Akondi et al. 2014), sulphation (Tyr) and isomerization to D-amino acids (Lewis and Garcia 2003). Rigid secondary structure formed by disulphide bridges exhibits enhanced stability and potency in conotoxins in contrast to nondisulphide peptides (Akondi et al. 2014). Post-translational modifications are essential for bioactive venom-peptides in order to survive chemical and enzymatic degradation by processing proteases existing in the venom itself (Milne et al. 2003).

1.6.4 Spider-venom peptides

The chemical diversity of spider-venom peptides is astonishing: It is estimated that one venom can contain 200 bioactive peptides (King and Hardy 2013). Disulphide-rich peptides are widely represented in spider venom and they are to a great extent responsible of the venom's insecticidal activity (Windley et al. 2012; King and Hardy 2013). Other peptides in the venom include linear and highly cationic cytolytic peptides (King and Hardy 2013) and they are active against Gram-negative and Gram-positive bacteria (Windley et al. 2012). It is however estimated that about 90 % of spider-venom toxins are globular disulphide-rich peptides, containing between one to seven disulphide bonds (Windley et al. 2012). Peptides in spider-venom are typically larger in size compared to cone snail venom peptides (Milne et al. 2003). In addition to the disulphide-rich peptide neurotoxins (King and Hardy 2013), spider venom peptides have antimicrobial, antiparasitic and enzyme inhibitory activity (Saez et al. 2010). Most of these peptides affect by modulating the activity of ion channels and receptors (Saez et al. 2010).

1.7 Cyanobacteria

Cyanobacteria are one of the most successful and oldest life forms still present on earth which can be found aerial, terrestrial, marine, brackish, glaciers and fresh water environments. They symbolize an exceptionally versatile but extremely specialized group of microorganisms adjusted to various ecological habitats (Dixit et al. 2013). Cyanobacteria can be divided taxonomically into five different classes: Chroococcales, Pleurocapsales, Oscillatoriales, Nostocales and Stigonematales. According to Castenholz (2001) subsection I, Chroococcales,

undergo exclusively binary fission in comparison Pleurocapsales (subsection II) which reproduce through multiple fissions to produce smaller daughter cells, baeocytes. Organisms belonging to Oscillatoriales (subsection III) reproduce solely vertical to the growing axis. Strains in subsection Nostocales and Stigonematales (subsections IV and V, respectively) are able to comprehend specific cells. In other words, heterocysts (for nitrogen fixation) and can form akinetes (dormant cells) and hormogonia. Branching filament formation is also a connective feature for subsection V (Castenholz 2001). Toxin-producing cyanobacteria are often those which constitute mass occurrences in water (Sivonen and Börner 2008). These toxins can be harmful to humans and animals.

1.7.1 Secondary metabolites of cyanobacteria

Cyanobacteria have been recognized as one of the most promising group of micro-organisms from which new and bioactive natural products are acquired (Burja et al. 2001; Singh et al. 2005). These secondary metabolites can be biologically active peptides and/or toxins (Burja et al. 2001). Biggest portion of secondary metabolites from cyanobacteria are isolated from organisms belonging to Oscillatoriales (49%), then Nostocales (26%), Chroococcales (15%), Pleurocapsales (6%) and Stigonematales (4%) (Gerwick et al. 2008). Because of their high chemical stability and water solubility, these compounds have important applications (Dixit et al. 2013) (Table 1.6.1) alongside with their many biological activities such as anticancer, antimalarial, anti-HIV, antibiotic and toxic effect (Burja et al. 2001). One cyanobacterial genome can contain multiple biosynthesis pathways for natural products, making them similar to some actinomycetes and myxobacteria (Gerwick et al. 2008) which are example groups of organisms that are as chemically productive as cyanobacteria (Nunnery et al. 2010). Cyanobacteria stand for essential source of new microbial secondary metabolites, alongside with above-mentioned bacteria and fungi, in search for new drug discoveries (Tan 2007). However, genes responsible for secondary metabolite production are found in most cyanobacterial genomes regardless of what environment they are extracted from (Shih et al. 2013).

Table 1.7.1. Examples of antimicrobial peptides from cyanobacteria

Peptide(s)	**Producer strain**	**Bioactivity**	**Reference**
Hassallidins	e.g. *Anabaena* sp.	Antifungal	Vestola et al. (2014)
Brunsvicamides	*Tychonema* sp.	Antibacterial	Müller et al. (2006)
Spiroidesin	*Anabaena spiroides*	Anti-cyanobacterial	Kaya et al. (2002)
Lyngbyazothrins A-D	*Lyngbya* sp.	Antibacterial	Zainuddin et al. (2009)
Tolybyssidins A, B	*Tolypothrix byssoidea*	Antifungal	Jaki et al. (2001)
Lyngbyabellin B	*Lyngbya majuscula*	Antifungal	Milligan et al. (2000)
Balticidins A−D	*Anabaena cylindrica* Bio33	Antifungal	Bui et al. (2014)
Kawaguchipeptin B	*Microcystis aeruginosa*	Antibacterial	Ishida et al. (1997)

1.8 Cyanobactins

Cyanobactins are ribosomally produced, linear and N-to-C macrocyclic peptides with lengths ranging between 3–20 amino acids (Arnison et al. 2013; Leikoski et al. 2013), and they are widespread within cyanobacteria (Donia et al. 2008; Leikoski et al. 2013). Over 100 cyanobactins have been isolated, making them one of the biggest and most important natural products in cyanobacteria (Schmidt and Donia 2009; Donia and Schmidt 2010). Various peptides belonging to this family are prenylated on Ser, Thr, or Tyr, while some are N-methylated on His (Arnison et al. 2013). Cyanobactins contain oxazolines, thiazolines, or their oxidized derivatives oxazoles and thiazoles (Sivonen et al. 2010). Some of these peptides (anacyclamides) comprise only of proteinogenic amino acids (Leikoski et al. 2010).

Cyanobactins are produced through the posttranslational modification and cleavage of short precursor peptides (Sivonen et al. 2010). The precursor peptides include very conserved leader peptides and recognition sequences, but have very versatile core peptide sequences (Oman and van der Donk 2010). The high conservation of the leader and recognition sequences might indicate conserved recognition factors for the biosynthetic enzymes, while the multivariate core peptides makes high chemical diversity possible in the products (Oman and van der Donk 2010). The leader sequence is flanked by recognition sequences which are expected to be the binding site for modifying enzymes responsible for post-translational modifications (Donia et al. 2006; Houssen et al. 2010).

1.8.1 Cyanobactin biosynthesis

Cyanobactin gene clusters are about 10 kb in size and contain 7 to 12 genes (Sivonen et al. 2010). The cyanobactin gene cluster encompasses two proteases which are responsible for cleavage of the precursor peptide and cyclization of the mature product, PatA and PatG (Lee et al 2009; Donia and Scmidt 2010; Sivonen et al. 2010). Eventually the 50–150 amino acids long precursor peptide is cleaved by proteases and modified to form the mature product (Donia and Schmidt 2010; Sivonen et al. 2010). PatA is responsible for proteolytic cleavage of the N-terminal, when PatG catalyzes cleavage of the C-terminal recognition sequence at the same time with cyclization (Lee et al 2009). Macrocyclization happens via the peptide backbone and not along amino acid side chains (Agarwal et al. 2012) and most of these compounds are further processed by heterocyclization (Arnison et al. 2013). PatA and PatG are both subtilisin-like serine proteases (Oman and van der Donk 2010) and they both have a domain of unknown function (Agarwal et al. 2012). The gene order in the gene cluster is not severely conserved (Sivonen et al 2010).

Some cyanobacteria have cyanobactin pathways, which produce highly modified short linear cyanobactins (Leikoski et al. 2013). Their length ranges from three to five amino acids with N-prenylated and O-methylated termini in cyanobacterial strains *Microcystis aeruginosa* PCC 9432 and *Oscillatoria nigroviridis* PCC 7112 (Leikoski et al. 2013).

2. RESEARCH AIMS

The objectives of this thesis project were 1) to map the distribution of cyanobactin-like gene clusters in different bacterial phyla; 2) Study and detect natural products from this novel pathway; 3) Production of detected peptides in heterologous host; 4) Bioactivity testing with synthetic analogs of peptides from this family (Figure 2).

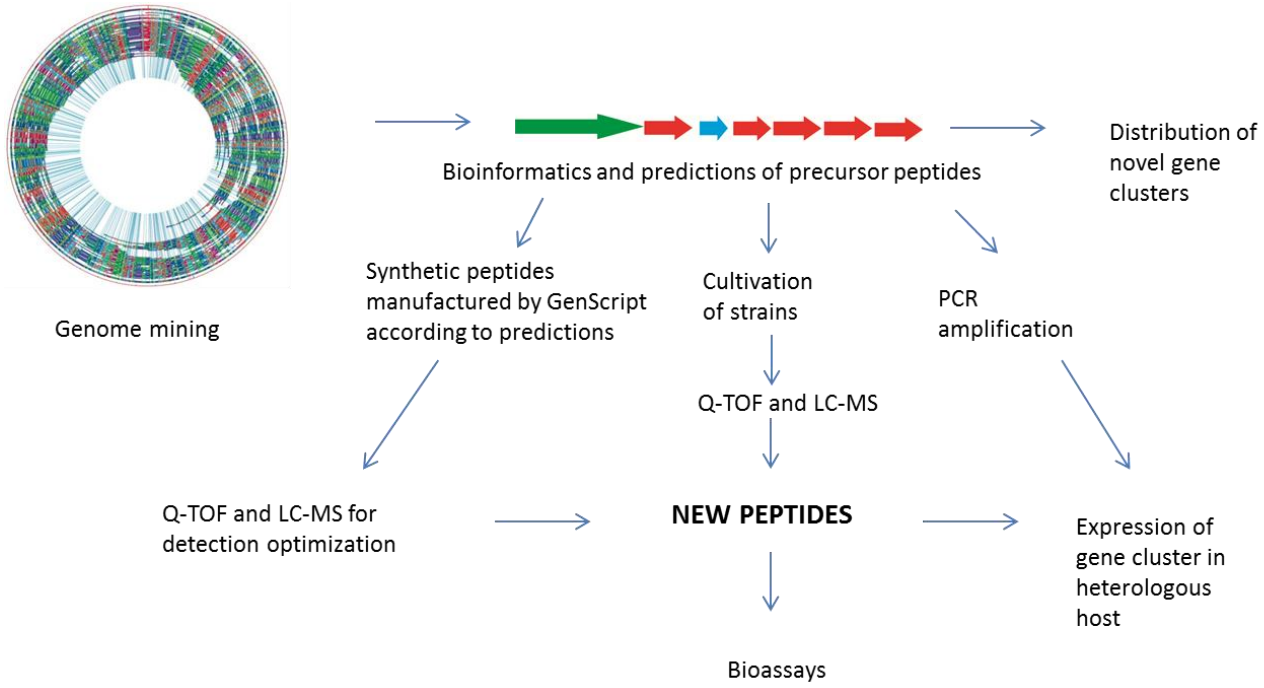


Figure 2. General workflow of natural product detection

3. MATERIALS AND METHODS

3.1 Bioinformatics

Other than known cyanobactin gene clusters were searched using cyanobactin protease PatA by Blastp (Altschul et al. 1997). Blastp searches for cyanobactin protease PatA revealed a single S8-protease together with 1–8 precursor genes and these novel, divergent gene clusters from cyanobactin gene clusters, were mapped and examined further using NCBI and JGI databases. Core peptides encoded to larger precursor peptides from these clusters are hypothesized to be natural products of new ribosomal pathway. These products were named from now on cyanobactin-like peptides.

Protease amino acid sequences of cyanobactin-like gene clusters from different bacterial strains were used to query the database at NCBI. The candidate cyanobactin-like gene clusters were collected and explored manually using Artemis (Rutherford et al. 2000). Artemis recognizes open reading frames from the selected genomic area. Open reading frames and annotated genes were manually examined for conserved amino acid sequences in the predicted leader sequence. Most of the predicted cyanobactin-like precursors were already annotated in the genome as hypothetical proteins, but open reading frames were additionally searched using conserved amino acid motifs present in the leader sequence (LPxQxxPVxR). Predicted precursor amino acid sequences were collected and aligned manually using the BioEdit sequence alignment editor. Hypothetical protease cleavage site was predicted in the precursor gene alignment. Proteases of the gene clusters were aligned using ClustalW for phylogenetic analysis.

Phylogenetic analysis was done including MUSCLE alignment and excluding using GBLOCKS algorithm (http://www.phylogeny.fr/version2_cgi/simple_phylogeny.cgi).

3.2 Synthetic peptides used for detection optimization

Synthetic peptides were ordered from Genscript (USA, Inc.) according to predictions. Stocks of synthetic peptides were prepared in water, 100 % methanol, 50 % methanol and 0.1 % formic acid with concentration of 0.6 mg/ml of the peptide and used for dilution series (600 ng/ml, 60 ng/ml, 6 ng/ml) in order to the detection limit of the peptides. Different peptide concentrations were analysed with cyanobacterial cell matrix (*Cylindrospermum stagnale* PCC 7417, 10.3 mg)

and without the cyanobacterial matrix to find out how the wanted peptide is detected from the bacterial backround. Both glass and plastic vials were used for LC-MS and Q-TOF to test how to prevent the attachment of the peptide to the vial.

3.3 Strains and cultivation

The bacterial strains included in the study were *Leptolyngbya* sp. PCC 7375, *Gloeocapsa* sp. PCC 7428, *Calothrix* sp. PCC 6303, *Oscillatoria* sp. PCC 10802, *Microcystis aeruginosa* PCC 9443, *Cylindrospermum stagnale* PCC 7417, *Pseudanabaena* sp. PCC 6802, *Mastigocladopsis repens* PCC 10914, *Tolypothrix* PCC 7910, *Streptomyces davawensis* JCM 4913, *Mesorhizobium alhagi* HAMBI 3019 and *Aurantimonas manganoxydans* DSM 21871. *S. davawensis* JCM 4913 is a member of actinobacteria, *M. alhagi* HAMBI 3019 and *A. manganoxydans* DSM 21871 belong to the proteobacteria phyla. Cyanobacteria (PCC-strains) were grown in 40 ml of Z8-media (Kotai 1972) in constant light with a photon irradiance of 4–10 µmol m^{-2} s^{-1} at 20–25 °C for 10−30 days. *S. davawensis* JCM 4913 was obtained from the Japan Collection of Micro-organisms and was grown in 50 ml of TY-broth (Tryptone 5,0 g/l, yeast extract 3,0 g/l, $MgSO_4$ x $7H_2O$ 2,0 g/l, distilled water 1,0 l) for 48 h. *M. alhagi* HAMBI 3019 was obtained from HAMBI culture collection and was grown in 50 ml of modified TY-broth (Tryptone 5,0 g/l, yeast extract 3,0 g/l, $CaCl_2$ x $2H_2O$ 0,7 g/l, distilled water 1 l) for 48 h. *A. manganoxydans* DSM 21871 was obtained from the German Collection of Micro-organisms and Cell Cultures and was grown in 50 ml of DSMZ-media 1326 (Peptone 2,0 g/l, yeast extract 0,5 g/l, artificial seawater 980 ml (artificial seawater: NaCl 17,55 g/l, KCl 0,75 g/l, $MgSO_4$ x $7H_2O$ 12,35 g/l, $CaCl_2$ x $7H_2O$ 1,46 g/l, distilled water 1 l)) for 48 h.

Pseudanabaena sp. PCC 6802 was grown in Z8-media in larger scale for bigger biomass production for chemical analyses. Cultivation was carried out in 3 L culture bottles 35.1 L in total, with aeration, under light with a photon irradiance of 4–10 µmol m^{-2} s^{-1} at 20–25 °C for 14 days. Cells were collected by centrifugation at 6000 × G for 10 min at 8–20 °C. The collected cells were freeze dried, which yielded to 6.17 g of dried cells.

3.3.1 Sulphur labeling

Pseudanabaena sp. PCC 6802 was grown with ^{34}S sulphur isotope instead of the normal ^{32}S sulphur in Z8-media. Z8-media is made from three stocksolutions (Z8I, Z8II and Z8III) and from micronutrientsolution. The main sulphur source in Z8I-media is $MgSO_4$ x 7 H_2O. Z8I-media was prepared without $MgSO_4$ x 7 H_2O and autoclaved. Autoclaved Z8I-media was added with sterile filtered (Whatman, Puradisc 25 mm, Sterile polyethersulfone membrane, pore size 0.2 µm, Kent, UK) magnesium sulphate containing sulphur isotope ^{34}S ($MgSO_4$, ICON, Stable isotopes, IS7080, 90 atom % ^{34}S) reaching stocksolution concentration of 0.1 mM, as it is described according to Kotai (1972). Both sulphur isotope labelled cell cultures were grown under light with a photon irradiance of 4–10 µmol m^{-2} s^{-1} at 20–25 °C for 14 days. *Pseudanabaena* sp. PCC 6802 ^{34}S-labeled cell extracts were compared with normal ^{32}S containing cell extracts with LC-MS.

3.3.2 Chemical analysis

Cells were collected from the 40 ml-cultures by centrifugation at 7000 × G for 7 min (Eppendorf Centrifuge 5804 R, Eppendorf). All collected cells and their supernatants were freeze dried (Christ® Beta 2–8 LSCplus, LyoCube 4–8, Osterode am Harz, Germany) yielding 10–55 mg. Freeze dried cells were extracted with 1 ml 50 % methanol and 200 µl of glass beads (Cell disruption media, 0.5 mm glass beads, Scientific industries, Inc.) by crushing mechanically with Fastprep cell disrupter (Bio 101, Thermo Electron Corporation, Qbiogene, Inc. Carlsbad, USA) at a speed of 6.5 m s^{-1} for 45 s. Suspensions were centrifuged at 20000 × G for 10 min and analyzed with LC-MS. Suspensions were filtered through 0.2 µm filter (13 mm syringe filter, PTFE, VWR international) for Q-TOF analysis.

Cell extracts in 50 % methanol from candidate strains were analyzed with high resolution liquidchromatogram (HPLC, Agilent 1100, Agilent Technologies, Santa Clara, USA) combined with mass-spectrometer (MS, Agilent XCT Plus Ion Trap, Agilent Technologies, Santa Clara, USA). Compounds were separated in C_8-column (Phenomenex® 100A Luna C_8), which particle size was 5 µm, inner-diameter 2 mm and length 150 mm. Moving phase consisted of solutions A and B, which were 0.1 % watery solution of formic acid (A) and 0.1 % formic acid in isopropanol (B). At the beginning of analysis moving phase consisted of 5 % liquid A, but during analysis relation was altered evenly that after 35 min proportion of liquid B was 100 %.

Flow rate of moving phase was 0.15 ml min^{-1} and temperature was 40 °C. Fractured compounds were formed into positive ions, which were analyzed by MS in m/z-range 50–2200.
Samples were also analyzed with high definition mass spectrometry (Q-TOF, Waters, SYNAPTG2-Si, USA). Moving phase consisted of solutions A and B, which were 0.1 % watery solution of formic acid (A) and 0.1 % formic acid in acetonitrile (B). At the beginning of analysis moving phase consisted of 95 % liquid A, but during analysis relation was altered evenly that after 5 min proportion of liquid B was 95 %, after 7 min proportion of liquid B was 95 % and after 10 min proportion of liquid A was 95 %. Injection volume was 1 µl with high pressure limit of 15000 psi. Additional instrumental parameters were as follows: Polarity: ES+; capillary: 3.0 kV; source temperature: 120 °C; sampling cone: 40.0; source offset: 80.0; source gas flow: 0.0 ml/min; desolvation temperature 600 °C; cone gas flow 50.0 L/Hr. Fractured compounds were formed into positive ions, which were analyzed by MS in m/z-range 500–2000. Calibration mass range was 622.7–1971.5.

Identification of cyanobactin-like peptides was based on the theoretical molecular weights calculated from the predicted core peptide from candidate strains using ChemBioDraw. Theoretical molecular weights were compared with observed molecular weights resulted from Q-TOF and LC-MS.

3.4 DNA extraction and determination of quality

Genomic DNA was extracted from the strains *Pseudanabaena* sp. PCC 6802 and *Streptomyces davawensis* JCM 4913 using commercial DNA extraction kit (E.Z.N.A.® SP Plant DNA Mini Kit Protocol - Fresh/Frozen Samples, Omega Bio-Tek, Doraville, GA). *Pseudanabaena* sp. PCC 6802 -cells were shaken with glass beads (acid washed, 425–600 µm or 710–1180 µm Sigma-Aldrich, St. Louis, USA) 6.5 m s^{-1} 30 s with shaker after addition of SP-buffer and RNAase-A. For other parts the work was carried out according to protocol for both bacteria. The DNA-yield was determined by NanoDrop (NanoDrop 1000 Spectrophotometer, Thermo Fisher Scientific, Wilmington, USA).

