Characterization of an Orphan Type III Polyketide Synthase Conserved in Uncultivated "Entotheonella" Sponge Symbionts

Silke Reiter,[a, b] Jackson K. B. Cahn,[a] Vincent Wiebach,[a] Reiko Ueoka,[a] and Jörn Piel*[a]

Uncultivated bacterial symbionts from the candidate genus "Entotheonella" have been shown to produce diverse natural products previously attributed to their sponge hosts. In addition to these known compounds, "Entotheonella" genomes contain rich sets of biosynthetic gene clusters that lack identified natural products. Among these is a small type III polyketide synthase (PKS) cluster, one of only three clusters present in all known "Entotheonella" genomes. This conserved "Ento-theonella" PKS (*cep*) cluster encodes the type III PKS CepA and the putative methyltransferase CepB. Herein, the characterization of CepA as an enzyme involved in phenolic lipid biosynthesis is reported. In vitro analysis showed a specificity for alkyl starter substrates and the production of tri- and tetraketide pyrones and tetraketide resorcinols. The conserved distribution of the *cep* cluster suggests an important role for the phenolic lipid polyketides produced in "Entotheonella" variants.

Introduction

Type III polyketide synthases (PKSs) generate a wide range of cyclic compounds, including chalcones, stilbenes, flavonoids, α-pyrones, and resorcinols.[1] Originally thought to be limited to plants, diverse type III PKSs from bacteria have been reported over the last two decades.[1a, 2] Their biosynthetic products have a multitude of important functions, for example in providing precursors for antibiotics, such as in vancomycin biosynthesis,[3] as alternative electron acceptors in mycobacterial respiration,[4] or in conferring antibiotic resistance.[5] Polyketides are assembled from simple acyl coenzyme A (CoA) building blocks in repeated rounds of decarboxylative Claisen-like condensations. For type III PKS products, structural diversity is defined by the choice of starter (e.g., acetyl, cinnamoyl, or fatty acyl) and extender units (most commonly malonyl or methylmalonyl), as well as by different cyclization modes through Claisen condensation, aldol condensation, or lactonization.[1a, 2, 6] In addition to these PKS-catalyzed reactions, auxiliary enzymes, such as methyltransferases (MTs) or oxidoreductases, can further modify the polyketides during or after chain elongation. Type III PKSs combine all basic PKS enzymatic functionalities in a single catalytic center of a small homodimeric 40–47 kDa enzyme. They do not necessarily require an acyl carrier protein (ACP), but can incorporate units directly from CoA, unlike other known PKS types.

We, and collaborators, have recently identified bacterial sponge symbionts of the candidate genus "Entotheonella" as rich producers of bioactive metabolites.[7] These unusual filamentous bacteria belong to the proposed candidate phylum "Tectomicrobia", which exclusively comprises uncultivated members, and represent the first chemically rich producer taxon from microbial dark matter. Marine sponges and their bacterial communities are in the focus of research as a drug-discovery resource and ancient models for animal–bacteria symbioses.[7a] For the sponge chemotype *Theonella swinhoei* Y, a variant with a yellow interior, we traced almost all of its many bioactive natural products to the producer "*Candidatus* Entotheonella factor" and its close relative "*Ca.* Entotheonella gemina" (Table 1).[7b, c] In contrast, the symbiont "*Ca.* Entotheonella serta" is a source of natural products in chemically dis-

[a] Dr. S. Reiter, Dr. J. K. B. Cahn, V. Wiebach, Dr. R. Ueoka, Prof. Dr. J. Piel
Department of Microbiology, Swiss Federal Institute of Technology
Vladimir-Prelog-Weg 4, 8093 Zürich (Switzerland)
E-mail: jpiel@ethz.ch

[b] Dr. S. Reiter
Institute for Insect Biotechnology, Justus Liebig University Giessen
Heinrich-Buff-Ring 26–32, 35392 Giessen (Germany)

Supporting information and the ORCID identification numbers for the
authors of this article can be found under https://doi.org/10.1002/
cbic.201900352.

Table 1. Sponges and their "Entotheonella" symbionts.

Sponge	Collection	"Entotheonella" phylotype	Known natural products
T. swinhoei Y	Japan	"E. factor", "E. gemina"	polytheonamides, onnamides, theopederins, orbiculamides, cyclotheonamides, pseudo-theonamides, nazumamide A
T. swinhoei WA	Japan	"E. serta" TSWA1	misakinolides, theonellamides
T. swinhoei WB	Israel	"E. serta" TSWB1, "Entotheonella" TSWB2	swinholides, theonellamides

tinct sponges *T. swinhoei* WA and WB (white interior).[7b–d,8] It is likely that many natural products of "Entotheonella" are involved in host defense, which is a role often suspected or shown for toxic compounds of sessile invertebrates.[8–10]

The "Entotheonella" genomes known to date contain large sets of orphan biosynthetic gene clusters (BGCs) that are mostly unique to one particular "Entotheonella" variant and lack characterized products.[7b–d,8d] Herein, we examine the BGCs that are conserved among "Entotheonella" and might contribute functions of broader importance to these poorly understood symbionts.[3–5,11] Of the three identified BGC families, we focus on functional analyses of the conserved "Entotheonella" PKS (*cep*) cluster; this is a type III PKS system of which we have identified five orthologous versions. In vitro enzyme characterization suggests that the products of this cluster are quinones with long-chain fatty acyl starter moieties, which are reminiscent of mycobacterial redox cofactors.[4]

Results and Discussion

The *cep* cluster is one of only three conserved "Entotheonella" BGCs in *T. swinhoei*

In our dataset on *T. swinhoei* Y, we detected only three gene clusters that were shared among its symbionts "E. factor" and close relative "E. gemina" (Figure 1).[7b] One was the *cep* BGC
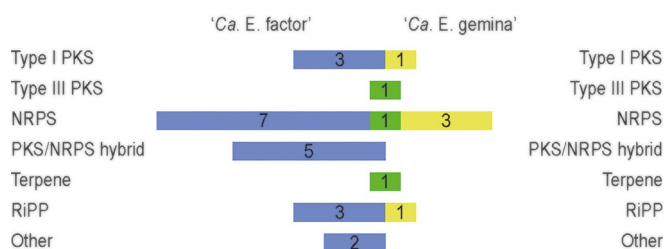

