Title: A Defined and Flexible Pocket Explains Aryl Substrate Promiscuity by the Cahuitamycin Starter Unit Activating Enzyme CahJ

Authors: David H. Sherman, Ph.D.; Ashootosh Tripathi, Ph.D.; Sung Ryeol Park, Ph.D.; Andrew Sikkema, Ph.D.; Hyo Je Cho, Ph.D.; Jianfeng Wu, Ph.D.; Brian Lee; Chuanwu Xi, Ph.D.; Janet L. Smith, Ph.D.

To be cited as: ChemBioChem 10.1002/cbic.201800233

Link to VoR: https://doi.org/10.1002/cbic.201800233

COMMUNICATION

A Defined and Flexible Pocket Explains Aryl Substrate Promiscuity by the Cahuitamycin Starter Unit Activating Enzyme CahJ

Ashootosh Tripathi[a,e¶], Sung Ryeol Park[a¶], Andrew Sikkema[a,b¶], Hyo Je Cho[d], Jianfeng Wu[c], Brian Lee[a], Chuanwu Xi[c], Janet L. Smith[a,b], David H. Sherman[a,e,f,g*]

[a] *Life Sciences Institute, University of Michigan, Ann Arbor, MI. Email: davidhs@umich.edu and ashtri@umich.edu*
[b] *Department of Biological Chemistry, University of Michigan, Ann Arbor, MI*
[c] *Department of Environmental Health Sciences, University of Michigan School of Public Health, Ann Arbor, MI*
[d] *Department of Pathology, University of Michigan, Ann Arbor, MI*
[e] *Department of Medicinal Chemistry, University of Michigan, Ann Arbor, MI*
[f] *Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109*
[g] *Department of Microbiology & Immunology, University of Michigan, Ann Arbor, Michigan.*

¶ Contributed equally to this manuscript.

Electronic Supplementary Information (ESI) available.

Abstract: Cahuitamycins are biofilm inhibitors assembled by a convergent non-ribosomal peptide synthetase pathway. Previous genetic analysis indicated that a discrete enzyme, CahJ, serves as a gatekeeper for cahuitamycin structural diversification. Herein, the CahJ protein was probed structurally and functionally to guide formation of new analogs by mutasynthetic studies. This analysis enabled in vivo production of a new cahuitamycin congener through targeted precursor incorporation.

Cahuitamycins, produced by *Streptomyces gandocaensis*, are a novel structural class that incorporate diverse aryl starter units, to generate potent biofilm inhibitors against the Gram-negative pathogenic bacterium, *Acinetobacter baumannii*[1]. This multidrug-resistant microorganism is responsible for a large number of nosocomial infections, including pneumonia, urinary tract infections, wound infections, and bacteremia with significantly high mortality rates (~60%). Biofilm formation contributes to the high rate of antimicrobial resistance (AMR) [2]. When in a biofilm, these microbes develop AMR up to 1,000 fold greater than planktonic forms of the bacterial cells[2]. Despite the significant role of biofilms in infectious diseases, there are currently no small molecule therapeutics in clinical use that specifically target biofilms[3].

Understanding cahuitamycin biosynthetic mechanisms and its key enzyme functions represents a critical step towards expanding structural diversity using pathway engineering (Fig. 1). As shown recently, the central role of starter unit selection in cahuitamycin diversification and biological activity (Fig. 1/ ESI Fig. S1), involves CahJ [1]. We were motivated to explore this essential adenylation (A) enzyme whose further manipulation could provide access to new cahuitamycin congeners.

