A unique biosynthetic pathway in bloom-forming cyanobacteria 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

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.

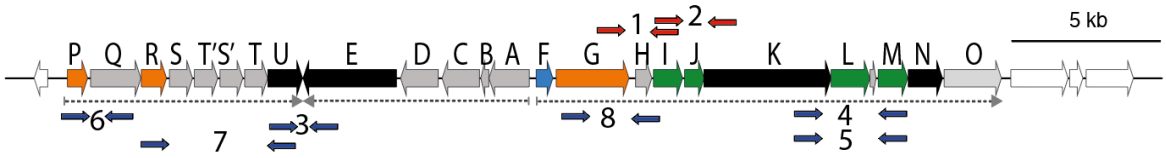


Table S2. *Microcystis* strains or genomes studied. 19 strains were cultured for metabolomics investigations.

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.

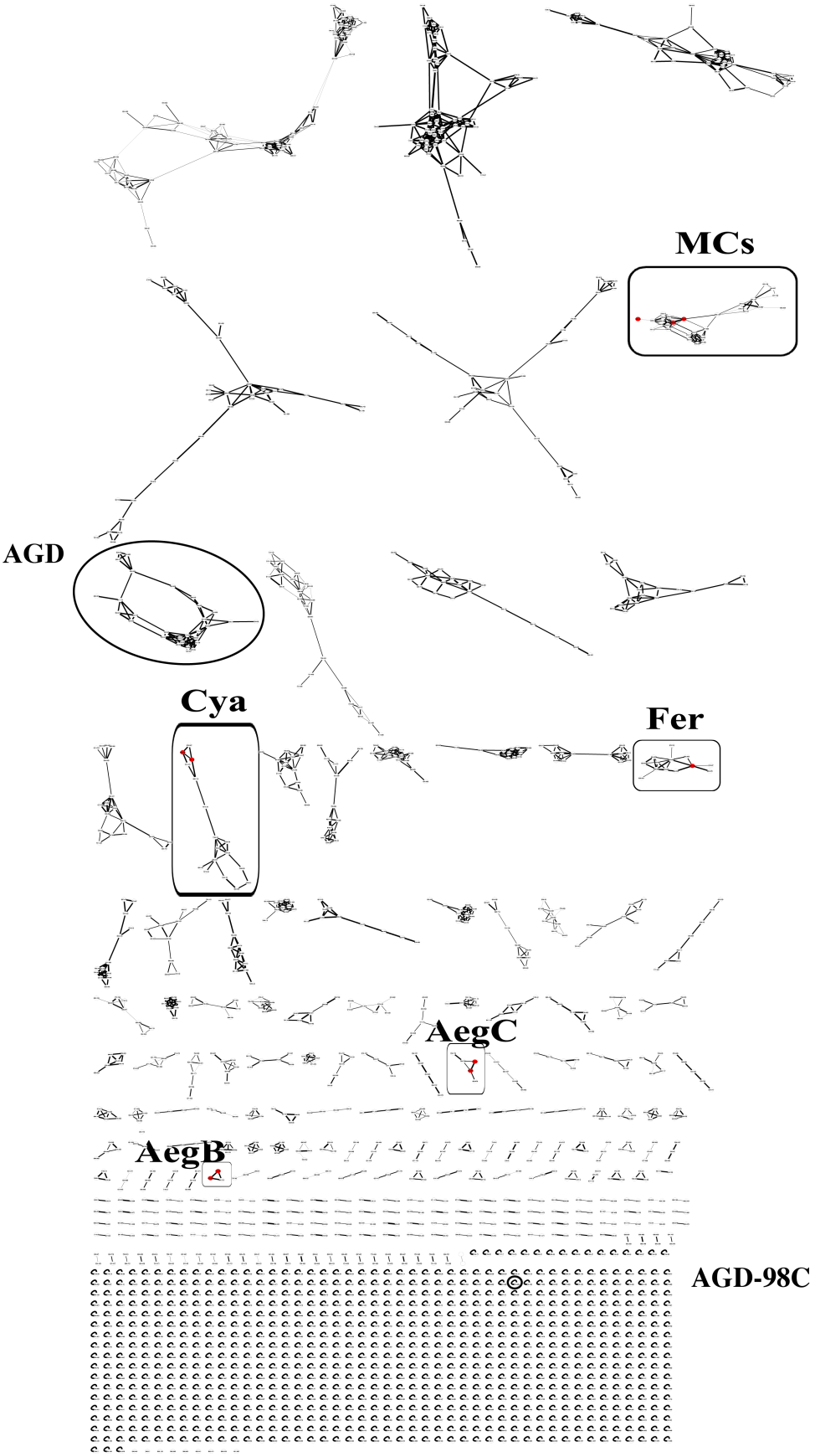


Figure S1. A

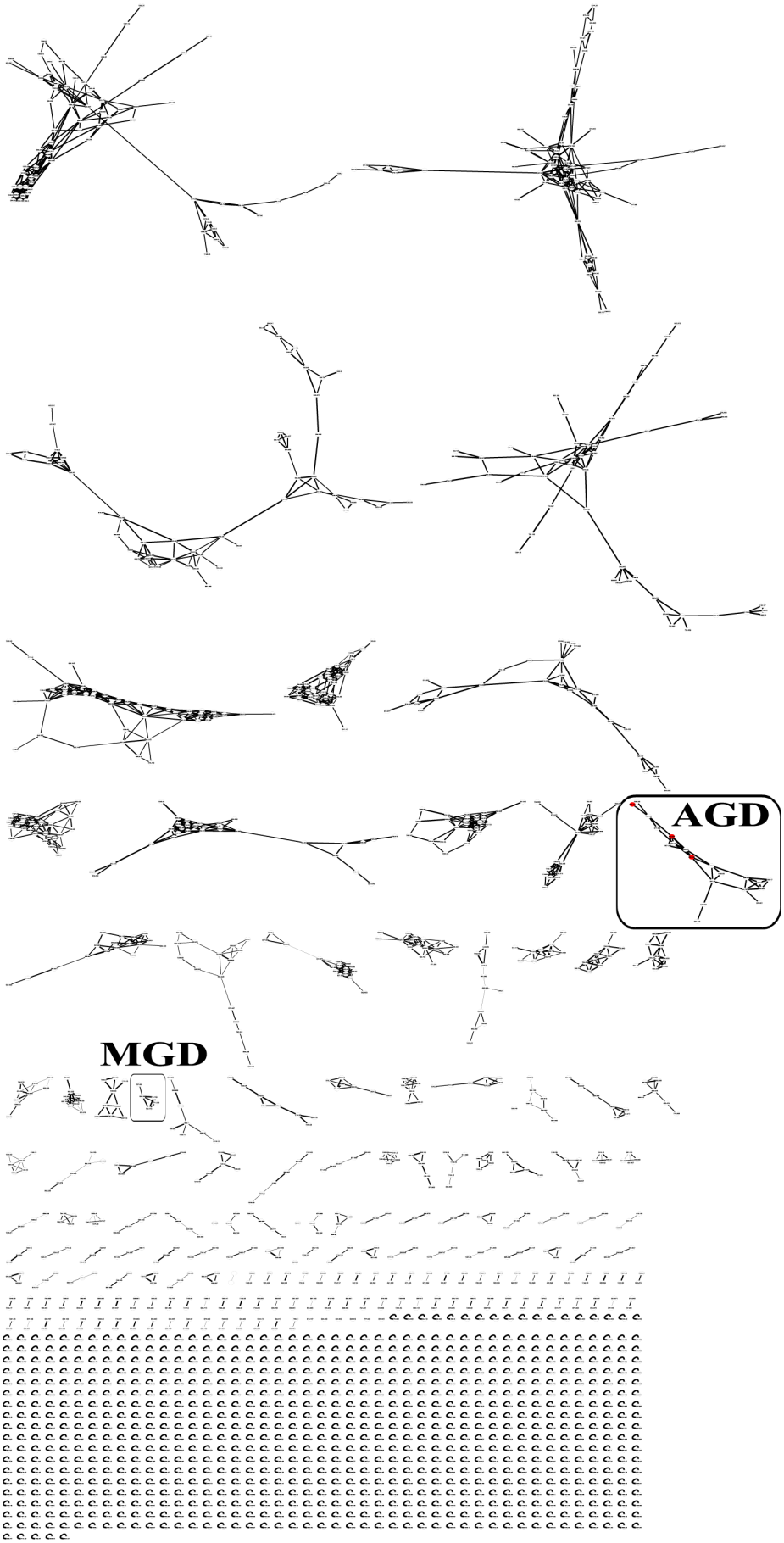
