JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

The reactivity of an unusual amidase may explain colibactin's DNA cross-linking activity

Yindi Jiang, Alessia Stornetta, Peter W. Villalta, Matthew R. Wilson, Paul D. Boudreau, Li Zha, Silvia Balbo, and Emily P. Balskus

J. Am. Chem. Soc., **Just Accepted Manuscript** • DOI: 10.1021/jacs.9b02453 • Publication Date (Web): 28 Jun 2019

Just Accepted

is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

The reactivity of an unusual amidase may explain colibactin's DNA cross-linking activity

Yindi Jiang,[†] Alessia Stornetta,[‡] Peter W. Villalta,[‡] Matthew R. Wilson,[†,§] Paul D. Boudreau,[†] Li Zha,[†,∥] Silvia Balbo,[‡] and Emily P. Balskus[*,†]

[†]Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138, United States

[‡]Masonic Cancer Center, University of Minnesota, 2231 Sixth Street Southeast, Minneapolis, MN 55455, United States

ABSTRACT: Certain commensal and pathogenic bacteria produce colibactin, a small molecule genotoxin that causes interstrand cross-links in host cell DNA. Though colibactin alkylates DNA, the molecular basis for cross-link formation is unclear. Here, we report that the colibactin biosynthetic enzyme ClbL is an amide bond-forming enzyme that links aminoketone and β-keto thioester substrates *in vitro* and *in vivo*. The substrate specificity of ClbL strongly supports a role for this enzyme in terminating the colibactin NRPS-PKS assembly line and incorporating two electrophilic cyclopropane warheads into the final natural product scaffold. This proposed transformation was supported by the detection of a colibactin-derived cross-linked DNA adduct. Overall, this work provides a biosynthetic explanation for colibactin's DNA cross-linking activity and paves the way for further study of its chemical structure and biological roles.

INTRODUCTION

Colibactin is a genotoxin made by human gut commensal and extraintestinal pathogenic *Escherichia coli* strains and other Proteobacteria.[1] Colibactin-producing *E. coli* (*pks*$^+$ *E. coli*) cause DNA double strand breaks,[1-2] affect progression of colitis-associated colorectal cancer (CRC) in mouse models,[3-4] and are more frequently detected in patients with colorectal cancer.[5-7] Recent studies have indicated colibactin's genotoxicity likely arises from a direct interaction with DNA, as we and others have reported the accumulation of interstrand cross-links in human cell lines incubated with *pks*$^+$ *E. coli*.[8,9] Cell lines with impaired interstrand cross-link repair are also more sensitive to colibactin exposure.[8] However, the underlying chemical and enzymatic basis for colibactin's DNA cross-linking activity remains unclear.

Colibactin is produced by a 54-kb gene cluster (the *pks* island) encoding a nonribosomal peptide synthetase-polyketide synthase (NRPS-PKS) hybrid assembly line,[1] and its biosynthesis involves a self-protection mechanism in which an *N*-acyl-D-asparagine scaffold (prodrug motif) is initially assembled and elaborated to produce an inactive metabolite (precolibactin) (**Figure 1A**).[10-12] Precolibactin is cleaved by the periplasmic peptidase ClbP to release the active colibactin genotoxin. This active species has not yet been identified, isolated, or structurally characterized, presumably due to its instability. We and others have isolated and structurally characterized several stable candidate precolibactins (**1**-**5**) from *pks*$^+$ *E. coli* strains lacking ClbP (**Figure 1B**).[13-20] Notably, metabolites **2**-**5** contain a cyclopropane ring, a structural feature found in some DNA-alkylating agents.[21,22] Formation of an α,β-unsaturated imine after cleaving the prodrug motif in precolibactin likely enhances the electrophilicity of the cyclopropane toward DNA bases, and the recent identification and structural characterization of colibactin-derived DNA adducts in human cells and colonic epithelial cells provided support for this functional group serving as an electrophilic warhead.[9,23] However, it was unclear how this structural motif could lead to generation of an interstrand cross-link.

To address this question, we focused on understanding previously uncharacterized components of the colibactin biosynthetic enzymes. Efforts to elucidate the biosynthetic origins of isolated candidate precolibactins have led to putative functional assignments for all of the enzymes in the biosynthetic pathway except for the putative amidase ClbL. This enzyme is essential for the genotoxicity of *pks*$^+$ *E. coli*[1] but is not required for biosynthesis of any isolated candidate precolibactins reported to date, indicating that these metabolites were likely not precursors of the final genotoxin(s). A recent finding that *pks*$^+$ *E. coli* strains with inactive ClbL were incapable of cross-linking DNA suggested that this enzyme participates in producing the active cross-linking agent.[24] However, the molecular basis for ClbL's contribution to colibactin's cross-linking activity is unknown.

