Examination of Substrate Specificity of the First Adenylation Domain in *mcyA* Module Involved in Microcystin Biosynthesis

Gozde Yaman [ID][1], **Mete Yilmaz** [ID][1,2,*]

[1]Program in Biotechnology, Graduate School of Natural and Applied Sciences, Bursa Technical University, Bursa, Turkey

[2]Department of Bioengineering, Faculty of Engineering and Natural Sciences, Bursa Technical University, Bursa, Turkey

Abstract: The cyanotoxin microcystin (MC) is a secondary metabolite, synthesized by nonribosomal peptide synthetase (NRPS) and polyketide synthase (PKS) enzymes. It has many isoforms and the mechanism of its diversity is not well understood. One of the MC synthetase genes, *mcyA,* codes for the McyA module containing two adenylation (A) domains. The first domain, McyA-A1, generally binds to L-serine (L-ser). Then the *N*-methyl transferase (NMT) domain converts L-Ser into *N*-methyldehydroalanine (Mdha), which usually occupies position 7 on the MC molecule. However, various other amino acids (AAs) might also be present at this position. In this study, bioinformatic analyses of selected cyanobacteria were performed to understand whether genetic information in the first adenylation domain of *mcyA* could explain incorporation of different AAs at position 7 of the MC molecule. Binding pocket signatures of McyA-A1 and putative activated AAs were determined via various bioinformatics tools. Maximum likelihood phylogenetic trees of full length *mcyA*, *mcyA*-A1 and 16S rRNA genes were prepared in Mega 6. Phylogenetic analysis of *mcyA*-A1 nucleotide sequences was in agreement with the predictions of activated AAs by McyA-A1. In comparison with the 16S rRNA and full length *mcyA* gene trees, *mcyA*-A1 phylogenetic trees suggested horizontal transfer of the A domain in either *Planktothrix agardhii* (Gomont) Anagnostidis & Komárek or *Planktothrix rubescens* (De Candolle ex Gomont) Anagnostidis & Komárek strains. Predictions of activated AAs were generally in agreement with the chemically determined position 7 AAs. However, there were exceptions suggesting the multispecificity of the first A domain of McyA in some cyanobacteria.

1. INTRODUCTION

There are approximately 500 natural amino acids [1] and they can be classified based on their R groups, stereoisomer structures (D- and L-form), being proteinogenic or nonproteinogenic. Proteinogenic amino acids are all 20 kinds of L-form amino acids, where mRNA, tRNA and ribosomes are involved in the protein synthesis pathway. Nonproteinogenic amino acids are D-form or β-amino acids and synthesized by either non-ribosomal peptide

synthetases (NRPSs), polyketide synthases (PKSs) or hybrid NRPS/PKS modular enzyme systems.

Cyanobacteria can produce a variety of toxic secondary metabolites (cyanotoxins) which are generally subdivided into neurotoxins, hepatotoxins, cytotoxins and dermatotoxins. Cyanotoxins are usually synthesized nonribosomally [2,3]. Microcystin (MC) is also a nonribosomally synthesized cyanotoxin, which shows hepatotoxic and neurotoxic effects, inhibiting eukaryotic serine/threonine protein phosphatases PP1 and 2 [4,5]. Its general structure is cyclo(-D-Ala^1-X^2-D-MeAsp^3-Z^4-Adda^5-D-Glu^6-Mdha^7), where X and Z are various L-amino acids (Figure 1) and D-Glu is D-iso-glutamic acid, Mdha is *N*-methyl-dehydroalanine, D-Ala is D-Alanine, D-MeAsp is D-*erythro*-β-methyl aspartic acid and Adda is (2*S*,3*S*,4*E*,6*E*,8*S*,9*S*)-3-amino-9-methoxy-2,6,8-trimethyl-10-phenyldeca-4,6-dienoic acid. Microcystin producing species are found in various genera including *Microcystis, Anabaena, Nostoc, Planktothrix (Oscillatoria), Phormidium, Fischerella* and *Hapalosiphon* [6,7].
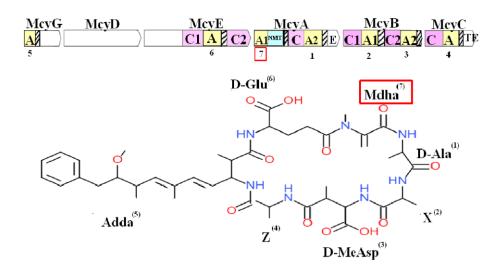

Figure 1. Schematic representation of MC synthetase operon (top) and MC structure (below). Position 7, which is investigated in this study, is marked. A, adenylation domain; C, condensation domain; NMT *N*-methyl transferase domain; TE, thioesterase domain. Peptidyl carrier protein is indicated in rectangular shape with lines.

