JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Communication

pubs.acs.org/JACS

A Prodrug Resistance Mechanism Is Involved in Colibactin Biosynthesis and Cytotoxicity

Carolyn A. Brotherton and Emily P. Balskus*

Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, United States

Supporting Information

ABSTRACT: Commensal *Escherichia coli* residing in the human gut produce colibactin, a small-molecule genotoxin of unknown structure that has been implicated in the development of colon cancer. Colibactin biosynthesis is hypothesized to involve a prodrug resistance strategy that entails initiation of biosynthesis via construction of an N-terminal prodrug scaffold and late-stage cleavage of this structural motif during product export. Here we describe the biochemical characterization of the prodrug synthesis, elongation, and cleavage enzymes from the colibactin biosynthetic pathway. We show that nonribosomal peptide synthetases ClbN and ClbB assemble and process an *N*-acyl-D-asparagine prodrug scaffold that serves as a substrate for the periplasmic D-amino peptidase ClbP. In addition to affording information about structural features of colibactin, this work reveals the biosynthetic logic underlying the prodrug resistance strategy and suggests that cytotoxicity requires amide bond cleavage.

Colibactin is a small-molecule genotoxin produced by certain strains of *Escherichia coli*, including commensals found in the human gut.[1] This secondary metabolite is biosynthesized by a hybrid nonribosomal peptide synthetase–polyketide synthase (NRPS–PKS) assembly line encoded within the *pks* genomic island (Figure 1A). *E. coli* harboring the *pks* island (*pks*$^+$) damage the DNA of eukaryotic cells both in vitro[1] and in vivo,[2] inducing double strand breaks and disrupting the cell cycle. This biological activity has raised concerns that colibactin production by gut microbes might contribute to human disease. This hypothesis is supported by recent work demonstrating that *pks*$^+$ *E. coli* promote tumorigenesis in a mouse model of colitis-associated colorectal cancer.[3] Understanding the mechanism and consequences of colibactin-induced DNA damage is critical, since an estimated 21% of humans harbor colibactin-synthesizing *E. coli* within their intestinal microbiota.[3,4] However, the structure of colibactin is currently unknown, as isolation attempts to date have been unsuccessful. Here we report the biochemical characterization of ClbN and ClbB, NRPS modules involved in the initiation of colibactin biosynthesis, and ClbP, a peptidase responsible for tailoring of the natural product scaffold during export. This work reveals structural features of colibactin and provides molecular evidence for the involvement of a prodrug strategy in its production and cytotoxicity.

We initiated our studies of colibactin biosynthesis by performing a detailed annotation and bioinformatic analysis of the genes in the *pks* island. We recognized that certain assembly line components and a tailoring enzyme encoded by the cluster closely resemble genes involved an antibiotic resistance mechanism that had been first postulated to occur in zwittermicin biosynthesis[5] and was recently shown using in vivo experiments to be operative in the assembly of xenocoumacin.[6] This resistance mechanism, which is known as the prodrug strategy, is hypothesized to be conserved and widespread, as homologues of the enzymatic machinery implicated in this process are encoded in many otherwise unrelated secondary metabolite biosynthetic gene clusters.[6,7a]

