Erschienen in: European Journal of Organic Chemistry ; 2017, 31. - S. 4566-4571
https://doi.org/10.1002/ejoc.201700882

Discovery of a Mosaic-Like Biosynthetic Assembly Line with a Decarboxylative Off-Loading Mechanism through a Combination of Genome Mining and Imaging

*Mahsa Mir Mohseni[+], Thomas Höver[+], Lena Barra, Marcel Kaiser, Pieter C. Dorrestein, Jeroen S. Dickschat, and Till F. Schäberle**

Abstract: *The biosynthetic gene cluster for the antiplasmodial natural product siphonazole was identified by using a combination of genome mining, imaging, and expression studies in the natural producer Herpetosiphon sp. B060. The siphonazole backbone is assembled from an unusual starter unit from the shikimate pathway that is extended by the action of polyketide synthases and non-ribosomal peptide synthetases with unusual domain structures, including several split modules and a large number of duplicated domains and domains predicted to be inactive. Product release proceeds through decarboxylation and dehydration independent of the thioesterase SphJ and yields the diene terminus of siphonazole. High variation in terms of codon-usage within the gene cluster, together with the dislocated domain organization, suggest a recent emergence in evolutionary terms.*

For decades natural products have played major roles as biological probes, inspiration for organic chemists and, most prominently, an important source of therapeutic compounds.[1,2] But natural product research currently struggles with two dilemmas. On the one hand, the development of resistance renders many drugs useless, while on the other hand, the development of innovative drugs has been deadlocked over the last two decades, since doubts have arisen concerning the usefulness of natural products as a basis for new leads. One reason is that known compounds are frequently re-isolated in bioactivity-based screening approaches. A still promising approach for today is offered by so far poorly studied groups of organisms, for example, microorganisms associated with insects or from marine habitats, or the diverse group of gliding bacteria that usually exhibit complex life cycles and harbor great potential to produce bioactive compounds.[3–7] Herein, we report on the discovery of the siphonazole biosynthetic gene cluster from *Herpetosiphon* sp. B060, a strain that we isolated from a soil sample from the intertidal zone, through a combination of genome mining, expression analysis, and imaging mass spectrometry (IMS). The gene cluster shows a mosaic-like structure, combines parts from the shikimate pathway and polyketide and non-ribosomal peptide biosynthesis,[8] and makes use of an unusual termination mechanism that was studied in detail.

Siphonazole was isolated from *Herpetosiphon* sp. B060 and its structure elucidated a decade ago,[8] and strategies for its total synthesis have also been developed,[9,10] but details of the biosynthesis, including the timing and the corresponding gene cluster, are unknown. Besides expression analysis, an interesting and fast method to follow the production of a natural product is offered by imaging mass spectrometry (IMS), which can be directly performed on an agar-plate culture. To investigate product distribution and to identify the siphonazole biosynthetic gene cluster, IMS analysis of *Herpetosiphon* sp. B060 grown on agar was performed (Figure 1 a). Siphonazole production was specifically detected at the outer edge of the swarming colonies, that is, only in young cells. Hence, it was shown that this method can be successfully used to investigate strains for which no genetic tools exist.

The draft genome of *Herpetosiphon* sp. B060 was screened for polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS) sequences, resulting in two contigs that encoded gene clusters for hybrid PKS/NRPS metabolites. One of these gene clusters showed high similarity to a pathway present in the related strain *Herpetosiphon aurantiacus*, which does not produce siphonazole. Furthermore, prediction of the adenylation (A) domain substrate specificities for this cluster pointed to the incorporation of ornithine, asparagine, and leucine, but none of these amino acids correlated to the structure of siphonazole (Scheme 1), which is assembled from a methylated 3,4-dihydroxybenzoate starter unit, one glycine and two threonines, three intact acetate units, and C2 of a fourth acetate (Figure S1 in the Supporting Information).[8] The second candidate PKS/NRPS gene cluster contained