Sequencing of MC biosynthesis genes was initially completed for *Microcystis aeruginosa* PCC 7806 [4], followed by *Planktothrix agardhii* NIVA CYA 126/8 [8] and *Anabaena* sp. 90 [9]. Within the MC synthetase operon, NRPS modules are expressed by *mcyA*, *-B* and *-C*; PKS modules are expressed by *mcyD*; hybrid NRPS / PKS modules are expressed by *mcyG* and *mcyE* genes. The *mcyJ, -F, -H, -I, -L* and *-T* genes do not have a modular structure and generally only contain a domain (Table 1). A module is a functional unit in NRPS or PKS and each module contains active units called domains. Adenylation (A), condensation (C) and peptidyl carrier protein (PCP) domains are the core domains in NRPS modules. The A domain recognizes the corresponding AA and activates it by using ATP. The PCP domain transfers the -SH (thiol) group to the activated residue. The C domain synthesizes the peptide bond between the -SH groups of the two activated residues. In addition, some modules may have modified domains called tailoring enzymes such as epimerization (E) and *N*-methyl transferase (NMT) domains [10].

The microcystin synthesis assembly line is started by McyG, which activates phenylacetate. Polyketide synthase modules of McyD, McyE and McyG complete the Adda structure at position 5. NRPS modules of McyA, McyB, McyC and McyG incorporate the

remaining AAs [4,5]. There are two A domains in McyA and the order of the domains is A1-NMT-PCP-C-A2-E in most MC producing species; however the NMT domain may not always be present (Table 1). Within these domains, the A1 domain usually activates L-Serine. Then, NMT domain transfers the methyl group from *S*-Adenosyl-L-methionine (SAM) to L-Serine followed or preceded by a dehydration reaction for the formation of Mdha at position 7 of the MC molecule. Various AAs can be incorporated in all 7 positions [11,12].

Table 1. Microcystin synthetase (*mcy*) genes and McyA domain order in cyanobacteria strains investigated in this study. Extra genes are marked in bold.

Cyanobacteria	mcy genes present	Domain order of McyA	Ref.
Anabaena sp. 90	A-B-C-D-E-F-G-H-I-J	A1-NMT-PCP-C-A2-T-E	[9]
Microcystis aeruginosa K-139	A-B-C-D-E-F-G-H-I-J	A1-NMT-PCP-C-A2-T-E	[9]
M. aeruginosa NIES-843	A-B-C-D-E-F-G-H-I-J	A1-NMT-PCP-C-A2-T-E	[3]
M. aeruginosa PCC7806	A-B-C-D-E-F-G-H-I-J	A1-NMT-PCP-C-A2-T-E	[4]
M. viridis NIES-102	A-B-C-D-E-F-G-H-I-J*	A1-NMT-PCP-C-A2-T-E*	
Fischerella sp. CENA161	A-B-C-D-E-F-G-H-I-J	A1-NMT-PCP-C-A2-T-E	[6]
Planktothrix agardhii NIVA-CYA 126/8	A-B-C-D-E-**G**-H-**J**-**T**	A1-NMT-PCP-C-A2-T-E	[8]
P. rubescens PCC7821	A-B-C-D-E-**G**-H-**J****	A1-PCP-C-A2-T-E	[11]
P. rubescens NIVA-CYA 98	A-B-C-D-E-**G**-H	A1-PCP-C-A2-T-E	
Nostoc sp.152	A-B-C-D-E-F-G-H-I-**L**	A1-NMT-PCP-C-A2-T-E	

*This study, Genbank: AP019314.1; **Genbank: CZCZ01000002. Abbreviations: A, adenylation domain; NMT, *N*-methyl-transferase domain; PCP, peptidyl carrier protein domain; C, condensation domain; E, epimerization domain.

The substrate specificity of A domains is important to understand the structure of secondary metabolites, including MC. The substrate specificity determination of A domains is usually carried out by determining of binding pocket residues of A domains via bioinformatics analyses. Ten residues constitute the binding pocket of an A domain and are numbered as "Positions-235, 236, 239, 278, 299, 301, 322, 330, 331 and 517" depending on their location in the GrsA module in *Bacillus brevis* (GenBank: CAA33603.1) [13,14]. Position-235 is generally occupied by Asp (Asp_{235}) and position-517 generally by a Lys residue (Lys_{517}). These residues are highly conserved, where Asp_{235} binds to $-NH_2$ group and Lys_{517} binds to -COOH group of activated AAs. Rest of the eight residues in the binding pocket can be variable and this variability is thought to be an important factor for A domain substrate specificity. It is suggested that A domain substrate specificity is associated with AA interactions between R groups of the activated AAs and R groups of the eight residues in the binding pocket [15,16].

