Chemistry & Biology
Article

Two Alternative Starter Modules for the Non-Ribosomal Biosynthesis of Specific Anabaenopeptin Variants in *Anabaena* (Cyanobacteria)

Leo Rouhiainen,[1] Jouni Jokela,[1] David P. Fewer,[1] Marina Urmann,[1] and Kaarina Sivonen[1,*]
[1]Department of Food and Environmental Sciences, P.O. Box 56, Viikki Biocenter, Viikinkaari 9, FIN-00014, University of Helsinki, Finland
*Correspondence: kaarina.sivonen@helsinki.fi
DOI 10.1016/j.chembiol.2010.01.017

SUMMARY

Anabaenopeptins are a diverse family of cyclic hexapeptide protease inhibitors produced by cyanobacteria that contain a conserved ureido bond and D-Lys moiety. Here we demonstrate that anabaenopeptins are assembled on a nonribosomal peptide synthetase enzyme complex encoded by a 32 kb *apt* gene cluster in the cyanobacterium *Anabaena* sp. strain 90. Surprisingly, the gene cluster encoded two alternative starter modules organized in separate bimodular proteins. The starter modules display high substrate specificity for L-Arg/L-Lys and L-Tyr, respectively, and allow the specific biosynthesis of different anabaenopeptin variants. The two starter modules were found also in other *Anabaena* strains. However, just a single module was present in the anabaenopeptin gene clusters of *Nostoc* and *Nodularia,* respectively. The organization of the *apt* gene cluster in *Anabaena* represents an exception to the established colinearity rule of linear non-ribosomal peptide synthetases.

INTRODUCTION

Cyanobacteria are a prolific source of small structurally variable bioactive peptides, which typically contain nonproteinogenic amino acids and are often cyclic (Moore, 1996; Namikoshi and Rinehart, 1996; Burja et al., 2001; Sivonen and Börner, 2008). Cyanobacterial peptides are organized into families based on structural characteristics and include protease inhibitors such as cyanopeptolins, scyptolins, and aeruginosins, and hepatotoxins such as the microcystins (Sivonen and Börner, 2008; Welker and von Döhren, 2006). Many of these peptides display a great chemical variation and are the end products of a non-ribosomal peptide synthetase (NRPS) pathway (Marahiel et al., 1997; Sieber and Marahiel, 2005).

NRPSs are large enzymes that display a modular structure, with each module catalyzing the activation and incorporation of an amino acid or in some cases a carboxylic acid into the growing peptide chain in a stepwise fashion (Marahiel et al., 1997; von Döhren et al., 1997). The modules have catalytic domains for amino acid activation, thiolation and condensation. Further modifications to the amino acids are common and include epimerization, heterocyclization and *N*-methylation (Marahiel et al., 1997; von Döhren et al., 1997; Finking and Marahiel, 2004; Sieber and Marahiel, 2005). NRPSs typically follow the colinearity rule in which the substrate specificity and the number of the modules reflect the composition of the product (Marahiel et al., 1997). The selection of amino acids in NRPS enzyme complexes is determined by the adenylation domains (Finking and Marahiel, 2004; Lautru and Challis, 2004). The condensation domains are also suggested to have a selective role in non-ribosomal biosynthesis of peptides (Sieber and Marahiel, 2005; Lautru and Challis, 2004; Fewer et al., 2007). Individual NRPS enzyme complexes can synthesize sets of peptides simultaneously (Lawen and Traber, 1993; Rouhiainen et al., 2004). A lack of substrate specificity and incomplete modification are usually assumed to explain such natural variation in non-ribosomal peptide chemical structures (Lawen and Traber, 1993; Marahiel et al., 1997).

Anabaenopeptins are a highly diverse family of cyclic hexapeptides which are potent protease inhibitors and produced by a broad range of morphologically and genetically diverse cyanobacteria (Sivonen and Börner, 2008). These compounds are *N*-methylated and contain a conserved ureido linkage connecting the side-chain amino acid residue to the D-lysine (Welker and von Döhren, 2006). *Anabaena* sp. strain 90 produces three major anabaenopeptins that differ in composition of the amino acid side chain (Figure 1). Relaxed substrate specificity is the only known mechanism to achieve variation in the unmodified amino acids introduced to peptides by the NRPSs. Here we report that *Anabaena* sp. strain 90 and other anabaenopeptin producing *Anabaena* strains bear two alternative starter modules resulting in the simultaneous and specific biosynthesis of the different anabaenopeptin variants.

RESULTS

Characterization of NRPS Genes of Anabaenopeptin Biosynthesis

Five open reading frames (ORFs) (*aptA*$_1$, *aptA*$_2$, *aptB*, *aptC,* and *aptD*) were identified in the 32.5 kb gene cluster in *Anabaena* sp. strain 90, coding for five NRPSs with the sizes of 2195, 2174, 1069, 2565, and 1403 amino acids, respectively (Table 1). AptA$_1$ and AptA$_2$ each contain two NRPS modules, both of which bear two adenylation and thiolation domains, one condensation and one epimerization domain (Figures 1 and 2A).