Figure 1. Gene clusters and gene cluster fragments present in the genomes of *E. factor* (blue) and *E. gemina* (yellow); thus suggesting minimal chemical overlap (green). Gene clusters were detected by using antiSMASH v4 and confirmed by manual inspection. In some cases, a shared cluster was not detected by antiSMASH due to assembly gaps, but could be clearly identified by a BLAST search of each gene cluster against the opposing genome.[7b] NRPS, nonribosomal peptide synthetase; RiPP, ribosomally synthesized and post-translationally modified peptide.

encoding a type III PKS (CepA) and a putative MT (CepB), the second a monomodular NRPS system, and the third a putative carotenoid-type terpene BGC (Figure S1 in the Supporting Information). The small BGC overlap is striking, considering the large number of natural product clusters. Focusing on the *cep* cluster for further analyses, we examined symbionts from a broader range of sponges for the occurrence of this BGC. This search identified three further *cep*-type clusters in the sponges *T. swinhoei* WA and WB that were collected in Japan and Israel, respectively (Table 1). The WA *cep* variant was present on a cosmid isolated from a metagenomic library of an enriched bacterial cell pellet of the sponge and assigned to the symbiont "E. serta" TSWA1.[7c] Two *cep* clusters existed in metage-

nomic sequence data of the Israeli WB chemotype.[7c] Single-copy phylogenetic markers suggested that this sponge contained an "E. serta" variant, denoted TSWB1, with about 99% nucleotide identity to "E. serta" TSWA1, and an additional new "Entotheonella" phylotype that is herein provisionally referred to as "Entotheonella" TSWB2 (Table S1). However, the functionally identical read coverage and nucleotide frequencies of the two "Entotheonella" phylotypes prevented separation of these genomes by binning methods. The *cep* clusters were identified by performing tblastn searches on the assembled contigs of the binned two-genome dataset; the putative Cep cluster of "E. serta" TSWB1 has >99% nucleotide identity to "E. serta" TSWA1, whereas that of "Entotheonella" TSWB2 has 78–79% identity (Table S2).

All five *cep* BGCs form 1.8 kb operons that contain the PKS-MT gene pair, but are otherwise located in different genomic environments (Figure 2).[7b] The *cepA* and *cepB* genes are either
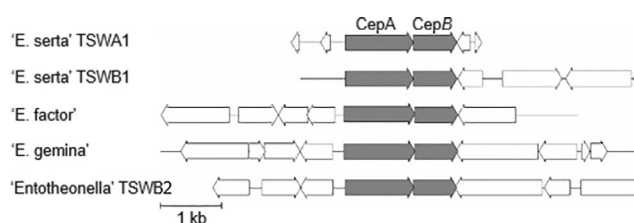

Figure 2. Schematic depiction of *cep* clusters and surrounding genomic regions in different "Entotheonella" variants. Gray arrows indicate *cepA* type III PKS genes and *cepB* MT genes; white arrows indicate putative non-PKS genes; black centered lines represent genome sequence that are 3 kb up- and downstream, if available. Accession numbers: "E. serta" TSWA1 LC043436; "E. serta" TSWB1 MG844359; "E. factor" KI931731; "E. gemina" KI929875; "Entotheonella" TSWB2 MK988560.

separated by one base pair ("E. serta" TSWA1, "E. serta" TSWB1) or overlap by 8 bp ("E. factor", "E. gemina", "Entotheonella" TSWB2). Amino acid and nucleotide sequence identities were >73% between all PKS and MT homologues, respectively (Table S2), strongly suggesting that the five *cep* clusters are orthologous. Their presence in all examined "Entotheonella" variants and the wide geographical distribution of the sponge hosts suggests that the PKS product is important for the bacteria. Examination of the genomic regions up- and downstream of the *cep* cluster (if the sequence was available) did not reveal other genes typically involved in polyketide biosynthesis and showed high variability, which indicated that the BGC consisted of only *cepA* and *cepB*. CepB resembles members of the *S*-adenosyl methionine (SAM)-dependent MT superfamily and contains the glycine-rich SAM binding motif conserved in these enzymes.[12] Querying the Swiss-Prot database with CepB yielded MenG from *Staphylococcus* species and PigF from *Serratia* sp. ATCC 39006 as the most similar proteins, with sequence identities between 26 and 28% and poor coverage (Table S3). MenG is a C-methyltransferase (C-MT) that converts 1,4-dihydroxy-2-naphthoate into menaquinol,[13] whereas PigF is an O-methyltransferase (O-MT) that methylates hydroxy-bipyrrole carbaldehyde to the methoxy product.[14] Thus, the moiety methylated by CepB remained unclear.

Sequence alignments of the CepA homologues show Cys, His, and Asn residues at positions equivalent to the conserved catalytic triad of type III PKSs (Figure S2), which is important for the condensation activity.[1a] The closest homologue in the Swiss-Prot/UniProt database was the Pks18 α-pyrone synthase conserved in the *Mycobacterium tuberculosis* complex (Table S4). Pks18 generates alkyl tri- and tetraketide pyrones (Figure 3) with a high specificity for long-chain starter substra-
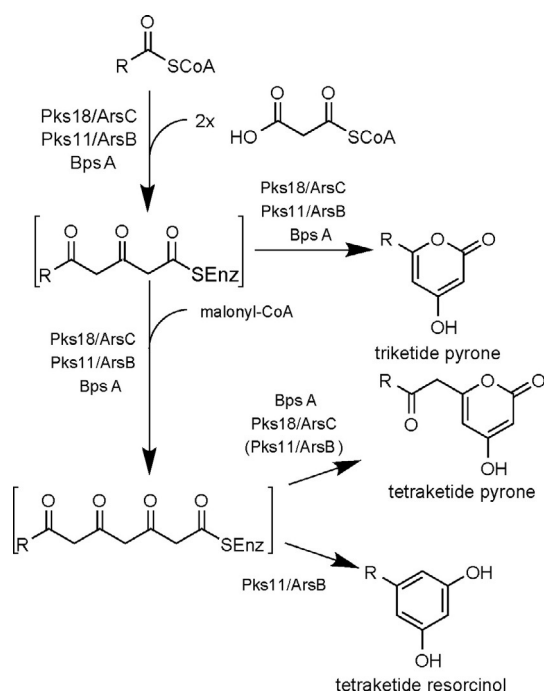

Figure 3. Reactions catalyzed by Pks18 and other type III PKSs that produce either pyrones (Pks18, BpsA, ArsC) or resorcinols and pyrones (Pks11, ArsB).[1c, 4, 17, 18] The Ars enzymes are two related type III PKSs, which both occur in the same *A. vinelandii* strain. R: long-chain alkyl residues, Enz, enzyme.

tes.[1c, 4, 15] In Pks18, a substrate-binding tunnel appears to facilitate the binding of long-chain alkyl starter units.[16] Three residues (Thr144, Cys205, and Ala209) are crucially involved in determining the tunnel volume and are therefore thought to influence substrate preference. A similar substrate profile was shown for the *Bacillus subtilis* enzyme BpsA,[17] which produces a mixture of alkyl pyrones and alkyl resorcinols and features a Pks18-like substrate-binding tunnel with Thr, Cys, and Phe residues at the corresponding positions.[17] The CepA enzymes contain identical residues to those of BpsA, with the exception of "E. factor", which features a Leu in place of Phe (Thr137, Cys198, Leu202; Figure S2). Therefore, we suspected a similarly structured tunnel, and thus, substrate range for CepA.