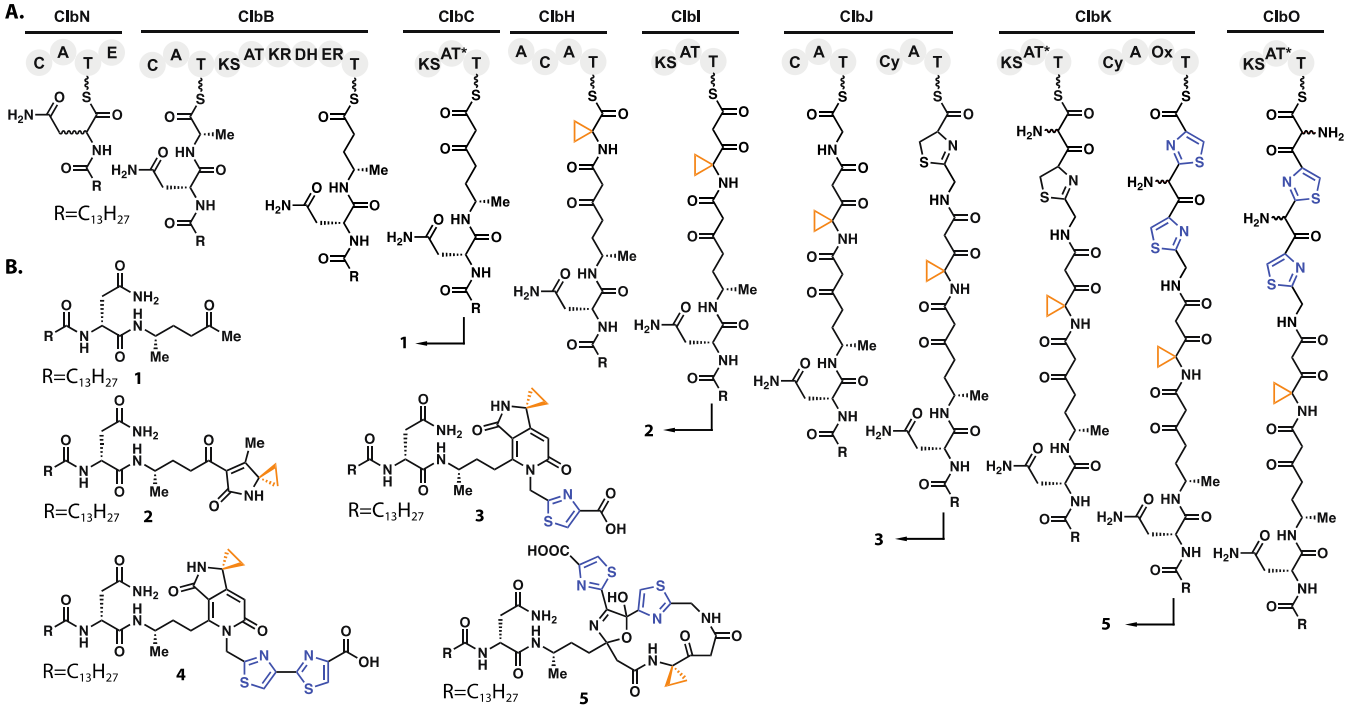

Figure 1. The *pks* enzymatic assembly line and the biosynthesis of characterized candidate precolibactins. (**A**) Hypothetical biosynthetic pathway for candidate precolibactins **1-3** and **5**. C, condensation; A, adenylation; T, thiolation; E, epimerization; KS, ketosynthase; AT, acyltransferase; KR, ketoreductase; DH, dehydratase; ER, enoyl reductase; AT*, atypical acyltransferase; Cy, cyclization; Ox, oxidase. Cyclopropane rings are highlighted in orange. Thiazoles are highlighted in blue. (**B**) Chemical structures of isolated and characterized candidate precolibactins **1-5**.

Here, we elucidate the function of ClbL in colibactin biosynthesis. By identifying, isolating, structurally characterizing a ClbL-dependent candidate precolibactin from *pks*$^+$ *E. coli* lacking ClbP and reconstituting its biosynthesis *in vitro*, we show that ClbL is an amide bond-forming enzyme. ClbL couples β-keto thioesters and aminoketones, structural motifs known to be generated by the colibactin assembly line. Based on this observation, we hypothesize that ClbL assembles a pseudo-dimeric precolibactin precursor possessing two electrophilic cyclopropane warheads which upon ClbP activation can cross-link DNA. This proposal is supported by the detection of two new colibactin-derived DNA adducts in *pks*$^+$ *E. coli*-treated DNA samples, including a putative cross-link. Overall, this work reveals the activity of ClbL and provides a molecular explanation for colibactin's DNA cross-linking activity.

Results

ClbL belongs to the amidase signature (AS) family of enzymes, which contain a Ser-*cis*Ser-Lys catalytic triad (**Figure S1**).[25] These enzymes catalyze a wide variety of hydrolytic reactions, including the breakdown of the neurotransmitter anandamide by fatty acid amide hydrolase (FAAH) in humans and hydrolysis of glutamine by Glu-tRNA^{Gln} amidotransferase subunit A (GatA) in bacteria.[26-27] Phylogenetic analysis of representative amidases indicated that ClbL is part of a distinct clade, potentially implying functional differences (**Figure S2**). Alignment of ClbL with structurally characterized amidases revealed the three conserved, essential catalytic residues (K80, S155, and S179) (**Figure S1**). To test whether these active site amino acids are essential for ClbL function and genotoxicity, plasmids expressing versions of ClbL with these residues individually mutated to alanine were constructed and used to complement a strain of *pks*$^+$ *E. coli* lacking *clbL* (BAC*pksΔclbL*). These mutant strains failed to cause cell cycle arrest in HeLa cells (**Figure S1**), suggesting that these three residues are crucial for the function of ClbL and confirming that ClbL's catalytic activity is critical for genotoxicity.

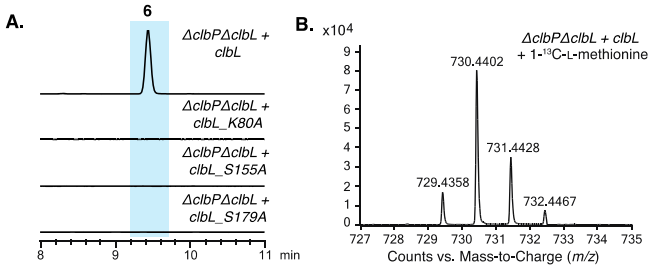

Figure 2. Discovery of a ClbL-dependent candidate precolibactin. (**A**) Extracted ion chromatograms (EICs) of metabolite **6** (*m/z* 729.4334) in the extracts of *E. coli* expressing wild-type or mutant ClbL. (**B**) LC-MS analysis of feeding of 1-^{13}C-L-methionine to ClbL expressing *E. coli* strains.

To identify candidate precolibactins linked to ClbL, we overexpressed wild-type ClbL or the ClbL active site mutants in *E. coli* DH10B BAC*pksΔclbP/ΔclbL*. Methanol extracts from these cultured strains were compared to identify mass features affected by the activity of ClbL. As amidases are typically amide bond-hydrolyzing enzymes, we expected the substrate for ClbL to accumulate in the mutant strains. Surprisingly, we were unable to detect any up-regulated metabolites in these strains. Consistent with the previous findings,[16] ClbL mutant strains produced detectable levels of **2** and other known candidate precolibactins (**Figure S3**), confirming that these metabolites are not ClbL-dependent. We did observe a novel metabolite (**6**, *m/z* 729.4334) accumulate only in strains expressing wild-type ClbL (**Figure 2A**, **Table S4**). To confirm metabolite **6** was *pks*-

associated, we fed wild-type ClbL-expressing strains L-[1-^{13}C]Met, the amino acid precursor of the cyclopropane ring.[15,16] Observation of a +1 mass shift demonstrated that methionine is incorporated into **6** (**Figure 2B**). Moreover, MS/MS fragmentation analysis indicated that **6** contains the prodrug scaffold (*m/z* 324.2437)[18] and a colibactin-characteristic fragment ion (*m/z* 231.1119)[17, 28] (**Figure S4**). These results suggested that metabolite **6** is a novel ClbL-dependent candidate precolibactin.

