

Article

## Uncovering the Formation and Selection of Benzylmalonyl-CoA from the Biosynthesis of Splenocin and Enterocin Reveals a Versatile Way to Introduce Amino Acids into Polyketide Carbon Scaffolds

Chenchen Chang, Rong Huang, Yan Yan, Hongmin Ma, Zheng Dai, Benying Zhang, Zixin Deng, Wen Liu, and Xudong Qu

*J. Am. Chem. Soc.*, **Just Accepted Manuscript** • DOI: 10.1021/jacs.5b00728 • Publication Date (Web): 12 Mar 2015

Downloaded from <http://pubs.acs.org> on March 16, 2015

### Just Accepted

“Just Accepted” manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides “Just Accepted” as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. “Just Accepted” manuscripts appear in full in PDF format accompanied by an HTML abstract. “Just Accepted” manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). “Just Accepted” is an optional service offered to authors. Therefore, the “Just Accepted” Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the “Just Accepted” Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these “Just Accepted” manuscripts.

# Uncovering the Formation and Selection of Benzylmalonyl-CoA from the Biosynthesis of Splenocin and Enterocin Reveals a Versatile Way to Introduce Amino Acids into Polyketide Carbon Scaffolds.

Chenchen Chang,<sup>§,†</sup> Rong Huang,<sup>§,†</sup> Yan Yan,<sup>§,‡</sup> Hongmin Ma,<sup>†</sup> Zheng Dai,<sup>†</sup> Benying Zhang,<sup>†</sup> Zixin Deng,<sup>†</sup> Wen Liu,<sup>‡</sup> and Xudong Qu<sup>\*,†</sup>

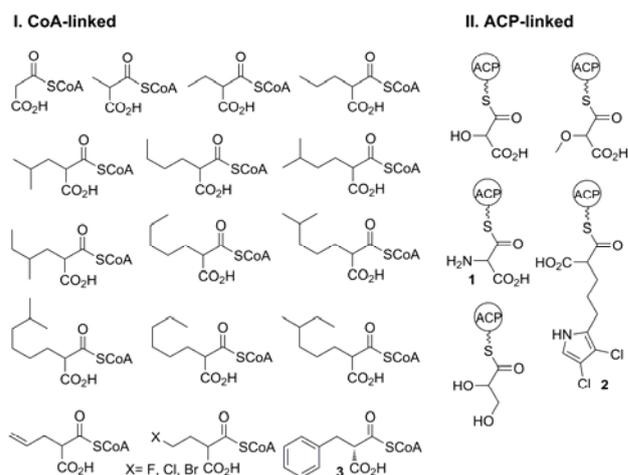
<sup>†</sup>Key Laboratory of Combinatorial Biosynthesis and Drug Discovery (Wuhan University), Ministry of Education, Wuhan University School of Pharmaceutical Sciences, 185 Donghu Rd., Wuhan 430071, China.

<sup>‡</sup>State Key Laboratory of Bioorganic and Natural Products Chemistry, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, 345 Lingling Road, Shanghai 200032, China.

**ABSTRACT:** Selective modification of carbon scaffolds via biosynthetic engineering is important for polyketide structural diversification. Yet, this scope is currently restricted to simple aliphatic groups due to 1) limited variety of CoA-linked extender units, which lack aromatic structures and chemical reactivity, and 2) narrow acyltransferase (AT) specificity, which is limited to aliphatic CoA-linked extender units. In this report, we uncovered and characterized the first aromatic CoA-linked extender unit benzylmalonyl-CoA from the biosynthetic pathways of splenocin and enterocin in *Streptomyces sp.* CNQ431. Its synthesis employs a deamination/reductive carboxylation strategy to convert phenylalanine into benzylmalonyl-CoA, providing a link between amino acid and CoA-linked extender unit synthesis. By characterization of its selection, we further validated that AT domains of splenocin and antimycin polyketide synthases are able to select this extender unit to introduce the phenyl group into their dilactone scaffolds. The biosynthetic machinery involved in the formation of this extender unit is highly versatile and can be potentially tailored for tyrosine, histidine and aspartic acid. The disclosed aromatic extender unit, amino acid-oriented synthetic pathway, and aromatic-selective AT domains provides a systematic breakthrough towards current knowledge of polyketide extender unit formation and selection, and also opens a route for further engineering of polyketide carbon scaffolds using amino acids.

Polyketides are a large family of structurally diverse natural products possessing various important pharmacological activities.<sup>1</sup> Indeed, many of these secondary metabolites have been used for treatment of human, animal or plant diseases. However, the synthetic accessibility of polyketide scaffolds, in the attempt to improve pharmacological performance, is made difficult due to their inert reactivity and structural complexity. Thus, alternative methods to modify polyketide structures, such as through biosynthetic engineering, can be a more efficient route. For example, the acyltransferase (AT) substrate specificity of polyketide synthase (PKS) can be altered through the use of domain swapping or site-directed mutagenesis that allows for the incorporation of different extender units into the polyketide carbon scaffold.<sup>2</sup> However, the ability to make changes to the polyketide scaffold via this method is currently limited by both the extender unit diversity and AT specificity.

For extender units, there are two varieties that can be selected by PKS: the CoA-linked and acyl carrier protein (ACP)-linked (Figure 1).<sup>3</sup> The most commonly observed CoA-linked extender units, malonyl-CoA and alkylmalonyl-CoAs, are synthesized by either  $\alpha$ -carboxylation of acyl-CoA,<sup>3a</sup> ligation of malonate with CoA<sup>4</sup>, or reductive carboxylation of  $\alpha$ ,  $\beta$ -unsaturated acyl-CoA.<sup>3b,3e-k</sup> ACP-linked extender units, on the other hand, are less prevalent and only used by few PKS systems.<sup>3a,3c,d,3j</sup> Substrates for CoA-linked extender units mostly originate from fatty acid metabolism. As a consequence, their

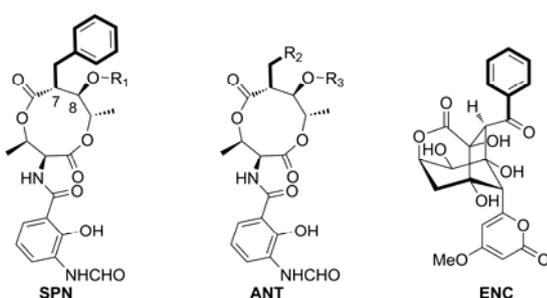


**Figure 1.** Structures of natural polyketide extender units. Aminomalonyl-ACP (1), dichloropyrrolylmalonyl-ACP (2), and benzylmalonyl-CoA (3) are derived from amino acids. Biosynthetic details of benzylmalonyl-CoA (3) are further elucidated in this study.

$\alpha$ -side groups mostly contain simple aliphatics with limited chemical reactivity and lack aromatic structures. In addition, the specificity of AT domains employ extender units as substrates is also limited in their specificity for aliphatic extender

units. Thus, due to limited extender unit diversity and narrow AT specificity, the current range of chemical space that can be introduced into polyketide scaffolds through extender unit selection is mostly restricted to simple aliphatics. Although *in vitro* extender unit synthesis<sup>5</sup>, or utilizing synthetic acyl-N-acetylcysteamine (acyl-NAC) thioesters as extender unit surrogates<sup>6</sup> can somewhat alleviate this predicament, their preparation cost, and low incorporation efficiency prevent their widespread utility.

