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

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

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

Cahuitamycins, produced by *Streptomyces gandocaensis*, are a structural class of biofilm formation inhibitors 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 respon-

sible for a large number of nosocomial infections, including pneumonia, urinary tract infections, wound infections, and bacteremia, with significantly high mortality rates ($\approx$60%). Biofilm formation contributes to the high rate of antimicrobial resistance (AMR).[2] When in a biofilm, these microbes develop AMR up to 1000 times greater than that of 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 cahuitamycins' biosynthetic mechanisms and their key enzyme functions is a critical step towards expanding structural diversity by using pathway engineering. As shown recently, the central role of starter unit selection in cahuitamycin diversification and biological activity involves CahJ (Scheme 1 and Figure S1 in the Supporting Information).[1] We were motivated to explore this essential adenylation (A) enzyme whose further manipulation could provide access to new cahuitamycin congeners.

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

Although A domains are generally known to be selective for a specific substrate, many also have the ability to catalyze adenylation across a range of structurally related molecules.[6] To investigate CahJ specificity toward acyl substrates, we employed a nonradioactive high-throughput malachite green colorimetric assay (Figure S3).[7] This indicated that CahJ possesses an innate ability to catalyze the activation of both salicylic acid (SA) and 6-methyl salicylic acid (6-MSA) for loading onto the N-terminal CahA aryl carrier protein (ArCP). Apparent steady-state kinetic parameters were determined by using this system, and the data for SA and 6-MSA were fit by using the Michaelis–Menten equation (Figure 1). SA had an apparent K_m of (3.5 ± 0.3) μM and a k_{cat} of (0.107 ± 0.002) min^{-1}, whereas 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 approximately 1000 times less than that of its closest structural and functional homologues.[8] This might be due to the slow release of the adenylated product, as evidenced by the persistent copurification of CahJ with bound salicyl adenylate identified during crystallographic studies. Ulti-

[a] Dr. A. Tripathi, Dr. S. R. Park, Dr. A. P. Sikkema, B. Lee, Prof. J. L. Smith, Prof. D. H. Sherman
Life Sciences Institute, University of Michigan
210 Washtenaw Avenue, Ann Arbor, MI 48109-2216 (USA)
E-mail: davidhs@umich.edu

[b] Dr. A. P. Sikkema, Prof. J. L. Smith
Department of Biological Chemistry, University of Michigan
1150 W. Medical Center Drive, Ann Arbor, MI 48109 (USA)

[c] Dr. J. Wu, Dr. C. Xi
Department of Environmental Health Sciences
University of Michigan School of Public Health
Ann Arbor, MI 48109 (USA)

[d] Dr. H. J. Cho
Department of Pathology, University of Michigan
Ann Arbor, MI 48109 (USA)

[e] Dr. A. Tripathi, Prof. D. H. Sherman
Department of Medicinal Chemistry, University of Michigan
Ann Arbor, MI 48109 (USA)

[f] Prof. D. H. Sherman
Department of Chemistry, University of Michigan
Ann Arbor, MI 48109 (USA)

[g] Prof. D. H. Sherman
Department of Microbiology and Immunology, University of Michigan,
Ann Arbor, MI 48109 (USA)

[h] Dr. S. R. Park
Baruch S. Blumberg Institute, Natural Products Discovery Institute
3805 Old Easton Road, Doylestown, PA 18902 (USA)

[i] Dr. A. P. Sikkema
Epigenetics and Stem Cell Biology Laboratory
National Institute of Environmental Health Sciences
National Institutes of Health
Research Triangle Park, NC 27709 (USA)