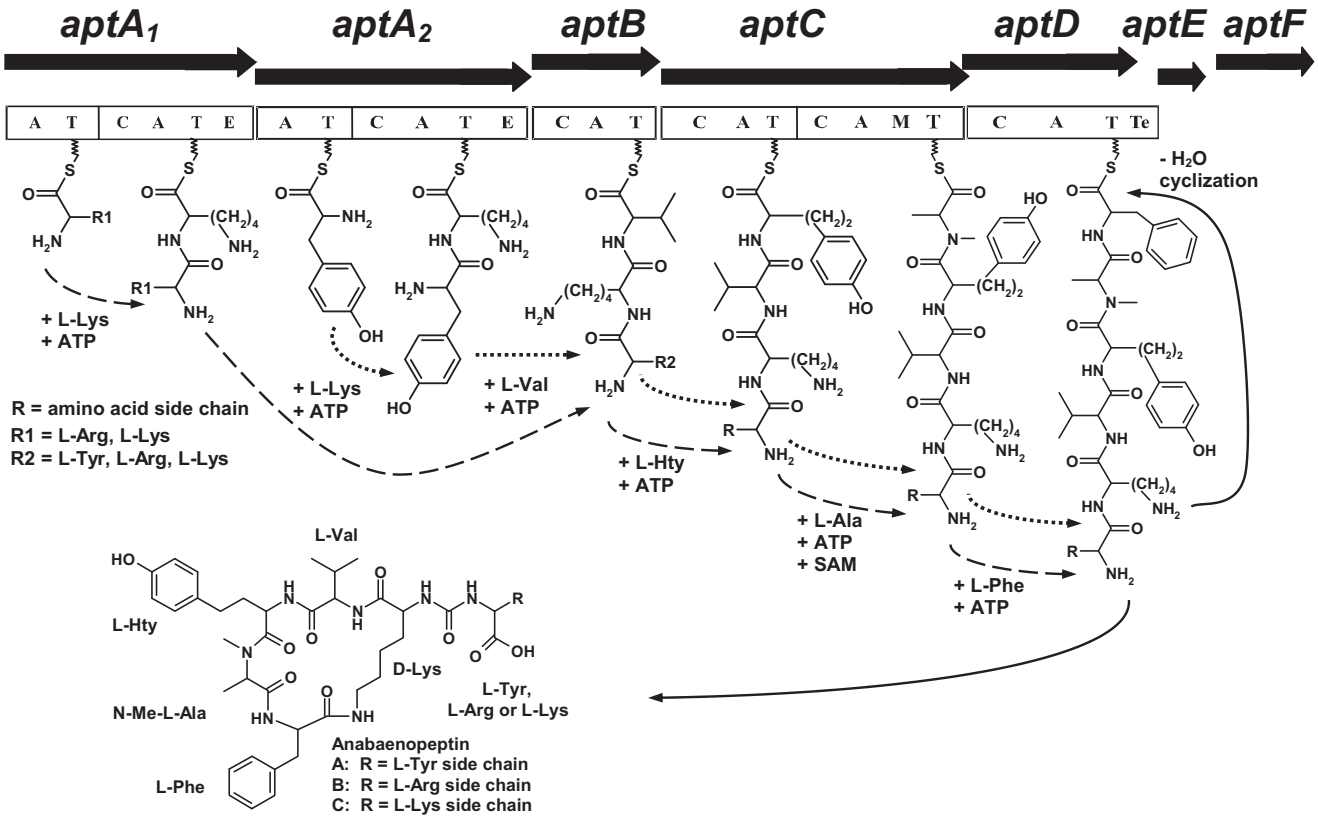

Figure 1. Biosynthesis of Anabaenopeptins A, B and C on the Peptide Synthetases with Alternative Bimodular Starter Proteins in *Anabaena* Sp. Strain 90

The pathways using either $AptA_1$ (dashed arrows) or $AptA_2$ (dotted arrows) are shown to yield arginine (B) and lysine (C) or tyrosine (A) anabaenopeptin variants, respectively. The substrate amino acids are loaded in L-form; L-lysine is isomerized to D-lysine by the epimerase domain in $AptA_1$ and in $AptA_2$. It is still unclear at which step the ureido bond is formed.

AptB and AptD were composed of a single module containing a condensation, adenylation, and thiolation, and in addition AptD included a thioesterase domain. AptC comprised two modules, each with a condensation, adenylation, and thiolation domain. The latter module also contained an *N*-methyltransferase domain embedded in the adenylation domain (Figures 1 and 2A). The cluster included two genes, *aptE* and *aptF*, which did not code for NRPS. The modular structure and the order of the catalytic domains match well the organization of a NRPS expected to assemble anabaenopeptins of *Anabaena* sp. strain 90,

Table 1. Comparison of Polypeptides Encoded by Anabaenopeptin Synthetase Gene Cluster of *Anabaena* Sp. Strain 90, *Nodularia spumigena* CCY 9414 and *Nostoc punctiforme* PCC 73102

Cyanobacterium	Gene/ORF, Length of the Encoded Protein (aa), Protein Identity/Similarity with *Anabaena* Sp. Strain 90 Protein (%)						
Anabaena sp. strain 90	*aptA*$_1$	*aptA*$_2$	*aptB*	*aptC*	*aptD*	*aptE*	*aptF*
	2195	2174	1069	2565	1403	392	765
Nodularia spumigena CCY9414	Missing	ORF1	ORF2	ORF3	ORF4	ORF5	ORF6
	-	2192	1075	2558	1394	392	738
		69/82	77/92	77/90	79/93	88/97	73/89
Nostoc punctiforme PCC3102	Missing	ORF1	ORF2	ORF3	ORF4	ORF5	ORF6
	-	2181	1076	2578	1406	392	675
		79/92	81/92	74/89	74/90	82/96	69/82

FASTA version 35.04 Oct. 20, 2008 (Pearson and Lipman, 1988). Algorithm: FASTA (3.5 Sept 2006) [optimized], Parameters: BL50 matrix http://fasta.bioch.virginia.edu/fasta_www2/fasta_www.cgi.

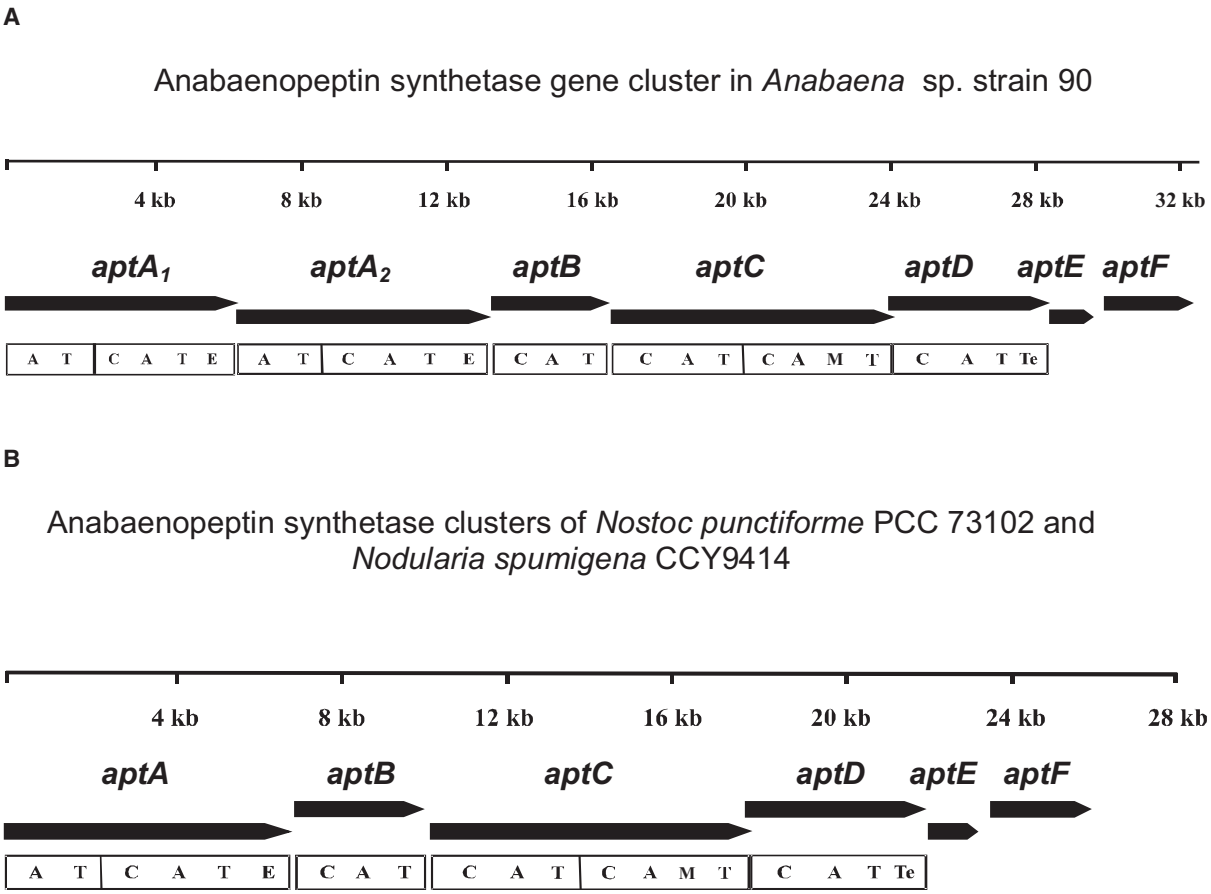

Figure 2. Anabaenopeptin Synthetase Gene Clusters in Cyanobacteria

(A) Anabaenopeptin synthetase gene cluster in *Anabaena* sp. strain 90.

