Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: an integrated peptide–polyketide synthetase system

Daniel Tillett [1], Elke Dittmann [2], Marcel Erhard [3], Hans von Döhren [3], Thomas Börner [2] and Brett A Neilan [1]

Background: Blooms of toxic cyanobacteria (blue-green algae) have become increasingly common in the surface waters of the world. Of the known toxins produced by cyanobacteria, the microcystins are the most significant threat to human and animal health. These cyclic peptides are potent inhibitors of eukaryotic protein phosphatases type 1 and 2A. Synthesized nonribosomally, the microcystins contain a number of unusual amino acid residues including the β-amino polyketide moiety Adda (3-amino-9-methoxy-2,6,8-trimethyl-10-phenyl-4,6-decadienoic acid). We have characterized the microcystin biosynthetic gene cluster from *Microcystis aeruginosa* PCC7806.

Results: A cluster spanning 55 kb, composed of 10 bidirectionally transcribed open reading frames arranged in two putative operons *(mcyA–C* and *mcyD–J)*, has been correlated with microcystin formation by gene disruption and mutant analysis. Of the 48 sequential catalytic reactions involved in microcystin synthesis, 45 have been assigned to catalytic domains within six large multienzyme synthases/synthetases (McyA–E, G), which incorporate the precursors phenylacetate, malonyl-CoA, *S*-adenosyl-L-methionine, glutamate, serine, alanine, leucine, D-methyl-isoaspartate, and arginine. The additional four monofunctional proteins are putatively involved in *O*-methylation (McyJ), epimerization (McyF), dehydration (McyI), and localization (McyH). The unusual polyketide amino acid Adda is formed by transamination of a polyketide precursor as enzyme-bound intermediate, and not released during the process.

Conclusions: This report is the first complete description of the biosynthesis pathway of a complex cyanobacterial metabolite. The enzymatic organization of the microcystin assembly represents an integrated polyketide–peptide biosynthetic pathway with a number of unusual structural and enzymatic features. These include the integrated synthesis of a β-amino-pentaketide precursor and the formation of β- and γ-carboxyl-peptide bonds, respectively. Other features of this complex system also observed in diverse related biosynthetic clusters are integrated *C*- and *N*-methyltransferases, an integrated aminotransferase, and an associated *O*-methyltransferase and a racemase acting on acidic amino acids.

[1]School of Microbiology and Immunology, The University of New South Wales, Sydney, NSW 2052, Australia
[2]Institute for Biology (Genetics), Humboldt University of Berlin, Berlin 10115, Germany
[3]Institute for Biochemistry and Molecular Biology, Technical University of Berlin, Berlin 10587, Germany

Correspondence: Brett A Neilan
E-mail: b.neilan@unsw.edu.au

Keywords: Cyanobacterium; Microcystin; *Microcystis aeruginosa*; Nonribosomal peptide synthetase; Polyketide synthase

Received: **21 March 2000**
Revisions requested: **18 April 2000**
Revisions received: **13 July 2000**
Accepted: **27 July 2000**

Published: **18 August 2000**

Chemistry & Biology 2000, **7**:753–764

1074-5521/00/$ – see front matter

PII: S1074-5521(00)00021-1

Introduction

Water blooms of cyanobacteria (blue-green algae) are of worldwide concern due to their production of a range of hepatotoxins and neurotoxins. One group of these toxins, the microcystins, is a remarkable family of more than 65 cyclic heptapeptides produced by a diverse range of cyanobacteria, including species of the genera *Microcystis*, *Anabaena*, *Nostoc* and *Oscillatoria* [1,2]. These potent eukaryotic serine/threonine protein phosphatases (PP) 1 and 2A inhibitors share the common structure cyclo(Adda-D-Glu-Mdha-D-Ala-L-X-D-MeAsp-L-Z-), where X and Z are variable L-amino acids, Adda is 3-amino-9-methoxy-2,6,8,-trimethyl-10-phenyl-4,6-decadienoic acid, D-MeAsp is 3-methylaspartic acid, and Mdha is *N*-methyl-dehydroalanine (Figure 1) [1,3]. Toxicity in vertebrates is mediated through the active transport of microcystin into hepatocytes by the bile acid organic anion transport system [4]. Acute poisoning, leading to death from massive hepatic hemorrhage, has been reported in both animals and humans [5–7]. Chronic ingestion of sub-lethal doses has been demonstrated to induce primary hepatocellular carcinoma in rodents [8] and has been epidemiologically linked to primary liver cancer in humans [9,10].

Biochemical and genetic studies have suggested a mixed polyketide synthase (PKS)/nonribosomal peptide synthe-

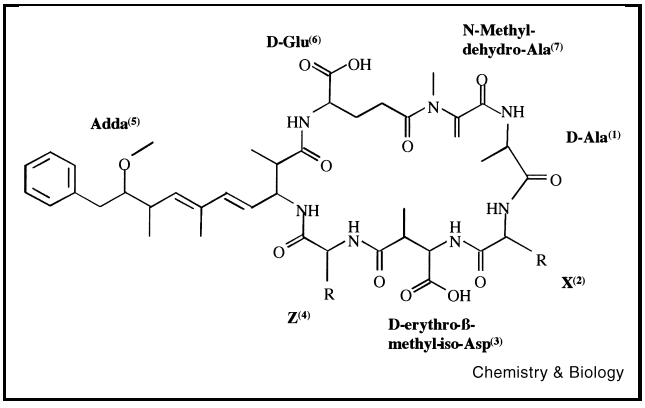

Figure 1. General structure of the microcystins. Variable L-amino acid residues are found at positions X and Z.

tase (NRPS) origin for the microcystins [11–13]. Feeding experiments using labeled precursors have shown the activation and incorporation of both amino acids and acetate units into microcystin-LR [11]. Genetic studies have demonstrated the involvement of a large NRPS gene cluster in MCYST biosynthesis [12]. Five domains of this cluster *(mcyABC)* have also been recently sequenced from *Microcystis* strain K139 [14]. While polyketides and nonribosomal peptides appear structurally unrelated, they are assembled in a remarkably similar manner. Both groups of metabolites are biosynthesized by large, multifunctional protein complexes that are organized into coordinate clusters of enzymatic sites termed modules, in which each module is responsible for one cycle of polyketide or polypeptide chain elongation [15–17]. The order of these modules, together with the number and type of catalytic domains within each module, determines the structure of the resulting polyketide or peptide product.

Despite the general interest in both NRP and PK synthesis, relatively few examples of metabolites of mixed origin have been biochemically or genetically characterized. These include rapamycin, FK506, mycobactin, epothilon, mycosubtilin, yersiniabactin, and the antibiotic TA [18–21]. We have undertaken the cloning of a large 55 kb gene cluster from *Microcystis aeruginosa* PCC7806 responsible for the biosynthesis of microcystin-LR. Analysis of this region revealed six large open reading frames (ORFs) (*mcyA–E* and *G*) of a mixed NRPS/PKS nature, together with a further four small ORFs (*mcyF* and *H–J*) with putative precursor and microcystin tailoring functions. Gene disruption studies provide evidence for the involvement of *mcyA* and *mcyD* in microcystin biosynthesis. Additionally, Western blot analysis of partially purified microcystin synthetase revealed homology to other characterized NRPS.

Results

Cloning and sequencing of the microcystin synthetase gene cluster

A 758 bp fragment of the microcystin synthetase gene, *mcyB*, was amplified using primers FAA and RAA [22]. This fragment was used to screen a λ Zap library of *M. aeruginosa* PCC7806 and isolate a clone containing a 7 kb fragment of the *mcy* cluster. As further attempts to isolate flanking clones from this library proved unsuccessful the remaining *mcy* sequence was obtained by various PCR gene walking approaches [22,23]. In total, 63.6 kb of the *mcy* gene cluster and flanking regions was isolated from *M. aeruginosa* PCC7806 (Figure 2). The G+C content of the nucleotide sequence *mcy* region (992–55 448; *mcyJ* to *mcyC*) is 39.2%. This is similar to the total G+C content of *M. aeruginosa* PCC7806 at 41.6% [24].

Sequence analysis of the *mcy* region revealed a bidirectional operonic structure (Figure 2). The larger of the two putative operons (*mcyD–J*) encodes the PKS–NRPS modules catalyzing the formation of the pentaketide-derived β-amino acid Adda and its linkage to D-glutamate, while the smaller (*mcyA–C*) encodes the NRPS modules for the extension of this dipeptidyl intermediate to the heptapeptidyl step and subsequent peptide cyclization.