3.4.1 PCR with 16S-primers

The quality of DNA was controlled by 16S rDNA-PCR. This was done to make sure that possible false results in cyanobactin-like gene cluster amplification are not the outcome of bad

quality template DNA. PCR was carried out in 1 x DyNAzyme II enzyme buffer (Thermo Scientific), 200 µM of dNTP (Thermo Scientific), 0.75 µM of F-primer pA (Edwards et al. 1989), 0.75 µM of R-primer B23S (Lepere et al. 2000), 0.4 U of DyNAzyme II DNA polymerase (Thermo Scientific) and 100 ng of template DNA in a final volume of 20 µl. PCR for *Streptomyces davawensis* JCM 4913 template DNA was carried out as described with one exception: Reverse primer was pH (Edwards et al 1989). The PCR was performed with an initial denaturation at 94 °C for 3 min; 30 cycles of denaturation at 94 °C for 30 s, annealing at 54 °C for 30 s, elongation 72 °C for 90 s and the final elongation step at 72 °C for 10 min.

The size and the content of DNA was controlled by gel electrophoresis in room temperature 100 V, 400 mA, 30 min, 0,9 % agarose in 0,5 × TAE-buffer (20 mM Tris, 10 mM acetic acid, 0.5 mM EDTA, pH 8.3)

3.4.2 Primer design for candidate strains and PCR

Spesific primers were designed to amplify the gene cluster containing the genes for cyanobactin-like biosynthesis (Table 3.4.2). The primers were designed from the beginning of the protease gene to the end of the last precursor gene. For *S. davawensis* the primers were designed from the beginning of the protease to the end of the V8-like endopeptidase, including the precursors in between. PCR was carried out in 1 x Phusion green HF buffer (Thermo Scientific), 200 µM of dNTP (Thermo Scientific), 0.75 µM of each primer, 0.4 U of Phusion Green High-Fidelity DNA Polymerase (Thermo Scientific) and 100 ng of template DNA in a final volume of 20 µl. The PCR was performed with an initial denaturation at 94 °C for 3 min; 30 cycles of denaturation at 94 °C for 30 s, annealing at 59.2 °C for 30 s, elongation 72 °C for 90 s and the final elongation step at 72 °C for 10 min. For *S. davawensis* JCM 4913 PCR was carried out as described above, but only specific primers JCM 4913_SpeIF and JCM 4913_SallR were used (Table 3.4.2). PCR-program was also the same, but the annealing temperature of the primers was 63.9 °C. The size and the content of DNA was controlled by gel electrophoresis as described above.

PCR-products were purified with commercial kit (Biotop, NucleoSpin® Gel and PCR Clean-up) and checked by NanoDrop (NanoDrop 1000 Spectrophotometer, Thermo Fisher Scientific, Wilmington, USA).

Table 3.4.2 Primer sequences for candidate strains for cyanobactin-like PCR amplification from *Pseudanabaena* sp. PCC 6802 and *S. dawavensis* JCM 4913.

Primer	**Sequence**
PCC 6802_SpeIF	5-AGGAACTAGTTGACTAGATGCAGTTACTTG-3
PCC 6802_SalIRcR	3-AAATGTCGACCCTTTATAGGCAATCAATG-5
JCM 4913_SpeIF	5-GGGACTAGTGGAACGGATGGACGTCCGG-3
JCM 4913_SallR	3-GCTCCGGCGGTCGACTGCCGTCGGTCAGT-5

3.5 Plasmid constructs

Insert (gene(s) of interest) and vector (plasmid) are first restricted enzymatically and ligated together forming recombinant plasmid (Figure 3.5) which is eventually transformed into *E. coli*.

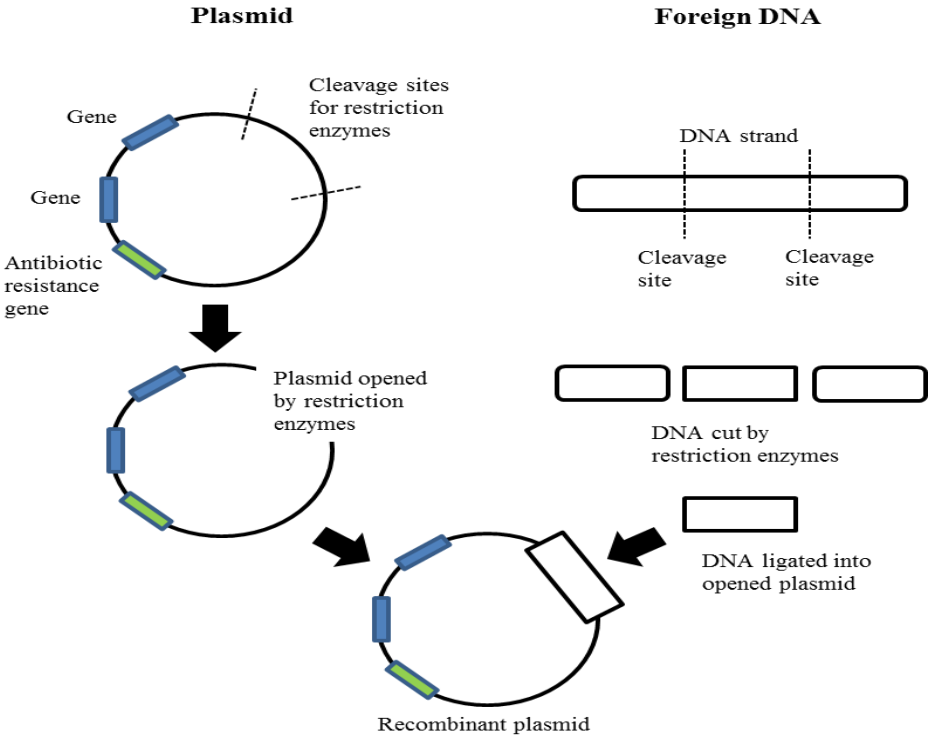


Figure 3.5. Procedure of using plasmid vectors to introduce foreign DNA in *E. coli* (modified from Goering et al. 2012).

3.5.1 Expression construct of *Pseudanabaena sp*. PCC 6802

Plasmids and PCR-product from *Pseudanabaena* sp. PCC 6802 were restricted with two restriction enzymes; SpeI (Promega, Madison WI USA, 10 U/µl, lot 00000731139) and SalI (Promega, Madison WI USA, 10 U/µl, lot 0000111825). Restricted PCR-product was ligated into two different plasmids; low copy number Pacyc84 and high copy number Puc57 (see appendix). Ligations were carried out with insert-to-vector molar ratio of 3:1.

Restriction digestion for insert (purified PCR-product, ~600 ng) was carried out with two restriction enzymes, SpeI and SalI as follows: DNA (purified PCR-product), water, SpeI (20 U), SalI (20 U) and 1 x Buffer (D, Promega, Madison WI USA, 10x, lot 29749708) ending with total volume of 40 µl. Vector (Pacyc184) was restricted similarly ending with total volume of 40 µl. Both restriction mixes were incubated in +37 °C for 3 hours. Ligation mixture was prepared as follows: Insert-to-vector molar ratio of 3:1, water, 1 × Buffer T4 (Promega, Madison WI USA, 10x, lot 0000030467), 3 U of ligase T4 (Promega, Madison WI USA, 3 U/µl, lot 0000081913) ending with total volume of 10 µl. Mixture was incubated overnight in room temperature. Puc57-vector was restricted as Pacyc184-vector ending with total volume of 40 µl. Ligation mixture was prepared as follows: Insert-to-vector molar ratio of 3:1, 1 × Buffer T4, 3 U ligase T4. Mixture was incubated for 2 h in room temperature.

3.5.2 Expression construct of *S. dawavensis* JCM 4913

Method for constructing recombinant plasmid of *S. dawavensis* JCM 4913 PCR-product was as described for *Pseudanabaena* sp. PCC 6802 in paragraph 3.5.1. Restriction digest for insert was carried out as follows: DNA (purified PCR product, ~3000 ng), water, SpeI (20 U), SalI (20 U) and 1 × Buffer (D) ending with total volume of 40 µl. Both vectors (Pacyc84 and Puc57) were restricted similarly ending with total volume of 40 µl. Ligation was performed in 1 × ligation buffer, 3 U of ligase and 3:1 ratio of insert to vector for both plasmids. Ligation mixtures were incubated overnight in room temperature

All restriction products were analysed on 0.9 % agarose gel (Bioline, Taunton MA, USA, lot ES520-BO17430) in 50 ml 0,5 × TAE- buffer (20 mM Tris, 10 mM acetic acid, 0.5 mM EDTA, pH 8.3) stained with Sybr Safe DNA gel stain (InVitrogen, lot 1531654, Carlsbad CA, USA).

Restriction mix (40 µl) was loaded in the well and in order to get the restriction products separated, DNA was controlled by gel electrophoresis in room temperature 100 V, 400 mA, 35 min and was visualized using a dark reader (Clare Chemical Research Inc.).

The gel was observed under UV-light and desired DNA-band was cut from the gel with sterile scalpel and placed to an Eppendorf-tube. DNA was purified from the gel with commercial kit (Qiagen, MinElute Gel Extraction kit) and yield was checked with NanoDrop.

3.6 Heterologous expression in *Escherichia coli*

Cloning and heterologous expression of the studied genes in *E. coli* was carried out in order to prove the existence of predicted cyanobactin-like peptides. In addition to *Pseudanabaena* sp. PCC 6802, *S. davawensis* JCM 4913 was chosen for the cloning. The plasmid constructs (described in 3.5.1 and 3.5.2) were transformed in chemically competent *E. coli* TOP10 -cells (Invitrogen, life technologies) according to manufacturer`s protocol.

Blue-whitescreening with X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) was used to detect the colonies containing the insert of interest. *E. coli* produces β-galactosidase, which fractures X-gal in the media, forming visually indicating blue colonies. Both of the used plasmid vectors, Puc57 and Pacyc84, contain *lac*Z-gene which produces β-galactosidase. These kind of cells fracture X-gal and form blue colonies. The multiple cloning site, where a desired insert was ligated into the vector, is located within the *lacZ*-gene. Successful ligation therefore modifies the open reading frame and disrupts the *lacZ*-gene. Then no functional β-galactosidase can form, resulting in white colonies.

E. coli TOP 10 -cells grew on selective LB-plates (Tryptone 10 g/l, yeast extract 5 g/l, NaCl 5 g/l, agar 15 g/l). Selectivity was determined by antibiotic selection: Chloramphenicol (1 ml/l) for Pacyc84 and ampicillin (2 ml/l) for Puc57. The result of cloning was checked by colony-PCR with specific primers for both strains. PCR-program was as in chapter 3.4.2, but denaturation in the beginning at 94 °C was 10 min. Template was prepared mixing one colony with 10 µl liquid LB-media containing antibiotic from where 2 µl was used for PCR-reaction. Blue colony was picked as a negative control. PCR-products were checked with gelelectrophoresis as described in section 3.4.1.

3.6.1 Plasmid isolation from *E.coli* TOP10 -cells

Colonies containing the ligated plasmid of interest, according to colony-PCR, were picked with sterile glass rod and inoculated in 5 ml LB with antibiotic in sterile glass tube. Glass tubes were grown in shaker +37 °C overnight. Plasmid was isolated with commercial kit (QiaPrep Spin Miniprep Kit, cat. nos 27104 and 27106) according to protocol from overnight cultures. Freezer stocks were prepared by adding 500 µl 85 % glycerol together with 500 µl cell culture and placed to -80 °C.

3.6.2 Plasmid transformation into *E.coli* -BL21 cells

Molecular cloning reaction and heterologous expression was done with Agilent Transformation Protocol (Catalog #200131 (BL21 (DE3) Competent Cells). Five colonies were picked with sterile glass rod and suspended to 5 ml of LB-media with selective antibiotic and grown overnight at +37 °C. The overnight cultures were used to inoculate fresh LB-media in sterile Erlenmeyer-flasks (2 × 50 ml) containing antibiotic to an OD_{600} 0.05–0.1. The cultures were grown until they reached OD_{600} ~0.4. Other flask was induced with IPTG to a final concentration of 0.5 mM and other was kept as a control without IPTG induction. Cultures were incubated for an additional 2–3 hours, collected at 7000 × G for 7 min, freeze dried, extracted with 50 % methanol and analyzed with LC-MS as described earlier.

3.7 Antimicrobial assays

The peptides (11 pc., Table 3.7) were synthesized by GenScript according to predictions and tested against selected bacterial strains (Table 3.7.1). Some of these peptides are somewhat similar to known antimicrobial peptides by their sequence and structure so antibacterial activity was consequently supposed and this hypothesis was tested. The purity of synthetic peptides used in this study was between 75 %–98.8 % (Table 3.7).

Stocks were prepared for antimicrobial peptides with concentration of 0.6 mg/ml in 50 % methanol. Bacterial strains were revived from -80 °C and grown on suitable media the time needed (Table 3.7.1). For antimicrobial testing a suspension of each bacteria was made reaching McFarland 0.5 in milli-Q water by picking 1–5 colonies. Sterile cotton bud was dipped into the

suspension and spread evenly on selective agar-plate. Each disk contained 60 µg of peptide. Peptide suspension was left to stand for 30-60 min so that the suspension would be adsorbed to the disk. Methanol (100 µl) was used as negative control and kanamycin (1000 µg/disk) as positive control. The plates were incubated 24 h or 48 h at temperatures 28 °C or 37 °C (Table 3.7.1). After incubation the radius from the center of the disk was measured to the outer circle of the inhibition zone and reported in millimetres. Experiment was repeated twice. Second time peptides A and B were pipetted on same disk, 60 µg/peptide. This was done to see if there is an enhancement in antibacterial activity when peptides are used in tandem.

Disk diffusion tests were also conducted with cysteine containing synthetic peptides (GASIS, NAVY, AAIF) which were used for detection optimization and they were 75 % pure (Table 3.7).

Table 3.7. Synthetic peptides included in disk diffusion experiment and MIC-test.

Origin of peptide (purity of peptide)	**Tests done**	**Length AA (core)**	**Cysteine (core)**	**Charge (core)**	**Exact mass**	**AA-sequence (core)**
Oscillatoria sp. PCC 10802 A (75%)	Disk	50 (25)	0	-1	2668.51	AWYDTVLDVVKTVAPIAAPILTSLI
Oscillatoria sp. PCC 10802 B (75%)	Disk + MIC	52 (24)	0	0	2558.44	AWYDTLLEVAKKVAPIATTVLGSL
Dickeya zeae Ech1591 (98.8 %)	Disk + MIC	49 (22)	0	2	2272.29	FWGEALKLLKKAGQGALSGVLS
Vibrio nigripulchritudo SOn1 A (98.1 %)	Disk + MIC	49 (22)	0	2	2325.75	FWGTAFDILKKAGKGALQGVMS
Vibrio nigripulchritudo SOn1 B (98.1 %)	Disk + MIC	49 (22)	0	2	2294.22	FWGGVFDVLKKAGKGALQGVMS
Oscillatoria sp. PCC 10802 C (75%)	Disk	50 (25)	2	1	2684.24	**GASIS**DYTSCYNLKGLARNLCMAAY
Leptolyngbya sp. PCC 7375 C (75%)	Disk	63 (31)	4	3	3670.76	IKRTHACR**NAVY**ECQTIGHWPACDYVRKHCL
Tolypothrix PCC 7910 B (75%)	Disk	54 (30)	2	-2	3140.55	**AAIF**GINLCDLMPEPERTACHIAGKLAGTE
Agarivorans albus MKT 106 (85.7 %)	MIC	49 (22)	0	1	2229.60	FWGGVLDVLKKAGQGALNGVLS
Roseibium sp. TrichSKD4 B (87.4%)	MIC	53 (26)	0	3	2674.23	FWGALASAAAPIAIDLGKKALRGLLR
Yersinia frederiksenii ATCC 33641 (96.4%)	MIC	49 (22)	0	1	2216.60	FWDSVLSVLKKAGQGALAGVLG

Table 3.7.1. Bacteria included in bioassay study.

Bacteria (HAMBI nr.)	**Media**	**Temp (°C)**	**Incubation time (h)**
Pseudomonas sp. (2796)	TGY	28	24
Pseudomonas aeruginosa (25)	TGY	37	24
Escherichia coli (396)	TGY	37	24
Burkholderia pseudanalbei (33)	Nutrient	37	24
Salmonella typhi (1306)	Nutrient	37	24
Xanthomonas campestris (104)	Nutrient	28	24
Burkholderia capacia (2487)	TSA	37	24
Bacillus cereus (1881)	TSA	28	24
Micrococcus luteus (2688)	TGY	28	48
Staphylococcus aureus (11)	TSA	37	24
Bacillus subtilis (251)	TGY	28	24

3.8 MIC (Minimum inhibitory concentration)

MIC-value is the lowest concentration of an antimicrobial agent that inhibits the visible growth of a micro-organism. The bacteria studied included *Staphylococcus aureus* HAMBI 11, *Bacillus subtilis* HAMBI 251, *Esherichia coli* HAMBI 396 and *Micrococcus luteus* HAMBI 2688. The test was performed according to Wiegand et al (2008) protocol E with modifications. Bacterial cells were cultured in Müller-Hinton-agar at 37 °C. Synthetic peptides were dissolved in DMSO with concentration of 10 mg/ml (weighed 1 mg/100 µl). This stock was used to make dilution series in ratio 1:2 (50 µl + 50 µl). Müller-Hinton-broth containing bacterial suspension and an antimicrobial peptide was prepared at various concentrations (100, 50, 25, 12.5, 6.25, 3.12, 1.56, 0.78 µg/ml) in 96-well polypropylene-plate (150 µl/well: MH-broth 98.5 µl, bacterial suspension 50 µl and antimicrobial peptide 1.5 µl). Two parallel positive controls were used: + ctrl A (150 µl/well: MH-broth 98.5 µl, bacterial suspension 50 µl and DMSO 1,5 µl) and + ctrl

B (150 µl/well: MH-broth 100 µl and bacterial suspension 50 µl). Negative control was MH-broth (150 µl/well).

Bacteria were first grown overnight and cell counts were calculated (cfu/ml). From these cultures a bacterial suspension was made by picking colonies with a glass rod in 2 ml of 0.9 % NaCl reaching McFarland 0.5. This suspension was then diluted in relation 1:10 or 1:100, depending on the cell density, reaching roughly 10^5 cfu/ml. Prior to this dilution, the concentration of bacteria was adapted to give an initial optical density (OD) reading of about 0.1–0.2 at 600 nm wavelength. Each bacteria was inoculated in its own 96-well plate to avoid contamination and incubated overnight at 37 °C. The MIC-value was taken as the concentration range of the antimicrobial peptide at which no bacterial growth was discovered with bare eye. In addition to sensory observation OD_{600}-values were measured.

4. RESULTS

4.1 Discovery of novel peptidefamily – cyanobactin-like peptides

An unusual cyanobactin-like gene cluster which encoded a single cyanobactin-like protease was found by mining draft and complete genomes, when cyanobactin proteases, PatA and PatG, were used to query the nonredundant database at NCBI. This resulted in a lot of hits many of which were cyanobactin gene clusters. Nonspesific hits to proteases were noticed which were not part of an actual cyanobactin gene cluster. The cyanobactin-like gene clusters were found in 38 strains from different bacterial phyla including cyanobacteria, proteobacteria, firmicutes, actinobacteria, bacteroidetes and planctomycetes (Figure 4.1). These cyanobactin-like gene clusters differed from cyanobactin gene clusters by lacking the proteins for heterocyclization, macrocyclization and prenylation. The cyanobactin-like gene clusters encoded a single S8 cyanobactin-like protease together with 1–8 precursor peptides with a varying length between 8–131 amino acids which most could be identified through the presence of conserved motif (LPxQxxPVxR) in the leader sequence which they shared with known cyanobactin leader sequences. Furthermore recognizable cyanobactin precursor peptide cleavage sites (GxxxS) were present in many of the cyanobactin-like precursor peptides. The cyanobactin-like precursor peptides lacked the terminal recognition sequence of the cyanobactin precursor peptides. This suggested, together with the absence of the cyanobactin macrocyclase and other enzymes for the post-translational modification of cyanobactins, that the product of the gene cluster was a linear peptide. This information, together with the cleavage sites, made it possible to predict the product of the pathway with some degree of precision. The predicted post-translational modifications for the cyanobactin-like precursor peptides include proteolytic cleavage and formation of disulphide bridges which lead to mature natural product. The protease and the precursor(s) were the only enzymes common to all gene clusters and therefore it is likely that there are no further post-translational modifications beyond cleavage. The product of the gene cluster has highly conserved amino acid residues in their leader sequence which share homology between bacterial strains. The core sequences share no homology between bacterial strains and they are very variable (Table 4.1.2.2).