[*] Dr. M. Mir Mohseni,[+] Dr. T. Höver,[+] Dr. T. F. Schäberle
Institut für Pharmazeutische Biologie, Universität Bonn
53115 Bonn (Germany)
E-mail: till.schaeberle@uni-bonn.de
L. Barra, Prof. J. S. Dickschat
Kekulé-Institut für Organische Chemie und Biochemie
Universität Bonn, 53121 Bonn (Germany)
M. Kaiser
Schweizerisches Tropen- und Public-Health-Institut (Swiss TPH)
Basel CH-4002 (Switzerland)
M. Kaiser
Universität Basel, CH-4003 Basel (Switzerland)
Prof. P. C. Dorrestein
Collaborative Mass Spectrometry Innovation Center
Skaggs School of Pharmacy and Pharmaceutical Sciences
University of California at San Diego, La Jolla, CA (USA)
[+] These authors contributed equally

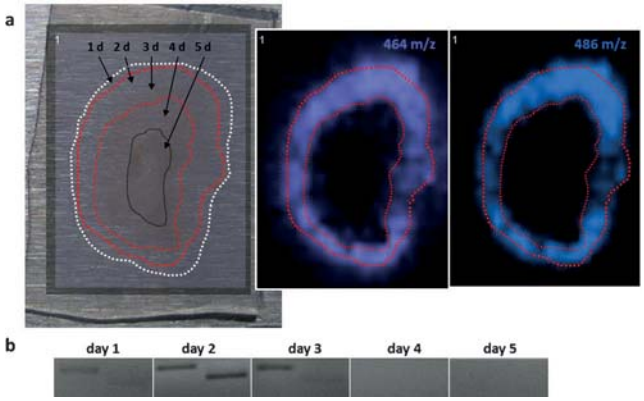
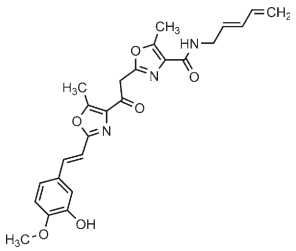


Figure 1. a) IMS analysis of an actively growing agar-plate culture of *Herpetosiphon* sp. B060. The detected ions *m*/*z* 464 and *m*/*z* 486 correspond to siphonazole $[M+H]^+$ and the sodium adduct $[M+Na]^+$. The culture was inoculated in the area indicated by the black line. The red lines indicate the area of the swarming colony, in which the cells are 2–3 days old. b) Time course of the expression of the siphonazole genes *sphA* and *sphB*.



Scheme 1. The structure of siphonazole.

