JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

COMMUNICATION

pubs.acs.org/JACS

Imaging Mass Spectrometry and Genome Mining via Short Sequence Tagging Identified the Anti-Infective Agent Arylomycin in *Streptomyces roseosporus*

Wei-Ting Liu,[†] Roland D. Kersten,[‡] Yu-Liang Yang,[§] Bradley S. Moore,[‡,§] and Pieter C. Dorrestein[*,†,‡,§]

[†]Department of Chemistry and Biochemistry, [‡]Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, [§]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California at San Diego, La Jolla, California 92093, Unites States

Supporting Information

ABSTRACT: Here, we described the discovery of anti-infective agent arylomycin and its biosynthetic gene cluster in an industrial daptomycin producing strain *Streptomyces roseosporus*. This was accomplished via the use of MALDI imaging mass spectrometry (IMS) along with peptidogenomic approach in which we have expanded to short sequence tagging (SST) described herein. Using IMS, we observed that prior to the production of daptomycin, a cluster of ions (**1**−**3**) was produced by *S. roseosporus* and correlated well with the decreased staphylococcal cell growth. With a further adopted SST peptidogenomics approach, which relies on the generation of sequence tags from tandem mass spectrometric data and query against genomes to identify the biosynthetic genes, we were able to identify these three molecules (**1**−**3**) to arylomycins, a class of broad-spectrum antibiotics that target type I signal peptidase. The gene cluster was then identified. This highlights the strength of IMS and MS guided genome mining approaches in effectively bridging the gap between phenotypes, chemotypes, and genotypes.

Natural products that are made by non-ribosomal peptide synthetases (NRPS) have an unrivaled track record as anti-infective agents in the clinic.[1,2] Penicillin, vancomycin, and daptomycin are examples of antibiotics that are NRPS-derived[3−6] (Figure 1). With the emergence of antibiotic-resistant microbes, there is a great interest in molecules that target drug resistant microbes.[7,8] However, the last broad-spectrum antibiotic introduced on the market was over 50 years ago.

Our laboratory has been interested in the development of mass spectrometric methodologies that interconnect phenotypes, chemotypes, and genotypes. A part of the motivation for these tools is to not only discover new biology, but also apply these tools to the discovery of antimicrobials. Here, we report the use of imaging mass spectrometry in combination with a short sequence tagging (SST)-based genome mining approach that connects phenotypes and chemotypes with genotypes. We applied this approach to the discovery of the arylomycins (**1**−**3**, Figure 1) and their biosynthetic pathway in *Streptomyces roseosporus*.

To connect phenotypes with chemotypes, we have recently developed methods to investigate microbial metabolic interactions via imaging mass spectrometry (IMS).[9−11] One of the goals is to enable the discovery of new therapeutic leads.

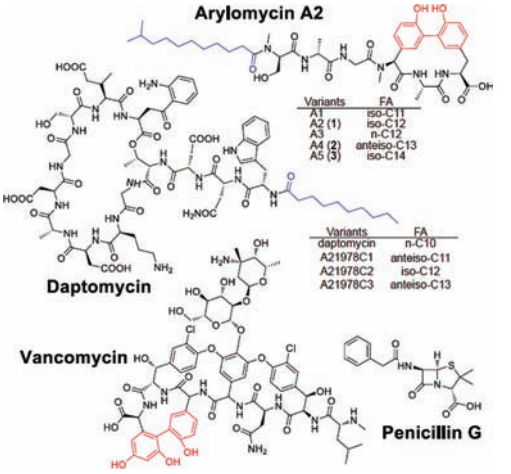


Figure 1. Structures of NRPS-derived compounds. Arylomycins and daptomycin have various components with different fatty acid (FA) (blue) chain lengths. Bis-aryl bridges are highlighted in red.