(B) Anabaenopeptin synthetase clusters of *Nostoc punctiforme* PCC 73102 and *Nodularia spumigena* CCY9414.

Abbreviations for domains: A, adenylation; C, condensation; T, thiolation; E, epimerization; M, *N*-methyltransferase and Te, thioesterase.

except that two extra modules were encoded in this gene cluster.

Operon Organization

We concluded that the anabaenopeptin synthetase genes are transcribed in four operons. Structural analysis predicted that *aptA*$_2$ and *aptB* are transcribed in one and *aptC*, *aptD,* and *aptE* in a second operon. The genes *aptA*$_1$ and *aptF* are predicted to be transcribed in distinct operons. Putative −10 and −35 elements (AAAAAT/GTCACA and TAAATT/TTGACA, respectively) were found in the 92 bp region between *aptA*$_1$ and *aptA*$_2$. The *aptA*$_2$ and *aptB* genes were separated by 57 bp, whereas *aptB* and *aptC* were 194 bp apart. The *aptC* and *aptD* genes overlap by four bases, *aptD* and *aptE* were 43 bp apart, and *aptE* and *aptF* were separated by 409 bp. The proposed start codon for *aptA*$_1$ is ATC, which is located 13 bases after a putative upstream ACGCAG Shine-Dalgarno sequence. The remaining genes, *aptA*$_2$, *aptB*, *aptC, aptD*, *aptE,* and *aptF,* start with ATG. Potential ribosome binding sites for *aptB*, *aptC*, *aptD,* and *aptE* were predicted as CAGAGG, AGCAAA, AAGAGG, and AATTGG, which precede the start of the corresponding gene with 7, 14, 10, and 9 bases, respectively. The presence of termination signals and promoters cannot be fully ruled out between *aptA*$_2$ and *aptB*, or between *aptD* and *aptE*.

Substrate Specificity of the Expressed Adenylation Domains

Anabaena sp. strain 90 is recalcitrant to genetic manipulation, and substrate specificities of expressed adenylation domains were studied in place of knockout mutagenesis experiments. Seven of the eight adenylation domains were overexpressed in *E. coli* BL21 Star (DE3) and their substrate specificities determined in an ATP-PPi exchange assay. The adenylation domains from six modules, modules 2-6 and 8 (Figures 1 and 2A), were expressed in a soluble form, and the substrate specificities were consistent with the structure and the order of the amino acid residues of L-Tyr-containing anabaenopeptin A (Figure 3). The first adenylation domain, AptA$_1$-1, was not obtained in soluble form under any conditions tested, and no expression of the first adenylation domain at the low temperatures (below 18°C) was detected with PAGE-runs. The second adenylation domains of AptA$_1$ and AptA$_2$ both activate L-Lys (Figure 3), and the second module in each protein carries an epimerase domain at the C terminus (Figure 1 and 2A). It was not possible

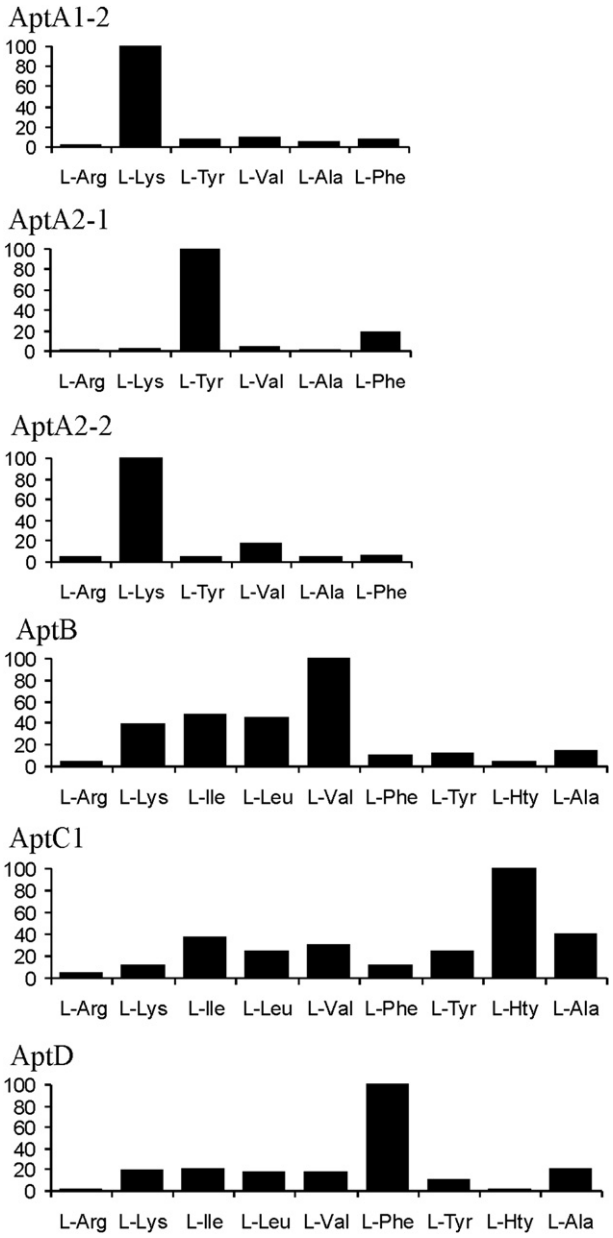

Figure 3. Substrate Specificity of *Anabaena* Sp

Strain 90 anabaenopeptin synthetase adenylation domains by ATP-pyrophosphate exchange assay. Amino acids used as substrates were those found in anabaenopeptins A, B, and C and the chemically related amino acids.

to test the substrate specificity of the first adenylation domain of the $AptA_1$ starter module. However, the specificity code of the first starter module, DVEDIGSVAK, differs only slightly from those for L-Lys, DTEDIGSVIK, and DTEDIGSVVK (Table 2).

Other than NRPS Genes of the Cluster

Two ORFs, *aptE* and *aptF*, present in the anabaenopeptin biosynthetic gene cluster and coding for the proteins with the sizes of 392 and 765 amino acids, were also identified (Table 1). AptE showed high similarity in the BLAST search only to the proteins of *Nodularia spumigena* CCY9414 and *Nostoc punctiforme* PCC 73102, annotated as 2-isopropylmalate synthase (accession, EAW45420) and pyruvate carboxyltransferase (accession, ACC81026) (82%/96% and 88%/97% identity/similarity in 392 amino acids), respectively. The highest similarity of AptF in BLAST searches was found with the ABC transporter-like proteins of *Nodularia spumigena* CCY9414 (accession, EAW45421) and *Nostoc punctiforme* PCC 73102 (accession, ACC81026), 73%/89% and 69%/82% identity/similarity, respectively (Table 1). InterProScan results suggest the presence of the pyruvate carboxyltransferase domain in AptE and that it belongs to the family of HMGL (hydroxymethylglutaryl-coenzyme A lyase)-like proteins. AptF belongs to a family of ABC transporter-like proteins.