We then isolated and structurally characterized this new metabolite, obtaining approximately 0.7 mg of purified **6** from 140 L of an *E. coli* DH10B BAC*pksΔclbP* strain overexpressing ClbL. The high-resolution mass of this compound ([M+H]$^+$ = 729.4334) yielded a molecular formula of $C_{41}H_{56}N_6O_6$. One-dimensional (1D) and two-dimensional (2D) NMR along with MS/MS fragmentation analysis allowed us to elucidate its chemical structure (**Figure 3A**, **Figure S5-S8**, **Table S5**).

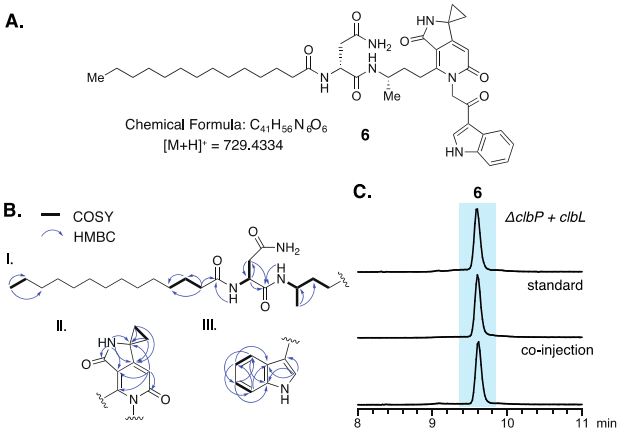

Figure 3. Structure determination of candidate precolibactin 6. (**A**) Chemical structure of metabolite **6**. (**B**) Key COSY and HMBC correlations that support this assignment. (**C**) EICs of metabolite **6** (*m/z* 729.4334) from the extracts of bacterial cultures, synthetic standard, and co-injection of synthetic standard and bacterial culture extracts.

Similarities between the NMR spectra of known precolibactins and metabolite **6** enabled us to assign fragments I and II easily (**Figure 3B**). Unexpectedly, COSY and HMBC correlations indicated the presence of an indole ring (**Figure 3B**). We were unable to assign the precise locations of a methylene and carbonyl within **6** using NMR, but MS/MS fragmentation analysis revealed a loss of indole (*m/z* 117.0607) from the parent ion, suggesting that the ketone is directly attached to the indole ring (**Figure S9**). To further support this structural assignment, we chemically synthesized a standard of metabolite **6**. This standard possessed the same exact mass, MS/MS fragmentation pattern, retention time, ^{1}H spectrum, and UV spectrum as **6** isolated from bacterial strains (**Figure 3C**, **Figure S10-S12**), confirming our structural assignment.

The chemical structure of **6** guided subsequent efforts to elucidate the function of ClbL. Metabolite **6** does not contain either a free amine or a carboxylic acid, which are the functional groups liberated by the activity of characterized amidases. Based on the known steps of colibactin biosynthesis, we envisioned two possible pathways for the formation of **6** (**Figure 4A**). In pathway A, the intermediate generated by ClbJ-NPRS1[17,18] could be intercepted and offloaded via nucleophilic attack by the known *E. coli* metabolite indole.[29] In pathway B, 2-amino-1-(1*H*-indol-3-yl)ethan-1-one (**7**) could act as a nucleophile to offload the β-keto thioester intermediate generated by the upstream biosynthetic enzyme ClbI.[30]

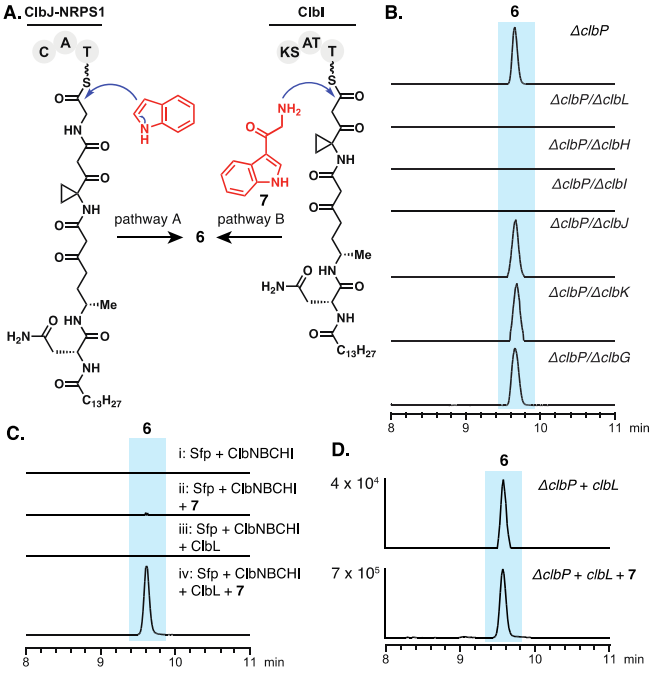

Figure 4. Investigating the formation of candidate precolibactin 6. (**A**) Two proposed pathways could generate **6**. (**B**) LC-MS analysis of extracts from double mutants of *E. coli* DH10B BAC*pks*. (**C**) LC-MS analysis of *in vitro* reconstitution assays. (**D**) LC-MS analysis of cell pellet extracts from *E. coli* DH10B BAC*pksΔclbP* + pTrc-*clbL* supplemented with **7**. Panels **B-D** show the EICs of **6** (*m/z* 729.4334).

To test the involvement of pathway A, we made multiple double mutants in *E. coli* DH10B BAC*pksΔclbP* (**Figure 4B**). Detection of **6** in the absence of ClbJ and ClbP indicated that ClbJ was not required for its biosynthesis. To confirm this result, we fed [1,2-$^{13}C_2$]Gly, the building block used by ClbJ-NPRS1, to a glycine auxotrophic BAC*pksΔclbP* strain. We observed a +2 mass shift in the known glycine-derived precolibactin **3**, but not in **6** (**Figure S13**), confirming that pathway A is not the source of **6**.

We next explored the involvement of pathway B by reconstituting the biosynthesis of metabolite **6** *in vitro*. We cloned, overexpressed, and purified C-terminal His$_6$-tagged ClbL (**Figure S14**). An assay mixture containing Sfp, ClbN, ClbB, ClbC, ClbH, ClbI, ClbL, myristoyl-CoA, malonyl-CoA, L-Asn, L-Ala, NADPH, ATP, SAM, and **7** generated **6** (**Figure 4C**, trace iv). To further confirm the involvement of **7** in this process, **7** was fed to an *E. coli* DH10B BAC*pksΔclbP* strain overexpressing ClbL. An approximately 20-fold increase in the production of **6** was observed (ion count of 4 x 10^4 vs. 7 x 10^5, **Figure 4D**), consistent with pathway B being operative *in vivo*. Notably, these results suggested that ClbL catalyzes amide bond formation rather than amide bond hydrolysis.