Supporting information and the ORCID identification numbers for the authors of this article can be found under https://doi.org/10.1002/cbic.201800233.
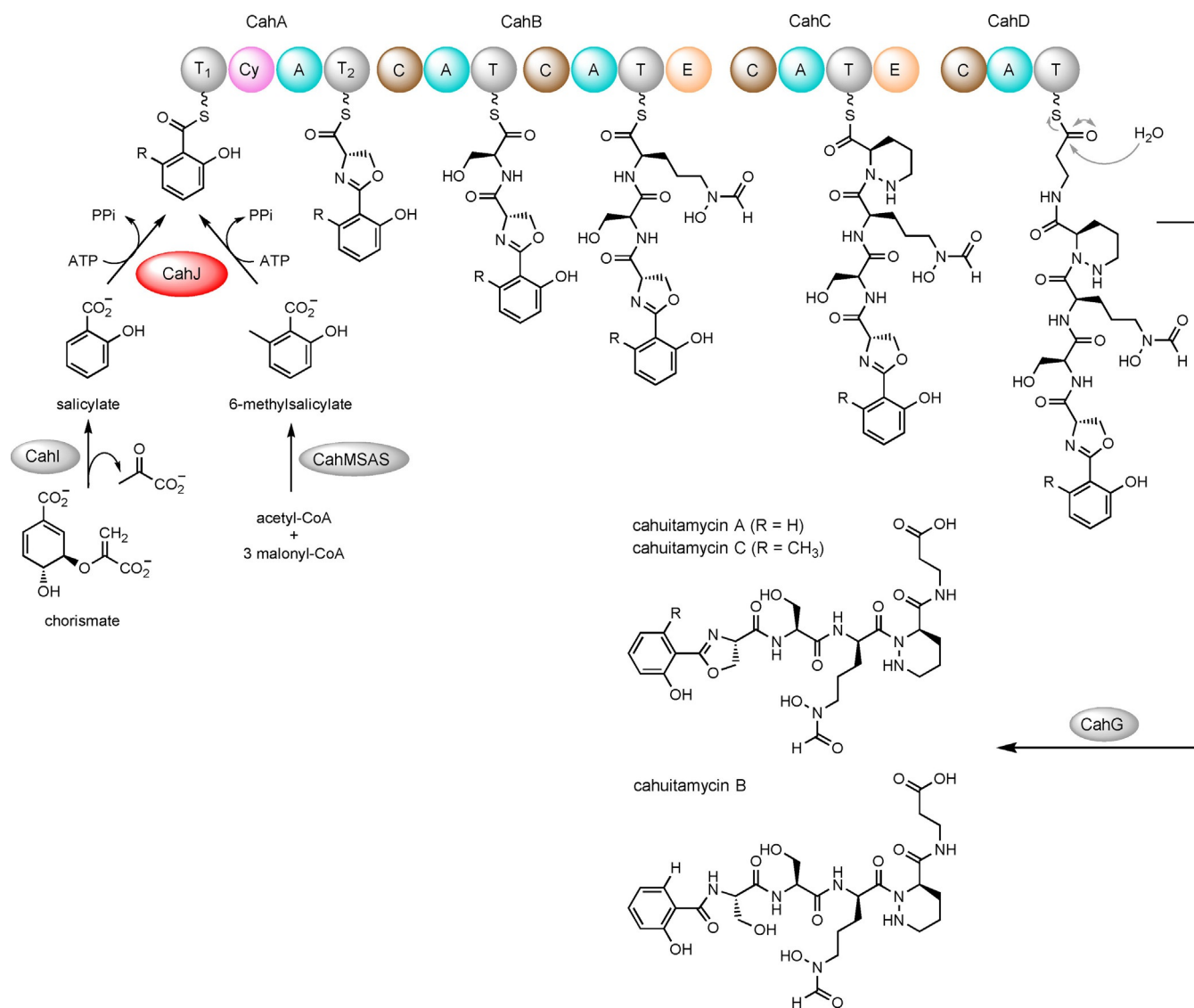
Scheme 1. Proposed biosynthesis of cahuitamycins A–C in *S. gandocaensis*. The biosynthetic gene cluster contains NRPS-encoding CahABCD along with putative genes involved in chain initiation (*cahI* and *cahMSAS*) and chain termination (*cahG*). The CahJ-dependent adenylation of salicylate and 6-methylsalicylate in cahuitamycin assembly is an ATP-dependent process that leads to the release of pyrophosphate (PPi). C: condensation domain, Cy: cyclization domain, E: epimerization domain, T: thiolation domain.