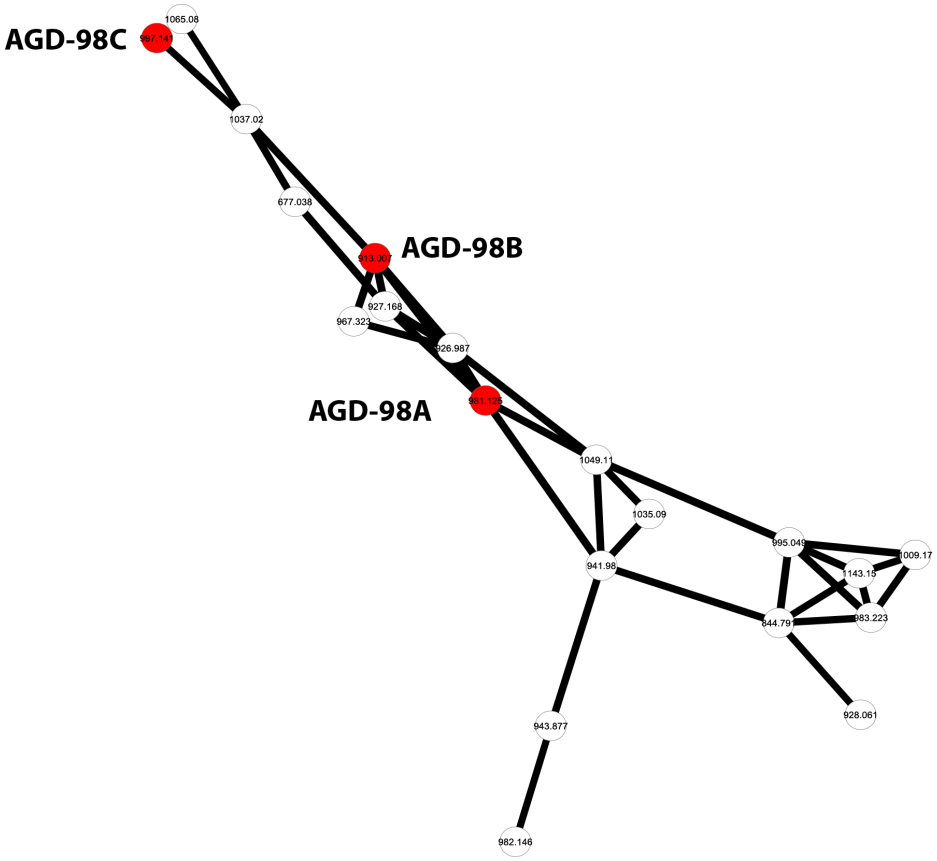


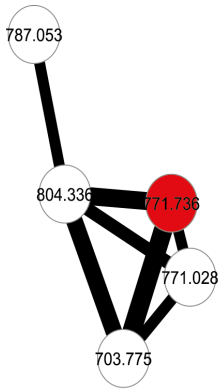
Figure S1. B

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



		Standard			Strains with candidate AGD gene cluster									
Parent mass	Assignment	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.



MGD AL772

		Standard			Strains with candidate AGD gene cluster									
Parent mass	Assignment	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).