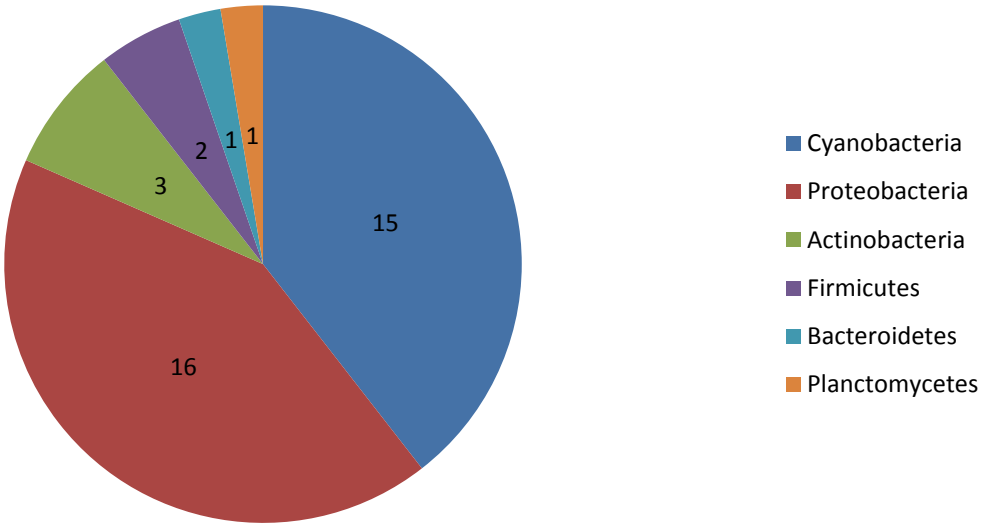


Figure 4.1. Widespread occurrence of cyanobactin-like gene clusters in 38 strains.

The cyanobactin proteases and the cyanobactin-like protease had 40–50% identity. The protease was bimodular encoding two of the domains found in PatA and PatG proteins. Further inspection identified 2–8 short hypothetical open reading frames associated with the protease. In order to determine the distribution of this gene cluster across the bacterial domain BLASTp searches using the S8 protease as a query against the NCBI- and JGI databases were performed. Genome mining strategies illustrated that the biosynthetic genes for cyanobactin-like peptides are distributed widely between bacterial phyla. The size of the gene clusters was 2.3–5.0 kb in the candidate strains. Genomic organization and their putative genes encoding cyanobactin-like peptides of a selection of gene clusters are schematically shown in (Figure 4.1.1). Genome mining resulted to the discovery of cyanobactin-like gene clusters from 38 bacterial strains representing six bacterial phyla including 81 precursor genes. Proteases and some of the precursors are presented in more detail in Table 4.1.2. Precursor peptides are aligned in Figure 4.1.2.1. Table 4.1.2.2 summarizes the predicted precursors from candidate strains and synthetic peptides used in this study.

In order to test these bioinformatics predictions we grew and extracted nine strains of cyanobacteria, two strains of proteobacteria and one strain of actinobacteria. Peptides corresponding to the bioinformatics predictions were detected from *Pseudanabaena* sp. PCC 6802. The cyanobactin-like gene cluster in *Pseudanabaena* sp. PCC 6802 consists of subtilisin-like protease, four precursor genes and one hypothetical protein. The size of the gene cluster is 3485 bp. Three of the four core peptides derived from bigger precursor peptides were found and their m/z values matched well with bioinformatic predictions and theoretical masses, error rate being $\leq$3.5 ppm analyzed by Q-TOF. These detected peptides are linear by structure. These peptides were named cyanobactin-like peptides and they can be broadly divided into cysteine-rich peptides and cysteine-poor peptides, with no disulphide bonds and they appear to be structurally different from each other.

Pseudanabaena sp. PCC 6802 (3500 bp)

Leptolyngbya sp. PCC 7375 (5000 bp)

Gloeocapsa sp. PCC 7428 (2300 bp)

Aurantimonas manganoxydans DSM 21871 (2600 bp)

Mesorhizobium alhagi HAMBI 3019 (2600 bp)

Streptomyces davawensis JCM 4913 (3500 bp)

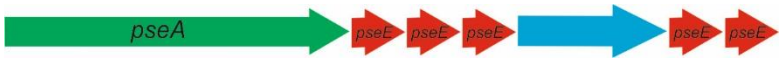
Calothrix sp. PCC 6303 (2700 bp)

Cylindrospermum stagnale PCC 7417 (2400 bp)

Gloeocapsa sp. PCC 7428 (2300 bp)

Tolypothrix sp. PCC 7910 (2700 bp)

Microcystis aeruginosa PCC 9443 (2700 bp)



Oscillatoria sp. PCC 10802 (4500 bp)

Mastigocladopsis repens PCC 10914 (2750 bp)

Figure 4.1.1. A selection of gene clusters found in the genomes of cyanobacteria, proteobacteria and actinobacteria. *PseA* illustrates protease and *pseE* precursor. Blue arrow indicates hypothetical protein. The abbreviation *pse* is derived from *Pseudanabaena* sp. PCC 6802.

Table 4.1.2. Proteases and precursors found by genome mining.

		Protease					Precursor				
Bacteria	**Taxonomy**	**Accession nr.**	**Length AA**	**Max identity (%)**	**Organism**	**Domains in protease**	**Accession nr.**	**Length AA**	**Max identity (%)**	**Organism**	**Precursors**
Leptolyngbya sp. PCC 7375 A	Cyanobacteria	WP_006517599.1	676	63	*Pleurocapsa* sp. PCC 7319	Peptidase S8, unknown	WP_006517600.1	63	63	*Leptolyngbya* sp. PCC 7375	8
Gloeocapsa sp. PCC 7428	Cyanobacteria	YP_007130541.1	632	74	*Calothrix* sp. PCC 6303	Peptidase S8, unknown	YP_007130542.1	82	48	*Cylindrospermum stagnale* PCC 7417	1
Calothrix sp. PCC 6303	Cyanobacteria	YP_007137041.1	636	74	*Gloeocapsa* sp. PCC 7428	Peptidase S8, unknown	YP_007137042.1	128	59	*Stanieria cyanosphaera* PCC 7437	2
Oscillatoria sp. PCC 10802	Cyanobacteria	WP_017721275.1	669	58	*Microcystis aeruginosa* PCC 9807	Peptidase S8, unknown	WP_008179623.1	82	44	*Moorea producens* 3L	5
Microcystis aeruginosa PCC 9443	Cyanobacteria	WP_002766681.1	371 (685)	71	*Cylindrospermum stagnale* PCC 7417	Peptidase S8, unknown	WP_002766678.1	poor hits			1
Cylindrospermum stagnale PCC 7417	Cyanobacteria	YP_007146141.1	646	66	*Calothrix* sp. PCC 6303	Peptidase S8, unknown	YP_007146140.1	81	48	*Gloeocapsa* sp. PCC 7428	1
Pseudanabaena sp. PCC 6802	Cyanobacteria	WP_019502196.1	656	62	*Calothrix* sp. PCC 6303	Peptidase S8, unknown	WP_019502198.1	125	38	*Microcystis aeruginosa* PCC 9443	4
Mastigocladopsis repens PCC 10914	Cyanobacteria	WP_017317508.1	658	71	*Acaryochloris marina* MBIC 11017	Peptidase S8, unknown	No BLAST hits				2
Streptomyces davawensis JCM 4913	Actinobacteria	YP_007526447.1	824	51	*Streptomyces* sp.	Peptidase S8, unknown	YP_007526446.1	59	38	*Sphingobium quisquiliarum* P25	5
Mesorhizobium alhagi HAMBI 3019	Proteobacteria	WP_008834107.1	657	51	*Mastigoclapsis repens*	Peptidase S8, unknown	WP_008834108.1	58	32	*Trichomonas vaginalis* G3	2
Aurantimonas manganoxydans DSM 21871	Proteobacteria	WP_009208976.1	709	46	*Hoeflea phototrophica* DFL-43	Peptidase S8, unknown	WP_009208977.1	130	38	*Magnaporthe oryzae* Y34	1
Yersinia enterocolitica subsp. Palearctica 105.5R	Proteobacteria	YP_004297516.1	575	99	*Yersinia enterocolitica* (Subtilisin DY)	Peptidase S8, unknown	YP_004297515.1	49	85	*Agarivorans albus* MKT 106	1
Agarivorans albus MKT 106	Proteobacteria	WP_016400604.1	648	57	*Dickeya dadantii* Ech586	Peptidase S8, unknown	WP_016400605.1	49	85	*Yersinia enterocolitica* subsp. *Palearctica* Y11	1
Roseibium sp. TrichSKD4 a	Proteobacteria	WP_009465765.1	687	40	*Dickeya dadantii* Ech586	Peptidase S8, unknown	WP_009465763.1	53	67	*Roseibium* sp. TrichSKD4	2
Clostridium sp. JC 122	Firmicutes	WP_010299408.1	1045	44	*Mastigoclapsis repens*	AAT_1 superfamily, Peptidase S8, unknown (Glyco hydro)	WP_010299410.1	79	34	*Naegleria gruberi*	2
Shewanella piezotolerans WP3	Proteobacteria	YP_002314062.1	755	33	*Candidatus Competibacter* phosphatis	Peptidase S8, unknown	YP_002314064.1	52	81	*Shewanella piezotolerans* WP3	2
Moorea producens 3L	Cyanobacteria	WP_008179627.1	1014	68	*Oscillatoria* sp. PCC 10802	Peptidase S8, unknown	WP_008179625.1	51	54	*Oscillatoria* sp. PCC 10802	6
Arthrospira platensis NIES 39	Cyanobacteria	YP_005071907.1	644	98	*Arthrospira platensis* C1	Peptidase S8, unknown	YP_005071906.1	71	99	*Arthrospira* sp. PCC 8005	3
Mycobacterium vulneris DSM 45247	Actinobacteria	CDO33419.1	689	45	*Arthrospira* sp. PCC 8005	Peptidase S8, unknown	CDO33418.1	61	44	*Arthrobacter* sp. 31y	3

Acaryochloris marina MBIC11017	Cyanobacteria	WP_012167256.1	653	71	*Mastigoclapsis repens*	Peptidase S8, unknown	No BLAST hits				1
Planctomycete KSU-1	Planctomycetes	WP_007222671.1	539	42	*Streptomyces* sp. Mg1	Peptidase C1, unknown	WP_007222670.1	131	33	*Nitrosospira multiformis* ATCC 25196	1
Microcystis aeruginosa PCC 9808	Cyanobacteria	WP_002792090.1	340 (667)	94	*Microcystis aeruginosa* NIES-44	Peptidase S8, unknown	WP_002766678.1	36	46	*Moorea producens*	1
Planktothrix PCC 8927	Cyanobacteria	AED99426.1	655	73	*Anabaena* sp. 90	Peptidase S8, unknown	AED99427.1	129	69	*Oscillatoria nigro-viridis* PCC 7112	4
Clostridium senegalense JC122 (DSM 25507)	Firmicutes	WP_026022449.1	640	46	*Oscillatoria* sp. PCC 10802	Peptidase S8, unknown	WP_010299410.1	56	41	*Paenibacillus terrigena*	2
Aurantimonas coralicida DSM 14790	Proteobacteria	WP_029702737.1	695	47	Hoeflea phototrophica	Peptidase S8, unknown	WP_024348656.1	70	36	*Azotobacter vinelandii*	1
Candidatus Competibacter denitrificans Run_A_D11	Proteobacteria	CDI04438.1	638	55	Mastigoclapsis repens	Peptidase S8, unknown	No BLAST hits	65	38	*Legionella longbeacheae*	1
Candidatus Entotheonella sp. TSY2	Proteobacteria	ETX08109.1	340	67	*Calothrix parietina*	Peptidase S8, thioredixin-like	No BLAST hits				2
Dickeya dadantii Ech586	Proteobacteria	WP_012886385.1	653	99	*Dickeya zeae*	Peptidase S8, unknown	WP_012886384.1	49	98	*Dickeya solani*	1
Dickeya zeae Ech1591	Proteobacteria	WP_012768058.1	654	92	*Dickeya zeae*	Peptidase S8, unknown	WP_012768059.1	49	98	*Dickeya dadantii*	1
Dickeya solani	Proteobacteria	WP_013319739.1	653	99	*Dickeya dadantii*	Peptidase S8, unknown	WP_022635150.1	49	98	*Dickeya dadantii*	1
Vibrio nigripulchritudo SOn1	Proteobacteria	WP_022611399.1	657	55	*Dickeya zeae*	Peptidase S8, unknown	WP_022611398.1	49	69	*Agarivorans albus*	2
Yersinia frederikseni ATCC 33641	Proteobacteria	WP_004708962.1	598	90	*Yersinia enterocolitica*	Peptidase S8, unknown	WP_004708955.1	49	85	*Agarivorans albus*	1
Yersinia enterocolitica subsp. enterocolitica 8081	Proteobacteria	WP_011816805.1	559	89	*Yersinia frederiksenii*	Peptidase S8, unknown	No BLAST hits				
Janthinobacterium sp. RA13 (JGI)	Proteobacteria	2585240732	637	50	*Calothrix* sp. PCC 6303	Peptidase S8, unknown	2585240733	58	54	*Aureispira* sp. CCB-QB1	1
Aureispira sp. CCB-QB1 (JGI)	Bacteroidetes	2587468798	645	41	*Oscillatoria* sp. PCC 10802	Peptidase S8, unknown	2587468797	57	53	*Janthinobacterium* sp. RA13	2
Acaryochloris marina	Cyanobacteria	YP_001514993.1	590	97	Acaryochloris sp.	Peptidase S8, unknown	YP_001514994.1	65	34	*Nematostella vectensis*	2
Tolypothrix sp. PCC 7910	Cyanobacteria	genome not public									2
Streptomyces scabiei 87.22	Actinobacteria	YP_003492207.1	290	85	*Streptomyces bottropensis* ATCC 25435	Peptidase S8, unknown	YP_003492205.1	66	83	*Streptomyces bottropensis* ATCC 25435	3

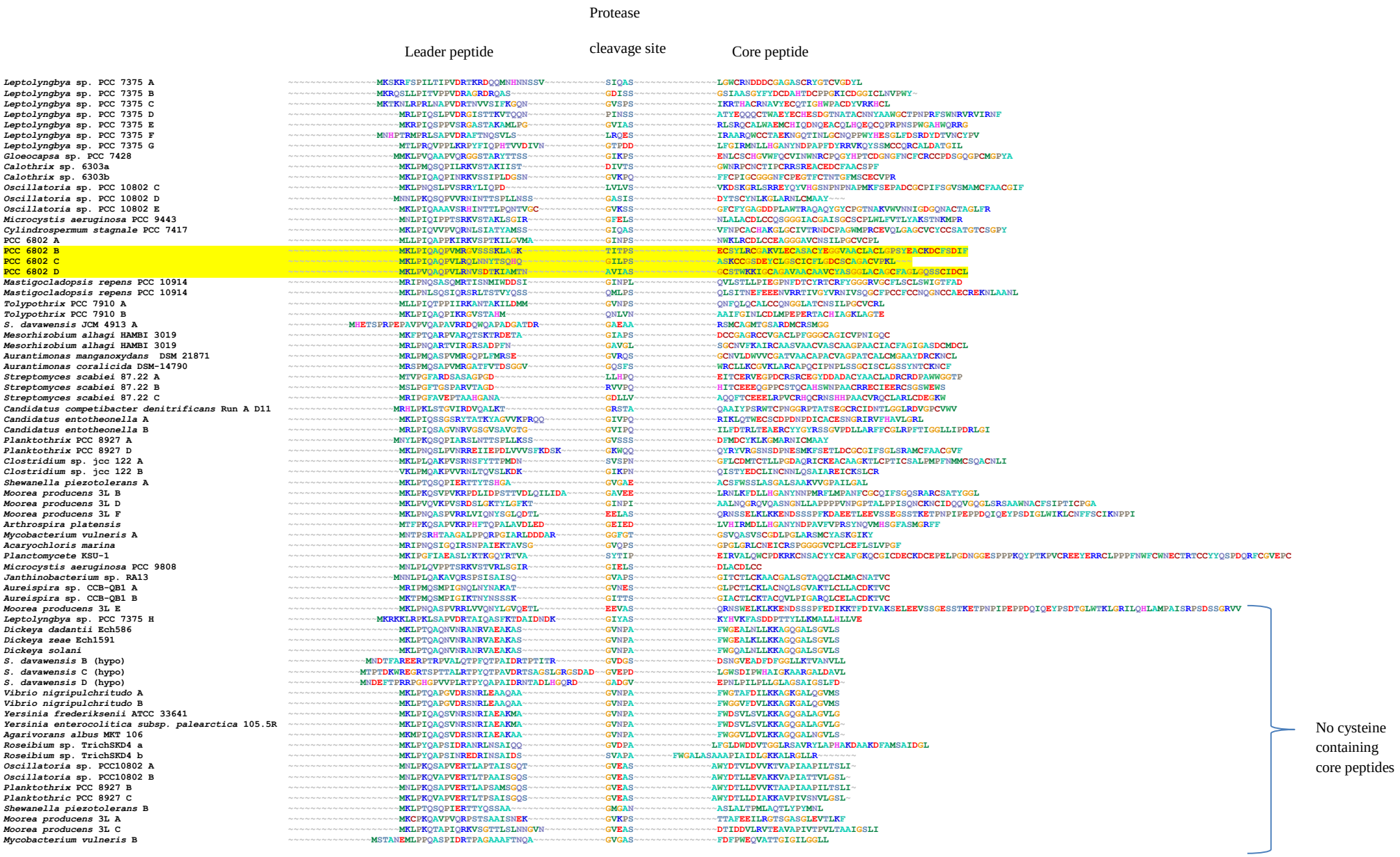


Figure 4.1.2.1. Cyanobactin-like precursor peptide alignment. Detected core peptides encoded in larger precursors are highlighted in yellow.

Table 4.1.2.2. Overview of the bacterial strains and their predicted core peptides in the study.

