How to cite: *Angew. Chem. Int. Ed.* **2023**, *62*, e202308540
doi.org/10.1002/anie.202308540

Cyclic Peptides | Hot Paper

Analysis of Rhizonin Biosynthesis Reveals Origin of Pharmacophoric Furylalanine Moieties in Diverse Cyclopeptides

*Friedrich J. Ehinger, Sarah P. Niehs, Benjamin Dose, Maria Dell, Jana Krabbe, Sacha J. Pidot, Timothy P. Stinear, Kirstin Scherlach, Claudia Ross, Gerald Lackner, and Christian Hertweck**

In memoriam Heinz-Günter Floss (1934-2022)

Abstract: Rhizonin A and B are hepatotoxic cyclopeptides produced by bacterial endosymbionts (*Mycetohabitans endofungorum*) of the fungus *Rhizopus microsporus*. Their toxicity critically depends on the presence of 3-furylalanine (Fua) residues, which also occur in pharmaceutically relevant cyclopeptides of the endolide and bingchamide families. The biosynthesis and incorporation of Fua by non-ribosomal peptide synthetases (NRPS), however, has remained elusive. By genome sequencing and gene inactivation we elucidated the gene cluster responsible for rhizonin biosynthesis. A suite of isotope labeling experiments identified tyrosine and L-DOPA as Fua precursors and provided the first mechanistic insight. Bioinformatics, mutational analysis and heterologous reconstitution identified dioxygenase RhzB as necessary and sufficient for Fua formation. RhzB is a novel type of heme-dependent aromatic oxygenases (HDAO) that enabled the discovery of the bingchamide biosynthesis gene cluster through genome mining.

Introduction

Peptide natural products constitute an extremely diverse group of ligands[1] that bind to an abundance of targets.[2] The versatile biological functions are not only attributed to the different peptide architectures, chain size and topology, but also depend on the range of amino acid residues incorporated.[3] Whereas ribosomal peptides are typically composed of a limited set of proteinogenic building blocks, a large number of non-canonical amino acids have been identified in the products of non-ribosomal peptide synthetases (NRPS).[4] Some of these moieties are pivotal in target binding. A remarkable example of such pharmacophoric amino acids is 3-furylalanine (Fua, Figure 1). Having been designed as a synthetic phenylalanine antimetabolite in the 1960s,[5] Fua is now also known to naturally occur in a number of specialized metabolites of diverse microbes.

The first peptidic natural products containing Fua were discovered while investigating the potential causes of

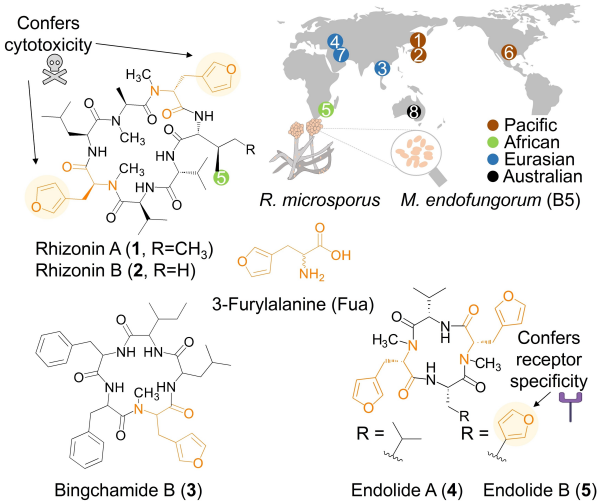

Figure 1. The rare amino acid 3-furylalanine (Fua) is a pharmacophore in natural products. The map indicates the distribution of the *Rhizopus-Mycetohabitans* symbiosis and rhizonin producer strain B5. Besides the rhizonins, two other groups of NRPs from *Streptomyces bingchenggensis* (bingchamide B, **3**) and *Burkholderia contaminans* (endolide A, **4**) contain Fua.

[*] F. J. Ehinger, Dr. S. P. Niehs, Dr. B. Dose, Dr. M. Dell, Dr. J. Krabbe, Dr. K. Scherlach, Dr. C. Ross, Dr. G. Lackner, Prof. Dr. C. Hertweck
Department of Biomolecular Chemistry, Leibniz Institute for Natural Product Research and Infection Biology (HKI)
Beutenbergstraße 11a, 07745 Jena (Germany)
E-mail: christian.hertweck@leibniz-hki.de

Dr. S. J. Pidot, Prof. T. P. Stinear
Department of Microbiology and Immunology, Doherty Institute
792 Elizabeth Street, Melbourne, 3000 (Australia)

Prof. Dr. C. Hertweck
Institute of Microbiology, Faculty of Biological Sciences, Friedrich Schiller University Jena
07743 Jena (Germany)

foodborne liver cancer in certain tribes in Africa.[6] The bioassay-guided screening of mycotoxins revealed a highly toxic mold growing on ground nuts in Mozambique.[7] Two cyclic heptapeptides, rhizonin A and B (**1** and **2**, Figure 1), were isolated from cultures of the contaminating fungus (*Rhizopus microsporus*) and proved to be severely hepatotoxic.[7] Rhizonins A and B differ only in one amino acid position (D-Ile/D-Val), yet each cyclopeptide contains two *N*-methylated Fua residues in both the D- and the L-forms.[8] A synthetic study uncovered a prominent role of Fua in the bioactivity of these peptides, as replacement of both Fua moieties of **1** with phenylalanine led to a dramatically reduced cytotoxicity.[9] Although rhizonins were named the first mycotoxins discovered from a "lower fungus" of the order Mucorales,[8] interestingly, it is not the fungus that has this unusual biosynthetic ability. The true producers are bacteria (*Mycetohabitans endofungorum*) residing in the fungal cells.[10]

The second Fua-derived cyclopeptide was discovered in the broth of *Streptomyces bingchenggensis*. Bingchamide B (3) (Figure 1) contains a single Fua building block and exhibits moderate cytotoxicity towards human colon carcinoma cell line HCT-116.[11] Endolides A and B (**4**, **5**, Figure 1) constitute the third family of Fua-containing natural products. The endolides were first isolated from the sponge-derived ascomycete *Stachylidium bicolor* but may be of symbiotic origin like the rhizonins.[12] Indeed, the fungus harbors yet unculturable bacteria of the order *Burkholderiales* that are distantly related to the rhizonin-producing *Rhizopus* endosymbionts.[13] Cyclopeptide **5** contains three L-Fua moieties (Figure 1), of which one is critical for target selectivity. Whereas **5** has affinity towards the serotonin receptor $5HT_{2b}$, congener **4**, in which the non-methylated Fua is replaced by L-Leu, binds to the vasopressin receptor 1 A.[13] These structure–activity relationships and the strong dependency of rhizonins toxicity indicate that Fua may play an important role as a pharmacophore (Figure 1).

