Colibactin: Understanding an elusive gut bacterial genotoxin

Journal:	*Natural Product Reports*
Manuscript ID	NP-PER-08-2015-000091.R1
Article Type:	Highlight
Date Submitted by the Author:	05-Sep-2015
Complete List of Authors:	Balskus, Emily; Harvard University,

Journal Name

ARTICLE

Colibactin: Understanding an elusive gut bacterial genotoxin

Emily P. Balskus[a]

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

This Highlight provides an overview of recent progress towards elucidating the structure, biosynthesis, and mode of action of colibactin, a genotoxic secondary metabolite synthesized by human gut bacteria. While isolating colibactin has been problematic, efforts to characterize its biosynthesis have provided critical information that has led to a rapid increase in our knowledge of this metabolite's structure and function. Major questions and gaps remain however, and the broader lessons learned from studying colibactin highlight central challenges to be faced in the genomics era of natural product research and in efforts to understand the human microbiome.

Introduction

An increased appreciation for the importance of commensal microorganisms in human health has highlighted the potential roles that specialized microbial metabolites might play in the environment of the human body.[1] Both the presence of secondary metabolite biosynthetic gene clusters in the genomes of human commensals[2] and the significance of secondary metabolites in pathogenic organisms[3] has led to growing interest in understanding whether members of the human microbiota synthesize complex small molecules that influence host biology.

Colibactin is a secondary metabolite produced by members of the human gut microbiota that has captured the attention of both biologists and chemists due to its intriguing biological activity and mode of discovery. The biosynthetic gene cluster that produces colibactin was identified in 2006 by Oswald and co-workers.[4] They found that certain *E. coli* strains induced the formation of DNA double strand breaks in transiently infected HeLa cells and used transposon mutagenesis to link this activity to a 54 kb gene cluster (the *pks* island). This gene cluster encoded a non-ribosomal peptide synthetase-polyketide synthase (NRPS-PKS) assembly line and was therefore hypothesized to produce a genotoxic secondary metabolite named colibactin. Elements of the *pks* island included three NRPSs (*clbH,J,N*), three PKSs (*clbC,I,O*), two NRPS/PKS hybrids (*clbB,K*), additional enzymes (*clbA,D-G,L,P*), and a transporter (*clbM*). Mutation of individual genes revealed that all of the genes encoding enzymes, including all components of the assembly line, were essential for activity. Transfer of the entire cluster also conferred genotoxicity to a non-genotoxic *E. coli* strain, suggesting that all of the genetic elements needed for colibactin production were found within the *pks* island.

[a] Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford St., Cambridge, MA 02138, USA

This discovery prompted further efforts to understand the implications of colibactin-producing gut bacteria for host biology, with an emphasis on exploring potential links between colibactin-mediated DNA damage and colorectal cancer (CRC). The majority of *E. coli* strains possessing the *pks* island (*pks+*) are members of the phylogenetic group B2, which includes both extraintestinal pathogens and intestinal commensals.[5,6] Notably, the *pks* island is present in the probiotic strain Nissle 1917, which has been used to treat chronic inflammatory bowel disease (IBD) for almost 100 years. Culture-based analyses have suggested that ~20% of healthy individuals harbor *pks+ E. coli*.[7] Strains with the *pks* island generated DNA double strand breaks in mouse colonic epithelial cells *in vivo*.[8] In 2012, Jobin and co-workers reported that *pks+ E. coli* enhanced tumor progression in azoxymethane-treated *IL10*$^{-/-}$ mice, a model for colitis-associated CRC.[7] This effect on tumor growth may proceed via induction of cellular senescence and secretion of factors that alter the growth of neighboring cells.[9,10] Together, these findings raise the possibility that colibactin-mediated DNA damage may influence risk for CRC in humans. This intriguing hypothesis would be consistent with the increased carriage of *pks+ E. coli* in CRC and IBD patients.[7,11] *Pks* island transcripts have also been observed in CRC tumor samples.[12] Other studies have suggested that the *pks* island can alter host immunity,[13] increase intestinal permeability,[14] affect lymphopenia,[15] and play a role in virulence.[16] Intriguingly, in *E. coli* Nissle 1917, deletion of the essential phosphopantethienyl transferase (PPTase) *clbA* diminishes this strain's anti-inflammatory activity in rodent models of colitis, suggesting that products of the *pks* island may influence the beneficial properties of this probiotic.[17] Finally, the *pks* island has been identified in the genomes of other Proteobacteria,[6] including strains that live in close association with sponges[18] and honeybees,[19] suggesting that colibactin might mediate evolutionarily conserved interactions with animal hosts.

Overall, these biological studies provide a rationale for wanting to understand the chemical structure of colibactin and

the molecular basis for its activity. However, the small molecule or molecules responsible for genotoxicity have not yet been isolated from any colibactin producer.