Strain	Length AA (core)	Cysteine (core)	Charge (core)	Exact mass	AA-sequence (core)	Application and/or result
Leptolyngbya sp. PCC 7375 A	61 (26)	4	-2	2769.1	LGWCRNDDDCGAGASCRYGTCVGDYL	
Leptolyngbya sp. PCC 7375 B	64 (35)	4	-3	3707.5	GSIAASGYFYDCDAHTDCPPGKICDGGICL NVPWY	
Leptolyngbya sp. PCC 7375 C	63 (31)	4	3	3670.76	IKRTHACRNAVYECQTIGHWPACDYVRKH CL	Synthetic peptide used for optimization
Leptolyngbya sp. PCC 7375 D	79 (51)	4	-1	5957.6	ATYEQQQCTWAEYECHESDGTNATACNN YAAWGCTPNPRFSWNRVRVIRNF	
Leptolyngbya sp. PCC 7375 E	73 (45)	4	1	5381.47	RLSRQCALWAEMCHIQDNQEACQLHQEQC QPRPNSPWGAHWQRRG	
Leptolyngbya sp. PCC 7375 F	77 (47)	4	-1	5418.45	IRAARQWCCTAEKNGQTINLGCNQPPWYH ESGLFDSRDYDTVNCYPV	
Leptolyngbya sp. PCC 7375 G	75 (44)	3	2	5020.42	LFGIRMNLLHGANYNDPAPFDYRRVKQYS SMCCQRCALDATGIL	
Leptolyngbya sp. PCC 7375 H	63 (26)	0	0	3044.64	KYHVKFASDDPTTYLLKMALLHLLVE	
Gloeocapsa sp. PCC 7428	82 (53)	10	-1	5873.31	ENLCSCHGVWFQCVINWNRCPQGYHPTCD GNGFNCFCRCCPDSGQGPCMGPYA	
Calothrix sp. PCC 6303 A	56 (29)	6	1	3291.37	GWNRPCNCTIPCRRSREACEDCFAACSPF	
Calothrix sp. PCC 6303 A	60 (32)	6	-1	3380.33	FFCPIGCGGGNFCPEGTFCTNTGFMSCECVP R	
Oscillatoria sp. PCC 10802 A	50 (25)	0	-1	2668.5	AWYDTVLDVVKTVAPIAAPILTSLI	
Oscillatoria sp. PCC 10802 B	52 (24)	0	0	2558.44	**AWYDTLLEVAKKVAPIATTVLGSL**	Antimicrobial activity (synthetic)
Oscillatoria sp. PCC 10802 C	50 (25)	2	1	2684.24	GASISDYTSCYNLKGLARNLCMAAY	Synthetic peptide used for optimization
Oscillatoria sp. PCC 10802 D	79 (55)	4	2	5969.75	VKDSKGRLSRREYQYVHGSNPNPNAPMKF SEPADCGCPIFSGVSMAMCFAACGIF	
Oscillatoria sp. PCC 10802 E	79 (49)	3	0	5243.39	GFCFYGAGDDPLAWTRAQAQYGYCPGTN AKVWVNNIGDGQNACTAGLFR	
Microcystis aeruginosa PCC 9443	72 (44)	6	2	4535.14	NLALACDLCCQSGGGIACGAISGCSCPLWL FVTLYAKSTNKMPR	
Cylindrospermum stagnale PCC 7417	80 (52)	10	1	5371.27	VFNPCACHAKGLGCIVTRNDCPAGWMPRC EVQLGAGCVCYCCSATGTCSGPY	
Pseudanabaena sp. PCC 6802 A	58 (29)	6	0	2993.35	NWKLRCDLCCEAGGGAVCNSILPGCVCPL	
Pseudanabaena sp. PCC 6802 B	72 (45)	8	-3	4718.01	**ECSYLRCGAKVLECASACYEGGVAACLA CLGPSYEACKDCFSDIF**	Detected
Pseudanabaena sp. PCC 6802 C	59 (32)	8	-1	3205.28	**ASKCCGSDEYCLGSCICFLGDCSCAGAC VPKL**	Detected

Pseudanabaena sp. PCC 6802 D	73 (45)	8	1	4216.85	**GCSTWKKIGCAGAVAACAAVCYASGGLACAGCFAGLGQSSCIDCL**	Detected
Mastigocladopsis repens PCC 10914 A	72 (44)	4	0	4862.31	QVLSTLLPIEGPNFDTCYRTCRFYGGGRVGCFLSCLSWIGTFAD	
Mastigocladopsis repens PCC 10914 A	82 (54)	8	-1	6087.75	QLSITNEFEEENVRRTIVGYVRNIVSQGCFPCCFCCNQGNCCAECREKNLAANL	
Tolypothrix PCC 7910 A	58 (30)	6	1	3169.44	QNFQLQCALCCQNGGLATCNSILPGCVCRL	
Tolypothrix PCC 7910 B	54 (30)	2	-2	3140.55	AAIFGINLCDLMPEPERTACHIAGKLAGTE	Synthetic peptide used for optimization
Streptomyces davawensis JCM 4913 A	59 (20)	2	2	2063.82	RSMCAGMTGSARDMCRSMGG	
Streptomyces davawensis JCM 4913 B	62 (23)	0	-3	2393.21	DSNGVEADFDFGGLLKTVANVLL	
Streptomyces davawensis JCM 4913 C	70 (23)	0	0	2416.3	LGWSDIPWHAIGKAARGALDAVL	
Streptomyces davawensis JCM 4913 D	66 (22)	0	-2	2206.22	EPNLPILPLLGLAGSAIGSLFD	
Mesorhizobium alhagi HAMBI 3019 A	58 (31)	8	0	2903.16	DCCGAGRCCVGACLPFGGGCAGICVPNIGQC	
Mesorhizobium alhagi HAMBI 3019 B	70 (45)	8	0	4241.84	SGCNVFKAIRCAASVAACVASCAAGPAACIACFAGIGASDCMDCL	
Aurantimonas manganoxydans DSM 21871	69 (43)	8	0	4233.86	GCNVLDWVVCGATVAACAPACVAGPATCALCMGAAYDRCKNCL	
Dickeya zeae Ech1591	49 (22)	0	2	2272.29	**FWGEALKLLKKAGQGALSGVLS**	Antimicrobial activity (synthetic)
Vibrio nigripulchritudo SOn1 A	49 (22)	0	2	2324.24	**FWGTAFDILKKAGKGALQGVMS**	Antimicrobial activity (synthetic)
Vibrio nigripulchritudo SOn1 B	49 (22)	0	2	2294.22	**FWGGVFDVLKKAGKGALQGVMS**	Antimicrobial activity (synthetic)
Agarivorans albus MKT 106	49 (22)	0	1	2228.23	**FWGGVLDVLKKAGQGALNGVLS**	Antimicrobial activity (synthetic)
Roseibium sp. TrichSKD4 B	53 (26)	0	3	2678.57	**FWGALASAAAPIAIDLGKKALRGLLR**	Antimicrobial activity (synthetic)
Yersinia frederiksenii ATCC 33641	49 (22)	0	1	2215.24	**FWDSVLSVLKKAGQGALAGVLG**	Antimicrobial activity (synthetic)

4.1.1 Phylogenetic analysis of proteases from candidate strains

The cyanobactin-like protease and the cyanobactin proteases were aligned using MUSCLE alignment and GBLOCKS curation. A PhyML phylogenetic tree was constructed in order to reconstruct the evolutionary history of cyanobactin genes and establish the relationships between cyanobactin and cyanobactin-like gene clusters. The tree shows complex evolutionary history and cyanobactin-like proteases show broader diversity than the cyanobactin proteases (PatA and PatG) (Figure 4.1.3). Majority of cyanobacteria can be differentiated as their own group highlighted in light grey, whereas another subgroup of proteases form a distinctive group highlighted in dark grey. Interestingly, cyanobactin proteases highlighted in yellow have clustered apart from each other (Figure 4.1.3).

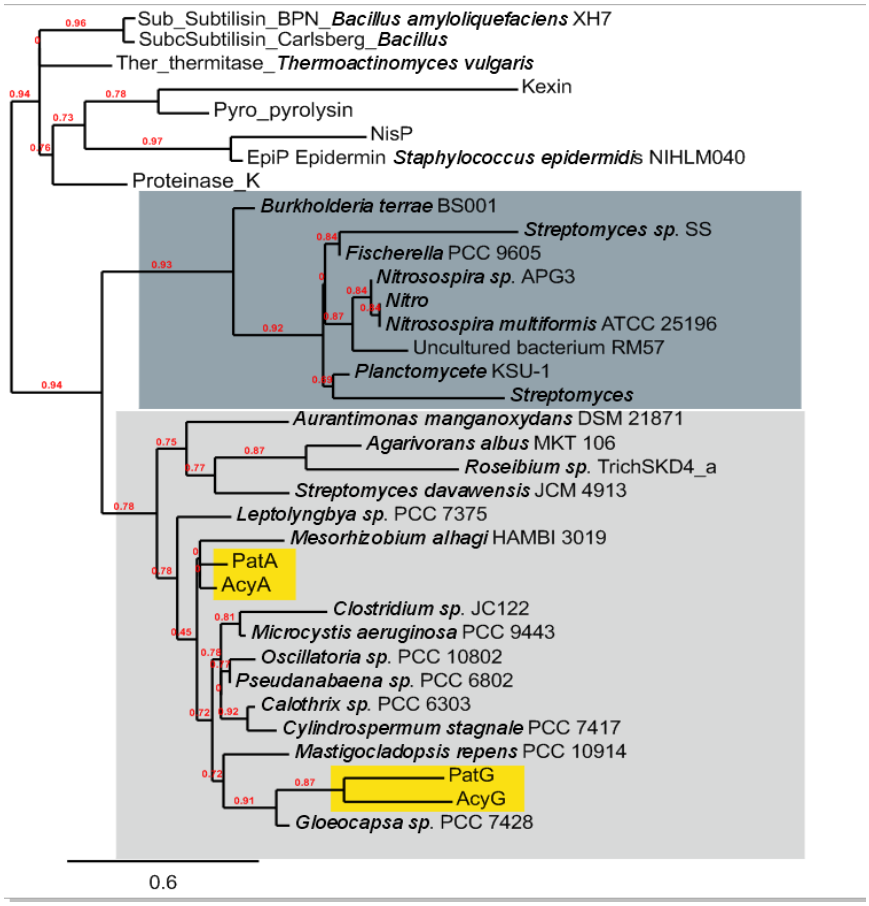


Figure 4.1.3. Phylogenetic tree showing the relationship between cyanobactin-like -and cyanobactin proteases. Cyanobactin proteases are highlighted in yellow.

4.2 Optimization for detection of synthetic peptide standards

Three peptides used for optimization were synthesized based on peptide predictions (Figure 4.2). These peptides contain 2–4 cysteines in their sequence and disulphide-bridge formation is supposed. Optimization for detection with synthetic peptides resulted that the best extraction method for bacterial samples is 50 % methanol. Other methods tested were pure methanol, pure water and 0.1 % formic acid. With these extraction methods, there were no differences between glass or plastic vials used for cyanobactin-like peptide detection. Detection limit was determined by dilution series to be 6–100 ng/ml, but more reliable limit was 60–100 ng/ml (Figure 4.2.1).

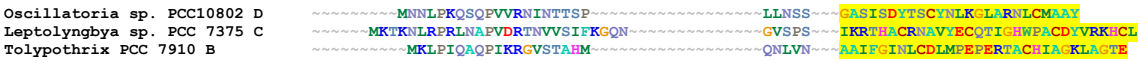


Figure 4.2. Synthetic peptides used for optimization highlighted in yellow.

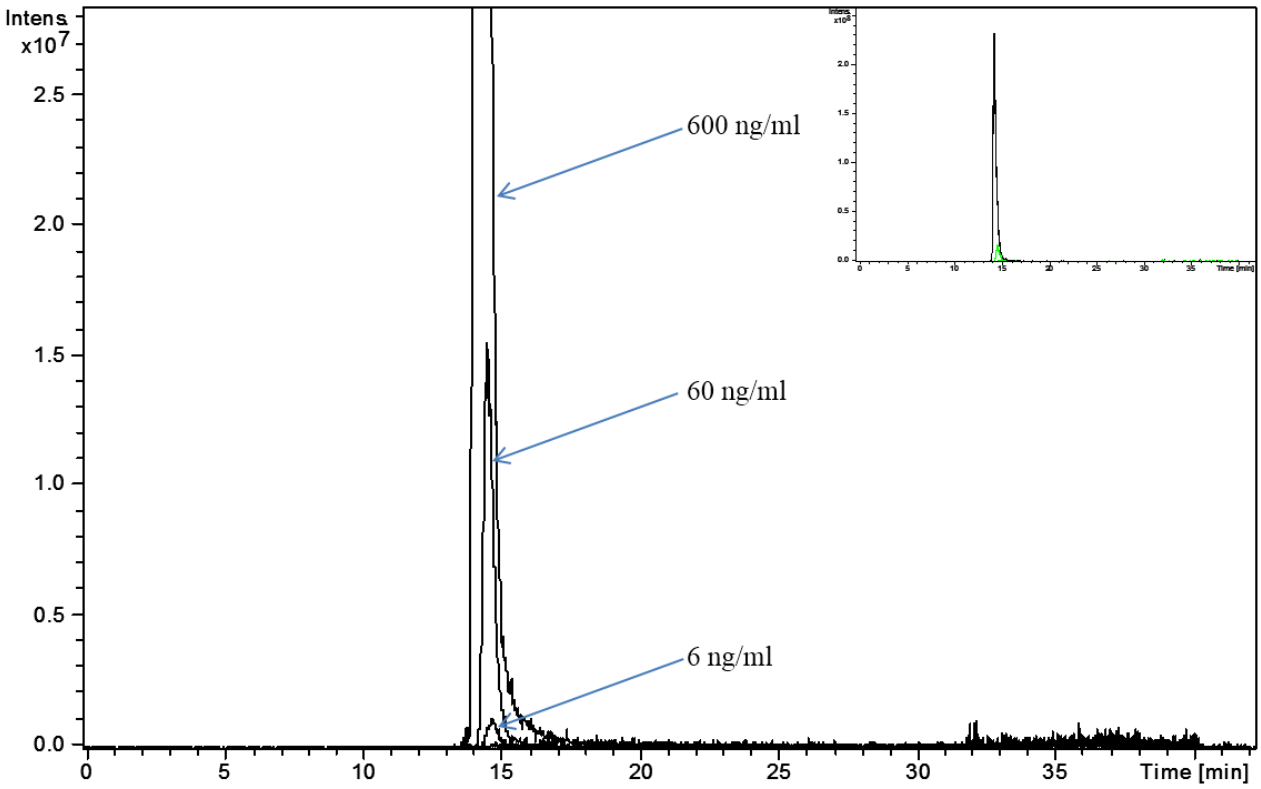


Figure 4.2.1. Dilution series of GASIS-peptide diluted in water.

4.3 Detection of the peptides from *Pseudanabaena* sp. PCC 6802 by mass using Q-TOF

Crude extract of *Pseudanabaena* sp. PCC 6802 was analyzed with Q-TOF. The flight times of peptide molecules and masses matched with predictions and *Pseudanabaena* sp. PCC 6802 produces these peptides (Figures 4.3 a-c). For all the detected peptides the error rate between theoretical m/z -values and detected m/z –values was ≤3.5 ppm. Detected masses for EGG-peptide were 1178.4963 and 1570.9902; for ASK-peptide 800.3132, 1066.7485 and 1599.6189; and for CAGA-peptide 1053.2001 and 1403.9377 (Table 4.3 and Figures 4.3 a-c).

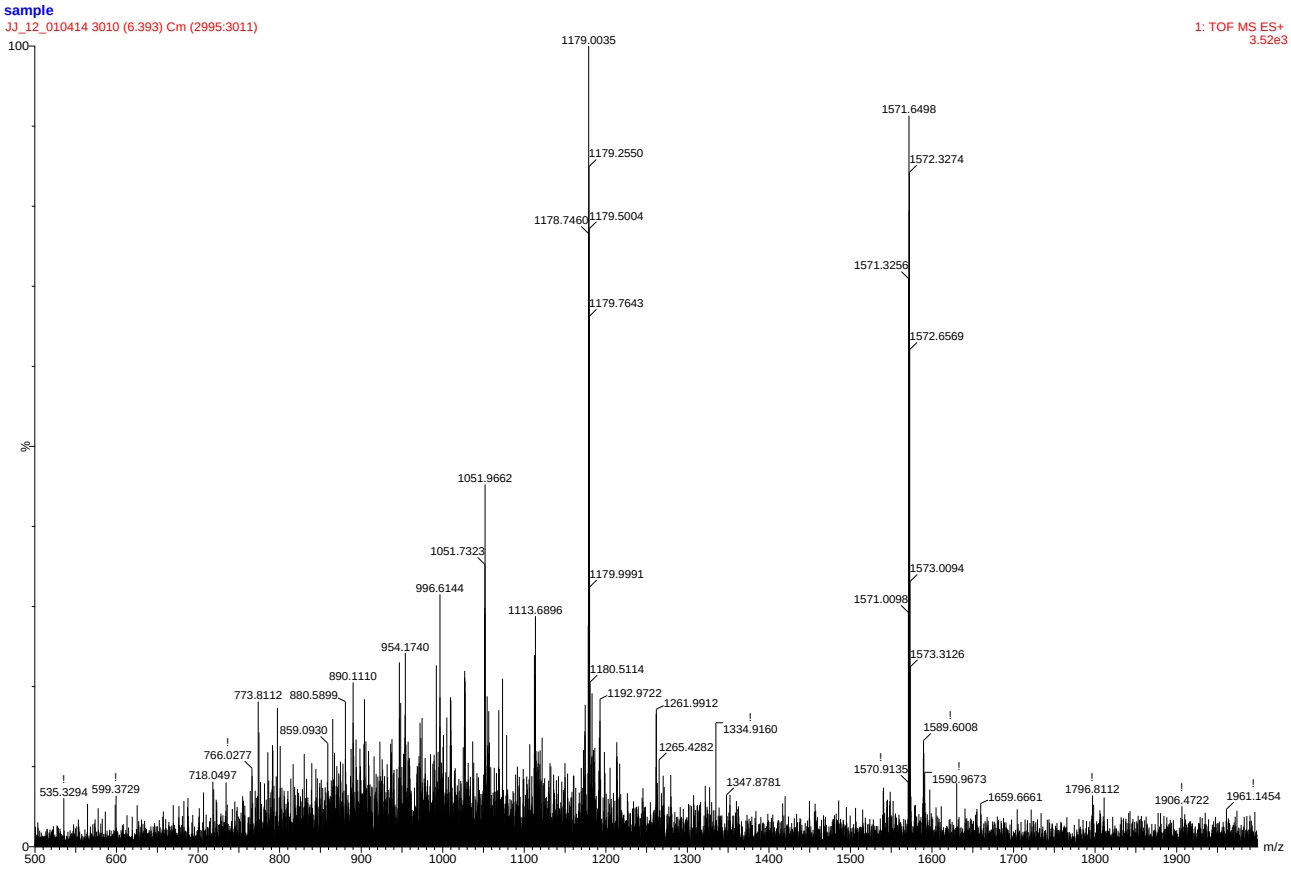


Figure 4.3 a. Q-TOF –profile of EGG-peptide. The two predominant peaks show the mass values of quadrapole (1178.4963 m/z) - and triple (1570.9902 m/z) charged forms.

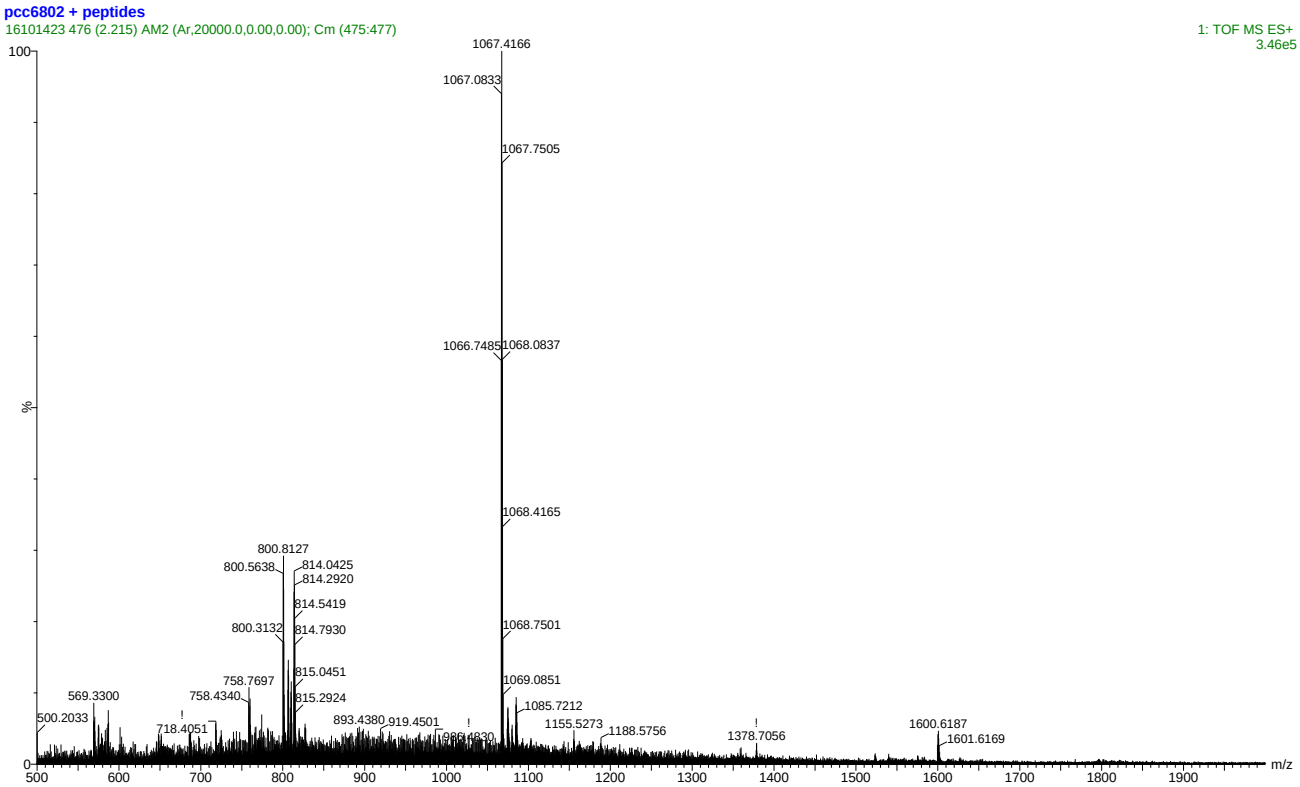


Figure 4.3 b. Q-TOF –profile of ASK-peptide. The predominant peak shows the mass value of triple (1066.7485 m/z) charged form.

