Halogenase Genes in Nonribosomal Peptide Synthetase Gene Clusters of *Microcystis* (Cyanobacteria): Sporadic Distribution and Evolution

Sabrina Cadel-Six, Catherine Dauga,† Anne Marie Castets,* Rosmarie Rippka,* Christiane Bouchier,‡ Nicole Tandeau de Marsac,* and Martin Welker*§*[1]

*Institut Pasteur, Unité des Cyanobactéries; Centre National de la Recherche Scientifique, Unité de Recherche Associée 2172, Paris, France; †Institut Pasteur, Genopole Ile de France—Plateforme 4, Paris, France; ‡Institut Pasteur, Genopole Ile de France—Plateforme 1, Paris, France; and §Technische Universität Berlin, Institut Chemie, AG Biochemie & Molekulare Biologie, Berlin, Germany

Cyanobacteria of the genus *Microcystis* are known to produce secondary metabolites of large structural diversity by nonribosomal peptide synthetase (NRPS) pathways. For a number of such compounds, halogenated congeners have been reported along with nonhalogenated ones. In the present study, chlorinated cyanopeptolin- and/or aeruginosin-type peptides were detected by mass spectrometry in 17 out of 28 axenic strains of *Microcystis*. In these strains, a halogenase gene was identified between 2 genes coding for NRPS modules in respective gene clusters, whereas it was consistently absent when the strains produced only nonchlorinated corresponding congeners. Nucleotide sequences were obtained for 12 complete halogenase genes and 14 intermodule regions of gene clusters lacking a halogenase gene or containing only fragments of it. When a halogenase gene was found absent, a specific, identical excision pattern was observed for both synthetase gene clusters in most strains. A phylogenetic analysis including other bacterial halogenases showed that the NRPS-related halogenases of *Microcystis* form a monophyletic group divided into 2 subgroups, corresponding to either the cyanopeptolin or the aeruginosin peptide synthetases. The distribution of these peptide synthetase gene clusters, among the tested *Microcystis* strains, was found in relative agreement with their phylogeny reconstructed from 16S–23S rDNA intergenic spacer sequences, whereas the distribution of the associated halogenase genes appears to be sporadic. The presented data suggest that in cyanobacteria these prevalent halogenase genes originated from an ancient horizontal gene transfer followed by duplication in the cyanobacterial lineage. We propose an evolutionary scenario implying repeated gene losses to explain the distribution of halogenase genes in 2 NRPS gene clusters that subsequently defines the seemingly erratic production of halogenated and nonhalogenated aeruginosins and cyanopeptolins among *Microcystis* strains.

Introduction

Cyanobacteria are a rich source of structurally diverse oligopeptides that are predominantly synthesized by nonribosomal peptide synthetases (NRPS). In *Microcystis*, a common genus in eutrophic freshwaters, numerous bioactive peptides have been identified that can be mostly classified as aeruginosins, microginins, microcystins, cyanopeptolins, and anabaenopeptins (Welker and von Döhren 2006). One intriguing property of these peptides is their structural diversity, achieved either by variability of amino acid moieties in particular positions or by modifications like glycosylation, sulfatation, methylation, or halogenation (Ishida et al. 1999, 2007; Rouhiainen et al. 2000; von Elert et al. 2005) giving rise to virtually hundreds of congeners in each peptide class.

The potential of *Microcystis* to produce a multitude of chlorinated peptides has been recognized in chemotyping studies of natural populations (Welker et al. 2006). Halogenated congeners have mainly been found among aeruginosin-, cyanopeptolin- (fig. 1), and microginin-type peptides, so far. It has to be noted that the names "aeruginosin" and "cyanopeptolin" refer to peptide types with unique features (see fig. 1) regardless whether individual peptide structures have been named differently. Cyanopeptolin-type peptides, for example, have been named cyanopeptolins, anabaenopeptilides, microcystilide, aeruginopeptins, or oscillapeptins in the original publications (for a review, see Welker and von Döhren 2006). Cyanopeptolin-type peptides have been reported from distant cyanobacterial taxa belonging to the Orders Chroococcales, Oscillatoriales, and Nostocales (Welker and von Döhren 2006). Most structural variants inhibit serine proteases (trypsin, chymotrypsin, or elastase) by preventing the hydrolytic attack on substrates by covering the active center with the rigid ring structure (Matern et al. 2003). Many structural variants of aeruginosins have been reported from *Microcystis* (Chroococcales) and *Planktothrix* (Oscillatoriales). Aeruginosins have also been shown to inhibit serine proteases, especially thrombin (Ishida et al. 1999). Despite the ubiquitous presence of these peptides (and others) in natural cyanobacterial consortia, a consistent hypothesis explaining their function in producing cells has not yet been proposed.

In microginins, linear tetra- or pentapeptides, chlorination occurs at an N-terminal aliphatic moiety (a modified decanoic acid) (Ishida et al. 1998), whereas in aeruginosins and cyanopeptolins, chlorination occurs at aromatic moieties. Expectedly, halogenating enzymes differ for these 2 types of chlorination. Aliphatic halogenation has been reported for the barbamide biosynthesis pathway of *Lyngbya majuscula* (Sitachitta et al. 2000), where a chlorine is transferred to a native leucine during biosynthesis by a nonheme FeII halogenase (Galonic et al. 2006). A similar enzyme has been reported for the syringomycin biosynthesis in *Pseudomonas syringae* (Vaillancourt et al. 2005).

Halogenation on aromatic moieties is thought to be catalyzed primarily by $FADH_2$-dependent halogenases (Vaillancourt et al. 2006; van Pée and Patallo 2006). In *Microcystis*, mono- or dichlorination of aeruginosins

[1] Present address: Anagnostec GmbH, Am Mühlenberg 11, Potsdam, Germany.

Key words: halogenase, cyanopeptolin, aeruginosin, DNA rearrangement, secondary peptide metabolite, chlorination, internal transcribed spacer, phylogeny.

E-mail: ntmarsac@pasteur.fr.

Mol. Biol. Evol. 25(9):2031–2041. 2008
doi:10.1093/molbev/msn150
Advance Access publication July 8, 2008

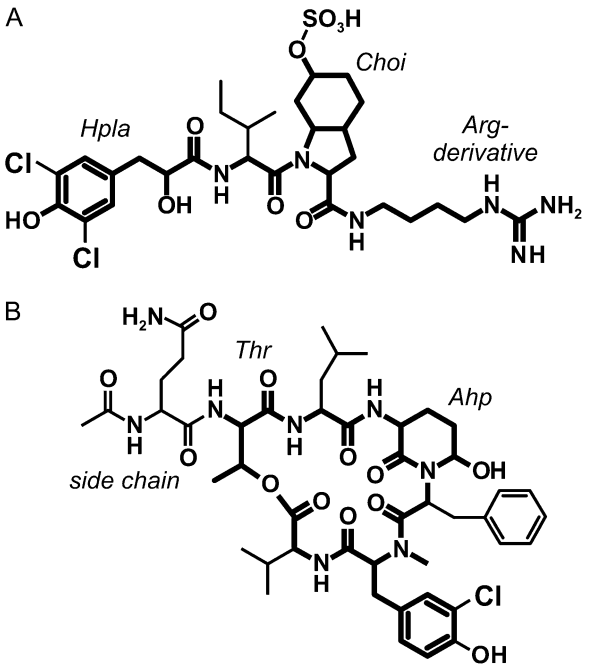

FIG. 1.—Flat structures of aeruginosin 101 (*A*; Ishida et al. 1999) and cyanopeptolin 954 (*B*; von Elert et al. 2005). Conserved parts of the structures of aeruginosins and cyanopeptolins, respectively, are indicated by bold lines. In other congeners, the chlorine atoms can be replaced by hydrogen atoms as it is the case for the sulfate group at Choi. Hpla: 4-hydroxyphenyl lactic acid; Choi: 2-carboxy-6-hydroxyoctahydroindole; Arg-derivative: agmatine (as shown), argininal, or argininol; Ahp: 3-amino-6-hydroxy-2-piperidone; and side chain: very variable consisting of amino acids, fatty acids, or (sulfated) glyceric acid (Welker and von Döhren 2006).

occurs at the Hpla moiety (3-(4-hydroxyphenyl) lactic acid) (Ishida et al. 1999) but not at the Choi moiety (2-carboxy-6-hydroxyoctahydroindole), where it has been found in *Planktothrix* (*Oscillatoria*) (Shin et al. 1997). Monochlorination in cyanopeptolins is reported for an N-methylated tyrosine or homotyrosine, respectively (Rouhiainen et al. 2000; von Elert et al. 2005).