There are nearly 280 variants of MCs [17] and the mechanisms that cause this diversity are not well understood. As more sophisticated techniques are used, both environmental samples and cyanobacteria cultures prove to produce more diverse and novel MC analogues [18]. Current data suggest that this diversity is based on either genetic diversity coded in the MC synthetase genes (e.g. insertions or deletions within various *mcy* genes) or biochemical status of MC producing cells [12]. For instance, point mutations for a few critical AAs in the A domain of GrsA were reported to change the substrate specificity of A domain in GrsA [13]. Fewer *et al*. [19] demonstrated recombination events between A domains of McyB1 and McyC. They reported that A domains of McyB1 and McyC were recombination hotspots for MC variants.

To our knowledge, there are no detailed bioinformatics studies regarding the first A domain of *mcyA*. Therefore, the main purpose of this study was to investigate whether genetic information in this domain could explain incorporation of different AAs at position 7 of the MC molecule. This is important since the toxicity of MCs might change with incorporation of

various AAs. Additionally, mechanisms explaining A domain AA selectivity will help in combinatorial biosynthesis of new NRPS and PKS enzymes [5]. For this purpose, ten MC producing cyanobacteria with sequenced MC synthetase genes were selected for bioinformatics analyses (i) to predict the binding pocket residues of the A1 domain in McyA (McyA-A1); (ii) to determine activated AAs by McyA-A1 based on the binding pocket residues; (iii) to investigate concordance between activated AAs and chemically determined position-7 AAs of MCs reported in the literature for the studied strains; (iv) to investigate concordance between activated AAs and phylogenetic clustering of *mcyA*-A1 sequences.

2. MATERIAL and METHODS

A detailed literature search was conducted to determine the best investigated cyanobacteria strains in terms of MC variant production to report the variation at position 7 of the MC molecule (Table 2). Ten cyanobacteria with sequenced MC synthetase genes and detailed MC characterizations were selected (Table 2) based on literature and GenBank searches. Full length *mcyA* and 16S rRNA gene nucleotide sequences of these cyanobacteria were downloaded from NCBI (National Center for Biotechnology Information) (https://www.ncbi.nlm.nih.gov) using Mega 6 [20]. Alignment was performed using Muscle [21] as implemented in Mega 6 [20] using codons for *mcyA* and nucleotides for 16S rRNA genes. Since *mcyA* is composed of domains, the borders of the first adenylation domain (mcyA-A1) in *mcyA* were determined using the NCBI conserved domain program [9]. Then the determined McyA-A1 domains were extracted from the full length *mcyA* alignment and a new codon-based alignment was performed after addition of the *B. brevis* GrsA domain nucleotide sequence. The nucleotide sequence alignment was converted to an AA alignment and binding pocket residues of McyA-A1 were predicted by comparison to the *B. brevis* GrsA domain AA sequences and by using an online tool [14]. Predictions of the activated AAs by the A domains were also performed using the same online tool.

In order to construct phylogenetic trees, best-fitting models among 24 different nucleotide substitution models were estimated in Mega 6 [20] using an automatic neighbor-joining tree. Models with the lowest Bayesian Information Criterion scores were assumed to describe the substitution patterns best. These models were used to construct phylogenetic trees of the full length *mcyA*, *mcyA* adenylation domain (*mcyA*-A1) and 16S rRNA gene nucleotide sequences of investigated cyanobacteria using Mega 6 [20]. Phylogenetic tree of the 16S rRNA genes of selected cyanobacteria (10 nucleotide sequences and a total of 1350 nucleotide positions) was constructed with the maximum likelihood (ML) method using the Hasegawa, Kishino, and Yano (HKY) model [22]. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.3331)). Unrooted phylogenetic tree for the full length *mcyA* gene (10 nucleotide sequences and 7077 nucleotide positions) was constructed with the ML method using the general time reversible (GTR) model [23]. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.8351)). Phylogenetic tree for the *mcyA*-A1 domain nucleotide sequences (11 sequences and 1152 nucleotide positions) was constructed with the ML method based on the GTR model [23]. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.1089)). For all phylogenetic trees, 1000 bootstrap replicates were used to test the stability of monophyletic groups. *GrsA* [24] and 16S rRNA gene nucleotide sequences of *Bacillus brevis* (Genbank: X15577.1) were chosen as the outgroups for *mcyA*-A1 and 16S rRNA gene-based phylogenetic trees, respectively.