Like fatty acids, amino acids are the other major carbon source that are widely used as biosynthetic building blocks for peptides and alkaloids synthesis. In polyketide assembly their involvement is limited, where only serine and proline are used in the synthesis of aminomalonyl-ACP (**1**) and dichloropyrrolypropylmalonyl-ACP (**2**) extender units (Figure 1).<sup>3c,3j</sup> Unlike fatty acids, amino acids are rich in functional groups (e.g. amino and thiol groups) that can further serve as chemical handles for semi-synthesis. Moreover, aromatic structures present in certain amino acids can be further diversified by available biosynthetic machinery that includes tryptophan halogenase<sup>7</sup> and phenylalanine hydroxylase<sup>8</sup>. The ability to biosynthetically incorporate and subsequently modify aromatic groups on polyketide scaffolds could help to enhance a molecule's bioavailability and bioactivity due to effectively increasing lipophilicity and/or form  $\pi$ - $\pi$  stacking interactions with the binding site of a target protein.<sup>9</sup> Therefore, finding a versatile strategy to convert amino acids into the more commonly used CoA-extender units would be highly desirable. Accordingly, based on *in vivo* and *in vitro* characterization, we herein report the formation and selection of an aromatic extender unit benzylmalonyl-CoA (**3**) (Figure 1) from the biosynthetic pathways of splenocin (SPN) and enterocin (ENC) (Figure 2). We found that this pathway employs a deamination/reductive carboxylation strategy to convert phenylalanine into a CoA-linked extender unit, and then it utilizes an aromatic-selective AT domain to select benzylmalonyl-CoA (**3**) for incorporation of a phenyl ring into the dilactone scaffold of SPNs. This versatile extender unit synthetic pathway and the broadly-selective AT described here, set the solid foundation for further expansion of polyketide structural diversity through the introduction of additional amino acids.



**Figure 2.** Structures of SPNs, ANTs and ENC. R<sub>1</sub>, R<sub>3</sub> are hydrogen or acyl groups, and R<sub>2</sub> denotes alkyl groups. SPN-C (R<sub>1</sub>= COCH(CH<sub>3</sub>)C<sub>2</sub>H<sub>5</sub>) and SPN-J (R<sub>1</sub>= H) are the focus of this study. For details of ANTs concerned in this study please see Figure S1.

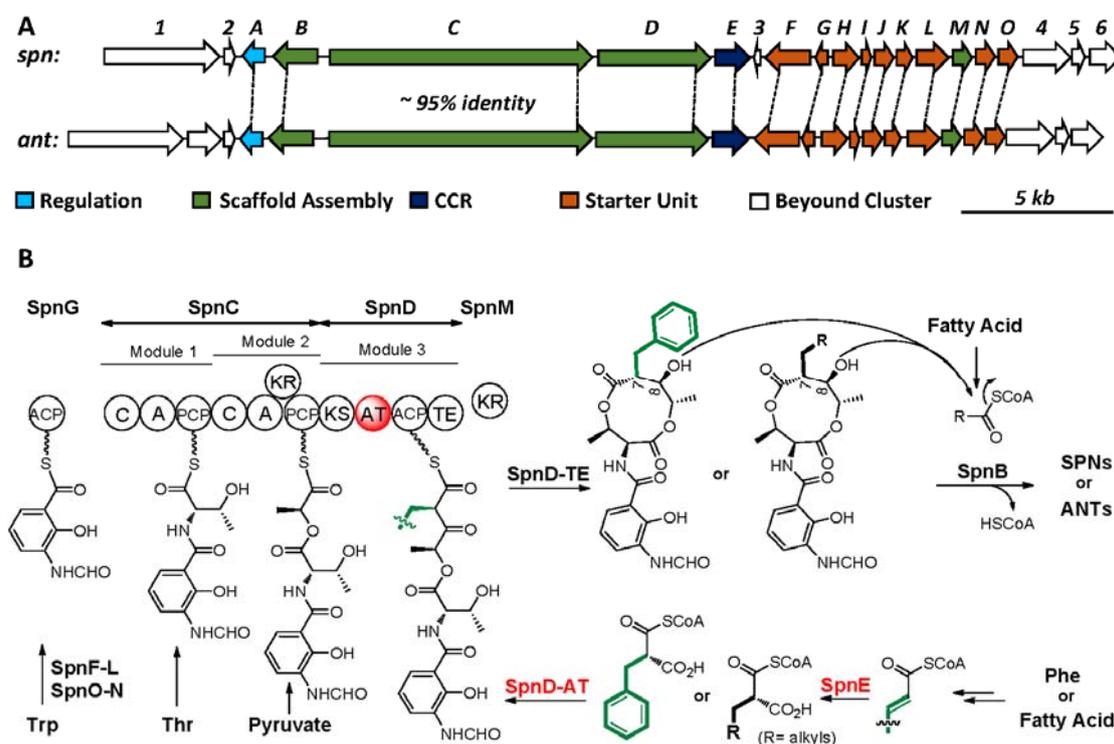
## RESULTS

**Discovering Polyketides with Amino Acid Origin.** We initially performed an extensive literature search to find polyketides of amino acid origin. We analyzed compounds using

retro biosynthetic analysis wherein we dissected each molecule into its corresponding extender units to determine whether it originated from an amino acid. From this analysis, we identified one group of natural products termed splenocins (SPNs, Figure 2). SPNs, produced by the marine bacteria *S. sp.* CNQ431, have potent anti-inflammatory activities.<sup>10</sup> Interestingly, they share an identical dilactone scaffold with antimycins (ANTs, Figure 2), but have an unusual benzyl group instead of alkyl groups at C7. In ANTs the C7-alkyls originate from extender unit that consist of alkylmalonyl-CoAs and are introduced by the AT domain of PKS AntD.<sup>11</sup> The resemblance of SPNs and ANTs, indicated to us that they may derive from similar biosynthetic pathways. Because a benzyl group is part of the scaffold in SPNs, this further led us to speculate that both a novel aromatic extender unit benzylmalonyl-CoA (**3**) and an aromatic-selective AT domain might be involved in the installation of a benzyl group into SPNs. Additionally, in living organisms, benzyl compounds are often derived from the aromatic amino acid phenylalanine, suggesting that a novel route to convert phenylalanine into the assumed CoA-linked extender unit is involved.