Anabaenopeptin Synthetase Gene Clusters from *Nodularia spumigena* CCY9414 and *Nostoc punctiforme* PCC 73102

BLAST searches using the *apt* gene cluster as a query against the nonredundant database at NCBI detected two gene clusters, which bear a great resemblance to the anabaenopeptin synthetase gene cluster: one from *Nodularia spumigena* CCY9414 (AAVW01000028) and the other from *Nostoc punctiforme* PCC 73102 (CP001037) (Figure 2B and Table 1). Both of these gene clusters encode four NRPSs containing six modules (Figure 2B). The substrate specificity codes for the same amino acids are highly similar in different anabaenopeptin synthetases (Table 2).

FASTA comparison of the Apt-proteins with the corresponding proteins of *Nodularia spumigena* CCY9414 and *Nostoc punctiforme* PCC 73102 showed high similarity, from 69% to 88% identity in amino acid sequences of NRPSs (Table 1). $AptA_1$ showed 58%/82% identity/similarity to $AptA_2$ and the NRPS coded by the first gene of *Nostoc punctiforme* PCC 73102, and 57%/81% identity/similarity to the corresponding protein of *Nodularia spumigena* CCY9414. Comparison of the putative lysine activating, epimerase domain-containing modules with that of $AptA_1$ gave slightly different results from those with the whole proteins: 62%/84% and 60%/83%, respectively. The corresponding module of $AptA_2$ showed 73%/89% and 79%/92% identity/similarity to these epimerase domain-containing modules of *Nodularia spumigena* CCY9414 and *Nostoc punctiforme* PCC 73102, respectively. The identity/similarity of epimerase domain-containing modules in $AptA_1$ and $AptA_2$ was 61%/85%.

Structures of Anabaenopeptins Produced by *Anabaena* Sp. Strain 90, *Nodularia spumigena* CCY9414 and *Nostoc punctiforme* PCC 73102

Anabaenopeptins A, B and C were the main anabaenopeptins detected in *Anabaena* sp. strain 90 and comprised 34%, 26%, and 34% of the total, respectively (Table 3). The relative amounts of other variants including anabaenopeptins F, oscillamide Y, and new variants were all under 2%. Only Tyr, Arg and Lys were found at position one. The third amino acid position of anabaenopeptins was the most relaxed and contained a ratio of Val to Ile/Leu of 96/4. Trace amount of homophenylalanine (Hph) was incorporated at position four, methylglycine (Mgly) and methylserine (Mser) at position five and homotyrosine (Hty) to

Table 2. Substrate Specificity Code of Anabaenopeptin Synthetase Modules and Amino Acids Incorporated

Anabaena sp. strain 90			*Nostoc punctiforme* PCC 73102			*Nodularia spumigena* CCY9414		
	Substrate specificity code[a]	Amino acids incorporated[b, c]		Substrate specificity code	Amino acids incorporated		Substrate specificity code	Amino acids incorporated
Module 1	DVEDIGSVAK	Arg/Lys	-			-		
Module 2	DTEDIGSVIK	D-Lys	-			-		
Module 3	DASTIAAVCK	Tyr	Module 1	DAWTIAAICK	Phe	Module 1	DAFFLGVTYK	Ile
Module 4	DTEDIGSVVK	D-Lys	Module 2	DAEDIGSVIK	D-Lys	Module 2	DTEDIGTVVK	D-Lys
Module 5	DALWLGGVFK	Val	Module 3	DALFMGVVFK	Ile	Module 3	DMWFMGGVIK	Met/MetO/Ile
Module 6	DLGFTGCVTK	Hty[d]	Module 4	DLGTIGCVIK	Hph/Hty	Module 4	DLGAIGCVIK	Hph
Module 7	DLFNNALTYK	MeAla	Module 5	DILQLGVIWK	MeGly	Module 5	DLGFSGCVT	MeHty/MeHph
Module 8	DAWTIAGVCK	Phe	Module 6	DLGFTGCVNK	Hty	Module 6	DVWTITAICK	AcSer/Ser

[a] Determined according to Stachelhaus et al. (1999).
[b] Based on the substrate specificity (*Anabaena* sp. strain 90) and the MS data of this work.
[c] Amino acids are in L- form except D-Lys based on the sequence analysis of this study: epimerization domain was found only in lysine incorporating module of the each synthetase.
[d] Hty is homotyrosine; Me is *N*-methyl-; Hph homophenylalanine; MetO methionine sulfoxide found in peptides produced by *Nodularia spumigena* CCY9414, but it most probably is the oxidation product of incorporated methionine; AcSer is *O*-acetyl serine.

position six. These results were consistent with the substrate specificities of the expressed adenylation domains (Figure 3).

Nodulapeptins B and C (31% and 27% of the total, respectively) containing L-Ile at the first position were the major anabaenopeptins found in the extracts from *Nodularia spumigena* CCY9414 (Table 3). Oxidized derivatives of methionine, methionine sulfoxide (MetO) and methionine sulfone (MeO_2), are probably noncatalytic oxidation products of incorporated methionine. When this is taken into account, position three incorporated about 90% methionine and 10% isoleucine/leucine. Valine was consistently incorporated in trace amounts at position one (<1%), Hph almost exclusively at position four, approximately 90% N-methyl-Hty and 10% N-methyl-Hph at position five, and about 80% O-acetylserine, 10% serine, and 10% methionine (MetO included) at position six. All but two of the identified variants (B, C) were previously unknown. Altogether 28 different variants were identified, but if MetO and $MetO_2$ are considered as noncatalytic oxidation products, the total number of different biosynthetically produced anabaenopeptins was 19.

Nostoc punctiforme PCC 73102 produced two anabaenopeptins containing L-Phe as the first amino acid. Anabaenopeptin NZ 857 (Grach-Pogrebinsky and Carmeli, 2008) and a new anabaenopeptin variant, nostamide A, were almost identical but the latter contained Hph instead of Hty. Product ion spectra from protonated molecular ions (MS^2) and from ions representing anabaenopeptin ring structure (MS^3) together with ion assignments confirming the structures are presented in Figure S1 and Table S1 (available online).

Extra Modules Are Common in Anabaenopeptin Producing *Anabaena* Strains

We designed polymerase chain reaction (PCR) primers spanning the intergenic region between $aptA_1$ and $aptA_2$. All the tested anabaenopeptin-producing *Anabaena* strains gave positive

Table 3. Major Anabaenopeptins Found in the Cell Extracts of *Anabaena* Sp. Strain 90, *Nodularia spumigena* CCY9414 and *Nostoc punctiforme* PCC 73102

		Amino Acids in Different Positions[a]						Relative Amount[b]
Producer	Peptide	1	2	3	4	5	6	(%)
Anabaena sp. strain 90	Anabaenopeptin A	L-Tyr	D-Lys	L-Val	L-Hty	MeAla	L-Phe	34
	Anabaenopeptin B	L-Arg	D-Lys	L-Val	L-Hty	MeAla	L-Phe	26
	Anabaenopeptin C	L-Lys	D-Lys	L-Val	L-Hty	MeAla	L-Phe	34
Nodularia spumigena CCY9414	Nodulapeptin B	L-Ile	D-Lys	MetO	L-Hph	MeHty	AcSer	31
	[MHph^5]Nodulapeptin B	L-Ile	D-Lys	MetO	L-Hph	MeHph	AcSer	5
	[Ser^6]Nodulapeptin B	L-Ile	D-Lys	MetO	L-Hph	MeHph	L-Ser	4
	Nodulapeptin C	L-Ile	D-Lys	L-Met	L-Hph	MeHty	AcSer	27
	[Met^6]Nodulapeptin C	L-Ile	D-Lys	L-Met	L-Hph	MeHty	L-Met	4
Nostoc punctiforme PCC73102	Anabaenopeptin NZ 857	L-Phe	D-Lys	L-Ile	L-Hty	MeGly	L-Hty	60
	Nostamide A	L-Phe	D-Lys	L-Ile	L-Hph	MeGly	L-Hty	40

[a] L-Hty is L-homotyrosine; Me means *N*-methyl-; L-Hph is L-homophenylalanine; MetO is methionine sulfoxide found in peptides produced by *Nodularia spumigena* CCY9414, but it most probably is the oxidation product of incorporated methionine; AcSer is *O*-acetyl serine.
[b] Quantitative analysis based on the positive ion chromatograms of the protonated peptides. Relative amounts of variants smaller than four percent of the total were not included.