A 'prodrug strategy' is involved in colibactin biosynthesis

Several findings from the original Oswald report suggested that isolating colibactin would be challenging. Genotoxicity was not observed when HeLa cells were treated with *E. coli* culture supernatants or cell lysates. This result could be consistent with a highly unstable metabolite. Activity was also lost when bacteria and human cells were separated by a 0.2 μm membrane, suggesting a requirement for cell-cell contact. However, subsequent studies of the transcription of the *pks* island found that the gene cluster was constitutively expressed and that expression levels were not altered by the presence of human cells.[20] All of these factors have complicated activity-based isolation efforts.

We took an unconventional approach to study colibactin, deciding to examine the activity of biosynthetic enzymes prior to isolation and structure elucidation (Fig. 1A). We envisioned that an understanding of enzymatic assembly line biosynthetic logic would allow us to analyze the gene cluster using bioinformatics and generate structural hypotheses that we could test through *in vitro* enzyme characterization. In their initial annotation of the colibactin NRPS-PKS assembly line, the Oswald group did not propose a product structure but noted the presence of NRPS domains (cyclization, oxidase) involved in forming thiazole heterocycles. In our analysis of the *pks* island, we recognized homologs of genes involved in the biosynthesis of the antibiotics zwittermicin and xenocoumacin: an initiating NRPS (ClbN) and a membrane-anchored, periplasmic peptidase (ClbP) (Fig. 1B).[21,22] Bode and co-workers had deleted the gene encoding the peptidase (*xcnG*) from xenocoumacin biosynthesis and found accumulation of larger metabolites in this mutant.[23] These inactive 'prexenocoumacins' contained an N-acylated D-asparagine residue that was presumably constructed by the first NRPS module of the assembly line and later removed by XcnG in the final stages of biosynthesis. Bode and co-workers postulated that this 'prodrug'-like logic served as a self-resistance mechanism.

We hypothesized related logic might be used in colibactin biosynthesis and tested this proposal by characterizing the *in vitro* activity of the predicted prodrug synthesizing, elongating, and cleaving enzymes from the *pks* island (Fig. 1C).[24] We found that the single-module NRPS ClbN accepted L-Asn as a substrate and generated an *N*-acylated 'prodrug motif' with various acyl-CoA co-substrates. A competition experiment revealed a preference for myristoyl-CoA. As ClbN contained an epimerization domain, we could identify the NRPS module that should elongate the prodrug motif by searching for condensation (C) domains containing residues characteristic of C domains that accept D-amino acids. The first module of the NRPS-PKS hybrid ClbB met this criterion, and the excised NRPS module was able to elongate the intermediate generated by ClbN, with a preference for L-alanine. This work was the first *in vitro* study of any prodrug motif synthesizing and elongating enzymes.

We also characterized the activity of the periplasmic peptidase ClbP *in vivo* using synthetic substrates containing the prodrug motif and *E. coli* strains expressing various versions of ClbP. We observed hydrolysis only in strains expressing the full-length enzyme. Peptidase mutants missing either the key active site serine or predicted transmembrane helices lacked activity. These experiments confirmed that ClbP did hydrolyze substrates containing the prodrug motif. Separate work by Bonnet and co-workers examined the *in vivo* role of ClbP and found that these ClbP mutants could not rescue genotoxicity in Δ*clbP E. coli* strains.[25,26] Together, these two studies indicated that ClbP-catalyzed hydrolysis of an inactive 'precolibactin' is essential for genotoxicity.

One major question regarding our strategy was the extent to which *in vitro* enzymatic activity would correspond to *in vivo* reactivity. This potential concern was mitigated when Müller, Zhang, and co-workers compared the metabolite profiles of wild-type and *Δpks* mutant strains of *E. coli* Nissle 1917 and identified N-myristoyl-D-asparagine (**1**) in the wild-type strain (Fig. 2A).[27] This metabolite was isolated and structurally characterized from an *E. coli* strain heterologously expressing the *pks* cluster. In addition to **1**, several related analogs with different chain lengths were detected but not isolated due to their low abundance. The isolated prodrug motif lacked genotoxicity and displayed only weak antibacterial activity. Moreover, the failure to observe any additional, higher molecular weight *pks*-associated metabolites in this study further highlighted the difficulty associated with isolating the active genotoxin.

Metabolomics guides the isolation of precolibactins

The discovery of ClbP's role in colibactin biosynthesis illuminated another approach for obtaining structural information: isolation of unprocessed and inactive 'precolibactins' from Δ*clbP* strains. The first study using this strategy was reported by Crawford and co-workers.[28] They combined comparative metabolomics with network analysis to identify metabolites associated with both *pks+ E. coli* and *clbP* mutants. As with the prexenocoumacins, variations within the fatty acyl substituents of the precolibactins were observed. None of the isolated molecules exhibited genotoxic activity. This observation, combined with the large number of distinct *pks*-associated masses, led to the as-yet-untested proposal that colibactin's activity may arise from a mixture of metabolites rather than one discrete compound.