Phenolic lipid synthases, such as Pks18 and BpsA, are a group of type III PKSs that produce pyrones or resorcinols by using long-chain acyl-CoAs rather than malonyl-CoA or aromatic molecules as starter building blocks (Figure 3).[2b] The presence or absence of a tryptophan residue in the active center has been proposed as the most important determinant of whether a resorcinol (e.g., Trp281 in ArsB) or a pyrone (Gly284

in ArsC, Leu266 in Pks18) is produced.[1c, 18] This site aligns with Trp243 in all CepA variants (Figure S2), which suggests resorcinols as conserved "Entotheonella" polyketides. The alignment also shows an Ala residue for all CepA variants (i.e., Ala309 CepA$_{factor}$) (Figure S2) at a position equivalent to an Arg residue involved in ACP binding in all three major PKS types (types I, II, and III).[1a, b, 19] Interestingly, in bacterial type III PKSs this residue is almost always an Arg or a Lys.[1a] At least one bacterial enzyme with an Arg site (SCO7671, *Streptomyces coelicolor*) was confirmed to use ACP-bound starters.[4] Bacterial type III PKSs featuring a Lys were reported to use ACP- and/or CoA-bound substrates or to directly cooperate with type I fatty acid synthases (ArsB and C), although ACP interactions have not been studied in every case.[1a, b, 20] By containing an Ala at this location, all five CepA variants resemble plant type III PKSs, which use CoA- rather than ACP-bound units.[1a] Among 324 bacterial type III PKS sequences used by Shimizu et al. to establish a functional classification based on phylogeny, only eight homologues showed this plantlike feature.[6c] However, none of these Ala-bearing bacterial type III PKSs have been functionally described, to date. In summary, based on our bioinformatic analyses, we suspected that CepA was most likely a resorcinol synthase that accepted long-chain fatty acid starters directly from CoAs. Because no orthologues with comparably high similarity scores were found in cultivable and genetically accessible organisms, we decided to assess the CepA function in vitro.

Expression and in vitro analysis of CepA

Due to the lack of culturable heterologous hosts related to "Entotheonella", we chose to express *cepA* from "E. factor" TSY1 (*cepA*$_{factor}$) in *Escherichia coli* for in vitro characterization. Sequences were amplified, as described below, from enriched metagenomic DNA by using PCR primers with appropriate restriction sites (Table S5). The amplified *cepA*$_{factor}$ gene was cloned into the expression vector pHis-8 to create recombinant genes that coded for N-terminally octahistidinyl-tagged proteins.[21] In preliminary tests, purified recombinant CepA$_{factor}$ was incubated with a range of starter test substrates and ma-
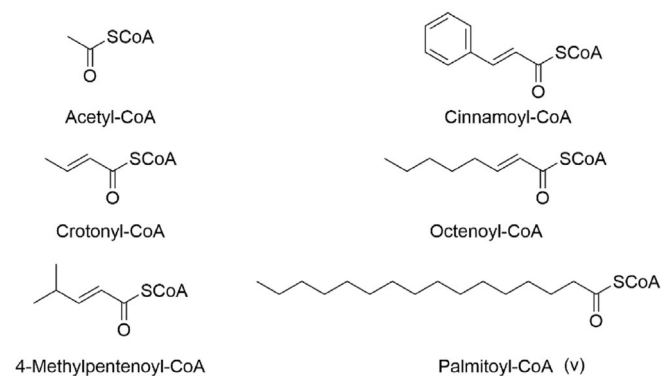

Figure 4. Structures of potential PKS starters used as test substrates in this study. SCoA: coenzyme A thioester."(✓)" indicates a starter substrate for which incorporation could be detected.

lonyl-CoA as an extender substrate. Saturated and unsaturated straight-chain CoA derivates, as well as aromatic CoA derivatives, were tested as starter substrates (Figure 4). The assay was monitored by detection of free thiol groups, arising from hydrolytic CoA release, with 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB).[22] In addition to different starter test substrates, the enzyme was assayed with either a mixture of malonyl-CoA and methylmalonyl-CoA or with only malonyl-CoA. It was shown that type III PKSs 10 and 11 from mycobacteria incorporated methylmalonyl-CoA and that tetraketides with incorporated methylmalonyl-CoA favored aldol condensation, relative to tetraketides elongated with malonyl-CoA alone.[4] For our enzyme, no difference was observed for DTNB assays with CoA mixtures versus malonyl-CoA alone (Figure S3H–I).

A moderate increase in free thiols was observed for reactions containing either acetyl-CoA, cinnamoyl-CoA, or palmitoyl-CoA as starters and malonyl-CoA as extenders (Figure S3). No conversion was detected with malonyl-CoA as the sole thioester (Figure S3E). Ethyl acetate extracts of the assays containing different starter test substrates (Figure 4) and malonyl-CoA as the elongation substrate were analyzed by means of HPLC–ESI-MS. Using a python script to analyze the resulting data, we detected new products, compared with the boiled and no-enzyme controls, only for the assay mixture containing

Table 2. Theoretical and measured high-resolution masses of detected products. The m/z values refer to the $[M-H]^-$ ions.

Compound	Formula	m/z measured	m/z calculated	Δ (ppm)
1	$C_{20}H_{33}O_3$	321.2435	321.2435	0
2	$C_{22}H_{35}O_4$	363.2541	363.2541	0
3	$C_{21}H_{35}O_2$	319.2642	319.2643	−0.31
4	$C_{21}H_{35}O_3$	335.2593	335.2591	0.6
5	$C_{21}H_{33}O_3$	333.2436	333.2435	0.3
1	$^{13}C_4C_{16}H_{33}O_3$	325.2560	325.2569	−2.77
2	$^{13}C_6C_{16}H_{35}O_4$	369.2731	369.2742	−2.98

palmitoyl-CoA as a starter (Figure S4). A possible reason for the acetyl-CoA and cinnamoyl-CoA samples being negative, in contrast to the DTNB assays, might be unspecific hydrolysis by the enzyme without successive elongation. In the reaction with palmitoyl-CoA (Figure S4C), new products were observed at m/z 321.2435 (**1**) and 363.2541 (**2**), which corresponded to molecular formulae of $C_{20}H_{33}O_3$ and $C_{22}H_{35}O_4$, respectively (Table 2 and Figure S5).

Taking into account the available substrates, these predicted formulae were in accordance with the use of malonyl-CoA as the sole elongation unit and palmitoyl-CoA (C_{16}) as the starter, corresponding to a cyclic triketide ($2 \times$ malonyl-CoA $+ 1 \times$ palmi-
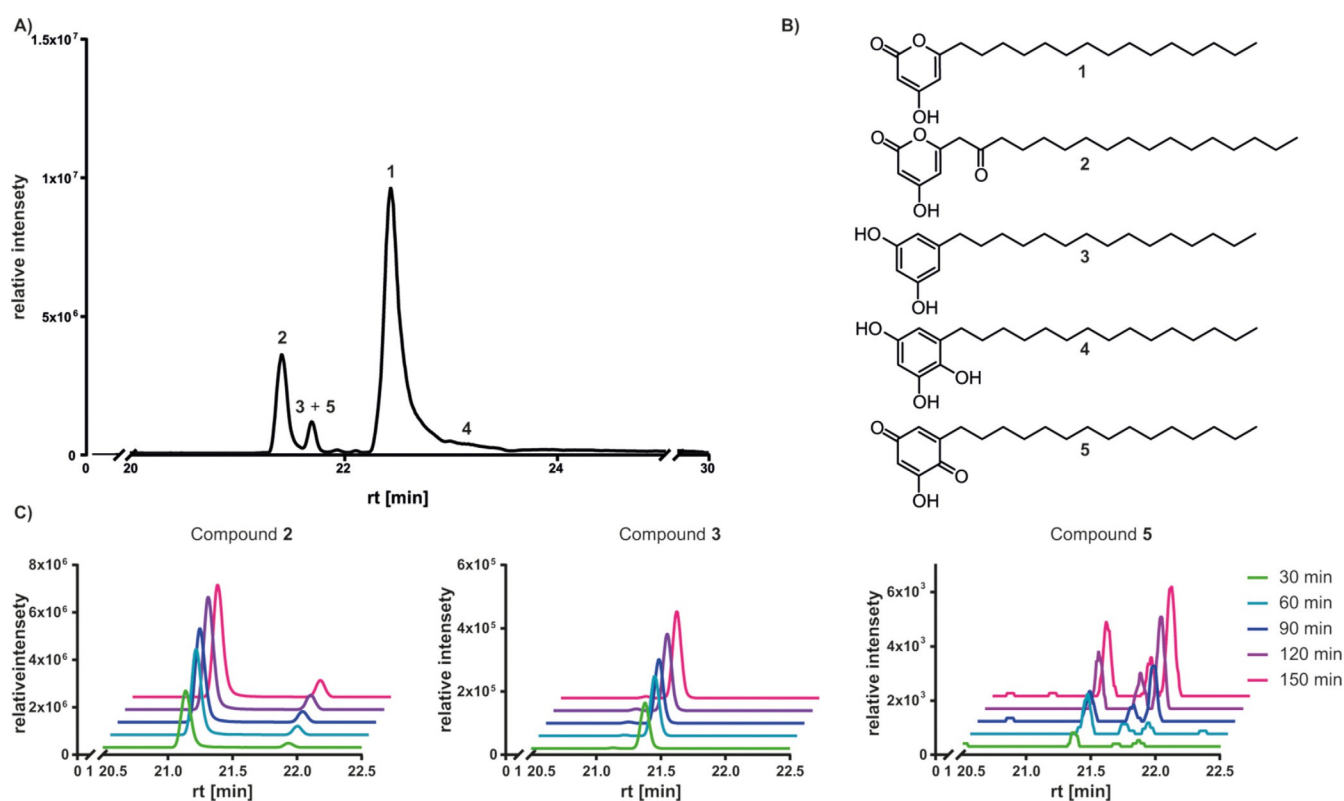

Figure 5. Pyrone and resorcinol formation by CepA. A) Summed extracted ion chromatograms (EICs) of all proposed product ions under optimized enzymatic reaction conditions with palmitoyl- and malonyl-CoA. B) Proposed structures for 1–5; **1** and **2** are major products detected in preliminary and optimized functional assays; **3**–**5** are minor products only detected under optimized reaction conditions. C) EICs for **2**, **3**, and **5** for time-course experiments, suggesting rapid formation of resorcinol (**3**) and the delayed occurrence of quinone (**5**). A sample of 100 μL was taken every 30 min over a total of 6 h. Multiple peaks in the chromatogram of **2** are most likely the two keto–enol tautomers.[23] The prominent peak corresponds to the proposed structure shown in B). Multiple peaks in the chromatogram of **5** are most likely a result of two possible regioisomers, namely, *ortho*- and *meta*-quinones, and their tautomers. For negative control and boiled enzyme control data, as well as data on compounds **1** and **4**, see Figures S17 and S18.

toyl-CoA) and a cyclic tetraketide ($3 \times$ malonyl-CoA $+ 1 \times$ palmitoyl-CoA), respectively (Figures 3 and 5B). The presence of three and four oxygen atoms, as well as the mass difference to the hypothetical linear polyketides, corresponding to the loss of one H_2O equivalent, supported the pyrone structure. At first glance, these results are inconsistent with the bioinformatic prediction of CepA being a resorcinol synthase. However, pyrone formation has also been observed to occur spontaneously,[1c] and resorcinol synthases often show a shifted product range towards pyrones in vitro, compared with that in vivo.[1c, 18] To obtain further evidence for the structure by means of NMR spectroscopy, we aimed to increase the activity of the enzyme by optimizing buffer conditions at different salt concentrations and pH values. Under the best conditions (pH 6 and 500 mM phosphate buffer), we observed, in addition to increased levels of **1** and **2**, new ions at m/z 319.2642 (**3**), 335.2593 (**4**), and 333.2436 (**5**; Figures 5A and S5). The predicted molecular formulae, $C_{21}H_{35}O_2$, $C_{21}H_{35}O_3$, and $C_{21}H_{33}O_3$, suggested the formation of tetraketides.