Beyond these remarkable structure–activity relationships, the presence of furan moieties in peptides is also of high interest for synthetic and medicinal chemists, as it might serve as a handle to allow selective modifications and crosslinks. This aim has prompted efforts to introduce the Fua regioisomer 2-Fua into peptides by in vitro translation[14] and solid phase peptide synthesis.[15] Similarly, it would be highly desirable to program the biosynthetic incorporation of Fua into peptides and to exploit the reactivity of the furan ring in vivo. However, such synthetic biology approaches have not yet been possible due to the lack of information on Fua-dependent assembly lines. Even our current understanding of the biosynthetic origin of Fua is very limited. Based on labeling experiments with the general carbon sources glycerol and glucose, it was only proposed that a shikimate pathway-derived precursor could serve as the substrate for Fua biosynthesis.[16] To date, no biosynthetic gene cluster (BGC) of a Fua-containing natural product has been identified, and the biosynthetic pathway, precursor(s), and biocatalyst(s) involved in Fua formation have remained elusive.

Herein we uncover the gene cluster for rhizonin biosynthesis and reveal the biochemical basis for Fua formation. We identify the biosynthetic precursors of this unusual amino acid and present a novel type of dioxygenase as a Fua synthase. By isotope labeling we also provide first insight into an unprecedented catechol cleavage reaction. Finally, in unearthing the elusive bingchamide BGC, we demonstrate that it is possible to discover Fua-incorporating peptide assembly lines by genome mining using the Fua synthase gene as a guide.

Results and Discussion

Identification of the Rhizonin Biosynthesis Gene Cluster

To gain insight into the biosynthetic origin of **1**, we sought to mine the genome sequence of *M. endofungorum* (strain B5). The initial *M. endofungorum* draft genome sequence (GenBank genome accession number: GCA_021991855.1), however, was fragmented and did not allow for a seamless bioinformatic analysis. Thus, we re-isolated *M. endofungorum* from the *Rhizopus microsporus* host and sequenced genomic DNA from the freshly prepared endosymbiont culture.[17] PacBio sequencing afforded sufficiently long reads to obtain a contiguous genome sequence of strain B5, a 2.6 Mb chromosome and a 0.8 Mb megaplasmid. In silico analysis by antiSMASH[18] revealed one polyketide synthetase (PKS)/NRPS gene locus (*rhi*, for rhizoxin biosynthesis),[19a] nine NRPS gene loci (including the known *hab* locus)[19b] and six additional BGCs (Figure 2A), including *bur* (burhizin) for lasso peptide formation.[19c] Compared to the previous draft genome sequence, our new analysis points to a substantially higher potential for natural product biosynthesis of *M. endofungorum* (Table S1).

To select candidates for the rhizonin (*rhz*) BGC among the nine NRPS gene loci (Figure 2B), we compared the deduced NRPS domain architectures to the chemical structure of **1**. Of these, only three NRPSs harbor methyltransferase (MT) domains for the biosynthesis of *N*-methylated amino acids as in **1**. Since the candidate NRPS BGC #6 codes for habitasporin biosynthesis,[19] it was excluded from further investigation. Bioinformatics analysis of the adenylation (A) domains[18,20] of the two octamodular NRPSs encoded by BGC #1 and #4 predicted the amino acid building blocks and their sequence: meX–Val–Thr–X–meGly–Val–X–X (NRPS BGC #1) and Val-meX–Leu–meX–meX–Ile–Val–Val (NRPS BGC #4, X for unknown). The number and positioning of MT domains, the unusual adenylation domains in modules 2 and 5, the epimerase domain in module 5, which would transform the second L-Fua unit into D-Fua, and the absence of homologous gene loci in rhizonin-negative *Mycetohabitans* spp. suggested that the NRPS encoded by BGC #4 is the best candidate for a rhizonin synthetase. The encoded assembly line, however, is not fully co-linear with the peptide backbone, as eight NRPS modules would yield an octapeptide, not a heptapeptide. The skipping of one module, which has been observed in other NRPS assembly lines,[21] could be a