mately, ligand-free CahJ was obtained by partial denaturation with urea. Based on genome annotation, *Escherichia coli* does not encode a salicylate synthase, 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 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 was that CahJ activates both 6-MSA and SA[1] (Figure S1); to explore CahJ substrate specificity, we selected a group of 28 structurally related carboxylic acid derivatives to test for activity. All data in this study were normalized relative to SA, with 6-MSA exhibiting 73% activity relative to SA. Other methylated SA derivatives, such as 4-methyl-SA (4-MSA) and 5-methyl-SA (5-MSA), showed similar activity with CahJ, but the activity for 3-methyl-SA was much lower (Figure 2). We found that CahJ lacked activity when the hydroxy group at the C2 position of salicylic acid was replaced with nitro or acetyl functional groups, as shown by 2-nitrobenzoic acid, O-acetyl salicylic acid (not shown), 4-methyl-2-nitrobenzoic acid, 5-methyl-2-nitrobenzoic acid, and 2-methyl-6-nitrobenzoic acid (Figure 2). Only 3-methyl-2-nitrobenzoic acid served as a CahJ substrate, albeit with relatively low (26%) conversion. Interestingly, halogenated benzoic acid substrates with chlorine/fluorine substituted for the 2-hydroxy group and 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-
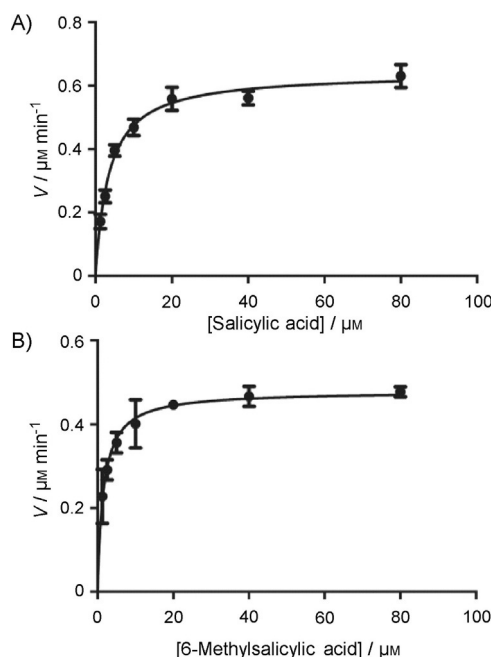
Figure 1. Determination of the kinetic parameters for CahJ. Steady-state kinetic data fit by using the Michaelis–Menten equation to determine the kinetic constants of CahJ for A) SA and B) 6-MSA.

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 Figure 3 below).

The ability of CahJ to transfer SA and 6-MSA onto its natural substrate, the CahA ArCP, was also evaluated by using intact-protein mass spectrometry analysis. CahJ effectively catalyzed loading of the CahA ArCP with both SA and 6-MSA (Figures S4 and S5). 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 (Figures 2, S4, and S5). Thus, although 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 behavior 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 (Table S1). The CahJ structure is similar to those of other members of the nonribosomal peptide synthetase (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; Figure 3 A). The most substantial difference observed between the structure of CahJ and other related enzymes occurs at residues 137–184, a region of the N-terminal domain positioned ≈20 Å from the designated active site and, therefore, unlikely to have an impact on the relative activity of the compared enzymes. The compact C-terminal domain comprises five β-strands and three α-helices (Figures 3 A and S2). There is a wide cleft between the C-terminal lid and the N-terminal