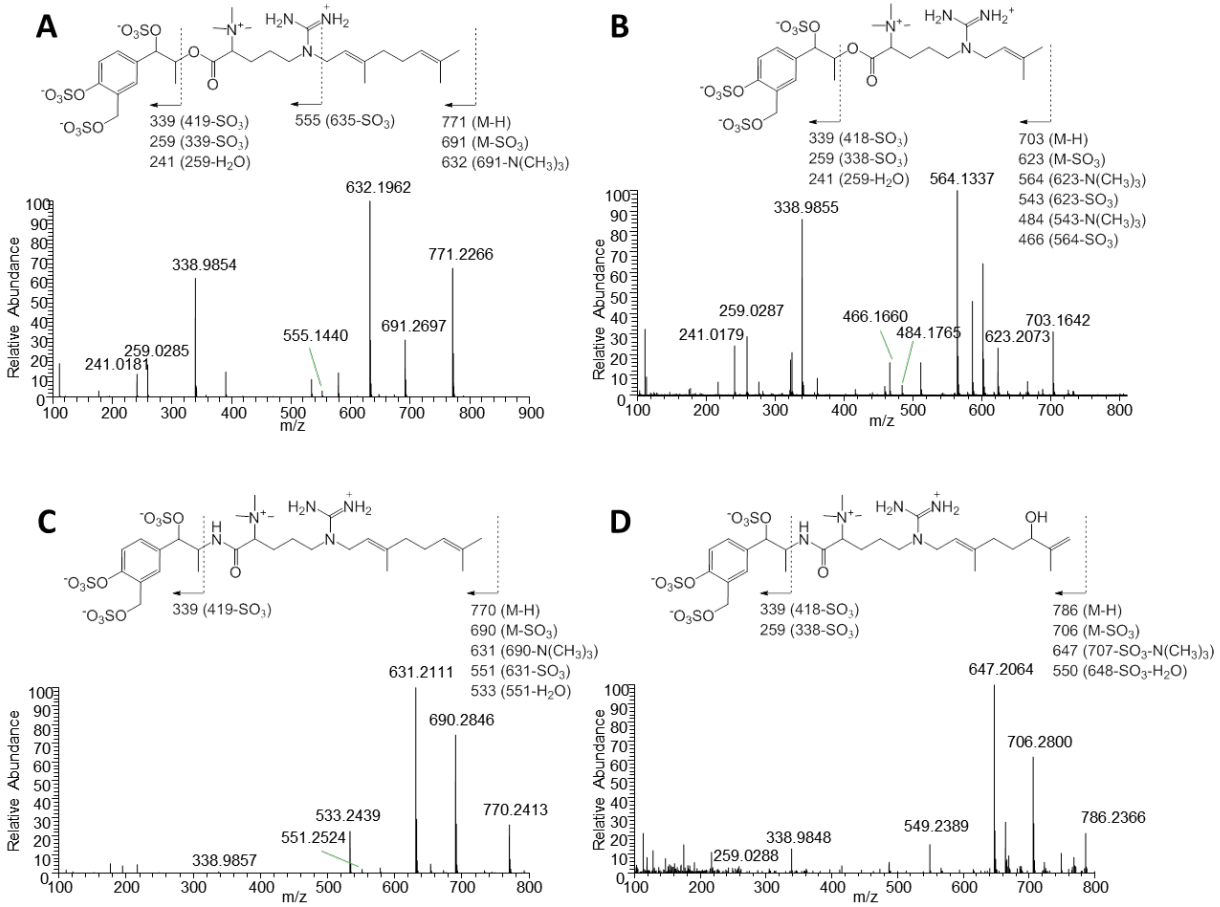


Figure S5. ^{1}H NMR Spectral comparison of AGD and MGD related compounds. The numbers on signals indicate the position in each compound. MGA-771 (**6**) is highlighted as a blue line with three chemical shifts indicated by red numbers.

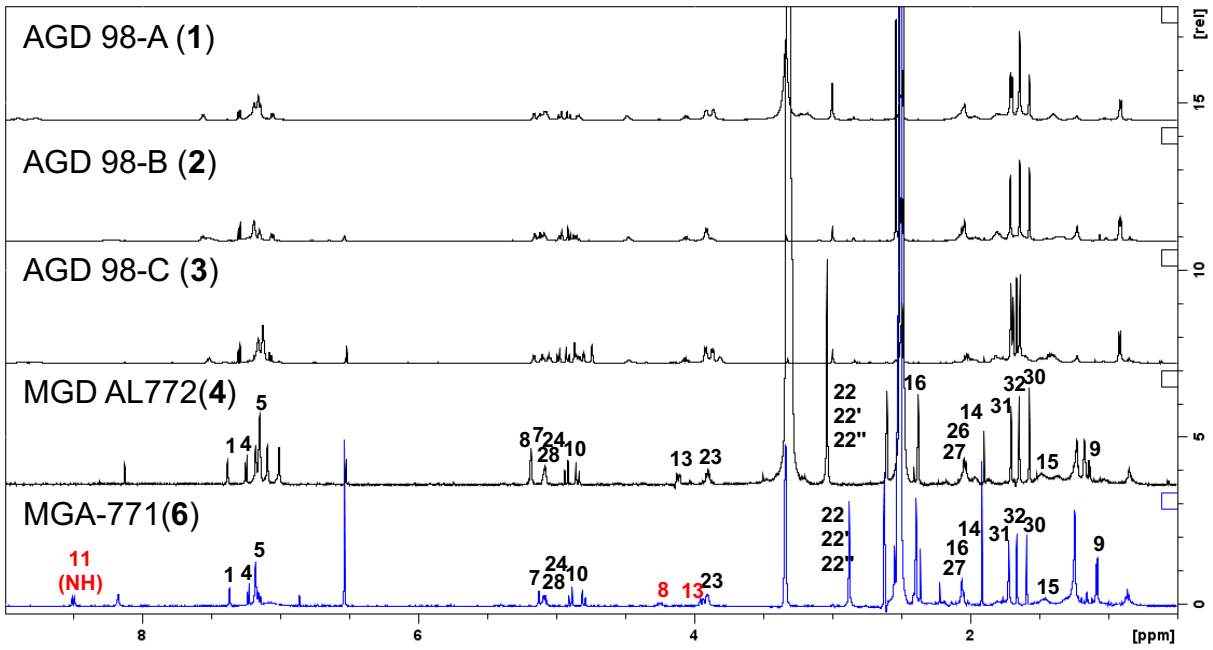


Figure S6. ^{1}H NMR spectrum of AGD 98-A (1) in DMSO-d_6 at 300 K.

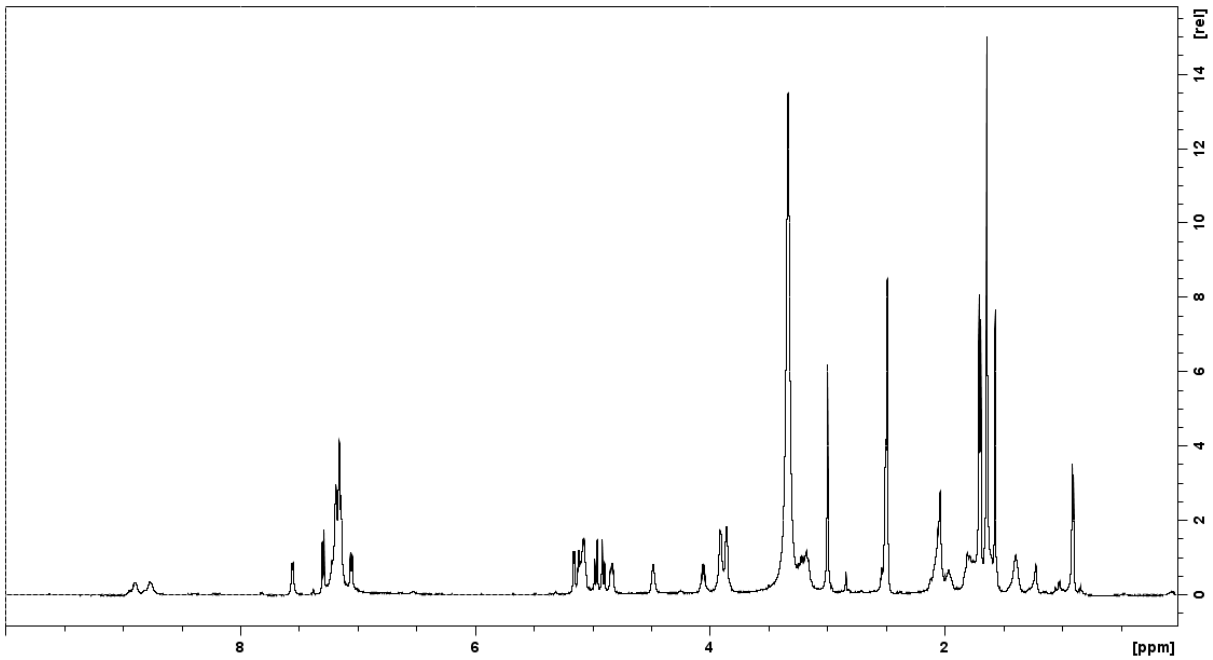


Figure S7. ^{1}H NMR spectrum of AGD 98-B (2) in DMSO-d_6 at 300 K.

