Published in final edited form as:

Nat Microbiol. ; 1: 15009. doi:10.1038/nmicrobiol.2015.9.

MATE transport of the *E. coli*-derived genotoxin colibactin

Jarrold J. Mousa^{1,†}, Ye Yang^{2,†}, Sarah Tomkovich², Ayaka Shima^{3,4,5,6,7}, Rachel C. Newsome¹, Prabhanshu Tripathi¹, Eric Oswald^{3,4,5,6,7}, Steven D. Bruner^{1,*}, and Christian Jobin^{2,8,*}

¹Department of Chemistry, University of Florida, Gainesville, Florida 32611, USA

²Department of Medicine, University of Florida, Gainesville, Florida 32611, USA

³INRA, USC 1360, F-31300 Toulouse, France

⁴Inserm, UMR1043, F-31300 Toulouse, France

⁵CNRS, UMR5282, F-31400 Toulouse, France

⁶Université de Toulouse, UPS, Centre de Physiopathologie de Toulouse Purpan (CPTP), F-31400 Toulouse, France

⁷CHU Toulouse, Hôpital Purpan, Service de bactériologie-Hygiène, F-31059 Toulouse, France

⁸Department of Infectious Diseases and Pathology, University of Florida, Gainesville, Florida 32611, USA

Abstract

Various forms of cancer have been linked to the carcinogenic activities of microorganisms^{1–3}. The virulent gene island polyketide synthase (*pks*) produces the secondary metabolite colibactin, a genotoxic molecule(s) causing double-stranded DNA breaks⁴ and enhanced colorectal cancer development^{5,6}. Colibactin biosynthesis involves a prodrug resistance strategy where an N-terminal prodrug scaffold (precolibactin) is assembled, transported into the periplasm and cleaved to release the mature product^{7–10}. Here, we show that ClbM, a multidrug and toxic compound extrusion (MATE) transporter, is a key component involved in colibactin activity and transport. Disruption of *clbM* attenuated *pks*⁺ *E. coli*-induced DNA damage *in vitro* and significantly decreased the DNA damage response in gnotobiotic *III10*^{−/−} mice. Colonization experiments

Reprints and permissions information is available online at www.nature.com/reprints.

* Correspondence and requests for materials should be addressed to S.D.B. and C.J. bruner@ufl.edu; christian.jobin@medicine.ufl.edu.

[†]These authors contributed equally to this work.

Author contributions

C.J., S.D.B. and E.O. conceived the project and designed the experiments. J.J.M., Y.Y., S.D.B. and C.J. wrote the manuscript. J.J.M. purified, crystallized, collected data and solved the structure of ClbM. J.J.M. and R.C.N. performed the ethidium fluorescence assay and LC-MS analysis of compound 1 accumulation. J.J.M., P.T. and Y.Y. performed the ClbM cellular localization experiment. Y.Y. made *clbM* and *acrA* in NC101 and performed the *in vitro* infection and mouse colonization studies. S.T. carried out the zebrafish colonization study. A.S. constructed MG*pks*⁺ and MG*pks*⁺ mutants and examined minimal inhibitory concentrations of various antibiotics and toxic compounds. Y.Y. performed statistical analyses. J.J.M. and Y.Y. contributed equally to this work.

Supplementary information is available online.

Competing interests

The authors declare no competing financial interests.

performed in mice or zebrafish animal models indicate that *clbM* is not implicated in *E. coli* niche establishment. The X-ray structure of ClbM shows a structural motif common to the recently described MATE family. The 12-transmembrane ClbM is characterized as a cation-coupled antiporter, and residues important to the cation-binding site are identified. Our data identify ClbM as a precolibactin transporter and provide the first structure of a MATE transporter with a defined and specific biological function.

Microbial dysbiosis has been associated with the development of colorectal cancer (CRC) in both human and mouse models, suggesting that bacterial populations represent an important component of intestinal pathology^{1,2}. Specific microbial genes are directly implicated in tumorigenesis, supporting the concept that microbial activities influence CRC³. The adherent invasive *Escherichia coli* strain NC101 promotes CRC development in azoxymethane-treated *Il10*^{-/-} mice through the action of colibactin, a natural product(s) derived from the genotoxic hybrid polyketide synthase/nonribosomal peptide synthetase (*pks*) gene island⁶. The *pks* island has been found in a higher prevalence in mucosa-adherent *E. coli* isolated from the intestine of CRC patients compared to non-CRC controls, suggesting a potential link between *pks* carrier isolates and tumorigenesis^{6,11}. The ability of *pks*⁺ *E. coli* strains to promote tumorigenesis is linked to the property of colibactin to induce double-stranded DNA breaks and cell senescence in various cell types, including intestinal epithelial cells⁴⁻⁶.

The secondary metabolite colibactin has yet to be fully defined chemically, although several small molecules, including a proposed 'warhead', have recently been linked to the biosynthetic pathway^{9,10,12-17}. The prodrug scaffold *N*-myristoyl-D-asparagine (**1**) has been identified as a periplasmic cleavage product^{9,10}. A transcriptomic analysis of *E. coli* NC101 gene expression using RNA-seq showed that the *pks* island was a highly responsive operon during the development of colitis-associated CRC in *Il10*^{-/-} mice⁵. Interestingly, *clbM* was among the five *pks* genes most responsive to inflammation/cancer⁵. ClbM is homologous to the MATE membrane-embedded transporters, a family largely believed to play a role in cell detoxification, evidenced by their general transport activity against antibiotics and cationic molecules of studied examples¹⁸⁻²¹. To assess the transport capacity of ClbM, we measured minimal inhibitory concentrations (MICs) of a panel of antibiotics and toxic compounds in *E. coli* KAM32 transformed with pk184 (*clbM*) or the empty vector. *E. coli* susceptibility to antibiotics and toxins was not improved by the presence of ClbM, suggesting a more specialized function for this transporter (Supplementary Table 1). Interestingly, overexpression of *clbM* in *E. coli* C43(DE3) *acrAB* prevented accumulation of ethidium, a common indicator of MATE function (Supplementary Fig. 1a). AcrA and AcrB are components of the AcrAB-TolC transport system, which is capable of exporting a wide range of small molecules²². Additionally, ethidium transport was responsive to the concentration of extracellular Na⁺, suggesting that ClbM-mediated transport is cation-coupled (Supplementary Fig. 1a).

Because *pks*⁺ *E. coli* mediate a host genotoxic response in mammalian cells, we focused our attention on the relationship between ClbM and the DNA damage response. To evaluate the effect of ClbM on genotoxicity, we deleted the transporter (*clbM*) in *E. coli* NC101, and

also generated a control transporter mutant (*acrA*) in order to determine the role of the AcrAB-TolC system in precolibactin transport. We infected IEC-6 cells with NC101, *clbM* or *acrA* (multiplicity of infection (MOI) = 30) for 4 h and examined the levels of phosphorylated histone H2AX (γ H2AX), a surrogate marker for DNA damage²³. Immunofluorescent staining analysis showed strong attenuation of γ H2AX in *clbM*-infected cells compared to NC101-infected cells (Fig. 1a). We then used flow cytometry to quantitatively assess γ H2AX induction and found that the DNA damage response was reduced by ~70% in *clbM*-infected cells compared to NC101-infected cells (Fig. 1b,c). Consistent with the attenuated γ H2AX induction, *clbM*-infected cells showed decreased DNA damage, as demonstrated by a single-cell gel electrophoresis assay (comet assay) (Fig. 1d). Similar results were observed using the colonic cell line HT29 (Supplementary Fig. 2). Decreased genotoxic activity was not due to impaired bacterial growth *in vitro* or bacterial attachment and invasion (Supplementary Fig. 3). Importantly, neither γ H2AX induction nor DNA damage appeared to be affected in *acrA* (Fig. 1a–d). Together, these data suggest that ClbM is important for colibactin genotoxic activity. It is worth noting that the partial requirement of *clbM* for the genotoxic effect was not observed at high infectious dose (MOI = 100) (Supplementary Fig. 4), consistent with the previous report⁴.

ClbP, located in the bacterial inner membrane, generates **1** from precolibactin^{7–10}. Western blot analysis of fractionated *E. coli* C43(DE3) membranes expressing deca-histidine-tagged ClbM showed that the transporter was located at the inner membrane (Supplementary Fig. 5). If ClbM is implicated in precolibactin transport to the site of cleavage (ClbP), deletion of *clbM* should impair production of **1**. Indeed, liquid chromatography–mass spectrometry (LC-MS) analysis of bacterial culture media extracts showed that **1** accumulation was significantly reduced in *clbM* compared with NC101 and *acrA* (Fig. 1e,f). To further evaluate the transport capacity of ClbM, we measured **1** accumulation in *E. coli* MG1655 harbouring BAC-*pks* (MG*pks*⁺). Similar to the observation in NC101, disruption of *clbM* reduced the amount of **1** detected in MG*pks*⁺ (Supplementary Fig. 1b). We then deleted *clbH*, a downstream NRPS gene involved in precolibactin assembly. Interestingly, accumulation of **1** was significantly reduced in the double mutant MG*pks*⁺ *clbM clbH* compared to MG*pks*⁺ *clbH* (Supplementary Fig. 1b). These data suggest that ClbM can transport derailed biosynthetic products and ClbP cleavable metabolites (products of *clbH*) as well as intact precolibactin.

To evaluate the role of *clbM* *in vivo*, we colonized germ-free *III0*^{−/−} mice with NC101 or *clbM* by oral gavage (10⁸ c.f.u. per mouse) and examined bacterial colonization and the host DNA damage response. Although *clbM* colonized to the same level as NC101 in the mouse gut 8 days post gavage (Fig. 2a), γ H2AX induction in the colonic epithelium of *clbM*-colonized mice was significantly lower than in NC101-colonized mice (Fig. 2b), suggesting the essential role of *clbM* for the NC101-induced genotoxic response *in vivo*. To visualize bacterial colonization in the gut, we took advantage of the transparent zebrafish larva model. NC101 or *clbM* carrying ptdTomato were introduced statically into zebrafish (5 days post fertilization (d.p.f.)). A period of 24 h later, consistent with the observation in mice, NC101 and *clbM* colonized equally well in zebrafish larvae (Fig. 2c and

Supplementary Videos 1 and 2). These observations suggest that *clbM* does not affect the capacity of NC101 to occupy a microbial niche but is important for the genotoxic effect.

The key role of *clbM* in *pks*-mediated genotoxicity led us to further probe the structure and mechanism of the transporter. ClbM was crystallized using lipid cubic phase crystallization²⁴ and the crystal structure was solved to 2.7 Å resolution (Supplementary Table 2). The overall topology consists of two pseudo-symmetrical lobes (*N*-lobe, TH1–TH6; *C*-lobe, TH7–TH12) and adopts a V-shaped conformation (Fig. 3a,b). Based on the ‘positive-inside’ rule attributed to membrane protein topology, the central cavity in our crystal form is open to the periplasmic environment²⁵. Only three unique homologue structures of the broad MATE family have so far been reported, including the cation transporter homologues NorM-VC and NorM-NG^{26,27} and proton transporters DinH and PfMATE^{28,29}. ClbM is topologically most similar to the PfMATE and NorM-VC/-NG transporters, yet shares low sequence identity with both (Supplementary Fig. 6). Several lipid molecules from the crystallization medium are bound in the inter-lobe cleft, revealing a large binding pocket (Fig. 3b). The binding pocket spans nearly 40 Å, noticeably larger than NorM and PfMATE (Supplementary Fig. 7), highlighting possible ClbM specificity for a postulated large 700–900 Da precolibactin molecule^{15,30}.

To identify the cation-binding site, key to the antiport mechanism, the protein was crystallized with Rb⁺ as the sole cation and the structure was solved to 3.3 Å. An isomorphous difference map revealed strong electron density on the C-terminal lobe, indicating a specifically bound Rb⁺ (Fig. 3c). No significant structural changes were observed between the native and Rb⁺-bound structure, consistent with the transporter in the drug-bound cation-bound state. The ClbM cation-binding site coordinates Rb⁺ remarkably differently from NorM-VC/-NG, appearing to be bound solely through π -cation interactions with Y277(TH7) and F292(TH8), although D299(TH8) is within 7 Å. ClbM TH8 is kinked to bring both F292 and D299 in close proximity to the bound cation. Compared to ClbM, NorM-VC binds Rb⁺ through Y367(TH10) oxygen–cation electrostatic interactions and F288(TH8) and F259(TH7) π -cation interactions; NorM-NG coordinates Cs⁺ through a homologous Y294(TH8) and a substrate-bound-shifted E261(TH7). Comparing ClbM and NorM-VC/-NG structures revealed this unique site for ClbM cation binding (Supplementary Fig. 8a–c). Although unique residues are present in ClbM, an evolutionary conserved mechanism of cation binding is present, whereby π -cation interactions are used as primary sites of cation coordination. Furthermore, ClbM shares a conserved TH1 proline residue (P38), which has been demonstrated to be a critical switch point for drug release in the proton-coupled PfMATE²⁹. TH1 may also be important for precolibactin export, where cation binding could trigger collapse of the drug-binding pocket. TH1 of ClbM maintains a straight conformation, similar to PfMATE(straight), further suggesting this structure of ClbM is in the drug-bound cation-bound state (Supplementary Fig. 8d).

Based on topological comparison between our structure, NorM-VC/-NG and PfMATE, a proposed mechanism for precolibactin transport by ClbM can be deduced. Our structure of ClbM is in the cation-bound/drug-bound state. Following this conformation, TH7 and TH8 can engage D299(TH7), which will collapse the drug-binding pocket, involving a TH1 conformational change around P38. Following drug release, a conformational change to the

intracellular open state triggers cation release and cytoplasmic precolibactin binding, leaving ClbM in the drug-bound intracellular open state. A conformational shift will return ClbM to the cation-free drug-bound state, where cation binding will provide the starting point. Precolibactin release in the periplasmic space will serve as substrate for the peptidase ClbP, which will generate activated colibactin and the cleavage product **1** (Supplementary Fig. 9). Speculatively, extracellular release of both compounds may be accomplished through porin or resistance-nodulation-division-based export.

Here, we demonstrate that the transporter ClbM plays a major role in mediating colibactin-induced genotoxicity. ClbM is the first MATE transporter structurally characterized that has been related to a specific biological function, namely transporting precolibactin from the cytoplasm to the periplasmic space. ClbM is a functionally unique MATE transporter, yet shares a predicted evolutionary conserved mechanism with the cation-coupled NorM-family of MATE transporters. With increasing evidence that microorganisms, including *pks+* *E. coli* strains, are involved in intestinal carcinogenesis, generating small-molecule inhibitors of MATE transporters could represent an attractive means with which to modulate tumorigenesis.

Methods

Ethics statement

All animal experiments were approved by the Institutional Animal Care and Use Committee of the University of Florida (protocols nos. 201308038 and 201308035).

Mutagenesis

NC101 *clbM* and NC101 *acrA* were generated using the λ RED recombination technique³¹. Briefly, the chloramphenicol resistant gene with flanking flippase recombination target (FRT) sites were polymerase chain reaction (PCR) amplified from the pKD3 plasmid using the primer sets *clbMKO* (forward, 5'-TGT TAA CGC AGG TGA CCT ATG CCA TCG CCG ATA TTT TTT GGC TGA GCC ACg tgt agg ctg gag ctg ctt c-3'; reverse, 5'-GGG AAA ACC AGC CCC ACA CCG GCA ATG ATG CCG CTG TTG GTG ACATCA Ata tgg gaa tta gcc atg gtc c-3') and *acrAKO* (forward, 5'-TTA ACT TTT GAC CAT TGA CCA ATT TGA AAT CGG ACA CTC GAG GTT TAC ATg tgt agg ctg gag ctg ctt c-3'; reverse, 5'-TCG GGC GAT CGATAA AGA AAT TAG GCATGT CTTAAC GGC TCC TGT TTA Aga tgg gaa tta gcc atg gtc c-3'). Each of the primers contains 50 nucleotides (indicated by capital letters) that are homologous to the upstream or downstream of the genetic regions to be deleted. The PCR products were gel-purified, digested with DpnI and transformed into *E. coli* NC101, which had previously been transformed with pKD46, which encodes the λ RED recombinase. Chloramphenicol-resistant clones were selected for successful genomic replacement of the first 177 nucleotides of *clbM* (to minimize the polar effect) or the entire *acrA*. pCP20, which encodes a thermally inducible flippase (FLP) recombinase, was then transformed into bacteria to remove the chloramphenicol-resistant cassette to generate NC101 *clbM* and NC101 *acrA*. Deletion of *clbM* (1–177 bp) and *acrA* on the *E. coli* NC101 chromosome was confirmed by PCR and sequencing.

E. coli MG1655 harbouring BAC-*pks* (MG*pks*⁺) was constructed as previously described³². MG*pks*⁺ *clbM*, MG*pks*⁺ *clbH* and MG*pks*⁺ *clbH clbM* were constructed by λ RED recombination using the primer set *clbM* (forward, 5'-TCA TGG TGC CAC CTC CAT ATC CAA GTG TTG GGA GCG CAC CGC CTC CTG ATc ata tga ata tcc tcc tta g-3'; reverse, 5'-TCATGG TGC CAC CTC CATATC CAA GTG TTG GGA GCG CAC CGC CTC CTG ATc ata tga ata tcc tcc tta g-3') and *clbH* (forward, 5'-ATG GA ACA GCA AGG GAT TAT GAG ACA GTT GCC TAC CGA CGA CCA AAC GAT gtg tag gct gga gct gct tc-3'; reverse, 5'-CGG TAT CCA CGG ATC TGC ACT TGG CGA TCT TGT CTG CCG AGA TAG AAC AAc ata tga ata tcc tcc tta g-3') and pKD4 as a PCR template. Kanamycin cassettes in *clbM* of MG*pks*⁺ *clbM* and MG*pks* *clbH clbM* and in *clbH* of MG*pks*⁺ *clbH* were removed by flp recombination.

Bacterial growth evaluation

To monitor bacterial growth, 1 ml overnight cultures of NC101, NC101 *clbM* and NC101 *acrA* were diluted 1:100 with fresh Luria-Bertani (LB) medium and shaken at 37 °C. Cultures (1 ml) were collected at various time points and the optical density at 600 nm (OD₆₀₀) was obtained using NanoDrop2000 (Thermo Scientific). Growth curves were generated and analysed using GraphPad Prism 5.0.

In vitro infection

IEC-6 [IEC6] (ATCC CRL1592) or HT29 (ATCC HTB38) were obtained from ATCC. Both cell lines were tested to be negative of Mycoplasma using the MycoAlert Mycoplasma Detection Kit (LT07–218, Lonza). HT29 was authenticated using the *GenePrint* 10 System (B9510, Promega). IEC-6 was not authenticated.

Cells were plated at 5×10^4 per well in eight-chamber culture slides for γ H2AX immunostaining, at 10^5 per well in 24-well tissue culture plates for bacterial attachment/invasion quantification (for infection of IEC-6 cells), or at 5×10^5 per well in six-well tissue culture plates for comet assay and γ H2AX flow cytometry. Cells were allowed to adhere and then serum-starved overnight (16–18 h). Cells were infected with log phase NC101, NC101 *pks*, NC101 *clbM* or NC101 *acrA* at the indicated MOI, determined by OD₆₀₀, and incubated for 4 h at 37 °C in a cell incubator under humidity and 5% CO₂.

To assess bacterial attachment, cells were washed with sterile PBS three times and lysed in deionized H₂O containing 1% Triton X-100 for 5 min in the incubator. Lysate was diluted in sterile PBS and plated on LB agar to enumerate adhering bacteria. To quantify invasion, infected cells were washed with PBS and incubated in media supplemented with gentamycin (200 μ g ml⁻¹) for 1 h in the incubator. Cells were washed with PBS again and lysed as described above. Lysate was diluted in sterile PBS and plated on LB agar to enumerate bacterial invasion. Data were analysed using GraphPad Prism 5.0.

Examination of γ H2AX in vitro

For γ H2AX immunostaining, infected cells were washed with sterile PBS three times and incubated in gentamycin-supplemented medium overnight in the incubator. Cells were fixed and permeabilized using ice-cold methanol (10 min), washed with PBS to remove methanol,

blocked with PBS containing 0.1% Tween 20 and 5% goat serum for 1 h at room temperature, and incubated with the Phospho-Histone H2AX (Ser139) (20E3) rabbit monoclonal antibody (9718, Cell Signalling) overnight at 4 °C. Cells were washed with PBS and incubated in the secondary Alexa Fluor 488 goat anti-rabbit IgG (H+L) antibody (A11034, Life Technologies) for 2 h at room temperature. Cells were washed again with PBS and mounted with VECTASHIELD mounting medium with DAPI (H-120, Vector Labs). Images were taken using a Leica DM6000B upright microscope.

γ H2AX flow cytometry was performed as previously described³³, with minor modifications. Briefly, after overnight incubation with gentamycin-supplemented medium, cells were lifted using trypsin and harvested. Cells were fixed in ice-cold 2% methanol-free formaldehyde solution (28908, Thermo Scientific Pierce) for 15 min, washed in PBS, and incubated in 70% ethanol overnight at -20 °C. Cells were washed with BSA-T-PBS (PBS containing 1% BSA and 0.2% Triton X-100), and incubated with the Alexa Fluor 647 anti-H2A.X-phosphorylated (Ser139) antibody (613407, BioLegend, diluted 1:100 in BSA-T-PBS) overnight at 4 °C. Cells were washed with BSA-T-PBS and PI stained. At least 20,000 cells from each sample were analysed using the Accuri C6 flow cytometer (BD Biosciences). Data were further processed using FCS Express 4 software. For estimation of NC101 *clbM* ability to induce γ H2AX, a binary system was established by assuming 0% induction in control uninfected cells and 100% in NC101-infected cells, and data from four independent infection/flow cytometry experiments were averaged. Data were analysed using GraphPad Prism 5.0.

Comet assay

Comet assay was performed following the manufacturer's protocol (Trevigen). Briefly, after overnight incubation with gentamycin-supplemented medium, infected cells were collected using trypsin. Cells were mixed with LMAgarose (Trevigen) and spread on 20-well CometSlides (Trevigen). CometSlides were incubated in the lysis buffer (Trevigen) for 1 h at 4 °C to lyse cells, and then incubated in alkaline electrophoresis solution (deionized H₂O containing 200 mM NaOH and 1 mM EDTA) for 20 min at room temperature. Electrophoresis was performed using the CometAssay Electrophoresis System II (Trevigen). DNA was stained with SYBR Gold nucleic acid gel stain (S-11494, Life Technologies) and visualized using the Leica DM6000B upright microscope.

LC-MS detection of *N*-myristoyl-D-asparagine

E. coli (100 ml) cells, grown overnight to OD₆₀₀ = 1, were pelleted by centrifugation. The medium was extracted with ethyl acetate (1:1, vol:vol) using a separatory funnel. The ethyl acetate fraction was removed by rotoevaporation, and dried extracts were redissolved in 3 ml methanol. Pelleted cells were extracted in 2 ml methanol, and combined extracts were run on an Agilent 6130 quadrupole LC-MS with an Agilent Zorbax SB-C18 1.8 μ m (2.1 $\times$ 50 mm) column. The running method was 0–2 min (H₂O/0.1% formic acid), 2–22 min (0–100% acetonitrile (ACN)/0.1% formic acid). Compound **1** eluted from the column at 20 min.

***E. coli* colonization in mice or zebrafish**

Germ-free *III0*^{-/-} mice (129/SvEv, 8–12 weeks old, mixed gender) were transferred to a specific-pathogen-free (SPF) room (CGRC 545, University of Florida) and randomly colonized with NC101 ($n = 6$) or *clbM* ($n = 8$) by oral gavage (10^8 per mouse). Ungavaged mice ($n = 5$) were used as controls. This sample size was based on a previous study showing higher γ H2AX in NC101 compared to NC101 *pks*-colonized AOM/*III0*^{-/-} mice⁶. Mice were killed 8 days after gavage and stools were collected. Mouse colon tissues were Swiss-rolled and fixed in formalin. Swiss rolls were embedded in paraffin and sectioned by the Molecular Pathology Core at the University of Florida. Stools were homogenized, diluted in sterile PBS, and plated on MacConkey agar plates to enumerate Enterobacteriaceae (blinded). Data were analysed using GraphPad Prism 5.0. The mouse experiment was repeated once.

Zebrafish (AB/Tü wild-type strain) embryos/larvae were treated with 0.015 mg l^{-1} n-phenylthiourea (PTU) (Sigma) to prevent pigmentation and were unfed for the duration of the experiment. 1 d.p.f. zebrafish were sorted into T25 flasks containing 30 ml autoclaved gnotobiotic zebrafish medium²⁸ (GZM; dH₂O, InstantOcean salt and Seachem liquid neutral regulator) at a density of 30 fish per flask. 5 d.p.f. zebrafish were left untreated or colonized with labelled bacteria (*E. coli* NC101 or *clbM* transformed with ptdTomato:ampicillin (632531, Takara Clontech)), grown from two overnight cultures, at a density of 6×10^7 c.f.u. ml⁻¹. 6 d.p.f. larvae (24 h post colonization) were euthanized in Tricaine-S (MS-222) (Western Chemical) at 300 mg ml^{-1} . Intestines were then dissected from each treatment group, homogenized by repeated passage with a 27-gauge insulin syringe, serially diluted and plated onto LB agar and LB agar with ampicillin ($100 \text{ } \mu\text{g ml}^{-1}$) to determine colonization levels (blinded). A total of 18 fish guts from three independent experiments were analysed for each group. This sample size was selected based on a previous study investigating gut microbe c.f.u. counting and imaging³⁴. Data were analysed using GraphPad Prism 5.0.

Examination of γ H2AX *in vivo*

γ H2AX in mouse colon Swiss-roll sections was examined by immunohistochemistry (IHC) as described previously⁶. Briefly, γ H2AX was detected using the Phospho-Histone H2A.X (Ser139) (20E3) rabbit monoclonal antibody (9718, Cell Signalling), followed Vectastain Elite ABC Rabbit IgG Kit (Vector Labs) for IHC. Images were taken at 500 $\times$ power using an Olympus CX41 upright microscope. Swiss-roll sections were divided into 13 microscopic fields under $\times 40$ objective on a Leica DM5500B upright microscope. γ H2AX positive cells in each field were counted (blinded) and averaged. Data were analysed using GraphPad Prism 5.0.

***E. coli* imaging in zebrafish**

For whole mount immunofluorescence, zebrafish were euthanized and fixed in 4% paraformaldehyde (methanol-free) for 24 h. After fixation, embryos were washed with PBS/Tween 20 and either dehydrated in methanol for storage or immediately stained for 30 min with 1:1,000 diamidino-2-phenylindole (DAPI) solution (Fisher, 1 mg ml^{-1}). Fish were incubated in FocusClear (CedarLane) overnight before imaging. Fish were mounted in 1%

lowmelting-point agarose on glass-bottomed dishes (MatTek Corporation) and imaged using a Leica TCS-SP5 confocal microscope ($\times 10$ objective; z slice size, 1 μm). The z -stack images were reconstructed using Volocity 6.3 software (PerkinElmer).

Drug susceptibility test

E. coli KAM32^{35,36} with *acrA* and *ydhD* deleted from the genome, transformed with pK184(*clbM*), pBACpks or empty vector controls, were used for a chemical susceptibility test. *E. coli* KAM32 harbouring pMEC2, which expressed efflux pump YdhE, was used as positive control⁷. The MIC was defined as the lowest concentration of a chemical that completely stopped visible growth of the bacteria. Bacteria were pre-cultured overnight at 37 °C in LB broth with the appropriate antibiotic. A 50 ml bacterial culture was inoculated into 4 ml LB broth and cultivated for 3 h. The bacterial culture was diluted with LB broth at a concentration of 10^7 c.f.u. ml^{-1} , and 5 ml diluted culture was added into 100 ml LB broth containing twofold serial diluted chemicals in 96 well plates. The growth of *E. coli* was examined by visual inspection after 24 h of incubation at 37 °C.

Ethidium fluorescence assay

The drug-sensitive strain *E. coli* C43(DE3) *acrAB* harbouring the *ClbM*-pET16b plasmid was grown to $\text{OD}_{600} = 0.6$, and *ClbM* expression was induced with 500 μM isopropyl β -D-thiogalactopyranoside (IPTG) for 2 h at 37 °C. Following this, 1 ml aliquots of cells were collected by centrifugation at 50,000 g and resuspended in 50 mM Tris-HCl pH 7.5. Cells were incubated with 200 μM ethidium bromide for 1 h. After incubation, the cells were isolated and washed three times with buffer, using centrifugation at 20,000 g between washes to isolate the cells. The cells were then resuspended in 3% trifluoroacetic acid to lyse the cells and release the fluorescent content. The cellular debris was removed by centrifugation at 20,000 g for 10 min and the fluorescence was measured on a FluoroMax-3 fluorometer (Horiba). The excitation and emission wavelengths were 485 nm and 595 nm, respectively²⁰. Transport activity was scaled to the lowest ethidium accumulation inside the cells. Three independent experiments were performed. Data were analysed using GraphPad Prism 5.0.

Cloning, expression and protein purification

The *clbM* gene was cloned from a bacterial artificial chromosome (BAC) harbouring the colibactin *pks* island and digested with NcoI/NheI restriction enzymes for ligation into a modified pET16b vector containing a C-terminal deca-histidine tag preceded by an enterokinase cleavage site. Primers used for cloning *clbM* into expression vectors are as follows: pk184-F, 5'-ATA TGA ATT CGC GGC GCA TTT AAT GAA TAA GG-3'; pk184-R, 5'-GGA ACA GTG CCA CAT CAT TGA TCG GAT CCA TA-3'; pET16b-F, 5'-CTA GTG GCT AGC TGG TGC CAC CTC CAT ATC CAA-3'; pET16b-R, 5'-CTA GTG GCT AGC TGG TGC CAC CTC CAT ATC CAA-3'.

C43(DE3) cells harbouring the *clbM*-pET16b plasmid were grown to an OD_{600} of 0.6 and protein expression was induced with 250 μM IPTG at 20 °C for overnight expression. Cells were collected by centrifugation at 8,000 g and lysed using a microfluidizer (M-110L Microfluidics). The insoluble material and inclusion bodies were removed by centrifugation at 15,000 g for 30 min. Following this, cell membranes were collected by centrifugation at

50,000g for 60 min. Membranes were solubilized in 0.5% *n*-dodecyl- β -D-maltopyranoside (DDM), 150 mM NaCl, 30% glycerol and 50 mM HEPES-NaOH pH 7.5 for 3 h. Insoluble material was removed by centrifugation at 50,000g for 60 min and the solubilized membranes were incubated with HisPur Ni-NTA resin (Thermo) overnight after the NaCl concentration was brought to 800 mM. After overnight incubation, the resin was washed with 5 $\times$ 10 ml buffer containing 0.05% DDM, 10% glycerol, 25 mM imidazole, 800 mM NaCl and 25 mM HEPES-NaOH pH 7.5. The protein was eluted in the same buffer with 500 mM imidazole. ClbM was immediately purified to homogeneity by size exclusion chromatography (Superdex 200 prep grade HiLoad 16/600, GE Healthcare) in buffer containing 0.05% DDM, 10% glycerol, 150 mM NaCl and 10 mM HEPES pH 7.5.

For a cell localization control, the *entS* gene was cloned from *E. coli* BL21(DE3) with an engineered C-terminal deca-His-tag and ligated into the NheI/NcoI sites of pET16b for overexpression. The *entS*-pET16b plasmid was transformed into *E. coli* BL21(DE3), which was grown to an OD₆₀₀ of 0.6 at 37 °C. Expression was initiated by the addition of 80 μ M IPTG and growth continued for 4 h at 37 °C before the cells were collected by centrifugation. Cell pellets were resuspended in 20 mM Tris-HCl pH 7.5 and 0.5 M NaCl. Primers used for cloning *entS* into expression vectors are as follows: C-EntS-NheI pET16b, 5'-GCG CTC GAG TTA ATG ATG ATG ATG ATG ATG ATG ATG ATG TGC GGC CGC AAG CTT GTC GTC GTC GTC GCT AGC ACT GTC GGA CGC TGT and N-EntS-NcoI pET16b, 5'-GCG CCA TGG ATA AAC AAT CCT GGC.

Inner and outer membrane fractionation and western blot

Inner and outer membranes were separated as described³⁷. Briefly, following IPTG induction, *E. coli* C43(DE3)³⁸ carrying *clbM*-pET16b or BL21(DE3) carrying *entS*-pET16b were lysed using a microfluidizer (M-110L, Microfluidics). Inclusion bodies were removed by centrifugation at 15,000g for 30 min at 4 °C. Bacterial membranes were collected by centrifuging the supernatant at 50,000g for 60 min at 4 °C. Bacterial membranes were homogenized (Dounce homogenizer) in 2% *N*-lauroylsarcosine and incubated for 30 min on ice. Outer membranes were separated by centrifuging at 50,000g for 60 min at 4 °C. Outer membranes were dissolved in 50 mM Tris-HCl pH 7.5 and 2% SDS. Sample protein concentrations were determined using Bio-Rad Protein Assay Dye Reagent Concentrate (#500-0006, Bio-Rad).

Cellular membrane proteins (5 or 10 μ g) were loaded onto 15% SDS polyacrylamide gel electrophoresis (PAGE) gels. The Precision Plus Protein Kaleidoscope Prestained Standard (#161-0375, Bio-Rad) was used as molecular weight standard. One SDS-PAGE gel was stained for 1 h using Coomassie Brilliant Blue G (B0770 Sigma; 0.1% Coomassie Brilliant Blue, 20% methanol and 10% acetic acid in diH₂O). The gel was destained in 50% methanol, 10% acetic acid in diH₂O. One gel was used for western blot. Briefly, proteins were transferred to a nitrocellulose membrane (no. 162-0115, Bio-Rad). The membrane was blocked in 5% milk Tris-buffered saline-Tween 20 (TBST) for 1 h and incubated with His? Tag Monoclonal Antibody (no. 70796-3, EMD Millipore; diluted 1:5,000 in 5% milk TBST) at 4 °C overnight. The membrane was washed with TBST and incubated with goat anti-mouse IgG-HRP (sc-2005, Santa Cruz Biotechnology; diluted 1:5,000 in TBST) for 1 h at

room temperature. The membrane was then washed five times with TBST and incubated with Western Lightning *Plus* ECL (no. NEL105001EA, PerkinElmer). Blot was performed using blue devil autoradiography film (30-101L, GeneseeScientific).

Crystallization

ClbM purified from size exclusion chromatography was concentrated to 30 mg ml⁻¹ with a 50 kDa molecular weight (MW) centrifugal concentrator (Millipore). Protein was mixed with melted monoolein (Sigma) by hand using the twin-syringe method at a 2/3 protein to lipid ratio (wt/wt). Crystallization trays were prepared manually using a 50× repeating dispenser coupled to a 10 µl syringe. LCP boluses of 200 nl were dispensed onto 96-well glass sandwich plates and overlaid with 1 µl of precipitant solution. Extensive crystal screening was conducted and one initial hit was obtained in 30% PEG 500 DME, 100 mM Li₂SO₄, 100 mM NaCl and 100 mM Tris pH 8.0. Optimization of this condition gave crystals that were 20 × 30 µm² in size from 30% PEG 500 DME, 100 mM Li₂SO₄, 400 mM NaCl and 100 mM sodium cacodylate pH 6.8. Crystals appeared after one day and grew to full size in three days (Supplementary Fig. 12A). For collection, crystals were obtained directly from the sponge phase and flash frozen in liquid nitrogen, using the surrounding sponge phase as a cryoprotectant. Crystals containing rubidium chloride were grown using similar precipitation conditions, but the sodium chloride was replaced with 400 mM rubidium chloride. Crystals grew to 10–20 µm² in size and were collected in a similar manner.

Data collection and structure determination

X-ray diffraction data were collected at the Advanced Photon Source GM/CA-CAT beamline 23-ID-D using a beam size of 10 × 10 µm². Approximately 50 crystals were screened for diffraction, with most crystals diffracting to 3.0–3.5 Å. One crystal diffracted to 2.7 Å, and this single crystal was used to solve the apo structure of ClbM (Supplementary Fig. 12B). The structure of ClbM was determined by molecular replacement using the PfMATE straight conformation as a starting search model in Phaser³⁹ (Protein Data Bank accession no. [3VVN](#)). The model for ClbM was improved through iterative refinements in Phenix and manual building in Coot⁴⁰, guided by unbiased composite omit maps. The ClbM-Rb structure was solved to 3.3 Å using diffraction data collected from a single crystal using the native ClbM structure as a search model. The structure was verified by a composite omit map.

Accession numbers

ClbM structures have been deposited in the Protein Data Bank under the accession codes [4Z3N](#) for native ClbM and [4Z3P](#) for the Rb⁺-bound structure.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by funding from R01 DK73338 (C.J.), R01DK47700 (C.J.), R01 GM0865700 (S.D.B.) and by the University of Florida (S.D.B.) and the UF DoM Gatorade Fund (C.J.). The authors thank V. de Crecy-Lagard for discussions.

References

1. Jobin C. Colorectal cancer: looking for answers in the microbiota. *Cancer Discov.* 2013; 3:384–387. [PubMed: 23580283]
2. Schwabe RF, Jobin C. The microbiome and cancer. *Nature Rev. Cancer.* 2013; 13:800–812. [PubMed: 24132111]
3. Sears CL, Garrett WS. Microbes, microbiota, and colon cancer. *Cell Host Microbe.* 2014; 15:317–328. [PubMed: 24629338]
4. Nougayrède J-P, et al. *Escherichia coli* induces DNA double-strand breaks in eukaryotic cells. *Science.* 2006; 313:848–851. [PubMed: 16902142]
5. Arthur JC, et al. Microbial genomic analysis reveals the essential role of inflammation in bacteria-induced colorectal cancer. *Nature Commun.* 2014; 5:4724. [PubMed: 25182170]
6. Arthur JC, et al. Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science.* 2012; 338:120–123. [PubMed: 22903521]
7. Dubois D, et al. ClbP is a prototype of a peptidase subgroup involved in biosynthesis of nonribosomal peptides. *J. Biol. Chem.* 2011; 286:35562–35570. [PubMed: 21795676]
8. Cougnoux A, et al. Analysis of structure–function relationships in the colibactin-maturing enzyme ClbP. *J. Mol. Biol.* 2012; 424:203–214. [PubMed: 23041299]
9. Brotherton CA, Balskus EP. A prodrug resistance mechanism is involved in colibactin biosynthesis and cytotoxicity. *J. Am. Chem. Soc.* 2013; 135:3359–3362. [PubMed: 23406518]
10. Bian X, et al. *In vivo* evidence for a prodrug activation mechanism during colibactin maturation. *ChemBioChem.* 2013; 14:1194–1197. [PubMed: 23744512]
11. Buc E, et al. High prevalence of mucosa-associated *E. coli* producing cyclomodulin and genotoxin in colon cancer. *PLoS ONE.* 2013; 8:e56964. [PubMed: 23457644]
12. Vizcaino MI, Engel P, Trautman E, Crawford JM. Comparative metabolomics and structural characterizations illuminate colibactin pathway-dependent small molecules. *J. Am. Chem. Soc.* 2014; 136:9244–9247. [PubMed: 24932672]
13. Brotherton CA, Wilson M, Byrd G, Balskus EP. Isolation of a metabolite from the pks island provides insights into colibactin biosynthesis and activity. *Org. Lett.* 2015; 17:1545–1548. [PubMed: 25753745]
14. Bian X, Plaza A, Zhang Y, Mueller R. Two more pieces of the colibactin genotoxin puzzle from *Escherichia coli* show incorporation of an unusual 1-aminocyclopropanecarboxylic acid moiety. *Chem. Sci.* 2015:3154–3160. [PubMed: 28706687]
15. Vizcaino MI, Crawford JM. The colibactin warhead crosslinks DNA. *Nature Chem.* 2015; 7:411–417. [PubMed: 25901819]
16. Li Z-R, et al. Critical intermediates reveal new biosynthetic events in the enigmatic colibactin pathway. *ChemBioChem.* 2015; 16:1715–1719. [PubMed: 26052818]
17. Brachmann AO, et al. Colibactin biosynthesis and biological activity depend on the rare aminomalonyl polyketide precursor. *Chem. Commun.* 2015; 3034:13138–13141.
18. Kaatz GW, McAleese F, Seo SM. Multidrug resistance in *Staphylococcus aureus* due to overexpression of a novel multidrug and toxin extrusion (MATE) transport protein. *Antimicrob. Agents Chemother.* 2005; 49:1857–1864. [PubMed: 15855507]
19. McAleese F, et al. A novel MATE family efflux pump contributes to the reduced susceptibility of laboratory-derived *Staphylococcus aureus* mutants to tigecycline. *Antimicrob. Agents Chemother.* 2005; 49:1865–1871. [PubMed: 15855508]
20. Long F, Rouquette-Loughlin C, Shafer WM, Yu EW. Functional cloning and characterization of the multidrug efflux pumps NorM from *Neisseria gonorrhoeae* and YdhE from *Escherichia coli*. *Antimicrob. Agents Chemother.* 2008; 52:3052–3060. [PubMed: 18591276]

21. Kuroda T, Tsuchiya T. Multidrug efflux transporters in the MATE family. *Biochim. Biophys. Acta*. 2009; 1794:763–768. [PubMed: 19100867]
22. Nikaido H, Zgurskaya HI. AcrAB and related multidrug efflux pumps of *Escherichia coli*. *J. Mol. Microbiol. Biotechnol.* 2001; 3:215–218. [PubMed: 11321576]
23. Kinner A, Wu W, Staudt C, Iliakis G. Gamma-H2AX in recognition and signaling of DNA double-strand breaks in the context of chromatin. *Nucleic Acids Res.* 2008; 36:5678–5694. [PubMed: 18772227]
24. Cherezov V. Lipidic cubic phase technologies for membrane protein structural studies. *Curr. Opin. Struct. Biol.* 2011; 21:559–566. [PubMed: 21775127]
25. Von Heijne G. Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. *J. Mol. Biol.* 1992; 225:487–494. [PubMed: 1593632]
26. He X, et al. Structure of a cation-bound multidrug and toxic compound extrusion transporter. *Nature*. 2010; 467:991–994. [PubMed: 20861838]
27. Lu M, et al. Structures of a Na⁺-coupled, substrate-bound MATE multidrug transporter. *Proc. Natl Acad. Sci. USA*. 2013; 110:2099–2104. [PubMed: 23341609]
28. Lu M, Radchenko M, Symersky J, Nie R, Guo Y. Structural insights into H⁺-coupled multidrug extrusion by a MATE transporter. *Nature Struct. Mol. Biol.* 2013; 20:1310–1317. [PubMed: 24141706]
29. Tanaka Y, et al. Structural basis for the drug extrusion mechanism by a MATE multidrug transporter. *Nature*. 2013; 496:247–251. [PubMed: 23535598]
30. Li Z, et al. Critical intermediates reveal novel biosynthetic events in the enigmatic colibactin pathway. *ChemBioChem*. 2015; 16:1715–1719. [PubMed: 26052818]
31. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA*. 2000; 97:6640–6645. [PubMed: 10829079]
32. Martin P, et al. Interplay between siderophores and colibactin genotoxin biosynthetic pathways in *Escherichia coli*. *PLoS Pathog.* 2013; 9:e1003437. [PubMed: 23853582]
33. Huang X, Darzynkiewicz Z. Cytometric assessment of histone H2AX phosphorylation: a reporter of DNA damage. *Methods Mol. Bio.* 2006; 314:73–80. [PubMed: 16673875]
34. Semova I, et al. Microbiota regulate intestinal absorption and metabolism of fatty acids in the zebrafish. *Cell Host Microbe*. 2012; 12:277–288. [PubMed: 22980325]
35. Chen J, et al. VmrA, a member of a novel class of Na⁺-coupled multidrug efflux pumps from *Vibrio parahaemolyticus*. *J. Bacteriol.* 2002; 184:572–576. [PubMed: 11751837]
36. Morita Y, et al. NorM, a putative multidrug efflux protein, of vibrio parahaemolyticus and its homolog in *Escherichia coli*. *Antimicrob. Agents Chemother.* 1998; 42:1778–1782. [PubMed: 9661020]
37. Filip C, Fletcher G, Wulff JL, Earhart CF. Solubilization of the cytoplasmic membrane of *Escherichia coli* by the ionic detergent sodium-lauryl sarcosinate. *J. Bacteriol.* 1973; 115:717–722. [PubMed: 4580564]
38. Miroux B, Walker JE. Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* 1996; 260:289–298. [PubMed: 8757792]
39. Adams PD, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr.* 2010; 66:213–221.
40. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr.* 2004; 60:2126–2132.

Figure 1. *clbM* is important for in vitro colibactin activity and production

a, Representative images of γ H2AX immunofluorescence staining show that γ H2AX induction is attenuated in *clbM*-infected cells compared to NC101-infected cells. IEC-6 cells were infected for 4 h (MOI = 30) and incubated in gentamycin-supplemented medium overnight before analyses. Images were taken from one of four independent experiments. Scale bars, 15 μ m. **b**, Representative γ H2AX flow cytometry histogram showing reduced γ H2AX induction in *clbM*-infected cells. **c**, *clbM* retains $34.05 \pm 7.32\%$ ($\pm$ s.d.) capability to induce γ H2AX *in vitro*. For the estimation, a binary system was established by assuming 0% induction in control uninfected cells and 100% in NC101-infected cells, and data from four independent infection/flow cytometry experiments were averaged. * $P < 0.0005$ compared to NC101 and ** $P < 0.005$ compared to control, one sample *t*-test. Error bars indicate s.d. **d**, *clbM* induces less DNA damage (visualized by shorter comet tails) than NC101. Scale bars, 30 μ m. **e**, LC-MS chromatogram of the SIM +343 *m/z* signal for compound 1 from one representative experiment. Three independent experiments were performed. **f**, Compound 1 accumulation is significantly reduced in *clbM*. NS, not significant. * $P < 0.01$, one-way ANOVA followed by Newman–Keuls multiple comparison test. Error bars indicate s.d.

Figure 2. *clbM* is essential for NC101-induced genotoxic response *in vivo*

a, NC101 ($n = 6$) and *clbM* ($n = 8$) colonize to similar levels in the mouse intestine. Eight days after gavage with NC101 or *clbM*, mouse stools were plated on MacConkey agar to enumerate Enterobacteriaceae. No Enterobacteriaceae were cultured from control mice ($n = 5$). **b**, Immunohistochemistry analysis reveals significantly less γ H2AX induction in the colon of mice colonized by *clbM* than in those colonized by NC101. Scale bars, 50 μ m. Right: each symbol represents data from one mouse. The mouse colonization experiment was repeated once. **c**, NC101 and *clbM* (red fluorescence-labelled) colonize to similar levels in the 6 d.p.f. zebrafish intestine. Blue is DAPI counterstain. Each symbol represents data from a pool of three zebrafish guts. A total of 18 fish guts from three independent experiments were analysed for each group. NS, not significant; * $P < 0.05$ and ** $P < 0.01$, two-tailed t -test. Error bars indicate s.d.

Figure 3. Structural and functional characterization of ClbM

a, ClbM spans 12-transmembrane helices (TH) isolated to the inner periplasmic membrane. Semi-symmetrical 6-TH lobes create the drug binding pocket. **b**, The drug-binding pocket spans nearly 40 Å across, and contains enough space inside to bind precolibactin. A model of a proposed precolibactin metabolite (compound 32 in ref. 13) is shown to scale. **c**, The C-terminal cation-binding site consists of Y277, F292 and D299 in close proximity to bound Rb⁺. The density around Rb⁺ was generated from an isomorphous difference map contoured at 4σ. **d**, A conserved transport mechanism similar to NorM-VC/-NG is predicted, using F292, Y277 and D299. The cation is coordinated primarily by Y277 in the cation-bound substrate-bound state. Following substrate release, D299 coordinates closer to the cation, allowing tighter binding. The intracellular-open state will proceed through a reverse mechanism, leading back to the extracellular-open state.