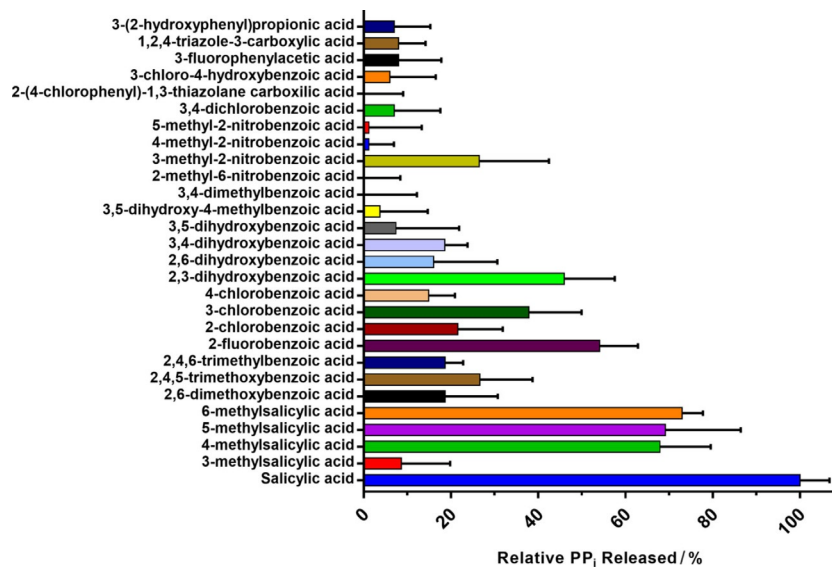
Figure 2. Determination of CahJ substrate scope. The relative specificity of CahJ towards 28 substrates is represented by the bars along with their respective s.d. The activity obtained from the reaction with SA is defined as 100 %.
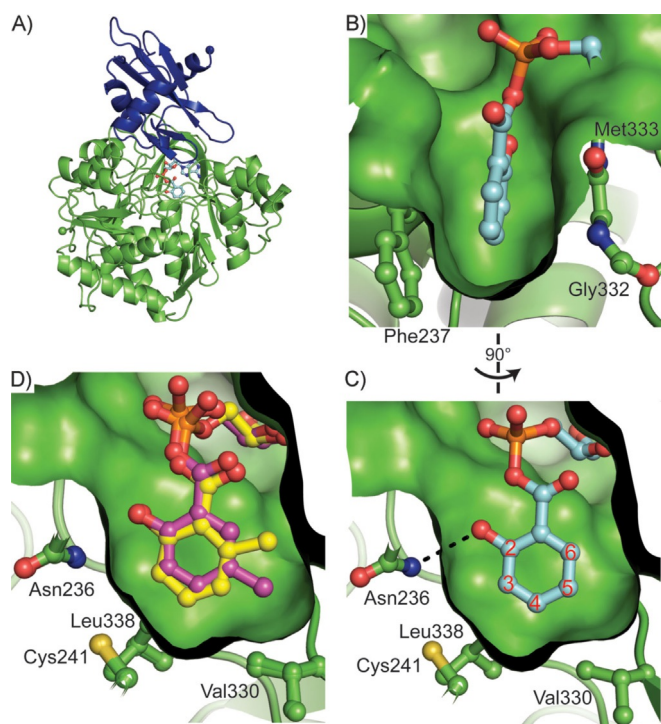
Figure 3. CahJ structure and substrate binding site. A) The overall structure of CahJ with the N-terminal domain in green and the C-terminal domain in blue. The N and C termini are shown as spheres (PDB ID: 5WM3). The bound SA adenylate is shown in cyan ball and stick. B) Side view of the flat, hydrophobic binding site for the substrate aromatic ring between Phe237 and the Gly332–Met333 peptide. The orientations of Phe237 and bound salicyl adenylate enable 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 of the substrate forms a hydrogen bond with Asn236 (3.0–3.2 Å among the three shown substrates SA, 5-MSA, and 6-MSA). 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: 5WM5 and 5WM4, 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.

domain, which are connected by only a short hinge devoid of regular secondary structure. ATP binds in the cleft at the interface of the N- and C-terminal domains (Figure 3 A). 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 (Figure 3 B). 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 (Figures 3 C and S6).

The 2-position pocket is formed by Asn236, which forms a hydrogen bond with the salicylate 2-hydroxy group. 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-hydroxy group decreased activity, for example for 2-fluorobenzoic acid (Figure 2). 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/Leu338 and Val330; 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 substituent. In the 5-MSA adenylate structure, the methyl fits snugly into the 5-position pocket (Figure 3 D). 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 ($\approx 5°$) of the aromatic ring, as observed in the 6-MSA adenylate structure (Figure 3 D). 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, thus 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-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 unnatural 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. We introduced an unnatural substrate, 4-MSA, exogenously to the Δcahl S. gandocaensis strain, which was incorporated to produce a new analogue, cahuitamycin F (**1**; Scheme 2). This new metabolite was isolated by reversed-phase (RP) HPLC, and the HRMS(ESI) $[M+H]^+$ ion peak at m/z 650.2706 provided a molecular formula of $C_{28}H_{39}N_7O_{11}$ (Figure S7), requiring 13 degrees of unsaturation. Extensive 1D and 2D NMR data were acquired for **1**, which indicated the expected structural similarity with cahuitamycin C (**2**; Figure S8), including eight methyl/methane carbons, ten methylene carbons and nine carbonyls/quaternary carbons, similar to the carbon backbone of reported cahuita-