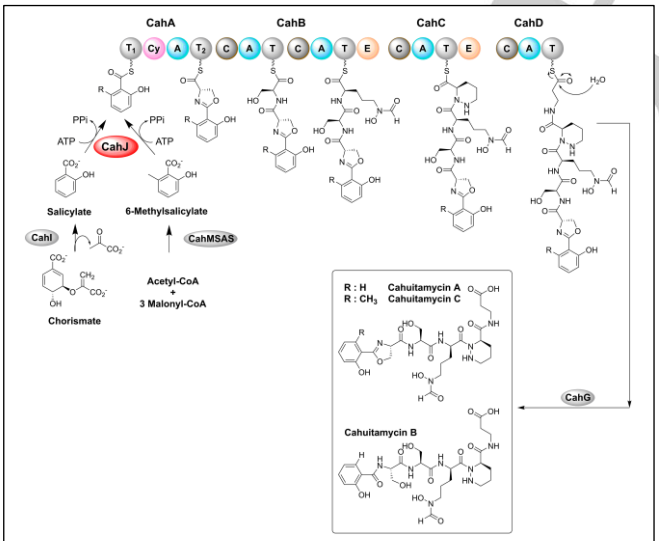

Figure 1: Proposed biosynthesis of cahuitamycins A-C in *S. gandocaensis*. The biosynthetic gene cluster contains nonribosomal peptide synthetase-encoding CahABCD along with putative genes involved in chain initiation (CahI and CahMSAS) and chain termination (CahG)). The CahJ-dependent adenylation of salicylate and 6-methylsalicylate in cahuitamycin assembly is an ATP-dependent process that leads to the release of pyrophosphate (PPi). Abbreviations: C, condensation domain; Cy, cyclization domain; E, epimerization domain T, thiolation domain.

Initial bioinformatics studies revealed that CahJ has highest overall amino acid sequence similarity to a single non-redundant salicylate-AMP ligase (WP_093824147) from *Streptomyces* sp. SolWspMP-5a-2. Further analysis using SMART[4] showed CahJ has an AMP-binding domain encompassing amino acid residues 31-437. The closest homolog of known structure is DhbE (58% identity, PDB ID: 1MD9) [5], which activates 2,3-dihydroxybenozic acid (DHB) (ESI Fig. S2).

Although A domains are generally known to be selective for a specific substrate, many also have the ability to catalyse adenylation across a range of structurally related molecules [6]. In an effort to investigate CahJ specificity toward acyl-substrates, we employed a nonradioactive high-throughput malachite green colorimetric assay (ESI Fig. S3)[7]. This indicated that CahJ possesses an innate ability to catalyse activation of both salicylic acid (SA) and 6-methyl salicylic acid (6-MSA) for loading onto the N-terminal CahA aryl carrier protein (ArCP). Using this system, apparent steady-state kinetic parameters were determined and the data for SA and 6-MSA were fit using the Michaelis-Menten equation (Fig. 2). SA had a apparent K_m of 3.5 ± 0.3 μM and k_{cat} of 0.107 ± 0.002 min^{-1}, while 6-MSA had a similar but slightly lower K_m (1.6 ± 0.2 μM) and k_{cat} (0.079 ± 0.002 min^{-1}) values. We found that the efficiency of CahJ in this assay was ~1000 fold less than that of its closest structural and functional homologs[8]. This may be due to the slow release of the adenylated product, as evidenced by the persistent co-purification of CahJ with bound salicyl adenylate identified during crystallographic studies. Ultimately, ligand-free CahJ was obtained by partial denaturation with urea. *E. coli* does not encode a salicylate synthase based on genome annotation, and the source of the salicylate was likely the bacterial culture medium. This also explains failed attempts at kinetic analysis of ligand-free CahJ by the colorimetric assay, as after a single adenylation, the enzyme would remain in the adenylated intermediate form. Turnover in the malachite green assay coupled to pyrophosphatase mimics the forward enzymatic

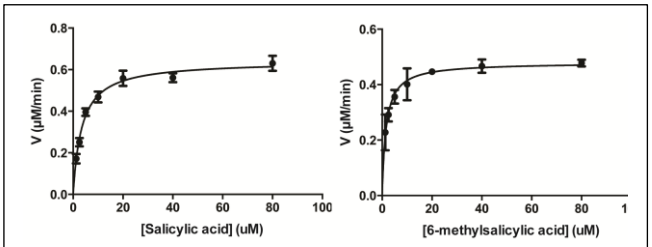

Figure 2. Determination of kinetic parameters for CahJ. Steady-state kinetic data fit using the Michaelis-Menten equation to determine kinetic constants of CahJ for (A) SA and (B) 6-MSA.

reaction, but this requires the dissociation of the adenylate intermediate, which, as expected, was very slow in absence of ArCP. This is consistent with the ArCP loading function of CahJ, which would require the enzyme bind to the high energy adenylated intermediate until an ArCP is available.