Structural organization of the *mcyD–J* region

The *mcyD–J* gene cluster contains seven ORFs all transcribed in opposite direction to the putative *mcyABC* oper-

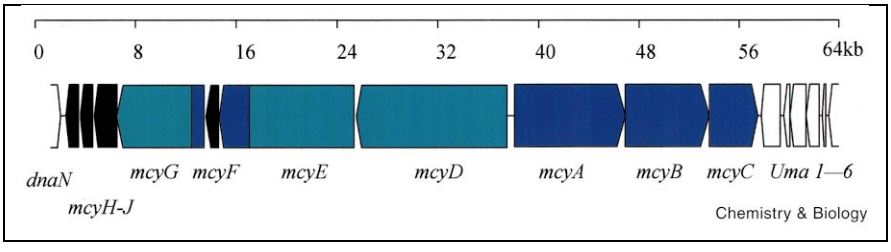

Figure 2. Organization of the gene cluster for microcystin biosynthesis. The direction of transcription and relative sizes of the ORFs deduced from the analysis of the nucleotide sequence are indicated. ORFs containing regions homologous to nonribosomal peptide synthetases or polyketide syntheses [15,16] are indicated in dark and light blue, respectively. Additional ORFs of putative microcystin tailoring function are indicated in black. Non-microcystin synthetase ORFs are shown in white.

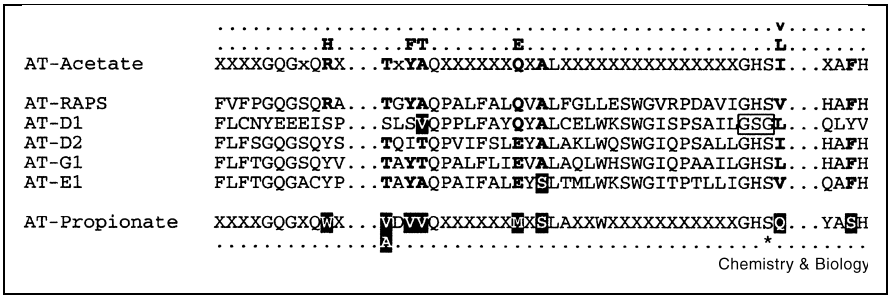

Figure 3. Alignment of the AT domains motifs and active sites from McyD, McyE and McyG. Included in the sequence alignment is the AT domain from rapamycin PKS (AT-RAPS) and the putative consensus acetate (ATAcetate) and propionate (AT-Propionate) AT domains sequences of Ikeda et al. [49] from the alignment of 57 AT domains of erythromycin, pikromycin, tylosin, niddamycin, rapamycin, and avermectin PKS. Bold and inverse letters indicate the proposed significant differences between acetate and propionate AT loading domains, respectively. The asterisk indicates the serine residue that is linked to the acyl-CoA in the acyl:acyltransferase complex. The highly divergent AT active site of McyD-1 is boxed.

on (Figure 2). The first ORF (11 721 bp) in this region (*mcyD*) is located 733 bp upstream of *mcyA* and encodes a large 435 714 Da polypeptide with high similarity to known type I PKS [15]. Alignment of McyD with PKS domains identified two type I modules, each consisting of a β-ketoacyl synthase (KS), an acyltransferase (AT), β-ketoacyl reductase (KR), a dehydratase (DH), and an acyl carrier protein (ACP). On the basis of conserved sequence motifs,

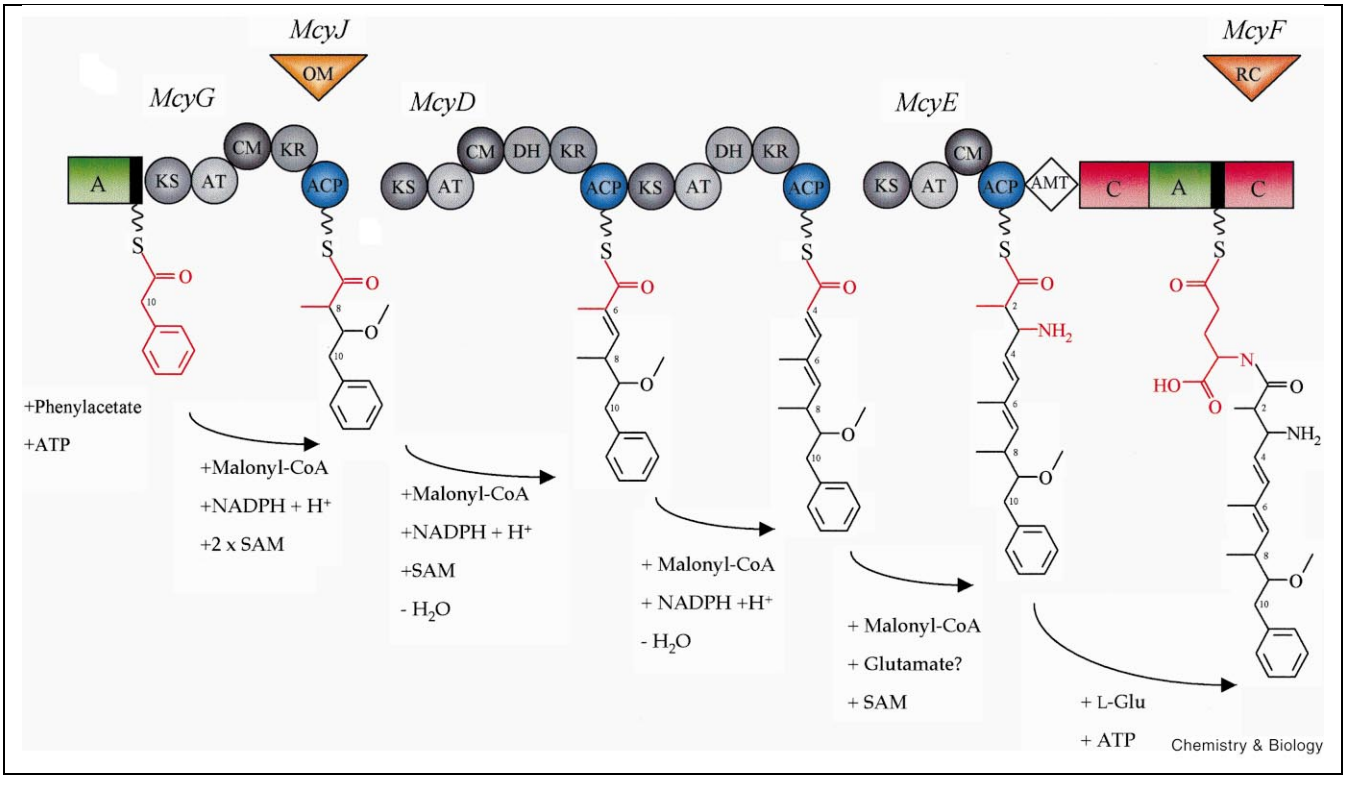

Figure 4. Model for the formation of Adda and predicted domain structure of McyG, McyD and McyE. Each circle and rectangle represents respectively a PKS or NRPS enzymatic domain. The putative aminotransferase domain is represented by a diamond. The activity of the tailoring ORFs, McyJ and F, are shown as inverted triangles. KS, β-ketoacyl synthase; AT, acyltransferase; ACP, acyl carrier protein; KR, ketoacyl reductase; DH, dehydratase; CM, *C*-methyltransferase; OM, *O*-methyltransferase; A, aminoacyl adenylation; C, condensation; AMT, aminotransferase; RC, racemase. The NRPS thiolation motif is shown in black. Reaction order shows transfer and condensation of Adda to D-glutamate.

both AT domains of McyD appear to accept malonyl-CoA. This is consistent with previous labeled precursor studies [11] and in comparison to the respective sequences for acetate and propionate providing modules (Figure 3). In addition, the amino-terminal module contains a putative *C*-methyltransferase (CM) domain of a type recently found in the yersiniabactin biosynthetic cluster [20] (Figure 4).

The ATG start codon and putative ribosome binding site (RBS) of the second ORF, *mcyE*, is located 167 bp downstream of the TAA stop codon of *mcyD*. This large 10 464 bp ORF encodes a 392 703 Da polypeptide product of mixed PKS and NRPS function. The amino-terminal region of McyE contains a PKS module consisting of a KS, AT, and ACP domain, linked to a CM domain (as found in McyD) and a putative aminotransferase (AMT) domain. This latter domain, of about 430 amino acids, shows approximately 30% identity to a large group of nonintegrated AMTs acting on glutamate semialdehyde or *N*-acetyl-ornithine. The pyridoxal phosphate binding motif, DE-(VI)(IMQ)TGFR with the adjacent conserved lysine residue, is present in this domain. The most logical role of this module in microcystin biosynthesis would be to supply the amino group to Adda. A similar domain has been identified recently in the mycosubtilin synthetase cluster, presumably also involved in the synthesis of a β-amino acid [21].