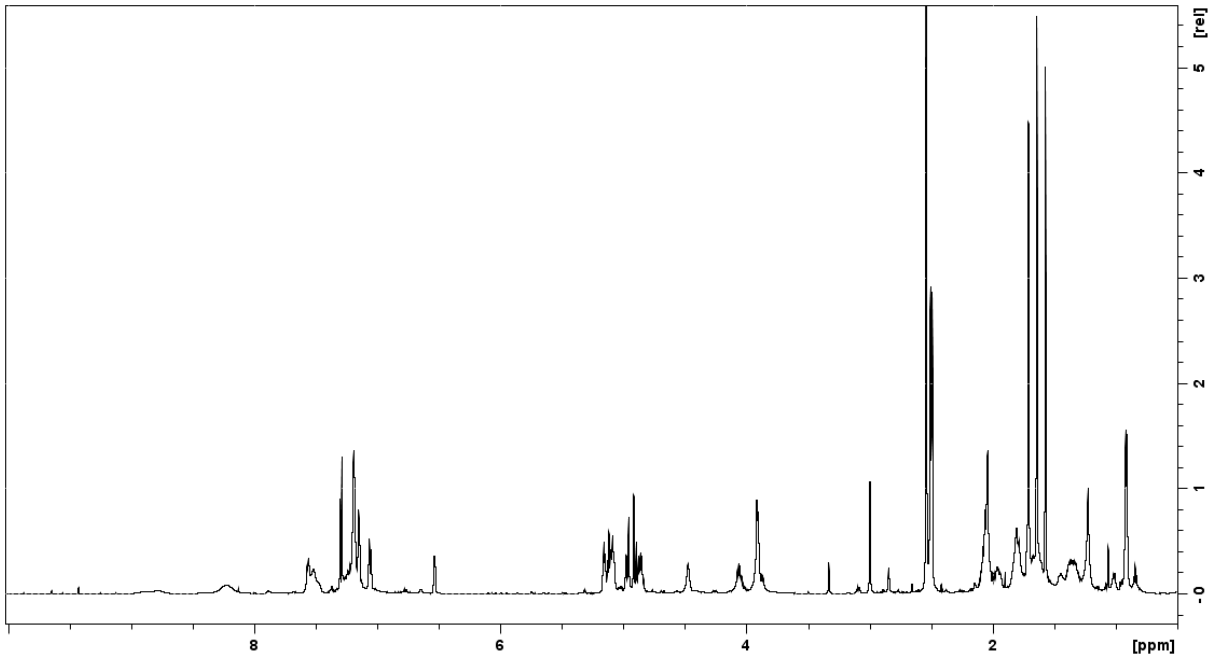


Figure S8. ^{1}H NMR spectrum of AGD 98-C (3) in DMSO-d_6 at 300 K.

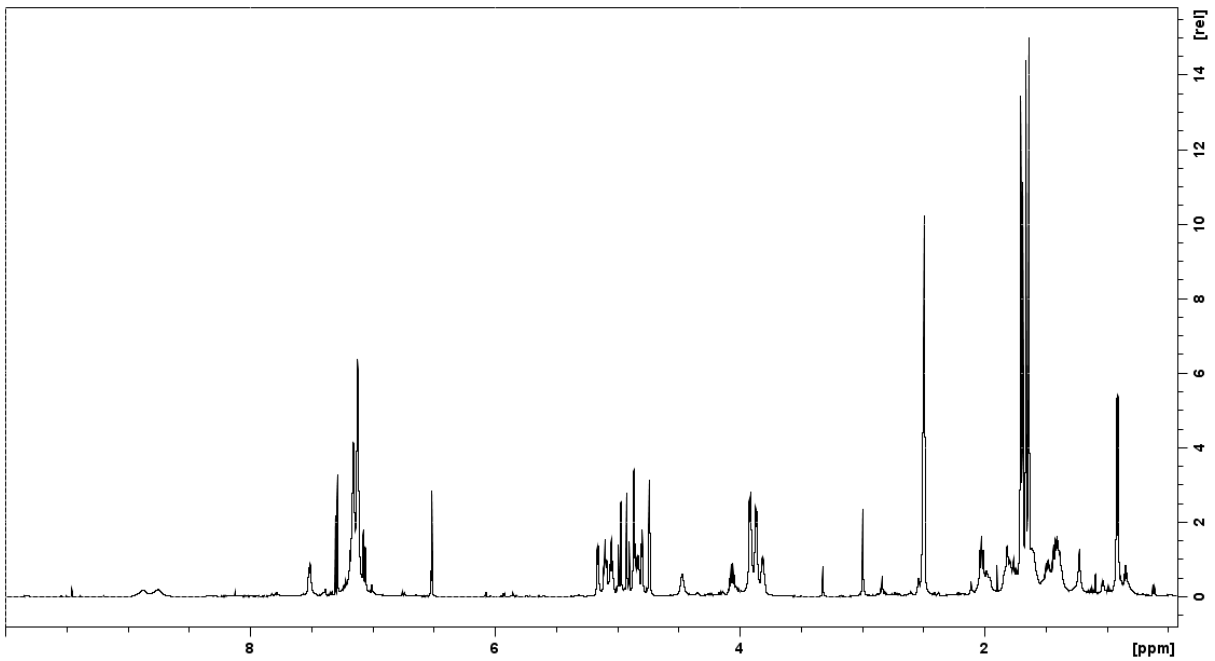


Figure S9. ^{1}H NMR spectrum of MGD AL772 (4) in DMSO-d_6 at 300 K.

