Unique Biosynthetic Pathway in Bloom-Forming Cyanobacterial Genus Microcystis Jointly Assembles Cytotoxic Aeruginoguanidines and Microguanidines

Claire Pancrace, Keishi Ishida, Enora Briand, Douglas Gatte Pichi, Annika R Weiz, Arthur Guljamow, Thibault Scalvenzi, Nathalie Sassoon, Christian Hertweck, Elke Dittmann, et al.

1

2

3

A Unique Biosynthetic Pathway in Bloom-Forming Cyanobacterial Genus *Microcystis* Jointly Assembles Cytotoxic Aeruginoguanidines and Microguanidines

Claire Pancrace[1,2,7], Keishi Ishida[3,7], Enora Briand[4,7], Douglas Gatte Pichi[5], Annika R. Weiz[5], Arthur Guljamow[5], Thibault Scalvenzi[1], Nathalie Sassoon[1], Christian Hertweck[3,6], Elke Dittmann[5,*], Muriel Gugger[1,*]

Affiliations

[1]Institut Pasteur, Collection des Cyanobactéries, 28 rue du Dr Roux, 75724 Paris Cedex 15, France

[2]UMR UPMC 113, CNRS 7618, IRD 242, INRA 1392, PARIS 7 113, UPEC, IEES Paris, 4 Place Jussieu, 75005, Paris, France

[3]Leibniz Institute for Natural Product Research and Infection Biology, Hans Knöll Institute, Beutenbergstr. 11a, 07745 Jena, Germany

[4]Ifremer, Laboratoire Phycotoxines, rue de l'Ile d'Yeu, 44311 Nantes, France

[5]Department of Microbiology, Institute of Biochemistry and Biology, University of Potsdam, 14476 Golm, Germany

[6] Faculty of Biological Sciences, Friedrich Schiller University Jena, 07743 Jena, Germany

Footnotes

[7]These authors contributed equally.

*Correspondence: mgugger@pasteur.fr, editt@uni-potsdam.de

Abstract

The cyanobacterial genus *Microcystis* is known to produce an elaborate array of structurally unique and biologically active natural products including hazardous cyanotoxins. Cytotoxic aeruginoguanidines represent a yet unexplored family of peptides featuring a trisubstituted benzene unit and farnesylated arginine derivatives. In this study, we aimed at assigning these compounds to a biosynthetic gene cluster by utilizing biosynthetic attributes deduced from public genomes of *Microcystis* and the sporadic distribution of the metabolite in axenic strains of the Pasteur Culture Collection of Cyanobacteria.

By integrating genome mining with untargeted metabolomics using liquid chromatography with mass spectrometry, we could link aeruginoguanidine (AGD) to a nonribosomal peptide synthetase gene cluster and co-assign a significantly smaller product to this pathway, microguanidine (MGD), previously only reported from two *Microcystis* blooms. Further, a new intermediate class of compounds named microguanidine amides was uncovered thereby further enlarging this compound family. The comparison of structurally divergent AGDs and MGDs reveals an outstanding versatility of this biosynthetic pathway and provides insights into the assembly of the two compound subfamilies.

Strikingly, aeruginoguanidines and microguanidines were found to be as widespread as the hepatotoxic microcystins, but the occurrence of both toxin families appeared to be mutually exclusive.

INTRODUCTION

Microcystis is a dominant bloom-forming cyanobacterium occurring in temperate freshwater ecosystems.[1] The genus is infamous for the production of the well-known hepatotoxin microcystin.[2] Both blooms and toxins cause ecosystem disturbance and public health threats, and constitute a growing concern in the frame of freshwater eutrophication and global warming. *Microcystis* has also been described as a producer of a multitude of bioactive natural products, some of interest for biotechnological and pharmaceutical application.[3-5]

Cytotoxic aeruginoguanidines (AGDs) represent one of the most remarkable families of compounds described for *Microcystis*.[6] The three AGD congeners reported for strain *M. aeruginosa* NIES-98 feature highly unprecedented characteristics such as a 1-(4-hydroxy-3-hydroxymethyl)-phenyl-1-hydroxy-2-propylamine-4',3',1-tri-*O*-sulfate (Hphpa trisulfate) moiety, along with geranylation and prenylation of arginines (Fig. 1A). While bloom-forming *Microcystis* belong to the most intensively studied cyanobacteria, AGDs were reported only twice from a bloom in Czech Republic and an isolate in Brazil,[7, 8] and never from any other cyanobacteria. Their intricate features confine AGDs into a unique compound family.[3]

Our recent genomic analysis of ten *Microcystis* strains revealed that the different genotypes share a highly similar core genome while their biosynthetic gene clusters (BGCs) involved in natural product (NP) formation show a sporadic distribution. Moreover, we uncovered three cryptic BGCs not associated with any cyanobacterial compound.[9] The continuously increasing number of publically available genomes of *Microcystis* further corroborates the high genetic diversity and patchy distribution of the NPs produced by this cyanobacterium.

Analysis of mass spectrometry (MS) data has been widely used for years in NP characterization efforts. Molecular networking computational approach uses tandem MS/MS data to group spectra based on their fragmentation patterns similarities, which gain strength in the frame of multi-strain comparison. Approaches combining molecular networking with genome mining highlight putative links between parent ions and pathways responsible for their biosynthesis. This combinatorial approach has been shown effective at linking NPs to their biosynthetic gene clusters in cyanobacteria and other prokaryotes such as *Salinospora*.[10, 11].

Here, we have utilized the sporadic distribution of BGCs in *Microcystis* to assign one of

85 the orphan BGCs to AGD. By integrating the genome sequence of the known AGD-
86 producing strain *Microcystis aeruginosa* NIES-98,[12] we screened *Microcystis* public
87 genomes and axenic PCC strains for the AGD and its candidate BGC using genome
88 mining, PCR and untargeted metabolomics. These data were further combined with
89 molecular networking and genome comparison to link AGD to its biosynthetic gene
90 cluster and study its diversity at the genetic and the metabolite level. The integrative
91 approach allowed to enlarge the AGD compound family with microguanidine amide
92 congeners (MGAs) and new variants of microguanidines (MGDs), and provides
93 comprehensive insights into the extraordinary versatility of this biosynthetic pathway.
94
95 **RESULTS AND DISCUSSION**
96 **Candidate synthesis BGC for sulfated, geranylated and prenylated compounds.**
97 Considering the chemical structure of aeruginoguanidine (Figure 1A), the BGC involved
98 in its synthesis was expected to encode nonribosomal peptide synthetase (NRPS)
99 modules with specificity for L-arginine and tailoring enzymes such as a
100 prenyltransferase and a sulfatase/sulfotransferase. The genome of the AGD-producing
101 strain *Microcystis aeruginosa* NIES-98 contained only one cluster with these features,
102 which was homologous to the MIC2 cluster previously described in the genomes of
103 *Microcystis aeruginosa* PCC 9806 and PCC 9717 and *Microcystis* sp. T1-4.[9] The candidate
104 BGC encoded two mono-modular NRPS, one of which comprising an integrated *N*-
105 methylation domain as anticipated for the *N*-methylation of the Arg moieties. Substrate
106 prediction of the second NRPS was more ambiguous without excluding Arg (Table 1).
107 The putative AGD BGC, which spans ~34kb in the genome of *Microcystis aeruginosa*
108 NIES-98, includes 25 genes (Table 1) organized in three operons (Figure 1B). The two
109 NRPS AgdE and AgdK are accompanied by a predicted hydroxybenzoate synthase
110 (AgdH), an AMP-dependent-ligase (AgdA), a peptidyl carrier protein (AgdB), a radical
111 SAM protein with decarboxylase function (AgdC) and two thioester reductases (AgdN
112 and AgdU). Several proteins consistent with tailoring enzymes involved in AGD
113 biosynthetic pathways are present such as two methyltransferases (AgdI, AgdM), an
114 aminotransferase (AgdL), an isoprenyltransferase (AgdJ), several
115 sulfatase/sulfotransferases (AgdD, AgdG, AgdP and AgdR), plus putative
116 permease/transporters (AgdF, AgdO), and thiamine pyrophosphatase (AgdQ) genes.
117 This candidate BGC for AGD present in seven genomes, including the public ones of

118 *Microcystis aeruginosa* TAIHU98, *Microcystis* sp. SPC777 and CACIAM03, was used to

119 optimize specific primers and PCR conditions to detect its presence in *Microcystis*

120 strains. The two primer pairs designed were targeting two genes of the candidate BGC

121 presumably involved in an early and a late stage of AGD biosynthesis. Both genes do not

122 share homologies with other NRPS BGCs in *Microcystis* (*agdH* and *agdJ*, Table S1). The

123 screening of these two selected genes revealed seven additional PCC *Microcystis* strains,

124 whose on-going genome sequences helped to better define the limits of this BGC (Table

125 S2). A close inspection of the 14 genomes revealed the candidate AGD BGC with 28 genes

126 in perfect synteny, without rearrangement, and expanded the initial MIC2 cluster with

127 conserved neighboring genes (Figure 1B). Noteworthy, the largest NRPS gene *agdK* of

128 *Microcystis* sp. PCC 10613 was reduced to a remnant fragment, as confirmed by PCR. In

129 addition, the gene *agdK* was split in two in the genomes of *Microcystis* sp. CACIAM03 and

130 TAIHU98. Similarly, the gene *agdQ* was split in the genome of PCC 9624, while a contig

131 border separated *agdP* and *agdQ* in the genomes of PCC 9624 and PCC 10613. The

132 predicted aminotransferase gene *agdL* was lacking in the genomes of PCC 9717 and PCC

133 9810, also confirmed by PCR. Finally, the genes *agdS* and *agdT*, without known function,

134 appeared duplicated in ten strains (Figure 1B).

135

136 **AGD and co-assignment of microguanidine by Molecular Networking.** Detection of

137 AGD was performed by LC-MS/MS to assess its presence in the AGD producer strain

138 NIES-98 and in ten strains of the PCC containing the candidate BGC, as well as in eight

139 PCC strains that did not contain it in their genomes. Two molecular networks (MNs)

140 were constructed from LC-MS/MS data, one in positive mode (MN(+)) and another in

141 negative mode (MN(–)). In order to dereplicate the complex dataset, signatures of NPs

142 previously found in some of these *Microcystis* strains were identified using high-

143 performance liquid chromatography electrospray ionization mass spectrometry (HPLC-

144 ESI-MS/MS). Specifically, MS/MS fragments were identified for the cyanopeptolins A, B

145 and C in PCC 7806, aeruginosamides B and C and ferintoic acid (anabaenopeptins) in

146 PCC 9432, and ferintoic acid in PCC 9701 as predicted from their genomes (Figure S1A).

147 [9, 13] The MN(+), consisting of 1998 nodes, was thus reliable in finding the expected

148 compounds. However, AGD was spread in several nodes of the MN(+) apart from each

149 other. Indeed, AGD had a better fragmentation pattern in negative mode as it was

150 collapsed into a single large node among the 1876 nodes of the MN(–) (Figure S1B). An

extraction of the AGD node in MN(–) encompassed all strains carrying the full candidate BGC for AGD synthesis, but neither the strain PCC 10613 nor the strains lacking this AGD candidate cluster (Figure 2A). Up to 20 different putative variants of AGD were found in these *Microcystis* strains, with strains NIES-98, PCC 9804, PCC 9805 and T1-4 able to produce the three known AGD standards, whereas the other strains produced one or two of those variants (Figure 2).

Strikingly, the MN(–) revealed that all the strains containing the AGD candidate BGC produced also a significantly smaller product of 772 Da (Figure 2B). Literature research revealed that a compound with this mass, microguanidine AL772, was previously reported for a *Microcystis* bloom.[14] Microguanidines (MGDs) share striking similarities with AGDs but display also considerable differences. Instead of the highly unusual Hphpa trisulfate moiety, MGDs contain 3-(4-hydroxy-3-hydroxymethylphenyl)-2-hydroxy-1-propanol (Hphpol). Further, MGDs feature a permethylation at the α-amino group of Arg that has not been observed in AGDs. Along with MGD AL772 (**4**, Figure 3**,** related Figures S3 and S4A), a new MGD variant, MGD-704 (**5**, Figure 3**,** related Figures S3 and S4B, and Table S3) was detected in the majority of strains differing from the two other characterized MGDs, KT636 and DA368.[14-16]

In addition, the structural elucidation of the MGD size range compounds by MS fragmentation and high-resolution MS analyses uncovered a novel intermediate class of metabolites mixing features of AGD and MGD. While both compounds contain the Hphpa trisulfate moiety linked with an amide bond to the arginine derivative as in AGDs they were lacking the second arginine moiety and carried the same permethylation at the α-amino group of Arg as in MGDs (**6** and **7**, Figure 3**,** related Figures S3 and S4C and D, and Table S3). To confirm the structure of **6**, several **4** and **6** producing strains were extracted and small amounts of **4** and **6** were purified by reversed-phase HPLC. The ^{1}H NMR spectra of AGD 98-A (**1**), AGD 98-B (**2**), AGD 98-C (**3**), MGD AL772 (**4**), and MGA (**6**) showed highly similar signals (Figure S5-S14). Detailed comparison of ^{1}H NMR signals between **4** and **6** revealed three notable differences, namely the appearance of new amide proton δ 8.48 (H11 in **6**), 1.02 and 0.18 ppm and high field shifted methine protons H8 δ 5.21 (**4**) to δ 4.23 (**6**) and H13 δ 4.11 (**4**) to δ 3.93 (**6**), respectively (Figure S5). The ^{1}H-^{1}H COSY correlation from H8 to H11 and HSQC analysis of **6** indicated that C8 (δ 49.8 in **6**, δ 75.3 in **4**) is adjacent to nitrogen (Figure S15-S18, Table S4). These results strongly supported that the predicted structure of **6** indeed possesses an amide

bond instead of the ester bond in **4**. As the low amount of **6** did not enable a sufficient quality of ¹³C NMR and other 2D NMR spectra, chemical shift assignment of **6** was performed by the comparison with NMR data of **4**. The stereochemistry of the geranyl group of **6** was determined as *Z*-form, judging from the close similarity of chemical shifts with **1**-**3** and ¹³C NMR data of geraniol (*E*-form) and nerol (*Z*-form) (www.chemicalbook.com/). This result further revealed that the stereochemistry of geranyl group of MGD AL772 (**4**) also has *Z*-form. The new intermediate class of compounds was designated microguanidine amide, with MGA-771 and MGA-787.

Indeed, the MGA peptides and the two MGD depsipeptides were observed simultaneously with AGDs in four strains (PCC 9804, PCC 9805, PCC 9811 and T1-4). Thus, *Microcystis* harboring the Agd BGC may build two different condensations between the modified Arg residue and the phenetylalcohol (ester bond) in MGD congeners or the phenetylamine (amide bond) in all AGD congeners (Figure 3).

The co-existence of AGD and MGD in the majority of Agd BGC positive strains, the existence of a new intermediate class and the large overlap in anticipated biosynthetic features lead us to conclude that AGD and MGD represent alternative products of the same biosynthetic pathway. Remarkably, strain PCC 10613 lacking the NRPS gene agdK was found to produce the MGDs in the MN(−) (Figure 2). Noteworthy, strain PCC 9624 in which the Agd BGC differed at the level of the *agdQ* produced only the AGD-98A and the MGD-AL772. Similarly, PCC 9810, PCC 9811 and PCC 9717 that lack the predicted aminotransferase *agdL* and several Agd genes of unknown function (*agdS'*, *agdT'*) produced a lower diversity of AGD variants under the same growing conditions than other AGD producing *Microcystis* strains. None of the other *Microcystis* strains analyzed, notably the ones containing the Mcy gene cluster, produced AGD, MGA or MGD.

Characterization of the BGC potentially involved in the AGD/MGD synthesis. One of the most striking findings of our study is the extraordinary diversity of products concurrently generated by the AGD/MGD pathway in single strains. Considering the variations detected even in the backbone of AGDs and MGDs and in the linkage of their individual moieties, the biosynthesis pathway cannot be considered as a classic assembly line of NRPS. This pathway is rather a toolkit of enzymes optionally producing a cocktail of metabolites that share the same precursors and similar tailoring modifications but combine the different building blocks to alternative products. At the

same time, the unprecedented diversity of products and intermediates and the existence of natural mutants lacking individual biosynthetic genes allows for conclusions regarding a number of biosynthetic steps of the complex pathway.

The presence of a putative *p*-hydroxybenzoate synthase (AgdH) in the AGD cluster indicates that the trisubstituted benzene unit of Hphpa and Hphpol might be derived from chorismate [17]. Given that Hphpa and Hphpol possess a rare *m*-hydromethyl residue in the benzene ring, AgdH might act in a similar way as isochorismate mutase, which has been reported to catalyze the transformation of isochorismate to *m*-carboxyphenylpyruvate.[18, 19] We cannot dissect all individual steps towards the Hphpa and Hphpol moieties, but we propose that the AMP-dependent ligase AgdA might activate the *o*-carboxylic acid group of a *p*-hydroxyphenylpyruvate intermediate followed by the transfer to the free-standing PCP AgdB (Figure 4). The resulting thioester is presumably reduced to the corresponding alcohol either by thioester reductase AgdN or U through reductive chain termination as shown for myxochelin biosynthesis in *Stigmatella aurantiaca*.[20] A yet unassigned hydroxylation step at the β-position of the *m*-hydroxymethyl-*p*-hydroxyphenylpyruvate yields 3-hydroxy-*m*-hydroxymethyl-*p*-hydroxyphenylpyruvate as the precursor of both Hphpa and Hphpol. We hypothesize that this precursor represents a branching point where further transformation of the α-keto group by aminotransferase AgdL yields Hphpa, while transformation by a reductase (e.g. AgdN or U) yields Hphpol (Figure 4). This hypothesis is supported by the fact that the lack of agdL in strains PCC 9717 and PCC 9810 still permits production of MGD variants containing the Hphpol moiety (**4** and **5**) but not the alternative Hphpa moiety as in MGAs (**6** and **7**). It is of note, that some of the predicted biosynthetic steps for Hphpa and Hphpol biosynthesis (Figure 4) share similarities to enzyme reactions involved in biosynthesis of the characteristic Choi moiety in the aeruginosin pathway [21]. In this context, it is worth mentioning that the majority of AGD/MGD producers also harbor aeruginosin biosynthesis genes in their genome (Figure 5), thus not excluding the possibility of a joint use of precursors and enzymes.