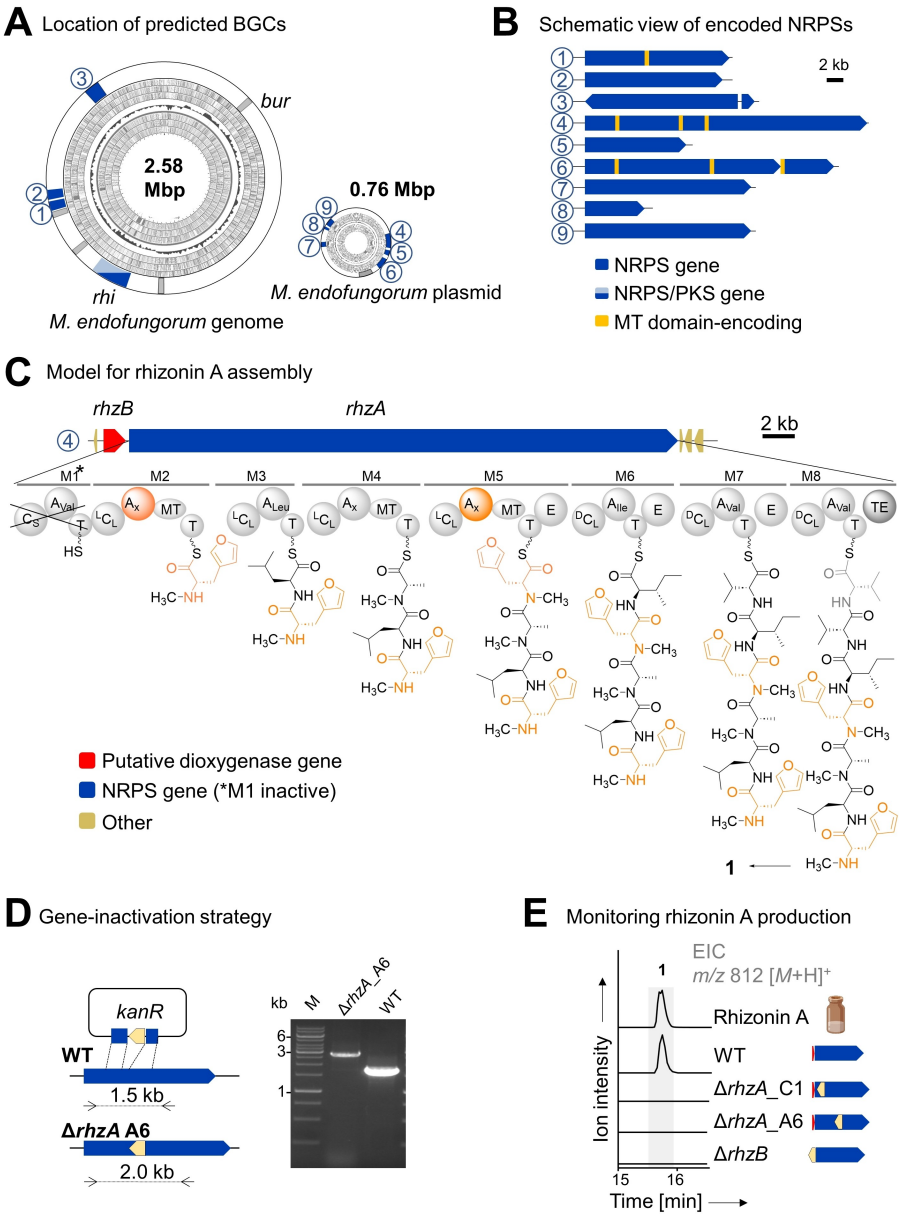

Figure 2. Biosynthetic origin of rhizonins. A) Biosynthetic potential predicted from the genome sequence of *M. endofungorum*. BGC: biosynthetic gene cluster. *rhi*, rhizoxin BGC, *bur*, burhizin BGC. B) Overview of encoded nonribosomal peptide synthetases (NRPSs), some containing methyltransferase (MT) domains. C) Biosynthetic model for rhizonin A assembly. *rhzB*, encoding a dioxygenase-like protein; *rhzA*, NRPS gene. A domain specificity indicated in lower case with X representing an unknown specificity. D) Gene-inactivation strategy used to create deletion mutants of the genes *rhzA* and *rhzB*. E) Rhizonin A production abrogated in mutants. Extracted ion chromatograms (EICs) of *m/z* 812 $[M+H]^+$ (**1**) in wild-type (WT) and mutant cultures.

plausible explanation for this non-colinearity (Figure 2C). Due to the incorporation of only two Val residues despite the three encoded Val-specific A domains (modules 1, 7, 8), one of these modules is a likely skipping point. Importantly, one Val is present as the D enantiomer in the final product, rendering module 7 and its epimerase domain crucial for biosynthesis of **1**. Thus, only module 1 and 8 remain dispensable. Active site-editing at the module 8 A domain, mutating the catalytic aspartate to alanine, fully abrogated production of all rhizonins (Figure S1). Taking this catalytic essentiality of module 8 into account, we expect that module 1 is skipped. Biosynthesis of **1** would begin at module 2, which installs an *N*-methylated Fua as the first residue. Primary backbone amine methylation by the first NRPS module also occurs in other cases.[22] Besides the NRPS (RhzA), the tentative rhizonin BGC (*rhz*, GenBank accession number: OQ975904) also encodes a dioxygenase (RhzB) and an MbtH-like protein (Figure 2C, Table S2,3).

To give additional evidence on the link between rhizonin biosynthesis and the *rhz* gene locus, we aimed to inactivate the NRPS gene *rhzA* and the predicted dioxygenase gene *rhzB* by construction of deletion mutants. Initial attempts

using a previously *pheS*-based double selection plasmid established for related *Mycetohabitans* strains (B1[17,23] and B8[24]), however, were not successful with this strain. The comparably weak growth of *M. endofungorum* prompted us to test alternative gene inactivation plasmids lacking the counterselection, thus reducing the metabolic burden on the bacterial strain. In this way we eventually succeeded in generating three verified deletion mutants (Figure 2D). Specifically, we disrupted the region within the *rhzA* gene that encodes the starter condensation (C) domain of module 1 (Δ*rhzA*_C1), the A domain of module 6 (Δ*rhzA*_A6), and the *rhzB* gene (Δ*rhzB*). Comparative metabolic profiling of the wild type and the mutants showed that **1** is only produced when RhzA and RhzB are intact (Figure 2E, Figure S2). Thus, we unequivocally identified *rhz* as the locus encoding the rhizonin assembly line.

Discovery of Rhizonin Congeners by Metabolic Profiling

When comparing the metabolic profiles of the wild type and the deletion mutants, we identified rhizonin congeners **6**–**8**, which eventually gave important clues on the biosynthetic origin of Fua. Since **6**–**8** were only produced in minor amounts (**6**: 0.15 mg L^{-1}, **7**: 0.26 mg L^{-1}, **8**: 0.33 mg L^{-1}) compared to **1** (3.2 mg L^{-1}), we inferred their structures from tandem mass spectrometry (MS/MS) fragmentation analyses and peptide hydrolysis followed by high-performance liquid chromatography high-resolution mass spectrometry (HPLC-HRMS)-based determination of the free amino acids. Mass analyses of the amino acid moieties and peptides, in particular MS/MS fragmentation of **1**, **6**, **7** and **8**, indicated that L-Fua was replaced with L-Phe in **6** (rhizonin C), D-Fua was replaced with D-Phe in **7** (rhizonin D), whereas both Fua residues were substituted with Phe in **8** (rhizonin E) (Figure S3–8). The substitution of Fua by Phe indicates substrate promiscuity of the respective A domains 2 and 5, reflecting the high similarity of these to the prototypic Phe-specific adenylation enzymes TycA (40.0 % and 40.5 %) and GrsA (37.0 % and 37.3 %).

To verify the amino acid substitutions, we first purified the congeners through size-exclusion chromatography on a Sephadex LH-20 column, followed by preparative HPLC. Then, we hydrolyzed the peptides and derivatized the free amino acids with Marfey's reagent. Using authentic amino acid references, we confirmed the deduced amino acid composition in **6**–**8** and noted that **8** harbors a Leu unit instead of an Ile residue (Figure 3A, Table S4). We confirmed the Fua substitutions by stable isotope labeling. We found that 1-^{13}C–Phe was incorporated once into **6** and **7**, and twice into **8** (Figure S9). These results underscore that RhzA possesses substrate flexibility towards pairs of Fua/Phe (modules 2 and 5) and Leu/Ile (module 6). While **6** and **7** are previously unknown congeners, compound **8** has been chemically synthesized before to reveal the crucial role of Fua for the toxicity of rhizonins.[9]

A Dioxygenase is Essential and Sufficient for Furylalanine Biosynthesis

Metabolic profiling showed that all five rhizonin congeners (**1**, **2**, **6**–**8**) are formed by the wild type, whereas only **8** is formed by the Δ*rhzB* mutant. Notably, both Fua moieties are replaced by Phe in congener **8** (Figure 3A). Thus, the Δ*rhzB* mutant is incapable of producing any of the Fua-harboring rhizonins, indicating that *rhzB* encodes an enzyme that plays an essential role in Fua biosynthesis. To support this assumption, we chemically complemented the Δ*rhzB* mutant by adding synthetic Fua to the culture broth. HPLC-HRMS analysis of the extract showed that the exogenously provided building block restored the formation of Fua-containing **1** in the Δ*rhzB* mutant (Figure S10). These results unequivocally demonstrate that RhzB is necessary to provide Fua for the respective rhizonins.