Continued isolation efforts recently led to the discovery of a larger candidate precolibactin (**3**) by multiple groups (Fig. 2B).[29-32] This metabolite contained a structural feature that

was unanticipated from bioinformatic analyses: a spiro cyclopropane ring. Cyclopropanes are found in several classes of DNA-alkylating natural products, including the duocarmycins and the illudins.[33,34] The presence of this functional group in **3** has led to the proposal that colibactin's mode of action may involve covalent modification of DNA or a target protein by an electrophilic 'warhead' (Fig. 2D). Vizcano and Crawford tested this hypothesis experimentally, exposing plasmid DNA to millimolar concentrations of **3** for 24 h and observing via gel electrophoresis small amounts of a higher molecular weight band that could potentially correspond to a cross-linked adduct.[31] Although intriguing, it is unclear if the reactivity of any candidate precolibactin will be relevant for colibactin's mode of action given that *clbP* mutants lack genotoxicity.

Isolation efforts have also led to the prediction and elucidation of the structures of several higher molecular weight candidate precolibactins (Fig. 2C). Vizcano and Crawford used a combination of mass spectrometry, feeding experiments, and biosynthetic information to predict the structure of a larger molecular weight species (**5**, *m/z* = 816).[31] This proposed structure contains the potential electrophilic warhead found in **3** and a thiazoline-thiazole. The presence of this heterocyclic framework, which is also found in the phleomycins,[35] could provide additional support for DNA as the direct target of colibactin. More recently, Qian and co-workers isolated several new candidate precolibactins, including a thiazole-containing metabolite (**4**), and predicted the presence of a bis-thiazole in a larger species (**6**, *m/z* = 796).[32] Interestingly, these precolibactins possess a cyclopropane-containing ring system that differs from that of **3** and **5**. The nature of the ring system fused to the cyclopropane should alter the reactivity of this group and may be important for genotoxicity. However, evidence that either of these two potential 'warheads' is actually incorporated into colibactin is still lacking.

The unusual chemistry of the colibactin NRPS-PKS assembly line

The successful isolation of these candidate precolibactins, in combination with both gene inactivation and feeding experiments, has shed light on the roles of additional colibactin biosynthetic enzymes and the order of events in the pathway. These studies have also revealed that this assembly line employs unusual biosynthetic logic.

The colibactin enzymatic assembly line contains a mixture of NRPS and PKS machinery that operates in a non-colinear fashion, including both cis-acyltransferase (AT) PKS modules and modules with 'deteriorated', potentially inactive (atypical) AT domains. The identities of many of the building blocks utilized by the assembly line could be readily predicted using bioinformatics, and this information guided both our initial biochemical experiments[24] and subsequent *in vivo* studies of the biosynthesis of larger metabolites.[30-32] Results thus far support the biosynthetic hypothesis outlined in Fig. 3A. Unusual features of this proposal include the use of both cis-AT (ClbB,I) and atypical-AT (ClbC) PKS machinery to perform elongations with malonyl-CoA and the incorporation of aminocyclopropane carboxylic acid (ACC) by an NRPS module (ClbH). The chemistry involved in ACC formation and utilization is not yet understood. This amino acid is also used in cytotrienin biosynthesis.[36] Though feeding experiments with cytotrienin and colibactin producers have found evidence for L-methionine as the source of ACC, additional experiments with the colibactin producer suggest that ACC is not generated as a discrete building block.[31]

Though this biosynthetic hypothesis could account for the production of isolated precolibactins, it does not include many additional enzymes encoded by the *pks* island that are essential for genotoxicity, suggesting that crucial events in the pathway are missing. Recently, *in vitro* biochemical characterization played a critical role in addressing this concern.[37] Piel and co-workers recognized that elements of the *pks* island (clbD-F and the first adenylation (A) domain of ClbH) resembled the enzymatic machinery that synthesizes the unusual PKS extender unit aminomalonyl-ACP from L-serine in zwittermicin biosynthesis. *In vitro* characterization of these enzymes supported the hypothesis that this building block is also generated in colibactin biosynthesis. *E. coli* mutants missing this biosynthetic machinery lacked genotoxicity, indicating an essential role for these enzymes in colibactin assembly. Interestingly, these mutants were still able to produce candidate precolibactins **3** and **4**, suggesting that aminomalonyl-ACP is used in a distinct part of the pathway that is not involved in the synthesis of known *pks*-associated metabolites. Identifying metabolites that incorporate this building block is a major challenge for future investigations.