The odd carbon number and lower relative oxygen content indicated intramolecular decarboxylative resorcinol formation (**3**) and, for **4** and **5**, spontaneous oxidation (Table 2 and Figure S6);[1c,4,17] this is in agreement with the bioinformatic prediction.

Although product levels for **1** and **2** were increased in the optimized assays, it was not possible to purify sufficient material for NMR spectroscopy based structure elucidation. To facilitate NMR spectroscopy studies, the assay was repeated with malonyl-CoA with a fully ^{13}C-labeled malonyl group. Correspondingly, ions with m/z 325.2560 and 369.2731, supporting the incorporation of two (**1**) and three (**2**) labeled extender units, respectively, were observed in the extract (Table 2 and Figure S7). NMR spectroscopy based structure elucidation in the case of ^{13}C-labeled compound **1** and MS/MS analysis in the case of compound **2** supported the tri- and tetraketide structures proposed in Figure 5B; analytical details are described in the next section. In these experiments, the structures of **3–5** could not be determined because of the very small amounts produced. The proposed structure for **3** was supported by a comparison of retention times obtained by means of HPLC–ESI-MS to a commercial standard of 5-pentadecylbenzene-1,3-diol (Figure S8A). Furthermore, the standard contained impurities attributed to autoxidation products **4** and **5**, which showed mass spectra and retention times identical to those of the assay products (Figure S8B and C). More details on structural elucidation are provided in the next section. Thus, the combined data support the CepA-catalyzed formation of resorcinols in addition to pyrones. The product range observed under optimized reaction conditions is similar to that described for BpsA.[17] Furthermore, time-course experiments showed a rapid increase in pyrone and resorcinol products not observed in the boiled enzyme and no enzyme controls (Figure S17). Quinol and quinone formation (**4** and **5**) were delayed, increasing at reaction times of around 90 to 150 min, respectively (Figures 5C and S18B), supporting a nonenzymatic autoxidation route (Figures S6 and S8).

Structure elucidation of in vitro products

For structure elucidation of major product **1** by means of NMR spectroscopy, it was purified by means of HPLC. It had a molecular formula of $C_{20}H_{34}O_3$, which was suggested by HPLC–ESI-MS (m/z 321.2435, $[M-H]^-$, $\Delta 0$ ppm; see Figure S5). The ^{1}H NMR spectrum and HSQC data suggested the presence of one ethyl group, two sp^2-methines, and a long methylene chain unit (Figures S9 and S10). Two units, **a** and **b**, were determined from COSY data (Figures S11 and S14). ^{13}C-labeling studies showed, by analysis of the ^{13}C NMR spectrum, the incorporation of four ^{13}C atoms in compound **1** (Figures S12 and S14). The multiplicity pattern of these ^{13}C NMR signals suggested the connection of C1−C4, which was also supported by HMBC correlations from H-2 to C-1 and C-4 and from H-4 to C-2 (Figures S13 and S14). Unit **a** was connected to C-4 through C-5 by HMBC correlations from H-6 to C-4 and C-5, and from H-7 to C-5. Finally, the length of the methylene chain and the pyrone cyclization were deduced from the molecular formula.

The structure of the less abundant tetraketide pyrone **2** was elucidated by means of MS/MS. The neutral loss of 238.2342 Da, corresponding to the linear part of the molecule ($C_{16}H_{30}O$, calculated mass 238.2297), gave the most abundant ion at m/z 125.0228, corresponding to the cyclic part ($C_6H_5O_3^-$, calculated mass 125.0244) of **2** (Figure S15). The structure of the cyclic section of **2** is further substantiated by additional fragments shown in Figure S15. Compounds **3–5** (m/z 319.26, 335.26, and 333.24, respectively) were compared with the retention times of a standard (5-pentadecylbenzene-1,3-diol, CAS 3158-56-3, Brunschwig, Switzerland) and its autoxidation products (Figure S6). The proposed structures of **4** and **5** appear more likely due to the observation of additional peaks in the standard spectra that are identical in retention times to those observed in the assay extracts.

Conclusion

The *cep* cluster is one of only three BGCs conserved among five different "Entotheonella" variants that inhabit sponges from distant geographical regions. By investigating the orthologous type III PKS clusters, we hoped to gain the first insights into the as-yet unknown chemical functions and conserved metabolic features of these hidden natural product factories. Our in vitro experiments demonstrate that CepA$_{factor}$ acts as a phenolic lipid synthase, processing long-chain fatty acid acyl-CoA and malonyl-CoA thioesters. The product range includes tetraketide resorcinols, as suggested by the presence of a tryptophan residue in the active site at a position critical for aldol condensation, as well as tri- and tetraketide α-pyrones. Resorcinol synthases often show a shifted product profile towards pyrones in vitro relative to in vivo activity.[1c, 18] The ratio of resorcinols to pyrone derailment products depends on the spatial constraints in the active-site cavity of the enzyme. In the active site of ArsB, a resorcinol synthase from *Azotobacter vinelandii*, Ser158, Leu219, and Thr235 form a steric wall responsible for strict resorcinol formation.[1c] It is possible that the divergence in "Entotheonella" PKSs from the ArsB residues to cysteine and

isoleucine/valine residues perturbs this organization, causing the more relaxed product range for CepA. In addition, the MT CepB might be needed to increase product fidelity of CepA. Unfortunately, extensive trials to establish functional CepB in vivo or in vitro were unsuccessful.

The in vivo biological function of CepA in "Entotheonella" species remains unknown, but a range of functions have been ascribed to phenolic lipids from plants, fungi, and cultivated organisms, including, but not limited to, anticancer, anti-inflammatory, genotoxic, and antibiotic activities.[24,25] Many other known functions of phenolic lipids are as primary, rather than secondary, metabolites; this hypothesis would be in alignment with the observed conservation of the *cep* gene cluster. In mycobacteria, alkylquinones act as alternative electron carriers in microaerophilic cell respiration.[4] The products of ArsB and another type III PKS, ArsC, are the main components of *Azotobacter* cyst membranes, and the activity of these enzymes is essential to the encystment process.[27] A related function of phenolic lipids is present in *Streptomyces griseus*, in which products of type III PKSs provide antibiotic resistance by changing the composition of the membrane.[5] Because all sequenced "Entotheonella" species are symbionts of marine sponges, the conserved CepA pathway may also serve to help facilitate symbiosis, either directly or indirectly, for example, as a signaling molecule, which is the case for Rhizobium–plant symbiosis mediated by flavonoids.[11]

Phenolic lipids are not unknown in the extracts of marine sponges; alkyl and alkenylresorcinols resembling those produced herein have been detected in *Haliclona* sponges.[28] However, they have not been detected in theonellid sponges, nor were we able to detect *cep*-type compounds in *Theonella* extracts, either directly or by molecular network analysis.[29] This might be due to low abundance in the overall sponge holobiont or structural differences, for example, regarding the as-yet unknown starter, that might also influence the cyclization pattern. For instance, Hug and co-workers recently demonstrated, in a myxobacterial system, that type III PKS compounds produced in vitro might have significant differences from the authentic product for precisely this reason.

"Entotheonella" have attracted interest as a chemically talented taxon rich in structurally and biosynthetically unusual metabolites.[30] The *cep* cluster is a rare deviation from the otherwise nonoverlapping natural product potential of known "Entotheonella" species. The functional information provided herein will help to elucidate the ecological role of this polyketide in this intriguing, yet poorly understood, group of symbiotic bacteria.

Experimental Section

Bioinformatic analysis: Blast searches to find *cep*-like clusters in other "Entotheonella" genomes were performed by using the tblastn search tool implemented in the molecular biology analysis suite Geneious version 7.1.8 (http://www.geneious.com) with the "*Ca.* E. factor" genes as a query sequence.[31] For comparison, multisequence alignments were created. Two consecutive alignment rounds were performed by using the Geneious alignment algorithm, followed by ClustalW implemented in the Geneious software. For alignments containing more than 100 sequences, a third round of ClustalW alignment was added. To determine sites correlated with the cyclization specificity, all five CepA sequences from "Entotheonella" were aligned with ArsB and ArsC from *A. vinelandii* and Pks18 from *M. tuberculosis*. CepA sequences were aligned with Pks18 and BspA (*B. subtilis* 168) for the analysis of residues possibly involved in substrate binding. To analyze ACP-binding capacity, sequences were aligned with CHS2 from *Medicago sativa*, as well as with 132 experimentally characterized type III PKSs or a representative set of 696 nonredundant type III PKSs from the KEGG database, as used by Shimizu et al.[1a,6c,19]

Overall cluster detection for comparison of BGCs was performed by using antiSMASH v4 and confirmed by manual inspection.[32]

Cloning, expression, and purification: The *cep*A gene from "*E. factor*" (*cep*A$_{factor}$) was PCR-amplified from filament-enriched metagenomic DNA by using the primers listed in Table S5.[7b,c] The gene was introduced by restriction digest and subsequent ligation into pHis8,[16] which yielded the plasmid pHis_*cep*A$_{factor}$ for the production of N-terminally His$_8$-tagged protein. After introduction into *E. coli* BL21 (DE3), cells were grown at 37 °C in terrific broth (TB) medium until an OD_{600} of 1.2 was reached. Cells were incubated on ice up to 2 h before protein expression was induced with isopropyl-β-ᴅ-thiogalactoside (IPTG; 0.1 or 0.5 mm). Cells were grown for 24 h or 3 days at 16 °C. Cells were harvested by centrifugation, snap frozen in liquid nitrogen for better lysis, and either used directly or stored at −80 °C. For lysis, cells were thawed on ice, resuspended in cell lysis buffer (50 mm sodium phosphate buffer with 300 mm, NaCl, 20 mm imidazole, and 10% glycerol), and lysed by sonication. Cell debris and insoluble components were sedimented by centrifugation. The supernatant was incubated with Ni^{2+}-NTA resin (Macherey–Nagel, Düren, Germany). The resin was washed sequentially with three different buffers containing 20, 30, and 40 mm imidazole before eluting the target protein with 250 mm imidazole (see Figure S16) The protein was dialyzed into 100 mm phosphate buffer, pH 8, overnight at 4 °C by using a 6–8 kDa cutoff membrane (Socochim SA, Lausanne, Switzerland). The resulting elution fractions were analyzed by means of SDS-PAGE and the concentration was determined with the Nanoquant protein quantitation reagent (Roth, Karlsruhe, Germany) by using a Thermo Nano-drop spectrophotometer and bovine albumin as the standard. The major band observed by SDS-PAGE was in good accordance with the predicted size of the recombinant protein of about 42 kDa.

Enzymatic assay for CepA$_{factor}$: The reactions to determine starter and elongation unit specificity contained 100 mm phosphate buffer, pH 8, and either malonyl-CoA (300 μm) or malonyl-CoA (300 μm) and methylmalonyl-CoA (300 μm) as extender units, as well as either a pool of starter units (100 μm each) or a single starter substrate (100 μm; Figure 4). Reactions were performed in a total volume of 200 μL and were incubated for 3 h at 37 °C. Starter units were kindly provided by the group of Erb.[33] Reactions for time-course experiments and isolation of polyketides contained phosphate buffer (500 mm, pH 6) with malonyl-CoA (300 μm) and palmitoyl-CoA (100 μm). The time-course experiments and larger scale reactions for preparative HPLC were performed in 3 mL. For the time-course experiments, a sample of 100 μL was taken every 30 min for a total of 6 h. A concentration of 0.425 mg μL^{-1} Ni^{2+}-purified enzyme was used in each reaction. Reactions were stopped by freezing at −20 °C. Negative controls contained either phosphate buffer (100 mm, pH 8) or boiled enzyme (10 min 99 °C), instead of enzyme. To increase the catalytic activity of the enzyme,

© 2019 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

buffer concentrations between 100 and 800 mᴍ were tested, as well as pH values ranging from 5.8 to 8.0. Optimal conditions were determined to be pH 6.0 and 500 mᴍ phosphate buffer. Malonyl-CoA, methylmalonyl-CoA, and palmitoyl-CoA were obtained from Sigma–Aldrich (USA). Detection of free thiols was performed as described by Ellman.[34] Aliquots (10 µL) of starter unit assay solution were mixed with 2.4 mᴍ DTNB (5 µL) in 100 mᴍ phosphate buffer. Adsorption was measured at $\lambda = 412$ nm by using a Thermo Nanodrop instrument.

Isolation and identification of compounds: Reactions were extracted with ethyl acetate (3× with either 200 µL or 3 mL, depending on the reaction volume). The organic layers were pooled and evaporated, and the residue was dissolved in methanol (60 µL) for reversed-phase HPLC–ESI-MS analysis. Samples were run on a Kinetex 2.6 µm XB-C18 100 Å column (4.6×150 mm; Phenomenex) with 5 min at 5% methanol+0.1% formic acid (FA) in water+0.1% FA followed by a gradient (5 to 100%) methanol+0.1% FA in water+0.1% FA over 10 min followed by 10 min at 100% methanol (0.1% FA) and 5 min of 5% methanol+0.1% FA in water+0.1% FA with a flow rate of 0.5 mL min⁻¹. HPLC–ESI-MS spectra were obtained from a Q Exactive Hybrid Quadrupole-Orbitrap mass spectrometer set to negative-ion mode and coupled to a Dionex UltiMate 3000 UHPLC system (Thermo Scientific). Preliminary HPLC–MS data were analyzed by using a Python script, created by Lackner, based on the eMZed[35] framework, by comparing the MS data of the active enzyme sample with controls to determine if any of the products were formed. A list of calculated masses for products likely to be formed from the added substrates was used as a query. Candidates for the list were narrowed down by preliminary substrate screening assays (see above). MS² spectra were obtained under the same conditions, with normalized collision energies (NCE) ranging from 30 to 85%. The triketide pyrone was purified by means of semipreparative HPLC on an Agilent 1260 infinity instrument coupled to an Agilent G1315D DAD by using a Kinetex C₁₈ column (10×250 mm; Phenomenex) with a gradient (5 to 100%) of methanol (0.1% FA) in water (0.1% FA) over 30 min followed by 40 min in 100% methanol+0.1% FA (1 mL min⁻¹). The compound was detected at $\lambda = 288$ nm and fractions checked for the presence of the compound by means of HPLC–ESI-MS. The solvents were evaporated and the residue was dissolved in methanol. The compound was further purified isocratically with 85% methanol+0.1% FA in water+0.1% FA (1 mL min⁻¹) by using a Kinetex C₁₈ column (4.6×250 mm; Phenomenex). Solvent was evaporated and the residue was dissolved in [D₆]DMSO.

NMR spectroscopy analysis: NMR spectra were recorded on a Bruker Avance III spectrometer equipped with a cold probe at 600 MHz for ¹H and 150 MHz for ¹³C at 298 K. Chemical shifts were referenced to the solvent signals at $\delta_H = 2.50$ ppm and $\delta_C = 39.51$ ppm for [D₆]DMSO. LC–ESI-MS was performed on a Thermo Scientific Q Exactive mass spectrometer (see Figures S5 and S7–S14).

Acknowledgements

Conflict of Interest

The authors declare no conflict of interest.

[1] a) M. B. Austin, J. P. Noel, *Nat. Prod. Rep.* **2003**, *20*, 79–110; b) S. Grüschow, T. J. Buchholz, W. Seufert, J. S. Dordick, D. H. Sherman, *ChemBioChem* **2007**, *8*, 863–868; c) R. Satou, A. Miyanaga, H. Ozawa, N. Funa, Y. Katsuyama, K. Miyazono, M. Tanokura, Y. Ohnishi, S. Horinouchi, *J. Biol. Chem.* **2013**, *288*, 34146–34157.

[2] a) B. S. Moore, J. N. Hopke, *ChemBioChem* **2001**, *2*, 35–38; b) Y. Li, R. Müller, *Phytochemistry* **2009**, *70*, 1850–1857; c) N. Funa, Y. Ohnishi, I. Fujii, M. Shibuya, Y. Ebizuka, S. Horinouchi, *Nature* **1999**, *400*, 897–899.

[3] a) H. Chen, C. C. Tseng, B. K. Hubbard, C. T. Walsh, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 14901–14906; b) V. Pfeifer, G. J. Nicholson, J. Ries, J. Recktenwald, A. B. Schefer, R. M. Shawky, J. Schroder, W. Wohlleben, S. Pelzer, *J. Biol. Chem.* **2001**, *276*, 38370–38377.

[4] A. Anand, P. Verma, A. K. Singh, S. Kaushik, R. Pandey, C. Shi, H. Kaur, M. Chawla, C. K. Elechalawar, D. Kumar, Y. Yang, N. S. Bhavesh, R. Banerjee, D. Dash, A. Singh, V. T. Natarajan, A. K. Ojha, C. C. Aldrich, R. S. Gokhale, *Mol. Cell* **2015**, *60*, 637–650.

[5] M. Funabashi, N. Funa, S. Horinouchi, *J. Biol. Chem.* **2008**, *283*, 13983–13991.

[6] a) B. S. Moore, C. Hertweck, J. N. Hopke, M. Izumikawa, J. A. Kalaitzis, G. Nilsen, T. O'Hare, J. Piel, P. R. Shipley, L. Xiang, M. B. Austin, J. P. Noel, *J. Nat. Prod.* **2002**, *65*, 1956–1962; b) Y. Katsuyama, Y. Ohnishi, *Methods Enzymol.* **2012**, *515*, 359–377; c) Y. Shimizu, H. Ogata, S. Goto, *ChemBioChem* **2017**, *18*, 50–65.

[7] a) M. W. Taylor, R. T. Hill, J. Piel, R. W. Thacker, U. Hentschel, *ISME J.* **2007**, *1*, 187; b) M. C. Wilson, T. Mori, C. Rückert, A. R. Uria, M. J. Helf, K. Takada, C. Gernert, U. A. Steffens, N. Heycke, S. Schmitt, C. Rinke, E. J. Helfrich, A. O. Brachmann, C. Gurgui, T. Wakimoto, M. Kracht, M. Crüsemann, U. Hentschel, I. Abe, S. Matsunaga, J. Kalinowski, H. Takeyama, J. Piel, *Nature* **2014**, *506*, 58–62; c) R. Ueoka, A. R. Uria, S. Reiter, T. Mori, P. Karbaum, E. E. Peters, E. J. Helfrich, B. I. Morinaka, M. Gugger, H. Takeyama, S. Matsunaga, J. Piel, *Nat. Chem. Biol.* **2015**, *11*, 705–712; d) T. Mori, J. K. B. Cahn, M. C. Wilson, R. A. Meoded, V. Wiebach, A. F. C. Martinez, E. J. N. Helfrich, A. Albersmeier, D. Wibberg, S. Datwyler, R. Keren, A. Lavy, C. Rückert, M. Ilan, J. Kalinowski, S. Matsunaga, H. Takeyama, J. Piel, *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 1718–1723.

[8] a) C. A. Bewley, D. J. Faulkner, *Angew. Chem. Int. Ed.* **1998**, *37*, 2162–2178; *Angew. Chem.* **1998**, *110*, 2280–2297; b) J. W. Blunt, B. R. Copp, W. P. Hu, M. H. Munro, P. T. Northcote, M. R. Prinsep, *Nat. Prod. Rep.* **2009**, *26*, 170–244; c) C. Gurgui, J. Piel, *Methods Mol. Biol.* **2010**, *668*, 247–264; d) G. Lackner, E. E. Peters, E. J. Helfrich, J. Piel, *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E347–E356.

[9] a) R. J. Capon, *Eur. J. Org. Chem.* **2001**, 633–645; b) V. J. Paul, M. P. Puglisi, R. Ritson-Williams, *Nat. Prod. Rep.* **2006**, *23*, 153–180; c) J. R. Pawlik, *BioScience* **2011**, *61*, 888–898.

[10] M. P. Puglisi, J. M. Sneed, R. Ritson-Williams, R. Young, *Nat. Prod. Rep.* **2019**, *36*, 410–429.

[11] X. Perret, C. Staehelin, W. J. Broughton, *Microbiol. Mol. Biol. Rev.* **2000**, *64*, 180–201.

[12] J. L. Martin, F. M. McMillan, *Curr. Opin. Struct. Biol.* **2002**, *12*, 783–793.

[13] A. Koike-Takeshita, T. Koyama, K. Ogura, *J. Biol. Chem.* **1997**, *272*, 12380–12383.

[14] A. K. Harris, N. R. Williamson, H. Slater, A. Cox, S. Abbasi, I. Foulds, H. T. Simonsen, F. J. Leeper, G. P. Salmond, *Microbiology* **2004**, *150*, 3547–3560.

[15] a) T. Garnier, K. Eiglmeier, J. C. Camus, N. Medina, H. Mansoor, M. Pryor, S. Duthoy, S. Grondin, C. Lacroix, C. Monsempe, S. Simon, B. Harris, R.

Atkin, J. Doggett, R. Mayes, L. Keating, P. R. Wheeler, J. Parkhill, B. G. Barrell, S. T. Cole, S. V. Gordon, R. G. Hewinson, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 7877 – 7882; b) R. Rukmini, V. M. Shanmugam, P. Saxena, R. S. Gokhale, R. Sankaranarayanan, *Acta Crystallogr. Sect. D* **2004**, *60*, 749 – 751.

[16] R. Sankaranarayanan, P. Saxena, U. B. Marathe, R. S. Gokhale, V. M. Shanmugam, R. Rukmini, *Nat. Struct. Mol. Biol.* **2004**, *11*, 894 – 900.

[17] C. Nakano, H. Ozawa, G. Akanuma, N. Funa, S. Horinouchi, *J. Bacteriol.* **2009**, *191*, 4916 – 4923.

[18] P. Saxena, G. Yadav, D. Mohanty, R. S. Gokhale, *J. Biol. Chem.* **2003**, *278*, 44780 – 44790.

[19] Y. M. Zhang, M. S. Rao, R. J. Heath, A. C. Price, A. J. Olson, C. O. Rock, S. W. White, *J. Biol. Chem.* **2001**, *276*, 8231 – 8238.

[20] A. Miyanaga, N. Funa, T. Awakawa, S. Horinouchi, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 871 – 876.

[21] J. M. Jez, J. L. Ferrer, M. E. Bowman, R. A. Dixon, J. P. Noel, *Biochemistry* **2000**, *39*, 890 – 902.

[22] G. L. Ellman, *Arch. Biochem. Biophys.* **1958**, *74*, 443 – 450.

[23] C. C. Zhou, D. R. Hill, *Magn. Reson. Chem.* **2007**, *45*, 128 – 132.

[24] a) M. Stasiuk, A. Kozubek, *Cell. Mol. Life Sci.* **2010**, *67*, 841 – 860; b) T. P. Martins, C. Rouger, N. R. Glasser, S. Freitas, N. B. de Fraissinette, E. P. Balskus, D. Tasdemir, P. N. Leão, *Nat. Prod. Rep.* **2019**, https://dx.doi.org/10.1039/C8NP00080H.

[25] J. J. Hug, F. Panter, D. Krug, R. Müller, *J. Ind. Microbiol. Biotechnol.* **2019**, *46*, 319 – 334.

[26] P. N. Leão, M. Costa, V. Ramos, A. R. Pereira, V. C. Fernandes, V. F. Domingues, W. H. Gerwick, V. M. Vasconcelos, R. Martins, *PLoS One* **2013**, *8*, e69562.

[27] N. Funa, H. Ozawa, A. Hirata, S. Horinouchi, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 6356 – 6361.

[28] R. Barrow, R. Capon, *Aust. J. Chem.* **1991**, *44*, 1393 – 1405.

[29] M. Wang, J. J. Carver, V. V. Phelan, L. M. Sanchez, N. Garg, Y. Peng, D. D. Nguyen, J. Watrous, C. A. Kapono, T. Luzzatto-Knaan, C. Porto, A. Bouslimani, A. V. Melnik, M. J. Meehan, W.-T. Liu, M. Crusemann, P. D. Boudreau, E. Esquenazi, M. Sandoval-Calderon, R. D. Kersten, et al., *Nat. Biotechnol.* **2016**, *34*, 828 – 837.

[30] M. Jaspars, G. Challis, *Nature* **2014**, *506*, 38 – 39.

[31] M. Kearse, R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, A. Drummond, *Bioinformatics* **2012**, *28*, 1647 – 1649.

[32] M. H. Medema, K. Blin, P. Cimermancic, V. de Jager, P. Zakrzewski, M. A. Fischbach, T. Weber, E. Takano, R. Breitling, *Nucleic Acids Res.* **2011**, *39*, W339 – W346.

[33] D. M. Peter, B. Vogeli, N. S. Cortina, T. J. Erb, *Molecules* **2016**, *21*, 517.

[34] G. L. Ellman, *Arch. Biochem. Biophys.* **1959**, *82*, 70 – 77.

[35] P. Kiefer, U. Schmitt, J. A. Vorholt, *Bioinformatics* **2013**, *29*, 963 – 964.