Herein, the pathogens *Staphylococcus aureus*[12] and *Staphylococcus epidermidis*[13] were co-cultured with *S. roseosporus* NRRL 15998, whose genome has been sequenced. This actinomycete produces daptomycin, an antibiotic used in clinic to treat Gram-positive bacterial infections.[4,6,14−17]

To demonstrate that IMS can be used to observe the molecules responsible for the inhibition of pathogens, we prepared lawns of *S. epidermidis* and *S. aureus* and then spotted *S. roseosporus* in the center (Figure 2, Figure S1). After 36 h incubation, inhibition zones were observed as expected in both staphylococcal lawns. Surprisingly, even though we determined that the IMS methodology can detect as little as 10 pmol of daptomycin, ions corresponding to daptomycin were not observed. Instead, a cluster of ions at m/z 863, 877, and 891, referred to as compounds **1**−**3** in this paper, were observed to localize at the zone of inhibition area. The absence of daptomycin-related signals in the zone-of-inhibition experiment suggested that *S. roseosporus* produced additional antibiotics.

A time course experiment of methanol extracts of *S. roseosporus* starter cultures revealed that compounds **1**−**3** were observed at 36 h (Figure S2), in agreement with the incubation time in the zone-of-inhibition experiment described above. Not

Received: May 3, 2011
Published: October 14, 2011

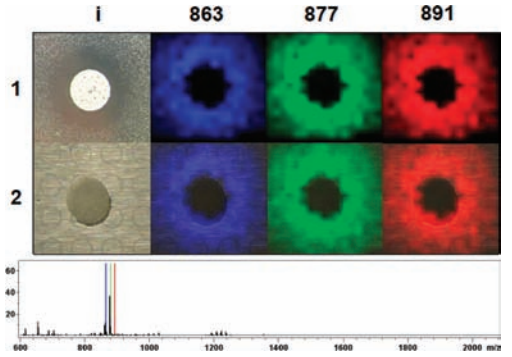


Figure 2. IMS of *S. roseosporus* spotted on top of a *S. epidermidis* lawn. (1) Ion distribution of compounds **1**−**3** (863, 877, 891) observed in IMS. 1i is a photograph showing *S. roseosporus* inhibition of *Staphylococci* growth. (2) Superimposition of the photograph with IMS data on top of MALDI target plate. Average mass spectrum of each IMS experiment was shown below IMS images with signals correlated to compounds **1**−**3** labeled with corresponding color as displayed in images.

until 48 h, the production of signals at *m*/*z* 1634.72, 1648.74, 1662.75, which correspond to daptomycin variants (A21978C1−3, Figure 1) was observed. That daptomycin is not produced until 48 h is consistent with the absence of daptomycin variants signals in the IMS data. MS-guided purification revealed that the molecules at *m*/*z* 863, 877, and 891 have monoisotopic masses of 825.439 (**1**), 839.455 (**2**), and 853.471 (**3**) Da, suggesting that the ion cluster observed in IMS exists as the potassium adduct. Compound **2** was purified and exhibited antibiotic activity against *S. epidermidis* with similar efficacy to daptomycin but milder activity toward *S. aureus*, in agreement with the smaller zone of clearing for the *S. aureus* observed in Figure 2 (Figures S1, S3).

To link to genotypes, we have recently developed a peptidogenomic mining approach to the discovery of peptidyl natural products.[18] The approach relies on the generation of peptide sequence tags from tandem mass spectrometric data to query genomes and to identify the biosynthetic genes. In turn, in an iterative fashion, the biosynthetic gene cluster supports the identification of a peptide as either a ribosomal or non-ribosomal product and facilitates the prediction of a (partial) structure. For ribosomally encoded peptides, a 5−6 consecutive amino acid residue sequence tag is often needed to successfully match to its precursor gene because of the larger proteomic search space. In this report, we show that for NRPS-derived peptides, this approach could be expanded to SST with only one or two amino acid residues to identify the candidate biosynthetic gene clusters as we suggested.[18] SST can be employed to carry out genome mining with molecules that are NRPS-derived. This is possible because the search tags can be more specific due to additional nonproteinogenic amino acids and the much smaller query space because of the small number (often <10) of NRPS gene clusters within a microbial genome. This scenario is similar to matching a peptide to a small database in a proteomics experiment where it also becomes possible to match to the correct peptide with minimal fragmentation data while much more fragmentation information is needed when a large database is used. Therefore, even with a very short sequence tag, we can still narrow down to the candidate biosynthetic gene cluster. The identification of candidate gene clusters, in turn, aids in the structural characterization of the molecule.

