Microbiology (2005), **151**, 1525–1533 DOI 10.1099/mic.0.27779-0

Genetic identification of microcystin ecotypes in toxic cyanobacteria of the genus *Planktothrix*

Rainer Kurmayer,[1] Guntram Christiansen,[2] Marlies Gumpenberger[1] and Jutta Fastner[3]

Correspondence
Rainer Kurmayer
rainer.kurmayer@oeaw.ac.at

[1]Austrian Academy of Sciences, Institute for Limnology, Mondseestraße 9, A-5310 Mondsee, Austria

[2]University of Hawaii at Manoa, Department of Chemistry, 2545 McCarthy Hall, Honolulu, HI 96822, USA

[3]Federal Environmental Agency, Corrensplatz 1, D-14195 Berlin, Germany

Microcystins (MCs) are toxic heptapeptides which are synthesized by the filamentous cyanobacterium *Planktothrix* and other genera via non-ribosomal peptide synthesis. MCs share the common structure cyclo(-D-Ala[1]-L-X[2]-D-erythro-β-iso-aspartic acid[3]-L-Z[4]-Adda[5]-D-Glu[6]-*N*-methyl-dehydroalanine[7]) [Adda; (2*S*, 3*S*, 8*S*, 9*S*)-3-amino-9-methoxy-2,6,8-trimethyl-10-phenyldeca-4,6-dienoic acid], in which numerous MC variants have been reported. In general, the variation in structure is due to different amino acid residues in positions 7, 2 and 4 within the MC molecule, which are thought to be activated by the adenylation domains *mcyA*Ad1, *mcyB*Ad1 and *mcyC*Ad, respectively. It was the aim of the study (i) to identify MC ecotypes that differed in the production of specific MC variants and (ii) to correlate the genetic variation within adenylation domains with the observed MC variants among 17 *Planktothrix* strains. Comparison of the sequences of *mcyA*Ad1 revealed two distinctive Ad-genotypes differing in base pair composition and the insertion of an *N*-methyl transferase (NMT) domain. The *mcyA*Ad1 genotype with NMT (2854 bp) correlated with *N*-methyl-dehydroalanine and the *mcyA*Ad1 genotype without NMT (1692 bp) correlated with dehydrobutyrine in position 7. Within *mcyB*Ad1, a lower genetic variation (0–4 %) and an exclusive correlation between one Ad-genotype and homotyrosine as well as another Ad-genotype and arginine in position 2 was found. The sequences of *mcyC*Ad were found to be highly similar (0–1 % dissimilarity) and all strains contained arginine in position 4. The results on adenylation domain polymorphism do provide insights into the evolutionary origin of adenylation domains in *Planktothrix* and may be combined with ecological research in order to provide clues about the abundance of genetically defined MC ecotypes in nature.

Received 23 November 2004
Revised 14 February 2005
Accepted 14 February 2005

INTRODUCTION

Planktothrix is considered to be an important genus of harmful cyanobacteria as its members are regularly found to produce the hepatotoxin microcystin (MC). During a field survey, all samples dominated by either the red-pigmented *Planktothrix rubescens* or the green-pigmented *Planktothrix agardhii* were shown to contain MCs (Fastner *et al.*, 1999). Both species are known to have a specific ecological niche in lakes of the temperate zone of the Northern hemisphere and are highly efficient in monopolizing resources and frequently dominate the phytoplankton community (Mur *et al.*, 1999; Scheffer *et al.*, 1997). Generally, red-pigmented, phycoerythrin-rich genotypes assigned to *P. rubescens* occur in deep, stratified and oligo- to mesotrophic waters, where they can build up metalimnetic layers. Green-pigmented, phycocyanin-rich genotypes frequently assigned to *P. agardhii* have a broader distribution and inhabit shallower, polymictic and mesotrophic to hypertrophic water bodies (Oliver & Ganf, 2000).

MCs are cyclic heptapeptides and share the common structure cyclo(-D-Ala[1]-L-X[2]-D-MeAsp[3]-L-Z[4]-Adda[5]-D-Glu[6]-Mdha[7]), where X and Z are variable L-amino acids (e.g. LR refers to leucine and arginine in the variable positions), D-MeAsp is D-erythro-β-iso-aspartic acid, Adda is (2*S*, 3*S*,

Abbreviations: Adda, (2*S*, 3*S*, 8*S*, 9*S*)-3-amino-9-methoxy-2,6,8-trimethyl-10-phenyldeca-4,6-dienoic acid; Dhb, dehydrobutyrine (2-amino-2-butenoic acid); MC, microcystin; Mdha, *N*-methyl-dehydroalanine; NMT, *N*-methyl transferase.

The GenBank/EMBL/DDBJ accession numbers for the sequences reported in this paper are AJ749248–AJ749302 and AJ863131–AJ863134, as indicated in Table 1.

8*S*, 9*S*)-3-amino-9-methoxy-2,6,8-trimethyl-10-phenyldeca-4,6-dienoic acid and Mdha is *N*-methyl-dehydroalanine (Carmichael *et al.*, 1988). MCs are synthesized, like other non-ribosomal peptides produced by bacteria and fungi, by the thiotemplate mechanism (Marahiel *et al.*, 1997). The large enzyme complex encoded by the *mcy* gene cluster is composed of peptide synthetases, polyketide synthases and tailoring enzymes (Christiansen *et al.*, 2003; Rouhiainen *et al.*, 2004; Tillett *et al.*, 2000). It has a modular structure, each module containing specific functional domains for activation, aminoacyl adenylation (Ad; adenylation domains) and thioesterification (thiolation domains) of the amino acid substrate and for the elongation (condensation domains) of the growing peptide (Tillett *et al.*, 2000).

The structural organization of MC biosynthesis has been elucidated and it has been postulated that McyA, McyB and McyC are responsible for the collinear activation and incorporation of Mdha[7], D-Ala[1], L-X[2], D-MeAsp[3] and L-Z[4] during biosynthesis (Tillett *et al.*, 2000). The first adenylation domain of McyA (*mcyA*Ad1) is expected to activate amino acids occurring in the variable position 7, where three different residues [dehydroalanine, dehydrobutyrine (Dhb) and serine] have been reported from *Planktothrix* strains (Luukkainen *et al.*, 1993; Sano & Kaya, 1995). *mcyB*Ad1 is responsible for the activation of residues in position 2, where three different amino acids [leucine, arginine and homotyrosine (Hty); Sivonen & Jones, 1999] have been described in *Planktothrix* strains. The adenylation domain of McyC (*mcyC*Ad) is correlated with the activation of amino acids in position 4, where only one amino acid (arginine) has been reported in *Planktothrix* strains (Sivonen & Jones, 1999). These reports are in agreement with field observations documenting that the most abundant variants are [Asp[3]]-variants of MC-LR, MC-RR and MC-HtyR (Henriksen & Moestrup, 1997; Fastner *et al.*, 1999).