Remaining questions

Though our knowledge of colibactin has increased dramatically over the past two years, there are still many critical gaps in our understanding of its structure, biosynthesis, and biological activity. Some of the important remaining questions include:

What is the structure of the active genotoxin(s)? Genetic evidence clearly indicates that none of the candidate precolibactins isolated or proposed to date could, on their own, actually be precursors to the active genotoxin(s). While it is possible that the ClbP-processed forms of one or more of these metabolites acts synergistically with one or more as-yet uncharacterized metabolites, it is equally possible that these metabolites are all derived from off-pathway intermediates that are generated only when *clbP* is removed and normal assembly line function is disrupted. This latter hypothesis is consistent with the failure to detect processed versions of these metabolites in *pks+ E. coli* extracts. It has also not yet been confirmed that any metabolite accumulating in a *clbP* mutant can be hydrolyzed to give a genotoxin. How will we ultimately establish the chemical structure of the active genotoxic product(s) of the *pks* island? Will isolation efforts

eventually succeed or will alternate approaches such as chemical synthesis or *in vitro* biosynthesis provide the answer?

How does the unusual NRPS-PKS assembly line function in colibactin biosynthesis? Though *in vivo* experiments have identified the components of the assembly line involved in synthesizing candidate precolibactins, an understanding of how these enzymes actually function is lacking. The colibactin NRPS-PKS clearly represents a rich source of unusual assembly line enzymology. *In vitro* biochemical experiments are needed to fully understand the collaboration between cis-AT and atypical-AT PKSs, the generation of ACC, and the potential use of aminomalonyl-ACP.

Why is prodrug cleavage required for genotoxicity? Though generation and cleavage of a 'prodrug motif' is used as a strategy in multiple bacterial biosynthetic pathways,[37] we don't yet have an understanding of how removal of the *N*-acyl D-Asn motif enhances the biological activity of any metabolite. Generation of a primary amine upon amide bond cleavage could influence activity by altering the molecule's overall charge. In the case of colibactin, it could also trigger a reaction(s) that changes metabolite structure in a way that enhances activity (Fig. 2D). Finally, it is possible that the presence of the prodrug motif alters activity *in vivo* simply by sequestering the active compound in the membrane of producing bacteria, limiting its access to cellular targets in other organisms. Elucidating the precise role of this tailoring step will require an understanding of the mode of action of these metabolites and the ability to systematically modify the structures of both the prodrug-containing and processed forms.

How does colibactin production lead to DNA damage? A central question regarding colibactin's activity is how exposure of host cells to *pks+ E. coli* leads to DNA damage. While it is clear that a downstream result is the formation of DNA double strand breaks, conclusive evidence for colibactin directly mediating DNA damage is still lacking. Certain features of characterized or proposed candidate precolibactin structures support the hypothesis that colibactin directly alkylates DNA, including the cyclopropane and the bis-thiazole/thiazolinyl-thiazole heterocycles. However, no colibactin- or precolibactin-DNA adduct has been structurally characterized and there is no evidence from studies of the host response to *pks+ E. coli* that conclusively implicates DNA crosslinking in colibactin's mode of action. Addressing this question will require additional *in vitro* and *in vivo* studies of colibactin and precolibactins' interactions with biomolecules and further examination of the events preceding the accumulation of DNA double strand breaks in *pks+ E. coli*-treated host cells.

Does colibactin production influence human biology? Even after identification of the active genotoxin(s), the question of whether colibactin contributes to CRC in humans will remain. One factor that complicates efforts to understand colibactin's biological role is the fact that almost all *in vivo* experiments have utilized *Δpks* or *ΔclbA* mutants.[7,8,10,15,17] Deletion of the PPTase likely modulates the activity of biosynthetic enzymes involved in other pathways,[40] so any effects seen cannot be definitively linked to colibactin production. Indeed, *clbA* and *clbP* mutants exhibit dramatically different phenotypes in a mouse model of infection.[16] Small molecule inhibitors capable of reducing colibactin production could help to elucidate the biological role of this metabolite. Although compounds that bind to ClbP have been identified and appear to prevent colibactin-mediated DNA damage when dosed at milimolar concentrations *in vitro* and *in vivo*, there is currently limited evidence that these molecules interfere with ClbP's activity.[41]

Ultimately it is important that we take an unbiased approach to studying colibactin's role in the human gut. While colibactin-mediated DNA damage and the potential link to cancer have provided important motivation, there is a need for future studies to examine additional facets of colibactin production, including its affects on gut community composition, the host immune response, and other aspects of gastrointestinal tract function.

Conclusions

There are many reasons why colibactin is a remarkable natural product. It is made by *E. coli*, our most extensively studied bacterial species, yet its isolation remains elusive. Colibactin producers inhabit the human gut, perhaps the most exciting microbial environment from the perspective of human biology, and may be connected to disease. But more generally, colibactin and the scientific activities it has inspired illustrate the challenges that both biologists and chemists who study natural products face in the age of genomics. With access to ever more extensive microbial genome sequencing data, the number of biosynthetic gene clusters encoding uncharacterized or 'cryptic' secondary metabolites grows, greatly outnumbering the pathways that generate structurally characterized molecules.[42] Moving forward, it is likely that uncharacterized natural products will increasingly be linked to their biological activities prior to isolation. The challenges associated with colibactin highlight the need to develop alternate strategies to study molecules that prove recalcitrant to isolation efforts.