As proof-of-principle for SST, we first demonstrate how this works with daptomycin. The ions at *m*/*z* 1634.73, 1648.74, and

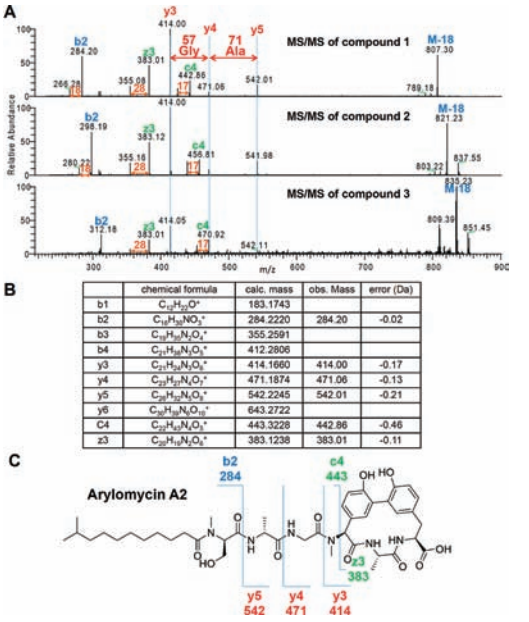
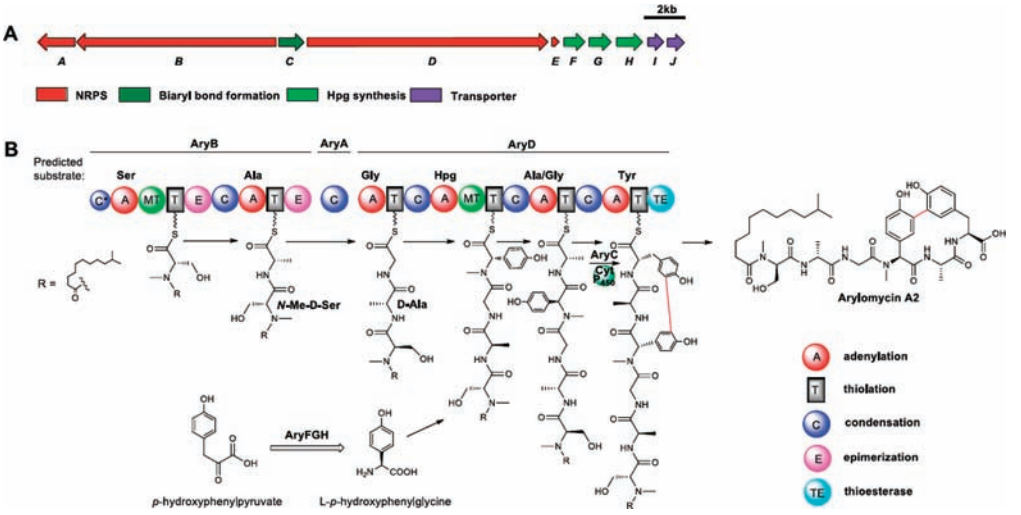


Figure 3. Correlating compounds **1**−**3** to arylomycins. (A) Alignment of IT MS/MS of compounds **1**−**3** revealed sequence tag Gly-Ala. (B) Annotated ion table corresponds to the IT MS/MS of compound **1**. (C) Ion map showing the fragmentation pattern of compound **1** correlates to arylomycin A2.

1662.75, corresponding to daptomycin variants, were subjected to tandem MS using collision-induced dissociation and resulted in fragment masses at *m*/*z* 1051.43, 1166.46, 1280.50 which suggested a sequence tag of Asp-Asn (Figure S4). Such a tag provides a minimal search unit that can be searched against all predicted NRPS biosynthetic pathways found on the *S. roseosporus* genome. All three tandem MS data sets of ions at *m*/*z* 1634.73, 1648.74, and 1662.75 resulted in an identical sequence tags (Figure S4). The combination of NP.searcher and NRPS predictor,[19,20] two programs designed to identify the amino acid specificity of non-ribosomal peptide synthetases, was utilized to predict all possible NRPS gene clusters and their amino acid codes in the *S. roseosporus* NRRL 15998 genome. Seven gene clusters that display NRPS features were found. Matching the sequence tag obtained from the tandem MS data of daptomycin variants against the NRPS predictor and NP.searcher predicted amino acids identified the daptomycin gene cluster. Therefore, the proof-of-principle experiment with daptomycin variants demonstrated that the correct gene cluster could be identified from the genome through the SST approach.