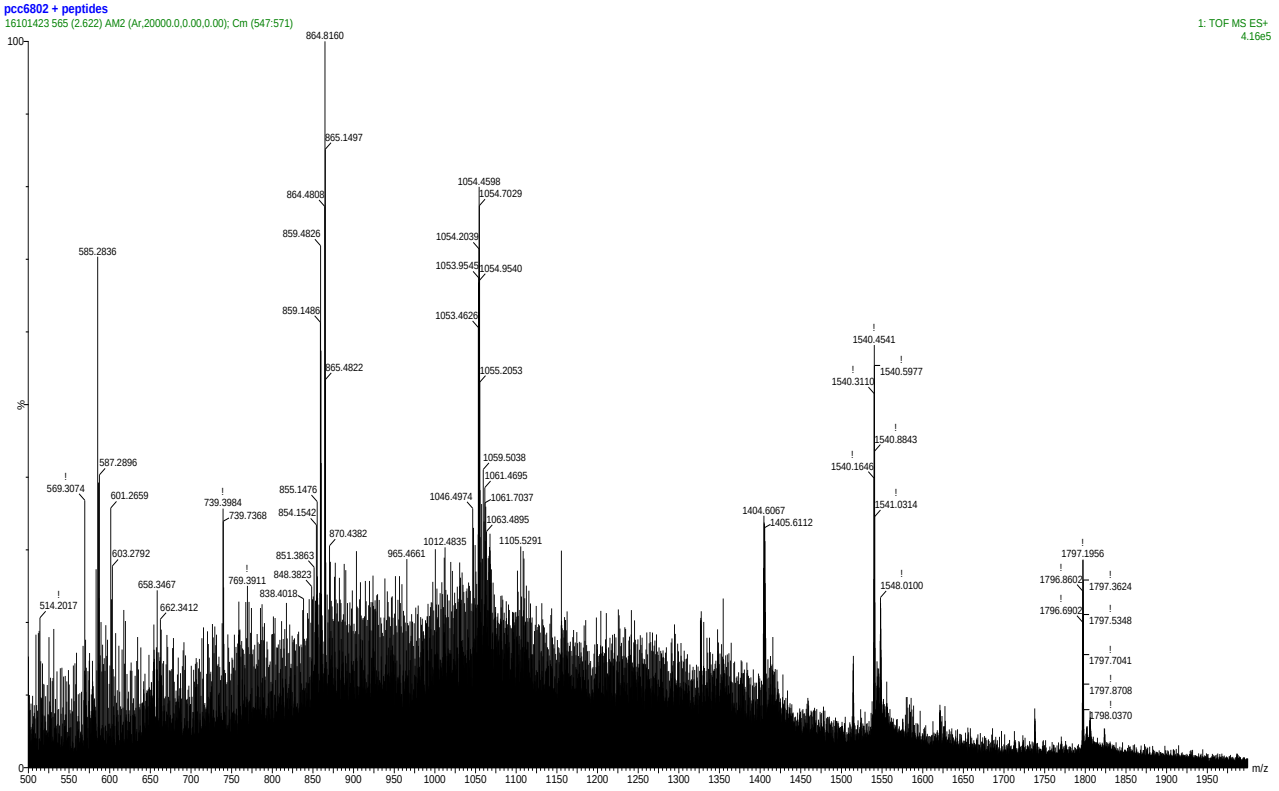


Figure 4.3 c. Q-TOF –profile of CAGA-peptide. The second predominant peak shows the mass value of quadrapole (1053.2001 m/z) charged form.

From every peptide their triple- and quadrapole charged forms were detected, and in addition double charged form from ASK-peptide (Table 4.3). Cyanobactin-like gene cluster in this strain contains four precursors from which three predicted core peptides were detected (Figure 4.3.1).

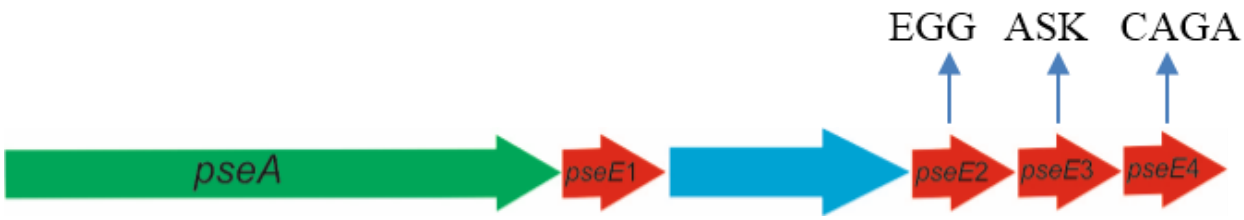


Figure 4.3.1. Cyanobactin-like gene cluster from *Pseudanabaena* sp**.** PCC 6802 illustrating the precursor genes from which predicted core peptides were detected.

Table 4.3. Cyanobactin-like precursor genes from *Pseudanabaena* sp**.** PCC 6802

Gene	Amino acid sequence (core peptide)	Elemental composition	Charge	Theoretical m/z	Detected m/z	Error (ppm)
*pseE*1	NWKLRCDLCCEAGGGAVCNSILPGCVCPL (Not detected)	$C_{125}H_{198}N_{36}O_{37}S_6$	5	598.4687		
			4	747.8339		
			3	996.7759		
			2	1494.66		
			1	2988.3121		
*pseE*2	ECSYLRCGAKVLECASACY**EGG**VAACLACLGPSYEACKDCFSDIF (Detected)	$C_{201}H_{300}N_{50}O_{65}S_8$	5	942.9973		
			4	1178.4946	1178.4963	1.4
			3	1570.9902	1570.9902	0.0
			2	2355.9814		
			1	4710.9551		
*pseE*3	**ASK**CCGSDEYCLGSCICFLGDCSCAGACVPKL (Detected)	$C_{130}H_{200}N_{34}O_{44}S_8$	5	640.4523		
			4	800.3134	800.3132	-0.2
			3	1066.7486	1066.7485	-0.1
			2	1599.619	1599.6189	-0.1
			1	3198.2302		
*pseE*4	GCSTWKKIG**CAGA**VAACAAVCYASGGLACAGCFAGLGQSSCIDCL (Detected)	$C_{176}H_{273}N_{49}O_{55}S_8$	5	842.7646		
			4	1053.2038	1053.2001	-3.5
			3	1403.9357	1403.9377	1.4
			2	2105.3997		
			1	4209.7916		

4.4 Sulphur labelling

The *pse* gene cluster containing *Pseudanabaena* sp. PCC 6802 was grown in media including normal sulphur ^{32}S- isotope and ^{34}S-isotope in order to find cysteine-rich peptides by LC-MS. Sulphur containing EGG-peptide was found by comparing ^{34}S-labeled LC-MS profile to the LC-MS profile of the unlabeled one. The aminoacid sequence of the EGG-peptide should contain eight cysteine residues. In the presence of ^{34}S, two sulphur-atoms are attached to each cysteine. This gives 16 mass unit addition to mass. The compound with bigger mass can be seen in the presence of ^{34}S, but not in the presence of ^{32}S (Figure 4.4).

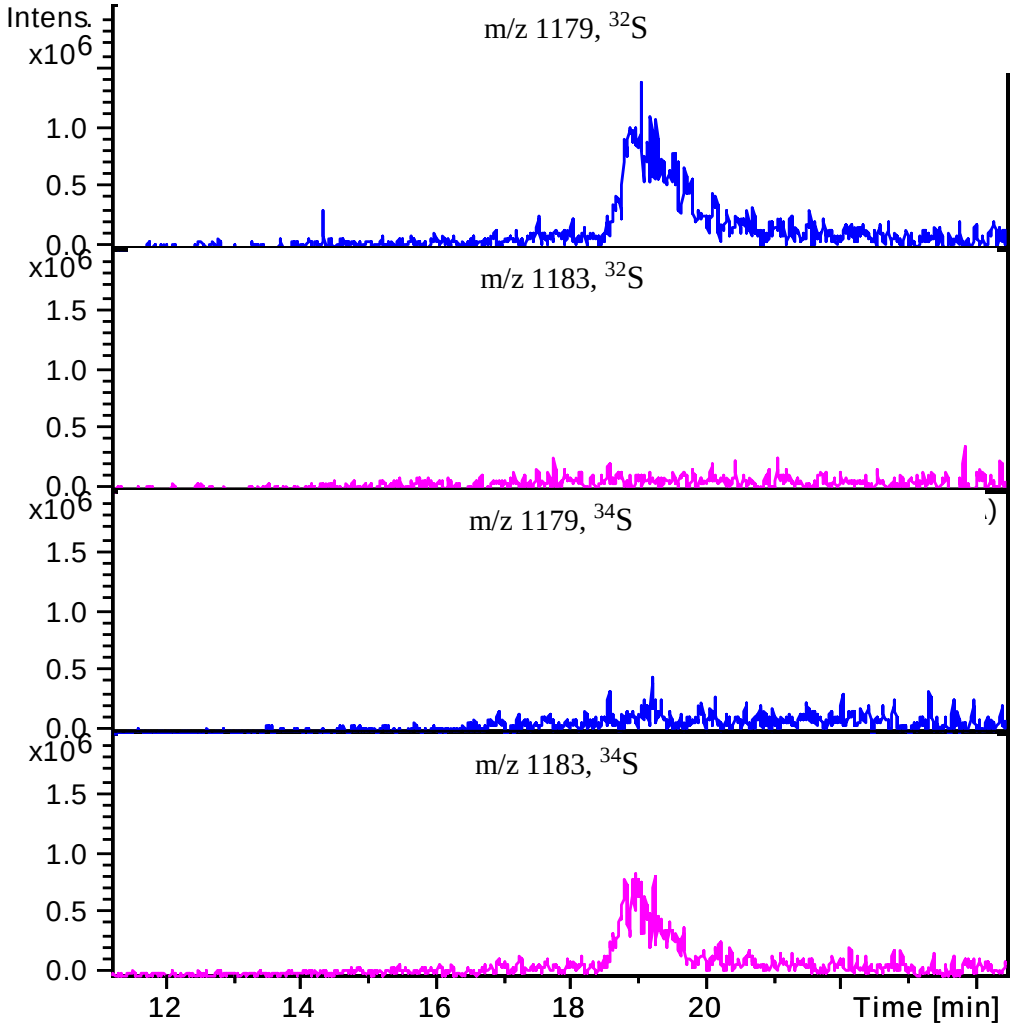


Figure 4.4. LC-MS –profile of the quadrapole charged EGG-peptide between PCC 6802 ^{32}S (1179 m/z) and PCC 6802 ^{34}S (1183 m/z).

4.5 Heterologous expression of the *pse* gene cluster in *Escherichia coli*

In order to support the result of finding predicted peptides by Q-TOF and sulphur labeling, the *pse* gene cluster from *Pseudanabaena* sp. PCC 6802 was amplified by PCR with specific primers, restricted with restriction enzymes, subcloned to two vectors and heterologously expressed in *E. coli*. The *pse* gene cluster from *S. davawensis* JCM 4913 was cloned as well. The 3485 bp long gene cluster from *Pseudanabaena* sp. PCC 6802 was successfully cloned and expressed in *E. coli* BL21-cells containing the genes needed for peptide expression. The quadrapole charged EGG-peptide in *E. coli* BL21-cells was detectable at m/z 1179.2 (Figure 4.5). Both of the vectors, high copy number plasmid Puc57 and low copy number plasmid Pacyc84 (data not shown), contained T7-promoter and ribosomal binding site and they both worked. Five parallel colonies were selected to the cloning procedure to make sure that at least one colony features the *pse* gene cluster and to compare expression levels between clones. All of the parallel colonies included the gene cluster, but in sample No. 2 the expression level of the gene product (EGG-peptide) was largest. Expression level in sample No. 2 is about six times larger compared with the *Pseudanabaena* sp. PCC 6802 -ctrl and expression level is about five times larger with IPTG-induced sample No. 2 compared with the crude extract *Pseudanabaena* sp. PCC 6802 -ctrl (Figure 4.5). Only EGG-peptide was detected from clone sample, ASK -and CAGA-peptides were not detected. The *pse* gene cluster from *S. dawavensis* JCM 4913 was not detected either. Surprisingly, there was no noteworthy enhancement of expression levels in any of the samples treated with IPTG-induction (Figure 4.5).

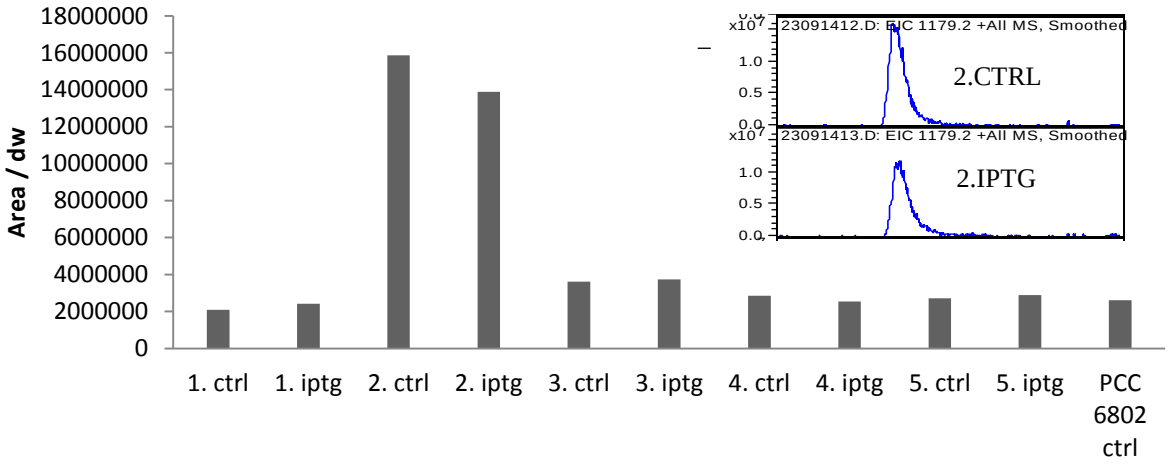


Figure 4.5. Heterologous expression of *Pseudanabaena* sp. PCC 6802 gene cluster in Puc57-plasmid in *E. coli* BL21-cells for quadrapole charged EGG-peptide at m/z 1179.2.

Clone samples, which expression levels of peptide were largest (2. CTRL and 2. IPTG), and *Pseudanabaena* sp. PCC 6802 crude extract were also analyzed by Q-TOF. Expression levels of EGG-peptide are larger in clone samples compared to crude extract sample (Figure 4.5.1).

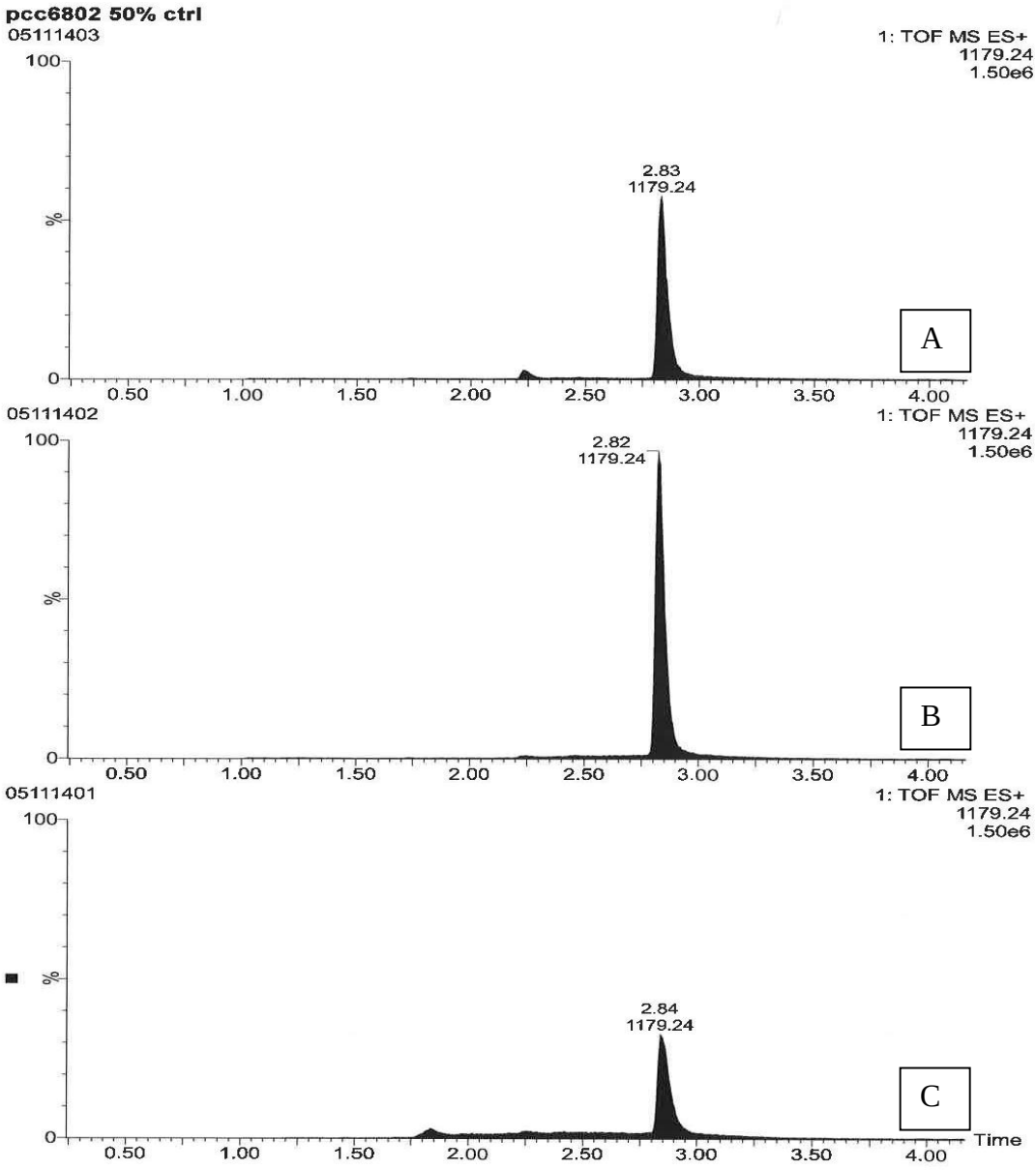


Figure 4.5.1. Q-TOF-profile of the quadrapole charged EGG-peptide from *Pseudanabaena* sp. PCC 6802 at m/z 1179.24. Sample A represents IPTG induced clone sample expressed in *E. coli* BL21-cells, sample B represents non-IPTG induced clone sample and sample C is *Pseudanabaena* sp. PCC 6802 crude extract.

4.6 Antimicrobial assays

Detection of the peptides from *Pseudanabaena* sp. PCC 6802 prove that protease cleavage site of the precursor was predicted right, confirmed by Q-TOF and sulphur labeling. Heterologous expression demonstrate that the boundaries of the gene cluster were correctly defined. This information mean that it was possible to accurately predict the cleavage site for all precursor peptides identified in the bioinformatics analysis. Further examination by bioinformatic analysis revealed a smaller group of cyanobactin-like peptides from various bacterial genomes which contain no cysteine in their core sequences and they resemble known α-helical antimicrobial peptides by their structure (Figure 4.6). Antimicrobial activity of these novel peptides was so hypothesized. Some of these peptides were synthesized and tested for their bioactivity.

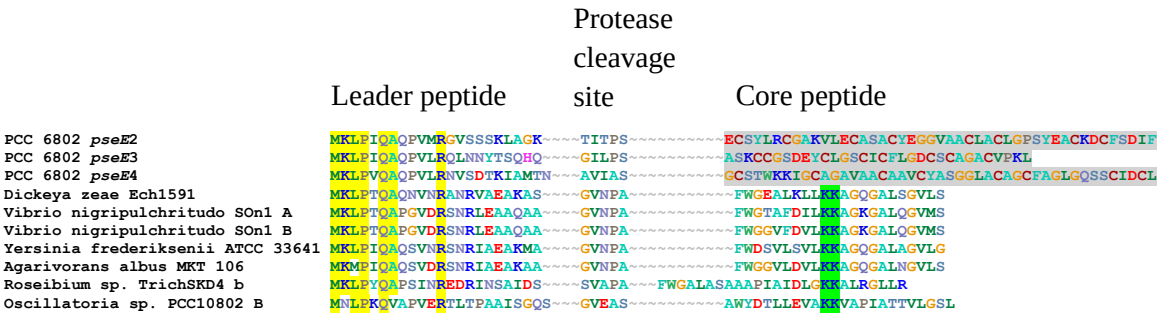


Figure 4.6. Amino acid sequence alignment between detected cysteine containing peptides from *Pseudanabaena* sp. PCC 6802 (highlighted in grey) and predicted core peptides with antimicrobial activity of the same family from different bacterial strains with double lysine-motif as a connective factor (highlighted in green). Conserved residues in the leader peptide are highlighted in yellow.