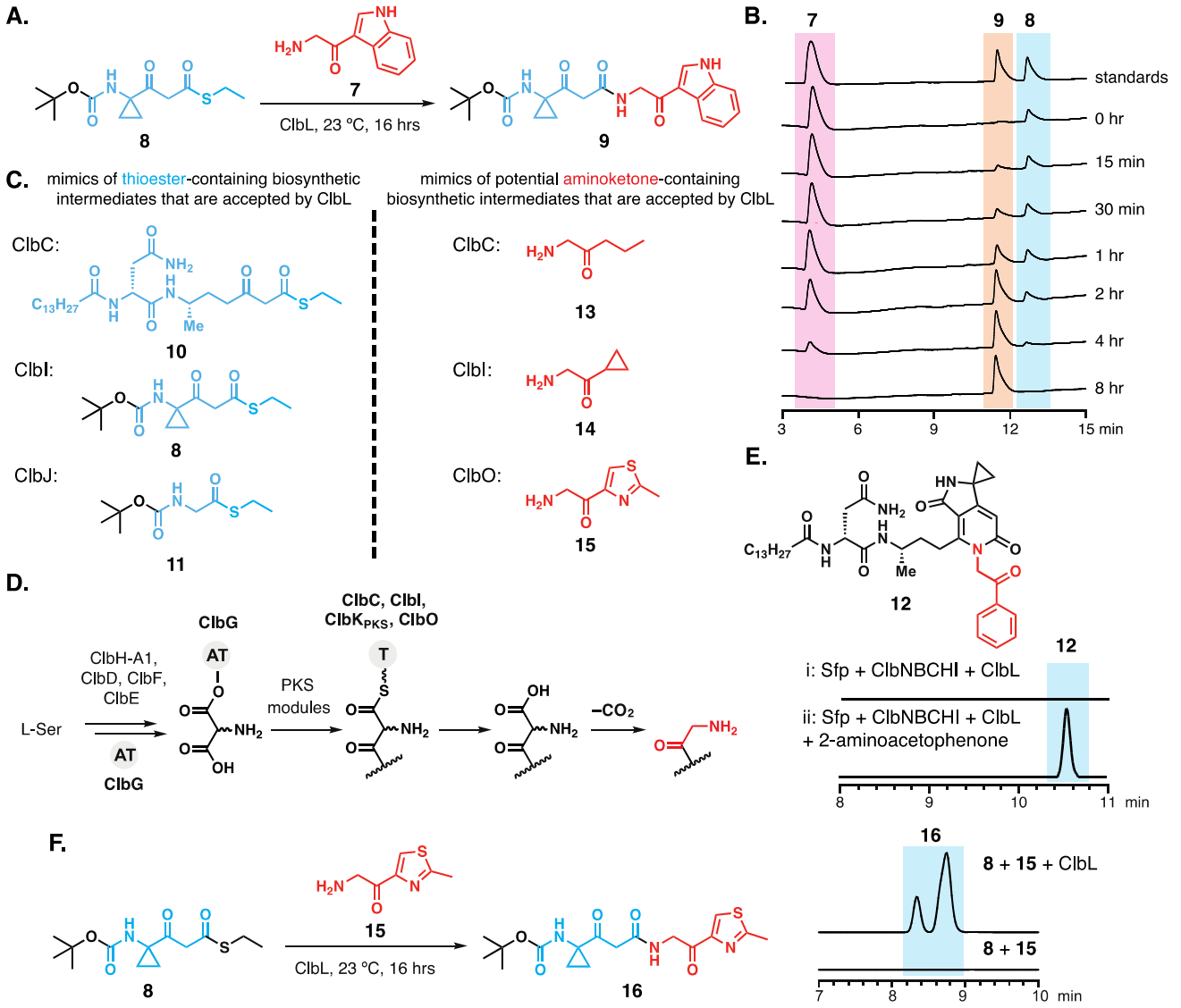

Figure 5. *In vitro* characterization of ClbL reveals its role in amide formation. (**A**) Reaction scheme of an *in vitro* assay to form **9**. (**B**) HPLC time course for the formation of **9** by ClbL (monitored at 250 nm). (**C**) Thioesters and α-aminoketones accepted by ClbL. (**D**) The *pks* assembly line may produce α-aminoketone nucleophiles. (**E**) EICs of **12** (*m/z* 690.4225) formed in an *in vitro* reconstitution assay. (**F**) ClbL catalyzes the formation of **16** *in vitro*. The two peaks likely correspond to two tautomers of the β-ketoamide product.

To further characterize the activity of ClbL *in vitro*, we synthesized ethyl thioester **8** as a surrogate for the ClbI-tethered biosynthetic intermediate (**Figure 1A** and **5A**). Incubating ClbL with **7** and **8** (**Figure 5A**) revealed a single new product peak accumulating during an HPLC time course experiment (**Figure 5B**). We confirmed this product was amide **9** by comparison to a synthetic standard (**Figure S15**).

To evaluate ClbL's preference for potential thioester substrates, we designed and synthesized ethyl thioester mimics of assembly line-tethered colibactin biosynthetic intermediates (**Figure S16**). ClbL and **7** were incubated with these mimics individually. In addition to **8**, we found that thioesters **10** (a mimic for the ClbC-bound intermediate) and **11** (a mimic for the ClbJ-bound intermediate) could be recognized by ClbL (**Figure 5C**, **Figure S17**). To assess ClbL's ability to discriminate among these substrates, we conducted a competition experiment. Mixing **8** and **11**, which are predicted to have very similar hydrophobicities, in equal molar amounts with ClbL and **7** produced more of the product arising from **8** compared to the product arising from **11** (**Figure S17**). This suggests ClbL preferentially recognizes the ClbI-bound intermediate. Furthermore, reexamination of extracts from ClbL-overexpressing strains did not reveal products of amide bond formation between **7** and ClbC- or ClbJ-tethered thioester intermediates (**Figure S17**). Together, these data indicate that ClbL prefers the ClbI-bound thioester as an electrophile for amide bond formation.