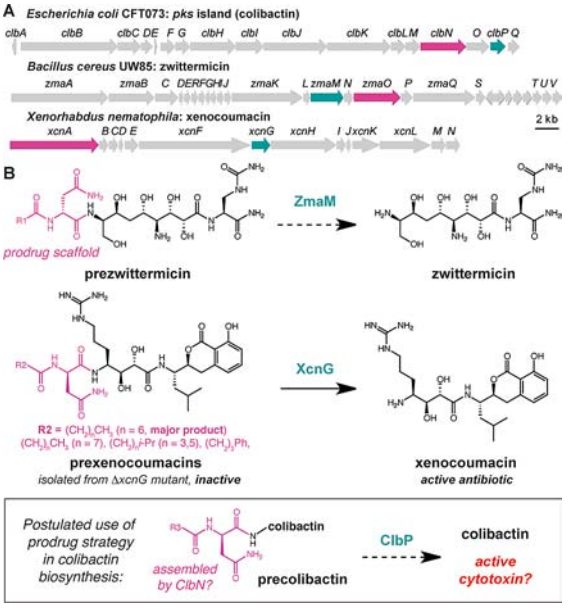

Figure 1. Colibactin biosynthesis may involve a prodrug resistance mechanism. (A) The *pks* gene cluster and other NRPS–PKS-containing gene clusters harboring putative prodrug scaffold synthesis and cleavage machinery. Genes encoding NRPSs predicted to synthesize prodrug scaffolds (pink) and peptidases that may cleave prodrugs (green) are highlighted. (B) Postulated biochemical logic underlying the prodrug resistance mechanism in the zwittermicin, xenocoumacin, and colibactin biosynthetic pathways.

The prodrug resistance strategy involves installation of a structural motif at the N-terminus of a secondary metabolite that is removed in the final stages of biosynthesis (Figure 1B). This prodrug scaffold, which consists of an *N*-acyl-D-asparagine residue, is thought to be assembled by an initiating NRPS

Received: December 12, 2012
Published: February 13, 2013

   dx.doi.org/10.1021/ja312154m | J. Am. Chem. Soc. 2013, 135, 3359–3362

module containing condensation (C), adenylation (A), thiolation (T), and epimerization (E) domains and extended by an enzymatic assembly line. After release of the nascent natural product from the assembly line, a periplasmic D-amino peptidase removes the prodrug scaffold to give the active molecule. This process has been previously characterized in xenocoumacin biosynthesis; deletion of the gene encoding the peptidase XcnG abolished antibiotic production and generated uncleaved "prexenocoumacins" that lacked activity.[6] Although there is in vivo evidence for the biochemical logic underlying this resistance mechanism, there has been no comprehensive in vitro characterization of the prodrug synthesis and cleavage enzymes.[8] We decided to study the activities of the putative initiating NRPS and D-amino peptidase from the *pks* cluster, anticipating that an understanding of the prodrug resistance mechanism in colibactin biosynthesis not only would illuminate important structural features of the natural product but also could reveal strategies for preventing cytotoxin production.

We first examined ClbN, the NRPS module predicted to be responsible for construction of the N-terminal prodrug motif (Figure 2A). The structures of prexenocoumacins revealed

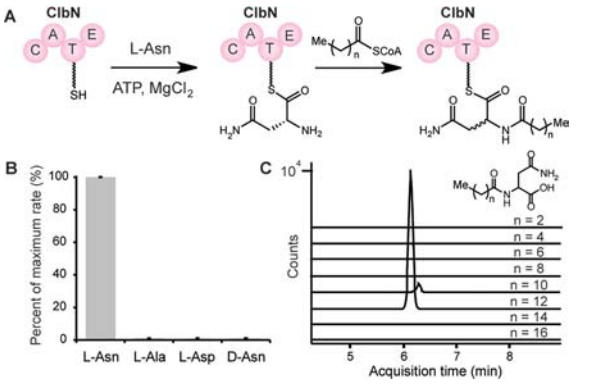

Figure 2. ClbN is an NRPS involved in constructing the prodrug scaffold. (A) Biosynthetic hypothesis for the activity of ClbN. (B) ATP–[^{32}P]PP$_i$ exchange assay for ClbN substrate specificity. (C) Extracted ion chromatograms of products obtained from the LC–MS competition assay with ClbN, L-Asn, and equal amounts of C_4–C_{18} fatty acyl-CoA substrates. All traces have been scaled to the same intensity.