A number of studies reported the presence and function of enzymes responsible for the halogenation of secondary metabolites of pharmaceutical interest, such as pentachloropseudilin (Wynands and van Pée 2004), clorobiocin (Eustáquio et al. 2003), pyrrolnitrin (Keller et al. 2000), or pyoluteorin (Dorrestein et al. 2005). Mechanistic studies indicated that chlorination is regioselective either on free substrates, like during pyrrolnitrin or rebeccamycin synthesis (Dong et al. 2005; Yeh et al. 2005), or while intermediates are tethered to NRPS enzymes like during clorobiocin or pyoluteorin synthesis (Eustáquio et al. 2003; Dorrestein et al. 2005). These studies suggest that specific halogenase genes are closely associated to particular NRPS gene clusters and that halogenase genes may be absent or dysfunctional in strains producing nonchlorinated peptide congeners that are very similar to chlorinated ones.

An initial mass spectrometrical screening of axenic *Microcystis* strains revealed the presence of aeruginosins and/or cyanopeptolins, both chlorinated and nonchlorinated. This raised the question whether the distribution of chlorinated congeners correlates with the presence of halogenase genes in corresponding NRPS operons that have been characterized for several cyanobacterial strains. Cyanopeptolin synthetase gene clusters have been described for *Microcystis* NIVA-CYA 172/5 (*mcn*; Tooming-Klunderud et al. 2007), *Anabaena* 90 (*apd*; Rouhiainen et al. 2000), and *Planktothrix* NIVA-CYA 116 (*oci*; Rounge et al. 2007), underlining the wide distribution of this type of nonribosomal peptides among cyanobacteria. The gene cluster consists of 3 (4 in *Microcystis*) genes coding for NRPS and, in *Planktothrix* and *Anabaena*, for a putative glyceric acid transferase domain and a putative formyl transferase domain (Rounge et al. 2007). Further, the gene clusters of *Anabaena* 90 and *Microcystis* NIVA-CYA 172/5 contain a gene for a halogenase (*apdC* and *mcnD*, respectively) that is absent in *Planktothrix* NIVA-CYA 116.

Aeruginosin (aeruginoside) synthetase gene clusters have been sequenced from *Planktothrix* NIVA-CYA 126-8 (*aer*; Ishida et al. 2007) and for 3 *Microcystis* strains, PCC 7806 (Frangeul et al. 2008), NIES-843 (Kaneko et al. 2007), and NIES-98 (Ishida K, Welker M, unpublished data). All known *aer* gene clusters possess 3 genes each coding for a complete NRPS module consisting of condensation, adenylation, and thiolation domains. In the second module, invariably an epimerization domain is found, consistent with the D-configuration of the second amino acid in most aeruginosins. In addition to the NRPS genes, particular gene clusters harbor genes for halogenases, glycosyltransferases, and sulfotransferases. Comparison of homologous gene clusters from these *Microcystis* strains revealed the genetic basis for the structural diversity of the final peptide products, with the presence or absence of halogenase genes being one of the most evident differences.

The present study reports the aeruginosin- and cyanopeptolin-type peptide production by 28 axenic *Microcystis* strains and describes the molecular organization of the corresponding NRPS gene clusters in the region where halogenase genes have been located. The phylogenetic history of the halogenase genes was reconstructed, as well as that of the individual strains, in order to explain the disparate distribution of halogenated peptides among strains of *Microcystis*.

Materials and Methods

Cyanobacterial Strains

Axenic strains of *Microcystis aeruginosa* (for details, see supplementary table 1, Supplementary Material online) were grown at 25 °C in 50 ml of $BG11_0$ supplemented with 2 mM $NaNO_3$ and 10 mM $NaHCO_3$ (Rippka and Herdman 1992). Continuous light was provided by Osram Universal White fluorescent tubes (30 $\mu E\, m^{-2}\, s^{-1}$). Cells in exponential growth phase ($OD_{750} = 0.5$) were harvested by centrifugation (10,000 × g, 10 min, 25 °C) and lyophilized.

Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry Analyses

Microcystis strains were analyzed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry directly from intact cells (Welker et al. 2004). Chlorinated peptides were evidenced by the intensity distribution of

isotopic peaks, that is, pseudomolecular ions at Δm/z 2 Da. Peaks with sufficient signal intensities ($>10^4$ counts) were further analyzed by postsource decay (PSD) fragmentation supported by collision-induced dissociation (Welker et al. 2006). Selected strains—generally those for which analysis of crude extracts did not yield unambiguous results—were further analyzed by high-performance liquid chromatography (HPLC) fractionation of lyophilized cells followed by off-line MS as described (Czarnecki et al. 2006).

Polymerase Chain Reaction Assays

DNA was extracted using the DNeasy Plant Mini Kit (QIAGEN, Courtaboeuf, France) with the following modifications: 5 mg of lyophilized cells were added to 450 μl of 50 mM Tris–HCl (pH = 8), 10 mM ethylenediaminetetraacetic acid, and 4 μl of RNase (100 mg ml^{-1}). The suspension was treated in Lysing Matrix A tubes (Q-BIOgene, Illkirch, France) for two 30 s runs at a speed setting of 4.0 in the FastPrep Instrument to disrupt the cells. The mixture was incubated for 10 min at 65 °C and processed further as described by the manufacturer. Two pairs of primers were designed based on the nucleotide sequence of the aeruginosin and cyanopeptolin synthetase gene clusters of strain PCC 7806, respectively. For the aeruginosin synthetase gene cluster, the forward primer aerA_F 5′-GAT AGC ACC CAG AAC GGA AGC-3′ is complementary to the 3′ end of *aerA*, and the reverse primer aerB_R 5′-CGT TAA ACG GAT GGT TAG AGC-3′ targets the 5′ end of *aerB*. For the cyanopeptolin synthetase cluster, the forward primer mcnC_F 5′-TAA GGA TAA TTT CTT TGA ATT GGG AG-3′ targets the 3′ end of *mcnC*, and the reverse primer mcnE_R 5′-GGG AAT AAT CTC TAA ATC AAC AGC-3′ targets the 5′ end of *mcnE*.

Polymerase chain reactions (PCRs) for amplifications of the *aerA–aerB* and *mcnC–mcnE* gene regions (100 μl) contained 10 μl *Taq* commercial buffer (10×), 2 mM $MgCl_2$, 50–100 ng of genomic DNA, 200 μM of each deoxynucleoside triphosphate, 1 μM of each primer, and 1U *Taq* polymerase (Promega Corporation, Charbonnières, France). Amplifications were performed with an initial denaturing step of 95 °C for 2 min followed by 40 amplification cycles (95 °C for 45 s, 50 °C for 45 s, and 72 °C for 1 min) and a final elongation step of 72 °C for 7 min in a Robocycler 40 Gradient temperature Cycler (Stratagene, Amsterdam, The Netherlands). Five microliters of each sample were analyzed by gel electrophoresis on 1% (w/v) agarose gels.

Amplicons of the aeruginosin and cyanopeptolin synthetase gene clusters were cloned into the pGEM-T vector (Promega Corporation) and the ligation mixture electroporated into JM 109 cells. Recombinants were selected on Luria Broth (LB) agar plates containing ampicillin (100 μg ml^{-1}) and isopropyl-beta-thio-galactoside (200 μmol). White colonies were picked and grown in LB-ampicillin (100 μg ml^{-1}) liquid medium overnight at 37 °C. Plasmid DNA was purified with the QIAprep Spin Miniprep Kit (QIAGEN). Purified plasmids were checked for the correct size of the inserts after *Eco*RI digestion followed by electrophoresis in 1% (w/v) agarose gels. The plasmid inserts were sequenced (Genome Express, Meylan, France), and potential coding sequences were translated and amino acid sequences compared with protein sequence databases by Blast search.