The extent to which the diversity of MC variants is genetically determined is not yet fully understood. Based on the gramicidin synthetase GrsA crystal structure from *Brevibacillus brevis*, the region forming the amino-acid-binding pocket of adenylation domains has been defined within the core motifs A3 to A6 and the role of critical side chains during substrate recognition in the adenylation domains has been demonstrated (Conti *et al.*, 1997). Using *in silico* analyses, eight specific critical amino acids (signature sequences) have been correlated with amino acid substrates and the so-called specificity-conferring code of adenylation domains could be defined (Stachelhaus *et al.*, 1999; Challis *et al.*, 2000). Point mutational investigations of a few critical amino acids of the adenylation domain of the peptide synthetase (GrsA) demonstrated a change of substrate specificity accompanied by losses in activity (Stachelhaus *et al.*, 1999). Another approach included the investigation of variations found in the *mcy* gene cluster and correlating this to the structural MC variants produced by natural strains (Kurmayer *et al.*, 2002; Mikalsen *et al.*, 2003). Those studies revealed considerable genetic variation within the *mcyB*Ad1 gene of the cyanobacterium *Microcystis* sp. that have been linked to recombination events (Mikalsen *et al.*, 2003), which have recently also been documented for *mcyA* (Tanabe *et al.*, 2004). Some genetic variants were suggested to correlate with the production of MC variants, i.e. the *mcyB* (C) genetic variant correlated with the production of MC-RR and its derivatives (Mikalsen *et al.*, 2003). However, analyses of more strains are needed in order to assess the contribution of genetic recombination events to the production of specific MC variants. It was the aim of this study to investigate whether specific MC variants are correlated with different adenylation domain (Ad) genotypes in cyanobacteria of the genus *Planktothrix*. This was done by aligning the translated amino acid sequences of *mcyA*Ad1, *mcyB*Ad1 and *mcyC*Ad isolated from 17 *Planktothrix* sp. strains from different lakes and correlating these data with the MC variants synthesized by the strains. If a relationship between Ad-genotypes within the *mcyABC* cluster and the occurrence of MC variants can be found, this knowledge could be used to identify specific MC ecotypes in nature. According to the EMBL nucleotide sequence database, an ecotype is defined as a distinct population of organisms of a widespread species that has adapted genetically to its own local habitat (Stoesser *et al.*, 2003). This knowledge is important to understand the wax and wane of specific *mcy* genotypes and the evolution of MC synthesis in our water bodies.

METHODS

Source and cultivation of cyanobacteria. Strains were isolated during a study by Kurmayer *et al.* (2004); in this study, 17 strains were selected to represent the two major groups of MC variants as described by Kurmayer *et al.* (2004): group 1, strains with demethylated variants of MC-RR; and group 2, strains with [Asp[3]]-MC-HtyR as the major variant. In addition, four strains (21/1, CCAP1459/14, CCAP1459/17, CCAP1459/31) assigned to group 2 by Kurmayer *et al.* (2004) were included for the sequencing of *mcyB*Ad1 only (Table 1). Strains with novel MC variants (group 3) isolated selectively from Lake Schwarzensee (Upper Austria) will be described elsewhere (R. Kurmayer, K. Ishida, J. Fastner and T. Hemscheidt, unpublished results). According to PCR analysis and sequencing of the internal transcribed spacer region of the phycocanin operon (Kurmayer *et al.*, 2004) and sequence information on the 16S rRNA gene provided by Suda *et al.* (2002), all of the strains of this study were assigned either to *P. agardhii* (green-pigmented) or *P. rubescens* (red-pigmented). All strains were cultivated in BG_{11} (Rippka, 1988) containing 2 mM $NaNO_3$ plus 10 mM $NaHCO_3$ at 15 °C and continuous light (5–10 μmol m^{-2} s^{-1}; Osram type L30W/77 Fluora).

MC analysis. Strains were grown simultaneously and analysed independently on three separate occasions (3 October 2002, 29 January 2004, 13 April 2004). All strains showed comparable growth and 3 weeks after inoculation cells were filtered on pre-weighed glass fibre filters (GF/C; Comesa), dried at 95 °C overnight and then reweighed to quantify the biovolume for extraction. On the day of harvest, the biovolume was between 0·1 and 0·3 mg dry weight l^{-1}. MCs were extracted using 75 % (w/v) aqueous methanol and the extracts were analysed for MC by HPLC with diode array detection (HPLC-DAD) as described by Kurmayer *et al.* (2003). MC variants were quantified at 240 nm by their characteristic absorption spectra (original spectrum and first-order derivative) and retention times

Table 1. *Planktothrix* strains used in this study

Species are abbreviated as Rub (*P. rubescens*; red-pigmented) and Aga (*P. agardhii*; green-pigmented). Country codes are given in ISO format (AT, Austria; DE, Germany; DK, Denmark; FI, Finland; JP, Japan; NO, Norway; UK, United Kingdom). 16S rRNA gene sequence accession numbers were taken from Suda *et al.* (2002) and phycocyanin internal transcribed spacer (PC-ITS) sequence accession numbers from Kurmayer *et al.* (2004). Groups with different MC variants as identified by Kurmayer *et al.* (2004) are identified as MC group 1, strains with [Asp^3]-MC-RR as the major variant, and group 2, strains with [Asp^3]-MC-HtyR as the major variant. ND, No data available.