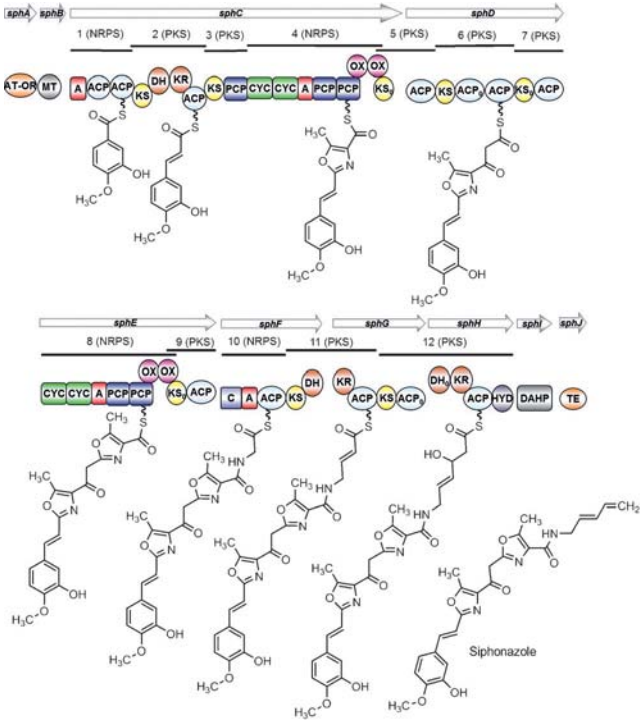


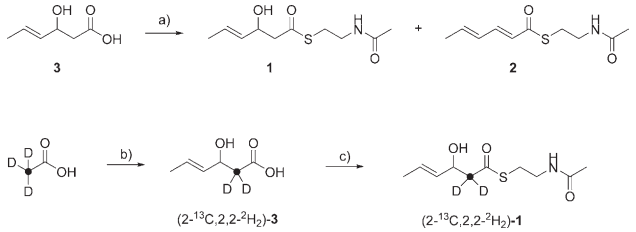
Figure 2. The biosynthetic gene cluster and a biosynthetic model for the assembly of siphonazole. A: adenylation domain, ACP: acyl carrier protein, AT-OR: acyltransferase–oxidoreductase, C: condensation domain, CYC: cyclisation domain, DAHP: 3-deoxy-D-arabinoheptulosonate-7-phosphate synthase, DH: dehydratase, HYD: hydrolase, KR: ketoreductase, KS: ketosynthase, MT: methyltransferase, OX: oxidation domain, PCP: peptidyl carrier protein, TE: thioesterase. Domains that are predicted through bioinformatics to be inactive are indicated by the index 0.

A domains predicted to be specific for glycine and threonine activation, as expected for siphonazole. Furthermore, oxidation domains were identified within the NRPS modules, which are required for oxazole ring formation. Expression level analysis for the cluster genes *sphA* and *sphB* revealed strong expression in young cells, with a maximum at day two (Figure 1 b), which is in line with the results of IMS analysis.

A fosmid library of *Herpetosiphon* sp. B060 genomic DNA was generated and screened with primers designed for specific parts of the cluster, resulting in the identification of three fosmids carrying partial cluster information. DNA sequencing of the three fosmids yielded large parts of the cluster sequence, and the gaps were closed by PCR amplification of genomic DNA, finally yielding the putative biosynthetic gene cluster for siphonazole biosynthesis (Figure 2). The hybrid NRPS/*trans*-AT PKS cluster consists of ten genes, *sphA* to *sphJ*, and spans 50 kb in total (Table S1 in the Supporting Information). RT-PCR with primer pairs covering the end of one gene and the start of the downstream following one revealed that *sphA* and *sphB*, as well as *sphC* to *sphJ*, are transcriptionally coupled units (Figure S2).

The core of the siphonazole biosynthetic gene cluster consists of twelve modules, including four NRPS and eight *trans*-AT PKS modules. Further, the *O*-methyltransferase SphB, a C-terminal hydrolase domain of SphH, and the aldolase SphI are integral parts of the cluster (Figure 2). The activities of these enzymes/domains, as well as activation of the starter unit, were analyzed in detail (see the Supporting Information).

Feeding experiments with (1-^{13}C)acetate revealed that the last acetate is decarboxylated after its incorporation.[8] The last module shows the unusual domain organization KS-ACP_0-DH_0-KR-ACP-HYD, with a C-terminal hydrolase that is unique to the siphonazole cluster. A bioinformatic analysis of the discrete thioesterase (TE) SphJ reveals that this enzyme may have a proof-reading function to release misprimed substrates from ACPs of the siphonazole pathway instead of being responsible for product release. To investigate how the domains of the last module participate in the processing of the fourth incorporated malonyl-CoA unit, and which domains of this module are relevant for product offloading, possibly together with the discrete TE SphJ, the C-terminal part of SphH (ACP-HYD, referred to as $SphH_{Cterm}$) and SphJ were heterologously expressed in *E. coli* and purified (Figures S21, S22). Since the natural substrate in its bound state with the ACP of $SphH_{Cterm}$ is difficult to obtain, enzyme incubation experiments were performed with short substrate mimics that were synthesized starting from 3-hydroxyhex-4-enoic acid (**3**).[11] The acid was transformed into the respective *N*-acetylcysteamine thioester (SNAC ester) **1** by using EDC as a coupling reagent (Scheme 2). As



Scheme 2. Synthesis of substrate analogues **1**–**3** used for the in vitro incubation experiments. a) EDC, DMAP, HSNAC, 70% of **1** and 4% of **2**; b) *n*BuLi (1 equiv), LDA (1 equiv);, then (*E*)-crotonaldehyde, 85%; c) DCC, DMAP, HSNAC, 47%. EDC = 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide, DMAP = 4-dimethylaminopyridine, HSNAC = *N*-acetylcystamine, LDA = lithium diisopropylamide, DCC = *N*,*N*′-dicyclohexylcarbodiimide.

a side product of this reaction the SNAC ester of sorbic acid (**2**) was isolated. Furthermore, for investigation of the enzyme mechanism, the isotopically labelled compound (2-^{13}C,2,2-^{2}H$_2$)-**1** was synthesized through aldol addition of (2-^{13}C,2,2-^{2}H$_3$)acetic acid to (*E*)-crotonaldehyde and conversion into the SNAC ester.