Table 4. Anabaenopeptins of *Anabaena* Strains and PCR Results with the *apt*-Primers Aptlys11f and Apt31r Indicating the Presence of Alternative Starter Modules

Strain	Origin	Anabaenopeptin A	B	C	D	$aptA_1$-$aptA_2$ PCR
90	Finland	+	+	+		+
66B	Finland		+	+	+	+
NIVA-CYA 83/1	Norway		+	+		+
202A1	Finland		+	+	+	+
202A2	Finland		+	+		+
246	Finland	+	+	+		+
250A	Finland	+	+	+		+
PH 256	Denmark	+	+	+		+
257	Finland	+	+	+		+
259	Finland		+	+	+	+
260	Finland	+	+	+		+
NRC 525	Canada	+	+			+

Anabaenopeptin A: L-Tyr - D-Lys - L-Val - L-Hty - Me-L-Ala - L-Phe
Anabaenopeptin B: L-Arg - D-Lys - L-Val - L-Hty - Me-L-Ala - L-Phe
Anabaenopeptin C: L-Lys - D-Lys - L-Val - L-Hty - Me-L-Ala - L-Phe
Anabaenopeptin D: L-Phe - D-Lys - L-Val - L-Hty - Me-L-Ala - L-Phe

results in PCR experiment to amplify the region of 1.4 kb, which corresponds the end part of $aptA_1$ (last 534 bp) and the beginning of $aptA_2$ (first 700 bp) (Table 4). The sequences of the PCR fragments (400–500 bp) were highly similar to $aptA_1$ (99% identity), but the similarity to $aptA_2$ was lower depending on which anabaenopeptins were found in the strain (77%–85% if anabaenopeptins B, C, and D and 99%–100% if A, B and C were found).

DISCUSSION

We characterized the 32.5 kb *apt* gene cluster from the cyanobacterium *Anabaena* sp. strain 90. Five genes code for the NRPSs comprising eight modules (Figures 1 and 2A), which direct the biosynthesis of anabaenopeptins according to the substrate specificity of the adenylation domains (Figure 3). A unique feature in this anabaenopeptin synthetase is the presence of two extra modules allowing the specific biosynthesis of anabaenopeptin variants (Figure 1). Except for the extra modules, the anabaenopeptin synthetase follows the colinearity rule of NRPSs (Marahiel et al., 1997), which means that the number of the modules and the order of the modular functions are in line with the structure of the peptide product. However, there are eight modules (Figures 1, 2A), whereas anabaenopeptin A, B, and C contain six amino acid residues (Figure 1 and Table 3) contradicting the colinearity rule. The substrate specificity assays showed that the adenylation domains of the second, third, fourth, fifth, sixth, and eighth modules, effectively activate L-Lys, L-Tyr, L-Lys, L-Val, L-Hty, and L-Phe, respectively (Figure 3). The second and fourth modules both activate L-Lys and bear epimerase domains to convert the L-amino acid to D-form (Figure 1). The adenylation domain of the first module was not expressed in soluble form and thus it was not possible to test the substrate specificity directly. However, the substrate specificity conferring amino acids of the second and fourth adenylation domains (Stachelhaus et al., 1999) are almost identical and both activating L-Lys (Table 2). The specificity code of the first module (DVEDIGSVAK) is very similar to that for L-Lys, (DTEDIGSVIK and DTEDIGSVVK) suggesting that the first module activates L-Lys, and also the other basic amino acid L-Arg (Table 2). The third module activated L-Tyr with high specificity (Figure 3). So, L-Arg and L-Lys, the alkaline amino acids, are most probably activated by the first module, and the third module activates only L-Tyr. Mass spectrometric analysis of the peptides produced by *Anabaena* sp. strain 90 (Table 3) and the high substrate specificity of these modules supports this conclusion.

Gene clusters similar to *Anabaena* sp. strain 90 *apt* genes were found in *Nodularia spumigena* CCY9414 and *Nostoc punctiforme* PCC 73102, and anabaenopeptins were also identified from these strains. The four NRPS genes code for six modules (Figure 2B and Table 1), which is in accordance with the number of amino acids in the produced peptides (Table 3) and different from the arrangement of the *apt* gene cluster in *Anabaena* sp. strain 90.

Anabaenopeptin synthetase gene cluster from *Planktothrix rubescens* NIVA-CYA 98 reported recently (Rounge et al., 2009) codes also for six modules in four genes (*anaA*, *anaB*, *anaC,* and *anaD*) as in *Nodularia spumigena* CCY9414 and *Nostoc punctiforme* PCC 73102. The polypeptides coded by the ORF5 in the gene clusters of *Nodularia spumigena* CCY9414 and *Nostoc punctiforme* PCC 73102 are very similar to AptE, 88%/97% and 82%/96% identity/similarity, respectively (Table 1). InterProScan search (Zdobnov and Apweiler, 2001) found in AptE similarity to the family of HMGL-like proteins. HMGL-family contains diverse enzymes including aldolases and a region of pyruvate carboxylase. However, the proteins highly similar to AptE are encoded in the gene clusters of *Nodularia spumigena* CCY9414 and *Nostoc punctiforme* PCC 73102 (Table 1), which strongly suggests a function related to the biosynthesis of anabaenopeptins. All the four peptide synthetases, $AptA_1$, $AptA_2$, and those coded by the first ORF in the clusters of *Nodularia spumigena* CCY9414, *Nostoc punctiforme* PCC 73102, and *Planktothrix rubescens* NIVA-CYA 98 contain typical adenylation and condensation domains of NRPSs with characteristic conserved motifs (Marahiel et al., 1997). Thus the ureido bond most likely results from modification taking place after the normal peptide bond is formed. We suggest that this modification is catalyzed by AptE in *Anabaena* sp. strain 90 and the corresponding proteins coded by the ORF5 in *Nodularia spumigena* CCY9414 and *Nostoc punctiforme* PCC 73102 (Table 1). The gene encoding protein similar to AptE was not reported in the cluster (AM990463) from *Planktothrix rubescens* NIVA-CYA 98 (Rounge et al., 2009) containing the anabaenopeptin synthetase genes. However, 251 bases upstream of *anaA* (in position 31981–33160), transcribed in the opposite direction, we found an ORF similar to *apt*E and with equal size. However, there was a frameshift, introduced by a single adenosine (in the position 32576) in a group of six adenosines. Removing this adenosine leads to a protein of 392 aa showing 72.2% identity/90.6% similarity to AptE in 392 aa overlap in FASTA comparison (Pearson and Lipman, 1988). It remains to be confirmed if this is a genuine frameshift or a sequencing error.