Adjacent to the PKS module of McyE is an NRPS module composed of two condensation domains, an adenylation domain, and a thiolation domain (Figure 4). The first condensation domain is predicted to catalyze the formation of the peptide bond between Adda and the α-amino group of D-glutamate. The adenylation domains were analyzed for substrate specificity employing a recently derived algorithm [25,26]. This procedure selects eight or nine putative pocket-lining residues from each adenylation domains, between the core motifs A4 and A5, to predict the respective amino acid substrate. Substrate binding pocket analysis of the McyE activation domain revealed similarity to the glutamate activation domain of surfactin synthetase [27], the glutamine activation domain of lichenysin synthetase [28], and the α-aminoadipate activating reductase involved in the lysine pathway of *Penicillium chrysogenum* [29]. The sequences of both condensation domains cluster together with both the second condensation domain of McyB and the condensation domain of McyC. These four domains contain divergent condensation domain motifs C3–C5 [16], possibly related to their involvement in the peptide bond formation with substrates containing activated β- and γ-carboxyl groups (Figure 5). The carboxy-terminal location of an NRPS condensation domain, as found in McyE, has only been described in the mycosubtilin cluster [21]. Similar carboxy-terminal condensation domains have also been identified within the rapamycin and FK506 clusters where they are putatively involved in the termination of macrolactone synthesis [30,31]. While the second McyE condensation domain appears to be involved in the peptide bond formation between D-glutamate and *N*-methyl-dehydroalanine, the mechanism of this unusual reaction remains unclear.

Located 32 bp downstream of the TAG stop codon of *mcyE* is a small ORF (*mcyF*) of 756 bp encoding a 28 192 Da polypeptide. This putative protein revealed 29% identity (51% similarity) to an aspartate racemase (*racD-1*) from *Pyrococcus abyssi*, and somewhat less similarity to a family of glutamate racemases. Although the role of McyF in microcystin biosynthesis has yet to be confirmed, it appears likely to be involved in either the supply of D-glutamate or D-MeAsp, or the peptidyl epimerization of L-glutamate.

The putative RBS and ATG codon of the fourth ORF (*mcyG*) is located 132 bp downstream of the TGA stop codon of *mcyF*. This large ORF (7896 bp) encodes a 294 266 Da polypeptide of mixed NRPS and PKS function. The amino-terminal region of McyG contains an NRPS

	C1	C2	C3	C4	C5	C6	C7
Consensus	SxAQxRLWxL	RHExLRTxF	MHHxISDGWS	YxDFAVW	IGxFVNTQCxR	HQDYPFE	RDxSRNPL
	MY		V	Y	V LA	N V	
McyD-C1(Adda-Glu)	SEAQRQLWLL	RHEALRTKI	THHIVADGWS	FRQYLTL	VGFCSQFLPLR	HQSYTLE	RDFSRSPL
McyD-C2(γGlu-Ser)	SHGQRRLWAL	RHEILRTFF	MHHIIFDGWS	YKDYTSW	IGFYVNTLVLR	HQDYPFD	RDPSRNAL
McyA-C1(mSer-Ala)	TALQLGMIFH	RHSVLRTLF	FHHSILDGWS	FRDFVAL	LGLFLNTLPLR	WRRYPLA	EIPAREVL
McyB-C1(DAla-Leu)	SPMQEGMLFH	RHPVLRTSF	HHHIILDGWS	YQDYIVW	VGLFINTLPRV	YAYVSLA	DIPPGVPL
McyB-C2(Leu-βAsp)	SHAQRRFYVL	RHESLRTSF	IHHIICDGWS	YKDYAGW	IGLFLNTLVIR	HSDYPFD	REINRTPL
McyC-C1(βAsp-Arg)	SHGQRRLWIL	RHEILRTTF	VHHIIGDAQS	YKDYAAW	IGFYVNTLALR	YRDYPFD	RDLSRNPL

Chemistry & Biology

Figure 5. Condensation domain core motifs in microcystin biosynthesis. The consensus sequences for motifs (C1–C7) of peptide synthetase condensation domains are indicated above the motifs identified in *M. aeruginosa* PCC7806 [16,17].

adenylation domain that clusters with the acyl-CoA synthetases, insect luciferases, and aryl-carrier protein synthetases of NRPS systems. Labeled precursor studies [11] suggest the direct activation of phenylacetate by this adenylation domain, however, binding pocket analysis gives no clear result. This domain would thus serve as a PKS starter supply domain, an organizational form also found in the rapamycin, FK506, and yersiniabactin systems [20,30,31]. The adjacent carrier domain sequence shows similarities to other NRPS carrier domains (26–32% identity). The carboxy-terminal PKS module of McyG consists of malonyl-specific KS, AT (Figure 3), CM, DH, KR, and ACP domains (Figure 4).

The second ATG codon, located 224 bp downstream of the TAA stop codon of *mcyG*, was selected as the initiation codon for McyH. This 1617 bp ORF encodes a putative 67 100 Da transmembrane protein belonging to the ABC transporter ATP binding family. McyH displays 38% identity (58% similarity) with an ABC transporter (sll0182) isolated from *Synechocystis* sp. PCC6803 [32]. No obvious biosynthetic function can be assigned to McyH, however, it is worth speculating that McyH may play a role in the thylakoid localization of microcystin previously observed in *M. aeruginosa* [33].

Located 39 bp downstream of the TAA stop codon of *mcyH* is the ATG start codon of the sixth ORF, *mcyI*. This 1014 bp ORF encodes a 36 838 Da polypeptide revealing a 41% identity to the catalytic region of D-3-phosphoglycerate dehydrogenase (SerA) from *Methanobacterium thermoautotrophicum* [34]. While no definite function can be assigned to McyI, its location within the *mcy* gene cluster suggests it may have a role in securing the required serine precursor of *N*-methyl-dehydroalanine, or in the synthesis of *N*-methyl-dehydroalanine following serine activation.

The final ORF in the cluster (*mcyJ*) is located 176 bp downstream of the TAA stop codon of *mcyI*. The lack of a potential RBS upstream of the first ATG in this 837 bp ORF suggests that the alternative start codon TTG, located 9 bp downstream, is used to initiate translation. This ORF encodes a 31 904 Da polypeptide with 35% identity (51% similarity) to the erythromycin synthase *O*-methyltransferase from *Saccharopolyspora erythraea* [35].

Analysis of the region 569 bp downstream of *mcyJ* revealed the presence of a partial ORF of >422 bp transcribed in the opposite orientation to the *mcyD–J* cluster (Figure 2). This ORF appears to encode a *M. aeruginosa* ortholog (70% identity and 83% similarity) of the DNA polymerase III β subunit (*dnaN*) from *Synechocystis* sp. PCC6803 [32].

Structural organization of the *mcyABC* and *uma1–6* regions

The smaller of these two putative *mcy* operons, *mcyABC*, contains three large ORFs. Translation of the first ORF, *mcyA* (8388 bp), appears to be initiated from the second ATG codon as the first ATG codon lacks a suitable RBS. The gene encodes an NRPS of 315 717 Da, containing two adenylation and thiolation domains, respectively, a condensation domain, an *N*-methyltransferase (NMT) domain, and an epimerization domain (Figure 6). Binding pocket analysis of the first adenylation domain of McyA revealed clustering of this domain with a group of domains known or suspected to activate L-serine.

The NMT domain (422 aa) is inserted between the core motifs A8 and A9, and reveals extensive homologies with the eight fungal (seven from cyclosporin synthetase [36] and one from enniatin synthetase [37]) and the two bacterial NMT domains (virginiamycin S synthetase and pristinamycin I [38]) known so far. Identities to the bacterial domains were 35.6% and 38.3%, while the fungal domains showed identities in the 26–29% range. The presence of an NMT domain allows the amino acid specificity of the first activation domain of McyA to be assigned, as microcystins possess only one *N*-methylated residue, *N*-methyl-dehydroalanine [39]. Interestingly, the bacterially derived NMT domains are significantly more similar to each other than to those of fungal origin, suggesting that these domains may be divided into fungal and bacterial families.

The first thiolation domain, presumably involved in the transport of the *N*-methyl-seryl residue, and the condensation domain of McyA, are not related to known domains involved in the condensation of *N*-methylated amino acids [17]. It is thus not certain, if, in analogy to the well studied systems of actinomycin, enniatin, and cyclosporin synthesis, the *N*-methyl-seryl thioester is processed, or if methylation follows dehydration or peptide bond formation.

The second adenylation domain of McyA, presumably activating L-alanine, shows a slightly reduced spacing of the pocket-lining residues, and is structurally related to the putative saframycin synthetase glycine adenylation domain, and the presumed serine adenylation domain of the DHBF protein of *Bacillus subtilis* [25].

The sequence of the second thiolation domain of McyA, if compared to available sequences of peptidyl carrier domains, groups with the large cluster of carriers having D-amino acids attached, or those adjacent to epimerization (Ep) domains. The terminating Ep domain is highly similar to other characterized integrated NRPS Ep domains and contains the conserved motifs E1–7 [16].