Furthermore, the strain *M. aeruginosa* PCC 10613 can be considered as a natural *agdK* mutant, thus allowing deducing the roles of the two NRPSs in the pathway. The fact that the lack of AgdK in PCC 10613 still enables MGD production strongly suggests that AgdE is the responsible NRPS activating Arg in the MGD and MGA pathways (Figure 6). On the other hand, the NRPS AgdK harbouring an N-methyltransferase domain is likely

incorporating *N*-Me-Arg in the AGD pathway. Whether or not AgdK acts iteratively or cooperates with AgdE to yield the MeArg-MeArg-Hphpa moiety of AGDs cannot be dissected based on the current dataset. The biosynthetic intermediate(s) might be methylated and decarboxylated by the radical SAM enzyme AgdC. Since AgdC shows close homology to the oxygen-independent coproporphyrinogen III oxidase of *E.coli* (HemN) we propose that it utilizes a 5`-deoxyadenosyl radical to trigger a decarboxylation reaction as demonstrated for the HemN enzyme family. [22] The intermediate may further be modified by several tailoring enzymatic reactions such as *N*-methylation (methyltransferase; AgdI or M) of Arg residue, to the tri-sulfation (sulfotransferases; AgdD, P and R, sulfatase; AgdG) of the Hphpa residue, and the *N*-alkylation (isoprenyltransferase; AgdJ) of *N*-MeArg residues. Some of the proposed biosynthetic steps may occur while substrates are tethered on PCP-domains of NRPSs or the standalone peptidyl carrier protein AgdB. The fact that no desulfated intermediates were observed in the MS/MS networking may suggest that sulfation of the aromatic moiety occurs in the PCP-bound state.

The distinct alkylation pattern at the guanidinyl group of *N*-trimethyl Arg (ω for AGDs and ε for MGDs) may derive from alternative substrate specificities of the isoprenyltransferase AgdJ (Figure 6). Comparison of the distinct AGD/MGD product profiles of individual *Microcystis* strains thus suggests an outstanding versatility of the pathway. A complete assignment of biosynthetic steps will require biochemical characterization of participating enzymes and targeted feeding studies, yet the analysis of natural agdK and agdL mutants led to definite conclusions regarding the role of these two enzymes.

The example of the joint AGD/MGD pathway further strengthens the paradigm that cyanobacteria have evolved unique mechanisms to produce diverse NPs of high complexity in single strains using limited genetic resources. Other cyanobacterial mechanisms include the utilization of alternative starter modules for NRPS as shown for the anabaenopeptin synthetase of strain *Anabaena* 90,[23] the integration of multispecific adenylation domains of NRPS as shown for the anabaenopeptin synthetase of *Planktothrix* NIVA-CYA 126,[24] and the microcystin synthetase in *Microcystis aeruginosa* NIES 843.[25] Recently, a simultaneous production of anabaenopeptins and namalides allowed to reveal a single pathway for their synthesis.[26] We can only speculate whether

282 AGDs and MGDs act synergistically or fulfill parallel independent functions in the
283 producing strains.

284 An interesting phenomenon observed during this study is that AGD/MGD production
285 and MC production are almost mutually exclusive among *Microcystis* strains. The only
286 exception was found in the genomes of two non-monoclonal Brazilian strains,[27, 28] that
287 carry both clusters and for which the production of these compounds is not yet
288 documented. There is increasing evidence that MCs are closely interfering with the
289 primary metabolism of *Microcystis* in addition to their toxicity.[29] Whether or not AGD
290 and MGD can complement for the loss of MC or reflect a different niche adaptation of
291 their respective producers remains elusive.

292 Our study further suggests that the rare detection of AGD and MGD in only two
293 *Microcystis aeruginosa* isolated in Japan and in Brazil (NIES 98[6] and NPCD-1[8]) and bloom
294 materials of *Microcystis* in Israel[14-16] respectively is not due to the scarce occurrence of
295 these metabolites among *Microcystis*, but rather to the lack of attention towards these
296 peculiar NPs in previous studies. Thus, the AGD/MGD producers seem to be as dispersed
297 worldwide as the MC producing strains, and therefore should be considered in future
298 screening of *Microcystis* blooms and isolates.

299

300 **CONCLUSIONS**

301 Cyanobacteria are infamous for worldwide bloom formation in freshwater bodies. Risk
302 assessment of *Microcystis* blooms primarily considers the hepatotoxin microcystin (MC).
303 The present study suggests that the neglected family of compounds, cytotoxic
304 aeruginoguanidines and microguanidines, is more frequently produced than previously
305 anticipated, mainly in non-MC producing *Microcystis* strains. Remarkably, the two
306 structurally divergent groups of compounds are products of a branched and versatile
307 biosynthetic pathway. The genetically constraint gene cluster generates a library of
308 diverse products in single strains and further strengthens the paradigm that
309 cyanobacteria have developed unique mechanisms to generate metabolic diversity.
310 These findings open new perspectives for future studies on orphan natural products and
311 evolution of their biosynthetic pathways.

312

313

314

MATERIALS AND METHODS

Strain cultures and detection of the cluster. Axenic *Microcystis* strains from the PCC and from the NIES collections were grown at 25 °C in 40 mL BG11$_0$ medium[30] supplemented with 2 mM NaNO$_3$ and 10 mM NaHCO$_3$ under continuous light (Table S2). For nucleic acid extraction, chemical and PCR analysis, the details are described in Supporting information.

Sequencing & genomics analysis. For the strains suspected to carry the *agd* gene cluster, whole genome sequencing was performed by the Mutualized Platform for Microbiology at Institut Pasteur. Genomes were integrated in the MicroScope platform[31] for further analysis. The genome sequencing is described in Supplemental information. The species tree was generated by a concatenation of 586 conserved proteins selected from the phylogenetic markers previously validated for Cyanobacteria.[32] Phylogenetic analysis is detailed in Supplemental information. AntiSMASH 3.0[33] was used to identify the targeted BGC in each genome sequence. In cases where the *agd* gene cluster spanned several contigs/scaffolds PCRs were performed to confirm the colocalization of the gene cluster parts in the same genomic locus (Table S1).

Cyanobacterial cell extraction. Lyophilized cyanobacterial cells from 200 mL cultures of 19 *Microcystis aeruginosa* strains were extracted with 80% aqueous methanol (v/v, 25 mL) using a sonicator (Sonoplus MS73, Bandelin, 30% power, 5 cycles for 2 min at room temperature). Each extract was centrifuged at 8,000 × g for 15 min at 15 °C. The residues were extracted with 80% aqueous methanol (v/v, 25 mL) and methanol (25 mL), respectively, as the above-mentioned procedure. The extracts were combined and dried under a reduced pressure. The crude residues were dissolved in 50% aqueous methanol (v/v, 1 mL) and kept in a fridge until analysis.

HPLC-MS measurement. LC-MS/MS measurements were carried out by Bruker HCT Ultra ion trap mass spectrometry (BrukerDaltonics, Bremen, Germany) coupled with an Agilent Technologies 1100 series liquid chromatogram system (Agilent, Waldbronn, Germany). The HR-LCMS measurements were performed by HPLC-HRMS series of Thermo Accela (LC) and Thermo Exactive (HRMS), an ESI source operating in both polarity mode and an orbitrap analyzer (Thermo Fisher Scientific, Bremen). The details of both measurements are described in Supporting Information.

Molecular networking. LC-MS/MS data acquired from Bruker instrument were used for molecular networking. Two molecular networks (MNs) were performed with LC-

MS/MS data, one in positive mode (MN(+)) and another with negative mode data (MN(–)) with LC-MS/MS data from *Microcystis* strains and AGD A, B and C standards. The steps followed for both MNs are described in Supporting Information.

REFERENCES

1. Harke, M. J., Steffen, M. M., Gobler, C. J., Otten, T. G., Wilhelm, S. W., Wood, S. A., and Paerl, H. W. (2016) A review of the global ecology, genomics, and biogeography of the toxic cyanobacterium, *Microcystis* spp, *Harmful Algae 54*, 4—20.

2. Merel, S., Walker, D., Chicana, R., Snyder, S., Baures, E., and Thomas, O. (2013) State of knowledge and concerns on cyanobacterial blooms and cyanotoxins, *Environ. Int. 59*, 303—327.

3. Welker, M., and von Dohren, H. (2006) Cyanobacterial peptides - nature's own combinatorial biosynthesis, *FEMS Microbiol. Rev. 30*, 530—563.

4. Kehr, J. C., Gatte Picchi, D., and Dittmann, E. (2011) Natural product biosyntheses in cyanobacteria: A treasure trove of unique enzymes, *Beilstein J. Org. Chem. 7*, 1622—1635.

5. Dittmann, E., Gugger, M., Sivonen, K., and Fewer, D. P. (2015) Natural product biosynthetic diversity and comparative genomics of the Cyanobacteria, *Trends Microbiol. 23*, 642—652.

6. Ishida, K., Matsuda, H., Okita, Y., and Murakami, M. (2002) Aeruginoguanidines 98-A-98-C: cytotoxic unusual peptides from the cyanobacterium *Microcystis aeruginosa*, *Tetrahedron 58*, 7645—7652.

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

8. Silva-Stenico, M., da Silva, C., Lorenzi, A., Shishido, T., Etchegaray, A., Lira, S., Moraes, L., and Fiore, M. (2011) Non-ribosomal peptides produced by Brazilian cyanobacterial isolates with antimicrobial activity, *Microbiol. Res. 166*, 161—175.

9. Humbert, J. F., Barbe, V., Latifi, A., Gugger, M., Calteau, A., Coursin, T., Lajus, A., Castelli, V., Oztas, S., Samson, G., Longin, C., Medigue, C., and de Marsac, N. T. (2013) A tribute to disorder in the genome of the bloom-forming freshwater cyanobacterium *Microcystis aeruginosa*, *PLoS One 8*, e70747.

380 10. Duncan, K. R., Crusemann, M., Lechner, A., Sarkar, A., Li, J., Ziemert, N., Wang, M.,
381   Bandeira, N., Moore, B. S., Dorrestein, P. C., and Jensen, P. R. (2015) Molecular
382   networking and pattern-based genome mining improves discovery of biosynthetic
383   gene clusters and their products from *Salinispora* species, *Chem. Biol. 22*, 460—471.

384 11. Moss, N. A., Bertin, M. J., Kleigrewe, K., Leao, T. F., Gerwick, L., and Gerwick, W. H.
385   (2016) Integrating mass spectrometry and genomics for cyanobacterial metabolite
386   discovery, *J. Ind. Microbiol. Biotechnol. 43*, 313—324.

387 12. Yamaguchi, H., Suzuki, S., Sano, T., Tanabe, Y., Nakajima, N., and Kawachi, M. (2016)
388   Draft genome sequence of *Microcystis aeruginosa* NIES-98, a non-microcystin-
389   producing cyanobacterium from Lake Kasumigaura, Japan, *Genome Announc. 4*,
390   e01187—01116.

391 13. Briand, E., Bormans, M., Gugger, M., Dorrestein, P. C., and Gerwick, W. H. (2016)
392   Changes in secondary metabolic profiles of *Microcystis aeruginosa* strains in
393   response to intraspecific interactions, *Environ. Microbiol. 18*, 384—400.

394 14. Gesner-Apter, S., and Carmeli, S. (2008) Three novel metabolites from a bloom of
395   cyanobacterium *Microcystis* sp., *Tetrahedron 64*, 6628—6634.

396 15. Adiv, S., and Carmeli, S. (2013) Protease inhibitors from *Microcystis aeruginosa*
397   bloom material collected from the Dalton Reservoir, Israel, *J. Nat. Prod. 76*, 2307—
398   2315.

399 16. Lifshits, M., and Carmeli, S. (2012) Metabolites of *Microcystis aeruginosa* bloom
400   material from Lake Kinneret, Israel, *J. Nat. Prod. 75*, 209—219.

401 17. Siebert, M., Severin, K., and Heide, L. (1994) Formation of 4-hydroxybenzoate in
402   *Escherichia coli*: Characterization of the ubiC gene and its encoded enzyme
403   chorismate pyruvate-lyase., *Microbiology 140*, 897-904.

404 18. Zamir, L. O., Nikolokakis, A., Bonner, C. A., and Jensen, R. A. (1993) Evidence for
405   enzymatic formation of isoprephenate from isochorismate, *Bioorganic Med. Chem.
406   Lett. 3*, 1441—1446.

407 19. Blasiak, L. C., and Clardy, J. J. (2010) Discovery of 3-formyl-tyrosine metabolites from
408   *Pseudoalteromonas tunicata* through heterologous expression, *Am. Chem. Soc. 132*,
409   926—927.

410 20. Li, Y., Weissman, K., and Müller, R. (2008) Myxochelin biosynthesis: Direct evidence
411   for two- and four-electron reduction of a carrier protein-bound thioester, *J. Am.
412   Chem. Soc. 130*, 7554—7555.

413   21. Ishida, K., Christiansen, G., Yoshida, W., Kurmayer, R., Welker, M., Valls, N., Bonjoch, J.,
414   Hertweck, C., Börner, T., Hemscheidt, T., and Dittmann, E. (2007) Biosynthesis and
415   structure of aeruginoside 126A and 126B, cyanobacterial peptide glycosides bearing
416   a 2-carboxy-6-hydroxyoctahydroindole moiety, *Chem. Biol. 14*, 565—576.

417   22. Layer, G., Pierik, A. J., Trost, M., Rigby, S. E., Leech, H. K., Grage, K., Breckau, D., Astner,
418   I., Jansch, L., Heathcote, P., Warren, M. J., Heinz, D. W., and Jahn, D. (2006) The
419   substrate radical of *Escherichia coli* oxygen-independent coproporphyrinogen III
420   oxidase HemN, *J. Biol. Chem. 281*, 15727—15734.

421   23. Rouhiainen, L., Jokela, J., Fewer, D. P., Urmann, M., and Sivonen, K. (2010) Two
422   alternative starter modules for the non-ribosomal biosynthesis of specific
423   anabaenopeptin variants in *Anabaena* (Cyanobacteria), *Chem. Biol. 17*, 265—273.

424   24. Kaljunen, H., Schiefelbein, S., Stummer, D., Kozak, S., Meijers, R., Christiansen, G., and
425   Rentmeister, A. (2015) Structural elucidation of the bispecificity of A domains as a
426   basis for activating non-natural amino acids, *Angew. Chem. Int. Ed. Engl. 54*, 8833—
427   8836.

428   25. Meyer, S., Kehr, J., Mainz, A., Dehm, D., Petras, D., Süssmuth, R., and Dittmann, E.
429   (2016) Biochemical dissection of the natural diversification of microcystin provides
430   lessons for synthetic biology of NRPS, *Cell Chem. Biol. 23*, 462—471.

431   26. Shishido, T. K., Jokela, J., Fewer, D. P., Wahlsten, M., Fiore, M. F., and Sivonen, K.
432   (2017) Simultaneous production of anabaenopeptins and namalides by the
433   cyanobacterium *Nostoc* sp. CEN543, *ACS Chem. Biol. 12*, 2746—2755.

434   27. Fiore, M. F., Alvarenga, D. O., Varani, A. M., Hoff-Risseti, C., Crespim, E., Ramos, R. T.,
435   Silva, A., Schaker, P. D., Heck, K., Rigonato, J., and Schneider, M. P. (2013) Draft
436   genome sequence of the Brazilian toxic bloom-forming cyanobacterium *Microcystis*
437   *aeruginosa* strain SPC777, *Genome Announc. 1*, e00547—00513.

438   28. Castro, W. O., Lima, A. R., Moraes, P. H., Siqueira, A., Aguiar, D., Baraúna, A., Martins,
439   L., Fuzii, H., de Lima, C., Vianez-Júnior, J., Nunes, M., Dall'Agnol, L., and Gonçalves, E.
440   (2016) Draft genome sequence of *Microcystis aeruginosa* CACIAM 03, a
441   cyanobacterium isolated from an Amazonian freshwater environment, *Genome*
442   *Announc. 4*, e01299.

443   29. Neilan, B., Pearson, L., Muenchhoff, J., Moffitt, M., and Dittmann, E. (2013)
444   Environmental conditions that influence toxin biosynthesis in cyanobacteria,
445   *Environ. Microbiol. 15*, 1239—1253.

446   30. Rippka, R., Deruelles, J., Waterbury, J. B., Herdman, M., and Stanier, R. Y. (1979)
447   Generic assignments, strain histories and properties of pure cultures of
448   Cyanobacteria, *J. Gen. Microbiol. 111*, 1—61.
449   31. Vallenet, D., Calteau, A., Cruveiller, S., Gachet, M., Lajus, A., Josso, A., Mercier, J.,
450   Renaux, A., Rollin, J., Rouy, Z., Roche, D., Scarpelli, C., and Medigue, C. (2017)
451   MicroScope in 2017: An expanding and evolving integrated resource for community
452   expertise of microbial genomes, *Nucleic Acids Res. 45*, D517—D528.
453   32. Pancrace, C., Barny, M., Ueoka, R., Calteau, A., Scalvenzi, T., Pedron, J., Barbe, V., Piel, J.,
454   and Humbert, J. (2017) Insights into the *Planktothrix* genus: Genomic and metabolic
455   comparison of benthic and planktic strains, *Sci. Rep. 7*, 41181.
456   33. Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H. U., Bruccoleri, R., Lee, S. Y., Fischbach,
457   M. A., Muller, R., Wohlleben, W., Breitling, R., Takano, E., and Medema, M. H. (2015)
458   antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic
459   gene clusters, *Nucleic Acids Res. 43*, W237—243.