prodrug scaffolds consisting of an *N*-acyl chain of variable length appended to an invariant D-asparagine.[6] In xenocoumacin biosynthesis, this structure is likely constructed by the first module of NRPS XcnA. ClbN, the homologous NRPS from the *pks* island, was cloned from *E. coli* CFT073, overexpressed as an N-His$_6$-tagged fusion, and purified. A boron dipyrromethene–coenzyme A (BODIPY–CoA) loading assay[9] confirmed that the T domain of apo-ClbN could be post-translationally modified to the holo protein [Figure S3 in the Supporting Information (SI)]. Examination of the ClbN A domain's substrate specificity using the ATP–[^{32}P]PP$_i$ exchange assay revealed selective activation of L-asparagine (Figure 2B). This selectivity matched that observed previously for ZmaO, the homologous NRPS module from the zwittermicin biosynthetic gene cluster.[5] Finally, a radiometric loading assay using ^{14}C-labeled L-asparagine demonstrated that the T domain of holo-ClbN is charged with L-asparagine in the presence of Sfp and ATP (Figure S5). Together, these results confirm that ClbN utilizes L-asparagine, consistent with its predicted role in prodrug scaffold construction.

We next investigated whether the ClbN C domain could acylate the amino substituent of the loaded L-asparagine. Multiple sequence alignments revealed that this domain possesses amino acid motifs characteristic of starter C domains involved in lipoinitiation (Figure S32).[10] Starter C domains employ multiple strategies for N-acylation, including the use of fatty acyl thioesters bound to trans-acting T domains[11] and the utilization of fatty acyl-CoA thioesters.[12] We hypothesized that ClbN would accept fatty acyl-CoA substrates because the *pks* island does not encode any homologues of fatty acid-activating enzymes. We were also interested in the specificity of this reaction, as the isolation of multiple prexenocoumacins suggested the potential for promiscuity in the initial N-acylation event.[6] To address these questions, we examined the activity of ClbN using a liquid chromatography–mass spectrometry (LC–MS)-based assay. Incubation of holo-ClbN with L-asparagine, ATP, and individual even-numbered straight-chain saturated fatty acyl-CoA substrates revealed that the enzyme. Accepted fatty acyl-CoAs with chain lengths ranging from 6 to 14 carbons (Figure S6). We further examined substrate specificity using a competition assay, incubating ClbN with a mixture of C_4 to C_{18} fatty acyl-CoAs. Strikingly, under these conditions, ClbN preferentially utilized myristoyl-CoA (C_{14}) for N-acylation (Figure 2C). To explore whether this substrate specificity might reflect in vivo activity, we used LC–MS to compare metabolites produced by strains expressing either ClbN or an empty expression vector. We identified both C_{12} and C_{14} *N*-acyl-asparagine in the strain expressing ClbN; these metabolites were not found in the control strain (Figure S11). Moreover, we were unable to detect shorter chain *N*-acyl-asparagine products in either strain. These preliminary results suggest that the selectivity observed in the in vitro competition assay may reflect ClbN's substrate preference in vivo. Finally, we verified that the ClbN E domain is functional by performing assays in D_2O and using LC–MS to confirm the incorporation of deuterium into L-Asn resulting from exchange during epimerization (Figures S13–S16).[13]

We then explored whether the *N*-acyl-asparagine scaffold synthesized by ClbN was processed further by the colibactin NRPS–PKS assembly line. The logic underlying the prodrug strategy requires that the next module be an NRPS in order to construct an amide bond that can serve as an eventual peptidase substrate. To find a candidate module, we used bioinformatic analyses to identify $^{D}C_L$ domains, which are C domains that follow E domains and accept upstream D-amino acyl thioester electrophiles.[10,14] The C domain of ClbB, a two-module NRPS–PKS hybrid, is the only $^{D}C_L$ domain in the *pks* island (Figure S33), indicating that it likely follows ClbN in the assembly line (Figure 3A). To test this hypothesis, we cloned, overexpressed, and purified the NRPS portion of ClbB (ClbB$_{NRPS}$), which contains C, A, and T domains, from *E. coli* CFT073. A BODIPY–CoA loading assay confirmed the successful posttranslational modification of apo-ClbB$_{NRPS}$ (Figure S17) and an ATP–[^{32}P]PP$_i$ exchange assay revealed that its A domain accepts multiple amino acids (Figure 3B). ClbB$_{NRPS}$ displayed the highest activity with L-alanine but also activated L-valine, L-serine, and glycine. This promiscuity extended to T domain loading, as radiometric assays showed tethering of both L-alanine and L-valine (Figure S18).