The ATG start codon and putative RBS of the second ORF, *mcyB*, is located 15 bp downstream of the TAA stop codon of *mcyA*. This 6318 bp ORF encodes a peptide synthetase of 242 334 Da containing two modules, each possessing adenylation, thiolation, and condensation do-

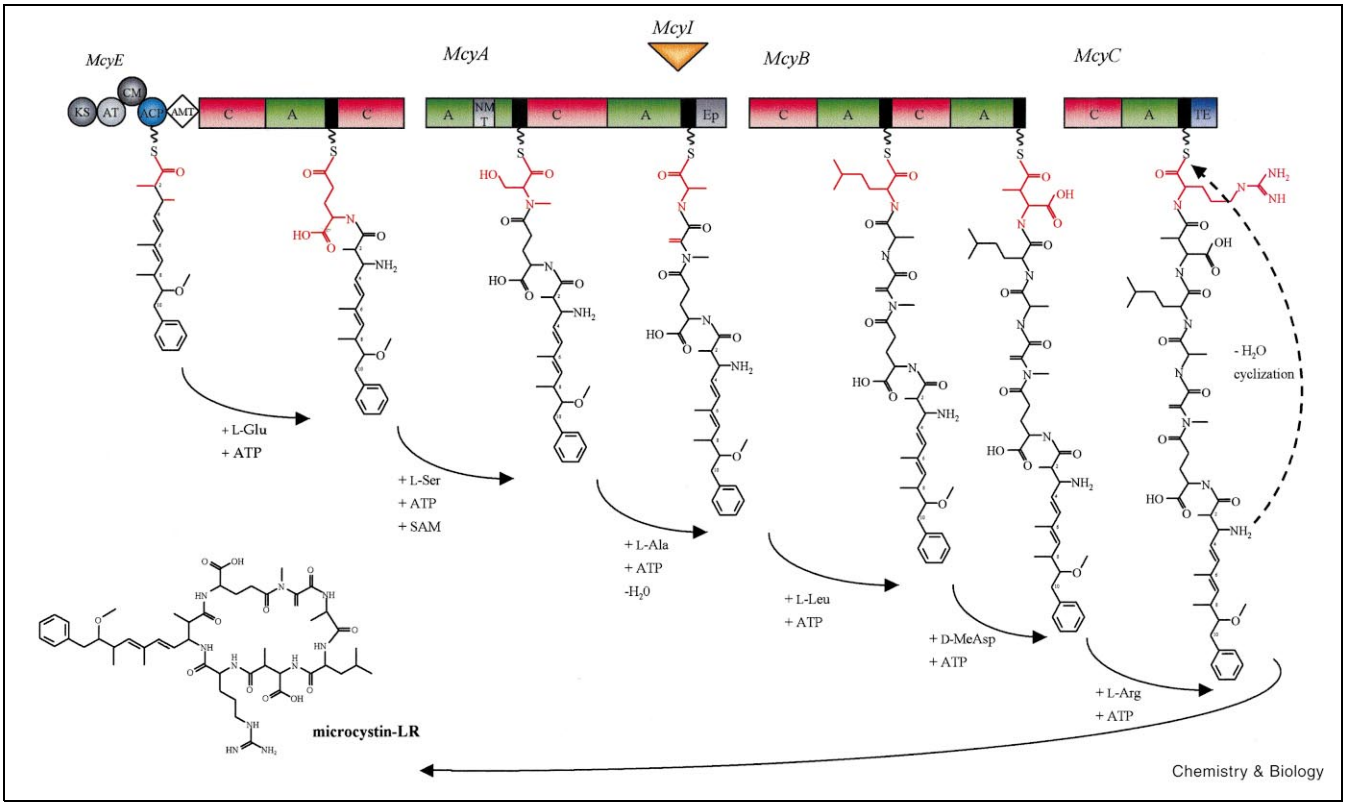

Figure 6. Biosynthetic model for microcystin-LR and predicted domain structure of McyE, McyA, McyB, and McyC. Each circle and rectangle represents, respectively, a PKS or NRPS enzymatic domain. The aminotransferase domain is represented by a diamond. The activity of the putative tailoring ORF, McyI, is shown as an inverted triangle. KS, β-ketoacyl synthase; AT, acyltransferase; CM, *C*-methyltransferase; ACP, acyl carrier protein; A, aminoacyl adenylation; C, condensation; NM, *N*-methyltransferase; Ep, epimerization; TE, thioesterase; AMT, aminotransferase. The NRPS thiolation motif is shown in black. Aminoacyl activation and condensation order is predicted: L-Z-Adda, L-glutamate, L-methylserine, D-alanine, L-leucine, D-methyl-aspartate, L-arginine, cyclization.

mains (Figure 6). The amino-terminal domain has been functionally identified by sequence alignment with known condensation domains as catalyzing peptide bond formation between L- and D-aminoacyl residues [17]. This is in agreement with the expected acceptance of the carboxy-terminal D-alanyl-peptidyl intermediate from McyA. From the structure of microcystin, the second activation domain would be expected to activate L-leucine and is in agreement with the predicted activation domain substrate pocket specificity. In *M. aeruginosa* PCC7806 the two major microcystins contain L-leucine and L-arginine at the variable amino acid positions (Figure 1) [40]. Minor forms of the toxin in this strain detected by MALDI-TOF mass spectrometry reveal either tyrosine or arginine replacement of the variable leucine residue [12].

The second condensation domain of McyB aligns closely to domains of surfactin and lichenysin synthetases [27,28] involved in peptide bond formation between leucyl and glutamyl or glutaminyl residues. The predicted binding pocket of the adjacent adenylation domain maps to a group of domains specific for aspartic acid or asparagine. Although the expected D-β-MeAsp acid pocket is well predicted, the activation of the β-carboxyl group cannot yet be related to sequence data. It should be noted that the α-carboxyl group is not activated, and thus the condensation domain does not belong to the cluster of L–D-peptide bond forming domains [26].

The third ORF, *mcyC*, is located 4 bp upstream of the TGA stop codon of *mcyB*, starting with an ATG codon 7 bp downstream of a putative RBS. This 3876 bp ORF encodes a 147 781 Da peptide synthetase with a carboxy-terminal thioesterase (TE) domain (Figure 6). The final activation domain is expected to activate predominately L-arginine, however, this domain fails assignment attempts. This residue is invariant in microcystins isolated from *M. aeruginosa* PCC7806 [40]. The specificity-conferring region has some similarity with serine-specific domains of McyA and syringomycin synthetase, as well as the glutamine-specific domain of tyrocidin synthetase [41]. The TE domain of about 240 amino acid residues is highly similar (26.6–

29.3% identity) to a group of integrated domains involved in the cyclization of various peptidyl intermediates. These include the systems forming surfactin, bacitracin, lichenysin, fengycin, tyrocidin, and gramicidin S. Additionally, the respective thioesterase motifs GXSXG and GXH [16] present in McyC (GHSAG and ATGIHREM) are unique.

Analysis of the region downstream of *mcyC* revealed the presence of six ORFs transcribed in the opposing direction to the *mcyABC* gene cluster (Figure 2). Although no function can be assigned to *uma1–6*, it appears unlikely that these ORFs are involved in microcystin synthesis as they are present in both toxic and nontoxic *Microcystis* strains [42]. Interestingly, *uma4* encodes a 47 465 Da peptide with 45% identity (65% similar) to TnpA, a transposase isolated from *Anabaena* sp. PCC7120. Speculatively, this ORF may have played a role in the acquisition of this biosynthetic cluster by *M. aeruginosa* PCC7806.

Disruption of *mcyA* and *mcyD*

To confirm the function of *mcyA* and *mcyD* in microcystin synthesis, we disrupted these genes by homologous integrative transformation. The *mcyA* and *mcyD* disruption constructs, pMCYA5C and pMCYD7C, were introduced into *M. aeruginosa* PCC7806 by both natural transformation and electroporation. Transformation of *M. aeruginosa* PCC7806 by electroporation resulted in approximately 150-fold more disruption mutants per μg of plasmid DNA than that obtained by natural transformation (data not shown). Two clones were randomly selected from each disruption experiment and studied further. PCR amplification of *mcyA* and *mcyD* from the respective PCC7806 Cm^r mutant clones showed the stable chromosomal integration of either pMCYA5C or pMCYD7C as single homologous double crossover recombination events (data not shown).

Analysis of both the *mcyA* and *mcyD* disruption mutants by both MALDI-TOF mass spectrometry and PP inhibition showed both mutants produced no detectable levels of microcystin (Figure 7). Neither disruption mutation affected the production of cyanopeptolin A and D, the other non-ribosomal peptides of *M. aeruginosa* PCC7806 [43]. MALDI-TOF mass spectrometry revealed not only the absence of all four wildtype microcystin isoforms (microcystin-LR, demethyl-LR, -RR and -YR), but also the absence of any partial, truncated or linear forms of microcystin (Figure 7). These results confirm the specific involvement of *mcyA* and *mcyD* in microcystin production. This is in accordance with the previously described *mcyB* mutants [12].