The colibactin story also embodies many larger issues facing human microbiota research. Colibactin's important biological context, interesting activity, and elusive structure have captured our attention, but we must not let this enthusiasm allow us to reach conclusions that are not yet fully supported by experimental data. Similar problems are seen more broadly in dealing with the excitement surrounding human microbiota research.[43] Success in characterizing this mysterious metabolite will certainly require creativity and ingenuity, but will also benefit from taking a critical and measured approach.

Finally, it is clear that techniques from multiple disciplines will be required to fully understand the roles of colibactin and other gut microbial metabolites in human biology. This again reflects a larger trend in research. By actively engaging diverse expertise in our efforts, we will not only enhance interest in

and appreciation for natural products among a broader scientific audience, but also gain the expanded perspective needed to tackle this challenging problem.

Acknowledgements

Studies of colibactin biosynthesis in the Balskus group have been supported by a Damon-Runyon-Rachleff Innovation Award, the Searle Scholars Program, and the Richard and Susan Smith Family Foundation (Newton, MA). Carolyn Brotherton, Matthew Wilson, and Yindi Jiang are acknowledged for their thoughtful comments on and contributions to this manuscript.

Notes and references

1 (a) G. Sharon, N. Garg, J. Debelius, R. Knight, P. C. Dorrestein and S. K. Mamanian, *Cell Metab.*, 2014, **20**, 719–730; (b) M. S. Donia and M. A. Fischbach, *Science*, 2015, **349**, DOI:10.1126/science.1254766.
2 M. S. Donia, P. Cimermancic, C. J. Schulze, L. C. Wieland Brown, J. Martin, M. Mitreva, J. Clardy, R. G. Linington and M. A. Fischbach, *Cell*, 2014, **158**, 1402–1414.
3 (a) D. H. Scharf, T. Heinekamp and A. A. Brakhage, *PLoS Pathog.*, **10**, e1003859; (b) J. B. Biggins, H.-S. Kang, M. A. Ternei, D. DeShazer and S. F. Brady, *J. Am. Chem. Soc.*, 2014, **136**, 9484–9490.
4 J.-P. Nougayrède, S. Homburg, F. Taieb, M. Boury, E. Brzuszkiewicz, G. Gottschalk, C. Buchrieser, J. Hacker, U. Dobrindt and E. Oswald, *Science*, 2006, **313**, 848–851.
5 J. R. Johnson, B. Johnston, M. A. Kuskowski, J.-P. Nougayrède and E. Oswald, *J. Clin. Microbiol.*, 2008, **46**, 3906–3911.
6 J. Putze, C. Hennequin, J.-P. Nougayrède, W. Zhang, S. Homburg, H. Karch, M. A. Bringer, C. Fayolle, E. Carniel, W. Rabsch, T. A. Oelschlaeger, E. Oswald, C. Forestier, J. Hacker and U. Dobrindt, *Infect. Immun*., 2009, **77**, 4696–4703.
7 J. C. Arthur, E. Perez-Chanona, M. Mühlbauer, S. Tomkovich, J. M. Uronis, T.-J. Fan, B. J. Campbell, T. Abujamel, B. Dogan, A. B. Rogers, J. M. Rhodes, A. Stintzi, K. W. Simpson, J. J. Hansen, T. O. Keku, A. A. Fodor and C. Jobin, *Science*, 2012, **338**, 120–123.
8 G. Cuevas-Ramos, C. R. Petit, I. Marcq, M. Boury, E. Oswald and J.-P. Nougayrède, *Proc. Natl. Acad. Sci. USA*, 2010, **107**, 11537–11542.
9 T. Secher, A. Samba-Louaka, E. Oswald and J.-P. Nougayrède, *PLoS One*, 2013, **8**, e77157.
10 A. Cougnoux, G. Dalmasso, R. Martinez, E. Buc, J. Delmas, L. Gibold, P. Sauvanet, C. Darcha, P. Dechelotte, M. Bonnet, D. Pezet, H. Wodrich, A. Darfeuille-Michaud and R. Bonnet, *Gut*, 2014, **63**, 1932–1942.
11 E. Buc, D. Dubois, P. Sauvanet, J. Raisch, J. Delmas, A. Darfeuille-Michaud, D. Pezet and R. Bonnet, *PLoS ONE*, 2013, **8**, e56964.
12 B. E. Dutilh, L. Backus, S. A. F. T. van Hijum and H. Tjalsma, *Best Pract. Res. Clin. Gastroenterol.*, 2013, **27**, 85–99.
13 T. Secher, D. Payros, C. Brehin, M. Boury, C. Watrin, M. Gillet, I. Bernard-Cadenat, S. Menard, V. Theodorou, A. Saoudi, M. Olier and E. Oswald, *Infect. Immun.*, 2015, **83**, 2420–2429.