**Cloning the Splenocin Biosynthetic Gene Cluster.** To provide further detail into the biosynthetic mechanism of SPNs, we first set out to clone the SPN biosynthetic gene cluster. A fosmid library of CNQ431 was constructed and screened by PCR amplification according to the nonribosomal peptide synthase (NRPS)/PKS gene *antC* of the ANT biosynthetic pathway from *S. sp.* NRRL 2288.<sup>11a</sup> Sequencing the positive fosmid allowed us to identify a cluster of 16-genes (Figure 3A and Table 1). With the exception of a redundant transposase gene (*orf1*), this cluster is highly conserved in sequence (~95 % overall identity) and organization to that of the *S*-form *ants*.<sup>11a</sup> suggesting that they share a similar biosynthetic mechanism. Further bioinformatic analysis of the genes within this locus revealed three genes (*spnC*, *D* and *M*) that encode a hybrid NRPS/PKS system putatively for dilactone scaffold formation, one gene (*spnE*) that encodes a crotonyl-CoA reductive carboxylase (CCR) putatively for synthesizing extender units, and eleven other genes that putatively encode the starter unit 3-formamidosalicyl-SpnG (*spnF-L*, *N* and *O*), C8 acylation (*spnB*) and pathway regulation (*spnA*). To correlate it with SPN production, the *nrps/pks spnC* was further selected for gene inactivation. LC-HRMS analysis of mutant strain fermentation confirmed it completely lost the ability to produce SPN-C (in our hands, we only observe SPN-C in the fermentation of CNQ431) as well as a few accompanied ANTs (A1-A5, A10 and A21, Figure 4 trace II and Figure S1), hence establishing this gene cluster is indeed responsible for encoding both of the SPNs and ANTs.

Based on the available model of ANT biosynthesis,<sup>11-12</sup> we next proposed the assembly logic of *spn* (Figure 3B). Tryptophan is first converted into the starter unit 3-formamidosalicyl-SpnG by successive participation of SpnN, F, G, H-L and O, and then sequentially condensed with L-threonine and pyruvate by the NRPS SpnC to form a peptidyl intermediate. SpnD, a PKS containing ketosynthase (KS), AT, ACP, and a thioesterase (TE) domains, selects benzylmalonyl-CoA (**3**) or alkylmalonyl-CoAs as the extender unit and catalyzes the condensation with the peptidyl intermediate to afford a C2-extension. The resulting polyketide-peptidyl intermediates are further reduced by SpnM at C8, and then cyclized by SpnD-TE to form the dilactone scaffold. Finally, the acyltransferase

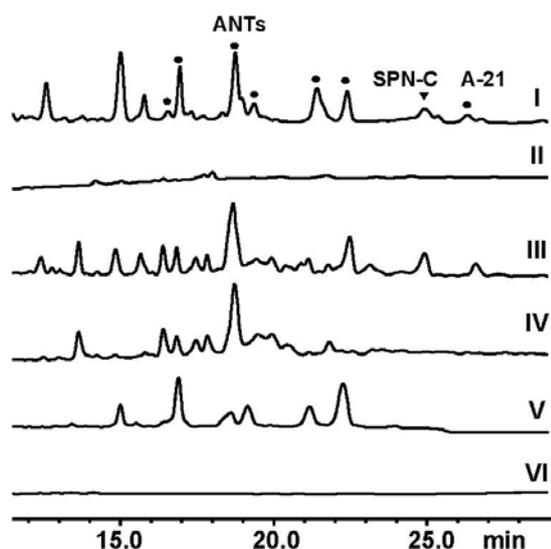


**Figure 3.** Biosynthetic gene clusters and proposed pathway. (A) Gene organization of the *spn* and *ant* (NRRL 2288) clusters, the deduced functions of the genes are labeled in color and summarized in Table 1. (B) Proposed biosynthesis of SPNs and ANTs in CNQ431. Domain notation: ACP, acyl carrier protein; C, condensation; A, adenylation; PCP, peptidyl carrier protein; KR, ketoreductase; KS, ketosynthase; TE, thioesterase.

**Table 1.** Deduced Functions of ORFs in *spn* Biosynthetic Gene Cluster (accession number KP719128).

Gene	Size (a.a.)	Protein Homolog and Origin	Identity (%)	Proposed function
1 <sup>a</sup>	1880	CaiC (WP_018469583), <i>S. sp.</i> LaPpAH-202	92	NRPS (incomplete)
2 <sup>a</sup>	145	(WP_026282526), <i>S. sp.</i> CNY228	83	Hypothetical protein
A	173	AntA (AGG37747), <i>S. sp.</i> NRRL 2288	100	Regulator
B	433	AntB (AGG37748), <i>S. sp.</i> NRRL 2288	92	Acyltransferase
C	2897	AntC (AGG37749), <i>S. sp.</i> NRRL 2288	95	NRPS (C-A-PCP-C-A-KR-PCP)
D	1273	AntD (AGG37750), <i>S. sp.</i> NRRL 2288	96	PKS (KS-AT-ACP-TE)
E	406	AntE (AGG37751), <i>S. sp.</i> NRRL 2288	99	Crotonyl-CoA reductive carboxylase
3 <sup>a</sup>	113	(WP_023422630); <i>S. sp.</i> GBA 94-10	93	Transposase
F	428	AntF (AGG37752), <i>S. sp.</i> NRRL 2288	97	AMP dependent CoA-ligase
G	82	AntG (AGG37753), <i>S. sp.</i> NRRL 2288	96	ACP
H	351	AntH (AGG37754), <i>S. sp.</i> NRRL 2288	97	Phenylacetate-CoA oxygenase, PaaA
I	64	AntI (AGG37755), <i>S. sp.</i> NRRL 2288	98	Phenylacetate-CoA oxygenase, PaaB
J	262	AntJ (AGG37756), <i>S. sp.</i> NRRL 2288	98	Phenylacetate-CoA oxygenase, PaaC
K	140	AntK (AGG37757), <i>S. sp.</i> NRRL 2288	98	Phenylacetate-CoA oxygenase, PaaD
L	367	AntL (AGG37758), <i>S. sp.</i> NRRL 2288	96	Phenylacetate-CoA oxygenase, PaaE
M	258	AntM (AGG37759), <i>S. sp.</i> NRRL 2288	97	3-oxoacyl-ACP reductase
N	282	AntN (AGG37760), <i>S. sp.</i> NRRL 2288	96	Tryptophan 2,3-dioxygenase
O	225	AntO (AGG37761), <i>S. sp.</i> NRRL 2288	97	Esterase
4 <sup>a</sup>	798	(WP_003946509), <i>S. albus</i>	97	Histidine kinase
5 <sup>a</sup>	134	(WP_030696821), <i>S. griseus</i>	98	Dynein regulation protein LC7
6 <sup>a</sup>	399	(WP_003946511), <i>S. albus</i>	97	ABC transporter

<sup>a</sup> orfs beyond *spn* gene cluster.



**Figure 4.** HPLC analysis of SPN-C and ANTs production in *S. sp.* CNQ431 and *S. coelicolor* recombinant strains. Peaks corresponding to SPN-C (▼) and ANTs (●) are indicated. I) Wild-type CNQ431; II) mWHU2450 ( $\Delta spnC$ ); III) mWHU2452 ( $\Delta encH$ ); IV) mWHU2451 ( $\Delta encP$ ); V) mWHU2453 (*S. coelicolor* carrying fosmid pWHU2405); and VI) wide-type *S. coelicolor*.