I had 11 synthetic analogs of peptide predictions (Table 3.6) from which seven were discovered to have antimicrobial activity (Table 4.7). Peptides from *Oscillatoria* sp. PCC 10802 (B), *Dickeya zeae* Ech1591 (D) and *Vibrio nigripulchritudo* SOn B (V) were shown to be active against Gram-negative and Gram-positive bacteria in disk diffusion assay (Table 4.6 and figure 4.6.1 a-f). Synthetic peptide A dissolved poorly in 50 % methanol and manifested no antimicrobial activity against tested strains, nor did the synthetic cysteine containing ones (GASIS, NAVY, AAIF) (Table 4.6). *Pseudomonas* sp. *and Pseudomonas aeruginosa* appeared to be resistant to tested peptides. Peptide A has a charge -1, peptide B has neutral charge and peptides D and V carry charge +2. The two cationic peptides (D and V) exhibited strong antimicrobial activities in contrast to the neutral peptide B and negatively charged peptide A.

There was no detectable amendment in antimicrobial activity when peptides A and B were used in tandem. For peptides D and V the results were the same so the experiment was reproductive.

Table 4.6. Antibacterial activity of peptides against various bacterial strains. Number stands for radius of inhibition in mm.

Bacteria (HAMBI nr.)	**Gram**	**Gasis**	**Navy**	**Aaif**	**PCC 10802 A**	**PCC 10802 B**	***Dickeya* (D)**	***Vibrio* (V)**	**+ ctrl**	**- ctrl**
Pseudomonas sp.(2796)	-	-	-	-	-	-	-	-	17	-
Pseudomonas aeruginosa (25)	-	-	-	-	-	-	-	-	12	-
Escherichia coli (396)	-	-	-	-	-	5	6	6	18	-
Burkholderia pseudanalbei (33)	-	-	-	-	-	-	5	4	15	-
Salmonella typhi (1306)	-	-	-	-	-	-	6	6	16	-
Xanthomonas campestris (104)	-	-	-	-	-	-	4	5	18	-
Burkholderia capacia (2487)	-	-	-	-	-	-	4	-	17	-
Bacillus cereus (1881)	+	-	-	-	-	4	-	-	18	-
Micrococcus luteus (2688)	+	-	-	-	-	5	6	6	20	-
Staphylococcus aureus (11)	+	-	-	-	-	-	4	3,5	15	-
Bacillus subtilis (251)	+	-	-	-	-	-	5	6	19	-

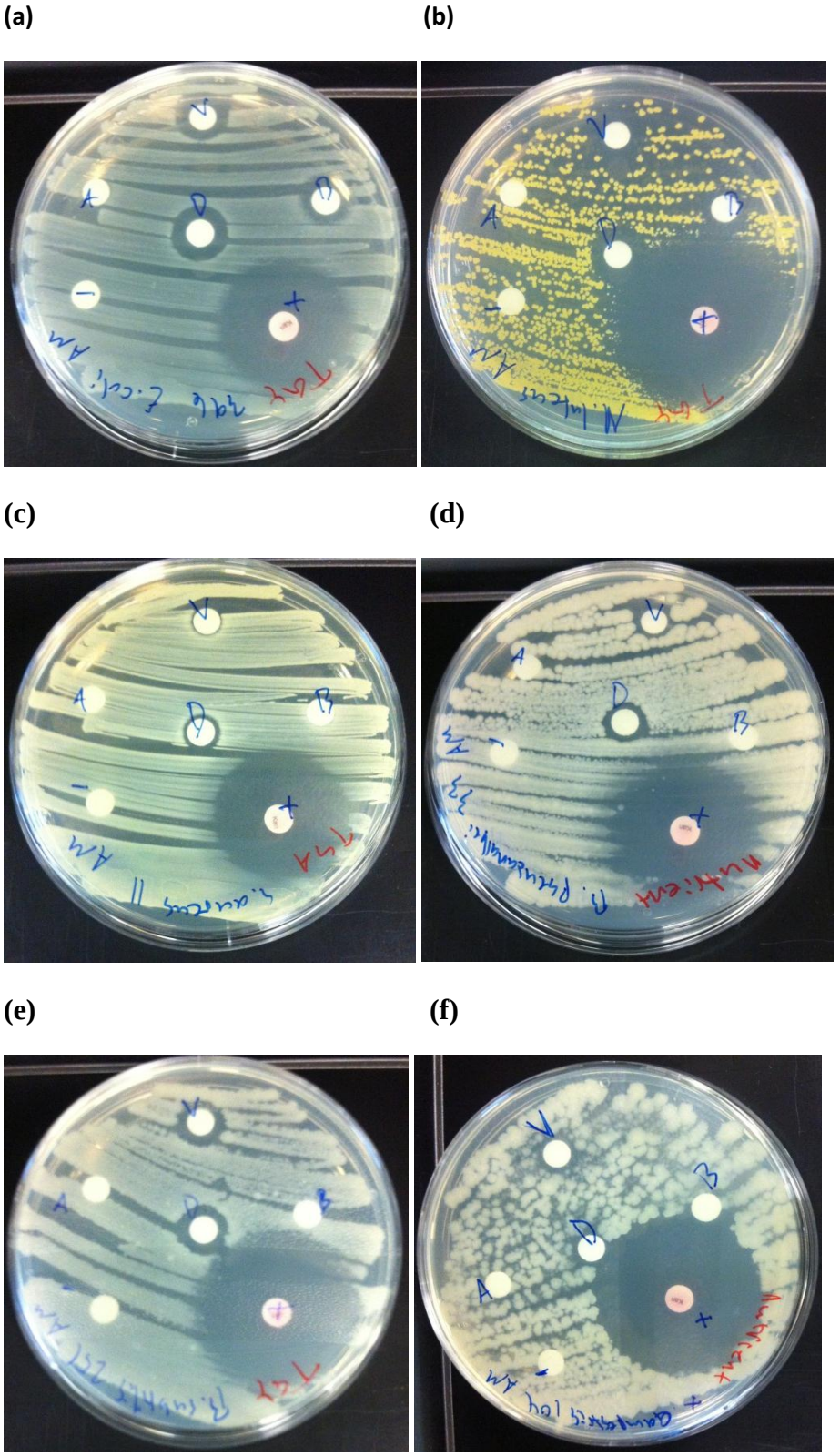


Figure 4.6.1. Disk diffusion assay illustrating the bioactivities of synthetic peptides against various bacterial strains. A and B are the peptides from *Oscillatoria* sp. PCC 10802, D is the peptide from *Dickeya zeae* ech1591 and V is the peptide from *Vibrio nigripulchritudo* SOn1. (a) *E. coli* (396), (b) *M. luteus*. (2688), (c) *S. aureus*. (11), (d) *B. pseudanalbei* (33), (e) *B. subtilis.* (251), (f) *X. campestris.* (104). Mass of each synthetic peptide/disk was 60 µg, negative control was methanol and positive control was kanamycin (1000 µg/disk).

4.6.1 Antimicrobial peptide database results

Amino acid sequences of the seven antimicrobial peptides were used to query the database at antimicrobial peptide database (Wang et al. 2009). The database search resulted in many hits from which the closest are reported in Table 4.6.1.

Table 4.6.1. Antimicrobial peptides in the study used to query known antimicrobial peptides from database.

Producer strain of peptide	**Peptide (Similarity percentage)**	**Origin**	**Bioactivity**	**Reference**
Dickeya zeae ech1591	Macropin 1 (47.82 %)	Bee venom	*Against B. subtilis, E. coli, S. aureus* and *C. albicans*	Monincová et al. (2014).
Vibrio nigripulchritudo SOn1 A	Magainin-MW1 (48 %)	Frogs, amphibians and animals	Against *E. coli*, *S. aureus* and *C. albicans*	Mechkarska et al. (2011).
Vibrio nigripulchritudo SOn1 B	Plasticin PD36K (48 %)	Synthetic	Against *S. aureus*, *B. megaterium, E. coli* and *E. cloacae*	El amri et al. (2006).
Yersinia frederiksenii ATCC 33641	Peptide PGQ (50 %)	Frogs, amphibians and animals	Against *E. coli, K. pneumonia, P. aeruginosa, S. aureus, S. faecalis* and *C. albicans*	Moore et al. (1991).
Agarivorans albus MKT 106	Peptide PGQ (51.85 %)	Frogs, amphibians and animals	Against *E. coli, K. pneumonia, P. aeruginosa, S. aureus, S. faecalis* and *C. albicans*	Moore et al. (1991).
Roseibium sp. TrichSKD4 B	Brevinin-1Ja (46 %)	Frogs, amphibians and animals	Against *S. aureus*, *E. coli* and mammalian COS7-cells	Koyama and Iwamuro (2008).
Oscillatoria sp. PCC10802 B	Dermaseptin-H13 (44.44 %)	Frogs, amphibians and animals	Against *S. aureus, P. aeruginosa*, and *E. coli*	Navon-Venezia et al. (2002).

4.7 MIC (minimum inhibitory concentration)

All the peptides, which had detected antimicrobial activity in disk diffusion assay, were tested for their MIC-values against Gram-positives *S. aureus, B. subtilis, M. luteus* and Gram-negative *E. coli*. Additionally four more peptides were ordered synthetic and these were from *Agarivorans albus* MKT 106, *Roseibium* sp. TrichSKD4, *Yersinia frederiksenii* ATCC 33641 and another peptide of the same gene cluster from *Vibrio nigripulchritudo* SOn1. Also the peptides from *Dickeya zeae* Ech1591 and *Vibrio nigripulchritudo* SOn1 were ordered with higher purity (98 %) in comparison with the purity (75 %) that was used in disk diffusion assay. Table 4.7 summarizes the MIC- values of the peptides used in this study, which are between 0.78-100 µg/ml. Peptide A from *Oscillatoria* sp. PCC 10802 and the synthetic cysteine

containing ones (GASIS, NAVY, AAIF) were not included in the test, because they showed no activity in disk diffusion assay. There was no detectable amendment in MIC-values when both peptides from *Vibrio nigripulchritudo* SOn1 were used in tandem. All the peptides are cationic with a net charge ranging between 1–3, except the peptide from *Oscillatoria sp.* PCC10802 B which had a net charge of 0. MIC-test was done in order to find out the lowest concentration of peptide inhibiting the growth of bacteria.

Table 4.7. Minimum inhibitory concentrations (MIC) of the synthetic antimicrobial peptides

Strain (Antimicrobial peptide)	**Peptide sequence**	**MIC (µg/ml)**			
		S. aureus	***B. subtilis***	***E. coli***	***M. luteus***
Agarivorans albus MKT 106 (85.7 %)	FWGGVLDVLKKAGQGALNGVLS	ND	ND	ND	50–100
Dickeya zeae Ech1591 (98.8 %)*	FWGEALKLLKKAGQGALSGVLS	25–50	12.5–25	12.5–25	3.125–6.25
Roseibium sp. TrichSKD4 b (87.4%)	FWGALASAAAPIAIDLGKKALRGLLR	6.25–12.5	3.125–6.25	25–50	0.78–1.5625
Vibrio nigripulchritudo SOn1 A (98.1 %)*	FWGTAFDILKKAGKGALQGVMS	ND	12.5–25	25–50	3.125–6.25
Vibrio nigripulchritudo SOn1 B (98.1 %)*	FWGGVFDVLKKAGKGALQGVMS	25–50	6.25–12.5	25–50	1.5625–3.125
Yersinia frederiksenii ATCC 33641 (96.4%)	FWDSVLSVLKKAGQGALAGVLG	50–100	25–50	ND	12.5–25
Vibrio nigripulchritudo SOn1 A + B*	FWGTAFDILKKAGKGALQGVMS + FWGGVFDVLKKAGKGALQGVMS	25–50	6.25–12.5	25–50	1.5625–3.125
Oscillatoria sp. PCC 10802 B (75%)	AWYDTLLEVAKKVAPIATTVLGSL	ND	ND	ND	12.5–25

*= Three peptides were synthesized with ~98 % purity, meaning that the net content of the peptide is 50 % according to the manufacturer (GenScript). Then the actual MIC-value would be half of the actual observed values (reported in the table). ND = not determined.

5. DISCUSSION

5.1 Novel antibacterial peptidefamily

Here I report a new group of antimicrobial peptides from bacteria. Cyanobactin-like gene clusters encoding peptides have been identified in the genomes of bacteria spanning six phyla.

Cyanobactins have been proven to be widespread among cyanobacteria (Leikoski et al. 2009), but the presence of cyanobactin gene clusters outside cyanobacterial phyla has not yet been proven. Our study suggests that the putative cyanobactin-like gene clusters are widespread in cyanobacteria and proteobacteria. Another as widely distributed group of peptides are microcins (Lee et al. 2008). Gene clusters encoding ribosomally produced microcin (produced by Gram-negative bacteria) homologs have been identified in the genomes of archaea as well as bacteria covering also six phyla, including cyanobacteria, actinobacteria, and proteobacteria (Lee et al. 2008). The widespread occurrence between various bacterial phyla of gene clusters encoding antimicrobial peptides suggest that they play valuable roles in microbial biology and physiology (Claesen and Bibb 2010).

Disulphide-bridge containing peptides and α-helical peptides are widely distributed in multicellular organisms with great diversity (Tossi et al. 2000; Zasloff 2002). Antimicrobial peptides with disulphide-bridges from bacteria are not well known, but some peptides with these characteristics are published (Duquesne et al. 2007; Singh et al. 2012; O'shea et al. 2013). Peptides with α-helical structure are not well known from bacteria either, except cecropins from *Helicobacter pylori* (Pütsep et al. 1999). Previously genome mining has revealed various structurally complex peptides with antimicrobial activity from different sources including micro-organisms and higher organisms (Challis 2008). These findings indicate that bacteria produce eukaryotic-like antimicrobial peptides.

5.2 Biosynthesis of cyanobactin-like peptides

The results presented here demonstrate that the gene clusters encoding cyanobactin-like peptides contain single protease and 1–8 precursor genes. Cyanobactin gene clusters by contrast consist of two proteases and 1–10 precursor genes (Lee et al. 2009; Donia and Scmidt 2010; Sivonen et al. 2010; Leikoski et al. 2013). Post-translational modifications for cyanobactin-like peptides include proteolytic cleavage and formation of sulphur bridges. In general the biosynthetic machinery of cyanobactin-like peptides seems less complex compared to cyanobactin machinery, but there could be post-translational enzymes in some of the gene clusters that have not been noticed. Genome mining can be targeted to tailoring (modifying) enzymes that give core peptide its exotic structure. This extended mining might give insights to genes encoding other modifying functions, e.g. methyltransferases, hydroxylases (Baltz 2008), export enzymes for external secretion of the peptide (Duquesne et al. 2007) and transferases that enact the transfer of some functional groups inside the compound. This would change the molecular weight of the peptide and new calculations might be needed for natural product detection by chemical analysis. So even if these compounds exist according to predictions and are being produced, they would not be detectable with false calculations. The detected cyanobactin-like peptides are found inside the cells, which supports the information that the cyanobactin-like gene cluster in *Pseudanabaena* sp. PCC 6802 does not encode an enzyme for external secretion of the peptide(s), but this does not necessarily apply to other candidate strains. Maybe some portion of these peptides could be found from the extracellular space. Thus it could be possible that in some of cyanobactin-like gene clusters there are other post-translational enzymes which modify the peptide structure.

The results presented here provide some insights into the evolutionary origins of the cyanobactin pathway. Phylogenetic analysis shows that the cyanobactin pathway may have emerged through fusion of two cyanobactin-like gene clusters (Figure 4.1.3). Natural product variety is driven among others by horizontal gene transfer, gene duplications, gene or partial gene deletions, replacements and homologous recombination (Baltz 2008). Horizontal gene transfer is likely to happen with cyanobactin-like peptides, because they are widespread between bacterial phyla. It is still unclear are the genes needed for cyanobactin-like peptide biosynthesis encoded in plasmids or chromosomally. They could be encoded either by plasmids or by the chromosome, like microcins, which are produced by enterobacteria and belonging to the large class of bacteriocins (Duquesne et al. 2007).

The detected cyanobactin-like peptides from *Pseudanabaena* sp. PCC 6802 are linear by structure. Most likely all cyanobactin-like peptides are linear, because they lack the macrocyclization motif in the precursor peptide. Another factor supporting linear structure of cyanobactin-like peptides is a single protease in the gene cluster. Presence of another protease would suggest cyclic structure of the compound, as it is with cyanobactins (Lee et al. 2009; Donia and Scmidt 2010; Sivonen et al. 2010).

5.3 Detection of new peptides and heterologous expression

Crude extracts from each candidate strain were screened by chemical analysis. Additionally, cyanobactin-like gene cluster from *Pseudanabaena* sp. PCC 6802 was cloned and heterologously expressed in *E. coli* in order to obtain further proof that the cyanobactin-like gene cluster was responsible for the production of the disulphide-bridge stabilized peptides detected from crude extracts. Three predicted core peptides out of four were detected from *Pseudanabaena* sp. PCC 6802 crude extract by Q-TOF. One of these peptides was detected from the *E. coli* -clone sample expressing the cyanobactin-like gene cluster by LC-MS and Q-TOF, which proved that the gene cluster is active and the peptide is being produced. The rest of the predicted core peptides from remaining candidate strains stayed yet undetected.

Trunkamide, anacyclamides, piricyclamides and patellamide are ribosomal peptides from cyanobacteria which all have been successfully cloned and expressed in *E. coli* (Schmidt et al. 2005; Donia et al. 2008; Leikoski et al. 2010; Leikoski et al. 2012). The *tru* gene cluster was used to produce peptide variants by introducing mutations into the *truE* gene and then expressing them with the rest of the pathway in *E. coli* (Tianero et al. 2012). In conclusion, there are not that many published expressions in heterologous host of ribosomally produced natural products from cyanobacteria (Micallef et al. 2014). Cyanobactin-like peptide (EGG) form an adequate addition to this group of cloned secondary metabolite gene clusters based on our results.

Many natural product gene clusters remain silent or are not well expressed under standard laboratory conditions (Winter et al. 2011). It would be really hard to produce, identify and test the bioactivity of gene product if the genes predicted using bioinformatics are inactive

(Scheffler et al. 2013). This issue can be resolved by altering the growth conditions of micro-organism to activate the relevant genes (Scherlach and Hertweck 2006). The growth conditions for candidate strains were not altered during this experiment. In order to detect peptides, these gene clusters could be cloned and heterologously expressed in *E. coli*. Heterologous expression is a potential method for gene clusters that are silent in their natural host under laboratory conditions (Challis 2008). Secondary metabolites lacking in the host organism but present in the heterologous host containing the cloned gene cluster could be a proof of active natural product production (Zerikly and Challis 2009). Thus, the gene clusters could turn active once cloned and the predicted natural product might be produced. On the other hand, gene clusters might be active in producer organisms, but the expression levels of the peptides are low and undetectable by chemical analysis in that sense. Expression level of detected EGG-peptide is quite large in the clone sample so it might overpower the other peptides by its relatively high expression level. Biological explanations for nondetection can be among others that the peptides are transported out of the cell, they do not exist and/or protease does not cleave every precursor. Consequently, gene clusters encoding cyanobactin-like peptides are maybe inactive in most of the candidate strains.

S. davawensis JCM 4913 was chosen to be cloned as well in order to prove peptide expression from another bacterial strain. However, the predicted peptides were not detected from this bacterium. Usually heterologous expression is a powerful method for exploring bacterial secondary metabolites, but inadequacies in background metabolism and different codon usage might alter pathways from the original organism incompatible with the heterologous host (Clardy et al. 2006). *S. davawensis* JCM 4913 has high GC-content in its genomic DNA. This affects codon usage in cloning which might have resulted in poor detection by chemical analysis. The cyanobactin-like gene cluster from *S. davawensis* JCM 4913 includes one V8-like protein Glu-specific endopeptidase-like protein. It is not known does this gene have some specific role in the protein expression. New heterologous hosts could expand the screening efforts to access novel natural products (Clardy et al. 2006).

