Published in final edited form as:

J Am Chem Soc. 2017 October 25; 139(42): 14817–14824. doi:10.1021/jacs.7b07807.

Molecular Basis of Gut Microbiome-Associated Colorectal Cancer: A Synthetic Perspective

Alan R. Healy^{†,‡}, Seth B. Herzon^{*,†,§}

[†]Department of Chemistry, Yale University, New Haven, Connecticut 06520, United States

[‡]Chemical Biology Institute, Yale University, West Haven, Connecticut 06516, United States

[§]Department of Pharmacology, Yale School of Medicine, New Haven, Connecticut 06520, United States

Abstract

A significant challenge toward studies of the human microbiota involves establishing causal links between bacterial metabolites and human health and disease states. Certain strains of commensal *Escherichia coli* harbor the 54-kb *clb* gene cluster which codes for small molecules named precolibactins and colibactins. Several studies suggest colibactins are genotoxins and support a role for *clb* metabolites in colorectal cancer formation. Significant advances toward elucidating the structures and biosynthesis of the precolibactins and colibactins have been made using genetic approaches, but their full structures remain unknown. In this Perspective we describe recent synthetic efforts that have leveraged biosynthetic advances and shed light on the mechanism of action of *clb* metabolites. These studies indicate that deletion of the colibactin peptidase ClbP, a modification introduced to promote accumulation of precolibactins, leads to the production of non-genotoxic pyridone-based isolates derived from the diversion of linear biosynthetic intermediates toward alternative cyclization pathways. Furthermore, these studies suggest the active genotoxins (colibactins) are unsaturated imines that are potent DNA damaging agents, thereby confirming an earlier mechanism of action hypothesis. Although these imines have very recently been detected in bacterial extracts, they have to date confounded isolation. As the power of “meta-omics” approaches to natural products discovery further advance, we anticipate that chemical synthetic and biosynthetic studies will become increasingly interdependent.

INTRODUCTION

The genomes of some strains of gut-commensal and extraintestinal pathogenic *Escherichia coli* (*E. coli*) contain a 54-kb non-ribosomal peptide synthetase–polyketide synthase (NRPS–PKS) gene cluster—referred to as *clb*—that codes for the enzymes responsible for the biosynthesis of metabolites named precolibactins and colibactins. Direct isolation of *clb* metabolites from wild-type *clb*⁺ *E. coli* has been challenging; consequently much of our understanding of the structures of these metabolites has depended upon the use of genetically modified strains, in vitro enzymology, and bioinformatics. Although significant

*Corresponding Author seth.herzon@yale.edu.

The authors declare no competing financial interest.

progress has been made, the full structures of the precolibactins and colibactins that account for all of the genes in the *clb* cluster remain unknown. The biosynthesis and structures of established *clb* metabolites have been reviewed.¹

Several recent studies implicate this *clb* cluster in colorectal cancer (CRC) progression. Two studies found that *clb*⁺ *E. coli* were more common in colorectal cancer patients (55 or 67% of CRC patients' gut microflora had the gene cluster) than in cancer-free subjects (19 or 21% of healthy subjects had the cluster).^{2,3} A third study found that colonization of the intestinal mucosa by *clb*⁺ *E. coli* was more frequent in patients with stage III or IV CRC than those with stage I CRC.⁴ In animal models, *clb*⁺ *E. coli* promoted colorectal cancer formation in multiple intestinal neoplasia (MIN) mice,⁴ in interleukin-10 (IL-10)-deficient MIN mice,⁵ and in IL-10-deficient mice treated with azoxymethane.² Eukaryotic cells colonized with *clb*⁺ *E. coli* underwent senescence and upregulated hepatocyte growth factor⁶ (HGF; dysregulation of HGF is associated with CRC tumor growth, scattering, and metastasis).⁷ On the other hand, a recent study in Japan did not find an increase in *clb* genes in CRC patients (43% and 46% of CRC and healthy patients, respectively, possessed *clb* genes).⁸ However, this last study isolated bacterial DNA by colonic lavage followed by PCR amplification, as opposed to the direct isolation and analysis of mucosa-associated *E. coli* from healthy and CRC patients used in refs 2 and 3.