Syringolin A produced by phytopathogenic *Pseudomonas syringae* pv. *syringae* strains is a cyclic tetrapeptide containing a ureido linkage between two side-chain valine residues, which are joined by a peptide bond to a dehydrolysine moiety, that closes the ring with 5-methyl-4-amino-2-hexenoic acid (Amrein et al., 2004). The gene cluster (*sylA*, *B*, *C*, *D,* and *E*) from *Pseudomonas syringae* pv. *syringae* B301D-R (Amrein et al., 2004) involved in the biosynthesis of syringolin A was proposed not to include a complete set of the genes required for the biosynthesis of syringolin A. The genes coding for a starter module and ureido bond formation were missing. *Pseudomonas syringae* pv. *syringae* B728a, which is completely sequenced (Feil et al., 2005), carries the *syl*-genes from *sylA* to *sylE*. However, BLAST search found no proteins similar to AptE coded by *syl* genes of *Pseudomonas syringae* pv. *syringae* B301D-R or in the genome of *Pseudomonas syringae* pv. *syringae* B728a. It was recently reported that SylC catalyzes the formation of a ureido bond (Ramel et al., 2009; Imker et al., 2009). SylC carries between the condensation and the adenylation domains a 239 amino acids section (aa 432 to aa 670), which is not present in $AptA_1$, $AptA_2$, or other Apt proteins, and shows only low similarity to any other protein in BLAST search. InterProScan search found structural similarity with acyltransferases in this part of SylC (superfamily: CoA-dependent acyltransferases). That SylC bears this unique domain not present in anabaenopeptin synthetase and *syl*-cluster does not contain AptE-like protein suggests a different mechanism for the formation of the ureido linkage in the biosyntheses of syringolin A and anabaenopeptins.

Nodularia spumigena CCY9414 and *Nostoc punctiforme* PCC 73102 produced two major anabaenopeptins with practically no variation in the first amino acid (Table 3). In *Anabaena* sp. strain 90 the structural variants of anabaenopeptins are achieved by the extra module, which results in the arginine and lysine variants, anabaenopeptin B and C. This implies that the nature of first residues has a selective value to the producing organism and is not simply a lack of substrate specificity.

PCR experiments indicated that the additional modules are common in anabaenopeptin-producing *Anabaena* strains (Table 4). The sequences of the PCR fragments (400 bp) were almost identical (99%) with the corresponding epimerase domain coding part of $aptA_1$, which has lower similarity to the first ORFs of *Nodularia spumigena* CCY9414 and *Nostoc punctiforme* PCC 73102 than $aptA_2$. This suggests that $aptA_1$ may be the result of lateral gene transfer or have resulted from gene duplication and subsequent intragenomic recombinations, or the corresponding gene was lost in *Nodularia spumigena* CCY9414 and *Nostoc punctiforme* PCC 73102. However, *anaA* of *Planktothrix rubescens* NIVA-CYA 98 (Rounge et al., 2009) was more similar to $aptA_1$ than to $aptA_2$, because this strain produces the arginine variant, anabaenopeptin B, which is the product of the $AptA_1$ containing synthetase.

Many peptides in anabaenopeptin class are protease inhibitors (Itou et al., 1999; Murakami et al., 2000; Sano et al., 2001). Inhibition of trypsin and the elastase activity by anabaenopeptin B has been shown (Repka et al., 2004; Bubik et al., 2008). The biological function of anabaenopeptins in cyanobacteria is still unknown. However, they share the following structural features: a single amino acid side chain; and a ureido bond connecting this to D-lysine, which initiates and closes the hydrophobic ring of five amino acids that typically include Hty or Hph in position four and an N-methylated amino acid in position five (Table 3). Anabaenopeptins exhibit considerable structural variation (Welker and von Döhren, 2006) and the mechanisms underlying this variation remain to be fully explored.

SIGNIFICANCE

Anabaenopeptin synthetase of *Anabaena* sp. strain 90 represents a novel strategy to create variation in the peptide structures produced by linear NRPSs. The *apt* synthetase of *Anabaena* spp. makes use of alternative starter modules with high substrate specificity to assemble a group of peptides with limited structural variance. This unique mechanism could be exploited to synthesize peptides with specified structures, and it provides a new model for producing peptides by combinatorial biosynthesis.

EXPERIMENTAL PROCEDURES

Cyanobacterial Strains and Growth Conditions

The cyanobacterium *Anabaena* sp. strain 90 was isolated from Lake Vesijärvi, Finland, in 1986, and made axenic (Rouhiainen et al., 1995). It was shown to produce three anabaenopeptins A, B, and C (Fujii et al., 2002). The other *Anabaena* strains used in this study are listed in Table 4. Anabaenopeptins produced by *Anabaena flos-aquae* NRC 525 were previously identified (Harada et al., 1995); other strains were analyzed in this study. *Anabaena* strains and *Nostoc punctiforme* PCC 73102 were grown in Z8 medium (Kotai, 1972). *Nodularia spumigena* CCY9414 was grown in artificial sea water made from one part ASN3 and two parts of BG11 medium (Rippka et al., 1979), both media were devoid of combined nitrogen. The strains were cultivated at 23°C ± 1°C with continuous illumination of 20–25 μmol $m^{-2}s^{-1}$.

DNA Methods

Cyanobacterial DNA was isolated following the method described earlier (Golden et al., 1988). The gene sequencing was performed using a cosmid library made in Lorist 6 (Rouhiainen, et al., 2000) and two shotgun libraries in pUC18 plasmid vector with Megabase 1000 (GE Healthcare) and ABI 3730 sequencer (Applied Biosystems) in the genome center of Beijing Genomics Institute, Chinese Academy of Sciences. *Escherichia coli* strain DH5α was used as the host for DNA cloning and manipulation.

Analysis of Anabaenopeptin Synthetase Gene Cluster Sequences

The complete genome of *Nostoc punctiforme* PCC 73102 and the unfinished genome of *Nodularia spumigena* CCY9414 are available in the GenBank database. The sequence analysis was done with the programs of EMBOSS (The European Molecular Biology Open Software Suite) (Rice et al., 2000). Sequence similarity searches in databases were carried out with BLAST (Altschul et al., 1990) and FASTA comparison (Pearson and Lipman, 1988). CLUSTALW (Thompson et al., 1994) was applied for multiple sequence alignments. The substrate specificity conferring amino acids of the peptide synthetase modules (Stachelhaus et al., 1999) were identified by using the NRPS predictor program (Rausch et al., 2005) and confirmed by aligning with GrsA (Stachelhaus et al., 1999). The InterProScan Sequence Search (Zdobnov and Apweiler, 2001) was used for the functional comparison of the proteins.