Our previous observation that CahJ activates both 6-MSA and SA[1] (ESI Fig S1), and to explore CahJ substrate specificity, we selected a group of 28 structurally related benzoic acid compounds to test for activity. All data in this study was normalized relative to SA, with 6-MSA exhibiting 73% activity relative to SA. Additional methylated SA derivatives, such as 4-

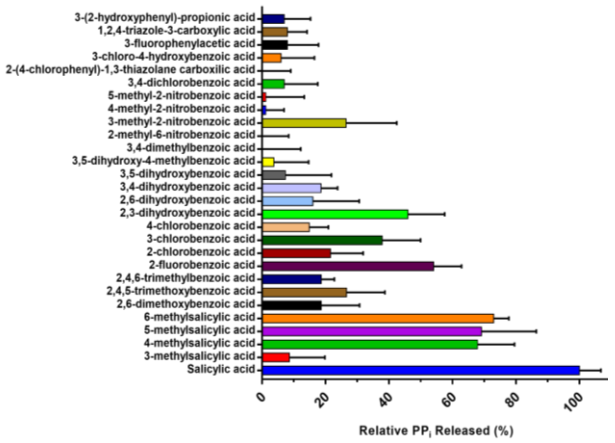

Figure 3. Determination of CahJ substrate scope toward diverse benzoic acid substrates. Relative substrate specificity of CahJ for 28 benzoic acid analogs is represented by the bars along with their respective s.d. The activity obtained from the reaction with SA is defined as 100%.

methyl-SA (4-MSA) and 5-methyl-SA (5-MSA), showed similar activity by CahJ with the exception of 3-methyl-SA (Fig. 3). However, we found that CahJ lacked activity when the hydroxyl group at the C2 position of salicylic acid was replaced with nitro or acetyl functional groups, as shown by 2-nitrobenzoic acid, acetyl salicylic acid, 4-methyl-2-nitrobenzoic acid, 5-methyl-2-nitrobenzoic acid, and 2-methyl-6-nitrobenzoic acid (Fig. 3). Only 3-methyl-2-nitrobenzoic acid served as a CahJ substrate, albeit with relatively low (26%) conversion (Fig. 3). Interestingly, halogenated benzoic acid substrates with chlorine/fluorine substituted for the 2-hydroxyl group and 2,3-dihydroxybenzoic acid (2,3-DHB) displayed appreciable reaction turnover. By contrast, CahJ showed no significant activity against substrates with extended ring systems (2-(4-chlorophenyl)-1,3-thiazolidine-4-carboxylic acid, 3-(2-hydroxyphenyl) propionic acid, etc.) or five membered ring structures (1,2,4-triazole-3-carboxylic acid). These observations can be explained by the limited size of the substrate-binding pocket revealed by structural studies (see below) (Fig. 4).

The ability of CahJ to transfer SA and 6-MSA onto its natural substrate, the CahA ArCP, was also evaluated using intact-protein mass spectrometry analysis. CahJ effectively catalysed loading of the CahA ArCP with both SA and 6-MSA (ESI Fig. S4-5). Our substrate scope study indicated that effective aryl transfer is limited to 5-MSA, 4-MSA, 2-fluorobenzoic acid and 2,3-DHB in addition to the natural SA and 6-MSA substrates. Surprisingly, CahJ transferred all substrates tested to the ArCP, including 3-methylsalicylate (3-MSA), which was a poor substrate in the malachite green assay (Fig. 3; ESI Fig. S4-5). Thus, while the malachite green assay discriminates substrate preferences, the results of the ArCP-dependent assay suggest that CahJ has the capacity to act *in vivo* on an even broader range of substrates.

The apparently greater activity with the natural ArCP acceptor is consistent with the high affinity of CahJ for aryl adenylate intermediates. Reaction with ArCP breaks the adenylate phosphoester bond by phosphoester-thioester exchange, releasing AMP and aryl-ArCP from the enzyme, whereas turnover in the malachite green assay requires dissociation of the acyl-adenylate. Taken together the assay results and purification behaviour indicate that both the nucleotide and aryl moieties of the adenylate contribute to high affinity binding by the enzyme.