3. RESULTS and DISCUSSION

Binding pocket analyses for McyA-A1 sequences were performed to predict which AA was incorporated at position 7 of the MC molecule. The predicted binding pocket residues and the predicted AAs potentially activated by McyA-A1 are summarized in Table 2. Of the ten cyanobacteria species, eight were predicted to activate L-Ser and two *Planktothrix rubescens* strains were predicted to activate L-Thr. Binding pocket signatures for all *Microcystis* strains

Table 2. Microcystin (MC) variants produced, predicted binding pocket residues in McyA-A1, and predicted activated amino acids (AAs) by McyA-A1 in cyanobacteria strains used in this study.

Cyanobacteria	MCs produced	Predicted binding pocket residues	Predicted AAs	Ref.
Anabaena sp. 90[1]	[D-Asp³, MeSer⁷]MC-LR; [MeSer⁷]MC-LR; MC-RR [Dha7]MC-LR*; [Dha7]MC-RR; [D-Asp³]MC-HillR; [D-Asp³] MC-LR; [D-Asp³]MC-RR; [DMAdda⁵] MC-LR; MC-LR; MC-HillR; MC-LR	DVWHLSLID	L-Ser	[9,12]
Planktothrix agardhii NIVA-CYA 126/8[2]	[D-Asp³]MC-RR; [D-Asp³]MC-LR	DVWHISLID	L-Ser	[7,12]
Planktothrix rubescens NIVA-CYA 98[3]	[Dha⁷]MC-LR; [Dha⁷]MC-RR	DFWNIGMVH	L-Thr	[25]
Planktothrix rubescens sp. PCC7821[4]	[Asp³, Dhb⁷]MC-RR; [Asp³]MC-LR*	DFWNIGMVH	L- Thr	[7,26]
Microcystis aeruginosa PCC7806[5]	[D-Asp³, MeSer⁷]MC-LR; [MeSer⁷]MC-LR; [Dha⁷]MC-LR; [D-Asp³, Dha⁷]MC-LR; [D-Asp³]MC-LR; MC-LR	DVWHFSLID	L-Ser	[12]
Microcystis aeruginosa NIES-843[6]	MC-LR; MC-RR; MC-YR	DVWHFSLID	L-Ser	[3]
Microcystis aeruginosa K-139[7]	[Dha⁷]MC-LR; [Asp³, Dha⁷]MC-LR	DVWHFSLID	L-Ser	[12]
Microcystis viridis NIES-102[8]	[Ser¹, D-Asp³,Dha⁷]MC-LR; [X,MeSer⁷] MC-(Hil)R; [Dha7]MC-LR; MC-RR; [D-Asp³]MC-YR; MC-FR [D-Asp³]MC-RR; MC-WR; [D-Asp³]MC-LR; MC-HillR;	DVWHFSLID	L-Ser	[12]
Nostoc sp.152[9]	[D-Asp³, ADMAdda⁵, Dha⁷] MC-(Hil)R; [DMAdda⁵] MC-LR; [Mdhb⁷] MC-LR; MC-YR; [ADMAdda⁵, Dha⁷]MC-LR; [D-Asp³,ADMAdda⁵, Dha⁷]MC-LR; [ADMAdda⁵]MC-LR [ADMAdda⁵, MeSer⁷]MC-LR;	DVWHISLID	L-Ser	[25,27]

	[D-Ser[1], ADMAdda[5]] MC-LR;			
	[ADMAdda[5]]MC-HillR;			
	[D-Asp[3],DMAdda[5]] MC-LR;			
	[D-Asp[3],DMAdda[5]] MC-VR;			
	[D-Asp[3],ADMAdda[5]] MC-LHar;			
	[D-Asp[3], ADMAdda[5]] MC-HillR			
Fischerella sp. CENA 161[10]	MC-LR	DVWHISLID	L-Ser	[6]

Genbank numbers: [1]AJ536156; [2]AJ441056.1; [3]AM990462.1; [4]CZCZ01000002; [5]AF183408; [6]AP009552; [7] AB019578; [8] AP019314.1; [9] KC699835.1; [10] KX891213. Abbreviations: Dhb, dehydrobutyrine; Dha, dehydroalanine; ADMAdda, *O*-acetyl-*O*-demethylAdda; DMAdda, *O*-demethylAdda; Hil, homoisoleucine; MeSer, *N*-Methylserine; D-Asp, desmethyl aspartic acid.
* trace amount

were DVWHFSLID. While binding pocket signatures for the nostocalean genera *Nostoc* sp. 152, *Fischerella* sp. CENA161 and the oscillatorian *Planktothrix agardhii* NIVACYA 126/8 were DVWHISLID; it was DVWHLSLID for *Anabaena* sp. 90, both with one AA change from the *Microcystis* binding pocket signature. Binding pocket signature for *P. rubescens* strains was DFWNIGMVH.