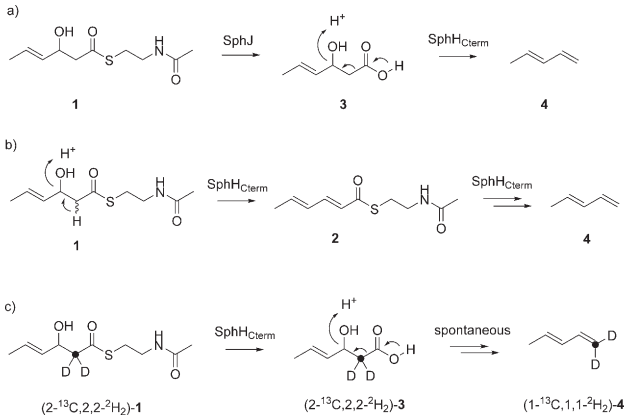
The synthetic substrate analogues were incubated with different enzyme preparations, and the production of pentadiene, which is formed analogously to the natural product siphonazole, was detected by capturing the volatiles from the incubation reaction on a solid-phase microextraction (SPME) fiber followed by GC/MS analysis.[12] Incubation of **1** with SphH$_{Cterm}$ and SphJ resulted in a strong production of pentadiene (**4**; Table 1 and Figure S14), which is in line with

Table 1: Results from incubation experiments with substrates **1**–**3**, (2-^{13}C,2,2-^{2}H$_2$)-**1**, and different enzyme combinations.

Substrate	SphH$_{Cterm}$ + SphJ	SphJ	SphH$_{Cterm}$	Control
1	+	−	+	−
(2-^{13}C,2,2-^{2}H$_2$)-**1**	n.d.	n.d.	+	−
2	−	−	−	−
3	n.d.	n.d.	+	+

+: pentadiene production, −: no pentadiene production, n.d.: not determined

a mechanism of initial hydrolysis of the thioester **1** to the acid **3** by SphJ, followed by a decarboxylation and dehydration by SphH (Scheme 3a), but this experiment does not always allow exclusion of the possibility that only one of the enzymes participates in the reaction. Therefore, the thioester **1** was also incubated with either SphJ or SphH$_{Cterm}$ alone, and indeed incubation of **1** with SphJ did not result in the production of **4** (only traces were detected, as in the control experiment without enzyme), while incubation with SphH$_{Cterm}$ alone resulted in the formation of **4** in similar amounts to those observed following incubation with both enzymes. These data call into question the involvement of the TE SphJ in the termination mechanism of siphonazole biosynthesis and further support a sole proof-reading function for this enzyme, while suggesting that product release, decarboxylation, and dehydration are all catalyzed by SphH$_{Cterm}$.



Scheme 3. Possible mechanisms for product release in siphonazole biosynthesis. a) Thioester hydrolysis by SphJ and dehydration/decarboxylation by SphH$_{Cterm}$. b) Initial dehydration by SphH$_{Cterm}$ followed by thioester hydrolysis and decarboxylation by the same enzyme. c) Thioester hydrolysis by SphH$_{Cterm}$ followed by spontaneous/concerted dehydration and decarboxylation.

The question of which of these three reactions occurs first was investigated through incubation experiments with substrates (2-^{13}C,2,2-^{2}H$_2$)-**1**, **2**, and **3**. If thioester hydrolysis were first, then the subsequent steps of dehydration and decarboxylation could be a concerted process (Scheme 3b). This was studied by using the isotopically labelled substrate (2-^{13}C,2,2-^{2}H$_2$)-**1**, which may first be hydrolyzed to the free acid (2-^{13}C,2,2-^{2}H$_2$)-**3**. Protonation of the hydroxy function in (2-^{13}C,2,2-^{2}H$_2$)-**3** and elimination of water could induce spontaneous decarboxylation, and in this mechanism, both deuterium atoms and the ^{13}C-labelling should be retained in the product (1-^{13}C,1,1-^{2}H$_2$)-**4**. As shown in an incubation experiment with SphH$_{Cterm}$ and (2-^{13}C,2,2-^{2}H$_2$)-**1**, this was the case (Figure S15; a mass spectrum of the obtained (1-^{13}C,1,1-^{2}H$_2$)-**4** is shown in in Figure S16). The results of this labelling experiment also speak against a mechanism in which the dehydration of **1** to **2** is the first step, because this would require the loss of one deuterium from C$_\alpha$. Furthermore, incubation of **2** with neither SphH$_{Cterm}$, SphJ, nor both yielded **4** (Table 1 and Figure S17), which is in agreement with the function of SphJ as proof-reading TE. If ester hydrolysis is the first step, then substrate **3** should be converted into **4** by SphH$_{Cterm}$, as was experimentally observed (Table 1 and Figure S18), but similar amounts of **4** were detected in the control experiment without enzyme, thus suggesting that SphH$_{Cterm}$ only catalyzes the thioester hydrolysis, while the decarboxylation–dehydration of **3** to **4** is likely a spontaneous reaction in water. However, it cannot be excluded that catalysis by SphH$_{Cterm}$ is required for conversion of the natural substrate.

Siphonazole was tested for its biological activity. Cytotoxic activity was observed, but the effect (the mean IC$_{50}$ value against a panel of 36 cancer cell lines was 5.90 μg mL^{-1}, 12.74 μM) was too weak for further evaluation as an anticancer agent. At non-cytotoxic concentrations, no antiviral activity was observed (tested: coxsackie virus B3, influenza virus A, herpes simplex virus type 1). Antibacterial or anti-

fungal effects were not observed. Instead, siphonazole showed activity against *Plasmodium falciparum* (IC_{50}: 0.59 μg mL^{-1}, 1.27 μM; Table S2).

The siphonazole biosynthetic gene cluster shows many idiosyncrasies, which is typical for *trans*-AT PKSs and hints at recombination events through horizontal gene transfer. Until now, most approaches to mutate a cluster in a target-oriented way have remained ineffective. *Herpetosiphon* sp. B060 is an example where nature has relatively recently performed recombinatorial biosynthesis, and obtained a natural product—the desired output for synthetic biology. Future investigation of the siphonazole cluster might provide new insight into this field to create "unnatural natural products". Furthermore, new tools for synthetic biology are desirable. The investigated thioester hydrolysis, which is a prerequisite for the terminal, presumably spontaneous decarboxylation–dehydration of the released molecule, by $SphH_{Cterm}$ represents such a novel functionality within PKS systems. A similar terminal alkene moiety was observed in the biosynthesis of the anticancer compound curacin A.[13] However, in curacin A biosynthesis, a sulfotransferase encoded in the terminal module first transfers a sulfonate group to the β-hxdroxy group, making it a better leaving group. The subsequent decarboxylative elimination is catalyzed by a TE. By contrast, in siphonazole biosynthesis, $SphH_{Cterm}$ is sufficient, and the simplified substrate mimic used indicates promiscuity of this enzyme, which could be used in further transformations. The promising antiplasmodial activity of siphonazole will be subject of future investigations, since additional economically priced treatment options are still needed to achieve reductions in the *Plasmodium falciparum* malaria burden.[14]

In this report, a work flow is illustrated that allows one to link a metabolite of interest directly to its biosynthetic gene cluster. Such approaches are needed to speed up gene cluster identification, even when only draft genomes are available. This holds especially true for organisms that are not genetically accessible. Using a mass spectrometry based technology such as IMS, which was established to observe the chemical output and metabolic exchange between an organism and the environment directly on agar plates,[15,16] facilitates the linking of a metabolite to its biosynthetic gene cluster. Increasing sequencing efforts make clear the discrepancy between computationally identified biosynthetic gene clusters and known metabolites, and this resource should be made use of. New microorganisms, preferably from underinvestigated ecological niches, should be domesticated, as recently shown for the teixobactin producer,[17] and metagenomics approaches should be intensified to create a basis for identifying gene loci for new lead compounds. Then, a mass spectrometry based analysis in combination with expression profiles should enable the prompt linking of genes and metabolites.