14 D. Payros, T. Secher, M. Boury, C. Brehin, S. Ménard, C. Salvador-Cartier, G. Cuevas-Ramos, C. Watrin, I. Marcq, J.-P. Nougayrède, D. Dubois, A. Bedu, F. Garnier, O. Clermont, E. Denamur, P. Plaisancié, V. Theodorou, J. Fioramonti, M. Olier and E. Oswald, *Gut Microbes*, 2014, **5**, 313–325.
15 I. Marcq, P. Martin, D. Payros, G. Cuevas-Ramos, M. Boury, C. Watrin, J.-P. Nougayrède, M. Olier and E. Oswald, *J. Infect. Dis.*, 2014, **210**, 285–294.
16 A. J. McCarthy, P. Martin, E. Cloup, R. Stabler, E. Oswald and P. W. Taylor, *Infect. Immun.*, 2015, IAI.00716-15.
17 M. Olier, I. Marcq, C. Salvador-Cartier, T. Secher, U. Dobrindt, M. Boury, V. Bacquié, M. Penary, E. Gaultier, J.-P. Nougayrède, J. Fioramonti and E. Oswald, *Gut Microbes*, 2012, **3**, 501–509.
18 V. Bondarev, M. Richter, S. Romano, J. Piel, A. Schwedt and H. N. Schulz-Vogt, *Environ. Microbiol.*, 2013, **15**, 2095–2113.
19 P. Engel, M. I. Vizcaino and J. M. Crawford, *Appl. Environ. Microbiol.*, 2015, **81**, 1502–1512.
20 S. Homburg, E. Oswald, J. Hacker and U. Dobrindt, *FEMS Microbiol. Lett.*, 2007, **275**, 255–262.
21 B. M. Kevany, D. A. Rasko and M. G. Thomas. *Appl. Environ. Microbiol.*, 2009, **75**, 1144–1155.
22 D. Reimer, E. Luxenburger, A. O. Brachmann and H. B. Bode, *ChemBioChem*, 2009, **10**, 1997–2001.
23 D. Reimer, K. M. Pos, M. Thines, P. Grün and H. B Bode, *Nat. Chem. Biol.*, 2011, **7**, 888–890.
24 C. A. Brotherton and E. P. Balskus, *J. Am. Chem. Soc*., 2013, **135**, 3359–3362
25 D. Dubois, O. Baron, A. Cougnoux, J. Delmas, N. Pradel, M. Boury, B. Bouchon, M. A. Bringer, J.-P. Nougayrède, E. Oswald and R. Bonnet, *J. Biol. Chem.*, 2011, **286**, 35562–35570.
26 A. Cougnoux, L. Gibold, F. Robin, D. Dubois, N. Pradel, A. Darfeuille-Michaud, G. Dalmasso, J. Delmas and R. J. Bonnet, *Mol. Biol.*, 2012, **424**, 203–214.
27 X. Bian, J. Fu, A. Plaza, J. Herrmann, D. Pistorius, A. F. Stewart, Y. Zhang and R. Müller, *ChemBioChem*, 2013, **14**, 1194–1197.
28 M. I. Vizcaino, P. Engel, E. Trautman and J. M. Crawford, *J. Am. Chem. Soc*., 2014, **136**, 9244–9247.
29 C. A. Brotherton, M. Wilson, G. Byrd and E. P. Balskus, *Org. Lett.*, 2015, **17**, 1545–1548.
30 X. Bian, A. Plaza, Y. Zhang and R. Müller, *Chem. Sci.*, 2015, **6**, 3154–3160.
31 M. I. Vizcaino and J. M. Crawford, *Nat. Chem.*, 2015, **7**, 411–417.
32 Z. R. Li, Y. Li, J. Y. Lai, J. Tang, B. Wang, L. Lu, G. Zhu, X. Wu, Y. Xu and P. Y. Qian, *ChemBioChem*, 2015, **16**, 1715–1719.
33 N. Gosh, H. M. Sheldrake, M. Searcey and K. Pors, *Curr. Top. Med. Chem.*, 2009, **9**, 1494–1524.
34 M. Tanasova and S. Sturla, *Chem. Rev.*, 2012, **112**, 3578–3610.
35 S. M. Hecht, *J. Nat. Prod.*, 2000, **63**, 158–168.
36 H. P. Zhang, H. Kakeya and H. Osada, *Tet. Lett.*, 1998, **39**, 6947–6948.
37 A. O. Brachmann, C. Garcie, V. Wu, P. Martin, R. Ueoka, E. Oswald and J. Piel, *Chem. Comm.*, 2015, **51**, 13138–13141.
38 D. Reimer and H. B. Bode, *Nat. Prod. Rep.*, 2014, **31**, 154–159.
39 S. Kraus and N. Arber, *Curr. Opin. Pharmacol.*, 2009, **9**, 405–410.
40 P. Martin, I. Marcq, G. Magistro, M. Penary, C. Garcie, D. Payros, M. Boury, M. Olier, J.-P. Nougayrède, M. Audebert, C. Chalut, S. Schubert and E. Oswald, *PLoS Pathog.*, 2013, **9**, e1003437.
41 A. Cougnoux, J. Delmas, L. Gibold, T. Faïs, C. Romagnoli, F. Robin, G. Cuevas-Ramos, E. Oswald, A. Darfeuille-Michaud, F. Prati, G. Dalmasso and R. Bonnet, *Gut*, 2015, DOI: 10.1136/gutjnl-2014-307241.
42 P. Cimermancic, M. H. Medema, J. Claesen, K. Kurita, L. C. Wieland Brown, K. Mavrommatis, A. Pati, P. A. Godfrey, M. Koehrsen, J. Clardy, B. W. Birren, E. Takano, A. Sali, R. G. Linington and M. A. Fischbach, *Cell*, 2014, **158**, 412–421.
43 W. P. Hanage, *Nature*, 2014, **512**, 247–248.