We next investigated the relevance of indole-containing nucleophile **7** to colibactin biosynthesis. **7** has only previously been found in termitomycamide B, which was isolated from the giant mushroom *Termitomyces titanicus* (**Figure S18**).[31] Intriguingly, **7** contains an α-aminoketone, a structural motif that could originate from decarboxylation of the essential colibactin building block aminomalonyl-ACP (**Figure 5D**). To examine the involvement of aminomalonyl-ACP biosynthetic enzymes in the production of **7**, we used an *E. coli* DH10B BAC*pksΔclbPΔclbG* strain, in which the transfer of aminomalonate to assembly line enzymes is abolished. This

strain still produced **6,** the candidate precolibactin derived from **7**, indicating that aminomalonyl-ACP is not required (**Figure 4B**). We further showed that **7** does not originate from any of the *pks* island enzymes as it was detected in *E. coli* DH10B (**Figure S19**). **7** was also absent from uninoculated growth medium, indicating it is produced in *E. coli* endogenously. Therefore, we questioned whether the incorporation of **7** was relevant for the biosynthesis of the active colibactin genotoxin. Supplementing **7** to *pks*$^+$ *E. coli* during infection of host cells did not boost genotoxicity (**Figure S20**), suggesting that **7** is likely not a relevant substrate for ClbL in colibactin biosynthesis. We therefore hypothesize that metabolite **6** arises from the promiscuous activity of ClbL and is not an on-pathway intermediate.

Next, we considered the possibility that an alternative nucleophile may be generated by colibactin biosynthetic enzymes. Assays with various amine substrates showed that the α-aminoketone of **7** is crucial for recognition by ClbL. No products were observed when tryptamine replaced **7** in the *in vitro* reconstitution of the biosynthesis of **6**. In contrast, we identified a new product **12** (*m/z* 690.4225) in assay mixtures containing 2-aminoacetophenone in place of **7** (**Figure 5E**). This result suggested the α-aminoketone and not the indole ring is important for substrate recognition by ClbL.

To identify the true nucleophilic coupling partner for ClbL, we examined its reactivity toward mimics of potential α-aminoketone substrates derived from the colibactin assembly line. We previously identified four PKS assembly line enzymes (ClbC, ClbI, ClbK, and ClbO) that accept aminomalonate *in vitro*.[18] Isolation of candidate precolibactin **5** also showed that the PKS module of ClbK accepts this building block *in vivo*, generating an α-aminoketone that is further oxidized to an α-iminoketone.[19] Once offloaded from PKS modules, the immediate products of extension with aminomalonyl-ACP should readily undergo decarboxylation to give rise to α-aminoketones resembling **7** (**Figure 5D**). To examine whether ClbL can use assembly line derived α-aminoketones for amide bond formation, we synthesized mimics of these potential substrates (**13-15**) and incubated them with ClbL and preferred thioester **8** (**Figure S21**). While ClbL can use all three mimics, it preferentially uses **15**, which mimics the α-aminoketone derived from ClbO, the final module of the colibactin assembly line (**Figure 5C** and **5F, Figure S22**). Together, these results suggest a potential role for ClbL in colibactin biosynthesis: forming an amide bond between a ClbI-bound thioester and the α-aminoketone derived from ClbO (**Figure 6**).

This proposed assembly line logic has important implications for colibactin's biological activity. Amide bond formation by ClbL could link the precursors of two electrophilic cyclopropane warheads together. Cleavage of the two prodrug motifs by ClbP would generate a species that would contain two electrophilic warheads and could be capable of generating interstrand cross-links (**17**) (**Figure 6**). This metabolite could be the active colibactin genotoxin. This scenario is consistent with previous observations that *pks*$^+$ *E. coli* promote interstrand cross-link formation in linearized plasmid DNA and human cell lines,[8,9] as well as a recent finding that a synthetic 'colibactin mimic' containing two cyclopropane rings cross-links DNA *in vitro*, while a monomer cannot.[32] This proposal is also consistent with our current understanding of colibactin-mediated DNA damage. We previously identified and characterized a pair of diasteromeric colibactin-derived DNA adducts (**18**) that likely arise from degradation of a larger lesion.[9] Depurination of the proposed colibactin-DNA cross-link **19** would give adduct **20**, which could undergo oxidative cleavage to give monoadducts **21** and **18** (**Figure 7A**). A similar cleavage has been shown to occur readily in model substrates and a synthetic colibactin mimic.[33,34]

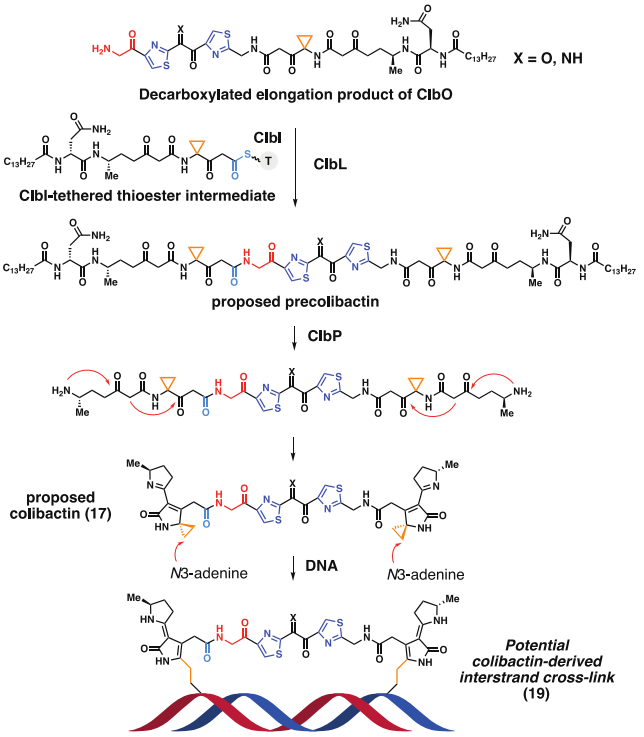

Figure 6. A model for ClbL-catalyzed formation of the active colibactin genotoxin). Key structural features colored as in Figures 1, 4, and 5.

To test this proposal, we used targeted selection ion monitoring (SIM), untargeted adductomics,[9] and MS^2 analyses to identify and support the structures of these proposed colibactin-derived DNA adducts in linearized plasmid DNA treated with *pks*$^+$ *E. coli*. We identified a large DNA adduct consistent with cross-link **20** (*m/z* 537.1719, *z* = 2) (**Figure 7B**). We also used untargeted DNA adductomics[9] to find a cross-link adduct lacking one of the adenine moieties **22** (*m/z* 938.2821, *z* = 1; *m/z* 469.6447, *z* = 2) (**Figure S23**). To confirm **20** and **22** are *pks*-associated, we supplemented assay mixtures with either L-[1-^{13}C]Met or L-[1-^{13}C]Cys. The observation of +2 mass shifts in **20** and **22** suggested that they derive from two methionines and two cysteines (**Figure 7C, Figure S23-24**), supporting our proposal that colibactin contains two electrophilic warheads and two thiazoles. We also detected the proposed decomposition product **21** (*m/z* 568.1721) (**Figure 7D-E, Figure S25-27**). Further characterization of colibactin-derived adducts will be necessary to clarify the identity of the active cross-linking agent.