To test whether RhzB is sufficient for the production of the rare amino acid, we sought to reconstitute Fua production in a heterologous host. Therefore, we cloned *rhzB* into the pET-28a vector and produced RhzB in *E. coli* BL21(DE3). The culture extract was treated with Marfey's reagent to enable a stereospecific assignment of the formed amino acids. Whereas the negative control strain containing the empty expression vector did not show L-Fua production, HPLC-HRMS analyses indicated that the *rhzB*-expressing *E. coli* gained the ability to produce L-Fua. Retention time, exact mass (*m/z* 408.1150 $[M+H]^+$), and MS/MS fragmentation pattern are identical with the Marfey adduct **9** of an authentic L-Fua reference (Figure 3B). Interestingly, the presence of RhzB in *E. coli* is sufficient to generate L-Fua, indicating that the precursor is a common metabolite.

Biosynthetic Origin of Furylalanine

To identify the substrate of RhzB, we performed stable isotope labeling experiments with plausible amino acid precursors. Specifically, Phe, Tyr, and Leu were added to *rhzB*-expressing *E. coli*, and the extract was derivatized with Marfey's reagent. Supplementation with 3,3-d_2-D,L–Phe or 1-^{13}C–L-Leu yielded adduct **9** exclusively in its unlabeled form (*m/z* 408.1150 $[M+H]^+$) (Figure 3C). In contrast, when supplementing the expression strain with 1-$^{13}C_1$–L-Tyr, we noted that predominantly $^{13}C_1$-**9** is formed (*m/z* 409.1183 $[M+H]^+$, Figure 3C). Likewise, addition of 1-^{15}N–L-Tyr to the recombinant RhzB producer led to the specific increase of the $^{15}N_1$-**9** peak (Figure S11). MS/MS analyses revealed that the fragment pattern fingerprint of $^{13}C_1$-**9**-and $^{15}N_1$-**9** agrees with the spectrum of the unlabeled reference (Figure S11). Specific MS/MS fragments showed a *m/z* 1 Da increase compared to the corresponding fragment in the reference measurement (Figure S11), reflecting the presence or absence of the labeled atom in the respective ions, which were inferred by Mass Frontier-assisted fragment prediction (Table S5). To test the biosynthetic origin of Fua in the wild-type producer we added 1-^{15}N–Tyr to cultures of *M. endofungorum*. HPLC-HRMS analysis of the culture extract derivatized with Marfey's reagent showed an enrichment of

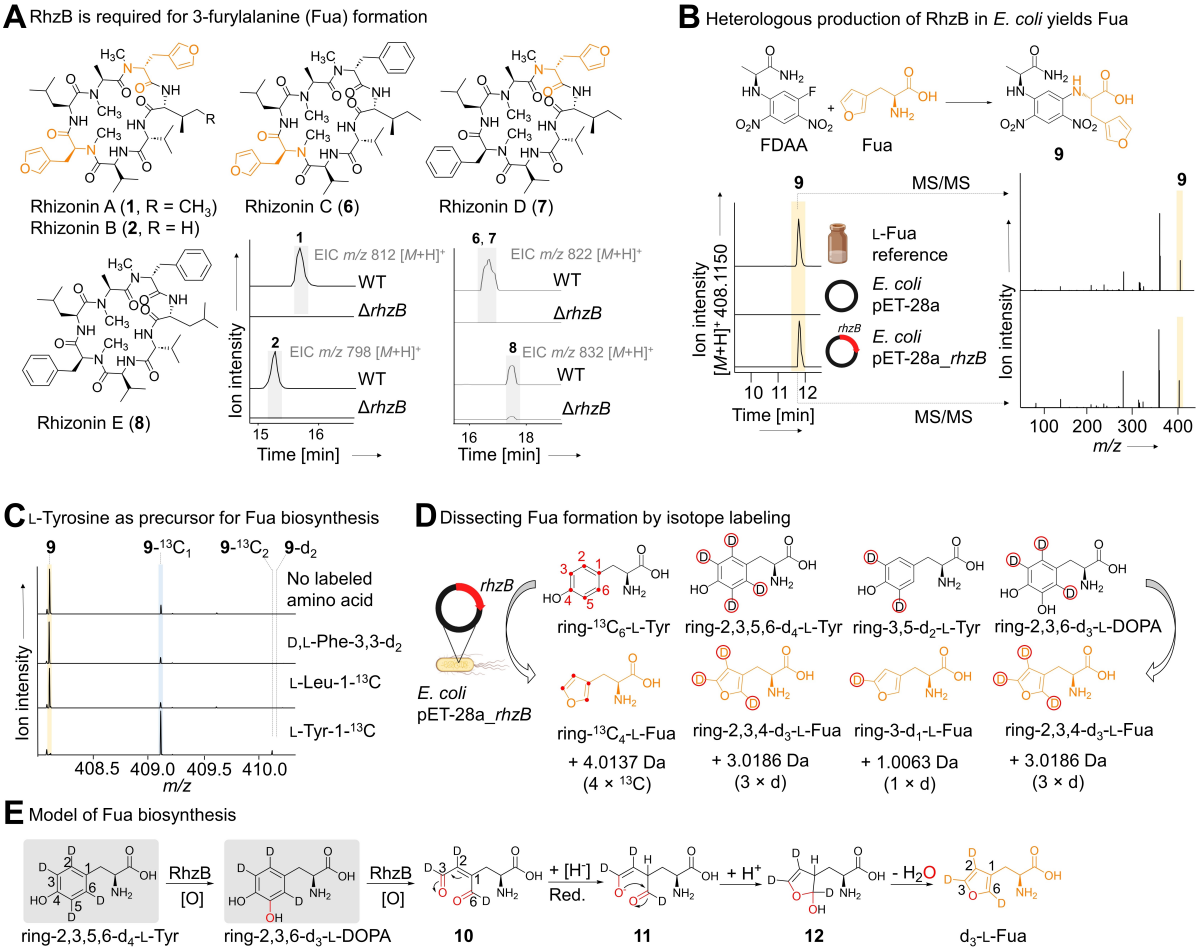

Figure 3. Biosynthetic origin of L-3-furylalanine (Fua). A) RhzB is required for Fua formation in rhizonin A biosynthesis. WT: wild type, EIC: extracted ion chromatogram. B) Heterologous production of RhzB in *E. coli* yields Fua. Derivatization of Fua with FDAA (1-fluoro-2–4-dinitrophenyl-5-L-alanine amide) to **9** facilitate HPLC-based separation of L- and D-amino acid adducts. HPLC–HRMS traces are representative of triplicate experiments. C) Stable isotope labeling reveals L-Tyr as precursor for Fua biosynthesis. Dotted lines in the HRMS spectrum indicate the mass of stable isotope-labeled **9** isotopologues. D) Fate of ring-labeled L-Tyr and L-DOPA in Fua production. E) Model of Fua biosynthesis. The deuteration indicates the expected labeling pattern when applying ring-2,3,5,6-d_4-L-Tyr and ring-2,3,6-d_3-L-DOPA, agreeing with experimental findings. The furan carbon atoms of Fua were assigned with the number of the respective phenol ring carbon they originate from. Red.: reduction.

the $^{15}N_1$-**9** species (Figure S12), indicating that Fua also arises from Tyr in the native pathway.

Model of RhzB-Catalyzed Furylalanine Formation

To identify the number of furan carbons originating from the phenol ring of Tyr, we added ring-ubi-$^{13}C_6$–Tyr to *E. coli* expressing *rhzB* (Figure 3C). HPLC-HRMS analyses of the extract showed that the +4 Da species (*m/z* 412.1287 $[M+H]^+$) is the most abundant isotopologue of the Marfey adduct **9** (Figure S13). Accordingly, four phenol ring carbons are incorporated into the furan ring, whereas two carbons are removed by RhzB en route to Fua. Such enzymatic C–C bond cleavage reactions in phenols often involve catechol intermediates that may undergo intradiol or extradiol ring openings (Figure S14),[25] followed by excision of oxidized carbons.[26] Ring hydroxylation of Tyr would generate L- 3,4-dihydroxyphenylalanin (L-DOPA), a common catechol substrate for aromatic ring-opening dioxygenases.[27] To test if L-DOPA is accepted as an on-pathway intermediate in Fua biosynthesis, we added ring-2,3,6-d_3–L-DOPA to *E. coli* producing RhzB. HPLC-HRMS analyses showed that the three deuterium labels are incorporated into Fua (Figure S13). Since *E. coli* cannot regenerate Tyr from L-DOPA,[28] the catechol must have served as a substrate for RhzB-mediated formation of Fua. Furthermore, addition of ring-3,5-d_2–Tyr to the recombinant *E. coli* strain resulted in a +1 Da (*m/z* 409.1213 $[M+H]^+$) shift of the mass of Marfey adduct **9** (Figure 3C, Figure S13), supporting a ring substitution at C-5 to form intermediary L-DOPA. Notably, the C-5 hydroxylation of Tyr must be catalyzed by RhzB, since *E. coli* cannot form L-DOPA with its autologous enzymes.[28] RhzB in vitro activity assays were so far unsuccessful (SI). While no free L-DOPA was detected during in vivo time course experiments (Figure S15), likely due to an immediate transformation by

RhzB, this underscores that L-DOPA is not intrinsically formed by *E. coli.*

The RhzB-catalyzed transformation of L-DOPA into Fua is remarkable because a single extradiol or intradiol ring cleavage of L-DOPA would not be sufficient for furan ring formation. An initial intradiol cleavage would require the subsequent removal of two individual carbons, e.g. by decarboxylation, which would be mechanistically rather unlikely (Figure S16). Additionally, the intradiol cleavage pathway would allow for deuterium exchange during furan ring formation. This would result in a 1:1 ratio of d_2-/d_3–Fua when supplying ring-2,3,6-d_3–L-DOPA or ring-2,3,5,6-d_4–Tyr, contrasting with d_3–Fua being the sole labeled species observed in our assays (Figure 3C, Figure S13). Considering that all deuterium atoms at C-2, C-3, and C-6 are conserved in the final product, a sequential extradiol C–C cleavage route appears to be more plausible. This cleavage of L-DOPA would remove both hydroxylated carbons (C-4 and C-5), possibly as oxalic acid. A 1,4-reduction would result in an enolate (**11**) that could undergo a Paal-Knorr-like heterocyclization (Figure S17).[29] Finally, elimination of water would provide the furyl heterocycle of Fua (Figure 3E). Although the precise mechanism of the dual C–C bond cleavage remains to be clarified, this general model considers all information gleaned from our stable isotope labeling experiments.

RhzB is an Unusual New Member of the HDAO Superfamily

The intriguing biocatalytic capability of RhzB raises questions on this enzyme's relationship to other dioxygenases. A similarity search indicated that RhzB is related to members of the L-tryptophan-2,3-dioxygenase (TDO) superfamily. Like RhzB, some members of this family utilize aromatic substrates beyond Trp,[30] and thus, the TDO superfamily has recently been redefined as heme-dependent aromatic oxygenases (HDAOs).[31] To further assign RhzB to the HDAO superfamily, we predicted the core architecture of RhzB by AlphaFold 2[32] and compared it to the prototypic *Xanthomonas campestris* TDO (*Xc*TDO) using the RCSB PDB structural alignment tool.[33] The α-helix-rich core architecture of RhzB and the *Xc*TDO are highly similar with a root-mean-square deviation (RMSD) of 3.09 Å across a stretch of 223 C_α atoms (Figure 4A). This is comparable to the structural alignment score of the *Xc*TDO and the human IDO1 (Figure S18C), which both catalyze an identical reaction. The richness in α-helices predicted for RhzB is also characteristic for HDAOs[31] and contrasts this group from the non-heme Fe(II) extra- and intradiol dioxygenases rich in β-sheets.[34]

Based on a superimposition of the cofactor-binding site of the RhzB and *Xc*TDO core structures, we identified the heme *b* histidine ligation site in RhzB, another hallmark of HDAOs[35] (Figure S19). To test if the heme *b* cofactor is also present in RhzB, we performed a pyridine hemachromagen assay[36] and recorded the differential absorption spectrum of purified RhzB (Figure 4B). All three peaks at 555, 524, and 419 nm are in agreement with the known values for the α, β, and Soret peak of heme *b*,[37] indicating that RhzB possesses the typical cofactor of HDAOs. Besides the conserved HDAO core domain, RhzB has an additional N-terminal domain of unknown function (DUF, Figure 4A). Structural superimposition showed that the RhzB DUF also has some architectural similarity with *Xc*TDO, albeit only covering about half of the *Xc*TDO stretch (Figure S18G). To investigate if the N-terminal DUF is essential in Fua formation, we created a truncated variant of RhzB by excising the 0.5 kbp DUF-encoding region and expressed the construct in *E. coli.* HPLC-HRMS analysis of the culture extract showed that the production of Fua was fully abrogated (Figure 4A, Figure S20), hinting at an involvement of the N-terminal DUF in the unusually complex Tyr to Fua transformation.

Since RhzB meets all biochemical and structural criteria for an assignment as a new HDAO, we sought to gain insight into its phylogenetic relationship to similar oxygenases. Thus, we constructed a phylogenetic tree using protein sequences of selected prototypic HDAOs (Table S6) and RhzB. We found that RhzB is most similar to the cross-kingdom group of TDOs and, unexpectedly, much distant to the group of tyrosine hydroxylases (Figure 4C).

Genome Mining Using RhzB as a Handle Reveals the Bingchamide BGC

Given the characteristic didomain architecture of RhzB, we assumed that its underlying gene sequence might serve as a handle to discover the unknown bingchamide BGC through genome mining. Indeed, submitting the *rhzB* gene to BLAST analyses[38] revealed the predicted ortholog *bngC* in the genome of the bingchamide producer *S. bingchenggensis.* Notably, *bngC* is located upstream of two NRPS genes (*bngA* and *bngB*) (Figure 4D). In silico analysis of this gene cluster by antiSMASH[18] revealed an NRPS assembly line that is colinear with the peptide backbone of **3** (Figure 4D, Table S7). Tertiary structure modeling of BngC indicated that it also possesses the N-terminal DUF unique for the RhzB clade of HDAOs (Figure S21). In the phylogenetic tree of HDAOs, BngC is located close to RhzB (Figure 4C). To test if BngC is indeed a Fua synthase, we cloned synthetic, codon-optimized *bngC* (Table S8) into pET-28a and generated a recombinant BngC-producing *E. coli* strain. The expected adduct **9** was observed by HPLC-HRMS in the cell extract after amino acid derivatization by Marfey's reagent (Figure 4D). On the basis of the NRPS analysis and the successful functional reconstitution of BngC, we uncovered the elusive bingchamide BGC. We performed further computational analyses applying RhzB as a handle for genome mining. Initial results, showing a high prevalence of RhzB homologs in diverse microbial taxa, indicate that it might be possible to discover novel Fua-bearing natural products by investigating microbes that encode a RhzB-like Fua synthase.