Mechanistic studies suggest colibactins (but not precolibactins, vide infra) are genotoxins that induce highly toxic double-strand breaks (DSBs) in DNA. Low levels of DNA DSBs create genomic instability and can lead to tumorigenesis.⁹ HeLa cells infected with *clb*⁺ *E. coli* at a multiplicity of infection (MOI, the ratio of bacteria to eukaryotic cells) equal to 100 suffered megalocytosis, persistent DNA DSBs, and G2/M arrest.¹⁰ *clb*⁺ *E. coli* induced DNA DSBs in a mouse intestinal loop model.¹¹ Consistent with DSB-associated cytotoxicity, cells deficient in the DNA DSB repair factor X-ray repair cross complementing 5 (Ku80) were sensitized to *clb*⁺ *E. coli*.¹¹ At lower MOIs (1–20), Chinese hamster ovary cells accumulated DNA damage and proliferated with increased mutational frequency and transformed phenotype.¹¹ Finally, the DNA damage induced by infection of a mouse intestinal loop model with 3×10^9 wild-type *clb*⁺ *E. coli* for 6 h was estimated as equivalent to ~0.5 Gy of ionizing radiation,¹¹ which is comparable to 5000 chest X-ray procedures.¹²

KEY STEPS IN COLIBACTIN TRANSPORT AND ACTIVATION

The links between the *clb* cluster and CRC have motivated extensive efforts to elucidate the biosynthesis, structures, and mechanism of action of precolibactins and colibactins. The biosynthesis of (pre)colibactins has been reviewed,¹ and we focus here only on their transport and activation, which are essential to the work discussed below. Precolibactins are assembled by the NRPS–PKS assembly line, transported to the periplasm by the 12-transmembrane MATE transporter ClbM,¹³ and deacylated by colibactin peptidase (ClbP), a pathway-dedicated serine protease anchored within the inner bacterial periplasmic membrane (Figure 1).¹⁴ This deacylation is required for cytopathic effects;¹⁰ it converts precolibactins to colibactins and likely constitutes a prodrug resistance mechanism in the bacteria.¹⁵ The colibactins are then trafficked from the periplasm to the target cells by a pathway that remains undefined.

STRUCTURES OF ADVANCED PRECOLIBACTIN METABOLITES OBTAINED FROM *clbP* DELETION STRAINS

The discovery and characterization of ClbP^{14,15} provided the foundation for all subsequent isolation efforts. Since colibactins per se are unstable, *clbP* deletion strains were produced under the hypothesis that eliminating this deacylation activity might promote the accumulation of the more stable precolibactin compounds within the cells, allowing researchers to obtain higher yields of metabolites from bacterial extracts. However, as discussed in the next section, this deletion produces surprising and initially counterintuitive structural changes in the resulting metabolites.

The three most complex known precolibactins **1**–**3** (obtained from *clbP* deletion strains) are shown in Figure 2 (for the structures of smaller *clb* metabolites, see ref 1). The structure of precolibactin A was predicted¹⁶ to be either of the isomers **1a** or **1b** based on the isolation of the shunt metabolite **4**,^{16,17} tandem MS analysis, bioinformatics, and biosynthetic logic. An HRMS signal corresponding to the exact mass of **1a** and **1b** was detected in *clbP* *E. coli* extracts,¹⁶ but precolibactin A could not be isolated due to apparent decomposition (as assessed by a reduction in ion intensity over time). Through synthesis, the correct structure was shown to be **1** (by LC/MS co-injection and tandem MS/MS; **1**, **1a**, and **1b** are isomers).¹⁸ Synthetic precolibactin A (**1**) undergoes disulfide bond formation rapidly under air. Thus, one plausible explanation for the apparent decomposition of natural precolibactin A (**1**) could be disulfide bond formation upon exposure to thiols in the bacterial extracts under air.

The structure of precolibactin C (**2**) was predicted in 2016;¹⁹ a compound with this structure was then isolated from a large-scale fermentation of a *clbP* *clbG* double mutant strain (0.5 mg of **2** were obtained from a 48-L fermentation).²⁰ Precolibactins A (**1**) and C (**2**) both contain a 2-pyridone residue and a spirocyclopropane embedded in a γ -lactam ring. The unusual macrocyclic 5-hydroxy-3-oxazoline named precolibactin-886 (**3**; 886 corresponds to the molecular weight of **3**) was isolated from a *clbP* *clbQ* double mutant strain and characterized in 2016.²¹ Precolibactin-886 (**3**) is the first chemical isolate resulting from the *clb* gene cluster known to contain an α -aminomalonate residue (red in **3**), likely resulting from the aminomalonoyl acyl carrier protein produced by ClbD, ClbE, ClbF, and ClbH²² and transferred to PKS modules by ClbG.²⁰ This subunit is believed to be important because the biosynthetic genes are essential for cytopathic effects.^{20,22}