For PCR amplifications of the 16S–23S rDNA intergenic transcribed spacers (ITS), the forward primer 322 5′-TGT ACA CAC CGC CCG TC-3′, complementary to the 3′ end of the 16S rRNA gene, and the reverse primer 340 5′-CTC TGT GTG CCT AGG TAT CC-3′, complementary to the 5′ end of the 23S rRNA gene, were used (Iteman et al. 2000). PCRs (50 μl) were performed as described above. Amplification was performed as described (Iteman et al. 2000) or using the following procedure with an initial denaturing step of 95 °C for 2 min followed by 30 amplification cycles (95 °C for 30 s, 60 °C for 30 s, and 72 °C for 1 min) and a final elongation step of 7 min at 72 °C in a GeneAmp PCR System 9700 (Applied Biosystem, Paris, France). Five microliters of reaction were analyzed by electrophoresis on 0.6% (w/v) agarose gel.

The ITS PCR products obtained were purified with the QIAquick PCR Purification Kit (QIAGEN) and sequenced directly (Genome Express).

All nucleotide sequences obtained in this study are available in the GenBank–EMBL–DDBJ database under the following accession numbers: ITS (AM773517–AM773544), *aerA–aerB* region (AM773654–AM773664), and *mcnC–mcnE* region (AM773665–AM773679).

Phylogenetic Analyses

ITS sequences were aligned and analyzed as described (Garrigues et al. 2005). The phylogenetic reconstructions based on maximum likelihood (ML) analysis were performed with the HKY85 model. Distance analysis was carried out with the same evolutionary model and BIONJ algorithm (Gascuel 1997). For maximum parsimony analysis, the Tree Bisection-Reconnection heuristic algorithm was used for searching through tree space.

Halogenase amino acid sequences collected from databases (supplementary table 2, Supplementary Material online) were aligned with ClustalX, and the alignment was manually refined by repositioning highly conserved residues in the halogenase superfamily (Dong et al. 2005) using GeneDoc version 2.6.002 (Nicholas KB and Nicholas HBJ 1997).

Trees were computed using PhyML with different substitution models: Dayhoff (Dayhoff et al. 1978), JTT (Jones et al. 1992), WAG (Whelan and Goldman 2001), and DCMut (Guindon and Gascuel 2003; Kosiol and Goldman 2005). For the amino acid sequence analysis, the best phylogenetic tree was obtained with the WAG substitution model (WAG model: log-likelihood value of −36273.005548).

Based on the amino acid sequence alignment, a nucleotide alignment of the halogenase genes of *Microcystis*, *Anabaena*, and 2 *Xanthomonas* strains was obtained using Tranalign, a reimplementation in EMBOSS of the program mrtrans. Trees were built with the PAUP software applying the general time reversible model to ML and BIONJ analyses.

Statistical confidence levels for all topologies were evaluated by the nonparametric bootstrap method (100 replicates).

GC contents were calculated with the program REVSEQ (EMBOSS), and analysis of synonymous or

Table 1
Presence and Lengths (base pair) of the *aerA–aerB* and *mcnC–mcnE*-Amplicons in Comparison to the Corresponding Oligopeptides Identified in *Microcystis* Strains

Strain	PCR, Base Pair *aerA–aerB*	Peptide Aeruginosins	PCR, Base Pair *mcnC–mcnE*	Peptide Cyanopeptolins
PCC 7005	991[a]	Aeruginosin 688[b]	**2,298**[a]	**Cyanopeptolin 954**
PCC 7806	**2,855**[a]	**Aeruginosin 686**	585[a]	Cyanopeptolin A
PCC 7813	**2,855**[a]	**Aeruginosin 636**[c]	585[a]	Aeruginopeptin 228
PCC 7820	**2,855**	**Aeruginosin 636**[c]	585	Aeruginopeptin 228
PCC 7941	991	n.d.	**2,298**[a]	**Cyanopeptolin 1040A**
PCC 9354	991	Aeruginosin[d]	**2,298**	**Cyanopeptolin 1040A**
PCC 9355	991	n.d.	**2,298**	**Cyanopeptolin 1040A**
PCC 9432	991	Aeruginosin 574	585	Cyanopeptolin[d]
PCC 9443	991	Aeruginosin 652	n.p.	n.d.
PCC 9603	n.p.	n.d.	**2,829**[a]	**Cyanopeptolin 989**[b]
PCC 9622	n.p.	n.d.	585[a]	Cyanopeptolin S[b]
PCC 9624	991	Aeruginosin 652	n.p.	n.d.
PCC 9701	**2,855**	**Aeruginosin 642**	585	Microcystilide A
PCC 9804	**2,855**[a]	**Aeruginosin 642**	n.p.	n.d.
PCC 9805	**2,855**	**Aeruginosin 642**	n.p.	n.d.
PCC 9807	991[a]	Aeruginosin[d]	681[a]	Cyanopeptolin 1020
PCC 9808	991[a]	n.d.	**2,298**[a]	**Cyanopeptolin 1040C**
PCC 9809	991[a]	Aeruginosin[d]	585[a]	Aeruginopeptin 228
PCC 9810	991	Aeruginosin 652	n.p.	n.d.
PCC 9811	991	Aeruginosin[d]	n.p.	n.d.
PCC 9812	**2,855**[a]	**Aeruginosin 608**	**2,298**[a]	Cyanopeptolin 1020
PCC 9905	991	Aeruginosin 688[b]	**2,298**[a]	**Cyanopeptolin 954**
PCC 10025	991	n.d.	585[a]	Cyanopeptolin[d]
NIES-89	**2,855**	**Aeruginosin 89**[b]	585[a]	n.d.
NIES-98	**2,855**[a]	**Aeruginosin 98-A**[b]	n.p.	n.d.
NIES-101	**2,855**	**Aeruginosin 101**[b]	n.p.	n.d.
NIES-102	991[a]	Aeruginosin 102[b]	585[a]	Cyanopeptolin[d]
NIES-298	**2,855**[a]	Aeruginosin 298-A	585[a]	n.d.

NOTE.—The presence of *aerA–aerB* is indicative of an aeruginosin NRPS gene cluster, whereas amplicons of *mcnC–mcnE* indicate a cyanopeptolin NRPS gene cluster. Sequences shown to contain a halogenase gene or deduced to contain it based on amplicon length and oligopeptides that are chlorinated are indicated in bold type. When amplicons were not sequenced, their length was deduced from sequences (see also fig. 2). If multiple congeners of one peptide type were detected, generally only the most abundant one (by peak intensity in mass spectra) is listed. n.d., no respective peptide detected; n.p., no PCR product. For more details on peptide structures, see supplementary table 3 (Supplementary Material online).

[a] Nucleotide sequence determined.

[b] Sulfated congener detected as M-SO_3+H^+ in positive ion extraction mode.

[c] Minor compound compared with the nonchlorinated congener.

[d] Only partially elucidated by the detection of characteristic fragments in PSD spectra.

nonsynonymous substitution rates was performed with SWAAP 1.0.1 (Pride 2004).