We examined the ability of ClbB$_{NRPS}$ to process substrates from ClbN by incubating both holo enzymes with L-asparagine, L-alanine, ATP, and a variety of fatty acyl-CoA substrates. We detected N-acylated dipeptide products in the presence of C_8 and C_{14} fatty acyl-CoAs using LC–MS (Figure 3C). This result indicates that ClbB does accept substrates from ClbN, validating

dx.doi.org/10.1021/ja312154m | J. Am. Chem. Soc. 2013, 135, 3359–3362

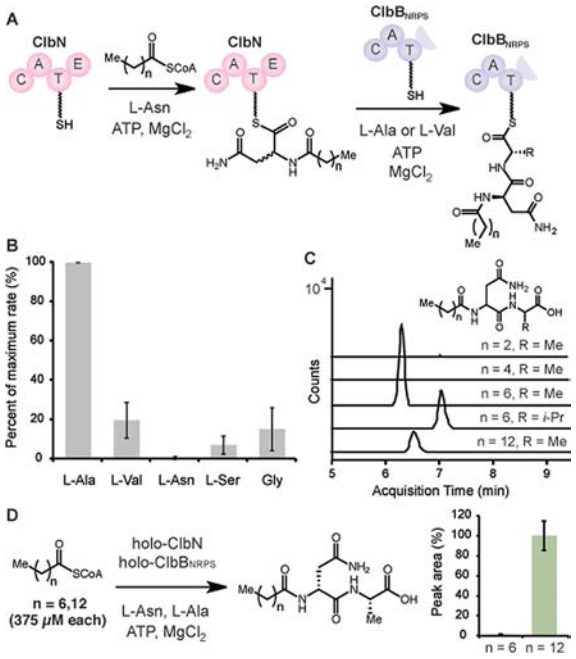

Figure 3. The NRPS module of ClbB elongates the prodrug scaffold. (A) Biosynthetic hypothesis for the activity of $ClbB_{NRPS}$. (B) ATP–[^{32}P]PP_i exchange assay for $ClbB_{NRPS}$ substrate specificity. (C) Extracted ion chromatograms of products obtained from individual LC–MS assays with ClbN; $ClbB_{NRPS}$; C_4, C_6, C_8, or C_{14} fatty acyl-CoA; L-Asn; and L-Ala or L-Val. All traces have been scaled to the same intensity. (D) LC–MS competition assay with equal amounts of C_8 and C_{14} fatty acyl-CoA substrates. Each bar represents the mean ± standard error of the mean (SEM) of normalized peak areas for extracted ion chromatograms from three experiments.

our prediction of assembly line order. $ClbB_{NRPS}$ did not generate products containing C_4 and C_6 *N*-acyl substituents. We also performed the assay with L-valine in place of L-alanine and were able to detect the corresponding N-acylated dipeptide (Figure 3C). Finally, we ran a competition experiment in which both enzymes were incubated with L-alanine and equal concentrations of octanoyl- and myristoyl-CoA; again the myristoylated product predominated (Figure 3D). Together, these data indicate that although ClbB is promiscuous with respect to both the chain length of the *N*-acyl group on the intermediate it accepts from ClbN and the amino acid it uses for elongation, the selectivity of the initial N-acylation by ClbN may control prodrug scaffold structure. Furthermore, the use of multiple amino acids by $ClbB_{NRPS}$ suggests the possibility that colibactin could be a set of related metabolites with different N-terminal amino acids.