Partial purification of microcystin synthetase

Partial purification of the microcystin synthetase was performed from both wildtype *M. aeruginosa* PCC7806 and the previously described *mcyB* disruption mutant cell extracts [12]. SDS–PAGE analysis revealed several high molecular

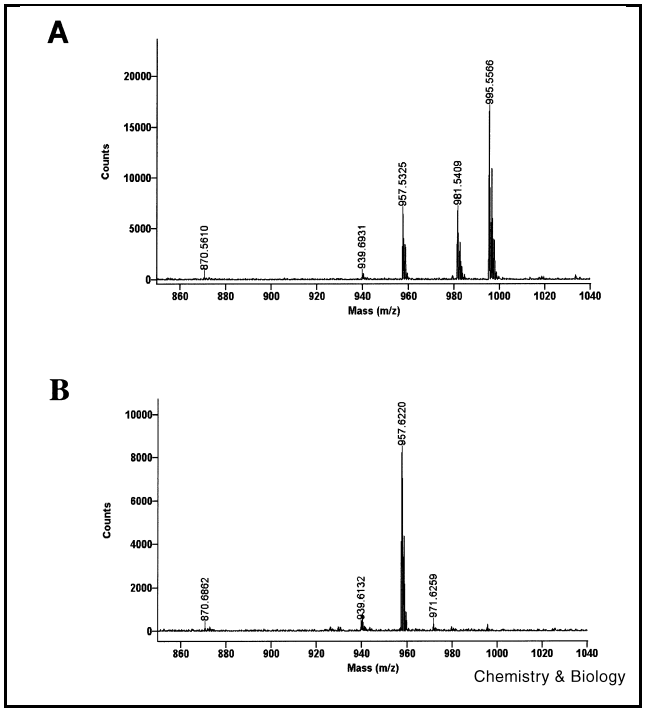

Figure 7. MALDI-TOF mass spectrometry analysis of whole cells from **(A)** wild-type and **(B)** *mcyA* mutant. Major peaks in the wildtype include microcystin-LR and demethylated microcystin-LR (*m/z* 995 and 981) and cyanopeptolins A and D (*m/z* 939 and 957) [43]. Both microcystin peaks are absent from the mutant strain, whereas the cyanopeptolins are still present. No partial peptides derived from microcystin were detected.

weight proteins absent from the *mcyB* mutant (Figure 8A). No protein above 200 kDa can be seen in the mutant. However, on the basis that the disruption is in the *mcyABC* cluster, McyD, McyE and McyG should be present.

Western blotting was performed with an antibody raised against gramicidin synthetase 2 [44]. This NRPS-specific antibody identified four wildtype proteins that are absent from the *mcyB* mutant (Figure 8B). While direct evidence for the involvement of these four proteins in microcystin synthesis requires amino-terminal sequencing, they do correspond approximately to the predicted masses of McyC, McyB, McyA, and McyE. Interestingly, the gramicidin synthetase 2 antibody did not appear to identify the NRPS domain of McyG. However, given the high sequence divergence identified within the McyG NRPS domain, this is unsurprising. Finally, the sixth high molecular weight protein encoded by the *mcy* gene cluster, McyD, was not identified by the gramicidin synthetase 2 antibody. This is not unexpected given the absence of any NRPS-like regions within the PKS modules of McyD.

Transcript analyses have shown that the transcription of the *mcyABC* operon is unaffected by disruption of *mcyB*

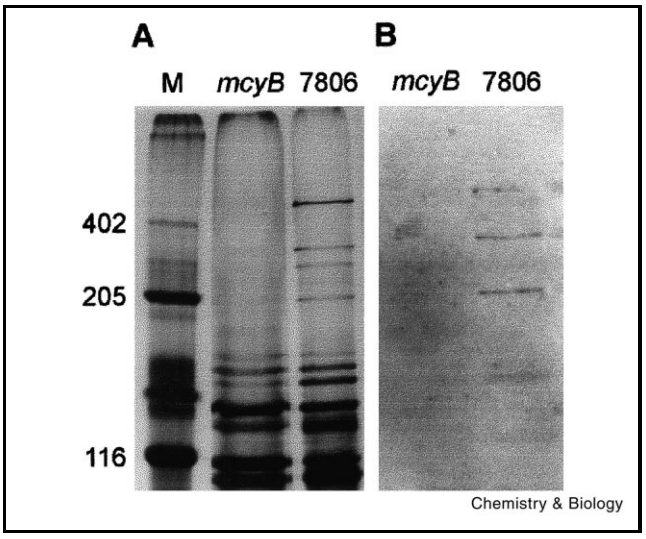

Figure 8. (A) Partial purification of microcystin synthetase by AcA22 gel filtration. Samples were run on a 5% SDS–PAGE gel. High molecular weight protein fractions of the wildtype PCC7806. The following sizes were determined for the high molecular weight bands: 210 ± 30 kDa; 270 ± 50 kDa; 310 ± 50 kDa; 440 ± 80 kDa; 500 ± 100 kDa. **(B)** Western blot of AcA22 filtrates probed with the GS2 antibody. Three of the high molecular weight bands and an additional band of 150 ± 10 kDa gave clear signals probably corresponding to McyE, McyA, McyB and McyC, respectively. The AcA22 fractions obtained from the *mcyB* mutant are missing these high molecular weight bands.

(data not shown). It is therefore somewhat surprising that all detectable multifunctional proteins of the *mcy* cluster are absent from the *mcyB* mutant (Figure 8). While in agreement with the lack of partial microcystin products within the *mcyB* [12], *mcyA* and *mcyD* mutants, this does suggest that the stability of the microcystin synthetase complex requires the formation of the complete complex. Such degradation of partial synthetase complexes has been previously observed in the rifamycin synthase system [45] and may be a general feature of these complex systems.

Discussion

M. aeruginosa PCC7806 produces the potent inhibitor of eukaryotic PP 1 and 2A, microcystin. In this report we describe the identification and sequencing of the gene cluster that specifies the synthesis of this cyclic peptide–polyketide hybrid. Genetic disruption of genes within the *mcy* cluster abolished microcystin production, but did not affect the production of the cyanopeptolins.

The *mcy* gene cluster contains 55 kb of DNA encoding 10 ORFs, *mcyA–mcyJ*. These ORFs are flanked on both sides by genes not involved in microcystin biosynthesis which show high similarity to genes localized on the chromosome of *Synechocystis* sp. PCC6803. This strongly suggests the *mcy* gene cluster is located on the chromosome and not on a large plasmid as previously proposed [46]. Although the overall structure of the *mcy* cluster resembles that of other NRPS and type I PKS, the six ORFs (McyA–E and G) exhibit a number of interesting features. The initiating adenylation domain of McyG is unusual. This domain appears to activate phenylacetate in an NRPS manner before feeding it into the Adda PKS pathway. Activation of starter aryl or polyketide-derived carboxylic acids has been found in both integrated (rifamycin, FK506, rapamycin, and a second uncharacterized PKS system from *Streptomyces hygroscopicus* [47]) and non-integrated systems (pristinamycin, actinomycin, yersiniabactin, mycobactin). The McyG adenylation domain displays only a limited similarity of less than 25% identity to the activation domains of these enzymes. This is comparable to the similarity found within the acyl-CoA synthetase superfamily [26]. However, this domain revealed higher identities with an uncharacterized adenylate domain of the PKSJ cluster of unknown function from *B. subtilis* (43.2%) and the putative serine activating domain of McyA (31.1%). Binding pocket analysis groups the McyG activation domain with the yet uncharacterized alanine/glycine activating domains [25]. Interestingly, the adjacent carrier domain appears more closely related to the NRPS thiolation domains than the PKS ACP domains.

The proposed assembly of Adda by McyG, McyD and McyE is illustrated in Figure 4. The first step involves the activation of phenylacetate by McyG in an ATP-dependent reaction. The activated phenylacetate is transferred to the 4-phosphopantetheine cofactor of the first carrier domain. The structural divergence of this domain from other ACPs, together with its assumed function in the transfer of an aromatic carboxylic acid, suggests that this thiolation domain could be termed an aryl carrier protein. The AT domain loads the second ACP domain with a malonyl group derived from malonyl-CoA, and is coupled to the phenylacetate thioester in a decarboxylative condensation reaction catalyzed by the KS domain. *S*-Adenosyl-L-methionine (SAM)-dependent *C*-methylation of the Adda C_8 and keto reduction are catalyzed by the CM domain and KR domain, respectively. Subsequent *O*-methylation of the C_9 hydroxyl is performed by McyJ in a SAM-dependent fashion. We assume this modification to take place at this early stage by interaction between the methyltransferase and the PKS since all known microcystins carry this modification. Chain elongation proceeds by transfer to the KS domain of McyD, followed by two rounds of malonyl additions and dehydration/keto reduction reactions. Both AT and DH domains from the first module of McyD contain significant deletions, but appear to be still functional. The SAM-dependent *C*-methylation of C_6 is catalyzed by the single CM domain of McyD. The β-ketoacyl chain is transferred to McyE where a fourth malonyl group is added before being converted to a β-aminoacyl thioester via a transamination reaction catalyzed by the AMT domain of McyE. A similar AMT domain has been recently reported

for the hybrid PK/NRPS mycosubtilin synthetase [21]. The origin of the transferred amino group is unknown, although glutamate is implicated. The formation of the Adda intermediate differs from the precursor synthesis of the PKS-derived β-amino acid Bmt ((4*R*)-4-[(*E*)-2-butenyl]-4-methyl-L-threonine) involved in cyclosporin biosynthesis. In the case of cyclosporin synthesis the Bmt residue is produced as a free intermediate, while the Adda residue is retained. In addition, microcystin synthetase combines the synthesis of a pentaketide in a type I PKS system with the addition of a further six amino acids. In the recently described mycosubtilin system an acyl-CoA precursor is elongated once before being transformed into the β-aminoacyl intermediate [21].