Results

Detection of Peptides and PCR Results

Mass spectral analysis of whole cells or HPLC fractions revealed the presence of various oligopeptides in all *Microcystis* strains. Aeruginosin- or cyanopeptolin-type peptides could be identified in 22 and 18 strains, respectively, whereas both peptide types were detected in 12 strains (table 1). Most congeners have been described previously or were very similar to previously described ones, for example, by addition of a sulfate group. In PCC 7005 and 9905, for example, a peptide (aeruginosin 688) was detected with a molecular mass and fragmentation pattern corresponding to microcin SF608 (Hpla-Phe-Choi-Agmatine; Banker and Carmeli 1999) that was, however, found to be sulfated, as evidenced by a mass difference of 80 Da when the samples were analyzed in negative ion extraction mode. Another aeruginosin, aeruginosin 686 (*Cl*Hpla-Tyr-Choi-Argininal) in PCC 7806, corresponds to a chlorinated aeruginosin 102 (*Su*Hpla-Tyr-Choi-Argininal) lacking a sulfate group. A corresponding nonchlorinated variant is aeruginosin 652 (Hpla-Tyr-Choi-Argininal), produced by 3 of the analyzed strains. Among cyanopeptolin-type peptides, 2 chlorinated congeners were identified with nearly identical mass but with different amino acid sequences, cyanopeptolins 1040A ([Arg-Ahp-Phe-*Cl*,MTyr-Val-O-Thr]-Asp-hexanoic acid; $M + H^+ = 1041.48$ Da) and 1040C ([Hty-Ahp-Leu-*Cl*,MTyr-Val-O-Thr]-Gln-hexanoic acid; $M + H^+ = 1041.51$ Da). The latter was identified in strain PCC 9808 together with a nonchlorinated congener (cyanopeptolin 1040B [Hty-Ahp-Phe-MTyr-Val-O-Thr]-Gln-hexanoic acid) with $M + H^+ = 1041.53$ Da in which a leucine is replaced by a phenylalanine, compensating the mass difference attributed to the lack of a chlorine atom ($\Delta m = 34$ Da). In cyanopeptolins, chlorination was found only in congeners with a tyrosine in position 5 of the ring (fig. 1*B*) and never when this position was occupied by a phenylalanine. For a number of strains, the aeruginosin- or cyanopeptolin-type peptides could not be fully characterized and only a classification could be achieved. For aeruginosins, a typical fragment in PSD spectra is the Choi-immonium ion (m/z 140), whereas for cyanopeptolins, no singular characteristic fragment can be identified but series

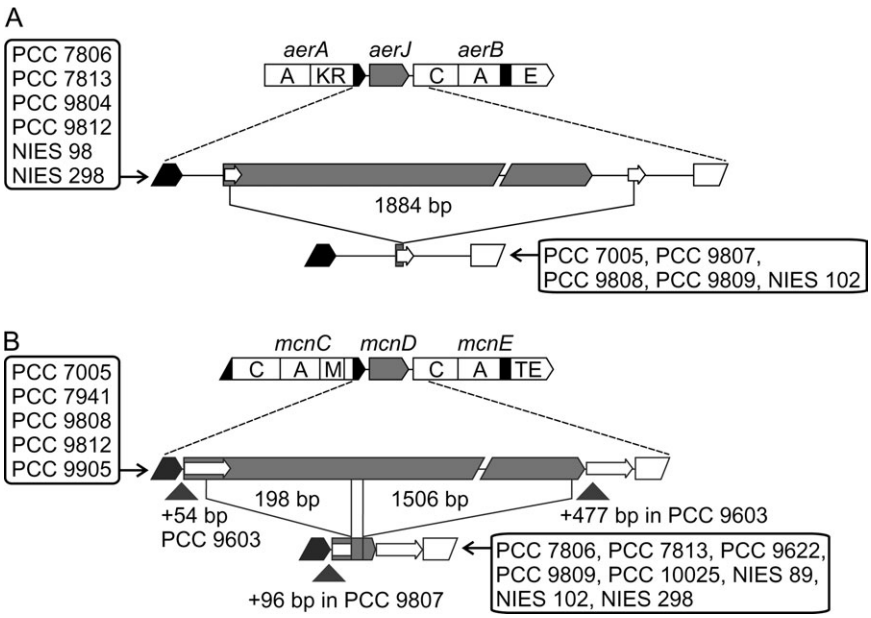

FIG. 2.—Schematic illustration of the halogenase genes (*aerJ* and *mcnD*) in segments of the aeruginosin (*aerA–aerB*; *A*) and cyanopeptolin (*mcnC–mcnE*; *B*) NRPS gene clusters of *Microcystis* strains. Full lines delimit deletions and triangles indicate insertions. Arrows indicate DRs as described in the text. Strains grouped by their numbers have the same gene arrangement in the respective regions; for further details, see text.

of fragments related to 3-amino-6-hydroxy-2-piperidone (Ahp) (Welker et al. 2006).

In agreement with previous findings (Czarnecki et al. 2006; Welker et al. 2006), several congeners of a particular peptide type could be identified in most strains, and only the most abundant ones, as assumed from peak height in mass spectra, are given in table 1 (with few exceptions as indicated). In the case of aeruginosins, congeners with a varying degree of chlorination have been detected in individual strains, for example, aeruginosins 98-B (Hpla-Ile-*Su*Choi-Agmatine), 98-A, and 101 in *Microcystis* NIES-98, which are non-, mono-, and dichlorinated at the Hpla moiety, respectively. Chlorination was never observed for both peptide types when these were coproduced by an individual strain.

PCR assays with primers specific for the *aerA–aerB* region gave positive results for all strains, except for strains PCC 9603 and PCC 9622, and single amplicons of either 2,855 or 991 bp were obtained (table 1). Sequences were not obtained for all amplicons, and hence, the length of non-sequenced amplicons could only be estimated. However, the high conservation of both sequence types, long and short amplicons, respectively (fig. 2), suggests that the length of non-sequenced amplicons is very close if not identical to sequenced ones. This also applies to the *mcn*-amplicons.

PCR assays with primers specific for the *mcnC–mcnE* region gave positive results for 20 strains. When positive, single amplicons of either 2,298 or 585 bp were obtained, except for strains PCC 9603 (2,829 bp) and PCC 9807 (681 bp) (table 1).

Positive PCR results were in agreement with the detection of corresponding aeruginosins and cyanopeptolins in most strains. However, no cyanopeptolins were detected in 2 strains for which the presence of a PCR amplicon indicates the existence of an *mcn* gene cluster (NIES-89 and NIES-298). Similarly, in 4 other strains (PCC 7941, PCC 9355, PCC 9808, and PCC 10025), *aer*-amplicons were detected but no corresponding aeruginosin-type peptides. The production of chlorinated congeners was clearly related to long amplicons for both peptide types with few exceptions, for example, in strain NIES-298 that produces only the non-chlorinated aeruginosins 298 A and B in detectable amounts (table 1).

Arrangement of Halogenase Genes in the *aer* and *mcn* Synthetase Gene Clusters

Nucleotide sequence analyses of the large amplicons (*aerA–aerB*: 2,855 bp; *mcnC–mcnE*: 2,298 bp) by Blast revealed in both cases an open reading frame (ORF) of 1878 bp (except for strain PCC 9812 whose *aerA–B* amplicon contains an ORF of 1,881 bp) named *aerJ* and *mcnD* (fig. 2*A*). The corresponding putative proteins (AerJ and McnD) did not show similarity to peptide synthetases by Blast but to various prokaryotic halogenases, including ApdC and McnD of *Anabaena* 90 and *Microcystis* NIVA-CYA 172/5, respectively. Detailed analysis of their deduced amino acid sequences showed the presence of motifs GxGxxG and WxWxIP, 2 highly conserved motifs in $FADH_2$-dependent halogenases (van Pée and Zehner 2003) (supplementary figure, Supplementary Material online).

In both halogenase gene regions, direct repeats (DRs) were identified. In the long *aerA–B* sequences (2,855 bp), 2 short DRs (GTTGA; $GTTG^{A}/_{C}$ in strain PCC 9812) were located 6 bp downstream of the putative start codon and 15 bp (9 bp in strain PCC 9812) downstream of the stop codon of *aerJ*. An identical deletion of 1,884 bp encompassing the totality of *aerJ* and leaving only one copy of the DR was observed in the corresponding sequences of strains that yielded short amplicons (991 bp) (fig. 2*A*). Intergenic regions upstream and downstream of *aerJ* shared by strains with and without *aerJ* showed more than 90% sequence identities.