Among the potentially L-Thr activating *P. rubescens* strains, NIVACYA 98 was reported to produce Dha[7] containing MCs [25] and PCC 7821 was reported to produce both Dhb[7] and Mdha[7] containing MCs [11,25] (Table 2). Three different residues, namely MeSer[7], Dha[7] and Mdha[7] were reported for *Anabaena* sp. 90, *M. viridis* NIES 102 and *M. aeruginosa* PCC7806 [12,27]. While production of Mdha[7] were reported for *Planktothrix agardhii* NIVA CYA 126/8 [12], *Microcystis aeruginosa* NIES 843 and *Fischerella* sp. CENA161 [3,6], Dha[7] was the only AA residue reported for *M. aeruginosa* K-139 [12]. Among the investigated species in this study, Mdhb[7] formation was only reported for *Nostoc* sp. 152 [27] that also produced Dha[7], MeSer[7] and Mdha[7] [12, 25] containing MCs (Table 2).

In most MC producing cyanobacteria, McyA-A1 domain usually activates L-Ser. In the case of L-Thr activation by the McyA-A1 domain, Dhb[7] formation may occur in species lacking NMT domain, as expected for the *P. rubescens* strains in this study (Table 1). However, in previous LC/MS analyses of *P. rubescens* NIVACYA 98, only Dha[7] formation was reported [7,25], which indicated L-Ser activation instead of L-Thr and conflicted with the binding pocket analysis results in this study (Table 2). Similarly, Kurmayer *et al*. [11] reported that *P. rubescens* PCC7821 produced MCs containing both Mdha[7] (5%) and Dhb[7] (95%), which indicated activation of both L-Ser and L-Thr. These results contradicted the proposition that polar AA activating A domains such as McyA-A1 had higher selectivity than hydrophobic AA activating A domains [11,28].

Another similar situation was observed for the McyA-A1 domain in *Nostoc* sp. 152, which was predicted to activate L-Ser. MCs containing Mdha[7], MeSer[7], Dha[7] [25,28] and Mdhb[7] [27] were reported for *Nostoc* sp. 152 (Table 2). Mdhb[7] formation (probably a minor variant [29]) indicates L-Thr activation and shows an active NMT, since it is a methylated form of Dhb. Apparently, L-Thr and L-Ser are simultaneously activated by McyA-A1 for *Planktothrix rubescens* PCC7821 and *Nostoc* sp. 152. This multispecificity suggests that there is a flexibility for McyA-A1 domain to activate various AAs during MC production.

The afore-mentioned A domain flexibility has been reported for various secondary metabolites. For example, the A domain involved in the biosynthesis of maremycin from *Streptomyces* sp. was predicted to activate Cys based on the specificity-conferring code, however *in vitro* and *in vivo* studies showed the activation of Cys, Me-Cys and Ser. It should be noted that in vitro catalytic efficiency of A domain towards Cys and Me-Cys were

comparable and 4 times higher than Ser, in agreement with the observation of the minor product maremycin D incorporating Ser *in vivo* [30]. In a comparable situation, L-Phenylalanine activating A domain for the antibiotic Tyrocidine A was shown to be flexible for various AAs *in vitro* [31]. The specificity of this A domain towards L-Phe in vivo suggested that gatekeeping function of the condensation (C) domain might be in place [31]. In fact, Meyer *et al*. [3] demonstrated flexibility of A domains and the gatekeeping and specificity-regulatory role of the C domains of *mcyB* and *mcyC* genes of the microcystin synthetases in *Microcystis* spp. Their *in vitro* experiments showed that the predicted AAs were used by A domains when the C domain was present. On the other hand, minor amounts of unanticipated AAs were incorporated in the MC molecules due probably to the leaky control of the C domain [3], which probably would explain the minor incorporation of un-predicted AAs by *mcyA*-A1 (e.g. Mdha^7 in *P. rubescens* PCC7821). However, it doesn't explain the sole production of Dha^7 containing MCs in *P. rubescens* NIVACYA 98, in which L-Thr activation was predicted (Table 2).

Anabaena sp. 90, *Microcystis* spp., *Planktothrix agardhii* NIVACYA 126/8 and *Fischerella* sp. CENA 161 investigated in this study were predicted to activate L-Ser (Table 2). Previous LC/MS analyses of these species reported MCs containing Mdha^7, MeSer^7 and Dha^7, supporting the activation of L-Ser by McyA-A1 [19]. MeSer^7 was suggested to form through an incomplete dehydration reaction involving the condensation domain of *mcyA* [18]. On the other hand, formation of Dha^7 might have resulted due to transient inactivity of NMT domain or SAM limitation during MC biosynthesis [12].
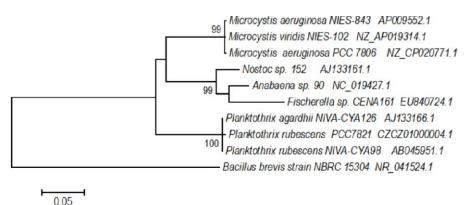