McyE forms an integrated NRPS/PKS structure, where both types of modules are fused into a single polypeptide. Three types of integrated NRPS/PKS systems have been described. The first type combines both PKS and NRPS multienzymes in direct transfer of intermediates. Examples of this type include rapamycin, FK506, mycobactin and epothilon [18]. The second type contains integrated genes where both NRPS and PKS domains are found within a single protein peptide. These systems include mycosubtilin [21], the antibiotic TA [19], yersiniabactin [20], and microcystin synthetase. Systems using polyketide-derived free intermediates as CoA derivatives, for example, Bmt residue in cyclosporin and the acyl precursors of surfactin, are a third type of interacting systems.

The PKS system of the *mcy* cluster shows a number of deviations from known PKS type I structures. The AT domains of type I modular PKS can be arranged, on the basis of conserved sequence motifs, into two substrate-specific groups which load either acetate or propionate precursors [48]. Ikeda et al. [49] compared 57 PKS AT domains involved in the synthesis of erythromycin, pikromycin, tylosin, niddamycin, rapamycin, and avermectin, and identified a number of invariant AT residues for either acetate or propionate activation (Figure 3). Previous labeled precursor studies suggested that acetate units are exclusively incorporated during Adda biosynthesis [11]. While examination of the *mcy* AT domains supports acetate activation, many of the proposed invariant residues are not conserved. Most striking is the first AT domain of McyD where 14 of the 19 proposed invariant residues are missing (Figure 3). This domain also contains a novel serine acyl-CoA complex motif where the highly conserved GHS motif is replaced by a GSG motif.

The three integrated *C*-methyltransferase domains (CM) show a high degree of sequence conservation between themselves, comparable to that between the CM domains from *Sorangium cellulosum* (epothilon cluster), *Aspergillus terreus* (lovastatin cluster), or *Yersinia pestis* (yersiniabactin cluster). Three core motifs have been identified by Kagan and Clarke in the large class of SAM-dependent methyltransferases [50]. Structurally, motif I and post-I have been shown to interact directly with SAM, with all three motifs involved in the formation of a central parallel β-sheet structure [51]. Alignments of the putative *O*-methyltransferase, McyJ, the three CM domains of McyD, McyE and McyG, and the NMT domain of McyA with other SAM-dependent methyltransferase sequences revealed that the three methyltransferase core motifs are slightly altered in form in all five putative methyltransferases (Figure 9).

Protein	Motif I	Post I	Motif II	Motif III
	.....			V
	...L.....		.....A..Y.L.....	.....L.I
	IVEV.C.P.....	.....E...	GTY..VIV.....	K...IIFL
	VLDIGGGTG...	hhXhD...	PQFDAIFC...	LLRPGGRLLI
McyJ	**LLDVGFGFA...**	**LNTTE...**	**NSFDKLTA...**	**VLQPGGRLAV**
McyD-MT	**ILEIGGGTG...**	**YIFTD...**	**GSFDIIIA...**	**LIAPKGLLIL**
McyE-MT	**ILEIGGGTG...**	**YSFSE...**	**RNYHIVVA...**	**LLRPGGYLVL**
McyG-MT	**ILEIGAGTG...**	**YTFTD...**	**HSYDLIIA...**	**LLRPGGHLLL**
McyA-NMT	**VLEIGCGTG...**	**YWGTD...**	**GRFDTIVL...**	**VLTPGGCIFL**

Chemistry & Biology

Figure 9. Alignment of the SAM-dependent methyltransferase domains of McyJ, McyD, McyE, McyG and McyA. The consensus SAM-dependent methyltransferase motifs of Kagan and Clarke are indicated [50]. Residues matching the consensus motifs are shown in bold.

The subsequent microcystin NRPS biosynthesis pathway is illustrated in Figure 6. According to this model, Adda is condensed with the γ-carboxyl-activated glutamate. The D-glutamate residue is presumably supplied by epimerization of L-glutamate by McyF. This racemase shows similarities to a group of aspartate/glutamate racemases, but is markedly different from the alanine racemases previously identified in the NRPS systems forming cyclosporin or HC toxin [52]. The activation of the γ-carboxyl group of D-glutamate has no related example in the pool of adenylate forming domains identified so far [25]. Pocket analysis surprisingly maps the specificity of this region to glutamate and glutamine activating domains. This suggests that stereospecificity may not be selected at the adenylation step, but at later stages of the NRPS process.

The condensation of Adda and L-glutamate is presumably catalyzed by the first condensation domain of McyE, linking the PKS and the NRPS modules. This condensation domain shows divergent core motifs C3–C5, a feature also shared by a number of other condensation domains of the *mcy* cluster (Figure 5). These divergent condensation domains are involved in the condensation of either β- or γ-amino acids, and are expected to differ structurally from domains involved in the condensation of α-amino acids. Similarly, the condensation domains of ACV synthetases acting on the Aad δ-carboxyl group cluster together in the alignment of NRPS condensation domains [17]. The carboxy-terminal location of the second condensation do-

main of McyE is unusual. This arrangement implies the formation of a complex between McyE and McyA to accomplish condensation between the γ-carboxyl group of L-glutamate and the following *N*-methyl-serine intermediate. A second unusual feature of this step is the dehydration of the seryl side chain to dehydroalanine. Catalysis and timing of this step remain unclear. Similarly, the related dehydrothreonine formation in the syringomycin system is likewise not understood. It had been proposed that dehydrothreonine is activated and condensed directly, but this appears unlikely with respect to reactivity of the of the enamine structure [25]. While the uncharacterized McyI is tentatively proposed as catalyzing this step, it remains to be shown if McyI, the condensation domain itself, or an additional protein is involved.

Two activation and a single peptide condensation reactions are predicted to be performed by McyA, incorporating L-serine and L-alanine. Analysis of the substrate specificity of both adenylation domains identifies the first domain as serine-specific, while the predicted contact side chains of the second domain are identical with those of the saframycin Mx1 synthetase B glycine activating domain, and show some similarity to several cysteine activating domains [25]. This is not unexpected, as the similar size of the side chains on glycine (H) and alanine (CH_3) suggested that these binding sites may lack a defined pocket structure. Current data available on the predicted pocket residues are unable to resolve this question [25]. It is worth noting that no microcystin analogs with glycine at this position have been observed [1,2].

N-Methylation of the serine residue is catalyzed by the NMT module within the first domain of McyA. This methyltransferase shows high similarity to the NRPS NMTs of the virginiamycin, pristinamycin, and less similarity to the fungal domains involved in biosynthesis of enniatin, cyclosporin, and SDZ214-103. L-Alanine is converted to the D-form by the Ep domain contained in the carboxy-terminal region of McyA.

Two further rounds of peptide chain elongation are catalyzed by McyB, which activates and condenses L-leucine and D-MeAsp into the growing peptide chain. The McyA–McyB interaction represents the common transfer of D-peptidyl intermediates involving a carboxy-terminal Ep domain and an N-terminal condensation domain. The condensation domain maps well by alignment with domains involved in D–L peptide bonds [17]. The predicted pocket of the putative leucine-specific adenylation domain is identical to a large number of leucine activating domains. Analysis of the second adenylation domain, which presumably binds D-MeAsp, shows only weak similarities with a number of acidic amino acid specific activation domains [25].

The final transfer of the hexapeptidyl intermediate is unusual, with the β-carboxyl group being the target of the condensation reaction preceding cyclization. Interestingly, the condensation domain of McyC does not cluster with previously described D–L condensing domains [17], but rather with the non-α–α-condensing domains of microcystin synthetase. This cluster of non-α–α-condensing domains from microcystin synthetase forms a subcluster within the larger group of peptide synthetase α–α-condensing domains (alignment not shown) and the core motifs of the condensation domains (C1–C7) are fairly well conserved allowing their designation (Figure 5) [16,17]. Significant deviations from the core motifs are found within the domains condensing the *N*-methylated or epimerized residues.

On the basis of the microcystin-LR structure, the final adenylation domain is predicted to activate L-arginine. Binding pocket analysis revealed no similarity to any known arginine activating domain, but weak similarity to those activating hydrophobic amino acids. This position is known to be frequently occupied by hydrophobic residues in microcystin analogs, indicating a possible change in specificity by mutational transition. However, in this case the predictive model failed.