PCR to Test the Presence of Extra Modules in Anabaenopeptin-Producing *Anabaena* Strains

The primers for the amplification of the end part of $aptA_1$ (last 534 bp) and the beginning of $aptA_2$ (first 700 bp), plus the sequence between (total length of 1.4 kb), were aptlys11 F, 5′-GATYTAGAGGGACATGGACG-3′ and apt31 R, 5′-AGAAGCAAACTGGAGAACACG-3′. Reaction conditions for PCR were as follows: one cycle with denaturation at 95°C for 2 min, primer annealing at

55°C for 40 s and extension at 70°C for 30 s; 25 cycles with denaturation at 95°C for 30 s, primer annealing at 55°C for 40 s and extension at 72°C for 30 s. Reactions were ended with the final extension step at 72°C for 5 min. DyNAzyme II polymerase (Finnzymes), 0.1 U, was used in 20 μl including buffer (10 mM Tris-HCl [pH = 8.8]; 1.5 mM $MgCl_2$; 50 mM KCl and 0.1% Triton X-100), 0.2 mM of each nucleotide, 0.5 μM primers, and 50 ng DNA. PCR was done with total DNA from 12 *Anabaena* strains. The sequences of the ends of PCR products were determined with cycle sequencing by ABI 310 Genetic Analyzer (Applied Biosystems).

Amplification and Expression of Adenylation Domains for Substrate Specificity Determination

Seven of the eight adenylation domains of the putative anabaenopeptin synthetase were amplified by PCR with *Pfu* DNA polymerase (Fermentas) as previously described (Fewer et al., 2009). The primers for the amplification of the adenylation domains are listed in Table S2. PCR products were purified with the "Quick Step 2" PCR purification kit (Edge Bio Systems) and cloned in pET101/D-TOPO expression vector (Invitrogen) according to the instructions of the manufacturer. *E. coli* strain TOP10 (Invitrogen) was used for transformation and to amplify the clones. The end sequences and correct orientation of the clones were verified by using cycle sequencing and ABI 310 Genetic Analyzer (Applied Biosystems).

The plasmids with the cloned adenylation domains were used to transform *E. coli* BL21 Star (DE3) (Invitrogen). Expression and purification of the adenylation domains were done as previously described (Fewer et al., 2009).

Substrate Specificity of the Adenylation Domains

The ATP pyrophosphate exchange assay (ATP PPi exchange assay) was performed as described earlier (Gruenewald et al., 2004), except that the reaction buffer was 50 mM Tris-HCl (pH 7.3), containing 10 mM $MgCl_2$ and 5 mM DTT, and the reaction time 15 min (at 37°C). All the reactions were done as duplicates. After adding 3 ml liquid scintillation fluid (Optiphase Hisafe3, Perkin Elmer), radioactivity was measured in Wallac 1411 Liquid Scintillation Counter.

Structure Analysis of Anabaenopeptins

Cyanobacterial cells were collected from the liquid cultures by centrifugation for 7 min at 7000 g, and freeze-dried. Dry cell material from *Anabaena* sp. strain 90 (100 mg), other *Anabaena* strains (5–50 mg), *Nostoc punctiforme* PCC 73102 (30 mg), and *Nodularia spumigena* CCY9414 (10 mg) was extracted for 20 s with 1 ml methanol:dimethylsulfoxide (7:3) or methanol in 2 ml plastic tubes containing approximately 200 μl 0.5 mm glass beads (Scientific Industries Inc.) using Fast Prep homogenizer (FP120, Bio 101, Savant) at speed value of 6. Extracts were centrifuged for 5 min at 10,000 g prior to liquid-chromatography mass spectrometry (LC-MS) analysis.

LC-MS analyses of extracts were performed with an Agilent 1100 Series LC / MSD Ion Trap XCT Plus System (Agilent Technologies, Palo Alto, CA) using a Phenomenex Luna C18(2) (150 × 2.0 mm, 5 μm, Phenomenex, Torrance, CA) LC column. The mobile phase was composed of 0.1% formic acid in water (A) and 2-propanol (B). The gradient run was from 10% B to 50% (B) over 60 min and then 10 min wash with 100% (B) at a flow rate of 0.15 ml min^{-1} at 40°C using 5 or 10 μl injections. Anabaenopeptins were detected with a diode array detector and with MS using electrospray ionization set in positive mode. The nebulizer gas (N_2) pressure was 30–35 psi (210–240 kPa), desolvation gas flow rate 8 l/min and the desolvation temperature 350°C. The capillary voltage was set to 5000 V (3200 V for *Nodularia spumigena* CCY9414), the capillary exit offset was 300 V, the skimmer potential was 85 V (42.0 V for *Nodularia spumigena* CCY9414), and the trap drive value was 144 (85.0 for *Nodularia spumigena* CCY9414). Spectra were recorded at a scanning range of 200 to 1100 m/z and rate of 26,000 m/z s^{-1}. Structures of anabaenopeptins were confirmed by generating product ion spectra using multiple reactions monitoring mode or auto MS^3 mode for *Nostoc punctiforme* PCC 73102.

ACCESSION NUMBERS

The nucleotide sequence of the anabaenopeptin synthetase gene cluster has been deposited in the GenBank under the accession number GU174493.

SUPPLEMENTAL INFORMATION

Supplemental Information includes one figure and two tables, and can be found with this article online at doi:10.1016/j.chembiol.2010.01.017.

ACKNOWLEDGMENTS

This work was supported by grants from the European Union (CYANOTOX, ENV4-CT98-0802) and the Academy of Finland (grant 214457 for academy professor, grants 53305 and 118637 for research centers of excellence) to K.S. We are grateful to Lyudmila Saari for her valuable help in handling the cultures.

Received: December 22, 2009
Revised: January 27, 2010
Accepted: January 28, 2010
Published: March 25, 2010

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Meyers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Amrein, H., Makart, S., Granado, J., Shakya, R., Schneider-Pokorny, J., and Dudler, R. (2004). Functional analysis of genes involved in the synthesis of syringolin A by *Pseudomonas syringae* pv. *syringae* B301D-R. Mol. Plant Microbe Interact. *17*, 90–97.

Bubik, A., Sedmak, B., Novinec, M., Lenarcic, B., and Lah, T.T. (2008). Cytotoxic and peptidase inhibitory activities of selected non-hepatotoxic cyclic peptides from cyanobacteria. Biol. Chem. *389*, 1339–1346.

Burja, A.M., Banaigs, B., Abou-Mansour, E., Burgess, J.G., and Wright, P.C. (2001). Marine cyanobacteria: a prolific source of natural products. Tetrahedron *57*, 9347–9377.

Feil, H., Feil, W.S., Chain, P., Larimer, F., DiBartolo, G., Copeland, A., Lykidis, A., Trong, S., Nolan, M., Goltsman, E., et al. (2005). Comparison of the complete genome sequences of *Pseudomonas syringae* pv. *syringae* B728a and pv. *tomato* DC3000. Proc. Natl. Acad. Sci. USA *102*, 11064–11069.

Fewer, D.P., Rouhiainen, L., Jokela, J., Wahlsten, M., Laakso, K., Wang, H., and Sivonen, K. (2007). Recurrent adenylation domain replacement in the microcystin synthetase gene cluster. BMC Evol. Biol. *7*, 183.

Fewer, D.P., Jokela, J., Rouhiainen, L., Wahlsten, M., Koskenniemi, K., Stal, L., and Sivonen, K. (2009). The non-ribosomal assembly and frequent occurrence of the protease inhibitors spumigins in the bloom-forming cyanobacterium *Nodularia spumigena*. Mol. Microbiol. *73*, 924–937.