To further expand our understanding of CahJ substrate selectivity, and its role in diversifying metabolites produced by the cahuitamycin pathway, crystal structures of CahJ as substrate complexes were solved (ESI Table S1). The CahJ structure is similar to those of other members of the NRPS A domain subfamily [9], particularly those that act on benzoic acid derivatives [5, 10]. However, the CahJ structure is the first for which salicylic acid is a natural substrate. The CahJ protein folds into two distinct domains, an N-terminal domain (amino acids 1-429), which contains the substrate binding site, and a smaller compact C-terminal domain (430-544) (Fig. 4A). The most substantial difference observed between the structure of CahJ and other related enzymes occur at residues 137-184, a region of the N-terminal domain positioned ~20 Å from the designated active site and, therefore, is unlikely to impact the relative activity of the compared enzymes. The compact C-terminal domain is comprised of five β-strands and three α-helices (Fig. 4A/ ESI Fig. S2). A wide cleft resides between the C-terminal lid and the N-terminal domain, which are connected by only a short hinge devoid of regular secondary structure. ATP binds in the cleft at

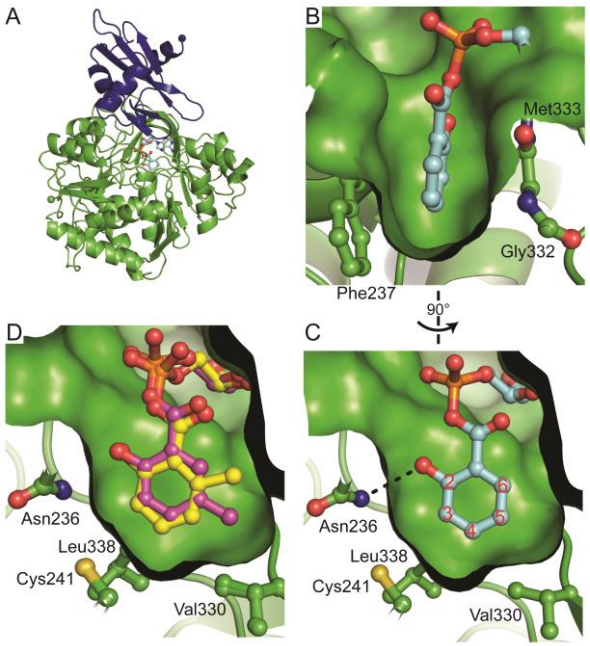

Figure 4. CahJ structure and substrate binding site. A) The overall structure of CahJ with the N-terminal domain in green and the C-terminal domain in blue. The N- and C-termini are shown as spheres (PDB ID 5WM3). The bound SA adenylate is shown in cyan ball and stick. B) Side view of the flat, hydrophobic binding site for the substrate aromatic ring between Phe237 and the Gly332-Met333 peptide. The orientations of Phe237 and bound salicyl adenylate enables π-stacking interactions. The protein surface is shown in green with the SA adenylate substrate in cyan. C) Face-on view of the substrate binding site. The 2-hydroxy group forms a hydrogen bond with Asn236 (3.0-3.2 Å among the three structures). Pockets for the binding of additional substituents are clear at the 4, 5, and 6 positions with Val330 creating a separation of the 4 and 5 position pockets. The 3-position pocket is restricted by Cys241 and Leu338. D) Overlay of 5-MSA and 6-MSA adenylates in the CahJ active site (PDB IDs 5MW5 and 5MW4, respectively). When overlaid, the rotation of 6-MSA relative to 5-MSA (and other substrates) is evident. The observed position of 6-MSA is incompatible with a methyl substituent at the 5-position.

the interface of the N- and C-terminal domains (Fig. 4A). The C-terminal lid domain of related enzymes is known to move during the two-step reaction, adopting one conformation for the adenylation reaction and another for the aryl transfer reaction[11]. The C-terminal domain of CahJ occupies the aryl transfer conformation in the structures.

With respect to substrate processing, CahJ binds aryl substrates in a flat, hydrophobic site between the Phe237 side chain and the Gly332-Met333 peptide. The site is ideally shaped to accommodate aromatic rings with Phe237 forming offset π-stacking interactions with the substrate aromatic ring (Fig. 4B). The narrow shape of the binding site dictates that only aromatic, and therefore flat, substrates can be accommodated. The periphery of the binding site can be defined as a number of pockets corresponding to the positions of the aromatic ring (Fig. 4C, ESI Fig. S6).