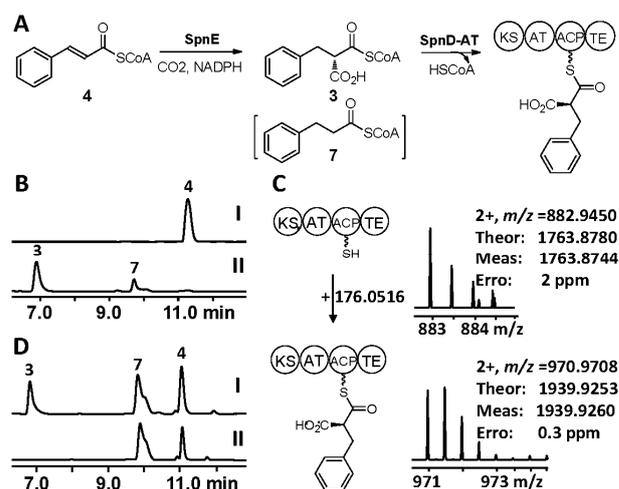
SpnB mediates the acylation of the C8 hydroxyl group to complete the biosynthesis of SPNs and ANTs.

**Characterization of Benzylmalonyl-CoA as an Extender Unit and the selectivity of the AT Domain.** In ANTs, the C7-alkyl groups are introduced by the AntD-AT from extender unit alkylmalonyl-CoAs which are synthesized through reductive carboxylation of  $\alpha$ ,  $\beta$ -unsaturated acyl-CoA precursors by the CCR protein AntE. We previously established that AntE is highly promiscuous and able to accept various substrates, including cinnamoyl-CoA (4), for conversion into their reductive carboxylated products.<sup>13</sup> In *spn*, we identified an AntE homologous protein SpnE (95 % identity), which is speculated to be responsible for producing benzylmalonyl-CoA (3) and alkylmalonyl-CoAs. To verify this assumption, we overexpressed SpnE and purified it from *E. coli* to examine its substrate specificity by LC-HRMS analysis. As anticipated, SpnE efficiently accepts both cinnamoyl-CoA (4) and alkyl- $\alpha$ ,  $\beta$ -unsaturated acyl-CoAs including crotonyl-CoA (5) and 5-methyl-2-hexenoyl-CoA (6) for conversion into corresponding reductive carboxylated products (Figure 5AB and Figure S2A-C). This provides direct evidence of benzylmalonyl-CoA (3) participating in SPN biosynthesis and establishes the capability of SpnE in providing a diverse array of extender units to this pathway.

CCRs are known to have broad selectivity, but their preference for aromatic substrates is not known. Kinetic analysis of SpnE revealed it indeed has a 2.3 fold preference of cinnamoyl-CoA (4) ( $k_{cat}/K_m = 1.42 \pm 0.008 \text{ min}^{-1} \text{ mM}^{-1}$ ) over the typical aliphatic substrate crotonyl-CoA (5) ( $k_{cat}/K_m = 0.614 \pm 0.041 \text{ min}^{-1} \text{ mM}^{-1}$ ) (Figure S2D). CCRs are typically conserved among different species (normally over 75% identity), which indicates that other CCRs may also have the ability to accept aromatic substrates to generate aromatic CoA-linked extender units.

Based on our assembly logic, SpnD-AT is assumed to select benzylmalonyl-CoA (3) or alkylmalonyl-CoAs for introduc-

tion at the C7-position on the dilactone scaffold. Amino acid sequence analysis of the specificity-conferring motifs within SpnD-AT revealed the presence of smaller amino acid residues compared to ATs that typically use malonyl-CoA, methylmalonyl-CoA or ethylmalonyl-CoA (Figure S3), and suggests the presence of a larger protein cavity that can accommodate a bulky benzyl group. To validate its function, we overexpressed the entire protein of SpnD in an *E. coli* BAP-1 strain which contains a *Bacillus* phosphopantetheinyl transferase (Sfp) gene to convert the ACP domain of SpnD into its *holo* form.<sup>14</sup> benzylmalonyl-CoA (3) was further prepared by SpnE and *in situ* incubated with the purified *holo*-SpnD to carry out the transacylation reaction. After the reaction, SpnD was purified by SDS-PAGE and subsequently trypsin digested to detect the presence of a benzylmalonyl moiety by LC-HRMS analysis (Figure 5C). As expected, a 176.0516 Da shift corresponding to the molecular weight of benzylmalonyl moiety was clearly detected in the phosphopantetheinylated serine peptide fragment (PVLVEIGPGDSLTK) of ACP. Meanwhile, we evaluated the change of benzylmalonyl-CoA (3), confirming that its consumption was consistent with the uploading activity of SpnD (Figure 5D). These results clearly validated that SpnD-AT can select benzylmalonyl-CoA (3) as an extender unit for introducing the benzyl group into the SPN dilactone scaffold.



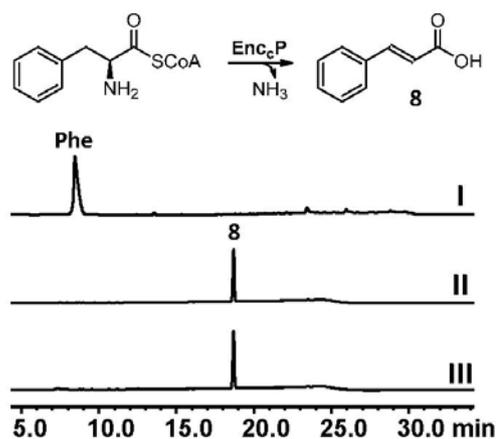
**Figure 5.** Characterization of SpnE and SpnD *in vitro*. (A) SpnE and SpnD-catalyzed reactions. (B) HPLC analysis of SpnE catalyzed reductive carboxylation of cinnamoyl-CoA (4) to produce the benzylmalonyl-CoA (3) as well as the shunt product phenylpropionyl-CoA (7) in the control (I, SpnE free) and reaction system (II). (C) HRMS analysis of molecular weight of the phosphopantetheinylated serine peptide fragment (PVLVEIGPGDSLTK) in the SpnD reaction. (D) HPLC analysis of the benzylmalonyl-CoA (3) consumption in the SpnD reaction before (I) and after (II).

**Correlation of Amino acid Origin to Benzylmalonyl-CoA Synthesis and Crosstalk of *enc* and *spn*.** Cinnamoyl-CoA (4) is a ubiquitous precursor for synthesis of phenylpropanoids in plants,<sup>15</sup> but its distribution in bacteria is rare and only known in a few pathways.<sup>16,17</sup> Both plant and bacteria systems utilize a phenylalanine ammonia-lyase (PLA) and a CoA ligase (CL) to sequentially convert phenylalanine to cinnamic acid (8) and then to cinnamoyl-CoA (4). However, in *spn*, no obvious *pal* and *cl* homologues are found. Flanking the *spn* locus are a few genes encoding NRPS (*orf1*), putative protein (*orf2*) and a

kinase (*orf4*) whose functions are not clear. To verify the potential role of *orf1*, *orf2*, or *orf4* in cinnamoyl-CoA (**4**) synthesis, the fosmid containing the *spn* locus and flanking genes (*orf1-6*) was chosen for heterologous expression. It was first engineered to bear elements for conjugation and integration, and then introduced into *S. coelicolor*. LC-HRMS analysis of the culture extract of the recombinant strain showed that it failed to produce SPNs while still being able to produce ANT<sub>s</sub> (Figure 4, trace V). This result shows that the genes involved in cinnamoyl-CoA (**4**) synthesis are not immediately flanking *spn* but, instead, are further outside of this locus.