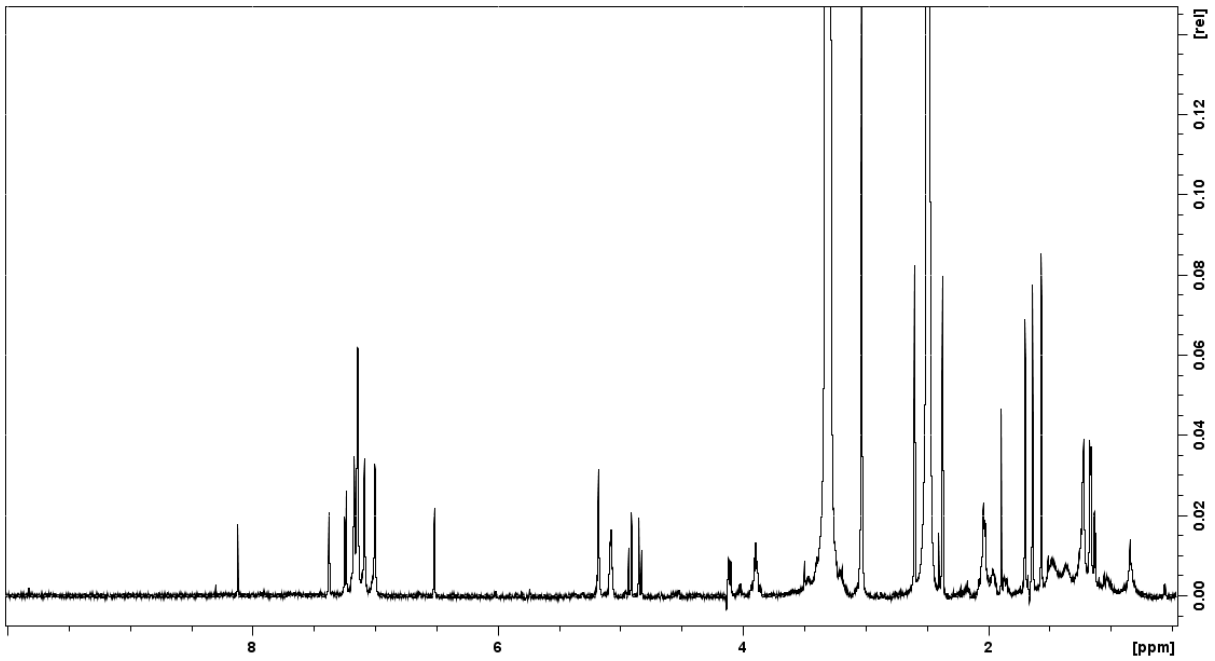


Figure S10. ^{13}C NMR spectrum of MGD AL772 (4) in DMSO-d_6 at 300 K.

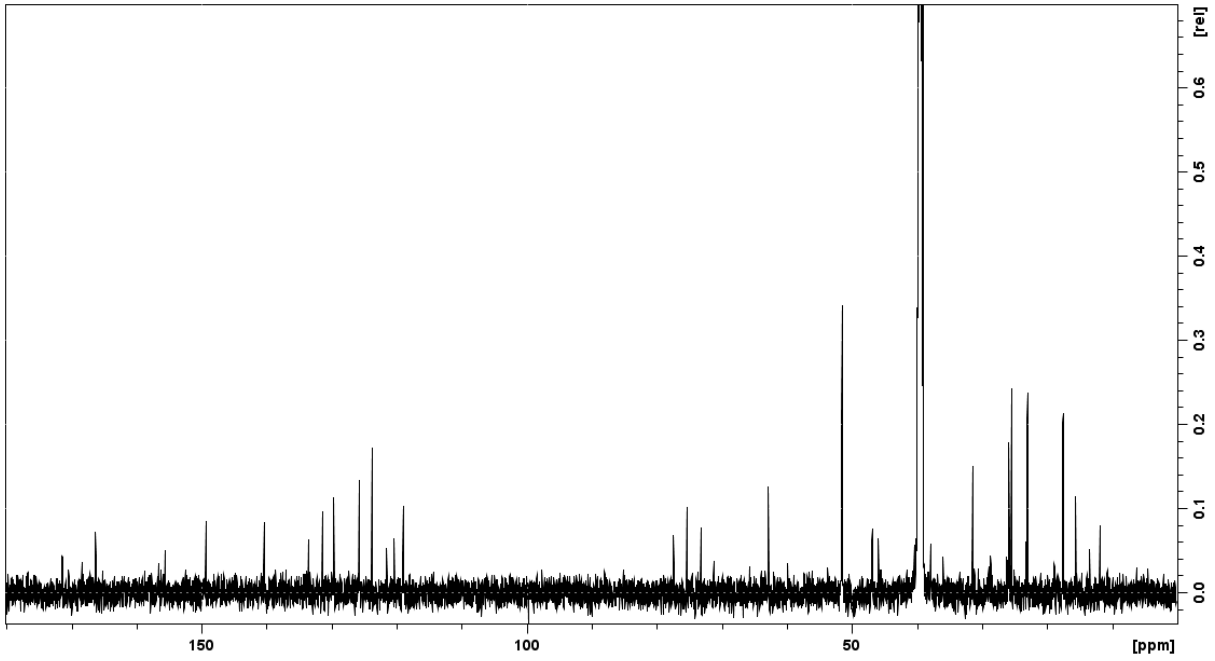


Figure S11. ^{1}H-^{1}H COSY spectrum of MGD AL772 (4) in DMSO-d_6 at 300 K.

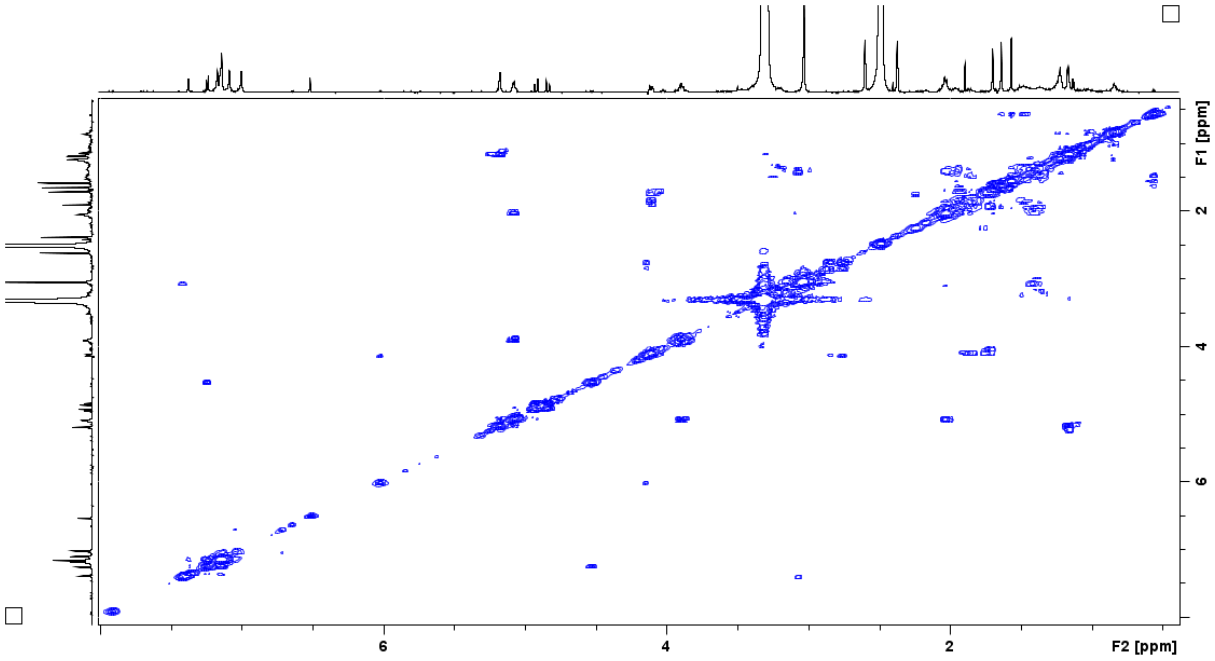


Figure S12. HSQC spectrum of MGD AL772 (4) in DMSO-d_6 at 300 K.