Journal Name

ARTICLE

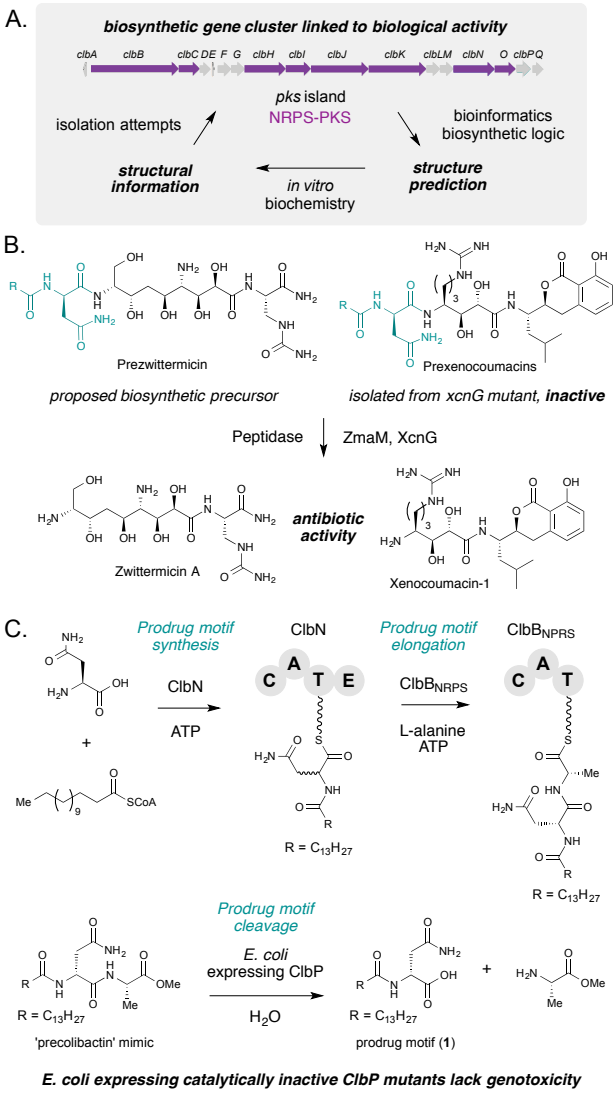

Fig. 1 A) Using bioinformatics and *in vitro* biochemistry to uncover structural features of colibactin. B) 'Prodrug' strategies in antibiotic biosynthesis. C) Characterization of the prodrug resistance mechanism in colibactin assembly.