Having established the roles of ClbN and ClbB in synthesizing and elongating the prodrug scaffold, we next examined whether peptidase ClbP is capable of prodrug activation. Full-length ClbP contains an N-terminal signal sequence that targets the protein to the inner membrane, a periplasmic peptidase domain containing the active site, and three C-terminal transmembrane helices.[7] Genetic studies in *E. coli* have shown that *clbP* is required for activity,[1] and previous characterization of ClbP in vivo has focused primarily on elucidating the elements of the protein required for restoring cytotoxicity to *clbP* mutants.[7b] These experiments revealed that the N-terminal signal sequence and all three transmembrane helices are necessary for complementation. Surprisingly, the periplasmic peptidase domain of ClbP ($ClbP_{pep}$), which has been overexpressed, purified, and crystallized,[7a] is unable to restore cytotoxicity.[7b] It was unclear whether this difference in cytotoxicity is directly linked to catalysis of amide bond cleavage, as the activities of ClbP and $ClbP_{pep}$ have not been examined with substrates containing the prodrug scaffold.

We sought to understand the chemistry of prodrug activation in more detail by evaluating the ability of ClbP to cleave substrates containing structural motifs generated by ClbN. Our initial efforts focused on $ClbP_{pep}$, which we cloned from *E. coli* CFT073 and expressed periplasmically. Using high-performance liquid chromatography (HPLC), we found that $ClbP_{pep}$ can cleave substrates containing *N*-octanoyl-D-asparagine and either L-alanine or L-valine (Figures S22 and S23). A version of $ClbP_{pep}$ containing a point mutation of the catalytic serine residue (S95A) was not able to cleave these substrates, confirming that activity is due to $ClbP_{pep}$. As expected, $ClbP_{pep}$ did not accept an *N*-octanoyl-L-asparagine-containing substrate (Figure S24). However, model substrates containing *N*-myristoyl-D-asparagine also were not cleaved by $ClbP_{pep}$ (Figure S27). This substrate specificity contrasted sharply with the product distribution observed in the ClbN competition experiment, leading us to hypothesize that the behavior of the truncated enzyme in vitro might not accurately reflect the activity of the full-length enzyme in vivo.

We investigated this possibility by examining the ability of *E. coli* expressing either $ClbP_{pep}$ or full-length ClbP to cleave model substrates in vivo. We incubated cultures expressing each of these proteins, as well as the S95A mutant of full-length ClbP and an empty vector control, with model substrates and used LC–MS to detect the cleavage products. Strikingly, we found that while both C_8 and C_{14} *N*-acyl-D-asparagine-containing substrates were processed by full-length ClbP, $ClbP_{pep}$ exhibited no activity (Figure 4). This behavior is consistent with the in vivo activities

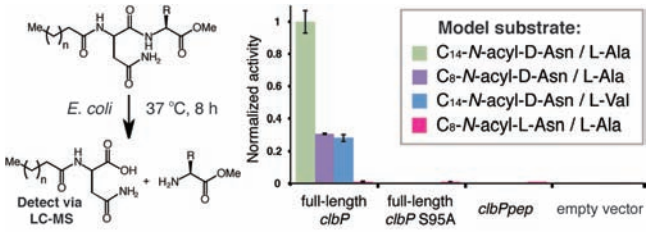

Figure 4. Peptidase ClbP cleaves the prodrug signal sequence in vivo. LC–MS assays detecting *N*-acyl-D- or -L-asparagine produced from cleavage of model peptidase substrates (100 μM) by *E. coli* BL21(DE3) strains are shown. Each bar represents the mean ± SEM of three experiments.

observed for cleavage of prexenocoumacin B by heterologously expressed XcnG and the isolated XcnG peptidase domain.[6] The ClbP S95A mutant was inactive, again indicating that cleavage is due to ClbP's peptidase function. Incubations with an *N*-octanoyl-L-asparagine-containing substrate resulted in negligible levels of hydrolysis and greater amounts of unreacted starting material compared with the corresponding D-asparagine-containing substrate (Tables S12–S14). As the extent of hydrolysis was similar across all of the cultures, including the ClbP S95A mutant, we attribute this reactivity to endogenous peptidases. The strong preference of ClbP for substrates containing D-asparagine over L-asparagine further confirms that ClbB elongates *N*-acyl-D-asparagine. Finally, we examined the activity of ZmaM, a homologous peptidase from the zwittermicin biosynthetic pathway that can restore cytotoxicity to a *clbP* mutant.[7a] *E. coli* expressing ZmaM also processed model