Next, we set out to identify the series of ions at *m*/*z* 825.439, 839.455, and 853.471 (**1**−**3**). These molecules are separated by 14.013 Da consistent with CH_2 mass shifts. Possible explanations to account for this 14 Da difference may arise from different length of fatty acid chain, amino acid substitution, or methylation. Each scenario is commonly found in NRPS biosynthetic pathways. Therefore, SST was employed to match these molecules to one of the remaining six NRPS gene clusters. To achieve *de novo* sequencing of peptides, it is often challenging to separate the ions belonging to y-ion series from the b-ion series. In this experiment however, all three ions were first subjected to low-resolution tandem MS, and the spectra were aligned (Figure 3A). This revealed a series of ions that displayed 14 Da mass shifts (shifting ions) and a series of ions that did not display the mass

Scheme 1. Arylomycin Biosynthetic Gene Cluster and Proposed Biosynthetic Pathway



shifts (nonshifting ions). We were able to retrieve a sequence tag from the nonshifting ions which displayed mass differences of 57 and 71 Da, suggesting glycine and alanine, respectively. There was only one predicted NRPS out of the remaining six NRPS gene clusters that were identified on the *S. roseosporus* genome that contained this sequence tag. That gene cluster is predicted to encode six amino acids, Ser, Ala/Gly, Gly, Hpg, Ala/Gly, and Tyr. Although high-resolution MS spectra could provide more unambiguous sequence tag, we show that SST works with high-resolution as well as low-resolution MS data. We consulted the NORINE database that contains greater than 1000 NRPS-derived molecules and enables users to input specific residue(s) to search for molecules that have specified structural units.[21] Searching the NORINE database for the Ser, Ala/Gly, Gly, Hpg, Ala/Gly, and Tyr tag resulted in one group of candidate molecules, the arylomycins.[22,23]

Next, the intact masses of arylomycins were compared and the fragmentation data were re-inspected (Figure 3B,C). The intact masses of the observed ions matched to the calculated masses of arylomycins within 0.5−2 ppm.[22] The analysis of the fragmentation of compound **1** revealed that the observed b, y, c, and z ions were within 1 ppm, in agreement with arylomycin A2 (Figure S5).

Arylomycins are an exciting set of biologically active molecules and a class of broad-spectrum antibiotics that target type I signal peptidases (SPase).[24−28] SPase is responsible for the cleavage of signal peptides from secreted protein and it is highly conserved among bacteria, located on the extracellular surface of the cytoplasmic membrane, and is essential for bacterial viability. Therefore, arylomycins have been suggested as promising therapeutic leads.[29,30] This promise has led to a 2010 start-up company, RQx Pharmaceuticals, that is aiming to develop arylomycin and its analogues for clinical use. Arylomycins were first discovered from *Streptomyces* sp. Tü 6075 isolated in the tropical rain forest at Cape Coast, Ghana.[22,23] Independently, using assays screened for novel signal peptidase I inhibitors, scientists at Eli Lilly & Company reported a similar group of compounds that share the same skeleton with arylomycins but with glycosylation on the hydroxyphenylglycine residue.[28] Although natural resistance has been reported,[31] arylomycins and their glycosylated congeners are effective against Gram-negative bacteria, such as *Helicobacter pylori*, *Yersinia pestis*, and Gram-positive bacteria *Streptococcus pneumonia*, *Streptococcus pyrogens*, *S. epidermidis* and *Staphylococcus hemolyticus* with MICs of 4−16 μg/mL.[22,23,28,31−34]