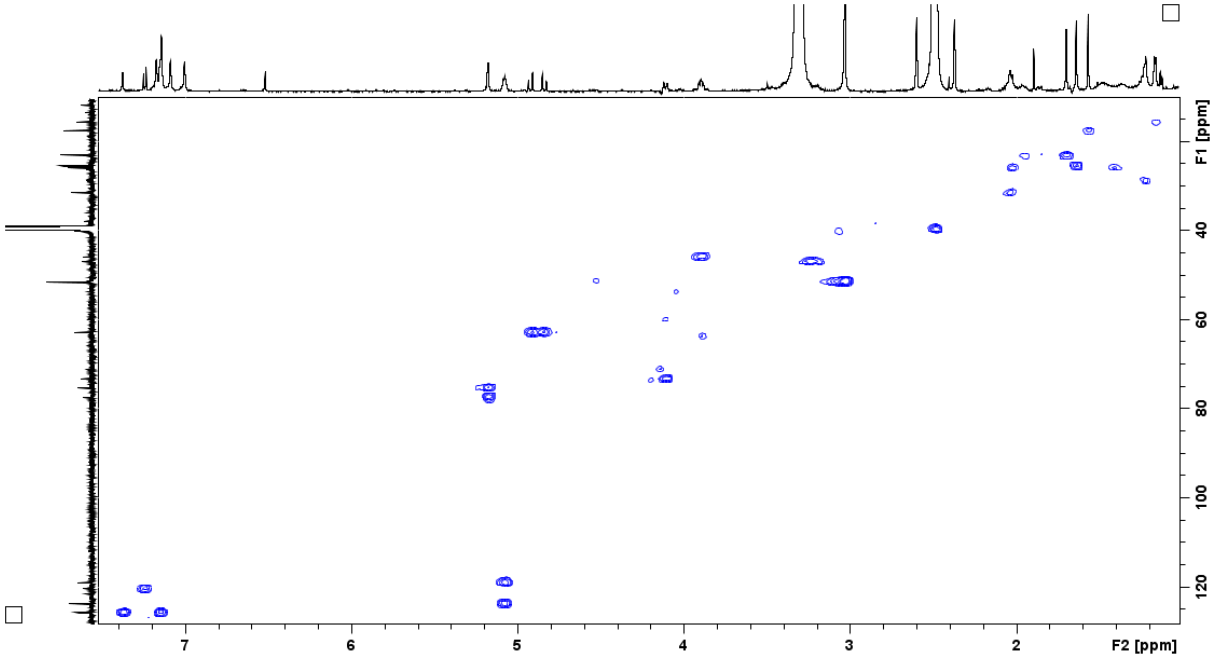


Figure S13. HMBC spectrum of MGD AL772 (4) in DMSO-d_6 at 300 K.

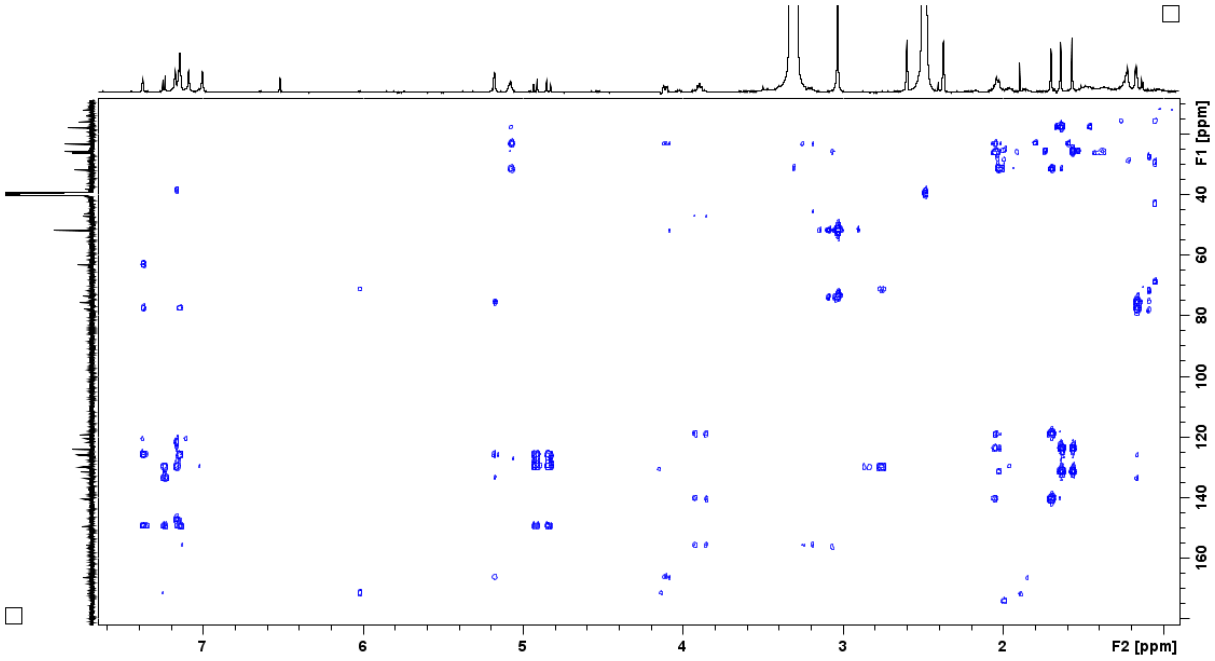


Figure S14. ^{1}H NMR of MGA-771 (6) in DMSO-d_6 at 300 K.