Species	Strain	MC group	Isolation		Accession numbers				
			Year	Origin	*mcyA*	*mcyB*	*mcyC*	PC-ITS	16S rRNA
Rub	3	1	2001	Mondsee, AT	AJ749248	AJ749276	AJ749285		
Rub	64	1	2001	Wörthersee, AT	AJ749252	AJ749277	AJ749286		
Rub	111	1	2001	Mondsee, AT	AJ749258	AJ749282	AJ749291		
Aga	31/1	1	2001	Wannsee, DE	AJ749249	AJ749267	AJ749294		
Aga	32	1	2001	Wannsee, DE	AJ749250	AJ749268	AJ749295		
Aga	39	1	2001	Wannsee, DE	AJ749251	AJ749269	AJ749296		
Aga	79	1	2001	L. Arresø, DK	AJ749253	AJ749270	AJ749297		
Aga	SAG 6.89	1	1969	Plußsee, Plön, DE	AJ749259	AJ749271	AJ749298	AJ558137	
Aga	CYA126/8	1	1984	L. Langsjön, FI	AJ441056	AJ441056	AJ441056	AJ558135	AB045914
Rub	82	1	2001	Ammersee, DE	AJ749255	AJ749279	AJ749288		
Rub	108	1	2001	Irrsee, AT	AJ749260	AJ749281	AJ749290		
Rub	PCC7821	1	1971	L. Gjersjoen, NO	AJ749261	AJ749283	AJ749292	AJ558154	AB045901
Rub	CCAP1459/30	1	ND	Plöner See, DE	AJ749265	AJ749284	AJ749293	AJ558147	
Aga	CCAP1459/11A	1	1975	L. Windermere, UK	AJ749262	AJ749272	AJ749299	AJ558160	AB045896
Aga	CCAP1459/21	1	1985	Esthwaite Water, UK	AJ749264	AJ749274	AJ749301	AJ558159	AB045900
Aga	CCAP1460/5	2	1983	L. Kasumigaura, JP	AJ749266	AJ749275	AJ749302		AB045954
Aga	CCAP1459/16	2	1979	Blelham Tarn, UK	AJ749263	AJ749273	AJ749300	AJ558158	AB045899
Rub	21-	2	1999	Figur, AT		AJ863131			
Rub	CCAP1459/14	2	1975	Loughrigg Tarn, UK		AJ863132			
Aga	CCAP1459/17	2	1981	Blelham Tarn, UK		AJ863133			
Aga	CCAP1459/31	2	1971	White Lough, UK		AJ863134			

(Fastner *et al.*, 1999). Fastner *et al.* (1999) reported that, using the gradient according to Lawton *et al.* (1994), [Asp^3, Mdha^7]-MC-RR eluted 0·8 min before [Asp^3, Dhb^7]-MC-RR (Fastner *et al.*, 1999). In order to test for the sensitivity of detection of [Asp^3, Mdha^7]-MC-RR in the presence of [Asp^3, Dhb^7]-MC-RR and vice versa, the proportion of one variant was gradually increased relative to the other during pilot experiments. Results showed that both [Asp^3]-MC-RR variants could be linearly detected down to a minimum proportion of <5 %. [MeAsp^3, Mdha^7]-MC-RR, MC-YR and MC-LR were used as external standards (Calbiochem). The concentrations of MC variants were determined as concentration equivalents of [MeAsp^3, Mdha^7]-MC-LR. In addition, dried HPLC fractions of putative MCs were dissolved in 20 µl 50 % aqueous methanol, sonicated for 10 min and left for 20 min. From 1 µl of this sample preparation, positive-ion mass spectra from 500 to 2000 Da were recorded using a MALDI-TOF mass spectrometer (Voyager DE-PRO; PerSeptive BioSystems) as described by Erhard *et al.* (1997). MC variants were identified by post-source decay (PSD) fragment structure analysis (Fastner *et al.*, 1999).

DNA amplification and sequencing. For DNA extraction, 2 ml culture was incubated for 1 h on ice and centrifuged at 13 000 r.p.m. for 10 min and the pellet was lyophilized in a vacuum centrifuge at 30 °C. DNA was extracted using a protocol described by Kurmayer *et al.* (2003). For PCR, DNA extracts were diluted 100-fold and 1·0 µl of the sample was pipetted into reaction tubes and incubated as described below. PCR amplifications were performed in a volume of 20 µl, containing 1× Qiagen PCR buffer, 3 mM $MgCl_2$ (Qiagen), 300 µM each dNTP (MBI Fermentas), 0·5 µM each primer, 0·5 units *Taq* DNA polymerase (Qiagen), 13·1 µl sterile Millipore water and 1·0 µl DNA extract. Primers used for PCR and sequencing are listed in Table 2. For *mcyA*Ad1 (product size 3022 bp), the PCR thermal cycling protocol included an initial denaturation step at 94 °C for 3 min, followed by 35 cycles at 94 °C for 30 s, annealing at 60 °C for 30 s and elongation at 72 °C for 3 min. For *mcyB*Ad1 (product size 1692 bp) the cycling protocol was identical, but annealing was at 52 °C for 30 s and elongation at 72 °C for 2 min. For *mcyC*Ad (product size 1416 bp) the cycling protocol was identical to that for *mcyA*Ad1, but the elongation time was 2 min at 72 °C. PCR products (4 µl of the reaction mixture) were visualized by electrophoresis in 1·0 % agarose in 0·5× TBE with ethidium bromide staining. The amplification products of *mcyABC* were sequenced directly by standard automated fluorescence techniques (Applied Biosystems). These sequence data have been submitted to the DDBJ/EMBL/GenBank databases under the accession numbers AJ749248–AJ749302 and AJ863131–AJ863134 (see Table 1).

Sequence alignment and analysis. Sequences were aligned using multiple sequence alignment (CLUSTAL W 1.8). Similarity values between amino acid sequences and the corresponding adenylation domain of strain CYA126/8 (GenBank accession no. AJ441056) were calculated using the program PROTDIST of the PHYLIP software package [version 3.6(alpha3); Felsenstein, 1993]. These values are the fractions of amino acid positions that are identical between the sequences (without weighting of protein positions).

Table 2. Oligonucleotide primers used for PCR (P) and sequencing (S)

All primers were designed during this study.