To verify that *S. roseosporus* NRRL 15998 produces arylomycins, the candidate arylomycin gene cluster was annotated (Scheme 1, Table S1) and analyzed for biosynthetic consistency with the proposed arylomycin product. As the candidate gene cluster in the NRRL 15998 strain contained a frameshift and sequencing gaps, we based our analysis on the complete, almost identical gene cluster sequence of *S. roseosporus* strain NRRL 11379, which also produces the same set of molecules (Figure S6). Arylomycins contain an *N*-acyl chain and six amino acids, Ser, Ala, Gly, Hpg, Ala, and Tyr. Ser1 and Hpg4 are N-methylated, Ser1 and Ala2 are in D-configuration, and finally Tyr6 and Hpg4 are cross-linked by a biaryl carbon−carbon linkage reminiscent of vancomycin (Figure 1). The gene cluster comprises 10 genes and is consistent with the arylomycins core structure. The assembly line NRPS contain 6 modules on 3 genes (*aryABD*) where all A domains have the predicted substrate specificity of the observed amino acids, Ser, Ala/Gly, Gly, Hpg, Ala/Gly, and Tyr, respectively.[35,36] The loading module has a C domain that clusters with starter C domains based on a phylogenetic analysis (Figure S7). This C domain incorporates the *N*-acyl group into arylomycins as starter C domains are known to catalyze initial N-acylation in NRP biosynthesis.[37] The two *N*-methyl groups at positions 1 (Ser) and 4 (Hpg) in arylomycins are in agreement with the methyltransferase domains in the corresponding NRPS modules. Furthermore, the two D-amino acid residues at positions 1 (D-Ser) and 2 (D-Ala) are consistent with the epimerization domains in corresponding NRPS modules. Finally, the gene cluster contains a cytochrome P450 enzyme with 49% similarity to the vancomycin OxyC protein that is predicted to form the bis-aryl carbon linkages in vancomycin[38] (Figure S8). Thus, AryC is likely to be responsible for the biaryl bond formation in arylomycins. Notably, we did not observe the nitrosated congeners corresponding to the arylomycin B series described in previous reports[22,23] in *S. roseosporus,* which is in agreement with the absence of nitrosating enzymes in the gene cluster.[39] Therefore, the data suggest that SST enabled the discovery of the promising anti-infective agent arylomycins from *S. roseosporus.*

dx.doi.org/10.1021/ja2040877 |*J. Am. Chem. Soc.* 2011, 133, 18010–18013

To finally confirm the production of the arylomycins lipopeptides from the daptomycin-producing *S. roseosporus* strains, a larger scale fermentation, extraction, and purification using reversed-phase HPLC was undertaken. Compounds **1** and **2** were subjected to NMR analysis. The resulting ^{1}H NMR spectra revealed the same peptide core as arylomycin by comparing with the NMR spectra described in the original arylomycin report[22,33] (Figure S9). The ^{1}H NMR spectra of **1** and **2** matched to arylomycin A2 and A4, respectively.[33]

The identification of arylomycins and their biosynthetic gene cluster from intensely studied microorganism of commercial importance highlights the strength of MS-guided genome mining approaches and IMS effectively bridging the gap between phenotypes, chemotypes, and genotypes. The SST approach enables matching of molecules identified through imaging mass spectrometry to NRPS biosynthetic machinery using only a minimal sequence tag. We anticipate that SST will also prove capable of identifying the biosynthetic machinery for molecules that contain nonstandard amino acids, which are often incorporated in NRPs. According to the NCBI genome database, there are now ~1700 fully sequenced bacterial genomes as assessed in July 2011 in contrast to ~1000 in October 2009[40] which represents a more than 70% increase in less than 2 years (and there are ~5000 bacterial genome sequencing projects in progress). Since the full repertoire of genome sequence continues to expand at a rapid pace, there is a need to increase our effectiveness in genome mining to identify new natural products. Current high-throughput approaches (e.g., metabolomic or proteomic) do not efficiently identify natural products and therefore there is a need to develop genome mining approaches that enable us to rapidly connect chemotypes to genotypes and ultimately phenotypes. Imaging mass spectrometry and/or peptidogenomics, especially with its expansion to smaller peptides and SSTs, are approaches that expedite genome mining for amino acid containing natural products and their connection to phenotypes. Furthermore, our findings will enable the discovery of the arylomycin biosynthetic gene cluster thereby enabling future bioengineering to produce novel arylomycin analogues.

■ ASSOCIATED CONTENT

Supporting Information. Materials, detailed experimental procedures and supplementary figures. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
pdorrest@ucsd.edu

■ ACKNOWLEDGMENT

Financial support was provided by the National Institute of Health (GM085770 to B.S.M. and GM086283 and AI095125 to P.C.D.) and the Beckman Foundation. W.-T.L. was supported, in part, by a study abroad grant SAS-98116-2-US-108 from Taiwan.

■ REFERENCES

(1) Newman, D. J.; Cragg, G. M. *J. Nat. Prod.* **2007**, *70*, 461.
(2) Li, J. W.; Vederas, J. C. *Science* **2009**, *325*, 161.
(3) Aharonowitz, Y.; Cohen, G.; Martin, J. F. *Annu. Rev. Microbiol.* **1992**, *46*, 461.
(4) Miao, V.; Coeffet-Legal, M. F.; Brian, P.; Brost, R.; Penn, J.; Whiting, A.; Martin, S.; Ford, R.; Parr, I.; Bouchard, M.; Silva, C. J.; Wrigley, S. K.; Baltz, R. H. *Microbiology* **2005**, *151*, 1507.
(5) van Wageningen, A. M.; Kirkpatrick, P. N.; Williams, D. H.; Harris, B. R.; Kershaw, J. K.; Lennard, N. J.; Jones, M.; Jones, S. J.; Solenberg, P. J. *Chem. Biol.* **1998**, *5*, 155.
(6) Robbel, L.; Marahiel, M. A. *J. Biol. Chem.* **2010**, *285*, 27501.
(7) Payne, D. J.; Gwynn, M. N.; Holmes, D. J.; Pompliano, D. L. *Nat. Rev. Drug Discovery* **2007**, *6*, 29.
(8) Fischbach, M. A.; Walsh, C. T. *Science* **2009**, *325*, 1089.
(9) Yang, Y. L.; Xu, Y.; Straight, P.; Dorrestein, P. C. *Nat. Chem. Biol.* **2009**, *5*, 885.
(10) Liu, W. T.; Yang, Y. L.; Xu, Y.; Lamsa, A.; Haste, N. M.; Yang, J. Y.; Ng, J.; Gonzalez, D.; Ellermeier, C. D.; Straight, P. D.; Pevzner, P. A.; Pogliano, J.; Nizet, V.; Pogliano, K.; Dorrestein, P. C. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 16286.
(11) Yang, Y. L.; Xu, Y.; Kersten, R. D.; Liu, W. T.; Meehan, M. J.; Moore, B. S.; Bandeira, N.; Dorrestein, P. C. *Angew. Chem., Int. Ed.* **2011**, *50*, 5839.
(12) Chambers, H. F.; Deleo, F. R. *Nat. Rev. Microbiol.* **2009**, *7*, 629.
(13) Otto, M. *Nat. Rev. Microbiol.* **2009**, *7*, 555.
(14) Gu, J. Q.; Nguyen, K. T.; Gandhi, C.; Rajgarhia, V.; Baltz, R. H.; Brian, P.; Chu, M. *J. Nat. Prod.* **2007**, *70*, 233.
(15) Baltz, R. H. *J. Antibiot.* **2010**, *63*, 506.
(16) Nguyen, K. T.; Ritz, D.; Gu, J. Q.; Alexander, D.; Chu, M.; Miao, V.; Brian, P.; Baltz, R. H. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 17462.
(17) Baltz, R. H.; Miao, V.; Wrigley, S. K. *Nat. Prod. Rep.* **2005**, *22*, 717.
(18) Kersten, R. D.; Yang, Y. L.; Xu, Y.; Cimermancic, P.; Nam, S.-J.; Fenical, W.; Fischbach, M. A.; Moore, B. S.; Dorrestein, P. C. *Nat. Chem. Biol.* **2011**; DOI: 10.1038/NChemBio.684.
(19) Rausch, C.; Weber, T.; Kohlbacher, O.; Wohlleben, W.; Huson, D. H. *Nucleic Acids Res.* **2005**, *33*, 5799.
(20) Li, M. H.; Ung, P. M.; Zajkowski, J.; Garneau-Tsodikova, S.; Sherman, D. H. *BMC Bioinf.* **2009**, *10*, 185.
(21) Caboche, S.; Pupin, M.; Leclere, V.; Fontaine, A.; Jacques, P.; Kucherov, G. *Nucleic Acids Res.* **2008**, *36*, 326.
(22) Holtzel, A.; Schmid, D. G.; Nicholson, G. J.; Stevanovic, S.; Schimana, J.; Gebhardt, K.; Fiedler, H. P.; Jung, G. *J. Antibiot.* **2002**, *55*, 571.
(23) Schimana, J.; Gebhardt, K.; Hoeltzel, A.; Schmid, D. G.; Suessmuth, R.; Mueller, J.; Pukall, R.; Fiedler, H. P. *J. Antibiot.* **2002**, *55*, 565.
(24) Paetzel, M.; Goodall, J. J.; Kania, M.; Dalbey, R. E.; Page, M. G. *J. Biol. Chem.* **2004**, *279*, 30781.
(25) Bockstael, K.; Geukens, N.; Van Mellaert, L.; Herdewijn, P.; Anne, J.; Van Aerschot, A. *Microbiology* **2009**, *155*, 3719.
(26) Powers, M. E.; Smith, P. A.; Roberts, T. C.; Fowler, B. J.; King, C. C.; Trauger, S. A.; Siuzdak, G.; Romesberg, F. E. *J. Bacteriol.* **2011**, *193*, 340.
(27) Luo, C.; Roussel, P.; Dreier, J.; Page, M. G.; Paetzel, M. *Biochemistry* **2009**, *48*, 8976.
(28) Kulanthaivel, P.; Kreuzman, A. J.; Strege, M. A.; Belvo, M. D.; Smitka, T. A.; Clemens, M.; Swartling, J. R.; Minton, K. L.; Zheng, F.; Angleton, E. L. *J. Biol. Chem.* **2004**, *279*, 36250.
(29) Clardy, J.; Fischbach, M. A.; Walsh, C. T. *Nat. Biotechnol.* **2006**, *24*, 1541.
(30) Butler, M. S.; Buss, A. D. *Biochem. Pharmacol.* **2006**, *71*, 919.
(31) Smith, P. A.; Roberts, T. C.; Romesberg, F. E. *Chem. Biol.* **2010**, *17*, 1223.
(32) Smith, P. A.; Powers, M. E.; Roberts, T. C.; Romesberg, F. E. *Antimicrob. Agents Chemother.* **2011**, *55*, 1130.
(33) Roberts, T. C.; Smith, P. A.; Cirz, R. T.; Romesberg, F. E. *J. Am. Chem. Soc.* **2007**, *129*, 15830.
(34) Dufour, J.; Neuville, L.; Zhu, J. *Chemistry* **2010**, *16*, 10523.
(35) Challis, G. L.; Ravel, J.; Townsend, C. A. *Chem. Biol.* **2000**, *7*, 211.
(36) Stachelhaus, T.; Mootz, H. D.; Marahiel, M. A. *Chem. Biol.* **1999**, *6*, 493.
(37) Imker, H. J.; Krahn, D.; Clerc, J.; Kaiser, M.; Walsh, C. T. *Chem. Biol.* **2010**, *17*, 1077.
(38) Pylypenko, O.; Vitali, F.; Zerbe, K.; Robinson, J. A.; Schlichting, I. *J. Biol. Chem.* **2003**, *278*, 46727.
(39) Kersten, R. D.; Dorrestein, P. C. *Nat. Chem. Biol.* **2010**, *6*, 636.
(40) Walsh, C. T.; Fischbach, M. A. *J. Am. Chem. Soc.* **2010**, *132*, 2469.