460

461   **FIGURE LEGENDS**

462   Figure 1. Aeruginoguanidines and the corresponding biosynthetic gene cluster. *(A)* The
463   structure of aeruginoguanidines (AGDs), **1**; AGD-98A, **2**; AGD-98B, **3**; AGD-98C; *(B)* AGD
464   biosynthetic gene cluster of *Microcystis aeruginosa* NIES-98 and its variation in 13 other
465   *Microcystis* genomes sharing 94 to 98% of similarity. The genes are color-coded with
466   orange for carbohydrate sulfotransferase, sulfotransferase and sulfatase; black for NRPS
467   and thioesterase; green for methyl-, isoprenyl- and aminotransferase; blue for
468   permease; grey for proteins with putative and unknown function. The blue line indicates
469   the span of the MIC2 cluster previously described.[9]

470   Figure 2. Molecular network of AGD (A) and of MGD with MGA (B). Characterized
471   structural variants are indicated as red-colored nodes and new congeners characterized
472   in this study are highlighted in green. Diversity and distribution of AGD and MGD
473   variants for each strain. Details of MN(+), MN(−), and the AGD and MGD networks are
474   presented in Figures S1-S3.

475   Figure 3. Microguanidine and microguanidine amide variants detected in strains
476   containing the AGD biosynthetic gene cluster. The MGD depsipeptides contain ester-
477   bonds, while the peptidic MGAs contain amide-bonds in their structures. Details of the
478   high-resolution MS data of MGA and MGD are presented in Table S3.

479 Figure 4. Proposed Hphpa and Hphpol biosynthesis. 3-hydroxy-*m*-hydroxymethyl-*p*-
480 hydroxyphenylpyruvate is synthesized as a precursor of both Hphpa and Hphpol via
481 several steps from isochorismate. The resulting intermediate is further transformed by
482 either an aminotransferase (AgdL) or a reductase (AgdN or U) to yield Hphpa or Hphpol,
483 respectively and further transferred to the free-standing PCP (AgdB) after adenylation
484 by AgdA. The AgdL enzyme for which natural mutants were identified in the course of
485 this study is highlighted in red.

486 Figure 5. Distribution of the known and unknown BGCs in the frame of the phylogeny of
487 the 23 *Microcystis* genomes based on maximum likelihood tree built upon 586 marker
488 genes. The known BGCs are involved in the synthesis of aeruginoguanidine (AGD) and
489 microguanidine (MGD) and/or MGA only, of microcystin, of cyanobactins including
490 aeruginosamide, of aeruginosin, of microviridin, of cyanopeptolin, of anabaenopeptin
491 including ferintoic acid, and of microginin. One BGC only predicted in one strain is
492 involved in synthesis of puwainaphycin.[12] The numbers indicate the unknown BGCs
493 detected in the genome; the origin of each strain is indicated in parenthesis.

494 Figure 6. Proposed AGD, MGA and MGD biosynthetic pathways. **Top line**; AGD
495 biosynthesis route: Hphpa which is linked to the free-standing PCP AgdB is transferred
496 to AgdE and condensed with the dipeptide, which is derived from AgdK and E. The
497 thioester-tethered intermediate is methylated by a radical SAM enzyme (AgdC) followed
498 by decarboxylation and released from the enzyme. The resulting molecule is further
499 modified by sulfation and farnesylation. **Middle line**: MGA (**6** and **7**) biosynthesis route,
500 almost the same pathway as AGD biosynthesis, but only AgdE is used and the α-amino
501 group of Arg is permethylated by AgdI or M. **Bottom line**; MGD (**4** and **5**) biosynthesis
502 route, almost the same pathway as MGAs, but using Hphpol as the intermediate instead
503 of Hphpa. The enzyme AgdK for which a natural mutant was identified in the course of
504 this study is highlighted in red.

505

506 **TABLE**

507 Table 1. Proposed function of proteins encoded in the AGD gene cluster and flanking
508 ORFs in *Microcystis aeruginosa* NIES-98. The strand position and the size of gene in
509 amino acids are indicated with the corresponding Best BLASTp hit and identity, all
510 found in *Microcystis* genomes. NRPS domains: C for condensation, A for adenylation with
511 substrate prediction, PCP for peptidyl carrier protein, and nMT for N-methyl transferase.

512

ASSOCIATED CONTENT

Supporting Information

The supporting Information is available free of charge *via* the ACS Publications website at DOI

Methods of preparation of the extracts and of recovering complete AGD cluster, HLPC-MS measurement and molecular networking; four supporting tables and 18 supporting figures on the detailed molecular network and the spectra of the new structures, as indicated in the text (PDF).

Accession Codes

New sequence data are archived in GenBank under accession numbers MH049490 to MH049500.

AUTHOR INFORMATION

Corresponding Author

* email: muriel.gugger@pasteur.fr

* email: editt@uni-potsdam.de

ORCID

Muriel Gugger: 0000-0001-6728-1976

Enora Briand: 0000-0001-8996-0072

Elke Dittmann: 0000-0002-7549-7918

Christian Hertweck: 0000-0002-0367-337X

Douglas Gatte Pichi: 0000-0001-9164-8969

Thibault Scalvenzi: 0000-0002-5760-1574

Notes

The author declare no competing financial interest

Table 1. Proposed function of proteins encoded in the AGD gene cluster and flanking ORFs in *Microcystis aeruginosa* NIES-98. The strand position and the size of gene in amino acids are indicated with the corresponding Best BLASTp hit and identity, all found in *Microcystis* genomes. NRPS domains: C for condensation, A for adenylation with substrate prediction, PCP for peptidyl carrier protein, and nMT for N-methyl transferase.
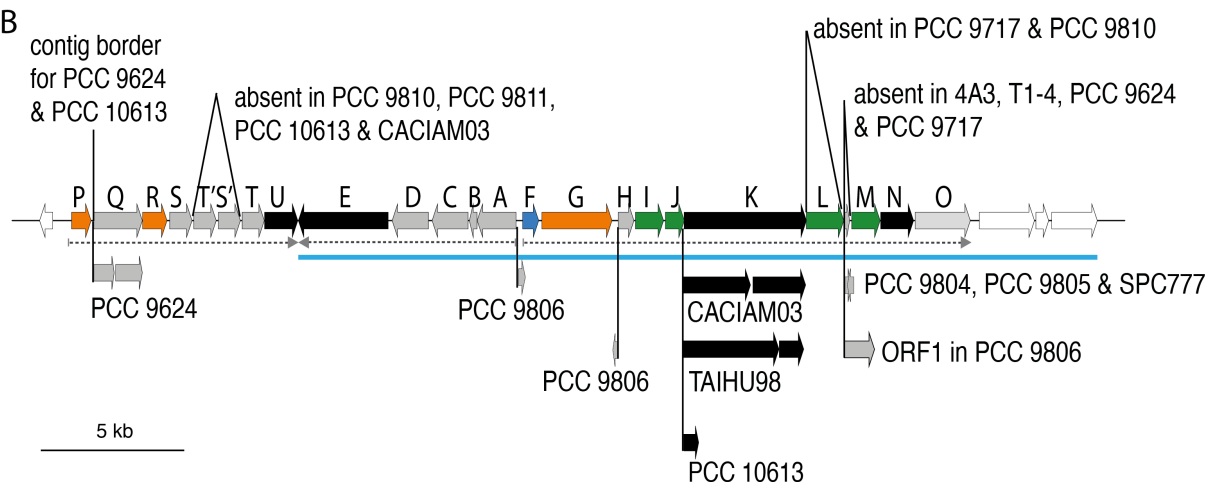
Gene (Strand)	Size (aa)	Proposed function (NRPS with substrat prediction)	Best BLASTp hit (Accession number)	Identity (%)
Orf (-)	160	Conserved protein of unknown function	Hypothetical protein O53_4696 (ELP52967.1)	100
agdP (+)	238	Carbohydrate sulfotransferase II	Hypothetical protein O53_4419 (ELP52967.1)	100
agdQ (+)	589	Thiamine pyrophosphate enzyme	Acetolactate synthase large subunit (EPF22845.1)	100
agdR (+)	296	Sulfotransferase I	Sulfotransferase domain protein (ELP52945.1)	100
agdS (+)	271	Conserved protein of unknown function	Hypothetical protein O53_4433 (ELP52708.1)	95
agdT (+)	274	Conserved protein of unknown function	Conserved hypothetical protein (CCH98454.1)	99
agdS' (+)	268	Conserved protein of unknown function	Hypothetical protein MAESPC_01420 (EPF22841.1)	99
agdT' (+)	270	Conserved protein of unknown function	Putative uncharacterized ORF3 domain protein (ELP52673.1)	99
agdU (+)	405	Thioester reductase	Polyketide synthase hetM (CCI12982.1)	98
agdE (-)	1093	NRPS ($A_{Arg,/Lys/Orn}$-PCP-C)	Linear gramicidin synthase subunit D (EPF22838.1)	98
agdD (-)	441	Sulfotransferase III	Zinc chelation protein SecC (WP_069474152.1)	100
agdC (-)	438	Radical SAM	Radical SAM superfamily protein (ELP52520.1) putative oxygen-independent coproporphyrinogen III synthase	100
agdB (-)	94	Peptidyl carrier protein	Phosphopantetheine attachment site family protein (ELP52599.1)	100
agdA (-)	473	AMP-dependent synthetase and ligase	AMP-dependent synthetase (WP_069474153.1)	100
agdF (+)	196	Permease	Conserved hypothetical protein (CCI31673.1)	97
agdG (+)	852	Sulfatase	Sulfatase family protein (ELP52537.1)	99
agdH (+)	191	4-Hydroxybenzoate synthetase	Hypothetical protein O53_4514 (ELP52787.1)	100
agdI (+)	342	O-Methyltransferase	Methyltransferase (WP_069474155.1)	100
agdJ (+)	231	Isoprenyl-transferase	Di-trans,poly-cis-decaprenylcistransferase (ELP52925.1)	99
agdK (+)	1588	NRPS (A_{Arg}-nMT-PCP-C)	Chondramide synthase cmdD (EPF22828.1)	99
agdL (+)	455	Aminotransferase	Uncharacterized aminotransferase yodT (CCI31679.1)	99
Orf (+)	71	Hypothetical protein	Hypothetical protein (WP_069474158.1)	100
agdM (+)	346	O-Methyltransferase	O-Methyltransferase family protein (ELP53140.1)	99
agdN (+)	401	Thioester reductase	Thioester reductase domain protein (ELP52682.1)	99
agdO (+)	671	ABC transporter	ABC Transporter transmembrane region 2 family protein (ELP52531.1)	99
Orf (+)	671	Conserved protein of unknown function	Hypothetical protein O53_4447 (ELP52722.1)	99
Orf (+)	156	Conserved protein of unknown function	Hypothetical protein O53_4299 (ELP52574.1)	100
Orf (+)	554	GUN4-like family protein	Hypothetical protein (WP_069474163.1)	100

Figure 1. Aeruginoguanidines and the corresponding biosynthetic gene cluster. *(A)* The structure of aeruginoguanidines (AGDs), **1**; AGD-98A, **2**; AGD-98B, **3**; AGD-98C; *(B)* AGD biosynthetic gene cluster of *Microcystis aeruginosa* NIES-98 and its variation in 13 other *Microcystis* genomes sharing 94 to 98% of similarity. The genes are color-coded with orange for carbohydrate sulfotransferase, sulfotransferase and sulfatase; black for NRPS and thioesterase; green for methyl-, isoprenyl- and aminotransferase; blue for permease; grey for proteins with putative and unknown function. The dashed arrows under the cluster indicate the three operons. The blue line indicates the span of the MIC2 cluster previously described.[9]

Figure 2. Molecular network of AGD (A) and of MGD with MGA (B). Characterized structural variants are indicated as red-colored nodes and new congeners charactized in this study are highlighted in green. Diversity and distribution of AGD and MGD variants for each strain. Details of MN(+), MN(−), and the AGD and MGD networks are presented in Figures S1-S3.

| Parent Mass	Strains										Assignment
m/z (M-H)⁻	NIES_98	PCC 9624	PCC 9717	PCC 9804	PCC 9805	PCC 9806	PCC 9810	PCC 9811	PCC 10613	T1-4	
667.038						■					
844.791											
913.007	■			■	■			■		■	AGD-98B (2)
926.987							■				
927.186										■	
928.061			■								
941.98						■					
949.877											
967.323	■		■	■							
981.125	■	■	■	■		■				■	AGD-98A (1)
982.146						■					
983.223											
995.049				■	■						
997.141	■			■	■		■			■	AGD-98C (3)
1009.17				■							
1035.09				■						■	
1037.02				■						■	
1049.11				■						■	
1065.08				■						■	
1143.15				■						■	
703.775			■	■		■					MGD-704 (5)
771.028	■		■	■		■	■		■		MGA-771 (6)
771.736			■	■		■					MGD-AL772 (4)
787.053			■	■		■					MGA-787 (7)
804.336			■	■							

Figure 3. Microguanidine and short aeruginoguanidine variants detected in strains containing the AGD biosynthetic gene cluster. The MGD depsipeptides contain ester-bonds, while the peptidic sAGDs contain amide-bonds in their structures. Details of the high-resolution MS data of sAGD and MGD are presented in Table S3.
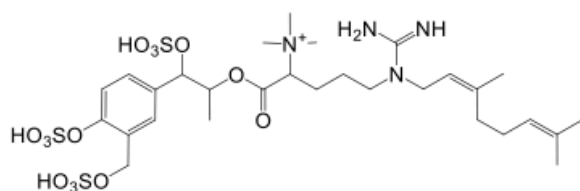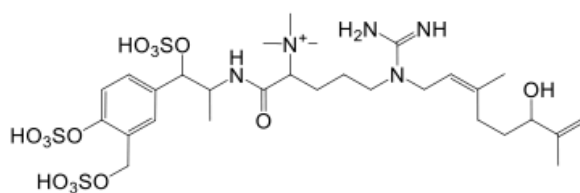

Microguanidine AL772 (4)
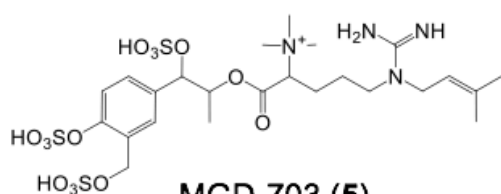
MGD-703 (5)
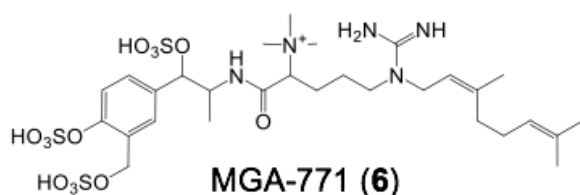
MGA-771 (6)

MGA-787 (7)

Figure 4. Proposed Hphpa and Hphpol biosynthesis. 3-hydroxy-*m*-hydroxymethyl-*p*-hydroxyphenylpyruvate is synthesized as a precursor of both Hphpa and Hphpol via several steps from isochorismate. The resulting intermediate is further transformed by either an aminotransferase (AgdL) or a reductase (AgdN or U) to yield Hphpa or Hphpol, respectively and further transferred to the free-standing PCP (AgdB) after adenylation by AgdA. The AgdL enzyme for which natural mutants were identified in the course of this study is highlighted in red.
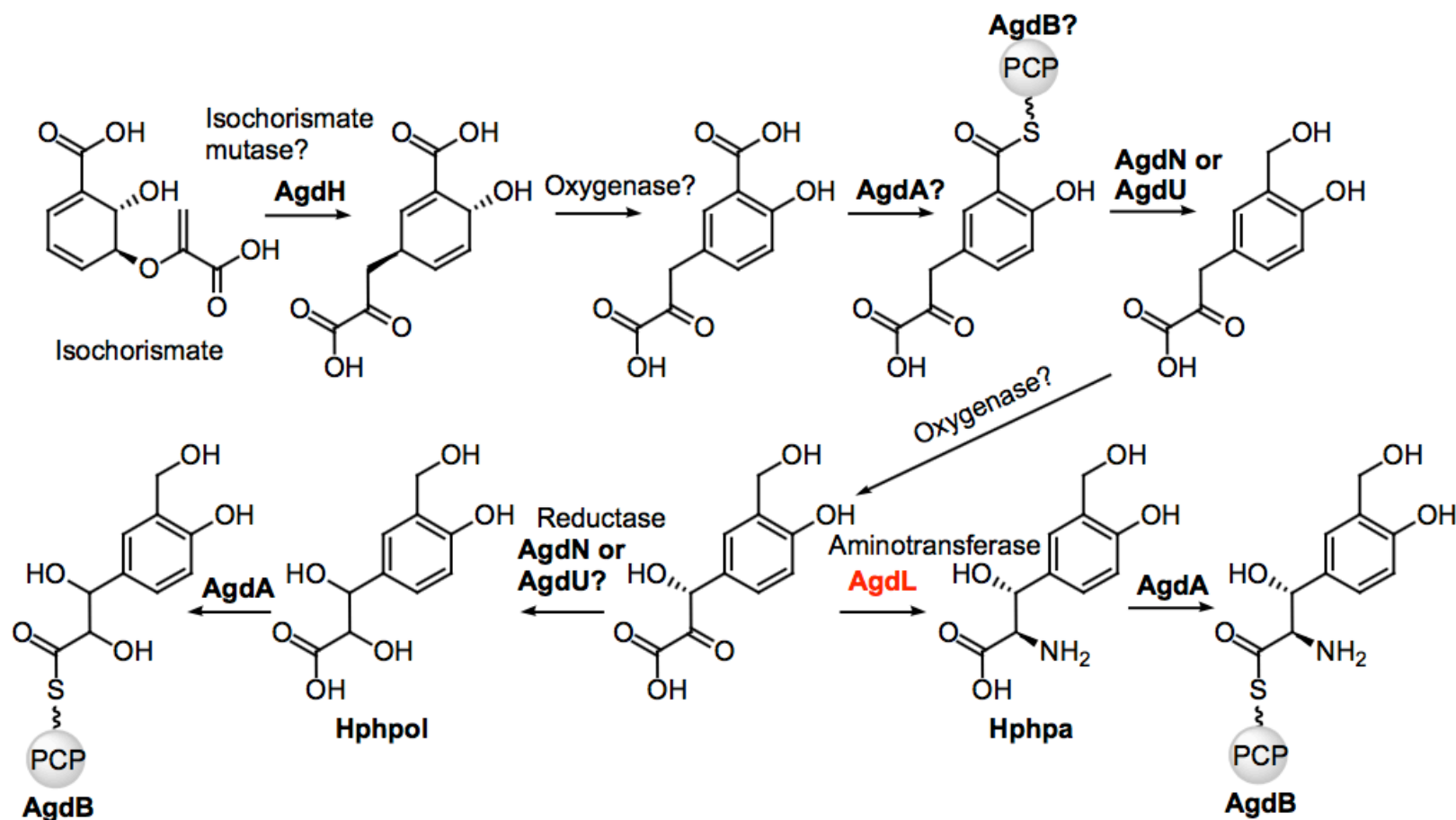
Figure 5. Distribution of the known and unknown BGCs in the frame of the phylogeny of the 23 *Microcystis* genomes based on maximum likelihood tree built upon 586 marker genes. The known BGCs are involved in the synthesis of aeruginoguanidine (AGD) and microguanidine (MGD) and/or microguanidine amide (MGA) only, of microcystin, of cyanobactins including aeruginosamide, of aeruginosin, of microviridin, of cyanopeptolin, of anabaenopeptin including ferintoic acid, and of microginin. One BGC only predicted in one strain is involved in synthesis of puwainaphycin.[12] The numbers indicate the unknown BGCs detected in the genome; the abbreviations in parenthesis after the name of the strain indicate its origin (Table S2).
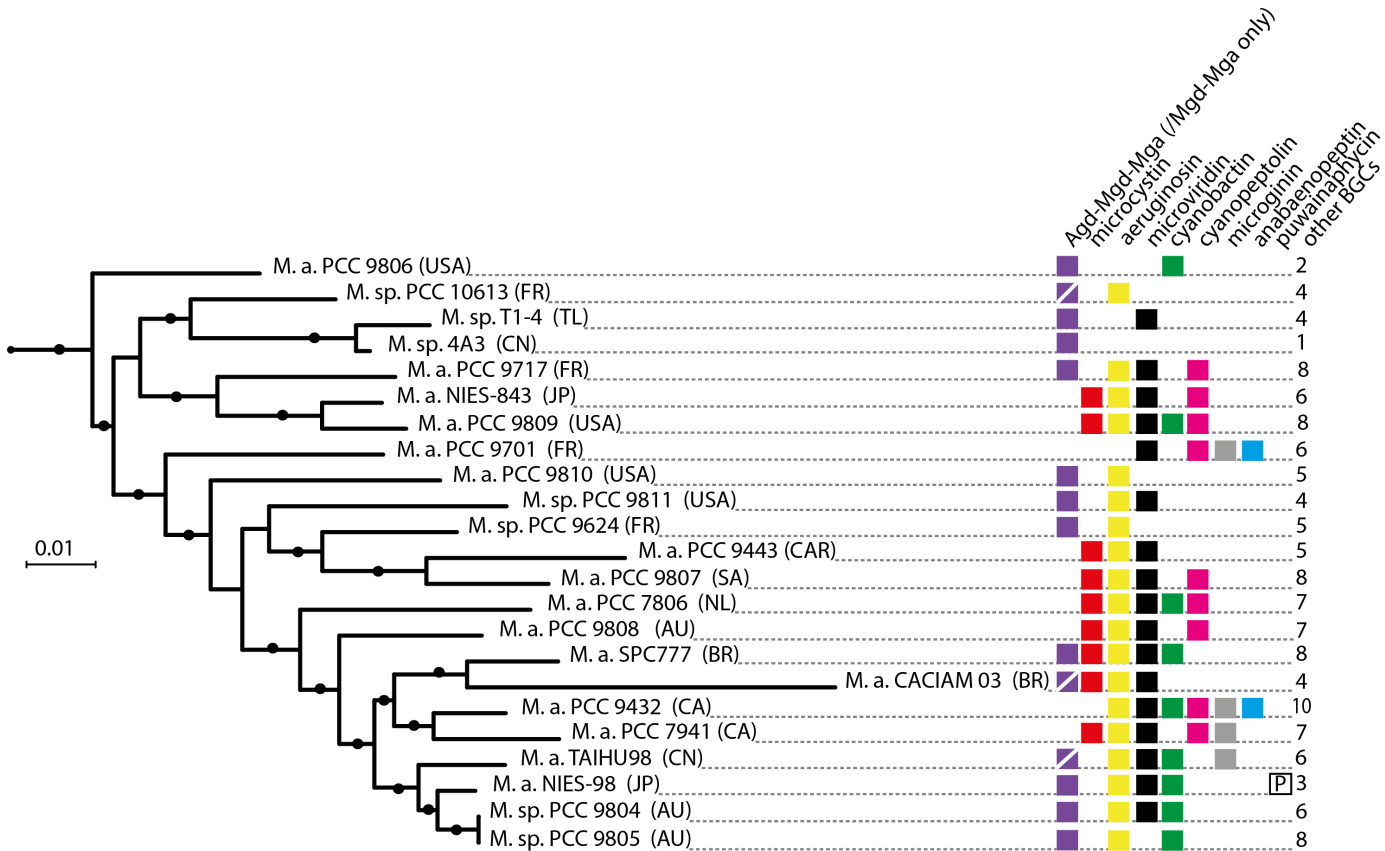
Figure 6. Proposed AGD, MGA and MGD biosynthetic pathways. **Top line**; AGD biosynthesis route: Hphpa which is linked to the free-standing PCP AgdB is transferred to AgdE and condensed with the dipeptide, which is derived from AgdK and E. The thioester-tethered intermediate is methylated by a radical SAM enzyme (AgdC) followed by decarboxylation and released from the enzyme. The resulting molecule is further modified by sulfation and farnesylation. **Middle line**: MGAs (**6** and **7**) biosynthesis route, almost the same pathway as AGD biosynthesis, but only AgdE is used and the α-amino group of Arg is permethylated by AgdI or M. **Bottom line**; MGD (**4** and **5**) biosynthesis route, almost the same pathway as sAGDs, but using Hphpol as the intermediate instead of Hphpa. The enzyme AgdK for which a natural mutant was identified in the course of this study is highlighted in red.
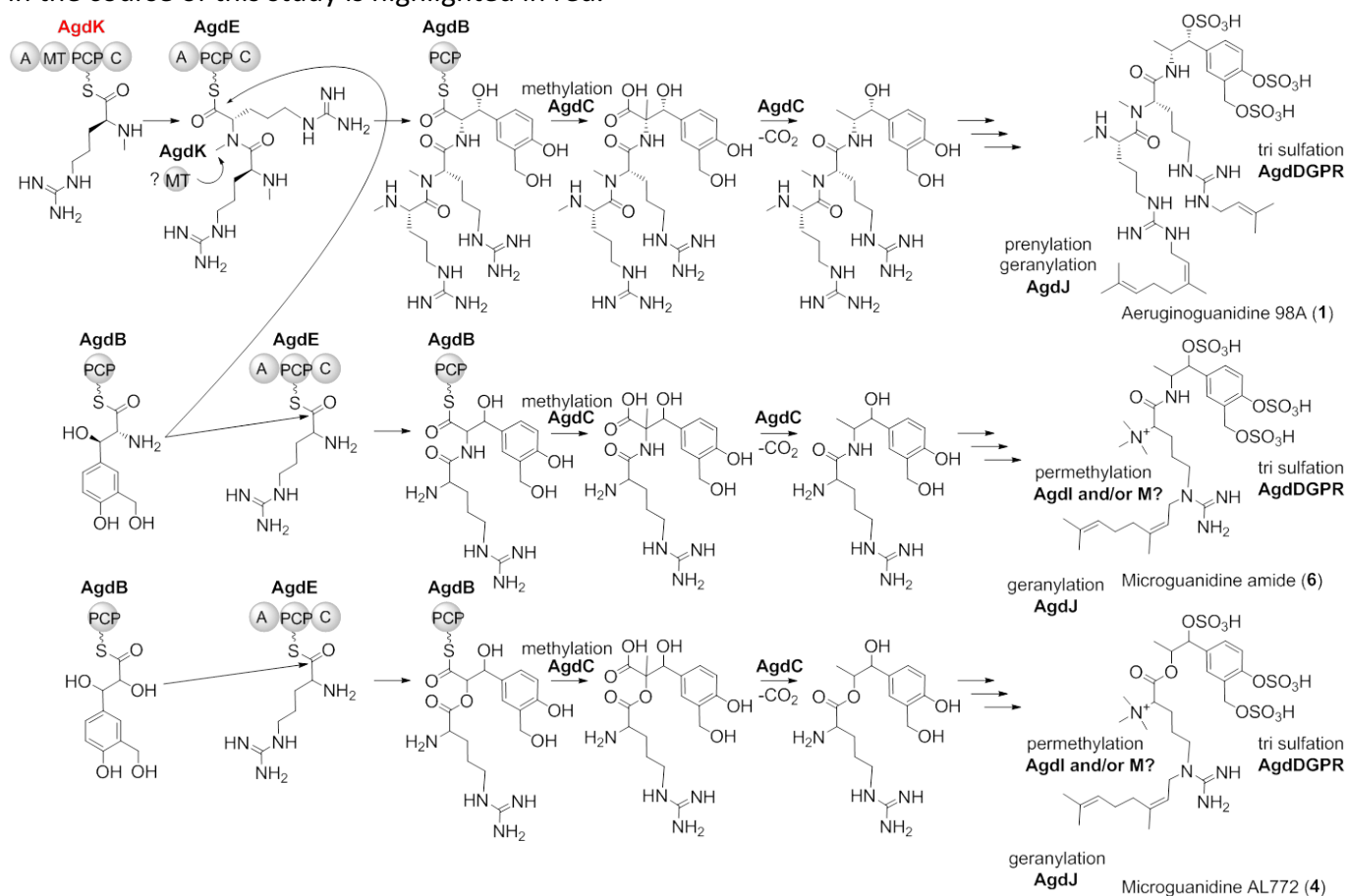
A unique biosynthetic pathway in bloom-forming cyanobacteria jointly assembles cytotoxic aeruginoguanidines and microguanidines

Claire Pancrace, Keishi Ishida, Enora Briand, Douglas Gatte Pichi, Annika R. Weiz, Arthur Guljamow, Thibault Scalvenzi, Nathalie Sassoon, Christian Hertweck, Elke Dittmann, Muriel Gugger

Supplemental Information

Additional materials and methods

Microcystis cultures for nucleic acid extraction, and chemical analysis. Nucleic acid extraction of cyanobacterial cells to obtain DNA were carried out as previously described[1]. For HPLC, MS and MS/MS analyses, cell pellets were centrifuged, rinsed with sterile water, flash frozen and lyophilized until further processing.

PCR screening for AGD cluster. Primer pairs targeting putative hydroxybenzoate synthase and prenyltransferase of MIC2 gene cluster were designed to amplify a 563b-long amplicon with 1F_agdH/1R_agdH, and a 686b-long amplicon with 2F_agdJ/2R_agdJ (Table S1). These two genes are detected concomitantly only in *Microcystis* strain containing this pathway. Screening of 30 *Microcystis* strains available at the PCC (http://cyanobacteria.web.pasteur.fr/) was performed by PCR using LA Taq TAKARA. PCR program was as follow: initial denaturation 2 min at 95°C, 35 cycles consisting of 30 s at 95°C, 30 s at 60°C for primer pair 1F_agdH/1R_agdH and 58°C for 2F_agdJ/2R_agdJ, and 1 min at 72°C, followed by a final elongation step 10 min at 95°C. Amplicons were visualized under UV light after electrophoresis on 1.5% agarose gel.

Genome sequencing. For the strains suspected to carry the Agd gene cluster, the whole genome sequencing was carried out using the Nextera XT DNA sample preparation kit (Illumina) for 2x150 bps paired-ends reads (insert size ~300 bps). All sequenced paired-ends reads were clipped and trimmed with AlienTrimmer[2] (v. 0.4.0), and subjected to a sequencing error correction with Musket[3] (v. 1.1) as well as a digital normalization procedure with khmer[4] (v. 1.3). For each sample, remaining processed reads were assembled with SPAdes[5] (v. 3.7.0).

Phylogenetic analysis. The species tree generated by a concatenation of 586 conserved proteins was performed as follow: Ambiguous and saturated regions were removed with BMGE v1.1242 (with the gap rate parameter set to 0.5). A Maximum-Likelihood phylogenetic tree was generated with the alignment using RAxML v7.4.343 with the LG amino acid substitution model. The genomes of *Cyanothece* sp. PCC 7422 and PCC 7822 were used as outgroup in order to root the phylogenetic tree with the closest relatives of the *Microcystis* in a cyanobacterial phylum wide phylogeny[6].

HPLC-MS measurement. LC-MS/MS measurements were carried out by Bruker HCT Ultra ion trap mass spectrometry (Bruker Daltonics) coupled with an Agilent Technologies 1100 series liquid chromatogram system (Agilent) consisting of binary pump G1312A, two degassers G1322A/G4225, well-plate sampler G1367A, diode array detector G1315A, and column thermostat G1316A. The ionization mode was electro-spray (ESI), polarity positive and negative separately, mass range mode ultra-scan, and nitrogen was used as a drying and nebulizer gas. The following parameters were applied: nebulizer 70 psi, dry gas 12 L/min, dry temperature 365 °C, scan range m/z 300–2000, No-of precursor ions 2. Ten μL of samples were subjected to a reversed-phase HPLC column Symmetry Shield RP18 (Waters, 3.5 μm, 4.6 × 100 mm) using a gradient system; solvent A; water containing 0.1% formic acid, solvent B; acetonitrile, 10%B for 10 min to 99%B in 25 min and kept 99%B for 4 min, to 10%B in 1 min.

The HR-LCMS measurements were performed by HPLC-HRMS series of Thermo Accela (LC) and Thermo Exactive (HRMS), an ESI source operating in both polarity mode and an orbitrap analyzer (Thermo Fisher Scientific). Five μL of samples were subjected to a reversed-phase HPLC column Betasil C18 (Waters, 3.0 μm, 2.1 × 150 mm) using a gradient system; solvent A; water containing 0.1% formic acid, solvent B; acetonitrile, 10%B for 2 min to 99.5%B in 20 min and kept 99.5%B for 7 min, to 10%B in 1 min.

Molecular networking. Two molecular networks (MNs) were performed, one with LC-MS/MS data in positive mode (MN(+)) and the second one with negative mode data (MN(−)). The following steps were done for both MNs. LC-MS/MS data from *Microcystis* strains and AGD A, B and C standards were converted to mzXML format using MSConvert, part of the ProteoWizard package[7] and were subjected to the molecular

networking workflow of Global Natural Products Social Molecular Networking web site[8] (GNPS at http://gnps.ucsd.edu) using the Group Mapping feature. The input data were searched against annotated reference spectra of the MS2 library within GNPS. Computationally, the algorithms compare MS2 spectra by their similarity and assign similarity scores[9]. For the networks presented in this paper, the parent mass peak tolerance was set to 2 Da and the ion tolerance for mass fragments was set to 0.95 Da. Pairs of consensus spectra were aligned if both spectra fell within the top 10 alignments for each of the respective spectra, the cosine of their peak match scores was ≥ 0.7 and the minimum matched peaks was 6. The maximum size of connected components allowed in the network was 100 and the minimum number of spectra to form a cluster was 2. For visualization, the created molecular networks were imported into the program Cytoscape[10] 2.8.3. Each node was labeled with their respective parent mass. The edges between nodes indicated the level of similarity between nodes, with thicker lines indicating higher similarity. Nodes created by solvent background were removed from the network. Each node that corresponded to detection of unclear or trace ions potentially related to AGD and MGD cluster of MN was confirmed by Thermo Exactive HR-HPLC and further validated running a fresh independent extraction through Bruker LC-MS/MS.

Extraction of cyanobacterial cells and isolation of MGD AL772 (4) and shortAGD (6). Lyophilized cells of *M. aeruginosa* PCC 9624 (132 mg), PCC 9804 (507, 400, 100, 355 mg), PCC 9805 (71 mg), PCC 9806 (34 mg), PCC9810 (79 mg), PCC 9811 (200, 40, 238 mg), PCC 10108 (191 mg) were extracted with 80% aqueous methanol (*v/v*, 40 mL) using a sonicator (Sonoplus MS73, Bandelin, 30% power, 5 cycles for 2 min at room temperature), respectively. Each extract was centrifuged at 8,000 × *g* for 15 min at 15 °C. The residues were further extracted with 80% aqueous methanol (*v/v*, 40 mL), respectively, as the above-mentioned procedure. The extracts were directly subjected to solid phase extraction Chromabond C18ec (1000 mg, Macherey-Nagel) and eluted with 80% aqueous methanol (*v/v*, 30 mL), respectively. Each of the flow-through and eluted fractions were combined and concentrated under a reduced pressure. The resulting residue was dissolved in *N,N*-dimethylformamide and filtered. This crude extract was subjected to reversed-phase HPLC (Phenomenex fusion RP, particle size 5 μm, pore size 80Å, 21.2 × 250 mm, Phenomenex) using a gradient system: solvent A, water containing

0.1% trifluoroacetic acid (TFA), solvent B, 83% aqueous acetonitrile (v/v), 20%B for 10 min, to 100%B in 30 min, at a flow rate 12 ml min^{-1}. Obtained fractions containing **4** and **6** were subjected to reversed-phase HPLC (Phenomenex fusion RP, particle size 5 μm, pore size 80Å, 10 × 250 mm, Phenomenex) using a gradient system: solvent A, water containing 0.1% TFA, solvent B, 83% aqueous acetonitrile (v/v), 10%B for 10 min, to 30%B in 10 min and kept for 30 min, at a flow rate 6 ml min^{-1}, respectively. The main fractions containing **4** and **6** were subjected to reversed-phase HPLC (Phenomenex Luna C18, particle size 10 μm, pore size 100Å, 4.6 × 250 mm, Phenomenex) using a gradient system: solvent A, water containing 0.1% formic acid, solvent B, acetonitrile, 0.5%B for 2 min, to 99.5%B in 20 min, at a flow rate 1 ml min^{-1} to yield crude **4** and **6**, respectively. These crudes **4** and **6** were further subjected to reversed-phase HPLC (Nucleodur sphinx, particle size 5 μm, pore size 100Å, 4.6 × 250 mm, Phenomenex) using a gradient system: solvent A, water containing 0.1% TFA, solvent B, acetonitrile, 5%B for 15 min, to 25%B in 5 min, to keep 25 min, to 99%B in 5 min) at a flow rate 1 ml min^{-1} to yield **4** (ca. 300 μg) and **6** (ca. 500 μg), respectively. NMR spectra of obtained peptides were measured on Bruker Avance 600 MHz spectrometers with cryo probe in DMSO-d_6. Spectra were referenced to the residual solvent peak.

Table S1. Primers used in this study. The list of primer pairs are indicated on the genetic locus scheme below, with primer in red used to detect the AGD locus, in blue primer to close gaps in genomic data.

Name	Sequence 5'-3'	Expected amplicon size
1F_agdH	CCAGCGAAACCAGCGAATCG	563
1R_agdH	GACGAAATAACTCTCAGGAAATT	
2F_agdJ	ACTAACCAACATCTCTACTAAAC	686
2R_agdJ	TTTTCCAAAGCGACGCTC	
3F_agdU	TAACAGAGCTATCTATCTCCTGTC	218
3R_agdE	TAACCGAGATTTCATGCAGATA	
3F_agdK	GTTCAACAGGAGATGCTTGCTG	1500-1662/3725$_a$
3R_agdM	ATAATCGAGATGTGGAAGGCAT	
4F_agdE	ATTCTCCTCAATTGGCTGTAAT	1344
4R_agdD	ACAGTTTAGCTCAGGTCCCACT	
5F_agdP	AACATCGTGATTATCGAGAATA	706
5R_agdQ	TCAGCATAAGCTGAGGCTAATC	
6F_agdR	TTGTCAACCATTATGTCAAGAG	1858/3616$_b$
6R_agdU	GTTGAGTCACAGGTTTAGTCAT	
7F_agdG	ACCGGTAAGGGCAGTAATGGCA	2276
7R_agdH	TGGAGTGTGCTTAACTCCGAA	

a. amplicon size in PCC 9810, in PCC 9717 and in PCC 9806
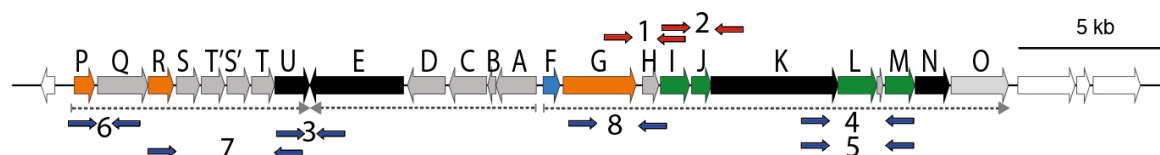*b.* Primer pair 6 is targeting gene duplication.

Table S2. *Microcystis* strains or genomes studied. 19 strains were cultured for metabolomics investigations.

Microcystis	Origin	Genome accession	Refs	Biomass analyzed
PCC 7806	The Netherlands, 1972	AM778843–958	[11]	+
PCC 7941	Ontario, Canada, 1954	CAIK00000000	[1]	+
PCC 9432	Canada, 1954	CAIH00000000	[1]	+
PCC 9443	Central African Republic, 1994	CAIJ00000000	[1]	+
PCC 9624	Seine, France, 1996		This study	+
PCC 9701	Guerlesquin, France, 1996	CAIQ00000000	[1]	+
PCC 9717	Rochereau, France, 1996	CAII00000000	[1]	+
PCC 9804	Camberra, Australia, 1985		This study	+
PCC 9805	Camberra, Australia, 1985		This study	+
PCC 9806	Oskosh, USA, 1975	CAIL00000000	[1]	+
PCC 9807	Pretoria, South Africa, 1973	CAIM00000000	[1]	+
PCC 9808	New South Wales, Australia, 1972	CAIN00000000	[1]	+
PCC 9809	Wisconsin, USA, 1982	CAIO00000000	[1]	+
PCC 9810	Alabama, USA, 1982		This study	+
PCC 9811	Wisconsin, USA, 1982		This study	+
PCC 10613	Orsonville, France, 2006		This study	+
4A3	Wuhan, China		This study	
CACIAM 03[a]	Tucuruí reservoir, Pará, Brazil,	MCIH00000000	[12]	
T1-4	Bangkok, Thaïland	CAIP00000000	[1]	+
NIES-98	Lake Kasumigaura Ibaraki, Japan, 1982	MDZH00000000	[13]	+
NIES-843	Lake Kasumigaura Ibaraki, Japan, 1997	AP009552.1	[14]	+
SPC777[a]	Billings reservoir, Sao Paulo, Brazil	ASZQ00000000	[15]	
TAIHU98[a]	Lake Taihu, China, 1997	ANKQ00000000.1	[16]	

a. Genome only

Table S3. High resolution MS data of microguanidine AL772 and its new congeners observed by Thermo Exactive (OrbiTrap) LCMS

Microguanidine	[M-H]$^-$ found	[M-H]$^-$ calculated	Element composition
AL772 (**4**)	771.2261	771.2245	$C_{29}H_{47}O_{14}N_4S_3$
5	703.1654	703.1619	$C_{24}H_{39}O_{14}N_4S_3$
6	770.2430	770.2405	$C_{29}H_{48}O_{13}N_5S_3$
7	786.2377	786.2354	$C_{29}H_{48}O_{14}N_5S_3$

Table S4. ^{1}H and ^{13}C NMR data of MGA-771 (6) and MGD AL772 (4) in DMSO-d_6

	MGA-771 (**6**)		MGD AL772 (**4**)	
Position	δ_C (mult)	δ_H (J=Hz)	δ_C (mult)	δ_H (J=Hz)
1	125.1 (d)	7.35 (brs)	125.7 (d)	7.38 (brs)
2	129.5 (s)		129.7 (s)	
3	149.1 (s)		149.3 (s)	
4	120.1 (d)	7.21 (d 8.5)	120.4 (d)	7.24 (d 8.6)
5	125.3 (d)	7.13 (dd 8.5, 2.0)	126.0 (d)	7.16 (m)
6	135.1 (s)		133.5 (s)	
7	78.3 (d)	5.11 (d 4.3)	77.4 (d)	5.16 (m)
8	49.8 (d)	4.23 (m)	75.3 (d)	5.21 (m)
9	17.9 (q)	1.06 (d 6.6)	15.6 (q)	1.17 (d 6.4)
10	62.7 (t)	4.78 (d 13.6), 4.88 (d 13.6)	62.8 (t)	4.84 (d 14.0), 4.92 (d 14.0)
11		8.48 (d 9.4)	-	-
12	nd		166.3 (s)	
13	72.5 (d)	3.93 (m)	73.2 (d)	4.11 (dd 11.5, 3.4)
14	23.1 (d)	1.56 (m), 1.73 (m)	23.0 (t)	1.86 (m), 1.96 (m)
15	22.4 (d)	1.43 (m), 1.78 (m)	23.2 (t)	1.36 (m), 1.50 (m)
16	47.0 (t)	3.22 (m)	46.9 (t)	3.24 (m)
18	nd		155.6 (s)	
19		nd		nd
20		nd		nd
22,22',22''	51.4 (q)	2.86 (s)	51.5 (q)	3.03 (s)
23	45.6 (d)	3.88 (m)	46.0 (t)	3.90 (m)
24	118.8 (d)	5.06 (m	119.0 (d)	5.08 (m)
25	140.4 (d)		140.3 (s)	
26	31.6 (t)	2.04 (m)	31.5 (t)	2.05 (m)
27	26.1 (t)	2.03 (m)	25.9 (t)	1.41 (m), 2.03 (m)
28	123.4 (d)	5.08 (m)	123.7 (d)	5.09 (m)
29	131.4 (s)		131.4 (s)	
30	17.8 (q)	1.57 (s)	17.6 (q)	1.57 (s)
31	23.1 (q)	1.70 (s)	23.0 (q)	1.70 (s)
32	25.4 (q)	1.64 (s)	25.5 (q)	1.64 (s)

nd: not determined.

Figure S1. Molecular network derived from positive mode (A) and negative mode (B) mass spectrometric analysis of extracts of the 19 *Microcystis* strains and the three AGD standards. Red nodes indicate consensus MS/MS spectra to compounds in a MS/MS library of known compounds. The respective name of identified class of compounds or molecule is given next to the black square. MCs: microcystins, Cya: cyanopeptolin, Fer: ferintoic acid, Aeg: aeruginosamide, AGD: aeruginoguanidine, and MGD: microguanidine.

MCs

AGD

Cya

Fer

AegC

AegB

AGD-98C

Figure S1. A

AGD

MGD
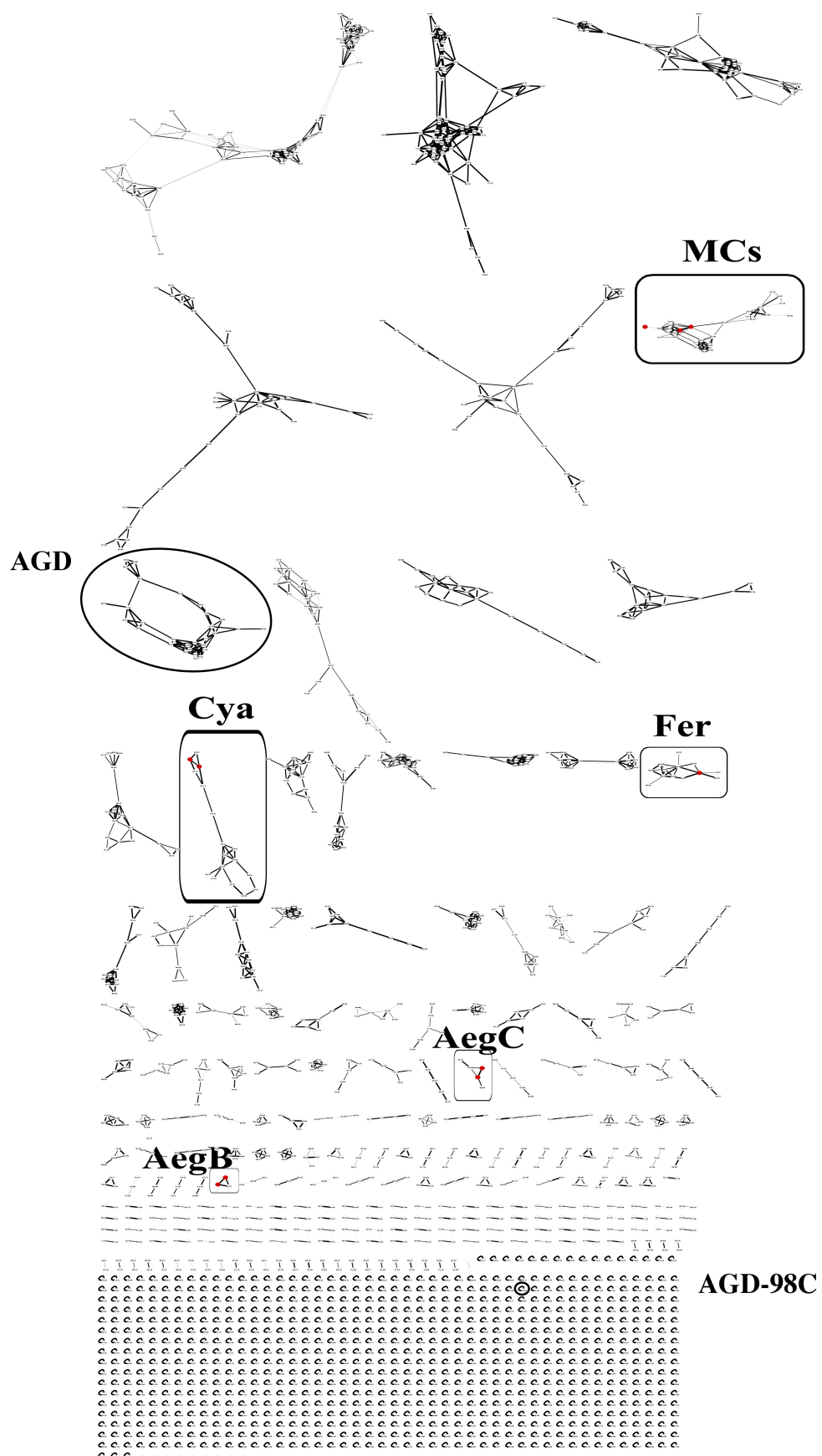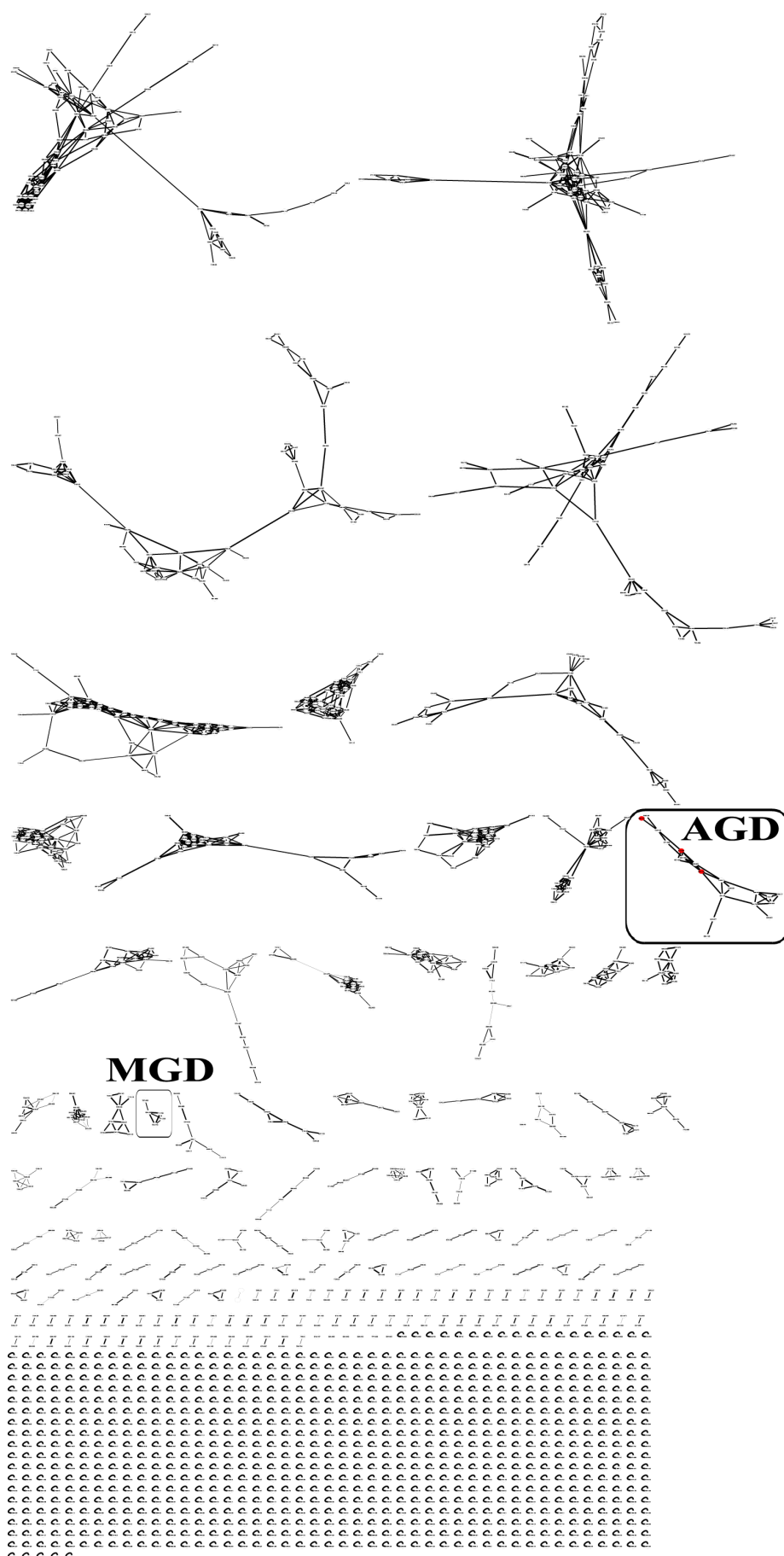
Figure S1. B

Figure S2. Molecular network of aeruginoguanidine molecular family and table of detected analogues in AGD standards and *Microcystis* PCC strains.

AGD-98C · AGD-98B · AGD-98A

(Molecular network nodes: 1065.08, 997.141, 1037.02, 677.038, 913.007, 927.168, 967.323, 926.987, 981.125, 1049.11, 1035.09, 995.049, 1009.17, 1143.15, 941.98, 983.223, 844.79, 928.061, 943.877, 982.146)

Parent mass	Assignment	Standard			Strains with candidate AGD gene cluster									
		AGD-98A	AGD-98B	AGD-98C	NIES-98	PCC 9624	PCC 9717	PCC 9804	PCC 9805	PCC 9806	PCC 9810	PCC 9811	PCC 10613	T1-4
677.038														
844.791								■	■					
913.007	AGD-98B[M-H]⁻		■					■	■			■		■
926.987														
927.186												■		■
928.061							■				■			
941.98											■			
949.877														
967.323														
981.125	AGD-98A[M-H]⁻	■			■	■		■		■				■
982.146										■				
983.223														
995.049					■			■	■					■
997.141	AGD-98C[M-H]⁻			■	■			■	■					■
1009.17														■
1035.09								■	■					■
1037.02								■	■	■				■
1049.11														■
1065.08								■	■					■
1143.15														■

Figure S3. Molecular network of microguanidine molecular family and table of detected analogues in AGD standards and *Microcystis* PCC strains.

787.053 — **804.336** — **771.736** MGD AL772 — **771.028** — **703.775**
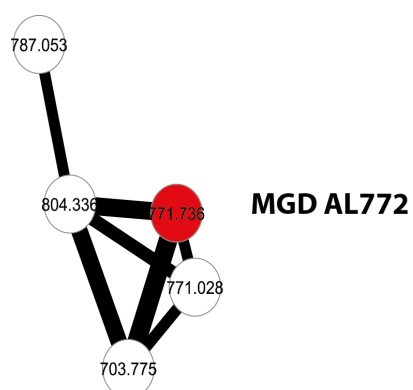
Parent mass	Assignment	Standard			Strains with candidate AGD gene cluster									
		AGD-98A	AGD-98B	AGD-98C	NIES-98	PCC 9624	PCC 9717	PCC 9804	PCC 9805	PCC 9806	PCC 9810	PCC 9811	PCC 10613	T1-4
703.775							■	■	■	■	■	■	■	■
771.028					■			■	■			■		■
771.736	MGD-AL772				■	■		■	■			■		■
787.053								■	■			■		■
804.336							■	■	■	■	■	■	■	■

Figure S4. Negative MS/MS spectrum obtained by OrbiTrap of microguanidine AL772 (A), of the new MGD 5 (B), and MGAs 6 (C) and 7 (D).
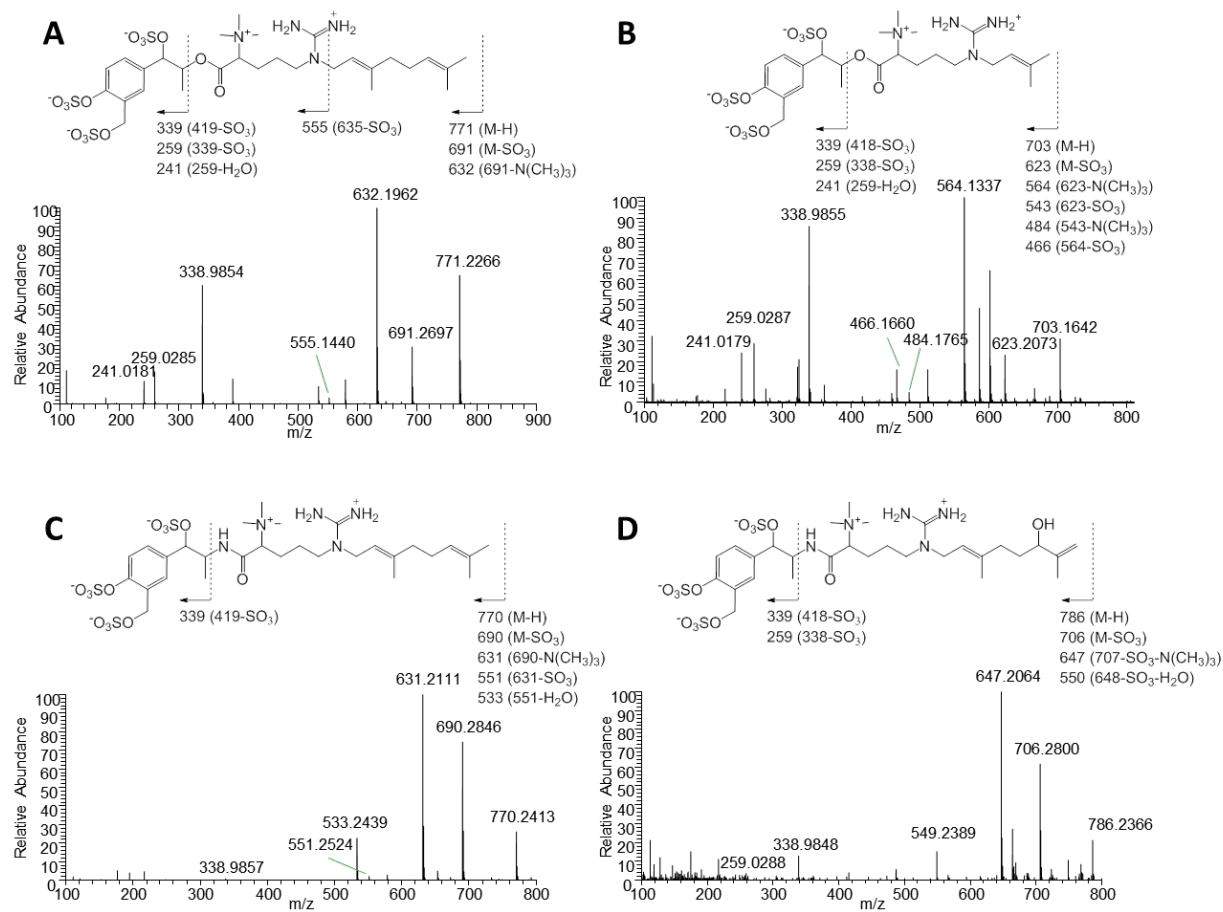
Figure S5. ¹H NMR Spectral comparison of AGD and MGD related compounds. The numbers on signals indicate the position in each compound. MGA-771 (**6**) is highlighted as a blue line with three chemical shifts indicated by red numbers.
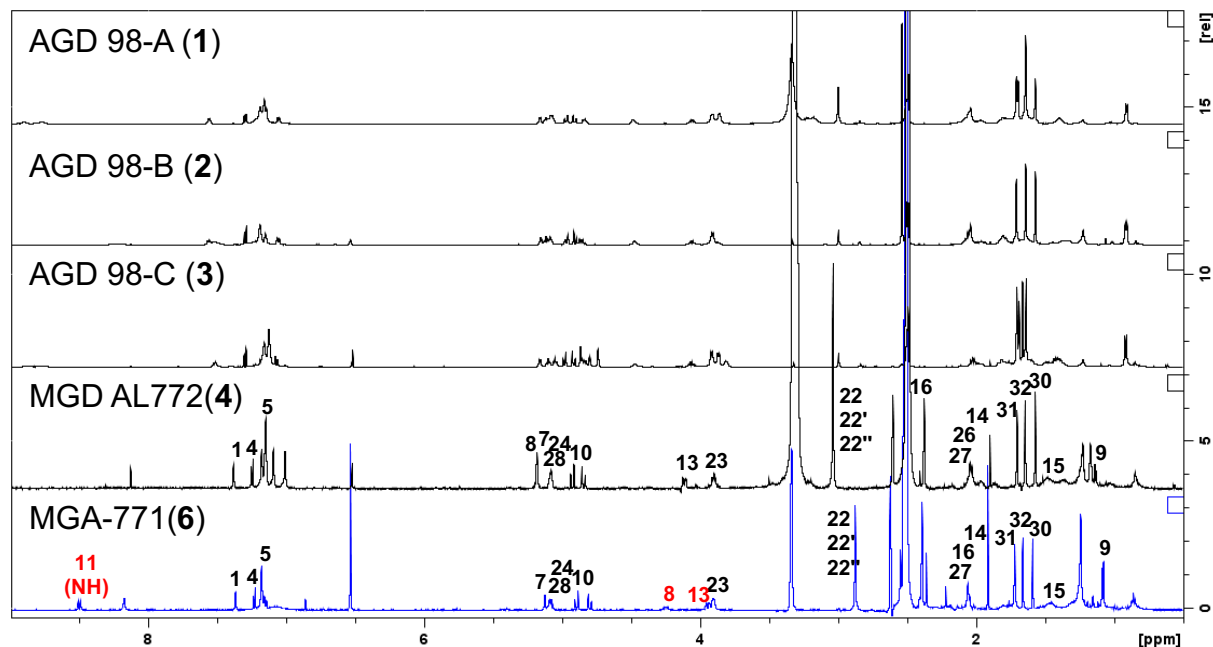
Figure S6. ¹H NMR spectrum of AGD 98-A (1) in DMSO-d_6 at 300 K.
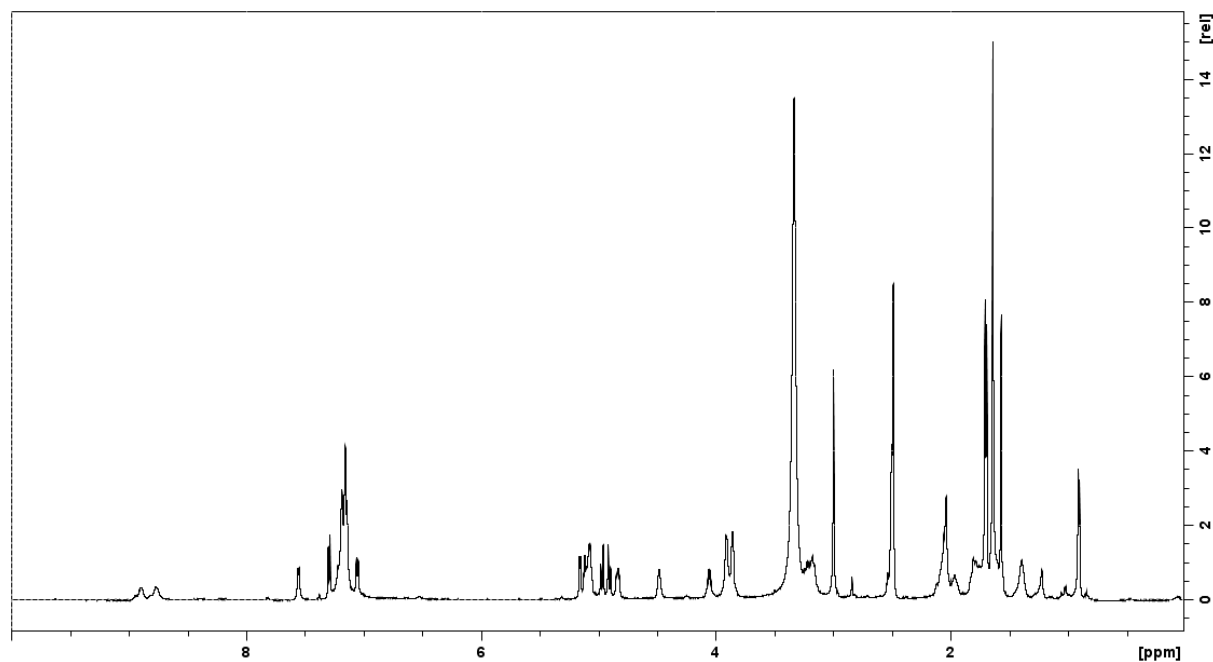
Figure S7. ^{1}H NMR spectrum of AGD 98-B (2) in DMSO-*d*₆ at 300 K.
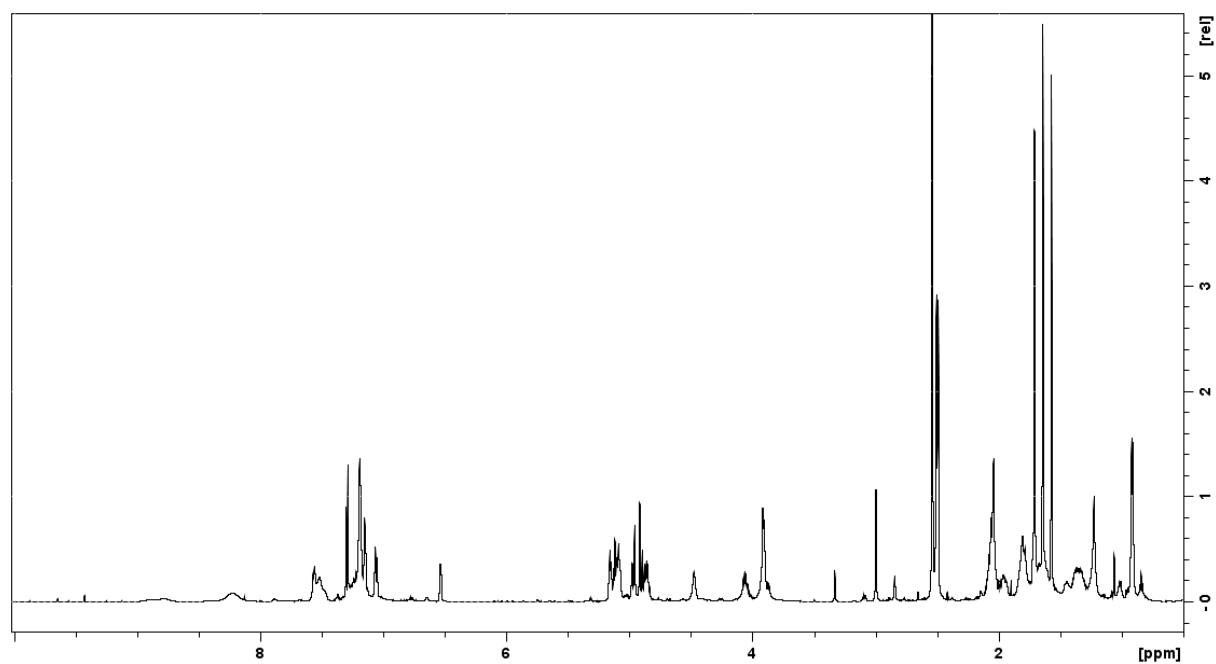
Figure S8. ^{1}H NMR spectrum of AGD 98-C (3) in DMSO-*d*₆ at 300 K.
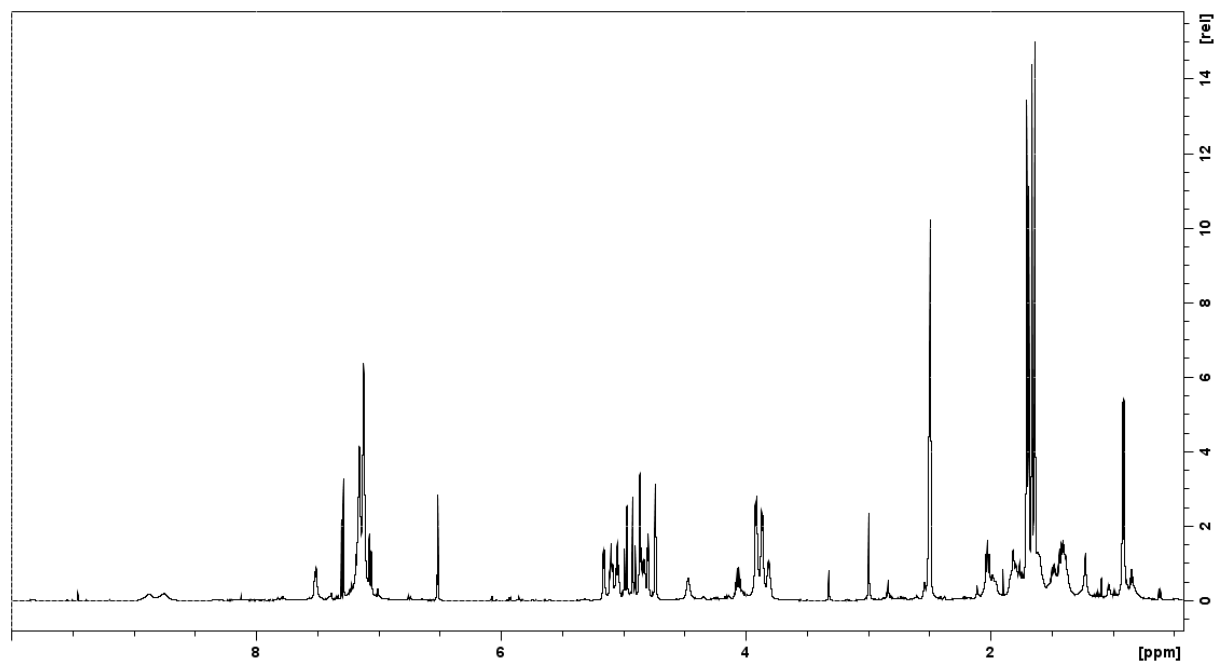
Figure S9. ^{1}H NMR spectrum of MGD AL772 (4) in DMSO-d_6 at 300 K.
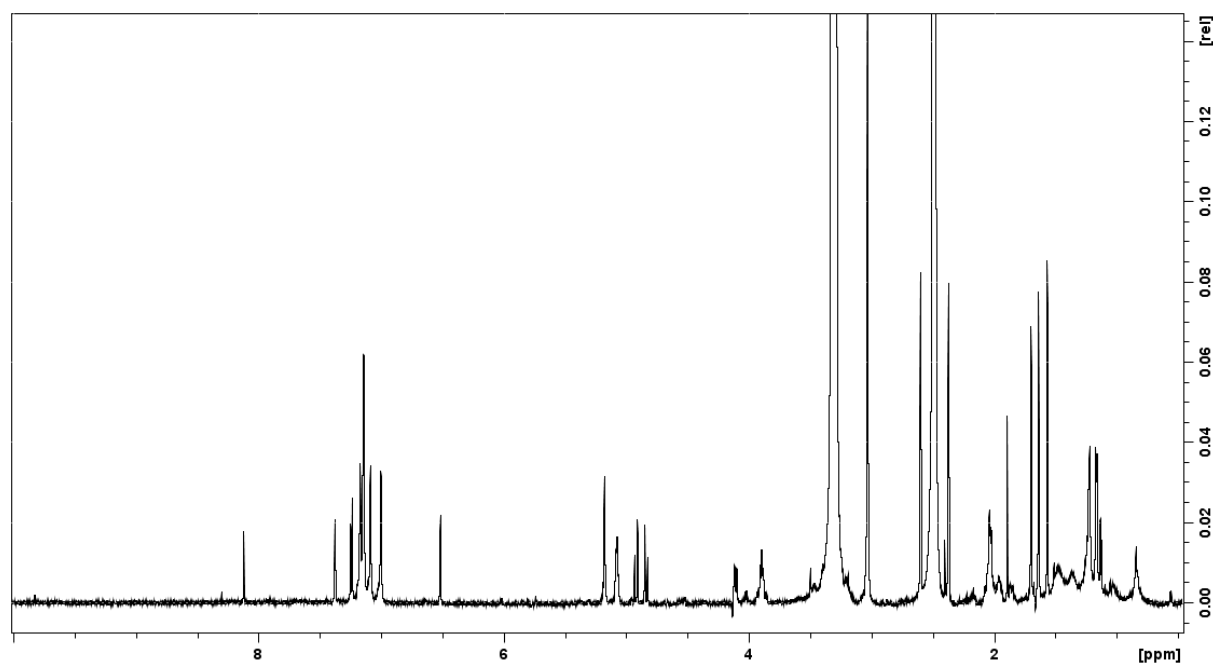
Figure S10. ^{13}C NMR spectrum of MGD AL772 (4) in DMSO-d_6 at 300 K.
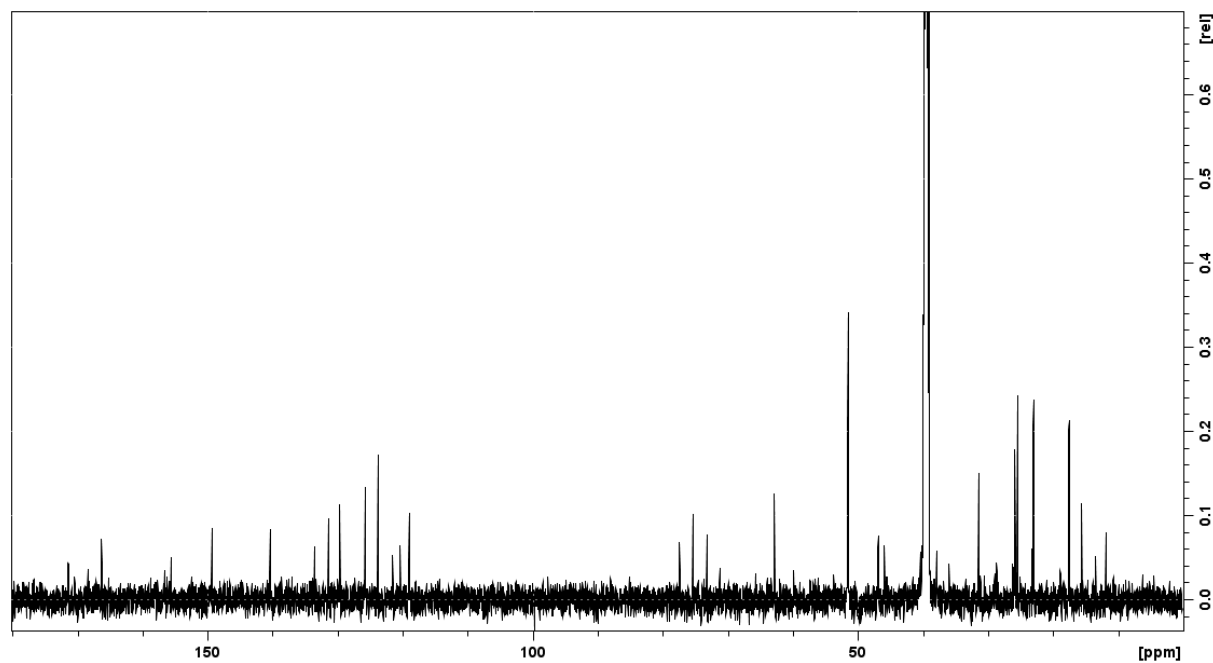
Figure S11. ^{1}H-^{1}H COSY spectrum of MGD AL772 (4) in DMSO-d_6 at 300 K.
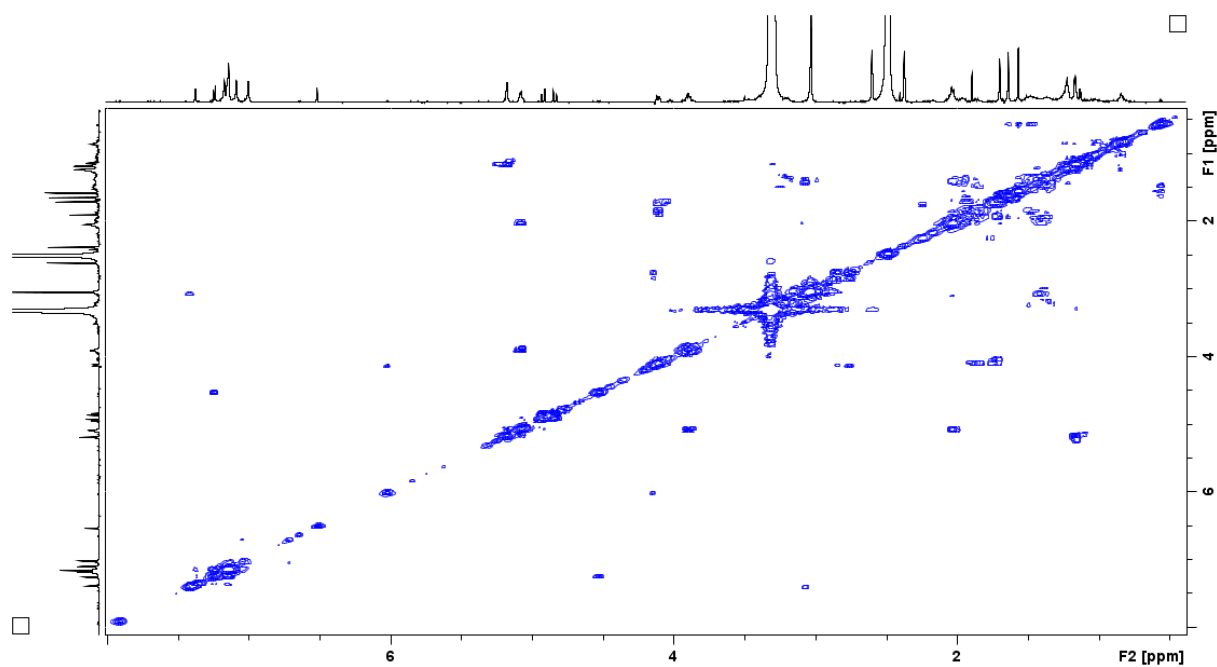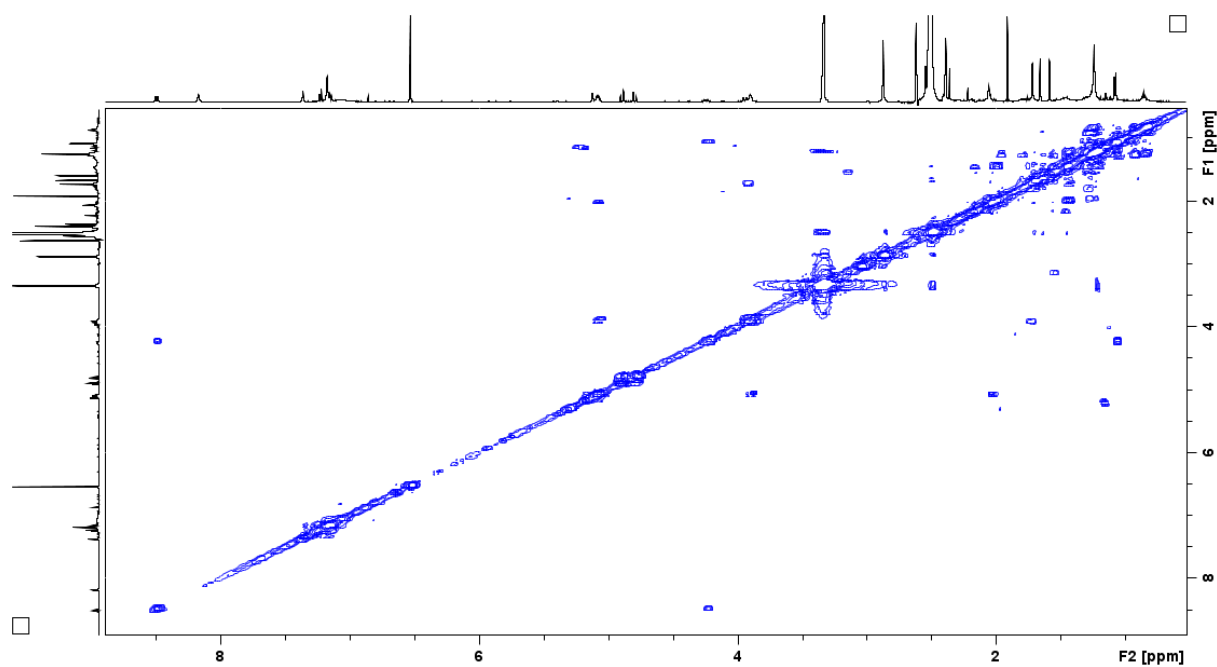

Figure S12. HSQC spectrum of MGD AL772 (4) in DMSO-d_6 at 300 K.
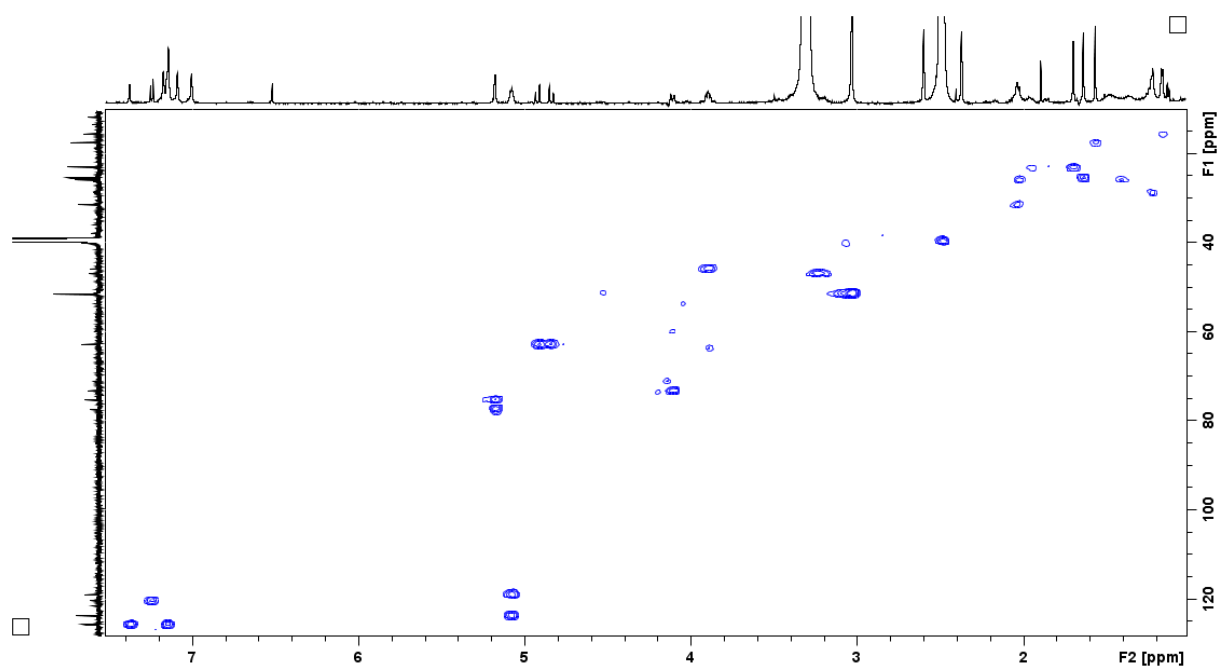
Figure S13. HMBC spectrum of MGD AL772 (4) in DMSO-d_6 at 300 K.
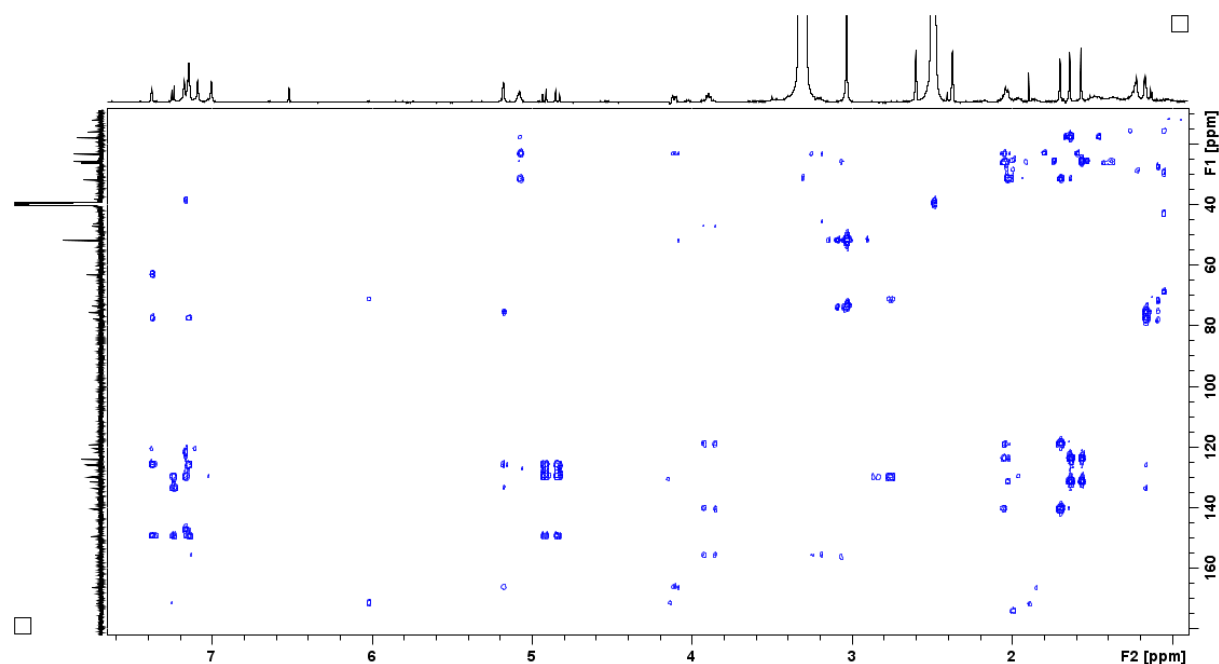
Figure S14. ^{1}H NMR of MGA-771 (6) in DMSO-d_6 at 300 K.
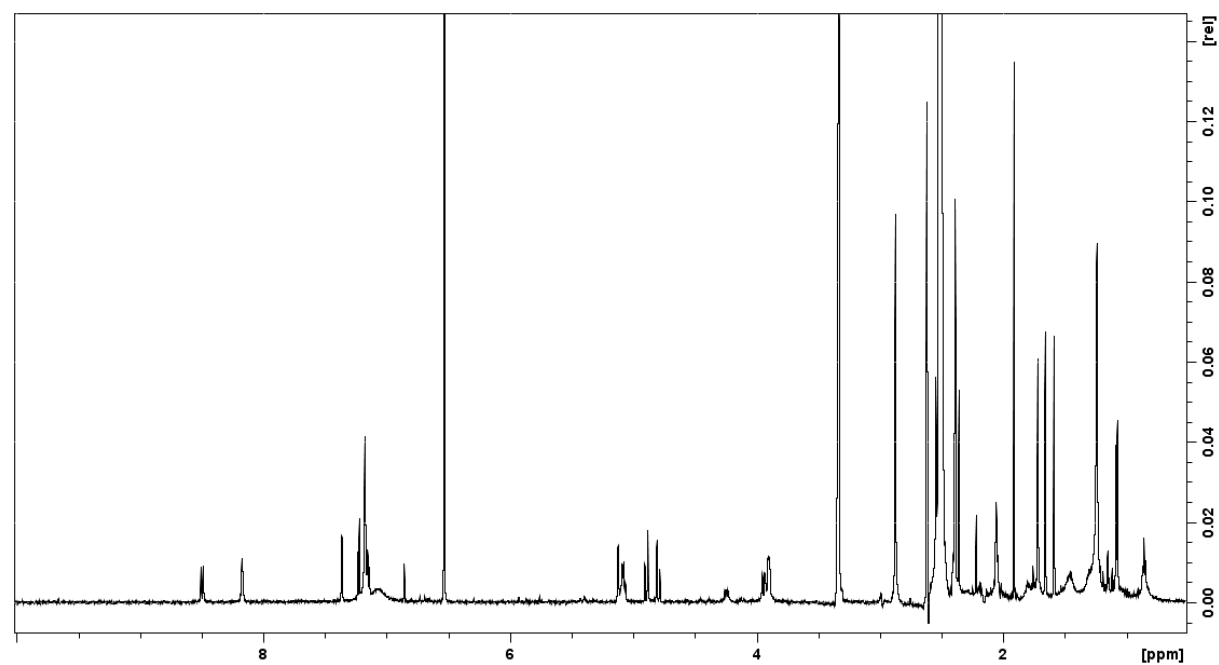
Figure S15. ^{1}H-^{1}H COSY spectrum of MGA-771 (6) in DMSO-d_6 at 300 K.