Scheme 2. Structure and absolute stereochemistry of cahuitamycin F (**1**) with key HMBC and COSY correlations.

mycins with a clear difference at the phenyl ring system. Analysis of gCOSY, gHSQCAD and gHMBCAD crosspeaks at $\delta_H = 6.77$ and 7.58 to $\delta_C = 146.5$ and 160.1 suggested a spin system consisting of an *ortho*-substituted phenol group. The presence of a singlet at $\delta_H = 6.82$ (H-26) with HMBC correlation to $\delta_C = 21.3$ (C-26) and $\delta_C = 121.4$ (C-24) suggested methylation at C-25 consistent with the hypothesized incorporation of 4-methylsalicylic acid by the $\Delta cahI$ strain of *S. gandocaensis*. In addition, correlations observed through long-range $^1H-^{13}C$ interactions between $\delta_H = 5.10$ (H-19) and $\delta_C = 168.7$ (C-21) as well as $^1H-^1H$ between $\delta_H = 5.10$ (H-19) and 4.68 (H-20b) indicated the moiety to be a N-terminal 2-phenyloxazoline group. Further analysis of the 2D NMR spectra indicated at least four more spin systems consisting of a serine, two modified ornithines (Orns) and a modified alanine (Figures S9–S13). The modified Orn was defined as N^δ-hydroxy-N^δ-formylornithine (*N*-OH-*N*-fOrn) based on the COSY correlations observed from $\delta_H = 4.30$ (H-10) to 3.45 (H-13) and a gHMBCAD correlation between H-13 and C-14 ($\delta_C = 163.1$). Similarly, the piperazic acid (Pip) moiety was deduced based on $^1H-^1H$ relay from $\delta_H = 3.57$, 3.62 (H-8) through 4.41 (H-5; Scheme 2). The C terminus of the peptide was identified as β-alanine (β-Ala) based on COSY correlation between H-3 ($\delta_H = 3.37$, 3.51) to H-2 ($\delta_H = 2.37$, 2.41) and an HMBC correlation from H-3 to C-1 ($\delta_C = 174.1$). All deduced moieties completed the planar structure of **1** (Scheme 2 and Figures S9–S13, Table S2). The absolute stereochemistry of cahuitamycin F was confirmed to be L-Ser, L-Ser, D-Pip and D-*N*-OH-Orn, and was in agreement with the stereochemistries of earlier reported cahuitamycins.[1]

Cahuitamycin F (**1**) was next tested for its ability to inhibit biofilm formation of *A. baumannii* by using a Crystal-Violet-based static biofilm assay followed by optical density measurements. The assay was conducted with cahuitamycin A (**3**) as a positive control, and the result showed that **1** is able to inhibit biofilm formation (Figure 4 A). 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**) made a negligible impact on the growth of *A. baumannii* (Figure 4 B); this is 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.

Figure 4. A) Inhibition of biofilm formation and B) growth of *A. baumannii* in the presence of cahuitamycins A (**3**) and F (**1**). 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).

This study describes the structural and biochemical characterization of CahJ, a promiscuous NRPS adenylating enzyme from *S. 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 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 developing 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 by in vivo mutasynthetic diversification. This study also provides an effective roadmap for future protein engineering to alter CahJ substrate selectivity to generate new cahuitamycin leads for further development.

[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*, 10710.

[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–D260.

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

[6] a) S. Garneau, P. C. Dorrestein, N. L. Kelleher, C. T. Walsh, *Biochemistry* **2005**, *44*, 2770–2780; b) G.-L. Tang, Y.-Q. Cheng, B. Shen, *J. Biol. Chem.* **2007**, *282*, 20273–20282; c) J. 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] a) A. M. Gehring, I. I. Mori, R. D. Perry, C. T. Walsh, *Biochemistry* **1998**, *37*, 17104; b) L. E. Quadri, T. A. Keating, H. M. Patel, C. T. Walsh, *Biochemistry* **1999**, *38*, 14941–14954; c) L. 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] a) E. J. Drake, B. P. Duckworth, J. Neres, C. C. Aldrich, A. M. Gulick, *Biochemistry* **2010**, *49*, 9292–9305; b) J. A. Sundlov, A. M. Gulick, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **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.