Final cyclization and release of the completed peptide chain in a condensation reaction is presumably performed by the TE domain encoded at the carboxy-terminus of McyC. A number of expected biosynthetic functions are missing from the *mcy* gene cluster. No discernible enzymatic functions are present within the cluster for the synthesis of phenylacetate, which is assumed to be synthesized from phenylalanine via phenylpyruvate in a manner similar to that identified in the non-microcystin producing cyanobacterium, *Synechocystis* [11]. Additionally, no pathway appears present for the biosynthesis of D-MeAsp. Based on labeled precursor studies Moore et al. [11] proposed a complex biosynthetic pathway involving condensation of acetyl-CoA and pyruvic acid to 2-hydroxy-3-methylsuccinic acid, oxidation to 2-oxo-3-methylsuccinic acid, and final transamination to D-MeAsp. Alternatively, the D-MeAsp could be the formed from glutamate by a methylaspartate mutase. Finally, no phosphopantetheine transferase is associated with the *mcy* gene cluster to catalyze the post-translational modification of the holo-NRPS thiolation and PK ACP domains [53].

The *mcy* gene cluster illustrates the remarkable modular nature of both the NRPS and PKS systems, with the various functional domains arranged in an almost 'mix-and-match' fashion to form the functional synthetase. The manner of integration of the PKS and NRPS via the AMT domain, together with the use of carboxy-terminal peptide bond condensation, opens new opportunities in the rational engineering of these systems to produce novel metabolites. Finally, the highly divergent nature of many

of the NRPS and PKS domains encoded by the *mcy* cluster reiterates how much we have to learn about these fascinating biosynthetic systems.

Significance

The microcystins are a family of polyketide/peptide-derived environmental toxins produced by diverse genera of cyanobacteria. Their complex biosynthetic pathway is one of many multienzyme-linked modular systems involving domains of PKS and NRPS. Cyanobacteria are a prominent source of such compounds, and will provide information on the functional organization of complex systems, their plasticity and evolution, and strategies for manipulation to generate new bio-products by combinatorial biology. At the same time possible clues are expected to emerge on the functional role of these metabolites in the ecosystem, and factors controlling their expression and thus promoting their environmental hazard.

We have analyzed the microcystin biosynthetic gene cluster of *M. aeruginosa* PCC7806, a producer of microcystin-LR, which is composed of the β-amino-polyketide moiety Adda, linked into the cyclic heptapeptide D-Glu-*N*-Me-dehydro-Ala-D-Ala-Leu-D-methyl-iso-Asp-Arg. This region spans 55 kb consisting of two oppositely transcribed gene clusters, encoding 10 proteins, six being multifunctional enzymes composed of PKS and NRPS domains. Two defined knock-out mutants were generated proving the involvement of this cluster in microcystin biosynthesis. Comparative sequence analysis assigns almost all of the 48 catalytic functions required for microcystin synthesis to this cluster and allows the identification of the primary precursors as phenylacetate, malonyl-CoA, SAM, glutamate, serine, alanine, leucine, D-methyl-iso-aspartate, and arginine. Among the reactions involved are peptide bonds between β- and γ-carboxyl groups, SAM-dependent *C*-, *N*- and *O*-methyl transfers, and dehydration of a seryl side chain. The modular architecture of microcystin synthetase provides new insights on the organization of these complex systems.

Materials and methods

Cloning and sequencing of the microcystin synthetase operon

Chromosomal DNA was isolated from *M. aeruginosa* PCC7806 as described [54]. Lambda Zap II library (Stratagene, La Jolla, CA) constructions and screenings were performed using the supplied protocol. PCR gene walking was performed using the hemidegenerate PCR technique to flanking peptide synthetase domains [22] and a modified version of the suppression PCR approach [23]. DNA sequencing was performed for both strands as described [55].

Insertional inactivation of mcyA and mcyD

The plasmid pMCYA5 was constructed by cloning a PCR-amplified 5 kb fragment of *mcyA* into the pGEM-T vector (Promega, Madison, WI, USA). The 1.4 kb *Bsa*AI fragment from pACYC184 containing the chloramphenicol resistance cassette was inserted into the *Bal*I site of pMCYA5. The plasmid pMCYD7 was constructed by cloning a PCR-amplified 7.6 kb fragment of *mcyD* into the pCR2.1-TOPO vector (Invitrogen, Carlsbad, CA, USA). The 0.8 kb *Hin*CII fragment from pUCBM20 containing the chloramphenicol resistance cassette was inserted into the *Xmn*I site of pMCYD7. The plasmids pMCYA5C and pMCYD7C were used to perform homologous recombinational inactivation of *mcyA* and *mcyD* in *M. aeruginosa* PCC7806 via natural transformation [12] and electroporation (14 kV/cm, 25 μF, 200 Ω), respectively.

Partial purification of microcystin synthetase and Western blotting

Partial purification of microcystin synthetase was performed essentially as described previously [12]. The Western blot was performed using standard procedures as described previously with the nitrocellulose membrane blocked with bovine serum albumin for 1 h at 37°C and the GS2 antibody as the primary antibody [44].

Microcystin analysis

Analysis of microcystin content was performed as described previously by the PP inhibition assay [56] and MALDI-TOF mass spectrometry [12].

Data deposition

The sequence reported in this paper has been deposited in GenBank under accession number AF183408.

Acknowledgements

We wish to thank R.A. Bass and M. Kaebernick for performing the PPi assays and B.R. Robertson for manuscript preparation. The authors are indebted to the reviewers for their critical analysis of this work. This research was supported by grants from the Deutsche Forschungsgemeinschaft to H.v.D. (Do270/8) and T.B. (Bo1045/13-3), Fonds der Chemischen Industrie to T.B., and the Australian Research Council and CRC for Water Quality to B.A.N.

References

1. 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.
2. Sivonen, K. (1996). Cyanobacterial toxins and toxin production. *Phycologia* **35**, S12–24.
3. Honkanen, R.E., et al., & Boynton, A.L. (1990). Characterization of microcystin-LR, a potent inhibitor of type 1 and type 2A protein phosphatases. *J. Biol. Chem.* **265**, 19401–19404.
4. Eriksson, J.E., Gronberg, L., Nygard, S., Slotte, J.P. & Meriluoto, J.A.O. (1990). Hepatocellular uptake of ^{3}H-dihydromicrocystin-LR, a cyclic peptide toxin. *Biochim. Biophys. Acta* **1025**, 60–66.
5. Beasley, V.R., Cook, W.O., Dahlem, A.M., Hooser, S.B., Lovell, R.A. & Valentine, W.M. (1989). Intoxication in livestock and water fowl. *Food Animal Pract.* **5**, 345–361.
6. Jochimsen, E.M., et al., & Jarvis, W.R. (1998). Liver failure and death after exposure to microcystins at a hemodialysis center in Brazil. *New Engl. J. Med.* **338**, 873–878.
7. Pouria, S.A., et al., & Codd, G.A. (1998). Fatal microcystin intoxication in haemodialysis unit in Caruaru, Brazil. *Lancet* **352**, 21–26.
8. Nishiwaki-Matsushima, R., et al., & Fujiki, H. (1992). Liver tumor promotion by the cyanobacterial cyclic peptide toxin microcystin-LR. *J. Cancer Res. Clin. Oncol.* **118**, 420–424.
9. Yu, S.Z. (1995). Primary prevention of hepatocellular carcinoma. *J. Gastroenterol. Hepatol.* **10**, 674–682.
10. Yu, S.Z. (1989). Drinking water and primary liver cancer. In *Primary Liver Cancer.* (Tang, Z.U., Wu, W.C.&Xia, S.S., eds.), pp. 30–37, Springer, Berlin.
11. Moore, R.E., Chen, J.L., Moore, B.S., Patterson, G.M.L. & Charmichael, W.W. (1991). Biosynthesis of microcystin-LR. Origin of the carbons in the Adda and Masp units. *J. Am. Chem. Soc.* **113**, 5083–5084.
12. Dittmann, E., Neilan, B.A., Erhard, M., von Döhren, H. & Börner, T. (1997). Insertional mutagenesis of a peptide synthetase gene that is responsible for hepatotoxin production in the cyanobacterium *Microcystis aeruginosa* PCC 7806. *Mol. Microbiol.* **26**, 779–787.