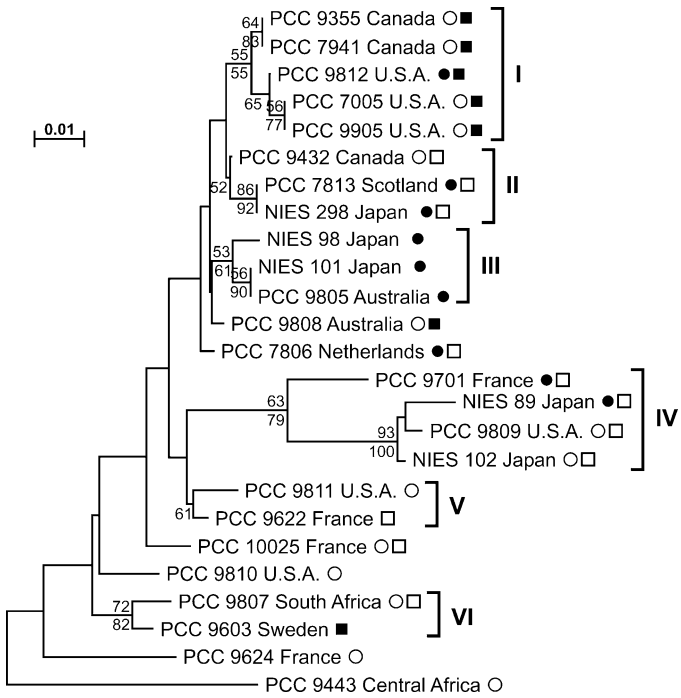

FIG. 3.—Distance tree of the ITS nucleotide sequences of *Microcystis* strains and distribution of the aeruginosin (*aer*, circles) and cyanopeptolin (*mcn*, squares) synthetase gene clusters, with their corresponding halogenase genes (*aerJ* and *mcnD*). Full circles: complete *aerJ*; empty circles: lack of *aerJ*; full squares: complete *mcnD*; and empty squares: truncated *mcnD*. ML (above) and distance (below) bootstrap values exceeding 50% are given at the nodes. The geographic origin of the strains is indicated.

Two long imperfect DRs (107 bp) were identified in the 2,298 bp *mcnC–mcnE*-sequences (fig. 2*B*). These include the last 2 nt of the putative start codon and the stop codon of *mcnD*. In strains that yielded short *mcnC–mcnE*-amplicons (585 bp), 2 deletions (198 and 1,506 bp) with exactly the same break points were identified, leaving fragments of *mcnD* with one truncated and one complete DR sequence (fig. 2*B*). These truncated *mcnD* sequences share more than 84% sequence identities with the corresponding segments in strains carrying a complete *mcnD* (*mcnD*$^{+}$).

The 2,829 bp *mcnC–mcnE*-sequence in strain PCC 9603 has the same imperfect DRs and a complete *mcnD* of 1,878 bp but carries 2 insertions, 1 of 54 bp and 1 of 477 bp upstream and downstream of *mcnD*, respectively (fig. 2*B*). In strain PCC 9807, the *mcnC–mcnE*-sequence (without a complete *mcnD*) had a size of 681 bp (compared with 585 bp in other strains) due to an insertion of 96 bp located 10 bp downstream of the *mcnC* stop codon. This insertion showed 96% sequence identities with the *mcyE–mcyG* intergenic region of the microcystin synthetase operon of *Planktothrix rubescens* (strain WAHN) (Mbedi et al. 2005).

Only in strain PCC 9812, the halogenase genes are present in both the *aer* and *mcn* NRPS gene clusters.

Distribution of the *aerJ* $^{+/-}$ and *mcnD* $^{+/-}$ Synthetase Gene Clusters and ITS Phylogeny

In phylogenetic trees based on full-length ITS sequences, the 28 *Microcystis* strains formed 6 distinct monophyletic groups (I–VI) that were supported by bootstrap values above 50%, irrespective of the 3 different methods employed for the analyses (fig. 3). These ITS groups did not reflect the geographic origin of strains, except ITS group I that includes 5 strains isolated in North America. In addition, 6 of the *Microcystis* strains (PCC 7806, PCC 9624, PCC 9443, PCC 9810, PCC 9808, and PCC 10025) seem to represent polyphyletic lineages, distinct from one another and more or less distant to the monophyletic ITS groups I–VI identified here.

Based on PCR results (table 1), the presence of both, the *aer* and *mcn* NRPS gene clusters, was frequently observed in individual strains (in 18 of the 28 strains) and consistently in all strains of ITS groups I, II, and IV. Strains PCC 9354 and PCC 7820 are, respectively, duplicates of strains PCC 7941 (ITS group I) and PCC 7813 (ITS group II) and thus are not included in figure 3 (see supplementary table 1, Supplementary Material online). Co-occurrence of the 2 NRPS gene clusters, however, is not restricted to these 3 ITS groups as it is also found in 1 strain of ITS group VI (PCC 9807) and in 3 of the polyphyletic strains (PCC 7806, PCC 9808, and PCC 10025).

Eight strains were positive only for the *aer* gene cluster (table 1), 4 of ITS group III (PCC 9805 is coidentic to PCC 9804; supplementary table 1, Supplementary Material online), 1 of ITS group V (PCC 9811), and the polyphyletic strains PCC 9810, PCC 9624, and PCC 9443.

Two strains were positive only for the *mcn* gene cluster in PCR assays, 1 of ITS group V (PCC 9622) and 1 of ITS group VI (PCC 9603).

The halogenase genes, *aerJ* or *mcnD*, appear variably present within corresponding NRPS gene clusters in strains belonging to the same ITS groups (fig. 3). The consistent

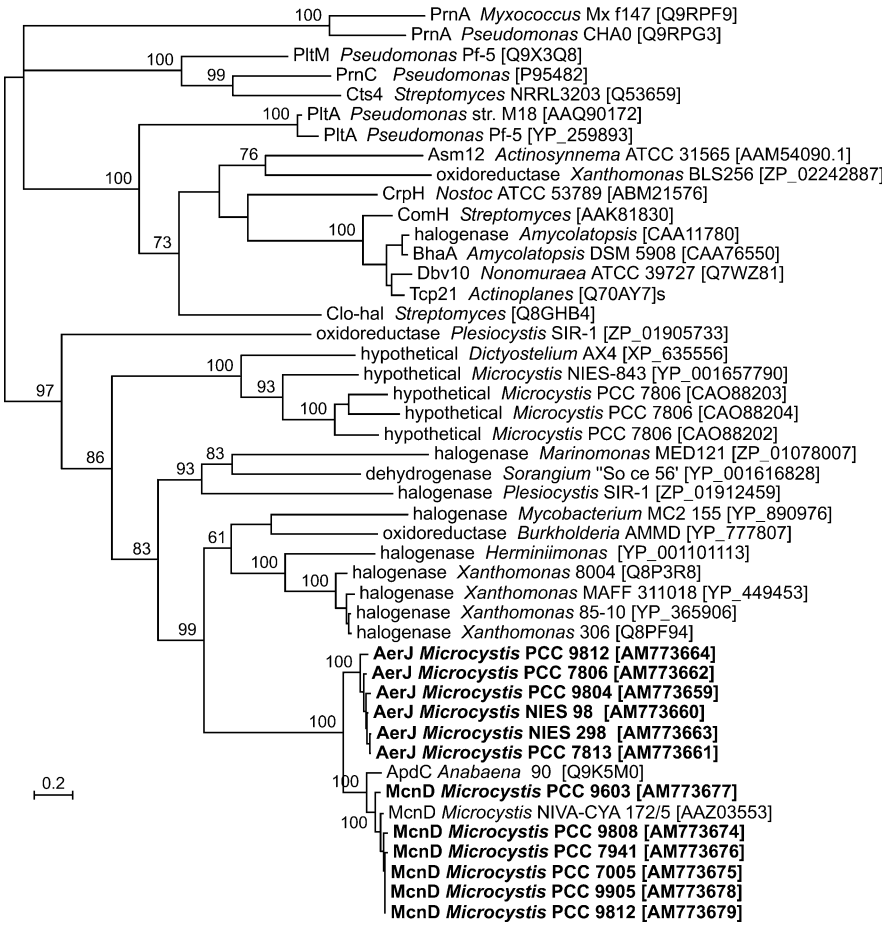

FIG. 4.—PHYML tree of the halogenase amino acid sequences (for further details, see supplementary table 2, Supplementary Material online) based on the WAG substitution model. Bootstrap values exceeding 50% are given at the nodes. Sequence names in bold are obtained in the present study.

presence of *aerJ* was observed only for the strains of ITS group III. A complete *mcnD* is also observed in all strains of ITS group I, whereas *aerJ* is only present in strain PCC 9812 within this group. In ITS groups II and IV, *aerJ* is present in some of the strains, none of which has a complete *mcnD*. In the remaining strains, including ITS groups V and VI, *aerJ* and *mcnD* are absent in all strains with the exception of strain PCC 9603 (ITS group VI). The presence of at least one complete halogenase gene is more frequently observed in strains of ITS groups I, II, and III than among all other strains.