Conclusions

ClbL belongs to the AS enzyme family, which participates in numerous biological processes, including the biosynthesis of the natural product actinonin in *Streptomyces* sp. ATCC 14903.[35] In addition to catalyzing amide bond hydrolysis, some amidases exhibit acyltransferase activity. For example, the amidase from *Rhodococcus sp.* R312 transfers the acyl group of an amide to

hydroxylamine at a 33-fold faster rate than the hydrolysis of the same amide, an activity that could be explained by the stronger nucleophilicity of hydroxylamine toward the acyl-enzyme intermediate compared to water.[36] Our discovery of ClbL's activity expands the known roles of amidases to include interfacing with enzymatic assembly lines. It is intriguing that ClbL preferentially recognizes α-aminoketones over the corresponding primary amines. Future crystallographic studies may shed further light on this specificity.

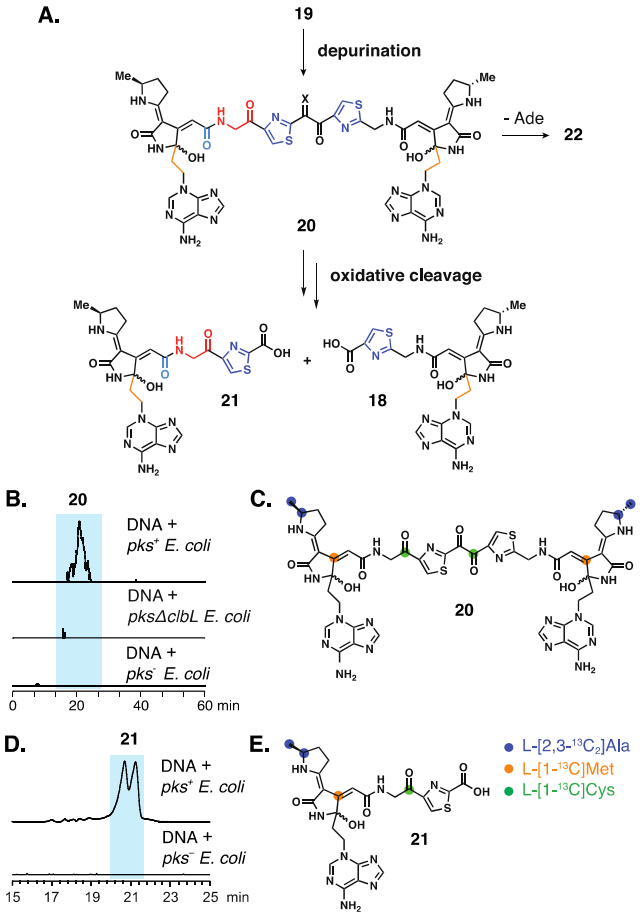

Figure 7. Detection of two new colibactin-derived DNA adducts. (**A**) Predicted structures of colibactin-derived DNA adduct **20** and its decomposition products **21** and **18**. (**B**) EICs of **20** (*m/z* 537.1719, *z* = 2) in *pks*$^+$ *E. coli* treated DNA samples. (**C**) Adduct **20** incorporates two alanines, two methionines, and two cysteines. (**D**) EICs of **21** (*m/z* 568.1721) in DNA samples treated with *pks*$^+$ *E. coli*. The two peaks likely arise from multiple diastereomers. (**E**) Adduct **21** incorporates one alanine, one methionine, and one cysteine.

We uncovered the activity of ClbL through studying the biosynthesis of metabolite **6**, an off-pathway candidate precolibactin that incorporates an endogenously produced α-aminoketone (**7**). A competition assay with **7** and **13-15** revealed that **7** is a preferred substrate for ClbL (**Figure S22**), explaining the presence of **6** in *E. coli* strains expressing the entire *pks* assembly line. However, we cannot conclude whether ClbL is likely to preferentially recognize **7** *in vivo* because we cannot examine its reactivity toward complex assembly line derived α-aminoketones. These proposed substrates are not readily accessed using chemical synthesis or *in vitro* reconstitution because of the instability of the α-aminoketone located between the two thiazole heterocycles.[33,34] Consistent with its instability, we were unable to detect a ClbO-derived α-aminoketone *in vivo*. Attempts to generate this substrate through *in vitro* reconstitution were also unsuccessful.

Nonetheless, ClbL's *in vitro* activity supports the hypothesis that this enzyme links two intermediates generated by the colibactin NRPS-PKS assembly line, a ClbI-bound thioester and a ClbO-derived α-aminoketone (**Figure 6**). This transformation may play a critical role in generating a DNA cross-linking agent that could represent the active colibactin genotoxin. This proposal is supported by ClbL's essential role in genotoxicity and cross-link formation, as well as our detection of adduct **20** in linearized plasmid DNA treated with *pks*$^+$ *E. coli*. We cannot exclude the possibility that ClbL accepts additional substrates. For example, we were unable to test whether ClbL could use a mimic of the putative ClbO-bound aminomalonate thioester intermediate as an electrophile due to its intrinsic instability. However, precedented colibactin biosynthetic logic and its DNA interstrand cross-linking activity allow us to establish two minimal criteria for the structure of colibactin: 1) it contains two electrophilic cyclopropane rings and 2) its construction requires all of the essential colibactin biosynthetic enzymes. Application of these conditions greatly narrows the potential structures for the active colibactin genotoxin(s) (**Figure S28**). These criteria also help in assessing the biological relevance of other reported candidate precolibactins. For example, Qian and Zhang recently disclosed a candidate precolibactin they proposed is the precursor to colibactin (**Figure S29**).[20] Though the corresponding ClbP cleavage product appeared to damage DNA *in vitro* and in cells in a Cu^{2+}-dependent manner, it did not form cross-links. It is also unclear whether ClbL is required for its biosynthesis. Together, these observations suggest this metabolite may not be a major contributor to genotoxicity.