EVIDENCE FOR COLIBACTIN PEPTIDASE-DEPENDENT REACTION PATHWAYS

Studies assessing the reactivity of synthetic precolibactin precursors revealed surprisingly that the presence or absence of a terminal *N*-acyl substituent (*N*-myristoyl-D-Asn in the natural precolibactins) controls the reaction pathways followed in the final stages of colibactin biosynthesis. The fully linear precursor **5** is liberated from the core NRPS–PKS machinery in the terminal step of precolibactin C (**2**) biosynthesis (Scheme 1A).²³ The stable linear precursor **5** was alternatively prepared by chemical synthesis and its reactivity was studied.¹⁸ In the presence of a mild base, **5** undergoes a reaction cascade involving

selective cyclodehydration to the unsaturated lactam **6**, followed by a second cyclodehydration to the pyridone precolibactin C (**2**). Under many conditions the second cyclization was faster than the first, indicating that this step is facile and making it initially difficult to detect the unsaturated lactam **6**.¹⁸ It is believed that a similar chemical pathway converts natural **5** to precolibactin C (**2**) in the *clbP* deletion strain.

Synthetic linear precursor **7**, which bears a *tert*-butoxycarbonyl substituent in place of the natural *N*-myristoyl-D-Asn side chain, was designed to assess if the amine liberated by *N*-deacylation would direct cyclization via an alternative reaction pathway. This cell- and enzyme-free deacylation provided a method to study the reactivity of colibactin products not accessible using the *clbP* deletion strains (Scheme 1B).²⁴ The behavior of the linear precursor **7** was distinct. Upon concentration from dilute acid and prior to cleavage of the carbamate group, the linear precursor **7** cyclized to the vinylogous imide **8**. The facile cyclization of **7** was attributed to the more electron-rich nature of the carbamate vs the amide in the natural system **5**. Removal of the carbamate followed by neutralization of the amine that was formed resulted in immediate cyclodehydration to provide the unsaturated imine **9**. The unsaturated imine **9** was stable and fully characterized.

The logical generalization from these studies is that the presence or absence of ClbP changes the reaction pathways followed by linear precursors such as **10** (Scheme 2). In wild-type strains, ClbP-mediated deacylation of **10** would be expected to form the amine **11**. As the amine in **11** is more nucleophilic than the carbamate in **7**, transformation to the unsaturated imine **13** is expected to be the predominant reaction pathway. The persistence of the *N*-myristoyl-D-Asn side chain in unnatural *clbP* deletion strains, then, effectively abolishes the nucleophilicity of the terminal amine. In the absence of amine participation, an artifactual cyclization pathway to generate pyridones **15**, via unsaturated lactams **14**, predominates. Following this discovery that the natural *N*-myristoyl-D-Asn side chain modulated the structure of the products, the unsaturated imine **9** was detected upon feeding the linear precursor **5** to a strain of *E. coli* expressing ClbP.²³

UNSATURATED IMINES DAMAGE DNA

The DNA-damaging activity of synthetic unsaturated imines and pyridone derivatives, identical to those produced in bacteria expressing or lacking ClbP, and resembling **13** and **15**, have now been evaluated in vitro.²⁴ The unsaturated imine **16** is representative of the structures studied (Scheme 3A). This imine damaged linearized, pBR322 DNA at nanomolar concentrations. Extensive cleavage of the DNA was observed by gel electrophoresis. The corresponding pyridone **17** was inactive at up to 0.5 mM concentration. The *N,N*-dimethylethylenediamine residue in **16** and **17** (shown in blue) was installed to mimic a currently undefined cationic residue encoded in the *clb* cluster (vide infra).²⁵ The ring-opened product **19** was obtained in 34% yield after heating the unsaturated imine **18** with propanethiol and *p*-toluenesulfonic acid (Scheme 3B). The dimeric imine **20** cross-linked linearized DNA,²⁶ while the *gem*-dimethyl derivative **21** did not damage DNA at up to 0.5 mM concentration (Scheme 3C). Although a discrete DNA alkylation product has not yet been isolated, the behavior of the compounds in Scheme 3A–C is consistent with a DNA-damaging mechanism involving nucleotide addition to the electrophilic cyclopropane (**22** →

23, Scheme 3D). The methylating agent methylmethanesulfonate is known to generate extensive strand breaks via the generation of thermally labile glycosidic linkages formed by nucleotide alkylation.²⁷ This reaction scheme is precedented; other natural products, such as CC-1065 and the duocarmycins, also alkylate DNA via nucleophilic cyclopropane ring-opening.²⁸ Extensive DNA alkylation can then lead to the production of DSBs via incomplete repair or replication fork collapse.²⁹

In earlier studies it was suggested that the precolibactins could alkylate DNA by ClbP-mediated deacylation to colibactins, unsaturated imine formation, and cyclopropane ring-opening, as shown in Scheme 3D.^{16,17a} Additionally it was discovered that the isolated metabolite **4** (Figure 1) weakly cross-linked linearized pBR322 DNA.¹⁶ The proposed cross-linking mechanism began with cyclopropane ring-opening analogous to that in Scheme 3D, followed by 1,4-addition of a second nucleotide. Surprisingly, DNA cross-links were not detected using monomeric synthetic unsaturated imines such as **16**. However, double alkylation products derived from **4** might be more stable than those derived from **16** due to the presence of a more electronegative carbonyl in **4**, rather than the imine in **16**. The unsaturated lactam **24** (which cannot cyclize to a pyridone owing to introduction of a methyl substituent to the amide, shown in red) is similar in structure to **4** and weakly alkylated DNA at 0.5 mM concentration. This suggests that iminium ion formation following ClbP deacylation significantly increases genotoxicity.

These data show that not only does *clbP* deletion change the structure of the metabolites produced by the *clb* cluster, this modification also modulates the electrophilicity of the products and changes the DSB phenotype associated with these *E. coli*. The pyridones resulting from *clbP* deletion mutants do not damage DNA and are not expected to be genotoxic, while the unsaturated imines produced in wild-type strains are potent DNA-damaging agents. This activity is likely to underlie the cytotoxic effects of the *clb* cluster.

SUMMARY OF CURRENT MODEL

It is now clear that wild-type *clb*⁺ *E. coli* produce electrophilic unsaturated imines **13** (Scheme 2) after ClbP deacylation of the corresponding linear precursors, and that deletion of *clbP*, a modification initially postulated to enhance precolibactin production, induces structural changes beyond removal of the terminal *N*-acyl substituent. In the absence of functional ClbP, the persistence of the *N*-myristoyl-D-Asn side chain diverts the intermediates toward non-genotoxic pyridones. Earlier reports determined that deletion of *clbP* abolishes cytopathic effects^{10,14b} and that ClbP inhibitors prevent DNA damage and *clb*⁺ *E. coli*-associated tumorigenesis.³⁰ The different metabolites produced in the presence and absence of ClbP provides a molecular-level explanation for these observations. In this model, precolibactins are conceptualized as structures derived from the persistence of the *N*-myristoyl-D-Asn side chain, while colibactins are the imines formed following ClbP deacylation.

UNRESOLVED ISSUES AND FUTURE DIRECTIONS

Several challenging issues surrounding the *clb* cluster, and its potentially causal link to CRC, remain unresolved. First, and perhaps of greatest significance to the work summarized here, the ability of unsaturated imines resembling **16** to recapitulate the cytopathic effects of the *clb* cluster in tissue culture, or to initiate or promote tumorigenesis in vivo, has not yet been demonstrated. Evaluation of homogeneous samples of **16** and related derivatives will be required to answer this question. These experiments will require a combination of chemical synthesis, cell culture, and animal studies.

Second, how precolibactin-886 (**3**) fits into the chemical model outlined above is not known with certainty, but there are some presently testable hypotheses about its role in this system. Evidence suggests that 5-hydroxy-oxazolidines can oxidize under air³¹ to 5-hydroxy-3-oxazolines, as in precolibactin-886 (**3**). Therefore, it is possible that “precolibactin-888” (**25**), which contains a 5-hydroxy-oxazolidine, is the precursor to precolibactin-886 (**3**, Scheme 4). If this is true, ClbP deacylation of precolibactin-888 (**25**) in wild-type strains would be expected to provide the amine **27**. The amine **27** might then transform to the unsaturated imine **30** via the vinylogous urea **29**. Alternatively, isomerization of precolibactin-888 (**25**) to the fully linear intermediate **26** may precede ClbP deacylation. Cyclodehydration of the deacylation product **28** would then provide the vinylogous urea **29**.³² Evaluation of the DNA damaging activity of structures derived from precolibactin-888 (**25**) or precolibactin-886 (**3**) will help to establish their relevance to *clb*-associated cytotoxicity. As only 2.8 mg of precolibactin-886 (**3**) were isolated from a 1000-L fermentation,²¹ chemical synthesis of **25** and derivatives may be the only technically feasible means to explore these hypotheses.

Third, while precolibactin-886 (**3**) is the most complex known precolibactin, its structure does not account for all of the enzymes in the *clb* cluster. Still unaccounted for are the putative protease ClbL^{15c} and amidase ClbO,^{17b} which do not have rigorously assigned activities. In addition, it was reported that genetically modified *E. coli* deficient in the enzymes responsible for synthesis of the aminomalonate residue in **3** (*clbD* or *clbDEF* strains) do not induce megalocytosis in HeLa cells.²² ClbG transfers the aminomalonyl-acyl carrier protein to several PKS enzymes in the cluster, suggesting this is an important building block.²⁰ This has led to the widespread notion that the aminomalonate is required for cytopathic effects, but further studies are needed to draw firm conclusions.³³

Fourth, the *clb* cluster is also surprisingly found in the probiotic strain Nissle 1917, which is used in Europe for the treatment of gastrointestinal disorders such as ulcerative colitis and diarrhea.³⁴ Based on the finding that a mutant Nissle *clbA* strain did not produce precolibactins or possess probiotic effects, it has been suggested that *clb* metabolites may be responsible for both phenotypes.^{34a} However, *clbA* contributes to the synthesis of siderophores, and a recent report established an iron-dependent interplay between precolibactin and siderophore biosynthesis.³⁵ Thus, it is possible that the loss of probiotic phenotype is due to a decrease in siderophore production, but further study is required to clarify this issue.

A final issue, whose answer may have bearing on the probiotic/genotoxic paradox highlighted above, involves the (apparent) mechanism transporting colibactins to eukaryotic cells and into their nuclei. Separation of *clb*⁺ *E. coli* and eukaryotic cells by a 0.2 μ m membrane completely prevents DNA damage,¹⁰ suggesting the compounds do not passively diffuse across cell membranes. One possibility is that the intestinal inflammation and reduced mucosa associated with *clb*⁺ *E. coli* cytopathic effects² allows greater cell-to-cell contact thereby promoting genotoxicity. This hypothesis is consistent with the observation that intestinal mucosa colonization by *clb*⁺ *E. coli* was more prevalent in patients with advanced CRC,⁴ and that intestinal inflammation promotes *clb*⁺ *E. coli*-associated tumorigenesis.² Outer membrane vesicles from Nissle 1917 undergo clathrin-mediated endocytosis in several human epithelial cell lines, suggesting probiotic effects may not require cell-to-cell contact.³⁶ Given the rich chemistry encoded in the linear biosynthetic products (e.g., **5**), an intriguing alternative hypothesis is that pathogenic and probiotic phenotypes arise from different modes of cyclization. Further investigations are required to resolve these issues.

In conclusion, perhaps the greatest challenge facing the field of human microbiome research involves establishing causal links between microbial metabolites and human health or disease states.³⁷ Progress has been mostly driven by “meta-omics”: metagenomics, metatranscriptomics, and metabolomics. In the context of precolibactins and colibactins, a chemical synthetic approach was used to augment recent insights into colibactin biosynthetic logic obtained through genetic approaches. Chemical synthesis provided experimental access to predicted isolates and biosynthetic intermediates, and facilitated insights into their structures, biosynthesis, and mechanism of action. As the full structures of the precolibactins were (and remain) undefined, engaging chemical synthesis at this stage required embracing a level of uncertainty unfamiliar to this field. However, the additional advances toward constructing a molecular model of *clb*⁺ *E. coli* genotoxicity and further elucidating their biosynthesis under-score the merits of this approach. As the power of genetic approaches to natural product discovery further advance, we predict that chemical synthetic and biosynthetic studies will become increasingly interdependent.³⁸ Chemical synthesis has long been used to validate or complete structural assignments of isolated metabolites. Integrating the synthesis of non-isolable or predicted natural product structures constitutes an exciting direction for the field.

ACKNOWLEDGMENTS

We are indebted to the colleagues and co-workers who contributed to the research described in this Perspective and whose names appear in the references. We also thank Professor Emily P. Balskus (Harvard University) and Professor Jason M. Crawford (Yale University) for comments on the manuscript. Financial support from the National Institutes of Health (GM110506) and Yale University is gratefully acknowledged.

REFERENCES

- (1) (a). Balskus EP Nat. Prod. Rep 2015, 32, 1534. [PubMed: 26390983] (b)Bode HB Angew. Chem., Int. Ed 2015, 54, 10408.(c)Trautman EP; Crawford JM Curr. Top. Med. Chem 2016, 16, 1705. [PubMed: 26456470] (d)Taieb F; Petit C; Nougayrede JP; Oswald E EcoSal Plus 2016, DOI: 10.1128/ecosalplus.ESP-0008-2016.

- (2). Arthur JC; Perez-Chanona E; Mühlbauer M; Tomkovich S; Uronis JM; Fan T-J; Campbell BJ; Abujamel T; Dogan B; Rogers AB; Rhodes JM; Stintzi A; Simpson KW; Hansen JJ; Keku TO; Fodor AA; Jobin C *Science* 2012, 338, 120. [PubMed: 22903521]
- (3). Buc E; Dubois D; Sauvanet P; Raisch J; Delmas J; Darfeuille-Michaud A; Pezet D; Bonnet R *PLoS One* 2013, 8, e56964. [PubMed: 23457644]
- (4). Bonnet M; Buc E; Sauvanet P; Darcha C; Dubois D; Pereira B; Déchelotte P; Bonnet R; Pezet D; Darfeuille-Michaud A *Clin. Cancer Res.* 2014, 20, 859. [PubMed: 24334760]
- (5). Tomkovich S; Yang Y; Winglee K; Gauthier J; Mühlbauer M; Sun X; Mohamadzadeh M; Liu X; Martin P; Wang GP; Oswald E; Fodor AA; Jobin C *Cancer Res.* 2017, 77, 2620. [PubMed: 28416491]
- (6). Cougnoux A; Dalmasso G; Martinez R; Buc E; Delmas J; Gibold L; Sauvanet P; Darcha C; Dechelotte P; Bonnet M; Pezet D; Wodrich H; Darfeuille-Michaud A; Bonnet R *Gut* 2014, 63, 1932. [PubMed: 24658599]
- (7). Stein U; Walther W; Arlt F; Schwabe H; Smith J; Fichtner I; Birchmeier W; Schlag PM *Nat. Med* 2009, 15, 59. [PubMed: 19098908]
- (8). Shimpoh T; Hirata Y; Ihara S; Suzuki N; Kinoshita H; Hayakawa Y; Ota Y; Narita A; Yoshida S; Yamada A; Koike K *Gut Pathog.* 2017, 9, 35. [PubMed: 28616082]
- (9). Gorgoulis VG; Vassiliou L-VF; Karakaidos P; Zacharatos P; Kotsinas A; Liloglou T; Venere M; DiTullio RA; Kastrinakis NG; Levy B; Kletsas D; Yoneta A; Herlyn M; Kittas C; Halazonetis TD *Nature* 2005, 434, 907. [PubMed: 15829965]
- (10). 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. [PubMed: 16902142]
- (11). Cuevas-Ramos G; Petit CR; Marcq I; Boury M; Oswald E; Nougayrède J-P *Proc. Natl. Acad. Sci. U. S. A* 2010, 107, 11537. [PubMed: 20534522]
- (12). National Council on Radiation Protection and Measurements. Ionizing Radiation Exposure of the Population of the United States, NCRP report no. 160; U.S. Department of Commerce, National Bureau of Standards: Washington, DC, 2009.
- (13). Mousa JJ; Yang Y; Tomkovich S; Shima A; Newsome RC; Tripathi P; Oswald E; Bruner SD; Jobin C *Nat. Microbiol* 2016, 1, 15009. [PubMed: 27571755]
- (14). (a). Dubois D; Baron O; Cougnoux A; Delmas J; Pradel N; Boury M; Bouchon B; Bringer MA; Nougayrède JP; Oswald E; Bonnet RJ *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. [PubMed: 23041299]
- (15). (a). Brotherton CA; Balskus EP *J. Am. Chem. Soc* 2013, 135, 3359. (b) Bian X; Fu J; Plaza A; Herrmann J; Pistorius D; Stewart AF; Zhang Y; Muller R *ChemBioChem* 2013, 14, 1194. [PubMed: 23744512] (c) Vizcaino MI; Engel P; Trautman E; Crawford JM *J. Am. Chem. Soc* 2014, 136, 9244. [PubMed: 24932672]
- (16). Vizcaino MI; Crawford JM *Nat. Chem* 2015, 7, 411. [PubMed: 25901819]
- (17). (a). Brotherton CA; Wilson M; Byrd G; Balskus EP *Org. Lett* 2015, 17, 1545. [PubMed: 25753745] (b) Bian X; Plaza A; Zhang Y; Müller R *Chem. Sci* 2015, 6, 3154. [PubMed: 28706687]
- (18). Healy AR; Vizcaino MI; Crawford JM; Herzon SB *J. Am. Chem. Soc* 2016, 138, 5426. [PubMed: 27025153]
- (19). Li ZR; Li Y; Lai JY; Tang J; Wang B; Lu L; Zhu G; Wu X; Xu Y; Qian PY *ChemBioChem* 2015, 16, 1715. [PubMed: 26052818]
- (20). Zha L; Wilson MR; Brotherton CA; Balskus EP *ACS Chem. Biol* 2016, 11, 1287. [PubMed: 26890481]
- (21). Li ZR; Li J; Gu JP; Lai JY; Duggan BM; Zhang WP; Li ZL; Li YX; Tong RB; Xu Y; Lin DH; Moore BS; Qian PY *Nat. Chem. Biol* 2016, 12, 773. [PubMed: 27547923]
- (22). Brachmann AO; Garcie C; Wu V; Martin P; Ueoka R; Oswald E; Piel J *Chem. Commun* 2015, 51, 13138.
- (23). Trautman EP; Healy AR; Shine EE; Herzon SB; Crawford JM *J. Am. Chem. Soc* 2017, 139, 4195. [PubMed: 28240912]

- (24). Healy AR; Nikolayevskiy H; Patel JR; Crawford JM; Herzon SB J. Am. Chem. Soc 2016, 138, 15563. [PubMed: 27934011]
- (25). Synthetic constructs lacking this residue are less potent but display parallel reactivity. See ref 24.
- (26). DNA cross-links were observed after a 3 h incubation of linearized pBR322 DNA with the dimer **20**. At later time points, degradation of the DNA (as observed with monomeric structures) occurred. See ref 24.
- (27). Lundin C; North M; Erixon K; Walters K; Jenssen D; Goldman ASH; Helleday T Nucleic Acids Res. 2005, 33, 3799. [PubMed: 16009812]
- (28). For a review, see: Tichenor MS; Boger DL Nat. Prod. Rep 2008, 25, 220. [PubMed: 18389136]
- (29) (a). Kondo N; Takahashi A; Ono K; Ohnishi TJ Nucleic Acids 2010, 2010, 543531. (b) Zhang J; Walter JC DNA Repair 2014, 19, 135. [PubMed: 24768452]
- (30). Cougnoux A; Delmas J; Gibold L; Fais T; Romagnoli C; Robin F; Cuevas-Ramos G; Oswald E; Darfeuille-Michaud A; Prati F; Dalmasso G; Bonnet R Gut 2016, 65, 278. [PubMed: 25588406]
- (31). Wessig P; Schwarz J; Lindemann U; Holthausen MC Synthesis 2001, 112, 1258.
- (32). It is also possible that the fully linear intermediate **26** undergoes oxidation and cyclization to generate precolibactin-886 (**3**), but this would seem less likely given the hydrolytic instability of *N*-H imines.
- (33). It seems plausible that the affinity for DNA of metabolites containing or lacking an aminomalonate residue (which is likely to be protonated under physiological conditions) will be distinct. In addition, the nature of the DNA lesions induced by these metabolites may differ. A comprehensive evaluation of the eukaryotic DNA damage and cell cycle response using *clbD* or *clbDEF* deletion strains is desirable.
- (34) (a). Olier M; Marcq I; Salvador-Cartier C; Secher T; Dobrindt U; Boury M; Bacquié V; Penary M; Gaultier E; Nougayrède J-P; Fioramonti J; Oswald E Gut Microbes 2012, 3, 501. [PubMed: 22895085] (b) Scaldaferri F; Gerardi V; Mangiola F; Lopetuso LR; Pizzoferrato M; Petito V; Papa A; Stojanovic J; Poscia A; Cammarota G; Gasbarrini A World J. Gastroenterol 2016, 22, 5505. [PubMed: 27350728]
- (35). Martin P; Tronnet S; Garcia C; Oswald E IUBMB Life 2017, 69, 435. [PubMed: 28295919]
- (36). Canas MA; Gimenez R; Fabrega MJ; Toloza L; Baldoma L; Badia J PLoS One 2016, 11, e0160374. [PubMed: 27487076]
- (37). Martinez KB; Leone V; Chang EB J. Biol. Chem. 2017, 292, 8553. [PubMed: 28389566]
- (38). For a prescient discussion of genetic approaches to natural product discovery, see: (a) Walsh CT; Fischbach MA J. Am. Chem. Soc 2010, 132, 2469. [PubMed: 20121095] Recently, nonribosomal peptides possessing antibiotic activity have been synthesized based solely on bioinformatics analysis of bacterial genomes, see: (b) Chu J; Vila-Farres X; Inoyama D; Ternei M; Cohen LJ; Gordon EA; Reddy BV; Charlop-Powers Z; Zebroski HA; Gallardo-Macias R; Jaskowski M; Satish S; Park S; Perlin DS; Freundlich JS; Brady SF Nat. Chem. Biol 2016, 12, 1004. [PubMed: 27748750] (c) Vila-Farres X; Chu J; Inoyama D; Ternei MA; Lemetre C; Cohen LJ; Cho W; Reddy BVB; Zebroski HA; Freundlich JS; Perlin DS; Brady SF J. Am. Chem. Soc 2017, 139, 1404. [PubMed: 28055186]