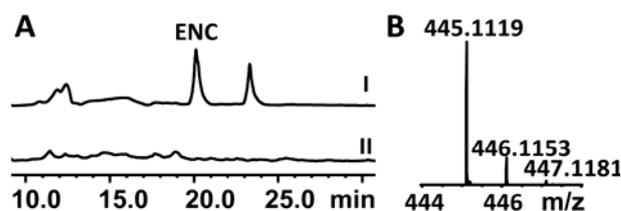
In order to get the entire genetic details, we further sequenced the CNQ431 genome.<sup>18</sup> Upon careful analysis, we identified a cluster of 19 genes which are highly identical (~96 % identity) to the cluster of ENC from *S. maritimus*.<sup>16</sup> (Figure S4, Table S2). Among the cluster, two genes (*encP*, and *encH*) are homologues to *encP* and *encH*, whose function are known as PLA and cinnamate CoA ligase, respectively.<sup>19</sup> We therefore speculated that these two genes might be responsible for supplying cinnamoyl-CoA (**4**) for benzylmalonyl-CoA (**3**) synthesis. To test this assumption, we individually inactivated the *encP* and *encH* genes, and analyzed their fermentation. Gratifyingly, LC-HRMS analysis revealed that inactivation of *encP* completely abolishes the production of SPN-C but not ANT<sub>s</sub> (except A-21) confirming the essential role of this gene in the synthesis of benzylmalonyl-CoA (**3**) (Figure 4 trace IV and Figure S1). The benzoyl group in the C8-hydroxyl of A-21 might be derived from the degradation of Enc<sub>c</sub>P-catalyzed deamination as in ENC.<sup>19a</sup> Inactivation of *encH* shows no influence on the biosynthesis of SPN-C or ANT<sub>s</sub> (Figure 4, trace III), suggesting that its function can be compensated by other CoA ligases in the genome as observed in the *enc* pathway.<sup>19a</sup> Indeed, CNQ431 has ten CoA-ligase genes in the genome, including two involved in the fatty acid degradation whose selectivity are very broad.<sup>13,20</sup> To further validate the function of Enc<sub>c</sub>P, we carried out *in vitro* studies. Enc<sub>c</sub>P was initially overexpressed by pET and MBP-fusion expression systems in *E. coli*. However, despite achieving soluble expressing of this protein in *E. coli*, no activity was detected. We therefore turned to a *Pseudomonas* expression system to successfully obtain active protein. Based on HPLC analysis, Enc<sub>c</sub>P was shown to convert phenylalanine into cinnamic acid (**8**) (Figure 6). The use of other amino acids as substrates, including tyrosine, histidine and tryptophan, did not result in the deaminated products (data not shown), suggesting Enc<sub>c</sub>P is a phenylalanine-specific ammonia lyase.

Crosstalk between secondary metabolites is not common, since it necessitates their expression coordinately which is often fermentation and temporal condition dependent. To confirm the crosstalk of *enc* and *spn*, culture extract of SPNs from CNQ431 was analyzed to see whether ENC is co-produced. LC-HRMS analysis however showed no ENC was produced, suggesting at least one gene except *encP* in the *enc* was not properly expressed under the SPN-producing condition. To further confirm this assumption, both CNQ431 and *encP*-inactivated strains were fermented under the ENC-producing condition. Based on LC-HRMS analysis, a peak with a molecular weight of ENC ( $[M+H]^+$  calcd. For C<sub>22</sub>H<sub>21</sub>O<sub>10</sub>, 445.1120; observed, 445.1119) was clearly detected in the fermentation of CNQ431, while not in the *encP*-inactivated mutant (Figure 7), confirming *enc* indeed encodes ENC and *encP* is essential for its synthesis. *enc* contains sev



**Figure 6.** HPLC analysis of Enc<sub>c</sub>P catalyzed reaction. I) Standard phenylalanine; II) Standard cinnamic acid (**8**); III) Enc<sub>c</sub>P reaction system.

en operons which controlled by promoters with different directions (Figure S4). These genes might be not coordinately expressed during SPN producing as similar to the biosynthesis of rhodochelin and erythrochelin where one of the essential genes is provided by an uncharacterized cluster.<sup>21-22</sup>

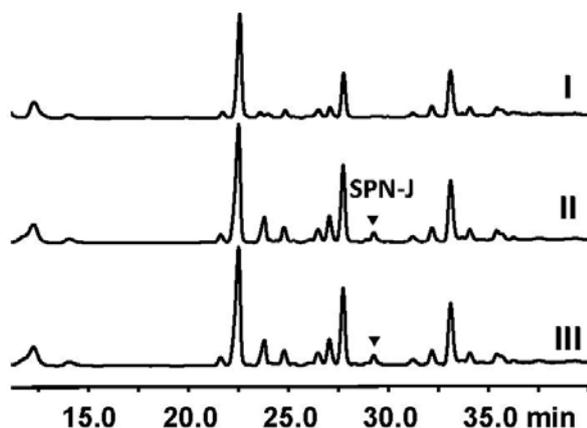


**Figure 7.** LC-HRMS analysis of ENC production in *S. sp.* CNQ431 (I) and mWHU2451 ( $\Delta encP$ ) (II). (A) HPLC trace of ENC. (B) HRMS spectrum of ENC.

**Engineering *ant* Pathway to Produce SPN.** We were next interested in whether this extender unit pathway can be applied to structural engineering. To simplify the process, we chose to modify the structurally similar ANT<sub>s</sub> via engineer the *ant* pathway from NRRL 2288.<sup>11a</sup> We first analyzed the AT domain of AntD to see if it needed to be swapped for SpnD-AT. Comparing the specificity-conferring motifs of SpnD-AT and AntD-AT, we surprisingly found they are indeed the same (Figure S3), suggesting AntD-AT is potentially able to accept benzylmalonyl-CoA (**3**). In a previous study we validated that AntE is able to catalyze the reductive carboxylation of cinnamoyl-CoA (**4**).<sup>13</sup> These results raise an interesting possibility that the non-production of SPNs by *ants* is probably due to the lack of cinnamoyl-CoA (**4**). This assumption was further supported by the fact that all of the ANT producers including *S. albus* J1074, *S. ambofaciens* ATCC 23877 and *S. sp.* NRRL 2288 do not contain genes encoding for cinnamoyl-CoA (**4**).<sup>11a</sup>

To test this assumption, we introduced the *ermE*\*-*encP* and *ermE*\*-*encP*-*encH* (*ermE*\* is a constitutive promoter) cassettes respectively into NRRL 2288-AL2110, in which the acyltransferase gene *antB* for C8-hydroxyl acylation was deleted.<sup>13</sup> This strain only produces C8-hydroxyl deacyl-ANT<sub>s</sub> that offers a cleaner metabolite background. Analysis of the recombinant strain revealed both *ermE*\*-*encP* and *ermE*\*-*encP*-*encH* can lead to production of a new compound (Fig-

ure 8), suggesting the function of *encH* can be compensated by an endogenous CoA ligase as we observed above. By scaling up the fermentation, this compound was further purified in enough quantity for spectrometric analysis. <sup>1</sup>H-NMR, <sup>13</sup>C-NMR and HRMS analysis confirmed its identity as SPN-J which contains a benzyl group in C7 (Figure 1).<sup>10</sup> This result demonstrates the feasibility of selectively engineering polyketide carbon scaffolds through this benzylmalonyl-CoA extender unit pathway. Moreover, it also demonstrates that the AT domain of AntD is able to accept benzylmalonyl-CoA (3) as an extender unit.