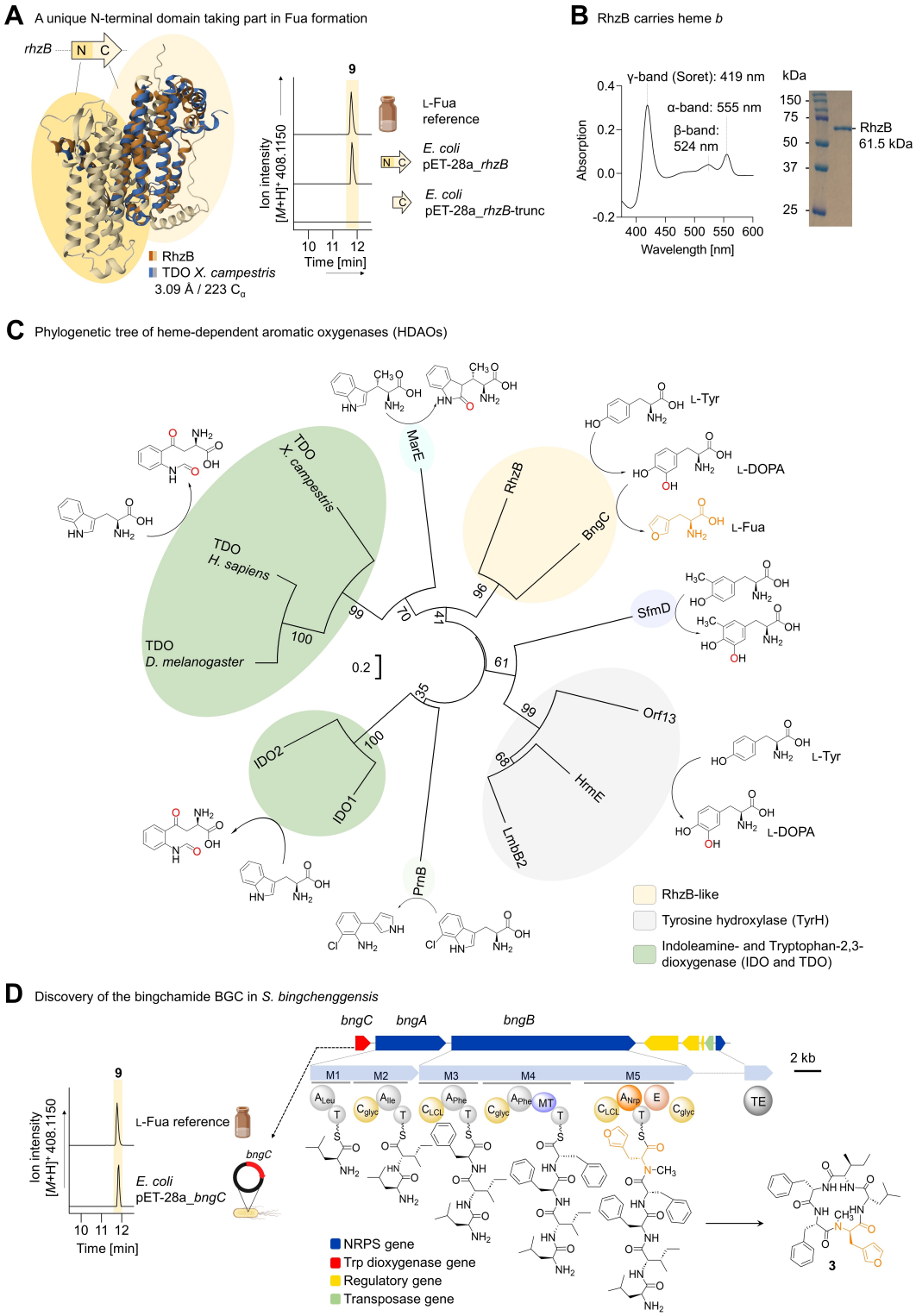

Figure 4. RhzB is a novel dioxygenase that enabled the discovery of the bingchamide BGC. A) Superimposition of RhzB and *Xc*TDO using the RCSB PDB pairwise structural alignment tool. Protein chains colored brown and blue at sites with high chain proximity, gold and gray for remote stretches. Alignment scores indicate root-mean-square deviation and stretch length of C_α atoms with high proximity. B) RhzB carries heme *b*. Differential UV/Vis absorption spectrum of purified RhzB (see SDS-PAGE). Pyridine hemachromagen assay, differential spectrum obtained by subtracting hemichrome (oxidized state) values from the hemachrome (reduced state) data. Dotted lines highlight absorption maxima. C) Phylogenetic tree of heme-dependent aromatic oxygenases (HDAOs). Protein sequences applied for phylogenetic tree construction are shown in Table S6. D) Discovery of the bingchamide BGC in *Streptomyces bingchenggensis*. The HPLC–HRMS traces in the extracted ion chromatogram (*m/z* 408.1150 $[M+H]^+$) are representative for a triplicate experiment. **9**: Marfey adduct of Fua. The methyltransferase (MT) domain located in module 4 is expected to act on the substrate of module 5. The configuration of **3** was inferred from the presence of epimerase domains in the NRPS architecture.

Conclusion

The biosynthetic origin of the infamous toxin rhizonin and its unusual furyl-substituted amino acid building blocks, L- and D-Fua, has posed a long-standing riddle. In this study we finally elucidate the biosynthetic code for rhizonin biosynthesis, shed light on Fua formation and reveal a highly unusual dioxygenase that plays the key role in this intriguing amino acid conversion. Our findings have important implications for various fields of research. First, knowledge of the genetic basis for rhizonin biosynthesis will aid in genetically detecting dangerous sources of rhizonin. Notably, these lethal, Fua-containing hepatotoxins are produced by bacteria that can live within the mycelium of *Rhizopus*, a fungus frequently used for food fermentation. Second, the intriguing RhzB-catalyzed biotransformation of a catechol into the pharmacophoric furan ring represents a previously unknown avenue to furan moieties in specialized metabolites.[39] RhzB is an important addition to the recently redefined group of heme-dependent aromatic oxygenases (HDAO),[30] markedly extending the catalytic scope of this enzyme family. Third, the successful heterologous reconstitution of Fua biosynthesis, along with the knowledge of the Fua-incorporating NRPS modules, paves the way to rationally engineer Fua-containing peptides. The position-selective introduction of Fua into peptide backbones will not only alter the biological activities but will also permit specific chemical derivatizations and cyclizations. Finally, by unearthing the elusive bingchamide BGC, we have demonstrated that RhzB serves as a guide for genome mining. This approach may also be employed to efficiently prioritize microbes for the discovery of novel Fua-carrying natural products in the future.

Author Contributions

F.J.E., S.P.N., B.D., M.D., K.S., and C.H. designed research; F.J.E., K.S., C.R., and G.L. performed stable isotope labeling and metabolic profiling; S.P.N. and B.D. identified the rhizonin BGC, constructed deletion mutants (with J.K.) and performed metabolic profiling of *M. endofungorum* (with F.J.E.), T.P.S and S.J.P. performed PacBio sequencing of *M. endofungorum*, F.J.E., S.P.N. and B.D. analyzed data; F.J.E. and M. D. performed cloning and heterologous expression, F.J.E. performed assays and in silico analyses, F.J.E., S.P.N., B.D., and C.H. wrote the paper.

Acknowledgements

We thank A. Perner and B. Bartels for HPLC-HRMS measurements, H. Heineke for NMR experiments, and V. Hänsch for helpful discussions on the Fua biosynthesis mechanism. This work was financially supported by the Deutsche Forschungsgemeinschaft (Project-ID 239748522, SFB 1127), and the Leibniz Award (to C.H.). Open Access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available in the supplementary material of this article.

Keywords: Amino Acids · Biosynthesis · Genome Mining · Stable Isotope Labeling · Toxins

[1] L. Wang, N. Wang, W. Zhang, X. Cheng, Z. Yan, G. Shao, X. Wang, R. Wang, C. Fu, *Signal Transduction Targeted Ther.* **2022**, *7*, 48.
[2] M. R. Naylor, A. T. Bockus, M. J. Blanco, R. S. Lokey, *Curr. Opin. Chem. Biol.* **2017**, *38*, 141–147.
[3] T. Dang, R. D. Süssmuth, *Acc. Chem. Res.* **2017**, *50*, 1566–1576.
[4] a) A. Stanišić, H. Kries, *ChemBioChem* **2019**, *20*, 1347–1356; b) G. H. Hur, C. R. Vickery, M. D. Burkart, *Nat. Prod. Rep.* **2012**, *29*, 1074–1098.
[5] D. E. Lewis, F. W. Dunn, *Arch. Biochem. Biophys.* **1964**, *107*, 363–366.
[6] S. J. V. Rensburg, I. F. H. Purchase, J. J. V. d. Watt, in *Symposium on Mycotoxins in Human Health* (Ed.: I. F. H. Purchase), Palgrave Macmillan, London, **1971**.
[7] T. Wilson, C. J. Rabie, J. E. Fincham, P. S. Steyn, M. A. A. Schipper, *Food Chem. Toxicol.* **1984**, *22*, 275–281.
[8] P. S. Steyn, A. A. Tuinman, F. R. van Heerden, P. H. van Rooyen, P. L. Wessels, C. J. Rabie, *J. Chem. Soc. Chem. Commun.* **1983**, *11*, 47–49.
[9] K. Shimokawa, K. Yamada, D. Uemura, *Bioorg. Med. Chem. Lett.* **2009**, *19*, 867–869.
[10] a) L. P. Partida-Martinez, C. F. D. Looß, K. Ishida, M. Ishida, M. Roth, K. Buder, C. Hertweck, *Appl. Environ. Microbiol.* **2007**, *73*, 793–797; b) L. P. Partida-Martinez, I. Groth, I. Schmitt, W. Richter, M. Roth, C. Hertweck, *Int. J. Syst. Evol. Microbiol.* **2007**, *57*, 2583–2590; c) P. Estrada-de Los Santos, M. Palmer, B. Chávez-Ramírez, C. Beukes, E. T. Steenkamp, L. Briscoe, N. Khan, M. Maluk, M. Lafos, E. Humm, M. Arrabit, M. Crook, E. Gross, M. F. Simon, F. B. Dos Reis Junior, W. B. Whitman, N. Shapiro, P. S. Poole, A. M. Hirsch, S. N. Venter, E. K. James, *Genes* **2018**, *9*, 389.
[11] W. S. Xiang, J. D. Wang, X. J. Wang, J. Zhang, *J. Antibiot.* **2009**, *62*, 501–505.
[12] C. Almeida, C. Silva Pereira, V. Gonzalez-Menendez, G. Bills, J. Pascual, M. Sánchez-Hidalgo, S. Kehraus, O. Genilloud, *Appl. Environ. Microbiol.* **2018**, *84*, e00660–00618.
[13] C. Almeida, F. E. Maddah, S. Kehraus, G. Schnakenburg, G. M. König, *Org. Lett.* **2016**, *18*, 528–531.
[14] K. W. Decoene, W. Vannecke, T. Passioura, H. Suga, A. Madder, *Biomedicine* **2018**, *6*, 99.
[15] K. Hoogewijs, D. Buyst, J. M. Winne, J. C. Martins, A. Madder, *Chem. Commun.* **2013**, *49*, 2927–2929.
[16] F. El Maddah, S. Kehraus, M. Nazir, C. Almeida, G. M. König, *J. Nat. Prod.* **2016**, *79*, 2838–2845.
[17] S. P. Niehs, B. Dose, K. Scherlach, S. J. Pidot, T. P. Stinear, C. Hertweck, *ACS Chem. Biol.* **2019**, *14*, 1811–1818.
[18] K. Blin, S. Shaw, K. Steinke, R. Villebro, N. Ziemert, S. Y. Lee, M. H. Medema, T. Weber, *Nucleic Acids Res.* **2019**, *47*, W81–W87.
[19] a) L. P. Partida-Martinez, C. Hertweck, *ChemBioChem* **2007**, *8*, 41–45; b) S. P. Niehs, K. Scherlach, B. Dose, Z. Uzum, T. P.