Primer	Purpose	Sequence (5′–3′)	Direction	T_m (°C)
***mcyA*Ad1**				
*mcyA*A1Gufwd	P, S	GCAACCAATTAGCCCATAATCTC	F	58·9
*mcyA*A1Gurev	P, S	GGCAGTTAGGGGGTAAGCATC	R	61·8
***mcyA*Ad1 (with NMT)**				
*mcyA*nA43kbfwd	P, S	CCGAAGGTCATAAGGACAGTGTT	F	60·6
*mcyA*vA93kbrev	P, S	GAGGTTGGAGCGGATGGTT	R	58·8
NachA6fwd	P, S	GGTATCTCCCCGATGGCAATA	F	55·2
VorM2rev	P	GCCATTGATCGACGGAGATATC	R	55·0
InternA6fwdseq	S	AATCGTCCTGAATTGACGGC	F	53·9
InternnA8fwdseq	S	CCTATTTCAAGTGGCTCCCCA	F	55·5
InternnA3fwdseq	S	TGCCATCGAGGCGTTGT	F	52·8
InternvA10fwd	S	GCGATTTTTGTGAAGGAGCAA	F	54·5
***mcyA*Ad1 (without NMT)**				
Nr80fwd	P, S	TAATCATTCCCATGTTGTTCGTCT	F	57·6
Nr80rev	P, S	TCTGTATTTTGGGTTGGTAAAGGAT	R	58·1
Nr80intern2fwd	P, S	TAATCAAACTCCCACCGCCT	F	53·6
Nr80intern2R	P, S	TTGCGGTACCACATAACCCAC	R	54·7
***mcyB*Ad1**				
*mcyB*A1totfwd	P, S	CACCTAGTTGAAGAACAAGTTCT	F	51·1
*mcyB*A1totrev	P, S	AGACTTGTTTAATAGCAAAGGC	R	51·5
Sample1,3,4fwd	P, S	TTCTTCTCTAGCTTTTGACG	F	49·4
Sample1-4rev	P, S	TCTGTTTGTATAATTACCACACTT	R	49·0
*mcyB*intern1	S	ACCCAGCAATCTTTAGTTACC	F	51·6
*mcyB*intern2	S	CACAGAAACCTCCGTTGAT	F	52·0
*mcyB*intern3	S	GACAATTCTAGACAGCCATCC	F	52·6
***mcyC*Ad**				
*mcyC*Afwd1	P, S	GCAGGCGAACCAATTAGCC	F	54·7
*mcyC*Arev1	P	GACTTCGGTTTGATTACCTGGG	R	54·2
RK_*mcyC*Aintfw	S	CTTTATGGGCCAACAGAAACAG	F	53·2
RK_*mcyC*Aintrev	S	AGTTCTCCAGCAACGCCAATAG	R	55·1

Phylogenetic trees were constructed by (i) average linkage clustering (UPGMA) from the amino acid distance matrix using the approximation of Kimura (1983) to the Dayhoff PAM matrix using the programs PROTDIST and NEIGHBOR, (ii) the maximum-likelihood method (ML) using the program PROML using the Jones–Taylor–Thornton model of change between amino acids (Jones *et al.*, 1992) and (iii) the maximum-parsimony (MP) method using PROTPARS from the PHYLIP software package. In general, sites were not weighted. The statistical significance of the branches was estimated by bootstrap analysis generating 1000 replicates of the original dataset. Finally, consensus trees following the 50 % majority rule were computed. For all of the genes, phylogenetic trees were congruent and the parsimonious trees and significant bootstrap values for all of the methods are presented.

RESULTS

MC variants

Analysis of the residue variation in position 7 revealed that either Mdha (nine strains) or Dhb (six strains) was part of the MCs (Table 3, Fig. 1). Residue variations in position 2 could be split into two groups: the first group exclusively produced one MC variant containing arginine (five strains). The second, larger group contained 10 strains, producing a mixture of MCs carrying arginine as the major variant ($\geqslant$67 %) and leucine ($<$33 %) and homotyrosine ($<$2 %). Two strains (CCAP1460/5, CCAP1459/16) showed no arginine, but homotyrosine ($\geqslant$69 %) and leucine ($<$30 %) were major variants. There were no residue variations seen in position 4 (arginine).

Genetic variation within the *mcyABC* cluster

Using the same primers for *mcyA*Ad1, a PCR product of either 3 kb (corresponding to the *mcyA*Ad1 sequence of strain CYA126/8; GenBank accession no. AJ441056) or 1·5 kb was obtained. Most of the reduction in PCR product size was due to the lack of a DNA sequence encoding an *N*-methyl transferase (NMT, 416 aa). In all strains showing a PCR product of 3 kb (950 aa), the NMT was inserted between the core motifs A8 (KIRGXRIELGEIE) and A9 (LPXYM) defined by Marahiel *et al.* (1997). Within the remaining *mcyA*Ad1 sequence (534 aa), the two distinctive Ad-genotypes of *mcyA* also differed significantly in

Table 3. Similarity in the amino acid composition of adenylation domains of *mcyA*Ad1, *mcyB*Ad1 and *mcyC*Ad and the production of MC variants measured for 17 strains of *P. agardhii* and *P. rubescens*

The presence (+) or absence (−) of the NMT within *mcyA*Ad1 is indicated. Similarity is given in comparison with strain CYA126/8, sequenced for the *mcy* gene cluster by Christiansen *et al.* (2003). For a complete list of strains and strain origins, see Table 1 in Kurmayer *et al.* (2004). MC variants were identified from their retention times in HPLC: [Asp^3, Mdha^7]-MC-RR, 13·4–13·9 min; [Asp^3, Dhb^7]-MC-RR, 14·3–14·8 min; [Asp^3]-MC-HtyR, 18·3–18·5 min; [Asp^3]-MC-LR, 19·3–19·6 min. The MC content is given in equivalents of [MeAsp^3, Mdha^7]-MC-LR (mean ± SEM).

Species	Strain	NMT	Amino acid sequence similarity (%)			MC variants (mean proportion ± SEM; $n=3$)					MC content [μg (mg dry weight)$^{-1}$]
			mcyA (534 aa)	*mcyB* (513 aa)	*mcyC* (448 aa)	[Asp^3, Mdha^7] -MC-RR	[Asp^3, Dhb^7] -MC-RR	[Asp^3]-MC-HtyR	[Asp^3]-MC-LR	Undetermined	
Rub	3	+	99	99^{d*}	100^a	99·2 ± 0·8				0·8 ± 0·8	1·4 ± 0·5
Rub	64	+	96	99^d	100	100				0	0·3 ± 0·1
Rub	111	+	98	99^d	99	98·4 ± 0·9				1·6 ± 0·9	1·8 ± 0·5
Aga	31/1	+	99	98	99	85 ± 1		1·9 ± 0·4	10·4 ± 0·6	2·6 ± 0·1	3·0 ± 0·6
Aga	32	+	100^c	97	100^a	88·5 ± 0·7			10·7 ± 0·2	0·8 ± 0·8	2·4 ± 0·1
Aga	39	+	100	98	100	91 ± 0·3			9·1 ± 0·3	0	2·6 ± 0·3
Aga	79	+	100	100^e	99	96·2 ± 0·6			3·4 ± 0·3	0·4 ± 0·4	1·6 ± 0·1
Aga	SAG6.89	+	98	98	100^a	79·8 ± 3			18·9 ± 2·3	1·3 ± 0·7	2·5 ± 0·5
Aga	CYA126/8	+	(100)^c	(100)^e	(100)^a	92·2 ± 1·2			7·1 ± 1·3	0·7 ± 0·9	2·1 ± 0·4
Rub	82	−	56^a	99^b	100		93·8 ± 0·7	0·6 ± 0·3	4·8 ± 1	0·8 ± 0·8	4·9 ± 0·6
Rub	108	−	56	99^b	100		96·7 ± 1·1		1·1 ± 0·04	2·2 ± 1·1	2·3 ± 0·4
Rub	PCC7821	−	56^b	98	100		94·9 ± 0·5		5·1 ± 0·5	0	1·0 ± 0·1
Rub	CCAP1459/30	−	56^b	98	100		67·4 ± 0·8		32·6 ± 0·8	0	2·0 ± 0·3
Aga	CCAP459/11A	−	56	96^c	100^a		100			0	2·7 ± 0·2
Aga	CCAP1459/21	−	56	96^c	100^a		100			0	1·9 ± 0·2
Aga	CCAP1460/5	−	56	97^a	100			74·9 ± 2·7	24·7 ± 2·3	0	2·1 ± 0·4
Aga	CCAP1459/16	−	56^a	97^a	100^a			68·8 ± 3·4	29·8 ± 3·2	1·5 ± 0·7	2·0 ± 0·2

*Superscripts indicate identical genotypes.

amino acid sequence (44 % dissimilarity when compared to CYA126/8; Table 3). Phylogenetic analyses of *mcyA*Ad1 sequences revealed a separation of *Planktothrix* strains that was found to be independent of the taxonomic distinction between *P. agardhii* and *P. rubescens*. Two main clades were observed: the *mcyA* (I) clade (lacking NMT) showed the lowest genetic dissimilarity (0–0·2 %), and the *mcyA* (II) clade (with NMT) showed higher genetic variation (0–3·2 %; Fig. 2). The *mcyA*Ad1 genotype with NMT showed 63 % sequence identity with *mcyA*Ad1 of *Microcystis aeruginosa* (BAA83992) and 65 % identity with *mcyA*Ad1 of *Anabaena* sp. 90 (AAO62586). In contrast, the *mcyA*Ad1 genotype without NMT (563 aa) revealed sequence identities (47–66 %) with putative non-ribosomal peptide synthetases from organisms of the cyanobacteria [*Nostoc punctiforme* PCC73102, *Anabaena* ATCC 2941, *Anabaena* sp. 90 (Rouhiainen *et al.*, 2000) and *Nodularia spumigena* (Moffitt & Neilan, 2004)], the bacillales (*Bacillus cereus*), the proteobacteria (*Pseudomonas* spp., *Chromobacterium violaceum* and *Ralstonia solanacearum*) and the actinobacteria (*Streptomyces avermitilis*) (data not shown).

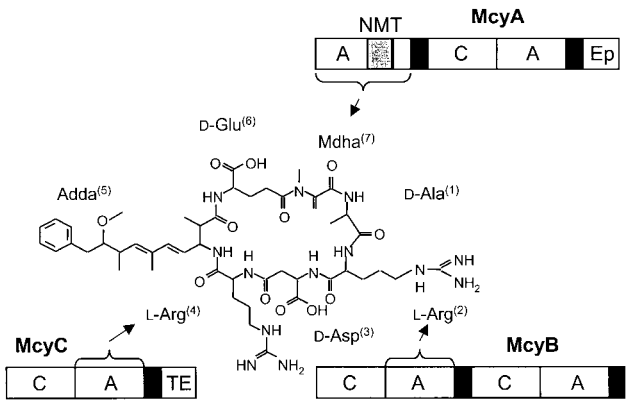

Fig. 1. Structure of [Asp^3, Mdha^7]-MC-RR and the multienzymes McyA, McyB and McyC believed to be responsible for the collinear activation and incorporation of Mdha^7, D-Ala^1, L-X^2, D-MeAsp^3 and L-Z^4 during biosynthesis (Tillett *et al.*, 2000). Each rectangle represents a non-ribosomal peptide synthetase enzymic domain: A, aminoacyl adenylation; C, condensation; NMT, *N*-methyl transferase; Ep, epimerase; TE, thioesterase. Thiolation domains are shown in black. Braces indicate the adenylation domains sequenced from *Planktothrix* strains (see Table 1) and arrows denote the corresponding amino acids activated during biosynthesis.

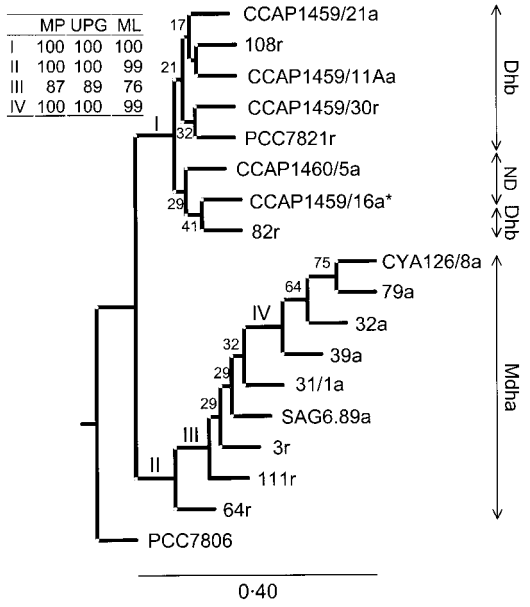

Fig. 2. Phylogenetic tree based on MP analysis calculated from *mcyA*Ad1 sequences (950 aa) from 17 *Planktothrix* strains. Significant bootstrap percentages obtained from 1000 replicates are indicated for *mcyA* clades I, II, III and IV at the top left for MP, UPGMA clustering (UPG) and ML. Bootstrap values for clades with lower support are given for MP only. *mcyA*Ad1 from *Microcystis* PCC7806 (AAF00960) was used as an outgroup. The suffixes r and a after the strain name indicate the species *P. rubescens* (r) and *P. agardhii* (a). The corresponding amino acids in position 7 determined in the MC molecule are given (ND, not determined). Bar, 4 substitutions per 10 amino acids. [Asp^3, Dhb^7]-MC-HtyR has been isolated from strain CCAP1459/16 (indicated by an asterisk) by Sano & Kaya (1998).