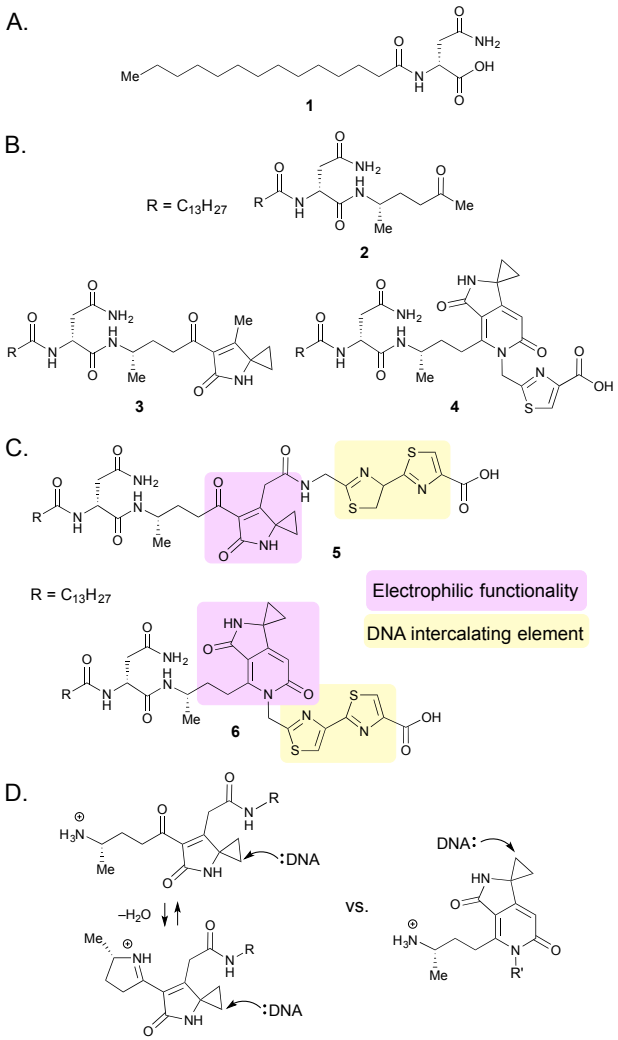

Fig. 2 A) The prodrug motif isolated from wild-type *pks+ E. coli*. B) Selected characterized candidate precolibactins isolated from *clbP* mutants. C) Predicted structures for candidate precolibactins detected in *clbP* mutants. D) Potential reactivity of electrophilic 'warheads' in candidate precolibactins.

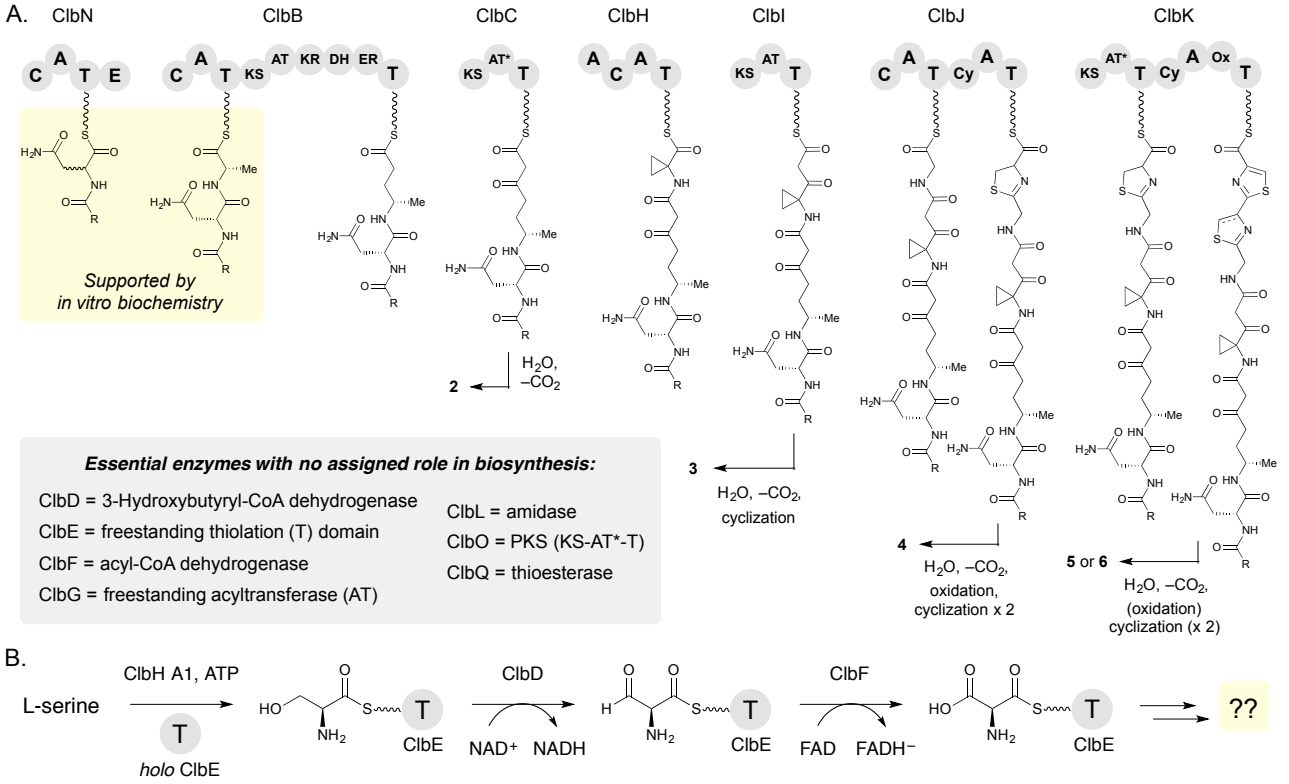

Fig. 3 A) Current hypothesis for the biosynthesis of candidate precolibactins (AT* = atypical-AT). B) Formation of unusual PKS extender unit aminomalonyl-ACP by essential colibactin biosynthetic enzymes.