Acknowledgments

M.M.M. was funded by NRW International Graduate Research School BIOTECH-PHARMA. Work in the lab of T.F.S. was funded by the Federal Ministry of Education and Research (BMBF). Boehringer Ingelheim Fonds is thanked for a travel grant to T.F.S.

Keywords: biosynthesis · genome mining · imaging mass spectrometry (IMS) · natural products · siphonazole

[1] C. T. Walsh, M. A. Fischbach, *J. Am. Chem. Soc.* **2010**, *132*, 2469 – 2493.
[2] D. J. Newman, G. M. Cragg, *J. Nat. Prod.* **2012**, *75*, 311 – 335.
[3] T. F. Schäberle, F. Lohr, A. Schmitz, G. M. König, *Nat. Prod. Rep.* **2014**, *31*, 953 – 972.
[4] S. C. Wenzel, R. Müller, *Curr. Opin. Drug Discovery Dev.* **2009**, *12*, 220 – 230.
[5] C. Beemelmanns, H. Guo, M. Rischer, M. Poulsen, *Beilstein J. Org. Chem.* **2016**, *12*, 314 – 327.
[6] A. Dávila-Céspedes, P. Hufendiek, M. Crüsemann, T. F. Schäberle, G. M. König, *Beilstein J. Org. Chem.* **2016**, *12*, 969 – 984.
[7] J. Korp, M. S. V. Gurovic, M. Nett, *Beilstein J. Org. Chem.* **2016**, *12*, 594 – 607.
[8] M. Nett, O. Erol, S. Kehraus, M. Köck, A. Krick, E. Eguereva, E. Neu, G. M. König, *Angew. Chem. Int. Ed.* **2006**, *45*, 3863 – 3867; *Angew. Chem.* **2006**, *118*, 3947 – 3951.
[9] J. Linder, C. J. Moody, *Chem. Commun.* **2007**, 1508 – 1509.
[10] J. Zhang, M. A. Ciufolini, *Org. Lett.* **2009**, *11*, 2389 – 2392.
[11] B. D. Ames, C. Nguyen, J. Bruegger, P. Smith, W. Xu, S. Ma, E. Wong, S. Wong, X. Xie, J. W. Li, J. C. Vederas, Y. Tang, S. C. Tsai, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 11144 – 11149.
[12] J. S. Dickschat, *Nat. Prod. Rep.* **2014**, *31*, 838 – 861.
[13] L. Gu, B. Wang, A. Kulkarni, J. J. Gehret, K. R. Lloyd, L. Gerwick, W. H. Gerwick, P. Wipf, K. Håkansson, J. L. Smith, D. H. Sherman, *J. Am. Chem. Soc.* **2009**, *131*, 16033 – 16035.
[14] P. G. Walker, J. T. Griffin, N. M. Ferguson, A. C. Ghani, *Lancet Glob. Health.* **2016**, Jun 3. pii: S2214-109X(16)30073-0.
[15] Y. L. Yang, Y. Xu, P. Straight, P. C. Dorrestein, *Nat. Chem. Biol.* **2009**, *5*, 885 – 887.
[16] a) Y. L. Yang, Y. Xu, R. D. Kersten, W. T. Liu, M. J. Meehan, B. S. Moore, N. Bandeira, P. C. Dorrestein, *Angew. Chem. Int. Ed.* **2011**, *50*, 5839 – 5842; *Angew. Chem.* **2011**, *123*, 5961 – 5964.
[17] L. L. Ling, T. Schneider, A. J. Peoples, A. L. Spoering, I. Engels, B. P. Conlon, A. Mueller, T. F. Schäberle, D. E. Hughes, S. Epstein, M. Jones, L. Lazarides, V. A. Steadman, D. R. Cohen, C. R. Felix, K. A. Fetterman, W. P. Millett, A. G. Nitti, A. M. Zullo, C. Chen, K. Lewis, *Nature* **2015**, *517*, 455 – 459.