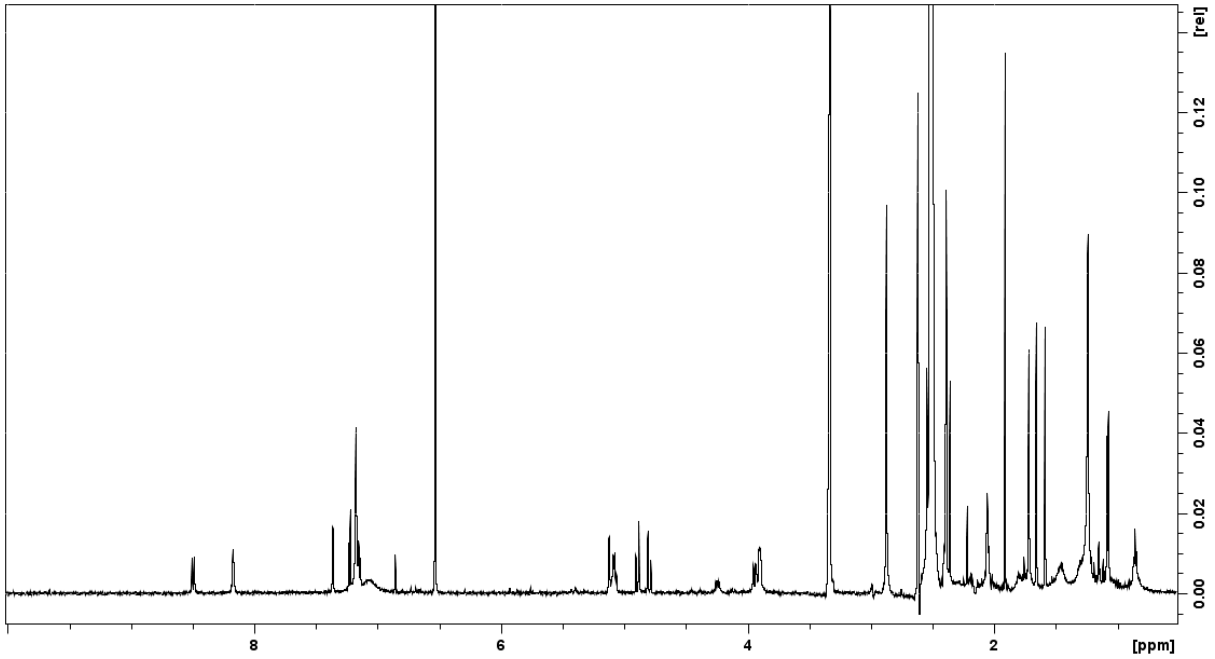


Figure S15. ^{1}H-^{1}H COSY spectrum of MGA-771 (6) in DMSO-d_6 at 300 K.

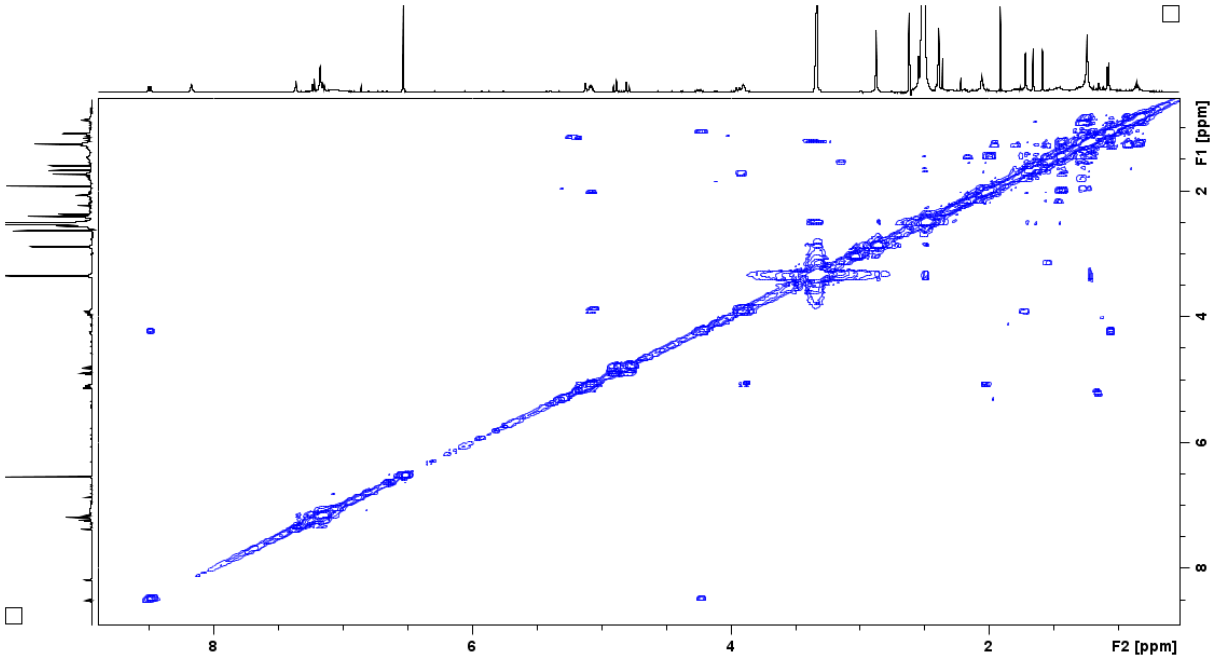


Figure S16. HSQC spectrum of MGA-771 (6) in DMSO-d_6 at 300 K.

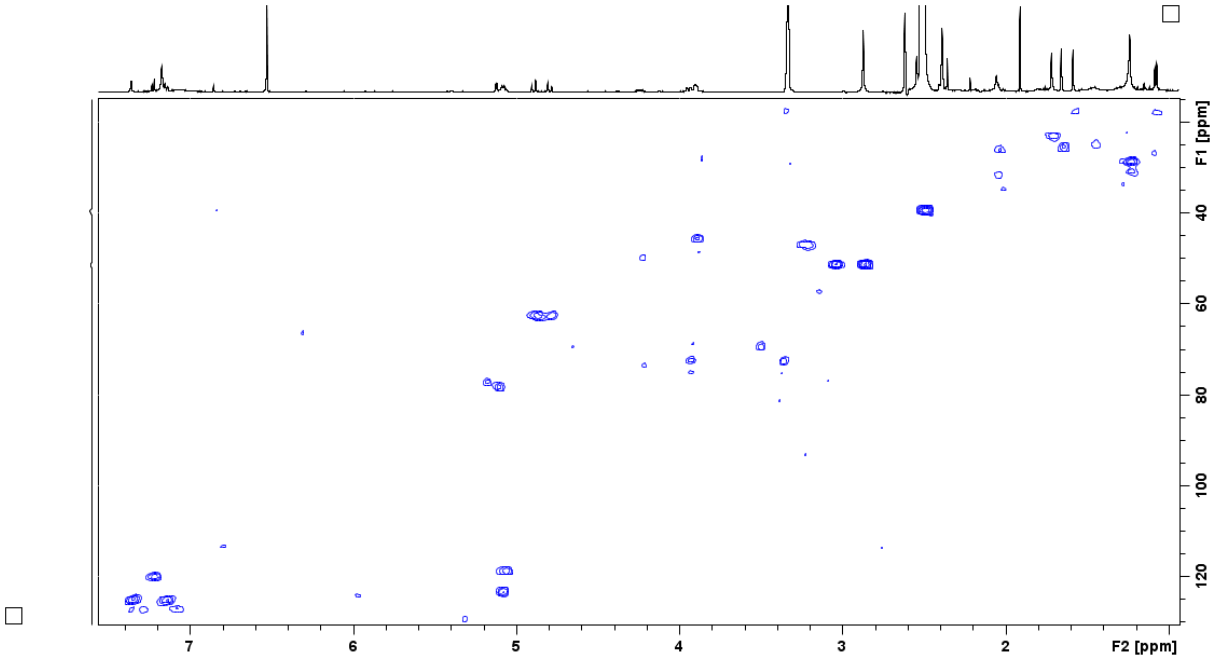


Figure S17. HMBC spectrum of MGA-771 (6) in DMSO-d_6 at 300 K.