While this work was in revision, the Herzon and Crawford groups independently identified colibactin-derived cross-link adduct **20** providing further support for our model.[34] They also chemically synthesized diketone **17** and showed that it cross-links linearized plasmid DNA at a concentration of 10 μM. This relatively low cross-linking reactivity suggests the active genotoxin(s) may contain an α-iminoketone motif. As our results suggest ClbL may be able to construct additional candidate genotoxins (**Figure S28**), we feel that the final structure(s) of colibactin still requires further investigation.

In summary, these studies provide a biosynthetic rationale for the formation of interstrand DNA cross-links by the gut bacterial genotoxin colibactin. As such cross-links represent the most toxic form of DNA damage in human cells, it is possible this activity could explain the effects of *pks*$^+$ *E. coli* on tumor development. Future efforts to characterize additional precolibactins, colibactins, and colibactin-DNA adducts will provide us with the additional molecular information needed to decipher how colibactin influences CRC development in patients.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Materials and Methods for enzyme purification, synthetic procedures, and additional experimental data (PDF)

AUTHOR INFORMATION

Corresponding Author

* balskus@chemistry.harvard.edu

Present Addresses

§Vertex Pharmaceuticals, 50 Northern Ave, Boston, MA 02210, United States.

‖ Boston Children's Hospital, 300 Longwood Avenue, Boston, MA 02115, United States.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

We thank Alex Sieg for help with cloning, Lihan Zhang for help with NMR experiments, and Matthew Volpe for comments on the manuscript. We acknowledge financial support from National Institutes of Health (R01CA208834-02) (E.P.B). Salary support for P.W.V. was provided by the U.S. National Institutes of Health and National Cancer Institute [Grant R50-CA211256]. Mass spectrometry for DNA adduct analysis was carried out in the Analytical Biochemistry Shared Resource of the Masonic Cancer Center, supported in part by the U.S. National Institutes of Health and National Cancer Institute [Cancer Center Support Grant CA-77598]. M.R.W. acknowledges support from the American Cancer Society-New England Division Postdoctoral Fellowship PF-16-122-01-CDD.

REFERENCES

(1) Nougayrède, J.-P.; Homburg, S.; Taieb, F.; Boury, Brzuszkiewicz, E.; Gottschalk, G.; Buchrieser, C.; Hacker, J.; Dobrindt, U.; Oswald, E. *Escherichia coli* induces DNA double-strand breaks in eukaryotic cells. *Science* **2006**, *313*, 848–851.

(2) Cuevas-Ramos, G.; Petit, C. R.; Marcq, I.; Boury, M.; Oswald, E.; Nougayrède, J.-P. *Escherichia coli* induces DNA damage *in vivo* and triggers genomic instability in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 11537–11542.

(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. Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science* **2012**, *338*, 120–123.

(4) Arthur, J. C.; Gharaibeh, R. Z.; Mühlbauer, M.; Perez-Chanona, E.; Uronis, J. M.; McCafferty, J.; Fodor, A. A.; Jobin, C. Microbial genomic analysis reveals the essential role of inflammation in bacteria-induced colorectal cancer. *Nat. Commun.* **2014**, *5*, 4724.

(5) Buc, E.; Dubois, D.; Sauvanet, P.; Raisch, J.; Delmas, J.; Darfeuille-Michaud, A.; Pezet, D.; Bonnet, R. High prevalence of mucosa-associated *E. coli* producing cyclomodulin and genotoxin in colon cancer. *PLoS ONE* **2013**, *8*, e56964.

(6) Eklof, V.; Löfgren-Burström, A.; Zingmark, C.; Edin, S.; Larsson, P.; Karling, P.; Alexeyev, O.; Rutegård, J.; Wikberg, M. L.; Palmqvist, R. Cancer-associated fecal microbial markers in colorectal cancer detection. *Int. J. Cancer*. **2017**, *141*, 2528–2536.

(7) Dejea, C. M.; Fathi, P.; Craig, J. M.; Boleij, A.; Taddese, R.; Geis, A. L.; Wu, X.; DeStefano Shields, C. E.; Hechenbleikner, E. M.; Huso, D. L.; Anders, R. A.; Giardiello, F. M.; Wick, E. C.; Wang, H.; Wu, S.; Pardoll, D. M.; Housseau, F.; Sears, C. L. Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. *Science* **2018**, *359*, 592–597.

(8) Bossuet-Greif, N.; Vignard, J.; Taieb, F.; Mirey, G.; Dubois, D.; Petit, C.; Oswald, E.; Nougayrède J.-P. The colibactin genotoxin generates DNA interstrand cross-links in infected cells. *mBio* **2018**, *9*, e02393–17.

(9) Wilson, M. R.; Jiang, Y.; Villalta, P. W.; Stornetta, A.; Boudreau, P. D.; Carrá, A.; Brennan, C. A.; Chun, E.; Ngo, L.; Samson, L. D.; Engelward, B. P.; Garrett, W. S.; Balbo, S.; Balskus, E. P. The human gut bacterial genotoxin colibactin alkylates DNA. *Science* **2019**, *363*, eaar7785.

(10) Dubois, D.; Baron, O.; Cougnoux, A.; Delmas, J.; Pradel, N.; Boury, M.; Bouchon, B.; Bringer, M. A.; Nougayrède, J.-P.; Oswald, E.; Bonnet, R. ClbP is a prototype of a peptidase subgroup involved in biosynthesis of nonribosomal peptides. *J. Biol. Chem*. **2011**, *286*, 35562–35570.

(11) Brotherton, C. A.; Balskus, E. P. A prodrug resistance mechanism is involved in colibactin biosynthesis and cytotoxicity. *J. Am. Chem. Soc*. **2013**, *135*, 3359–3362.

(12) Bian, X.; Fu, J.; Plaza, A.; Herrmann, J.; Pistorius, D.; Stewart, A. F.; Zhang, Y.; Müller, R. *In vivo* evidence for a prodrug activation mechanism during colibactin maturation. *ChemBioChem* **2013**, *14*, 1194–1197.

(13) Vizcaino, M. I.; Engel, P.; Trautman, E.; Crawford, J. M. Comparative metabolomics and structural characterizations illuminate colibactin pathway-dependent small molecules. *J. Am. Chem. Soc*. **2014**, *136*, 9244–9247.

(14) Brotherton, C. A.; Wilson, M.; Byrd, G.; Balskus, E. P. Isolation of a metabolite from the *pks* island provides insights into colibactin biosynthesis and activity. *Org. Lett.* **2015**, *17*, 1545–1548.

(15) Vizcaino, M. I.; Crawford, J. M. The colibactin warhead crosslinks DNA. *Nat. Chem*. **2015**, *7*, 411–417.