Figure 1.

Abbreviated representation of precolibactin transport and activation. Reproduced with permission from ref 1d. Copyright 2016 ASM Press.

Figure 2. Structures of precolibactin A (**1**), precolibactin C (**2**), precolibactin-886 (**3**), and the shunt metabolite **4**. The originally predicted structures of precolibactin A are shown as **1a** and **1b**.

Scheme 1. Reaction Pathways of Synthetic Linear Precolibactin and Colibactin Precursors Depend on the Terminal *N*-Acyl Substituent^a

^a(A) Linear intermediate **5**, an established product of the *clb* pathway,²³ undergoes cyclodehydration to precolibactin C (**2**) under mild conditions. (B) Synthetic *N*-(*tert*-butoxycarbonyl) derivative **7** undergoes cyclization to the vinylogous imide **8** upon concentration from dilute acid. Removal of the carbamate followed by neutralization results in immediate transformation to the unsaturated imine **9**.

Scheme 2. Mechanistic Model for the Production of *clb* Metabolites^a

^aIn wild-type (ClbP proficient) strains, *N*-deacylation followed by sequential cyclization reactions provides unsaturated imines **13**. In the absence of ClbP, unnatural cyclization pathways to generate pyridones **15** predominates. The gray sphere denotes the variable heterocyclic fragment of the metabolites.

Scheme 3. Key Control Compounds and Mechanism of Action of Colibactins^a

^a(A) Unsaturated imine **16** damaged linearized pBR322 DNA at nanomolar concentrations, while the corresponding pyridone **17** did not generate detectable damage at up to 0.5 mM concentration. (B) Ring-opened product **19** was obtained in 34% yield after heating the unsaturated imine **18** with propanethiol and *p*-toluenesulfonic acid. (C) Dimeric unsaturated imine **20** cross-linked linearized pBR322 DNA, while the *gem*-dimethyl derivative **21** did not damage DNA at up to 0.5 mM concentration. (D) Results of the studies summarized in panels A–C are consistent with a mechanism involving DNA alkylation by nucleotide addition to the electrophilic cyclopropane (**22** → **23**). (E) The unsaturated lactam **24** weakly damaged DNA at 0.5 mM concentration. The grey spheres in reaction D denote the variable heterocyclic fragment of the metabolites.

Scheme 4.

Possible Pathways for the Transformation of Precolibactin-888 (25) to Precolibactin-886 (3) and Unsaturated Imine 30