13. Arment, A.R. & Carmichael, W.W. (1996). Evidence that microcystin is a thio-template product. *J. Phycol.* **32**, 591–597.
14. Nishizawa, T., Asayama, M., Fujii, K., Harada, K. & Shirai, M. (1999). Genetic analysis of the peptide synthetase genes for a cyclic heptapeptide microcystin in *Microcystis* spp. *J. Biochem.* **126**, 520–529.
15. Cane, D.E., Walsh, C.T. & Khosla, C. (1998). Harnessing the biosynthetic code: Combinations, permutations and mutations. *Science* **282**, 63–68.
16. Marahiel, M.A., Stachelhaus, T. & Mootz, H.D. (1997). Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chem. Rev.* **97**, 26511–26573.
17. von Döhren, H., Dieckmann, R. & Pavela-Vrancic, M. (1999). The nonribosomal code. *Chem. Biol.* **6**, R273–R279.
18. Molnar, I., et al., & Ligon, J.M. (2000). The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from *Sorangium cellulosum* SO ce90. *Chem. Biol.* **7**, 97–109.
19. Paitan, Y., Alon, G., Orr, E., Ron, E.Z. & Rosenberg, E. (1999). The first gene in the biosynthesis of the polyketide antibiotic TA of *Myxococcus xanthus* codes for a unique PKS module coupled to a peptide synthetase. *J. Mol. Biol.* **286**, 465–474.
20. Gehring, A.M., et al., & Perry, R.D. (1998). Iron acquisition in plague: Modular logic in enzymatic biogenesis of yersiniabactin by *Yersinia pestis*. *Chem. Biol.* **5**, 573–586.
21. Duitman, E.H., et al., & Vater, J. (1999). The mycosubtilin synthetase of *Bacillus subtilis* ATCC6633: A multifunctional hybrid between a peptide synthetase, an amino transferase, and a fatty acid synthase. *Proc. Natl. Acad. Sci. USA* **96**, 13294–13299.
22. Neilan, B.A., et al., & Börner, T. (1999). Nonribosomal peptide synthesis and toxigenicity of cyanobacteria. *J. Bacteriol.* **181**, 4089–4097.
23. Siebert, P.D., Chenchik, A., Kellogg, D.E., Lukyanov, K.A. & Lukyanov, S.A. (1995). An improved PCR method for walking in uncloned genomic DNA. *Nucleic Acids Res.* **23**, 1087–1088.
24. Neilan, B.A., et al., & Goodman, A.E. (1997). rRNA sequences and evolutionary relationships among toxic and nontoxic cyanobacteria of the genus *Microcystis*. *Int. J. Syst. Bacteriol.* **47**, 693–697.
25. 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.
26. 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.
27. Cosmina, P., et al., & van Sinderen, D. (1993). Sequence and analysis of the genetic locus resposible for surfactin synthesis in *Bacillus subtilis*. *Mol. Microbiol.* **8**, 821–831.
28. Yakimov, M.M., Kroger, A., Slepak, T.N., Giuliano, L., Timmis, K.N. & Golyshin, P.N. (1998). A putative lichenysin A synthetase operon in *Bacillus licheniformis*: initial characterization. *Biochim. Biophys. Acta* **1399**, 141–153.
29. Gutierrez, S., Diez, B., Montenegro, E. & Martin, J.F. (1991). Characterization of the *Cephalosporium acremonium pcbAB* gene encoding alpha-aminoadipyl-cysteinyl-valine synthetase, a large multidomain peptide synthetase: linkage to the *pcbC* gene cluster of early cephalosporin biosynthetic genes and evidence of multiple functional domains. *J. Bacteriol.* **173**, 2354–2365.
30. Schwecke, T., et al., & Leadlay, P.F. (1995). The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. *Proc. Natl. Acad. Sci. USA.* **92**, 7839–7843.
31. Motamedi, H. & Shafiee, A. (1998). The biosynthetic gene cluster for the macrolactone ring of the immunosuppressant FK506. *Eur. J. Biochem.* **256**, 528–534.
32. Kaneko, T., et al., & Tabata, S. (1996). *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* **3**, 109–136.
33. Shi, L., Carmichael, W.W. & Miller, I. (1995). Immuno-gold localization of hepatotoxins in cyanobacterial cells. *Arch. Microbiol.* **163**, 7–15.
34. Smith, D.R., et al., & Reeve, J. (1997). Complete genome sequence of *Methanobacterium thermoautotrophicum* delta-H: functional analysis and comparative genomics. *J. Bacteriol.* **179**, 7135–7155.
35. Haydock, S.F., Dowson, J.A., Dhillon, N., Roberts, G.A., Cortes, J. & Leadlay, P.F. (1991). Cloning and sequence analysis of genes involved in erythromycin biosynthesis in *Saccharopolyspora erythraea*: sequence similarities between EryG and a family of S-adenosylmethionine-dependent methyltransferases. *Mol. Gen. Genet.* **230**, 120–128.
36. Weber, G., Schorgendorfer, K., Schneider-Scherzer, E. & Leitner, E. (1994). The peptide synthetase catalyzing cyclosporin production in *Tolypocladium niveum* is encoded by a giant 45.8-kilobase open reading frame. *Curr. Genet.* **26**, 120–125.
37. Haese, A., Schubert, M., Herrmann, M. & Zocher, R. (1993). Molecular characterization of the enniatin synthetase gene encoding a multifunctional enzyme catalysing N-methyldepsipeptide formation in *Fusarium scirpi*. *Mol. Microbiol.* **7**, 905–914.
38. de Crecy-Lagard, V., et al., & Blanc, V. (1997). Streptogramin B biosynthesis in *Streptomyces pristinaespiralis* and *Streptomyces virginiae* – Molecular characterization of the last structural peptide synthetase gene. *Antimicrob. Agents Chemother.* **41**, 1904–1909.
39. Rinehart, K.L., et al., & Carmichael, W.W. (1988). Nodularin, microcystin, and the configuration of Adda. *J. Am. Chem. Soc.* **110**, 8557–8558.
40. Dierstein, R., Kaiser, I., Weckesser, J., Matern, J., König, W.A. & Krebber, R. (1990). Two closely related peptide toxins in axenically grown *Microcystis aeruginosa* PCC 7806. *Syst. Appl. Microbiol.* **13**, 86–91.
41. Mootz, H.D. & Marahiel, M.A. (1997). The tyrocidine biosynthesis operon of *Bacillus brevis*: Complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *J. Bacteriol.* **179**, 6843–6850.
42. Tillett, D., Parker, D.L.&Neilan, B.A. (2000). Population structure of the genus *Microcystis*. *Appl. Environ. Microbiol.* (submitted).
43. Martin, C., Oberer, L., Ino, T., Konig, W.A., Busch, M. & Weckesser, J. (1993). Cyanopeptolins, new depsipeptides from the cyanobacteruim *Microcystis* sp. PCC 7806. *J. Antibiot.* **46**, 1550–1556.
44. Wessels, P., von Döhren, H. & Kleinkauf, H. (1996). Biosynthesis of acylpeptolactones of the daptomycin type: a comparitive analysis of peptide synthetases forming A21978C and A54145. *Eur. J. Biochem.* **242**, 665–673.
45. Yu, T.-W., et al., & Floss, H.G. (1999). Direct evidence that the rifamycin polyketide synthase assembles polyketide chains processively. *Proc. Natl. Acad. Sci. USA* **96**, 9051–9056.
46. Bolch, C.J.S., Blackburn, S.I., Jones, G.J., Orr, P.T. & Grewe, P.M. (1997). Plasmid content and distribution in the toxic cyanobacterial genus *Microcystis* Kuetzing ex Lemmermann (Cyanobacteria: Chroococcales). *Phycologia* **36**, 6–11.
47. Ruan, X., Stassi, D., Lax, S.A. & Katz, L. (1997). A second type-I PKS gene cluster isolated from *Streptomyces hygroscopicus* ATCC 29253, a rapamycin-producing strain. *Gene* **203**, 1–9.
48. Haydock, S.F., et al., & Leadlay, P.F. (1995). Divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA-acyl carrier protein transacylase domains in modular polyketide syntheses. *FEBS Lett.* **374**, 246–248.
49. Ikeda, H., Nonomiya, T., Usami, M., Ohta, T. & Omura, S. (1999). Organization of the biosynthetic gene cluster for the polyketide anthelmintic macrolide avermectin in *Streptomyces avermitilis*. *Proc. Natl. Acad. Sci. USA* **96**, 9509–9514.
50. Kagan, R.M. & Clarke, S. (1994). Widespread occurrence of three sequence motifs in diverse *S*-adenosylmethionine-dependent methyltransferases suggests a common structure for these enzymes. *Arch. Biochem. Biophys.* **310**, 417–427.
51. Schluckebier, G., O'Gara, M., Saenger, W. & Cheng, X. (1995). Universal catalytic domain structure of AdoMet-dependent methyltransferases. *J. Mol. Biol.* **247**, 16–20.
52. Cheng, Y.Q. & Walton, J.D. (2000). A eukaryotic alanine racemase gene involved in cyclic peptide biosynthesis. *J. Biol. Chem.* **275**, 4906–4911.
53. Lambalot, R.H., et al., & Walsh, C.T. (1996). A new enzyme superfamily – the phosphopantetheinyl transferases. *Chem. Biol.* **3**, 923–936.
54. Tillett, D. & Neilan, B.A. (2000). Xanthogenate nucleic acid isolation from cultured and environmental cyanobacteria. *J. Phycol.* **36**, 251–258.
55. Tillett, D. & Neilan, B.A. (1999). *n*-Butanol purification of Dye Terminator sequencing reactions. *BioTechniques* **26**, 606–610.
56. An, J. & Carmicheal, W.W. (1994). Use of a colorimetric protein phosphatase inhibition assay and enzyme linked immunosorbent assay for the study of microcystins and nodularins. *Toxicon* **32**, 1495–1507.