Phylogenetic Relationships of Halogenases

For a broader view on the distribution of halogenases among bacteria, we collected sequences from accessible databases based on a Blast search or on the literature. This revealed a number of homologues of the *Microcystis* halogenases in unrelated prokaryotic taxa (see supplementary table 2, Supplementary Material online), also including a sequence of the slime mold *Dictyostelium discoideum* (Protista). Noteworthy, for a majority of the sequences, the information on function is only derived in silico and refers to putative enzymes.

In the phylogenetic tree obtained, the NRPS-related halogenases of *Microcystis* (AerJ and McnD) and *Anabaena* strain 90 (ApdC) form a strongly supported monophyletic group rooted by putative halogenases from Proteobacteria representative of the γ-subdivision (4 *Xanthomonas* species and *Marinomonas* sp. MED121) and β-subdivision (*Burkholderia ambifaria* AMMD and *Herminiimonas arsenicoxydans*) and one halogenase from Actinobacteria (*Mycobacterium smegmatis* MC2) (fig. 4 and supplementary table 2, Supplementary Material online). Three other, far more distant, putative halogenase sequences (CAO88202, CAO88203, and CAO88204) were identified in the genome of *Microcystis* PCC 7806 in a gene cluster encoding a polyketide synthase. These sequences group with a putative halogenase identified in the genome of *Microcystis* NIES-843 and a halogenase of the protist *D. discoideum* AX4. A halogenase of the filamentous cyanobacterium *Nostoc* sp. ATCC 53789 is closely related to a group of halogenases involved in glycopeptide synthesis in Actinobacteria (*Streptomyces lavendulae*, *Amycolatopsis orientalis*, *Amycolatopsis balhimycina* DSM 5908, *Actinoplanes teichomycetus*, and *Nonomuraea* sp. ATCC 39727). The relationship of cyanobacterial halogenases to those in diverse prokaryotic lineages (including a protist) indicates that a horizontal gene

transfer (HGT) may be involved in their present taxonomic distribution. Furthermore, the average genomic GC content of *Microcystis* is 41.9 ± 0.4% (n = 9 strains; Rippka and Herdman 1992), whereas that of *aerJ* and *mcnD* halogenase genes is slightly lower (37.7 ± 0.5%) and displays higher GC deviations in the third position of codons (GC3: 26.8 ± 0.04%).

The NRPS-related halogenases of the *Microcystis* strains are separated into 2 well-defined subgroups, one including McnD/ApdC sequences and the other AerJ sequences. This subdivision may result of a convergent evolution due to the functional specialization of each gene to putatively chlorinate similar substrates or reflect the presence of ancient paralogues that, in the course of evolution, diverged in the respective NRPS clusters. Pairwise comparison of the synonymous (K_S) and nonsynonymous (K_A) nucleotide substitutions between the *aerJ* and *mcnD* (*apdC*) data sets showed different rates of mutations for each subgroup, with a higher degree of divergence in the *mcnD* subgroup (data not shown). However, no evidence of accelerated evolution (positive selection pressure) promoting amino acid divergence was found along the gene. The K_A/K_S ratios below 1 obtained for both subgroups indicate that the genes were subject to purifying selection. These results support an ancient origin of the halogenase genes, followed by their progressive divergence in the *aer* and *mcn* NRPS gene clusters.

Discussion

For most of the *Microcystis* strains examined, the detection of the NRPS gene clusters and halogenase genes by PCR assays is consistent with the detection of corresponding peptides by chemical analyses, as well as with the chlorination of particular congeners. The apparent lack of "expected" peptides in some of the strains may be the result of cellular concentration levels below the limit of detection, dysfunctional halogenase genes, or impairment of other regions of the NRPS clusters such as that reported for the microcystin synthetase gene cluster in *P. rubescens* (Christiansen et al. 2006).

In the present study, as well as in more than 1,000 single colonies and isolate samples of *Microcystis*, no chlorinated congeners of both, aeruginosins and cyanopeptolins, could be detected (Welker et al. 2004, 2006). In agreement, a chlorinated aeruginosin, but no chlorinated cyanopeptolin, was detected in strain PCC 9812, in spite of the fact that a halogenase gene is present in both of the 2 corresponding NRPS gene clusters. A reason for the apparent mutual exclusion of chlorinating enzymes or chlorinating activity, respectively, associated with the 2 peptide synthetases remains, however, to be elucidated.

The chlorinating activity of AerJ and McnD apparently can result in the production of congeners with varying degrees of chlorination, both in axenic strains (see above) and in field samples (Welker et al. 2006). In *Microcystis* NIES-98, for example, non-, mono-, and dichlorinated aeruginosins are detectable as it is correspondingly the case with cyanopeptolins in PCC 9808. Both the aeruginosin and cyanopeptolin halogenases evidently are functional in general in individual strains although apparently not absolutely necessary for the peptides' biosynthesis in the NRPS assembly line, that is, a lacking chlorination does not interrupt the further biosynthetical steps. This was also found for the clorobiocin synthesis in *Streptomyces*, for example, where the disruption of the halogenase gene (*clo-hal*) resulted in the production of a nonhalogenated congener at higher cellular concentrations than compared with clorobiocin in the wild type (Eustáquio et al. 2003).

Chlorination of a small peptide has been shown to modify the properties of the molecule, that is, by electronic and steric effects, with consequences on its biological activity (Harris et al. 1985; Eustáquio et al. 2003). In cyanobacteria, the function of chlorinated or nonchlorinated oligopeptides remains, however, to be elucidated. The strong potential of cyanopeptolins, for example, to inhibit serine proteases is not influenced by chlorination per se as evidenced by X-ray studies of a nonchlorinated and a chlorinated congener cocrystallized with proteases (Lee et al. 1994; Matern et al. 2003). Individual congeners, on the other hand, have been shown to differently inhibit protease activity of a potential grazer, the common freshwater herbivore *Daphnia* (von Elert et al. 2005; Czarnecki et al. 2006). The multitude of bioactive peptides in a population could increase the efficiency as grazing protection compared with a few peptides by hampering the physiological adaption of a grazer to "inhibitory" diet through the expression of specific proteolytic enzymes.

However, the high structural diversity provides evidence that corresponding NRPS gene clusters have an intrinsic plasticity leading to the production of various analogous peptide structures in natural mixtures of clones or chemotypes. Structural diversity of natural products synthesized by modular enzyme complexes is largely based on recombination of nucleotide sequences coding for individual domains, like in fungal polyketide synthase systems (Jenke-Kodama et al. 2005). Accordingly, recombination between cyanobacterial NRPS gene modules, involving exchanges of substrate-specific domains, has been shown for the microcystin synthetase genes (Mikalsen et al. 2003; Tanabe et al. 2004).

In the phylogenetic tree inferred from halogenase amino acid sequences, some unexpected groupings were found, encompassing "alien sequences" as judged by the phylogeny of the respective organisms from which the sequences were obtained. This is clearly the case for the halogenase of the protist *D. discoideum*, which groups with 3 cyanobacterial halogenases, and thus, the corresponding genes may have been acquired by HGT from a cyanobacterium. A similar conclusion was made by phylogenomic analysis of protistan polyketide synthases (John et al. 2008).

Similarly to the halogenase of *M. smegmatis*, the *Microcystis* halogenases could have been acquired from an ancestor of Proteobacteria. Sequences acquired through HGT generally tend to accumulate mutations that progressively lead to the acquisition of sequence characteristics corresponding to their new genetic environment. The *Microcystis aerJ* and *mcnD* halogenase genes display GC contents only slightly lower than that of total genomic DNAs. Therefore, it seems likely that their presence in modern *Microcystis* strains is the result of an ancient HGT.

Halogenase genes in the *mcn* and *apd* clusters in *Microcystis* (Chroococcales) and *Anabaena* (Nostocales), and chlorinated cyanopeptolins in other phylogenetically distant cyanobacterial taxa, support the view that *mcn*/*apd* NRPS gene clusters containing a halogenase gene occurred early in the cyanobacterial lineage. Chlorinated aeruginosins, on the other hand, have been reported for *Planktothrix* strains, but the chlorination occurs on another residue than in *Microcystis* (at the Choi moiety; Shin et al. 1997), and the *aer* gene cluster of *Planktothrix* NIVA-CYA 126-8 does not contain *aerJ* (Ishida et al. 2007). As long as no further data are available, it can be assumed that the possession of *aerJ* is a characteristic of *Microcystis* strains.

Considering the topology of the halogenase phylogenetic trees and the relatively high sequence identity between *aerJ* and *mcnD*, an ancient gene duplication of *mcnD* and its integration into the *aer* cluster before their diversification may have occurred in an ancestor leading to modern *Microcystis*. An independent acquisition of halogenase genes by HGT cannot be excluded but seems less likely because *aerJ* and *mcnD* are rarely present together in an individual strain.

For the evolutionary scenario, we hence propose that an ancestral *Microcystis* had both, the *aer* and *mcn* gene clusters, each containing a halogenase gene that was subsequently lost in some of the phylogenetic lineages. The presence of *mcnD* fragments between *mcnC* and *mcnE* in strains lacking a complete *mcnD* gene clearly indicates that the presence of a halogenase gene in the *mcn* gene cluster represents the ancestral state. The loss of *mcnD* might then have occurred either repeatedly with identical break points in a number of strains or as a singular loss event inherited to *Microcystis* lineages. Because the distribution of *mcnD* in the *mcn* cluster is consistent for all the monophyletic ITS groups (except ITS group VI), a singular loss event appears to be more likely.

On the other hand, the genomic region around *mcnD* is bordered by long imperfect DRs that possibly present a region of high susceptibility to specific DNA rearrangements. This might be the reason for the peculiar sequences in both strains of ITS group VI (PCC 9603 and PCC 9807) that respectively contain unique types of insertions in the *mcnC–mcnE* region, suggesting that in these strains particular mechanisms modified the gene arrangement possibly more recently than in other lineages (figs. 2 and 3).

In the *aer* gene clusters, no representatives carrying fragments of halogenase genes were found, but the strictly identical deletions observed and the DRs presumably representing break points suggest that this lack may also be due to gene loss. If so, the lack of *aerJ* in *Microcystis* strains could be the result of repeated losses as suggested by the inconsistent distribution of *aerJ* among strains of the ITS groups I, II, and IV. Indeed, the involvement of short DRs, of which only one is left in *aerJ*$^{-}$ strains (fig. 2*A*), has frequently been observed in gene deletions and is explained by a slipped-strand mispairing mechanism (Levinson and Gutman 1987).

The mechanism leading to the loss of one or the other complete NRPS gene clusters (*aer* or *mcn*) in some *Microcystis* lineages (fig. 3) remains to be explored. A similar sporadic distribution of the microcystin synthetase gene cluster, not concordant with phylogenies based on housekeeping genes, was proposed to be the result of repeated gene losses by a yet unknown mechanism, whereas horizontal transfer of entire gene clusters was considered unlikely (Rantala et al. 2004).

The present study highlights the genetic basis for natural product structural diversity exemplarily for NRPS-associated halogenase genes and chlorination of peptide products. Mechanisms such as HGT, recombination, and gene losses apparently result in a high plasticity of the architecture and functionality of strain-specific NRPS gene clusters, leading to a vast structural diversity of corresponding peptide products in closely related cyanobacteria. How the evolution of and rearrangements in cyanobacterial NRPS gene clusters actually come to pass is poorly understood yet, especially with regard to the lack of evidence for HGT in the genus *Microcystis*. An improved understanding of cyanobacterial combinatorial biosynthesis in vivo could be the base for the design and production of new pharmaceuticals in the future.

Supplementary Material

Supplementary figure and tables 1–3 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/). Sequences deposited in the GenBank–EMBL–DDBJ database: ITS (AM773517–AM773544), *aerA–aerB* region (AM773654–AM773664), *mcnC–mcnE* region (AM773665–AM773679).

Acknowledgments

We are very grateful to A. Boussier and S. Gribaldo for very valuable discussions. We acknowledge L. Via-Ordorika for making available unpublished strains, D. Parker for depositing the UWOCC strains in the PCC, T. Coursin for maintenance of the stock cultures, and T. Laurent for providing cell material. This study was supported by the Institut Pasteur, the Centre National de la Recherche Scientifique (URA 2172), the Ministère de l'Education Nationale, de la Recherche et de la Techologie, and the EU-project "PEPCY"—Toxic and Bioactive PEPtides in CYanobacteria (European commission research grant QLRT-2001-02634). Useful comments by 2 anonymous reviewers were appreciated.

Literature Cited

Banker R, Carmeli S. 1999. Inhibitors of serine proteases from a waterbloom of the cyanobacterium *Microcystis* sp. Tetrahedron. 55:10835–10844.

Christiansen G, Kurmayer R, Liu Q, Börner T. 2006. Transposons inactivate biosynthesis of the nonribosomal peptide microcystin in naturally occurring *Planktothrix* spp. Appl Environ Microbiol. 72:117–123.

Czarnecki O, Lippert I, Henning M, Welker M. 2006. Identification of peptide metabolites of *Microcystis* (Cyanobacteria) that inhibit trypsin-like activity in planktonic herbivorous *Daphnia* (Cladocera). Environ Microbiol. 8:77–87.

Dayhoff MO, Schwartz RM, Orcutt BC. 1978. A model of evolutionary change in proteins. In: Dayhoff MO, editor.