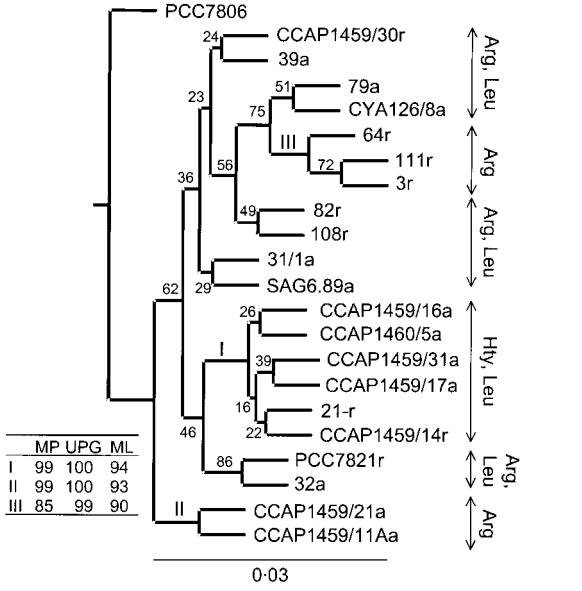

Fig. 3. Phylogenetic tree based on MP analysis calculated from *mcyB*Ad1 sequences (513 aa) from 21 *Planktothrix* strains. Bootstrap percentages obtained from 1000 replicates are indicated for *mcyB* clades I, II, III at the bottom left for MP, UPGMA clustering (UPG) and ML. Bootstrap values for clades with lower support are given for MP only. *mcyB*Ad1 from *Microcystis* PCC7806 (AAF00961) was used as an outgroup. The suffixes r and a after the strain name indicate the species *P. rubescens* (r) and *P. agardhii* (a). The corresponding amino acids in position 2 determined in the MC molecule are given. Bar, 0·3 substitutions per 10 amino acids.

Within *mcyB*Ad1, the genetic variation was lower and no polymorphism in sequence length was observed (0–4 %; Table 3, Fig. 3). Phylogenetic analysis of *mcyB*Ad1 sequences revealed the occurrence of three clades (I, II, III), with the *mcyB* (I) clade consisting of both *P. rubescens* and *P. agardhii*, the *mcyB* (II) clade consisting of *P. agardhii* and the *mcy*B (III) clade consisting of *P. rubescens* only. The *mcyB* (II) clade (strains CCAP1459/11A and CCAP1459/21) showed the most significant deviation in amino acid sequence (4 %). Genetic variation was lowest within *mcyC*Ad (0–1 %; Table 3), and no consistent branching using any of the three phylogenetic methods was detected (data not shown).

In summary, no correlation was found between *mcy* genotype and taxonomic distinction between *P. agardhii* and *P. rubescens*. In addition, no correlation between the *mcy* genotype distribution and the origin of isolation was found. For example, identical *mcyB*Ad1 genotypes were isolated from habitats with the largest geographical distance observed in this study, i.e. Blelham Tarn, UK (strain CCAP1459/16) and Lake Kasumigaura, Japan (strain CCAP1460/5).

Genetic variation, signature sequences and MC variants

The *mcyA*Ad1 genotype with NMT had a signature sequence DVWHISLI that matched exactly (identity 8/8) with the reference sequence of nostopeptolide synthetase (gb|AAF15891.2|) and the prediction of serine as the amino acid substrate. This prediction correlated perfectly with the presence of Mdha in [D-Asp^3, Mdha^7]-MC-RR in nine of the nine strains (Table 3, Fig. 2). The *mcyA*Ad1 genotype without NMT had a signature sequence DFWNIGMV that matched exactly (identity 8/8) with the reference sequence of exochelin synthetase (gb|AAC82550.1|), pyoverdine synthetase D (gb|AAB60198.1|), fengycin synthetase (emb|CAA09819.1|, emb|CAA84361.1|) and coelichelin synthetase (gi|5763943), all predicting threonine as the amino acid substrate. This prediction correlated with Dhb occurring in [D-Asp^3, Dhb^7]-MC-RR (six strains).

Within *mcyB*Ad1, the signature sequences had no clear precedent in the database. The *mcyB* (I) clade (six strains) was derived from one genotype only and showed the

signature sequence DALLFGFV. An exclusive correlation with homotyrosine and leucine, but with no arginine, as major residues in position 2 was recognized (Fig. 3). Notably, all other strains reported to contain homotyrosine and leucine in position 2 as major residues by Kurmayer *et al.* (2004) (21/1, CCAP1459/14, CCAP1459/17, CCAP1459/31) were found exclusively to have one *mcyB*Ad1 genotype. The *mcyB* (II) clade (two strains) showed the signature sequence DAWAFGLV and the *mcyB* (III) clade (three strains) showed the signature sequence DALFFGVV, and both clades correlated exclusively with arginine. The remaining 10 strains were found to be without clear genetic differentiation, showed the signature sequence DALFFGLV and produced a mixture of MCs carrying arginine and leucine as major variants.

Genetic differentiation was lowest within *mcyC*Ad; one signature sequence, DPWGFGLV, without a precedent in the database was found and no variation in position 4 was found (Table 3). In summary, both Ad genotypes of *mcyA*Ad1 were found to be specific for the amino acid composition in position 7. The Ad genotypes of *mcyB*Ad1 were found to be both specific and unspecific, the latter correlating with the activation of two (or three) amino acids during MC biosynthesis.

DISCUSSION

Genetic variation within adenylation domains and correlation with the synthesis of MC variants

In this study, a significant correlation between *mcyA*Ad1 genotypes with NMT and *mcyA*Ad1 genotypes without NMT and the occurrence of either Mdha or Dhb in position 7 of the MC molecule was found. According to Tillett *et al.* (2000), *mcyA*Ad1 is responsible for the incorporation of L-serine into the growing molecule. L-Serine is *N*-methylated by the corresponding NMT domain and transformed by dehydration into Mdha prior to or following the condensation reaction. Analogous to L-serine, L-threonine is transformed by dehydration into Dhb (Rinehart *et al.*, 1994). Whether strains CCAP1460/5 and CCAP1459/16, containing *mcyA*Ad1 without NMT, indeed produced [Asp^3, Dhb^7]-MC-HtyR was not tested explicitly in this study, but [Asp^3, Dhb^7]-MC-HtyR has been isolated from strain CCAP1459/16 by Sano & Kaya (1998).

Within *mcyB*Ad1, a significant correlation between *mcyB* (I) clade and homotyrosine in position 2 was found. Even the strains from Lake Schwarzensee (Kurmayer *et al.*, 2004) were of the same *mcyB*Ad1 genotype and contained homotyrosine as the major residue in position 2 (R. Kurmayer, K. Ishida, J. Fastner and T. Hemscheidt, unpublished results). In addition, two *mcyB* clades (II, III) correlated with arginine exclusively in position 2. *mcyB* (III) clade was found in 12 strains ($<1{\cdot}5$ % dissimilarity) and all strains contained arginine in position 2 (Gumpenberger, 2004). In summary, the genetic differences in the *mcyB*Ad1 genotypes correlated only partly with the observed differences in the structure of the MC variants. Corresponding to this study, the majority of *mcyB*Ad1 sequences from a number of single colonies from *Microcystis* sp. did not correlate with differences in MC amino acid composition (Kurmayer *et al.*, 2002). This result corresponds to the observation that *Microcystis* strain HUB524, with the same *mcyB*Ad1 sequence, produces three different MCs simultaneously containing either leucine, arginine or tyrosine in position 2 (Fastner *et al.*, 1999). Those results indicate the potential of *mcyB*Ad1 to activate a variety of amino acids during MC biosynthesis. The results correspond to the general view that adenylation domains activating hydrophobic amino acids (e.g. *mcyB*Ad1) possess a lower selectivity when compared to adenylation domains activating polar amino acids (e.g. *mcyA*Ad1; Challis *et al.*, 2000).

It has been suggested that the synthesis of specific MC variants in a particular strain depends on the physiological conditions; for instance, Rapala *et al.* (1997) found an increasing proportion of MC-RR at the expense of MC-LR with increasing temperature. According to the present authors' unpublished measurements, at a higher temperature along with high light conditions (20 °C, 40 μmol m^{-2} s^{-1}) as opposed to the culture conditions used in this study (15 °C, 5–10 μmol m^{-2} s^{-1}), unaltered synthesis of either [D-Asp^3, $Mdha^7$] or [D-Asp^3, Dhb^7] variants of MC-RR was observed. In addition, the occurrence of MC-RR only in some of the strains (CCAP1459/11A, CCAP1459/21, 3, 64, 111) was found to be unaltered at higher temperature and high light conditions (20 °C, 40 μmol m^{-2} s^{-1}; R. Kurmayer, unpublished).

Genetic recombination of the *mcyABC* cluster

In this study, the *mcyA*Ad1 genotype without NMT (563 aa) revealed extensive sequence identity (47–66 %) with non-ribosomal peptide synthetase genes from other cyanobacterial and bacterial genera. The genetic variation was lowest among strains of this genotype (0–0·2 %), suggesting a relatively recent recombination event. Recombinations involving adenylation domains of the *mcy* gene cluster in *Microcystis* have been suggested for *mcyB* (Mikalsen *et al.*, 2003) and the NMT region of *mcyA* (Tanabe *et al.*, 2004). Recombinations and deletions involving the condensation domain in *ndaA* have been reported by Moffitt & Neilan (2004). Recombination therefore seems to be a general feature in *mcy* genes, and these findings may be important in understanding how new structural variants of MCs are created.

The same *mcyA*Ad1 primers were used to amplify *mcyA*Ad1 genotypes both with and without NMT in *Planktothrix* spp. Notably, *ndaA* (AAO64403), which shows 61 % identity to the *mcyA*Ad1 genotype without NMT, consists of not only a threonine adenylation domain but also an NMT domain (Moffitt & Neilan, 2004). So far, an *mcyA*Ad1 threonine adenylation domain with NMT has not been

found in *Planktothrix* strains. It is possible that the NMT has been lost after the transfer of the *mcyA*Ad1 threonine adenylation domain into *mcyA* of *Planktothrix*.

Ecological implications

In this study, it could be shown that DNA polymorphisms within specific regions of adenylation domains are associated with the synthesis of specific MC variants. The process by which secondary metabolic pathways evolve is probably a result of modifications and combinations of reactions from existing pathways. Through the process of natural selection, the producer of new structures will only increase in number relative to the producers of the older structures if the production of the new structure is advantageous. Assuming that the observed recombination within *mcyA*Ad1 and the resulting [D-Asp^3, Dhb^7]-MC-RR ecotype originated from a single DNA recombination event, it must be assumed that natural selection favoured the increase of [D-Asp^3, Dhb^7]-MC producers relative to [D-Asp^3, Mdha^7]-MC producers. Notably, the first quantitative results showed that (i) in a few lakes, the *mcyA*Ad1 genotype without NMT dominated, while, in other populations, *mcyA*Ad1 genotypes both with and without NMT were found to co-occur over several years and (ii) the *mcyB*Ad1 genotype producing [Asp^3]-MC-HtyR occurred much less frequently in Lake Irrsee (Upper Austria) when compared with Lake Mondsee (Upper Austria; R. Kurmayer, unpublished). In the field, [D-Asp^3, Dhb^7]-MC-RR has been reported as the dominant variant in *P. rubescens* (Blom *et al.*, 2001; Fastner *et al.*, 1999), while [D-Asp^3, Mdha^7]-MC-RR was found to be most abundant in phytoplankton samples dominated by *P. agardhii* (Fastner *et al.*, 1999). It is speculated that specific environmental conditions may not only influence the absolute abundance of MC-producing genotypes via the dominance of *P. rubescens* over *P. agardhii* (e.g. Kurmayer *et al.*, 2004) but may also influence the proportion of specific MC ecotypes. The quantification of MC ecotypes as well as transplantation in their natural context will deliver important clues on the function of MCs in ecosystems.

ACKNOWLEDGEMENTS

Johanna Schmidt provided valuable assistance in the culturing of strains and microcystin analysis in the laboratory. We would also like to thank Martin Meixner for sequencing. We appreciate the comments of two anonymous reviewers. This study was supported by the Austrian Science Fund (P15709) CYTOGENE (Linking CYanTOxin production to GENEtic diversity) and by the EU project PEPCY (QLK4-CT-2002-02634). The European Community is not responsible for any use that might be made of data appearing herein.

REFERENCES

Blom, J. F., Robinson, J. A. & Jüttner, F. (2001). High grazer toxicity of [D-Asp^3, (*E*)-Dhb^7]microcystin-RR of *Planktothrix rubescens* as compared to different microcystins. *Toxicon* **39**, 1923–1932.

Carmichael, W. W., Beasly, V., Bunner, D. L. & 11 other authors (1988). Naming of cyclic heptapeptide toxins of cyanobacteria (blue-green algae). *Toxicon* **26**, 971–973.

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**, 211–224.

Christiansen, G., Fastner, J., Erhard, M., Börner, T. & Dittmann, E. (2003). Microcystin biosynthesis in *Planktothrix*: genes, evolution, and manipulation. *J Bacteriol* **185**, 564–572.

Conti, E., Stachelhaus, T., Marahiel, M. A. & Brick, P. (1997). Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J* **16**, 4174–4183.

Erhard, M., von Döhren, H. & Jungblut, P. (1997). Rapid typing and elucidation of new secondary metabolites of intact cyanobacteria using MALDI-TOF mass spectrometry. *Nat Biotechnol* **15**, 906–909.

Fastner, J., Erhard, M., Carmichael, W. W., Sun, F., Rinehart, K. L., Rönicke, H. & Chorus, I. (1999). Characterization and diversity of microcystins in natural blooms and strains of the genera *Microcystis* and *Planktothrix* from German freshwaters. *Arch Hydrobiol* **145**, 147–163.

Felsenstein, J. (1993). PHYLIP – phylogeny inference package, version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle, USA.

Gumpenberger, M. (2004). *Phänotypische und genetische Charakterisierung toxischer Cyanobakterien aus der Gattung Planktothrix spp.* MSc thesis, University of Salzburg, Austria (in German).

Henriksen, P. & Moestrup, O. (1997). Seasonal variations in microcystin contents of Danish cyanobacteria. *Nat Toxins* **5**, 99–106.

Jones, D. T., Taylor, W. R. & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* **8**, 275–282.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution.* Cambridge: Cambridge University Press.

Kurmayer, R., Dittmann, E., Fastner, J. & Chorus, I. (2002). Diversity of microcystin genes within a population of the toxic cyanobacterium *Microcystis* spp. in Lake Wannsee (Berlin, Germany). *Microb Ecol* **43**, 107–118.

Kurmayer, R., Christiansen, G. & Chorus, I. (2003). The abundance of microcystin-producing genotypes correlates positively with colony size in *Microcystis* sp. and determines its microcystin net production in Lake Wannsee. *Appl Environ Microbiol* **69**, 787–795.

Kurmayer, R., Christiansen, G., Fastner, J. & Börner, T. (2004). Abundance of active and inactive microcystin genotypes in populations of the toxic cyanobacterium *Planktothrix* spp. *Environ Microbiol* **6**, 831–841.

Lawton, L. A., Edwards, C. & Codd, G. A. (1994). Extraction and high-performance liquid chromatographic method for the determination of microcystins in raw and treated waters. *Analyst* **119**, 1525–1530.

Luukkainen, R., Sivonen, K., Namikoshi, M., Färdig, M., Rinehart, K. L. & Niemelä, S. I. (1993). Isolation and identification of eight microcystins from thirteen *Oscillatoria agardhii* strains and structure of a new microcystin. *Appl Environ Microbiol* **59**, 2204–2209.

Marahiel, M. A., Stachelhaus, T. & Mootz, H. D. (1997). Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chem Rev* **97**, 2651–2674.

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 *mcyABC* and impact on microcystin production in *Microcystis* strains. *J Bacteriol* **185**, 2774–2785.

Moffitt, M. C. & Neilan, B. A. (2004). Characterization of the nodularin synthetase gene cluster and proposed theory of the evolution of cyanobacterial hepatotoxins. *Appl Environ Microbiol* **70**, 6353–6362.

Mur, L. R., Skulberg, O. M. & Utkilen, H. (1999). Cyanobacteria in the environment. In *Toxic Cyanobacteria in Water: a Guide to their Public Health Consequences, Monitoring and Management*, pp. 15–40. Edited by I. Chorus & J. Bartram. London & New York: E. & F. N. Spon.

Oliver, R. L. & Ganf, G. G. (2000). Freshwater blooms. In *The Ecology of Cyanobacteria: their Diversity in Time and Space*, pp. 149–194. Edited by B. A. Whitton & M. Potts. Dordrecht: Kluwer Academic.

Rapala, J., Sivonen, K., Lyra, C. & Niemelä, S. I. (1997). Variation of microcystins, cyanobacterial hepatotoxins, in *Anabaena* spp. as a function of growth stimuli. *Appl Environ Microbiol* **63**, 2206–2212.

Rinehart, K. L., Namikoshi, M. & Choi, B. W. (1994). Structure and biosynthesis of toxins from blue-green algae (cyanobacteria). *J Appl Phycol* **6**, 159–176.

Rippka, R. (1988). Isolation and purification of cyanobacteria. *Methods Enzymol* **167**, 3–27.

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

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**, 686–692.

Sano, T. & Kaya, K. (1995). A 2-amino-2-butenoic acid (Dhb)-containing microcystin isolated from *Oscillatoria agardhii*. *Tetrahedron Lett* **36**, 8603–8606.

Sano, T. & Kaya, K. (1998). Two new (*E*)-2-amino-2-butenoic acid (Dhb)-containing microcystins isolated from *Oscillatoria agardhii*. *Tetrahedron* **54**, 463–470.

Scheffer, M., Rinaldi, S., Gragnani, A., Mur, L. R. & van Nes, E. H. (1997). On the dominance of filamentous cyanobacteria in shallow, turbid lakes. *Ecology* **78**, 272–282.

Sivonen, K. & Jones, G. (1999). Cyanobacterial toxins. In *Toxic Cyanobacteria in Water: a Guide to their Public Health Consequences, Monitoring and Management*, pp. 41–112. Edited by I. Chorus & J. Bartram. London & New York: E. & F. N. Spon.

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

Stoesser, G., Baker, W., van den Broek, A. & 13 other authors (2003). The EMBL Nucleotide Sequence Database: major new developments. *Nucleic Acids Res* **31**, 17–22.

Suda, S., Watanabe, M. M., Otsuka, S., Mahakahant, A., Yongmanitchai, W., Nopartnaraporn, N., Liu, Y. & Day, J. G. (2002). Taxonomic revision of water-bloom-forming species of oscillatorioid cyanobacteria. *Int J Syst Evol Microbiol* **52**, 1577–1595.

Tanabe, Y., Kaya, K. & Watanabe, M. M. (2004). Evidence for recombination in the microcystin synthetase (*mcy*) genes of toxic cyanobacteria *Microcystis* spp. *J Mol Evol* **58**, 633–641.

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**, 753–764.