**Figure 8.** HPLC analysis of the SPN-J production in AL2110 recombinant strains. Peaks corresponding to SPN-J (▼) is indicated. I) Wide-type AL2110; II) mWHU2454 (AL2110 carrying *ermE\*<sup>-</sup>encP*) and III) mWHU2455 (AL2110 carrying *ermE\*<sup>-</sup>encP-encH*).

## DISCUSSION

### Novel Strategy to Convert an Amino Acid into Polyketide Extender Unit.

Fatty acids and amino acids are the major carbon sources in living organisms. Their utility in secondary metabolite synthesis rarely overlap. Amino acids are mostly used in peptide and alkaloid synthesis, while fatty acids are used for polyketide synthesis. The major reason for this is due to their distinct chemical properties. For instance, extender units of polyketides contain a carboxylic group in the *C $\alpha$*  of a thioesterified carboxylate that serves as a driving force in the Claisen condensation reaction when it is decarboxylated.<sup>23</sup> In contrast, this is not the case for amino acids due to the presence of an amino group in the *C $\alpha$*  that decreases the acidity of *C $\alpha$*  proton and prevents it from undergoing the same chemistry as with *C $\alpha$*  carboxylates. Thus, to convert amino acids into polyketide extender units, Nature utilizes a different approach. For example, in the synthesis the aminomalonyl-ACP (1), serine is first activated by loading onto an ACP, then two dehydrogenases are recruited to sequentially oxidize its *C $\beta$* -hydroxyl group into a carboxyl group.<sup>3c</sup> In the synthesis of dichloropyrrolypropylmalonyl-ACP (2), proline serves as a starter unit and is assumed to be activated and modified to form the dichloropyrroly-ACP, which further undergoes two rounds of C-2 extension with two molecules of malonyl-CoA before being converted into an extender unit.<sup>3j</sup> Since these pathways require unique biosynthetic machinery to mediate highly specific chemical reactions, it is challenging for both of Nature and engineering efforts to tailor them to suit other amino acids.

Interestingly, in *spn* Nature employs a more direct strategy to produce benzylmalonyl-CoA (3). By employing an amino acid ammonia lyase, it performs the transformation of the *C $\alpha$*  amino group of phenylalanine into a classic functional group found in fatty acids—the  $\alpha$ ,  $\beta$ -double bond. The resulting deaminated product  $\alpha$ ,  $\beta$ -unsaturated acid can be recognized by a promiscuous fatty acid CoA ligase to convert it into its CoA-form, which can subsequently go into the CCR synthetic pathway like other  $\alpha$ ,  $\beta$ -unsaturated acyl-CoAs from fatty acid metabolism to generate a CoA-linked extender unit. Distinct from the previously known ACP-linked models, this strategy unifies amino acid degradation and fatty acid derived extender unit synthesis, and is more energy-saving and concise. It can also be tailored to fit several other amino acids. Moreover, compared with ACP-linked extender units which are used by only a few PKS systems, CoA-linked extender units are more useful for polyketide engineering. The benzylmalonyl-CoA (3) pathway allows for the first time the ability to convert other major carbon source-amino acids into CoA-linked extender units, opening up a new avenues for polyketide extender unit synthesis.

### Versatility for Converting Other Amino Acids into CoA-linked Extender Units, and Significance for Expanding Extender Unit Variety.

CoA-linked extender units are majorly sourced from fatty acid metabolism and therefore mostly contain aliphatics. Currently, there are no CoA-linked extender units that have aromatic structures. Haloethylmalonyl-CoA,<sup>3e,24</sup> and allylmalonyl-CoA<sup>3f</sup> are extender units known to bear active groups on their side chains (Figure 1). These functionalities are introduced onto the extender units by unusual biosynthetic machineries making them poorly amenable as platforms for creating novel extender units. Although chemo-enzymatic approaches result in the synthesis of unusual extender units or acyl-NAC surrogates, either their preparation are costly or incorporation efficiency are low, making them unsuitable for larger scale production. Hence, finding a more versatile way to increase both the structural diversity and chemical reactivity of CoA-linked extender units is a long-term goal in the field of polyketide engineering.

Amino acids, which are abundant *in vivo* and have plenty of structural diversity and chemical reactivity, are an ideal source for CoA-linked extender unit synthesis. In this study, we disclosed an unprecedented strategy to convert an amino acid into a CoA-linked extender unit. The conversion is simple and concise. More significantly, the bio-machinery involved in the process are either readily available or highly promiscuous and make this strategy highly versatile for incorporating additional amino acids. Genes encoding amino acid deamination, such as phenylalanine, tyrosine, histidine and aspartic acid are well studied.<sup>25</sup> Some of them are broadly-selective and can even accept modified amino acids, such as *m*-tyrosine which can be generated *in vivo* by introduction of a phenylalanine *m*-hydroxylase.<sup>8,25d</sup> Bacterial CoA ligases, in particular those involved in fatty acid degradation are very promiscuous, and deaminated amino acids can be effectively converted by them or their specific CoA-ligases, such as 4-hydroxycoumarate-CoA ligase into  $\alpha$ ,  $\beta$ -unsaturated CoAs.<sup>13,20,26</sup> We and others have revealed several CCRs that are extremely promiscuous and can tolerate diverse substitutions on the *C $\alpha$* , including alkyl groups of up to ten carbons, halogenated alkyl-, cyclohexenyl-, alkynyl-substituted alkyl- and even benzyl-groups.<sup>3b,3i,13,26</sup> Although AntE and SpnE are unable to turnover 4-hydroxycoumarate-CoA and 3-indoleacrylic-CoA which

are deaminated products of tyrosine and tryptophan (data not shown), sequence analysis with the well-studied CCR CinF revealed that their binding pockets have the potential for enlargement (Figure S5). By replacing the selectivity-conferring residues (V163 and V362, numbers corresponding to CinF) with smaller residues such as glycines, these CCRs may be able to accept bulky aromatic substrates. Currently, several structures of broad-selective CCRs have been made available.<sup>31,27</sup> Based on the protein structures, directed evolution at these sites could be highly effective in expand CCR's selectivity scope. It is believed, the strategy disclosed here can be readily tailored to convert other amino acids including tyrosine, histidine and aspartic acid into CoA-linked extender units. Extender units derived from amino acids, particularly those with aromatic moieties and active groups will greatly expand their chemical diversity, and facilitate polyketide structural engineering.

### Broad Selectivity of the AT Domain and Potential to Introduce Structural Diversity into the Polyketide Scaffold.

With ATs serving as a gatekeeper for extender unit selection, they are other critical factors for structural engineering. Currently introducing aromatic groups into polyketide scaffolds through extender unit selection is not yet possible, mostly because ATs lack aromatic-selectivity.<sup>28</sup> In this study, we characterized the first example of an AT domain (SpnD-AT) that can effectively select an aromatic extender unit to introduce a benzyl group into the SPN scaffold. Through *in vivo* study we further validated that the AT domain of AntD is also competent in accepting benzylmalonyl-CoA (**3**). These discoveries overcome the current limitation of AT selection to now include aromatics.