The 2-position pocket is formed by Asn236, which forms a hydrogen bond with the salicylate 2-hydroxyl. Asn236 is invariant in SA and 2,3-DHB-utilizing enzymes and forms the only hydrogen bond between aryl substrates and CahJ. Lack of a 2-hydroxyl group decreased activity, for example with 2-fluorobenzoic acid (Fig. 3). The 3-position pocket is defined by the side chains of Cys241 and Leu338. These side chains form a steric block that restricts the binding of substrates bearing a substituent at the 3-position, consistent with the low activity of 3-MSA in the malachite green assay.

The 4-position pocket lies between Cys241 and Leu338 and Val330, while the 5-position pocket is located between Val330 and Gly307. The 4- and 5-position pockets are both hydrophobic and large enough to accommodate a methyl group, or similarly sized substituents. In the 5-MSA adenylate structure, the methyl fits snugly into the 5-position pocket (Fig. 4D). In addition, substrates with substituents at either the 4- or 5-position had generally high activity in the malachite green assay. Thus, it is anticipated that the Cl substituent of 3-chlorobenzoic acid does not bind in the restricted 3-position, but rather occupies the 5-position.

The CahJ active site 6-position pocket is bordered by invariant Gly307 and Gly308, and is not large enough to accommodate a methyl group without a slight rotation (~5°) of the aromatic ring, as observed in the 6-MSA adenylate structure (Fig. 4D). This rotation places the 6-methyl substituent close to the 5-position pocket, and would create a steric clash with Val330 if a 5-methyl substituent were present simultaneously, indicating a likely mutually exclusive relationship between the two sites. A similar relationship appears to exist between the 4- and 5-positions, in which CahJ would accommodate a C-4 substituent by a slight rotation of the aryl ring. This would result in the binding site poorly accommodating a C-5 substituent.

The CahJ active site is virtually identical to those of adenylating enzyme/domain family members that act on 2,3-dihydroxybenzoic acid (DHB), including DhbE[5, 12], BasE[10a] and EntE[10b, 13]. The active sites of all four proteins contain pockets at the 4-, 5-, and 6-positions of the aromatic ring. Other enzymes in the SA/DHB family are expected to possess similar substrate flexibility, but only CahJ has been interrogated with a wide panel of substrates. EntE possessed substrate flexibility in tests with two non-natural substrates, but neither contained methyl substituents[14].

We next decided to address whether the unnatural substrates identified in our *in vitro* assays would also serve as substrates in the intact cahuitamycin pathway. Therefore, we introduced an unnatural substrate 4-MSA exogenously to the Δ*cahI Streptomyces gandocaensis* strain, which was incorporated to produce a new analog, cahuitamycin F (**1**; Fig. 5). This new metabolite was isolated by reverse-phase HPLC (RPHPLC) and the HRESIMS $[M+H]^+$ ion peak at *m/z* 650.2706 provided a molecular formula of $C_{28}H_{39}N_7O_{11}$ (ESI Fig. S7), requiring 13° of unsaturation. Extensive 1D and 2D NMR data were acquired for **1**, which indicated the expected structural similarity with cahuitamycin C (**2**) (ESI Fig. S8), including eight methyl/methane carbons, ten methylene carbons and nine carbonyls/quaternary carbons, similar to the carbon backbone of reported cahuitamycins with a clear difference at the phenyl ring system. In brief, analysis of gCOSY, gHSQCAD and gHMBCAD cross peaks at δ_H 6.77 and 7.58 to δ_C 146.5 and 160.1 suggested the spin system consisting of an *ortho*- substituted phenol group. The presence of a singlet at δ_H 6.82 (H-26) with HMBC correlation to δ_C 21.3 (C-26) and δ_C 121.4 (C-24) suggested methylation at C-25 consistent with the hypothesized incorporation of 4-methylsalicylic acid by the Δ*cahI* strain of *S. gandocaensis*. In addition, correlations observed through long-range ^1H–^{13}C between δ_H 5.10 (H-19) and δ_C 168.7 (C-21) as well as ^1H–^1H between δ_H 5.10 (H-19) and 4.68 (H-20b) interactions indicated the moiety to be a *N*-terminal 2-phenyl-oxazoline group. Further analysis of the 2D NMR spectra indicated at least four more spin systems consisting of a serine (Ser), two modified ornithines (Orns) and a modified alanine (Ala) (ESI Fig. S9-13). The modified Orn was defined as N^{δ}-hydroxy-N^{δ}-formylornithine (*N*-OH-*N*-fOrn) based on the COSY correlations observed from δ_H 4.30 (H-10) to 3.45 (H-13) and a gHMBCAD correlation between H-13 and C-14