Stinear, S. J. Pidot, C. Hertweck, *PNAS Nexus* **2022**, *1*, 152; c) E. V. Bratovanov, K. Ishida, B. Heinze, S. J. Pidot, T. P. Stinear, J. D. Hegemann, M. A. Marahiel, C. Hertweck, *ACS Chem. Biol.* **2020**, *15*, 1169–1176.
[20] B. O. Bachmann, J. Ravel, *Methods Enzymol.* **2009**, *458*, 181–217.
[21] a) S. C. Wenzel, P. Meiser, T. M. Binz, T. Mahmud, R. Müller, *Angew. Chem. Int. Ed.* **2006**, *45*, 2296–2301; b) M.-S. Kim, M. Bae, Y.-E. Jung, J. M. Kim, S. Hwang, M. C. Song, Y. H. Ban, E. S. Bae, S. Hong, S. K. Lee, S.-S. Cha, D.-C. Oh, Y. J. Yoon, *Angew. Chem. Int. Ed.* **2021**, *60*, 19766–19773; c) S. Yu, E. Fiss, R. Jacobs William, *J. Bacteriol.* **1998**, *180*, 4676–4685.
[22] a) E. J. Kim, Y. E. Du, Y. H. Ban, Y.-H. Shin, *Front. Microbiol.* **2021**, *12*, 626881; b) X. Cai, L. Zhao, H. B. Bode, *ACS Synth. Biol.* **2023**, *12*, 203–212.
[23] S. P. Niehs, B. Dose, K. Scherlach, M. Roth, C. Hertweck, *ChemBioChem* **2018**, *19*, 2167–2172.
[24] a) S. P. Niehs, B. Dose, S. Richter, S. J. Pidot, H. M. Dahse, T. P. Stinear, C. Hertweck, *Angew. Chem. Int. Ed.* **2020**, *59*, 7766–7771; b) *Angew. Chem.* **2020**, *132*, 7840–7845.
[25] A. M. Burroughs, M. E. Glasner, K. P. Barry, E. A. Taylor, L. Aravind, *J. Biol. Chem.* **2019**, *294*, 10211–10235.
[26] S. L. Guimarães, J. B. Coitinho, D. M. Costa, S. S. Araújo, C. P. Whitman, R. A. Nagem, *Biochemistry* **2016**, *55*, 2632–2645.
[27] a) Y. Wang, I. Shin, Y. Fu, K. L. Colabroy, A. Liu, *Biochemistry* **2019**, *58*, 5339–5350; b) K. L. Colabroy, I. R. Smith, A. H. S. Vlahos, A. J. Markham, M. E. Jakubik, *Biochim. Biophys. Acta* **2014**, *1844*, 607–614.
[28] M. Kanehisa, S. Goto, *Nucleic Acids Res.* **2000**, *28*, 27–30.
[29] V. Amarnath, K. Amarnath, *J. Org. Chem.* **1995**, *60*, 301–307.
[30] Y. Wang, I. Davis, I. Shin, D. J. Wherritt, W. P. Griffith, K. Dornevil, K. L. Colabroy, A. Liu, *ACS Catal.* **2019**, *9*, 4764–4776.
[31] I. Shin, Y. Wang, A. Liu, *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2106561118.
[32] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, *Nature* **2021**, *596*, 583–589.
[33] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **2000**, *28*, 235–242.
[34] F. H. Vaillancourt, J. T. Bolin, L. D. Eltis, *Crit. Rev. Biochem. Mol. Biol.* **2006**, *41*, 241–267.
[35] R. Dick, B. P. Murray, M. J. Reid, M. A. Correia, *Arch. Biochem. Biophys.* **2001**, *392*, 71–78.
[36] I. Barr, F. Guo, *Bio-Protoc.* **2015**, *5*, e1594.
[37] E. A. Berry, B. L. Trumpower, *Anal. Biochem.* **1987**, *161*, 1–15.
[38] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J. Mol. Biol.* **1990**, *215*, 403–410.
[39] a) N. W. Gaikwad, K. M. Madyastha, *Biochem. Biophys. Res. Commun.* **2002**, *290*, 589–594; b) Y.-M. Chiang, E. Szewczyk, A. D. Davidson, N. Keller, B. R. Oakley, C. C. C. Wang, *J. Am. Chem. Soc.* **2009**, *131*, 2965–2970; c) S. Zhou, N. R. Malet, L. Song, C. Corre, G. L. Challis, *Chem. Commun.* **2020**, *56*, 14443–14446.

Manuscript received: June 21, 2023
Accepted manuscript online: August 31, 2023
Version of record online: September 13, 2023