Previously, we demonstrated AntD can accept various types and lengths of substituted extender units.<sup>13</sup> Most of these substrates are bulkier than extender units derived from amino acids, suggesting AntD-AT is able to select other amino acid-derived extender units. By coupling the versatile extender unit synthetic strategy and AT swapping, a complete pathway to introduce amino acid varieties into polyketide carbon scaffolds can likely be constructed. In addition, besides aromatic groups from amino acids, these biosynthetic proteins can also be tailored to introduce synthetic aromatics into the polyketide scaffold. CoA ligases converting cinnamate analogues into their CoA forms are available.<sup>5e,20</sup> These genes can be co-expressed with selectivity-improved CCRs to constitute an artificial pathway for converting synthetic cinnamates into CoA-linked extender units. With the selection of SpnD-AT or AntE-AT, introduction of synthetic cinnamate varieties into polyketide scaffold is potentially feasible. Currently, elucidation of the SpnE-AT structure is underway. With structural-based direct evolution, it is believed this selectivity can be engineered to allow for more diverse aromatic groups.

### CONCLUSION:

In summary, we have characterized the formation and selection of the aromatic extender unit benzylmalonyl-CoA (**3**) from the biosynthesis of splenocin and enterocin. The synthesis of this extender unit is distinct from the previous known CoA-linked extender units. By employing a deamination/reductive carboxylation strategy, phenylalanine is converted into the CoA-linked extender unit. More significantly, this pathway is concise and highly versatile which can be applied to a few amino acids including tyrosine, histidine and

aspartic acid via exchanging individual biosynthetic machinery. Additionally, their selection and incorporation into PKS assembly can be performed by the AT domains of SpnD and AntD, which overcome the current limitation of extender unit selection. This versatile extender unit synthetic pathway broadens the specificity of the AT domain and paves the way for introducing additional amino acids and synthetic aromatics with diverse structural diversity and chemical reactivity into the polyketide carbon scaffold.

### ASSOCIATED CONTENT

**Supporting Information.** Experimental details and characterization data. This material is available free of charge via the Internet at <http://pubs.acs.org>.

### AUTHOR INFORMATION

#### Corresponding Author

[quxd@whu.edu.cn](mailto:quxd@whu.edu.cn)

#### Author Contributions

§These authors contributed equally.

#### Funding Sources

This work was supported by NSFC grants (31322002, 31270119, and 31400052), Program of NCET (NCET-12-0423), the State Key Laboratory of Materials-Oriented Chemical Engineering (KL13-19), and China Postdoctoral Science Foundation (2014M560620 and 2014M552077).

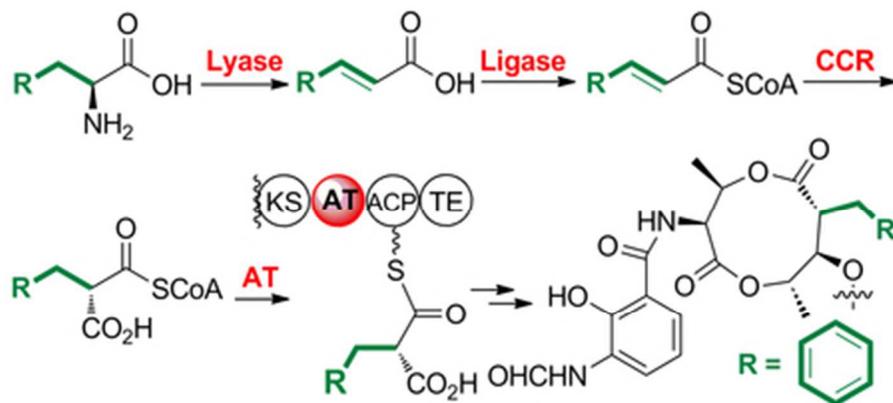
### ACKNOWLEDGMENT

We thank Prof. William Fenical at Scripps Research Institute for provision of the strain CNQ431; Prof. Ikuro Abe and Dr. Lihan Zhang at Tokyo University for gifting CoA substrates; Prof. Yizheng Zhang at Shichuan University for provision of the plasmid pCIBhis; Prof. Chao Peng at National Center for Protein Science Shanghai for performing the Mass analysis and Prof. Kui Hong at Wuhan University for provision of the ENC fermentation condition.

### REFERENCES

1. Newman, D. J.; Cragg, G. M. *J. Nat. Prod.* **2012**, *75*, 311-335.
2. (a) Williams, G. J. *Curr. Opin. Struct. Biol.* **2013**, *23*, 603-612; (b) Dunn, B. J.; Khosla, C. *J. R. Soc. Interface* **2013**, *10*, 20130297.
3. (a) Chan, Y. A.; Podevels, A. M.; Kevany, B. M.; Thomas, M. G. *Nat. Prod. Rep.* **2009**, *26*, 90-114; (b) Wilson, M. C.; Moore, B. S. *Nat. Prod. Rep.* **2012**, *29*, 72-86; (c) Chan, Y. A.; Boyne, M. T.; Podevels, A. M.; Klimowicz, A. K.; Handelsman, J.; Kelleher, N. L.; Thomas, M. G. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 14349-14354; (d) Jia, X.; Tian, Z.; Shao, L.; Qu, X.; Zhao, Q.; Tang, J.; Tang, G.; Liu, W. *Chem. Biol.* **2006**, *13*, 575-585; (e) Liu, Y.; Hazzard, C.; Eustáquio, A. S.; Reynolds, K. A.; Moore, B. S. *J. Am. Chem. Soc.* **2009**, *131*, 10376-10377; (f) Mo, S.; Kim, D. H.; Lee, J. H.; Park, J. W.; Basnet, D. B.; Ban, Y. H.; Yoo, Y. J.; Chen, S.-w.; Park, S. R.; Choi, E. A.; Kim, E.; Jin, Y.-Y.; Lee, S.-K.; Park, J. Y.; Liu, Y.; Lee, M. O.; Lee, K. S.; Kim, S. J.; Kim, D.; Park, B. C.; Lee, S.-g.; Kwon, H. J.; Suh, J.-W.; Moore, B. S.; Lim, S.-K.; Yoon, Y. J. *J. Am. Chem. Soc.* **2010**, *133*, 976-985; (g) Wilson, M. C.; Nam, S.-J.; Gulder, T. A. M.; Kauffman, C. A.; Jensen, P. R.; Fenical, W.; Moore, B. S. *J. Am. Chem. Soc.* **2011**, *133*, 1971-1977; (h) Xu, Z.; Ding, L.; Hertweck, C. *Angew. Chem. Int. Ed.* **2011**, *50*, 4667-4670; (i) Quade, N.; Huo, L.; Rachid, S.; Heinz, D. W.; Müller, R. *Nat. Chem. Biol.* **2012**, *8*, 117-124; (j) Mantovani, S. M.; Moore, B. S. *J. Am. Chem. Soc.* **2013**, *135*, 18032-18035.
4. An, J. H.; Kim, Y. S. *Eur. J. Biochem.* **1998**, *257*, 395-402.