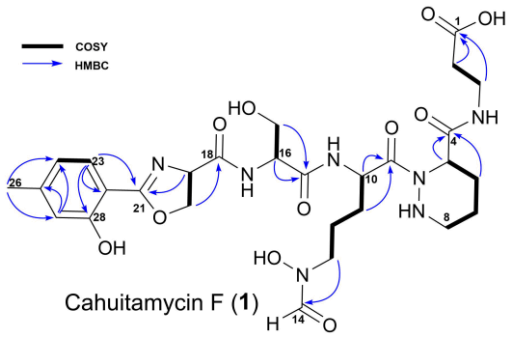

Figure 5. Structure of cahuitamycin F (**1**) with absolute stereochemistry. Cahuitamycins F structure with key HMBC and COSY correlations.

(δ_C 163.1). Similarly, the piperazic acid (Pip) moiety was deduced based on ^1H-^1H relay from δ_H 3.57, 3.62 (H-8) to 4.41 (H-5) (Fig. 6). The C terminus of the peptide was identified as β-alanine (β -Ala) based on COSY correlation between H-3 (δ_H 3.37, 3.51) to H-2 (δ_H 2.37, 2.41) and an HMBC correlation from H-3 to C-1 (δ_C 174.1). All deduced moieties completed the planar structure of **1** (Fig. 5; ESI Table S2; ESI Fig. S9-13). The absolute stereochemistry of cahuitamycin F was confirmed to be L-Ser, L-Ser, D-Pip and D-*N*-OH-Orn, respectively, and was in agreement with earlier reported cahuitamycins[1].

Cahuitamycin F (**1**) was next tested for its ability to inhibit biofilm formation of *A. baumannii* using crystal violet based static biofilm assay followed by optical density measurements. The assay was conducted using cahuitamycin A (**3**) as a positive control and the result from this assay showed that **1** is able to

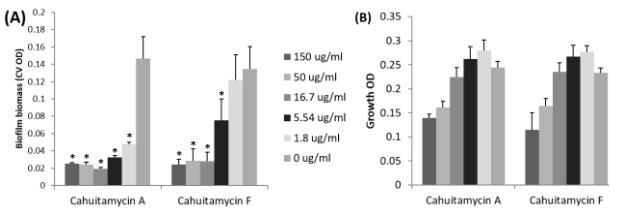

Figure 6. Biological activity of cahuitamycins A (3) and F (1). A) Inhibition of biofilm formation and (B) growth of *A. baumannii* in the presence of cahuitamycins A and F. Results are the average of three replicates ±s.d. Student's t-test was used for statistical analysis. *P<0.001 compared with control (no addition of compounds).

inhibit biofilm formation (Fig 6A). The calculated half-maximal inhibitory concentration (IC_{50}) value for **1** was 18.3 µM, which is similar to the IC_{50} of **3** (15.6 µM) against *A. baumannii* biofilm formation. Cahuitamycin F (**1**) showed negligible impact on the growth of *A. baumannii* (Fig. 6B), consistent with the activity of other cahuitamycins. These data suggest that the terminal 2-hydroxybenzoyl-oxazoline group represents a key pharmacophore of the cahuitamycins. This insight is important for future medicinal chemistry efforts focused on increasing efficacy in this new class of natural product biofilm inhibitors.