Finking, R., and Marahiel, M.A. (2004). Biosynthesis of nonribosomal peptides. Annu. Rev. Microbiol. *58*, 453–488.

Fujii, K., Sivonen, K., Nakano, T., and Harada, K.-I. (2002). Structural elucidation of cyanobacterial peptides encoded by peptide synthetase gene in *Anabaena* species. Tetrahedron *58*, 6863–6871.

Golden, J.W., Carrasco, C.D., Mulligan, M.E., Scneider, G.J., and Haselkorn, R. (1988). Deletion of a 55-kilobase-pair DNA element from the chromosome during heterocyst differentiation of *Anabaena* sp. strain PCC 7120. J. Bacteriol. *170*, 5034–5041.

Grach-Pogrebinsky, O., and Carmeli, S. (2008). Three novel anabaenopeptins from the cyanobacterium *Anabaena* sp. Tetrahedron *64*, 10233–10238.

Gruenewald, S., Mootz, H.D., Stehmeier, P., and Stachelhaus, T. (2004). In vivo production of artificial nonribosomal peptide products in the heterologous host *Escherichia coli*. Appl. Environ. Microbiol. *70*, 3282–3291.

Harada, K., Fujii, K., Shimada, T., Suzuki, M., Sano, H., Adachi, K., and Carmichael, W.W. (1995). Two cyclic peptides, anabaenopeptins, a third group of bioactive compounds from the cyanobacterium *Anabaena flos-aquae* NRC 525-17. Tetrahedron Lett. *36*, 1511–1514.

Imker, H.J., Walsh, C.T., and Wuest, W.M. (2009). SylC catalyzes ureido-bond formation during biosynthesis of the proteasome inhibitor syringolin A. J. Am. Chem. Soc. *191*, 18263–18265.

Itou, Y., Suzuki, S., Ishida, K., and Murakami, M. (1999). Anabaenopeptins G and H, potent carboxypeptidase A inhibitors from the cyanobacterium *Oscillatoria agardhii* (NIES-595). Bioorg. Med. Chem. Lett. *9*, 1243–1246.

Kotai, J. (1972). Instructions for Preparation of Modified Nutrient Solution Z8 for Algae. Publication B-11/69, Norwegian Institute for Water Research (Oslo: Blindern).

Lautru, S., and Challis, G.L. (2004). Substrate recognition by nonribosomal peptide synthetase multi-enzymes. Microbiology *150*, 1629–1636.

Lawen, A., and Traber, R. (1993). Substrate specificities of cyclosporin synthetase and peptolide SDZ 214-103 synthetase. Comparison of the substrate specificities of the related multifunctional polypeptides. J. Biol. Chem. *268*, 20452–20465.

Marahiel, M.A., Stachelhaus, T., and Mootz, H.D. (1997). Modular peptide synthetases involved in nonribosomal peptide synthesis. Chem. Rev. *97*, 2651–2673.

Moore, R.E. (1996). Cyclic peptides and depsipeptides from cyanobacteria: a review. J. Ind. Microbiol. *16*, 134–143.

Murakami, M., Suzuki, S., Itou, Y., Kodani, S., and Ishida, K. (2000). New anabaenopeptins, potent carboxypeptidase-A inhibitors from the cyanobacterium *Aphanizomenon flos-aquae*. J. Nat. Prod. *63*, 1280–1282.

Namikoshi, M., and Rinehart, K.L. (1996). Bioactive compounds produced by cyanobacteria. J. Ind. Microbiol. *17*, 373–384.

Pearson, W.R., and Lipman, D.J. (1988). Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA *85*, 2444–2448.

Ramel, C., Tobler, M., Meyer, M., Bigler, L., Ebert, M.-O., Schellenberg, B., and Dudler, R. (2009). Biosynthesis of the proteasome inhibitor syringolin A: the ureido group joining two amino acids originates from bicarbonate. BMC Biochem. *10*, 26.

Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W., and Huson, D.H. (2005). Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using Transductive Support Vector Machines (TSVM). Nucleic Acids Res. *33*, 5799–5808.

Repka, S., Koivula, M., Harjunpää, V., Rouhiainen, L., and Sivonen, K. (2004). Effects of phosphate and light on growth and bioactive peptide production of a cyanobacterium, *Anabaena* strain 90 and its anabaenopeptilide-mutant. Appl. Environ. Microbiol. *70*, 4551–4560.

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet. *16*, 276–277.

Rippka, R., Deruelles, J., Waterbury, J.B., Herdman, M., and Stanier, R.Y. (1979). Generic assignments, strain histories and properties of pure cultures of cyanobacteria. J. Gen. Microbiol. *111*, 1–61.

Rouhiainen, L., Sivonen, K., Buikema, W.J., and Haselkorn, R. (1995). Characterization of toxin-producing cyanobacteria by using an oligonucleotide probe containing a tandemly repeated heptamer. J. Bacteriol. *177*, 6021–6026.

Rouhiainen, L., Paulin, L., Suomalainen, S., Hyytiäinen, H., Buikema, W.J., Haselkorn, R., and Sivonen, K. (2000). Genes encoding synthetases of cyclic depsipeptides, anabaenopeptilides, in *Anabaena* strain 90. Mol. Microbiol. *37*, 156–167.

Rouhiainen, L., Vakkilainen, T., Lumbye Siemer, B., Buikema, W.J., Haselkorn, R., and Sivonen, K. (2004). Genes coding for the synthesis of hepatotoxic heptapeptides (microcystins) in the cyanobacterium *Anabaena* strain 90. Appl. Environ. Microbiol. *70*, 686–692.

Rounge, T.B., Rohrlack, T., Nederbragt, A.J., Kristensen, T., and Jakobsen, K.S. (2009). A genome-wide analysis of nonribosomal peptide synthetase gene clusters and their peptides in a *Planktothrix rubescens* strain. BMC Genomics *10*, 396.

Sano, T., Usui, T., Ueda, K., Osada, H., and Kaya, K. (2001). Isolation of new protein phosphatase inhibitors from two cyanobacteria species, *Planktothrix* spp. J. Nat. Prod. *64*, 1052–1055.

Stachelhaus, T., Mootz, H.D., and Marahiel, M.A. (1999). The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. Chem. Biol. *6*, 493–505.

Sieber, S.A., and Marahiel, M.A. (2005). Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. Chem. Rev. *105*, 715–738.

Sivonen, K., and Börner, T. (2008). Bioactive compounds produced by cyanobacteria. In The Cyanobacteria: Molecular Biology, Genomics and Evolution, A. Herrero and E. Flores, eds. (Linton, UK: Caister Academic Press), pp. 159–197.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighing, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res. *22*, 4673–4680.

von Döhren, H., Keller, U., Vater, J., and Zocher, R. (1997). Multifunctional peptide synthetases. Chem. Rev. *97*, 2675–2705.

Welker, M., and von Döhren, H. (2006). Cyanobacterial peptides-Nature's own combinatorial biosynthesis. FEMS Microbiol. Rev. *30*, 530–563.

Zdobnov, E.M., and Apweiler, R. (2001). InterProScan - an integration platform for the signature-recognition methods in InterPro. Bioinformatics *17*, 847–848.