- 1  
2  
3 5. (a) Pohl, N. L.; Hans, M.; Lee, H. Y.; Kim, Y. S.; Cane, D. E.;  
4 Khosla, C. *J. Am. Chem. Soc.* **2001**, *123*, 5822-5823; (b) Hughes,  
5 Amanda J.; Keatinge-Clay, A. *Chem. Biol.* **2011**, *18*, 165-176; (c)  
6 Koryakina, I.; Williams, G. J. *ChemBioChem* **2011**, *12*, 2289-2293; (d)  
7 Koryakina, I.; McArthur, J.; Randall, S.; Draelos, M. M.; Musiol, E.  
8 M.; Muddiman, D. C.; Weber, T.; Williams, G. J. *ACS Chem. Biol.*  
9 **2012**, *8*, 200-208; (e) Go, M. K.; Chow, J. Y.; Cheung, V. W. N.; Lim,  
10 Y. P.; Yew, W. S. *Biochemistry* **2012**, *51*, 4568-4579.  
11 6. Kennedy, J. *Nat. Prod. Rep.* **2008**, *25*, 25-34.  
12 7. Runguphan, W.; Qu, X.; O'Connor, S. E. *Nature* **2010**, *468*, 461-  
13 464.  
14 8. Qu, X.; Jiang, N.; Xu, F.; Shao, L.; Tang, G.; Wilkinson, B.; Liu,  
15 W. *Mol. Biosyst.* **2011**, *7*, 852-861.  
16 9. (a) Ritchie, T. J.; Macdonald, S. J. *Drug Discov. Today* **2009**, *14*,  
17 1101-1020; (b) Ward, S. E.; Beswick, P. *Expert Opin. Drug Discov.*  
18 **2014**, *9*, 995-1003.  
19 10. Strangman, W. K.; Kwon, H. C.; Broide, D.; Jensen, P. R.;  
20 Fenical, W. *J. Med. Chem.* **2009**, *52*, 2317-2327.  
21 11. (a) Yan, Y.; Zhang, L.; Ito, T.; Qu, X.; Asakawa, Y.; Awakawa,  
22 T.; Abe, I.; Liu, W. *Org. Lett.* **2012**, *14*, 4142-4145; (b) Sandy, M.;  
23 Rui, Z.; Gallagher, J.; Zhang, W. *ACS Chem. Biol.* **2012**, *7*, 1956-  
24 1961.  
25 12. (a) Sandy, M.; Zhu, X.; Rui, Z.; Zhang, W. *Org. Lett.* **2013**, *15*,  
26 3396-3399; (b) Liu, J.; Zhu, X.; Seipke, R. F.; Zhang, W. *ACS Synth.*  
27 *Biol.* **2014**; doi: 10.1021/sb5003136 (c) Seipke, R. F.; Patrick, E.;  
28 Hutchings, M. I. *Peer J.* **2014**, *2*, e253.  
29 13. Yan, Y.; Chen, J.; Zhang, L.; Zheng, Q.; Han, Y.; Zhang, H.;  
30 Zhang, D.; Awakawa, T.; Abe, I.; Liu, W. *Angew. Chem. Int. Ed.*  
31 **2013**, *52*, 12308-12312.  
32 14. Quadri, L. E. N.; Weinreb, P. H.; Lei, M.; Nakano, M. M.;  
33 Zuber, P.; Walsh, C. T. *Biochemistry* **1998**, *37*, 1585-1595.  
34 15. Hahlbrock, K.; Scheel, D. *Annu. Rev. Plant Physiol. Plant Mol.*  
35 *Biol.* **1989**, *40*, 347-369.  
36 16. Piel, J.; Hertweck, C.; Shipley, P. R.; Hunt, D. M.; Newman, M.  
37 S.; Moore, B. S. *Chem. Biol.* **2000**, *7*, 943-955.  
38 17. (a) Bezanson, G. S.; Desaty, D.; Emes, A. V.; Vining, L. C.  
39 *Can. J. Microbiol.* **1970**, *16*, 147-151; (b) Hill, A. M.; Thompson, B.  
40 L.; Harris, J. P.; Segret, R. *Chem. Commun.* **2003**, 1358-1359.  
41 18. Yu, M.; Dai, Z.; Qu, X.; Gao, X. *Genome Announc.* **2015**, *3*,  
42 e01383-14.  
43 19. (a) Xiang, L.; Moore, B. S. *J. Bacteriol.* **2003**, *185*, 399-404; (b)  
44 Xiang, L.; Moore, B. S. *J. Bacteriol.* **2005**, *187*, 4286-4289.  
45 20. Kaneko, M.; Ohnishi, Y.; Horinouchi, S. *J. Bacteriol.* **2003**,  
46 *185*, 20-27.  
47 21. Bosello, M.; Robbel, L.; Linne, U.; Xie, X.; Marahiel, M. A. *J.*  
48 *Am. Chem. Soc.* **2011**, *133*, 4587-4595.  
49 22. Lazos, O.; Tosin, M.; Slusarczyk, A. L.; Boakes, S.; Cortés, J.;  
50 Sidebottom, P. J.; Leadlay, P. F. *Chem. Biol.* **2010**, *17*, 160-173.  
51 23. Hertweck, C. *Angew. Chem. Int. Ed.* **2009**, *48*, 4688-4716.  
52 24. Eustáquio, A. S.; O'Hagan, D.; Moore, B. S. *J. Nat. Prod.* **2010**,  
53 *73*, 378-382.  
54 25. (a) Fibriansah, G.; Veetil, V. P.; Poelarends, G. J.; Thunnissen,  
55 A.-M. W. H. *Biochemistry* **2011**, *50*, 6053-6062; (b) Kyndt, J. A.;  
56 Meyer, T. E.; Cusanovich, M. A.; Van Beeumen, J. J. *FEBS Lett.*  
57 **2002**, *512*, 240-244; (c) Schwede, T. F.; Rétey, J.; Schulz, G. E. *Bio-*  
58 *chemistry* **1999**, *38*, 5355-5361; (d) Turner, N. J. *Curr. Opin. Chem.*  
59 *Biol.* **2011**, *15*, 234-240; (e) Raj, H.; Szymański, W.; de Villiers, J.;  
60 Rozeboom, H. J.; Veetil, V. P.; Reis, C. R.; de Villiers, M.; Dekker, F.  
J.; de Wildeman, S.; Quax, W. J.; Thunnissen, A.-M. W. H.; Feringa,  
B. L.; Janssen, D. B.; Poelarends, G. J. *Nat. Chem.* **2012**, *4*, 478-484.  
26. Zhu, X.; Liu, J.; Zhang, W. *Nat. Chem. Biol.* **2015**, *11*, 115-120.  
27. Zhang, L.; Chen, J.; Mori, T.; Yan, Y.; Liu, W.; Abe, I. *Acta.*  
*Cryst. Section F* **2014**, *70*, 734-737.  
28. Koryakina, I.; McArthur, J. B.; Draelos, M. M.; Williams, G. J.  
*Org. Biomol. Chem.* **2013**, *11*, 4449-4458



38x17mm (300 x 300 DPI)