5.4 Antimicrobial characteristics of cyanobactin-like peptides

Some peptides from smaller group of cysteine-free cyanobactin-like peptides originating from different bacterial strains were synthesized and tested for their bioactivity. Cyanobactin-like peptides without cysteines in their core sequence resemble known α-helical amphiphilic antimicrobial peptides by their structure, so antimicrobial activity for this smaller group of cyanobactin-like peptides was supposed. This hypothesis was proven by seven synthetic peptides manufactured according to predictions. Amphiphilic properties can be seen by the alternating order of hydrophobic and hydrophilic amino acids in the core sequence (Tossi et al. 2000). When the compound is laterally displayed, the negatively charged ends are on the other side and the positively charged ends on the opposite side, suggesting that the peptide is able to adapt α-helical form at the moment of interaction with bacterial membranes (Tossi et al. 2000; Brogden 2005). Currently, the mode of action of cationic hydrophobic antimicrobial peptides is assumed as follows: 1) cationic peptides are electrostatically attracted to negatively charged microbial surfaces; 2) antimicrobial peptides interact with the negatively charged phospholipid head groups of the microbial cytoplasmic membrane; and 3) an amphipathic helical conformation may be adapted, which allows peptides to insert the hydrophobic face into the bilayer (Tossi et al. 2000). The seven antimicrobial peptides discovered in this study may act on microorganisms in a similar manner.

MIC-values for each peptide are within publishable range, albeit we did not use clinically important pathogenic strains in antibacterial testing. Cationic antimicrobial peptides have showed MIC-values as low as 0.25–4 µg/ml (Moore et al. 1991; Hancock and Lehrer 1998) by contrast to our MIC-values which were between 0.78–100 µg/ml. With high probability the MIC-values in this study are however underestimated. According to the manufacturer Genscript, the net content of the peptide is usually 50–80 %. This means that 50–80 % out of reported percentage (75 %–98.8 % in this study) is actually pure peptide. Other substances in the dry peptide powder are for example water, absorbed solvents, counter ions and salts. MIC-values are likely lower against tested strains.

Antimicrobial peptide magainin 2 is 23 amino acids long and it has a double lysine-motif in the residues 10 and 11 (Zasloff 1987). Magainin 2 induces *E. coli* cell death by permeabilizing the cell membrane and causes the efflux of intracellular K+ (Matsuzaki et al. 1997). Synthetic antimicrobial peptides from *Dickeya zeae* Ech1591, *Vibrio nigripulchritudo* SOn1 (A and B),

Yersinia frederiksenii ATCC 33641, *Agarivorans albus* MKT 106 are all 22 amino acids long and they have double lysine-motif in the same place as magainin 2. When these connective factors are taken into account, it can be hypothesized, that these antimicrobial cyanobactin-like peptides have a similar mode of action as magainins.

The majority of cyanobactin-like peptides contain cysteine residues in their core sequences, so disulphide bridge formations between cysteines is expected as post-translational modifications. Post-translational modications involving cysteine residues happen especially frequently in RiPPs (Arnison et al. 2013). The antimicrobial activity for these peptides stays yet hypothetical and unrevealed. Bactofencin A is a cationic disulphide bridge containing bacteriocin produced by *Lactobacillus salivarius* DPC6502, is active against *S. aureus* and *Listeria innocua*. It is an unusual bacteriocin resembling eukaryotic defensins, containing two cysteine residues that form an intramolecular disulfide bond (O'Shea et al. 2013; O'Connor et al. 2015). This suggests that disulphide-bridge containing cyanobactin-like peptides might have antimicrobial activity against clinical pathogens. Singh et al. (2012) reported a broadspectrum antimicrobial peptide, laterosporulin. It was shown to be resistant to proteolytic enzymes, thermo stable and pH tolerant (Singh et al. 2012). Later laterosporulin was proven to inhibit the growth of both Gram-positive and Gram-negative bacteria (Singh et al. 2015). It is cysteine-rich 49 amino acids long peptide produced by Brevibacillus *sp*. strain GI-9 (Singh et al. 2012). According to Singh et al. (2015) laterosporulin adopts a four-strand tangled β-sheet structure stabilized by three disulphide-bridges. The intramolecular disulphide-bridges partly determine the intense stability for such a short peptide (Singh et al. 2015). The detected EGG-peptide from *Pseudanabaena* sp. PCC 6802 is 45 amino acids long and contains eight cysteines, supposedly forming four intramolecular disulphide-bridges. Considering the short length and disulphide-bridges, it can be hypothesized that the structure of EGG-peptide is even more rigid and more resistant to chemical forces compared to laterosporulin. Additionally, it might have inhibiting effect to the growth of Gram-positive and Gram-negative bacteria.

5.5 Classification of cyanobactin-like peptides

Cyanobactin-like peptides can be classified into two different classes according to our findings: 1) majority with 2–5 disulphide bonds and 2) minority without disulphide bonds. In this sense, the classification can be in outline supposed similarly with conus peptides, from which two major classes are determined: 1) majority with multiple disulphide bonds, indicated as

conotoxins and 2) minority with a single or no disulphide bond (Wang and Chi 2004; Kaas et al. 2010; Akondi et al. 2014), indicated as conopeptides (Terlau and Olivera 2004). Cyanobactin-like peptides have much more similarities to the antimicrobial peptides produced by eukaryotes. However, conus peptides are of eukaryotic origin and cyanobactin-like peptides arise from bacteria, so the appropriate classification scheme for the novel peptides detected here remains unclear.

Bacteriocins are ribosomally produced antimicrobial peptides from bacteria (Cotter et al. 2005, 2013) and they are found in all major lineages of bacteria (Riley and Wertz 2002). Bacteriocins are miscellaneous group and they can usually be classified into two classes: 1) Peptides that go through post-translational modifications; 2) Unmodified peptides (Cotter et al. 2013). The cysteine-rich cyanobactin-like peptides go through some post-translational modifications, which are proteolytic cleavage and formation of disulphide-bridges. The cysteine-poor, α-helical cyanobactin-like peptides, are proteolytically cleaved according to our bioinformatic results. Bacteriocins can be recognized by a double-glycine (GG) motif in the leader peptide sequence (Oman and van der Donk 2010; Micallef et al. 2014). From cyanobactin-like peptides, only predicted precursors from strains *Gloeocapsa* sp. PCC 7428 and *Aurantimonas coralicida* DSM-14790 contain this motif in the leader peptide. In this regard, cyanobactin-like peptides can not be classified as bacteriocins. However, more data should be obtained regarding the post-translational modifications in order to enlighten the possible classification together with bacteriocins.

Cyanobactins share some characteristics with compounds in bacteriocin superfamily, but they can be clearly differentiated as a subgroup based on chemical structures and biosynthetic pathway genes (Donia and Schmidt 2010). For cyanobactin-like peptides the classification can be similarly thought as with cyanobactins, because they are structurally and biosynthetically related to each other. Cyanobactin-like peptides from cyanobacteria have a conserved motif (LPxQxxPVxR) in their leader sequence, which is a connective factor with cyanobactins (Oman and van der Donk 2010; Leikoski et al. 2013). However, it appears to be more challenging to characterize subgroups inside a family that is as multiform as the cyanobactin-like peptides, which is also the challence with microcins (Duquesne et al. 2007).

Wang et al. (2011) showed that genes for the biosynthesis of bacteriocins are widespread in cyanobacteria. A total of 145 putative bacteriocin gene clusters were recognized in the genomes

of 43 cyanobacteria, from which a total of 290 precursor proteins were mapped with size varying between 28 to 164 amino acids (Wang et al. 2011). Cyanobactin-like peptides from *Pseudanabaena* sp. PCC 6802 could represent a novel class of bacteriocins from cyanobacteria, as the genetic machinery for bacteriocin production is widespread in cyanobacteria. Database searches (Wang et al. 2009) for detected cyanobactin-like peptides (EGG, ASK and CAGA) have revealed that these peptides do not share prominent homology with previously characterized bacteriocins but more closely resemble eukaryotic antimicrobial peptides. It would thus appear that cyanobactin-like peptides together with bactofencin A (O'Shea et al. 2013) are the first novel group of bacteriocins, if classified together with bacteriocins.

5.6 Ecological role, significance and future research of cyanobactin-like peptides

Natural products from microorganisms influence the performance and survival of other organisms (Demain and Sanchez 2009). It is suggested that bacteria produce secondary metabolites to match against competing microorganisms in their natural environment (Baltz 2008). Cyanobactin-like peptides might give competitive advantage to its host in an environment where there are other bacterial populations in the fight for ecological niches. These peptides might represent bacterial analogs for the ones that are present in vertebrates.

Sequences of these peptides are novel which raises the impact of these findings. Cysteine-rich cyanobactin-like peptides appear to represent a small class of peptides produced by bacteria, together with laterosporulin and bactofencin A (Singh et al. 2012; O'shea et al. 2013). Linear peptides are generally poor drug candidates because of their fast degradation and insufficient distribution (Velásquez and van der Donk 2011). Albeit, linear peptides with free termini are exposed to proteolytic digestion and ribosomal peptides have many ways to occlude the ends of the peptide (Leikoski et al. 2013). Thus, the cysteine-free α-helical cyanobactin-like peptides, which have antimicrobial activity, might pave the way for new possible drugs derived from natural products.

The novel pathway reported in this study probably presents only one fraction of micro-organisms true metabolic potential. Whole genome sequencing projects have shown that the natural product productivity potential of a microorganism is seriously underestimated (Winter et al. 2011) and only fraction of these are known products. Structure prediction, structure

determination and compound isolation for bioactivity testing could be done for the detected peptides. Compound isolation could be done from the heterologous host carrying the cyanobactin-like gene cluster from *Pseudanabaena* sp. PCC 6802, because the peptide expression was more extensive compared to crude sample. More work could be done with candidate strains in order to detect more natural products corresponding predictions from this novel peptide family, for instance cloning of the gene clusters. Bioinformatics could be extended aiming to identify all the genes in the gene clusters and to determine the borders of the gene clusters, and examine what genes are common to the clusters in addition to the protease and precursor genes. These could possibly be additional post-translational enzymes. It can be without a doubt presumed that cyanobacteria are and will become more valuable research targets of ribosomally produced and post-translationally modified peptides (Velasquez and van der Donk 2011; Wang et al. 2011).

6. CONCLUSIONS

Cyanobactin-like biosynthetic gene clusters were discovered from a wide selection of bacteria distributed across the bacterial domain. This novel gene cluster is present in the genomes of 38 bacteria representing six different phyla. From these we had 12 candidate strains in the study, nine cyanobacteria, two proteobacteria and one actinobacteria. Novel disulphide-bridge stabilized linear peptides were detected from *Pseudanabaena* sp. PCC 6802, and the gene cluster was successfully expressed in a heterologous host. After the detection of cysteine-rich peptides from cyanobacterium *Pseudanabaena* sp. PCC 6802 corresponding to bioinformatics predictions, more accurate predictions could be done in relation to protease cleavage sites. This led to the discovery of peptides with an α-helical structure. These peptides had antimicrobial activity between 0.78–100 μg/ml and they have the same ribosomal pathway as the ones in *Pseudanabaena* sp. PCC 6802, excluding formation of disulphide-bridges as post-translational modifications.

This study has led to the identification of a new type of antimicrobial peptide produced through the proteolytic cleavage of short precursor proteins and provided insights into the origin of the cyanobactin biosynthetic pathway. Novel cysteine containing peptides were discovered from cyanobacteria with yet hypothetical antimicrobial activity. Interestingly, cysteine-rich- and α-helical antimicrobial peptides are traditionally believed to be produced only by eukaryotes. This

study shows an extensive expansion of natural products produced by bacteria and that their true metabolic potential is most likely much wider than anticipated. Genome mining revealed novel pathways of antimicrobial peptides from unexplored sources. They are ribosomally produced, post-translationally modified, linear and they have antimicrobial activity.

ACKNOWLEDGMENTS

I want to express my deepest gratitude to David Fewer and Niina Leikoski for the guidance and motivational teaching through the study. I would like to thank Matti Wahlsten and Jouni Jokela for introducing me to the world of chemical analyses and teaching me many new things. I would also like to thank Professor Kaarina Sivonen who kindly informed me before the project that David is looking for a master student. Anirudra Parajuli will be thanked for his valuable help in cloning steps. I want to thank the whole cyanobacteria group for great and inspiring working environment. Being a part of this group has been a privilege.

REFERENCES

Agarwal V, Pierce E, McIntosh J, Schmidt EW and Nair SK (2012). Structures of cyanobactin maturation enzymes define a family of transamidating proteases. Chem. Biol. 19: 1411–22.

Akondi KB, Muttenthaler M, Dutertre S, Kaas Q, Craik DJ, Lewis, RJ, and Alewood PF (2014). Discovery, synthesis, and structure-activity relationships of conotoxins. Chem. Rev. 114: 5815–5847.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25: 3389–3402.

Arnison PG, Bibb MJ, Bierbaum G, Bowers AA, Bugni TS, Bulaj G, Camarero JA, Campopiano DJ, Challis GL, Clardy J, et al. (2013). Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. Nat. Prod. Rep. 30: 108–160.

Bachmann BO, Van Lanen SG and Baltz RH (2014). Microbial genome mining for accelerated natural products discovery: is a renaissance in the making? J. Ind. Microbiol. Biotechnol. 41: 175–184.

Baltz RH (2008). Renaissance in antibacterial discovery from actino-mycetes. Curr. Opin. Pharmacol. 8: 557–563.

Berdy J (2005). Bioactive microbial metabolites. A personal view. J. Antibiot. 58: 1–26

Berry CL, Brassinga AKC, Donald LJ, Fernando WGD, Loewen PC, et al. (2012) Chemical and biological characterization of sclerosin, an antifungal lipopeptide. Can. J. Microbiol. 58: 1027–1034. doi: 10.1139/w2012-079.

Boman HG (1998). Gene-encoded peptide antibiotics and the concept of innate immunity: an update review. Scand. J. Immunol. 48: 15–25.

Brogden KA (2005). Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? Nat. Rev. Microbiol. 3: 238–250.

Bui TH, Wray V, Nimtz M, Fossen T, Preisitsch M, Schröder G, Wende, K, Heiden SE and Mundt S (2014). Balticidins A-D, antifungal hassallidin-like lipopeptides from the Baltic Sea cyanobacterium *Anabaena cylindrica* Bio33. J. Nat. Prod. 77: 1287–1296.

Burja AM, Banaigs B, Abou-Mansour E, Burguess JG and Wright PC (2001). Marine cyanobacteria — a profilic source of natural products. Tetrahedron 57: 9347–9377.

Castenholz RW (2001). Cyanobacteria, oxygenic photosynthetic bacteria. In: Boone DR, Castenholz RW, Garrity GM, (eds). Bergey´s manual of systematic bacteriology vol 1. 2nd ed. New York, United States: Springer. pp. 473-599.

Challis GL (2008). Genome mining for novel natural product discovery. J. Med. Chem. 51: 2618–2628.

Claesen J and Bibb M (2010). Genome mining and genetic analysis of cypemycin biosynthesis reveal an unusual class of posttranslationally modified peptides. Proc. Natl. Acad. Sci. USA 107: 16297–16302.

Clardy J, Fischbach MA and Walsh CT (2006). New antibiotics from bacterial natural products. Nat. Biotechnol. 24: 1541–50.

Cotter PD, Hill C, and Ross RP (2005). Bacteriocins: developing innate immunity for food. Nat. Rev. Microbiol. 3: 777–788.

Cotter PD, Ross RP and Hill C (2013). Bacteriocins - a viable alternative to antibiotics? Nat. Rev. Microbiol. 11: 95–105.

Demain AL, and Sanchez S (2009). Microbial drug discovery: 80 years of progress. J. Antibiot. (Tokyo) 62: 5-16.

Donia MS, Hathaway BJ, Sudek S, Haygood MG, Rosovitz MJ, Ravel J and Schmidt EW (2006). Natural combinatorial peptide libraries in cyanobacterial symbionts of marine ascidians. Nat. Chem. Biol. 2: 729–735. doi: 10.1038/nchembio829.

Donia MS, Ravel J and Schmidt EW (2008). A global assembly line for cyanobactins, Nat. Chem. Biol. 4: 341–343.

Donia MS and Schmidt EW (2010). Cyanobactins – ubiquitous cyanobacterial ribosomal peptide metabolites. In: Comprehensive natural products II chemistry and biology. Moore BS and Crews P (Eds.). Elsevier, Oxford. pp. 539–558.

Dixit RB and Suseela MR (2013). Cyanobacteria: potential candidates for drug discovery, Antonie van Leeuwenhoek 103 (5): 947–961.

Dunbar KL and Mitchell DA (2013). Revealing nature´s synthetic potential through the study of ribosomal natural product biosynthesis. ACS Chem. Biol. 8: 473–487.

Duquesne S, Destoumieux-Garzon D, Peduzzi J and Rebuffat S (2007). Microcins, gene-encoded antibacterial peptides from enterobacteria. Nat. Prod. Rep. 24: 708–734.

Edwards U, Rogall T, Blocker H, Emde M and Bottger EC (1989). Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. Nucleic Acids Res. 17: 7843–7853.

El Amri C, Lacombe C, Zimmerman K, Ladram A, Amiche M, Nicolas P and Bruston F (2006). The plasticins: membrane adsorption, lipid disorders, and biological activity. Biochemistry 45: 14285–14297.

Finking R and Marahiel MA (2004). Biosynthesis of nonribosomal peptides. Annu. Rev. Microbiol. 58: 453–488.

Fischbach MA, Walsh CT (2006). Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. Chem. Rev. 106: 3468–3496.

Fox JL (2013). Antimicrobial peptides stage a comeback. Nat. Biotechnol. 31: 379–382.

Gerwick WH, Coates RC, Engene N, Gerwick L, Grindberg RV, Jones AC and Sorrels CM (2008). Giant marine cyanobacteria produce exciting potential pharmaceuticals. Microbe 3: 277–284.

Giuliani A, Pirri G, Nicoletto SF (2007). Antimicrobial peptides: an overview of a promising class of therapeutics. Cent. Eur. J. Biol. 2: 1–33.

Goering RV, Dockrell HM, Wakelin D, Zuckerman M, Chiodini PL, Roitt IM, et al. (2012). Mims' medical microbiology. 5th ed. Philadelphia, PA: Mosby Elsevier. p. 17.

Hancock REW and Lehrer R (1998). Cationic peptides: a new source of antibiotics. Trends Biotechnol. 16: 82–88.

Hancock RE and Sahl HG (2006). Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies. Nat. Biotechnol. 24: 1551–1557.

Hassan M, Kjos M, Nes IF, Diep DB and Lotfipour F (2012). Natural antimicrobial peptides from bacteria: Characteristics and potential applications to fight against antibiotic resistance. J. Appl. Microbiol, 113: 723–736.

Houssen WE, Wright SH, Kalverda AP, Thompson GS, Kelly SM, Jaspars M (2010). Solution structure of the leader sequence of the patellamide precursor peptide, PatE1-34 ChemBioChem 11: 1867–1873 .

Ishida K, Matsuda H, Murakami M and Yamaguchi K (1997). Kawaguchipeptin B, an antibacterial cyclic undecapeptide from the cyanobacterium Microcystis aeruginosa. J. Nat. Prod. 60: 724–726.

Jaki B, Zerbe O, Heilmann J and Sticher O (2001). Two novel cyclic peptides with antifungal activity from the cyanobacterium *Tolypothrix byssoidea* (EAWAG 195). J. Nat. Prod. 64: 154–158.

Jin Y, Hammer J, Pate M, Zhang Y, Zhu F, Zmuda E and Blazyk J (2005). Antimicrobial activities and structures of two linear cationic peptide families with various amphipathic beta-sheet and alpha-helical potentials, Antimicrob. Agents Chemother. 49: 4957–4964.

Kaas Q, Westermann JC and Craik DJ (2010). Conopeptide characterization and classifications: an analysis using ConoServer. Toxicon. 55: 1491–1509

Kaya K, Mahakhant A, Keovara L, Sano T, Kubo T, Takagi H (2002). Spiroidesin, a novel lipopeptide from the cyanobacterium *Anabaena spiroides* that inhibits cell growth of the cyanobacterium *Microcystis aeruginosa*. J. Nat. Prod. 65: 920–921.

King GF and Hardy MC (2013). Spider-venom peptides: structure, pharmacology, and potential for control of insect pests. Annu. Rev. Entomol. 58: 475–496.

Kopp F and Marahiel MA (2007). Where chemistry meets biology: the chemoenzymatic synthesis of nonribosomal peptides and polyketides. Curr. Opin. Biotechnol. 18(6): 513–520.