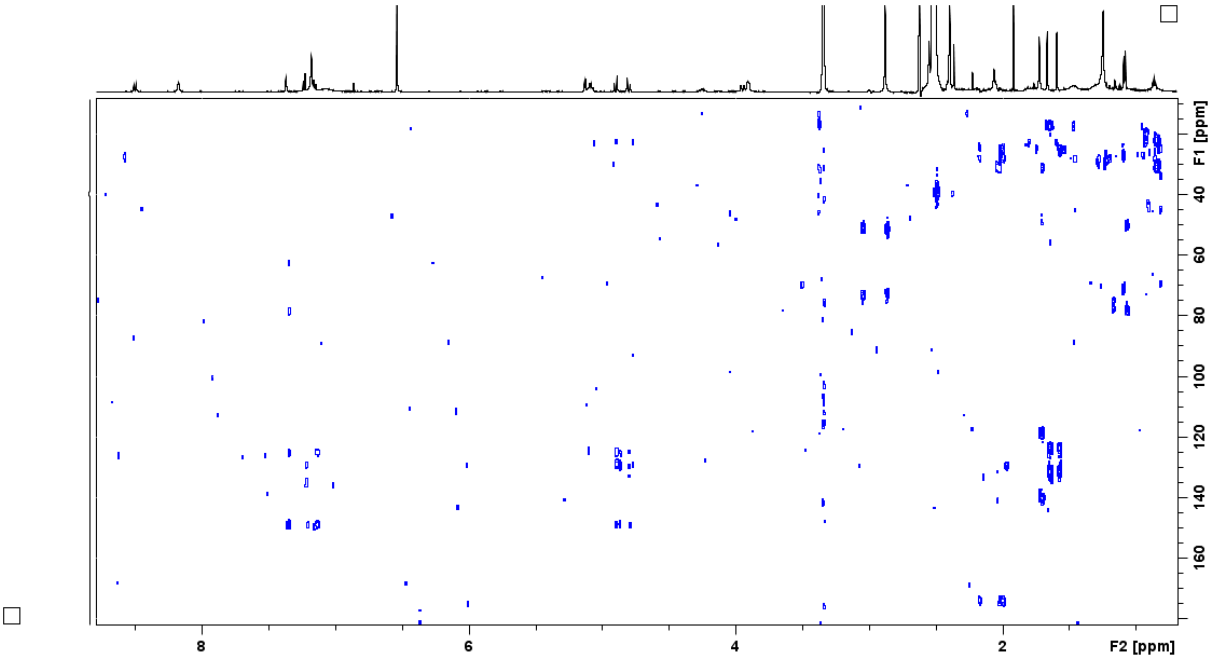
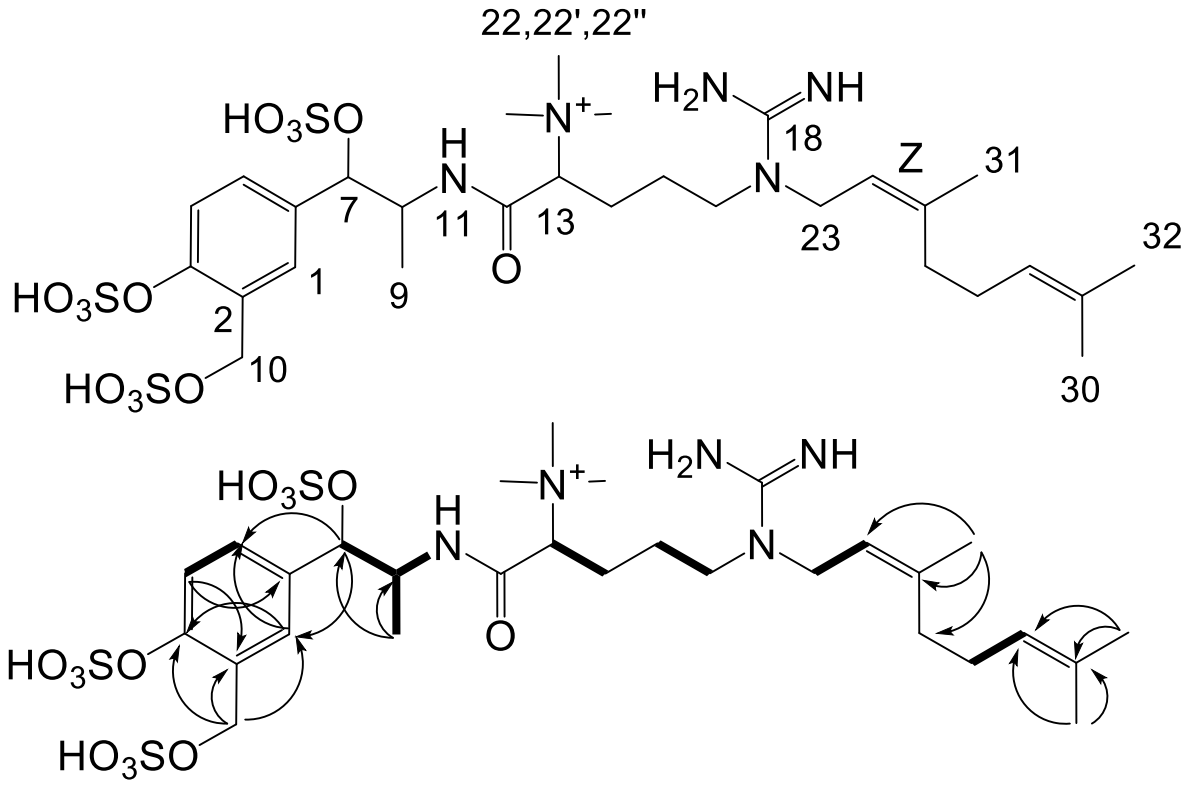


Figure S18. Observed ^{1}H-^{1}H COSY (bold line) and HMBC (arrow) correlations.



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.