Atlas of protein sequence structure. Vol. 5. 345–352. Washington (DC): National Biomedical Research Foundation.
Dong C, Flecks S, Unversucht S, Haupt C, van Pée KH, Naismith JH. 2005. Tryptophan 7-halogenase (PrnA) structure suggests a mechanism of regioselective chlorination. Science. 309:2216–2219.
Dorrestein PC, Yeh E, Garneau-Tsodikova S, Kelleher NL, Walsh CT. 2005. Dichlorination of a pyrrolyl-S-carrier protein by FADH2-dependent halogenase PltA during pyoluteorin biosynthesis. Proc Natl Acad Sci USA. 102:13843–13848.
Eustáquio AS, Gust B, Luft T, Li SM, Chater KF, Heide L. 2003. Clorobiocin biosynthesis in Streptomyces: identification of the halogenase and generation of structural analogs. Chem Biol. 10:279–288.
Frangeul L, Quillardet P, Castets AM, et al. (20 co-authors). 2008. Highly plastic genome of *Microcystis aeruginosa* PCC7806, an ubiquitous toxic freshwater cyanobacterium. BMC Genomics. 9:274.
Galonic DP, Vaillancourt FH, Walsh CT. 2006. Halogenation of unactivated carbon centers in natural product biosynthesis: trichlorination of leucine during barbamide biosynthesis. J Am Chem Soc. 128:3900–3901.
Garrigues T, Dauga C, Ferquel E, Choumet V, Failloux AB. 2005. Molecular phylogeny of *Vipera Laurenti*, 1768 and the related genera *Macrovipera* (Reuss, 1927) and Daboia (Gray, 1842), with comments about neurotoxic *Vipera aspis aspis* populations. Mol Phylogenet Evol. 35:35–47.
Gascuel O. 1997. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Mol Biol Evol. 14:685–695.
Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate phylogenies by maximum likelihood. Syst Biol. 52:696–704.
Harris CM, Kannan R, Kopecka H, Harris TM. 1985. The role of the chlorine substituents in the antibiotic vancomycin: preparation and characterization of mono- and didechloro-vancomycin. J Am Chem Soc. 107:6652–6658.
Ishida K, Christiansen G, Yoshida WY, Welker M, Bonjoch J, Hertweck C, Börner T, Hemscheidt TK, Dittmann E. 2007. Biosynthetic pathway and structure analysis of aeruginoside 126A and B, cyanobacterialpeptides bearing an unusual 2-carboxy-6-hydroxyoctaindole moiety. Chem Biol. 14: 565–576.
Ishida K, Matsuda H, Murakami M. 1998. Four new microginins, linear peptides from the cyanobacterium *Microcystis aeruginosa*. Tetrahedron. 54:13475–13484.
Ishida K, Okita Y, Matsuda H, Okino T, Murakami M. 1999. Aeruginosins, protease inhibitors from the cyanobacterium *Microcystis aeruginosa*. Tetrahedron. 55:10971–10988.
Iteman I, Rippka R, Tandeau de Marsac N, Herdman M. 2000. Comparison of conserved structural and regulatory domains within divergent 16S rRNA-23S rRNA spacer sequences of cyanobacteria. Microbiol SGM. 146:1275–1286.
Jenke-Kodama H, Sandmann A, Müller R, Dittmann E. 2005. Evolutionary implications of bacterial polyketide synthases. Mol Biol Evol. 22:2027–2039.
John U, Beszteri B, Derelle E, van de Peer Y, Read B, Moreau H, Cembella AD. 2008. Novel insights into evolution of protistan polyketide synthases through phylogenomic analysis. Protist. 159:21–30.
Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci. 8:275–282.
Kaneko T, Nakajima N, Okamoto S, et al. (23 co-authors). 2007. Complete genomic structure of the bloom-forming toxic cyanobacterium *Microcystis aeruginosa* NIES-843. DNA Res. 14:247–256.
Keller S, Wage T, Hohaus K, Hölzer M, Eichhorn E, van Pée KH. 2000. Purification and partial characterization of tryptophan 7-halogenase (PrnA) from *Pseudomonas fluorescens*. Angew Chem Int Ed Engl. 39:2300–2302.
Kosiol C, Goldman N. 2005. Different versions of the Dayhoff rate matrix. Mol Biol Evol. 22:193–199.
Lee AY, Smitka TA, Bonjouklian R, Clardy J. 1994. Atomic structure of the trypsin-A90720A complex: a unified approach to structure and function. Chem Biol. 1:113–117.
Levinson G, Gutman GA. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol. 4:203–221.
Matern U, Schleberger C, Jelakovic S, Weckesser J, Schulz GE. 2003. Binding structure of elastase inhibitor scyptolin A. Chem Biol. 10:997–1001.
Mbedi S, Welker M, Fastner J, Wiedner C. 2005. Variability of *mcy*-genes in the genus *Planktothrix* (Oscillatoriales, Cyanobacteria). FEMS Microbiol Lett. 245:299–306.
Mikalsen B, Boison G, Skulberg OM, Fastner J, Davies W, Gabrielsen TM, Rudi K, Jakobsen KS. 2003. Natural variation in the microcystin synthetase operon *mcyABC* and impact on microcystin production in *Microcystis* strains. J Bacteriol. 185:2774–2785.
Nicholas KB, Nicholas HBJ. 1997. GeneDoc: Analysis and visualization of genetic variation. [Internet]. Available from: http:www.nrbs.org/. (Last accessed June 24, 2008).
Pride DT. 2004. Swaap 1.0.1: A tool for analysing substitutions and similarity in multiple alignments. [Internet]. Available from: http://www.bacteriamusem.org/SWAAP/SwaapPage/htm.
Rantala A, Fewer D, Hisbergues M, Rouhiainen L, Vaitomaa J, Börner T, Sivonen K. 2004. Phylogenetic evidence for the early evolution of microcystin synthesis. Proc Natl Acad Sci USA. 101:568–573.
Rippka R, Herdman M. 1992. Catalogue and taxonomic handbook, vol. 1: catalogue of strains. Paris (France): Institut Pasteur.
Rouhiainen L, Paulin L, Suomalainen S, Hyytiainen H, Buikema W, Haselkorn R, Sivonen K. 2000. Genes encoding synthetases of cyclic depsipeptides, anabaenopeptilides, in *Anabaena* strain 90. Mol Microbiol. 37:156–167.
Rounge TB, Rohrlack T, Tooming-Klunderud A, Kristensen T, Jakobsen KS. 2007. Comparison of cyanopeptolin genes in *Planktothrix, Microcystis*, and *Anabaena* strains: evidence for independent evolution within each genus. Appl Environ Microbiol. 73:7322–7330.
Shin HJ, Matsuda H, Murakami M, Yamaguchi K. 1997. Aeruginosins 205A and -B, serine protease inhibitory glycopeptides from the cyanobacterium *Oscillatoria agardhii* (NIES-205). J Org Chem. 62:1810–1813.
Sitachitta N, Marquez BL, Williamson RT, Rossi J, Roberts MA, Gerwick WH, Nguyen V-A, Willis CL. 2000. Biosynthetic pathway and origin of the chlorinated methyl group in barbamide and dechlorobarbamide, metabolites from the marine cyanobacterium *Lyngbya majuscula*. Tetrahedron. 56:9103–9113.
Tanabe Y, Kaya K, Watanabe MM. 2004. Evidence for recombination in the microcystin synthetase (*mcy*) genes of toxic cyanobacteria *Microcystis* spp. Mol Evol. 58:633–641.
Tooming-Klunderud A, Rohrlack T, Shalchian-Tabrizi K, Kristensen T, Jakobsen KS. 2007. Structural analysis of a non-ribosomal halogenated cyclic peptide and its putative operon from *Microcystis*: implications for evolution of cyanopeptolins. Microbiol SGM. 153:1382–1393.
Vaillancourt FH, Yeh E, Vosburg DA, Garneau-Tsodikova S, Walsh CT. 2006. Nature's inventory of halogenation catalysts: oxidative strategies predominate. Chem Rev. 106:3364–3378.

Vaillancourt FH, Yin J, Walsh CT. 2005. SyrB2 in syringomycin E biosynthesis is a nonheme FeII α-ketoglutarate- and O2-dependent halogenase. Proc Natl Acad Sci USA. 102: 10111–10116.

van Pée KH, Patallo EP. 2006. Flavin-dependent halogenases involved in secondary metabolism in bacteria. Appl Microbiol Biotechnol. 70:631–641.

van Pée KH, Zehner S. 2003. Enzymology and molecular genetics of biological halogenation. In: Gribble GW, editor. Natural production of organohalogen compounds. Berlin (Germany): Springer. p. 171–199.

von Elert E, Oberer L, Merkel P, Huhn T, Blom JF. 2005. Cyanopeptolin 954, a chlorine-containing chymotrypsin inhibitor of *Microcystis aeruginosa* NIVA Cya 43. J Nat Prod. 68:1324–1327.

Welker M, Brunke M, Preussel K, Lippert I, von Döhren H. 2004. Diversity and distribution of *Microcystis* (Cyanobacteria) oligopeptide chemotypes from natural communities studied by single colony mass spectrometry. Microbiol SGM. 150:1785–1796.

Welker M, Marsalek B, Sejnohova L, von Döhren H. 2006. Detection and identification of oligopeptides in *Microcystis* (cyanobacteria) colonies: toward an understanding of metabolic diversity. Peptides. 27:2090–2103.

Welker M, von Döhren H. 2006. Cyanobacterial peptides—nature's own combinatorial biosynthesis. FEMS Microbiol Rev. 30:530–563.

Whelan S, Goldman N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol. 18:691–699.

Wynands I, van Pée KH. 2004. A novel halogenase gene from the pentachloropseudilin producer *Actinoplanes* sp. ATCC 33002 and detection of in vitro halogenase activity. FEMS Microbiol Lett. 237:363–367.

Yeh E, Garneau S, Walsh CT. 2005. Robust in vitro activity of RebF and RebH, a two-component reductase/halogenase, generating 7-chlorotryptophan during rebeccamycin biosynthesis. Proc Natl Acad Sci USA. 102:3960–3965.

Claudia Schmidt-Dannert, Associate Editor

Accepted July 3, 2008