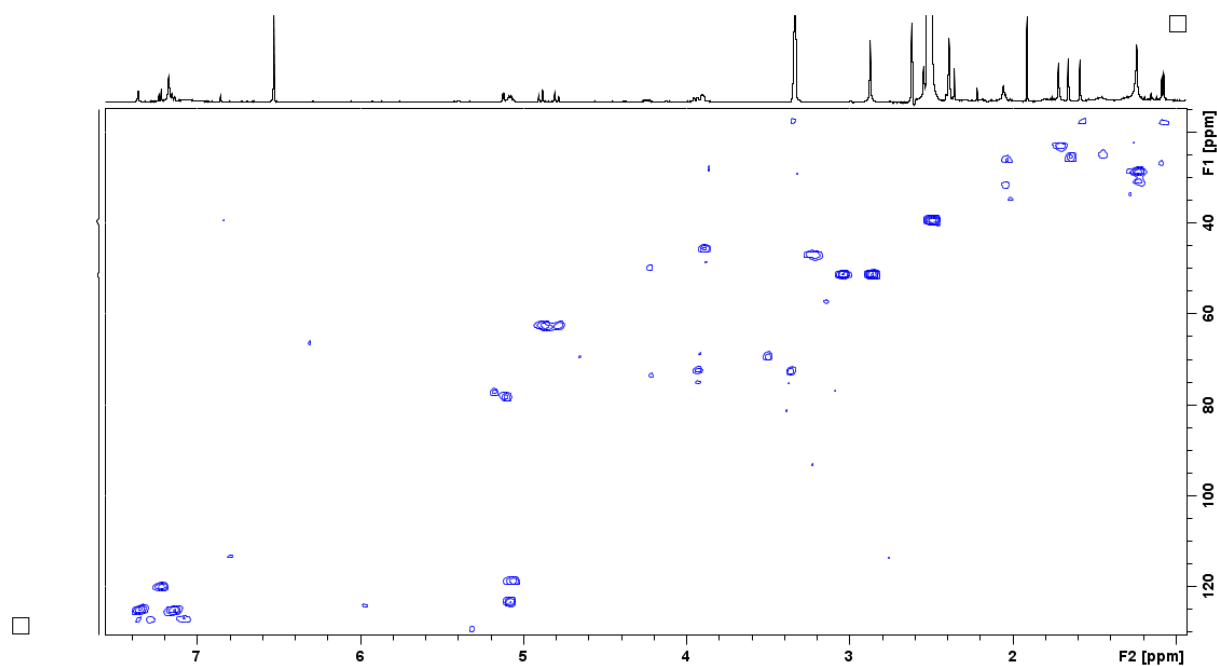**Figure S16.** HSQC spectrum of MGA-771 (6) in DMSO-d_6 at 300 K.

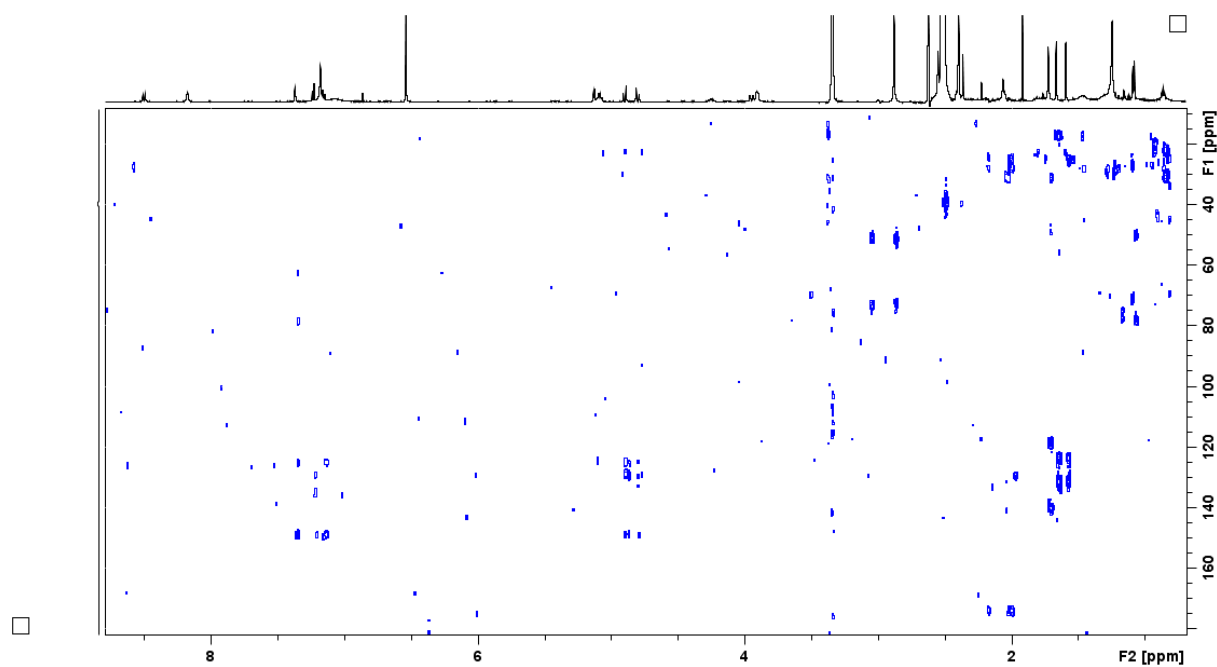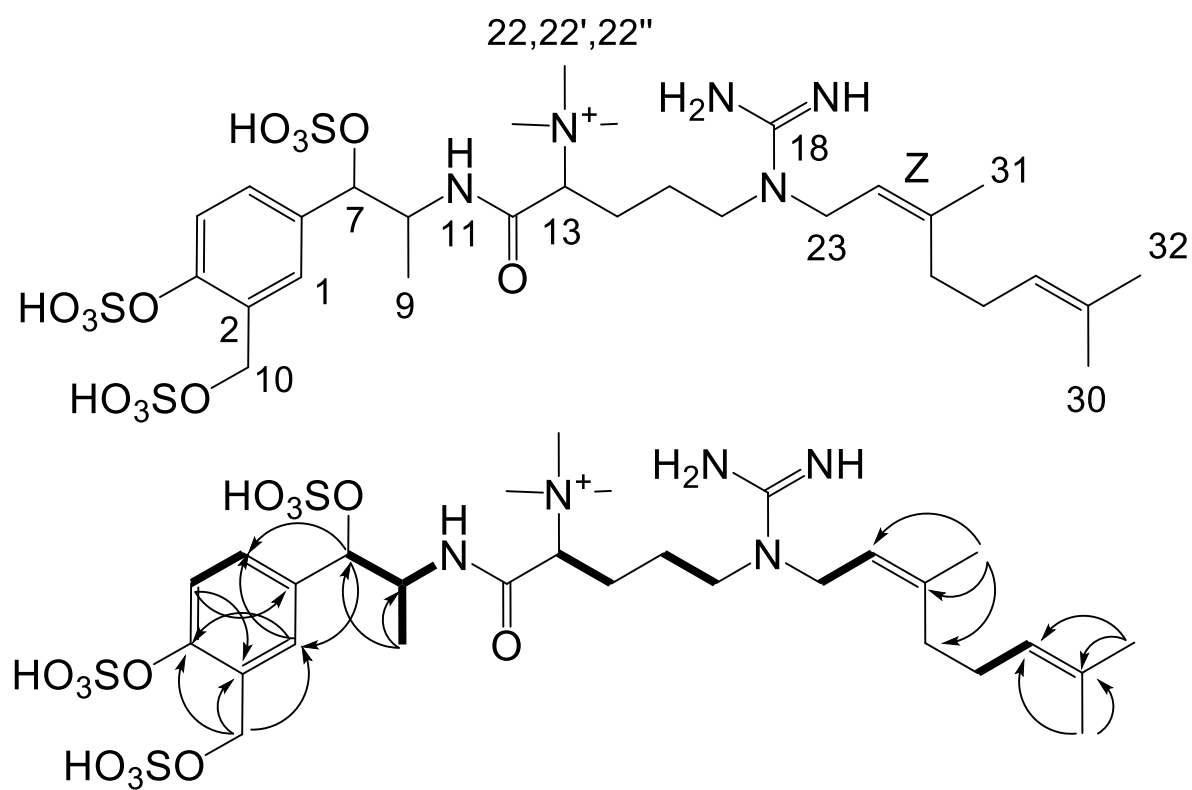**Figure S17. HMBC spectrum of MGA-771 (6) in DMSO-d_6 at 300 K.**

Figure S18. Observed ¹H-¹H COSY (bold line) and HMBC (arrow) correlations.

References

1. Humbert, J. F., Barbe, V., Latifi, A., Gugger, M., Calteau, A., Coursin, T., Lajus, A., Castelli, V., Oztas, S., Samson, G., Longin, C., Medigue, C., and de Marsac, N. T. (2013) A tribute to disorder in the genome of the bloom-forming freshwater cyanobacterium *Microcystis aeruginosa*, *PLoS One 8*, e70747.

2. Criscuolo, A., and Brisse, S. (2013) AlienTrimmer: A tool to quickly and accurately trim off multiple short contaminant sequences from high-throughput sequencing reads, *Genomics 102*, 500—506.

3. Liu, Y., Schroder, J., and Schmidt, B. (2013) Musket: A multistage k-mer spectrum-based error corrector for Illumina sequence data, *Bioinformatics 29*, 308—315.

4. Crusoe, M. R., Alameldin, H. F., Awad, S., Boucher, E., Caldwell, A., Cartwright, R., Charbonneau, A., Constantinides, B., Edvenson, G., Fay, S., Fenton, J., Fenzl, T., Fish, J., Garcia-Gutierrez, L., Garland, P., Gluck, J., Gonzalez, I., Guermond, S., Guo, J., Gupta, A., Herr, J. R., Howe, A., Hyer, A., Harpfer, A., Irber, L., Kidd, R., Lin, D., Lippi, J., Mansour, T., McA'Nulty, P., McDonald, E., Mizzi, J., Murray, K. D., Nahum, J. R., Nanlohy, K., Nederbragt, A. J., Ortiz-Zuazaga, H., Ory, J., Pell, J., Pepe-Ranney, C., Russ, Z. N., Schwarz, E., Scott, C., Seaman, J., Sievert, S., Simpson, J., Skennerton, C. T., Spencer, J., Srinivasan, R., Standage, D., Stapleton, J. A., Steinman, S. R., Stein, J., Taylor, B., Trimble, W., Wiencko, H. L., Wright, M., Wyss, B., Zhang, Q., Zyme, E., and Brown, C. T. (2015) The khmer software package: enabling efficient nucleotide sequence analysis, *F1000Research 4*, 900.

5. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., and Pevzner, P. A. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing, *J. Comput. Biol. 19*, 455—477.

6. Shih, P. M., Wu, D., Latifi, A., Axen, S. D., Fewer, D. P., Talla, E., Calteau, A., Cai, F., Tandeau de Marsac, N., Rippka, R., Herdman, M., Sivonen, K., Coursin, T., Laurent, T., Goodwin, L., Nolan, M., Davenport, K. W., Han, C. S., Rubin, E. M., Eisen, J. A., Woyke, T., Gugger, M., and Kerfeld, C. A. (2013) Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing, *Proc. Natl Acad. Sci. U. S. A. 110*, 1053—1058.

7. Kessner, D., Chambers, M., Burke, R., Agus, D., and Mallick, P. (2008) ProteoWizard: open source software for rapid proteomics tools development, *Bioinformatics 24*, 2534—2536.

8. Wang, M. X., and Carver, J. J., and Phelan, V. V., and Sanchez, L. M., and Garg, N., and Peng, Y., and Nguyen, D. D., and Watrous, J., and Kapono, C. A., and Luzzatto-Knaan, T., and Porto, C., and Bouslimani, A., and Melnik, A. V., and Meehan, M. J., and Liu, W. T., and Criisemann, M., and Boudreau, P. D., and Esquenazi, E., and Sandoval-Calderon, M., and Kersten, R. D., and Pace, L. A., and Quinn, R. A., and Duncan, K. R., and Hsu, C. C., and Floros, D. J., and Gavilan, R. G., and Kleigrewe, K., and Northen, T., and Dutton, R. J., and Parrot, D., and Carlson, E. E., and Aigle, B., and Michelsen, C. F., and Jelsbak, L., and Sohlenkamp, C., and Pevzner, P., and Edlund, A., and McLean, J., and Piel, J., and Murphy, B. T., and Gerwick, L., and Liaw, C. C., and Yang, Y. L., and Humpf, H. U., and Maansson, M., and Keyzers, R. A., and Sims, A. C., and Johnson, A. R., and Sidebottom, A. M., and Sedio, B. E., and Klitgaard, A., and Larson, C. B., and Boya, C. A., and Torres-Mendoza, D., and Gonzalez, D. J., and Silva, D. B., and Marques, L. M., and Demarque, D. P., and Pociute, E., and O'Neill, E. C., and Briand, E., and

Helfrich, E. J. N., and Granatosky, E. A., and Glukhov, E., and Ryffel, F., and Houson, H., and Mohimani, H., and Kharbush, J. J., and Zeng, Y., and Vorholt, J. A., and Kurita, K. L., and Charusanti, P., and McPhail, K. L., and Nielsen, K. F., and Vuong, L., and Elfeki, M., and Traxler, M. F., and Engene, N., and Koyama, N., and Vining, O. B., and Baric, R., and Silva, R. R., and Mascuch, S. J., and Tomasi, S., and Jenkins, S., and Macherla, V., and Hoffman, T., and Agarwal, V., and Williams, P. G., and Dai, J. Q., and Neupane, R., and Gurr, J., and Rodriguez, A. M. C., and Lamsa, A., and Zhang, C., and Dorrestein, K., and Duggan, B. M., and Almaliti, J., and Allard, P. M., and Phapale, P., and Nothias, L. F., and Alexandrovr, T., and Litaudon, M., and Wolfender, J. L., and Kyle, J. E., and Metz, T. O., and Peryea, T., and Nguyen, D. T., and VanLeer, D., and Shinn, P., and Jadhav, A., and Muller, R., and Waters, K. M., and Shi, W. Y., and Liu, X. T., and Zhang, L. X., and Knight, R., and Jensen, P. R., and Palsson, B. O., and Pogliano, K., and Linington, R. G., and Gutierrez, M., and Lopes, N. P., and Gerwick, W. H., and Moore, B. S., and Dorrestein, P. C., and Bandeira, N. (2016) Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking, *Nat. Biotechnol. 34*, 828—837.

9. Watrous, J., Roach, P., Alexandrov, T., Heath, B. S., Yang, J. Y., Kersten, R. D., van der Voort, M., Pogliano, K., Gross, H., Raaijmakers, J. M., Moore, B. S., Laskin, J., Bandeira, N., and Dorrestein, P. C. (2012) Mass spectral molecular networking of living microbial colonies, *Proc. Natl Acad. Sci. U. S. A. 109*, E1743—1752.

10. Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L., and Ideker, T. (2011) Cytoscape 2.8: New features for data integration and network visualization, *Bioinformatics 27*, 431—432.

11. Frangeul, L., Quillardet, P., Castets, A., Humbert, J., Matthijs, H., Cortez, D., Tolonen, A., Zhang, C., Gribaldo, S., Kehr, J., Zilliges, Y., Ziemert, N., Becker, S., Talla, E., Latifi, A., Billault, A., Lepelletier, A., Dittmann, E., Bouchier, C., and de Marsac, N. T. (2008) Highly plastic genome of *Microcystis aeruginosa* PCC 7806, a ubiquitous toxic freshwater cyanobacterium, *BMC genomics 9*, 274.

12. Castro, W. O., Lima, A. R., Moraes, P. H., Siqueira, A., Aguiar, D., Baraúna, A., Martins, L., Fuzii, H., de Lima, C., Vianez-Júnior, J., Nunes, M., Dall'Agnol, L., and Gonçalves, E. (2016) Draft genome sequence of *Microcystis aeruginosa* CACIAM 03, a cyanobacterium isolated from an Amazonian freshwater environment, *Genome Announc. 4*, e01299.

13. Yamaguchi, H., Suzuki, S., Sano, T., Tanabe, Y., Nakajima, N., and Kawachi, M. (2016) Draft genome sequence of *Microcystis aeruginosa* NIES-98, a non-microcystin-producing cyanobacterium from Lake Kasumigaura, Japan, *Genome Announc. 4*, e01187—01116.

14. Kaneko, T., Nakajima, N., Okamoto, S., Suzuki, I., Tanabe, Y., Tamaoki, M., Nakamura, Y., Kasai, F., Watanabe, A., Kawashima, K., Kishida, Y., Ono, A., Shimizu, Y., Takahashi, C., Minami, C., Fujishiro, T., Kohara, M., Katoh, M., Nakazaki, N., Nakayama, S., Yamada, M., Tabata, S., and Watanabe, M. (2007) Complete genomic structure of the bloom-forming toxic cyanobacterium *Microcystis aeruginosa* NIES-843, *DNA Res. 14*, 247—256.

15. Fiore, M. F., Alvarenga, D. O., Varani, A. M., Hoff-Risseti, C., Crespim, E., Ramos, R. T., Silva, A., Schaker, P. D., Heck, K., Rigonato, J., and Schneider, M. P. (2013) Draft genome sequence of the Brazilian toxic bloom-forming cyanobacterium *Microcystis aeruginosa* strain SPC777, *Genome Announc. 1*, e00547—00513.

16. Yang, C., Lin, F., Li, Q., Li, T., and Zhao, J. (2015) Comparative genomics reveals diversified CRISPR-Cas systems of globally distributed *Microcystis aeruginosa*, a freshwater bloom-forming cyanobacterium, *Front. Microbiol. 6*, 394.