Figure 2. Phylogenetic tree of the 16S rRNA genes of selected cyanobacteria. Bootstrap values above 50 percent from 1000 replicates are at the nodes. Branch lengths represent the number of substitutions per site. All positions with less than 95% site coverage were eliminated. *B. brevis* 16S rRNA gene was used as the outgroup.
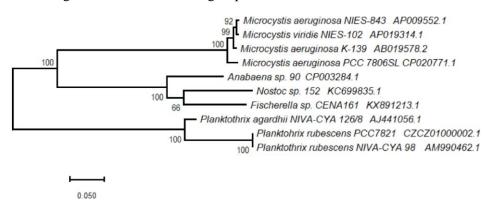

Figure 3. Unrooted phylogenetic tree for the full length *mcyA* gene. Bootstrap values above 50 percent from 1000 replicates are at the nodes. All positions with less than 95% site coverage were eliminated. Nucleotide accession numbers are given at the ends of sequence names. Branch lengths represent the number of substitutions per site.
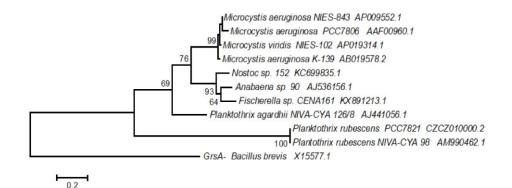
Figure 4. Phylogenetic tree for the *mcyA*-A1 domain nucleotide sequences. Bootstrap values above 50 percent from 1000 replicates are at the nodes. All positions with less than 95% site coverage were eliminated. *GrsA* was used as the outgroup. Nucleotide accession numbers are given at the ends of sequence names. Branch lengths represent the number of substitutions per site.

Phylogenetic analysis of the 16S rRNA gene of the investigated cyanobacteria species clearly separated strains belonging to Nostocales, Oscillatoriales and Chroococcales with high bootstrap support (Figure 2). A similar tree, with high bootstrap support, was obtained when the full length *mcyA* nucleotide sequences were analyzed (Figure 3). However, *P. agardhii* NIVACYA 126/8 did not cluster with *P. rubescens* strains when *mcyA*-A1 domain was analyzed (Figure 4). Clustering of other strains were in agreement with the clusters obtained in 16S rRNA and *mcyA* gene phylogenetic analyses.

In general, phylogenies of full length *mcyA* gene and *mcyA*-A1 domain sequences followed that of 16S rRNA gene of the analyzed strains. The exception was clustering of *P. agardhii* NIVACYA126/8 *mcyA*-A1 domain sequences with Nostocales and Chrococcales sequences, albeit with low bootstrap support. This suggests that either *P. agardhii* or *P. rubescens* acquired its A1 domain via horizontal gene transfer after the separation of both species [5,11]. Phylogenies of the *mcyA* condensation domains of the investigated strains had similar topology to full length *mcyA* and 16S rRNA gene trees (data not shown), further strengthening the idea that some *Planktothrix* strains might have acquired their A domains via horizontal transfer.

4. CONCLUSION

The phylogenetic tree and clustering of A1 domains are in agreement with the prediction of activated AAs by A1 domains. L-Thr activating *P. rubescens* A1 domains are clearly separated from the L-Ser activating domains. These results suggest that phylogenetic analyses of A domains might potentially give information on the activated AAs by these domains. On the other hand, MC congener analysis of these strains show that binding pocket or phylogenetic analysis alone may not explain the AA flexibility of A domains in some strains. Further work involving heterologous expression of *mcyA*-A1 and -C domains together or separately, followed by protein purification and *in vitro* substrate specificity assays might help to explain variation of position 7 AAs in MCs.

Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJSM belongs to the author(s).

Orcid

Gozde Yaman https://orcid.org/0000-0001-6044-3194
Mete Yilmaz https://orcid.org/0000-0002-0982-727X

5. REFERENCES

[1]. Walsh, C.T., O'Brien, R.V., Khosla, C. (2013). Nonproteinogenic amino acid building blocks for nonribosomal peptide and hybrid polyketide scaffolds. *Angew Chem Int Ed Engl.*, *52*(28), 7098–7124.
http://dx.doi.org/10.1002/anie.201208344

[2]. Dittmann, E., Gugger, M., Sivonen, K., Fewer, D.P. (2015). Natural product biosynthetic diversity and comparative genomics of the cyanobacteria. *Trends Microbiol.*, *23*(10), 642–652.
http://dx.doi.org/10.1016/j.tim.2015.07.008

[3]. Meyer, S., Kehr, J.C., Mainz, A., Dehm, D., Petras, D., Süssmuth, R.D., Dittmann, E. (2016). Biochemical dissection of the natural diversification of microcystin provides lessons for synthetic biology of NRPS. *Cell Chem Biol.*, *23*(4), 462–471.
http://dx.doi.org/10.1016/j.chembiol.2016.03.011

[4]. Tillett, D., Dittmann, E., Erhard, M., von Döhren, H., Börner, T., Neilan, B.A. (2000). Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: an integrated peptide–polyketide synthetase system. *Chem Biol.*, *7*(10), 753–764.
http://dx.doi.org/10.1016/s1074-5521(00)00021-1

[5]. Dittmann, E., Fewer, D.P., Neilan, B.A. (2013). Cyanobacterial toxins: biosynthetic routes and evolutionary roots- Review Article. *FEMS Microbiol Rev.*, *37*, 23-43.
http://dx.doi.org/10.1111/j.1574-6976.2012.12000.x

[6]. Heck, K., Alvarenga, D.O., Shishido, T.K., Varani, A.M., Dörr, F.A., Pinto, E., Rouhiainen, L., Jokela, J., Sivonen, K., Fiore, M. F. (2018). Biosynthesis of microcystin hepatotoxins in the cyanobacterial genus *Fischerella*. *Toxicon.*, *141*, 43–50.
http://dx.doi.org/10.1016/j.toxicon.2017.10.021

[7]. Shishido, T.K., Jokela, J., Humisto, A., Suurnäkki, S., Wahlsten, M., Alvarenga, D.O., Sivonen, K., Fewer, D.P. (2019). The biosynthesis of rare Homo-amino acid containing variants of microcystin by a benthic cyanobacterium. *Mar Drugs*, *17*(5), 271.
http://dx.doi.org/10.3390/md17050271

[8]. Christiansen, G., Fastner, J., Erhard, M., Borner, T., Dittmann, E. (2003). Microcystin biosynthesis in *Planktothrix*: Genes, evolution, and manipulation. *J. Bacteriol.*, *185*(2), 564–572.
http://dx.doi.org/10.1128/jb.185.2.564-572.2003

[9]. Rouhiainen, L., Vakkilainen, T., Siemer, B.L., Buikema, W., Haselkorn, R., Sivonen, K. (2004). Genes coding for hepatotoxic heptapeptides (microcystins) in the cyanobacterium *Anabaena* strain 90. *Appl Environ Microbiol.*, *70*(2), 686–692.
http://dx.doi.org/10.1128/aem.70.2.686-692.2004

[10]. Mootz, H.D, Schwarzer, D., Marahiel, M.A. (2002). Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *Chem BioChem.*, *3*, 490–504.
http://dx.doi.org/10.1002/1439-7633(20020603)3:6<490::AID-CBIC490>3.0.CO;2-N

[11]. Kurmayer, R., Christiansen, G., Gumpenberger, M., Fastner, J. (2005). Genetic identification of microcystin ecotypes in toxic cyanobacteria of the genus *Planktothrix*. *Microbiol.*, *151*(5), 1525–1533.
http://dx.doi.org/10.1099/mic.0.27779-0

[12]. Fewer, D.P., Tooming-Klunderud, A., Jokela, J., Wahlsten, M., Rouhiainen, L., Kristensen, T., Rohrlack, T., Jakobsen, K., Sivonen, K. (2008). Natural occurrence of microcystin synthetase deletion mutants capable of producing microcystins in strains of the genus *Anabaena* (Cyanobacteria). *Microbiol.*, *154*(4), 1007–1014.
http://dx.doi.org/10.1099/mic.0.2007/016097-0

[13]. Stachelhaus, T., Mootz, H.D., Marahiel, M.A. (1999). The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem Biol.*, *6*(8), 493–505. http://dx.doi.org/10.1016/s1074-5521(99)80082-9

[14]. Ansari, M.Z., Yadav, G., Gokhale, R.S., Mohanty, D. (2004). NRPS-PKS: a knowledge-based resource for analysis of NRPS/PKS megasynthases. *Nucleic Acids Res.*, *1*(32), 405-413.
http://dx.doi.org/10.1093/nar/gkh359

[15]. Lautru, S., Challis, L.G. (2004). Substrate recognition by nonribosomal peptide synthetase multi-enzymes. *Microbiol.*, *150*(6), 1629–1636.
http://dx.doi.org/10.1099/mic.0.26837-0

[16]. Kudo, F., Miyanaga, A., Eguchi, T. (2019). Structural basis of the nonribosomal codes for nonproteinogenic amino acid selective adenylation enzymes in the biosynthesis of natural products. *J. Ind Microbiol Bio.*, *46*, 515-536.
http://dx.doi.org/10.1007/s10295-018-2084-7

[17]. Bouaïcha, N., Miles, C.O., Beach, D.G., Labidi, Z., Djabri, A., Benayache, N.Y., Quang, T.N. (2019). Structural Diversity, Characterization and Toxicology of Microcystins. *Toxins*, *11*(12), 714.
http://dx.doi.org/10.3390/toxins11120714

[18]. Yilmaz, M., Foss, A.J., Miles, C.O., Özen, M., Demir, N., Balcı, M., Beach, D.G. (2019). Comprehensive multi-technique approach reveals the high diversity of microcystins in field collections and an associated isolate of *Microcystis aeruginosa* from a Turkish lake. *Toxicon*, *167*, 87-100.
http://dx.doi.org/10.1016/j.toxicon.2019.06.006

[19]. Fewer, D.P., Rouhiainen, L., Jokela, J., Wahlsten, M., Laakso, K., Wang, H., Sivonen, K. (2007). Recurrent adenylation domain replacement in the microcystin synthetase gene cluster. *BMC Evol Biol.*, *7*(1), 183.
http://dx.doi.org/10.1186/1471-2148-7-183

[20]. Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.*, *30*(12), 2725-2729.
http://dx.doi.org/10.1093/molbev/mst197

[21]. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, *32*(5), 1792–1797.
http://dx.doi.org/10.1093/nar/gkh340

[22]. Hasegawa, M., Kishino, H., Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.*, *22*(2), 160-174.
http://dx.doi.org/10.1007/bf02101694

[23]. Nei, M., Kumar, S. (2000). Molecular Evolution and Phylogenetics. Oxford University Press: New York, USA, 332 pp, ISBN-10: 0195135857.

[24]. Mikalsen, B., Boison, G., Skulberg, O.M., Fastner, J., Davies, W., Gabrielsen, T.M., Rudi, K., Jakobsen, K.S. (2003). Natural variation in the microcystin synthetase operon *mcy*ABC and impact on microcystin production in microcystis strains. *J. Bacteriol.*, *185*(9), 2774-2785.
http://dx.doi.org/10.1128/jb.185.9.2774-2785.2003

[25]. Gehringer, M.M., Adler, L., Roberts, A.A., Moffitt, M.C., Mihali, T.K., Mills, T.J. T., Fieker, C., Neilan, B.A. (2012). Nodularin, a cyanobacterial toxin, is synthesized in planta by symbiotic *Nostoc* sp. *ISME J.*, *6*(10), 1834–1847.
http://dx.doi.org/10.1038/ismej.2012.25

[26]. Tooming-Klunderud, A., Fewer, D.P., Rohrlack, T., Jokela, J., Rouhiainen, L., Sivonen, K., Kristensen, T., Jakobsen, K.S. (2008). Evidence for positive selection acting on microcystin synthetase adenylation domains in three cyanobacterial genera. *BMC Evol.*

Biol., *8*(1), 256.
http://dx.doi.org/10.1186/1471-2148-8-256

[27]. Puddick, J., Prinsep, M., Wood, S., Kaufononga, S., Cary, S., Hamilton, D. (2014). High levels of structural diversity observed in microcystins from *Microcystis* CAWBG11 and characterization of six new microcystin congeners. *Mar. Drugs*, *12*(11), 5372-5395.
http://dx.doi.org/10.3390/md12115372

[28]. Challis, G.L., Ravel, J., Townsend, C.A. (2000). Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem Biol.*, *7*(3), 211–224.
https://doi.org/10.1016/S1074-5521(00)00091-0

[29]. Sivonen, K., Carmichael, W.W., Namikoshi, M., Rinehart, K.L., Dahlem, A.M., Niemela, S.I. (1990). Isolation and characterization of hepatotoxic microcystin homologs from the filamentous freshwater cyanobacterium *Nostoc* sp. strain 152. *Appl. Environ. Microbiol.*, *56*(9), 2650-2657.
http://dx.doi.org/10.1128/AEM.56.9.2650-2657.1990

[30]. Huang, T., Duan, Y., Zou, Y., Deng, Z., Lin, S.b(2018). NRPS protein MarQ catalyzes flexible adenylation and specific S-methylation. *ACS Chem. Biol.*, *13*(9), 2387-2391.
http://dx.doi.org/10.1021/acschembio.8b00364

[31]. Schaffer, M.L., Otten, L.G. (2009). Substrate flexibility of the adenylation reaction in the Tyrocidine non-ribosomal peptide synthetase. *J. Mol. Catal B-Enzym.*, *59*(1-3), 140-144.
http://dx.doi.org/10.1016/j.molcatb.2009.02.004