(16) Bian, X. Y.; Plaza, A.; Zhang, Y. M.; Müller, R. Two more pieces of the colibactin genotoxin puzzle from *Escherichia coli* show incorporation of an unusual 1-aminocyclopropanecarboxylic acid moiety. *Chem. Sci*. **2015**, *6*, 3154–3160.

(17) Li, Z. R.; Li, Y.; Lai, J. Y.; Tang, J.; Wang, B.; Lu, L.; Zhu, G.; Wu, X.; Xu, Y.; Qian, P. Y. Critical intermediates reveal new biosynthetic events in the enigmatic colibactin pathway. *ChemBioChem* **2015**, *16*, 1715–1719.

(18) Zha, L.; Wilson, M. R.; Brotherton, C. A.; Balskus, E. P. Characterization of polyketide synthase machinery from the *pks* island facilitates isolation of a candidate precolibactin. *ACS Chem. Biol*. **2016**, *11*, 1287–1295.

(19) Li, Z. R.; Li, J.; Gu, J. P.; Lai, J. Y. H.; Duggan, B. M.; Zhang, W. P.; Li, Z. L.; Li, Y. X.; Tong, R. B.; Xu, Y.; Lin, D. H.; Moore, B. S.; Qian, P. Y. Divergent biosynthesis yields a cytotoxic aminomalonate-containing precolibactin. *Nat. Chem. Biol*. **2016**, *12*, 773–775.

(20) Li, Z. R.; Li, J.; Cai, W.; Lai, J.; McKinnie, S.; Zhang, W. P.; Moore, B.; Zhang, W.; Qian, P. Y. Macrocyclic colibactin induces DNA double-strand breaks via copper-mediated oxidative cleavage. *bioRxiv* **2019**, doi: 10.1101/530204.

(21) Tanasova, M.; Sturla, S. J. Chemistry and biology of acylfulvenes: sesquiterpene-derived antitumor agents. *Chem. Rev*. **2012**, *112*, 3578-3610.

(22) Ghosh, N.; Sheldrake, H. M.; Searcey, M.; Pors, K. Chemical and biological explorations of the family of CC-1065 and the duocarmycin natural products. *Curr. Top. Med. Chem*. **2009**, *9*, 1494-1524.

(23) Xue, M.; Shine, E.; Wang, W.; Crawford, J. M.; Herzon, S. B. Characterization of natural colibactin-nucleobase adducts by tandem mass spectrometry and isotopic labeling. Support for DNA alkylation by cyclopropane ring opening. *Biochemistry* **2018**, *57*, 6391–6394.

(24) Shine, E. E.; Xue, M.; Patel, J. R.; Healy, A. R.; Surovtseva, Y. V.; Herzon, S. B.; Crawford, J. M. Model colibactins exhibit human cell genotoxicity in the absence of host bacteria. *ACS Chem. Biol.* **2018**, *13*, 3286–3293.

(25) Shin, S.; Lee, T. H.; Ha, N. C.; Koo, H. M.; Kim, S. Y.; Lee, H. S.; Kim, Y. S.; Oh, B. H. Structure of malonamidase E2 reveals a novel Ser-*cis*Ser-Lys catalytic triad in a new serine hydrolase fold that is prevalent in nature. *EMBO J*. **2002**, *21*, 2509–2516.

(26) Cravatt, B. F.; Giang, D. K.; Mayfield, S. P.; Boger, D. L.; Lerner, R. A.; Gilula, N. B. Molecular characterization of an enzyme that degrades neuromodulatory fatty-acid amides. *Nature* **1996**, *384*, 83–87.

(27) Curnow, A. W.; Hong, K. W.; Yuan, R.; Kim, S. I.; Martins, O.; Winkler, W.; Henkin, T. M.; Söll, D. Glu-tRNA^{Gln} amidotransferase: a novel heterotrimeric enzyme required for correct decoding of glutamine codons during translation. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 11819–11826.

(28) Trautman, E. P.; Healy, A. R.; Shine, E. E.; Herzon, S. B.; Crawford, J. M. Domain-targeted metabolomics delineates the heterocycle assembly steps of colibactin biosynthesis. *J. Am. Chem. Soc.* **2017**, *139*, 4195–4201.

(29) Li, G.; Young, K. D. Indole production by the tryptophanase TnaA in *Escherichia coli* is determined by the amount of exogenous tryptophan. *Microbiology* **2013**, *159*, 402-410.

(30) Zha, L.; Jiang, Y.; Henke, M. T.; Wilson, M. R.; Wang, J. X.; Kelleher, N. L.; Balskus, E. P. Colibactin assembly line enzymes use *S*-adenosylmethionine to build a cyclopropane ring. *Nat. Chem. Biol.* **2017**, *13*, 1063–1065.

(31) Choi, J. H.; Maeda, K.; Nagai, K.; Harada, E.; Kawade, M.; Hirai, H.; Kawagishi, H. Termitomycamides A to E, fatty acid amides isolated from the mushroom *Termitomyces titanicus*, suppress endoplasmic reticulum stress. *Org. Lett.* **2010**, *12*, 5012–5015.

(32) Healy, A. R.; Nikolayevskiy, H.; Patel, J. R.; Crawford, J. M.; Herzon, S. B. A mechanistic model for colibactin-induced genotoxicity. *J. Am. Chem. Soc.* **2016**, *138*, 15563–15570.

(33) Healy, A. R.; Wernke, K. M.; Kim, C. S.; Lees, N. R.; Crawford, J. M.; Herzon, S. B. Synthesis and reactivity of precolibactin 886. *ChemRxiv* **2019**, doi: 10.26434/chemrxiv.7849151.v1.

(34) Xue, M.; Kim, C. S.; Healy, A. R.; Wernke, K. M.; Wang, Z.; Frischling, M. C.; Shine, E. E.; Wang, W.; Herzon, S. B.; Crawford, J. M. Structure elucidation of colibactin. *bioRxiv* **2019**, doi: 10.1101/574053.

(35) Wolf, F.; Leipoldt, F.; Kulik, A.; Wibberg, D.; Kalinowski, J.; Kaysser, L. Characterization of the actinonin biosynthetic gene cluster. *ChemBioChem* **2018**, *19*, 1189–1195.

(36) Thiéry, A.; Maestracci, M.; Arnaud, A.; Galzy, P. Acyltransferase activity of the wide spectrum amidase of *Brevibacterium* sp. R312. *J. Gen. Microbiol.* **1986**, *132*, 2205–2208.

TOC graphic: