JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Communication

pubs.acs.org/JACS

Biosynthetic Timing and Substrate Specificity for the Thiopeptide Thiomuracin

Zhengan Zhang,[†,§] Graham A. Hudson,[†,§] Nilkamal Mahanta,[†,‡] Jonathan I. Tietz,[†] Wilfred A. van der Donk,[*,†,‡] and Douglas A. Mitchell[*,†,‡]

[†]Department of Chemistry, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States

[‡]Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States

Supporting Information

ABSTRACT: The biosynthesis of the thiopeptide thiomuracin is a well-orchestrated process involving a multitude of posttranslational modifications. We show that six Cys residues of a precursor peptide are first cyclodehydrated and oxidized to thiazoles in an ordered, but nonlinear fashion that is leader-peptide-dependent. Then four alcohols are glutamylated and converted to alkenes in a C-to-N terminal directional process that is leader-peptide-independent. Finally, two of these alkenes undergo a formal [4 + 2] cycloaddition to form a trithiazole-substituted pyridine macrocycle. We describe here the factors that govern the substrate specificity and order of biosynthetic events that turn a ribosomal peptide into a powerful antibiotic.

Thiopeptides are a structurally complex class of antibiotics that belong to the rapidly expanding group of ribosomally synthesized and posttranslationally modified peptides (RiPPs).[1] With rare exceptions, RiPP precursor peptides consist of a leader peptide (LP), which are important for recognition by select biosynthetic proteins,[2] and a core peptide where the posttranslational modifications take place. Because of the direct link between the gene encoding the precursor peptide and the natural product structure, RiPPs have much bioengineering potential.[3−5] We recently reported the in vitro biosynthesis of thiomuracin GZ (**4**),[6] which only lacks the ancillary tailoring posttranslational modifications found in naturally produced thiomuracins.[7] The overall process converts a linear precursor peptide (TbtA, **1**) into a potent antibiotic by cyclodehydration by TbtFG and dehydrogenation by TbtE of the six Cys residues to the corresponding thiazoles. Subsequently, four alcohol groups are dehydrated in a process that involves glutamylation with Glu-tRNA by TbtB and subsequent Glu elimination by TbtC. Finally, two of the dehydrated residues react in a formal [4 + 2] cycloaddition (TbtD), followed by elimination of water and the LP to yield **4** (Figure 1A).

In this work, we determined the molecular details underlying this well-orchestrated set of enzymatic reactions. We show the order of thiazole and alkene formation, the minimal structural changes to render TbtA a substrate for dehydration, and reveal the parts of the TbtA peptide that are recognized by the various enzymes. Our results show that thiazole formation requires the LP, but dehydration surprisingly does not. These insights uncover which analogs of **4** will be accessible by in vitro biosynthesis using altered TbtA substrates.

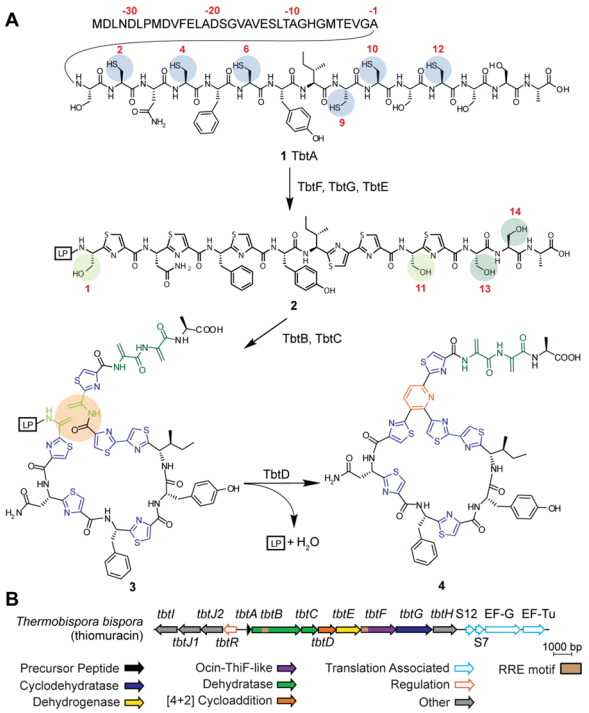

Figure 1. Thiomuracin biosynthesis. (A) Enzymatic route to the thiomuracin core scaffold **4**. (B) Thiomuracin biosynthetic gene cluster. RRE, RiPP recognition element. LP, leader peptide.

We first focused on the substrate permissiveness of the TbtEFG thiazole synthetase. A series of TbtA variants were prepared (Figure S1) in which each Cys residue was systematically substituted with Ala starting from the C-terminus (Figure S2). These variants were treated with purified TbtEFG as previously reported.[6] A complementary set of variants with Cys substituted beginning at the N-terminus of the core peptide was also prepared (Figure S3). All Ala-substituted TbtA variants were converted to the pan-thiazole species with incompletely processed peptides typically not observed (Figures S2 and S3). Six additional variants of TbtA were prepared that each contained a single Cys at the wild-type locations. TbtEFG

Received: August 27, 2016
Published: October 4, 2016

DOI: 10.1021/jacs.6b08987
J. Am. Chem. Soc. 2016, 138, 15511−15514

processed each variant to the expected monoazole (Figure S4), indicating that the installation of each thiazole is independent of the presence of other thiazoles.

With the substrate permissiveness of TbtEFG established, we next investigated the order of thiazole formation. After treating wild-type TbtA with TbtEFG, aliquots were removed and quenched at various times with excess iodoacetamide to alkylate unmodified Cys residues. TbtA mono-, di-, and triazole intermediates were intercepted whereas the tetra- and pentazole species were not, suggesting that the first three thiazoles are distributively installed with the remaining three processively introduced under our reaction conditions. Electrospray ionization (ESI) MS/MS analysis of these intermediates revealed a nonlinear processing order: Thz4 → Thz2 → Thz12 (Figure 2).

Given the observed substrate tolerance, we reasoned that TbtEFG relied heavily on the LP to drive substrate binding. Previous studies on other RiPPs have shown that the LP is important and that it is bound by a specific RiPP recognition element (RRE) that exhibits homology to PqqD.[8−11] For thiomuracin, PqqD-like domains are clearly present in the TbtF and TbtB proteins (Figure 1B) but the region(s) of the LP recognized are not known. Also, the [4 + 2] cycloaddition enzyme is LP-dependent,[6,12] yet TbtD lacks an identifiable RRE. Previous studies have shown that TbtF avidly binds the TbtA LP ($K_d \approx 70$ nM).[8] We thus evaluated binding to maltose-binding protein (MBP) tagged TbtB and MBP-TbtD using a synthetic TbtA LP derivatized with fluorescein isothiocyanate (FITC) at the N-terminus (FITC-TbtA$_{LP}$). Despite having PqqD-like domain, TbtB only weakly bound the LP ($K_d \approx 20$ μM, Figure S5), while the RRE-less TbtD bound the same peptide considerably tighter ($K_d \approx 140$ nM, Figure S6).

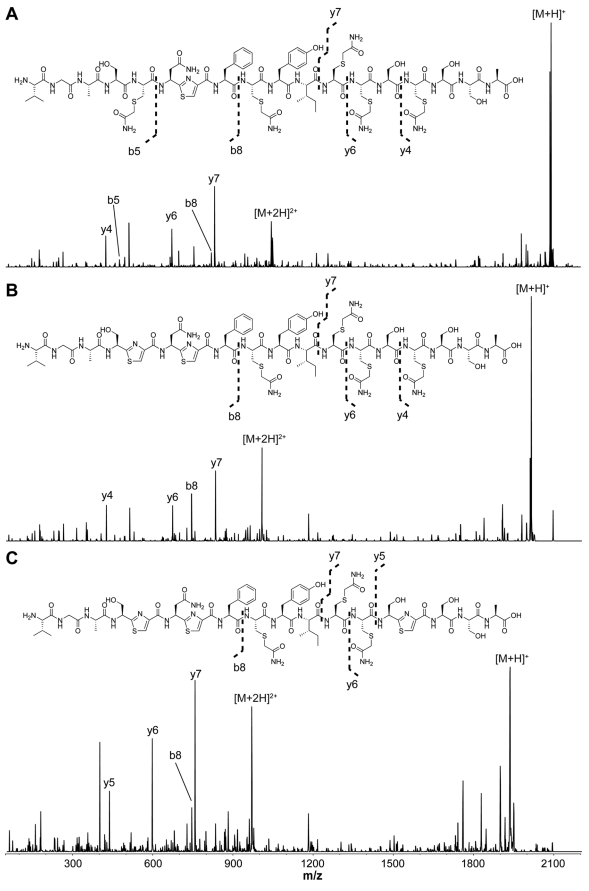

Figure 2. Tandem MS analysis of intermediates intercepted from TbtEFG reactions. (A) monoazole, (B) diazole, and (C) triazole. Predicted masses and associated errors are in Table S2.

To elucidate which parts of the LP are important for TbtF and TbtD binding, we generated a panel of TbtA variants and performed competition fluorescence polarization (FP) assays with FITC-TbtA$_{LP}$ to determine relative IC$_{50}$ values. This revealed a set of residues important for TbtF binding (Leu−29, Asp−26, Phe−24) and TbtD binding (Leu−29 as well, but also a number of acidic residues: Asp−30/−24/−20 and Glu−14, Table 1). Interestingly, many thiopeptide LPs are rich in Asp/Glu;[13−15] we hypothesize that TbtD will display a binding site to electrostatically complement a polyanionic LP with a hydrophobic pocket to accommodate LP position −29.

The regions of TbtA and TbtF important for binding were confirmed by measuring the affinity of the first 16 residues of TbtA (LP positions −34 to −19) versus the next 17 residues (−18 to −3). The N-terminal portion was bound by TbtF with near wild-type affinity while the C-terminal half showed no binding (Figure S9). Additionally, the RRE of TbtF (residues 1−85) retained most of the binding affinity ($K_d \approx 100$ nM) whereas the TbtF protein lacking the RRE (residues 86−628) did not detectably interact with FITC-TbtA$_{LP}$ (Figure S10).

We next evaluated the TbtBC-catalyzed dehydration reaction (Figure 1A). Dehydration is a two-step process in which TbtB first glutamylates a targeted Ser hydroxyl of TbtA; subsequently, TbtC eliminates the glutamate to yield the alkene. These dehydrations take place at positions 1/11/13/14 of the hexazole **2**.[6] To determine if a thiazoline-bearing TbtA would be accepted as a TbtBC substrate, the dehydrogenase component (TbtE) of the thiazole synthetase was omitted from reactions with TbtA. These reactions cleanly produced the expected hexazoline of TbtA, which, upon reaction with TbtBC, resulted in at most two dehydrations (Figure S11). These data show that although TbtFG does not require TbtE for activity, the hexazoline-containing TbtA is incompletely processed by TbtBC.

We next used **2** as a substrate for TbtB in the absence of TbtC, which resulted in mostly one glutamylation event.[6] Analysis by ESI MS/MS localized the glutamate to Ser14 (Figure 3). Similarly, assays that included TbtB and TbtC that were quenched prior to full tetradehydration showed that Ser13/14 were dehydrated faster than Ser1/11 (Figure S12). Thus, TbtBC prefers to modify the two isolated Ser residues over those adjacent to a thiazole where the Ser carbonyl group has been removed (Figure 1A). Alternatively, since dehydration of Ser1/11 sets up the [4 + 2] cycloaddition reaction, which we have shown is efficient,[6] TbtB may have evolved an increased level of substrate specificity at these positions, such that dehydration at Ser13/14 is a prerequisite for dehydration at Ser1/11.

We next focused on the glutamate elimination step. TbtC is homologous to the C-terminal domain of LanB proteins, which exhibit reduced activity upon mutation of a conserved Arg.[16] Therefore, we replaced this residue in TbtC with Ala (R58A) and carried out reactions with **2**. Mass spectral analysis showed a tridehydrated and monoglutamylated intermediate, with the latter localized to position 1 (Figure S13). Thus, Ser1 is acted on last, and glutamylation (and hence dehydration) progresses in a C-to-N terminal direction.

These enzymatic preferences do not inform on whether earlier dehydrations are important for subsequent ones. Therefore, a series of TbtA variants were again investigated. Hexazole-bearing TbtA-S13A and -S14A each were dehydrated predominantly twice by TbtBC, with Ser1/11 largely escaping dehydration

DOI: 10.1021/jacs.6b08987
J. Am. Chem. Soc. 2016, 138, 15511−15514

Table 1. Competitive Binding IC_{50} Values of TbtA Ala Variants[a]

TbtA LP variant	IC_{50} TbtF (μM)	IC_{50}TbtD (μM)	TbtA LP variant	IC_{50} TbtF (μM)	IC_{50} TbtD (μM)
Wild-type	0.67 ± 0.03	0.61 ± 0.03	D(−20)A	1.45 ± 0.46	4.13 ± 1.62
D(−33)A	0.99 ± 0.22	0.72 ± 0.07	S(−19)A	1.57 ± 0.46	0.91 ± 0.11
L(−32)A	1.58 ± 0.36	1.71 ± 0.47	V(−17)A	0.49 ± 0.05	0.69 ± 0.04
N(−31)A	0.57 ± 0.07	1.13 ± 0.08	V(−15)A	0.63 ± 0.06	1.06 ± 0.06
D(−30)A	0.94 ± 0.20	7.28 ± 0.77	E(−14)A	1.37 ± 0.15	4.47 ± 0.96
L(−29)A	≫10	≫10	S(−13)A	1.41 ± 0.07	2.76 ± 0.18
P(−28)A	0.85 ± 0.14	2.66 ± 0.55	L(−12)A	0.70 ± 0.04	3.29 ± 0.68
M(−27)A	2.75 ± 0.55	1.09 ± 0.19	T(−11)A	0.51 ± 0.06	0.45 ± 0.02
D(−26)A	8.23 ± 1.45	6.00 ± 0.41	H(−8)A	0.98 ± 0.06	3.68 ± 0.31
V(−25)A	0.62 ± 0.04	2.25 ± 0.20	M(−6)A	0.45 ± 0.03	0.49 ± 0.07
F(−24)A	>10	1.51 ± 0.09	E(−4)A	0.53 ± 0.03	0.70 ± 0.02
E(−23)A	1.27 ± 0.15	3.46 ± 0.57	V(−3)A	0.58 ± 0.05	1.35 ± 0.14

[a]Actual FP traces and K_d values are shown in Figures S7−S8.

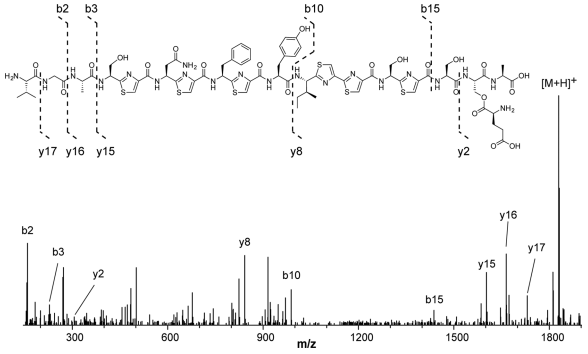

Figure 3. MS/MS of monoglutamylated **2**. The b15 ion localizes glutamylation to either Ser13 or Ser14. The low intensity y2 ion supports Ser14 glutamylation. Predicted masses are in Table S2.

(Figures S13−14). This observation suggests that dehydration at Ser1/11 is strongly promoted by prior dehydration at Ser13/14. In contrast, similarly treated hexazole-bearing TbtA-S1A and -S11A peptides each resulted in a tridehydrated product, underscoring that dehydration at Ser1/11 is not important for TbtBC to process Ser13/S14 (Figures S15−S16). These data are consistent with C-to-N terminal dehydration.

Given the comparatively weak interaction between $TbtA_{LP}$ and TbtB, we next investigated whether TbtBC might act in an LP-independent fashion. Peptide **2** was digested with endoproteinase GluC, providing a hexazole core peptide that retains only three residues of the LP (Figure 1).[6] Upon treatment with TbtBC, this peptide was tetradehydrated (Figure 4), in accord with the above-mentioned weak binding to $TbtA_{LP}$ ($K_d \approx 20$ μM). Strikingly, the affinity of TbtB toward FITC-labeled **2** was

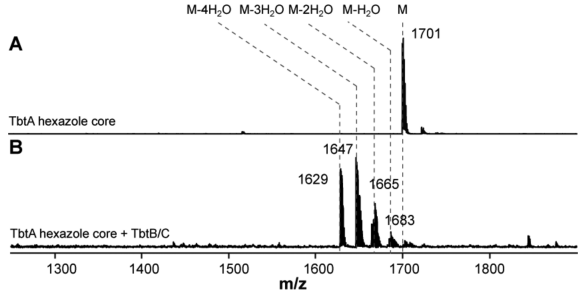

Figure 4. MALDI-TOF mass spectra of (A) the TbtA core peptide containing six thiazoles and (B) treated with TbtBC.

nearly 3 orders of magnitude tighter ($K_d \approx 30$ nM, Figure S5). GluC-digested **2** also effectively competed for TbtB binding to FITC-labeled **2**. These data demonstrate that the PqqD-like domain of TbtB is a nonfunctional RRE and that TbtB recognizes a specific conformation of the core peptide of TbtA induced by the posttranslational modifications. Ungapped sequence analysis confirms that the defunct RRE of TbtB does not contain conserved motifs, whereas the RRE of TbtF contains several, including locations known to physically contact the LP (Figures S17−18).[8,9,16,17] Motif conservation analysis was also performed on TbtD, revealing two polybasic regions potentially involved in RRE-independent LP binding (Figure S19).

The precise modifications that confer TbtBC activity were next investigated using the aforementioned panel of TbtA variants systematically replaced Cys with Ala starting from the C- or N-terminus (Figures S2 and S3). Each pan-thiazole TbtA variant was reacted with TbtBC. Substituting Cys2 and Cys4 with Ala, yielding the corresponding pentazole and tetrazole species, did not greatly affect dehydration at Ser13/14; however, dehydration at Ser11, and especially Ser1, was reduced (Figures 5A−B and S20A). Triple Ala replacement at positions 2/4/6 (triazole species) abolished dehydration at positions 1/11, but again, Ser13/14 underwent dehydration (Figures 5C and S21A). A similar trend was observed for tetra- and pentasubstituted TbtA variants (diazole and monoazole species, respectively, Figures 5D−E and S21B). These data suggest that a thiazole at position 12 might be the gatekeeper modification to render TbtA a substrate for TbtB. However, the TbtA-C12A pentazole was still dehydrated twice at Ser13/14 (Figures 5F and S20B). Conversely, Ala replacement at positions 10 and 12 (tetrazole species) abolished processing and the Cys10 TbtA monoazole was also not a substrate (Figures S22, S23). Hence, it is a thiazole at Cys10 that elicits TbtBC activity.

Lastly, we investigated the substrate requirements of TbtD. Since many of the variants were incompletely dehydrated, precluding the [4 + 2] cycloaddition reaction, we focused on further defining LP requirements. We previously reported that treatment of **3** with GluC, which removes all but three residues of the LP, resulted in loss of TbtD activity.[6] In this work, we tested if retaining a larger portion of the LP would still be processed by TbtD. Thus, we treated **3** with endoproteinase AspN, retaining 20 residues of the LP, and found that TbtD accepted this peptide as a substrate (Figure S24). We next tested a TbtA variant where Glu(−4) was substituted with Ala. This variant was efficiently converted to the hexazole tetradehyrate species, analogous to **3**, after reaction with TbtEFG and TbtBC. We then treated this

DOI: 10.1021/jacs.6b08987
J. Am. Chem. Soc. 2016, 138, 15511−15514

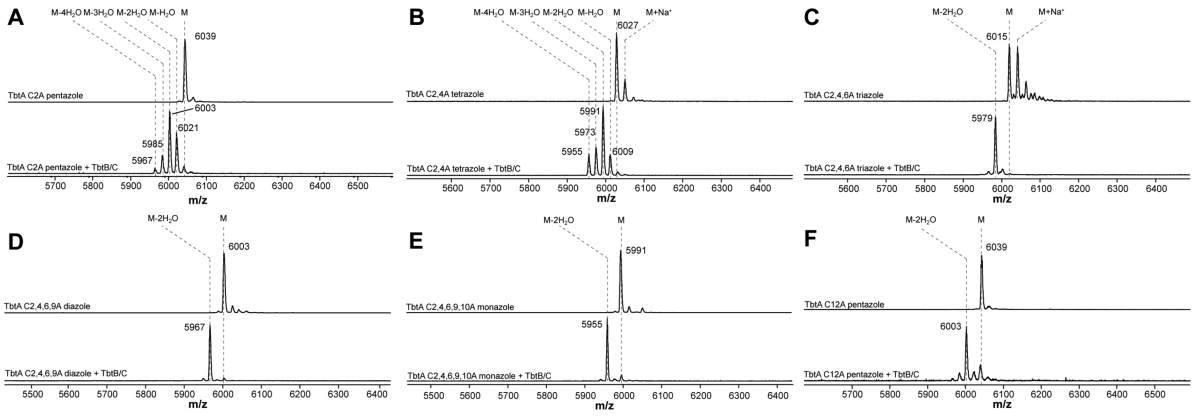

Figure 5. MALDI-TOF mass spectra of (A) TbtA C2A pentazole, (B) TbtA C2A/C4A tetrazole, (C) TbtA C2A/C4A/C6A triazole, (D) TbtA C2A/C4A/C6A/C9A diazole, (E) TbtA C2A/C4A/C6A/C9A/C10A monoazole, and (F) TbtA C12A pentazole, treated with TbtBC.

peptide with GluC, which, owing to the E(−4)A substitution, now retained 13 residues of the LP. This variant was also a substrate for TbtD (Figure S25), demonstrating that the minimal LP requirement of TbtD is between TbtA residues (−13) and (−3), consistent with a very recent report.[18] These findings may appear to contradict the binding data in Table 1, but it is notable that the binding experiments were performed with linear TbtA variants whereas the TbtD activity experiments with truncated TbtA variants were performed with a modified core peptide. We hypothesize that TbtD recognizes both the aforementioned N-terminal residues of the LP and the posttranslationally modified core peptide. At present, we cannot readily measure the latter since the enzyme rapidly processes **3**.

In summary, this work presents the first insights into the substrate specificities, directionality, and timing of catalysis by six proteins involved in thiopeptide formation. TbtEFG is remarkably tolerant in thiazole formation and the specificity is governed only by the presence of the LP. Leu(−29) of TbtA was found to be the most critical residue for binding by the RRE-containing TbtF protein. The order of thiazole formation is unusual in that it is neither N-to-C nor C-to-N directional. Conversely, the subsequent four dehydrations installed by TbtBC are LP-independent and installed in a strict C-to-N fashion. TbtBC appears to recognize a specific conformation introduced into the TbtA core peptide, imparted by thiazole formation at Cys10. Despite the conformation-restricting nature of peptidic cyclodehydration, a hexathiazoline-containing TbtA peptide is an incompetent substrate for TbtBC. Akin to thiazole formation, the [4 + 2] cycloaddition reaction is also LP-dependent with Leu(−29) again important for LP binding, in addition to several acidic residues. Despite requiring particular residues found within the N-terminal portion of the LP, TbtD also requires C-terminal LP residues, which we have narrowed down to between positions (−13) and (−3). The data reported here define the substrate scope for the in vitro biosynthesis of thiomuracin and will guide future efforts to improve the properties of thiopeptide natural products.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b08987.

Experimental details and supporting figures (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*douglasm@illinois.edu

*vddonk@illinois.edu

Author Contributions

§G.A.H. and Z.Z. contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported in part by the National Institutes of Health (GM097142 to D.A.M.; GM058822 to W.A.V.).

■ REFERENCES

(1) Arnison, P. G.; Bibb, M. J.; Bierbaum, G.; Bowers, A. A.; Bugni, T. S.; Bulaj, G.; Camarero, J. A.; Campopiano, D. J.; Challis, G. L.; Clardy, J.; et al. *Nat. Prod. Rep.* **2013**, *30*, 108.

(2) Yang, X.; van der Donk, W. A. *Chem. - Eur. J.* **2013**, *19*, 7662.

(3) Sardar, D.; Lin, Z.; Schmidt, E. W. *Chem. Biol.* **2015**, *22*, 907.

(4) Pan, S. J.; Link, A. J. *J. Am. Chem. Soc.* **2011**, *133*, 5016.

(5) Li, C.; Zhang, F.; Kelly, W. L. *Mol. BioSyst.* **2011**, *7*, 82.

(6) Hudson, G. A.; Zhang, Z.; Tietz, J. I.; Mitchell, D. A.; van der Donk, W. A. *J. Am. Chem. Soc.* **2015**, *137*, 16012.

(7) Morris, R. P.; Leeds, J. A.; Naegeli, H. U.; Oberer, L.; Memmert, K.; Weber, E.; Lamarche, M. J.; Parker, C. N.; Burrer, N.; Esterow, S.; Hein, A. E.; Schmitt, E. K.; Krastel, P. *J. Am. Chem. Soc.* **2009**, *131*, 5946.

(8) Burkhart, B. J.; Hudson, G. A.; Dunbar, K. L.; Mitchell, D. A. *Nat. Chem. Biol.* **2015**, *11*, 564.

(9) Koehnke, J.; Mann, G.; Bent, A. F.; Ludewig, H.; Shirran, S.; Botting, C.; Lebl, T.; Houssen, W. E.; Jaspars, M.; Naismith, J. H. *Nat. Chem. Biol.* **2015**, *11*, 558.

(10) Dunbar, K. L.; Tietz, J. I.; Cox, C. L.; Burkhart, B. J.; Mitchell, D. A. *J. Am. Chem. Soc.* **2015**, *137*, 7672.

(11) Latham, J. A.; Iavarone, A. T.; Barr, I.; Juthani, P. V.; Klinman, J. P. *J. Biol. Chem.* **2015**, *290*, 12908.

(12) Wever, W. J.; Bogart, J. W.; Baccile, J. A.; Chan, A. N.; Schroeder, F. C.; Bowers, A. A. *J. Am. Chem. Soc.* **2015**, *137*, 3494.

(13) Li, J.; Qu, X.; He, X.; Duan, L.; Wu, G.; Bi, D.; Deng, Z.; Liu, W.; Ou, H.-Y. *PLoS One* **2012**, *7*, e45878.

(14) Zhang, Q.; Liu, W. *Nat. Prod. Rep.* **2013**, *30*, 218.

(15) Just-Baringo, X.; Albericio, F.; Alvarez, M. *Mar. Drugs* **2014**, *12*, 317.

(16) Ortega, M. A.; Hao, Y.; Zhang, Q.; Walker, M. C.; van der Donk, W. A.; Nair, S. K. *Nature* **2014**, *517*, 509.

(17) Bailey, T. L.; Williams, N.; Misleh, C.; Li, W. W. *Nucleic Acids Res.* **2006**, *34*, W369.

(18) Wever, W. J.; Bogart, J. W.; Bowers, A. A. *J. Am. Chem. Soc.*, Just Accepted Manuscript, DOI: 10.1021/jacs.6b05389.

DOI: 10.1021/jacs.6b08987
J. Am. Chem. Soc. 2016, 138, 15511−15514