dx.doi.org/10.1021/ja312154m | J. Am. Chem. Soc. 2013, 135, 3359–3362

substrates (Figure S30), providing further support for the involvement of prodrug cleavage in generating active cytotoxin. Notably, *E. coli* expressing ZmaM cannot cleave prexenocoumacin B.[6] Our results suggest that this discrepancy arises from differences in metabolite structure, indicating that these peptidases have evolved to recognize specific substrates.

Together with previous in vivo complementation experiments,[7] these results provide evidence that colibactin formation involves amide bond hydrolysis. Importantly, the differential activity of ClbP and $ClbP_{pep}$ toward model substrates in vivo parallels their ability to complement *clbP* mutants and restore cytotoxicity to *pks*$^+$ *E. coli*.[7b] This observation strongly suggests that generation of active cytotoxin requires cleavage of the prodrug scaffold generated by ClbN (Figure 5). The difference in

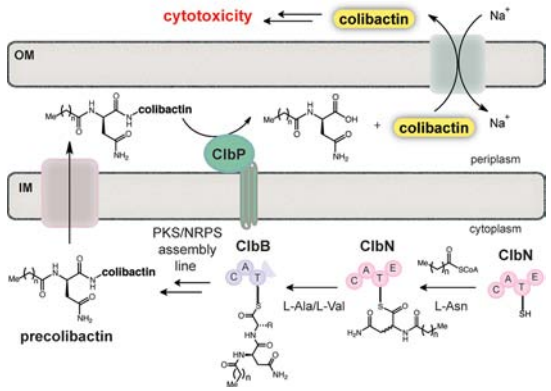

Figure 5. Proposed biochemical logic underlying the prodrug resistance mechanism in colibactin biosynthesis. IM = inner membrane, OM = outer membrane.

the activities of the full-length enzyme and the truncated peptidase domain is intriguing and could indicate that the transmembrane helices influence the catalytic activity or that interactions of both substrate and enzyme with the inner membrane are important. In support of the latter hypothesis, examination of the putative substrate binding pocket of ClbP revealed a negative electrostatic potential,[7b] which suggests that colibactin is positively charged. Positively charged amino acids facilitate binding of myristoylated proteins to cell membranes as a result of strong electrostatic interactions with acidic phospholipids.[15] If precolibactin contains a myristoylated prodrug scaffold attached to a positively charged core, it may be similarly membrane-associated. Our work also confirms that the biochemical logic employed by the colibactin prodrug resistance mechanism is conserved across multiple biosynthetic pathways, as ClbN produced a prodrug scaffold resembling that of the prexenocoumacins and peptidase ZmaM processed model substrates containing the colibactin prodrug scaffold.

In summary, we have gained insights into the biochemical events involved in colibactin assembly by characterizing the activities of ClbN, ClbB, and ClbP, the enzymes involved in synthesizing, elongating, and cleaving the colibactin prodrug scaffold. By elucidating the biosynthetic logic underlying this resistance strategy, we have obtained information about structural features of colibactin, which should have an immediate practical impact on efforts to isolate and characterize this natural product. More importantly, we have also gained an understanding of the chemistry employed in generating the mature cytotoxin. Ultimately, this knowledge could also guide the design of small molecules that interfere with colibactin production by inhibiting ClbP. Such compounds would be useful for elucidating the role that cytotoxin production plays in gut communities containing *pks*$^+$ *E. coli* and could provide a means of preventing colibactin-mediated carcinogenesis.

■ ASSOCIATED CONTENT

Supporting Information

Experimental methods and additional data. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

balskus@chemistry.harvard.edu

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Alex Sieg for help with initial experiments; Alan Saghatelian and Sunia Trauger for assistance with LC−MS; Jo Handelsman for providing *Bacillus cereus* UW85; and Harvard University, the Corning Foundation, the Searle Scholars Program, the Richard and Susan Smith Family Foundation (Newton, MA), and the Milton Fund for support. C.A.B. acknowledges a fellowship from Novartis.

■ REFERENCES

(1) Nougayrède, J. P.; Homburg, S.; Taieb, F.; Boury, M.; Brzuszkiewicz, E.; Gottschalk, G.; Buchrieser, C.; Hacker, J.; Dobrindt, U.; Oswald, E. *Science* **2006**, *313*, 848.

(2) Cuevas-Ramos, G.; Petit, C. R.; Marcq, I.; Boury, M.; Oswald, E.; Nougayrède, J. P. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 11537.

(3) Arthur, J. C.; Perez-Chanona, E.; Mühlbauer, M.; Tomkovich, S.; Uronis, J. M.; Fan, T. J.; Campbell, B. J.; Abujamel, T.; Dogan, B.; Rogers, A. B.; Rhodes, J. M.; Stintzi, A.; Simpson, K. W.; Hansen, J. J.; Keku, T. O.; Fodor, A. A.; Jobin, C. *Science* **2012**, *338*, 120.

(4) Johnson, J. R.; Johnston, B.; Kuskowski, M. A.; Nougayrède, J. P.; Oswald, E. *J. Clin. Microbiol.* **2008**, *46*, 3906.

(5) Kevany, B. M.; Rasko, D. A.; Thomas, M. G. *Appl. Environ. Microbiol.* **2009**, *75*, 1144.

(6) Reimer, D.; Pos, K. M.; Thines, M.; Grün, P.; Bode, H. B. *Nat. Chem. Biol.* **2011**, *7*, 888.

(7) (a) Dubois, D.; Baron, O.; Cougnoux, A.; Delmas, J.; Pradel, N.; Boury, M.; Bouchon, B.; Bringer, M.; Nougayrède, J.; Oswald, E.; Bonnet, R. *J. Biol. Chem.* **2011**, *286*, 35562. (b) Cougnoux, A.; Gibold, L.; Robin, F.; Dubois, D.; Pradel, N.; Darfeuille-Michaud, A.; Dalmasso, G.; Delmas, J.; Bonnet, R. *J. Mol. Biol.* **2012**, *424*, 203.

(8) $ClbP_{pep}$ has previously been reported to cleave Gly-, D-Ala-, L-Ala-, and L-Ala-L-Ala-*p*-nitroanilides with low catalytic efficiencies (k_{cat}/K_m = 0.005, 0.018, 0.020, and 0.040 μM^{-1} min^{-1}, respectively).[7a] We were unable to detect any cleavage activity with D-Ala- and L-Ala-*p*-nitroanilides using $ClbP_{pep}$ purified from periplasmic expressions.

(9) La Clair, J. J.; Foley, T. L.; Schegg, T. R.; Regan, C. M.; Burkart, M. D. *Chem. Biol.* **2004**, *11*, 195.

(10) Rausch, C.; Hoof, I.; Weber, T.; Wohlleben, W.; Huson, D. H. *BMC Evol. Biol.* **2007**, *7*, 78.

(11) Wittmann, M.; Linne, U.; Pohlmann, V.; Marahiel, M. A. *FEBS J.* **2008**, *275*, 5343.

(12) Kraas, F. I.; Helmetag, V.; Wittmann, M.; Strieker, M.; Marahiel, M. A. *Chem. Biol.* **2010**, *17*, 872.

(13) Stachelhaus, T.; Walsh, C. T. *Biochemistry* **2000**, *39*, 5775.

(14) Clugston, S. L.; Siber, S. A.; Marahiel, M. A.; Walsh, C. T. *Biochemistry* **2003**, *42*, 12095.

(15) Sigal, C. T.; Zhou, W. J.; Buser, C. A.; McLaughlin, S.; Resh, M. D. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 12253. (b) McLaughlin, S.; Aderem, A. *Trends Biochem. Sci.* **1995**, *20*, 272.