This study describes the structural and biochemical characterization of CahJ, a promiscuous NRPS adenylating enzyme from *Streptomyces gandocaensis.* CahJ natively selects salicylic acid and 6-methylsalicylic acid as starter units for the cahuitamycin biosynthetic pathway. First, CahJ steady-state kinetic parameters were determined for SA and 6-MSA. We then demonstrated that CahJ is capable of activating a range of benzoic acid derivatives and transferring them to the N-terminal ArCP domain of CahA. Crystal structures of CahJ complexed with both natural and unnatural substrates have provided new insights toward substrate flexibility and development of a structure-based rationale for enzyme flexibility. The combined biochemical and structural studies were employed to guide efforts to generate new cahuitamycin congeners, leading to the generation of, cahuitamycin F, via in vivo mutasynthetic diversification. This study also provides an effective roadmap for future protein engineering to alter CahJ substrate selectivity for generation of new cahuitamycin leads for further development.

Acknowledgements

This work was supported by the National Institutes of Health (NIH) grant R35 GM118101, and the Hans W. Vahlteich Professorship (D.H.S.). The contents of this manuscript are that of the authors and do not necessarily represent the views of the NIH.

There are no conflicts to declare.

Keywords: Natural Product • Biosynthesis • X-ray Protein Structure • Adenylation Domain • Kinetics

[1] S. R. Park, A. Tripathi, J. Wu, P. J. Schultz, I. Yim, T. J. McQuade, F. Yu, C.-J. Arevang, A. Y. Mensah, G. Tamayo-Castillo, *Nat. Commun.* **2016**, *7*.

[2] H. W. Boucher, G. H. Talbot, J. S. Bradley, J. E. Edwards, D. Gilbert, L. B. Rice, M. Scheld, B. Spellberg, J. Bartlett, *Clin. Infect. Dis.* **2009**, *48*, 1-12.

[3] D. J. Newman, G. M. Cragg, *J. Nat. Prod.* **2016**, *79*, 629-661.

[4] I. Letunic, T. Doerks, P. Bork, *Nucleic Acids Res.* **2015**, *43*, D257-260.

[5] J. J. May, N. Kessler, M. A. Marahiel, M. T. Stubbs, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12120-12125.

[6] aS. Garneau, P. C. Dorrestein, N. L. Kelleher, C. T. Walsh, *Biochemistry* **2005**, *44*, 2770-2780; bG.-L. Tang, Y.-Q. Cheng, B. Shen, *J. Biol. Chem.* **2007**, *282*, 20273-20282; cJ. Neres, D. J. Wilson, L. Celia, B. J. Beck, C. C. Aldrich, *Biochemistry* **2008**, *47*, 11735-11749.

[7] T. J. McQuade, A. D. Shallop, A. Sheoran, J. E. Delproposto, O. V. Tsodikov, S. Garneau-Tsodikova, *Anal. Biochem.* **2009**, *386*, 244-250.

[8] aA. M. Gehring, I. I. Mori, R. D. Perry, C. T. Walsh, *Biochemistry* **1998**, *37*, 17104; bL. E. Quadri, T. A. Keating, H. M. Patel, C. T. Walsh, *Biochemistry* **1999**, *38*, 14941-14954; cL. E. Quadri, J. Sello, T. A. Keating, P. H. Weinreb, C. T. Walsh, *Chem. Biol.* **1998**, *5*, 631-645.

[9] E. Conti, T. Stachelhaus, M. A. Marahiel, P. Brick, *EMBO J.* **1997**, *16*, 4174-4183.

[10] aE. J. Drake, B. P. Duckworth, J. Neres, C. C. Aldrich, A. M. Gulick, *Biochemistry* **2010**, *49*, 9292-9305; bJ. A. Sundlov, A. M. Gulick, *Act Cryst.* **2013**, *69*, 1482-1492.

[11] A. M. Gulick, *ACS Chem. Biol.* **2009**, *4*, 811-827.

[12] K. Zhang, K. M. Nelson, K. Bhuripanyo, K. D. Grimes, B. Zhao, C. C. Aldrich, J. Yin, *Chem. Biol.* **2013**, *20*, 92-101.

[13] J. A. Sundlov, C. Shi, D. J. Wilson, C. C. Aldrich, A. M. Gulick, *Chem. Biol.* **2012**, *19*, 188-198.

[14] F. Rusnak, W. S. Faraci, C. T. Walsh, *Biochemistry* **1989**, *28*, 6827-6835.