Kotai J (1972). Instructions for preparation of modified nutrient solution Z8 for algae. Norwegian Institute for Water Research, Oslo, http://www.sccap.dk/media/freshwater/7.asp

Koyama T and Iwamuro S (2008). Molecular cloning of a cDNA encoding atypical antimicrobial and cytotoxic brevinin-1Ja from the skin of the Japanese brown frog, *Rana japonica*. Zoolog. Sci. 25: 487–491.

Lee SW, Mitchell DA, Markley AL, Hensler ME, Gonzalez D, Wohlrab A, Dorrestein PC, Nizet V and Dixon JE (2008). Discovery of a widely distributed toxin biosynthetic gene cluster. Proc. Natl. Acad. Sci. USA. 105: 5879–5884.

Lee J, McIntosh J, Hathaway BJ and Schmidt EW (2009). Using marine natural products to discover a protease that catalyses peptide macrocyclization of diverse substrates. J. Am. Chem. Soc. 131: 2122–2124.

Leikoski N, Fewer DP and Sivonen K (2009). Widespread occurrence and lateral transfer of the cyanobactin biosynthesis gene cluster in cyanobacteria. Appl. Environ. Microbiol. 75: 853–857.

Leikoski N, Fewer DP, Jokela J, Wahlsten M, Rouhiainen L and Sivonen K (2010). Highly diverse cyanobactins in strains of the genus Anabaena. Appl. Environ. Microbiol. 76: 701–709.

Leikoski N, Fewer DP, Jokela J, Alakoski P, Wahlsten M and Sivonen K (2012). Analysis of an inactive cyanobactin biosynthetic gene cluster leads to discovery of new natural products from strains of the genus microcystis. PLoS ONE 7(8): e43002. doi:10.1371/journal.pone.0043002.

Leikoski N, Liu L, Jokela J, Wahlsten M, Gugger M, Calteau A, Permi P, Kerfeld CA, Sivonen K and Fewer DP (2013). Genome mining expands the chemical diversity of the cyanobactin family to include highly modified linear peptides. Chem. Biol. 20: 1–11.

Lepere C, Wilmotte A and Meyer B (2000). Molecular diversity of *Microcystis* strains (Cyanophyceae, Chroococcales) based on 16S rRNA sequences. Syst. Geogr. Plants. 70: 275–283.

Lewis RJ and Garcia ML (2003). Therapeutic potential of venom peptides. Nat. Rev. Drug Discov. 2: 790−802.

Marahiel MA (2009). Working outside the protein-synthesis rules: insights into non-ribosomal peptide synthesis. J. Pept. Sci. 15: 799–807.

Matsuzaki K, Sugishita K, Harada M, Fujii N and Miyajima K (1997). Interactions of an antimicrobial peptide, magainin 2, with outer and inner membranes of Gram-negative bacteria. Biochim. Biophys. Acta 1327: 119–130.

Mechkarska M, Ahmed E, Coquet L, Leprince J, Jouenne T, Vaudry H et al (2011). Peptidomic analysis of skin secretions demonstrates that the allopatric populations of *Xenopus muelleri* (Pipidae) are not conspecific. Peptides. 32: 1502–1508.

Micallef ML, D'Agostino PM, Al-Sinawi B, Neilan BA and Moffitt MC (2015). Exploring cyanobacterial genomes for natural product biosynthesis pathways. Mar. Genomics. 21: 1–12.

Milligan KE, Marquez BL, Williamson RT and Gerwick WH (2000). Lyngbyabellin B, a toxic and antifungal secondary metabolite from the marine cyanobacterium *Lyngbya majuscula*. J. Nat. Prod. 63: 1440–1443.

Milne TJ, Abbenante G, Tyndall JD, Halliday J and Lewis RJ (2003). Isolation and characterization of a cone snail protease with homology to CRISP proteins of the pathogenesis-related protein superfamily. J. Biol. Chem. 278: 31105–31110.

Monincová L, Veverka V, Slaninová J, Buděšínský M, Fučík V, Bednárová L, Straka J and Čeřovský V (2014). Structure–activity study of macropin, a novel antimicrobial peptide from the venom of solitary bee *Macropis fulvipes* (Hymenoptera: Melittidae). J. Peptide Sci. 20: 375–384. doi: 10.1002/psc.2625.

Moore KS, Bevins CL, Brasseur MM, Tomassini N, Turner K, Eck H and Zasloff M (1991). Antimicrobial peptides in the stomach of *Xenopus laevis*. J. Biol. Chem. 266: 51 -57.

Müller D, Krick A, Kehraus S, Mehner C, Hart M, Küpper FC, Saxena K, Prinz H, Schwalbe H, Janning P, Waldmann H and König GM (2006). Brunsvicamides A–C: sponge-related cyanobacterial peptides with *Mycobacterium tuberculosis* protein tyrosine phosphatase inhibitory activity. J. Med. Chem. 49: 4871–4878.

Navon-Venezia S, Feder R, Gaidukov L, Carmeli Y and Mor A (2002). Antibacterial properties of dermaseptin S4 derivatives with *in vivo* activity. Antimicrob. Agents Chemother. 46: 689–694.

Newman DJ and Cragg GM (2007). Natural products as sources of new drugs over the last 25 years. J. Nat. Prod. 70: 461–477.

Nunnery JK, Mevers E and Gerwick WH (2010). Biologically active secondary metabolites from marine cyanobacteria. Curr. Opin. Biotechnol. 21: 787–793. doi: 10.1016/j.copbio.2010.09.019.

Nissen-Meyer J. and Nes IF (1997). Ribosomally synthesized antimicrobial peptides: their function, structure, biogenesis, and mechanism of action. Arch. Microbiol. 167: 67–77.

Nolan EM and Walsh CT (2009). How nature morphs peptide scaffolds into antibiotics. ChemBioChem. 10: 34–53. doi: 10.1002/cbic.200800438.

O'Connor PM, Ross RP, Hill C and Cotter PD (2015). Antimicrobial antagonists against food pathogens; a bacteriocin perspective. Curr. Opin. Food Sci. 2: 51–57.

Olivera BM and Cruz LJ (2001). Conotoxins, in retrospect. Toxicon 39(1): 7–14.

Oman TJ and van der Donk WA (2010). Follow the leader: the use of leader peptides to guide natural product biosynthesis. Nat. Chem. Biol. 6: 9–18.

O'Shea EF, O'Connor PM, O'Sullivan O, Cotter PD, Ross RP and Hill C (2013). Bactofencin A, a new type of cationic bacteriocin with unusual immunity. MBio. 4: 00498-00413.

Papagianni M (2003). Ribosomally synthesized peptides with antimicrobial properties: biosynthesis, structure, function, and applications. Biotechnol. Adv. 21: 465–499.

Powers JP and Hancock REW (2003). The relationship between peptide structure and antibacterial activity. Peptides. 24: 1681–1691.

Pütsep K, Brändén CI, Boman HG and Normark S (1999). Antibacterial peptide from *H. pylori*. Nature. 398: 671–672.

Riley MA and Wertz JE (2002). Bacteriocins: Evolution, ecology, and application. Annu. Rev. Microbiol. 56: 117–137.

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA and Barrell B (2000). Artemis: sequence visualization and annotation. Bioinformatics 16: 944–945.

Saez NJ, Senff S, Jensen JE, Er SY, Herzig V, et al. (2010). Spider-venom peptides as therapeutics. Toxins 2: 2851–71.

Sang Y and Blecha F (2008). Antimicrobial peptides and bacteriocins: alternatives to traditional antibiotics. Anim. Health Res. Rev. 9: 227–235.

Scheffler RJ, Colmer S, Tynan H, Demain AL and Gullo VP (2013). Antimicrobials, drug discovery, and genome mining. Appl. Microbiol. Biotechnol: 97: 969–978.

Scherlach K and Hertweck C (2006). Discovery of aspoquinolones A–D, prenylated quinoline-2-one alkaloids from *Aspergillus nidulans*, motivated by genome mining. Org. Biomol. Chem. 4: 3517–3520.

Schmidt EW, Nelson JT, Rasko DA, Sudek S, Eisen JA, Haygood MG and Ravel J (2005). Patellamide A and C biosynthesis by a microcin-like pathway in *Prochloron didemni*, the cyanobacterial symbiont of *Lissoclinum patella*. Proc. Natl. Acad. Sci. USA. 102: 7315–7320. doi: 10.1073/pnas.0501424102.

Schmidt EW and Donia MS (2009). Cyanobactin ribosomally synthesized peptides — a case of deep metagenome mining Methods. Enzymol, 458: 575–596.

Shih PM, Wu DY, Latifi A, Axen SD, Fewer DP, Talla E, Calteau A, Cai F, de Marsac NT, Rippka R, et al (2013). Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. Proc. Natl. Acad. Sci. USA. 110: 1053–1058.

Singh S, Bhushan KN and Banerjee UC (2005). Bioactive compounds from cyanobacteria and microalgae: an overview. Crit. Rev. Biotechnol. 25: 73–95.

Singh PK, Chittpurna A, Sharma V, Patil PB and Korpole S (2012). Identification, purification and characterization of laterosporulin, a novel bacteriocin produced by *Brevibacillus* sp. strain GI-9. PLoS One 7: e31498.

Singh PK, Solanki V, Sharma S, Thakur KG, Krishnan B and Korpole S (2015). The intramolecular disulfide-stapled structure of laterosporulin, a class IId bacteriocin, conceals a human defensin-like structural module. FEBS Lett. 282: 203–214.

Sivonen K and Börner T (2008). Bioactive compounds produced by cyanobacteria. In A. Herraro and E. Flores (ed.). The cyanobacteria: molecular biology, genomics and evolution. Caister Academic Press,Norfolk, United Kingdom. pp. 159–197.

Sivonen K, Leikoski N, Fewer DP and Jokela J (2010). Cyanobactins ribosomal cyclic peptides produced by cyanobacteria. Appl. Microbiol. Biotechnol. 86: 1213–1225.

Tan LT (2007). Bioactive natural products from marine cyanobacteria for drug discovery. Phytochemistry 68: 954–979.

Terlau H and Olivera BM (2004). Conus venoms: a rich source of novel ion channel-targeted peptides. Physiol. Rev. 84: 41–68.

Tianero MD, Donia MS, Young TS, Schultz PG and Schmidt EW (2012). Ribosomal route to small-molecule diversity. J. Am. Chem. Soc. 134: 418–425.

Tossi A, Sandri L and Giangaspero A (2000). Amphipathic, alpha-helical antimicrobial peptides. Biopolymers 55: 4–30.

Velásquez JE and van der Donk WA (2011). Genome mining for ribosomally synthesized natural products. Curr. Opin. Chem. Biol. 15: 11–21.

Vestola J, Shishido TK, Jokela J, Fewer DP, Aitio O, Permi P, Wahlsten M, Wang H, Rouhiainen L and Sivonen K (2014). Hassallidins, antifungal glycolipopeptides, are widespread among cyanobacteria and are the end-product of a nonribosomal pathway. Proc. Natl. Acad. Sci. USA. 111: 1909–1917.

Wang CZ and Chi CW (2004). Conus peptidessa rich pharmaceutical treasure, Acta Biochim. Biophys. Sin. 36: 713–23.

Wang G, Li X and Wang Z (2009). APD2: the updated antimicrobial peptide database and its application in peptide design. Nucleic Acids Res. 37: 933–937.

Wang H, Fewer DP and Sivonen K (2011). Genome mining demonstrates the widespread occurrence of gene clusters encoding bacteriocins in cyanobacteria. PLoS ONE 6(7): e22384. doi:10.1371/journal.pone.0022384.

Wang H, Fewer DP, Holm L, Rouhiainen L and Sivonen K (2014). Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes Proc. Natl. Acad. Sci. USA. 111: 9259–9264.

Wang G (2014). Post-translational modifications of natural antimicrobial peptides and strategies for peptide engineering. Curr. Biotechnol. 1: 72–79.

White WC and Wimley ME (1995). Selsted structure, function, and membrane integration of defensins. Curr. Opin. Struct. Biol. 5: 521–527.

Wiegand I, Hilpert K and Hancock RE (2008). Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. Nat. Protoc. 3; 163–175.

Wimley WC and Hristova K (2011). Antimicrobial peptides: Successes, challenges and unanswered questions. J. Membr. Biol. 239: 27–34.

Windley MJ, Herzig V, Dziemborowicz SA, Hardy MC, King GF and Nicholson GM (2012). Spider-venom peptides as bioinsecticides. Toxins 4: 191–227.

Winter JM, Behnken S and Hertweck C (2011). Genomics-inspired discovery of natural products. Curr. Opin. Chem. Biol. 15: 22–31.

Zainuddin NE, Jansen R, Nimtz M, Wray V, Preisitsch M; Lalk, M and Mundt S (2009). Lyngbyazothrins A–D, antimicrobial cyclic undecapeptides from the cultured cyanobacterium *Lyngbya* sp. J. Nat. Prod. 72: 1373–1378.

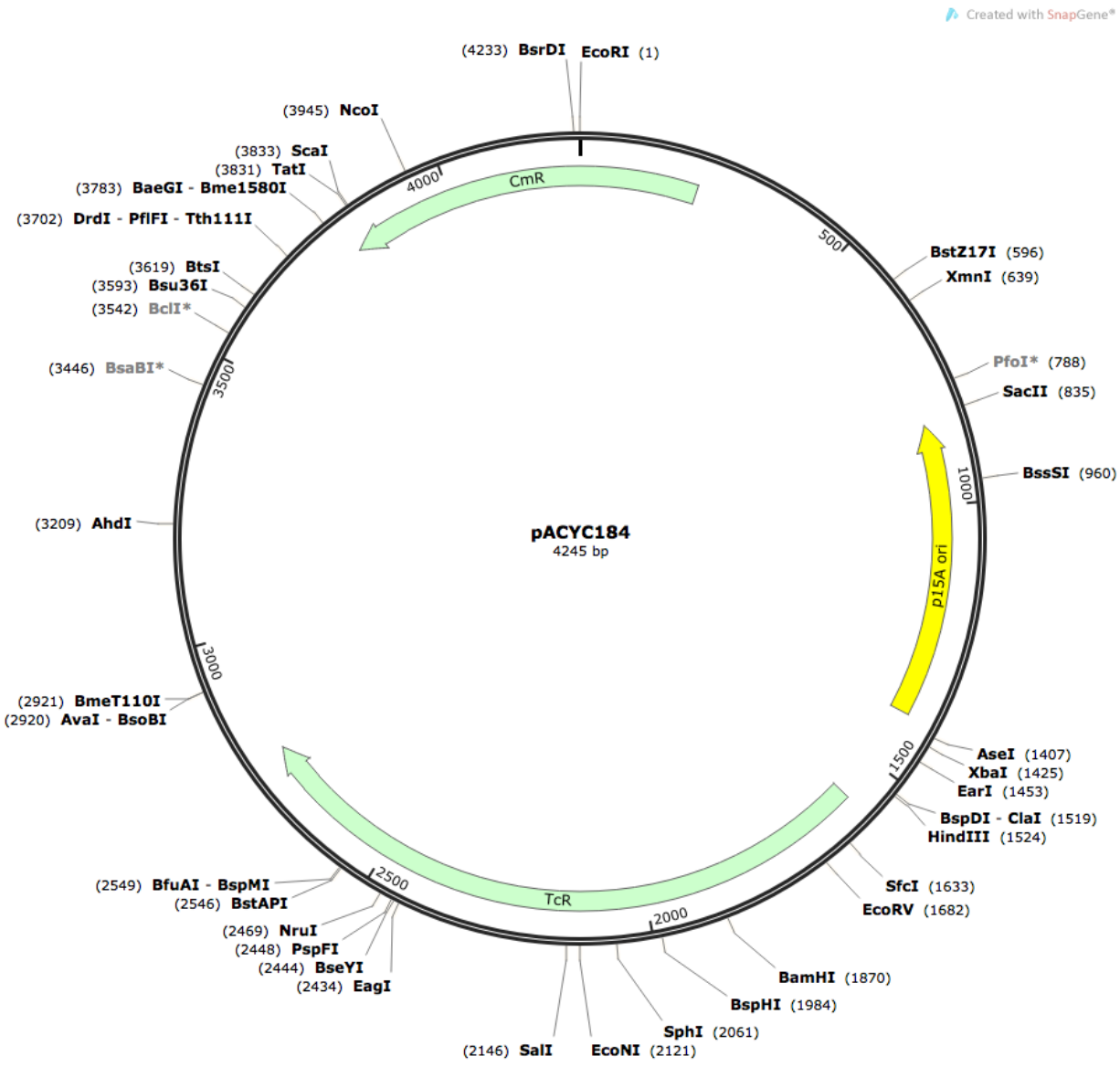
Zasloff M (1987). Magainins, a class of antimicrobial peptides from Xenopus skin: isolation, characterization of two active forms, and partial cDNA sequence of a precursor. Proc. Natl Acad. Sci. USA. 84: 5449–5453.

Zasloff M (2002). Antimicrobial peptides of multicellular organisms. Nature 415: 389–395.

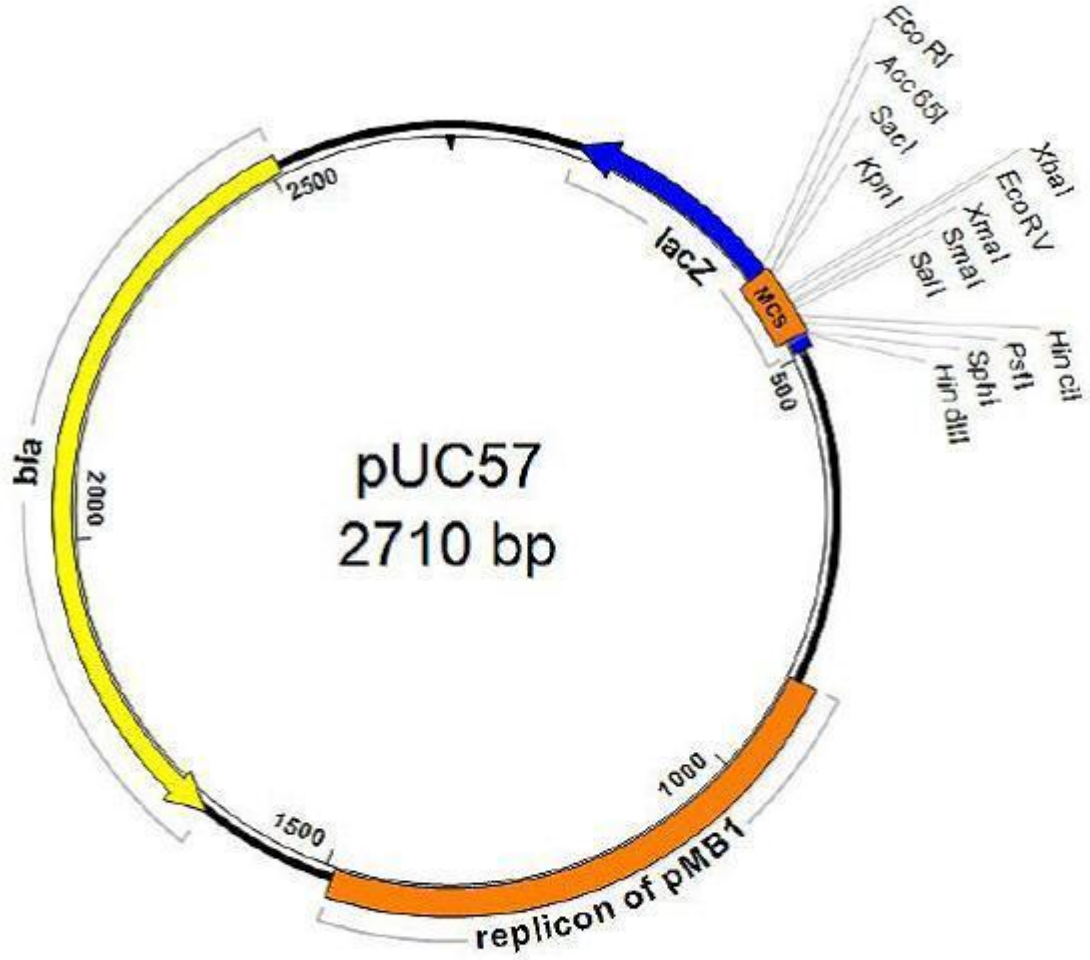
Zerikly M and Challis GL (2009). Strategies for the discovery of new natural products by genome mining. ChemBioChem 10: 625−633.

APPENDIX

Plasmid vectors used for subcloning



http://www.snapgene.com/resources/plasmid_files/basic_cloning_vectors/pACYC184/



Multiple Cloning Site:

http://www.centic-biotec.com/pUC57/en